

**Electronic Transport Properties of Silicon-Germanium
Single Photon Avalanche Detectors**

Helen Rafferty

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

**The University of Leeds
School of Electronic and Electrical Engineering**

October 2017

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

This work was undertaken within the EPSRC-funded SPADs project.

Publications

Conference Presentations

N.J. Pilgrim, **H.Rafferty**, Z. Ikonic, R.W. Kelsall *Ge/Si Single Photon Avalanche Diodes*. European Optical Society Annual Meeting, Aberdeen, UK, 25-28 September 2012.

H.Rafferty, A.D. Burnett, Z. Ikonic, R.W. Kelsall *Electronic Structure of Misfit Dislocations in Germanium*. International Workshop on Computational Electronics, Nara, Japan, 4-6 June 2013.

H.Rafferty, A.D. Burnett, Z. Ikonic, R.W. Kelsall *Electronic Structure of Misfit Dislocations in Germanium*. UK Semiconductors, Sheffield, UK, 3-4 July 2013.

H.Rafferty, A.D. Burnett, Z. Ikonic, R.W. Kelsall *DFT Calculation of the Electronic Structure of Defects in Germanium*. Theory, Modelling and Computational Methods for Semiconductors, Salford, UK, 22-24 January 2014.

H.Rafferty, A.D. Burnett, Z. Ikonic, R.W. Kelsall *DFT Calculation and Spectral Analysis of Electronic Structure of Dislocations in Germanium*. International Workshop on Computational Electronics, Paris, 3-6 June France, 2014.

H.Rafferty, Z. Ikonic, R.W. Kelsall *Calculation of Band-to-band Tunnelling Currents in Germanium*. International Workshop on Computational Electronics, Indiana, USA, 2-4 September 2015.

Acknowledgements

I would like to thank many people who have assisted me throughout the course of this work. Primarily, my heartfelt thanks go to my supervisors, Robert Kelsall and Zoran Ikonc, for their support, advice, guidance, knowledge and moral support over the course of my studies. I would also like to thank EPSRC for funding this research, Matt Probert, Phil Hasnip, Peter Brommer and Stewart Clark for support with CASTEP and complementary code, and Vineet Unni for taking time to assist me with facilities at Sheffield University.

I'm indebted to Andy Burnett, Alex Valavanis, Leon Lever, Brent Murphy and Will Furnass; this thesis would not have been possible without your input and expertise. Ian Bolton and Alex Barton provided invaluable help with graphics and figures, and I would not have made it through this work overall without the support and care of a huge number of friends; I'm very grateful to you all.

Finally, my heartfelt gratitude also goes to my parents, Patrick and Julie Rafferty, for their love and support throughout the past few years.

Abstract

Single photon avalanche detectors (SPADs) have uses in a number of applications, including time-of-flight ranging, quantum key distribution and low-light sensing. Germanium has an absorption edge at the key communications wavelengths of 1.3-1.55 μm , and can be grown epitaxially on silicon, however, SiGe SPADs exhibit a number of performance limitations, including low detection efficiencies, high dark counts and afterpulsing. Unintentional doping may affect electronic performance, and band-to-band tunnelling at high operational voltages SPADs may lead to noise currents. Additionally, defects in the Si/Ge interface lead to trap states within the bandgap and contribute to afterpulsing. This work investigates a range of critical performance parameters in SiGe SPADs. The effect of intentional and unintentional doping in SPADs on electric fields, potential profiles and carrier transport in the device is investigated, and optimal dopant profiles for a SiGe SPAD discussed. The dependence of band-to-band tunnelling currents in Ge on bias voltage, Ge thickness and temperature is investigated, and these currents are compared to other sources of noise currents in SPADs. DFT calculations of misfit dislocation structures in Ge are undertaken, to establish electronic bandstructures and optimised geometries for these defects, and identify trap states in the bandgap, which may contribute to afterpulsing and dark counts in SPADs. A number of directions for continuing work are identified, to progress understanding of noise currents and afterpulsing in SPADs.

Contents

| | |
|--|-------------|
| Acknowledgements | v |
| Contents | viii |
| List of Tables | xi |
| List of Figures | xiii |
| 1 Introduction | 1 |
| 1.1 Motivation and Applications | 2 |
| 1.2 Single Photon Detection Methods | 4 |
| 1.2.1 Photomultiplier Tubes | 4 |
| 1.2.2 Superconducting Nanowires | 4 |
| 1.2.3 APDs and SPADs | 5 |
| 1.3 SPAD Performance | 6 |
| 1.3.1 Silicon-only SPADs | 6 |
| 1.3.2 Germanium-only SPADs | 6 |
| 1.3.3 SPADs in III-V Materials | 7 |
| 1.3.4 SiGe SPADs | 7 |
| 1.4 Structure of the Thesis | 8 |
| 2 Theory of Silicon-Germanium Single Photon Avalanche Detectors | 11 |
| 2.1 Avalanche Photodiodes and SPADs | 11 |

| | | |
|----------|--|-----------|
| 2.1.1 | <i>p-n</i> Junctions and Photodiodes | 11 |
| 2.1.2 | Impact Ionisation | 14 |
| 2.1.3 | Avalanche Breakdown | 16 |
| 2.1.4 | Band-to-Band Tunnelling | 20 |
| 2.2 | Silicon-Germanium Heteroepitaxy | 21 |
| 2.2.1 | Silicon and Germanium Crystal Structures | 21 |
| 2.2.2 | Growth Techniques | 24 |
| 2.2.3 | Dislocations in Silicon-Germanium Heteroepitaxial Interfaces | 27 |
| 2.2.4 | Interface Traps | 30 |
| 2.3 | Bandstructure Models for Simulation | 32 |
| 2.3.1 | Analytical Bandstructures | 33 |
| 2.3.2 | Periodic Structures and the Brillouin Zone | 34 |
| 2.3.3 | Full Bandstructure Models | 35 |
| 2.3.4 | Pseudopotentials | 36 |
| 2.3.5 | Density Functional Theory | 40 |
| 2.4 | Electron Transport in Semiconductors | 43 |
| 2.4.1 | The Drift Diffusion Approximation | 45 |
| 2.5 | Conclusion | 47 |
| 3 | Electron Transport in SPADs | 49 |
| 3.1 | Introduction | 49 |
| 3.2 | Investigation of Doping Profiles | 50 |
| 3.2.1 | Breakdown and Punchthrough Voltage | 52 |
| 3.2.2 | Electric Field Profiles | 54 |
| 3.2.3 | Potential Profiles and Carrier Concentrations | 57 |
| 3.2.4 | Experimental Measurements | 59 |
| 3.3 | Conclusion | 63 |
| 4 | Band-to-Band Tunnelling in Germanium | 87 |
| 4.1 | Band-to-Band Tunnelling Model | 88 |

| | | |
|----------|---|------------|
| 4.2 | Calibration and Optimisation of Simulation Grids | 92 |
| 4.3 | Variation of Total Tunnelling Current with Bias Voltage | 94 |
| 4.4 | Total Tunnelling Current with Changing Thickness | 97 |
| 4.5 | Temperature Dependence of Tunnelling Current | 99 |
| 4.6 | Conclusion | 104 |
| 5 | Electronic Structure of Defects in Bulk Germanium | 107 |
| 5.1 | The CASTEP Software | 109 |
| 5.2 | Simulation Models | 109 |
| 5.2.1 | Geometry Optimisation Parameters | 117 |
| 5.2.2 | Geometry Optimisation Results | 117 |
| 5.3 | Bandstructure Calculations | 122 |
| 5.3.1 | The Brillouin Zone and Folded Bandstructures | 122 |
| 5.3.2 | The Unfolding Algorithm | 124 |
| 5.4 | Conclusion | 132 |
| 6 | Conclusion | 135 |

List of Tables

| | | |
|-----|--|-----|
| 3.1 | Doping profiles for SPAD device simulations | 51 |
| 3.2 | Simulated breakdown and punchthrough voltages | 53 |
| 4.1 | Simulation parameters for band-to-band tunnelling calculations on bulk Ge | 91 |
| 4.2 | Fitting parameters applied to the Varshni formula | 100 |
| 4.3 | Coefficients calculated for each bandgap by changing formfactors | 101 |
| 5.1 | Comparison of experimental and simulated bandgaps for bulk Ge | 113 |
| 5.2 | Unfolding results for the 12×8 supercell | 128 |
| 5.3 | Unfolding results for the 16×10 supercell | 130 |

List of Figures

| | | |
|------|---|----|
| 2.1 | Charge profile, electric field and potential for a pn -junction | 12 |
| 2.2 | IV characteristic For a pn -junction diode | 13 |
| 2.3 | Avalanche breakdown | 17 |
| 2.4 | Dark current and photocurrent in an APD | 19 |
| 2.5 | Band-to-band tunnelling under the influence of an electric field | 21 |
| 2.6 | Unit cells for silicon/germanium | 22 |
| 2.7 | he Brillouin zone and symmetry directions for a diamond lattice | 24 |
| 2.8 | Screw dislocation | 28 |
| 2.9 | Edge dislocation | 29 |
| 2.10 | Burgers vector and Burgers circuits for edge and screw dislocations | 30 |
| 2.11 | Misfit and threading dislocation | 31 |
| 2.12 | TEM image showing threading dislocations | 32 |
| 2.13 | Typical bandstructure for silicon | 43 |
| 2.14 | Typical bandstructure for germanium | 44 |
| 3.1 | Schematic of the band offsets in a SiGe SPAD | 66 |
| 3.2 | PiPiN SPAD structure | 67 |
| 3.3 | IV curves for simulations including high charge layer doping | 68 |
| 3.4 | IV curves for simulations including medium charge layer doping | 69 |
| 3.5 | IV Curves for simulations including low charge layer doping | 70 |
| 3.6 | Electric field profile for SPAD device 1 | 71 |

| | | |
|------|--|-----|
| 3.7 | Electric field profile for SPAD device 2 | 71 |
| 3.8 | Electric field profile for SPAD device 5 | 72 |
| 3.9 | Electric field profile for SPAD device 6 | 73 |
| 3.10 | Electric field profile for SPAD device 9 | 74 |
| 3.11 | Electric field profile for SPAD device 10 | 74 |
| 3.12 | Potential profile for SPAD device 10 | 75 |
| 3.13 | Electron concentration profile for SPAD device 5 | 76 |
| 3.14 | Hole concentration profile for SPAD device 9 | 77 |
| 3.15 | Electron concentration profile for SPAD device 1 | 78 |
| 3.16 | Hole concentration profile for SPAD device 1 | 79 |
| 3.17 | Electron concentration profile for SPAD device 2 | 80 |
| 3.18 | Hole concentration profile for SPAD device 2 | 81 |
| 3.19 | Schematic of the 1st set of SPAD devices | 81 |
| 3.20 | Current–voltage (IV) characteristics for the first set of SPAD devices . . . | 82 |
| 3.21 | Schematic of the second set of SPAD devices | 83 |
| 3.22 | current–voltage (IV) characteristics for the second set of SPAD devices . . | 84 |
| 3.23 | Schematics of SPAD devices fabricated in 2013, with target doping profiles | 85 |
| 4.1 | Energy range selected for band-to-band tunnelling simulations | 92 |
| 4.2 | Energy grid calibration curve for band-to-band tunnelling simulations . . | 93 |
| 4.3 | k-point grid calibration curve for band-to-band tunnelling simulations . . | 94 |
| 4.4 | Digitizing step calibration curve for band-to-band tunnelling simulations . | 95 |
| 4.5 | Band-to-band tunnelling current with changing bias voltage | 96 |
| 4.6 | Band-to-band tunnelling current with changing thickness | 98 |
| 4.7 | Band-to-band tunnelling current with increasing temperature | 103 |
| 5.1 | Primitive and cubic cells for silicon and germanium | 110 |
| 5.2 | Optimized bandstructure for bulk Ge | 112 |
| 5.3 | Optimized bandstructure for bulk Ge (2) | 113 |
| 5.4 | Germanium supercells with misfit dislocations used for DFT calculations . | 115 |

| | | |
|------|---|-----|
| 5.5 | $12 \times 10 \times 1$ supercell of Ge containing a misfit dislocation | 116 |
| 5.6 | Geometry optimised $8 \times 6 \times 1$ supercell of Ge | 117 |
| 5.7 | Geometry optimised $12 \times 8 \times 1$ supercell of Ge | 119 |
| 5.8 | Geometry optimised $16 \times 10 \times 1$ supercell of Ge | 120 |
| 5.9 | Geometry optimised $20 \times 12 \times 1$ supercell of Ge | 121 |
| 5.10 | Folded bandstructure for the $12 \times 8 \times 1$ supercell of Ge | 123 |

Chapter 1

Introduction

This work discusses the development and optimisation of silicon-germanium (SiGe) single photon avalanche detectors (SPADs). A single-photon detection device is a crucial component in a number of applications including quantum key distribution, and in applications where resolution of very low-level light or photon-counting is required, such as in a number of imaging and sensing technologies. including time-of-flight ranging, time-resolved photoluminescence and optical-time-domain reflectometry [1, 2].

Challenges to the development of viable high-efficiency SiGe SPADs arise from a number of sources. There is a 4.2% lattice mismatch between silicon and germanium, and defects within the SiGe heterointerface disrupt theoretical descriptions of bandstructure and electron transport, leading to localised band gap states and trapping and detrapping of carriers, with effects on device performance such as amended generation-recombination rates, missed detection events and afterpulsing [1].

The need to operate SPADs at high bias voltage precipitates sources of current other than those related to a detection event. Band-to-band tunnelling currents, as first described by Zener [3] become more significant at higher bias voltages. Doping densities and distributions in the material system lead to further non-linear fields from which a tunnelling current may arise.

This work investigates the contributions made to the performance of SiGe SPADs from field distributions caused by varying doping densities and types in the material, from the effect of band-to-band tunnelling at high bias voltages, and from electronic states due to crystalline defects. A number of simulation methods are employed to evaluate impacts on the performance of SPAD devices in this material system, from full-device carrier models, to quantum simulations and atomistic calculations.

1.1 Motivation and Applications

Single photon detection systems play an important role in a number of applications. Components for quantum communication systems, such as quantum repeaters and qubit amplifiers, and the implementation of optical quantum computing all require a single-photon source and a subsequent detector, which must be integrated into circuitry for successful implementation; furthermore, monolithic integration of a single-photon detector is required for large-scale reproducible systems [1, 4]. Quantum key distribution is the most developed quantum technology, and involves encoding of information on a single particle — such as a photon — and then subsequent detection and decoding [5]. Quantum computing, where calculations are performed using superposed states of quantum particles, promises significantly faster computation, but also demands highly sensitive and accurate integrated detectors for successful implementation at usable scales [6–9].

Single-photon detectors also have utility in applications where very low level light detection. Fluorescent dyes are frequently used in biological imaging, including tracking of neuronal activity, drug uptake and bloodflow, and are increasingly being used in novel microscopy techniques. Detection of the rate of decay of a fluorophore or investigation of fluorescence at the molecular level requires the use reliable single photon detectors [10, 11].

Time-correlated single-photon counting is a technique that has been applied to a number of fields, including fluorescence lifetime imaging microscopy [10, 12–14] and time-of-flight ranging for sensitive low-light imaging at distance. Time-of-flight ranging depends on direction of a photon source, typically a pulsed laser, towards a non-reflective target, and time-sensitive detection of the scattered photon. From this information, detailed imaging of the target can be achieved. The technology has applications across fields such as manufacturing, defence and environmental monitoring [1, 15–18],

Traditionally, photonic technologies have relied on the use of bulk devices, however, more recently, integrated circuits on a number of material systems have improved the stability and scalability of these applications [19]. Integration of electronic circuits provides improved device performance through reduced use of interconnects and greater reliability, as components are fabricated in the same controlled conditions, and losses due to electrical circuitry transmission may be reduced, leading to few delays and larger bandwidth density [20] Silicon complementary metal–oxide–semiconductor (CMOS) is the dominant technology for microelectronics production and there are advantages associated with leveraging this mature bulk production route for photonics components, including low-cost and high production volumes [21–23].

Development of photodetection systems that operate at wavelengths of 1.3 and 1.55 μm provides compatibility with existing telecommunications infrastructure. Silicon becomes transparent at wavelengths above $\approx 1.1 \mu\text{m}$, however, and photoabsorption at higher wavelengths requires either modification of silicon or use of some other material. Germanium has an absorption edge at a photon energy of around 0.8eV, corresponding to a wavelength of 1.55 μm and can be grown on a silicon substrate within established CMOS fabrication techniques; germanium on silicon therefore presents a suitable material system for investigation and development of single photon detectors at these relevant infrared wavelengths. A number of other material systems, including III-V SPADs us-

ing InGaAs/InP — which has an absorption edge at $1.55 \mu\text{m}$ — have been applied to development of single photon detectors at infrared wavelengths [1].

1.2 Single Photon Detection Methods

1.2.1 Photomultiplier Tubes

A number of methods for detection single photons have been developed, amongst these, the most mature technology is the photo multiplier tube (PMT). PMTs work on the principle of emission of an electron from a photocathode following a collision with an incident photon, and subsequent acceleration of emitted electrons in an electric field, and then amplification of the current when these electrons collide with a second dynode, where multiple electrons are knocked out. PMTs provide low noise and low timing jitter, with very high current amplification and large optical collection areas, however they have limited detection efficiency and are physically bulky and mechanically fragile [1, 24]. They are, nonetheless, widely used in a number of applications including medical and particle physics [25–27]. The timing jitter of PMTs may be improved by combining them with an avalanche photodiode (APD), to provide current amplification, in a device known as a hybrid photodetector [1].

1.2.2 Superconducting Nanowires

Superconducting nanowire single photon detectors (SNSPDs) have recently emerged as a promising photon detection technology for the target infrared wavelengths. Their mechanism of action involves a nanowire maintained below the critical temperature for superconduction; a single incoming photon can create a nanoscopic region of resistivity in the wire, and joule heating in this region will expand the resistive region, leading to a measurable voltage pulse. SNSPDs show excellent detection efficiencies — consistently over 70% and as high as 90% — low noise and precise timing resolution [28–31]. However, the need to operate SNSPDs at cryogenic temperatures and their small absorption area limits their applicability in integrated circuits [24].

1.2.3 APDs and SPADs

The principle mode of operation of an APD is that of a p - n junction under a reverse bias voltage; light incident on the depleted region of the junction will generate electron-hole pairs, and translation of these pairs into a measurable current will create a photodetector (photodiode) [32]. An APD operates through accelerating photogenerated electrons or holes sufficiently that they create additional carrier pairs, and create an avalanche current [33]. Single Photon Avalanche Detectors (SPADs) consist of an APD operated above breakdown voltage. If a sufficiently high reverse bias voltage is applied across the diode, avalanche multiplication will occur from a single incoming photon, and a self-sustaining avalanche current generated. The device is then 'reset' to enable a subsequent detection event, known as 'Geiger mode' operation. Alternatively, the device may be operated in 'gated mode' whereby the bias voltage is pulsed such that the diode is only above breakdown voltage for the times where a photon is expected, to limit runaway thermal generation of carriers and decrease dark counts and afterpulsing effects. This may be exploited in applications such as QKD, where the arrival time of a photon is known with reasonable accuracy [1, 24].

The detection efficiency of a SPAD increases with excess bias voltage, however this comes at the expense of excess noise, since high bias voltages are more likely to lead to generation of carriers through tunnelling current, and hence high dark currents [24]. One implementation of an SPAD intended to mitigate this is the separate absorption and charge multiplication (SACM-SPAD) device, with photon absorption and generation of an electron/hole pair taking place in one region of material which has an absorption edge at the required energy, and a second multiplication layer providing a region for avalanche multiplication. In this implementation a doped charge layer is usually present at the interface between the two regions, which regulates the electric field between the absorption layer (in which the electric field should ideally be low to limit dark current) and the multiplication layer which requires high electric field for impact ionisation to occur. Implementations of SACM-SPAD devices include those where the

absorption/multiplication regions are of the same material, but more commonly different materials are used for each region to optimise device performance for each of these processes. Material structures for SACM-APD devices in the infrared include use of the InGaAs infrared absorption edge alongside InP as a multiplication layer, and germanium absorption with charge multiplication in silicon. Other alloy systems include AlGaAs/GaAs, AlGaSb/GaSb and InGaAsP/InP [33].

1.3 SPAD Performance

A number of parameters are employed to assess the performance of SPADs. The detection efficiency is of primary importance and measures the likelihood that an incoming photon will generate a current pulse. Timing jitter measures the uncertainty between the arrival time of an incoming photon and the corresponding current pulse, and may have implications in applications such as time-of-flight ranging and QKD. The dark count rate (DCR) arises from thermally generated or field- or impurity- induced carriers causing an avalanche current outside of a detection event. Finally, afterpulsing occurs when mid-band trap states are occupied with carriers, which later detrap causing secondary current pulses after the detection event [1, 24].

1.3.1 Silicon-only SPADs

Silicon-only SPADs have been used commercially for a number of years in the 0.4-1 μm wavelength region. High detection efficiencies of up to 65-70% have been reported, and low timing jitter of around 20ps. Dark count rates can be obtained under 100 counts s^{-1} at room temperature, and as low as 5 100 counts s^{-1} after cooling, however Si-only SPADs can not detect photons in the infrared region between 1.3-1.55 μm [1, 24, 34].

1.3.2 Germanium-only SPADs

Germanium-only SPADs have been produced commercially for a number of years. However, these demonstrate relatively low detection efficiencies of under 10% when operated

in gated mode under cryogenic conditions, and suffer from high dark counts and afterpulsing. Timing jitters of under 100ps have been reported [1, 35]

1.3.3 SPADs in III-V Materials

A number of III-V materials have been used for development of SACM-SPADs, particularly utilising InGaAs as an absorber in the 1.55 μm region and InP as a multiplication region. There is a band discontinuity in the interface between these materials, which may be mitigated by growth of an InGaAsP grading layer to aid transport of carriers over the interface [1]. InGaAs/InP SPADs have been demonstrated with detection efficiencies of 10-20%, [2, 36–38], however one limitation of InGaAsP SPADs is high DCR — 8000 counts s^{-1} have been reported for 20% detection efficiency under cryogenic conditions [38], and typical rates can be of the order of hundreds to thousands of counts per second in Geiger mode. Timing jitter is usually high, in the region of hundreds of picoseconds [2], although jitter as low as 50ps has been reported with high performance circuits [39]. Additionally further work is required to limit defects in growth structures and reduce afterpulsing [1].

1.3.4 SiGe SPADs

The SiGe SPAD utilises a separate absorption and multiplication approach. A low electric field in the Ge absorber suppresses dark counts, while a high electric field in the Si region promotes carrier multiplication by impact ionisation. The regions are separated by a highly doped 'charge layer', regulating the optimal electric field in each operating region. Germanium on silicon, as a material system, also offers the possibility of on-chip integration with photonics components. However, SiGe SPADs currently demonstrate high DCR, and low detection efficiencies. The devices currently developed within the SPADs project have demonstrated detection efficiencies of around 10% at 1.31 μm and 100K, dropping to 0.15% at 1.55 μm . Timing jitter was measured at 420ps and dark counts of $10_6 - 10^7 s^{-1}$ were observed, increasing with increasing excess bias voltage [40]. Challenges to the development of SiGe SPADs include reducing sources of dark counts,

improvement of the detection efficiency and reducing defects, and hence trap states which contribute to afterpulsing.

1.4 Structure of the Thesis

The aim of this thesis is to investigate critical performance characteristics of SiGe SPADs. A number of investigations are undertaken and models developed which may be applied to further characterisation of dark counts, afterpulsing and electronic transport in germanium. The structure of the thesis is as follows:

In Chapter 2, the relevant theory and physics underpinning the operation of SPADs is introduced, as well as a discussion of the sources of dark currents and afterpulsing events. Theoretical models for device simulation, tunnelling and modelling of the electronic structure of defects are described.

In Chapter 3, the effect of variable types and densities of charge and absorption layer doping is investigated, for a typical SPAD device. IV characteristics, electric field and potential profiles, and carrier concentrations are generated using semiconductor device simulation software, to establish optimal intentional and background doping profiles for SPAD devices, to promote photon detection limit dark counts.

Chapter 4 investigates the effect of band-to-band tunnelling in bulk Ge, in relation to a number of parameters. Since SPADs are operated at high bias voltages, tunnelling between the valence band and conduction band within the Ge absorber may result in generation of carriers outside of photon detection events and lead to additional noise currents. The effect of increasing bias voltage, increasing Ge thickness and increasing temperature on the band-to-band tunnelling current are simulated and discussed.

In Chapter 5, the electronic structure of defects in bulk Ge is investigated. Density

functional theory is used to perform *ab initio* calculations of Ge bandstructures for cells including misfit dislocations. Energy minimisation is applied to optimise the cell geometry, and a preliminary model to identify bandgap states due to misfit dislocations in bulk Ge — potential sources of afterpulsing in SiGe SPADs — is developed.

Finally, conclusions and suggestions for further work are discussed in Chapter 6.

Chapter 2

Theory of Silicon-Germanium Single Photon Avalanche Detectors

2.1 Avalanche Photodiodes and SPADs

2.1.1 p - n Junctions and Photodiodes

A p - n junction is formed when a material or region with an excess of electrons (n -type material) and a material with an excess of holes (p -type material) form an abrupt boundary. At this p - n junction, the gradient of chemical potential causes a transfer of electrons to the p -type region where they will recombine with holes — resulting in a ‘depletion layer’ containing very few free carriers. The n -region of the junction will be positively charged and the p -region will be negatively charged, and the junction hence has a ‘built-in potential’ caused by this shift in carrier types.

The carrier concentrations and subsequent charge density profiles lead to the field and potential profiles for a p - n junction shown in Figure 2.1 If a positive voltage (opposing the built-in voltage) is applied across the p - n junction (forward bias), the depletion

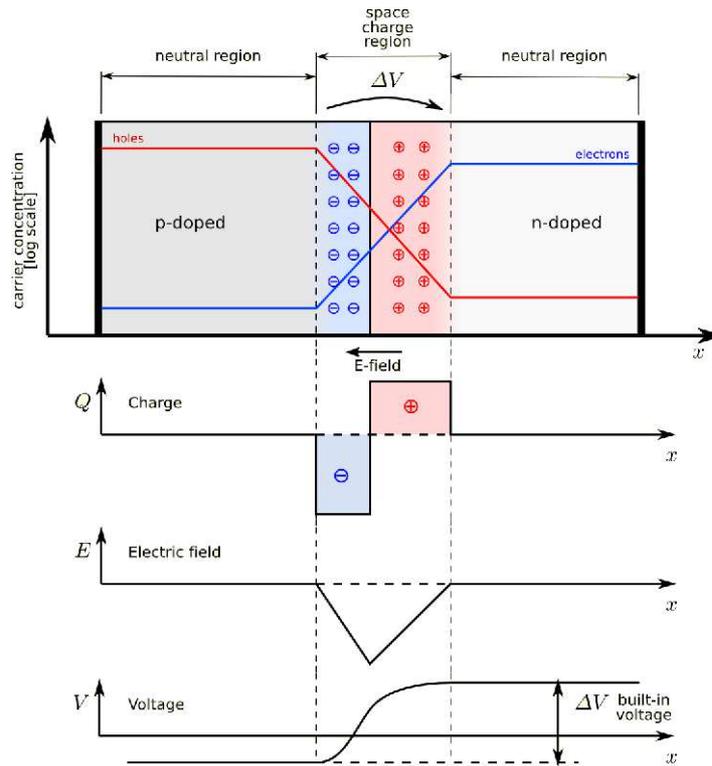


Figure 2.1: Charge profile, electric field and potential for a pn -junction [41]

region will decrease in size and under sufficient bias voltage a current will flow. In reverse bias, the depletion region will be widened and the p - n junction will be resistive. Hence, a p - n junction may be used as a basic diode—a directional current device with a characteristic I-V profile wherein a current may flow under an applied voltage in one direction, but no current flows under the reverse voltage (up to some breakdown point). The characteristic I-V profile for a p - n junction diode is shown in Figure 2.2. If light of frequency $\hbar\omega > E_g$ is incident upon a semiconductor material, electron-hole pairs will be generated by promotion of electrons across the band gap of the material. To create a photodetector (photodiode), these photogenerated carriers must translate into a measurable current. If the semiconductor material forms part of a pn junction, the built-in potential will separate the photogenerated carriers, and an electromotive force will be generated, in a phenomenon known as the photovoltaic effect [32]. Hence, a pn junction may be used as a simple photodetector.

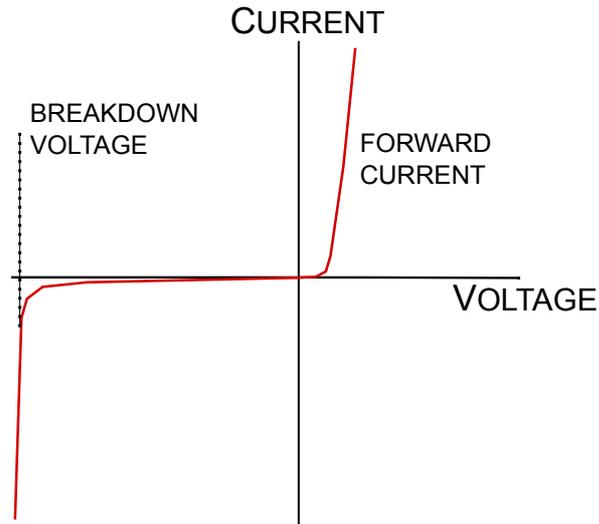


Figure 2.2: IV characteristic For a pn -junction diode

Illumination of a pn junction with constant light intensity would lead to a steady state, homogenous carrier generation rate $G(x, t) = G_0$, from which it is possible to derive a current relationship for the diode. The photocurrent I is linearly proportional to G_0 which is proportional to the optical power P_0 . Carriers will be generated both within the depletion layer and within a diffusion distance L from the edge of the depletion region. The diffusion distance (for electrons or holes) is given by:

$$L = \sqrt{D\tau} \quad (2.1)$$

where D is the diffusion co-efficient for electrons or holes, and τ is the lifetime of the electron/hole.

Hence, for a generation rate $G_e x t$ per unit volume of depletion layer, the photocurrent is given by:

$$I_{photo} = qA_j G_e x t (w_d + L_p + L_n) \quad (2.2)$$

where A_j is the cross-sectional area of the diode, q is the unit charge of the carriers, w_d represents the width of the depletion region and L_n and L_p the diffusion distance on the

n and p sides of the junction respectively [42].

Quantum efficiency η is a useful parameter for understanding photocurrent, as it quantifies the likelihood that an incoming photon will be detected by the diode. The internal quantum efficiency η_i is the probability of an incoming photon generating an electron-hole pair, and this is related by the device characteristics to the external quantum efficiency η_e , which is the probability of a incoming photon resulting in a measurable electron-hole pair [43]. The responsivity R of a photodiode is defined as the ratio of the current to the optical power [33]:

$$R = \frac{\eta_e q}{\hbar\nu} = \frac{\eta_e \lambda (\mu m)}{\hbar\nu} \quad (2.3)$$

In a pn junction diode the usable region for optical absorption leading to photocurrent is too narrow for practical use. To create a usable photodiode, the responsivity of the diode may be increased by adding an additional layer of intrinsic semiconductor between the n - and p -doped regions to make a p - i - n diode. This maintains the basic field and potential profiles present in pn -junction diode, but creates a considerably wider region of photoabsorption in the intrinsic layer, resulting in an area of high carrier injection.

A key figure of merit for photodiodes is the noise equivalent power (NEP), which is useful for characterising the signal-to-noise ratio of a detector, and gives the optical power required to achieve an rms signal-to-noise ratio of 1 at a bandwidth of 1 Hz:

$$NEP = \frac{hc}{\lambda \eta_e} \sqrt{\frac{2I_{eq}}{q}} \quad (2.4)$$

where λ is the wavelength of the incident photon, and I_{eq} is the dark current [43]

2.1.2 Impact Ionisation

Impact ionisation — a generation process — occurs where a sufficiently energetic carrier promotes an electron across the band gap, creating an electron-hole pair and thus generating further carriers. If these carriers are themselves sufficiently energetic they

may promote further electrons across the bandgap and initiate an avalanche of carriers. Impact ionisation processes proceed while conserving both energy and momentum. The mechanism and likelihood of impact ionisation occurring is characterised by a number of parameters including the threshold energy E_{th} for impact ionisation to occur and the related impact ionisation rate W_{ii} .

The threshold energy must exceed the band gap energy E_g for an indirect band gap material. Furthermore, values for E_{th} and W_{ii} depend on the availability of initial and final states for carriers [44] and numerical calculation methods for impact ionisation parameters are therefore dependent on the bandstructure.

One simple formulation for the threshold energy uses a weighted average of band gaps for symmetry directions:

$$E_{th} = \frac{1}{8}(E_g^\Gamma + 3E_g^X + 4E_g^L) \quad (2.5)$$

where E_g^Γ , E_g^X , and E_g^L are the band gaps between the Γ , X , and L -valleys and the valence band maximum at Γ [45].

The Keldysh model also provides a straightforward formulation for threshold energy; this assumes a simple parabolic bandstructure model and is suitable for low electron energies and direct band gap materials. The approximation yields a quadratic dependence of impact ionisation rate on the energy E of the initiating particle:

$$W_{ii} = pW_{E_{th}} \left(\frac{E - E_{th}}{E_{th}} \right)^2 \quad (2.6)$$

where $W_{E_{th}}$ is the total phonon scattering rate at the threshold energy E_{th} and p is a prefactor obtained by fitting to experimental data. However, this model does not account for momentum dependence in energy transitions and deviation of the band structure from a simple parabolic model. For higher energies, where bandstructures are no longer well-approximated by parabolic models, impact ionisation parameters must be calculated using methods which include full band structure calculations and conserve

both energy and momentum in transitions. Fermi's Golden Rule is often employed for calculation of impact ionisation rates by integration over all possible initial and final states where energy and momentum are conserved, and involves integration over all possible initial and final states for the process [46].

Impact ionisation threshold energies have been studied numerically in silicon ([47], [48], [49]) [50]) but less so in germanium ([45]). The electron impact ionisation rate in silicon was studied in Ref [47]. An empirical pseudopotential method was employed to calculate the full band structure for silicon, and Fermi's Golden Rule was used to calculate the impact ionisation rate. A 'soft' threshold for impact ionisation was identified, whereby the impact ionisation rate increased gradually with increasing electron energy and no sharp threshold was identified. The impact ionisation rate was found to be strongly anisotropic particularly at low energies when conservation rules for both energy and momentum were accounted for. A later calculation of the impact ionisation rate and primary hole energy threshold for silicon found a threshold energy of around 1.5 eV for hole-initiated impact ionisation—although it was noted again that this was strongly anisotropic. Impact ionisation rate was calculated using 1st order perturbation theory for direct impact ionisation only, and the impact ionisation rate was found to be an exponent of 3.4 of the hole energy, in contrast to the simple Keldysh model which uses an exponent of 2 [48]. A soft-threshold energy for silicon of 3 eV has also been calculated [50].

2.1.3 Avalanche Breakdown

The process of avalanche breakdown occurs when a particle initiating impact ionisation is sufficiently energetic, that the particle it ionizes itself gains enough energy to ionize another particle, and an avalanche of current results. The process of avalanche multiplication is shown schematically in Figure 2.3. If a semiconductor material is subjected to a high reverse bias, a single electron-hole pair may be sufficiently energetic that the process of impact ionisation leads to a self-sustaining avalanche current, with

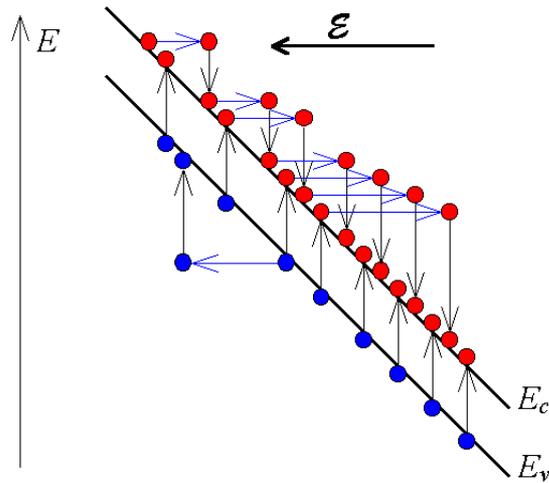


Figure 2.3: Avalanche breakdown [51]

both electrons and holes contributing to the avalanche gain. This results in a feedback mechanism which couples the gain from each carrier type, providing significant non-linear amplification and a sudden and significant increase in current. This effect may be exploited in an avalanche photodiode—APD— where a photodiode is operated at voltages close to the breakdown voltage to provide amplification of the photocurrent [32].

Within an APD, if amplification occurs with only one carrier type a linear response between the incident light intensity and photocurrent for a uniform electric field and dopant profile results. The likelihood of impact ionisation can be described by α_n and α_p , the ionisation co-efficients for electrons and holes respectively, defined as the number of electron-hole pairs generated by a single carrier per unit distance, and strongly related to the electric field. α is generally different for holes and electrons so avalanche multiplication with one carrier type only can be observed within a selected field regime. The avalanche gain is defined by a multiplication factor M determined by the carrier ionisation co-efficient, and, formally, breakdown occurs where the multiplication factor M approaches infinity; this can be seen in the I-V characteristic for a p - n junction shown in Figure 2.2 for higher reverse bias voltage. The expression for responsivity of

an APD is modified, such that the ratio of output current to optical power includes the multiplication factor [33]:

$$R = M \frac{\eta_e q}{\hbar\nu} = \frac{\eta_e \lambda (\mu m)}{\hbar\nu} \quad (2.7)$$

Under the right bias conditions, the generation of a single electron–hole pair through photoabsorption can result in an individual large and measurable current pulse. Here, an APD may be used to provide a current response to a single incoming photon, and is hence termed a single photon avalanche detector, or SPAD. SPADs are typically operated above breakdown voltage and may be considered to have a digital response, in contrast to a device which simply provides amplification of photocurrent. The depletion layer initially contains no free carriers; an incoming photon will create an electron-hole pair which will be rapidly accelerated in the high-field region, and provide a self-sustaining avalanche current. This current must then be ‘quenched’ by an external quenching circuit to prepare the diode for the next detection event. This operating regime is known as ‘Geiger mode’ [24].

A SPAD may be characterised by a detection efficiency η , defined as the probability that an incoming single photon results in a measurable current spike. Dark current for a SPAD is described by the dark count rate (DCR) - the number of counts registered with no incident light, N_D . The detection efficiency in Geiger mode increases with increasing excess bias (above breakdown) voltage, although this is concomitant with an increase in dark count due to increased avalanche triggering probability and field enhanced dark count generation. Hence, the choice of optimum bias depends on the application for which the detector is designed. The expression for noise-equivalent power for a SPAD may be re-rendered in terms of N_D [1]:

$$NEP = \frac{hc}{\lambda \eta_i} \sqrt{\frac{2N_D}{q}} \quad (2.8)$$

Timing jitter is commonly employed in measures of repetition rate and resolution in SPADs, and characterises the delay time between the arrival of a photon at the detector and the output of a measurable electrical pulse. Ordinarily a number of measurements

of this delay will be taken and the timing jitter will be given as the full width at half maximum (FWHM) of the resultant histogram [1].

A typical current–bias voltage relationship for an APD is shown in Figure 2.4, indicating the breakdown region in which the device may be operated as a SPAD and the associated increase in DCR. The operation mode of a SPAD, biased above breakdown,

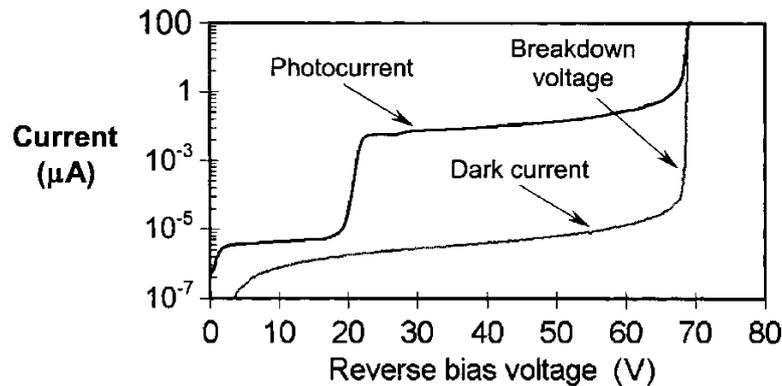


Figure 2.4: Dark current and photocurrent in an APD [1]

can generate additional sources of noise current. Additional available energy states within the bandgap, caused by, for example, defects in the crystal structure, may act as carrier traps; carriers trapped during detection pulses may become subsequently de-trapped and trigger a further avalanche, resulting in ‘afterpulsing’, with a probability dependent on the location of the additional energy states [24].

SPADs must be operated at a voltage above which avalanche breakdown can occur and trigger a current pulse from a single detection event; however, this voltage may give rise to a number of ‘false’ detection events and dark currents, as discussed above. One mechanism for circumventing these performance issues is to fabricate SPADs with different absorption and multiplication regions, constructed of different materials, separated by a highly-doped ‘charge layer’ which may enable different fields to be applied across each material. Hence, an absorption material may be selected with the bandgap required for the target wavelength, and then maintained under a field which is sufficient to allow

generated photoelectrons to pass into the multiplication region, but still sufficiently low to suppress noise currents and dark events. In the multiplication region, the second material may be maintained under high field conditions to initiate the avalanche current. The silicon-germanium SPAD is one such separate multiplication and absorption region device.

2.1.4 Band-to-Band Tunnelling

At the high operating fields associated with SPADs, a significant noise current may be generated through Zener (tunnelling) breakdown, where quantum mechanical tunnelling of carriers from the valence band to the conduction band leads to a rapid increase in current [32]. The mechanism of tunnelling as a source of electrical breakdown was described by Zener in 1934 [3]. The rate at which an electron may escape from one energy band into a higher energy band under the influence of an electrical field was calculated for the one-dimensional case. A simple representation of electron tunnelling from one band to another under the influence of an electric field is shown in Figure 2.5.

The mechanism discussed by Zener assumed electron wavefunctions with the periodicity of the crystal lattice, which have real solutions when electrons reside inside a permitted energy band. Upon encountering a potential barrier – in this case the forbidden band gap – the electron wave is largely reflected, however a small proportion of the wave will pass through the potential barrier. By considering an applied electric field, Zener was able to calculate the probability of an electron wave passing through a barrier of given width to the upper band. Solutions to the Schrodinger equation within the band gap are found to be complex, implying an exponentially decreasing wave function, with an associated probability of transition to the upper band dependent on the decay rate of the wave function. Currents due to this effect were found to become considerable above a characteristic threshold ‘breakdown field’. Breakdown fields for a number of semiconductor materials are established in the literature. Ioffe [52] cite the breakdown field for silicon to be $3 \times 10^5 \text{V/cm}$ and that of germanium to be 10^5V/cm . This, however, only

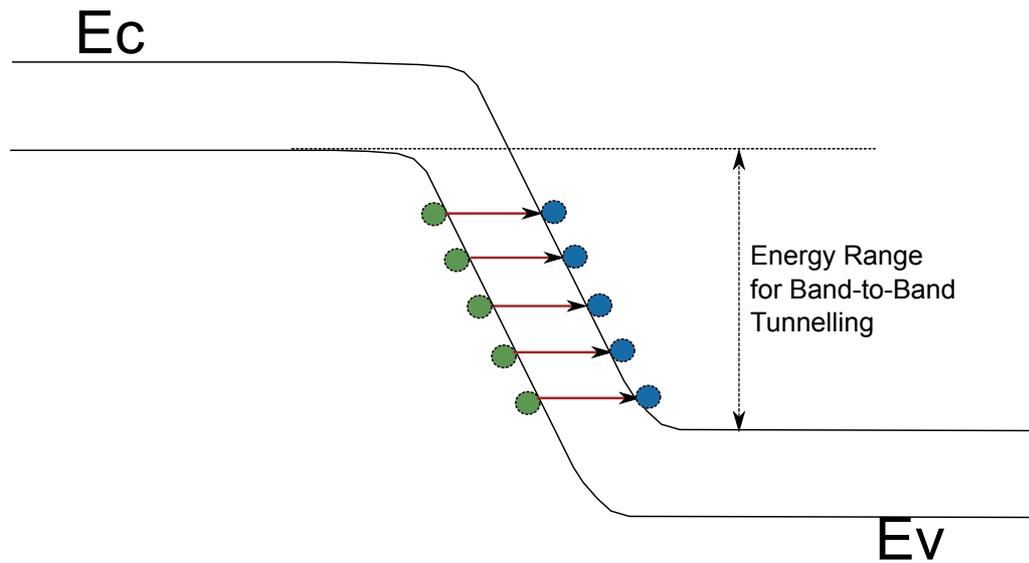


Figure 2.5: Band-to-band tunnelling under the influence of an electric field

represents the field at which a sudden and dramatic increase in current is observed; as quantum mechanical tunnelling is a probabilistic effect, noise currents due to tunnelling may be observed at lower fields.

Direct tunnelling can occur between electronic states which are matched in momentum; alternatively, ‘indirect’ tunnelling occurs where the initial and final states of the electron do not possess the same momentum; in this case, the difference in momentum must be provided by phonons or impurities [53]. Only the direct case will be approached in this work.

2.2 Silicon-Germanium Heteroepitaxy

2.2.1 Silicon and Germanium Crystal Structures

Both silicon and germanium have a diamond crystalline structure, consisting of two interlocking face centred cubic (FCC) cells, and characterised by a primitive rhombohedral cell, or a conventional cubic cell, as shown in 2.6.

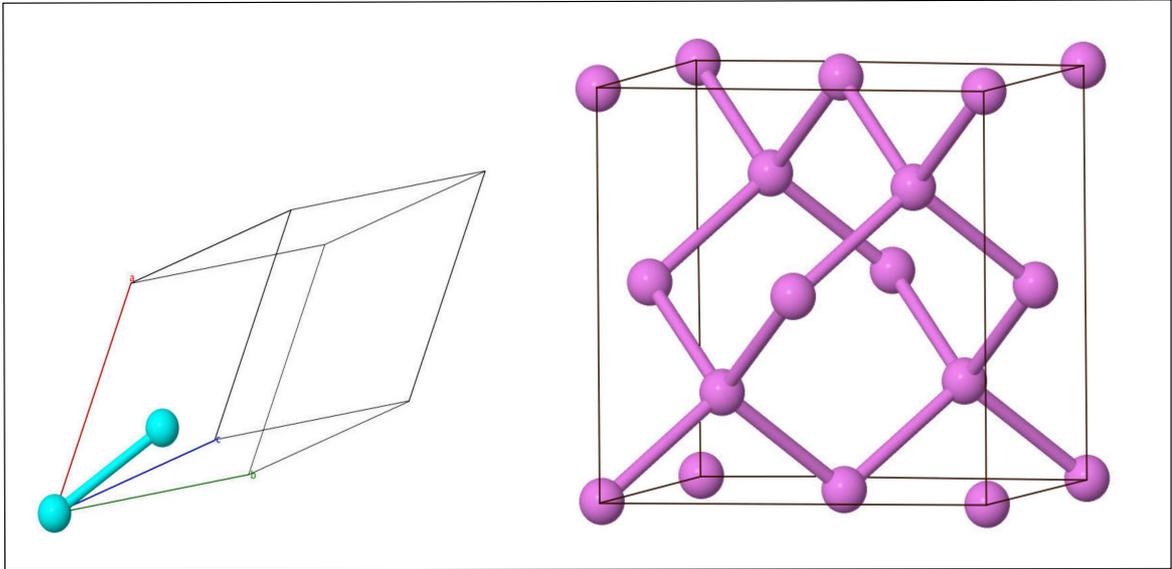


Figure 2.6: The rhombohedral primitive unit cell of silicon and germanium, and the conventional cubic cell

A Bravais lattice describes repeated points over all space that may be used to construct an entire crystal; these points may represent atoms, groups of atoms, ions or molecules, and in the case of elemental silicon or germanium, the positions of each primitive unit cell are represented. The general form of a Bravais lattice is:

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (2.9)$$

where \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 are any three vectors not in the same plane, the n span over all integers and each R is a lattice vector. Silicon and germanium are constructed of lattice

vectors

$$\mathbf{a}_1 = \frac{a_0}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{a}_2 = \frac{a_0}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{a}_3 = \frac{a_0}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \quad (2.10)$$

where a_0 represents the lattice constant: 5.434 Å and 5.658 Å for silicon and germanium, respectively, and the basis is a two-atom one, with atoms at (0, 0, 0) and $(\frac{1}{4}a_1, \frac{1}{4}a_2, \frac{1}{4}a_3)$.

The reciprocal lattice is a useful formulation for crystal lattices that plays a role across studies of crystalline materials, from diffraction to calculation of periodic wavefunctions for electrons. In general, the reciprocal lattice vectors are defined by

$$\mathbf{b}_1 = 2\pi \frac{a_2 \times a_3}{a_1 \cdot (a_2 \times a_3)} \quad \mathbf{b}_2 = 2\pi \frac{a_3 \times a_1}{a_2 \cdot (a_3 \times a_1)} \quad \mathbf{b}_3 = 2\pi \frac{a_1 \times a_2}{a_3 \cdot (a_1 \times a_2)} \quad (2.11)$$

and any general reciprocal lattice vector can be given by:

$$\mathbf{G} = n_1 \mathbf{b}_1 + n_2 \mathbf{b}_2 + n_3 \mathbf{b}_3 \quad (2.12)$$

in common with the description of a lattice vector provided above. More formally

$$e^{\mathbf{G} \cdot \mathbf{R}i} = 1 \quad (2.13)$$

The unit cell in reciprocal space is known as the Brillouin Zone, and takes the shape of a truncated octohedron for the primitive unit cells of silicon and germanium. A number of other conventions are applied to the description of crystal planes and spaces. The integers n_1 , n_2 and n_3 from the definition of a lattice vector R provided above can be used to define a ‘Miller Index’ for a plane; a plane with a Miller Index $(n_1 n_2 n_3)$ is in the direction normal to the lattice vector R given by these integers. Directions in the lattice are specified using square brackets, as $[n_1 n_2 n_3]$ [54], and those in the reciprocal lattice using curved brackets: $n_1 n_2 n_3$

The Brillouin zone has a number of characteristic symmetry directions and points. The Γ point lies at the centre of the Brillouin zone 0 0 0 and there are 6 X points, which

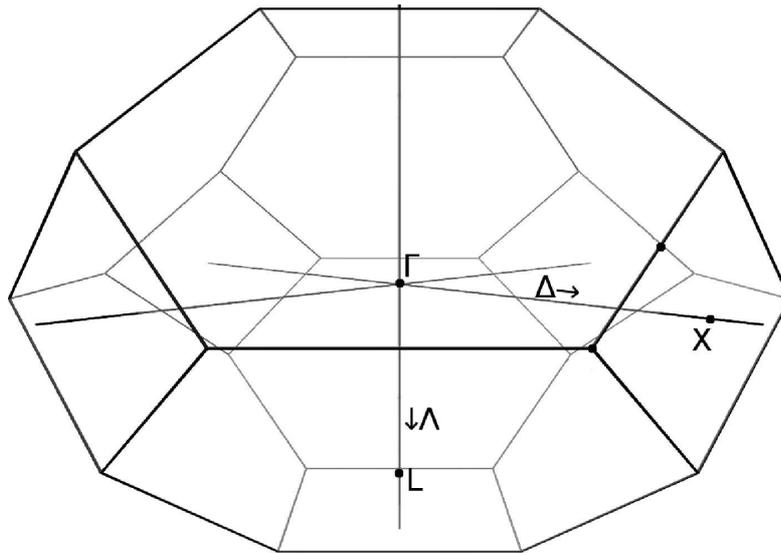


Figure 2.7: The Brillouin zone and symmetry directions for a diamond lattice [55]

lie in each of the $1\ 0\ 0$ directions (including the negative directions) in 3 dimensions. The 8 L points lie in each of the $1\ 1\ 1$ directions. Each point has a corresponding direction. The Δ direction lies along the line from the Γ point and the X point and the Λ direction lies along the line from the Γ point to the L point. The symmetry points and directions for the silicon and germanium Brillouin Zone are shown in Fig. 2.7.

2.2.2 Growth Techniques

Germanium on silicon SPADs are grown epitaxially using molecular beam epitaxy (MBE) or chemical vapour deposition (CVD). CVD is now the preferred method for fabrication of silicon devices, due to its superior uniformity and reproducibility [56]. CVD is a process for creating materials by reaction of chemical constituents in the vapour phase, to form solid films, and consists of six key phases: diffusion of gaseous reactants to the surface; adsorption of the reacting species onto surface sites, often after migration on the surface; surface chemical reaction between the reactants, usually catalysed by the surface; desorption of the reaction by-products; diffusion of the by-products away from the surface and incorporation of the condensed solid product into the microstructure of

the growing film [57].

For fabrication of silicon-on-germanium heterostructures, silane gas (SiH_4) as a silicon source and germane gas (GeH_4) as a germanium source are the gaseous reactants. Silicon and germanium are adsorbed onto their respective surfaces, possibly after diffusion across the surface, and in both cases, hydrogen is desorbed after reaction. For the simplest case, where silicon is grown upon a silicon reaction surface - the substrate - silicon atoms diffuse and react across the silicon wafer, and form a growth layer, without strain. The diffusion distance depends on the kinetic energy of the atoms, and as such, substrate temperatures of between 100°C and below 1250°C are required. However, at temperatures above 700°C diffusion of dopants may result, and temperatures below 300°C may result in poor growth, with islands forming rather than layers. Germanium has a similar structure to silicon, and although these elements have a lattice mismatch of 4.2%, germanium layers may be grown on a silicon substrate or silicon thin film, with increasing strain within the germanium layer as the thickness of the germanium layer increases [56].

One method for reducing the required epitaxy temperature and increasing the growth speed is the use of plasma-enhanced CVD (PE-CVD). Here, charged plasmas rather than gaseous particles are used for epitaxy, and a voltage discharge arc is used to dissociate SiH_4 and GeH_4 . PE-CVD was primarily developed for use in fabrication of SiGe alloy layers, where the lattice mismatch necessitated growth of layers with low surface roughness and low density of defects [58]. Diffusion and defects are suppressed at the lower temperatures used for PE-CVD [59], and the use of PE-CVD offers a number of other advantages over thermal CVD, including control of the microstructure of the film, directionality in film deposition and independence of growth speed from substrate temperature. PE-CVD removes H_2 efficiently, resulting in fast growth rates, and the composition of SiGe layers is controlled only by the gas composition, hence, sharp interfaces can be formed [60, 61], although layer thickness precision may be poorer.

Doping of growth layers may be unintentional, from impurities in the CVD atmosphere, or intentional, where dopant atoms are introduced to growth layers for engineering of the material properties, such as doping of the charge layer in a SPAD. Doping may be undertaken using in-situ doping, where dopant atoms are introduced to the gaseous epitaxy environment. This process may be used to give localised doping, where particular layers are doped. However, dopants may not remain uniformly distributed in the target regions and may become segregated, forming dopant islands on the surface of the layer, particularly for *n*-type dopants. Diffusion doping occurs when dopant atoms are introduced to the epitaxially grown wafer; however, dopant diffusion varies strongly with temperature, and hence high temperatures are required, In addition, dopants are diffused from the top down and hence the highest concentrations are usually found in surface layers and there is limited control over the depth and dose of doping [62].

Ion implantation may also be used for doping. In this technology, dopant atoms are accelerated through a voltage towards the wafer. Ion implantation may cause a dopant profile which decreases in concentration through the doped layer, and the high speed of the dopant atoms can cause significant damage to the wafer [62], making this method unsuitable for use in SPADs, where a low defect density is required to prevent dark counts and other noise phenomena. Hence, although diffusion doping has broadly fallen out of use for most applications, it is still the preferred method for SPAD doping.

The 4.2% lattice mismatch between silicon and germanium, with germanium having the slightly larger lattice constant, means that germanium grown epitaxially on silicon will be strained for germanium growth layers below some critical thickness, and there will be defects within some distance from the interface for layers thicker than the critical thickness, where the strain is no longer able to accommodate the additional energy introduced [63, 64]. The lattice constant of germanium is 5.658 Å and that of silicon is 5.431 Å; from energy balance calculations, it can be found that a thickness of Ge greater

than ≈ 6.5 nm will result in defect formation along the silicon–germanium interface, with defect structures formed according to the energy relaxation of the interface [65]. Operation of a Ge–on–Si SPAD requires an absorption layer which is typically at least an order of magnitude larger than this critical thickness for efficient collection of photons, so the formation and density of defects at the silicon–germanium interface is of critical importance to the operation of a Ge–Si SPAD. Although growth on silicon substrates may be conducted on a number of possible planes ((001), (110) and (111)) substrates oriented to the (001) plane are most commonly used and offer the most well developed techniques for heteroepitaxy [66].

Heteroepitaxy of germanium on silicon has proceeded in a number of directions to limit the formation of defects at growth thicknesses above the critical thickness. Techniques include growth of a silicon–germanium graded alloy layer, where a progressively higher concentration of germanium in each growth layer is included up to some useable structure on which pure Ge layers may be grown. Graded layers allow relaxation to occur over an extended distance, and permit dislocations to glide, although they also increase surface roughness, requiring further processing such as chemical polishing or annealing [66]. Alternative techniques involve growth of a high–defect Ge buffer layer—the seed layer—at low temperature, on which a further layer of low defect Ge can be added at high growth temperatures, to contain defects within a known region and relax strain progressively. Further processing of epitaxially grown layers, by thermal cycling or annealing can ‘glide’ defects towards particular regions of the material, most usefully towards the wafer edges where they may annihilate, providing a low–defect region within a heterojunction which may be sufficient for a given application [67], [66, 68].

2.2.3 Dislocations in Silicon-Germanium Heteroepitaxial Interfaces

Although point defects may be formed in crystalline structures during growth, defects formed at the interface due to lattice mismatch are most likely to be dislocations—lines or planes of defects propagating through the structure which form as a result of built-up

strain greater than the lattice is able to absorb. Dislocations are characterised by their topology as screw or edge dislocations, although in reality the full form of an individual dislocation may be a combination of these two structures. A screw dislocation occurs where shear stress in a crystal plane causes a displacement between adjacent sections of crystal lattice (Figure 2.8) and a helical path is traced around the resultant linear defect by the surrounding atoms. Pure screw dislocations thus propagate at 90° to the interface plane for a heteroepitaxial interface. Edge dislocations are significant in lattice-mismatched epitaxial growth, and occur when an unmatched plane terminates between two completed planes, as shown in Figure 2.9 [69, 70]. Dislocations may

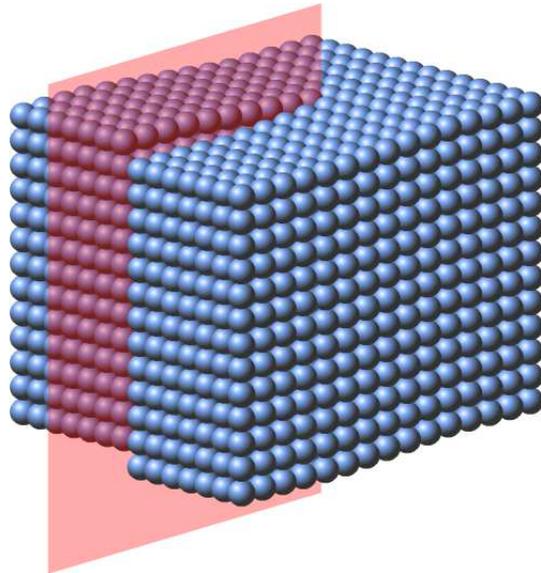


Figure 2.8: Screw dislocation [71]

be characterised mathematically using Burgers vectors and Burgers circuits. A Burgers circuit is an atom-to-atom path taken through a crystal which contains one or more dislocations. The Burgers vector is then the vector difference between this circuit for the defected and undefected crystal – the additional path needed to ‘close’ the circuit and provide the identical atom-to-atom path in the undefected crystal. Figure 2.10 shows the Burgers circuits and Burgers vector describing an edge and a screw dislocation, [69].

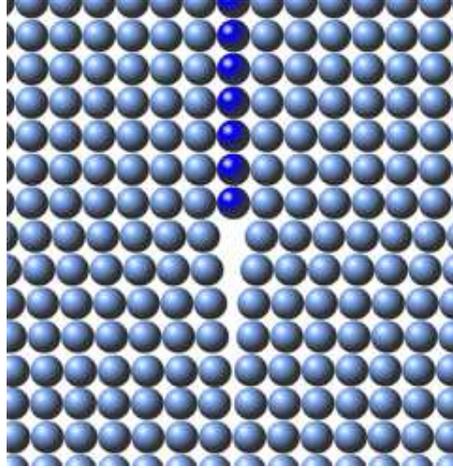


Figure 2.9: Edge dislocation [72]

For practical Si–Ge heteroepitaxy, dislocations are usually characterised in the form of misfit and threading dislocations. Misfit dislocations are lines of dangling bonds (in 3–dimensions) where the basal line of an edge dislocation plane terminates, parallel to the heterointerface, caused by the built-up strain in the mismatched interface. Threading dislocations then occur where the side edges of the edge dislocation plane terminate, and usually glide at 60° to the interface plane, in the $(1\ 1\ 1)$ plane of the lattice for diamond cubic structures such as silicon and germanium. Threading dislocations will terminate the total defect structure either at a surface or at another threading dislocation, and are structurally a combination of both screw and edge dislocations. [74], [65]. An idealised diagram of misfit and threading dislocations for germanium grown on silicon is shown in Figure 2.11. Since a threading dislocation is then an inherent rectifying structure for any lattice-mismatched heteroepitaxy, the total threading dislocation density—or TDD—is a useful measure of the quality of a heterointerface [70]. Threading dislocations may be visualised in heterointerfaces through tunnelling electron microscopy imaging, see Figure 2.12 and achievable defect densities for germanium–on–silicon growth are of the order of below $1 \times 10^7 \text{cm}^{-2}$ [75].

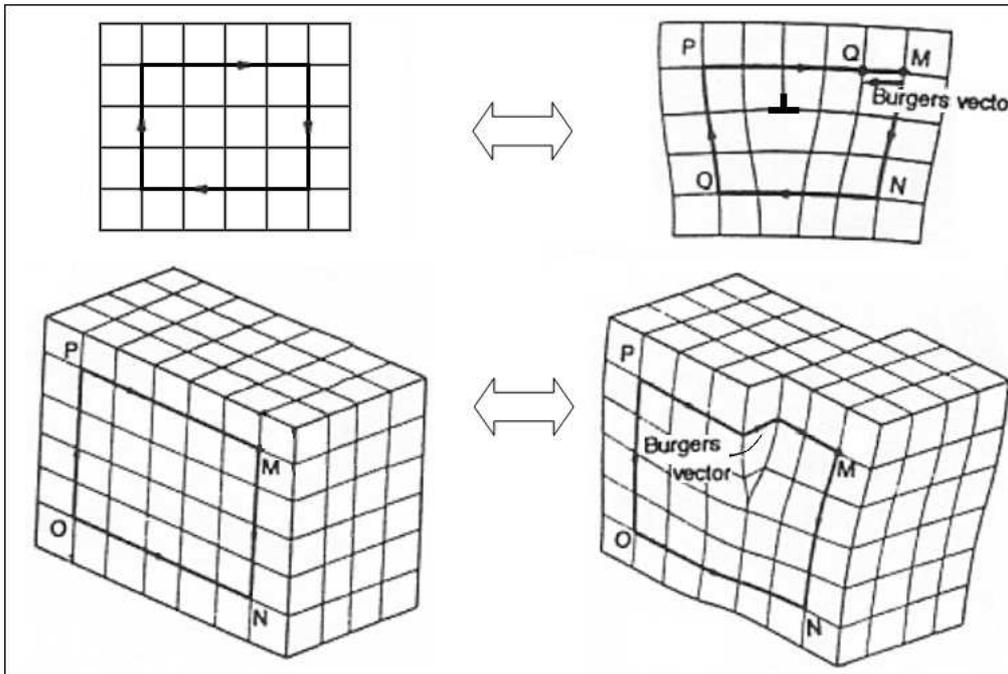


Figure 2.10: Burgers vector and Burgers circuits for edge and screw dislocations [73]

2.2.4 Interface Traps

Dislocations leave a number of atoms with a mismatched dangling bond, which can create a one-dimensional energy band within the band gap of semiconductor materials, from hybridization of the orbitals of the unbonded atoms. These can act as ‘trap states’ in the band gap, with consequences for the electronic and photoelectric properties of a material, and the effects of these states have been widely studied, both experimentally and theoretically. Changes in conduction and absorption/emission with increasing plastic deformation (i.e. greater dislocation density) of semiconductor materials has confirmed that these defects can be critical to device performance, and studies using photoemission and photoluminescence spectra, and Hall measurements have been used to characterise defect states [76–80].

Trap states are characterised by their energy relative to the valence and conduction band edges and are usually defined in terms of a density N_t and characteristic energy

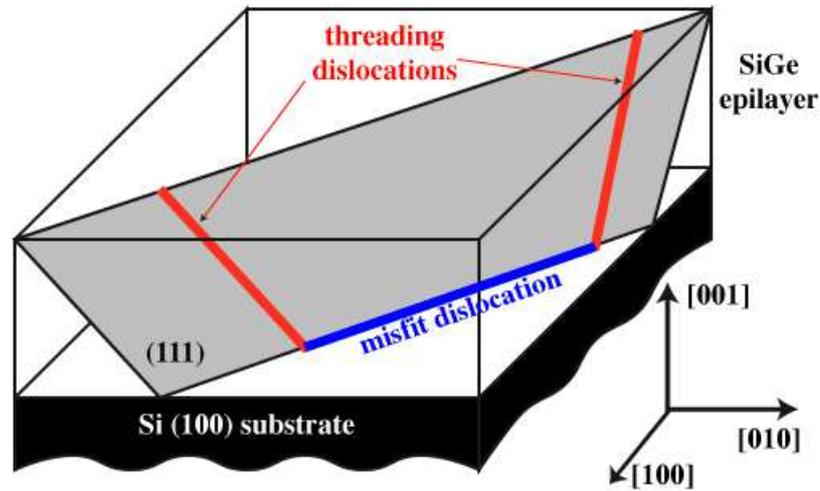


Figure 2.11: Misfit and threading dislocation [74].

E_t [81]. In general, traps closer to the conduction band will tend to trap and re-emit electrons from the conduction band, while traps close to the valence band will tend to trap and re-emit electrons into the valence band. Traps close to the band edges are denoted ‘shallow’ states, from the smaller energies needed for electrons and holes to interact with these energy levels, and defect states near the middle of the band gap are known as ‘deep’ states, and can perform both functions – trapping and re-emission of electrons or holes from either band. Hence, these can also act as recombination centres, since a two-level recombination process mediated by a trap is more energetically likely than a single step process, and have significant effects on the recombination rates of carriers within a semiconductor material or device [24]. In addition, deep states impact on optical absorption and emission characteristics [77]. Occupied trap states may also result in space charge, affecting the local electric field.

Efficient operation of a SiGe SPAD relies on photo-absorption, followed by impact ionization and rapid conduction of generated carriers over the SiGe interface and then quenching of this current; defects at the heterointerface may contribute significantly to performance deterioration in SPADs across these stages, and the possibility of defect

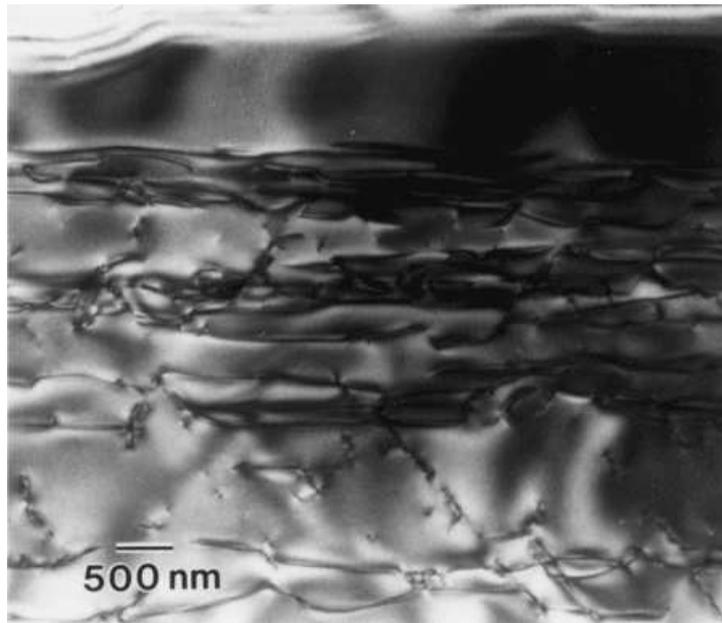


Figure 2.12: TEM image showing threading dislocations [74].

states in the band gap, and their energies, should be included in any model attempting to identify the effects of dislocations on SPAD performance.

2.3 Bandstructure Models for Simulation

The simplest formulation for electronic bandstructure is one where two permitted energy bands are assumed, separated by a band-gap, and this may be employed in commercial software (e.g. Silvaco), with single values for the energies of the valence and conduction band edges for a given material. This approach is computationally straightforward and adequate for simulation of processes where we are concerned with what happens at band edges, and with direct transfer of electrons from band-to-band, however, there are a number of limitations. For many materials (including silicon and germanium) the minimum energy in the conduction band exists at a different electron momentum to the maximum energy of the valence band; the smallest energy gap between the valence

and conduction bands is 'indirect', and transfer of electrons between bands requires a change in momentum as well as energy. If we wish to calculate processes between the conduction and valence band that do not involve a change in momentum then the 'direct' gap must be used (and hence the appropriate energy values for the valence and conduction band edges). Additionally electrons/processes away from band edges (i.e. high-energy processes) may not be adequately characterised by a single energy for each of the conduction and valence band. In these cases a more complex bandstructure model is required.

2.3.1 Analytical Bandstructures

For a free electron moving in a constant uniform electrostatic potential, the E - \mathbf{k} dispersion relation relating the energy E to the momentum wavevector \mathbf{k} is quadratic:

$$E(\mathbf{k}) = \frac{\hbar^2 |\mathbf{k}|^2}{2m_0} \quad (2.14)$$

The most straightforward deviation from a simple band model towards a description of multiple bands and valleys uses this quadratic relationship and assumes that close to the band edges the dispersion relationship maintains this parabolicity; the curvature of individual bands is then amended by amending the dispersion relation obtained for free electrons through the use of a variable 'effective mass' m^* (which may be smaller or larger than m_e) for each valley or band [82]:

$$E(\mathbf{k}) = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*} \quad (2.15)$$

where the effective mass is derived from the dispersion relation:

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2 E(\mathbf{k})}{d\mathbf{k}^2} \quad (2.16)$$

In general, the effective mass may also vary in different lateral directions i, j and so the full form is tensorial with components [33]:

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\delta^2 E(\mathbf{k})}{\delta \mathbf{k}_i \delta \mathbf{k}_j} \quad (2.17)$$

At low energies (i.e. close to the band edges) this parabolic approximation works well; however further away from band edges (e.g. in a high energy/high field regime) a non-parabolicity factor α may be included to account for the deviation from purely parabolic band behaviour:

$$E(\mathbf{k})(1 + \alpha E(\mathbf{k})) = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*} \quad (2.18)$$

where α has units of inverse energy and relates to the degree of admixture of valence band and conduction band states. α has an analytical expression derived from the effective mass and the width of the material band gap:

$$\alpha = \frac{(1 - \frac{m^*}{m_0})^2}{E_{gap}} \quad (2.19)$$

This non-parabolic expression for the band structure introduces a small degree of extra complexity, but may be suitable for high-energy device simulation (where the $\alpha E(\mathbf{k}^2)$ term in Equation 2.18 may dominate) or for materials with a small band gap where α becomes larger.

The methods discussed thus far offer relatively easily-calculable analytical band structures for bulk material or device simulations. However, for simulations where the only the entire energy dispersion relation provides the required accuracy, a full band model should be employed. Full electronic bandstructure calculations are usually complex, and numerical methods for these calculations can be broadly categorised into empirical methods which include fitting parameters from experimental data and *ab initio* atomistic methods, which are computationally expensive for anything more than a few atoms [82].

2.3.2 Periodic Structures and the Brillouin Zone

Calculation of a full bandstructure for a crystalline material in principle requires solution of the Schrodinger equation in 3 dimensions, and consideration of the the full potential under which electrons move. However, since, as discussed above, crystals may be described as a periodic repetition of units, the potential experienced by an electron in a periodic structure may be amended to take the form of a periodic potential with the

same periodicity as the the crystal lattice. Bloch's theorem states that a wavefunction for a single electron moving in this periodic potential takes the form of a Bloch function, composed of a plane wave envelope function and a periodic function with the periodicity of the lattice. Using index numbers n for electrons of identical momentum vector \mathbf{k} but different energies, the Bloch eigenstates may be written as:

$$\Psi_{n,\mathbf{k}} = u_{n,\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \quad (2.20)$$

where the wavefunctions $\Psi_{n,\mathbf{k}}$ and $u_{n,\mathbf{k}}$ are periodic in the direct lattice. Proofs of this theorem [83] show that the wavefunctions in a position \mathbf{r} are equal to those at positions $\mathbf{r}+\mathbf{R}$, where \mathbf{R} is a lattice vector. The energy solutions are also periodic in the reciprocal lattice [33, 54], and as a result it is only necessary to consider the electronic structure in the Brillouin Zone.

This, still, however, leads to a Schrodinger equation with a number of potential terms — those describing the interaction between the electrons and the periodic potential given by the nuclei, those describing nucleus–nucleus and electron–electron interactions, and exchange/correlation interactions arising from the Pauli Exclusion Principle. The problem is a many-body one, with no straightforward analytical solution [84].

2.3.3 Full Bandstructure Models

A number of approximations have been developed to simplify this many–body problem and yield computationally–achievable solutions that nonetheless provide sufficient accuracy in predicting the behaviour of real systems. The Born–Oppenheimer approximation assumes that, since electrons are much lighter than ions, they move much more quickly and follow ionic cores adiabatically, essentially treating the ions as static. This allows separation of ionic and electronic wavefunctions for solution of the Hamiltonian. The problem is still, however, complex, and must take account of ion–ion interactions,

ion–electron interactions, and electron–electron interactions [84, 85].

A further approximation involves treating the system as a set of independent electrons, acting under some overall potential provided by all other electrons in the system, and yielding the electronic Hamiltonian:

$$H\Psi_n = \left(-\frac{\hbar^2}{2m} + V_{ext} + V_{eff} \right) \Psi_n = \epsilon_i \Psi_n \quad (2.21)$$

where the Ψ_n are the wavefunctions of n independent electrons, V_{ext} is the external potential under which the electrons move (due to the ionic cores) and V_{eff} is the effective potential given by the average interaction with all other electrons. The Hartree potential treats V_{eff} as the average of the Coloumb potential of all other electrons, and this was expanded by inclusion of Fermi–statistics to provide the Hartree–Fock potential, which includes the effect of electron exchange — the requirement that electrons remain antisymmetric under exchange, a form of additional non–local potential under which electrons act. This still, however, omits the effects of correlation — the potential arising from screening of exchange effects, and provides only approximate solutions [84, 85]. Overall, *ab initio* methods rely on schemes to approximate the forms of these two potentials V_{eff} and V_{ext} while still yielding accurate calculation results.

2.3.4 Pseudopotentials

The use of a pseudopotential – an approximate potential describing the most significant interactions experienced by an electron – was first introduced by Fermi, and almost simultaneously by Hans Hellman [86, 87]. The basis for the method is that we are generally interested only in the behaviour of the valence electrons, as these are the most chemically active, and that some approximations can be used to deal with the core electrons, and the pseudopotential method provides approximate solutions, which nonetheless have provided sufficiently accurate results in real systems for considerably less computational effort.

Development of the form of this pseudopotential, and of computationally-efficient schemes for calculation of electronic structures has proceeded in a number of directions. Calculation of the form of the pseudopotential may be purely *ab initio*, or empirical, where fitting to experimental data is used to adjust the pseudopotential.

***Ab Initio* Pseudopotentials**

The pseudopotential method originated from the orthogonal plane wave (OPW) method developed by Herring [88]. This method relies on separation of core and valence states, and construction of valence electron wavefunctions from a basis set of plane waves that must be orthogonal to those of the core electron wavefunctions [54]. The core wavefunctions are assumed to be known, and under the ‘true potential, the rapidly oscillating inner core potential creates a large kinetic energy pressure on the valence electrons from the need to maintain the orthogonality. This generates a form of Pauli repulsion term, which approximately cancels out the Coulomb potential, leading to an overall weak repulsive core potential: the frozen core approximation [86, 89]. The complex and varying core potential can then be replaced, to a good approximation, with a simpler form, and Phillips and Kleinman [90] further developed this method by suggesting that the core term in the orthogonalisation equation could be replaced by smooth analytic functions, with limited impact on the calculation results for valence electrons.

Construction of an *ab initio* pseudopotential in practice proceeds by first selecting a reference atomic configuration, then performing an all-electron calculation using an atomic simulation. Next, a pseudo-wavefunction is constructed that matches that from the all-electron wavefunction as closely as possible: the parameters for matching depend on the requirements for the pseudopotential, but must include that the valence eigenvalues match those from the all-electron calculation, and that the continuity of the wavefunctions and their first derivatives are matched across the core-valence boundary. Once these parameters have been satisfied the Schrodinger equation is then inverted to obtain the pseudopotential [84, 91].

The choice of cutoff radius r_c for the pseudopotential, i.e. the radial distance r at which the potential is assumed to be core for $r < r_c$ and valence for $r > r_c$ will determine the number of plane waves used to construct the pseudopotential, and hence the computational burden. Pseudopotentials should ideally be as ‘soft’ as possible, i.e. use as few plane waves in their construction as possible, while reproducing the valence charge density as accurately as possible. Two significant formulations for the calculation of pseudopotentials have been developed to meet these requirements. Norm-conserving pseudopotentials are based on the principle that the pseudo wavefunctions and eigenvalues, and the logarithmic derivatives and energy derivatives of the wavefunctions must match those from the all-electron calculation for $r > r_c$, and that the integrals from 0 to r_c of the charge densities must match (norm-conservation). Norm-conserving pseudopotentials maintain the same scattering properties at valence energies as the ‘true’ core potentials that they are approximating [84, 91, 92]. Ultrasoft pseudopotentials were developed for systems with highly-localised valence orbitals, and relax the norm-conservation requirement, and the scattering properties and energy derivatives of the valence wavefunctions match only at a set of chosen reference valence energies. Ultrasoft pseudopotentials use a limited number of plane waves and are soft as possible within the core, with an additional core augmentation charge term [84, 91, 93, 94].

In practice, pseudopotentials for many materials have been calculated and tabulated under a range of conditions, and it is possible to perform electronic structure calculations either using this existing data or by constructing a pseudopotential *ab initio* with existing software.

Empirical Pseudopotential Method

In the Empirical Pseudopotential Method (EPM), pseudopotentials are constructed based on empirical data (measured bandstructures, spectral data) and adjusted using a

series of fitting parameters. The Hamiltonian for solution is

$$H = \frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \quad (2.22)$$

where $V(\mathbf{r})$ is a weak position-dependent potential. The potential can be expressed as the product of a structure factor $S(\mathbf{G})$ and a form factor $V(|\mathbf{G}|)$:

$$V(\mathbf{r}) = \sum_{\mathbf{G}} S(\mathbf{G}) V(|\mathbf{G}|) e^{i\mathbf{G}\cdot\mathbf{r}} \quad (2.23)$$

The structure factor $S(\mathbf{G})$ is determined from the structure of the crystal lattice, and the form factors $V(|\mathbf{G}|)$ are determined empirically. From here, the EPM proceeds through a matrix solution of the Hamiltonian, with initial values of these form factors used to calculate quantities, such as direct or indirect bandgaps, reflectivities, or densities of states that have been observed experimentally. The form factors are then adjusted until the calculated quantities are in closest possible agreement with the experimental values. In practice, only a certain number of the form factors need be used and adjusted. The contributions from the first five \mathbf{G} vectors, with squared magnitudes of 0, 3, 4, 8 and 11, are useful to consider; the structure factor for $|\mathbf{G}_2|^2 = 4$ is zero, leaving only form factors V_3 , V_8 and V_{11} to be determined. Although further form factors can be determined, these are likely to represent smaller contributions to the overall potential, and can usually be ignored; for example, for silicon, the V_3 , V_8 and V_{11} form factors have yielded band structure results to within a few percent accuracy [86, 95]. The method is well-developed and has been widely applied, and significant work in calculating form factors for individual semiconductor materials now means that the method can be applied using literature values for the form factors [86].

This formulation is, however, local — the pseudopotential is assumed to be a simple function of position and the energy and angular momentum dependence of the core region is ignored. A further non-local formulation takes account of the first three angular momentum terms and applies a correction factor [96].

2.3.5 Density Functional Theory

Density Functional Theory (DFT) is a powerful method for solution of the many-body problem in solid state physics. The basis for DFT is that the total ground state energy of the system of n electrons that we are concerned with is a unique functional of the electron density; this also implies the converse, that if the electron density is known, then the ground state energy of the system can be determined. This basis, however, gives no indication of the form of the functional. The second, variational principle, of DFT, states that the electron density that minimizes the energy of the overall functional is the true electron density; hence, if the form of the functional was known, it would be possible to vary the electron density until the energy was minimised, giving the true electron density [97].

The relationship between the energy and the electron density takes the form

$$E = E[\rho(r)] = \int dr V_{ext} \rho(r) + F[\rho(r)] \quad (2.24)$$

where ρ is the electron density, V_{ext} is the external potential, and F is the functional term given by the effective potential in the Hamiltonian.

The Kohn-Sham Equations

The effective tool for solution in DFT calculations usually takes the form of the Kohn-Sham equations. In this formulation, the density is written in terms of the wavefunctions of the n non-interacting electrons:

$$\rho(r) = \sum N n = 1 \Psi_n^*(r) \Psi_n(r) \quad (2.25)$$

F may be further broken down into three components:

$$F[\rho(r)] = E_K[\rho(r)] + E_{xc}[\rho(r)] + E_H[\rho(r)] \quad (2.26)$$

where E_K , E_{xc} and E_H represent the kinetic energy of the electron, the energy due to the exchange–correlation potential, arising from the Pauli principle and electron statistics,

and the energy due to the Hartree potential, respectively. The effective potential V_{eff} in Equation 2.21 is given by the Kohn–Sham potential V_{KS} :

$$V_{KS}(r) = \frac{\partial}{\partial \rho(r)}(E_{xc}[\rho(r)] + E_H[\rho(r)]) = V_H(r) + \frac{\partial E_{XC}[\rho(r)]}{\partial \rho(r)} \quad (2.27)$$

Solutions using this formulation are then found by solving the one–electron wavefunctions Ψ_n and the charge density self–consistently over a number of iterations using the Kohn–Sham equations, until agreement is found to within some defined convergence tolerance. [84, 85, 97]. In practice, calculations proceed using wavefunctions expanded in a plane wave basis set as described above, and using pseudopotentials to describe the interactions between electrons and the ionic core.

XC Functionals

Although a self–consistent formulation for calculating the electron density and one–particle wavefunctions is described above, this still requires knowledge of the form of E_{XC} , or the electron–electron interactions due to the effects of the exchange and correlation forces. In principle, there is no straightforward *ab initio* formulation for this interaction for most systems, however, in practice, a number of simple formulations provide sufficiently accurate results and act as a ‘correction factor’ in DFT calculations.

The Local Density Approximation (LDA) assumes that the electron density behaves as that from a uniform electron gas of density ρ [84, 85]:

$$E_{XC} = \int dr \rho(r) \epsilon_{XC}(\rho(r)) \quad (2.28)$$

Though this approximation ignores all spatial variations in density, it frequently provides good results in many systems. An alternative formulation is the Generalised Gradient Approximation (GGA). Functionals based on this semi–local approximation are based on a slowly–varying electron density, and improve on the simplicity of the LDA by including the gradient of the density [98].

Despite the successes of these functionals, they have a number of limitations, in particular for DFT calculations with semiconductors. The use of a uniform or slowly-varying electron density introduces artificial electron self-interactions, and does not take account of occupation numbers of electrons in bands. These effects create an additional deocalisation in occupied states, which increases their energy, and hence, GGA and LGA functionals usually inherently underestimate the band gap in semiconductor materials, or provide calculations with zero band gap when used without correction [77] A number of corrections to this self-interaction have been applied to mitigate the zero-bandgap problem [99].

Cut-off Energy and Sampling Grids

Bloch's Theorem (Equation 2.20) also allows construction of electronic wavefunctions in terms of a basis set of plane waves with coefficients $u_{n,\mathbf{k}}(\mathbf{G})$:

$$\Psi_{n,\mathbf{k}} = \sum_{\mathbf{G}} u_{n,\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (2.29)$$

where the \mathbf{G} are the general reciprocal lattice vectors described above. In this case, \mathbf{G} may relate to the reciprocal lattice of an entire supercell used for a DFT calculation, rather than simply any primitive crystal cell. In principle, there are an infinite number of \mathbf{k} -vectors in the expansion, however, in practice, the \mathbf{k} -vector grid can be sampled, since two \mathbf{k} -vectors very closely spaced will lead to only a negligible change in $\Psi_{n,\mathbf{k}}$ and some plane waves with an energy greater than some cutoff-energy can be excluded, with limited effects on the accuracy of the calculation. The choice of cut-off energy and sampling grid is therefore critical to DFT calculations. The cutoff energy determines the number of plane waves used in the basis set, and the size of the Fourier transform (FFT) grid used in calculations.

A number of \mathbf{k} -point sampling methods exist. The Monkhorst-Pack formulation is commonly used, which identifies a set of special points in the reciprocal lattice that improve the accuracy of the calculation with minimum sampling, by meeting a number

of conditions, which guarantee that quantities such as charge densities or energies can be more accurately averaged over the Brillouin Zone [100, 101].

Typical bandstructures for silicon and germanium are shown in Figures 2.13 and 2.14.

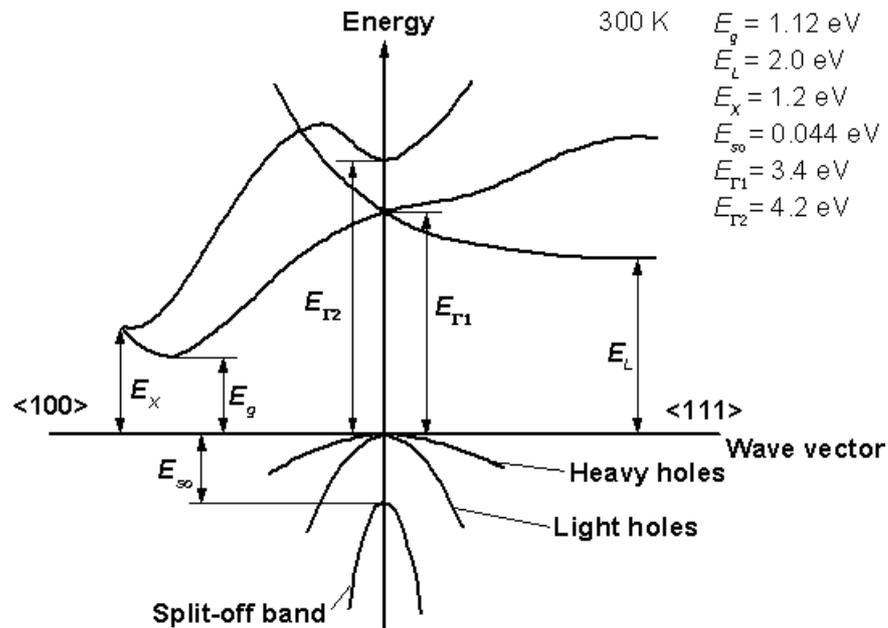


Figure 2.13: Typical bandstructure for silicon (with direct and indirect bandgap energies) [102]

2.4 Electron Transport in Semiconductors

Modelling electron transport in a semiconductor device requires a self-consistent solution for a number of equations [104]. Poisson's equation relates the electrostatic field to the space charge density:

$$\nabla \cdot (\epsilon \nabla \Psi) = -\rho \quad (2.30)$$

where Ψ is the potential, ϵ is the local permittivity and ρ is the local space charge density. The space charge density includes all contributions from mobile and fixed charge and

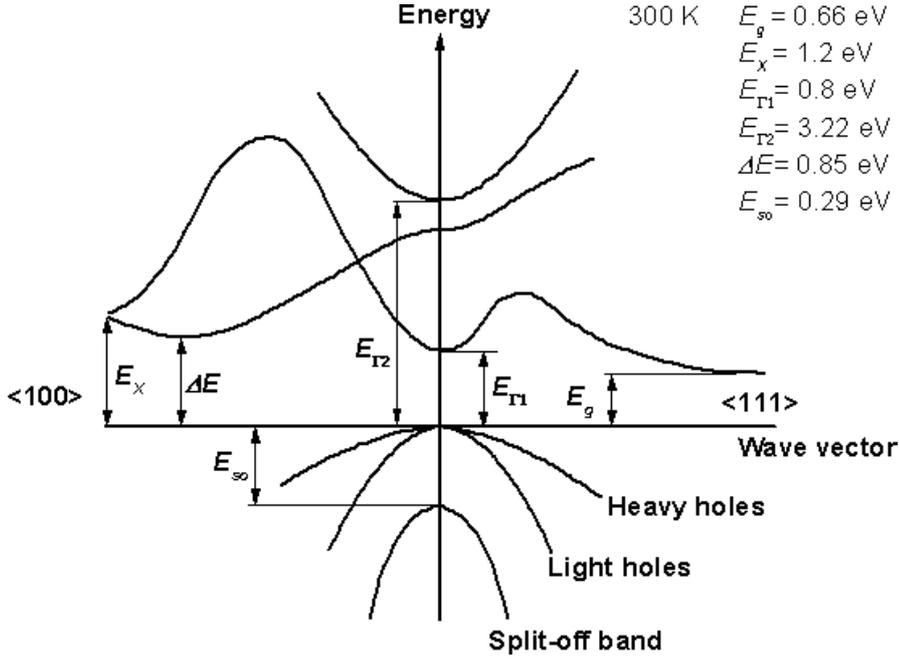


Figure 2.14: Typical bandstructure for germanium (with direct and indirect bandgap energies) [103]

other sources such as ionized impurities. The electric field E can then be obtained from the potential by:

$$E = -\nabla\Psi \quad (2.31)$$

The continuity equations describe the conservation of carriers:

$$\frac{\delta n}{\delta t} = \frac{1}{q} \nabla \cdot J_n + G_n - R_n \quad (2.32)$$

$$\frac{\delta p}{\delta t} = -\frac{1}{q} \nabla \cdot J_p + G_p - R_p \quad (2.33)$$

where J represents the current densities of electrons and holes respectively and G and R are the carrier generation and recombination rates; q is the electronic charge.

For modelling the distribution of carriers, a kinetic theory approach is usually adopted for carrier densities, where the probability of finding a carrier with a given momentum \mathbf{k} at position \mathbf{r} at time t is calculated stochastically using a distribution function. This is

the basis of the Boltzmann transport equation (BTE), which considers the total number of particles entering and leaving a region of phase space, and is given in totality by:

$$\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} + \frac{1}{\hbar} \nabla_{\mathbf{k}} \cdot E(\mathbf{k}) \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t) + \frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t) = \frac{\partial f}{\partial t} |_{coll} + s(r, k, t) \quad (2.34)$$

The first, second and third terms on the left hand side represent, respectively, the rate of change of the particle distribution due to the action of the Lorentz force, the rate of change of the distribution due to the action of diffusion, from temperature, concentration or other thermodynamic gradients and the rate of change of distribution due to the action of collisions—the total of in-collision forces and out-collision forces. A detailed derivation of the BTE and a breakdown of terms can be found in Reference [82]. In general a full solution of the BTE is not practical and the collision term in particular may take an integro-differential form which is computationally complex to solve [105]

2.4.1 The Drift Diffusion Approximation

The most straightforward method for approximate solution of the BTE is the drift–diffusion approach. This approach is attractive, since no independent variables are introduced beyond Ψ , n , and p [104]. Some further simplifying assumptions are applied to the BTE to obtain the drift–diffusion approximation:

- It is assumed that all collisions are elastic.
- It is assumed that the carrier velocities adjust instantaneously in the local electric field, hence detail about short-scale interactions is neglected.
- Since the model uses near-equilibrium conditions, it is only valid for small perturbations from equilibrium.
- Only a value rather than a distribution function for the velocity is extracted, hence the model does not take account of noise well.
- The model cannot describe transients.

A mobility term μ is introduced, given by:

$$v = \mu E \quad (2.35)$$

which relates the velocity of carriers v to the electric field E . At low field, μ may be assumed constant and the relationship between carrier velocity and electric field is linear. At higher fields, a number of physical effects will affect this linear relationship, including scattering from optical phonons and ionised impurities and the carrier velocity may saturate. More complex models of mobility may be developed analytically or empirically to take account of high fields [82].

The drift diffusion approximation is then developed by expressing the current densities using the quasi-Fermi levels for electrons and holes, which are then linked to the carrier concentrations and potential through the Boltzmann approximations. Effective electric fields, E_n and E_p are defined, and amended expressions for the the current densities can be derived:

$$J_n = qn\mu_n E_n + qD_n \nabla_n \quad (2.36)$$

$$J_p = qp\mu_p E_p - qD_p \nabla_p \quad (2.37)$$

These are the drift-diffusion equations, where the first term in each represents the effect of drift on carrier densities and the second gives the effect of diffusion, where D_n and D_p are the diffusivities of electrons and holes, respectively. A full derivation is given in [104]. These expressions for current density can then be applied to the continuity equations 2.32 and 2.33, and then solved self-consistently with Poisson's equation 2.30 to model electron transport in a semiconductor device. The choice of timestep for

Self-consistent solutions of this set of equations requires separation of the device or material in which transport is being studied into a mesh of spatial units. The total charge in the device must be conserved, and the current on the boundaries of each mesh box is defined, with the requirement that the current entering each box must be equal to the current exiting within each simulation time step. The choice of timestep must be

sufficient to meet the relaxation condition relating to the approximation that interactions are instantaneous, and the choice of mesh size must be smaller than the range of electrostatic interaction of each particle. [82].

A number of simulation packages can be used to solve these equations self-consistently and provide calculations of carrier densities, electric fields and other parameters on the device scale.

2.5 Conclusion

This chapter introduces the basic solid state physics and electronics theory relevant to development of SPADs, including the principles of photodiodes, and the operating conditions specific to use of photodiodes as single photon detection devices. A number of sources of performance limitation for SPADs are discussed, including noise currents from Zener tunnelling, arising as a result of the high fields needed for operation of a photodiode as a SPAD, and defects in the crystal structure due to the lattice mismatch between silicon and germanium. A number of considerations for modelling the performance limitations of SPADs are discussed, as well as the equations required to simulate semiconductor materials and full semiconductor devices. In the following sections, these models and equations will be applied to investigate the performance of SPADs, including modelling of fields and carrier densities in a SPAD, an assessment of band-to-band tunnelling currents in germanium under a range of field conditions and the electronic and geometric structure of defects in bulk germanium.

Chapter 3

Electron Transport in SPADs

3.1 Introduction

In this chapter, a number of SPAD devices are investigated, using the Silvaco simulation software. The software enables simulation of whole devices and an extensive number of thermal, electrical and optical models and inputs are available for device characterisation. The approach followed that discussed in the previous chapter, applying the drift-diffusion model and solving the carrier continuity equations self-consistently with the Poisson equation for whole SPAD devices. From here, IV curves, electric field profiles, carrier density profiles and other device parameters can be calculated. Silvaco's ATLAS solver was used to undertake modelling of the device in two dimensions - the longitudinal direction of the growth axis and one lateral axis perpendicular to this across the width of the device. Device characteristics in the third width direction, also perpendicular to the growth axis, were assumed to be identical to those in the second lateral direction.

The SPAD devices under investigation are silicon-germanium SACM-APDs operated under breakdown conditions to act as Geiger-mode single photon detectors. The device follows the mode of operation of the $p-i-n$ photodiode discussed earlier, with an additional highly-doped charge layer which regulates the electric field between the germanium absorption layer (where a low field is desirable to limit noise and dark count due

to carrier tunnelling and other effects) and the silicon multiplication layer, where a high field is required to achieve avalanche multiplication. The doping profile of the SPAD devices may be of a ‘PiPiN’ scheme, (with highly p^{++} and n^{++} doped end regions in the germanium and silicon respectively, and a p doped charge layer) or a ‘NiNiP’ scheme, with the reverse doping profile. Previous SPAD fabrication has indicated that a PiPiN structure is easier to grow, as p -type dopants are less susceptible to surface segregation in silicon and germanium, hence maintaining the doping density in the charge layer as close to the target density as possible, and reducing the possibility of background doping in the ideally-intrinsic germanium absorber. Transport from the germanium to the silicon layer may be inhibited by the band offset between these materials, particularly for transport between the L and Γ valleys, where a potential barrier is presented between the germanium and silicon layers, as shown in Figure 3.1. A typical PiPiN SPAD structure is shown in Figure 3.2, including indicative design thicknesses. In principle, the absorption and multiplication layers are intrinsic, but may in reality be unintentionally doped. Here, the effects of background doping type in the absorption layer are also investigated, to establish IV-curves, carrier distributions and the influence of background doping in the absorber layer on transport over the heterojunction interface.

3.2 Investigation of Doping Profiles

Although the germanium absorption layer is ideally intrinsic, some background doping of unknown concentration is likely. Background doping in the germanium absorption layer may be n -type or p -type. This background doping will affect the electric fields and carrier transport across the device and have an impact on device performance. Additionally, the intentional doping in the silicon charge layer can be varied to optimise fields and device performance. A number of permutations of these possible doping profiles have been simulated and device characteristics (electric field, potential, carrier concentrations) calculated for each. It was assumed that background doping in the silicon multiplication layer is low ($1 \times 10^{15} \text{cm}^{-2}$) and of n -type.

Table 3.1: Doping profiles for SPAD device simulations

| Simulation | Charge Layer Doping Density (cm^{-2}) | Absorption Layer Doping Density (cm^{-2}) | Absorption Layer Doping Type |
|------------|---|---|---------------------------------|
| 1 | 5×10^{17} (high) | 1×10^{16} (high) | <i>p</i> -type |
| 2 | 5×10^{17} (high) | 1×10^{16} (high) | <i>n</i> -type |
| 3 | 5×10^{17} (high) | 1×10^{15} (low) | <i>p</i> -type |
| 4 | 5×10^{17} (high) | 1×10^{15} (low) | <i>n</i> -type |
| 5 | 2×10^{17} (medium) | 1×10^{16} (high) | <i>p</i> -type |
| 6 | 2×10^{17} (medium) | 1×10^{16} (high) | <i>n</i> -type |
| 7 | 2×10^{17} (medium) | 1×10^{15} (low) | <i>p</i> -type |
| 8 | 2×10^{17} (medium) | 1×10^{15} (low) | <i>n</i> -type |
| 9 | 1×10^{17} (low) | 1×10^{16} (high) | <i>p</i> -type |
| 10 | 1×10^{17} (low) | 1×10^{16} (high) | <i>n</i> -type |
| 11 | 1×10^{17} (low) | 1×10^{15} (low) | <i>p</i> -type |
| 12 | 1×10^{17} (low) | 1×10^{15} (low) | <i>n</i> -type |

The combinations of doping profile in the charge layer and absorption layer simulated are listed in Table 3.1; these permutations of dopant type and density result in 12 distinct simulations. Two-dimensional device simulations were carried out using Silvaco's ATLAS device simulator, implementing a drift diffusion model with field-density-dependent and carrier-density-dependent mobility terms, and including the effects of Shockley-Read-Hall generation/recombination processes. The device width was $20 \mu\text{m}$ and simulations in this direction were conducted using a regular $2 \mu\text{m}$ mesh. The total height of the device was $2.2 \mu\text{m}$, and a $0.01 \mu\text{m}$ mesh size was used along this length, to fully elucidate the carrier concentrations, and field and potential effects along the critical transport direction for electrons. The Okuto-Crowell model [107] was used to simulate the effects of impact ionisation; this method is based on pseudolocal impact ionisation co-efficients for

electrons and holes that depend on both electric field and position, and is particularly suitable for high electric fields.

3.2.1 Breakdown and Punchthrough Voltage

IV profiles for the devices under increasing reverse bias voltage were produced for PiPiN devices with the range of doping profiles listed in Table 3.1. As a SPAD operates above reverse breakdown voltage, the breakdown voltage for each doping profile across the device was a critical parameter of interest from the simulations; ideally, the IV curves should show an identifiable breakdown point with a sudden increase in current. In addition, the voltage at which the device showed full carrier depletion of the absorber layer was also sought from the simulation results. This has a critical impact on the ability of the device to act as a SPAD: insufficient carrier depletion in the absorber layer will lead to additional dark currents. This effect can be observed from a small but noticeable increase in the current at some voltage below the breakdown voltage, commonly referred to as the ‘punchthrough’ voltage from the phenomenon observed in FET structures, although the mechanism in a SPAD is not identical.

IV curves for all combinations of doping density and both n - and p -type doping in the absorption layer are provided in Figures 3.3, 3.4 and 3.5, and the breakdown and punchthrough (where observed) voltages are provided in Table 3.2.

Broadly, the breakdown voltage increased with decreasing charge layer doping. This is to be expected, since the charge layer doping distributes the total potential difference between the absorption layer and the multiplication layer. A higher dopant concentration in the charge layer increases the proportion of the total potential difference which appears across the silicon multiplication layer, meaning that carriers in the multiplication layer are already more energetic than in the lower-field case, and a smaller reverse bias voltage is required to initiate avalanche breakdown.

The relationship between breakdown voltage and absorption layer dopant density and

Table 3.2: Simulated breakdown and punchthrough voltages for SPAD device simulations

| Simulation | Breakdown Voltage (V) | Punchthrough Voltage (V) |
|-------------------|----------------------------------|-------------------------------------|
| 1 | 35 | (none) |
| 2 | 35 | (none) |
| 3 | 35 | (none) |
| 4 | 35 | (none) |
| 5 | 38 | (none) |
| 6 | 48 | 30 |
| 7 | 40 | 40 |
| 8 | 40 | 30 |
| 9 | 48 | 18 |
| 10 | 60 | 13 |
| 11 | 52 | 16 |
| 12 | 52 | 14 |

type is slightly more complex. p -type doping and n -type doping will tend to change the direction of the electric field gradient across the absorption layer; assuming the highly-doped charge layer produces (for any given doping density) a fixed difference in electric field between the absorption and multiplication regions, the electric field at the interface of the absorption and charge layers (and hence the direction of the electric field gradient across the absorber) determines the electric field across the multiplication region. From the results in Table 3.2, it can be seen that n -type absorption layer doping produces a significantly higher breakdown voltage than p -type doping, with a higher doping density in the absorption region leading to a greater difference between breakdown voltages for each type.

A number of other characteristics can be seen in the IV curves for each doping profile. The phenomenon of punchthrough was observed in a number of simulations, and the voltages at which this was observed are summarised in Table 3.2. This current increase is only observed for medium and low charge layer doping and is occurs at a higher voltage when the absorption layer is p -doped than when it is n -doped. This sudden increase in current below the breakdown voltage is most likely related to carrier transport in the charge layer interface and is discussed in more detail later in alongside simulations of electric field and carrier concentrations.

The IV curve for simulation 2 shows an additional increase in current at a higher voltage than the breakdown voltage—at 38 V reverse bias voltage compared to a breakdown voltage of 35 V. This may be related to breakdown in the Ge absorption layer.

3.2.2 Electric Field Profiles

Electric field profiles at increasing bias voltages have been produced for all doping profiles. The direction of growth is from left to right along the horizontal axis (silicon to germanium) of the graphs, and the silicon–germanium interface is marked as a thick black line. Broadly, they show the effect of the charge layer doping density in distribut-

ing the electric field between the absorption and multiplication layers, and the impact of n -type and p -type doping in the absorption layer on the electric field gradient. The breakdown field in germanium is 10^5Vcm^{-1} and that in silicon is $3 \times 10^5 \text{Vcm}^{-1}$. Fields in the absorption and multiplication regions should be distributed with these breakdown fields in mind to remain below breakdown in the germanium layer, and above it in the silicon layer. For high charge layer doping of $5 \times 10^{17} \text{cm}^{-2}$ a number of characteristics can be observed in the electric field profiles:

- Electric fields in the silicon multiplication layer are high (relative to the absorption layer) for all permutations of absorption layer doping. The electric field is consistently above the silicon breakdown field for bias voltages above the breakdown voltage observed in the IV curves.
- Electric fields in the absorber layer are consistently below the breakdown field of germanium.
- Electric fields in the absorber layer are zero for all bias voltages below breakdown when the absorber is p -doped, regardless of the doping density in the absorption layer, and also for the condition where the absorber is n -doped at the lower density of $1 \times 10^{15} \text{cm}^{-2}$.
- When the absorption layer is n -doped at the higher density ($1 \times 10^{16} \text{cm}^{-2}$), there is a shallow electric field gradient in the absorption layer even at bias voltages below breakdown, nonetheless, this is small and remains below the breakdown field for germanium.
- These field profiles are reflected in the low breakdown voltages for these simulations, and the similar breakdown voltages for all simulations in which high charge layer doping has been implemented.

Electric field profiles for the case of high charge layer ($5 \times 10^{17} \text{cm}^{-2}$) and high absorption layer ($1 \times 10^{16} \text{cm}^{-2}$) doping are shown in Figures 3.6 and 3.7, demonstrating the effects

discussed.

Notably, for the case of medium charge layer doping density ($2 \times 10^{17} \text{cm}^{-2}$, simulations 5-8), the density and type of absorption layer doping begins to play a more significant role in device characteristics. For low absorption layer doping ($1 \times 10^{15} \text{cm}^{-2}$, simulations 7 and 8) fields in the absorption layer are close to zero, fields in the silicon multiplication region are relatively high and breakdown voltages are similar for both the n - and p -doped absorption layers, as in the case of high charge layer doping. However, for high absorption layer doping ($1 \times 10^{16} \text{cm}^{-2}$) of n -type (simulation 6), the electric field due to the doping profile of the absorption region becomes significant. At increasing bias voltage, the depletion of carriers in the absorber leads to an extending region of negative space charge; this in turn creates an electric field which opposes that set up by the bias voltage, and in which the field is lower at the interface between the charge layer and the absorption layer than at the top contact. In addition, medium charge layer doping produces a smaller field difference between the absorption layer and the multiplication than for higher charge layer doping. As a result electric fields in the multiplication layer are significantly lower than for the doping profiles discussed so far, and there is a significant increase in breakdown voltage. In contrast, electric fields in the absorption layer begin to approach the breakdown field in germanium; for higher bias voltages, in cases where the absorption layer doping is high ($1 \times 10^{16} \text{cm}^{-2}$), the breakdown field is exceeded at sharp points in the germanium layer for p -type doping, and exceeded consistently at for n -type doping. The electric field profiles for medium charge layer doping and high absorption layer doping (simulations 5 and 6) are shown in Figures 3.8 and 3.9.

For low charge layer doping ($1 \times 10^{17} \text{cm}^{-2}$, simulations 9-12) this effect is more pronounced still: breakdown voltages are higher overall compared to devices with higher dopant densities in the charge layer (since a smaller proportion of the total potential difference is distributed across the multiplication layer and thus a higher bias voltage

is required to initiate breakdown). For the case where the charge layer doping is low and the absorption layer doping is high, dopant type has a significant impact on the breakdown voltage due to the space charge field set up in the absorption layer— with a decreasing absorption layer field gradient for p -type and an increasing field gradient for n -type absorption layer doping. Electric field profiles for low charge layer doping and high absorption layer doping (simulations 9 and 10) are shown in Figures 3.10 and 3.11. The field in the germanium layer exceeds the breakdown field for higher bias voltages in all cases where the charge layer doping is low (simulations 11 and 12).

3.2.3 Potential Profiles and Carrier Concentrations

A potential profile for a typical PiPiN device in which the absorption and multiplication regions are purely intrinsic will contain a potential well for holes in the interface between the charge layer and multiplication layer, which will inhibit the movement of holes when the electric field is not sufficiently high to overcome this well, and cause removal of holes to ionized acceptors in the depleted p -type region. Increasing bias voltage might be expected to flatten the potential barrier entirely for some doping profiles and at this voltage a small increase in current due to movement of electrons and holes with uninhibited transport would be observed. This phenomenon is observed in MOSFET devices, where it is denoted ‘punchthrough’. At significant bias voltage (towards the region of punchthrough for those profiles it is present in) the punchthrough voltage is consistent with the voltage at which the potential barrier is removed, and this effect is shown in Figure 3.12 for the simulation with low charge layer doping and high n -type absorption layer doping (simulation 10). Alternatively, this ‘punchthrough’ voltage may represent the bias at which the device becomes fully depleted, and depletion could also be the source of the small current increases (this depletion is also known as ‘punchthrough’ when applied to SPADs) observed in the IV curves for some doping profiles. In general, depletion is likely to occur in a device where there is sufficient electric field across the absorption layer to remove all carriers before the device breaks down; hence, under this definition, the simulations including high charge layer doping (simulations 1-4), and

that including medium charge layer doping with high p -type absorption layer doping (simulation 5) will have an insufficient field in the germanium for full depletion before breakdown. Any device in which breakdown occurs before punchthrough will not function effectively as a SPAD device: the residual carriers in the device will lead to large dark currents [40]. In contrast to the trend observed between the multiplication layer electric field and the breakdown voltage, where a higher field results in a lower breakdown voltage, punchthrough is only observable where the electric field across the multiplication layer is relatively low, from a combination of a number of doping profiles, as listed in Table 3.1, which distribute the electric field in this way:

- low charge layer doping density ($1 \times 10^{17} \text{cm}^{-2}$, simulations 1-4), meaning there are fewer carriers to deplete from the charge layer, and punchthrough can be achieved at a lower voltage.
- medium doping density in the charge layer ($1 \times 10^{17} \text{cm}^{-2}$), and a low dopant density in the absorption layer ($1 \times 10^{15} \text{cm}^{-2}$, simulations 7 and 8);
- medium doping density in the charge layer ($1 \times 10^{17} \text{cm}^{-2}$), and high absorption layer doping density ($1 \times 10^{16} \text{cm}^{-2}$) of n -type (simulation 6), opposing the field in the multiplication layer.

Doping profiles which reduce the electric field across the multiplication region and increase the field in the absorber will act to lower the bias voltage required to remove the potential barrier in the charge layer and achieve punchthrough.

Electron and hole concentrations were calculated for a range of bias voltages for all simulation cases. Breakdown is evident in electron concentration profiles for simulations 1-5, 10 and 12 where a significant increase in electron concentration in the multiplication layer at voltages above the breakdown field can be seen. Figures 3.13 and 3.15 demonstrate this effect. For simulations 6-8, the simulation voltage did not exceed the breakdown voltage observed in the IV curves, and this effect can not be seen. For simulations 9 and 11, which both included the lowest charge layer doping density simulated,

a drop in electron concentration was observed in the multiplication layer, close to the charge layer, at voltages above breakdown, however, the electron concentrations are so low that these data are likely not significant, and may be the product of numerical noise. An interesting effect is observed in all simulations, when comparing electron concentrations to hole concentrations. Within the first section of the multiplication layer, a ‘spike’ in hole concentration is observed, indicating that the effect of the field and charge layer is to sweep holes into the absorption layer. These peaks can be observed even when the applied bias voltage is below the breakdown field. A corresponding drop in electron concentration is seen at this locale, indicating recombination of electrons and holes in this space charge region. Further into the multiplication layer, the electron and hole concentration profiles are dependent on the dopant type and density in the multiplication layer, as may be expected. n -type doping produces a flattened hole concentration profile, and an electron concentration profile wherein electrons are distributed according to the potential profile, while the converse is true for p -type doping. These effects can be seen in Figures 3.15, 3.15, 3.17 and 3.18. In the IV curve for high charge layer doping and high n -type absorption layer doping (simulation 2), there is what appears to be an additional breakdown event at around 38 V reverse bias voltage, after the initial breakdown. The high hole concentration in the absorption layer for biases (Figure 3.18) above this voltage indicate that this secondary breakdown may be caused by hole-initiated impact ionisation in the germanium absorber.

3.2.4 Experimental Measurements

A number of SPAD devices have been fabricated, processed and characterised based on device design undertaken as part of the SPADs project. The breakdown characteristics and doping profiles of these devices have been investigated and can be compared to the target profiles simulated here.

A first set of SPAD measurements were taken for three fabricated devices at Heriot-Watt in November 2011; the target charge layer doping for these PiPiN devices corresponded

to the low, medium and high doping densities simulated above. A schematic of these devices is shown in Fig 3.19. Reverse bias current-voltage curves were produced for these devices, and are shown in Figure 3.20. The reverse bias characteristics of device 11-141 with medium charge layer doping were measured twice, once immediately after fabrication and again in the main characterisation run. The reverse bias characteristics of this device changed between these two measurements. None of the devices showed any breakdown up to 40 V, and punchthrough was only observed in device 11-141 during the first measurement, at a voltage of around 10 V. This punchthrough voltage correlates most closely to the simulation for low doping in the charge layer with high background n -type doping in the intrinsic germanium (simulation 10), that predicted a punchthrough voltage of 13 V. This indicates the possibility that the target doping concentrations in these devices have not been met, although the lack of replication of this IV curve for the second measurement on 11-141 suggests that the devices are not stable. It also is predicted from the simulation results that breakdown would be measured in the device with high charge layer doping (Device 11-167, simulations 1-4) at a voltage of below 40 V—the absence of measured breakdown in this device re-enforces the conclusion that the absorption layer has a background concentration of n -type dopants, possibly opposing the high electric field across the multiplication layer. Overall, these devices were not stable, and this limits any further conclusive comparison of device performance with simulation results.

Two further SPAD devices were fabricated and characterised in 2012, with low and medium charge layer doping in the charge layer — $1 \times 10^{17} \text{cm}^{-2}$ and $2 \times 10^{17} \text{cm}^{-2}$, respectively. Schematic diagrams of these devices are shown in Figure 3.21. Reverse-bias IV characterisation for these devices is shown in Figure 3.22. The curves in Figure 3.21 show a distinct breakdown, but at significantly lower voltage than would be expected from the simulations. Device 12-026 with a target charge layer doping density of $1 \times 10^{17} \text{cm}^{-2}$ demonstrated breakdown at around 22 V and device 12-027 with a target charge layer doping density of $2 \times 10^{17} \text{cm}^{-2}$ at around 27 V. From the target

design doping profiles and the earlier simulations, it would be expected that breakdown be observed in these devices at a minimum of 52 V for low doping in the charge layer (Simulations 11 and 12, with low background doping in the absorption layer) and a minimum of 40 V for a high doping density in the charge layer (Simulation 5, with high p -type background doping in the absorption layer). In addition, the higher of these two measured breakdown voltages corresponds to the higher charge layer doping density — the opposite effect to that observed in the simulations. No punchthrough was observed in either of these devices, though the simulations predicted punchthrough at a maximum voltage of 18 V for the devices with a low charge layer doping density. The simulations for the medium charge layer doping density predicted punchthrough at voltages higher than the measured breakdown voltage here.

This lack of correlation between simulated IV profiles and experimental measurements suggests that target doping profiles may not have been obtained during device fabrication, and in particular, from the observed breakdown voltages in these devices, that the charge layer doping density was significantly higher than the target density — the earlier simulations show lower breakdown voltages for high charge layer doping. In addition, a higher-than-expected density of p -type doping in the absorber layer would suppress the emergence of punchthrough before breakdown in the device with a low charge layer doping: Simulation 9 shows that the highest punchthrough voltage was obtained with low charge layer doping for high p -type doping, and Simulations 1-8 show, in general, the increase in voltage for, or complete loss of punchthrough (i.e. full depletion of the absorber layer) with increasing charge layer doping density. It is hence likely, from examining the phenomena identified from the simulations, that target doping profiles were not present in the fabricated devices. In addition to the effects on punchthrough and breakdown voltages, this has knock-on effects on the electric field profiles and, hence, detection efficiency and SPAD performance.

A further generation of SPAD devices was fabricated in 2013, and those with dop-

ing profiles expected to produce working SPAD devices are shown schematically in Figure 3.23. Secondary ion mass spectrometry (SIMS) measurements were conducted on these devices and those from the second set discussed above (shown in Figure 3.21, to establish the doping profiles and investigate any deviation from target doping density values. A number of conclusions were drawn from the SIMS analysis results:

- There was a low concentration of n -type (Arsenic) dopant in the absorber layer of the second set of SPAD devices — $1 \times 10^{15} \text{cm}^{-2}$.
- There was significant diffusion of Boron (p -type dopant) from the top contact into the Ge absorber layer in the third set of devices, with variable concentrations of between $1 \times 10^{16} \text{cm}^{-2}$ and $1 \times 10^{17} \text{cm}^{-2}$. It is reasonable to assume, as they were fabricated at similar growth temperatures, that this diffusion also occurred in the second set of devices.
- The overall doping in the absorber layer in these devices was hence high, and of p -type.
- The doping profile in the absorber layers was not uniform, and varied throughout this layer through uncontrolled dopant diffusion.
- Significant diffusion of n -type dopants (Phosphorus) was also observed from the substrate into the silicon multiplication layer, further affecting the segregation of the electric field across the absorption and multiplication regions, and limiting the depletion of the absorber layer.

The presence of a high concentration of p -type dopants in the absorber and diffusion of n -type dopants into the multiplication layer would indicate that the devices fabricated were not fully depleted, and hence could not demonstrate efficient SPAD operation. From the IV results from Simulation 1, it can be seen that such a profile is likely to lead to low breakdown voltages prior to full depletion of the device, and poor SPAD operation.

Although a total of six batches of SPAD devices were fabricated and processed across

four institutions within the SPAD project, characterisation results varied significantly between devices, fabrication of devices to the target dopant profiles provided by the simulations proved difficult, and significant dopant diffusion from the substrate and the top contact hindered the development of a working SiGe SPAD with good single-photon detection efficiency. Although a detection efficiency of 4 % was eventually achieved in Geiger mode using the second set of devices (comparable to the best results for Ge-only SPAD devices), this required measurement at 77K, and could only be obtained at a wavelength of 1310nm; efficiencies at the target wavelength of 1550nm were over a factor of ten lower, due to reduced absorption in this range at low temperature. Very high dark count rates of the order of 10^6 to $10^7 s^{-1}$ were also observed [40]. Overall, good performance in a SPAD device is critically dependent on a number of factors – dopant distributions in the operational layers and precise layer thicknesses, and the electrical characteristics arising from these parameters, and although the simulation results can provide good target doping profiles and thicknesses to achieve these characteristics, fabrication of devices to this specification is more challenging.

3.3 Conclusion

A number of doping profiles for a PiPiN SPAD device were simulated, to establish IV curves, field and potential profiles, and carrier concentrations for these devices under a range of doping conditions. The charge layer doping is intentional: doping of this separation feature distributes the electric field across the device in a number of ways depending on the density and type of charge layer doping. In addition, background doping in the ideally-intrinsic absorption layer may affect the characteristics of a SPAD device. It was found that breakdown voltage increases with decreasing charge layer doping, which is to be expected since the a higher charge layer doping density increases the overall proportion of the electric field across the device which occurs across the multiplication layer, and hence decreases the voltage required to achieve breakdown. At high charge layer doping, the type of absorption layer doping (*p*-type or *n*-type) plays

no significant role in the breakdown voltage. At lower charge layer doping densities, p -type absorption layer doping results in a lower breakdown voltage than n -type, since the gradient of the electric field in the absorber determines the size of the electric field in the multiplication layer, and these different doping types produce opposite field gradients in the absorber. This effect becomes more pronounced when the absorption layer doping density is higher. ‘Punchthrough’ — the effect where the potential barrier at the interface is removed for higher bias voltages — was observed for devices with low charge layer doping, with medium charge layer doping but low absorption layer doping, and with medium charge layer doping and high, n -type absorption layer doping. This is, as expected, in line with those simulations where a higher breakdown voltage is observed.

Potential profiles show a potential well in the multiplication layer, which is removed after punchthrough. Electron and hole concentrations showed the effect of breakdown in most simulations, where a significant increase in electron concentration was observed after breakdown voltage. For some simulations which included the lowest charge layer doping density simulated, a drop in electron concentration was observed in the multiplication layer close to the charge layer, indicating that hole-initiated impact ionisation may have occurred in the absorption layer in these simulations. This is consistent with the high fields shown in the multiplication layer for these simulations, where the electric field segregation between the absorption and multiplication layer was not ideal for SPAD operation.

Effective operation of a SACM-SPAD requires control of fields between the absorption and multiplication layers. From these simulations, it can be seen that low absorption-layer fields and high multiplication layer fields can best be achieved with high charge layer doping. However, the effect of punchthrough should also be considered. If punchthrough, as defined from the application to SPADs, is the voltage at which carriers are depleted, then a device in which punchthrough is achieved is necessary for effective operation without dark counts or noise. From these simulations, this is achieved only when the

charge layer doping is at a medium ($2 \times 10^{17} \text{cm}^{-2}$) density and fields in the germanium absorber are sufficient to deplete the region. In this case, a low absorption layer doping density ($1 \times 10^{15} \text{cm}^{-2}$) of *n*-type is desirable to limit dark counts due to excess fields in the germanium, while ensuring full depletion of the absorber before breakdown.

Experimental fabrication of Si/Ge SPADs proved difficult throughout the project, and experimental measurements frequently varied significantly from the simulated characteristics, indicating that target dopant profiles had not been achieved. SIMS analysis of samples from the project confirm this conclusion, indicating that dopant diffusion from the substrate and bottom contact and wide spatial variations in dopant distributions within each operational layer severely inhibited the performance of the Si/Ge SPAD fabricated, leading to low detection efficiencies and high dark count rates.

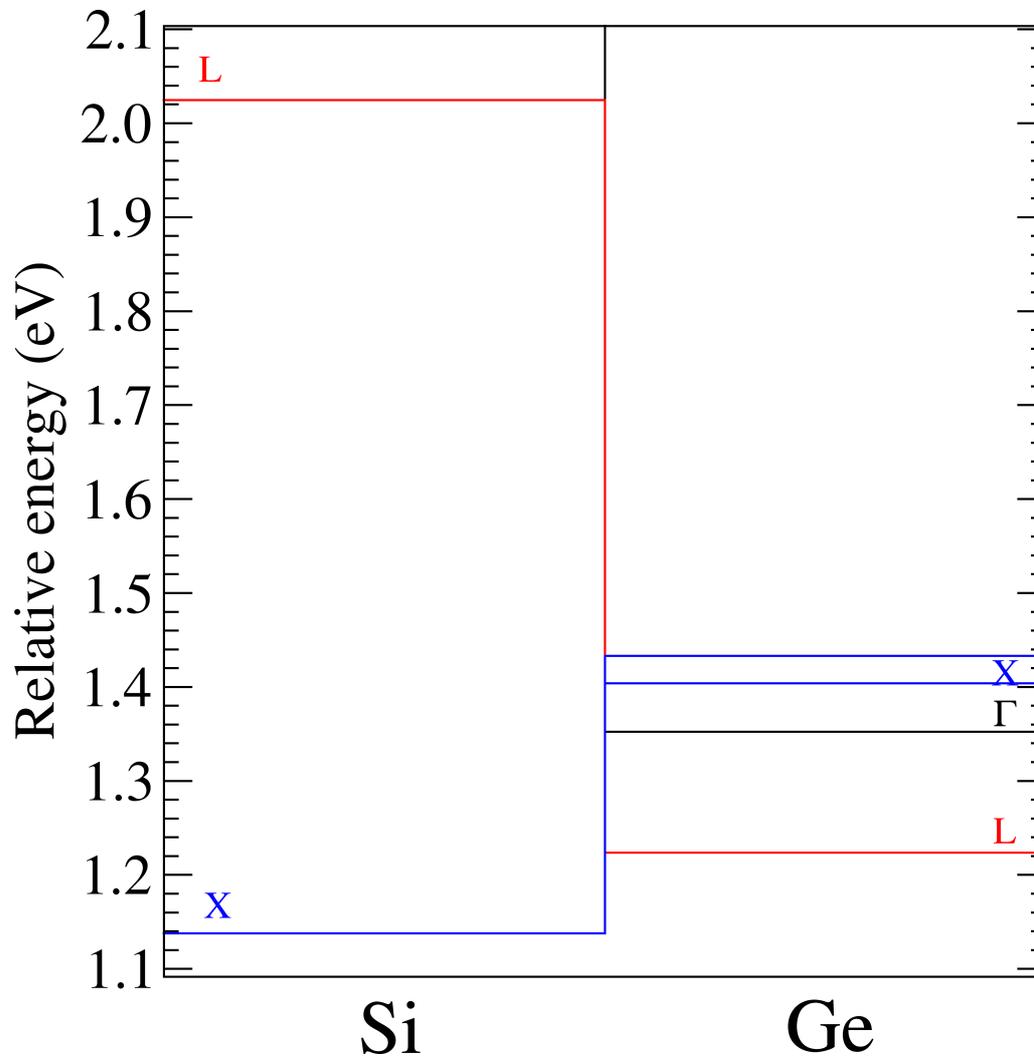


Figure 3.1: Schematic of the band offsets between the silicon and germanium layer in a SiGe SPAD in the flat-band case, highlighting the potential barriers for electrons traversing from Ge into Si in the L or Γ minima [106]

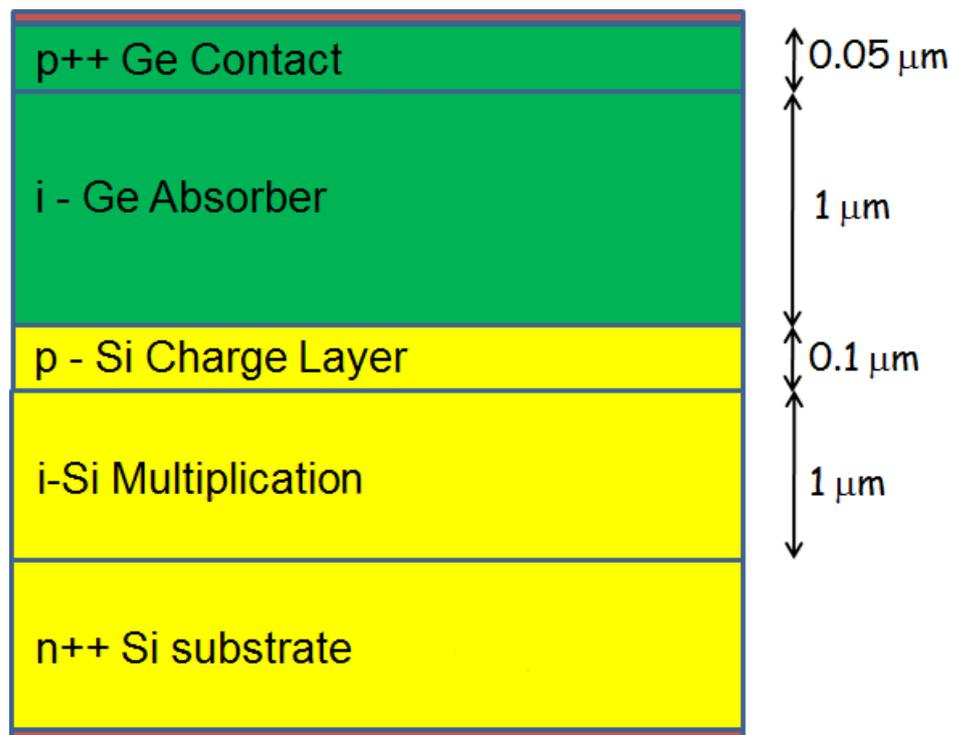
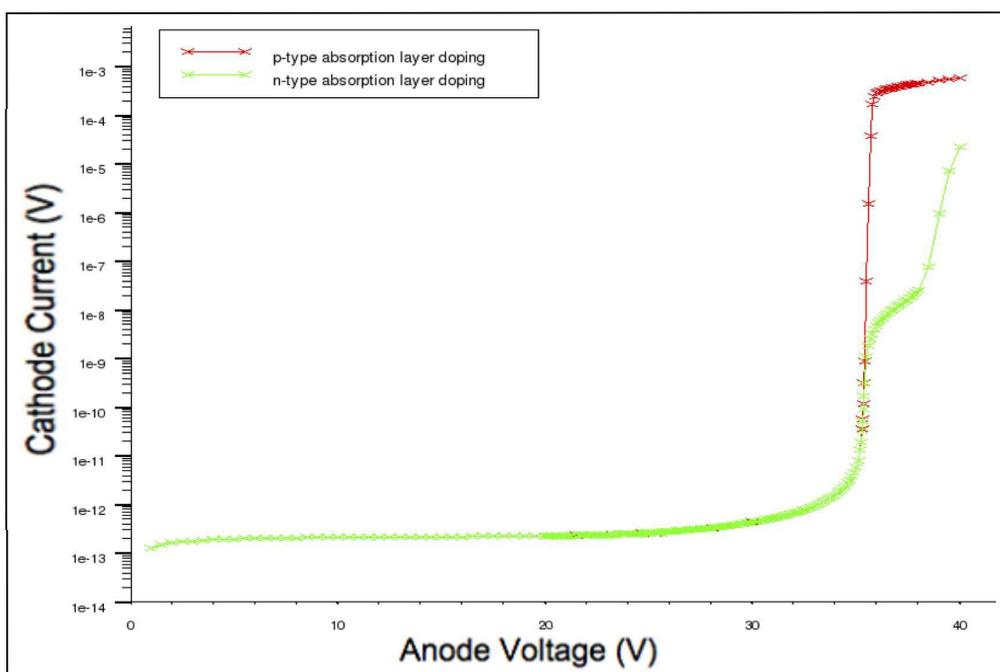
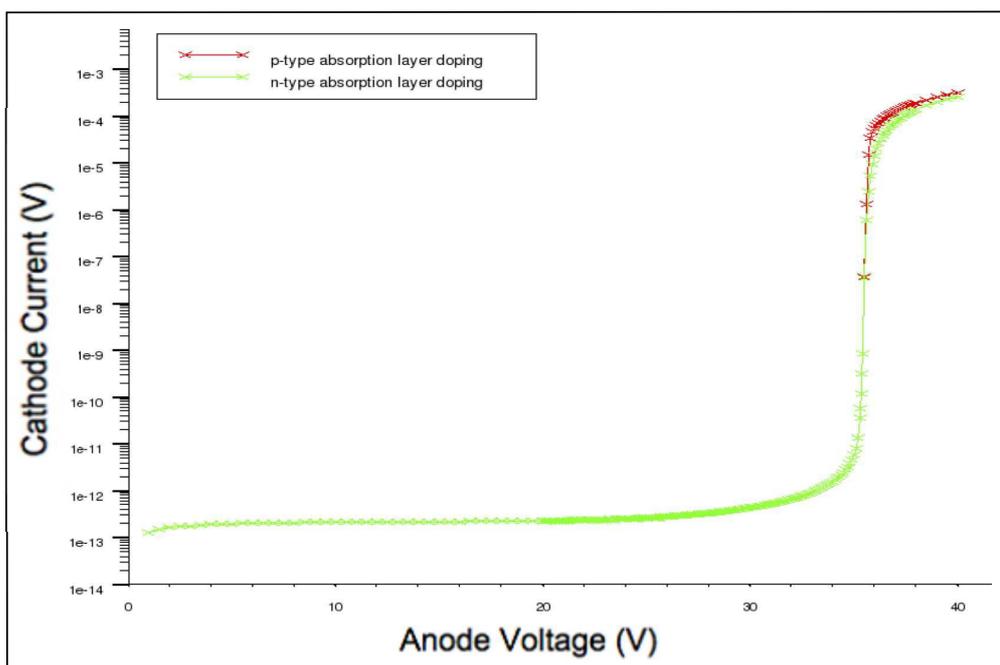


Figure 3.2: PiPiN SPAD structure

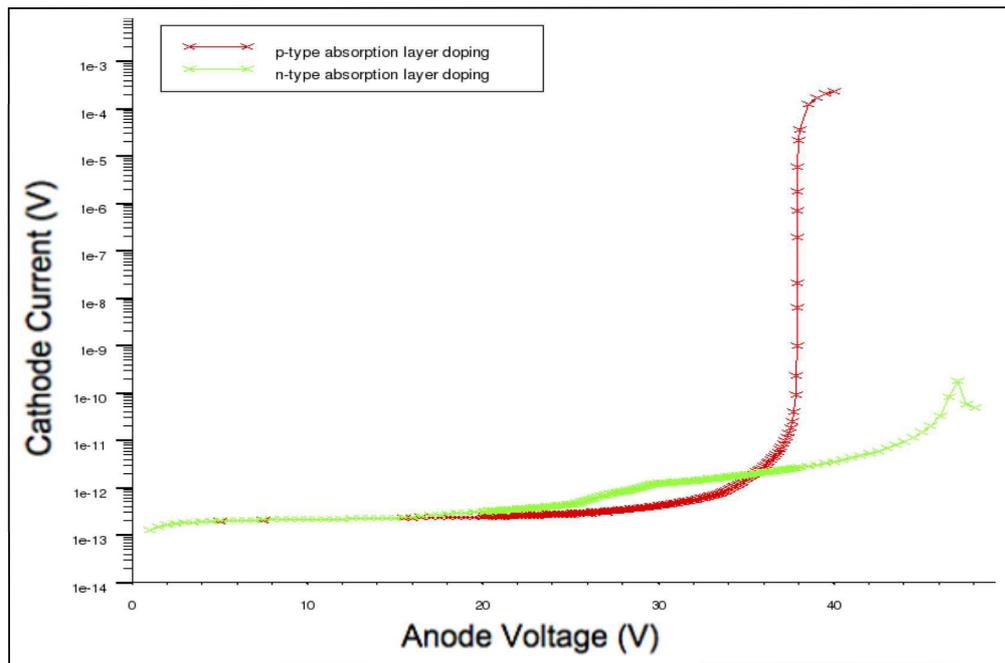


(a) IV curve for high ($5 \times 10^{17} \text{cm}^{-2}$) charge and high ($1 \times 10^{16} \text{cm}^{-2}$) absorption layer doping – simulations 1 and 2

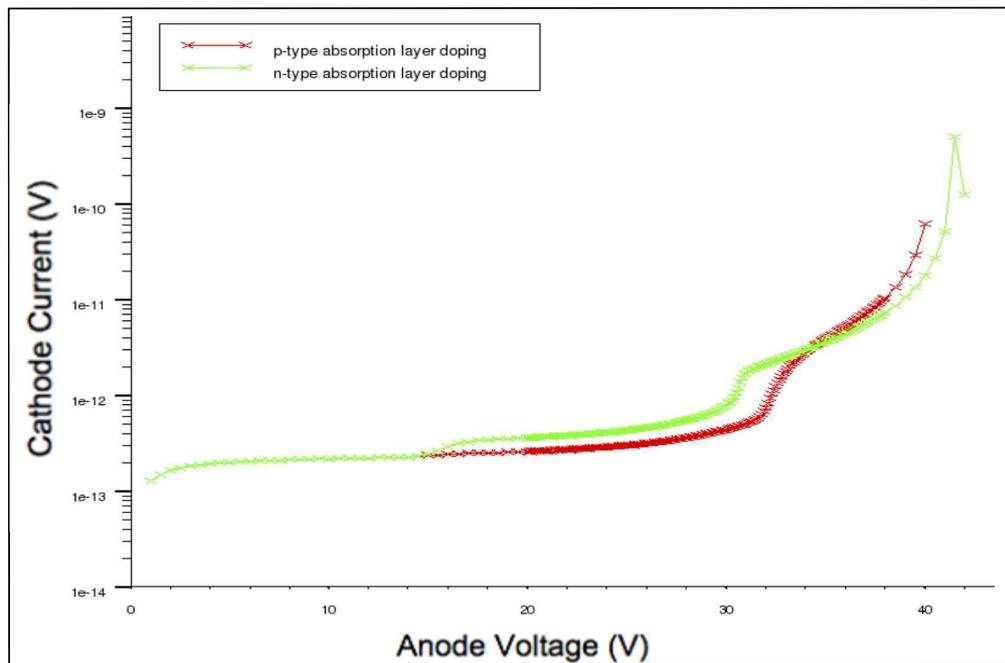


(b) IV curve for high ($5 \times 10^{17} \text{cm}^{-2}$) charge layer doping and low ($1 \times 10^{15} \text{cm}^{-2}$) absorption layer doping – simulations 3 and 4

Figure 3.3: IV curves for simulations including high charge layer doping

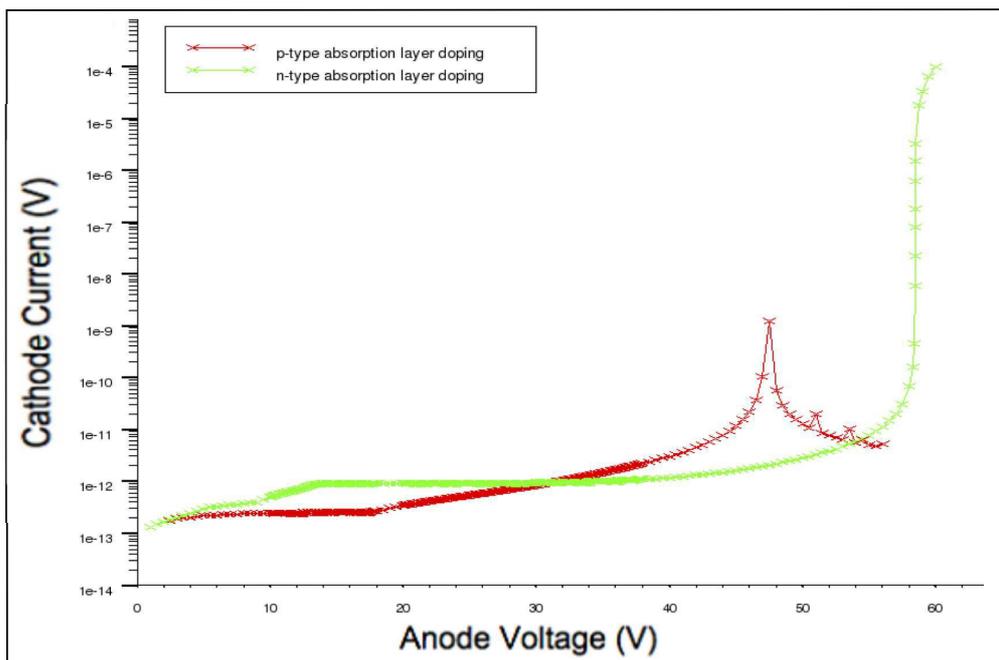


(a) IV curve for medium ($2 \times 10^{17} \text{cm}^{-2}$) charge layer doping and high ($1 \times 10^{16} \text{cm}^{-2}$) absorption layer doping – simulations 5 and 6

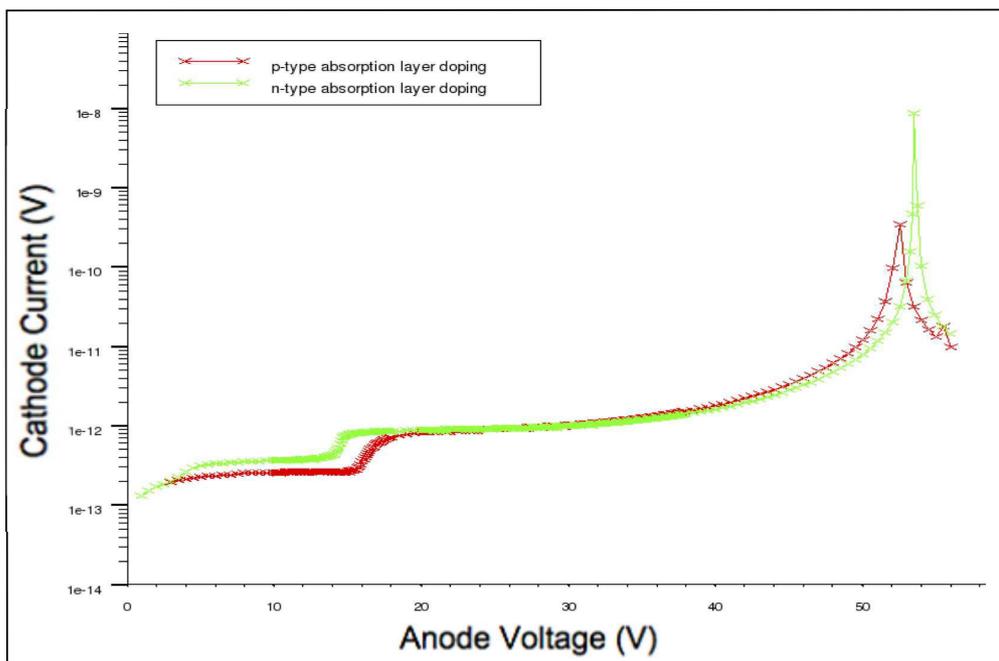


(b) IV curve for medium ($2 \times 10^{17} \text{cm}^{-2}$) charge layer doping and low ($1 \times 10^{15} \text{cm}^{-2}$) absorption layer doping – simulations 7 and 8

Figure 3.4: IV curves for simulations including medium charge layer doping



(a) IV curve for low ($1 \times 10^{17} \text{cm}^{-2}$) charge layer doping and high ($1 \times 10^{16} \text{cm}^{-2}$) absorption layer doping – simulations 9 and 10



(b) IV curve for low ($1 \times 10^{17} \text{cm}^{-2}$) charge layer doping and low ($1 \times 10^{15} \text{cm}^{-2}$) absorption layer doping – simulations 11 and 12

Figure 3.5: IV Curves for simulations including low charge layer doping

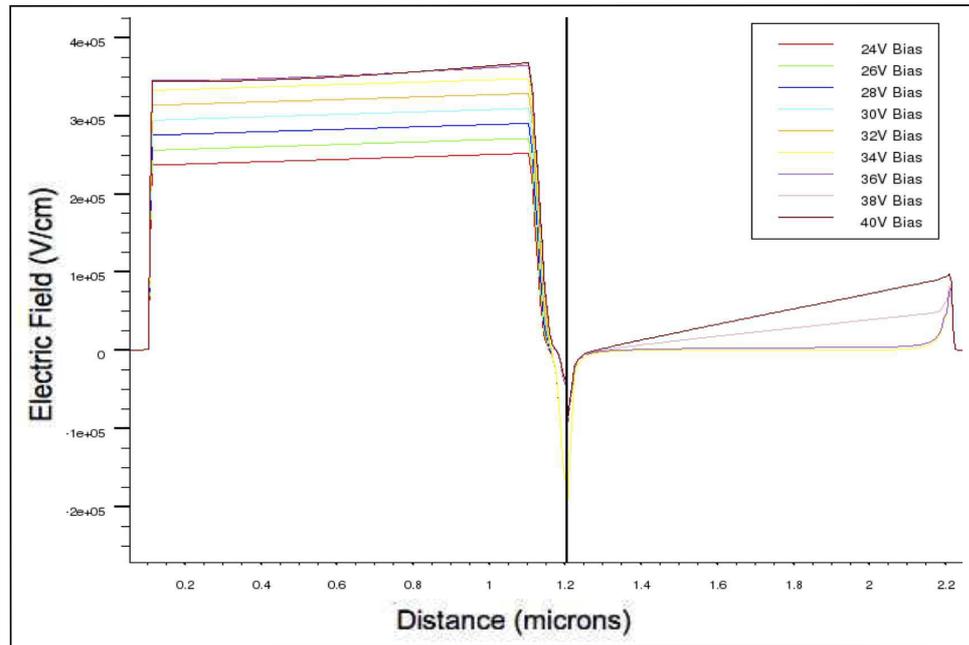


Figure 3.6: Electric field profile for SPAD device with high charge layer doping and high p -type absorption layer doping — simulation 1

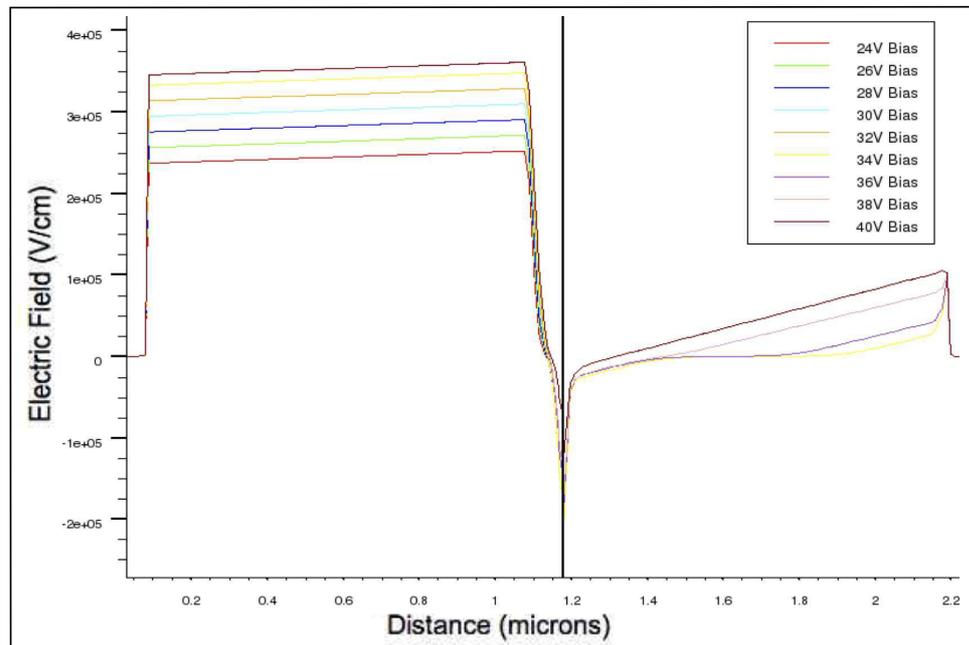


Figure 3.7: Electric field profile for SPAD device with high charge layer doping and high n -type absorption layer doping — simulation 2

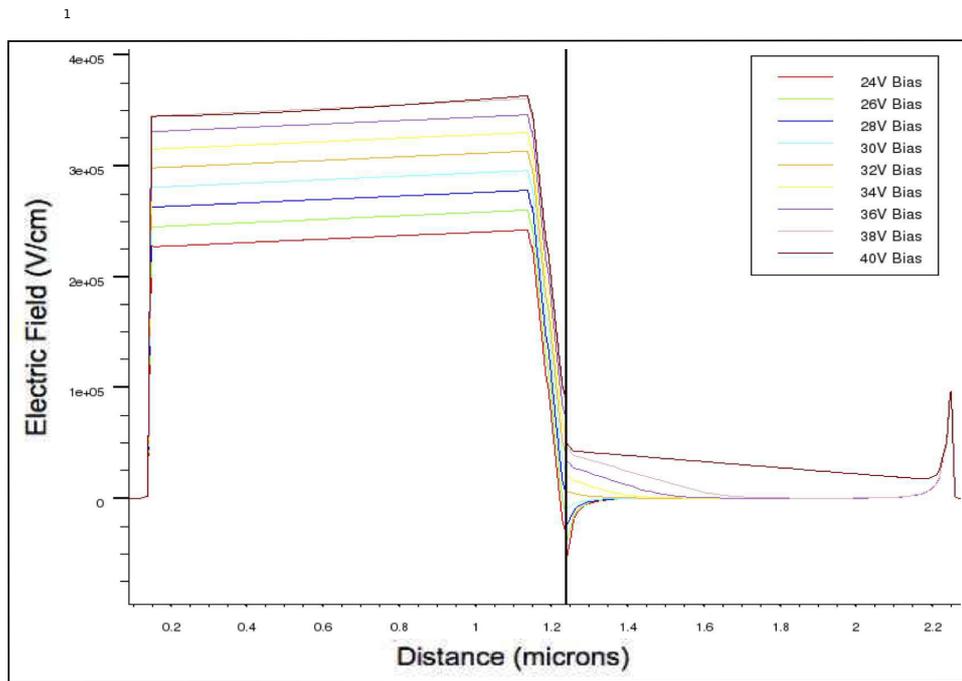


Figure 3.8: Electric field profile for SPAD device with medium charge layer doping and high *p*-type absorption layer doping — simulation 5

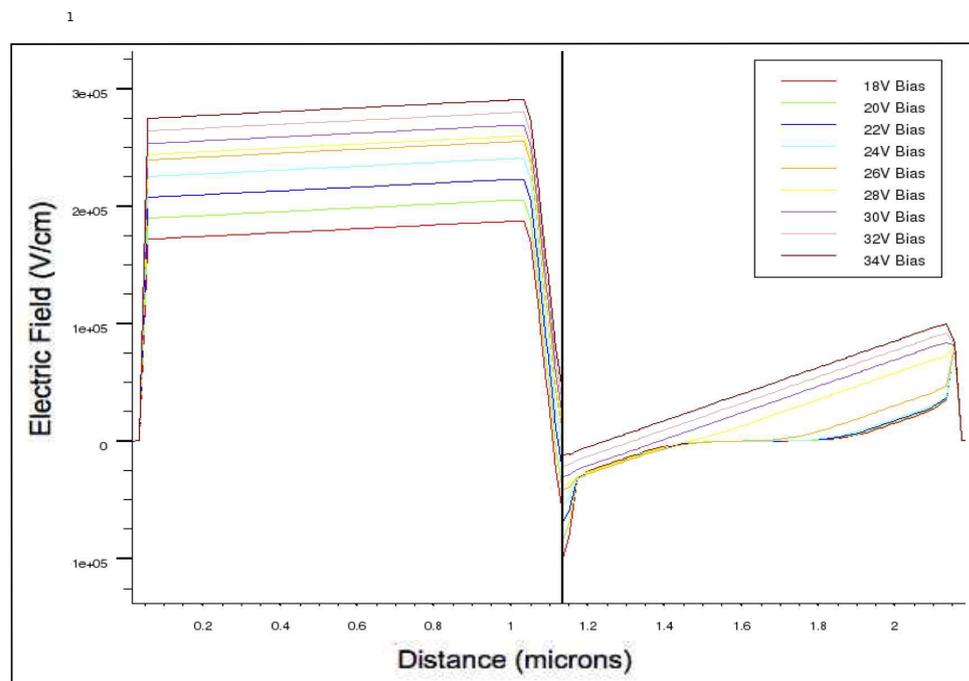


Figure 3.9: Electric field profile for SPAD device with medium charge layer doping and high n -type absorption layer doping — simulation 6

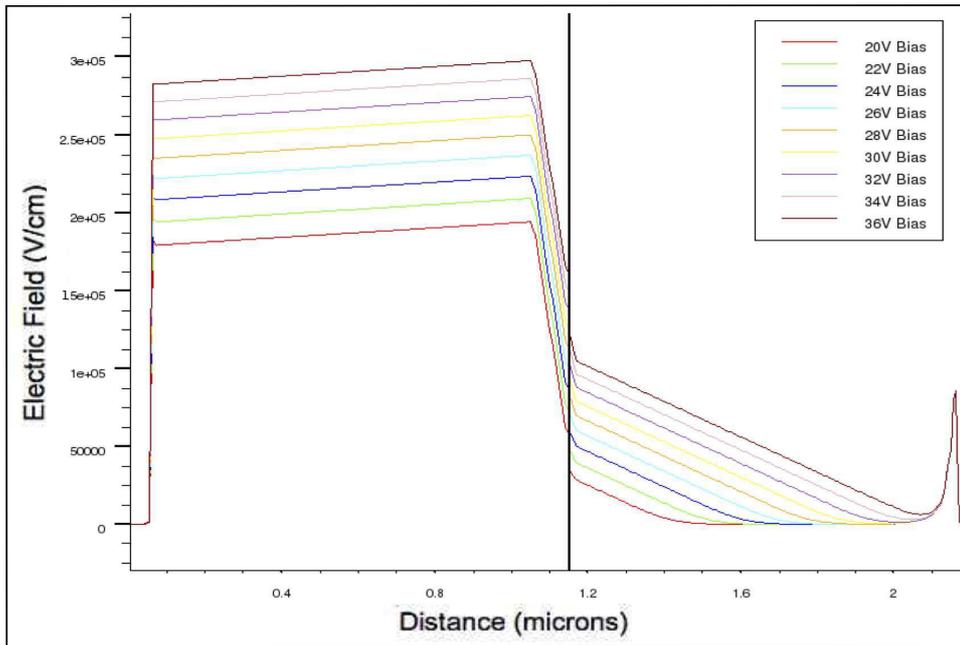


Figure 3.10: Electric field profile for SPAD device with low charge layer doping and high p -type absorption layer doping — simulation 9

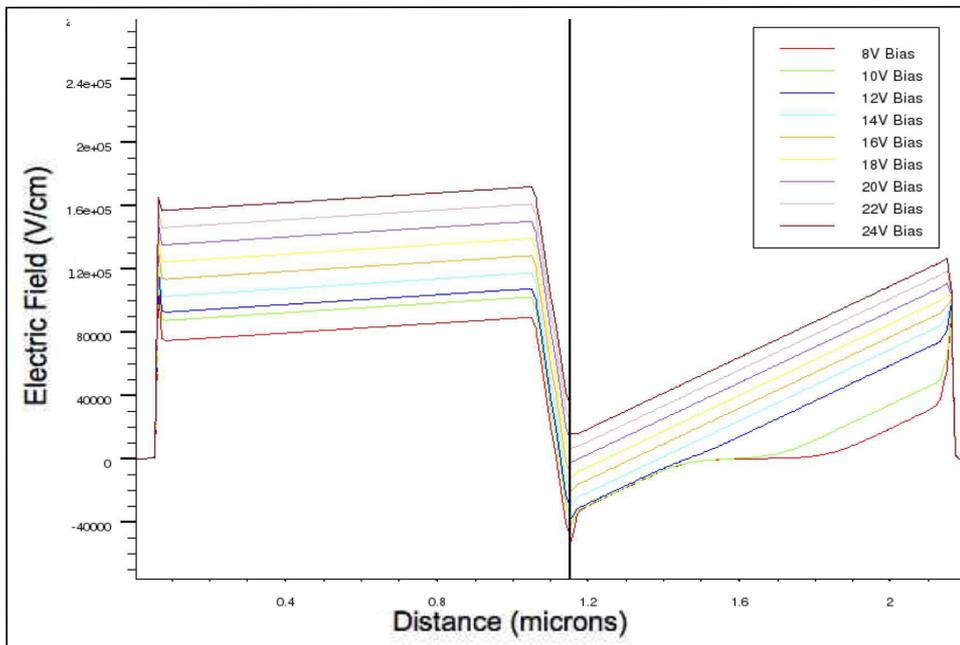


Figure 3.11: Electric field profile for SPAD device with low charge layer doping and high n -type absorption layer doping — simulation 10

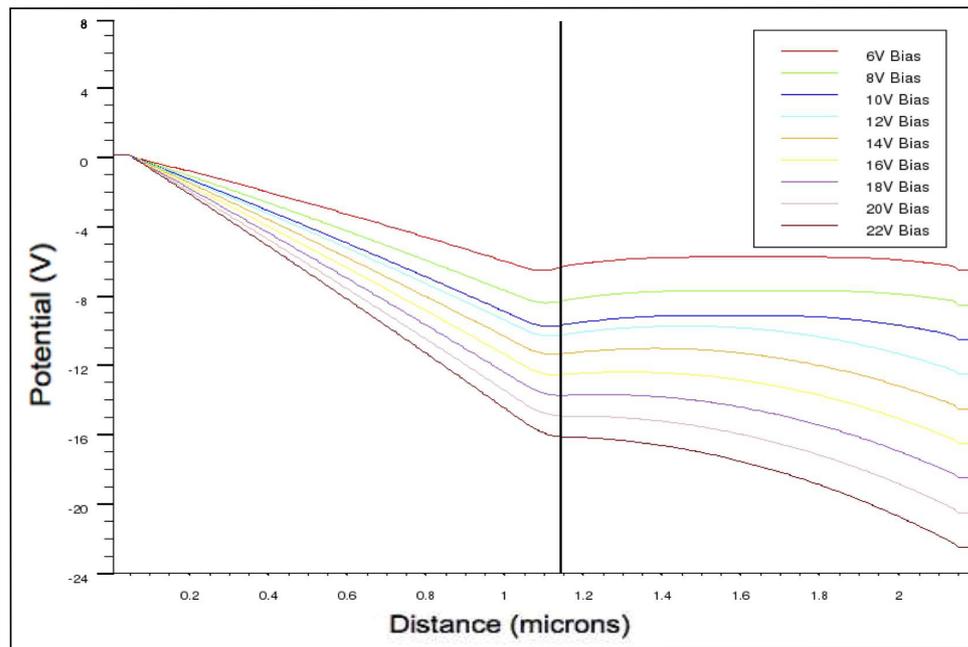


Figure 3.12: Potential profile for SPAD device with low charge layer doping and high n -type absorption layer doping, showing the removal of the potential barrier at the interface at the punchthrough voltage (13V) — (simulation 10)

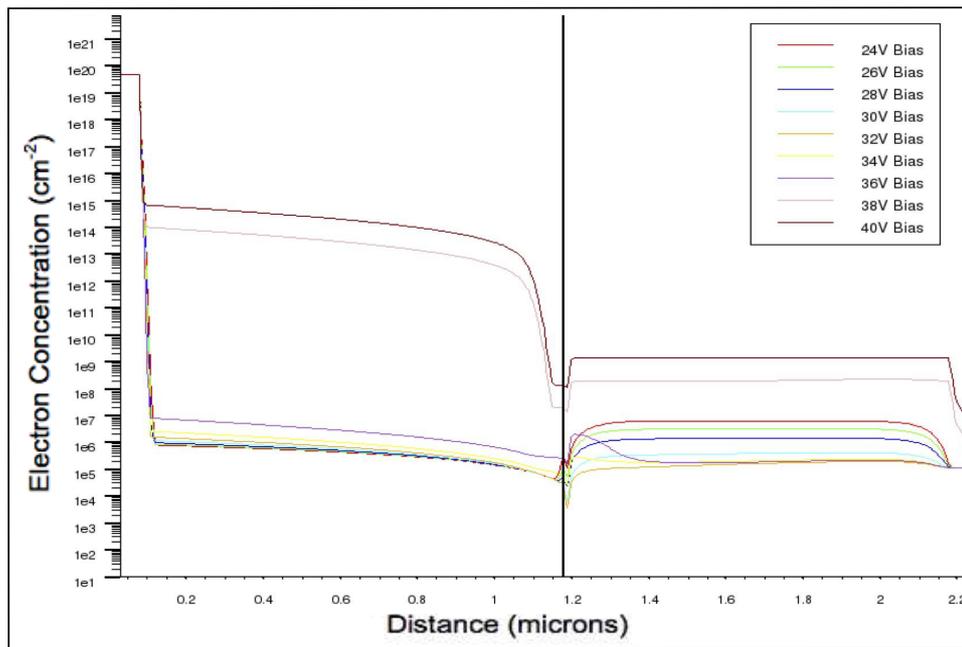


Figure 3.13: Electron concentration profile for SPAD device with medium charge layer doping and high *p*-type absorption layer doping, showing an increase in electron concentration at the breakdown voltage — simulation 5

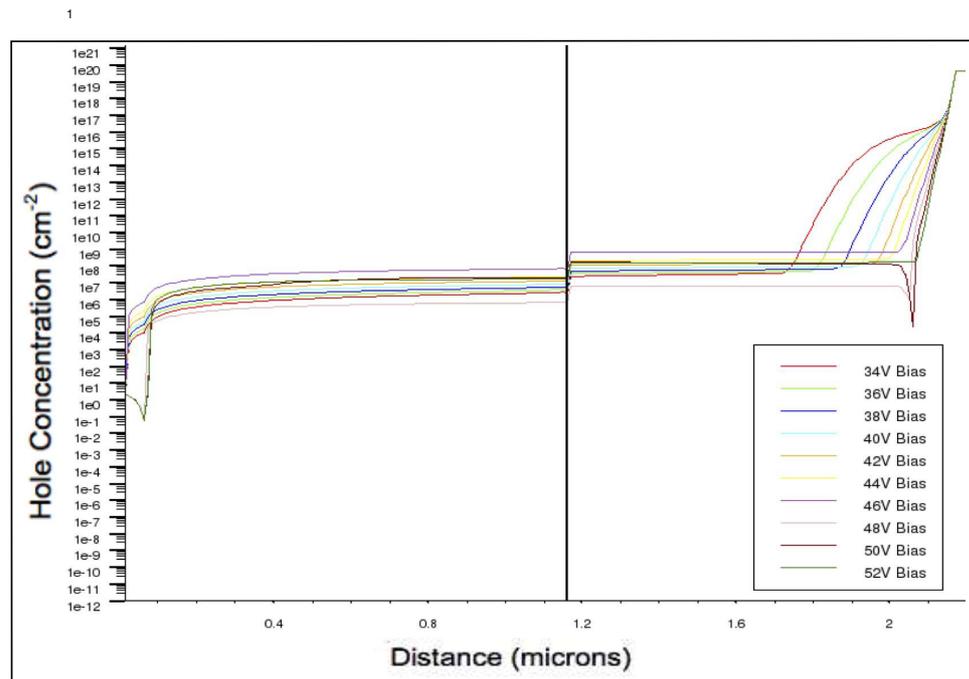


Figure 3.14: Hole concentration profile for SPAD device with low charge layer doping and high p -type absorption layer doping, for voltages in the range 34-52 V - breakdown voltage is 48 V — simulation 9

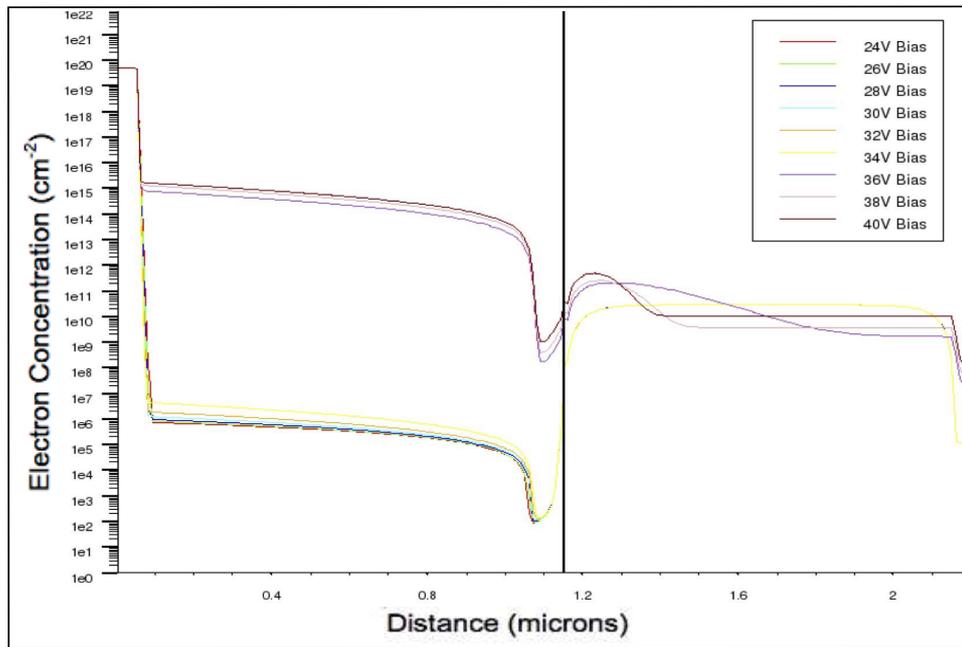


Figure 3.15: Electron concentration profile for SPAD device with high charge layer doping and high *p*-type absorption layer doping — simulation 1

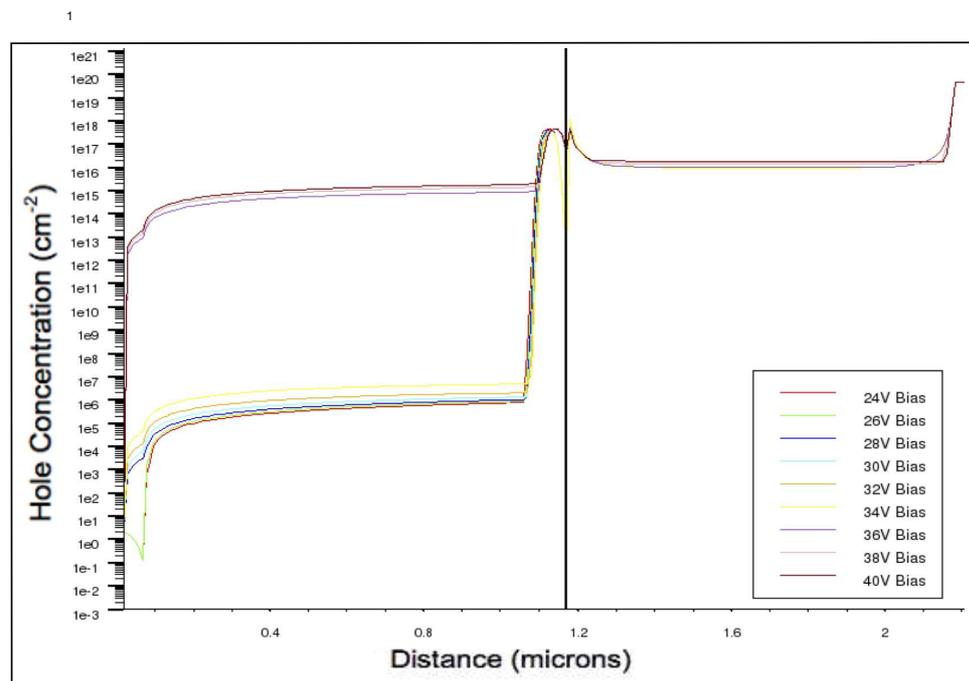


Figure 3.16: Hole concentration profile for SPAD device with high charge layer doping and high *p*-type absorption layer doping — simulation 1

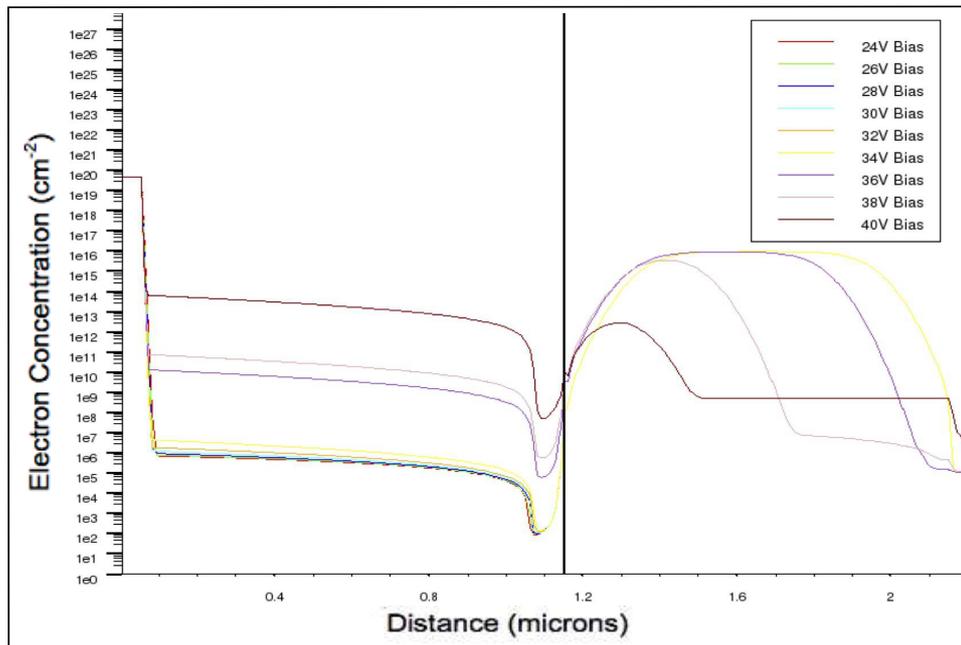


Figure 3.17: Electron concentration profile for SPAD device with high charge layer doping and high *n*-type absorption layer doping — simulation 2

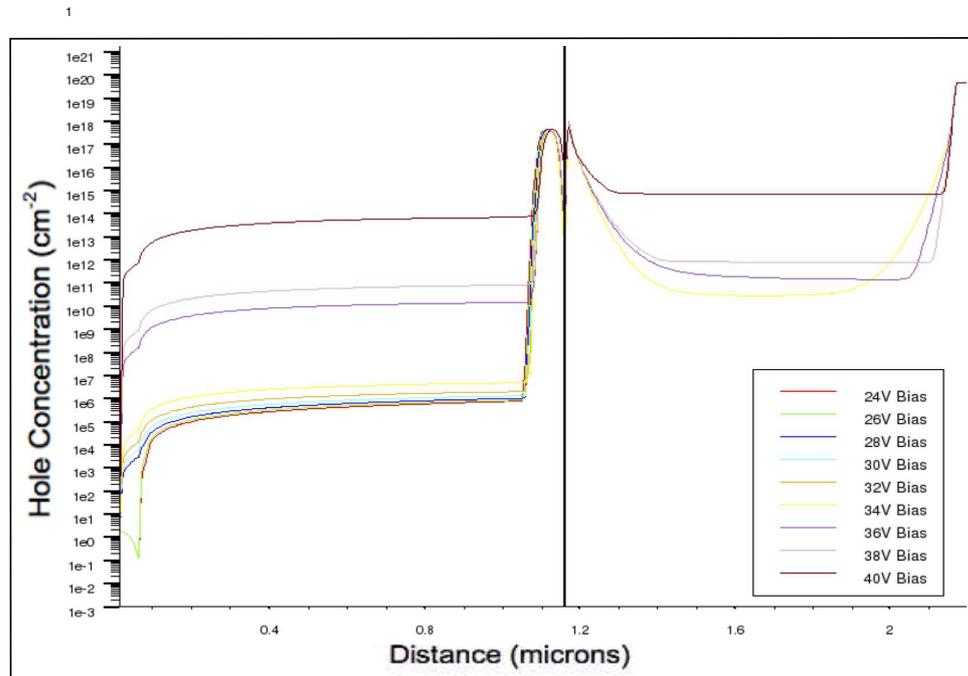


Figure 3.18: Hole concentration profile for SPAD device with high charge layer doping and high *n*-type absorption layer doping — simulation 2

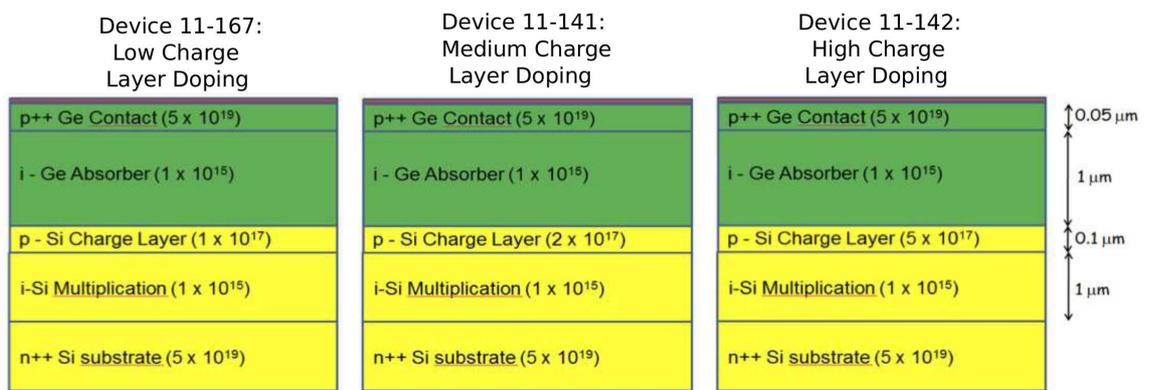


Figure 3.19: Schematic of the 1st set of SPAD devices fabricated and characterised, in November 2011

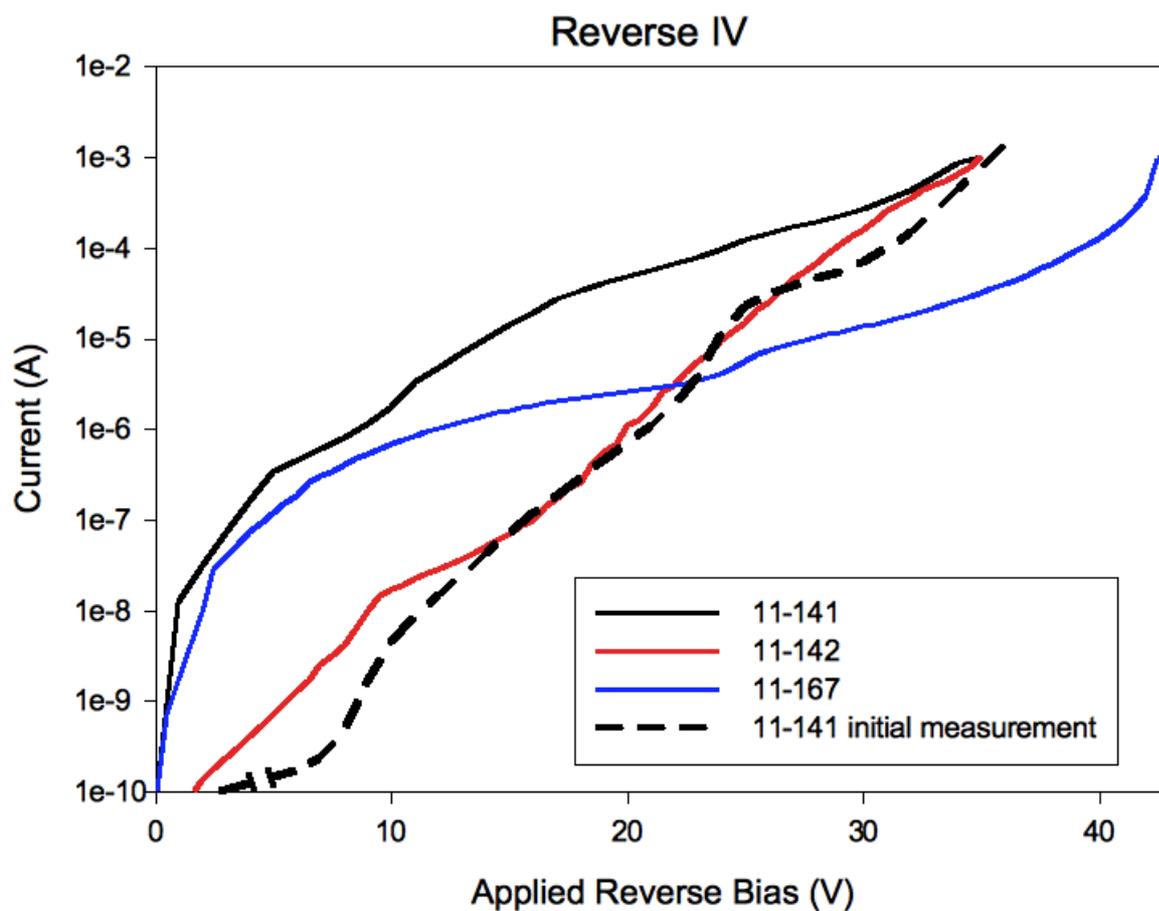


Figure 3.20: Reverse biased current–voltage (IV) characteristics for the 1st set of SPAD devices fabricated and characterised in November 2011

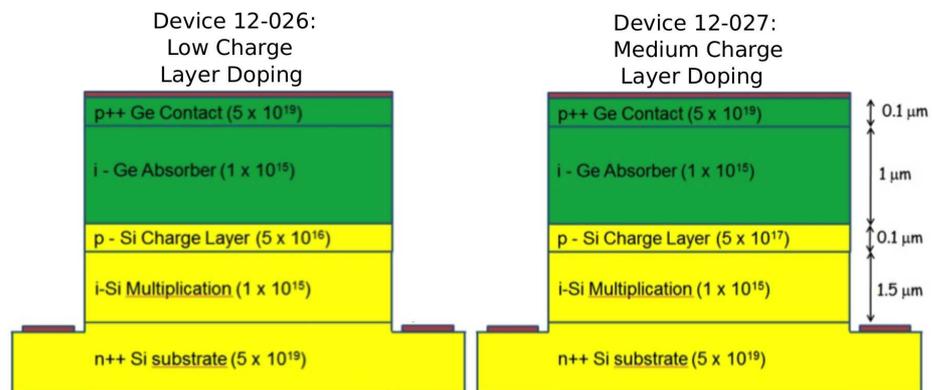


Figure 3.21: Schematic of the second set of SPAD devices fabricated and characterised in 2012

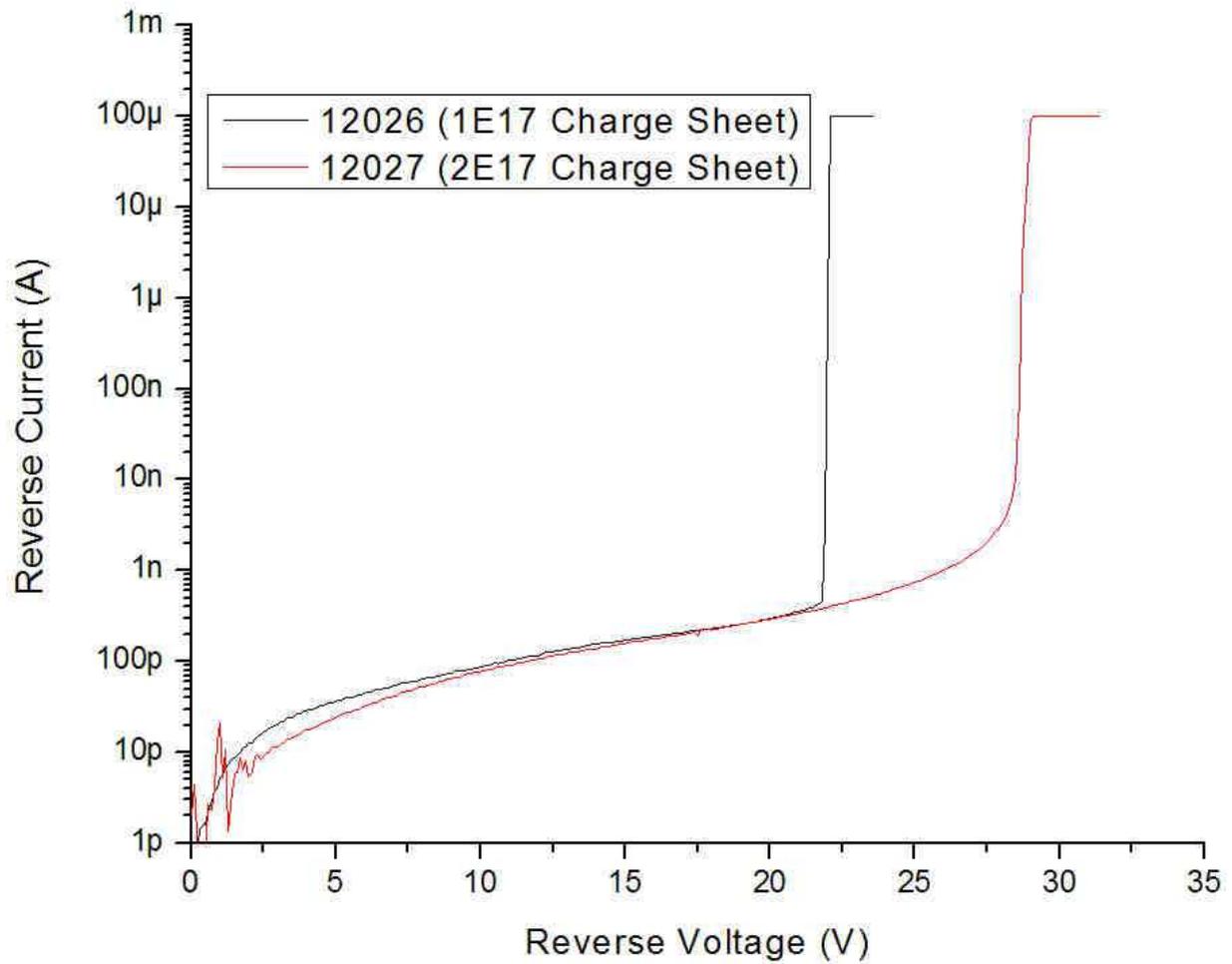


Figure 3.22: Reverse biased current–voltage (IV) characteristics for the second set of SPAD devices fabricated and characterised in 2012. Device identification and doping profiles correspond to the schematic structures shown in 3.21

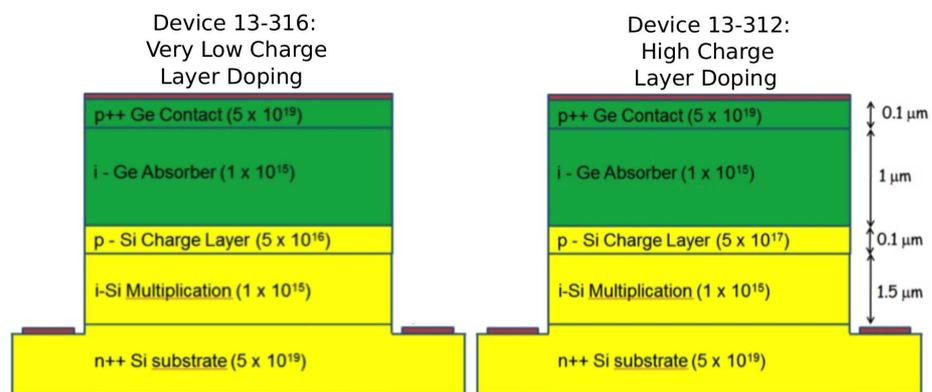


Figure 3.23: Schematics of SPAD devices fabricated in 2013, with target doping profiles

Chapter 4

Band-to-Band Tunnelling in Germanium

At the high bias voltages associated with SPAD operation, there is a significant possibility of band-to-band tunnelling, wherein an electron can pass through the forbidden band gap under the influence of a high electric field, and contribute to noise currents in the absence of an incoming photon [1, 32]. Whereas tunnelling currents in silicon have been studied previously [108, 109] and a number of studies have reported tunnelling currents under a range of conditions in Ge and Si based MOSFETs and TFETs,[110–113], there are limited studies on the tunnelling currents generated in bulk germanium [114]. Although in the ideal SPAD device, the field across the germanium layer is sufficiently low that noise currents through tunnelling and other mechanisms are suppressed, an understanding of the contribution of the tunnelling current to noise current in SiGe SPADs is useful for development of performance. In this chapter, the band-to-band tunnelling contributions in bulk germanium are studied under a number of conditions, using the empirical pseudopotential bandstructure calculation method described in Chapter 2 and a band-matching technique.

4.1 Band-to-Band Tunnelling Model

The band-to-band tunnelling model used combines calculation of wavefunctions using an empirical pseudopotential method, and band-matching to identify states which may propagate through the bandgap to the valence band.

Calculation of band-to-band tunnelling currents in a single semiconductor material requires consideration of the initial (valence band) states and the final (conduction band) states, and the potential barrier between them, in this case the ‘forbidden’ band gap. However, for consideration of a bulk semiconductor under the influence of an electric field, a number of wave functions must be calculated, under each potential condition. Though the potential is, in practice, continuous, the simulation model considers a discretized number of equally-spaced sections of the material and of the band gap, with a step-size determined by considering the largest step size at which the results of the calculation converge to an acceptable limit. Each step is considered to be at some potential offset relative to the previous step, and the Schrodinger equation within this step is solved with in-plane wavefunctions $e^{ik_{||}c_{||}}$ assumed to have wholly real $k_{||}$, and wavefunctions $e^{ik_{||}c_{||}}$ in the z -direction permitted to have complex $k_{||}$; the potential term is given by the constant potential within this step. The appropriate form of the wavefunctions is given by Bloch’s theorem, discussed later, and the Schrodinger equation must include a potential with the periodicity of the lattice, given by the atomic spacing of the crystalline material [115].

For the one-dimensional case, in the z -direction, states propagate in the bands, or exponentially decay within the barrier in both the forward and backward directions, with coefficients in these directions given by a and b , respectively. Matching between states in the n th and $n+1$ th step is given by an interfacial matrix, I , which matches states at the interface between steps, and is given by [116]:

$$\begin{bmatrix} \mathbf{a}_{n+1} \\ \mathbf{b}_{n+1} \end{bmatrix} = I(n+1) \begin{bmatrix} \mathbf{a}_n \\ \mathbf{b}_n \end{bmatrix} \quad (4.1)$$

Where, here, \mathbf{a}_n and \mathbf{b}_n represent the vector of coefficients of all incoming and outgoing states of the n th layer, respectively. For material under a further external potential, incoming and outgoing wavefunctions must be coupled at the interface, which may be undertaken using a transfer matrix, $T(0, N)$ for N total steps, which couples all states on the left and all states on the right:

$$\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix} = T(0, N) \begin{bmatrix} \mathbf{a}_N \\ \mathbf{b}_N \end{bmatrix} \quad (4.2)$$

However, this formulation may lead to numerical instability, and an alternative formulation, using a scattering matrix $S(0, n)$ couples incoming states to outgoing states from the 0th layer to the n th layer successively:

$$\begin{bmatrix} \mathbf{a}_n \\ \mathbf{b}_n \end{bmatrix} = S(0, n) \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix} \quad (4.3)$$

The scattering matrix after adding each successive layer n may be combined with the interface matrix, to yield the total scattering matrix $S(0, N)$ across the system of N layers. The total transmission and reflection coefficients may then be calculated from the coefficients of the incoming and outgoing states for the whole system using the scattering matrix from the 0th to the N th layer [116], by setting an arbitrary value to the initial incoming state \mathbf{a}_0 and calculating the right propagating states \mathbf{a}_N , which are transmitted. Similarly, the left propagating state \mathbf{b}_N is assumed to be zero, from here \mathbf{b}_0 gives the total reflection coefficient:

$$\begin{bmatrix} \mathbf{a}_N \\ \mathbf{b}_0 \end{bmatrix} = S(0, N) \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_N \end{bmatrix} \quad (4.4)$$

Following calculation of one-dimensional transmission coefficients across all energies in the bandstructure, and across a grid of all possible values of the in-plane wave vector in the remaining two dimensions, the total tunnelling current may be calculated using [117]:

$$j = \frac{(-e)}{(2\pi)^3 \hbar} \int_{BZ_{\parallel}} d\mathbf{k}_{\parallel} \int_{-\infty}^{+\infty} dE \times \sum_{k_j^-} \sum_{k_i^+} T_{(L, k_j^- \rightarrow R, k_i^+)}(k_{\parallel}, E) [f_R(E) - f_L(E)] \quad (4.5)$$

where the outer integral extends over all 2-dimensional wavevectors in the plane perpendicular to the direction of current flow, within the first Brillouin Zone. $k_{||}$ represents the integral across in-plane k-vectors in the x and y directions, E represents the integral across all energies in theory, but is applied to the energy range used in the simulations in practice, T are the transmission coefficients between states on the left and states on the right, and f_L and f_R are the Fermi-Dirac distributions for the left and right states, respectively. The difference between the Fermi-Dirac distributions of these states may be set to 1, since the summation in Equation 4.5 applies only across all propagating states on the left and all propagating states on the right. These states can be assumed to have an occupancy of 1 on the left and 0 on the right for an undoped semiconductor, and are the only current carrying states; all other states may be assumed to have equal occupancies, and hence lead to a value of $[f_R(E) - f_L(E)]$ of zero. Although some difference in Fermi occupancy of the left and right states may exist at temperatures of above

In summary, in this formulation, the calculated, ‘allowed’ states in the conduction and valence bands will take the form of plane waves, and the tunnelling states in the bandgap will be evanescent states decaying in either the forward or reverse direction. Matching of these wavefunctions across steps in the semiconductor structure, using the technique described above [116], leads to calculation of the probability of tunnelling under a given set of conditions, and the associated transmission coefficient for each energy in the bandstructure.

The complex band structure of Ge was calculated using the local empirical pseudopotential method [115]. Although addition of the nonlocal part in real band structure calculations can improve the accuracy of the calculations somewhat, in typical semiconductors, such as Ge, good accuracy can be obtained using a local potential only; in addition, for calculation of complex bandstructures adding nonlocality results in a nonlinear eigenvalue problem (with complex wavevectors in the z -direction), which is much more difficult to solve than the linear problem arising from the use of a local

| Simulation Parameter | Setting |
|----------------------|---|
| k-parallel Range | $0.0/0.0 - \frac{0.2\pi}{a} / \frac{0.2\pi}{a}$ |
| Lattice Constant | 5.658 Å |
| Monolayer Thickness | 2.829 Å |

Table 4.1: Simulation parameters for band-to-band tunnelling calculations on bulk Ge

pseudopotential. The scattering-matrix method discussed above was used to calculate transmission co-efficients [116]. The calculation method used to generate currents from transmission coefficients is described in equation 4.5 and explained in detail in [117]. The spin-orbit interaction was included in the calculation, as this is significant in Ge, and does not add significantly to the computational burden.

A number of simulation parameters were applied to the calculations, including the types, parameters and thicknesses of material layers, the applied reverse bias voltage, the in-plane k-grid point (expressed as fractional coordinates in the x - and y - directions), the energy range over which which transmission coefficients are to be calculated and pseudopotential formfactors. Spin-orbit interaction parameters and the V3, V8 and V11 form factors were obtained from [118], and layer thicknesses are expressed in numbers of monolayers, equivalent to 0.5 of a lattice constant for a cubic cell of germanium - 2.829 Å. The basic parameters used in the simulations are shown in Table 4.1. For simulation of direct band-to-band tunnelling, where no change in momentum occurs, the total bias voltage should exceed the bandgap. Initial high-resolution calculations on bulk germanium with a thickness of 10 monolayers (28.29 Å) and no applied voltage were undertaken to identify the empirical location of the bandgap for bulk germanium. This was identified to be between 9.429eV and 10.550eV at Γ - a gap of 0.713eV. Hence, the applied bias voltage lower limit was set to this bandgap to ensure that band-to-band tunnelling was possible. Simulations were conducted within an energy range up to the valence band edge. Figure 4.1 shows the range of energies selected for simulation schematically. The software required calibration over a number of parameter grids, to

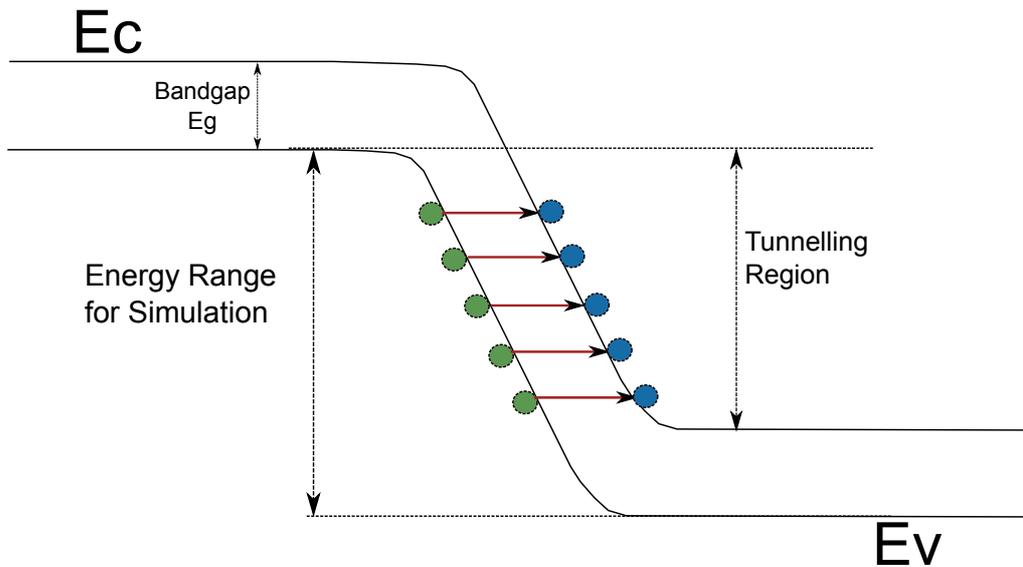


Figure 4.1: Schematic of the energy range selected for band-to-band tunnelling simulations

identify the appropriate ranges and grid sizes for simulation.

4.2 Calibration and Optimisation of Simulation Grids

The calculation requires calibration over a number of parameters: energy grid size, k-grid size and digitising step size, which determines the number of ‘slices’ into which the total thickness of the simulation cell is divided, by defining an energy step over which the applied potential is discretized. Calibration of energy grid size was undertaken by calculating the total tunnelling current for a single in-plane k-grid point (0, 0), for 354 monolayers with a total bias voltage across this thickness of 1V, for a range of energies from 8.5-9.5 eV. The calibration was undertaken for a range of energy divisions from 21 to 201, and the result is shown in Figure 4.2. The calculation was performed for tunnelling current in arbitrary units – physical constants were not included. The calibration converged at an energy division of around 91, where a less than 2% variation between results was observed. An energy grid density of 101 energy points per 1V of bias voltage, rounded to the nearest odd number for each bias voltage used, was chosen for further simulations.

4.2. C

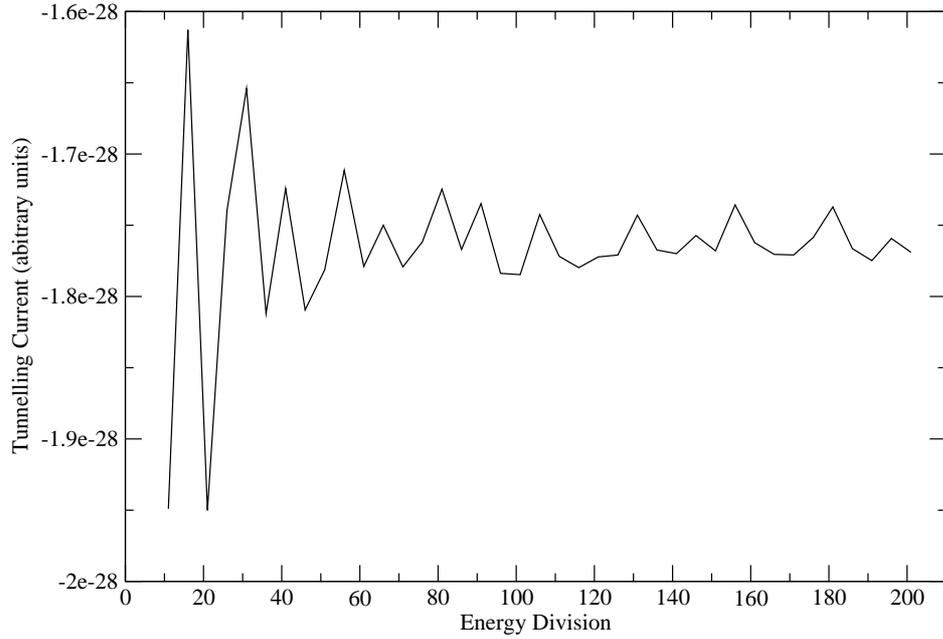


Figure 4.2: Energy grid calibration curve for band-to-band tunnelling simulations

Calibration of k-grid size was undertaken by calculating the total tunnelling current for all in-plane k-grid points between $0.0/0.0$ and $\frac{0.2\pi}{a}/\frac{0.2\pi}{a}$ (where a is the lattice constant of germanium) with a grid spacing of between 0.01 and 0.001 , for 80 monolayers with a total bias voltage across this thickness of $2V$, and a range of energies from 8.0 - 9.5 eV. The result is shown in Figure 4.3. The calibration showed some convergence at a k-grid size of 0.005 , where the variation between this result and subsequent results was around 12% . As the simulation time scales quadratically with k-grid size, maintaining the largest possible grid spacing is essential for computational efficiency. The k-grid size of 0.005 is approximately similar to that used in other studies ([117] and was chosen to optimise calculation time and result precision.

Calibration of digitizing step size was undertaken by calculating the total tunnelling current for one in-plane k-grid point ($0.0/0.0$) for 354 monolayers with a total bias voltage across this thickness of $1V$ (giving an overall field equal to the breakdown field),

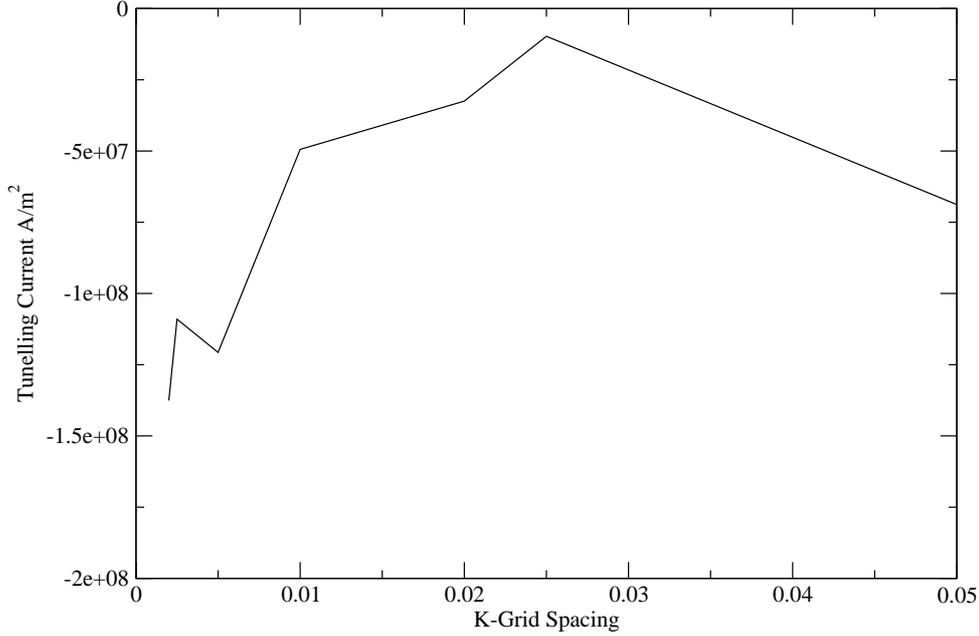


Figure 4.3: k-point grid calibration curve for band-to-band tunnelling simulations

for a range of energies from 9.0-9.5 eV. The calibration was undertaken for a range of digitizing step sizes from 0.01 eV to 0.10 eV, and the result is shown in Figure 4.4. A digitising step size of 0.01 eV was used in the calculations.

4.3 Variation of Total Tunnelling Current with Bias Voltage

The total tunnelling current over a range of bias voltages was simulated, for 800 monolayers ($0.226 \mu\text{m}$) of bulk germanium, in a range from 1.5 to 6.4V in 0.1V steps, across an energy range equivalent to the size of the bias voltage, as indicated in Figure 4.1, to just over the valence band edge. Within this range, as shown in Figure 4.1, the transmission current calculated will be that only due to band-to-band tunnelling (rather than those due to transport wholly within the valence or conduction bands) as the Fermi occupancies of any other initial and final states would be equal, giving zero contribution to the Fermi occupancy difference in Equation 4.5. This calculation was

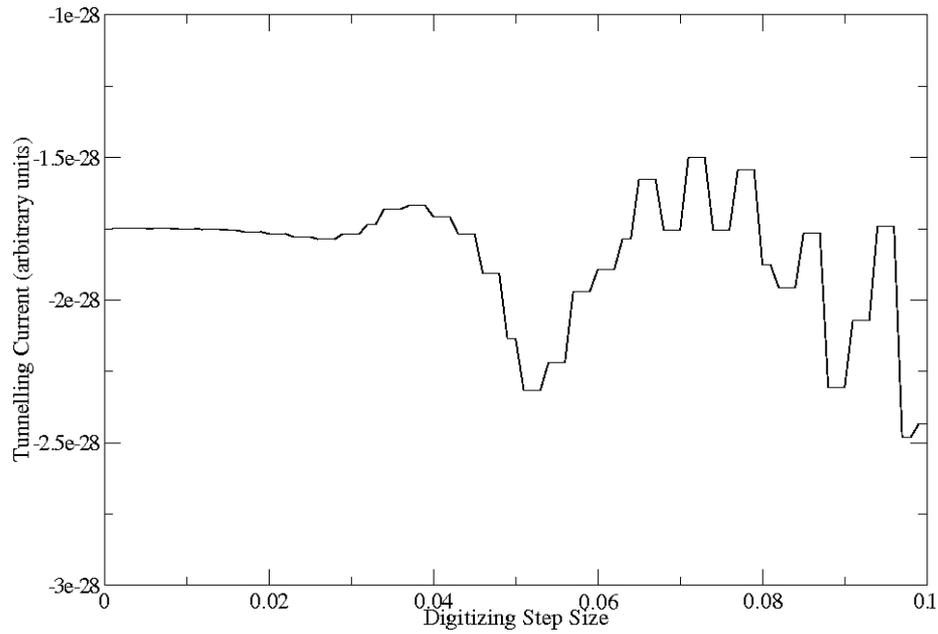


Figure 4.4: Digitizing step size calibration curve for band-to-band tunnelling simulations

undertaken at a fixed energy-grid size of 101 points, for computational efficiency, as calculation of transmission coefficients becomes more computationally expensive for high bias voltages. The tunnelling current with respect to bias voltage is shown in Figure 4.5, plotted on a logarithmic scale. Tunnelling currents were found to be negligible for lower bias voltages, falling consistently below $1 \times 10^{-4} \text{ Acm}^{-2}$ for voltages up to 3.5V — 150% of the breakdown field in germanium — with only minor numerical instability in an otherwise slowly rising current. The tunnelling current rises rapidly above around 5V — corresponding to electric fields more than double that of the breakdown field — and continues to rise steeply to significant currents up to the final simulated voltage of 6.4V, or close to 300% of the breakdown field. The current showed a steadily increasing relationship when plotted on a logarithmic scale, which approached near linearity above 4V, indicating an exponential dependence of the current on bias voltage. Hence, operation of a SPAD at voltages significantly above the breakdown field may be expected to lead to significant tunnelling currents in bulk germanium. Effective partition of the electric field, as discussed in Chapter 3, is essential to maintain low fields in the germanium

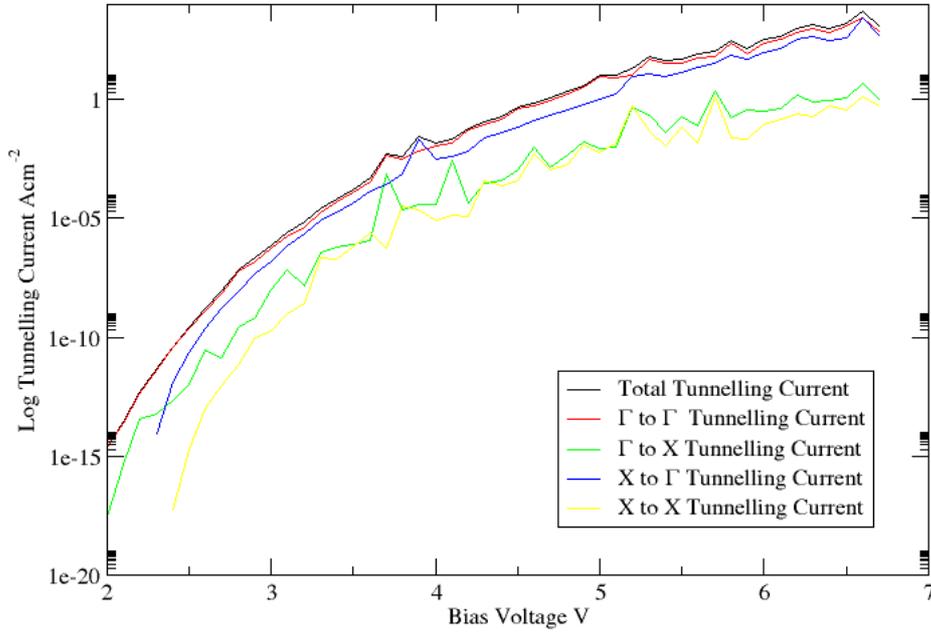


Figure 4.5: Band-to-band tunnelling current with changing bias voltage for a $0.266 \mu\text{m}$ thickness of bulk Ge

absorption layers and prevent high dark counts through tunnelling currents.

The total current between different regions of the bandstructure was calculated, by extracting transmissions between regions from the full set of transmission data. Here, a Γ state in the valence or conduction band was defined to be any state where the real part of k_z was less than $\pm\frac{\pi}{a}$ and an X-state was one where the real part of the k_z was greater than $\frac{\pi}{a}$ or less than $-\frac{\pi}{a}$. In either case, the complex part of the k-vector should be zero, indicating a plane wave in the valence or conduction band rather than an evanescent tunnelling wave, as discussed above. Although the entire bandstructure was divided evenly between Γ and X states using these definitions, the energy range used, shown schematically in Figure 4.1, covering the bandgap region, excludes states too far distant

from the Γ and X valleys at lower bias voltages as these would fall, in the conduction band, at an energy higher than the simulation range. The tunnelling between different bandgap regions is shown in Figure 4.5.

The analysis showed that the largest contribution to the tunnelling current calculated above was due to transmission between Γ valence states and Γ conduction band states, as may be expected. The next greatest contribution came from current between X and Γ , and tunnelling currents arose almost entirely from the contributions between these two sets of regions. The X to Γ contributions increased sharply after a bias of around 2.2V. The Γ to Γ bandgap is around 0.713eV, and the X valence band to Γ conduction band gap is 0.781 eV, hence these transitions are easily accessible for tunnelling, with the former having a higher tunnelling probability at lower voltages. Significantly smaller (by one or two orders of magnitude) contributions from the valence band at Γ to the conduction band at X, and from the valence band at X to the conduction band at X were observed, reflecting the larger energy gap for these transitions — over 1eV. Broadly, the transitions between individual bandstructure regions followed the same exponential trend as the overall tunnelling current. A number of small peaks are observed across all curves, but may be most likely attributed to numerical noise arising from the lower resolution of this calculation. No transitions to the L-valley were observed in these simulations; the 100 germanium orientation used in the simulations precludes any transitions in this direction.

4.4 Total Tunnelling Current with Changing Thickness

The total tunnelling current over a range of thicknesses of bulk germanium from 550 to 1550 monolayers (0.156 μm to 0.438 μm) in 50-monolayer steps, was calculated, for an energy range equivalent to the applied bias voltage, as discussed above. The bias voltage for the simulations was adjusted such that the total field across the germanium thickness was maintained at 100% of the breakdown field for germanium. The tunnelling

current with respect to germanium thickness is shown in Figure 4.6, plotted on a logarithmic scale. The total tunnelling current in Figure 4.6 shows a broadly linear trend.

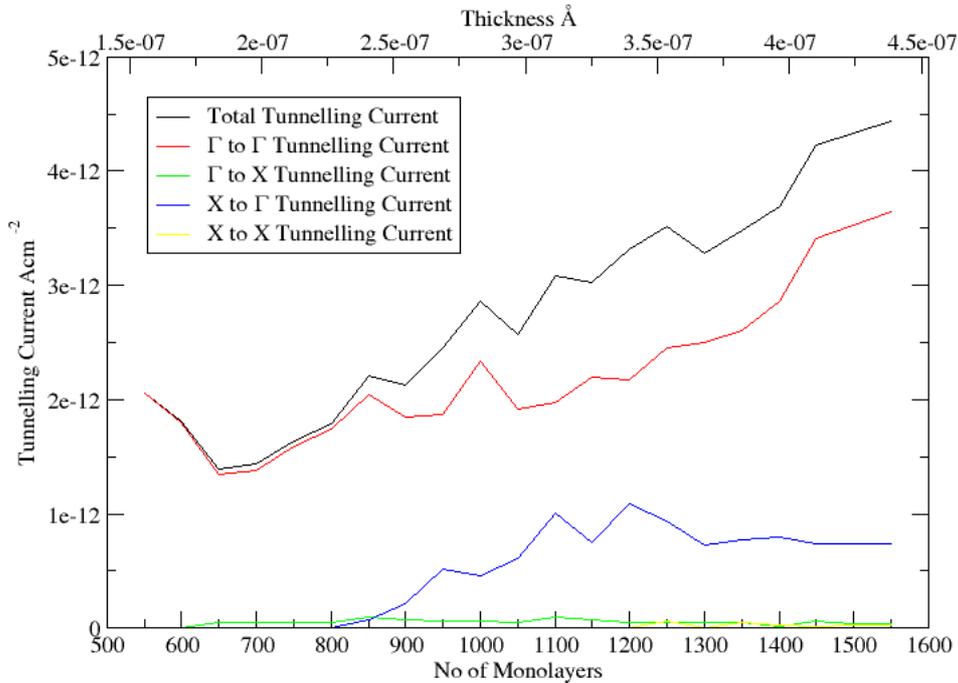


Figure 4.6: Band-to-band tunnelling current with changing thickness at 100% of Ge breakdown field

Increasing the thickness introduces a greater number of states for tunnelling on both the valence band and conduction band sides of the calculation, and this effect results in this upward trend. An analysis of tunnelling between bandstructure regions was conducted for thickness dependence, and is shown in Figure 4.6. For thicknesses below 850 monolayers, the current arises almost entirely from Γ to Γ transitions. Between 850 and 1550 monolayers — bias voltages of 2.26 to 4.38 V — an increasing contribution from X to Γ transitions is observed, indicating the availability of further valence band states for tunnelling at increasing thickness. Γ to Γ and X to Γ transitions are easily accessible for tunnelling at this field. Deep valence band states and high conduction band states

are only at higher thicknesses – reflected in the small contributions made by Γ to X and X to X transitions to total tunnelling current. The X to X tunnelling current was zero up to a thickness of 850 monolayers, indicating that tunnelling between these valence and conduction band states is only instigated above a critical thickness of germanium. For sufficiently thick samples, a saturation in the tunnelling current should be observed, where no further valence band states become available for tunnelling. Additionally, the transmission coefficients will reduce with increasing thickness. Ideally, the calculation should be run for a greater number of monolayers.

4.5 Temperature Dependence of Tunnelling Current

The temperature dependence of the tunnelling current was calculated. As the simulation code has no direct method for adjusting temperature data, a method based on adjustment of form factors and calculation of the associated change in bandgap was used. The formula developed by Varshni [119] calculates the adjustment in bandgap due to a change in temperature:

$$E_g(T) = E_0 - \frac{\alpha T^2}{T + \beta} \quad (4.6)$$

Where α and β are empirical fitting parameters, and E_0 denotes the bandgap at zero temperature. Hence, the change in bandgap ΔE_g arising from a change in temperature may be given by:

$$\Delta E_g(T_2 - T_1) = \frac{\alpha T_2^2}{T_2 + \beta} - \frac{\alpha T_1^2}{T_1 + \beta} \quad (4.7)$$

Applying Varshni coefficients for germanium obtained from literature [52] to Equation 4.7 provides a relationship between the temperature change and the change in bandgap, and allows modelling of the effects of change in temperature, by changing the bandgap. The fitting parameters used in Equation 4.6 are listed in Table 4.2. The indirect bandgap obtained using the original form factors corresponds to a temperature of $T_1=270\text{K}$, using these fitting parameters. As the temperature dependence of the tunnelling current was sought, the bandgap shift, and hence the formfactor adjustments, were performed relative to this reference temperature. The range of temperatures T_2 over which the bandgap

Table 4.2: Fitting parameters applied to the Varshni formula for calculating bandgap dependence on temperature

| | |
|----------------------|-----------------------|
| α at Γ | 5.82×10^{-4} |
| β at Γ | 296 |
| α at X | 4.73×10^{-4} |
| β at X | 636 |
| α at L | 4.77×10^{-4} |
| β at L | 235 |

adjustment was made corresponded to $T=170-470$ K in 20 K steps. The change in bandgap due to variations in temperature was implemented by determining the appropriate change in form factors required to implement this bandgap in the pseudopotential code. Initial inputs to the code included the V3, V8 and V11 form factors, and any change ΔV in these can be related to a change in the bandgap at X, L or Γ using the following matrix relation:

$$\begin{bmatrix} \Delta V_3 \\ \Delta V_8 \\ \Delta V_{11} \end{bmatrix} = \begin{bmatrix} a_3^\Gamma & a_8^\Gamma & a_{11}^\Gamma \\ a_3^X & a_8^X & a_{11}^X \\ a_3^L & a_8^L & a_{11}^L \end{bmatrix}^{-1} \begin{bmatrix} \Delta E_g^\Gamma \\ \Delta E_g^X \\ \Delta E_g^L \end{bmatrix} \quad (4.8)$$

Here, the coefficients a of the matrix are determined by examining the simulated band gap change observed when making a small change in each form factor individually. To determine these coefficients, each form factor was changed separately by 1%, and the bandgap changes at X, L and Γ for each case were examined. The coefficients a_n^x may then be calculated using the following formula:

$$a_n^x = \frac{\Delta E_g^x}{\Delta V_n} \quad (4.9)$$

Where ΔE_g is given by:

$$\Delta E_g^x = \Delta E_c^x - \Delta E_{vb}^\Gamma \quad (4.10)$$

Table 4.3: Coefficients calculated for each bandgap change by changing V3, V8, V11 formfactors

| | Γ | X | L |
|----------|----------|---------|---------|
| a_3 | -1.2800 | -12.994 | -15.566 |
| a_8 | -93.061 | -148.98 | -117.96 |
| a_{11} | -102.17 | -79.924 | -81.709 |

and ΔE_c^x , ΔE_{vb}^x , with $x = L, X, \Gamma$, refer to the conduction and valence band edges at different symmetry points in k-space, respectively. The coefficient inputs to equation 4.8 derived from the bandstructure calculations undertaken after changing each form factor individually are shown in Table 4.3: From here, the form factors may be adjusted using Equation 4.8, to obtain the required change in bandgap. It is notable that the bandgaps assigned to the original form factors, obtained from literature values, were calculated for an arbitrary value of temperature. However, form factors represent empirical fitting data which may be specific to the computational method originally used to obtain them. For example, a limited plane-wave basis set, used for computational efficiency, may be adjusted by the use of appropriate form factors such that a simulated bandstructure matches experimental data as closely as possible. Furthermore, the number of form factors used (in this case 3) reflects the accuracy of the fit to experiment, and the use of more form factors may reflect the need to fit more closely to experimental parameters, while the use of fewer may indicate that the data is fitted to a limited number of degrees of freedom of the experimental data, though it may be sufficiently accurate for the required use.

A change in temperature would result in a change in the the Fermi occupancies f_L and f_R given in Equation 4.5. However, within this temperature range, for an undoped semiconductor, these changes in occupancies may be considered insignificant. A calculation of the change in the Fermi occupancy difference at the highest temperature simulated, 470K, at the Γ point yielded a difference $\Delta(f_R - f_L)$ of around 0.03%, in-

dicating that the assumption of filled valence band states and empty conduction band states throughout the temperature range considered.

A number of other factors may be affected by a change in temperature, including lattice constant. The low thermal expansion coefficient of germanium ($5.9 \times 10^{-6} \text{C}^{-1}$, linear) suggests that, within the temperature range studied, the change in lattice constant will be negligible compared to the order of change of bandgaps applied; nonetheless the lattice constant was adjusted for simulations at each temperature using this linear expansion coefficient. Hence the method used here, changing only the bandgap in response to a change in temperature, represents a simple but fast technique for investigating the tunnelling dependence on temperature.

The results of the calculation of the tunnelling current vs temperature dependence on a logarithmic scale are shown in Figure 4.7, for a temperature range of 270–470K under a bias voltage of 2.362V over 800 monolayers, a field equivalent to 100% of the breakdown field for germanium. The tunnelling current shown on Figure 4.7 shows a linear trend, indicating an exponential relationship between increasing temperature and tunnelling current. The effect of increasing temperature, given by Equation 4.6 and considering the range of coefficients α and β used, should be a first-order decrease in bandgap. Hence, this result also implies an inverse exponential relationship between the bandgap size and tunnelling current. A comparison between the calculated target bandgap changes for each temperature, given by Equation 4.6 and the bandgap output from the simulations indicated a maximum deviation between these numbers of 0.87%, indicating that this provides an accurate method of simulating the effect of temperature on tunnelling current.

An analysis of tunnelling between bandstructure regions was conducted for thickness dependence, and is shown in Figure 4.7. By far the greatest contribution to total tunnelling current was transitions between the valence bands and conduction bands at Γ .

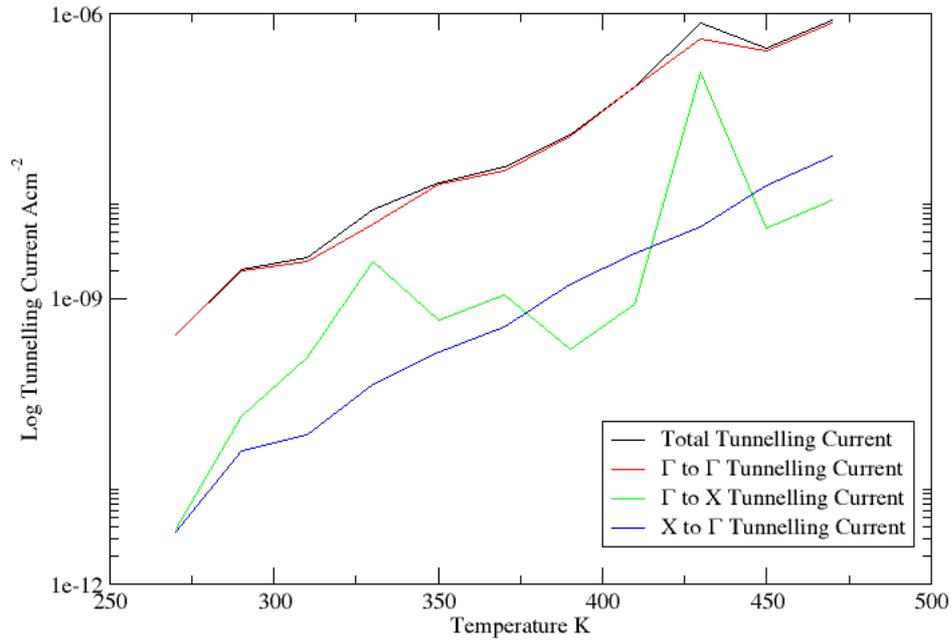


Figure 4.7: Band-to-band tunnelling current with increasing temperature at 100% of Ge breakdown field

From the previous discussion of the effect of the bandgap at these regions, and given that these simulations were undertaken at a bias voltage at which Γ to Γ transitions have dominated in previous calculations, this result is to be expected. A smaller contribution from the X to Γ region, following broadly the same trend as the total tunnelling current, was also observed. No X to X transitions were observed at this bias voltage, and this is again consistent with previous calculations — in the calculation of tunnelling current with bias voltage given in Figure 4.5, no X to X transitions were observed until a voltage of 2.4V, and the narrowing of the bandgap with increasing temperature does not appear sufficient to initiate X to X transitions at this bias voltage, for this range of temperatures. Γ to X transitions showed significant fluctuations, and the current between these regions frequently lay above that observed from the X to Γ transitions, in contrast to

the previous results, where this transition showed only a small contribution to the total tunnelling current. The Γ to X transmission current rises to peaks at 330K and 430K and drops to troughs at 390K and 450K. Assigning states to either Γ or X relies on an arbitrary separation of states either side of the mid-point between Γ and X; if there are a large number of states close to the mid-point, this may result in fluctuations in the current calculated for each transition.

4.6 Conclusion

In this chapter, band-to-band tunnelling currents for bulk Ge were calculated for a range of parameters, including the bias voltage across the germanium, changing Ge thickness, and the change in bandgap due to a change in temperature. Since SPADs are operated above breakdown voltage, band-to-band tunnelling within the Ge absorption layer may present a significant source of noise currents and dark counts. Bias voltage was found to have a significant impact on band-to-band tunnelling currents, with an exponential increase in tunnelling current with increasing bias voltage. As bias voltage was increased, tunnelling between deeper states, such as valence band X to conduction band Γ was observed. At above 200% of the breakdown field for Ge, the tunnelling current was found to rise rapidly, hence, operation of a SPAD considerably above breakdown voltage may lead to significant noise currents.

The variation in tunnelling current with increasing thickness of the Ge layer was calculated, for a fixed electric field. The tunnelling current was found to rise broadly linearly with increasing layer thickness, indicating that a greater number of valence and conduction band states become available for tunnelling with increasing thickness. In particular, deeper transitions such as those between valence band X and conduction band Γ were found to increase at higher thicknesses. The tunnelling current is expected to plateau

above a certain thickness of Ge, where no further states are available for tunnelling, however this effect was not observed within the thickness range simulated. The thicknesses simulated here — up to $0.44 \mu\text{m}$ — remain below the indicative thickness of the Ge absorption layer for the current SPAD designs of $1\mu\text{m}$ indicating that a significant contribution to tunnelling current from all regions of the bandstructure may be expected in all regions of the Ge absorber if the electric field is sufficiently high.

The temperature dependence of the tunnelling current was calculated, by adjusting the formfactors used in the empirical pseudopotential calculation to generate an adjusted bandgap at each temperature, according to Varshni's [119] formula. An exponentially increasing relationship between temperature and tunnelling current was observed, indicating that the decrease in bandgap with increasing current results in a higher probability of tunnelling between all regions of the Brillouin Zone. The greatest contribution to the tunnelling current in this case arose from valence band Γ to conduction band Γ transitions. This method represents a simple formulation for calculation the temperature dependence of the tunnelling current.

From the IV curves given in Chapter 3, simulated avalanche currents in SPADs range from orders of 10^2 to 10^{-4} Acm^{-2} . Band-to-band tunnelling currents simulated here at the Ge breakdown field are of the order of $10^{-12} \text{ Acm}^{-2}$, and increase with bias voltage beyond this point, approaching the lowest orders of the avalanche current for fields at around 150% of the nominal breakdown field for Ge. Hence, band-to-band tunnelling currents will become significant contributors to noise currents at voltages significantly above breakdown, but are likely to make only a small contribution to noise current pulses at lower fields. However, band-to-band tunnelling may be important at high fields in samples thinner than the characteristic ionisation length, where carriers are not in the sample for a sufficient length of time to undergo impact ionisation processes.

[40] characterise dark count rates for SiGe SPADs grown within the SPADs project.

At room temperature, dark currents ranged from 0.04 to 0.4 Acm⁻² with increasing bias voltage. Again, band-to-band tunnelling currents will be observed within an order of magnitude of these values only at 160% of breakdown field — band-to-band tunnelling currents may contribute significantly to the overall dark current at very high bias voltages, considerably above breakdown, but are likely to make only a small contribution to the dark current at lower fields.

Chapter 5

Electronic Structure of Defects in Bulk Germanium

In the work presented in previous chapters, models were developed to describe the electronic characteristics of SPADs with different doping profiles, and to quantify the effect of band-to-band tunnelling in germanium. However, these models used a germanium layer that was assumed to be free from defects, with an ideal bandstructure generated on this assumption. As discussed in Chapter 2, in realistic silicon-germanium heteroepitaxy the lattice mismatch between silicon and germanium results in strain within the germanium layer up to some critical thickness [65] and defects within the germanium for layers grown above this thickness [66] — a situation that is applicable to the germanium thicknesses used in SPADs.

The structure of defects has been discussed in Chapter 2. A number of previous studies have been conducted to establish the atomic and electronic behaviours of defects in semiconductor materials, and point defects and interstitials have been studied extensively both experimentally and theoretically [120–128]. Broadly, work in this area has provided significant background on the nature of point defects and interstitials, confirming the presence and energetics of the dangling bonds around these defects in diamond

structures, and describing the effect on strain, lattice distortions, dopant diffusion and electronic properties, particularly in silicon.

A number of works have discussed the structure and electronic properties of dislocations in diamond structures, using *ab initio* and experimental approaches [129, 130]. [131] notes that studies on the electronic structure of defects in Ge are limited, although some experimental work has been undertaken on irradiation induced point defects and a trap centre 0.31eV below the conduction band identified [132–134]. Other studies have identified bandgap states 0.82 eV above the valence band minimum, from an *ab initio* study of germanium interstitials [135], and [136] identified defect levels in germanium at 0.05 eV and 0.11 eV above the valence band edge at the Γ point for hydrogen-passivated dangling bonds. However, the *ab initio* method used in the latter calculations was found to underestimate the Ge band gap, and a number of empirical corrections were applied. These values have nonetheless been used in device-level simulations of SiGe SPADS [137].

DFT has attracted attention as a highly-powerful and computationally efficient technique for calculating electronic structures for a range of materials [138], however, there have been limitations to the application of DFT to semiconductor materials, particularly Ge, because of the narrowing or disappearance of the band gap, a phenomenon discussed in 2. A number of techniques have been applied to address this limitation, including the use of fitting parameters, adjustments to the self-interaction introduced by the use of LDA or GGA functionals, or development of physically ‘good’ potentials for the system [99, 136, 138–141]. A further difficulty in calculating bandstructures for crystalline materials containing defects arises from the periodic supercell formulation used in DFT: insufficiently small supercells can lead to interaction between adjacent cells and interference in energy minimisation or electronic structure calculations [77].

In this chapter, DFT simulations using the CASTEP software are conducted, to identify optimised geometric configurations of misfit dislocations in Ge, and the additional band

gap states that arise from these.

5.1 The CASTEP Software

The CASTEP software [142] was used to perform geometry optimisations and calculate electronic bandstructures for Ge cells containing a misfit dislocation. The DFT method implementation within CASTEP involves formulation of a set of one-electron Kohn-Sham equations, which are then solved using the plane-wave pseudopotential approach. As described in [2], Bloch's theorem is applied to expand the wavefunctions in a plane-wave basis set, the electron-ion potential is described using a pseudopotential, and interactions between each electron are encapsulated in an exchange–correlation functional. A self-consistent energy-minimisation algorithm is used to solve the electronic wavefunctions and the corresponding charge density, and a number of other corrections and parameters may be applied in the software [142].

One initial step for development of a simulation model for bulk germanium containing defects is the development of a set of parameters applicable to the simulation. These were developed using a primitive cell of undefected germanium.

5.2 Simulation Models

Germanium has a diamond structure with a primitive rhombohedral cell and a conventional cubic unit cell, as shown in Figure 5.1. The CASTEP software uses periodic boundary conditions, and hence any supercell provided for simulation will be replicated periodically across all 3-dimensional space. Simulation of bulk germanium therefore requires only provision of a single primitive cell to reproduce the full structure of bulk germanium. The primitive cell was provided to the CASTEP input file using Cartesian lattice vectors and the relative positions of the two Ge atoms within the cell, and a number of parameters were established using this initial cell.

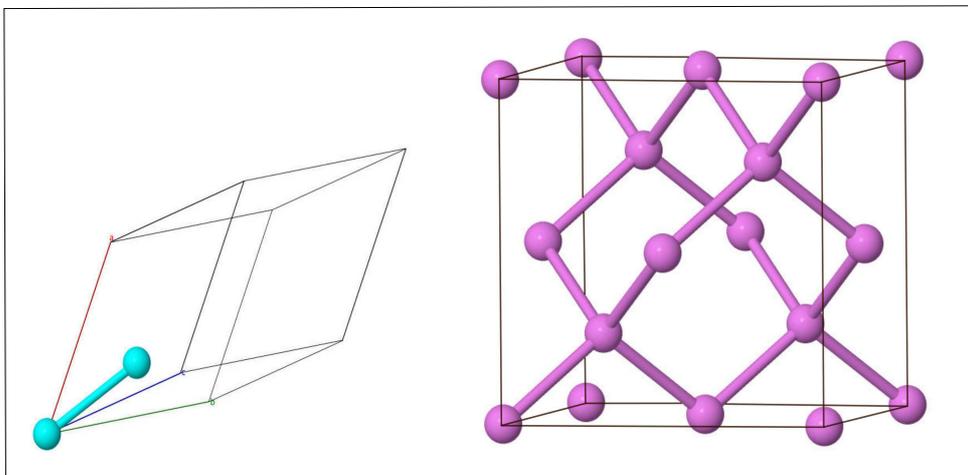


Figure 5.1: The rhombohedral primitive unit cell of silicon and germanium, and the conventional cubic cell

The cutoff energy which would yield sufficient accuracy while minimising computational requirements was determined using a calibration on the primitive cell. A single-point energy minimisation was calculated for the primitive cell using a Monkhorst-Pack grid [100] with a spacing of 0.04 \AA^{-1} (usually sufficient for most calculations), using default parameters and a range of cutoff energies to either side of the default value of 300 eV. The cutoff energy at which the single-point energy calculation showed convergence was found to be 400 eV. Further calculations were carried out using a ‘FINE’ basis set precision, indicating a convergence of atomic energies of about 0.1eV/atom.

Since the number of plane waves used in the basis set will vary with cutoff energy, depending on the density of the k-point grid defining reciprocal space, a correction factor may be applied, accounting for the difference in the number of plane waves in a basis

set using an infinitely dense k-point grid, and that using a discrete k-point grid. The correction is based on the derivative of the total energy of the system with respect to the natural logarithm of the cutoff energy [143]:

$$\frac{dE_{tot}}{d\ln E_{cut}} \quad (5.1)$$

Applying this basis set correction was found to yield slightly more accurate results and was therefore applied to all further simulations. The convergence tolerance for simulations — the difference in energy between successive self-consistent calculations at which the simulation was assumed to be fully converged — was set to 1×10^{-4} eV for geometry optimisation calculations, and 1×10^{-7} eV for electronic structure calculations, and calculations were performed without spin polarisation.

As discussed, loss of bandgap is common problem in DFT calculations performed for Ge. Here, the choice of pseudopotential and exchange–correlation functional was considered to minimise this problem and produce a Ge bandstructure with bandgaps matching experimental values as closely as possible. Pseudopotentials may be norm–conserving, wherein the requirement that the calculation of valence orbital energies must match those achieved from an all–electron calculation is maintained, or ultrasoft, which relaxes this requirement. Initial calculations were undertaken using the norm–conserving pseudopotential Ge.00.recpot provided with CASTEP. This pseudopotential treats the 3d10 electrons as core, and has been previously successfully applied to calculations of Ge–dimer relaxation on Si(100) [144] and Ge bonding on the Si(100) surface [145]. The treatment of the 3d10 electrons as core makes this pseudopotential extremely computationally efficient, however, it may also introduce significant inaccuracies into the calculation of electronic valence wavefunctions, by incorporating valence electrons into the core part of the pseudofunctions.

A number of LDA and GGA exchange–correlation functionals are available within the CASTEP software. Here, the Revised-Perdew-Burke-Ernzerhof (RPBE) functional [146]

was used. The original Perdew-Burke-Ernzerhof functional [147] was designed to satisfy conditions that are energetically significant for a uniform electron gas, and neglecting conditions, such as second-order gradient co-efficients in the slowly-varying limit, that have little effect on energetics of the system, in contrast to earlier functionals. The RPBE functional was further developed along the same principle as PBE, but with parameterisations applied to match chemabsorption data.

The bulk Ge bandstructure produced using these parameters is shown in Figure 5.2 and Table 5.1 compares the established experimental bandgaps for bulk Ge with those obtained from the simulation using these parameters. Using this set of parameters, the

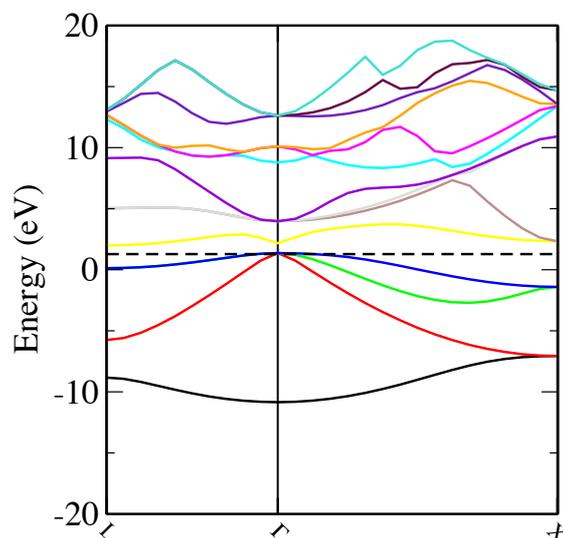


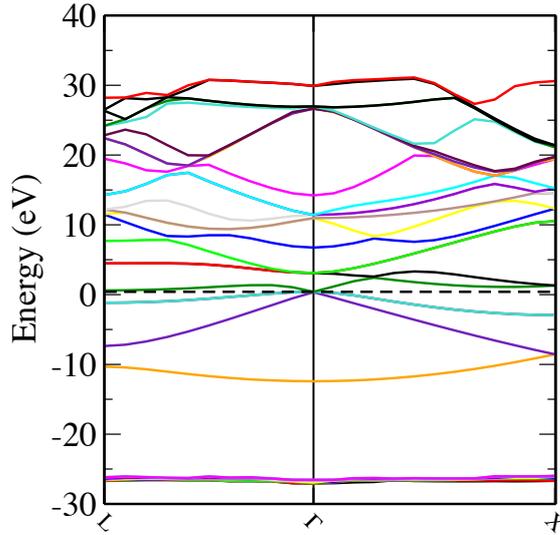
Figure 5.2: Optimized bandstructure for bulk Ge

bandgap found at Γ varied by within 1% of the experimental values and that at L by around 5%. There was a greater discrepancy in the bandgap at X, of around 23%.

Further simulations were undertaken with a more computationally-expensive pseudopotential in which the 3d orbital electrons are treated as valence, and these results are shown in Figure 5.3. Use of this pseudopotential leads to loss of the bandgap at Γ

Table 5.1: Comparison of experimental and simulated bandgaps for bulk Ge [103]

| Bandgap | Experimental eV | Simulated eV | % Change |
|------------|-----------------|--------------|----------|
| E_L | 0.66 | 0.628 | 5% |
| E_X | 1.2 | 0.972 | 23% |
| E_Γ | 0.8 | 0.808 | 1% |

**Figure 5.3:** Optimized bandstructure for bulk Ge, using a pseudopotential in which the 3d electrons are treated as valence

and a significant increase in computational time — from around 200 seconds to in excess of 3 hours for this primitive cell. The discrepancies in the bandgaps at X and L were broadly of the same order as those observed using the Ge_00.recpot potential. Although treatment of the 3d10 orbitals as core rather than valence may introduce some inaccuracies to a full-bandstructure calculation, here, the representation of the bandgap, and the ability to complete calculations with limited computational burden were key components of the modelling approach, and the more computationally-efficient Ge_00.recpot pseudopotential was therefore retained for further simulations.

Simulation of misfit dislocations in bulk germanium requires development of a cell containing an artificially –placed dislocation. This was obtained by removing atoms from a supercell of bulk germanium. The periodic boundary conditions used by the CASTEP software mean that this cell will be replicated across 3–dimensional space, so one requirement for the simulation supercell was ideally that the cell should be sufficiently large that adjacent cells, containing a dislocation each, replicated the defect density found in physical systems. The dislocation density for realistic heteroepitaxially–grown germanium is of the order of $1 \times 10^7 \text{cm}^{-2}$ [75] or about one every million atoms. However, DFT calculations can be computationally–expensive and require, even ordinarily, high performance computing (HPC) resources for anything other than very basic simulations. Additionally, the presence of the defect within the cell disrupts the inherent symmetry found in bulk germanium, and further increases the computational load, as symmetric cells may be computed more efficiently by applying symmetry operations to simplify the calculation. Hence, simulation of Ge with physically–realistic defect densities is not possible at this level of theory.

Instead, 2–dimensional supercells of Ge, in which as many atoms as possible are used in two lattice directions, (designated x and y) while maintaining a cell only one lattice vector deep, and with a missing atom, in the third dimension, designated z were used. Tessellation of the cell through space to maintain periodic boundary conditions will propagate a misfit dislocation in this z –direction in a physically–realistic way. Hence, the limitation placed on the number of atoms for practical simulation may be applied optimally to separation of adjacent dislocations, in the x – and y – directions, since no spacing is required in the z – direction.

A number of supercells of germanium containing a line of missing atoms across half the x –length of the cell were simulated. The supercell sizes are described by the number of primitive unit cells they were composed of in the x – and y – directions, and are shown in Figure 5.4, and an of a representative supercell propagated through space is

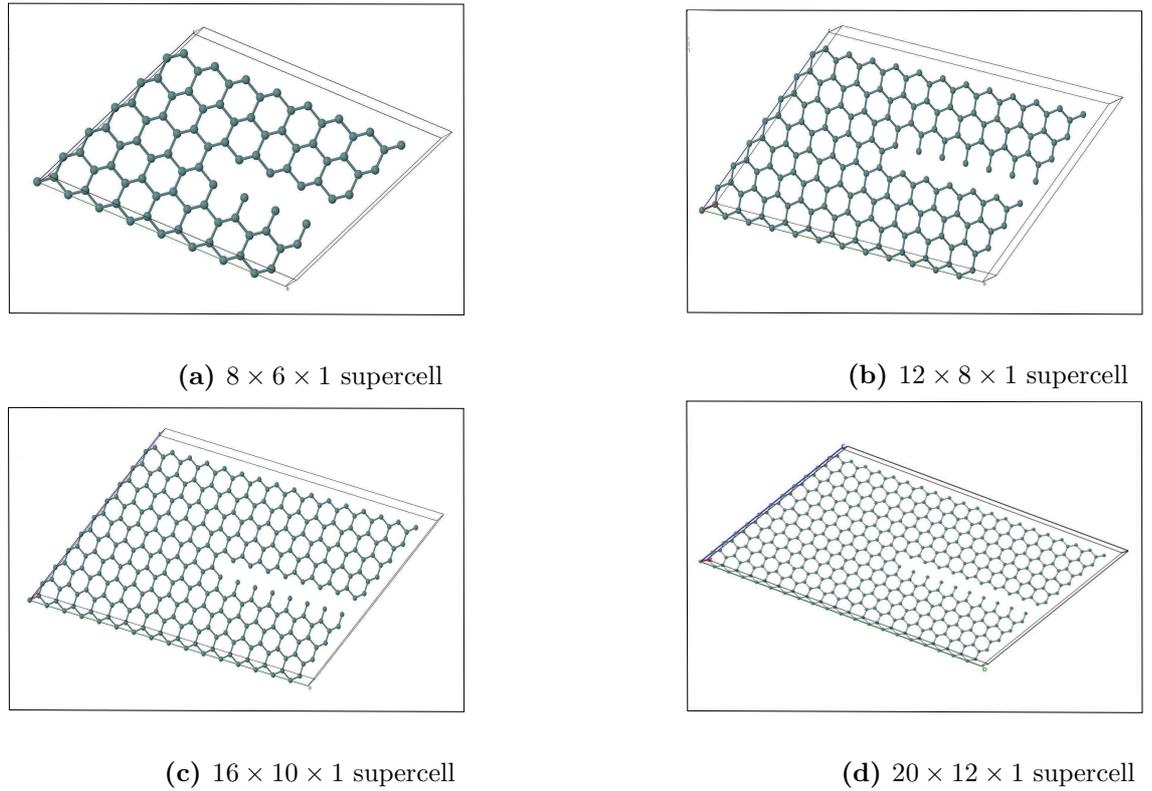


Figure 5.4: Germanium supercells with misfit dislocations used for DFT calculations

shown in Figure 5.5. A number of calculations were performed using these cells and the parameters discussed above: geometry optimization was conducted for each cell, to establish what cell size began to yield convergent optimized geometries and provide an amended structure for subsequent bandstructure calculations. These geometry optimized cells were then used to calculate bandstructures for the defected cells. From here, a post-processing step was used to identify electronic states arising from defects in each cell at the Γ point, and calculate the energies of each defect state.

Geometry optimisation in CASTEP establishes a ground state energy for the electrons, then optimises the ionic positions, within, in this case, a fixed unit cell size and shape. The minimum energy and enthalpy configuration for the ionic system corresponds to zero force and stress; the forces arise from the pseudopotential applied to the calculation,

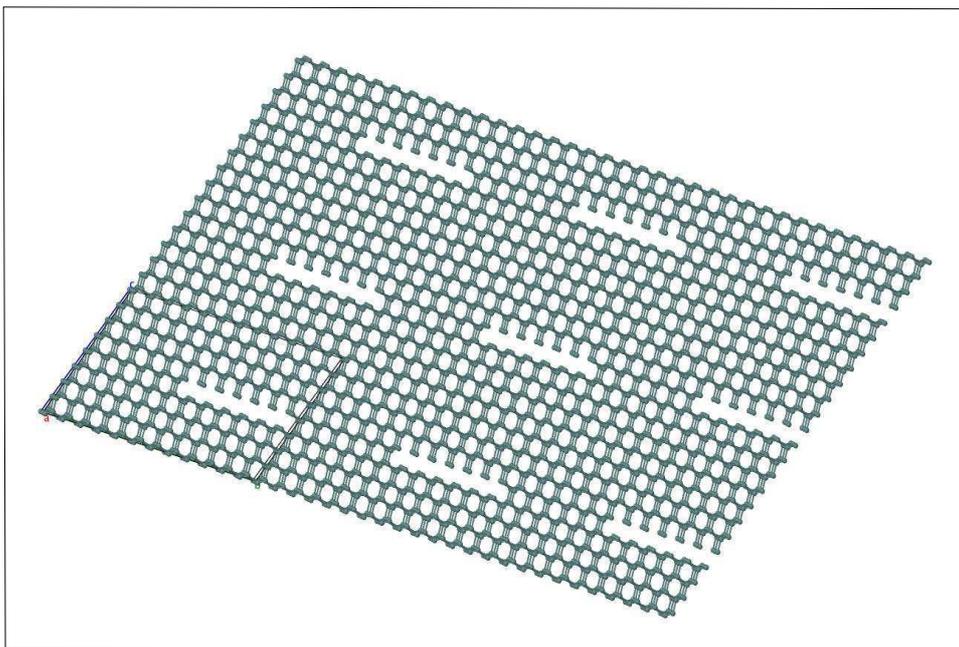


Figure 5.5: $12 \times 10 \times 1$ supercell of Ge containing a misfit dislocation, propagated across 3-dimensional space

and from the ion-ion Coulomb interaction, while contributions to stresses arise from the kinetic energy of the ions and the Hartree energy. Within the DFT framework, the optimised geometry is obtained by finding a global minimum in the overall multi-dimensional potential energy surface constructed from the basis sets discussed previously [148].

Geometry optimisation was performed for all cells containing a defect to obtain an energy-relaxed structure for the manually-placed misfit dislocations. It was necessary to maintain the size and shape of the unit cell (i.e. the lattice vectors) and optimise within this structure, as the post-processing algorithm used for calculation of Γ -point energies requires that the supercell lattice vectors are integer multiples of the unit cell lattice vectors.

5.2.1 Geometry Optimisation Parameters

The BroydenFletcherGoldfarbShanno (BFGS) algorithm was used for the geometry optimisation calculations. This is a quasi-Newtonian method that uses the approximation that the energy surface is quadratic about an energy minimum, and thus attempts to minimise the second derivative of the energy potential surface to find optimised atomic configurations [148].

5.2.2 Geometry Optimisation Results

The result of the geometry optimisation for the $8 \times 6 \times 1$ supercell containing a misfit dislocation is shown in Figure 5.6. A small number of atoms around the misfit dislocation show significant movements, while others have remained close to their original positions. Overall, the $8 \times 6 \times 1$ geometry optimisation yielded large atomic movements around the

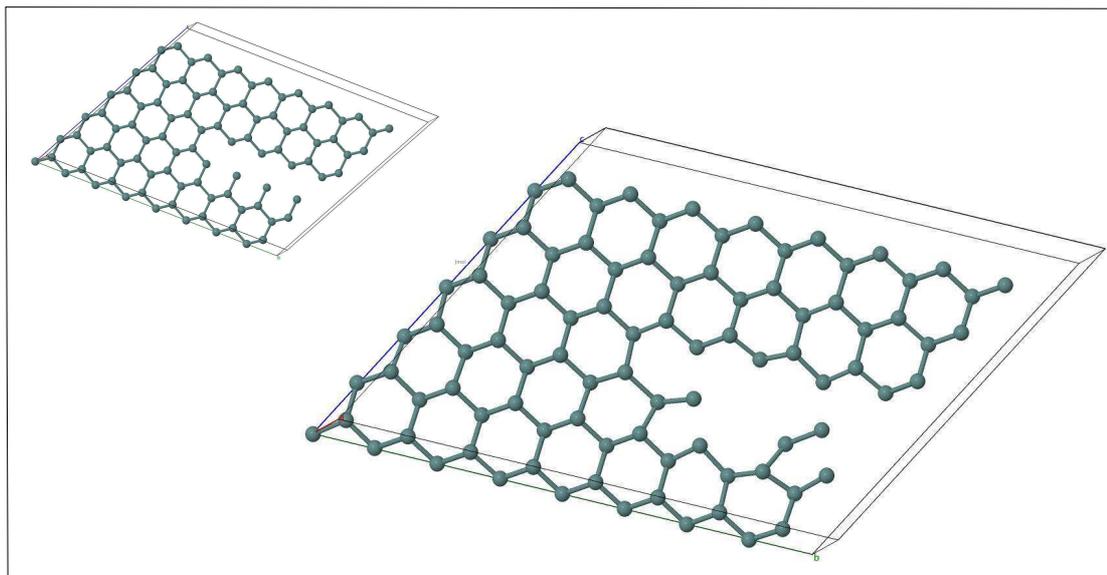


Figure 5.6: Geometry optimised $8 \times 6 \times 1$ supercell of Ge containing a misfit dislocation, with original input cell inset

dangling germanium atoms found along the misfit line, and smaller movements in the remainder of the cell. Qualitatively, the optimised geometry for this supercell showed that dangling bonds were retained within the dislocation core, and little reconstruction

was observed.

A calculation of the changes in bond angles from the theoretical crystalline structure for Ge, with a bond angle of 109.47° , showed that all bonds showed some deviation from the theoretical value, although for the majority of these — 345 of the total of 518 bond sets — this deviation was less than 1° , and for 53 bonds the deviation was less than 0.1° . The largest deviations were observed around the dislocation line, with angles ranging from a maximum of 123.22° (a 12.6% deviation) to a minimum of 79.2° (28% deviation) for one bond within the dislocation core. These maximum and minimum values are both lower than those observed in previous studies of dislocation structures in diamond structures for well-reconstructed dislocations (around $135\text{--}138^\circ$ maximum and $95\text{--}96^\circ$ minimum), however bond-angles as low as 50° have been observed for ‘kink’ structures in dislocations in diamond materials [149–151].

The deviations in bond length for this configuration were calculated, and a bond stretched by 11.5% compared to the theoretical value for bulk Ge of 2.45 \AA was observed for one bond in the dislocation core, with a number of other bonds stretched by up to around 5-6% within the dislocation core for Ge atoms with dangling bonds. Compressed bonds were also observed, generally adjacent to the stretched bonds, but with a maximum deviation from the theoretical value of 2.7%. Compared to previous calculations of changes in bond lengths in dislocation structures in diamond materials, which generally showed discrepancies of less than 3%, [149–151], a small number of bond-length changes calculated here were considerably higher, and most likely indicate that the cell is not relaxed to physically-realistic structures around the dislocation at this cell size. It has been observed that incomplete reconstruction of dislocation cores in small supercells can lead to additional stresses within the cell, and formation of stress-induced electronic bands in electronic calculations [130], and this cell size is therefore unlikely to yield electronic structure results that adequately reproduce the band gap states arising from dislocations in bulk Ge.

The geometry-optimised $12 \times 8 \times 1$ supercell is shown in Figure 5.7. Again, for

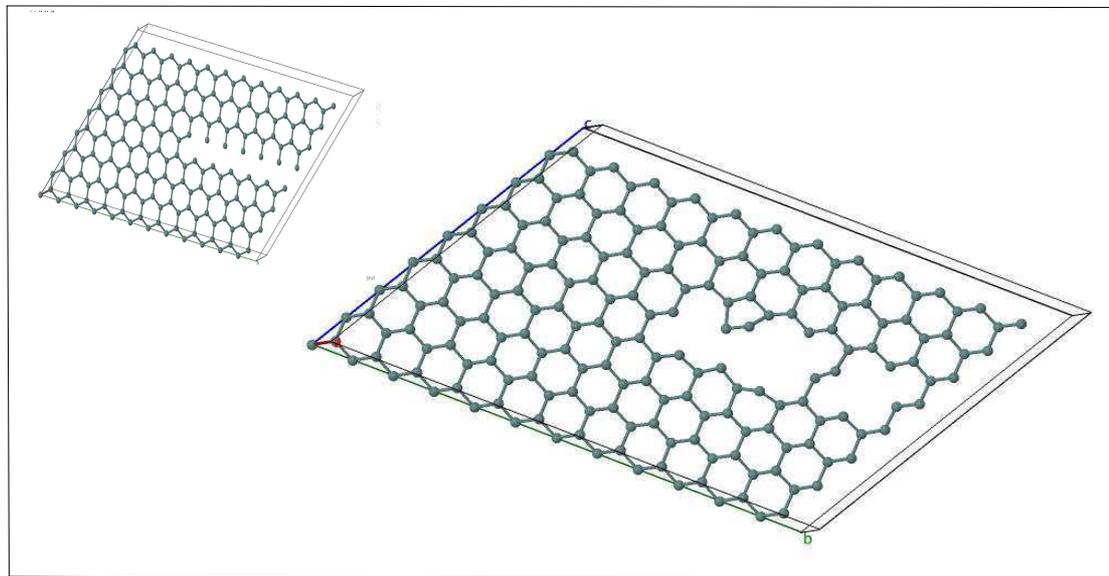


Figure 5.7: Geometry optimised $12 \times 8 \times 1$ supercell of Ge containing a misfit dislocation, with original input cell inset

the $12 \times 8 \times 1$, a number of distorted bonds were found around the dislocation core, with a maximum bond angle of 125° and a minimum of 72° . However, this was for a small number of angles, and 868 out of the total of 1082 angles showed a less than 1% deviation from the theoretical value for bulk Ge. The deviation in bond lengths was also smaller for this cell size, with a maximum stretch of 2.93% and a maximum compression of 4.36%, beginning to approach values observed in previous studies. Qualitatively, a closed pentaring — a structure formed of a five-membered ring of atoms — was observed in this structure alongside stretched bonds and evidence of reconstruction across the dislocation core.

The geometry-optimised structure for the $16 \times 10 \times 1$ supercell is shown in Figure 5.8. Here, the dislocation structure can be seen to take on a regular converged pat-

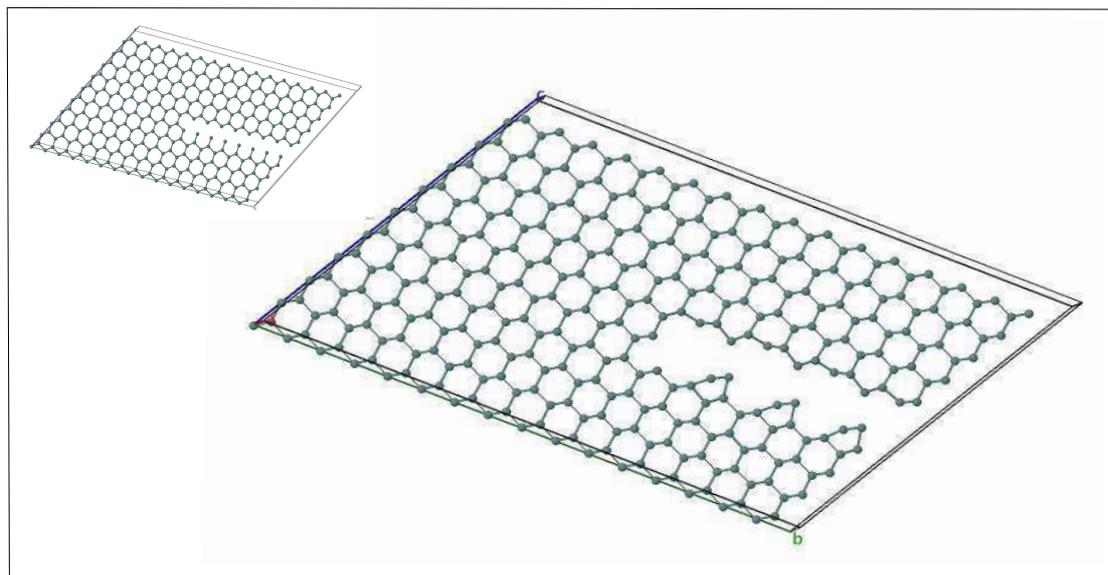


Figure 5.8: Geometry optimised $16 \times 10 \times 1$ supercell of Ge containing a misfit dislocation, with original input cell inset

tern, with repeated pentaring-structured germanium, alongside alternate stretched and compressed bonds surrounding the dislocation. The maximum bond angle here was 123° and the minimum was 71° . The maximum bond compression was by 4.7% however the maximum stretch was only 2%, indicating that, in these simulations, relaxation around the dislocation can generate structures similar to ‘kink’ structures found in other studies of dislocations and faults in diamond structures. Generally, the configurations of atoms around the dislocation showed the formation of pentarings and 60° glide and shuffle sets commonly observed in previous studies on the structure of diamond materials [129, 130, 149–151].

The geometry optimised structure for the $20 \times 12 \times 1$ supercell is shown in Figure 5.9. This dislocation structure mirrors the regular converged pattern observed for the $16 \times 10 \times 1$ cell, extended along the greater dislocation length. The repeated pentaring-structured germanium pattern is observed, alongside alternate stretched and compressed

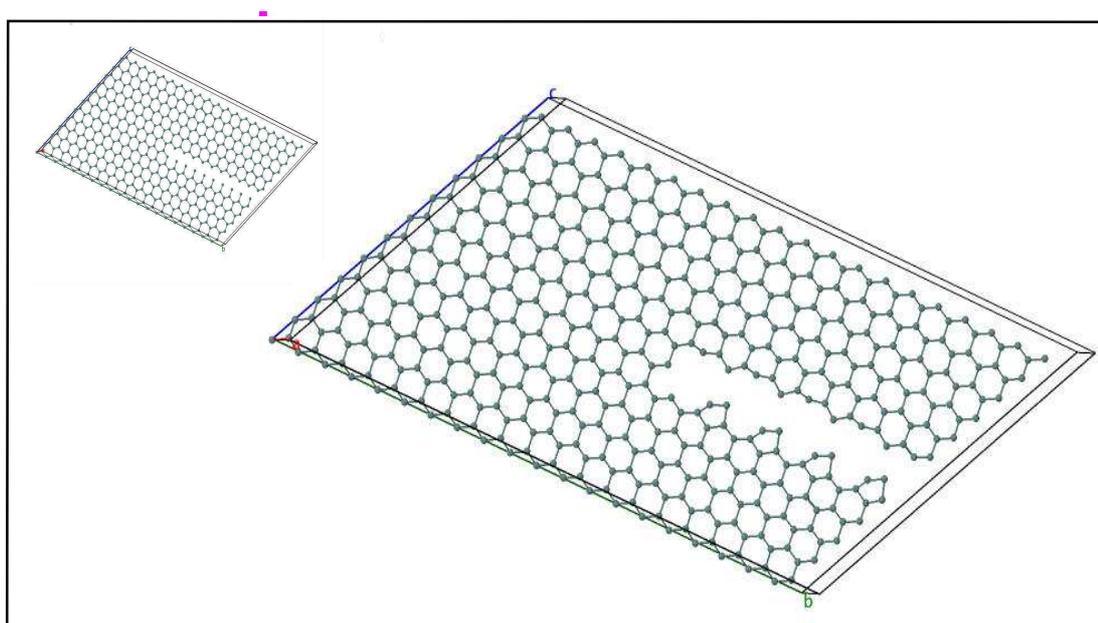


Figure 5.9: Geometry optimised $20 \times 12 \times 1$ supercell of Ge containing a misfit dislocation, with original input cell inset

bonds surrounding the dislocation. The maximum bond angle here was 123° , again, and the minimum was 71° , consistent with the previous results. The maximum bond compression was by 4.5% with a maximum stretch of 2.2%, again, consistent with the results of the $16 \times 10 \times 1$ simulation, and indicating, that, for this dislocation structure in a cell with fixed dimensions, the geometry optimisation results converged at the $20 \times 12 \times 1$ cell size for these simulation parameters.

The structures in these geometry optimisation simulations replicate those observed in a number of other studies on dislocations in diamond structures, including the presence of pentarings and glide and shuffle sets [129, 130, 149–151], and convergence of the structure was observed with increasing cell size, although no complete reconstructions across the dislocation core were observed. This indicates that, within the parameters employed within these simulations, the 20×12 cell represents a fully-relaxed structure; however, the need to maintain the supercell lattice dimensions during geometry optimi-

sation may have induced additional non-physical stresses within the calculation, leading to incomplete reconstruction of the dislocation core and indicating that these may not yet be fully-relaxed structures. Ideally, simulations on larger supercells still should be conducted to identify or confirm a fully-relaxed structure with these parameters.

5.3 Bandstructure Calculations

Electronic bandstructure calculations were performed for the geometry optimised cells, using the parameters described earlier, and taking a k-point pathway along the direction of propagation of the misfit dislocation — equivalent to the [100] direction in the cubic cell. The electronic bandstructure calculation for the $6 \times 8 \times 1$ geometry optimised supercell failed to converge, even after increasing the number of SCF iterations to 200, and decreasing the convergence tolerance to 1×10^{-6} , possibly indicating that the relaxed structure at this cell size represents a cell so disordered that obtaining an energy minimisation and bandstructure is not possible.

5.3.1 The Brillouin Zone and Folded Bandstructures

A supercell bigger than the primitive unit cell for germanium will have a set of reciprocal lattice vectors, and hence a Brillouin Zone, smaller than that of the primitive cell. In bandstructure simulations, the effect of this BZ contraction is folding of the bands at the BZ edge, with the number of folds dependent on the size of the supercell in 3-dimensions relative to the primitive cell. The folded bandstructure for the $12 \times 8 \times 1$ supercell after geometry optimisation is shown in Figure 5.10 alongside the bandstructure between Γ and X for the undefected primitive cell of Ge. The effect of the contracted Brillouin Zone on the simulated bandstructure can clearly be observed: clusters of folded bands obscure any attempt to visualise the conventional E - k relationship, and the identity of any additional bandgap states arising from the dislocation structure is also lost. Recovering an unfolded bandstructure from this folded structure requires mapping of folded band sections back to their origin in the expanded primitive Brillouin Zone. A number

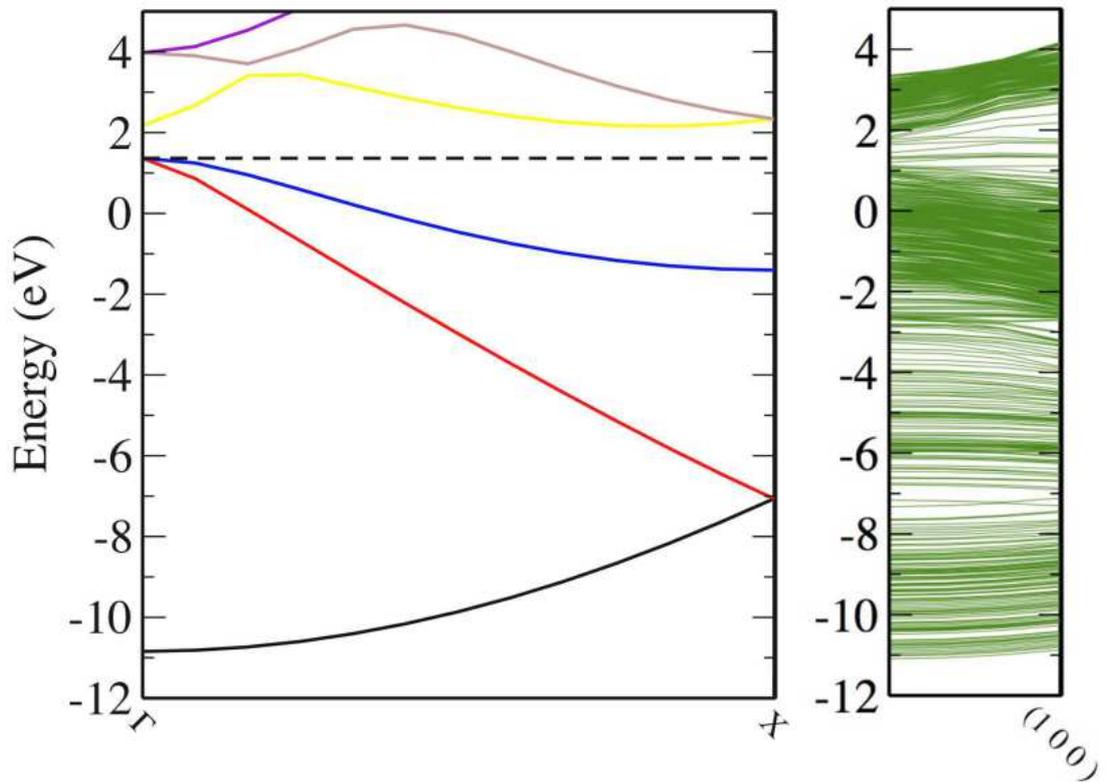


Figure 5.10: Folded bandstructure in the $[100]$ direction for the $12 \times 8 \times 1$ supercell of Ge containing a misfit dislocation after geometry optimisation. The bandstructure for undefected bulk germanium is also presented

of methods for obtaining an unfolded bandstructure from a supercell calculation have been proposed [152–155]. This may be achieved exactly for an undefected supercell, and the bulk Ge bandstructure shown in Figure 5.2 can be recovered exactly for a supercell of bulk Ge simulated with these parameters.

Unfolding of defected cells is more complicated—the defect structures within the supercell mean that these cells are not exact multiples of any primitive unit cell, and translational symmetry is lost. In this case, reproduction of a typical E - k dispersion relationship directly is not possible, however the folded bandstructure can be mapped

probabilistically back to states referenced from a primitive unit cell to produce a ‘effective bandstructure’ (EBS) comparable to that of undefected Ge [155–157].

5.3.2 The Unfolding Algorithm

[155] describes a theoretical method for unfolding bandstructures for alloy materials or materials with defects, to produce an EBS that can be used for comparison with ‘classic’ bandstructure pictures. Using a capitalisation convention wherein the real and reciprocal lattice vectors of the primitive cell are denoted as \mathbf{k} and \mathbf{g} , respectively, and those in the supercell as \mathbf{K} and \mathbf{G} , respectively, the \mathbf{g} -vectors associated with the primitive cell will be a subset of the supercell lattice \mathbf{G} -vectors. A wavevector \mathbf{k} of the primitive cell will fold into a wavevector \mathbf{K} of the supercell if there is a single \mathbf{G} -vector \mathbf{G}_0 which may be subtracted from \mathbf{k} to obtain \mathbf{K} . However, a given wavevector of the supercell \mathbf{K} unfolds into the primitive cell Brillouin zone by addition of a number of possible supercell \mathbf{G} -vectors, to yield a number of possible \mathbf{k} states. Hence, while folding maps a given state in the primitive cell directly to a single state in the supercell, the reverse process of unfolding has no such unique relationship.

A given supercell, with lattice vectors described by $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ is constructed from a set of primitive cell vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ by stacking in 3-dimensions, with a matrix transformation given by:

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} \quad (5.2)$$

where the m_{ij} are obtained from the fact that the supercell lattice vectors should be integer multiples of the primitive cell vectors, and represent the stacking of the unit cell in 3-dimensions to form the supercell. The Brillouin Zone of the supercell may be mapped to the Brillouin Zone of the primitive cell using the inverse matrix, where, similarly $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are the reciprocal lattice vectors of the primitive cell and $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$

are those of the supercell:

$$\begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_3 \end{bmatrix} = M^{-1} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \quad (5.3)$$

In the case used for simulations here, the primitive cell and supercell are collinear, and the matrix will be diagonal.

The unfolding process used here requires calculation of the eigenvalues of both a reference primitive cell (pc) and of the supercell (SC) containing defects. Any SC eigenvector can be expressed as a linear combination of pc eigenvectors, and projection of each supercell eigenstate $\langle \mathbf{K}m |$ onto all primitive Bloch states of a given \mathbf{k} -vector $| \mathbf{k}_i n \rangle$ at the same energy will allow calculation of the amount of ‘Bloch character’ of each pc state $\mathbf{k}_i n$ that is preserved in the SC states $\langle \mathbf{K}m |$ at an energy $E_m = E_n$, or, which of the set of pc $\mathbf{k}_i n$ that map to a single SC $\langle \mathbf{K}m |$ show contributions to an energy eigenvalue. The ‘spectral weight’ $P_{\mathbf{K}m}(\mathbf{k}_i)$ is given by:

$$P_{\mathbf{K}m}(\mathbf{k}_i) = \sum_n \langle \mathbf{K}m | \mathbf{k}_i n \rangle^2 \quad (5.4)$$

and this quantity represents the probability of finding a set of pc states contributing to a given SC state. From here, a spectral function of the energy E as a continuous variable may be calculated, by summing across all energies E_m at each value of \mathbf{k}_i :

$$A(\mathbf{k}_i, E) = \sum_m P_{\mathbf{K}m}(\mathbf{k}_i) \delta(E_m - E) \quad (5.5)$$

to produce an effective $E - k$ bandstructure relationship for the supercell [155, 157].

This method is particularly applicable to SC calculations based on plane-wave pseudopotential methods, where eigenfunctions represented as plane wave coefficients and band energies as eigenvalues are readily accessible. It has hence been implemented as the `bs_sc2pc` tool in `CASTEP`, drawing the relevant information from bandstructure calculations, and has been demonstrated thus far for supercells of defected graphene

[157]. This mapping of the spectral function $A(\mathbf{k}_i, E)$ can provide a visual picture of the bandstructure that can be directly compared to the classic bulk bandstructure, however, here, the objective was to identify individual bandgap states that may indicate the energies of defect levels arising from misfit dislocations in Ge. To achieve this, the `bs_sc2pc` tool was amended at the code level to calculate the spectral weights of individual folded Γ states, and extract potential ‘true’ Γ states from the folded bandstructures for the defected supercells. Broadly, this required identification of the possible pc states k_i that may fold into a specific Γ state K , and unfolding of these K individually into these k_i states, to identify those with a high spectral weight at Γ in the pc, compared to those with a high spectral weight at other pc locations, and hence identify those SC states likely to represent ‘true’ Γ states.

Figure 5.10 shows that there are number of states in the nominal bandgap at Γ , based on the primitive undefected calculation, for this electronic structure calculation. However, these energy levels that appear to exist in the bandgap at Γ may have folded from other regions of k -space. The amended `bs_sc2pc` tool was used to produce a spectral weight for these individual folded Γ states for the 12×8 and 16×10 supercells, with respect to Γ states in the pc. This was achieved using the following procedure: The set of reference primitive-cell \mathbf{k} -points to which the folded Γ states should be matched was obtained by calculating the set of locations in the primitive cell which could fold exactly to Γ in the supercell. In general, for an $x \times y \times z$ supercell, the number of points m which may fold into Γ from the primitive cell is $m = x \times y \times z$. The foldpoints were calculated by establishing which \mathbf{G} -vectors of the reciprocal space of the supercell fell inside the first Brillouin zone of the primitive cell. This may be achieved by calculating \mathbf{G} -vectors from integral numbers of reciprocal lattice vectors of the supercell, and finding the equivalent vectors in the basis of the reciprocal lattice vectors of the primitive cell:

$$\mathbf{b} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \mathbf{B} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix} \quad (5.6)$$

Here, the array N represents iteration over a range of positive and negative integers used to generate the g-vectors of the supercell from the reciprocal lattice vectors B_1, B_2, B_3 of the supercell, given in the matrix \mathbf{B} . Similarly, \mathbf{b} represents the matrix formed from the primitive cell reciprocal lattice vectors, and the vector n_1, n_2, n_3 gives the relative position of each supercell g-vector within the primitive cell reciprocal lattice for any given set of N . Hence:

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \mathbf{b}^{-1}\mathbf{B} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix} \quad (5.7)$$

If the elements n_1, n_2, n_3 are in the range $0 \leq n_x < 1$ then the supercell \mathbf{G} -vector falls within the Brillouin Zone of the primitive cell. The modified `bs_sc2pc` tool was used to match all folded SC Γ states within the bandgap of the pc at Γ to all possible pc eigenstates within this range of \mathbf{G} -vectors.

The range of bands to be unfolded was identified from the bandgap of the primitive cell, which lay between 1.364 eV and 2.173 eV. Using a convention where bands are indexed from 1 for the lowest valence band state, bands 374 to 390 at Γ were unfolded for the geometry optimised 12×8 supercell, with energies from 1.444 to 2.163 eV and bands 624 to 647 were unfolded for the 16×10 supercell with energies from 1.382 to 2.157 eV.

The results of the unfolding calculation on these bandgap states for the 12×8 supercell is shown in Table 5.2. The energy of the band above the valence band edge is provided, as a conventional reference for bandgap states, and the spectral weight of each band at Γ (i.e. the proportion of pc Bloch character at Γ found in each SC Γ state) is identified, alongside an indication of whether the spectral weight at pc Γ was the highest spectral weight found for each bandgap state, or whether higher weights were found for other pc states, indicating that these bandgap states were more likely folded from elsewhere. Band 374, at 0.080 eV above the pc valence band edge, and band 382, at 0.625 eV above

Table 5.2: Unfolding results for the 12×8 supercell

| Band Number | Energy above VB edge eV | Spectral Weight at Γ | Highest at Γ ? |
|-------------|-------------------------|-----------------------------|-----------------------|
| 374 | 0.0795 | 0.01147 | YES |
| 375 | 0.2794 | 0.00102 | no |
| 376 | 0.3300 | 0.00511 | no |
| 377 | 0.4029 | 0.00329 | no |
| 378 | 0.4343 | 0.00128 | no |
| 379 | 0.4788 | 0.00103 | no |
| 380 | 0.5780 | 0.00106 | no |
| 381 | 0.5950 | 0.00726 | no |
| 382 | 0.6246 | 0.05204 | YES |
| 383 | 0.6522 | 0.00359 | no |
| 384 | 0.6722 | 0.00497 | no |
| 385 | 0.7008 | 0.00492 | no |
| 386 | 0.7212 | 0.00149 | no |
| 387 | 0.7358 | 0.00089 | no |
| 388 | 0.7446 | 0.00061 | no |
| 389 | 0.7566 | 0.00284 | no |
| 390 | 0.7985 | 0.00041 | no |

the valence band edge (or 0.183 eV below the conduction band edge at Γ , exhibited their highest spectral weights at Γ , possibly indicating that these may be ‘true’ Γ states. For other bands, spectral weights were often widely distributed, with results suggesting a number of possible other pc locations for the origin of the folded Γ states. For example, four different pc locations away from Γ were the highest-weighted states for band 381, with equal spectral weights each of 0.05148, and, similarly, for both bands 378 and 385, four possible pc states away from Γ were suggested with spectral weights greater than 0.028. It was noted in the geometry optimisation results, however, that the 12×8 supercell is unlikely to represent a fully-relaxed structure, and the analysis of the mapping of SC Γ states to pc Γ locations was repeated for the 16×10 supercell. The results are shown in Table 5.3. Here, the highest-weighted Γ states were found at band 625, at 0.069 eV above the valence band edge, and band 644, at 0.755 eV above the valence band edge (or 0.053 eV below the conduction band edge). There is some consistency in the results — both cells exhibited two likely ‘true’ Γ states, one close to the valence band edge and one closer to the conduction band edge — however there was a significant discrepancy in the energies of these states between the two cells. This is perhaps to be expected, as no convergence in geometry had been identified between these two cells based on earlier results, and this is likely to lead to significant discrepancies in electronic structure calculation results. In addition, [130] notes that insufficiently large supercells generate artificial stresses that lead to additional stress-induced electronic bands, and it is likely that this effect is present in the 12×8 supercell, at least.

Convergence in these results could be demonstrated through completing the analysis using the geometry-optimised 20×12 supercell, however, the cell of this size presented a significant computational burden to the CASTEP software, and it was not possible to produce and electronic bandstructure for this cell size using this method.

Comparison with other results obtained for bandgap states due to defects in Ge shows limited consistency. [136] identified defect levels in germanium at 0.05 eV and 0.11 eV

Table 5.3: Unfolding results for the 16×10 supercell

| Band Number | Energy above VB edge eV | Spectral Weight at Γ |
|-------------|-------------------------|-----------------------------|
| 624 | 0.0183 | 0.00222 |
| 625 | 0.0691 | 0.01988 |
| 626 | 0.0940 | 0.00754 |
| 627 | 0.2674 | 0.00109 |
| 628 | 0.2707 | 0.00096 |
| 629 | 0.3806 | 0.00028 |
| 630 | 0.4043 | 0.00016 |
| 631 | 0.4168 | 0.00022 |
| 632 | 0.4391 | 0.00089 |
| 633 | 0.4613 | 0.00016 |
| 634 | 0.4905 | 0.00013 |
| 635 | 0.5255 | 0.00221 |
| 636 | 0.5431 | 0.00006 |
| 637 | 0.6279 | 0.00052 |
| 638 | 0.6453 | 0.00448 |
| 639 | 0.6539 | 0.00008 |
| 640 | 0.6681 | 0.00325 |
| 641 | 0.6714 | 0.00012 |
| 642 | 0.6755 | 0.00119 |
| 643 | 0.7211 | 0.00022 |
| 644 | 0.7545 | 0.07083 |
| 645 | 0.7620 | 0.00238 |
| 646 | 0.7870 | 0.00004 |
| 647 | 0.7922 | 0.00722 |

above the valence band edge at the Γ point for hydrogen-passivated dangling bonds, and the results for both supercells here identified a state close to the valence band edge, with that at 0.069 eV for the 16×10 supercell lying closest to this previous result. It is notable that the increase in cell size from 12×8 to 16×10 had the effect of lowering the energy of the state nearest the valence band edge and increasing the energy of the state closest to the conduction band edge, and it would be useful to investigate whether this separation of states, and identification of states closer to the band edges is observed for larger supercell sizes.

The investigation of bandgap states in diamond with defects presented in [129] indicated that mixed-type dislocations, with reconstruction of the dislocation core, yielded stress-induced bandgap states close to the band edges, while pure edge dislocations containing unsaturated dangling bonds lead to more deep states within the centre of the bandgap. It is possible that with a larger supercell size, further dislocation core relaxation and pairing of dangling bonds during geometry optimisation may lead to a similar result, of fewer deep states, and defect states closer to the band edges. [158] also provides electronic structure results for diamond, and a similar pattern is observed – that poorly-reconstructed dislocations with dangling and stretched bonds yield deep states close the centre of the bandgap, while reconstructions wherein atomic co-ordination numbers are maintained demonstrate perturbations in electronic structure close to the band edges. The method described here should ideally be applied to larger Ge cells containing a misfit dislocation to confirm this effect, and establish the electronic structure in the bandgap for defect densities approaching physically-realistic values.

Although reasonable-to-good parallelisation efficiencies were often obtained in the calculations, simulation of cells larger than the 16×10 defected cell used here were not possible due to the excessive computational time, and limitations to the amount of parallel computing resources available. The use of the pseudopotential discussed above was extremely computationally efficient, and, allowed calculations up to this cell-size.

In addition, development of the CASTEP software has proceeded in a number of directions recently to address the need for greater memory efficiency and to calculate larger supercells, including implementation of process threading, band parallelism and more efficient memory allocation, and these developments, in combination with greater supercomputing power may enable the method to be developed to simulate larger defected supercells to identify convergence in the calculation of bandgap states at Γ or simulate SiGe interfaces containing defects.

Future work could progress these larger supercells or interfaces, and continue to investigate the utility of the unfolding method provided in [156] and [157], and the `bs_sc2pc` tool implementation in CASTEP for investigation of the gap states arising from the defects inherent in SiGe epitaxy.

5.4 Conclusion

In this chapter, DFT implemented using the CASTEP software was used to perform geometry optimisation on supercells of germanium containing a misfit dislocation, to obtain a minimum-energy structure for the defects, and to calculate bandstructures for these relaxed cells. The simulations were calibrated against a number of criteria, and simulation parameters were selected based on those which gave the best agreement between calculated and experimental values for a primitive cell of bulk germanium, while maintaining attainable computational efficiency.

Geometry optimisation yielded structures with some deviation from the manually-placed defect cells, including both distortions of the crystal structure — changes in bond angle and length — and a number of other structures including pentarings and glide and shuffle structures, which broadly correlated with previous experimental results.

An unfolding algorithm described by [156] for alloys and implemented in CASTEP as

the `bs_sc2pc` tool was used to unfold the defected supercell bandstructures, and identify the likelihood that states found at Γ in the folded supercell bandstructure represented ‘true’ Γ bandgap states in the unfolded regime. A number of gap states at Γ were identified and compared to experimental values, and some consistency was observed with increasing supercell size: both calculations yielded two likely ‘true’ Γ states, one close to the valence band edge and one closer to the conduction band edge. However, the energies of these states were not highly consistent with each other or with previous experimental results, and it is likely that the method needs to be applied to larger supercell sizes to confirm the relaxation of the dislocation structure and identify bandgap states arising from dislocations in Ge.

Trap states contribute to afterpulsing in SPADs, by providing sites for carriers in the bandgap which may detrapp outside of a detection event, and should be included in device simulations. This chapter investigates a preliminary method for identifying such states, however, the cell sizes used here may be insufficiently large to provide a physically-realistic separation between adjacent defects, and the method should ideally be progressed with bigger supercell sizes and SiGe interfaces to establish the utility of the CASTEP software (in light of further developments in memory and parallelism capacity) and the `bs_sc2pc` tool for identifying bandstructure effects in Ge and SiGe interfaces containing misfit dislocations.

Chapter 6

Conclusion

A range of simulations effecting electron transport in SPADs has been investigated across a range of scales from simulation of full-device-thicknesses, to *ab initio* calculations of defect states in bulk Ge. Doping profiles in the charge and absorption layers were simulated, to investigate the effect on IV curves, electric field and potential profiles and carrier concentrations, to establish ideal intentional and background doping types and densities for these layers, to promote carrier transport and suppress noise currents. The effect of increasing bias voltage, layer thickness and temperature on band-to-band tunnelling currents was simulated, and compared to calculations using other methods, and to other sources of noise currents in SPADs, to establish the likely contribution of these parameters to noise currents in SPADs. Finally, DFT calculations of the electronic structure of misfit dislocations in bulk Ge, as well as symmetric energy minima for these defects was simulated, to gain insight into the energies of trap states introduced by defects, which are significant in dark counts and particularly afterpulsing effects in SPADs.

Simulations on SPADs structures using the ATLAS solver within the Silvaco software [104] It was found that breakdown voltage increases with decreasing charge layer doping, which is to be expected since the a higher charge layer doping density increases the overall proportion of the electric field across the device which occurs across the multiplication

layer, and hence decreases the voltage required to achieve breakdown. For high charge layer doping, the type of absorption layer doping played no significant role in the breakdown voltage, however at lower charge layer doping densities, p -type absorption layer doping results in a lower breakdown voltage than n -type, because different doping types produce opposite field gradients in the absorber. Punchthrough, indicating full depletion of the device, and removal of the potential well for carriers at the interface was observed for a number of doping profiles. Ideally, device doping should facilitate punchthrough, as full depletion of the device enables operation as a SPAD, where the only carriers should be generated through incidence of an incoming photon — carriers remaining within the device will contribute to the DCR. From this results, the ideal doping profile for a SPAD would be one with a charge layer doping density of $2 \times 10^{17} \text{cm}^{-2}$ and an absorption layer doping density of $1 \times 10^{15} \text{cm}^{-2}$) of n -type, to limit dark counts due to excess fields in the germanium, while ensuring full depletion of the absorber before breakdown. Fabrication of SiGe SPADs is, however, difficult, and experimental measurements frequently varied significantly from the simulated characteristics, indicating that target dopant profiles had not been achieved.

The dependence of band-to-band tunnelling currents in Ge on bias voltage, sample thickness and temperature was simulated, using an empirical pseudopotential method and a band-matching technique. Band-to-band tunnelling currents were found to increase exponentially with bias voltage, indicating that operation of a SPAD at increasing reverse biases will result in increasing noise current contributions from band-to-band tunnelling. A breakdown of tunnelling currents between different regions in the Brillouin Zone indicated that, at lower biases, tunnelling was entirely between valence band Γ and conduction band Γ , whereas at higher biases, a further contribution from valence band X to conduction band Γ was observed, in keeping with the slightly larger energy gap for these transitions. Other transitions made only small contributions. Tunnelling currents increased with increasing thickness, since the number of available states for tunnelling increases with sample thickness, this was confirmed by the increasing contributions from

deeper valence band states with increasing thickness. The tunnelling current may be expected to saturate with thickness, since at some point no further states for tunnelling will be available, and transmission currents will decrease with increasing germanium thickness. Tunnelling currents were found to increase with increasing temperature — equivalent, in this method, to a decreasing bandgap. Calculated tunnelling currents were composed almost entirely of valence band Γ to conduction band Γ transitions. Comparison of simulated band-to-band tunnelling currents indicated that, at voltages below breakdown, band-to-band tunnelling was likely to be insignificant relative to avalanche currents and dark currents. However, at fields considerably higher than breakdown — above 150% of the breakdown field — the simulated band-to-band currents begin to approach the order of these other currents, indicating that at voltages significantly higher than breakdown, band-to-band tunnelling may result in noise currents. However, at lower voltages, it is likely that noise currents arise from sources other than band-to-band tunnelling.

Defects in semiconductor materials induce perturbations to bandstructures that can lead to additional energy states in the bandgap, which may contribute to afterpulsing in SPADs. DFT implemented using the CASTEP software was used to perform geometry optimisation on supercells of germanium containing a misfit dislocation, to obtain a minimum-energy structure for the defects, and to calculate bandstructures for these relaxed cells. A number of gap states at Γ were identified and compared to experimental values, and some consistency was observed with increasing supercell size, however, the method should be applied to larger supercell sizes to identify bandgap states arising from dislocations in Ge at physically-realistic densities. Additional energy states in the bandgap can contribute to afterpulsing in SPADs, by providing sites for carriers in the bandgap which may detrapp outside of a detection event, and should be included in device simulations.

The work presented here addresses a number of challenges and performance param-

eters in development of SiGe SPAD performance: control of field properties and carrier behaviour through doping profiles across the device, calculation of the voltage, thickness and temperature dependence of band-to-band tunnelling, and its contribution to overall noise currents, and development of a method for establishing the energies of trap states in Ge, which contribute to afterpulsing. Further work can progress in a number of directions. The *ab initio* calculations of the energies of bandgap states in Ge should be applied to larger cell sizes to establish whether the geometries, dislocation reconstructions and energies of the bandgap states converge with increasing cell size and smaller dislocation densities. Further studies could also focus on defect structures in a SiGe interface model. A number of alternative calculation methods could be investigated — though DFT provides one theoretical approach, it is computationally expensive, even with the use of a very efficient pseudopotential, and other methods, including empirical pseudopotential methods, may yield accurate results for larger cell sizes, approaching those needed for physically-realistic defect densities. Definitive values for bandgap energies induced by defects will enhance the full-device model considered in Chapter 3, and allow trap energies to be incorporated into the model, as well as into other studies of devices based on epitaxially grown germanium.

Band-to-band tunnelling has been shown here to provide only a small contribution to noise currents at lower bias voltages. However, the simulations considered only undoped Ge, and hence little change in Fermi occupancy from 1 in the valence band to 0 in the conduction band. A more complex model could consider the effect of varying doping types and densities, incorporating the effects of both changing occupancies and of doping on the electric field across the sample. This work is also applicable to other devices, particularly tunnel FETs, and the model may be extended to consider band-to-band tunnelling currents in these structures. This would require modelling of a *pn* or *pin* structure, and require a 2- or 3-dimensional transport and electrostatics simulation to determine potential profiles. Here, band-to-band tunnelling would depend on gate bias as well as source-drain bias.

More investigation into other sources of noise currents would progress understanding of SPAD performance. Avalanche breakdown is of particular interest, and calculations of impact ionisation for Ge — which has been only briefly studied compared to the same phenomenon in Si — would be a useful theoretical area for future study. The full-zone empirical pseudopotential method for bandstructure calculations used in this thesis could be applied to calculations of impact ionisation in Ge. Additionally, edge breakdown has been suggested as a possible source of noise currents in SPADs, and further investigations into this effect should be undertaken.

References

- [1] G. S. Buller and R. J. Collins, “Single-photon generation and detection,” *Measurement Science and Technology*, vol. 21, p. 012002, Jan. 2010.
- [2] S. Pellegrini, R. Warburton, L. Tan, J. Ng, A. Krysa, K. Groom, J. David, S. Cova, M. Robertson, and G. Buller, “Design and Performance of an InGaAs-InP Single-Photon Avalanche Diode Detector,” *IEEE Journal of Quantum Electronics*, vol. 42, pp. 397–403, Apr. 2006.
- [3] C. Zener, “A Theory of the Electrical Breakdown of Solid Dielectrics,” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 145, pp. 523–529, July 1934.
- [4] J. P. Sprengers, A. Gaggero, D. Sahin, S. Jahanmirinejad, G. Frucci, F. Mattioli, R. Leoni, J. Beetz, M. Lermer, M. Kamp, S. Hfling, R. Sanjines, and A. Fiore, “Waveguide superconducting single-photon detectors for integrated quantum photonic circuits,” *Applied Physics Letters*, vol. 99, p. 181110, Oct. 2011.
- [5] R. H. Hadfield, “Single-photon detectors for optical quantum information applications,” *Nature Photonics*, vol. 3, pp. 696–705, Dec. 2009.
- [6] J. L. O’Brien, “Optical Quantum Computing,” *Science*, vol. 318, pp. 1567–1570, July 2007.
- [7] E. Knill, R. Laflamme, and G. J. Milburn, “A scheme for efficient quantum computation with linear optics,” *Nature*, vol. 409, pp. 46–52, Jan. 2001.

- [8] H. J. Kimble, "The quantum internet," *Nature*, vol. 453, pp. 1023–1030, June 2008.
- [9] T. Jennewein, M. Barbieri, and A. G. White, "Single-photon device requirements for operating linear optics quantum computing outside the post-selection basis," *Journal of Modern Optics*, vol. 58, pp. 276–287, Feb. 2011.
- [10] D. Schwartz, E. Charbon, and K. L. Shepard, "A Single-Photon Avalanche Diode Array for Fluorescence Lifetime Imaging Microscopy," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 2546–2557, Nov. 2008.
- [11] D. Palubiak, M. El-Desouki, O. Marinov, M. Deen, and Q. Fang, "High-Speed, Single-Photon Avalanche-Photodiode Imager for Biomedical Applications," *IEEE Sensors Journal*, vol. 11, pp. 2401–2412, Oct. 2011.
- [12] S. Kumar, C. Dunsby, P. A. A. De Beule, D. M. Owen, U. Anand, P. M. P. Lanigan, R. K. P. Benninger, D. M. Davis, M. A. A. Neil, P. Anand, C. Benham, A. Naylor, and P. M. W. French, "Multifocal multiphoton excitation and time correlated single photon counting detection for 3-D fluorescence lifetime imaging," *Optics Express*, vol. 15, no. 20, p. 12548, 2007.
- [13] H. E. Grecco, P. Roda-Navarro, and P. J. Verveer, "Global analysis of time correlated single photon counting FRET-FLIM data," *Optics Express*, vol. 17, p. 6493, Apr. 2009.
- [14] W. Becker, A. Bergmann, M. Hink, K. Knig, K. Benndorf, and C. Biskup, "Fluorescence lifetime imaging by time-correlated single-photon counting," *Microscopy Research and Technique*, vol. 63, pp. 58–66, Jan. 2004.
- [15] G. S. Buller and A. Wallace, "Ranging and Three-Dimensional Imaging Using Time-Correlated Single-Photon Counting and Point-by-Point Acquisition," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, pp. 1006–1015, July 2007.

- [16] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernandez, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Applied Optics*, vol. 48, p. 6241, Nov. 2009.
- [17] F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements," *IEEE Photonics Journal*, vol. 4, pp. 795–804, June 2012.
- [18] R. E. Warburton, A. McCarthy, A. M. Wallace, S. Hernandez-Marin, R. H. Hadfield, S. W. Nam, and G. S. Buller, "Subcentimeter depth resolution using a single-photon counting time-of-flight laser ranging system at 1550 nm wavelength," *Optics Letters*, vol. 32, no. 15, p. 2266, 2007.
- [19] D. Bonneau, E. Engin, K. Ohira, N. Suzuki, H. Yoshida, N. Iizuka, M. Ezaki, C. M. Natarajan, M. G. Tanner, R. H. Hadfield, S. N. Dorenbos, V. Zwiller, J. L. O'Brien, and M. G. Thompson, "Quantum interference and manipulation of entanglement in silicon wire waveguide quantum circuits," *New Journal of Physics*, vol. 14, p. 045003, Apr. 2012.
- [20] K. Ohashi, K. Nishi, T. Shimizu, M. Nakada, J. Fujikata, J. Ushida, S. Torii, K. Nose, M. Mizuno, H. Yukawa, M. Kinoshita, N. Suzuki, A. Gomyo, T. Ishi, D. Okamoto, K. Furue, T. Ueno, T. Tsuchizawa, T. Watanabe, K. Yamada, S. Itabashi, and J. Akedo, "On-Chip Optical Interconnect," *Proceedings of the IEEE*, vol. 97, pp. 1186–1198, July 2009.
- [21] G. T. Reed, W. R. Headley, G. Z. Mashanovic, F. Y. Gardes, D. J. Thomson, and M. M. Milosevic, "Silicon Photonics - The Evolution of Integration," in *Silicon Photonics for Telecommunications and Biomedicine*, CRC, 2011.
- [22] D. Marris-Morini, L. Vivien, G. Rasigade, J.-M. Fedeli, E. Cassan, X. Le Roux, P. Crozat, S. Maine, A. Lupu, P. Lyan, P. Rivallin, M. Halbwax, and S. Laval,

- “Recent Progress in High-Speed Silicon-Based Optical Modulators,” *Proceedings of the IEEE*, vol. 97, pp. 1199–1215, July 2009.
- [23] L. Tsybeskov, D. Lockwood, and M. Ichikawa, “Silicon Photonics: CMOS Going Optical [Scanning the Issue],” *Proceedings of the IEEE*, vol. 97, pp. 1161–1165, July 2009.
- [24] “Single-Photon Generation and Detection, 1st Edition | Alan Migdall, Sergey Polyakov, Jingyun Fan, Joshua Bienfang | ISBN 9780123876966.”
- [25] R. Abbasi et al, “Calibration and characterization of the IceCube photomultiplier tube,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 618, pp. 139–152, June 2010.
- [26] P. H. MacArthur, S. Shiva, and M. T. Gladwin, “Measurement of circulating nitrite and S-nitrosothiols by reductive chemiluminescence,” *Journal of Chromatography B*, vol. 851, pp. 93–105, May 2007.
- [27] P. K. Chattopadhyay, D. A. Price, T. F. Harper, M. R. Betts, J. Yu, E. Gostick, S. P. Perfetto, P. Goepfert, R. A. Koup, S. C. De Rosa, M. P. Bruchez, and M. Roederer, “Quantum dot semiconductor nanocrystals for immunophenotyping by polychromatic flow cytometry,” *Nature Medicine*, vol. 12, pp. 972–977, Aug. 2006.
- [28] C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, “Superconducting nanowire single-photon detectors: physics and applications,” *Superconductor Science and Technology*, vol. 25, p. 063001, June 2012.
- [29] E. A. Dauler, M. E. Grein, A. J. Kerman, F. Marsili, S. Miki, S. W. Nam, M. D. Shaw, H. Terai, V. B. Verma, and T. Yamashita, “Review of superconducting nanowire single-photon detector system design options and demonstrated performance,” *Optical Engineering*, vol. 53, no. 8, pp. 081907–081907, 2014.

-
- [30] S. Chen, L. You, W. Zhang, X. Yang, H. Li, L. Zhang, Z. Wang, and X. Xie, "Dark counts of superconducting nanowire single-photon detector under illumination," *Optics Express*, vol. 23, p. 10786, Apr. 2015.
- [31] H. Shibata, K. Shimizu, H. Takesue, and Y. Tokura, "Ultimate low system dark-count rate for superconducting nanowire single-photon detector," *Optics Letters*, vol. 40, p. 3428, July 2015.
- [32] M. Fukuda, *Optical Semiconductor Devices*. Wiley Series in Microwave and Optical Engineering, John Wiley & Sons, 1999.
- [33] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. John Wiley & Sons, 2007.
- [34] M. Lpez, H. Hofer, and S. Kck, "Detection efficiency calibration of single-photon silicon avalanche photodiodes traceable using double attenuator technique," *Journal of Modern Optics*, vol. 62, pp. S21–S27, Dec. 2015.
- [35] A. Lacaita, M. Mastrapasqua, M. Ghioni, and S. Vanoli, "Observation of avalanche propagation by multiplication assisted diffusion in p-n junctions," *Applied Physics Letters*, vol. 57, no. 5, p. 489, 1990.
- [36] Y. Liang, Y. Jian, X. Chen, G. Wu, E. Wu, and H. Zeng, "Room-Temperature Single-Photon Detector Based on InGaAs/InP Avalanche Photodiode With Multichannel Counting Ability," *IEEE Photonics Technology Letters*, vol. 23, pp. 115–117, Jan. 2011.
- [37] X. Wang, G. Zhang, R. Chen, Y. Gao, L. Xiao, and S. Jia, "Detection efficiency enhancement of single-photon detector at 1.55- μm by using of single photons lock-in and optimal threshold," *Optics & Laser Technology*, vol. 44, pp. 1773–1775, Sept. 2012.
- [38] J. Cheng, S. You, S. Rahman, and Y.-H. Lo, "Self-quenching InGaAs/InP single

- photon avalanche detector utilizing zinc diffusion rings,” *Optics Express*, vol. 19, p. 15149, July 2011.
- [39] M. A. Itzler, X. Jiang, M. Entwistle, K. Slomkowski, A. Tosi, F. Acerbi, F. Zappa, and S. Cova, “Advances in InGaAsP-based avalanche diode single photon detectors,” *Journal of Modern Optics*, vol. 58, no. 3-4, pp. 174–200, 2011.
- [40] R. Warburton, G. Intermite, M. Myronov, P. Allred, D. Leadley, K. Gallacher, D. Paul, N. Pilgrim, L. Lever, Z. Ikonc, R. Kelsall, E. Huante-Ceron, A. Knights, and G. Buller, “Ge-on-Si Single-Photon Avalanche Diode Detectors: Design, Modeling, Fabrication, and Characterization at Wavelengths 1310 and 1550 nm,” *IEEE Transactions on Electron Devices*, vol. 60, pp. 3807–3813, Nov. 2013.
- [41] Wikipedia contributors, “pn junction.” <https://bit.ly/2v5XIO3> Page Version ID: 503714556 Accessed 24th March 2012.
- [42] S. Dimitrijević, *Principles of Semiconductor Devices*. Oxford University Press, 2012.
- [43] S. L. Chuang, *Physics of Optoelectronic Devices*. Wiley, Sept. 1995.
- [44] M. Levinshtein, J. Kostamovaara, and S. Vainshtein, *Breakdown phenomena in semiconductors and semiconductor devices*. World Scientific, Nov. 2005.
- [45] C. K. Chia, “Numerical simulation of impact ionization in Ge/Al_xGa_{1-x}As avalanche photodiode,” *Applied Physics Letters*, vol. 97, pp. 073501–073501–3, Aug. 2010.
- [46] J. Kolnk, Y. Wang, I. H. Ouzman, and K. F. Brennan, “Theoretical investigation of wavevectordependent analytical and numerical formulations of the interband impactionization transition rate for electrons in bulk silicon and GaAs,” *Journal of Applied Physics*, vol. 76, pp. 3542–3551, Sept. 1994.
- [47] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi, “A Monte Carlo simulation of anisotropic elec-

- tron transport in silicon including full band structure and anisotropic impact ionization model,” *Journal of Applied Physics*, vol. 75, pp. 297–312, Jan. 1994.
- [48] T. Kunikiyo, M. Takenaka, M. Morifuji, K. Taniguchi, and C. Hamaguchi, “A model of impact ionization due to the primary hole in silicon for a full band Monte Carlo simulation,” *Journal of Applied Physics*, vol. 79, pp. 7718–7725, May 1996.
- [49] J. S. Marsland, “Impact ionization in semiconductors: recent progress on non-local effects RID E-9948-2010,” *Physica Status Solidi a-Applications and Materials Science*, vol. 206, pp. 874–879, May 2009. WOS:000266472100020.
- [50] A. Spinelli, A. Pacelli, and A. Lacaita, “Dead space approximation for impact ionization in silicon RID A-5675-2008,” *Applied Physics Letters*, vol. 69, pp. 3707–3709, Dec. 1996. WOS:A1996VW75000029.
- [51] B. Van Zeghbroeck, *Principles of Semiconductor Devices*. New Jersey: Prentice Hall, 2011.
- [52] Ioffe Institute, “Electrical Properties of Germanium,” 2015. <https://bit.ly/2Hdogwi> Accessed 17th March 2015.
- [53] E. O. Kane, “Theory of Tunneling,” *Journal of Applied Physics*, vol. 32, no. 1, p. 83, 1961.
- [54] N. W. Ashcroft and N. D. Mermin, *Solid state physics*. Rochester: Saunders College, 1976.
- [55] A. Valavanis, “n-type Silicon-Germanium based Terahertz Quantum Cascade Lasers,” 2009.
- [56] D. J. , “Si/SiGe heterostructures: from material and physics to devices and circuits,” *Semiconductor Science and Technology*, vol. 19, p. R75, Oct. 2004.
- [57] S. Sivaram, *Chemical Vapor Deposition*. Boston, MA: Springer US, 1995.

- [58] G. Isella, D. Chrastina, B. Rssner, T. Hackbarth, H. J. Herzog, U. Knig, and H. von Knel, "Low-energy plasma-enhanced chemical vapor deposition for strained Si and Ge heterostructures and devices," *Solid-State Electronics*, vol. 48, pp. 1317–1323, Aug. 2004.
- [59] C. Rosenblad, H. R. Deller, A. Dommann, T. Meyer, P. Schroeter, and H. v. Knel, "Silicon epitaxy by low-energy plasma enhanced chemical vapor deposition," *Journal of Vacuum Science & Technology A*, vol. 16, pp. 2785–2790, Sept. 1998.
- [60] M. Kummer, C. Rosenblad, A. Dommann, T. Hackbarth, G. Hck, M. Zeuner, E. Mller, and H. von Knel, "Low energy plasma enhanced chemical vapor deposition," *Materials Science and Engineering: B*, vol. 89, pp. 288–295, Feb. 2002.
- [61] M. Bosi and G. Attolini, "Germanium: Epitaxy and its applications," *Progress in Crystal Growth and Characterization of Materials*, vol. 56, pp. 146–174, Sept. 2010.
- [62] M. J. Cooke, *Semiconductor Devices*. New York: Pearson Education Limited, Oct. 1990.
- [63] G. Abstreiter, K. Eberl, E. Friess, W. Wegscheider, and R. Zachai, "Silicon/germanium strained layer superlattices," *Journal of Crystal Growth*, vol. 95, pp. 431–438, Feb. 1989.
- [64] C. K. Maiti, L. K. Bera, S. Maikap, S. K. Ray, N. B. Chakrabarti, R. Kesavan, and V. Kumar, "Growth of Silicon- Germanium Alloy Layers," *Defence Science Journal*, vol. 50, pp. 299–315, Jan. 2013.
- [65] R. People and J. C. Bean, "Calculation of critical layer thickness versus lattice mismatch for $\text{Ge}_x\text{Si}_{1-x}/\text{Si}$ strainedlayer heterostructures," *Applied Physics Letters*, vol. 47, pp. 322–324, Aug. 1985.
- [66] V. Shah, A. Dobbie, M. Myronov, and D. Leadley, "Effect of layer thickness on

- structural quality of Ge epilayers grown directly on Si(001),” *Thin Solid Films*, vol. 519, pp. 7911–7917, Sept. 2011.
- [67] L. Colace and G. Assanto, “Germanium on Silicon for Near-Infrared Light Sensing,” *IEEE Photonics Journal*, vol. 1, pp. 69–79, Aug. 2009.
- [68] J. M. Hartmann, A. Abbadie, A. M. Papon, P. Holliger, G. Rolland, T. Billon, J. M. Fdli, M. Rouvire, L. Vivien, and S. Laval, “Reduced pressurechemical vapor deposition of Ge thick layers on Si(001) for 1.31.55- μ m photodetection,” *Journal of Applied Physics*, vol. 95, pp. 5905–5913, May 2004.
- [69] D. Hull, *Introduction to dislocations*. Oxford: Butterworth-Heinemann, 4th ed ed., 2001.
- [70] J. E. Ayers, *Heteroepitaxy of semiconductors: theory, growth, and characterization*. Boca Raton: CRC Press, 2007.
- [71] ESA, “Screw Dislocation.” <https://bit.ly/2HvkCvC>, Accessed 26th July 2012.
- [72] ESA, “Edge Dislocation.” <https://bit.ly/2HtbpDQ> Accessed 26th July 2012.
- [73] Wikipedia contributors, “Burgers vector.” <https://bit.ly/2GVWVLu> Page Version ID: 675053228 Accessed 17th August 2017.
- [74] D. Paul, “Misfit Dislocation.” <https://bit.ly/2HvuS6I> Accessed 22nd January 2012.
- [75] V. A. Shah, A. Dobbie, M. Myronov, and D. R. Leadley, “High quality relaxed Ge layers grown directly on a Si(0 0 1) substrate,” *Solid-State Electronics*, vol. 62, pp. 189–194, Aug. 2011.
- [76] T. Figielski, “Dislocations as electrically active centres in semiconductors - half a century from the discovery,” *Journal of Physics-Condensed Matter*, vol. 14, pp. 12665–12672, Dec. 2002. WOS:000180091100002.

- [77] R. M. Nieminen, “Issues in first-principles calculations for defects in semiconductors and oxides,” *Modelling and Simulation in Materials Science and Engineering*, vol. 17, no. 8, p. 084001, 2009.
- [78] D. Holt and B. Yacobi, “Extended Defects in Semiconductors - Electronic Properties, Device Effects and Structures - 3.3 Light Microscopy - Knovel.”
- [79] W. Read, “LXXXVII. Theory of dislocations in germanium,” *Philosophical Magazine Series 7*, vol. 45, no. 367, pp. 775–796, 1954.
- [80] K. D. Usadbl and W. Schrter, “The influence of hybridization on the electronic states at dislocations in semiconductors,” *Philosophical Magazine Part B*, vol. 37, pp. 217–232, Feb. 1978.
- [81] R. Jaszek, “Carrier scattering by dislocations in semiconductors,” *Journal of Materials Science: Materials in Electronics*, vol. 12, pp. 1–9, Jan. 2001.
- [82] D. Vasileska and S. M. Goodnick, “Computational Electronics,” *Synthesis Lectures on Computational Electromagnetics*, vol. 1, pp. 1–216, Jan. 2006.
- [83] J. R. Hook and H. E. Hall, *Solid state physics*. Wiley, July 1991.
- [84] M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, “First-principles simulation: ideas, illustrations and the CASTEP code,” *Journal of Physics: Condensed Matter*, vol. 14, pp. 2717–2744, Mar. 2002.
- [85] G. P. Srivastava, *Theoretical Modelling of Semiconductor Surfaces: Microscopic Studies of Electrons and Phonons*. World Scientific, 1999.
- [86] M. L. Cohen, “Electronic structure of solids,” *Physics Reports*, vol. 110, pp. 293–309, Aug. 1984.
- [87] H. Hellmann, “A New Approximation Method in the Problem of Many Electrons,” *The Journal of Chemical Physics*, vol. 3, pp. 61–61, Jan. 1935.

- [88] C. Herring, "A New Method for Calculating Wave Functions in Crystals," *Physical Review*, vol. 57, pp. 1169–1177, June 1940.
- [89] P. Schwerdtfeger, "The Pseudopotential Approximation in Electronic Structure Theory," *ChemPhysChem*, vol. 12, pp. 3143–3155, Dec. 2011.
- [90] J. C. Phillips and L. Kleinman, "New Method for Calculating Wave Functions in Crystals and Molecules," *Physical Review*, vol. 116, pp. 287–294, Oct. 1959.
- [91] R. Magri, "Pseudopotentials for Band Structure Calculations," <https://bit.ly/2qqRZrf> Accessed 19th March 2015.
- [92] D. R. Hamann, M. Schlter, and C. Chiang, "Norm-Conserving Pseudopotentials," *Physical Review Letters*, vol. 43, pp. 1494–1497, Nov. 1979.
- [93] D. J. Singh and L. Nordstrom, *Planewaves, Pseudopotentials, and the LAPW Method*. Springer Science & Business Media, Oct. 2006.
- [94] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," *Physical Review B*, vol. 41, pp. 7892–7895, Apr. 1990.
- [95] M. L. Cohen and T. K. Bergstresser, "Band Structures and Pseudopotential Form Factors for Fourteen Semiconductors of the Diamond and Zinc-blende Structures," *Physical Review*, vol. 141, pp. 789–796, Jan. 1966.
- [96] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Physical Review B*, vol. 14, pp. 556–582, Jul 1976.
- [97] D. Sholl and J. A. Steckel, *Density Functional Theory: A Practical Introduction*. New Jersey: John Wiley & Sons, Sept. 2011.
- [98] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of

- the generalized gradient approximation for exchange and correlation,” *Physical Review B*, vol. 46, pp. 6671–6687, Sept. 1992.
- [99] J. P. Perdew, “Density functional theory and the band gap problem,” *International Journal of Quantum Chemistry*, vol. 28, pp. 497–523, Mar. 1985.
- [100] H. J. Monkhorst and J. D. Pack, “Special points for Brillouin-zone integrations,” *Physical Review B*, vol. 13, pp. 5188–5192, June 1976.
- [101] D. J. Chadi and M. L. Cohen, “Special Points in the Brillouin Zone,” *Physical Review B*, vol. 8, pp. 5747–5753, Dec. 1973.
- [102] Ioffe Institute, “Silicon Bandstructure at 300k.” <https://bit.ly/2EBOgvP> Accessed 17th March 2015.
- [103] Ioffe Institute, “Germanium Bandstructure.” <https://bit.ly/2H9vNfJ> Accessed 17th March 2015.
- [104] Silvaco, *Atlas User’s Manual*. Santa Clara: Silvaco Inc., 2012.
- [105] C. Moglestue, *Monte Carlo simulation of semiconductor devices*. London: Chapman & Hall, 1993.
- [106] N. J. Pilgrim, Z. Ikonik, and R. W. Kelsall, “Influence of absorber layer dopants on performance of Ge/Si single photon avalanche diodes,” *Journal of Applied Physics*, vol. 113, p. 144508, Apr. 2013.
- [107] Y. Okuto and C. R. Crowell, “Threshold energy effect on avalanche breakdown voltage in semiconductor junctions,” *Solid-State Electronics*, vol. 18, pp. 161–168, Feb. 1975.
- [108] A. Schenk, “Rigorous theory and simplified model of the band-to-band tunneling in silicon,” *Solid-State Electronics*, vol. 36, pp. 19–34, Jan. 1993.

- [109] M. Luisier and G. Klimeck, "Atomistic Full-Band Design Study of InAs Band-to-Band Tunneling Field-Effect Transistors," *IEEE Electron Device Letters*, vol. 30, pp. 602–604, June 2009.
- [110] T. Krishnamohan, D. Kim, C. D. Nguyen, C. Jungemann, Y. Nishi, and K. Saraswat, "High-mobility low band-to-band-tunneling strained-Germanium double-gate heterostructure FETs: Simulations," *IEEE Transactions on Electron Devices*, vol. 53, pp. 1000–1009, May 2006.
- [111] O. Nayfeh, C. Chleirigh, J. Hennessy, L. Gomez, J. Hoyt, and D. Antoniadis, "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions," *IEEE Electron Device Letters*, vol. 29, pp. 1074–1077, Sept. 2008.
- [112] P. Adell, H. Barnaby, R. Schrimpf, and B. Vermeire, "Band-to-Band Tunneling (BBT) Induced Leakage Current Enhancement in Irradiated Fully Depleted SOI Devices," *IEEE Transactions on Nuclear Science*, vol. 54, pp. 2174–2180, Dec. 2007.
- [113] M. Shin, "Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the kp method," *Journal of Applied Physics*, vol. 106, p. 054505, Sept. 2009.
- [114] K.-H. Kao, A. Verhulst, W. Vandenberghe, B. Soree, G. Groeseneken, and K. de Meyer, "Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs," *IEEE Transactions on Electron Devices*, vol. 59, pp. 292–301, Feb. 2012.
- [115] P. Dzwig, M. G. Burt, J. C. Inkson, and V. Crum, "Optical matrix elements and cross sections for deep levels in GaAs: the impurity superlattice model," *Journal of Physics C: Solid State Physics*, vol. 15, pp. 1187–1198, Feb. 1982.
- [116] D. Y. K. Ko and J. C. Inkson, "A new matrix method for tunnelling in heterostructures: σ , X effects in single-barrier systems," *Semiconductor Science and Technology*, vol. 3, p. 791, Aug. 1988.

- [117] A. Di Carlo, P. Vogl, and W. Ptz, “Theory of Zener tunneling and Wannier-Stark states in semiconductors,” *Physical Review B*, vol. 50, pp. 8358–8377, Sept. 1994.
- [118] M. Ferhat, A. Zaoui, B. Khelifa, and H. Aourag, “Band structure calculations of $\text{Ge}_x\text{Si}_{1-x}$,” *Solid State Communications*, vol. 91, pp. 407–411, Aug. 1994.
- [119] Y. P. Varshni, “Temperature dependence of the energy gap in semiconductors,” *Physica*, vol. 34, no. 1, pp. 149–154, 1967.
- [120] J. Bernholc, N. O. Lipari, and S. T. Pantelides, “Self-Consistent Method for Point Defects in Semiconductors: Application to the Vacancy in Silicon,” *Physical Review Letters*, vol. 41, pp. 895–899, Sept. 1978.
- [121] S. T. Pantelides, “The electronic structure of impurities and other point defects in semiconductors,” *Reviews of Modern Physics*, vol. 50, pp. 797–858, Oct. 1978.
- [122] M. J. Puska, S. Pykk, M. Pesola, and R. M. Nieminen, “Convergence of supercell calculations for point defects in semiconductors: Vacancy in silicon,” *Physical Review B*, vol. 58, pp. 1318–1325, July 1998.
- [123] Y. Bar-Yam and J. D. Joannopoulos, “Electronic structure and total-energy migration barriers of silicon self-interstitials,” *Physical Review B*, vol. 30, pp. 1844–1852, Aug. 1984.
- [124] A. M. Stoneham, *Theory of Defects in Solids: Electronic Structure of Defects in Insulators and Semiconductors*. Clarendon Press, Jan. 2001.
- [125] P. M. Fahey, P. B. Griffin, and J. D. Plummer, “Point defects and dopant diffusion in silicon,” *Reviews of Modern Physics*, vol. 61, pp. 289–384, Apr. 1989.
- [126] S. Goedecker, T. Deutsch, and L. Billard, “A Fourfold Coordinated Point Defect in Silicon,” *Physical Review Letters*, vol. 88, p. 235501, May 2002.
- [127] G. D. Watkins and J. R. Troxell, “Negative-U Properties for Point Defects in Silicon,” *Physical Review Letters*, vol. 44, pp. 593–596, Mar. 1980.

- [128] Y.-H. Lee and J. W. Corbett, “EPR studies of defects in electron-irradiated silicon: A triplet state of vacancy-oxygen complexes,” *Physical Review B*, vol. 13, pp. 2653–2666, Mar. 1976.
- [129] N. Fujita, A. T. Blumenau, R. Jones, S. berg, and P. R. Briddon, “Theoretical studies on 100 dislocations in single crystal CVD diamond,” *physica status solidi (a)*, vol. 203, pp. 3070–3075, Sept. 2006.
- [130] A. T. Blumenau, M. I. Heggie, C. J. Fall, R. Jones, and T. Frauenheim, “Dislocations in diamond: Core structures and energies,” *Physical Review B*, vol. 65, p. 205205, May 2002.
- [131] E. Igumbor, C. N. M. Ouma, G. Webb, and W. E. Meyer, “Abinitio study of germanium di-interstitial using a hybrid functional (HSE),” *Physica B: Condensed Matter*, vol. 480, pp. 191–195, Jan. 2016.
- [132] J. Fage-Pedersen, A. N. Larsen, and A. Mesli, “Irradiation-induced defects in Ge studied by transient spectroscopies,” *Physical Review B*, vol. 62, pp. 10116–10125, Oct. 2000.
- [133] F. D. Auret, W. E. Meyer, S. Coelho, and M. Hayes, “Electrical characterization of defects introduced during electron beam deposition of Pd Schottky contacts on n-type Ge,” *Applied Physics Letters*, vol. 88, p. 242110, June 2006.
- [134] F. D. Auret, S. M. M. Coelho, G. Myburg, P. J. Janse van Rensburg, and W. E. Meyer, “Electronic and annealing properties of the E0.31 defect introduced during Ar plasma etching of germanium,” *Physica B: Condensed Matter*, vol. 404, pp. 4376–4378, Dec. 2009.
- [135] M. D. Moreira, R. H. Miwa, and P. Venezuela, “Electronic and structural properties of germanium self-interstitials,” *Physical Review B*, vol. 70, p. 115215, Sept. 2004.
- [136] P. Broqvist, A. Alkauskas, and A. Pasquarello, “Defect levels of dangling bonds

- in silicon and germanium through hybrid functionals,” *Physical Review B*, vol. 78, no. 7, p. 075203, 2008.
- [137] J. Hayes, F. Gity, B. Corbett, and A. Morrison, “Modeling the effects of interface traps on passive quenching of a Ge/Si geiger mode avalanche photodiode,” *Optical and Quantum Electronics*, vol. 44, no. 3, pp. 119–124, 2012.
- [138] K. Wada and L. C. Kimerling, *Photonics and Electronics with Germanium*. John Wiley & Sons, May 2015.
- [139] C. E. Ekuma, M. Jarrell, J. Moreno, and D. Bagayoko, “Re-examining the electronic structure of germanium: A first-principle study,” *Physics Letters A*, vol. 377, pp. 2172–2176, Nov. 2013.
- [140] D. Bagayoko, “Understanding density functional theory (DFT) and completing it in practice,” *AIP Advances*, vol. 4, p. 127104, Dec. 2014.
- [141] T. W. Hollins, S. J. Clark, K. Refson, and N. I. Gidopoulos, “Optimized effective potential using the Hylleraas variational method,” *Physical Review B*, vol. 85, p. 235126, June 2012.
- [142] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. Payne, “First principles methods using CASTEP,” *Z. Kristall.*, vol. 220, pp. 567–570, 2005.
- [143] G. P. Francis and M. C. Payne, “Finite basis set corrections to total energy pseudopotential calculations,” *J. Phys.-Condes. Matter*, vol. 2, pp. 4395–4404, 1990.
- [144] J.-M. Jin and L. J. Lewis, “Ge-dimer relaxation on Si(100),” *Physical Review B*, vol. 49, pp. 2201–2204, Jan. 1994.
- [145] V. Milman, D. E. Jesson, S. J. Pennycook, M. C. Payne, M. H. Lee, and I. Stich, “Large-scale ab initio study of the binding and diffusion of a Ge adatom on the Si(100) surface,” *Physical Review B*, vol. 50, pp. 2663–2666, July 1994.

- [146] B. Hammer, L. B. Hansen, and J. K. Nørskov, “Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals,” *Physical Review B*, vol. 59, pp. 7413–7421, Mar. 1999.
- [147] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple,” *Physical Review Letters*, vol. 77, pp. 3865–3868, Oct. 1996.
- [148] M. Probert, “Geometry Optimization.” <https://bit.ly/2IMsZ5k> Accessed 20th June 2017.
- [149] R. W. Nunes, J. Bennetto, and D. Vanderbilt, “Atomic structure of dislocation kinks in silicon,” *Physical Review B*, vol. 57, pp. 10388–10397, May 1998. arXiv: cond-mat/9707152.
- [150] K. W. Lodge, A. Lapicciarella, C. Battistoni, N. Tomassini, and S. L. Altmann, “The 90 partial dislocation in silicon: Geometry and electronic structure,” *Philosophical Magazine A*, vol. 60, pp. 643–651, Nov. 1989.
- [151] A. S. Nandedkar and J. Narayan, “Atomic structure of dislocations in silicon, germanium and diamond,” *Philosophical Magazine A*, vol. 61, pp. 873–891, June 1990.
- [152] L.-W. Wang, L. Bellaiche, S.-H. Wei, and A. Zunger, “Majority representation of alloy electronic states,” vol. 80, pp. 4725–4728, 05 1998.
- [153] P. B. Allen, T. Berlijn, D. A. Casavant, and J. M. Soler, “Recovering hidden Bloch character: Unfolding Electrons, Phonons, and Slabs,” *Physical Review B*, vol. 87, Feb. 2013. arXiv: 1212.5702.
- [154] W. Ku, T. Berlijn, and C.-C. Lee, “Unfolding first-principles band structures,” *Physical Review Letters*, vol. 104, May 2010. arXiv: 1002.4218.
- [155] V. Popescu and A. Zunger, “Extracting e vs k effective band structure from supercell calculations on alloys and impurities,” *Physical Review B*, vol. 85, p. 085201, Feb. 2012.

-
- [156] V. Popescu and A. Zunger, “Effective band structure of random alloys,” *Physical Review Letters*, vol. 104, p. 236403, June 2010.
- [157] P. Brommer and D. Quigley, “Automated effective band structures for defective and mismatched supercells,” *Journal of Physics: Condensed Matter*, vol. 26, no. 48, p. 485501, 2014.
- [158] C. J. Fall, A. T. Blumenau, R. Jones, P. R. Briddon, T. Frauenheim, A. Gutierrez-Sosa, U. Bangert, A. E. Mora, J. W. Steeds, and J. E. Butler, “Dislocations in diamond: Electron energy-loss spectroscopy,” *Physical Review B*, vol. 65, p. 205206, May 2002.