CARDIFF UNIVERSITY

CARDIFF UNIVERSITY

PRIFYSGOL CAERDYDD

DOCTORAL THESIS

Transport Effects in InSb Quantum Well Heterostructures

Author:

Christopher J. McIndo

Supervisor:

Dr. P.D. Buckle

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

in the

Condensed Matter and Photonics Group,
Cardiff School of Physics and Astronomy
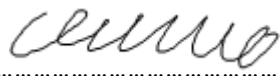
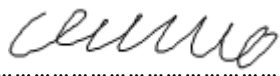Monday, 14 January 2019

Blank Page

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.
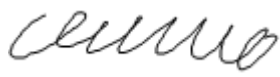
Signed ....   Date   14/01/2019

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ....   Date   14/01/2019

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ....   Date   14/01/2019

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ....   Date   14/01/2019

Blank Page

ABSTRACT

Doctor of Philosophy

TRANSPORT EFFECTS IN INSB QUANTUM WELL HETEROSTRUCTURES

By Christopher McIndo

In this thesis, a number of transport phenomena observed in n-type InSb quantum wells have been investigated. These include the low temperature ($3\ K$) mobility behaviour as a function of carrier density, the upper mobility limits observed at this temperature, and descriptions of the measured mobilities at a range of temperatures from $3\ K$ to $300\ K$. Following observations performed on material surfaces through use of Nomarski imaging, as well as analysis of detailed longitudinal magnetoresistance measurements, a proposed potential barrier model has been used to describe the low temperature upper mobility limit. Following this, modelling of advanced quantum structures has been performed, with theoretical and experimental results presented.

It is shown that at lower carrier densities ($< 3 \times 10^{11}\ cm^{-2}$), interface roughness scattering dominates, whilst at intermediate densities ($3 \times 10^{11}\ cm^{-2} < n_{2D} < 4 \times 10^{11}\ cm^{-2}$), scattering related to surface features dominates, where these features have an average diameter of $l_e = 2.43\ \mu m$. These surface features have been observed through optical Nomarski imaging, and their presence confirmed through detailed magnetoresistance measurements. This scattering occurs due to potential barriers formed at the boundaries of the uniform crystal growth that occurs around a central threading dislocation, where these barriers are on average pinned at a height 77% of the mid-gap value ($\Phi' \approx 115\ meV$). This scattering gives the upper mobility limit attained of $\approx 250{,}000\ cm^2/Vs$. Further increases in mobility cannot be realised without a decrease in surface feature density, achievable through buffer redesigns.

Initial investigations have been performed into exploiting the extreme material parameters inherent to InSb through use of surface bar gated structures to create a quantised conductance with a spin polarised current. Potential modelling has been performed showing this aim is theoretically possible, and preliminary measurements show evidence for good channel modulation and pinch off, though quantised conductance has yet to be observed.

Blank Page

This thesis is dedicated to the people who mean the most to me, my family, without whom this would not have been possible. Specifically:

♦ *it is dedicated to my darling wife, Laura, who gave me unending support despite the stresses of her own PhD research, who kept me happy even in the hard times, and who kept my work and life on track despite organising a wedding mid-PhD;*

♦ *to her parents, Martin and Helen, and her extended family who have made me one of their own, giving support and encouragement;*

♦ *and finally, to my parents, Lynda and Kevin, and grandparents, Len and Jean (and all of the dogs and horses too), who supported me through my whole academic life, from primary and secondary school, and through my life as an undergraduate, master's, and postgraduate student, and were always there for me whenever needed.*

To Laura: I love you, and to all of my family: thank you, I wouldn't be here without you.

There are also so many others who need specific acknowledgment for their patience, help and support as I completed this work. Firstly, I must acknowledge my primary supervisor, Dr Phil Buckle, who has helped guide me, and this work, from the very start to now, three and a half years later. There have been many detailed discussions, and without his valuable insights, much off this would not have been possible.

I must also take the time to specifically thank his research group, my fellow PhDs and PostDocs in the Cardiff Condensed Matter and Photonics Group, for all of their time, effort, and support given to me. I must thank Dr. Craig Allford, Dr. George Smith, Dr. Dave Hayes, Dr. Fadwa Alshaeer and Shawkat Jubair, for their insight into difficult discussions, for the help in experimental measurements, and for providing me with data to work with, and for assistance in fabrication of samples to measure. I could not have completed this research without all of their assistance. I must also explicitly thank Dr. Andreas Papageorgiou for his specific help with programming image analysis techniques.

I must also thank the staff of the School of Physics and Astronomy, both those from my Undergraduate time and during my Postgraduate research. From the lecturers and postgraduate teaching assistants, to my project and CUROP supervisors, and to the technical and admin staff, there has been unmatched support and assistance from, I have been made to feel part of the family. Specifically, and in no particular order I would like to acknowledge Nicola Hunt, Steven Baker, Prof. Carole Tucker, Prof. Emyr Macdonald, Dr. Sean Giblin, Dr. Joe Wilkes, Prof. Mark Hannam and Prof. Patrick Sutton, who have helped me immeasurably in my 7 years at Cardiff.

I cannot forget to acknowledge all of the friends I have made in my time at the Cardiff Physics department; Daniel O'Neill, Marie Anderson, Pete Millington, Phil Whybra, and those friends I have made outside of physics; Rosie Hughes, Gabi Cahil and Tom

McManus, you have kept life fun and more than interesting, and kept me sane during this process, the value of which cannot be understated.

Thank you Laura, and thank you all.

CONTENTS

Journal Articles

- *C. J. McIndo, D. G. Hayes, A. Papageorgiou, L. A. Hanks, G. V. Smith, C. P. Allford, S. Zhang, E. M. Clarke and P. D. Buckle, "Determination of the transport lifetime limiting scattering rate in $InSb/Al_xIn_{1-x}Sb$ quantum wells using optical surface microscopy", Physica E: Low-dimensional Systems and Nanostructures, vol. **91**, pp. 169-172, 2017.*

- *D. G. Hayes, C. P. Allford, G. V. Smith, C. McIndo, L. A. Hanks, A. M. Gilbertson, L. F. Cohen, S. Zhang, E. M. Clarker and P. D. Buckle, "Electron transport lifetimes in $InSb/Al_{1-x}In_xSb$ quantum well 2DEGs", Semiconductor Science and Technology, vol. **32**, p. 085002, 2017.*

- *C. J. McIndo, D. G. Hayes, A. Papageorgiou, L. A. Hanks, G. V. Smith, C. P. Allford, S. Zhang, E. M. Clarke and P. D. Buckle, "Optical Microscopy as a probe of the rate limiting transport lifetime in $InSb/Al_{1-x}In_xSb$ quantum wells", Journal of Physics: Conference Series, vol. **964**, p. 012005, 2018.*

- *C. J. McIndo, L. A. Hanks, G. V. Smith, C. P. Allford, S. Zhang, E. M. Clarke and P. D. Buckle, "Potential barrier model for rate limiting transport lifetime in $InSb/Al_xIn_{1-x}Sb$ quantum wells: Landauer-Büttiker and Monte Carlo modelling of magnetoresistance", in preparation.*

Conference Presentations

- *UK Semiconductors 2017, The University of Sheffield, Sheffield, U.K., 12th-13th July 2017, "Determination of the transport lifetime limiting scattering rate in $InSb/Al_xIn_{1-x}Sb$ quantum wells using optical surface microscopy" – oral presentation.*

- *Advances in Quantum Transport in Low Dimensional Systems, University College London, London, U.K., 4th-5th September 2017, "Determination of the transport lifetime limiting scattering rate in $InSb/Al_xIn_{1-x}Sb$ quantum wells using optical surface microscopy" – poster presentation.*

- *International Symposium on Quantum Hall Effects and Related Topics, Max Planck Institute for Solid State Research, Stuttgart, Germany, 27th-29th July 2018, "Advances in electron transport in $InSb/Al_xIn_{1-x}Sb$ quantum wells" – poster presentation.*

- *UK Semiconductors 2018, The University of Sheffield, Sheffield, U.K., 4th-5th July 2017, "Advances in electron transport in $InSb/Al_xIn_{1-x}Sb$ quantum wells: magnetoresistance measurements and transport lifetime modelling" – oral presentation.*

- *ICSNN (20th International Conference on Superlattices, Nanostructures and Nanodevices), Spanish National Research Council (CSIC), Madrid, Spain, 23rd-27th July 2018, "Advances in electron transport in $InSb/Al_xIn_{1-x}Sb$ quantum wells: magnetoresistance measurements and transport lifetime modelling" – poster presentation.*

XXIV

Blank Page

1 : INTRODUCTION

1.1 Motivation

Electronic devices are ubiquitous in the modern world, with their spread and complexity ever increasing. The constant demand for smaller devices capable of running at greater speeds with lower running costs has been present since the invention of the transistor by Bardeen and Brittain in late 1948 [1]. Since the first realisation of the transistor, advances have led to the invention of the integrated circuit and the development of modern computing and electronics as we know it today. The desire for more complex devices at reduced costs, coupled with the dependence on the transistor, led to the observation known as Moore's Law, an observation that noted the rapid decrease in size of transistor components and the rapid increase in the number of components per integrated circuit, with these roughly doubling every year [2]. Traditionally the chip and the transistor have relied on silicon technology, adapted to ever more complex needs. Currently the number of transistors on an integrated chip is in the billions, with gate lengths of 14 nm [3] or below. This ever decreasing size has led some to claim that Moore's law is coming to an end [4, 5, 6].

Whilst silicon has been the underpinning technology that has allowed the rapid growth of electronic devices, it is approaching its limits [7]. As the desire for ever faster speeds, reduced operating powers, smaller structures and other specialised needs is nearing, research is branching out into other semiconductor structures, specifically that of III-V and II-VI semiconductors [8]. These are semiconductors made by combining elements from group III (II) of the periodic table with those from group V (IV) to form a new crystal structure. These new compound semiconductors allow a much greater range of research and application, overcoming some of the many undesirable properties of silicon technology. The significance of these materials was recognized in the awarding of the Nobel Prize in Physics in 2000 "*for developing semiconductor heterostructures used in high-speed- and opto-electronics*" [9].

Of particular interest are those materials and devices based on the III-V compound indium antimonide (InSb). InSb is an extreme of the III-V semiconductors, having the largest lattice constant, the smallest band gap (which is also direct) and the lightest effective mass [10, 11, 12, 13] (see Table 1.1 for a comparison between InSb and other common semiconductors).

1 : Introduction

Table 1.1 Physical and electrical properties of InSb, including band gap, effective mass, relative dielectric constant and lattice constant. The values for the corresponding parameters in the more common semiconductors silicon (Si) and gallium arsenide (GaAs) are given as a comparison, showing that InSb is very favourable for high speed, low power electrical devices. Data from references [10, 11, 12, 13].

Property		Si	GaAs	InSb
Band gap E_g (300 K)	(eV)	1.11	1.43	0.17
Band gap E_g (0 K)	(eV)	1.17	1.519	0.235
Electron effective mass m_e^*		0.98	0.063	0.014
Heavy hole effective mass m_{hh}^*	(m_e)	0.49	0.51	0.43
Light hole effective mass m_{lh}^*		0.16	0.082	0.015
Relative Dielectric Constant ε_r	(ε_0)	11.7	12.91	16.8
Lattice Constant a	(Å)	5.431	5.653	6.479

This table shows how the parameters of InSb compare very favourably with other more commonly used semiconductors (silicon and gallium arsenide as two of the most common used for electrical and optoelectronic devices), and when coupled with its large Landé g-factor ($g{\sim}50$) [14] and the highest reported room temperature electron mobility ($\mu_e = 78{,}000\ cm^2/Vs$) [11], it is clear that InSb is a possible candidate for use in a range of advanced electronic devices. These range from low power, high frequency electronics, to mm-wave devices, high electron mobility transistors and other advanced quantum devices. InSb quantum wells (QWs) are also an ideal candidate to use for investigation into, and exploitation of, Majorana fermions, where the possible planar fabrication techniques allow for integration and scalability of multiple devices [15, 16, 17, 18, 19, 20, 21].

All of these devices require good quality material to achieve the best mobilities possible, resulting in (e.g.) the lowest power consumptions or longest electron ballistic lengths. However, this is where the immaturity of research into InSb suffers when compared to other semiconductors.

1 : Introduction

One of the most significant complications arises from inherent strain in epitaxially grown InSb quantum wells (used to form confined two-dimensional electron gasses, 2DEGs) where there are generally no suitable lattice matched materials for use as substrates or confining potential barriers. Instead the most common substrate used is semi-insulating GaAs, with this introducing a large strain due to the ~14.6% lattice mismatch [22].

The effect of this strain and the introduction of associated defects on the electron mobility has been shown to be significant [23, 24, 25, 26, 27, 28], however the effect has only been shown as a correlation, not as a direct explanation for the electron transport observed. In determining the properties and mechanisms of scattering for electrical transport, the strain has only traditionally been considered in terms of the secondary change in parameters (i.e. a strain modified effective mass, relative dielectric constant, etc.), and not directly in terms of defects and scattering. Transport lifetime modelling using the relaxation time approximation has been performed in InSb previously [17], however many of the required parameters used in the modelling had a large associated uncertainty, and to match observed data, the extremes of these ranges tended to be used. Recent improvements in growth quality, combined with further measurements and extended modelling has allowed for the refinement of many of these parameters, with these values now showing a discrepancy between measured and predicted mobilities [29]. This discrepancy is believed to be related to strain, and the resultant threading dislocations and other defects, not previously accounted for in transport modelling of InSb. These defects manifest at the surface of the material as visible hillocks, with a characteristic size strongly correlated to the largest mean free paths measured [30]. Accounting for scattering due to these defects that are related to the surface features gives excellent agreement between the measured mobilities and those predicted by adjusted transport modelling [30, 31].

The quality of this epitaxially grown InSb and related scattering is the subject of this thesis, specifically focussing on modelling of the resultant observed upper limit to mobility, with the determination of limiting scattering mechanisms across a wide range of sample carrier densities. Following this, designs for advanced bar gated quantum structures are considered, with the possibility of using these as a field-effect transistor exploiting Zener tunnelling in a reduced dimension system [32, 33, 34, 35]. Due to the large Landé g-factor and spin orbit coupling, these devices have the potential to exploit the large spin splitting, exhibiting quantised conduction of a spin polarised current.

1.2 Historical Perspective

Initially, research into semiconductor structures and devices was dominated by elemental semiconductors such as silicon and germanium. Compound III-V semiconductors such as GaAs were studied, but these were limited to bulk 3D structures due to the limitations of growth methods. Growth was generally performed via the Bridgman technique [36], growing a single crystal structure from a melt. However since the invention of Molecular Beam Epitaxy (MBE) by J. R. Arthur and A. Y. Cho in the late 1960s [37, 38], research into more advanced material systems, and more advanced heterostructures, has exploded. Of the increased research into III-V systems, GaAs and it's heterostructures are by far the most mature field. This has been due to relative ease of growth of GaAs heterostructures with advantageous lattice matched materials for use as substrates and electronic potential barriers (see Chapter 2 : Material and Electrical Properties of Indium Antimonide) [37, 38]. This has led to a huge improvement of the mobility of GaAs heterostructures over time, as shown in Figure 1.1 [39], with mobilities ranging from ~10^4 cm^2/Vs in 1978 to ~10^7 cm^2/Vs by the year 2000.
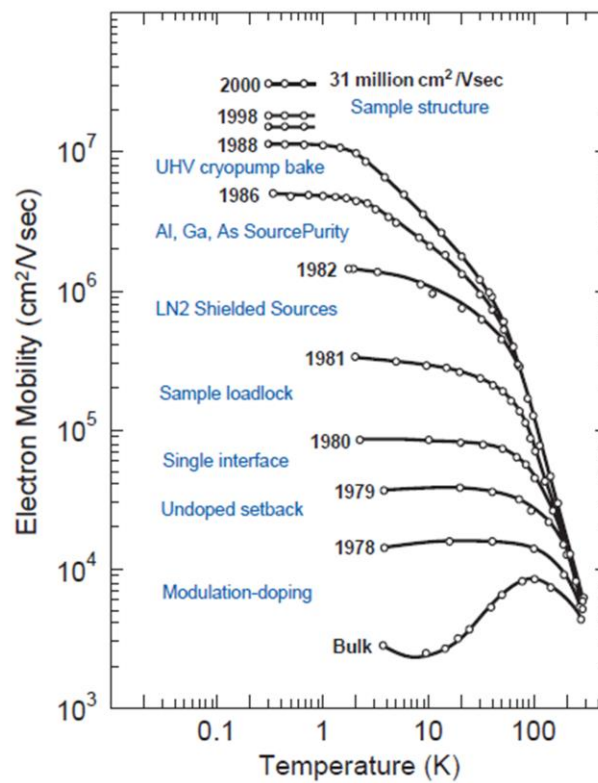
Figure 1.1 Improvement of mobility in bulk and quantum well GaAs systems over time, with brief explanations for the main contribution to each increase given. The majority of improvements have arisen due to improvements in growth via MBE, whilst the more recent advancements have arisen as a result of sample structure optimisation. Figure reproduced from Fig. 1 in reference [39].

1 : Introduction

These large mobilities have been achieved through modulation doping, where the donor dopant atoms are physically separated from the electron charge carriers in the well, reducing the associated remote ionised impurity scattering. Further advances in mobility have been also achieved through the use of quantum well structures, where electrons are confined in the growth direction, either by sandwiching a layer of a lower band gap material (GaAs), between higher band gap material (typically AlGaAs), or using a single GaAs/AlGaAs heterojunction to create a triangular well. This causes the electron energy levels to be confined in the growth direction, but free to move in the other two orthogonal directions. This is called a two-dimensional electron gas (2DEG).

These massive improvements, giving mobilities $> 35 \times 10^6 \ cm^2/Vs$ [40] have occurred in a material system with an effective mass ~5 times larger than in InSb. If these similar advances could be replicated in a material with a lighter effective mass, this could potentially lead to mobilities significantly in excess of those seen in GaAs.

A large amount of research into the properties of InSb was performed by Putley and others in the 1960s [41, 42]. Following on from this, early work into InSb quantum wells was performed at QinetiQ, Malvern[1] and Imperial College London, where (amongst other reasons) there was particular interest in InSb based devices due to the band gap being in the infrared [43, 44, 45, 46, 47]. Here, many developments were achieved, including physical improvements in sample doping and growth quality, as well as optimisation of device structures for practical purposes, giving a deeper understanding of the underlying physics.

This work continued in the late 2000s and early 2010s [48, 49, 50], leading to reports of the low temperature mobilities in excess of 200,000 cm^2/Vs [15], and a Landé g-factor of ~50 [14], as well as further understanding of the non-parabolic nature of InSb [16, 49]. Fundamental to the understanding and improvement of these structures was accurate knowledge of the electronic transport properties, and so it was at this point that the relaxation-time approximation transport model mentioned previously was researched and implemented [17]. Whilst there was generally good agreement between measured data and theoretical predictions [17], values used in the model tended to be on the extreme end of those plausible, and when more reasonable and refined values are used, these models failed to match the experimental data [29]. This lack of agreement showed that despite

---

[1] QinetiQ, Saint Andrews Road, Malvern, Worcestershire WR14 3PS, United Kingdom.

the many advances and observations in InSb, a critical limiting scattering mechanism was missing.

At a similar time during the 2000s, InSb crystal quality was being investigated at the University of Oklahoma, led by M. B. Santos. This work involved extensive use of a range of techniques (e.g. plan-view transmission electron microscopy (TEM), cross-sectional TEM (X-TEM), atomic force microscopy (AFM)) to study the morphology of epitaxially grown InSb on GaAs substrates [51]. This included correlating structural defects with measured mobilities [23, 24], with these defects primarily consisting of microtwin defects [25, 26] and threading dislocations [52]. Investigations were performed with the aim of reducing these defect densities, using dislocation filtering interfaces [52], or varying the Al content of $Al_xIn_{1-x}Sb$ layers [27]. Following this, 2° offcut substrates were shown to give a reduction in the microtwin defect density, with this leading to an improved room temperature mobility [28]. The significant impact of microtwin defects on the physical structure of typical quantum wells was demonstrated through dark field X-TEM, where these defects were shown to create physical well misalignments [53].

Despite this wealth of research into the link between structural quality and mobility, and the increased understanding of electrical transport properties, the unaccounted for limiting mobility scattering mechanism remained. Despite this, advancements in device designs have occurred, with recent reports of quantum point contact and g-factor measurements in InSb quantum well systems [19, 18]. There is a desire to advance these quantum structures further, using small area bar gated structures to create a transistor where operation is through exploitation of Zener tunnelling (an interband tunnel transistor [32, 33, 34, 35]). In this regime, the large g-factor and spin-orbit interaction of InSb can be used to preferentially increase or decrease the tunnelling current of a certain spin direction, leading to a polarised spin tunnelling current, which would be of use in the goal of achieving quantum computing.

1.3 Outline of Thesis

This thesis describes the work undertaken to fabricate, measure and analyse data from Hall measurements performed across a range of InSb quantum well (QW) samples. The aim of this work is to understand the physics of electron transport in these samples, in terms of limiting scattering mechanisms, with the overall goal of overcoming these current limits. To achieve this goal requires detailed knowledge of the properties of InSb, as well as the physical heterostructures studied. With this knowledge, more advanced quantum structures can be realised, an important step in the goal of reaching lower power, higher

frequency devices, as well as devices exploiting spin polarised currents and Majorana Fermion observations.

Chapter 1 : Introduction gives an overview of the research performed and the motivations behind the work, as well as placing the work in an historical context.

Chapter 2 : Material and Electrical Properties of Indium Antimonide gives a more detailed look at the physical crystal properties, as well as the resulting electronic properties, of the binary InSb and its ternary alloy AlInSb. The concept of periodic lattice structure is introduced, with a discussion of how this gives rise to electronic band theory. The formation of bands and band gaps in the free and nearly free electron models is discussed, as well as the modifications necessary to account for the non-parabolicity of InSb via $\boldsymbol{k} \cdot \boldsymbol{p}$ theory, with equations for the effective mass, dispersion relations and band gap given. Following this, the density of states for 3D systems, as well as reduced dimensionality systems, is derived, and the effect of temperature on electronic state population is considered via the Fermi-Dirac distribution, with the equation for the non-parabolic Fermi energy considered.

Chapter 3 : Realisation and Characterisation of InSb Quantum Wells details the processes involved in the physical realisation of complex heterostructures such as InSb QW samples, as well as detailing the experimental methods and the underlying theory used to electrically investigate the InSb QW samples studied in this thesis, explaining how their electrical properties (carrier density and mobility) are determined. The chapter begins with a discussion on the theory of alloying to create a ternary semiconductor, with the associated change of parameters such as lattice constant and band gap discussed. Details of band alignments between InSb and AlInSb are discussed, with values for band offsets given. The realisation of these heterostructures via Molecular Beam Epitaxy (MBE) is then described, detailing the growth of the sample from a basic substrate to creating complex structures, with a typical sample structure given, as well as a table of samples studied in this thesis along with example Schrödinger-Poisson solutions for the band edge profile. The associated difficulties with this growth technique are introduced, including dopant dragging, strain, and dislocations. Following this, the processes involved in device fabrication are described, including the standard techniques used to define Hall bars, the basic device used in this work, to study the InSb material and its properties. The methods of device characterisation are discussed, with a description of the physical experimental set up (source-measurement units, electromagnet and pulse tube cryostat), measurement methods (constant current δ-mode) and control software (python) used to perform electrical measurements as a function of temperature and magnetic field. Finally, the chapter ends with a derivation of the 2D transverse Hall effect, the primary physical effect

used to determine the carrier density, and how this combines with a longitudinal resistance measurement to derive the sample mobility. Typical experimental results are shown.

In Chapter 4 : Schrödinger-Poisson and Transport Modelling, the results of these electrical measurements across the range of samples are discussed, with the broad explanations for observed trends given. Schrödinger-Poisson modelling of the quantum well samples is performed, and basic transport lifetime modelling is outlined, with the primary limiting scattering mechanisms derived and implemented to analyse the results shown previously. The reasons for the discrepancies between the model and measured data are discussed, with the need for an extra scattering mechanism demonstrated. Following this need for an extra scattering mechanism, the sample quality is discussed, specifically through the use of Differential Interference Contrast DIC (Nomarski) imaging. Characteristic surface roughness related to structural defects is shown to exist on all samples imaged, with the image analysis techniques and extraction of the characteristic size of these surface features discussed. This average surface feature size is shown to correlate very strongly with the longest electron mean free paths, and through a simple Drude model, scattering from defects related to these surface features is included in the standard transport model and is shown to match the measured data well.

Chapter 5 : Monte Carlo Modelling and Magnetoresistance extends this analysis further, proposing a potential barrier model for scattering from these defects. Material within a feature is treated as uniform potential, whereas the boundaries between features are treated as potential barriers. Current flow is achieved via tunnelling through barriers with this modelled using Landauer-Büttiker theory. The mobility is then analysed using a combination of a Monte-Carlo (MC) and Drude model. Effects of variation in barrier characteristics (pinned height and separation) on the results of the model are discussed. Detailed longitudinal magnetoresistance measurements are performed and analysed, showing evidence for scattering at circular electronic orbit sizes matching those extracted from Nomarski imaging, as well as giving a background dopant density. Combining the Monte-Carlo model with the magnetoresistance measurements allows for the determination of the pinned barrier height and shape, with this analysis extended across the range of samples, confirming this surface feature related scattering as the limiting scattering mechanism at high carrier densities and mobilities. Interface roughness is shown to be the limiting scattering mechanism at lower densities.

Chapter 6 : Gated Structure Modelling is concerned with more advanced quantum structures that could be possible given advances in the material described previously, specifically, small area triple bar gated structures. The aim of these devices is to use the bar gates to change band alignments, giving rise to Zener tunnelling from the 2DEG

conduction band to a confined state in the valence band and back to the conduction band. This confined state should give rise to a well-defined quantised conductance, where exploitation of the large g-factor in InSb could further refine the current to be spin polarised. This chapter covers the fabrication, and modelling of these structures, through electrostatic effects on the band structure and Schrödinger calculations of the bound states.

The full results and implications from this thesis are then summarised in Chapter 7 : Conclusions and Future Work, which includes a brief description of initial e-beam patterned triple bar gated sample measurements and possible future work.

Two appendices are included, where Appendix A discusses further the Schrödinger-Poisson modelling, including a description of the numerical methods used in calculating solutions to the differential equations (shooting method and finite difference method). Appendix B discusses the speed improvements and associated run-time reductions possible through the use of the Cython programming language, as opposed to pure Python or NumPy, for the Monte Carlo model implemented in Chapter 5.

2 : Material and Electrical Properties of Indium Antimonide

2.1 Introduction

Semiconducting material is realised in two main forms, elemental (such as germanium and silicon), or compound (such as InSb, GaAs or CdTe), in which a group II or III element is alloyed with a group VI or V element to form a II-VI or III-V semiconductor [54]. Traditionally Si has been the most widely used semiconducting material due to several advantageous properties, including natural abundance, a SiO_2 native oxide, and the relative ease of growth of pure, low defect density material. However, the electrical properties of Si are often far from ideal (such as the indirect bandgap), and it is due to these reasons that research and industrial interest into compound semiconductor materials has grown [55, 56].

The potential combinations of compound semiconductors are vast, including from simple binaries such as InSb, to ternaries, AlInSb or quaternaries AlInAsSb. All of these different compounds will have a different physical structure, with corresponding different electrical properties and behaviours with temperature (as in Table 1.1 [10, 11, 12, 13]). It is therefore critical to understand the basic physical structure of the material studied here and how this gives rise to the corresponding electrical behaviour.

Much of the information given subsequently on crystal structure, Brillouin zones and E vs k dispersion relations (band diagrams) can be found in most condensed matter and semiconductor text books, many of which will go into greater mathematical detail than is covered here (examples include Hoffmann's "*Solid State Physics: An Introduction*" [57], Kittel's "*Introduction to Solid State Physics*" [13] and Ashcroft and Mermin's "*Solid State Physics*" [58]). In this thesis, a description of the crystal structure and reciprocal lattice of InSb will be given, along with a brief qualitative description of band structure formation. An introduction to the free and nearly free electron models will be given, with an explanation of how this arrives at a band structure. Following this, important physical properties such as the density of states and Fermi distributions will be examined.

2.2 Properties of InSb

2.2.1 Crystal structure

Both elemental and III-V semiconductors form crystalline material[2], where a basis set of atoms (the simplest repeating atom or motif of atoms) are repeated over a regular set of points (the Bravais lattice) defining the positions of these constituent basis atoms. Together the basis atoms and Bravais lattice fully describe the regular repeating 3D nature of the crystal. In InSb, like most III-V compound semiconductors, the basic crystalline structure is a zincblende structure. Zincblende is a diamond like structure consisting of two intersecting face-centred-cubic (fcc) Bravais lattices, a lattice for each constituent element (In and Sb), with the two elements covalently bonded together at an offset of $a/4$ (where a is the face-centred-cubic edge vector length). For the fcc lattice, the primitive translation vectors point to the face-centred atoms, giving a 60° angle between vectors.

(a)   *(b)*
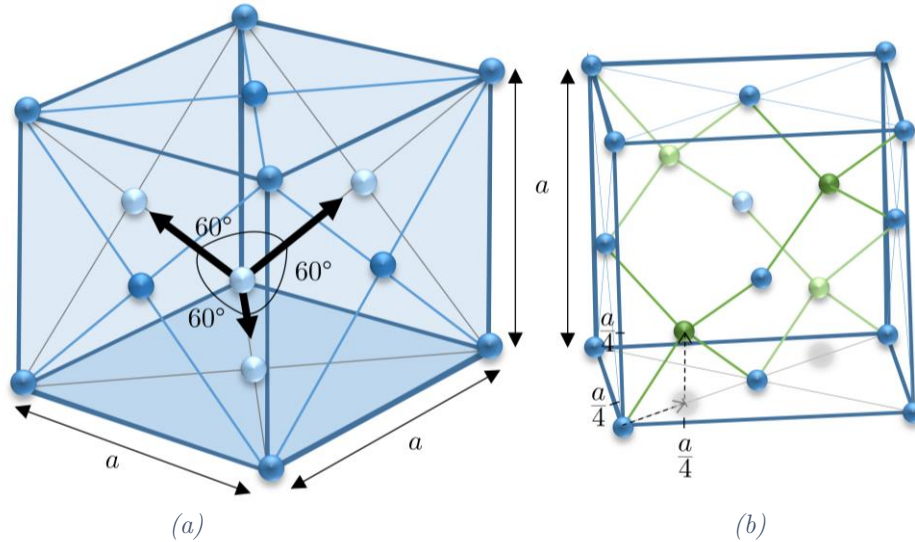
Figure 2.1 (a) The face-centred-cubic (fcc) Bravais lattice with atoms located at each vertex and the centre of each face. The edge vector length a is shown, as are the 3 primary lattice vectors (bold arrows), at an angle of 60° to each other. (b) Zincblende structure of InSb, consisting of two interpenetrating fcc lattices for each constituent species (In, blue) and (Sb, green), offset by a/4. Covalent bonds are formed between the four nearest neighbour atoms in a tetrahedral arrangement (only the bonds within the unit cell are shown).

---

[2] Semiconductors can also form amorphous material, where this is exploited for applications such as in silicon solar cells, though this type of material is not considered here.

In a perfect crystal this basic unit cell would be repeated indefinitely, with all the atoms in their perfect arrangement, however in any real crystal there will be defects which will alter the perfect repeating nature. These defects can occur in many forms, including unwanted interstitial impurity atoms, missing atoms (vacancies) or atoms in incorrect lattice sites. Defects also include purposefully created changes to the perfect crystal, such as creating a ternary compound by replacing a correct atom with a different element atom (such as replacing In with Al in AlInSb) or when doping such as with tellurium (discussed in section 3.2). The effect of defects present in the lattice is to change the electrical properties of the crystal, whilst also possibly introducing strain and other associated effects. This variation in physical and electronic structure between the basic binary and ternary compounds allows for the refinement of specific desired properties, and is the reason much research is performed into ternary compounds and their incorporation into devices.

The physical structure of the crystal can be visualised in real space as in Figure 2.1, however many of the important electrical properties of the material can be more readily considered in reciprocal space. This is known as momentum space or k-space, and has many direct analogies with real space. In k-space, the reciprocal lattice vectors ($\boldsymbol{G}$) are related to the real space Bravais lattice vectors ($\boldsymbol{R}$), such that $\boldsymbol{R}_i \cdot \boldsymbol{G}_j = 2\pi n \delta_{ij}$, where n is an integer, and $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$, where i and j represent lattice vector directions. If the real space lattice spacing is given by a, then the reciprocal lattice will be spaced by $2\pi/a$ (the specific values of these $\boldsymbol{G}$ vectors for the Bravais lattice structures are not relevant here).

As the face-centred-cubic Bravais lattice defined the possible positions of atoms in real space, the equivalent reciprocal lattice is the body-centred-cubic lattice, which determines the possible electronic momentum ($\boldsymbol{k}$) states allowed. There is also an equivalent unit cell that can be repeated to describe the whole structure, with this known as the 1ˢᵗ Brillouin zone. Due to the repeating nature of the crystal, any point outside of this zone can be mapped to a point inside, with the significance that this zone contains all the information necessary to describe the electronic properties of the crystal. The 1ˢᵗ Brillouin zone is given by the Wigner-Seitz crystal, which is defined as all the points in reciprocal space closest to any given lattice point (shown in Figure 2.2).

Due to the nature of the geometrical shape, there are several important points that are invariant to rotations and reflections about the $\boldsymbol{k} = 0$ zone centre (the Γ point). These points are denoted as L, X, K and W and are labelled in the figure alongside the three primary directions [010] (Δ), [111] (Σ) and [110] (Λ). The band structure of InSb along these directions is shown in Figure 2.5a (reproduced from Fig. 8 in reference [59]).

(a)   *(b)*
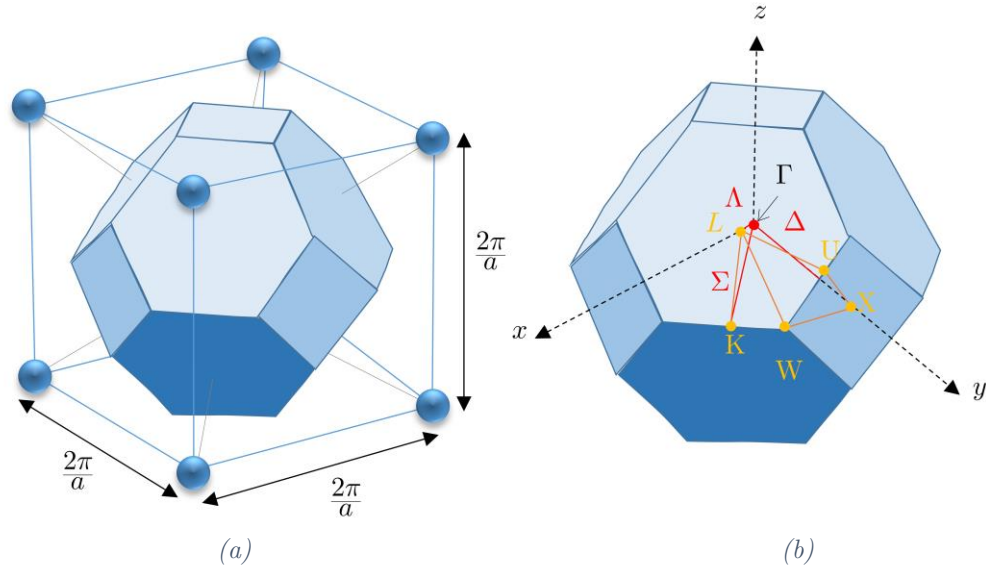
Figure 2.2 (a) The Wigner-Seitz unit cell defining the 1st Brillouin zone of the body-centred-cubic Bravais lattice (the reciprocal of the fcc lattice), where the unit cell width is given by $2\pi/a$. (b) The 1st Brillouin zone, showing the high symmetry points that are invariant to rotations and reflections (orange points) as well as the $\boldsymbol{k}=0$ zone centre point (Γ, red), and three primary directions (red lines).

2.2.2 Band structure

The band structure of any semiconducting material (and all of its other associated electronic properties) arise as a direct result of the crystal structure and the arrangement of the bound positive ionic cores. The concept of band structure formation is described mathematically using many models and refinements, such as starting from the tight binding approximation or the simple free and nearly free electron models [60]. These different models have different starting points, and different intuitive understandings, however in general, the results are broadly the same under standard conditions. The basis of many models involves solving the time independent Schrödinger equation[3] for the electrons to determine their allowed energy states, where this is performed for one and then many atomic orbitals (the tight binding model), or for a crystal with either a uniform (free electron model) or periodic (nearly free electron model) potential energy distribution. The results from these models take the form of energy dispersion relations, with "forbidden

---

[3] The Schrödinger equation is a mathematical equation describing the quantum mechanical state of the object concerned, and how this evolves with time. For special cases standing wave solutions can occur, with this simplifying the full Schrödinger equation to the time independent form, where the eigenfunction solutions give the wavefunctions, with the corresponding energy given by the eigenvalue.

energy" band gaps formed due to the difference in potential near and between the periodic ionic cores.

2.2.3 Origin of band structure (free electron model)

Following reference [13], a qualitative description of energy bands and band gap formation can be derived by consideration of the splitting of the atomic energy levels due to the close proximity and overlap of the atomic orbitals.

As a pair of atoms with well-defined atomic orbital energy levels are brought into close proximity, their electronic wavefunctions will overlap and interfere, and due to the Pauli exclusion principle, this overlap will necessarily affect their associated energy levels. A pair of energy states will form, with a lower energy bonding state due to constructive interference (with the electron density focused on and between the nuclei), and a higher energy anti-bonding state due to destructive interference (with a depletion between the nuclei). This concept readily extends to N atoms as in a real crystal, with the large N giving rise to a vanishingly small difference between similar states (a band), separated by a large energy from other bands (a band gap).

This is the (qualitative) basis of the (quantitative) tight binding model, with this giving an intuitive description of how bands are formed, however, to be of any realistic use, a quantitative description is needed. The actual energy dispersion relations (E vs $\boldsymbol{k}$) need to be determined for all states in the 1$^{\text{st}}$ Brillouin zone to be able to understand the electronic properties of the crystal. Here $\boldsymbol{k}$ is a wavenumber, but is also used as a quantum number to describe the specific wavefunction and energy solutions. Two models will be discussed here, the free electron model, and the nearly free electron model.

An initial attempt at calculating the dispersion relations resulted in the Drude-Sommerfeld model (also known as the free electron model) [60], where the crystal is already formed, and the electronic states then have to be determined. Initially the electrons are treated as free in the sense that they do not interact with the ionic cores, nor with other electrons. The problem is then reduced to finding the energy eigenstates of the time independent Schrödinger equation (TISE) for a single electron for a given potential, with all other electrons in the crystal then populating these eigenstates according to the Pauli exclusion principle. The TISE for a given electronic $\boldsymbol{k}$ state is then given by

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + V(\boldsymbol{r}) \right)\psi_{\boldsymbol{k}}(\boldsymbol{r}) = E_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) \qquad (2.1)$$

where $\hbar$ is the reduced Planck constant ($\hbar = 1.055 \times 10^{-34}$ Js), m_e is the mass of the free electron ($m_e = 9.11 \times 10^{31}$ kg), ∇ is the gradient operator, $V(\boldsymbol{r})$ is the position dependant potential, ψ is the electronic wavefunction and E is the corresponding energy. To be able to solve equation (2.1) and determine the E vs $\boldsymbol{k}$ dispersion relation, the correct form of the potential must first be chosen, however due to the periodicity of the lattice, for any Bravais lattice vector $\boldsymbol{R}$, the potential must be invariant under a translation by $\boldsymbol{R}$, i.e. $V(\boldsymbol{r}) = V(\boldsymbol{r} + \boldsymbol{R})$. For the simple case of the free electron model, it is assumed this potential is constant, where the specific value amounts to a relative energy shift, and so can be assumed to be $V = 0$.

If the electrons are considered bound to a macroscopic box (the crystal), length L, then the simplest boundary conditions would be to require the wavefunction to vanish at the boundaries of this box. This would however correspond to standing waves in the crystal which would not be convenient to describe electrical conductivity. Instead, Born-von-Kármán boundary conditions are chosen, with this describing periodic boundary conditions. Now, the boundary conditions are given by

$$\psi(\boldsymbol{r}) = \psi(x,y,z) = \psi(x+L,y,z) = \psi(x,y+L,z) = \psi(x,y,z+L). \qquad (2.2)$$

The particular form of wavefunctions satisfying the Schrödinger equation for these boundary conditions are plane waves, normalised so there is a unity probability of finding the electron within the box, volume $V = L^3$, such that

$$\psi(\boldsymbol{r}) = \frac{1}{\sqrt{V}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \qquad (2.3)$$

where in each dimension, $\boldsymbol{k}$ must be some integer multiple (n) half wavelengths of the reciprocal box length, such that in each dimension

$$k = \pm\frac{\pi n}{L}, \qquad (2.4)$$

where $n \geq 0$.

Shifting this wavefunction by a Bravais lattice vector $\boldsymbol{R}$ has the effect of shifting the wavefunction via a phase factor $e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$, but as the electron density is proportional to $\psi^*\psi = |\psi|^2$, the probability of finding an electron at a particular location remains unchanged. Therefore, the electron density at all lattice points for a given $\boldsymbol{k}$ is equal, as would be expected.

The energy eigenstates are found by applying the TISE to ψ, giving the energies as

$$E = \frac{\hbar^2 \boldsymbol{k}^2}{2m_e} = \frac{\hbar^2}{2m_e}\left(k_x^2 + k_y^2 + k_z^2\right), \tag{2.5}$$

with energy separation between states of the order of

$$\Delta E = \frac{\hbar^2}{2m_e}\left(\frac{2\pi}{L}\right)^2, \tag{2.6}$$

which is small due to the macroscopic size of L. At 0 K, each of these energy ($\boldsymbol{k}$) states will begin to fill from the lowest energy, with 2 electrons per state due to the Pauli exclusion principle. The exact number of filled states will depend on the exact density of states available, whilst at higher temperatures, the Fermi-Dirac distribution becomes important. This filling is described in more detail in section 2.3. The E vs k dispersion relationship for the free electron model in a one-dimensional system is shown in the extended, reduced and periodic (repeated) zone schemes in Figure 2.4.

2.2.4 The nearly free electron model

Whilst the free electron model is simplistic, it does create parabolic bands similar to those seen in real crystal structures (Figure 2.5a, [59]), however it completely fails to describe the observed band gaps, inherent for the definition of semiconductors and insulators. Previously the potential due to the ion cores was neglected, however this must be included to understand the origin and formation of band gaps. As the electrons now experience a non-zero potential, they are no longer considered free, however the potential is still considered to be small, a perturbation to the free electron case. This new refined model is known as the nearly free electron model.

Whilst this model considers contributions from the ionic core potential, the strength of the potential experienced by the electrons is altered compared to that from a pure nucleus charge. This is as a result of screening from inner orbital electrons, with the strength of the potential also modified due to interaction from neighbouring lattice site potentials [57]. A schematic representation of the potential from the ion cores is shown in Figure 2.3.
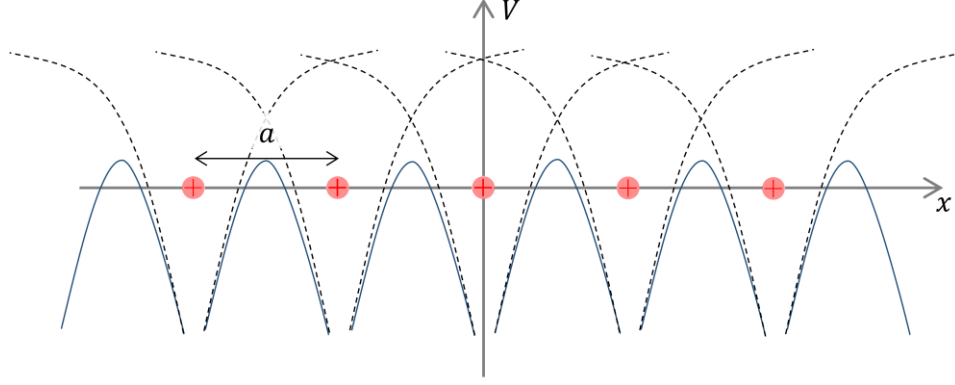
Figure 2.3 A 1D lattice of positive ionic cores (red), separated by a distance a. The potential (V) is shown as a grey dashed line for the unscreened potential that an electron would experience due to the pure ionic charge, where the strength of this potential is altered in the crystal (blue, solid) due to inner electron orbital screening and interactions from neighbouring lattice sites.

As in the free electron model, the specific form of the wavefunction solutions to the TISE in the nearly free electron model depend on the choice of the potential $V(\boldsymbol{r})$. Realistic atomic potentials are very complex and can almost never be solved analytically [57], however the restriction stated previously that the potential must be periodic allows the formation of more general solutions for any periodic potential. It was shown by F. Bloch that solutions for wavefunctions in a periodic potential are themselves periodic, taking the form

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = U_{\boldsymbol{k}}(\boldsymbol{r})e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \qquad (2.7)$$

where $U_{\boldsymbol{k}}(\boldsymbol{r})$ is another periodic function with the same periodicity as the lattice (i.e. $U_{\boldsymbol{k}}(\boldsymbol{r}) = U_{\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R})$). Alternatively, the Bloch solutions can be stated as

$$\psi_{\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R}) = \psi_{\boldsymbol{k}}(\boldsymbol{r})e^{i\boldsymbol{k}\cdot\boldsymbol{R}}, \qquad (2.8)$$

where the significance of this result is the same as for the free electron model (i.e. the probability density of finding an electron $|\psi|^2$ at a given $\boldsymbol{k}$ is equal at all lattice points).

The justification of the Bloch equation as solutions for periodic potentials will not be proved here, nor will a form for $U_{\boldsymbol{k}}$ be given, however examples of these can be found in many text books, such as "*Solid State Physics: An Introduction*" by Philip Hofmann [57]. Bloch solutions can be thought of as a fast varying periodic function associated with the potential of the ion, or equivalently the atomic orbitals ($U_{\boldsymbol{k}}$), modulated by a plane wave envelope solution extending throughout the crystal ($e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$). The fast varying function ($U_{\boldsymbol{k}}$)

is known as the Bloch function, where for the majority of physics concerned in the following sections of this thesis, the Bloch contribution to the wavefunction can be neglected, however for a full understanding of band formation and dispersion relations, it must be considered. The reason for neglecting the Bloch contribution subsequently is that the majority of the electronic properties (such as transport properties), are governed by the envelope function for the free electrons. This is known as the envelope function approximation, and is valid when the envelope is varying slowly on the scale of the lattice [60]. This means only the simpler envelope rather than the more complex Bloch wavefunction need be considered[4].

The solution for ψ given by equation (2.7) is qualitatively very similar to the free electron case (equation (2.3)) in that it is a plane wave solution, and therefore spread throughout the crystal, however it is now modified by the periodic nature of the lattice. The modulation by the ion cores explains how electrons are free to travel many multiples of the interatomic distance without scattering; instead of scattering from the cores, they are instead modulated by them to form a periodic wavefunction. It is only departures from this perfect periodicity that causes scattering events, whether they are from impurities, crystal atom defects, or lattice vibrations (phonons). From here onwards the vector notation will be dropped and a simple 1D case considered (i.e. $\boldsymbol{k} = (k_x, 0, 0) = k$).

The non-zero periodic potential in the nearly free electron model will have the effect of creating band gaps by not allowing travelling waves to form at specific k values (the Brillouin zone boundaries). These specific k values occur at multiples of $n\pi/a$, where a is the interatomic spacing. A wavenumber of $n\pi/a$ is the equivalent of having n half wavelengths between atomic sites (in k-space the lattice sites are separated by $2\pi/a$), therefore giving the periodicity of the wavefunction matching the periodicity of the potential. At these $k = n\pi/a$ values, standing wave solutions are formed rather than travelling waves, with the wave unable to propagate thorough the lattice. The band gap arises due to the fact that these localised wavefunction solutions are not unique, depending on whether the wavefunction is focussed on or between the lattice points (i.e. sin or cos like solutions), with a gap in energy between the two states proportional to the strength of the potential at the lattice points. This localisation on or between the lattice sites is analogous to bonding and antibonding states in atomic orbitals. As k increases beyond a

---

[4] This is important in the case of heterostructures where often across a boundary, the form of the Bloch wavefunction remains the same whilst the envelope changes. This modifies the boundary conditions to include the change in effective mass if only considering the envelope function.

zone boundary, the states are now considered in a higher band (similar to changing between s and p orbitals [60]).

Now, the solutions for k between $n\pi/a$ $(= nG/2)$ and $(n+1)\pi/a$, (and so also $-(n+1)\pi/a$ to $-n\pi/a$ $(= -nG/2)$) form a set of solutions, collectively known as a band, or equivalently a Brillouin zone (the 1ˢᵗ Brillouin zone is given by $n = 0$). The number of distinct states per band is given by the total number of unit cells in the lattice, N. This can be seen from the boundary conditions as described in the free electron model (namely $k = \pm\pi n/L$), and substituting $L = Na$, then $k = \pm\pi n/Na$.

The result of this lattice potential and bad gap formation is that now the dispersion is not completely parabolic, as the dispersion relation approaches a zone boundary, it deviates from the parabolic case and begins to flatten out. This non-parabolicity is particularly important in narrow band gap materials like InSb [61].

2.2.5 Energy bands and E vs k dispersion relations

Figure 2.4 shows the 1D band structure obtained from the free and nearly free electron models for the reduced, periodic (repeated) and extended zone schemes.

As can be seen from all of the representations, the dispersions for the free and nearly free electron models (black and coloured lines in Figure 2.4) are remarkably similar, with the differences due to the periodic potential most prominent at the zone boundaries (multiples of $k = \pi/a = G/2$). At these points band gaps have formed and the dispersion differs from the free electron parabola, flattening out.

In the extended zone scheme, only a single dispersion relation focussed on the $k = 0$ point in the first Brillouin zone is considered, with this plotted crossing many zone boundaries (increasing values of $k = n\pi/a = nG/2$). However this point and this zone are not special or unique. Due to the periodicity of the reciprocal lattice, the mathematical derivations could easily have started at any other reciprocal lattice point and the same result would be achieved, therefore this dispersion should occur at every lattice point. This is the periodic (repeated) zone scheme.

Alternatively, as can be seen from the periodic scheme, every dispersion curve crosses the 1ˢᵗ Brillouin zone at increasing energies (bands). The consequence of this is that the first zone contains all the necessary information to describe the band structure of the crystal. This is the reduced zone scheme.
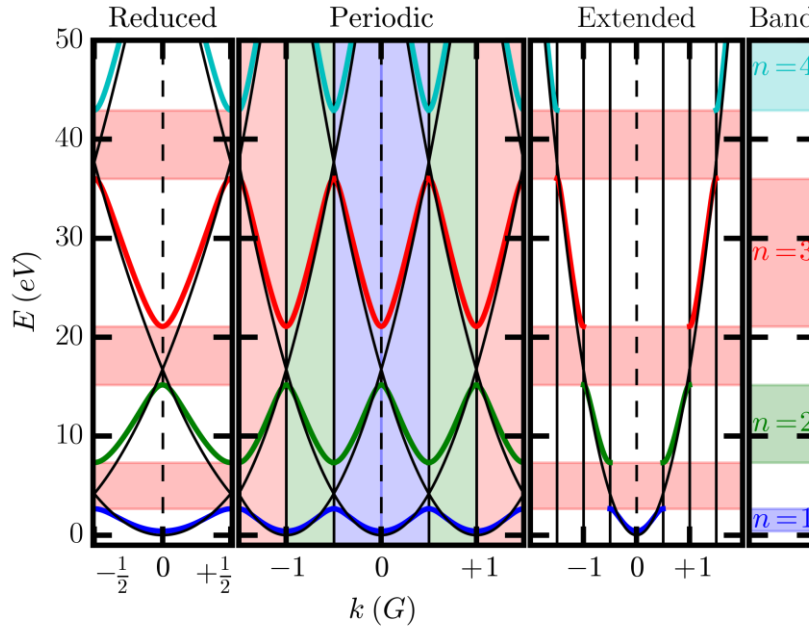
Figure 2.4 Band structures for the free (thin black curves) and nearly free (thick coloured curves) electron models, showing the opening of band gaps and forbidden energy regions (red horizontal shading) at Brillouin zone boundaries (vertical black lines). The left of Figure 2.4 shows the 1D dispersion relations in the reduced zone scheme, the centre dispersion is shown in the periodic (repeated) zone scheme, and the right dispersion is shown in the extended zone scheme. The different Brillouin zones shown in the periodic zone scheme are coloured such that zone 1 is blue, 2 green, 3 red and 4 cyan. The bands in all three schemes are coloured corresponding to their origin Brillouin zone (see text for further description). The x-axis for all dispersion relations is plotted as k in units of the reciprocal lattice vector G ($= 2\pi/a$). The differences between the parabolic free electron model and the nearly free electron model are clear, and most prominent at the Brillouin zone boundaries where the nearly free electron model flattens out and band gaps form.

Figure 2.4 shows bands plotted in only one-dimension (i.e. $k = k_x$; $k_y, k_z = 0$). If non-zero k values in the other dimensions are considered the dispersion relations can appear further complicated due to energy degeneracies of bands (see, e.g., Fig. 8. in Chapter 7 of reference [13]), however the basic understanding remains, and the bands are merely shifted in energy. This is how the band structure of InSb is shown in Figure 2.5a [59], where the band structure in the reduced zone scheme is shown in the principle reciprocal lattice directions.

In reality, this band structure shown in Figure 2.5a is more complex than is needed for most applications. To understand electronic transport in InSb and most other semiconductors, only the valence and conduction bands in the energy region around the band gap need to be considered. In InSb this occurs at the zone centre Γ point, where

these bands consist of an *s*-like conduction band, and *p*-type light and heavy hole valence bands, as well as a split-off hole valence band [41]. These are shown in Figure 2.5b.

(a)   (b)
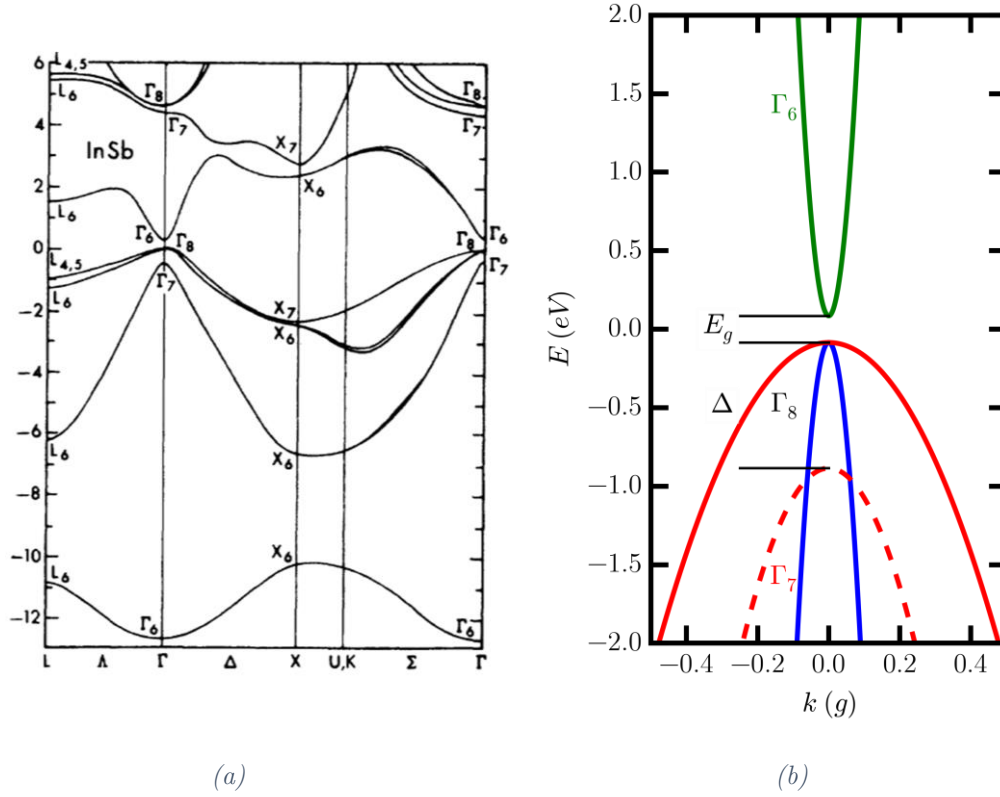
Figure 2.5 (a) Band diagram of InSb in the principle lattice directions shown in Figure 2.2b reproduced from Fig. 8 in reference [59]. The y-axis is energy in eV. The region of interest is the zone centre Γ, where the band gap is direct, with the conduction band given by Γ_6, the degenerate light and heavy hole bands by Γ_8 and the split off band by Γ_7. (b) shows this region of interest as a 4-band image, with the band gap E_g and split off energy Δ labelled. The conduction band (Γ_6) is shown in green, the light and heavy hole bands (Γ_8) are shown in blue and red respectively, and the split off band (Γ_7) is shown by a red dashed line.

2.2.6 Corrections to band structure calculations (effective mass approximation and $k \cdot p$ theory for non-parabolic bands)

As can be seen from the nearly free electron model as well as from the band diagram of InSb, near the band centre the dispersion is parabolic, however at higher energies and closer to the Brillouin zone edges, there is a large deviation. This deviation is known as non-parabolicity, and due to the extreme nature of InSb with its narrow band gap, this onset occurs at lower energies as compared to more conventional semiconductors such as GaAs. It can also be seen that the different bands curve at differing rates, with this having a significant effect on the transport.

The effective mass of the system is a way of describing how the electron's response to an applied force is modified due to the interaction with, and periodicity of, the lattice. This modification takes the form of an effective mass m^*, the apparent mass of the electron compared to the rest electron mass m_e, where m^* can make the electron mass appear larger or smaller than the free electron. This effective mass is given by the curvature of the dispersion relation such that

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\left(\frac{\partial^2 E}{\partial k^2}\right). \tag{2.9}$$

In a perfectly parabolic band the effective mass is independent of energy, however as InSb is highly non-parabolic, the subsequent change in mass with k must be included in any detailed analysis. The band edge effective masses in InSb are given in Table 1.1, along with a comparison to GaAs and Si. This table shows that InSb has a significantly reduced effective mass (the lightest of all the III-Vs [62]), and shows the different masses for the light and heavy hole corresponding to the different curvatures seen in the band structure. This effective mass then replaces the free electron mass in the equations derived previously. A schematic view of how the effective mass can differ in differing directions, and for the different bands is given in Figure 2.6, (following reference [63]).

(a)   (b)
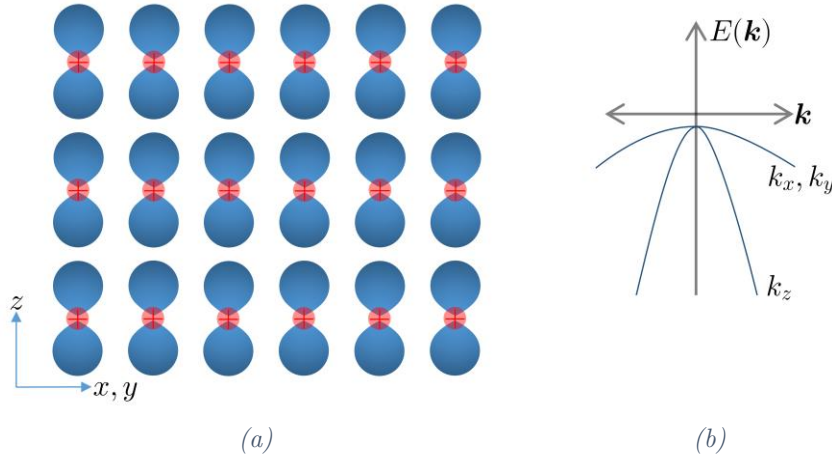
Figure 2.6 A lattice of p_z orbitals (a) forming a valence band, with a large orbital overlap in the z-direction and little overlap in the x- and y-directions, with the corresponding "heavy" bands in the k_x and k_y directions, and "light" k_z band (b), with the curvature giving the differing effective masses. The light and heavy bands are degenerate at $k = 0$.

To account for the non-parabolicity, a more detailed model than that employed for the nearly free electron model is needed, with this realised in the form of $\boldsymbol{k}\cdot\boldsymbol{p}$ theory using Kane's eight-band model (the eight bands are the conduction, light and heavy hole, and

split off bands, with 2 spin directions for each band). The details and implementation of this model will not be discussed here but can be found elsewhere [64, 65, 66].

The important results for this model are given in the form of a modified dispersion relation, such that

$$E(1 + \lambda E) = \frac{\hbar^2 k^2}{2m^*} \tag{2.10}$$

where $\lambda = 1/E_g$, and the effective mass is taken at the band edge [49]. The effect of this first order correction is shown in Figure 2.7 alongside a comparison to the much more parabolic GaAs.
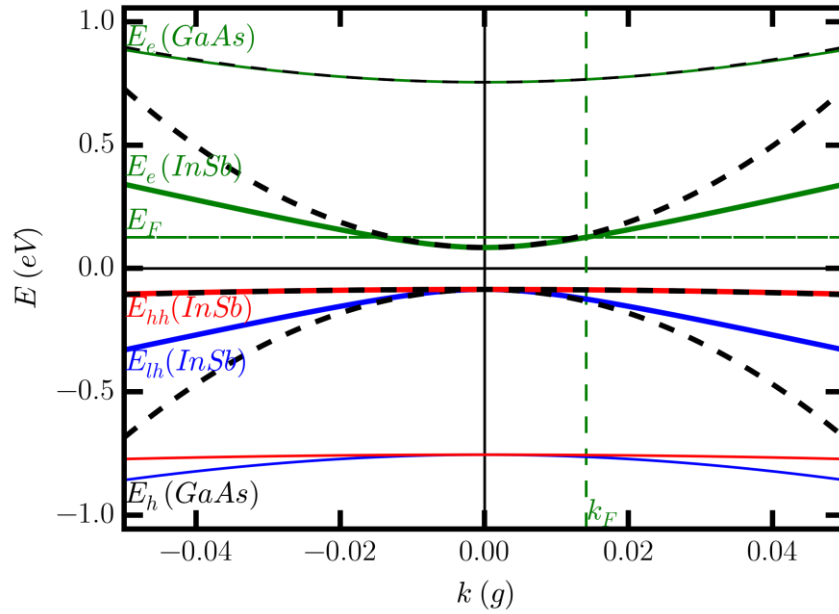
Figure 2.7 Comparison of InSb non-parabolic band structure (thick, solid lines), the parabolic approximation (black, dashed lines), and GaAs band structure (thin, solid lines). The conduction bands (E_e) are shown in green, the heavy hole valence band (E_{hh}) in red, and the light hole valence band (E_{lh}) in blue. The Fermi energy (E_F) and wavevector (k_F) for a carrier density $n_{2D} = 3 \times 10^{11}$ cm^{-2} are shown by green dashed lines. For simplicity the 0 of the energy scale is set to be mid-gap for both InSb and GaAs.

This non-parabolicity is also important for many things such as the density of states (section 2.3.3) and will give rise to parameters such as a non-parabolic effective mass.

When considering the variation of the band gap energy of InSb with temperature (T), as a result of the change in interatomic distance, the band gap is broadly empirically found to follow the relationship [49]

$$E_g = 0.24 - \frac{6 \times 10^{-4} \; T^2}{T + 500}. \qquad (2.11)$$

2.3 Fermi energy (E_F), Density of states (D.O.S.) and the Fermi-Dirac distribution $(f(E))$

Knowing the band parameters of InSb is vital to understanding the electronic transport, however the picture is not complete. To understand the electronic carrier concentration in a sample and the associated carrier energies, as well as how this varies with temperature, the density of states (D.O.S.) and the Fermi-Dirac distribution $(f(E))$ must first be understood. The density of states describes the number of available states in a given energy range, whilst the Fermi-Dirac distribution gives the probability of a state being occupied at a given energy for a given temperature. Together the two determine how the energy states of a sample are filled and, given a measured free carrier density, determine the Fermi energy, E_F (for continuous states, the energy of the highest occupied states at $0 \; K$, or alternatively the energy at which the Fermi-Dirac distribution is equal to 0.5).

2.3.1 3D density of states

It was shown in the section 2.2.3 that for an electron in a crystal (neglecting non-parabolicity), the energy of the states is given by $E = \hbar^2 k^2 / 2m^*$ (equation (2.5)), where the k state values are quantised in steps of $2\pi/L$ for a cube of side length L, such that

$$k_x = \frac{n_x 2\pi}{L}, k_y = \frac{n_y 2\pi}{L}, k_z = \frac{n_z 2\pi}{L}, \qquad (2.12)$$

where n_x etc. are positive integers (including 0).

Consideration of k in more than one dimension has clearly introduced an energy degeneracy, and so to be able to determine the energy of the highest filled state (the Fermi energy at $0 \; K$), the number of k states in a given energy range must be determined, and from there it is possible to determine how many are filled.

In 3D, this density of states in k space is determined by considering a spherical shell of inner radius k and thickness dk, where the radius of the filled states is given by k_{max}. The volume of the shell is then

$$V_{\text{shell}} = 4\pi k^2 dk. \qquad (2.13)$$

The k space volume of each state is given by

$$V_{\text{state}} = \left(\frac{2\pi}{L}\right)^3. \qquad (2.14)$$

The density of states, given by the number of available states in the range k to $k + dk$, is then given by the ratio of these two values, such that

$$g_{3D}(k)dk = \frac{V_{\text{shell}}}{V_{\text{state}}} = 2 \times \frac{k^2 L^3}{2\pi^2} dk \qquad (2.15)$$

where a factor of 2 has been introduced for spin degeneracy. It is then simple to show that for an energy range E to $E + dE$, the density of states (per unit volume) is given by

$$g_{3D}(E)dE = \frac{1}{(2\pi)^2} \left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}} dE. \qquad (2.16)$$

Integration of equation (2.16) over an energy range will give the number of available states in that range, and when combined with the Fermi-Dirac distribution will give the energy of the highest filled states (the Fermi Energy at $0\ K$) for a given electron carrier density.

2.3.2 Reduced dimensions (2D, 1D and 0D)

To calculate the density of states in reduced dimensions, where the k values are further restricted (quantised) in a given direction, a similar procedure is followed to the 3D spherical case. In 2D, where the k_z value is quantised such as in a quantum well, travelling wave solutions only exist in the k_x and k_y directions, so to determine the density of states involves determining the area of a circular annulus. In 1D such as in a quantum wire, this reduces to the length of a line segment whilst in 0D such as in a quantum dot, the k states are fully quantised in all directions, and so the density of states becomes a delta-function.

The density of states for each dimension is summarised in Table 2.1 for a single subband, along with the energy dependence. This is also summarised in Figure 2.8.

Table 2.1 The density of states $(g_D(E))$ in 3-, 2-, 1- and 0-dimensional systems, as well as the energy dependence (E^n) of the density of states in each dimesnion.

Dimension (D)	$g_D(E)$	E^n dependance
3D	$\frac{1}{(2\pi)^2}\left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}$	$E^{\frac{1}{2}}$
2D	$\frac{m^*}{\pi\hbar^2}$	E^0
1D	$\frac{\sqrt{2m^*}}{\pi\hbar} E^{-\frac{1}{2}}$	$E^{-\frac{1}{2}}$
0D	$2\delta(E)$	$\delta(E)$

As the samples studied in this thesis are based on InSb quantum wells, the 2D density of states is of primary importance. In 2D there is no energy dependence, meaning that all available states occur at the band edge (or subband energy if considering multiple states in a quantum well). Therefore this density of states should be extended to a sum over available subbands (or well energy states), such that for quantized states in the z-direction, numbered as n_z,

$$g_{2D}(E)dE = \sum_{n_z}\frac{m^*}{\pi\hbar^2}\theta\big(E - E_{n_z}\big)dE. \tag{2.17}$$

Here θ is the Heaviside step function [67], where $\theta\big(E - E_{n_z} < 0\big) = 0$ and $\theta\big(E - E_{n_z} \geq 0\big) = 1$. The density of states given by equation (2.17) is shown in Figure 2.8 as a staircase like function with an envelope following that of the 3D density of states due to the quadratic spacing of states in the quantum well (as can be seen from equation (2.5)).

Finally, to account for non-parabolicity, the density of states given by equation (2.17) is modified such that

$$g_{2D}(E)dE = \sum_{n_z} \frac{m^*}{\pi\hbar^2}(1+2\lambda E)\theta\big(E - E_{n_z}\big) \qquad (2.18)$$

where $\lambda = 1/E_g$ as previously [49].
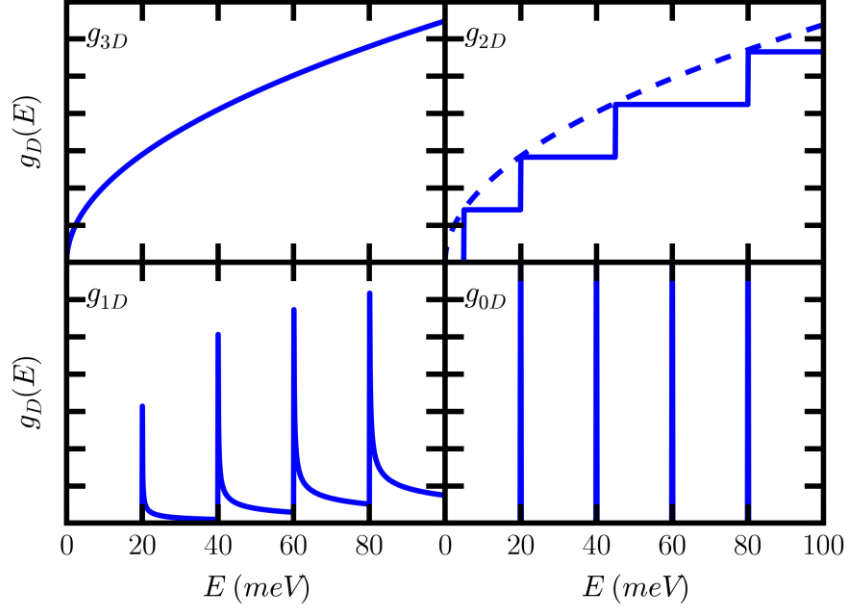
Figure 2.8 The density of states $(g_D(E))$ as a function of energy for 3D, 2D, 1D and 0D, illustrating the relationships given in Table 2.1. The 2D D.O.S. follows a staircase like shape (solid) with a 3D like envelope (dashed).

2.3.3 Fermi-Dirac distribution, Fermi energy and electron density

To determine the number of free carriers (n) and the Fermi energy (E_F) in a sample at a given temperature T, the density of states needs to be combined with the Fermi-Dirac distribution and integrated over an increasing energy range from the band edge (set to 0 for convenience) to infinity, such that

$$n = \int_0^\infty g_D(E)f(E)dE, \qquad (2.19)$$

where $f(E)$ is the Fermi-Dirac distribution given by

$$f(E) = \frac{1}{e^{\frac{E-E_F}{k_B T}} + 1}, \qquad (2.20)$$

where $k_B = 1.38 \times 10^{-21} J/K$ is the Boltzmann constant. The Fermi energy E_F has been used here as opposed to the chemical potential μ as the two are exactly equivalent at $0\ K$, and the difference at higher temperatures is not significant [60].

The Fermi-Dirac distribution gives the probability of a state at a given energy being occupied, where at $0\ K$ this distribution simplifies to a step function such that $f(E) = 0$ for $E < E_F$ and $f(E) = 1$ for $E > E_F$. When $E = E_F$ the distribution is equal to 0.5, with this definition of Fermi energy valid across temperature.
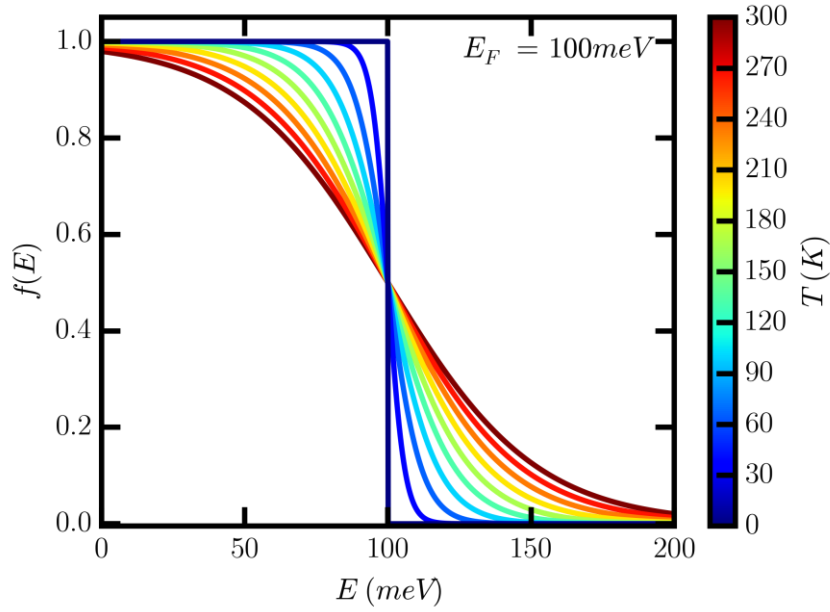
Figure 2.9 The Fermi-Dirac distribution $f(E)$ as a function of temperature from $0\ K$ to $300\ K$ for a Fermi energy of $E_F = 100\ meV$. At $0\ K$ the distribution is a step function with a step at E_F. As the temperature is increased, the step begins to broaden and smooth out, where the value of E_F gives an occupation of $f(E_F) = 0.5$.

In the case of a 2D quantum well at $0\ K$, and neglecting non-parabolicity, the integral in (2.19) can be simplified as there is no energy dependence in the density of states, giving

$$n_{2D} = \int_0^{E_F} g_{2D}(E) f(E) dE = \frac{m^*}{\pi \hbar^2} E_F \qquad (2.21)$$

where n_{2D} is the 2D carrier density. This can be rearranged to give the Fermi energy of the sample as

$$E_F = \frac{\pi \hbar^2}{m^*} n_{2D}. \qquad (2.22)$$

This definition of energy is equivalent to that given previously (equation (2.5)), such that (replacing k with the Fermi wavevector k_F), the Fermi energy is now given by

$$E_F = \frac{\hbar^2 k_F^2}{2m^*} \qquad (2.23)$$

which can be used to give two alternative definitions for k_F,

$$k_F = \frac{\sqrt{2m^* E_F}}{\hbar} = \sqrt{2\pi n_{2D}}. \qquad (2.24)$$

When accounting for non-parabolicity, the situation necessarily becomes more complex, and so the results will merely be stated here for the non-parabolic Fermi energy and effective mass (the definition for k_F does not change when accounting for non-parabolicity) [16, 49].

The Fermi energy is given by

$$E_F = \sqrt{\frac{E_g^2}{4} + \frac{E_g \pi \hbar^2 n_{2D}}{m^*}} - \frac{E_g}{2}. \qquad (2.25)$$

The non-parabolic energy dependant effective mass (m^*_{NP}) is given by

$$\frac{1}{m^*_{NP}} = \frac{1}{\hbar^2 k}\frac{\partial E(k)}{\partial k} \qquad (2.26)$$

which combined with the non-parabolic E vs k relation gives

$$m^*_{NP} = m^*\left(1 + \frac{2E}{Eg}\right), \qquad (2.27)$$

where m^* is the standard conduction band edge effective mass, as has been used previously. In a QW, the subband edge effective mass should be used instead of the conduction band edge effective mass, where this is modified from the conduction band edge value due to the confinement energy, however this correction is small and so has been neglected.

3 : REALISATION AND CHARACTERISATION OF INSB QUANTUM WELLS

3.1 Introduction

In this chapter, the realisation of the InSb QW structures studied will be discussed, where these are formed by "sandwiching" a thin layer of InSb between higher band gap material, in this case, AlInSb. The basics of semiconductor alloying and band gap alignments will be discussed, with a description of the physical process of material growth (via molecular beam epitaxy, MBE) given. A description of how this alloying is achieved to create the InSb material samples used in this thesis is given, and the effect on the electronic nature of the InSb, as well as the effect of induced strain, is discussed. This is followed by a detailed description of the device processing and fabrication steps performed to realise experimentally testable electronic devices.

This is followed by a description of the measurement techniques used to characterise these samples (the Hall effect and longitudinal resistance measurements to determine mobility and carrier density), as well as the methods of electrically measuring these samples as a function of temperature and magnetic field.

3.2 Growth of InSb Quantum Wells (Sample Material)

3.2.1 Semiconductor alloying (Vegard's Law)

All binary III-V or II-IV semiconductor compounds have well defined band gaps and effective masses, however, if intermediate values of these are desired (such as to absorb or emit light at a specific wavelength), or more complex heterostructures are desired, then the binary must be alloyed with a third element to form a ternary. In this case, InSb can be alloyed with aluminium to form the ternary, AlInSb. To achieve this alloying, a certain fraction of the indium atoms in the lattice are replaced with smaller Al atoms (where this fraction is known as the alloy fraction, x). This changes the properties of the crystal,

giving a smaller lattice parameter and therefore a larger band gap[5] [62]. In this manner very specific values can be achieved by carefully controlling the alloy fraction to achieve a material with properties that are somewhere intermediate between the binaries of InSb and AlSb.

The values for the lattice constant, band gap, and relative dielectric constant of the ternary $Al_xIn_{1-x}Sb$ can be determined to a first approximation by Vegard's law [68], which states that the properties of a semiconductor ternary material vary linearly between the two binary compounds. In reality the variation is not completely linear for all parameters however, where this is accounted for by introducing an empirical bowing parameter, b, such that (for example) the band gap is then given by

$$E_g(Al_xIn_{1-x}Sb) = xE_g(AlSb) + (1-x)E_g(InSb) + bx(1-x). \qquad (3.1)$$

The bowing parameter for the band gap of AlInSb is given by $b = -0.43$ [69] which introduces a small correction to the simple linear variation. At $300\ K$ this gives band gaps for the two $Al_xIn_{1-x}Sb$ compositions used in this thesis ($x = 0.1$ and $x = 0.15$) of $0.34\ eV$ and $0.43\ eV$ respectively, significantly larger than the $0.17\ eV$ InSb band gap (for reference, $E_g(AlSb) = 2.27\ eV$) [11]. This difference of ~170 meV (combined with a favourable conduction band offset, described in the following section) is why $Al_{0.1}In_{0.9}Sb$ can be used as a confining barrier material, with this value being significantly larger than the thermal energy of the electrons at room temperature ($k_BT \approx 25\ meV$).

InSb can of course be alloyed with elements other than Al, leading to different physical and electrical properties, and other binaries can be used as a different basis depending on the properties required. It is useful to visualise this in diagram form, as is shown in Figure 3.2 [62].

---

[5] Due to bowing, when alloying InSb to form InAsSb, a smaller band gap can be achieved (as seen in Figure 3.2 [62]).
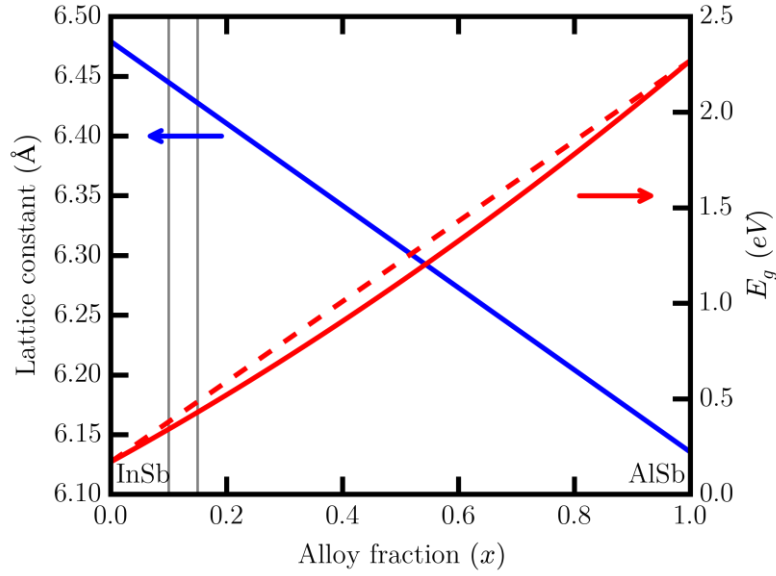
Figure 3.1 Vegard's law for determining the lattice constant (blue) and band gap (red) of the ternary $Al_xIn_{1-x}Sb$ as a function of alloy fraction x at 300 K. The effect of bowing is included for the band gap value (solid line), whilst the dashed line shows how the effect of neglecting the bowing is small. The two vertical grey lines denote the two alloy fractions used in this work ($x = 0.1$ and $x = 0.15$). Alloy fractions of $x = 0$ represents pure InSb and $x = 1$ represents pure AlSb.
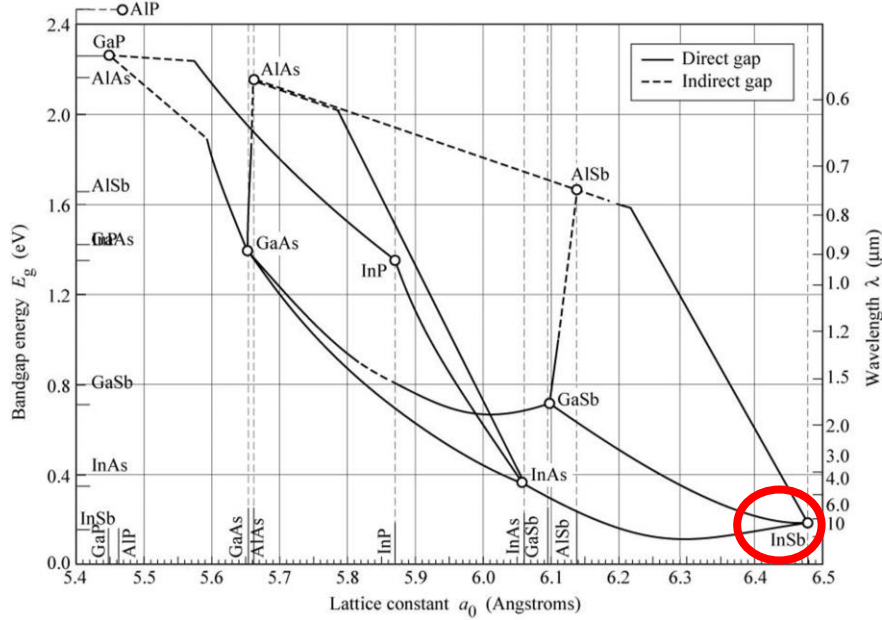
Figure 3.2 Lattice constant and band gap energy (and corresponding wavelength) for several III-V materials and their alloys [62]. Solid lines indicate direct band gap alloys, whereas dashed lines indicate indirect band gap alloys. InSb has the largest lattice constant and smallest band gap, however there is a large lattice mismatch between its nearest possible alloy material AlSb, as well as with common substrate materials such as semi-insulating GaAs.

Figure 3.2 shows the band gap energies and lattice constants for many III-V semiconductors, confirming the statement made in Chapter 1 that InSb is the most extreme of the III-V semiconductors, having the largest lattice constant and smallest band gap. It also shows how there are no lattice matched alloys to use in InSb heterostructures, and shows there is a large lattice mismatch with GaAs, often used as a semi-insulating (S.I.) substrate for MBE grown InSb samples. This mismatch introduces strain into the system, with this possibly leading to the introduction of defects. The effect of strain is briefly discussed in section 3.2.3, with further details given in section 3.3.

This induced strain changes the lattice constant of the InSb (see Figure 3.6 in section 3.2.3), where it is assumed that the lattice constant of the 30 nm InSb quantum well is strained to match that of the lower $Al_{0.1}In_{0.9}Sb$ layer as it is less than the critical thickness[6] (this is justified in section 3.3.1). The derivation of the strained parameters will not be discussed here as the justifications can be found elsewhere (e.g. [49, 70]), however the resultant values are given in Table 3.1 for a temperature of 10 K.

Table 3.1 Various parameters for the three main semiconductor matierals used in the samples studied in this work, $Al_{0.1}In_{0.9}Sb$, $Al_{0.15}In_{0.85}Sb$ and the InSb QW. The InSb lattice in the QW is strained to match that of the $Al_{0.1}In_{0.9}Sb$ layer. The concept of conduction band offsets are explained in the following section, where the specific values are referecned to the InSb band edge.

Alloy fraction x	Band gap E_g (meV)	Effective mass m^* (m_e)	Dielectric constant ε_r (ε_0)	Conduction band offset ΔE (meV)
0.1	445.9	0.024	17.05	123
0.15	548.9	0.029	16.75	191
0	255.4	0.014	17.7	0

---

[6] The critical thickness is the maximum thickness a strained epilayer can be without relaxing. Above this thickness, the strain cannot be accommodated and the layer relaxes, introducing defects.

3.2.2 Heterojunctions and band alignments

An important consideration when designing heterostructures is how the different band gap materials will align, as this alignment of the valence and conduction bands will affect the confinement of electrons and holes respectively. The alignment has the possibility to drastically change the electrical properties of any heterostructure, and so must be accounted for in heterostructure designs.

There are 3 main types of band gap alignment (denoted type I, II and III) as shown in Figure 3.3, where the different types are defined by whether the gaps are straddling, staggered or broken [63].
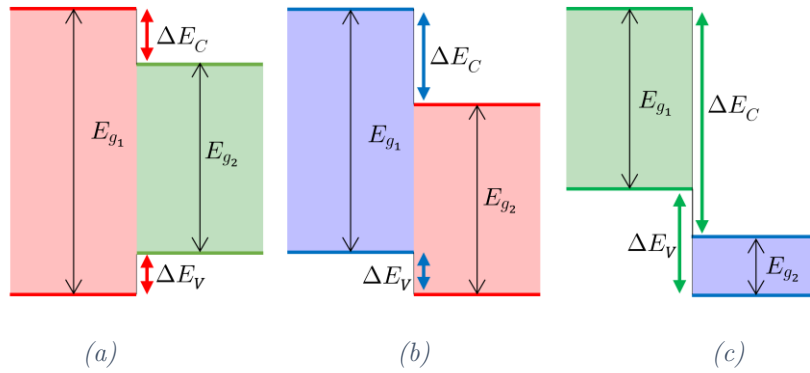
Figure 3.3 The three types of band gap alignment, (a) type I, straddling gap, (b) type II, staggered gap, and (c) type III, broken gap. The band gaps of each material are denoted as E_{g_1} and E_{g_2} for the larger and smaller band gap materials respectively, and the conduction and valence band offsets are shown (ΔE_C and ΔE_V respectively).

In type I straddling band gaps, the band gap of the larger band gap material (E_{g_1}) fully encompasses the band gap of the narrower band gap material (E_{g_2}), with the valence and conduction band energies of the narrow gap material within those of the larger gap. The band alignments between InSb/AlInSb for the alloy fractions used in this thesis are of this type, with the values for the conduction band offsets (ΔE_C) given in Table 3.1.

Type II and type III alignments occur when the narrow band gap material does not align completely within the gap of the larger band gap material, and is instead partially offset so either the conduction or valence band lies in the gap (type II, staggered gap), or the band alignments are completely offset so both the conduction and valence bands of the narrow gap material lie wholly within either the conduction or valence bands of the larger gap material (type III, broken gap).

The straddling gap as seen in InSb/AlInSb allows the formation of a quantum well for both electrons and holes, where a layer of InSb is sandwiched between two barrier layers

of AlInSb. This means free electrons residing in the conduction band of the InSb layer are at a lower energy than the surrounding regions, with these barriers giving rise to confinement in the perpendicular (z-) direction. If the width of the quantum well is of the order of the deBroglie wavelength of the electrons, then quantum mechanical effects dominate, and in the z-direction, discrete electronic energy levels will form, whilst still being free in the other two dimensions. This reduces the system to a two dimensional plane, where the electrons can be considered a two dimensional electron gas (2DEG), and the physics discussed earlier for 2D systems (such as the density of states) becomes relevant. The deBroglie wavelength (λ) is given by

$$\lambda = \frac{h}{\sqrt{2m^* k_B T}} \qquad\qquad (3.2)$$

where h is Planck's constant ($h = 6.63 \times 10^{-34}\ Js$), giving $\lambda \sim 65\ nm$ at $300\ K$ in InSb. The well dimensions used in this thesis ($30\ nm$) are well below this value.

The specific values for band alignments and offsets (i.e. ΔE_C and ΔE_V) are generally given by Anderson's rule [71], which states that for the two semiconductor materials combined at a heterojunction, the vacuum energy levels are equal and aligned. From here, each material's electron affinity (χ, the energy difference between the top of the conduction band and the vacuum level) and specific band gap determine the alignment of the bands between the two materials. Many semiconductor materials such as GaAs and InSb fail to follow Anderson's rule however, where they instead follow a more empirical conduction band : valence band offset ratio, where quoted values often range between ~60% : 40% [72] to ~70% : 30% [73]. Where relevant in this work, the middle of this range, i.e. 65% : 35%, is used. This means that if the mid-gap of InSb is taken as a reference (this is E_F at $0\ K$ for an undoped system, set at 0 for ease), the mid-gap of the AlInSb barrier will be offset higher relative to the InSb. Therefore the specific conduction band offsets and conduction band energies are given by

$$\Delta E_g = E_g(\text{AlInSb}) - E_g(\text{InSb}) \qquad\qquad (3.3)$$

$$\Delta E_g = \Delta E_C + \Delta E_V \qquad\qquad (3.4)$$

$$\Delta E_C = 0.65 \Delta E_g = \frac{\Delta E_g}{1 + 35/65} \qquad\qquad (3.5)$$

$$E_C(\text{AlInSb}) = \frac{E_g(\text{InSb})}{2} + \Delta E_C \qquad (3.6)$$

where E_C is the conduction band energy. This is how the band offsets given in Table 3.1 were determined.

The InSb layer sandwiched between two AlInSb layers now forms a QW, where the ground state energy (E_{QW}) is raised above the conduction band edge, with the energy of this state (for an infinite potential approximation) given by equations (2.4) and (2.5), such that

$$E_{QW} = \frac{\pi^2 \hbar^2}{2m^* L^2} \qquad (3.7)$$

where L is width of the QW (30 nm). For a well of this width, the ground sate energy is ~30 mev above the conduction band edge. A comparison between this infinite well approximation and a full Schrödinger solution [74] is given in Figure 3.4.
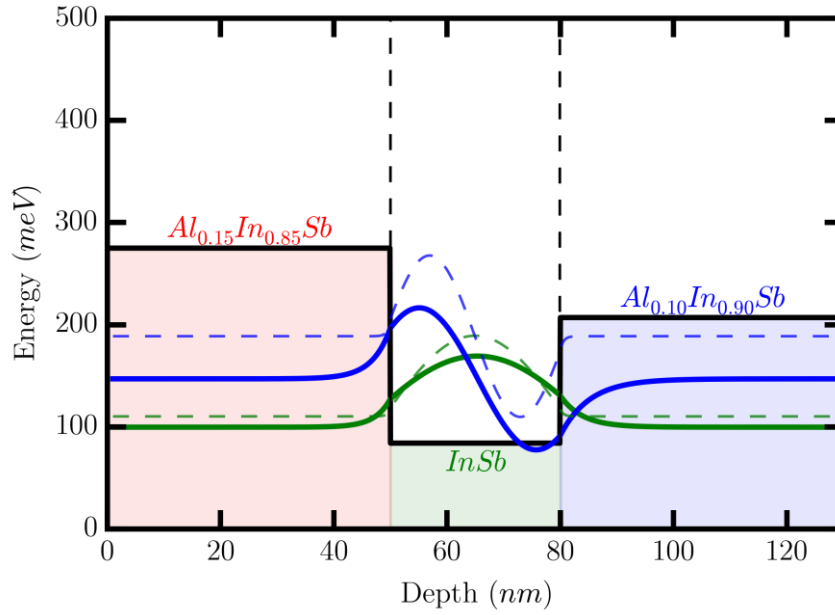
Figure 3.4 Comparison between Schrödinger wavefunction solutions for a finite well (thick, solid lines) [74] and an infinite well approximation given by equation (3.7) (thin, dashed lines). For the ground state wavefunction the infinite well solution gives a good approximation to the ground state energy, however this diverges for higher energy states.

The final correction to the band alignments of InSb is given by the alignments relative to the surface. Previously the mid-gap Fermi energy of the InSb was set to 0 for

convenience, however in InSb/AlInSb there is a significant surface pinning. This is due to the termination and reconstruction of the crystal at the surface, leading to a large number of states located at approximately the mid-gap value of the upper $Al_{0.15}In_{0.85}Sb$ barrier[7]. Therefore the Fermi energy in the upper barrier is pinned to the mid-gap of the AlInSb, with the band alignments of the InSb and lower barrier layers referenced to this value. This is shown in Figure 3.5. In Figure 3.5 the band bending due to charge movement to align the Fermi levels of each material has been neglected, as has any doping (intentional or otherwise) in the upper barrier and lower layers.
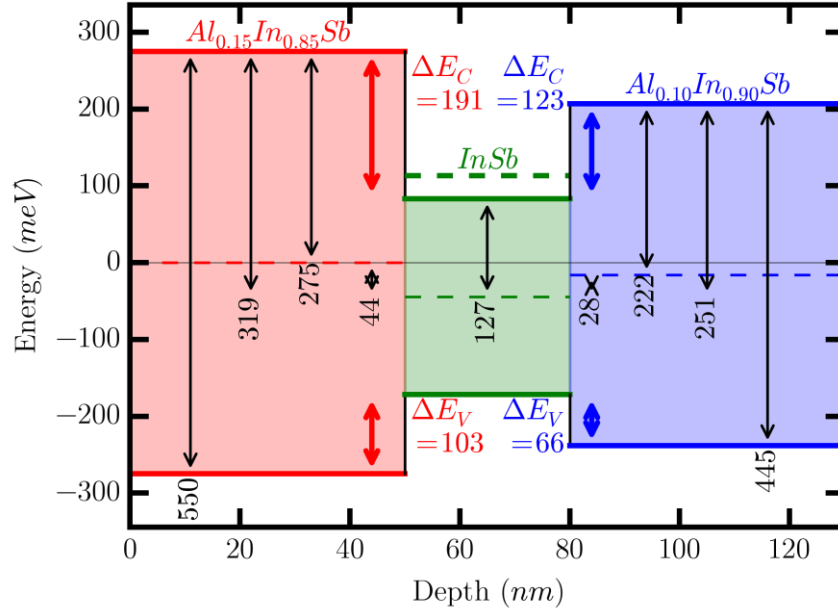
Figure 3.5 Band alignments and offsets for $Al_{0.15}In_{0.85}Sb$, InSb and $Al_{0.1}In_{0.9}Sb$ layers respectively, neglecting all doping and band bending due to charge movement to align the respective Fermi levels. All energy values are in meV. The surface layer Fermi level is mid-gap pinned at 0 (the mid-gap of each material is marked by a thin dashed line). The conduction and valence band offsets (ΔE_C and ΔE_V) for each layer are marked, as well as other relative energy values. The ground state quantum well energy (E_{QW}) of ~30 meV is denoted by the thick, green dashed line.

---

[7] A range of values for this pinning have been reported, including from mid-gap pinning [131, 132], to ~1/3rd gap pinning [148] for the band structure under a Schottky contact. Mid-gap pinning only has been considered here.

3.2.3 Physical growth (MBE, doping and strain)

The samples studied in this thesis are InSb Quantum well samples primarily provided by the national III-V semiconductor centre [8] in Sheffield (though historical samples produced by QinetiQ, Malvern have also been studied [17, 16, 75, 76]). The standard basis of each sample is a semi-insulating (S.I.) GaAs (100) wafer[9], onto which different layers of InSb or AlInSb are grown to produce a final sample wafer. The process by which all samples are grown is molecular beam epitaxy (MBE), a process in which layers are grown one atomic layer at a time, and where the growth rate and material composition properties can be precisely controlled. By varying the growth (substrate) temperature and the relative amounts of indium, antimony and aluminium in the growth chamber (via cell temperature and corresponding flux rate), the precise alloy fraction can be controlled. It is not straight forward to grow InSb on GaAs however as there is a large lattice mismatch between the two compounds (approximately 14.6% for InSb on GaAs [22], see Figure 3.2), which if not accounted for, would lead to strain and growth related defects which could adversely affect the physical and electrical quality of the sample.

Strain in a crystal heterojunction arises due to the fact the regular spaced atoms in each of the two layers of the heterojunction cannot perfectly line up to form crystal bonds directly across the junction [63]. It is assumed that the GaAs substrate is sufficiently thick compared to the MBE grown epilayer on top, that the GaAs lattice constant remains unchanged after material is added. Therefore either the epitaxial layer must reduce its lattice constant to match that of GaAs (and so become strained) or a bond must be missed (a misfit dislocation or defect). A missing bond partially relaxes the built in strain in the crystal, though disrupts the regular periodicity of the crystal, adversely affecting the sample quality. For now it is assumed the buffer layer will be completely relaxed due to the large lattice mismatch between GaAs and $Al_{0.1}In_{0.9}Sb$, but the InSb quantum well will be strained to match that of the $Al_{0.1}In_{0.9}Sb$ due to the smaller mismatch and thinner width of the layer. The critical thickness of epilayers, and the associated dislocations and defects formed will be discussed further in section 3.3.

---

[8] Formerly the EPSRC National Centre for III-V Technologies, presently the EPSRC National Epitaxy Facility, North Campus, University of Sheffield, Sheffield S3 7HQ, United Kingdom.

[9] A semi-insulating wafer substrate is material that has been specifically doped with deep level trap states, reducing the free carrier density and therefore the conductivity. For the GaAs substrates here, the suppliers, Wafer Technology Ltd, state the free carrier density is typically ~10^7 cm^{-3}, whislt the resistivity is typicall ~10^8 Ωcm with a corresponding mobility of ~6000 $cm^2 V^{-1} s^{-1}$.

(a)   (b)   (c)
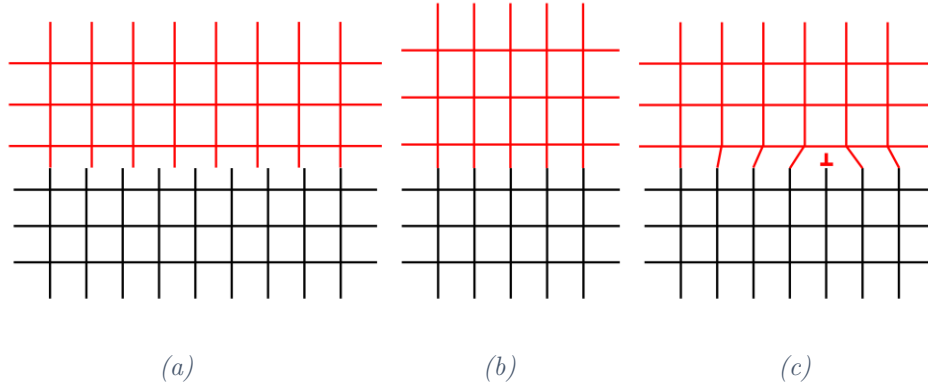
Figure 3.6 Schematic example of strain due to lattice mismatch at a heterojunction interface. The bottom smaller lattice constant material represents the GaAs substrate, whilst the red upper layer represents the larger lattice constant $Al_{0.1}In_{0.9}Sb$ layer (the spacing of lattice points is directly proportional to the actual lattice constants of the two materials. In (a), the two materials are brought together, showing the clear difference in lattice constant. In (b), the AlInSb layer is strained to match that of the GaAs, reducing in the horizontal direction and extending in the vertical. In (c) some of the strain has been released by the formation of a misfit dislocation ($\perp$). In this diagram this appears as a single point, however in a real crystal, this will form a line of misfit dislocations.

The process of growing a sample via MBE involves placing an initial substrate into the evacuated growth chamber (Figure 3.7), where ultrahigh vacuum pressures of the order of 10^{-11} *mbar* are achieved [77]. Facing the surface of the substrate are several Knudsen effusion cells, where each cell holds a very high purity element (such as Al, In, Ga, etc.). These cells are heated to very high temperatures such that the material begins to evaporate from the surface, where the material then enters the growth chamber as a beam of fast moving molecules, hitting the substrate surface. The temperature of the substrate, as well as the energy of the incoming atoms determines the incorporation of each element as well as the crystal reconstruction of the surface. These also determine how rapidly an atom impinging on the substrate surface will form a chemical bond, locking in its position in the crystal. Therefore growth temperatures and growth rates during MBE growth are critical to achieve the desired sample quality and layer thicknesses, where sample growth rates are typically approximately 1 $\mu m/hr$ or 1 ML/s (where 1 ML is one monolayer) [77]. The substrates are also typically rotated at a constant rate so as to aid in the uniformity of growth across the whole surface. The use of simple mechanical shutters in front of the effusion cells allows for the interruption of the flux of a given element, allowing for atomically sharp elemental transitions in the growth layers. Finally, monitoring of the progress of the growth is achieved via reflected high-energy electron diffraction (RHEED), whereby a beam of electrons is directed at the surface of the grown layers at a shallow angle, where it is then diffracted onto a fluorescent screen. The resultant diffraction pattern gives information on the growth of the epilayer such as the crystal reconstruction

of the surface, whilst periodic changes in the intensity of the pattern represent monolayer growth. It is in this way that complex heterostructures such as QWs can be achieved.
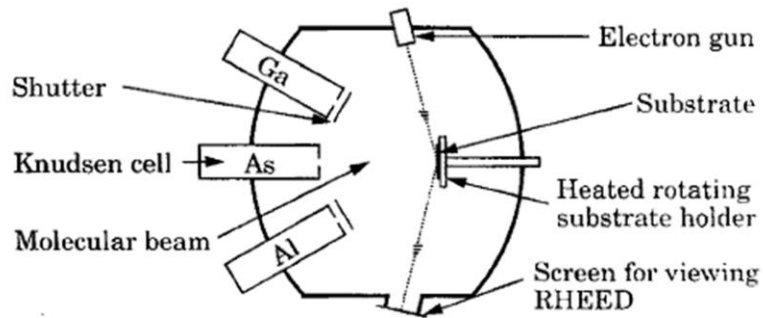
Figure 3.7 Simplified schematic diagram of a molecular beam epitaxy (MBE) chamber, reproduced from Figure 3.2 in reference [63]. This diagram is significantly simplified, though shows the relevant aspects for material growth. See the text for a description of the parts labelled and their operation.

Intentional doping of samples is achieved in the same way, where for InSb, n-type doping can be achieved by inclusion of tellurium (Te) atoms during growth. A Te atom naturally takes the place of an Sb atom in the crystal lattice and therefore has an extra non-bonding valence electron that it can donate, giving a free carrier in the conduction band. The incorporation of Te into InSb structures grown by MBE is low however (i.e. much of the flux onto the surface results in atoms that do not remain in the lattice), however the activation (the number of Te atoms that are included in the lattice that donate an electron to the conduction band) is very high [17, 78]. An alternative dopant is silicon, where silicon has a significantly higher incorporation than Te, however in these samples, Si is amphoteric, meaning it can sit on both the In and the Sb lattice sites, acting as both a donor and an acceptor [79]. This reduces the electron density donated from the silicon dopant for a given level of doping, and is why only Te doped samples have been considered in this work.

Dopant can be incorporated throughout the whole sample, such that the charged ionic cores of the dopant atoms are in close proximity to the conducting electrons, or the doping can be physically separated from the conducting electrons. This separation doping technique is known as modulation doping, where in these QW samples, the upper $Al_{0.15}In_{0.85}Sb$ barrier is Te doped, whilst conduction occurs in the InSb QW layer. This helps to realise high mobility samples by physically separating the charged ions that act as scattering centres from the electrons in the conducting channel.

Doping can be further characterised into 3D (bulk or slab doping) and 2D (δ-doping). In slab doped samples the dopant is spread evenly through the doped layer, spread over a finite 3D thickness, whereas for δ-doping all the dopant is included into a highly doped single layer of growth (ideally a 2D plane or sheet of charge) [80]. An advantage of δ-doping as compared to slab doping is that a larger amount of charge can be placed in a single layer without exceeding the critical thickness of the layer and introducing structural defects [81, 82]. Further to this, the δ-layer can be placed in close proximity to the well, giving a larger contribution of charge carriers to the well for a given level of doping. Finally this method of doping results in a more uniform positively charged layer after the donation of electrons as compared to the slab doped case, however a proportion of electrons can remain in this layer, acting to partially screen the dopant ion cores. As this is a single plane, the screening in the δ-doped case is more significant than the slab doped case, reducing the level of scattering. As it is possible to achieve higher carrier densities using δ-doping whilst also introducing screening, δ-doping is used in many practical devices, with all the QW samples studied in this work Te δ-doped. This doping is practically achieved by briefly interrupting the flux of Al and In atoms[10] and instead having a flux of pure Te. This Te flux is typically only maintained for a few minutes compared to the multiple hour growth times for the whole sample.

Ideally the δ-doping layer would be a perfectly sharp atomic layer, however as the incorporation of Te is low, the dopant has a tendency to drag on the growth plane [17, 48], i.e. if growth is in the z-direction, such that the surface is at $z = 0$, then the dopant is not fixed at a depth ($z = -25\ nm$ in these samples), but instead a proportion drags during growth to $z + dz$. This can be seen in Figure 3.8, showing SIMS (secondary ion mass spectroscopy) measurements performed on a typical AlInSb/InSb sample. The amount of doping is clearly peaked at the intended δ-dopant plane, however a significant proportion of dopant is present in higher layers closer to the surface, with a ratio of dopant at the surface compared to the δ-plane of approximately 1/9th. In subsequent Schrödinger-Poisson and transport modelling it is necessary to include the effect of this dopant distribution.

---

[10] A low level flux of the group V element (Sb) is often maintained as this has a higher rate of evaporation from the sample surface, so the low flux maintains the correct atomic ratios on the surface [77].
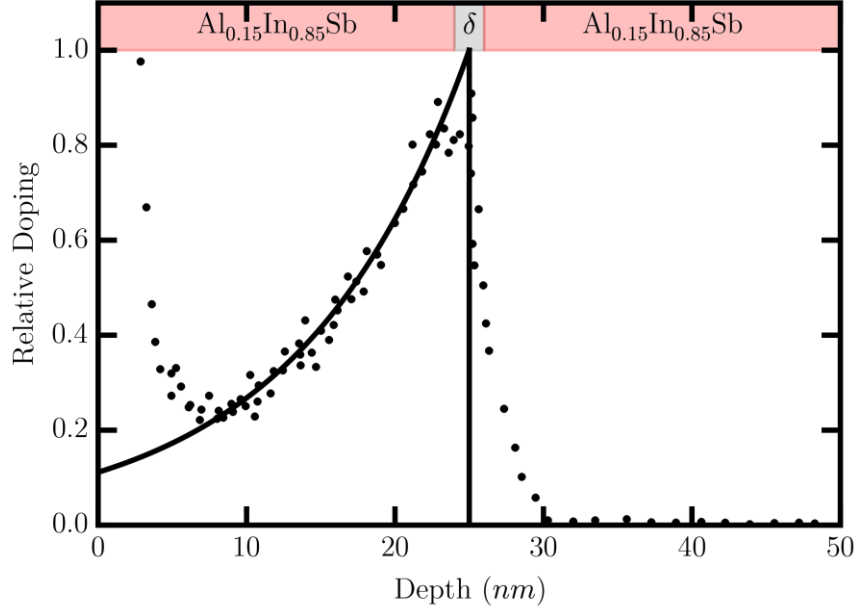
Figure 3.8 Secondary ion mass spectrometry (SIMS) measurements of relative level of doping in the top cap of a typical δ-doped QW sample with the δ-layer located 25 nm below the surface. The measured data (black points) show a clear peak at a depth of 25 nm, with this decreasing exponentially towards the surface (the increase near the surface is a side-effect of the SIMS measurement and not representative of the actual distribution [17, 48]). An exponential fit to the data (black line) shows a decrease to a value of $1/9^{th}$ at the surface as compared to the δ-plane.

3.2.4 Sample structure

The InSb QW samples studied in this work are based on one primary structure, with each sample varying from this base structure in either the level of δ-doping, the thickness of the spacer layer between the well and the δ-plane, or both. This basic structure is outlined in Table 3.2, where a description of the individual layers is given in the following paragraphs, and a summary of the primary samples is given in Table 3.3. Where any sample studied has a different structure, this will be noted when relevant.

In growth order, an initial accommodation layer of AlSb is grown. This is grown "hot", and acts as a "plastic" layer, helping to alleviate the strain due to the lattice constant mismatch between the GaAs substrate and subsequently grown layers [83]. The plastic nature allows the surface atoms to be mobile, and glide defects out of the sample, rather than up through the growth planes.

To further accommodate for the lattice mismatch, a ~3 μm thick buffer layer of $Al_{0.1}In_{0.9}Sb$ is then grown, with this layer also acting as the lower barrier for the quantum well. This wide layer theoretically ensures that at the surface, the lattice constant is that

of the fully relaxed, unstrained AlInSb. The buffer has to be significantly thick so as to allow time for the threading dislocations to either self-annihilate, annihilate with other threading dislocations, or terminate at the edges of the sample. Whilst thicknesses above $3 \ \mu m$ can marginally reduce the number of defects present at the surface further, the thicker buffer introduces negative impacts on device performance, such as device isolation and alloy (interface) roughness. The thicker buffer also increases the difficulty in device processing, such as with air-bridged gates, as well as increasing the growth time and cost for MBE. As an added benefit however, a thick buffer layer acts to absorb impurities present in the MBE growth chamber, helping to improve the quality of growth in the active region of the samples [84].

On top of this, the 30 nm quantum well is grown, followed by the 50 nm top cap of $Al_{0.15}In_{0.85}Sb$, with this acting as the upper barrier, completing the structure. The top barrier additionally includes a δ-doping layer a set distance (spacing) above the quantum well, where for the majority of samples studied, this spacer layer is 25 nm thick.

Table 3.2 Ideal standard sample structure for the AlInSb/InSb quantum well structures studied in this work. A description of the layers and their purpose is given in the text.

Layer		Composition	Thickness
Surface		–	–
	Top cap	$Al_{0.15}In_{0.85}Sb$	25 nm
Upper Barrier	δ-doping	Te	1 $monolayer$
	Spacer	$Al_{0.15}In_{0.85}Sb$	25 nm
Quantum Well		InSb	30 nm
Lower Barrier / Buffer		$Al_{0.1}In_{0.9}Sb$	3 μm
Accommodation		'Hot' AlSb	0.5 μm
Substrate		GaAs	–

Samples provided from the national epitaxy facility were provided over 4 growth runs (batches), where in each batch samples were grown sequentially so as to maintain MBE conditions (such as background impurity density). In each batch, a sample with a standard level of doping was grown for comparison to previous batches, whilst between batches growth conditions were altered to determine optimal parameters. The results of this optimisation were realised in batch IV, with this batch representing the majority of samples studied in this thesis. Batch IV consists of a range of samples doped at various levels to examine the trend of mobility with carrier density, with variations of spacer also realised in this batch. Table 3.3 lists the various samples from this growth batch, whilst the following section details how the conduction band profile and carrier density of these samples is investigated via Schrödinger-Poisson modelling.

Table 3.3 Summary of batch IV samples provided by the national epitaxy facility and studied in this work. There are two variants, doping variants, where the spacer is maintained at 25 nm and the level of dopant is varied, and spacer variants where the level of doping is maintained and the spacer is altered. For all samples the total top cap thickness is 50 nm. The measured values are obtained via Hall effect measurements with 2 carrier fitting at 3 K. Where there are two values for a measurement these are repeat measurements for a different device made from the same sample.

Variant	Sample name	Intended doping level $(\times 10^{11} \ cm^{-2})$	Spacer Thickness (nm)	Measured 3 K Carrier Density n_{2D} $(\times \ 10^{11} \ cm^{-2})$	Measured 3 K mobility μ (cm^2/Vs)
Doping	SF1054	5	25	3.54 3.08	243,000 239,000
	SF1055	3.5	25	2.87	255,000
	SF1056	2	25	2.14	195,000
	SF1057	6.5	25	4.04 4.01	226,000 238,000
	SF1058	8	25	4.33 4.40	207,000 206,500
Spacer	SF1066	5	30	2.93 2.96	250,000 242,000
	SF1067	5	35	1.08 1.55	151,000 119,000
	SF1068	5	20	4.85	174,000

3.2.5 Schrödinger-Poisson

The band structure of a QW sample determines many of its electrical properties, including determining the electron carrier density and the subband energy level spacing and filling. To a first approximation, the band structure is given by the band alignments outlined earlier, however in a real sample, the Fermi levels must also align, with this achieved through movement of charge within the sample. The specific shape of the band structure is then determined by the spatial distribution of charge, whilst the distribution of charge is determined by the shape of the band structure and the confined wavefunction (ψ). Therefore to determine the band structure a self-consistent method is needed, where an initial band structure is given (as in Figure 3.5), and the wavefunction solution to the Schrödinger equation is determined. The charge density distribution (ρ) is then given by $q|\psi|^2$, where q is the charge of the carriers (-1.6×10^{-19} C for electrons), following which Poisson's equation,

$$\nabla^2 V = -\frac{\rho}{\varepsilon} \text{ or } \nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon} \, , \tag{3.8}$$

is solved to determine the potential (V) due to the charge density ($\boldsymbol{E}$ is the electric field, and $\varepsilon = \varepsilon_r \varepsilon_0$ is the permittivity of the material). This new potential is in addition to the potential profile given by the conduction band edge, and so will cause a redistribution of charge (according to the solution to Schrödinger's equation). To determine the final band structure and well wavefunction states, this process of solving the Schrödinger and Poisson (S.P.) equations must be performed iteratively until a stable condition is reached. The specific details of how the Schrödinger equation is solved numerically are discussed further in Appendix A.

The change of the charge distribution and the associated change in potential has the effect of creating electric fields within the sample, most prominent between the accumulation of electrons in the well and the positive ion cores of the donor atoms in the δ-doped plane. These electric fields cause the straight conduction bend edges to "bend", with the sharpness of the bend related to the strength and extent of the charge distribution. An example of this is shown in Figure 3.9 where the band structure for a 1 nm δ-doped plane is compared to an exponential decay of dopant as given by SIMS measurements (as in Figure 3.8).
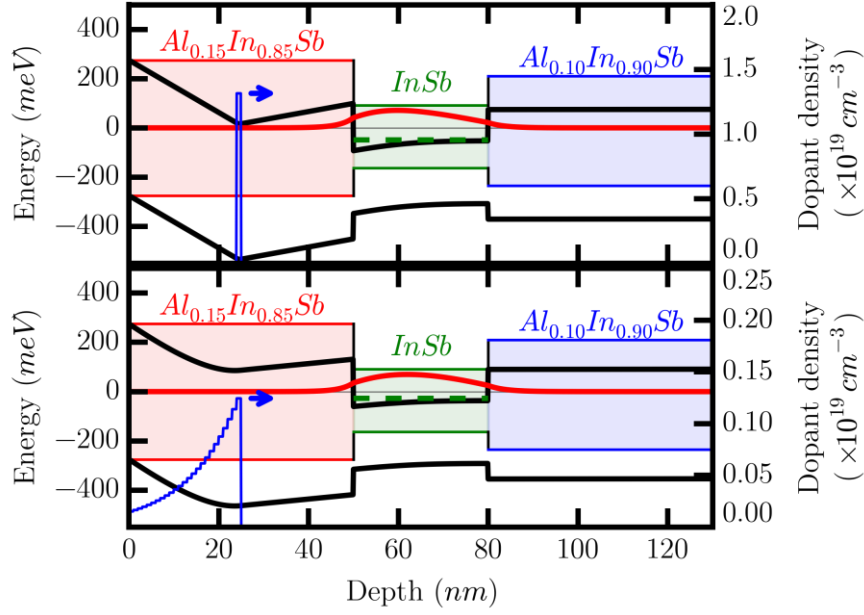
Figure 3.9 Self-consistent Schrödinger-Poisson solutions [74] for a 30 nm QW with AlInSb barriers with perfect δ-doping (top) and including dopant dragging following an exponential decay (bottom) for the same level of doping. The faint lines and shading show the 1ˢᵗ order band alignments with no doping or charge movement (as Figure 3.5). The thick black lines denote the calculated band structure including dopant and charge movement, whilst the red line gives the electron wavefunction in the well at a ground state energy (E_{QW}) given by the green dashed line (all left axis). The Fermi energy (E_F) is set to 0 meV and pined at the mid-gap of the surface layer. The thin blue line shows dopant profile on the right axis (note the change of scale). For the perfect δ-doping, a carrier density of $n_{2D} = 3 \times 10^{11}$ cm^{-2} is achieved, whereas for the exponential decay a reduced $n_{2D} = 1.6 \times 10^{11}$ cm^{-2} is achieved.

It is clear from Figure 3.9 that the distribution of dopant charge drastically affects the conduction band profile, altering the shape of the barrier to the surface and the amount of charge donated to the well, as well as the scattering rates due to these remote ionised impurities. Chapter 4 : Schrödinger-Poisson and Transport Modelling analyses these effects further.

Finally, the Schrödinger-Poisson solution also determines the carrier density in the sample. For a single filled subband, this is calculated by taking the difference between the energy of the subband (the ground state energy of the QW (E_{QW}), given by the Schrödinger solution, approximated by equation (3.7)) and the Fermi energy E_F, and dividing by the density of states (given by equation (2.18)), such that

$$n_{2D} = \frac{E_F - E_{QW}}{g_{2D}(E)} \qquad (3.9)$$

In the S.P. solutions in this work, the Fermi energy is generally fixed at 0 meV and the band energies are measured relative to this. This carrier density will be less than the number of donor atoms input into the sample due to charge donation to the surface states mentioned previously.

3.3 Standard Defects in MBE Grown InSb on GaAs

This section is primarily concerned with the structural defects that can occur in crystal structures during MBE growth due to strain mismatched layers. This includes the formation and propagation of these defects through the crystal, including misfit and threading dislocations, screw-like growth, and stacking faults and microtwin defects. The information in the following section can be found in many references (many of which go into far more detail than is given here), where the primary sources used here are given in references [77, 85, 86, 87, 88, 89, 90, 91, 92, 93].

3.3.1 Critical thickness

The maximum thickness a strained layer can achieve before a misfit dislocation occurs is known as the critical thickness. This thickness is a balance between the energy required to break multiple bonds and relieve strain (proportional to the sample area), and the elastic potential energy built into the strained system (proportional to thickness) [63]. Below a certain thickness it is energetically favourable for the lattice to remain strained, however above this thickness, it is favourable for the lattice to relax by introducing misfit dislocations. This energy balance is the basis of the Matthews and Blakeslee model for critical thickness [94], and whilst many other models exist, this model is generally considered to give a conservative estimation, and so any layer with a thickness less than the critical thickness from this model should remain strained. The critical thickness (h_c) for a heterojunction according to this model given by [87, 95]

$$h_c = \frac{b}{8\pi f \cos(\lambda)} \left(\frac{1 - \nu \cos^2(\theta)}{1 + \nu} \right) \left(\ln\left(\frac{h_c}{b}\right) + 1 \right), \qquad (3.10)$$

where ν is Poisson's ratio, f is the lattice mismatch and b is the magnitude of the Burgers vector of the dislocation in the crystal (Figure 3.10). In crystals, the slip planes are often the planes of densest packing, which for face-centred-cubic lattices are the {111} family planes. The Burgers vectors for dislocations are then often directed to the nearest

neighbour atoms, which for face-centred-cubic lattices such as InSb are in the $\langle 110 \rangle$ direction, giving

$$b = a_s \frac{\langle 110 \rangle}{2} = a_s \frac{\sqrt{2}}{2} = \frac{a_s}{\sqrt{2}}, \qquad (3.11)$$

where a_s is the lattice constant of the substrate material. Burgers vectors are discussed further in the following section (section 3.3.2), also see Figure 3.14.
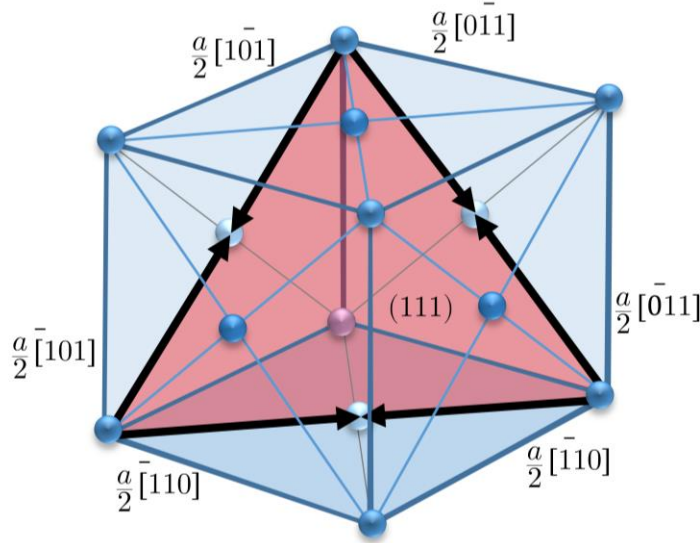
Figure 3.10 The planes of densest packing are given by the family of {111} planes (red), where the Burgers vectors (bold arrows) bound this plane, pointing to the nearest neighbour atoms given by the vectors a/2 $\langle 110 \rangle$.

The mismatch between lattice constants of the substrate and the epilayer (a_e) is given by f, such that

$$f = \frac{|a_s - a_e|}{a_s}. \qquad (3.12)$$

For $Al_{0.1}In_{0.9}Sb$ grown on GaAs ($a_s = 5.6535$ Å, $a_e = 6.445$ Å), the lattice mismatch is ~14%, however for InSb on $Al_{0.1}In_{0.9}Sb$ ($a_s = 6.445$ Å, $a_e = 6.479$ Å) this is only ~0.5%.

Poisson's ratio (ν) gives the ratio of the transverse expansion of the crystal to the amount of axial compression, as shown in Figure 3.6 ($\nu = 0.35$ in InSb [12, 96]). The angles λ and α are the angles between the slip direction and that direction in the film plane which is perpendicular to the line of intersection of the slip plane and the interface, and the angle between the dislocation line and its Burgers vector respectively. These are both 60° for InSb, giving $\cos(\theta) = \cos(\lambda) = 1/2$.

There is no analytical solution for the critical thickness, instead solutions can be found graphically, plotting the lines $y = h_c$, and $y = \frac{b}{8\pi f \cos(\lambda)}\left(\frac{1-\nu\cos^2(\theta)}{1+\nu}\right)\left(\ln\left(\frac{h_c}{b}\right)+1\right)$, with the intersection giving the critical thickness. Due to the large mismatch between $Al_{0.1}In_{0.9}Sb$ grown on GaAs, there is no intersection, implying that any growth will be immediately relaxed due to misfit dislocations. This is the reason for inclusion of the "hot" AlSb layer, to minimise these defects and glide them out of the sample perpendicular to the growth direction. For InSb on $Al_{0.1}In_{0.9}Sb$ however, the mismatch is more reasonable, with the strained InSb layer having a second strained interface with the top cap aiding in maintaining the strain. To account for this, a factor of 2 can be included in the right hand side of equation (3.10) [95], with this giving a more reasonable value of $h_c = 37.4\ nm$. This is shown in Figure 3.11.
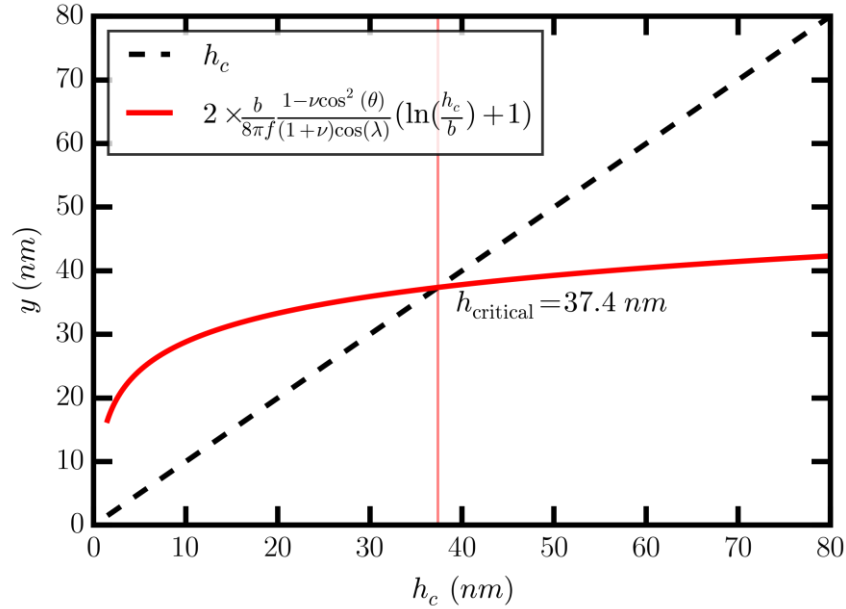
Figure 3.11 Graphical solution of the Matthews and Blakeslee model for the critical thickness of an InSb quantum well layer grown on $Al_{0.1}In_{0.9}Sb$.

This value of the critical thickness is larger than the width of the QWs studied in this work and so the wells can be considered to be strained to match the lattice constant of the $Al_{0.1}In_{0.9}Sb$ layer. This thickness is however only marginally larger than the well thickness. An alternative model for critical thickness was proposed by People and Bean following similar energetics arguments [97]. Their equation for critical thickness of the quantum well is given by

$$h_c = 2 \times \frac{1}{16\pi\sqrt{2\pi}} \left(\frac{1-\nu}{1+\nu}\right) \frac{b^2}{a_0 f^2} \ln\left(\frac{h_c}{b}\right). \qquad (3.13)$$

The result of this model is shown in Figure 3.12, where the critical thickness is significantly larger than from the conservative Matthews and Blakeslee model, so it can be safely assumed the well is not defected due to the lattice constant mismatch.
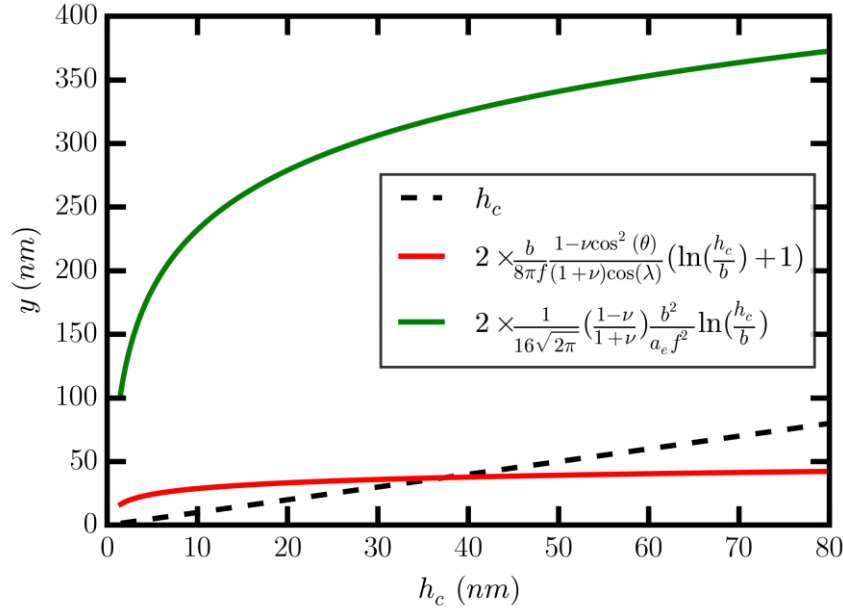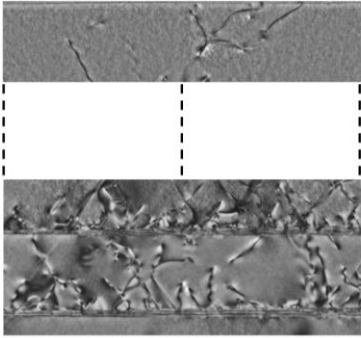
Figure 3.12 Graphical solution of the People and Bean (green) and Matthews and Blakeslee (red) models for the critical thickness of an InSb quantum well layer grown on $Al_{0.1}In_{0.9}Sb$. The People and Bean model gives a significantly larger estimate for the critical thickness (not shown) as compared to the Matthews and Blakeslee model ($h_c = 37$ nm).

3.3.2 Threading dislocations and screw dislocations

Dislocations in strained crystal growth form along the interface between the substrate and the epilayer. A misfit dislocation in this plane is known as a 90° dislocation, and is efficient at relieving strain, however in the fcc lattice, this dislocation is unable to glide and extend. However, dislocations in fcc lattices also often form at an angle of 60°, and whilst this is less favourable for strain relief, the dislocation can glide in this direction and extend. In this way under standard conditions, dislocations are of this 60° type, extending away from the substrate and "climbing" into the sample. These are known as threading dislocations and are clearly visible in a transmission electron microscopy (TEM) image of a typical sample studied in this work (see Figure 3.13).

Layer	Composition	Thickness
Surface	–	–
Upper Barrier	$Al_{0.15}In_{0.85}Sb$	50 nm
Quantum Well	InSb	30 nm
Buffer	$Al_{0.10}In_{0.90}Sb$	3 μm
Accommodation	"Hot" AlSb	0.5 μm
Substrate	GaAs	–

Figure 3.13 Layer composition with corresponding TEM images of a cross-section through a typical sample, showing clear threading dislocations at interfaces with the substrate, accommodation and lower buffer layers, and a significantly reduced density of defects at the surface $Al_{0.15}In_{0.85}Sb$ layer. TEM imaging was performed by Dr Richard Beanland at the Department of Physics, Warwick University.

To understand the formation and termination of defects, it is intuitive to consider a defect as a boundary between regions of crystalline material, with these regions displaced relative to each other. In this way, a dislocation cannot simply terminate in the lattice, and must either reach a free surface (the top or edge surfaces of the sample) or must annihilate with another dislocation (forming a loop).

When the Burgers vector of a dislocation is aligned perpendicular to the line of the dislocation, as in the 90° misfit, the dislocation is known as a purely edge dislocation (where $n \pm 1$ atomic planes are attempting to bond onto n planes, leaving an edge line of a plane un-bonded). However if the Burgers vector is aligned parallel to the line of the dislocation, this is known as a screw dislocation (as now the atomic planes are not truly parallel, but a single atomic plane is distorted in a helical fashion). If the Burgers vector is between parallel and perpendicular, this is known as a mixed type dislocation, and this is the case for standard 60° threading dislocations. Screw and edge dislocations are illustrated in Figure 3.14 [98].
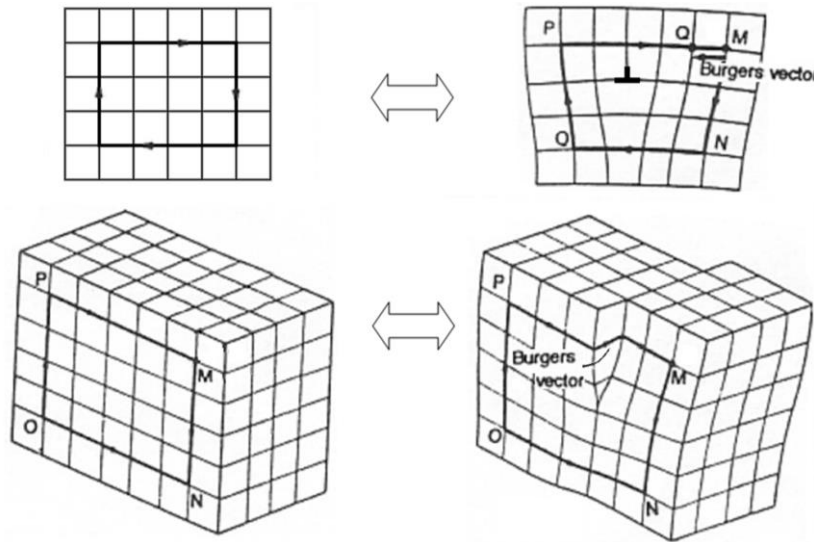
Figure 3.14 Examples of a purely edge misfit dislocation (top), and a purely screw dislocation (bottom) [98]. To determine the Burgers vector of a defect, a circuit from point M clockwise through points N, O, P and back to M, is defined on an unperturbed lattice (left). Then following these same steps on the dislocated lattice (right), the finish point (Q) will be different to the starting point (M), where the Burgers vector is the vector between these two points (i.e. the extra step in the circuit introduced by the dislocation).

Growth around screw dislocations occurs at the stepped interface of the screw, where atoms adsorbed on the surface during MBE migrate to the step edge, bonding to it and advancing the step. However as the atomic planes are now in a helical arrangement, a complete atomic layer covering the sample surface cannot be formed, instead advancing continuously helically in the growth direction around the central screw dislocation. This dislocation is pinned in the lattice, so the atomic step rotates about this point. As the radius of a section of the step increases away from the screw dislocation, this section has to advance further to maintain equal advance of the spiral growth, however the rate of growth is also inversely proportional to the radius of curvature. These factors mean that at steady state growth, (rate of advance)/(distance from dislocation) = (angular velocity) = (revolution per unit time) is constant for all points on the step and uniform spiral growth is achieved.

3.3.3 Stacking faults and microtwins

A final type of dislocation considered is a stacking fault (and the closely related microtwin defect). The stacking of a crystal is the two-dimensional ordering of the lattice planes, and in a regular crystal the pattern of these in a given direction would repeat periodically. In a face-centred-cubic lattice, the stacking in the {111} plane occurs via the sequence "ABCABCABC…" (see Figure 3.15).
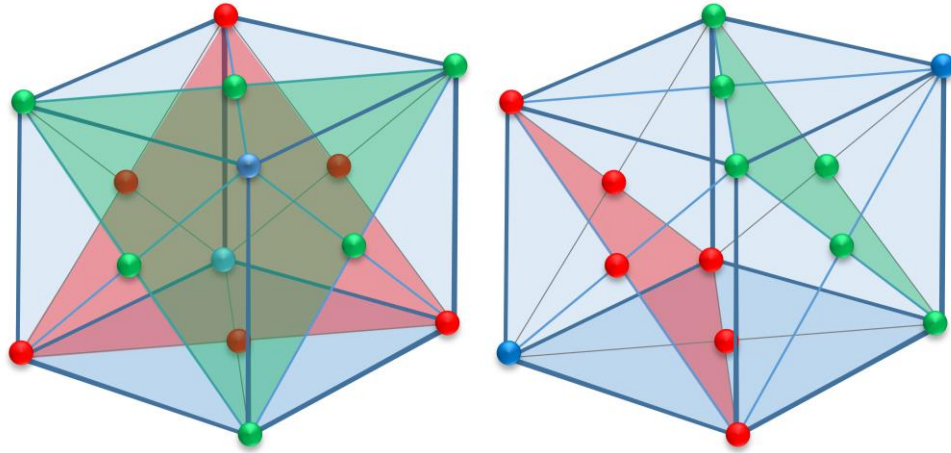
Figure 3.15 Stacking of {111} planes in a face-centred-cubic lattice, where the planes are ordered A (blue) (only a single atom in this plane is shown), B (red), C (green), A, B, C etc. in a perfect crystal. A stacking fault occurs when this order is altered, by either exclusion of a plane, or inclusion of an extra plane.

A stacking fault occurs when one of these planes slips, changing the regular ordering (for example to "ABCACABC…"), with this stacking fault known as a partial dislocation. This can occur in the form of a missing plane, or the inclusion of an additional plane, altering the stacking sequence. The 60° threading dislocations discussed previously can often split into 90° and 30° partial dislocations, generating stacking faults. Stacking faults are often not isolated, with several occurring in parallel planes. Figure 3.16 shows a schematic example of stacking faults, alongside a TEM of a crystal with stacking faults.

(a)   *(b)*   *(c)*
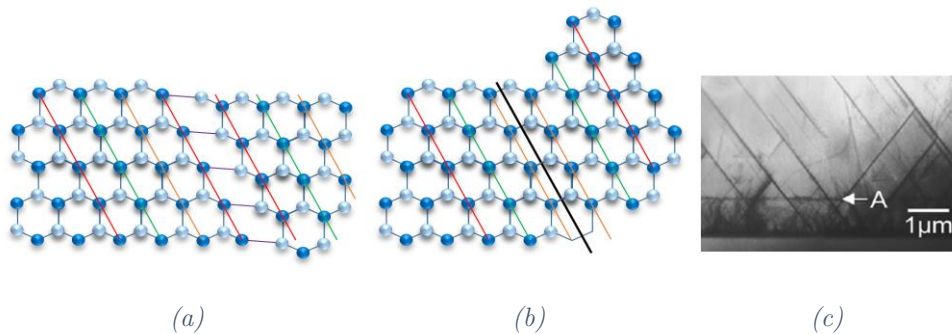
Figure 3.16 Examples of stacking faults in a crystal lattice. The stacking sequence is represented by coloured lines, ordered red, green, orange, red, etc., with (a) showing insertion of an extra plane (red), and (b) showing a microtwin defect with mirroring of the stacking sequence about the thick black line. (c) shows a TEM image from reference [24] of the generation of threading dislocations (thin lines) and microtwin defects (thick angled lines) in an AlInSb layer due to lattice mismatch with an interface (bottom).

A special subset of stacking faults occurs when the stacking order becomes reversed (or mirrored) about the stacking fault (e.g. "ABCACBA…"). These can occur in zincblende

crystals if part of the crystal is rotated at 60° around the ⟨111⟩ axis (Figure 3.16). This type of stacking fault is known as a microtwin defect, and again these are often not found in isolation, and have been shown to have significant impacts on crystal quality and the associated electrical quality of the material [24, 25, 26, 28, 53].

3.3.4 Effect of defects on mobility

Much of the work examining the correlations between material quality (assessed through measured defect densities) and electrical quality (quantified through measured mobility) was performed by the group of M. B. Santos at the University of Oklahoma during the 2000s.

This work focussed on using a range of techniques (e.g. TEM and AFM) to analyse the surface morphology of epitaxially grown InSb on GaAs substrates, often correlating this with measured mobilities.

An initial paper from the year 2000 ("*Improving the surface morphology of InSb quantum-well structures on GaAs substrates*" [51]) considered the effect of growth temperatures and V/III ratios on the abruptness of interfaces, with this qualified through analysis of the morphology of the surface (analysed via AFM). In this work, pyramidal (hillock) formations due to spiral like growth around threading dislocations were observed, with typical diameters ~1 μm to ~3 μm, heights ~20 nm, and inclination angles of ~1.5°. This growth around the spiral defects was relatively uniform, however abrupt oriented steps and square defects were also observed, where growth around these feature was not uniform, leading to large discontinuities either side of the step or square edge. In this work it was noted that hillock like features were always observed during growth of InSb, however square mounds were only observed during growth of AlInSb. It was shown that the size of these square defects could be reduced via thicker buffer layers, where for the buffer thickness corresponding to the samples studied in this work (3 μm), lateral edge lengths of ~5 μm were observed. This work did not examine the mobility of any of the samples analysed, nor the effect of the density of hillocks, steps or square mounds on the electrical quality.

A subsequent paper titled "*Mobility anisotropy in InSb/Al$_x$In$_{1-x}$Sb single quantum wells*" [23] did examine the effect of these abrupt steps on the mobility however. Specifically, the relationship between the anisotropy of the orientation of steps, and the corresponding anisotropy in the measured mobility was investigated. In this work it was theorised that the step would act as a scattering centre in two ways; the steps acting as the termination of a plane of crystal dislocation scattering centres propagating through

the QW, and the features creating a potential step via narrowing of the QW. Through subsequent measurement and analysis, it was shown that the physical quantum well morphology had a critical effect on the observed mobility, with the scattering related to the potential step dominating. This work showed that the step like morphology of the surface, and so the well, can have a large impact on the mobility observed due to variations in potential, though again did not further analyse this scattering mechanism. The crystal quality was merely shown to have an impact on the measured mobility, with this not quantified or analysed further.

Following this, work was undertaken examining the effect of microtwin (MT) defects on the mobility, where these were shown to have drastic effects [24]. It was shown that an MT can offset and incline a QW either side of the defect, with this effect still observable ~0.5 μm from the defect. Scattering from these defects is believed to be due to the difference in crystallographic orientation between the microtwin and surrounding material, causing either diffraction of electrons, or direct scattering in the QW. It was also shown that the previously examined abrupt oriented steps found to have a limiting effect on the mobility are formed as the terminal edges of microtwin defects threading through the sample [25], with these microtwins forming from the substrate/epilayer interface [26].

To try and reduce these defect densities, investigations were performed into using multiple $Al_xIn_{1-x}Sb/Al_yIn_{1-y}Sb$ interfaces [52], and varying Al compositions [27], in an attempt to trap threading dislocations, preventing them propagating through the QW and to the surface. It was shown that using multiple interlayers can act to reduce the threading dislocation density as a function of buffer thickness at a much greater rate than simply using a thicker buffer. For the samples studied, dislocation densities for samples with no interlayers varied between ~5×10^9 cm^{-2} to ~1×10^9 cm^{-2} for buffers between ~1 μm and ~3 μm respectively. Over a comparable thickness range, dislocation densities for samples with multiple interlayers varied between ~1×10^{10} cm^{-2} to ~5×10^8 cm^{-2}. Again this improved dislocation density was correlated with improved mobility, though a direct quantitative analysis was not performed. For comparison, for the samples studied in this work, the threading dislocation density is ~7.5×10^8 cm^{-2} for samples with no interlayers and a 3 μm buffer.

Expanding on this work, investigations were performed into the effect of miscut substrates on microtwin densities, where a 2° offcut was found to be effective in reducing the densities observed [28]. However, despite all these findings, it was still stated that "*the physics behind the electron mobility reduction due to MTs has not been fully understood yet*" [53].

More recently, work has been undertaken led by the University of Waterloo (March and November 2017) [99, 100], and the Chinese Academy of Sciences (July 2017) [101]. Again this work focused on possible methods to reduce the density of threading dislocations and hillock formations. These involved investigating samples grown with various interlayers (references [99, 100]), and using AFM and TEM to quantify the defect densities. As previously, it was found that increasing the number of interlayers reduced the associated defect densities, and gave rise to larger hillock formations (typically a few μm in size), though again the associated effect on electrical quality was not considered. Specific values for defect densities achieved, and a comparison to those measured for samples in this study, are given in Table 4.2.

Similarly, the effects of superlattice buffers were also considered (reference [101]), with these again reducing the defect densities observed. In this study however, the effect on mobility was analysed, where the reduction in defect density was shown to have a direct link with the associated change in mobility. This analysis was performed between 90 K and 180 K for bulk 3D samples, with mobilities ranging between ~5,000 cm^2/Vs and ~30,000 cm^2/Vs respectively. A simplified form of dislocation scattering was considered, broadly, deformation potential scattering and depletion potential scattering were considered. This analysis did not determine which was the limiting method of scattering due to these dislocations, and did not expand on the physical significance of the values determined from the fitting of mobilities performed. This work also considered only a relatively limited temperature range with relatively low mobility samples, and did not consider more complex structures such as quantum well samples.

As can be seen, there has been a wealth of research performed into the structural properties of epitaxially grown InSb, with the aim of understanding and reducing defect densities. A small proportion of this work has considered the associated effect on the electrical quality of the material grown, primarily qualitatively correlating defect densities with measured mobilities. A limited number of studies have performed a more quantitative analysis, though there has been little attempt at explaining the underlying physics relating the defects observed to the underlying electron scattering processes. Similarly, there has been no major work on understanding how the defect (and specifically hillock) density affects the mobility of a 2D quantum well system.

3.4 Device Fabrication

The basic device used to study the samples described above is that of a Hall bar (used to exploit the Hall Effect, described in section 3.7), allowing the extraction of the 2D sheet carrier density (n_{2D}) and mobility (μ) via measurements of the longitudinal and transverse

resistances as a function of B-Field. Images of example Hall bar devices used in this thesis are shown in Figure 3.19.

The fabrication process flow for a Hall bar is given below, with all fabrication steps performed in the class 1000 and class 100 cleanrooms[11] at the School of Physics and Astronomy in Cardiff. The cleanroom environment actively regulates temperature and humidity, maintaining these between acceptable levels, whilst also maintaining a low air particulate count. This ensures sample cleanliness is maintained, and allows for efficient, reliable and repeatable processing of devices, whilst also ensuring any fabrication by-products are dealt with in a safe manner, protecting the users and the environment.

3.4.1 Cleaving, cleaning and photolithography

The initial stage of the fabrication process is to cleave a small (~1 cm^2) sample from the larger InSb wafer (with this eventually processed and packaged to give the final desired measurable devices). The sample must then be cleaned before any processing can occur, this is either to remove dirt and contaminants from the surface, or to remove any protective resist coating. This is performed by soaking the sample in 80 $°C$ Microposit Remover 1165 for 20 minutes (if the sample is coated in resist), followed by 5 minutes in acetone in an ultrasonic bath and a final rinse in IPA (isopropyl alcohol) to remove all acetone. Finally the sample is blow dried using nitrogen gas, and any moisture removed by baking on a hot plate at 95 $°C$ for 3 minutes.

After cleaning the sample, the next step is to define the shape of the devices, using photolithography to pattern on the sample the shape of the mesas. A mesa is a raised area of the sample material that defines the shape of the device. This is achieved by selectively etching the surrounding material to leave just the mesa. This is achieved by first spin coating the sample with a layer of photoresist (a UV light sensitive polymer) by placing a few drops of the liquid resist in the centre of the sample, and spinning at a rate of 5000 revolutions per minute (rpm) for 5 seconds. The sample and resist are then baked at 80 $°C$ for 5 minutes to harden the resist. For this stage, Microposit S1813 resist is used, with these parameters giving a uniform sample coating approximately 1.5 μm thick. This is a positive tone resist, meaning that areas exposed to UV light are removed upon

---

[11] According to USA Federal Standard 209D, the cleanroom class designation is determined by the number of 0.5 μm or larger particles in a cubic foot of air. For a class 1000 designation, a maximum of 1000 0.5 μm particles per cubic foot are permitted, for a class 100, a maximum of 100 0.5 μm particles are permitted, etc.

developing, as opposed to a negative tone resist where only the areas exposed to light remain.

A Karl Suss Contact Mask Aligner is used alongside a patterned glass mask to expose the resist to UV light in the desired pattern. This is achieved by placing the mask containing the desired pattern into very close proximity to the surface, and exposing for a set time, altering the chemistry of the resist exposed to the light. The mask is a square of clear soda-lime or quartz glass, selectively coated in opaque metal to give a desired pattern (i.e. the mesa pattern or the metal deposition pattern). Typically a mask will not contain a single device pattern, but will be grouped into fields, with each field containing many devices of either the same or varying patterns. This ensures a fabricated sample contains many devices for experimental testing.

The first photolithography step involves removing the resist edge bead present around the edge of the sample after spinning. This is a slightly thicker region formed due to the surface tension of the resist, which if not removed would lead to a lower quality of lithography for the devices along with other potential subsequent problems. To remove the edge bead, a simple square metal pattern is used to protect the centre of the sample, leaving the outer edges to be exposed. An exposure time of 45 s, followed by development in Microposit MF319 for 30 s and rinsing in deionised water for 30 s ensures the edge bead is removed and photolithography for the actual devices can begin.

The initial device photolithography stage defines the shape of the mesa, including the central Hall bar, the contacting legs and the Ohmic contact pad locations (shown in Figure 3.19). For this pattern the exposure time is set at 8 s, followed by development for 30 s and rinsing for 30 s. A final post exposure bake is performed at 80 $°C$ for 5 minutes to harden the resist and prepare the sample for chemical wet etching.

3.4.2 Chemical wet etching

Once the sample has been developed and the photoresist that was exposed to UV light has been removed, the next step is to perform a wet chemical etch to form a mesa, outlining the shape of the device, and isolating each device from the surrounding material and other devices. This aids in reducing current leakage paths, and helps to ensure the electrically measured values are representative of the material measured.

To etch Hall bar samples fabricated at Cardiff, a wet chemical etch is used, allowing for a relatively rapid etch rate compared to dry etching techniques, whilst maintaining device integrity. The specific etchant used for the InSb/AlInSb layers studied is a combination of lactic and nitric acids, in a ratio of 50 ml : 8 ml, with typical etch rates

on the order of ~1 $\mu m/minute$. As mixing the acids is exothermic, the mixture is made at least 30 minutes prior to use to allow the acids to cool, mix thoroughly and equalise in temperature. This mixture ensures an even etch in the exposed areas, whilst the unexposed and undeveloped photoresist protects the areas of the sample defining the shapes of the devices.

As etching occurs both vertically and horizontally, it is important to regularly check the etch depth using a Bruker Dektak Stylus Profiler and to examine devices under an optical microscope to ensure the acid does not fully etch through the Hall bar legs, as this would render the device unusable. It is desirable to etch sufficiently far below the quantum well (preferably down to the GaAs substrate) to ensure the mesa is electrically isolated from any nearby devices, and so to enable this, the leg widths have been designed to include an extra 6 μm width, to account for etching through the 3 μm buffer. Once the required etch depth has been reached the sample is washed in deionised water to remove any acid remnants, and the remaining resist is removed by washing in acetone, rinsing in IPA and drying with nitrogen gas.

3.4.3 Ohmic contact metal deposition

The next processing step is to form Ohmic contacts to allow for good electrical contact to the devices and the 2DEG. These are ideally low resistance contacts that allow a current to be passed to the devices, and allow for voltage drops across the devices to be measured. To achieve this, a similar photolithography process to that previous is used to define the areas in which the metal will be deposited.

To achieve metallisation, a uniform coating of metal is deposited across the whole surface, and will therefore need to be removed in select places, whilst attaching to the sample surface in others. To enable this, a bi-layer resist lithography step is used, where this bi-layer gives rise to an undercut, ensuring the metal can stick sufficiently in the exposed areas forming the Ohmic contacts, and aiding in removing unwanted metal from the sample without also removing the Ohmic contact metal.

Before spinning the resist, the sample is again moisture baked at 95 °C for 3 minutes. The initial layer of resist is PMGI SF11 (again a positive resist), with this spun on at a speed of 5000 rpm for 20 s, followed by baking at 95 °C for 10 minutes. After waiting for the sample to cool, the second layer of resist is applied, this time S1813 is used again, with this spun on at 5000 rpm for 20 s. A final bake is performed at 80 °C for 5 minutes. Following this, the photolithography steps are similar to previous, with the edge bead removed first, followed by the photolithography for the Ohmic contacts. At this stage it

is important to align the contact pattern on the mask to the sample below, ensuring there is a slight overlap of the contacts over the edge of the mesa. This alignment gives as many working devices as possible, with the contact overlap shorting any eddy currents that may form around the contacts, giving more reliable results.

The different resists used have different properties with regards to exposure and development times. The surface layer of S1813 is exposed and developed as in the single resist layer process, however once the developer solution has removed the surface S1813 in the desired pattern, this exposes the SF11 underneath. This resist is more readily removed than the S1813, creating an undercut. This undercut ensures the metal can stick to the sample in the desired locations to form the contacts but also introduces a physical disjoint with the metal coating the rest of the surface resist. This means the undesired excess metal can be removed easily later without removing the contact metal also.

To coat the sample surface in metal, an Edwards thermal evaporator is used. This method works by mounting the sample face down in a high vacuum chamber above crucibles holding the desired metals for deposition. A large current is then passed through the crucibles, melting and then evaporating the metal within. This evaporated metal forms a beam, coating the inside of the chamber and the sample. The thickness of the metal films is measured using a crystal thickness monitor placed in close proximity to the sample, and when the desired thicknesses are achieved, a mechanical shutter is closed to stop further evaporation onto the sample. The structure of the contacts used is a simple combination of a 10 nm zinc keying layer (used to help the contacts stick to the surface), followed by a thicker 300 nm gold layer. To externally contact the devices once processing is finished, fine gold wires are bonded to this gold layer. To aid in the formation of good contacts, the sample is heated to 100 $°C$ for this process.

It is common in semiconductor device fabrication to anneal contacts at high temperatures to improve their properties, however InSb has a very low melting point, and previous attempts at annealing have shown no benefit, instead damaging device quality.

The final step is to remove the excess metal, with this achieved by removing the underlying resist and "floating off" the metal above. To do this, the sample is soaked in acetone until the metal loosens, and then rinsed in acetone and IPA and dried with nitrogen gas.

(a) Photoresist coating for
    mesa lithography

(b) Mesa photolithography for
    Hall bar pattern

(c) Wet chemical etching for
    Hall bar pattern

(d) Bi-layer resist for Ohmic
    contact lithography

(e) Photolithography for
    Ohmic contacts

(f) Metal evaporation for
    Ohmic contacts

(g) Finished Hall bar with mesa and Ohmic contacts
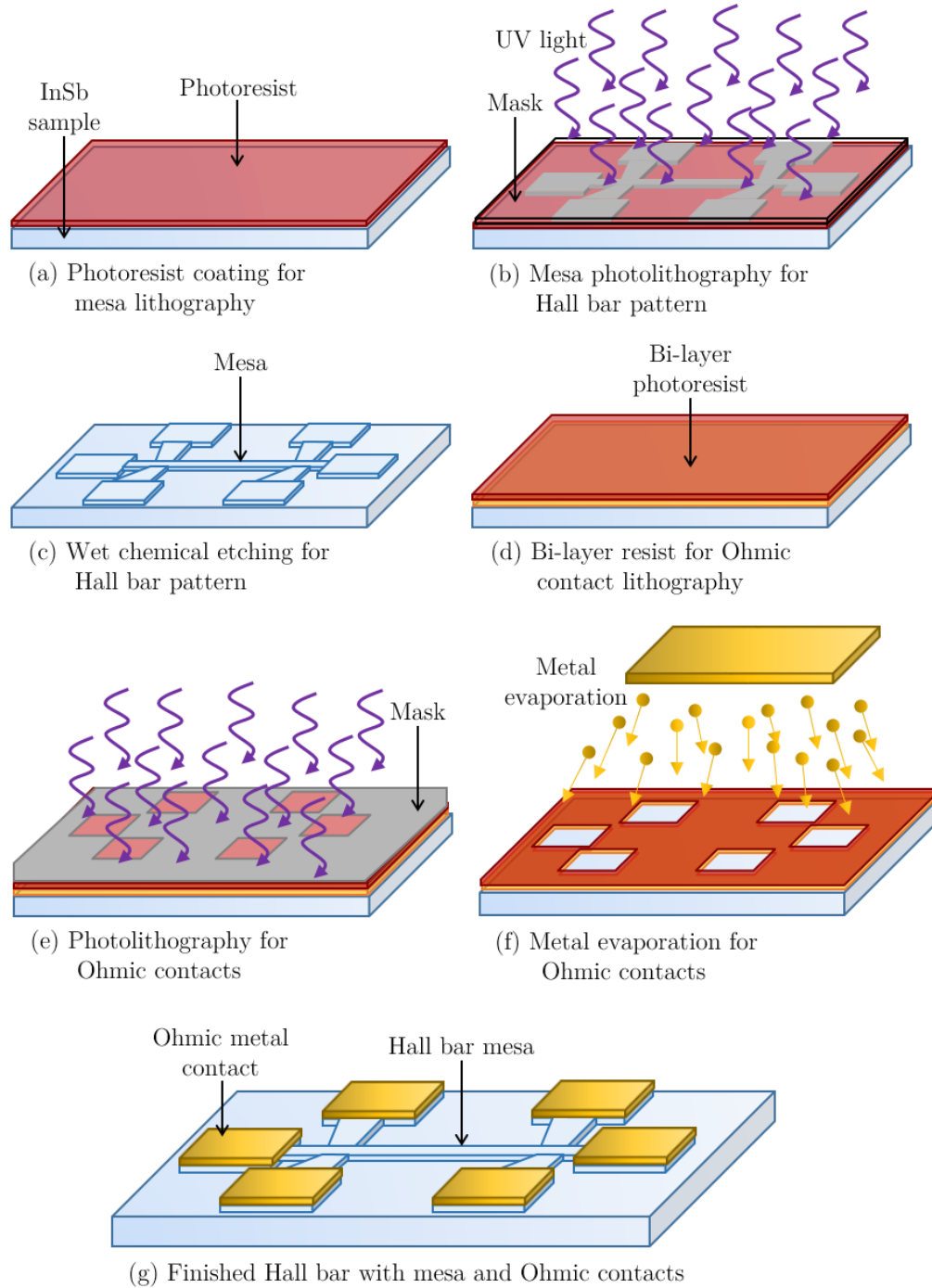
Figure 3.17 Summary of the process flow fabrication steps for a standard Hall bar fabricated in the cleanroom at the School of Physics and Astronomy in Cardiff. Steps shown included photolithography, wet chemical etching and Ohmic contact metal evaporation. See the text for a more detailed explanation of the steps and processes involved.

3.4.4 Packaging and bonding

Individual sections of the sample can then be chipped up, and placed into a 20-way ceramic chip package (Figure 3.18). The package has 20 individual gold contacts, allowing for multiple devices to be connected and tested at one time, whilst also allowing for quick and easy exchange of samples. The sample is mounted to the package using low temperature GE varnish, ensuring the sample is securely held, whilst also ensuring a good thermal contact. To contact the individual Hall bar devices, a West Bond 7400A wedge-wedge wire bonder is used with 12.5 μm fine gold wire. This method involves heating the sample to ~100 °C, and using a heated needle tip at ~250 °C to apply the gold wire to the surface contact pad, with the combination of heat and force used to make the wire stick. The wire is then bonded to a contact pin on the package, with this process repeated for all the contacts desired.

(a)

(b)
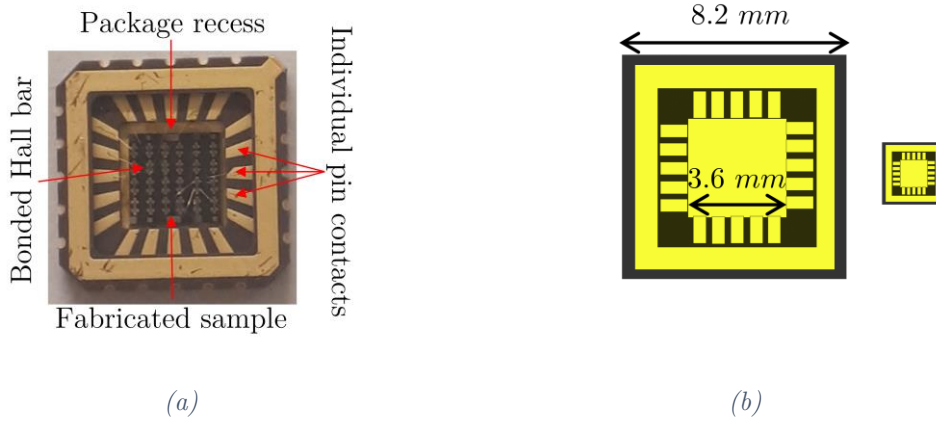
Figure 3.18 A photograph of a standard 20-way ceramic chip package containing a fabricated sample with bonded Hall bars (a) and a schematic diagram showing the inner and outer dimensions, as well as a "real-size" representation (b).

A typical Hall bar device is shown in Figure 3.19, where the width of the Hall bars studied in this work are ~40 μm, and the length between the central contacting legs is ~200 μm, giving an aspect ratio of ~5.
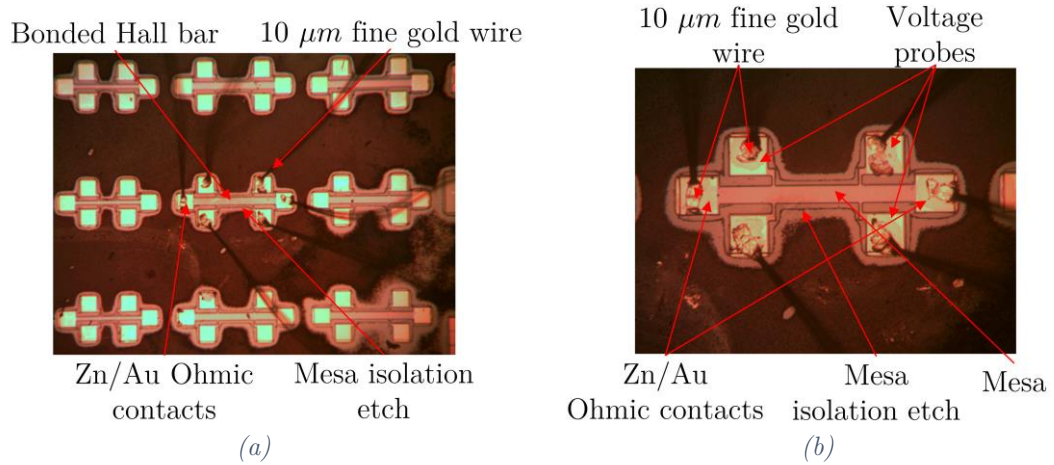
Figure 3.19 Optical microscope image of a fabricated sample (a) and an individual Hall bar (b). Each Hall bar consists of a mesa with gold Ohmic contacts for current injection at either end and 4 central contacts for measuring longitudinal and transverse voltages. These contacts act as the non-current carrying probes for 4-point resistance measurements. Also shown are the fine gold wires used for bonding to the package for electrical testing.

3.5 Control Software and Electrical Measurement

All experimental measurements performed in this work used computer controlled (python) software to interface to the physical hardware. The software was written in house and modified over several years, including as part of the work performed during this PhD. The software consists of a graphical user interface (GUI) where experimental measurement types can be selected (i.e. constant current, constant voltage, etc.), and experimental parameters can be input (i.e. set temperatures to perform measurements at, magnetic field values, current values, etc.). The software extracts these values, then using the pyVISA module with a National Instruments Visa backend, sends commands to the physical hardware (temperature controller for the cryostat, power supply for the electromagnet, source measurement units, etc.) either via a GPIB[12] connection or a USB connection, where the specific commands sent are SCPI commands[13]. Once the commands have been performed, and measurements acquired, the data can be imported directly into python for further analysis and plotting.

The main hardware used for experimental measurements are a Keithley 6221 Current Source Meter, a Keithley 2182A Nanovoltmeter, an Oxford Instruments Mercury iTC

---

[12] General Purpose Interface Bus

[13] Standard Commands for Programmable Instruments

intelligent Temperature Controller and an Oxford Instruments Optistat AC-V12 pulse tube cryostat system, consisting of a Cryomech, Inc. PT403 cold head, and water cooled Cryomech, Inc. CP830 helium gas compressor. The sample space within the cryostat is kept under vacuum maintained by an Œrlikon Leybold TURBOLAB 80 basic turbomolecular pump, backed by a dry compressing backing pump. The sample space is situated between the poles of a simple electromagnet with an Agilent N5769A power supply allowing for variable magnetic field measurements.

To perform the electrical measurements required for the simple Hall characterisation of the samples, the Current Source Meter is used in tandem with the Nanovoltmeter to supply a known current and measure the resultant longitudinal and transverse voltages (V_{xx} and V_{xy}). A series of $I(V)$ repeat measurements at a set current are performed and averaged to determine the voltage, where this is then repeated as a function of B-field to obtain a B-V curve for both V_{xx} and V_{xy}. These curves can be analysed as described in section 3.7 to give the sample carrier density and mobility, with this process then repeated across temperature.

Passing a current through a sample for any length of time can result in heating of the sample and thermal drift of the measured voltage. This linear voltage increases independently of the current direction and will result in inaccurate measurements. To combat this, a "δ-mode" measurement has been implemented. This δ-mode is a pseudo-AC current technique where the current is cycled between two constant values, a forward and reverse value, in discrete steps. For simplicity this is taken to be a forward and equal reverse current, though the technique can also be applied with a constant offset. With time, the measured voltage will drift linearly by an amount δV, which for a constant cycle rate will lead to a constant drift for each cycle. This offset can be removed via a simple average over 3 cycles (i.e. forward current, reverse, and forward again). This is illustrated in Figure 3.20. The δ-mode implemented is a modified version of inbuilt δ-mode present on the Keithley 6221 and Keithley 2182A, instead using the python software to instruct the Keithley 2182A to measure on both A and B channels (for V_{xx} and V_{xy}), as opposed to the inbuilt version which is limited to channel A.

(a)   *(b)*
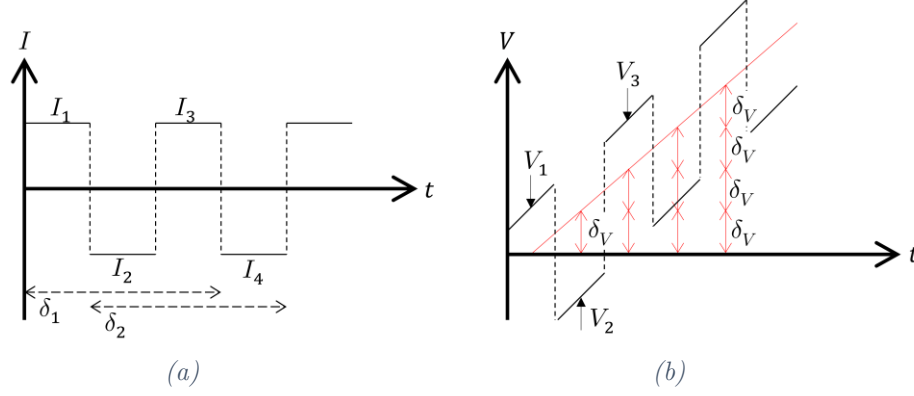
Figure 3.20 Example of the δ-mode measurement scheme for (a) set current (I) and (b) measured voltage (V) as a function of time for 4 cycles (labeled 1 to 4) where the voltage is measured at the points indicated by arrows. For simplicity $I_2 = -I_1$, though this technique will also work with a constant offset to both the forward and reverse current values. The measured voltage shows a linearly increasing thermal voltage, increasing by a value δV per cycle. A set of 3 set currents and measured voltages forms a single delta mode measurement ($δ_1$, etc.).

For the 3 cycles required to exploit the δ-mode technique (following Figure 3.20), in cycle 1, a current I_1 is applied and voltage V_1 measured; in cycle 2, a current $I_2 = -I_1$ is applied and a voltage $V_2 = -V_1 + δV$ is measured; and finally in cycle 3, current $I_3 = I_1$ is again applied and voltage $V_3 = V_1 + 2δV$ measured. Now if the difference is taken between cycle 1 and 2, and again between cycle 3 and 2, and these two resultant values averaged, the voltage offset $δV$ can be removed, and the true voltage, V, corresponding to a current I_1, obtained. This process is summarised in equations (3.14) through (3.17).

$$V = \frac{1}{2}\left[\frac{V_1 - V_2}{2} + \frac{V_3 - V_2}{2}\right] \tag{3.14}$$

$$V = \frac{1}{2}\left[\frac{V_1 + V_1 - δV}{2} + \frac{V_1 + 2δV + V_1 - δV}{2}\right] \tag{3.15}$$

$$V = \frac{1}{2}\left[V_1 - \frac{δV}{2} + V_1 + \frac{δV}{2}\right] \tag{3.16}$$

$$V = V_1 \tag{3.17}$$

This method will also work if the current values are not exactly equal and opposite signs ($I_1 \neq -I_2$, i.e. there is an offset). The voltage measured will correspond to a current value of half the range between I_1 and I_2, such that

$$I = I_1 - \frac{I_1 + I_2}{2}. \qquad (3.18)$$

If I_1 and I_2 are both positive, the equivalent I will be smaller than either I_1 or I_2 individually.

This process can then be repeated for a number of averages by simply extending the number of cycles as desired. Despite the use of the δ-mode technique, experimental currents should still be kept small to minimise any excess heating, where experimental currents used in this work were typically $\leq 1\ \mu A$.
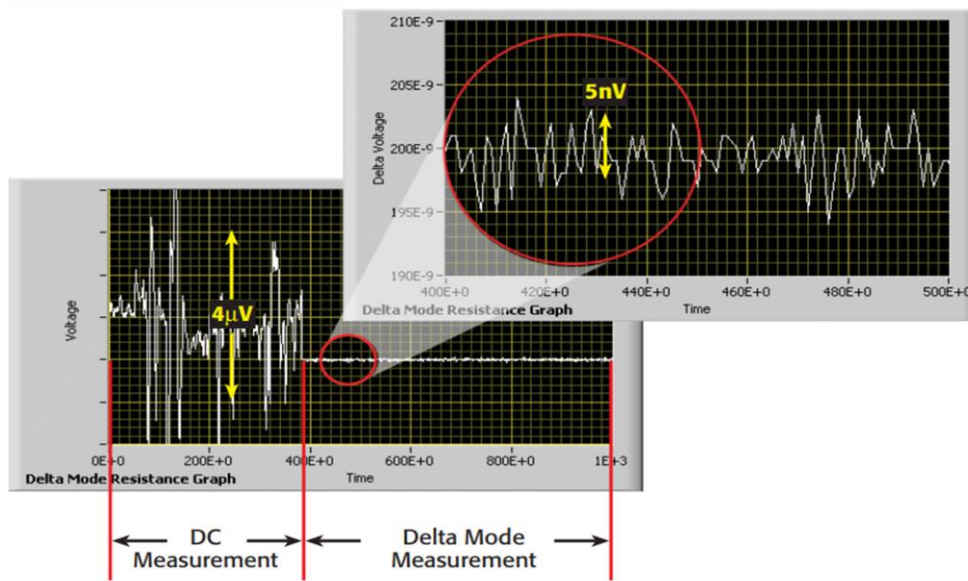
Figure 3.21 Example comparison of standard DC voltage measurements and δ-mode voltage measurements taken from the Keithley 2182A Nanovoltmeter datasheet, showing the clear advantage of reduced noise in δ-mode measurements [102].

To ensure the voltage measurements are representative of the sample device and not of the leads or the system, 4-point measurements are performed. In this scheme, the current is passed into and out of the device via a pair of outer contacts, whilst the voltage measurement is performed between a central pair of contacts (Figure 3.22). Due to the high impedance of the voltmeter, no current flows through these central contacts, and the only voltage measured is due to the device. The resistance of the current carrying leads and contacts then does not affect the measured voltage. If the resistance of a lead is given by R_l, the resistance of a contact given by R_c and the resistance of the sample given by R_s, then the 2-point (4-point) resistance $R_{2(4)}$ is given by

$$R_2 = 2R_l + 2R_c + R_s, \tag{3.19}$$

$$R_4 = R_s. \tag{3.20}$$

For all the Hall characterisation measurements performed in this work, 4-point measurements have been used.
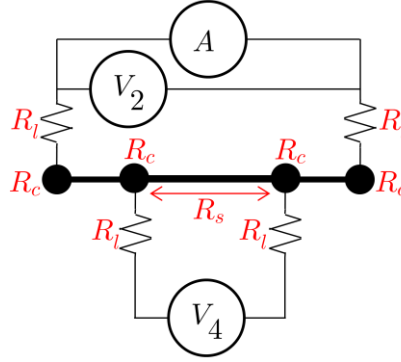
Figure 3.22 Schematic diagram depicting the 2-point and 4-point measurement schemes. A current is passed between the end contacts, where each wire has a resistance R_l and each contact has a resistance R_c. In the 2-point scheme, the voltage (V_2) is measured at these same contacts, measuring a resistance $R = 2R_l + 2R_c + R_s$. In the 4-point scheme, a different, central pair of contacts are used to probe the voltage (V_4), where due to the high impedance of the voltmeter, no current flows through these contacts or wires. The resistance measured is then $R = R_s$.

3.6 Cryostat and Electromagnet

As mentioned previously, cold temperature measurements were performed using an Oxford Instruments Optistat AC-V12 pulse tube cryostat system, with a Cryomech, Inc. PT403 cold head and water cooled Cryomech, Inc. CP830 helium gas compressor. A schematic diagram of the main components of a pulse tube type cryostat are shown in Figure 3.23.

Cooling in a pulse tube style cryostat is achieved via an oscillating helium pressure wave expanding and contracting, and so exchanging heat. The helium compressor first creates an increased pressure and causes gas to flow through the system, this higher pressure helium gas then is allowed to expand in 2 individual pulse tubes. As the gas expands, its temperature decreases, decreasing the temperature of the copper cold plate stages thermally anchored to the tubes. The 1st cold stage is cooled to a temperature between ~40 K to ~75 K, whilst the second stage is cooled further, to ~2.8 K. A

regenerator heat exchanger is used to aid cooling by cooling gas entering the pulse tubes, whilst warming the gas exiting the tubes.

Attached to the 2nd cold stage is a copper cold finger, the end of which houses a sample holder where the 20 way packages containing samples can be mounted and electrical connection made without the need for any further wire bonding. At the base of the cold finger, mounted to the 2nd stage is a resistive heater and rhodium-iron thermometer, calibrated to control temperatures at the end of the cold finger. When changing temperatures during a measurement across temperature, a wait time is included to ensure time for the cold finger and cold stage to equalise in temperature. This wait time is only triggered after ≥ 20 temperature readings (taken over $\geq 100\ s$) matching the set temperature within a set tolerance have been achieved, further ensuring correct temperature readings.

As the sample space of the cryostat does not use any cryogenic liquids or gasses, the space must be maintained under a high vacuum to minimise the condensation and freezing of the gasses present in ambient air (H_2O, N_2, etc.), ensuring there is no build-up of ice that may affect the temperature or quality of the sample. This vacuum is maintained by a simple turbomolecular pump where pressures of $< 1 \times 10^{-5}\ mbar$ within the sample space are achieved. The operational measurement range of the cryostat is then between ~2.8 K and ~300 K.

To achieve magnetic field measurements, the cold finger is placed between two coils of a large electromagnet, where the sample is situated such that is located between the poles of the electromagnet, and the magnetic field lines are focussed vertically through the sample (as in Figure 3.24). The electromagnet is capable of achieving B-fields of ~0.6 T with non-focussing poles, where focussing poles increase this value to ~0.85 T. The current and voltage supplied to the coils are controlled via a N5769A power supply, where this is again controlled via the python software. The value of magnetic field is calibrated at the position of the sample, and when ramping in field, the magnet is ramped at a constant rate, ensuring electrical measurements are not performed until the correct field is established.
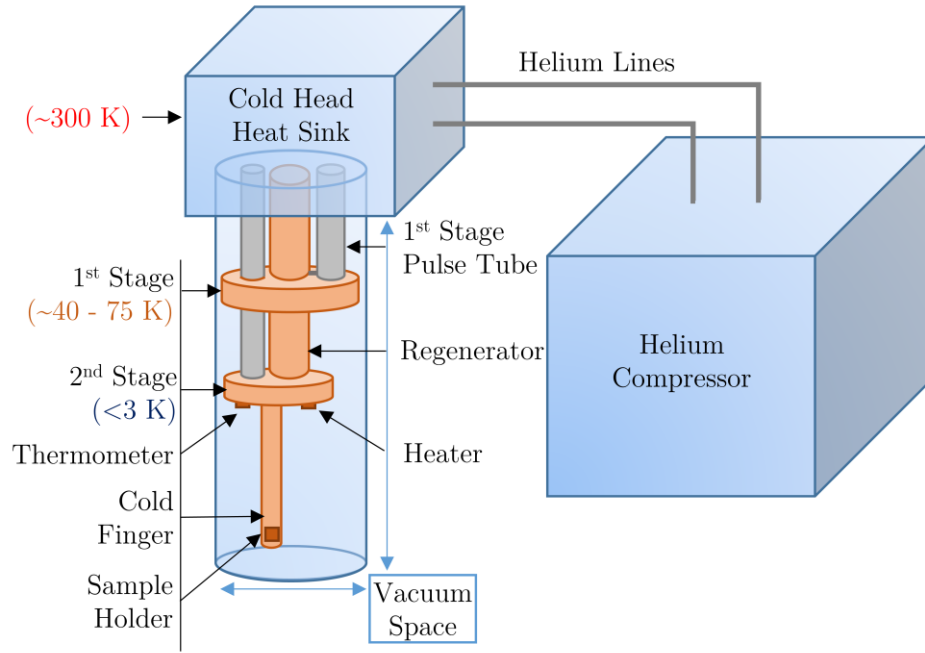
Figure 3.23 Schematic diagram of the main components of a pulse tube type cryostat. The operation of such a cryostat is described in the text. In short, a helium compressor creates an oscillating pressure wave, where this is allowed to expand and cool within the pulse tubes, cooling the cold stages. The attached cold finger houses the sample holder and is maintained under vacuum.

3.7 Hall Effect

The experimental hardware and techniques described above are all designed to exploit the Hall effect to characterise samples. The discovery of the Hall Effect was first published by Edwin H. Hall in 1879 [103], and since then, Hall's suggestion that "*a more complete and accurate study of the phenomenon will probably occupy me for some months to come*" has certainly come true. The Hall effect has found extensive practical use in electrical device characterisation as it allows the determination of the majority charge carrier (holes or electrons) and of the carrier density. Combining the transverse measurement of resistance required for the Hall effect with a longitudinal resistance measurement allows for the experimental determination of the resistivity and the mobility. In this way, at a given temperature, a sample can be characterised via two simple resistance measurements to determine its carrier density and mobility. These are used as the primary sample characterisation technique in this work.

3.7.1 Single Carrier

The Hall Effect describes the response of a flow of current when subjected to an external magnetic field (in this case the current is internal to the semiconductor). This is summarised in Figure 3.24. In the Drude model, which nicely describes the effects seen, a carrier (hole or electron) is treated as a quasi-free particle, only interacting with the crystal lattice. When an electric field (E) is applied, by the application of a voltage across the semiconductor, the carriers begin to drift due to the force given by the equation

$$\boldsymbol{F} = q\boldsymbol{E} \tag{3.21}$$

where q is the charge of the carrier ($q = -e$ for electrons and $q = +e$ for holes, where $e = 1.6 \times 10^{-19}\ C$). This force causes the carriers to accelerate, such that

$$\boldsymbol{F} = \frac{m^*\boldsymbol{v}}{\tau} \tag{3.22}$$

where $\boldsymbol{v}$ is the mean velocity vector of the carriers, proportional to electric field and temperature of the crystal lattice, and τ is the transport lifetime, which is the average lifetime of carriers between scattering events which completely randomise the carrier's velocity. This in effect only considers large angle scattering, in contrast to the quantum lifetime τ_q which is the average time between any scattering event (both large and small angle), or alternatively, a measure of the mean lifetime of electrons in cyclotronic motion between momentum scattering events [63]. The differences between τ and τ_q will be discussed further in the following chapter.

The mobility is a measure of the response of the carriers to this force, given as the velocity per unit electric field,

$$\mu = \frac{e\tau}{m^*} \tag{3.23}$$

where μ is measured in units of m^2/Vs or cm^2/Vs ($= (m/s)(V/m)^{-1}$) which is equivalent to $1/T$ (where T is Tesla) and is conventionally positive irrelevant of carrier charge, giving

$$\boldsymbol{F} = \frac{e\boldsymbol{v}}{\mu}. \tag{3.24}$$

The current density is given by the number of carriers (charge e) passing through the cross sectional area of a sample per second, (in 2D, this is the number of carriers crossing a line spanning the width of the sample), giving

$$\boldsymbol{j} = n_{2D}q\boldsymbol{v}. \qquad (3.25)$$

A conventional positive current is given by positively charged holes with a velocity in the same direction as the current, or negatively charged electrons with a velocity negative to the current.

The current and the force are linked via the velocity, such that

$$\boldsymbol{v} = \frac{\boldsymbol{j}}{n_{2D}q} = \frac{\boldsymbol{F}\mu}{e} = \frac{q\boldsymbol{E}\mu}{e}. \qquad (3.26)$$

This results in a current parallel to the electric field, where the current density is given by

$$\boldsymbol{j} = n_{2D}q\boldsymbol{v} = \frac{n_{2D}q^2\mu\boldsymbol{E}}{e} = n_{2D}e\mu\boldsymbol{E} = \sigma\boldsymbol{E} \qquad (3.27)$$

where the conductivity is given by $\sigma = n_{2D}e\mu$.

When a magnetic field ($\boldsymbol{B}$) is also applied, the electrons feel a Lorentz force in addition to the force from the electric field, giving a total force of

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \qquad (3.28)$$

If the magnetic field is applied perpendicular (e.g. in the z-direction) to the electric field (e.g. the x-direction), then carriers will feel a force in the third orthogonal (y-) direction due to the cross product term (see Figure 3.24), given by

$$F_y = -qv_xB_z, \qquad (3.29)$$

causing the carriers to follow a curved path. If the length between the contacts applying the electric field to the sample (x-direction) is sufficiently large compared to the width of the material (y-direction), then charge will begin to collect on the side faces of the material, with opposing sign charges collecting on opposite faces. This build-up of charge creates a potential difference between the two side faces, the Hall electric field (E_H), with this field opposing the Lorentz force, giving

$$F_y = qE_H. \tag{3.30}$$

Eventually these two forces will reach equilibrium, and the net force experienced by a carrier will reduce to the zero magnetic field limit. Equating equations (3.29) and (3.30) to give $F_y = 0$, and substituting equation (3.26) (and setting $B = B_z$) then gives

$$E_H = v_x B_z = \frac{j_x B_z}{n_{2D} q}. \tag{3.31}$$

Multiplying both sides of (3.31) by the width of the sample W converts the Hall field E_H to the Hall voltage V_H, and converts the current density j_x to the current through the sample I, giving in 2D

$$V_H = \frac{IB}{n_{2D} q}, \tag{3.32}$$

where following Figure 3.24, $V_H = V_{xy} = (V_A - V_B)$.

A simple rearrangement of equation (3.32) (taking the magnitude of the Hall voltage) gives the sheet carrier density as

$$n_{2D} = \frac{IB}{V_H e} = \frac{1}{R_{xy}} \frac{B}{e} \tag{3.33}$$

where R_{xy} is the transvers Hall resistance. The units of n_{2D} are then m^{-2} or cm^{-2}, and the sign of the Hall voltage relative to the sign of the current gives the sign of the charge of the majority carriers. If the conventional current is in the positive x-direction, and V_H is positive (negative), the majority carriers are holes (electrons), whereas if the current is in the negative x-direction and V_H is positive (negative), the majority carriers are electrons (holes).

If the length of the sample is too short compared to the width, then a shorting of the transverse Hall voltage will occur due to charge entering the end contacts as opposed to collecting on the side faces [41, 104]. This will result in a reduced Hall voltage, and a correspondingly increased longitudinal resistance due to the curved trajectories of the carriers giving increased path lengths. To achieve a representative Hall voltage, a length : width ratio of $\gtrsim 4$ is required. For the Hall bar samples fabricated in this work, this ratio is 5.
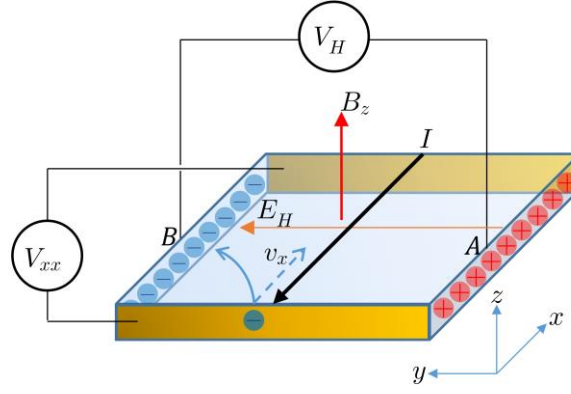
Figure 3.24 Schematic diagram showing the Hall effect in a sample for a conventional current, I (black), flowing in the negative x-direction (electrons moving in the positive x-direction), with a magnetic field pointing in the positive z-direction, B_z (red). Initially, an electron entering through the end contact will experience a force due to the E-field, accelerating the electron in the positive x-direction, and a Lorentz force in the positive y-direction (blue curved arrow). Eventually a negative charge will form on the B side of the sample, with a (relatively) positive charge forming on the A side. This charge build up causes a transverse (Hall) field, E_H, that opposes the Lorentz force. At equilibrium these forces balance and the electrons follow a straight path (blue dashed arrow). A measurement of the Hall voltage $V_H = V_{xy} = V_A - V_B$ will give a positive value, corresponding to electron majority carriers, and their density. A measurement of V_{xx} will then give the mobility.
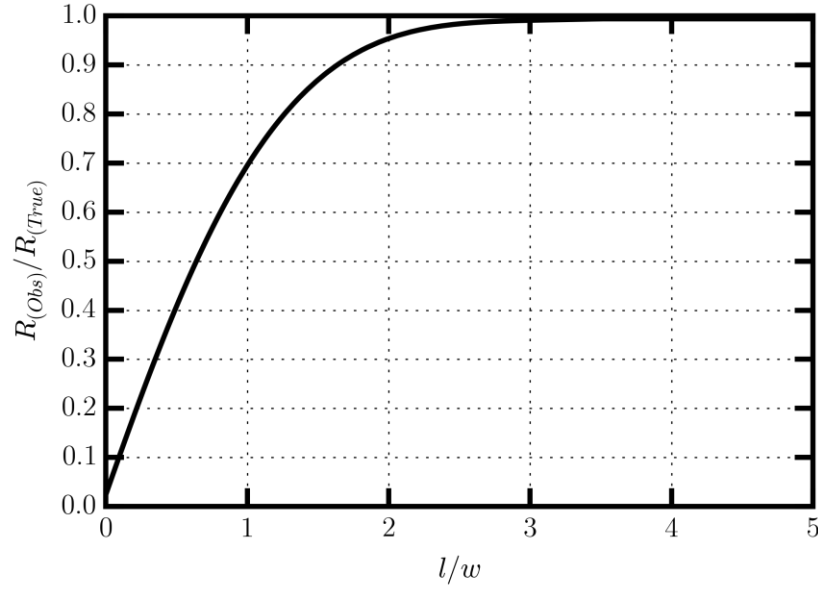
Figure 3.25 Shorting of the transverse Hall resistance comparing the "True" resistance that would be obtained for an ideal infinitely long Hall bar following equation (3.32) (R_{True}) to the resistance observed on a non-ideal Hall bar (R_{Obs}) as a function of the aspect length : width ratio (l/w). For ratios of ~4 or greater, the value of $R_{Obs} = R_{True}$. The aspect ratio used to fabricate Hall bars in this work is ~5. Figure reproduced from Figure 2.12 in reference [41], following reference [105].

To determine the longitudinal resistivity of the material and so the mobility, the measurement of the transverse Hall resistance (R_{xy}) must be combined with a measurement of the longitudinal resistance (R_{xx}), such that

$$\frac{1}{\rho_{xx}} = \frac{5}{R_{xx}} = ne\mu. \tag{3.34}$$

This result is derived from the simple Drude model, and where in 2D, the sheet resistance and resistivity are related by a factor of the length to width ratio. In 3D ρ is given by the resistance, multiplied by the ratio of the cross sectional area to the length, $\Omega A/L$, measured in units of $\Omega m = \Omega m^2/m$, whereas in 2D the sheet resistivity is measured in units of $\Omega/\square = (\Omega m)/m = \Omega$. Dividing the longitudinal resistance by the aspect ratio gives the number of "squares" comprising the 2D sample, converting from the resistance to resistivity (the width of the sample is one square, giving $R_{xy} = \rho_{xy}$). The mobility is then given by

$$\mu = \frac{\rho_{xx}}{n_{2D}e} = \frac{1}{5}\frac{R_{xx}}{n_{2D}e}. \tag{3.35}$$

If measured as a function of B-field, equations (3.33) and (3.35) would give a linearly increasing R_{xy} where the gradient is inversely proportional to the carrier density, and a field independent R_{xx} which is directly proportional to the mobility.

The 4-point measurements of R_{xx} and R_{xy} are performed between the central contacts of the Hall bar device (Figure 3.19) as shown schematically in Figure 3.26. The current I is passed into the Hall bar through the end contact (contact 1), whilst the far end (contact 2) is grounded, acting as the exit point for current from the device. Channel A of the voltmeter is connected between contacts 3 and 4 or between contacts 5 and 6, measuring V_{xx}, whilst channel B is connected between contacts 3 and 5 or 4 and 6 to measure V_H. In this way the voltage measuring contacts are separated from any current carrying contacts or leads, giving a 4-point measurement of only the device resistance.
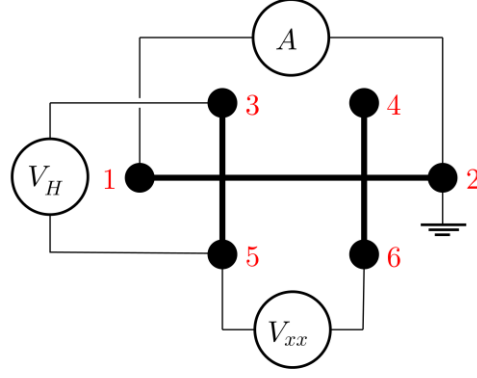
Figure 3.26 Schematic diagram for the 4-point measurement of the Hall voltage ($V_H = V_{xy}$) to determine the carrier density and of the longitudinal voltage (V_{xx}) to determine the resistivity and mobility. A known current is passed into the device through contact 1, and exits through contact 2. The voltages are then measured between contacts 5 and 6 for the longitudinal voltage, and between 3 and 5 for the Hall voltage.

3.7.2 Multi Carrier

Equations (3.33) and (3.35) are valid expressions when the contributions to resistance arise due to a single carrier type (i.e. carriers of a single charge type, all having the same effective mass, the same energy, same mobility etc.). If this is not the case, either due to carriers of a different type (holes or electrons), or carriers with a different mobility (i.e. carriers in the well vs the δ-doped plane), then more complex relations must be derived. These relations are realised in the form of conductivity (resistivity) tensors, such that (for 2 carriers, species 1 and 2) [106]

$$\rho_{xx} = \frac{\dfrac{\tau_1}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{\tau_2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}}{\left[\dfrac{\tau_1}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{\tau_2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}\right]^2 + \left[\dfrac{\omega_{c_1}\tau_1^2}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{\omega_{c_2}\tau_2^2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}\right]^2} \tag{3.36}$$

and

$$\rho_{xy} = \frac{\dfrac{-\omega_{c_1}\tau_1^2}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{-\omega_{c_2}\tau_2^2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}}{\left[\dfrac{\tau_1}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{\tau_2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}\right]^2 + \left[\dfrac{\omega_{c_1}\tau_1^2}{1+\omega_{c_1}^2\tau_1^2}\dfrac{n_1 e^2}{m_1^*} + \dfrac{\omega_{c_2}\tau_2^2}{1+\omega_{c_2}^2\tau_2^2}\dfrac{n_2 e^2}{m_2^*}\right]^2} \tag{3.37}$$

where $\omega_c = eB/m^*$ is the cyclotron frequency of the carrier (the frequency of the circular orbit in a magnetic field). These expressions can be simplified with the relation $\omega_c \tau = \mu B$, giving

$$\rho_{xx} = \frac{\dfrac{n_1 e\mu_1}{1+\mu_1^2 B^2} + \dfrac{n_2 e\mu_2}{1+\mu_2^2 B^2}}{\left[\dfrac{n_1 e\mu_1}{1+\mu_1^2 B^2} + \dfrac{n_2 e\mu_2}{1+\mu_2^2 B^2}\right]^2 + \left[\dfrac{n_1 e\mu_1^2 B}{1+\mu_1^2 B^2} + \dfrac{n_2 e\mu_2^2 B}{1+\mu_2^2 B^2}\right]^2} \tag{3.38}$$

and

$$\rho_{xy} = \frac{\dfrac{-n_1 e\mu_1^2 B}{1+\mu_1^2 B^2} + \dfrac{-n_2 e\mu_2^2 B}{1+\mu_2^2 B^2}}{\left[\dfrac{n_1 e\mu_1}{1+\mu_1^2 B^2} + \dfrac{n_2 e\mu_2}{1+\mu_2^2 B^2}\right]^2 + \left[\dfrac{n_1 e\mu_1^2 B}{1+\mu_1^2 B^2} + \dfrac{n_2 e\mu_2^2 B}{1+\mu_2^2 B^2}\right]^2}. \tag{3.39}$$

The conductivity (σ) and resistivity are linked such that $\rho_{xx} = \sigma_{xx}/(\sigma_{xx}^2 + \sigma_{xy}^2)$, and similarly for ρ_{xy}. Now R_{xy} is no longer linear, instead curving up or down depending on the sign and ratio of carrier density of the two carriers, and R_{xx} is no longer field independent, instead increasing as a function of B. To determine the values for $n_{1(2)}$ and $\mu_{1(2)}$ from measured $R_{xx}(B)$ and $R_{xy}(B)$ data, a simultaneous least squares fitting must be performed. This is then done across a range of temperatures to characterise the sample, and in turn, the material. An example of typical measured $R_{xx}(B)$ and $R_{xy}(B)$ (converted to ρ_{xx} and ρ_{xy}) data as a function of temperature is shown in Figure 3.27, and the results from 2 carrier fitting of this data are shown in Figure 3.28.
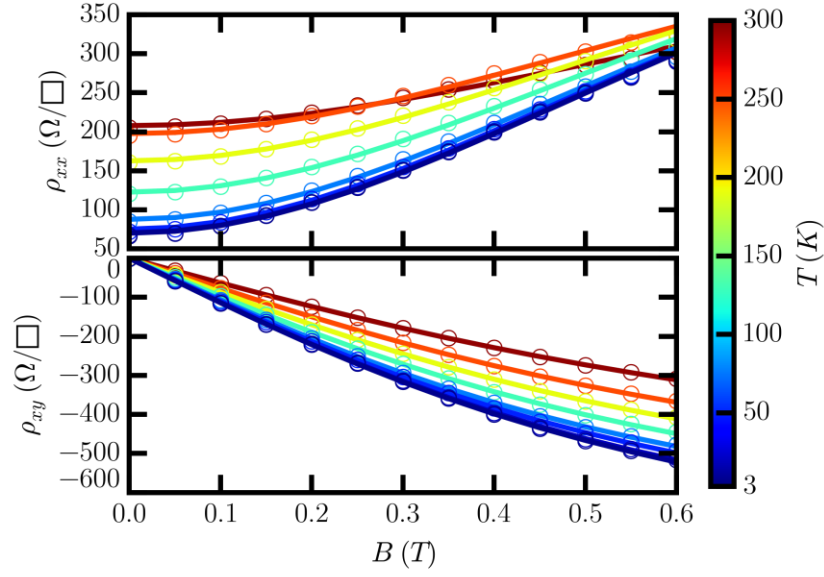
Figure 3.27 ρ_{xx} vs B (top) and ρ_{xy} vs B (bottom) data (○), lines correspond to 2 carrier fits of n_{2D} and μ to the data following equations (3.38) and (3.39) (where colour represents temperature). The multi-carrier nature can clearly be seen due to the curving of the lines.

(a)

(b)
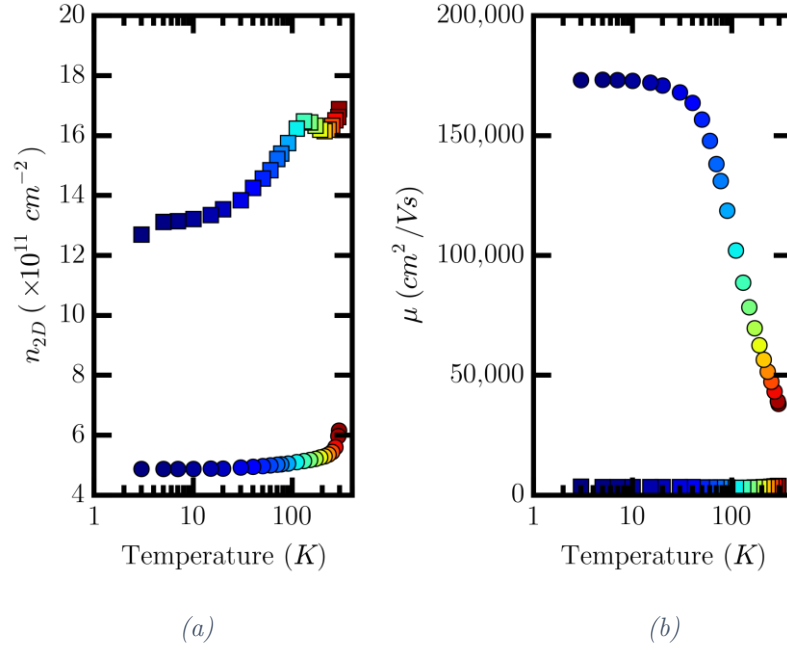
Figure 3.28 (a) Extracted carrier density (n_{2D}) and (b) mobility (μ) as a function of temperature for 2 carriers (○ and □ marker symbols) from the data and fits shown in Figure 3.27 (where colour represents temperature). The ○ data points correspond to 2D carriers in the well, having a characteristic plateau at low temperatures in both n_{2D} and μ (described further in the following chapter), whereas the □ data points comprise either carriers in the δ-doped plane or in the bulk of the semiconductor (or more likely both), where this can be inferred by the low mobility values.

4 : SCHRÖDINGER-POISSON AND TRANSPORT MODELLING

4.1 Introduction

In this chapter, the samples described in the previous chapter will be studied, including analysis of temperature dependant Hall measurements (used to determine the carrier densities and mobilities given in Table 3.3). From these measurements it is evident there are trends present in the mobility as a function of carrier density (see section 4.2). To investigate these further, it is important to understand how the band structure and energy level filling changes as a function of carrier density, with this achieved through self-consistent Schrödinger-Poisson modelling, discussed in section 4.3. With knowledge of how the band structure and properties of the samples vary as a function of carrier density, transport lifetime modelling is performed with the aim of understanding the scattering mechanisms limiting the observed low temperature mobilities (section 4.4). Applying this model using standard scattering mechanisms (and using more refined values than those used previously [16, 17]), it is shown that there is a clear discrepancy between the expected and measured mobility values. To explain this, a new, unaccounted for scattering mechanism is required. The mechanism is shown to be related to surface features (hillocks) present on all samples, and observed via differential interference contrast (DIC) microscopy (Nomarski) imaging. The physics of Nomarski imaging, as well as description of the image analysis techniques exploited to determine a mean surface feature size is given in section 4.5. This section continues to bring together this surface analysis with the transport lifetime model developed in section 4.4, and shows that there is a strong correlation between the surface feature sizes observed and the largest measured mean free paths. Including this scattering in the transport modelling gives good agreement with observed mobility values for multiple samples. Section 4.5.1 then examines the origin of these surface hillocks, including work in the literature aimed at reducing their density and improving the material quality [99, 100, 101]. Finally the chapter is then summarised in section 4.6.

4.2 Mobility as a Function of Carrier Density

For the InSb QWs studied here, the variation in structure between samples is either given by an altered level of doping or an altered spacer between the well and the δ-dopant

plane. These samples were fabricated into Hall bar devices and measured as a function of temperature between $3\,K$ and $300\,K$, and between B-fields of $0\,T$ and $0.6\,T$, to determine a sample sheet carrier density (n_{2D}) and sample mobility (μ). The results of these measurements at $3\,K$ were summarised in Table 3.3 and are plotted in Figure 4.1, alongside data from previous growth batches (with nominally similar structures [16, 17, 75, 76]).
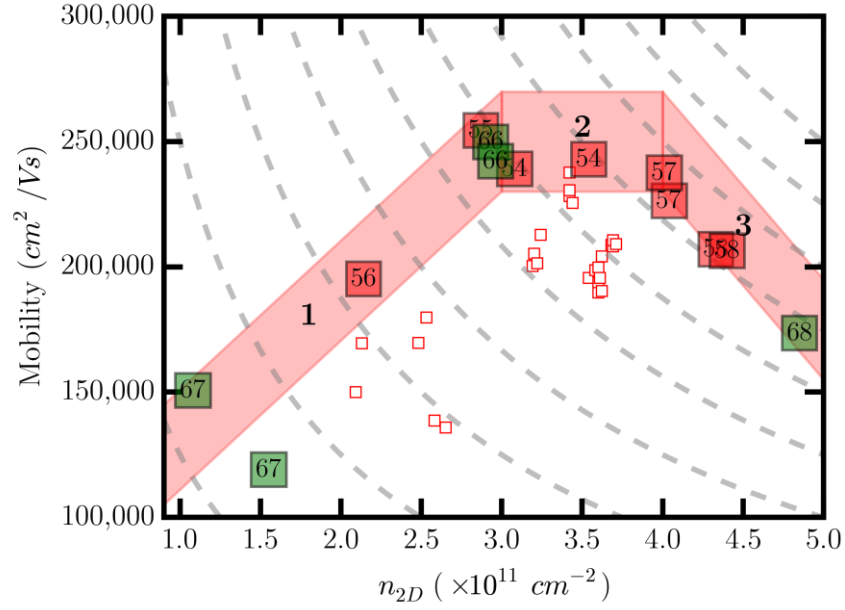
Figure 4.1 Measured $3\,K$ mobility as a function of carrier density (from 2 carrier fitting) for the batch IV samples studied in this work (large, filled squares), labelled with sample name (SF10xx) as given in Table 3.3, and historical samples (small, unfilled squares [16, 17, 75, 76]). Samples with varying dopant are shown by red filled squares whilst green filled squares show the varied spacer samples. Clear trends are shown, with these regions highlighted and numbered. In region 1, interface roughness scattering dominates, where the increasing mobility is due to increased screening, including Thomas-Fermi screening of the charged impurity centres. In region 2, the mobility is dominated by scattering related to features visible on the surface. In region 3, multisubband occupancy and intersubband scattering become dominant. Samples are predominantly single carrier until region 3 where multicarrier behaviour is observed. Grey dashed lines are contours of constant conductivity from $2\,mS$ to $20\,mS$.

Figure 4.1 shows that at a given carrier density, the sample batch studied here (solid, labelled squares) have a consistently higher mobility than previous samples, (empty squares) with this attributed to improved growth conditions and optimised growth parameters for MBE. Figure 4.1 also shows clear trends evident across this batch of samples, where there is initially an increasing mobility observed for an increasing carrier density (labelled region 1). Following this, there is a small plateau of mobility, occurring

at a value of ~250,000 cm^2/Vs for a range of carrier densities from ~3×10^{11} cm^{-2} to ~4×10^{11} cm^{-2} (labelled region 2). Finally, for an increase in carrier density, a corresponding decrease in mobility is observed (region 3). The samples measured are all predominately single carrier until region 3, where multiple carriers are observed.

The following sections focus on describing these 3 regimes using Schrödinger-Poisson and transport lifetime modelling to determine the scattering mechanisms involved, attributing the principle physical cause for each region [29, 30]. It will be shown subsequently that in region 1, the dominant mechanism for the increasing mobility trend observed is screening of the interface roughness scattering, as well as Thomas-Fermi screening of the charged impurity centres (i.e. remote ionised impurities and background impurities). Region 2 is dominated by a surface feature related scattering mechanism, giving an absolute limit to the mobilities observed, whilst in region 3, multisubband occupancy and intersubband scattering become important. In region 3, samples with measured carrier densities above ~4×10^{11} cm^{-2} become visibly 2 carrier in nature (as shown in Figure 3.27 for sample SF1068).

4.3 Schrödinger-Poisson Band Structure Modelling

To analyse the trends shown in Figure 4.1, knowledge of the band structure and scattering mechanisms is needed. To determine the band structure of each structure, and how this differs with carrier density, self-consistent Schrödinger-Poisson modelling was performed, using the basic structure and band alignments given previously [74]. This section describes this modelling, and investigates the level of doping required to achieve the observed carrier densities.

Whilst tellurium dopant has a high activation in InSb based samples, not all of the donated electrons will enter the well to act as charge carriers. Instead a number will transfer to the abundant surface states that cause mid-gap pinning. This non-trivial relationship is important for understanding the ratio of the filling of levels in the well compared to the δ-dopant plane, where if this plane is doped too heavily the conduction band edge will be below the Fermi level and the plane will be populated at all temperatures. Low lying states in the dopant plane can act as available states for electrons from the well to scatter into, but with a significantly reduced mobility due to alloy scattering and the non-remote ionic cores of the dopant atoms.

The amount of dopant required to achieve the carrier density measured in the well is also important for scattering related to this dopant (remote ionised impurity scattering),

where the effect on the transport lifetime depends on the amount of charge as well as the distribution.

To determine this amount and distribution of charge, and the subsequent band structure, energy levels, and filling, self-consistent Schrödinger-Poisson (S.P.) modelling is required [74]. For the subsequent modelling, following references [29] and [49] (and for brevity), only a temperature of $T = 10\ K$ was considered[14], using the values given in Table 3.1, as the aim of this aim of this section is then to help explain the cold temperature limiting scattering mechanisms. At this temperature, the thermal energy is given by $k_B T \approx 0.85\ meV$, where this energy determines whether electrons will occupy electronic states above the Fermi energy. If a state is several $k_B T$ above E_F, the state can be considered remote and unoccupied.

Each layer (l) input to the S.P. solver must contain a finite thickness, t (here in Å), and a dopant level $N_{\delta l}$ in cm^{-3}. For the simulations performed here, each layer of the dragged dopant distribution (Figure 3.8) is taken to be 10 Å thick, and the total input dopant is converted into a 2D density n_δ by summing over these layers, such that

$$n_\delta\ (cm^{-2}) = \sum_d N_{\delta 0} e^{\frac{d}{\lambda}} \times \frac{t}{10} \times \frac{1\ cm}{10^7\ nm}. \qquad (4.1)$$

Here, the final multipliers represent a conversion of the layer thickness into nm, and a conversion from nm to cm. The parameter d in equation (4.1) denotes the distance from the initial dopant plane ($d = 0$) to the surface (where this is the top cap thickness minus the spacer, typically $d = 25\ nm$), whilst λ represents the decay length. As shown by the SIMS data in Figure 3.8, the level of doping at the surface drops to a value of ~1/9th that in the initial dopant plane, giving $\lambda \approx -11.4\ nm$. The level of dopant in a sample can then either be characterised by the 3D value in the first dopant plane ($N_{\delta 0}$), or the total 2D dopant density (n_δ), where for a standard 25 nm distribution of dopant,

$$n_\delta\ (cm^{-2}) = N_{\delta 0}\ (cm^{-3}) \times (1.01 \times 10^{-6}\ cm). \qquad (4.2)$$

---

[14] Simulations were performed at a temperature of $T = 10\ K$ as there are reliable parameterisations for this temperature [49], however any small change in this temperature will have little impact on these values, nor on the subsequent results.

An illustration of how the dopant density affects the band structure is shown in Figure 4.2 (for a fixed N_{bkg}), whilst Figure 4.3 illustrates how the S.P. solutions are sensitive to the level of 3D background charge (N_{bkg}) present throughout the sample (for a fixed n_δ).
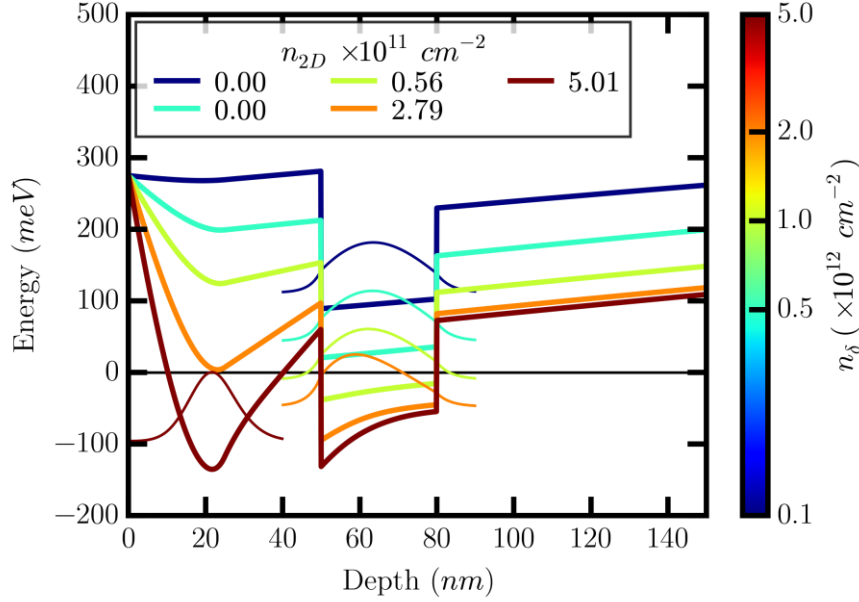
Figure 4.2 Self-consistent Schrödinger-Poisson solutions [74] for the conduction band edge (thick lines) and ground state wavefunctions (thin lines) for a range of dopant densities, n_δ (the colour of each curve corresponds to the level of n_δ). The Fermi energy is set to 0 meV and the corresponding carrier density (n_{2D}) for each n_δ is given in the legend. This figure shows that for low values of n_δ, the ground state wavefunction is above the Fermi energy and there are no carriers in the well. Increasing n_δ increases the carrier density in the well until the dopant plane becomes occupied at the highest n_δ. The background dopant density is set at $N_{bkg} = 5 \times 10^{14}$ cm^{-3} p-type for all solutions.

It is clear from Figure 4.2 that if the dopant level is too low, then there will be little charge transfer to the well, and the band bending will not be significant enough to lower the conduction band edge and the ground state energy, shown by the wavefunction, below the Fermi level (set to $0\ meV$). This low dopant level is shown in Figure 4.2 for $n_\delta \approx 0.1 \times 10^{12}\ cm^{-2}$, and $n_\delta \approx 0.5 \times 10^{12}\ cm^{-2}$, where in this instance, there are no free carriers in the well (i.e. the carrier density is 0 cm^{-2}). However, if the dopant level is increased further, the energy of the ground state is pulled below the Fermi energy and a carrier density is achieved, where this occurs at $n_\delta \approx 10^{12}\ cm^{-2}$. Further increases lower the conduction band edge in the well further, but simultaneously lower the conduction band edge in the presence of the dopant. Eventually the conduction band edge here also is lowered below the Fermi level, and so a carrier population will arise in this dopant plane

(as well as in the well). This occurs at $n_\delta \approx 2 \times 10^{12} \; cm^{-2}$. Finally, as the dopant level increases again, the conduction band edge in the plane is dragged lower than in the well, and the ground state energy of the system shifts to this dopant plane. This is shown in Figure 4.2 for a value of $n_\delta \approx 5 \times 10^{12} \; cm^{-2}$. What is also clear from this figure is how at higher dopant levels, the well becomes increasingly asymmetrical, and by extension, so does the wavefunction. This change in shape of the wavefunction moves the peak (the location of the majority of charge) closer to the dopant plane, where this will marginally increase scattering.
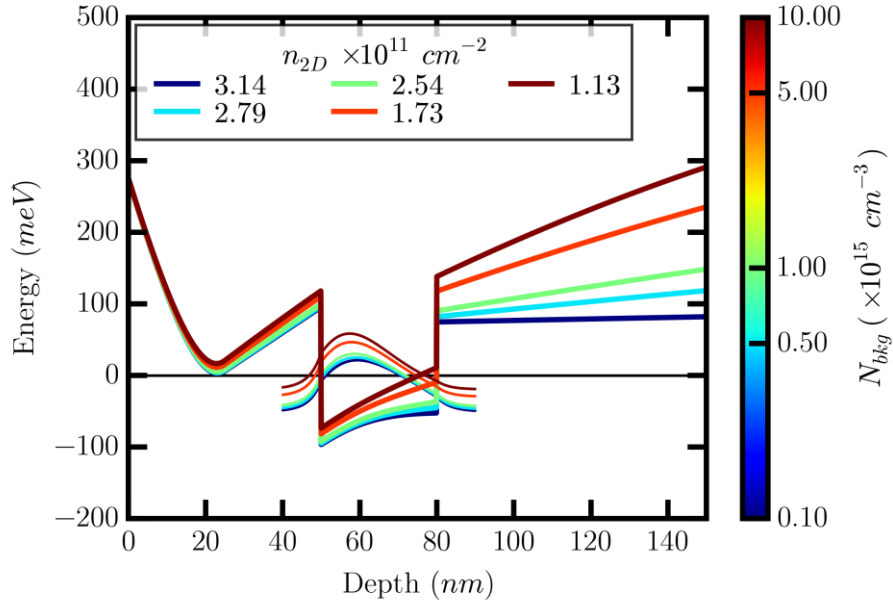
Figure 4.3 Similar to Figure 4.2, self-consistent Schrödinger-Poisson solutions [74] for varying p-type N_{bkg} levels, with colour representing N_{bkg}. For all solutions $n_\delta = 2 \times 10^{12} \; cm^{-2}$. This figure shows that, as expected, higher N_{bkg} raises the band structure below the well (bringing E_F closer to the valence band), with this giving corresponding lower carrier densities in the well.

Figure 4.3 shows that the S.P. solutions are also sensitive to the background level of charge, where the majority of this charge is predominantly based in the $3 \; \mu m$ buffer layer. In previous works investigating the transport properties of similar InSb samples [16, 17], the background dopant has been assumed p-type with levels as high as $N_{bkg} \sim 1 \times 10^{16} \; cm^{-3}$. More recent measurements of samples consisting of just an $Al_{0.1}In_{0.9}Sb$ buffer layer (i.e. without the well or top cap, termed a flat layer), grown during the same growth batch as the samples shown in Figure 4.1, has revealed a background acceptor concentration of no more than $N_{bkg} < 5 \times 10^{14} \; cm^{-3}$ [29], significantly less than the values used previously.

To analyse the effect of the level of N_{bkg} and n_δ on the energy levels and carrier densities achieved in the well S.P. modelling was performed for a range $0 < n_\delta \leq 5 \times 10^{12}\ cm^{-2}$ and $10^{14}\ cm^{-3} \leq N_{bkg} \leq 10^{16}\ cm^{-3}$. Figure 4.4 illustrates how the carrier density in the well (n_{2D}) varies as a function of dopant density (n_δ) whilst Figure 4.5 shows how the ground state energy (E_1), and higher energy levels vary as a function of n_{2D} due to varying dopant levels for 3 different N_{bkg} values. The specific value corresponding to the band structures shown in shown in Figure 4.2 are marked with $\square$ markers, whilst those corresponding to Figure 4.3 are marked with $\bigcirc$, the values common between both are marked with a $\diamondsuit$ marker. An example of the wavefunction solutions for the band structure corresponding to the $\diamondsuit$ marker ($N_{bkg} = 5 \times 10^{14}\ cm^{-3}$, $n_\delta = 2 \times 10^{12}\ cm^{-2}$) is shown in Figure 4.6.
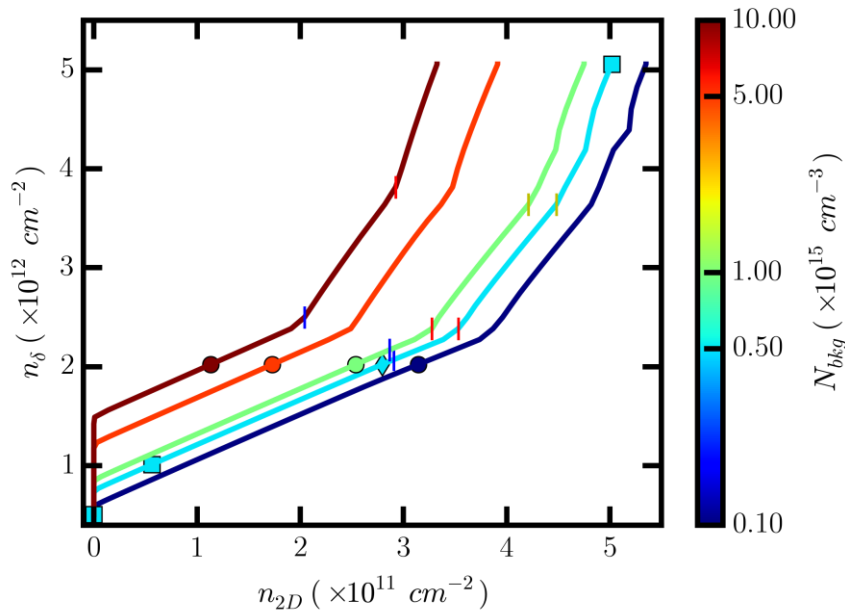
Figure 4.4 Carrier density in the quantum well (n_{2D}) vs dopant density (n_δ) for a varying N_{bkg} level (the colour of each curve represents the level of N_{bkg}). The specific value corresponding to the band structures shown in shown in Figure 4.2 are marked with $\square$ markers, whilst those corresponding to Figure 4.3 are marked with $\bigcirc$, the values common between both are marked with a $\diamondsuit$ marker (with this band structure shown in Figure 4.6. Small vertical lines denote the carrier density at which the corresponding energy levels shown in Figure 4.5 cross the Fermi energy (i.e. become second and third carrier at the blue and red vertical lines respectively.).
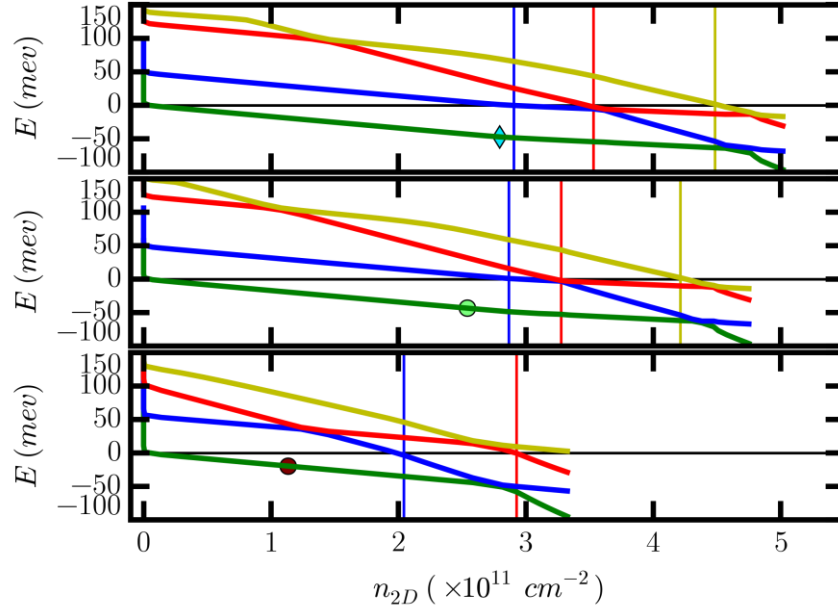
Figure 4.5 Energy level of wavefunction states E1 (green), E2 (blue), E3 (red) and E4 (yellow) as a function of the carrier density in the well (n_{2D}) for 3 levels of N_{bkg}: $N_{bkg} = 5 \times 10^{14}\ cm^{-3}$ (top), $N_{bkg} = 1 \times 10^{15}\ cm^{-3}$ (middle) and $N_{bkg} = 1 \times 10^{16}\ cm^{-3}$ (bottom). Vertical lines denote the carrier density at which the corresponding energy level crosses the Fermi level (i.e. becomes multi carrier).

Figure 4.4 (along with Figure 4.5) clearly shows the statement made previously that for a background acceptor level of ~$5 \times 10^{14}\ cm^{-3}$, below $n_\delta \approx 0.7 \times 10^{12}\ cm^{-2}$, the ground state energy associated with the well, E_1, is above the Fermi level (> 0), and there are no carriers in the well ($n_{2D} = 0$). Above this dopant level, n_{2D} begins to increase smoothly with n_δ, and E_1 decreases smoothly below 0, giving only a single energy level populated. This trend continues until $n_\delta \approx 2 \times 10^{12}\ cm^{-2}$ ($n_{2D} \approx 2.8 \times 10^{11}\ cm^{-2}$), when carriers first begin to populate the second energy level in the well, E_2, though this is still a high mobility state. A further small increase in dopant causes carriers to also populate the dopant plane, E_3, where this occurs at $n_{2D} \approx 3.6 \times 10^{12}\ cm^{-2}$ and is visible as a clear change in the relationship between n_δ and n_{2D}. This is a low mobility state, where the scattering rate and effective mass are significantly higher than in the well, with this agreeing with the reduced mobility and 2 carrier nature of the highest carrier density samples shown in Figure 4.1.

In Figure 4.5, several anti-crossings are present, where at these points the states associated with each energy level swap, i.e. for the anti-crossing between state $E2$ and $E3$, at low n_{2D} these are associated with the well and the dopant plane respectively, whilst at higher n_{2D}, $E3$ is associated with the well and $E2$ with the dopant plane. At all n_{2D}, the energy level with the steepest gradient as a function of n_{2D} is associated with the dopant

plane, as is evident when $n_{2D} {\sim} 5 \times 10^{12}\ cm^{-2}$ for $N_{bkg} {\sim} 5 \times 10^{14}\ cm^{-3}$, where the ground state energy is associated with the dopant plane, as shown in Figure 4.2.

The effect of the background charge level is also clear from the curves in Figure 4.4 and Figure 4.5, with higher N_{bkg} levels requiring increased values of n_{δ} to achieved comparable carrier densities to the lower N_{bkg} values. This increased doping has the effect of causing energy level crossings, the onset of multilevel filling, and occupation of the dopant plane to occur at comparatively lower carrier densities. These are not observed in Hall effect measurements of samples, showing that the increased values for N_{bkg} assumed previously were overestimates. This also confirms that the measured value of $N_{bkg} < 5 \times 10^{14}\ cm^{-3}$ p-type is consistent between the flat layer measurements and Schrödinger-Poisson simulation solutions of quantum well samples.

For the level of dopant required for a given carrier density, a straight line relationship can be fitted to the curves shown in Figure 4.4 relating carrier density to the amount of dopant. This can then be used in transport modelling across multiple samples with varying carrier densities, investigating the remote ionised impurity scattering for these samples. Using a least squares fitting to the curve for $N_{bkg} = 5 \times 10^{14}\ cm^{-3}$ then gives this simple relationship as

$$n_{\delta} = 0.45 \times n_{2D} + 0.76. \qquad (4.3)$$

The other important parameter required to calculate the remote ionised impurity scattering rate that can be extracted from S.P. modelling is the distance from the peak of the wavefunction (the location of charge in the well) to the initial dopant plane, where this is labelled d. This relationship is shown in Figure 4.7 for multiple values of N_{bkg}, where the $\square$, $\bigcirc$ and $\Diamond$ markers have the same meaning as previously. Again, a least squares fit can be performed to give d as a function of carrier density (with the fit performed to the region where the ground state remains in the well, i.e. $d > 30\ nm$). For $N_{bkg} = 5 \times 10^{14}\ cm^{-3}$,

$$d\ (nm) = -1.36 \times n_{2D} + 37.96. \qquad (4.4)$$

where n_{2D} is in units of $10^{11}\ cm^{-2}$.
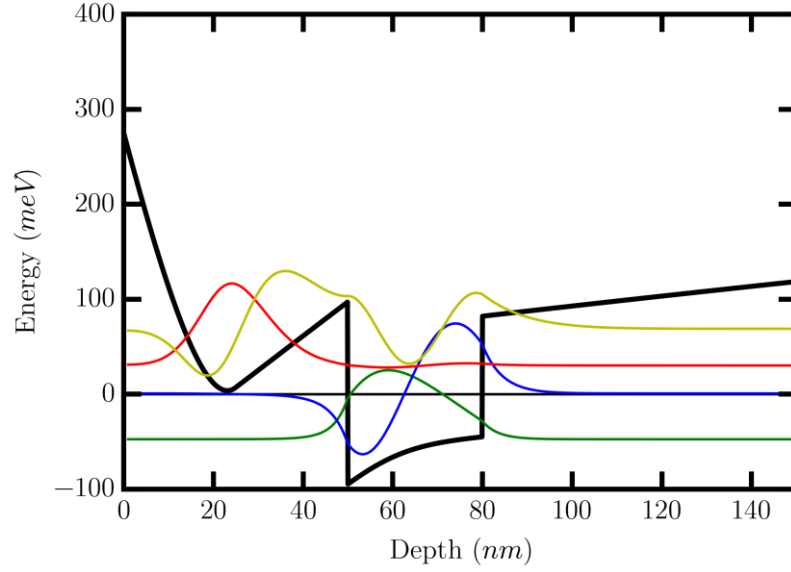
Figure 4.6 Self-consistent Schrödinger-Poisson solution for $N_{bkg} = 5 \times 10^{14}$ cm^{-3} and $n_\delta = 2 \times 10^{12}$ cm^{-2}, corresponding to the diamond markers in Figure 4.4 and Figure 4.5. The conduction band edge is shown in black, the ground state (E1) wavefunction in green, E2 in blue, E3 localised in the dopant plane in red and E4 in yellow. The colours of the energy level correspond to those in Figure 4.5. The Fermi level is set at 0 meV.
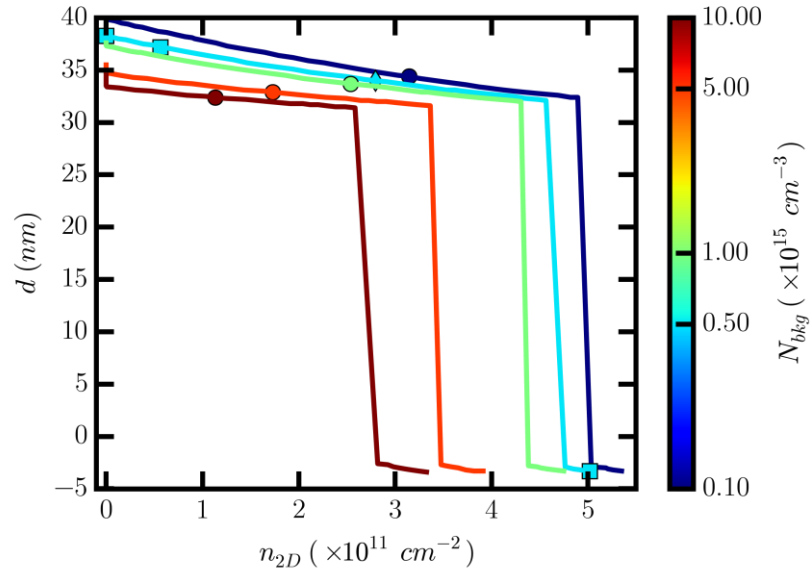
Figure 4.7 Separation between the peak of the ground state wavefunction and the δ-doped layer (d) as a function of carrier density in the well (n_{2D}) for various N_{bkg} values (the colour of the lines corresponds to the N_{bkg} level). The sharp decrease at higher n_{2D} corresponds to the ground state localising in the dopant plane rather than the well.

Using a combination of equations (4.3) and (4.4), and the other results from this S.P. modelling, the scattering rates for remote ionised impurities and other scattering mechanisms can be determined.

4.4 Transport Lifetime Modelling

To model the mobilities measured as a function of temperature for the samples shown in Figure 4.1, and to therefore determine the limiting scattering rates and the overall high-mobility limiting scattering mechanism, a relaxation time approximation transport lifetime model was implemented. This transport model follows closely those described previously by Orr et al. in their work on similar InSb devices [17, 48], and also includes the Schrödinger-Poisson modelling performed above. Non-parabolicity is included via a 4-band $\boldsymbol{k} \cdot \boldsymbol{p}$ model for a non-parabolic effective mass [16, 49].

Several predominant scattering mechanisms present in III-V heterostructures are considered, including the effects of phonons (both optical and acoustic) which give rise to non-elastic scattering, ionised impurity scattering (remote ionised scattering from dopant and scattering from background p-type charge) and interface roughness of the walls of the QW. These latter two mechanisms both change the electron momentum, but not the energy, giving rise to elastic scattering [63]. These various scattering mechanisms are then combined using Mathieson's rule (equation (4.43)). The effect of dopant dragging in the top cap is also considered.

The transport model material parameters were taken from references [11, 12, 17, 69, 107, 108].

4.4.1 Phonon scattering

In a rigid crystal lattice, such as in InSb, vibrations of the constituent atoms can occur due to thermal energy, however due to the crystal bonding, these vibrations are not independent. Instead, only certain modes of vibration are allowed, and a phonon is then a quasi-particle with energy $\hbar\omega_k$, quantifying this vibration. As the thermal energy is increased, the number of phonons will correspondingly increase, and so phonon scattering is often the dominant scattering mechanism at high temperatures, regardless of sample structure or design. Due to this, the phenomenon of phonon scattering in semiconductors is well documented [109, 110, 111, 112, 113, 114, 115]. For InSb heterostructures, the two main vibrational modes limiting transport lifetimes are the deformation-potential transverse-acoustic mode, and the polar-optical mode.

The acoustic phonon scattering rate ($1/\tau_{Ac}$) is then given by [110, 116]

$$\frac{1}{\tau_{Ac}} = \frac{3m^*\Xi^2 k_B^2 T}{2\hbar^3 \rho_d v_s^2 w},$$

(4.5)

where Ξ is the deformation potential (related to strain), ρ_d is the crystal density, v_s is the longitudinal sound velocity and w is the quantum well width. For the samples studied here, $w = 30\ nm$, $\rho_d = 5790\ kg/m^3$, and $v_s = 3700\ m/s$. There is uncertainty in the literature over the value of the deformation potential in InSb, where the quoted values vary between $7.2\ eV$ and $30\ eV$. Whilst this value will have a large effect due to the inclusion of Ξ^2 in the acoustic phonon scattering rate, this effect will be minimal in the cold temperature, high mobility regime of interest here. Therefore a single value of $\Xi = 20\ eV$ has been used for simplicity.

Scattering from polar optical phonons is more complex, but following Ridley [111] and Price [115], the scattering rate ($1/\tau_{Op}$) is given by

$$\frac{1}{\tau_{Op}} = \frac{e^2 \omega_0 N(\omega_0) m^* w}{4\pi \varepsilon_p \hbar^2}$$

(4.6)

where $\varepsilon_p^{-1} = \varepsilon_\infty^{-1} - \varepsilon_s^{-1}$, and ε_∞ and ε_s are the high-frequency and static dielectric constants respectively ($\varepsilon_\infty = 15.7\varepsilon_0$, $\varepsilon_s = 16.85\varepsilon_0$ and $\varepsilon_0 = 8.85 \times 10^{-12}\ F/m$), and ω_0 is the frequency of the phonons, related to the optical phonon energy given by $\hbar\omega_0 = 25\ meV$ in InSb. The number of phonons occupying the ω_0 mode ($N(\omega_0)$) is given by the Boltzmann distribution

$$N(\omega_0) = \left(\exp\left(\frac{\hbar\omega_0}{k_B T}\right) - 1\right)^{-1},$$

(4.7)

assuming there is no intersubband scattering. This is a reasonable assumption in lightly doped samples due to the large energy level spacing of states in the well, ensuring a phonon cannot scatter an electron from one energy level to another.

4.4.2 Remote ionised impurity scattering

Where phonon scattering is related to temperature and lattice vibrations, impurity scattering is directly linked to the coulomb interaction between charged impurities and electrons in the well. When the impurities are located in a 2D plane, such as with ideal δ-doping, and separated from the charge carriers, the scattering is known as remote ionised impurity (Rii) scattering. However, if the charged impurities are unwanted atoms

incorporated during MBE, and are spread equally across 3 dimensions, including in the plane of the QW, this is referred to as background impurity scattering (Bkg).

For both impurity scattering mechanisms, the scattering rate can be derived using a combination of Fermi's golden rule, for the rate of scattering from an initial state i to a final state f, and the Born approximation, relating this scattering rate to the Fourier transform of the perturbation potential. Much of the following derivations are taken from reference [63], chapters 8.2 and 9.5 (and 9.6 for background impurity scattering).

Fermi's golden rule states that the transition rate from a state i to state f is given by

$$W_{fi} = \frac{2\pi}{\hbar}|V_{fi}|^2\delta(\varepsilon_f - \varepsilon_i) \tag{4.8}$$

where $\varepsilon_{i(f)}$ is the energy of state $i(f)$, and V_{fi} is the matrix element of the perturbation potential $\hat{V}$, given by

$$V_{fi} = \int \phi_f^* \hat{V} \phi_i. \tag{4.9}$$

$\phi_{i(f)}$ are the initial (final) wavefunction states. Due to the δ-function in equation (4.8), Fermi's golden rule states that the initial and final state energies must be the same, so conserving energy, and giving elastic scattering.

For plane wave solutions, with $\boldsymbol{k}$ states separated by a momentum change $\hbar\boldsymbol{q}$, bounded to an area A,

$$\phi_i = \frac{1}{\sqrt{A}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{4.10}$$

$$\phi_f = \frac{1}{\sqrt{A}}e^{i(\boldsymbol{k}+\boldsymbol{q})\cdot\boldsymbol{r}}, \tag{4.11}$$

and for a perturbation from a simple point potential given by $\hat{V} = V(\boldsymbol{r})$, equation (4.9) gives

$$V_{fi} = \frac{1}{A}\int e^{-i(\boldsymbol{k}+\boldsymbol{q})\cdot\boldsymbol{r}} V(\boldsymbol{r})e^{i\boldsymbol{k}\cdot\boldsymbol{r}}d^2\boldsymbol{r} = \frac{1}{A}\int V(\boldsymbol{r})e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} d^2\boldsymbol{r}. \tag{4.12}$$

Equation (4.12) is a statement that the matrix element for the perturbation potential is proportional to the Fourier transform of the scattering potential ($V_{fi} = A^{-1}\tilde{V}(\boldsymbol{q})$), where

by inserting this into equation (4.8), the rate of scattering for an electron from state $\boldsymbol{k}$ to $\boldsymbol{k} + \boldsymbol{q}$ can be determined, such that

$$W_{\boldsymbol{k}+\boldsymbol{q},\boldsymbol{k}} = \frac{1}{A^2}\frac{2\pi}{\hbar}|\tilde{V}(\boldsymbol{q})|^2\delta(\varepsilon(\boldsymbol{k}+\boldsymbol{q})-\varepsilon(\boldsymbol{k})). \tag{4.13}$$

This is a statement of the Born approximation, that the scattering rate is proportional to the square of the modulus of the Fourier transform of the scattering potential. To determine the scattering rate for any $\boldsymbol{q}$, equation (4.13) must be summed over all $\boldsymbol{q}$, where this can be converted to an integral with each $\boldsymbol{q}$ state having an "area" $A/(2\pi)^2$, such that

$$\frac{1}{\tau} = \frac{A}{(2\pi)^2}\int W_{\boldsymbol{k}+\boldsymbol{q},\boldsymbol{k}}d^2\boldsymbol{q}. \tag{4.14}$$

This is then the scattering rate for a single impurity. In 2D, the total number of impurities in an area A is given by $N_{2D}^{imp} = An_{2D}^{imp}$, where n_{2D}^{imp} is the average 2D impurity density. Multiplying equation (4.14) by this factor, and substituting equation (4.13), the total scattering rate from 2D impurities can be derived (as a function of the potential $V(\boldsymbol{r})$), giving

$$\frac{1}{\tau} = n_{2D}^{imp}\frac{2\pi}{\hbar}\int|\tilde{V}(\boldsymbol{q})|^2\delta(\varepsilon(\boldsymbol{k}+\boldsymbol{q})-\varepsilon(\boldsymbol{k}))\frac{d^2\boldsymbol{q}}{(2\pi)^2}. \tag{4.15}$$

It is important to note that τ here is the single-particle or "quantum" lifetime (τ_q) and not the transport lifetime (τ_{tr}). The quantum lifetime describes the time an electron stays in a state before any scattering event, whereas the transport lifetime is related to current flow and conductivity, where small angle scattering has a much smaller effect than large angle scattering. If an electron is scattered through an angle θ, the component of the motion parallel to the electrons original motion is proportional to $\cos(\theta)$, so the scattering rate determining the transport lifetime must be weighted by a factor $1 - \cos(\theta)$ compared to equation (4.15). The relationship between $\boldsymbol{k}$, $\boldsymbol{k} + \boldsymbol{q}$ and θ can be understood via Figure 4.8.

As energy must be conserved, then the magnitude of $\boldsymbol{k}$ must equal $|\boldsymbol{k} + \boldsymbol{q}|$, which in $\boldsymbol{k}$ space can be visualised by two vectors of equal length and equal origin pointing to the initial and final states on a circle.
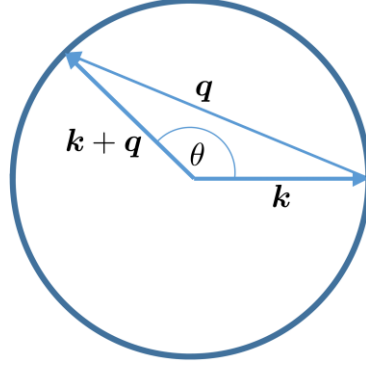
Figure 4.8 Relationship between $\boldsymbol{k}$-space wavevector $\boldsymbol{k}$ for an initial state, and wavevector $\boldsymbol{k} + \boldsymbol{q}$ for a final state. The scattering angle between initial and final states is θ, where $|\boldsymbol{k}| = |\boldsymbol{k} + \boldsymbol{q}|$ and $\sin(\theta/2) = q/2k$.

The angle between these vectors is then θ, where from Figure 4.8

$$\sin\left(\frac{\theta}{2}\right) = \frac{q/2}{k} \; ; q = 2k\sin\left(\frac{\theta}{2}\right). \tag{4.16}$$

Using the trigonometry identity $\cos(2\theta) = 1 - 2\sin^2(\theta)$, the transport scattering rate weighting factor is given by

$$1 - \cos(\theta) = 2\sin^2\left(\frac{\theta}{2}\right) = \frac{q^2}{2k^2}, \tag{4.17}$$

giving the scattering rate as

$$\frac{1}{\tau_{tr}} = n_{2D}^{imp}\frac{2\pi}{\hbar}\int\frac{q^2}{2k^2}|\tilde{V}(\boldsymbol{q})|^2\delta(\varepsilon(\boldsymbol{k}+\boldsymbol{q})-\varepsilon(\boldsymbol{k}))\frac{d^2\boldsymbol{q}}{(2\pi)^2}. \tag{4.18}$$

For simplicity, the subscript will be dropped subsequently, and where relevant, the distinction between the quantum and transport lifetimes will be made by labelling the quantum lifetime τ_q.

To simplify equation (4.18) further, the integration can be rewritten in terms of the final sate $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q}$, and expanded in cylindrical coordinates, substituting equation (2.5) for the energy, such that

$$\frac{1}{\tau} = \frac{n_{2D}^{imp}}{2\pi\hbar}\int_{-\pi}^{+\pi}d\theta\int_0^{\infty}k'dk'\frac{|\boldsymbol{k}'-\boldsymbol{k}|^2}{2k^2}|\tilde{V}(|\boldsymbol{k}'-\boldsymbol{k}|)|^2\delta\left(\frac{\hbar^2\boldsymbol{k}'^2}{2m^*}-\frac{\hbar^2\boldsymbol{k}^2}{2m^*}\right) \tag{4.19}$$

where

$$|\boldsymbol{k}' - \boldsymbol{k}|^2 = k'^2 + k^2 - 2k'k\cos(\theta). \tag{4.20}$$

Changing the variable of integration to enable integrating over the δ-function introduces a factor of the inverse of the derivative of the argument of the δ-function, namely $m^*/\hbar^2 k'$, with the integration then giving $k' = k$, $|\boldsymbol{k}' - \boldsymbol{k}| = q = 2k\sin(\theta/2)$ and

$$\frac{1}{\tau} = \frac{n_{2D}^{imp} m^*}{2\pi\hbar^3} \int_{-\pi}^{+\pi} d\theta \frac{2k^2(1-\cos(\theta))}{2k^2} \left| \tilde{V}\left(2k\sin\left(\frac{\theta}{2}\right)\right) \right|^2, \tag{4.21}$$

or with a reduced range of integration

$$\frac{1}{\tau} = \frac{n_{2D}^{imp} m^*}{\pi\hbar^3} \int_{0}^{+\pi} d\theta \left(1-\cos(\theta)\right) \left| \tilde{V}\left(2k\sin\left(\frac{\theta}{2}\right)\right) \right|^2. \tag{4.22}$$

Equation (4.22) can then be rewritten in terms of q as

$$\frac{1}{\tau} = \frac{n_{2D}^{imp} m^*}{2\pi\hbar^3 k^3} \int_{0}^{2k} |\tilde{V}(q)|^2 \frac{q^2 dq}{\sqrt{1-\left(\frac{q}{2k}\right)^2}}, \tag{4.23}$$

where the limit of integration is $2k$ corresponding to complete backscattering of the electron as the largest wavevector satisfying conservation of energy. The remote ionised impurity quantum lifetime scattering rate $(1/\tau_q)$ is then given by removing a factor of $(1 - \cos(\theta))$ from (4.22) or $q^2/2k^2$ from (4.23). Finally, as the majority of conduction occurs at the Fermi level in the cold temperature limit, k in equations (4.22) and (4.23) can be replaced by the Fermi wavevector k_F.

Whilst these equations are generic, to calculate the actual scattering rate, the matrix element $\tilde{V}(q)$ must be determined. For this, again following Davies [63], the unscreened potential from a charged impurity at a distance d_s (the spacer) is given by

$$\tilde{V}(q) = \int_{0}^{\infty} r dr \int_{0}^{2\pi} d\theta \frac{e^2}{4\pi\varepsilon\sqrt{r^2+d^2}} e^{iqr\cos(\theta)} = \frac{e^2}{2\varepsilon} \frac{e^{-qd_s}}{q}. \tag{4.24}$$

To account for screening by the 2DEG, $\tilde{V}(q)$ must be divided by the Thomas-Fermi dielectric function,

$$\epsilon_{TF} = 1 + \frac{q_{TF}}{q}, \tag{4.25}$$

where q_{TF} is the Thomas-Fermi wavevector, given by

$$q_{TF} = \frac{m^*e^2}{2\pi\epsilon\hbar^2} = \frac{2}{a_B^*}, \tag{4.26}$$

and a_B^* is the effective Bohr radius. This gives the screened matrix element and Rii scattering rate as

$$\tilde{V}(q) = \frac{e^2}{2\varepsilon}\frac{e^{-qd_s}}{q + q_{TF}} \tag{4.27}$$

and

$$\frac{1}{\tau_{Rii}} = \frac{n_{2D}^{imp}m^*}{2\pi\hbar^3 k_F^3}\left(\frac{e^2}{2\varepsilon}\right)^2 \int_0^{2k_F} \frac{e^{-2q|d_s|}}{(q + q_{TF})^2}\frac{q^2 dq}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}}, \tag{4.28}$$

respectively.

One final refinement can be made to equation (4.28) for scattering from a single plane of remote ionised impurities by consideration of the extent of the wavefunction out of the plane of the 2DEG (i.e. the width of the wavefunction in the z-direction). This correction depends on the Fang-Howard parameter b (related to the shape of the wavefunction), with the scattering rate given by:

$$\frac{1}{\tau_{Rii}} = \frac{n_{2D}^{imp}m^*}{2\pi\hbar^3 k_F^3}\left(\frac{e^2}{2\varepsilon}\right)^2 \int_0^{2k_F} \frac{e^{-2q|d_s|}}{(q + q_{TF}G(q))^2}\left(\frac{b}{b+q}\right)^6 \frac{q^2 dq}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}} \tag{4.29}$$

$$G(q) = \frac{1}{8}\left(2\left(\frac{b}{b+q}\right)^3 + 3\left(\frac{b}{b+q}\right)^2 + 3\left(\frac{b}{b+q}\right)\right) \tag{4.30}$$

$$b = \left(\frac{33m^*e^2 n_{2D}}{8\hbar^2\varepsilon}\right)^{\frac{1}{3}}. \tag{4.31}$$

To extend equation (4.29) to include the effect of dopant dragging in the top cap, the variation in n_{2D}^{imp} as a function of distance must be considered. In a perfect δ-doped sample, $n_{2D}^{imp}(d) = n_{2D}^{imp}\delta(d - d_s)$, where d_s is the spacer distance, and this can be included via integrating equation (4.29) from $d = d_s$ to $d = d_0$ at the surface, such that

$$\frac{1}{\tau_{Rii}} = \frac{n_{2D}^{imp}m^*}{2\pi\hbar^3 k_F^3}\left(\frac{e^2}{2\varepsilon}\right)^2 \int_{d_s}^{d_0}\int_0^{2k_F} \frac{\delta(d - d_s)e^{-2q|d|}}{(q + q_{TF}G(q))^2}\left(\frac{b}{b+q}\right)^6 \frac{q^2 dq\ dd}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}}. \quad (4.32)$$

By integrating over the δ-function, the only non-zero value is achieved when $d = d_s$ and equation (4.32) is then equivalent to equation (4.29). If dopant is present in more than one plane p, with $d_p \geq d_s$, then $n_{2D}^{imp}(d) \propto \sum_p \delta(d - d_p)$, and the amount of charge in each plane is inversely proportional to the number of planes, giving

$$\frac{1}{\tau_{Rii}} \propto \frac{m^*}{2\pi\hbar^3 k_F^3}\left(\frac{e^2}{2\varepsilon}\right)^2 \sum_p \int_{d_s}^{d_0}\int_0^{2k_F} \frac{\delta(d - d_p)e^{-2q|d|}}{(q + q_{TF}G(q))^2}\left(\frac{b}{b+q}\right)^6 \frac{q^2 dq\ dd}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}}. \quad (4.33)$$

In the transport model here, these planes have been taken as individual monolayers of 2D dopant density, where the dopant level in each plane follows the exponential decay given by SIMS measurements (Figure 3.8 and equation (4.1)). However, for a purely continuous distribution of dopant in d, this becomes instead a 3D dopant distribution (with a value N_{3D} at $d = d_s$), and the continuous form of equation (4.1) can be used, giving the final scattering rate as

$$\frac{1}{\tau_{Rii}} = \frac{m^*}{2\pi\hbar^3 k_F^3}\left(\frac{e^2}{2\varepsilon}\right)^2 \int_{d_s}^{d_0}\int_0^{2k_F} \frac{N_{3D}e^{-\frac{d}{\lambda}}e^{-2q|d|}}{(q + q_{TF}G(q))^2}\left(\frac{b}{b+q}\right)^6 \frac{q^2 dq\ dd}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}}. \quad (4.34)$$

4.4.3 Background charged impurity scattering

To determine the scattering rate from background p-type impurities, it must first be assumed that these impurities are a uniform density throughout the sample, including in the plane of the 2DEG. With this assumption, and using the simplified form of the remote ionised impurity scattering rate (equation (4.28)), an integration over d can be performed to determine the scattering rate from 3D impurities. The simplified equation is used due to the associated uncertainties in background impurity level, meaning more refined

equations do not increase the accuracy of the result. Then the terms in equation (4.28) that are dependent on d become

$$n_{2D}^{imp} \rightarrow n_{3D}^{imp} = N_{bkg}, \tag{4.35}$$

$$e^{-2q|d_s|} \rightarrow \int_{-\infty}^{+\infty} e^{-2q|d|} dd = \frac{1}{q}, \tag{4.36}$$

and the final scattering rate from charged background impurities is given by

$$\frac{1}{\tau_{Bkg}} = \frac{N_{bkg} m^*}{2\pi \hbar^3 k_F^3} \left(\frac{e^2}{2\varepsilon}\right)^2 \int_0^{2k_F} \frac{q}{(q+q_{TF})^2} \frac{dq}{\sqrt{1 - \left(\frac{q}{2k_F}\right)^2}}. \tag{4.37}$$

4.4.4 Interface roughness scattering

The final principle scattering mechanism often considered in these materials systems is interface roughness, where this is scattering due to irregularities in the interface between the AlInSb and InSb layers. These irregularities create a fluctuating potential, altering the width of the well and the associated energy levels. The method of determining the scattering due to interface roughness is similar to that performed for remote ionised impurities, where a scattering matrix element is determined (V_{fi} or $\tilde{V}(q)$ as in equation (4.12)) for a 2D density of fluctuations, as opposed to a 2D impurity density, and the scattering rate calculated using equation (4.23).

The analysis to determine $\tilde{V}(q)$ has been performed by Penner et al. [117], with this summarised here. Using the assumption that the barriers of the QW can be considered infinite with respect to the ground state energy level of the well, the ground state energy level is given by equation (3.7)

$$E_{QW} = \frac{\pi^2 \hbar^2}{2m^* L^2}. \tag{3.7}$$

The magnitude of perturbation potential associated with a fluctuation of the well width due to roughness of the interfaces is then given by (replacing L with the well width w)

$$\hat{V} = \left| \frac{\partial E_{QW}}{\partial w} \right| = \frac{\pi^2 \hbar^2}{m^* w^3} \Delta(\boldsymbol{r}) \tag{4.38}$$

where $\Delta(\boldsymbol{r})$ is the rms height variation (roughness), Δ, as a function of position (see Figure 4.9). The scattering matrix element is then given by the 2D Fourier transform of the perturbation potential, giving

$$|\tilde{V}(q)|^2 = \frac{1}{A} \left(\frac{\pi^2 \hbar^2}{m^* w^3} \right)^2 S(q) \tag{4.39}$$

where $S(q)$ is the Fourier transform of the autocorrelation function of the well width fluctuations, $S(r)$. The function $S(r)$ describes the probability that the roughness at a position $\boldsymbol{r} + \boldsymbol{r}'$ will be the same as at position $\boldsymbol{r}$, where this is described using a Gaussian function such that

$$S(r) = \frac{1}{A} \int \Delta(\boldsymbol{r}' - \boldsymbol{r})\Delta(\boldsymbol{r}')d^2\boldsymbol{r}' = \Delta^2 \exp\left(-\frac{r^2}{\Lambda^2} \right) \tag{4.40}$$

where Λ is the characteristic extent of the roughness (the correlation length, see Figure 4.9). The scattering matrix element is then proportional to the Fourier transform $S(q)$, giving [118]

$$S(q) = \pi \Delta^2 \Lambda^2 \exp\left(-\frac{q^2 \Lambda^2}{4} \right). \tag{4.41}$$

To determine the final scattering rate, equation (4.41) must be substituted into equation (4.39) and then using equation (4.23) (neglecting n_{2D}^{imp} due to the inclusion of the 2D nature of the interface roughness), the interface roughness scattering rate is given by

$$\frac{1}{\tau_{Int}} = \frac{\pi^4 \hbar \Delta^2 \Lambda^2}{2m^* w^6 k_F^3} \int_0^{2k_F} \exp\left(-\frac{q^2 \Lambda^2}{4} \right) \frac{q^2 dq}{\sqrt{1 - \left(\frac{q}{2k_F} \right)^2}}. \tag{4.42}$$

For typical MBE grown InSb quantum wells, Δ is typically of the order of a few monolayers, whilst Λ is typically ~20 nm [48, 119, 120].
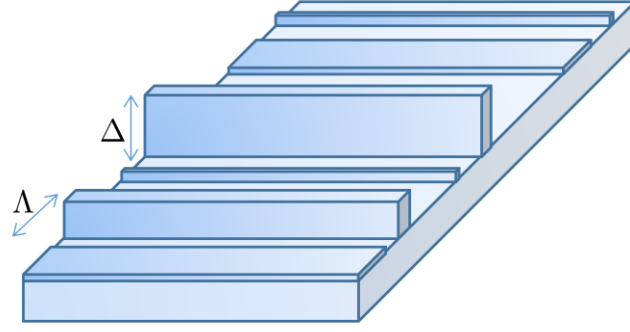
Figure 4.9 Schematic example of interface roughness in the plane of the well, with a typical rms height variation (roughness) Δ, and correlation length Λ.

4.4.5 Combined model

To determine the overall mobility as a combination of the transport lifetimes given by these scattering mechanisms, the average transport lifetime must first be combined using Mathieson's rule, where this is given by

$$\frac{1}{\tau} = \frac{1}{\tau_{Op}} + \frac{1}{\tau_{Ac}} + \frac{1}{\tau_{Rii}} + \frac{1}{\tau_{Bkg}} + \frac{1}{\tau_{Int}} \qquad (4.43)$$

for optical phonons, acoustic phonons, remote ionised impurities, background charged impurities and interface roughness respectively. The mobility is then given by equation (3.23), $\mu = e\tau/m^*$. Table 4.1 lists the InSb specific values important for these scattering mechanisms, as well as comparing to the other III-V materials GaAs and InAs.

Using the parameters given in Table 4.1, combined with the scattering rate equations given above, and the results from the Schrödinger-Poisson modelling, the transport model can be used to attempt to match measured carrier densities and mobilities, as a function of temperature, with the associated limiting scattering mechanisms. The less well known parameters (e.g. N_{bkg}, Δ and Λ) can then be used to adjust this fit, refining the values for those parameters.

Figure 4.10 shows the above transport model, including a non-parabolic effective mass, applied to a region 1 sample with a 3 K carrier density of $n_{2D} = 2.14 \times 10^{11}$ cm^{-2} and 3 K mobility of $\mu = 195,000$ cm^2/Vs (sample SF1056, see Table 3.3). This model uses both reasonable values for background impurities and interface roughness, and using more extreme values to match measured data. For both reasonable and extreme parameters, at high temperatures, phonon scattering is dominant, where this has a strong temperature

dependence, leading to a sharp increase in mobility with decreasing temperature. As temperature reduces below ~77 K, the effects of phonons on limiting the mobility is reduced, with temperature independent scattering mechanisms beginning to dominate. These lead to a plateau in mobility, characteristic of 2D systems, and shows that room-temperature mobility values are not necessarily indicative of the largest cold temperature values achievable.

Table 4.1 Parameters used for scattering mechanisms considered in the transport lifetime scattering model (including acoustic and optical phonons, remote and background impurities and interface roughness). Values are given for InSb as well as GaAs and InAs for comparison. Deformation potentials have been reported for InSb ranging from 7.2 eV to 30 eV, so a value of 20 eV has been used as the approximate middle of this range. Values from references [11, 12, 17, 69, 107, 108].

Property		InSb	GaAs	InAs
Electron effective mass m^*	(m_e)	0.014	0.067	0.023
Deformation potential Ξ	(eV)	20	12	9.5
Lattice Constant a	(Å)	6.479	5.653	6.06
Optical phonon energy $\hbar\omega_0$	(meV)	25	35	30
High frequency dielectric constant ε_∞	(ε_0)	15.7	10.89	12.3
Static dielectric constant ε_s	(ε_0)	16.85	12.9	15.15
Thomas-Fermi wavevector q_{TF}	(nm^{-1})	0.029	0.192	0.057
Effective Bohr radius a_B^*	(nm)	68.4	10.4	35.0
Longitudinal velocity of sound v_s	(m/s)	3700	4730	3830
Crystal density ρ_d	(kg/m^3)	5790	5320	5680

Figure 4.10a uses the values given in Table 4.1, as well as a dopant level determined from Schrödinger-Poisson modelling (section 4.3) and typical values for N_{bkg}, Δ and Λ of 5×10^{14} cm^{-3}, 3 ML and 20 nm respectively. The predicted mobility from the transport model is ~600,000 cm^2/Vs, 3 times the measured value. Figure 4.10b shows the transport model predictions matching well to the data, however to achieve this, parameters of N_{bkg}, Δ and Λ of 1×10^{16} cm^{-3}, 6 ML and 20 nm respectively were required. This roughness is significantly higher than would be expected from MBE grown material, and roughness of this level has not been observed previously in TEM cross-sections of InSb QWs. Also, this density for N_{bkg} is significantly higher than the measured value of $< 5 \times 10^{14}$ cm^{-3}, and at this level, S.P. modelling shows that the level of doping required to compensate causes the dopant plane to become occupied at a carrier density significantly less than that observed.

(a) (b)
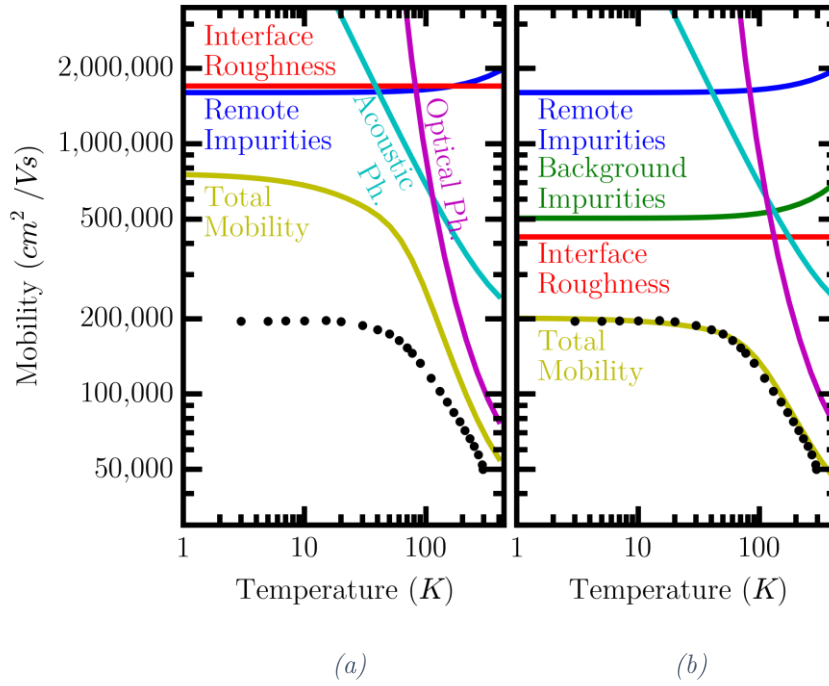
Figure 4.10 Transport modelled mobility as a function of temperature including standard scattering mechanisms including acoustic and optical phonons, background and remote ionised impurities and interface roughness (lines) as well as data extracted from a 2 carrier fit from a sample with $n_{2D} = 2.14 \times 10^{11}$ cm^{-2} (SF1056, points). Using reasonable values (a) of $N_{bkg} = 5 \times 10^{14}$ cm^{-3}, $\Delta = 3$ ML and $\Lambda = 20$ nm, the total mobility predicted is far greater than that measured. The measured mobility can be matched (b) though this requires more extreme values of N_{bkg}, Δ and Λ of 1×10^{16} cm^{-3}, 6 ML and 20 nm respectively. This indicates an unaccounted for scattering mechanism must be included.

This analysis shows that using standard scattering mechanisms, the measured mobility values are significantly lower than those predicted, and whilst these models have been applied previously, to match the measured data requires use of extreme values. Using refined values of these parameters, including through further modelling, it is clear that there is another, unaccounted for, scattering mechanism that must be invoked to explain this discrepancy.

This statement is also true when considering the ratio of transport to quantum lifetimes (τ_{tr}/τ_q). The quantum lifetime can be extracted from the amplitude of the low field Shubnikov–de Haas oscillations through analysis of "Dingle plots" [121, 122], where this has been performed by Hayes et al. [29] for the samples studied in this work. This analysis resulted in an upper limit [123] quantum lifetime of $\tau_q \approx 0.11$ ps for a sample with a carrier density of $n_{2D} = 2.14 \times 10^{11}$ cm^{-2}. The corresponding transport lifetime, extracted from the mobility, measured via the Hall effect as described previously, was found to be $\tau_{tr} = 1.86$ ps. The Dingle plot reproduced from reference [29] is shown in Figure 4.11.
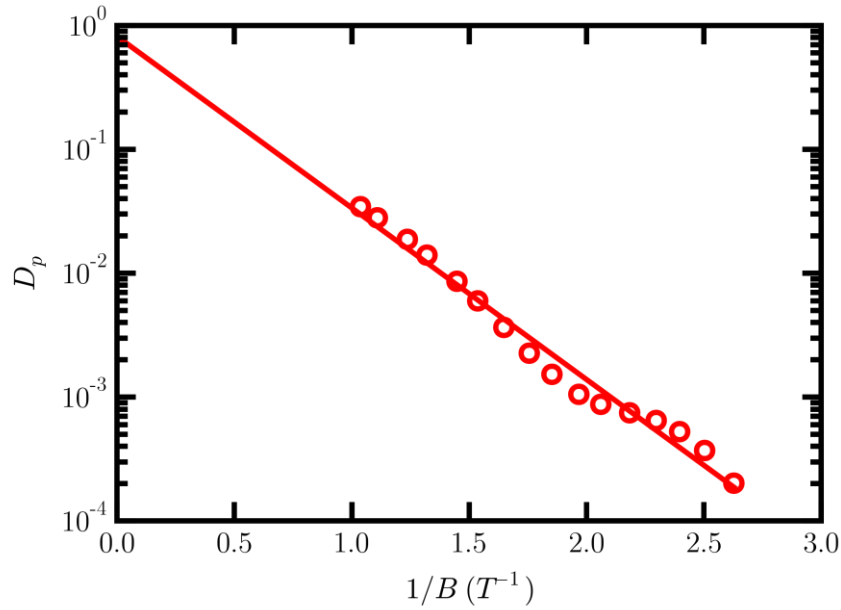
Figure 4.11 Example Dingle plot reproduced from reference [29]. Data points extracted from analysis of peak amplitudes of Shubnikov-de-Haas oscillations with a least squares fit to determine quantum lifetime. D_p is the "Dingle parameter" [122].

As the quantum and transport lifetimes were measured from the same sample, they can be directly compared, where using these extracted values, a ratio of $\tau_{tr}/\tau_q \approx 17$ is achieved. This is comparable to ratios extracted in GaAs and InGaAs systems (where $\tau_{tr}/\tau_q \approx 5$ to 20) [121, 122]. Often the dominant scattering mechanism in similar QW samples is considered to be remote ionized impurity scattering, however, using the analysis

of Das Sarma and Stern for screened ionized impurity scattering in a 2DEG [124], a ratio of $\tau_{tr}/\tau_q > 100$ is predicted. Reduced values of this lifetime ratio have been accounted for through additional scattering from background impurity scattering [121], however when considering the previous S.P. modelling, increased values of background impurities are inconsistent with the mobility and carrier density trends observed. Increasing the level of background impurities and interface roughness to the level required to achieve the limiting transport mobility (i.e. as in Figure 4.10b), the lifetime ratio achieved is ≈ 54, significantly greater than that measured. To achieve the correct ratio, the required values are so extreme as to predict a transport lifetime significantly less than that measured [29].

This inability to simultaneously explain the measured τ_{tr}/τ_q ratio whilst simultaneously explaining the measured mobility, even when using extreme and unreasonable values, indicates than a previously unaccounted for scattering mechanism is required.

It is also worth noting that in the Dingle plots analysed, there is a clear oscillation of the extracted data points about the best fit line, where this has previously been attributed to inter-subband scattering when observed in GaAs samples [125]. When considering the previous S.P. modelling, the second subband is separated by an energy $\gg k_B T$ and should not be populated, with the origin of this oscillation therefore remaining uncertain.

4.5 Surface Image Analysis (Nomarski Imaging)

To help examine the limiting factors affecting the higher mobility samples, as well as the discrepancy between the predicted mobilities from transport modelling, and those measured through Hall effect measurements, the surface morphology of the samples was considered using optical differential interference contrast DIC (Nomarski) imaging. This allows for direct imaging of a sample surface to visually inspect the material quality [85].

The technique of DIC (Nomarski) imaging relies on the phase interference produced due to the varying height of a sample surface, visualising features that would otherwise be too shallow to observe using a simple light microscope. This is achieved by first passing the light produced by a source through a polariser, and then directing this through a Wollaston prism (and angled wedge prism) and onto the sample surface. Small variations in height on the surface will cause small phase shifts between the incoming and reflected light when passed back through the prism and a second polariser, with this visible as surface features through the eyepiece or attached camera. This interference technique gives a large sensitivity to surface feature height variations, of the order of nm, but gives no quantitative height information. The illumination level, thickness of the prism, and

the objective lens magnification can all be used to adjust the apparent level of sensitivity to the surface feature height variations observed, but again, this is only qualitative. A schematic diagram of a Nomarski microscope is shown in Figure 4.12 [85].
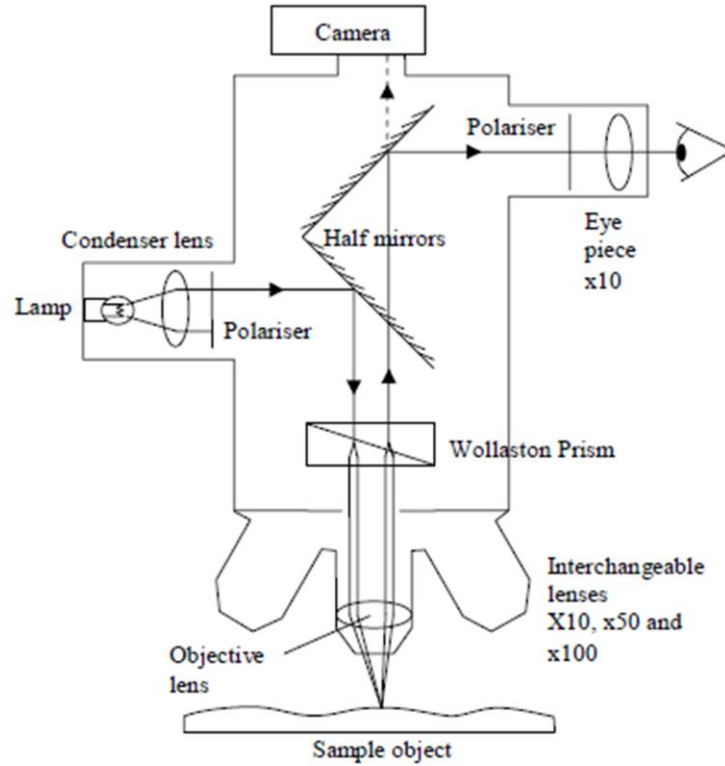
Figure 4.12 Schematic diagram showing the operation of a typical Nomarski microscope (reproduced from reference [85]). Light from a source is first polarised before being passed through an angled wedge (Wollaston) prism and an objective lens (typically × 50) before being shone on the sample. The reflected light is then passed back through this prism, where height variations on the sample surface give rise to interference. The light is then passed through an eye piece to the observer and to a digital camera used to record images of the sample.

To analyse the surfaces of the samples studied here using Nomarski imaging, multiple microscope images were taken of the various unprocessed wafers in a spread of positions across the surface. Images were taken at an optical magnification of × 50 as this was found to be the most appropriate magnification for subsequent image analysis, giving the best quality image whilst also a wide enough field to allow reasonable analysis. A raw Nomarski image of a standard wafer with $n_{2D} \sim 3 \times 10^{11}$ cm^{-2} and $\mu \sim 200,000$ cm^2/Vs is shown in Figure 4.13.
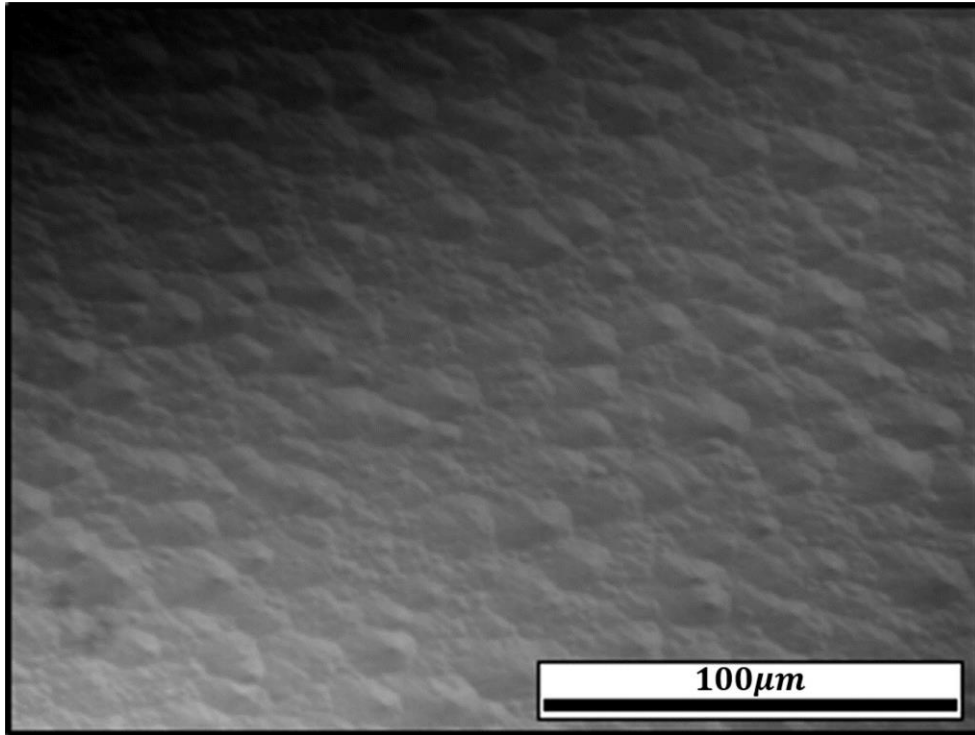
Figure 4.13 Optical Nomarski image of a typical wafer surface at a magnification of ×50. The surface is clearly textured, with approximately circular surface features consisting of various sizes, and with well-defined boundaries between features.

The Nomarski image shown in Figure 4.13 is typical of all samples imaged, showing a clear surface roughness present across the whole surface, consisting of approximately circular features with clear boundaries separating features. Due to the close proximity of the 2DEG to the surface (50 nm) it is reasonable to assume that this surface roughness, and in particular the boundaries, have a severe impact on the transport of electrons through the quantum well.

To analyse the Nomarski images, and so to extract an average feature size for each sample, multiple grey scale images were taken in a spread across the sample surface. These were then processed using a self-written GUI (graphical user interface) and analysis program. The processing steps required to extract an average feature size are described subsequently (expanding the description given in reference [30]), with a summary given in the flow chart in Figure 4.14. Figure 4.15 shows an example of the GUI being used to analyse a Nomarski image of the wafer used to produce the sample analysed in the previous section (sample SF1056), showing the full image taken and a zoomed in region near the centre. The processed images for select steps are shown subsequently in Figure 4.16.
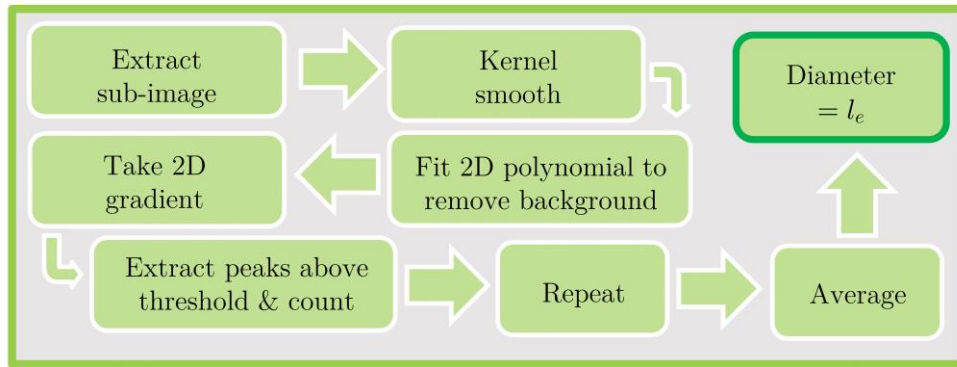
Figure 4.14 Flow chart describing the steps required to analyse Nomarski images of sample surfaces, and from these images extract an average feature size (diameter), l_e. A more full description of each step is given in the text.
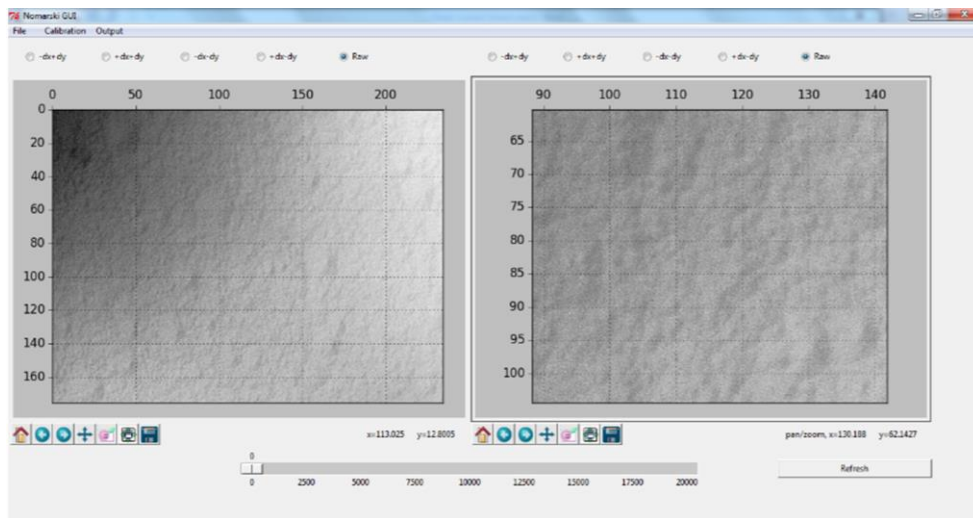
Figure 4.15 Graphical user interface showing an image of wafer SF1056, as analysed in the transport model in Figure 4.10. The image on the left consists of the raw, full size image, whilst the figure on the right shows a magnified portion of the image (a sub-image). The radio buttons (tick boxes) above each image allow the 2D gradient to be taken in the 4 possible directions, whilst the slider underneath the images allows for the threshold to be varied. With desired settings selected, the Refresh button re-processes the image and displays this overlaid on the original, allowing comparison of highlighted peaks to the raw image. Once the correct parameters are found, the full image can be analysed and the average feature size and error output. A more detailed description of each step is given in the text.

The initial step of analysis is to obtain multiple grey scale images of each wafer, allowing the averaging of results from each image and ensuring the feature size obtained is representative of the whole wafer. For each image, a small square sub-image is then extracted from the original image using a window typically 200 pixels by 200 pixels, ensuring each sub-image contains many features. To remove noise from the image which may give false positives during later analysis steps, each sub-image is kernel smoothed

before a 2D polynomial is fitted to remove any background trend (blush) in the image. The background is calculated using a 2D least squares fit of the function

$$f(x,y) = a + bx^2 + cx + dxy + ey + fy^2,$$

(4.44)

where a, b, c, d, e and f are constants. Subtracting the fitted $f(x,y)$ from the image then results in a uniformly illuminated image, with the variations in brightness then corresponding only to the variations in height due to the features (the peaks and boundaries).

To determine the number of these features, and so the average feature size, a 2D gradient is calculated by first determining gradients in the x- and y-directions, and depending on the original direction of illumination, calculating the sum of or difference in these gradients. The 2D gradient then either accentuates the boundaries between features, or the peaks of the features, where the latter is desirable to distinguish between separate features, and so allowing extraction of an average feature size.

A variable threshold is then applied so as to match the regions highlighted by the 2D gradient to the original image, selecting and labelling features above the threshold and rejecting those below. The threshold is adjusted until there is one labelled region per surface feature, with the total number of labels giving the total number of features. This analysis is then repeated for multiple windows on each image (typically 40 sub images), covering the whole image, with results combined to give a mean feature count, with the error given by the standard deviation.

Figure 4.16 shows an example of an extracted sub image alongside labelled regions for a given threshold, where this enables matching a labelled region with a peak. The threshold can then be varied to better match this labelling to the corresponding peaks.

This process is then repeated for all images taken of each wafer, and corresponding feature counts averaged, with errors combined in quadrature. If there are N images of a wafer, with the average number of features in an image given by $n \pm \sigma_n$, then the average number of features for the wafer, $f_N \pm \delta_N$, is given by

$$f_N = \frac{1}{N} \sum_N n \;\; ; \;\; \delta_N = \frac{1}{N} \sqrt{\sum_N (\sigma_n^2)}.$$

(4.45)
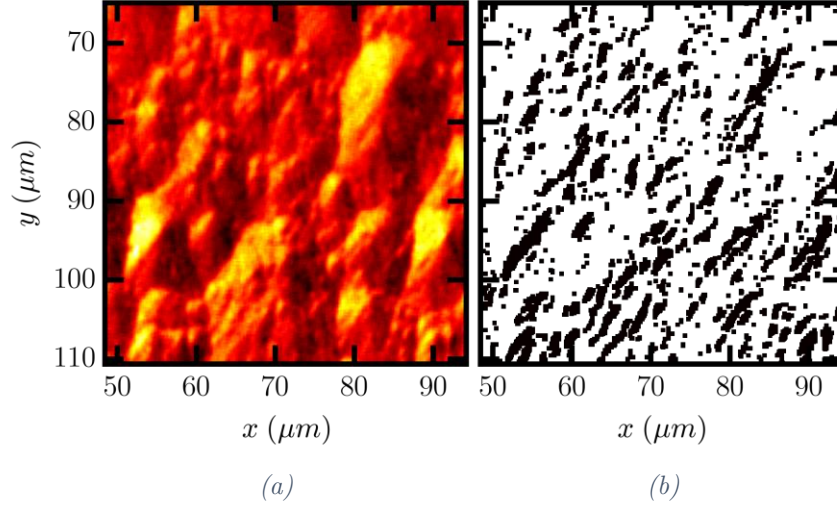
Figure 4.16 (a) Extracted sub-image of a Nomarski image of wafer SF1056 (as shown in Figure 4.15), and (b) processed image showing the labelled peaks for a given threshold and gradient direction. By matching the labelled peaks to the original sub-image, the threshold can be adjusted, or the gradient direction altered, to give one labelled region per peak (or a close as possible).

To turn the average feature count into an average feature size, it is assumed that the features fill all space, as can be seen through the whole image being textured and rough in Figure 4.13 and again in Figure 4.15. Therefore, the average area of each feature is given by the area of the sub-image divided by the mean feature count. To transform from an area into a diameter, the length the electrons will traverse when within a feature, the features are assumed to be approximately hexagonal, where the edge-to-edge diameter of a regular hexagon, l_e, is related to its area by

$$l_e = \sqrt{\frac{2}{\sqrt{3}}\frac{A}{f_N}} = \sqrt{\frac{2}{\sqrt{3}}A_f}, \qquad (4.46)$$

where A is the area of the image containing f_N features, with each feature having an average area A_f.

To analyse the effect of this average surface feature size on the electronic transport, and determine if this is the extra scattering mechanism required, the measured mobility can be converted into a ballistic length λ (the mean free path), using a simple Drude model, such that [17]

$$\lambda = \left(\frac{h}{e\sqrt{2\pi}}\right)\mu\sqrt{n_{2D}}. \qquad (4.47)$$

Figure 4.17 shows a comparison of the ratio of the mean feature diameter calculated from Nomarski imaging analysis to the corresponding mean free path length deduced from this basic Drude model. The mean free paths are calculated using mobilities from the standard $3\ K$ Hall measurements of samples shown in Figure 4.1, as well as from historical samples from references [16, 17, 75, 76].
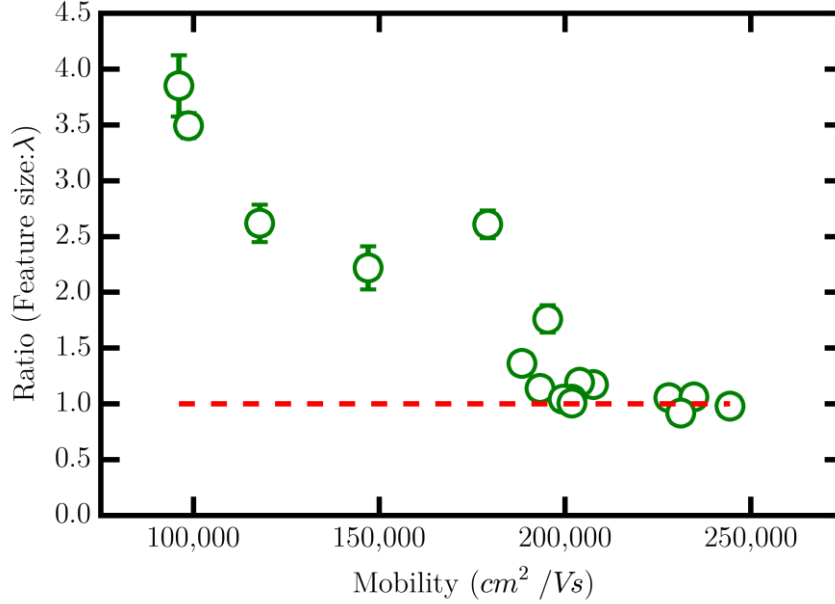
Figure 4.17 Ratio of average surface feature size extracted from Nomarski imaging (with one standard deviation error determined from size distributions given by multiple image sampling) to mean free path (λ) measured from low temperature Hall effect measurements, as a function of mobility. At larger mobilities, the ratio approaches a limiting value of ≈ 1 (red dashed line), showing the relationship between feature size and upper mobility (and mean free path) limiting scattering.

Figure 4.17 shows there is a clear trend in the ratio of feature size to λ as a function of mobility, approaching a value of 1 for the highest mobility samples. The measured mean free paths reach a limit of ~2.5 μm for the highest mobility samples, whilst the average feature size determined from Nomarski imaging is $2.43 \pm 0.13\ \mu m$. This is in excellent agreement with the maximum mean free paths, with this limiting ratio of 1 strongly suggesting that these features relate to the low temperature transport lifetime limiting scattering mechanism.

To include the effect of these surface features in the transport model, a first approximation can be made via a rearrangement of equation (4.47), assuming the feature diameters are a ballistic length, then the resultant scattering rate is given by

$$\frac{1}{\tau_l} = \left(\frac{h}{m^*\sqrt{2\pi}}\right)\left(\frac{l_e}{\sqrt{n_{2D}}}\right)^{-1}, \tag{4.48}$$

where this can then be included via Mathieson's rule.

Figure 4.18 shows the transport model with the effect of these features included, using a value of $l_e = 2.43 \ \mu m$, and using the more realistic values of $N_{bkg} = 5 \times 10^{14} \ cm^{-3}$, $\Delta = 3$ ML and $\Lambda = 20 \ nm$ respectively. The predicted mobility now matches closely to the measured values, confirming the statement that at low temperatures, when phonon effects are reduced, an electron travelling in the quantum well may travel ballistically through a feature until it reaches a boundary where it scatters, limiting the transport lifetime. Further advancement in mobility and mean free path without significant buffer redesign will be impossible.
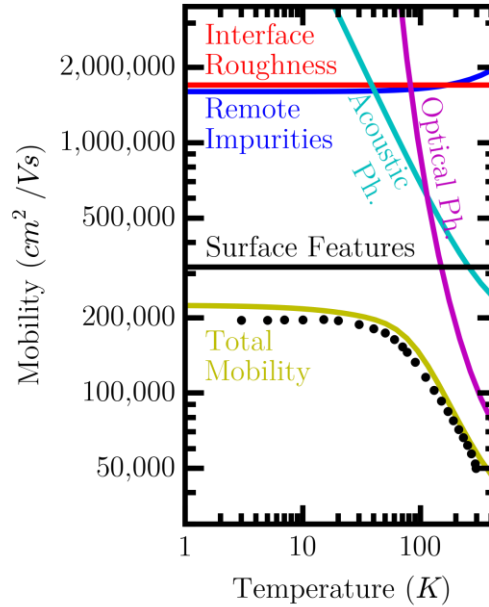
Figure 4.18 As Figure 4.10, transport modelled mobility as a function of temperature for acoustic and optical phonons, background and remote ionised impurities and interface roughness, using reasonable values of $N_{bkg} = 5 \times 10^{14} \ cm^{-3}$, $\Delta = 3 \ ML$ and $\Lambda = 20 \ nm$. Also included is scattering related to 2.43 μm surface features following equation (4.48), with this total mobility giving excellent agreement with data extracted from a 2 carrier fit from a sample with $n_{2D} = 2.14 \times 10^{11} \ cm^{-2}$ (SF1056, points). Incorporation of this scattering mechanism shows this is the unaccounted for scattering mechanism required.

Figure 4.18 shows that scattering related to these surface features can be used to explain the mobility limit for lower carrier density, region 1 samples. Repeating the above analysis for a higher carrier density region 2 sample (sample SF1055,

$n_{2D} = 2.89 \times 10^{11} \ cm^{-2}$) then allows for the verification of the significance of this scattering mechanism for samples spanning a range of carrier densities. Figure 4.19 then shows the results of this transport model with the same reasonable values ($N_{bkg} = 5 \times 10^{14} \ cm^{-3}$, $\Delta = 3$ ML and $\Lambda = 20 \ nm$) and same average feature size. Now, the predicted mobility is again very close to that measured, though slightly lower. This shows that these features have a significant effect across a range of carrier densities, and the parameters of the limiting scattering mechanisms are also approximately consistent for samples within the same batch. Again further advancements in mobility will require a buffer redesign to increase this mean feature size.
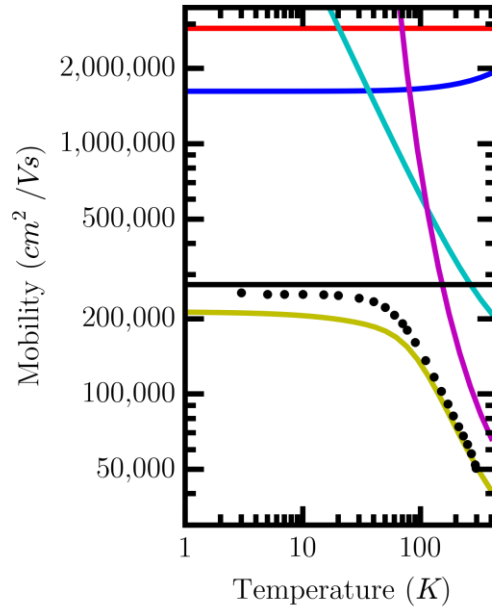
Figure 4.19 As Figure 4.18, transport modelled mobility using reasonable parameters and surface feature scattering for an increased carrier density sample ($n_{2D} = 2.89 \times 10^{11} \ cm^{-2}$, SF1055, points). The total predicted mobility is marginally lower than that measured, though still in very close agreement, showing that this surface feature related scattering for ~2.43 μm features is the upper mobility limiting scattering mechanism across a range of samples.

4.5.1 Origin of surface feature hillocks

These surface hillocks have historically been observed on GaAs (e.g. in reference [126] (1984) and reference [127] (2001)) and InSb (e.g. reference [128] (1985)), where the size and distribution of these hillocks was linked to the growth temperatures and pressures, and in particular, the group V flux onto the surface. The distribution of hillocks observed in InSb in reference [128] were typically a few thousand Angstroms in height and several microns in lateral dimension, though these appeared as triangular or hexagonal structures

with a flat top protruding from the surface. The hillocks observed in GaAs however where more pyramidal in shape, as observed in Nomarski images of our material, and had lateral dimensions of ~2.5 μm.

More recently, interest in InSb has increased due to its unique and extreme properties amongst III-V semiconductors, making it the ideal candidate for the search for Majorana fermions [15, 18]. Due to this interest, several groups have recently published work on improving growth of InSb with the aim of reducing threading dislocation densities (TDDs) and hillock densities (HDs), using these as proxies for material quality. The hillock density has also been correlated with mobility observed in bulk InSb grown on superlattice buffers [99, 100, 101, 129]. Though there has been work performed aimed at reducing hillock densities in InSb, only correlations have been made with material and electrical quality, as opposed to a direct link as shown in section 4.5.

These hillocks are believed to form as a result of screw like growth forming around a threading dislocation propagating through the structure to the surface [130]. Threading dislocations that do not form hillocks then emerge as pinned interfacial steps on the stepped surfaces of these screw like hillocks. As the growth around an individual screw dislocation is relatively uniform, the crystal structure within a hillock will be regular, however this is not necessarily true for the boundary between two adjacent hillocks, where the screw edges may not align exactly. This leads to a crystal deformation at the boundary, and acts as a scattering centre surrounding the hillock. This is illustrated schematically in Figure 4.20 [86].

With these boundaries then acting as scattering centres, large improvements in mobility cannot reasonably be achieved without increasing the size of these hillocks (and so reducing the threading dislocation (TDD) and hillock (HD) densities). Several methods have been employed to achieve this including graded or stepped buffers, interlayers and offcut substrates, and superlattice structures.

An example of the results from reference [100] for various numbers of interlayers on the TDD, HD, and the TDD:HD ratio is given in Table 4.2, alongside typical values determined for the samples studied in this work.

Table 4.2 shows that for the samples in reference [100], modifying the structure from a simple buffer with no interlayer (as in the samples studied in this work), to include a single or double interlayer, has a large effect on the density and size of the hillocks observed. However whilst the TDDs and HDs observed in reference [100] are comparable to those observed here, their samples did not include a QW or cap layer, and so cannot be compared electrically.
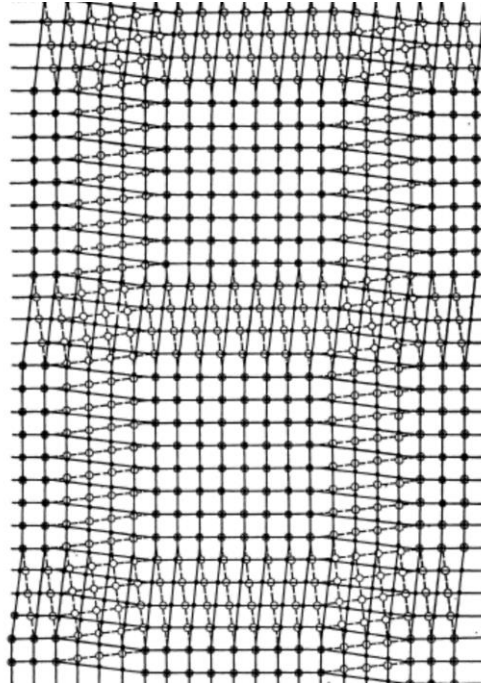
Figure 4.20 Simplified example of the crystal structure in the boundary between two regions of uniform material grown around two screw dislocations [86]. The two regions of uniform growth represent the surface features observed through Nomarski imaging, whilst at the boundary, the crystal becomes non-uniform, with this disruption acting as a scattering centre bounding each feature.

Table 4.2 Comparison between threading dislocation density (TDD), hillock density (HD), and hillock diameter (l_e), for samples from reference [100] with varying numbers of interlayers used to reduce defect density, and typical densities averaged across samples studied in this work. The average HDs and sizes in this work are comparable to those from reference [100] for 2 interlayers, though the for samples studied in this work, there are approximately twice as many threading dislocations per hillock as compared to those in reference [100].

Sample Source	Number of Interlayers	TDD ($\times 10^9\ cm^{-2}$)	HD ($\times 10^7\ cm^{-2}$)	l_e (μm)	TDD:HD
	0	2.6	120	0.3	2.2
Reference [100]	1	1.1	5	1.5	22
	2	0.25	1.7	2.6	15
This work (InSb QW)	0	0.75	1.95	2.4	38

4.6   Conclusions

Schrödinger-Poisson modelling and transport lifetime modelling has been performed with the aim of determining the transport limiting scattering mechanisms in the InSb QW samples studied.

The S.P. modelling investigated the level of dopant required to achieve measured carrier densities, accounting for dopant dragging in the top cap as determined through SIMS measurements. The effect of background charged impurities on the level of dopant required was investigated, with a background dopant density of $N_{bkg} \leq 5 \times 10^{14}\ cm^3$ consistent between electrical measurements of AlInSb layers, and the onset of 2 carrier nature in QW samples. With this relationship between dopant density and carrier density determined, the separation between the dopant plane and the peak of the wavefunction was investigated, again, with the relationship determined.

Following this a transport lifetime model was implemented, including the standard scattering mechanisms of acoustic and optical phonons, remote ionised and charged background impurities, and interface roughness. This model was then implemented to determine the limiting scattering mechanisms for a sample with a $3\ K$ carrier density of $n_{2D} = 2.14 \times 10^{11}\ cm^{-2}$ and mobility $\mu \approx 200,000\ cm^2/Vs$. It was shown that using reasonable values for N_{bkg} and the interface roughness, as well as values for remote ionised impurity scattering determined from S.P. modelling, the predicted mobility is significantly greater than that measured. To match the measured values, extreme values of roughness and background charge are required, with this incompatible with S.P. modelling and independent measurements. An unaccounted for scattering mechanisms is therefore required.

To explain this discrepancy, Nomarski imaging of wafer surfaces was performed, with these surfaces shown to have a characteristic roughness (hillocks). Analysis of these Nomarski images, involving smoothing, taking 2D gradients, and extracting peaks above a threshold allows for the determination of an average feature size, l_e. The average surface feature size extracted across multiple sample is $l_e = 2.43 \pm 0.13\ \mu m$, where comparison to extracted mean free paths gives a limiting ratio of $l_e/\lambda \approx 1$. Incorporation of this scattering mechanism into the transport model using a simple Drude model shows that the predicted mobility can be matched to that measured using reasonable values consistent with the previous modelling. This shows that scattering related to these surface features is the previously unaccounted for scattering mechanism.

5 : MONTE CARLO MODELLING AND MAGNETORESISTANCE

5.1 Introduction

It was shown in the previous chapter that scattering related to surface roughness (hillocks) observed through Nomarski imaging results in a clear impact on the mobility observed in the quantum well. Previous studies have correlated the apparent material quality (quantified via threading dislocation and hillock densities) with the electrical quality (quantified by the mobility) [99, 100, 101, 129], whereas transport modelling shows a direct numerical causal link between the two [30]. This simple modelling assumed that these surface features correspond to regions of uniform crystal potential where conduction occurs ballistically. Under this assumption, the boundaries between features act as hard scattering centres, limiting the mean free path and mobility. Figure 4.20 [86] showed these boundaries are regions of non-conformity of the crystal structure over a finite width, where this non-regular structure will give rise to a variation in potential, resulting in a barrier. An electron crossing a barrier will therefore have a finite probability of scattering, limiting the lifetime, the mobility, and the mean free path.

The following sections extend this idea further, treating these boundaries as potential barriers and analysing the current transport mechanisms, and associated mobilities, to determine a range of plausible barrier shapes [31]. The current transport is calculated using Landauer theory, whilst a Monte Carlo (MC) model is implemented in Python/Cython to determine the corresponding mobility. This is then incorporated into the transport model described in the previous chapter, enabling the determination of the individual mobility limiting scattering rates and the limiting scattering mechanisms. Variation is subsequently introduced into the model to better represent the non-uniformity in feature sizes as observed through Nomarski imaging (Figure 4.13), and the model is refined using more physically realistic, back-to-back Schottky-like barriers pinned in the band gap [93, 131, 132]. Following this, detailed longitudinal magnetoresistance measurements are performed, where these are shown to have features at length scales corresponding to macroscopically observed values for parameters such as surface feature size, background impurity density and threading dislocation density [133]. Combining the results of these physical measurements with the barrier shapes and sizes determined from the Monte Carlo modelling allows for the determination of the limiting scattering rates

and mechanisms across the full range of single carrier samples shown in Figure 4.1 (i.e. regions 1 and 2). Scattering rates in multicarrier samples are not considered here.

5.2 Determination of Potential Barrier Heights

To determine the range of possible potential barrier shapes, the model implemented must match experimentally measured values (boundary conditions) for a sample, where these are the experimental current and voltage. Matching this, the model must then give a reasonable limiting mobility as given by the transport lifetime modelling performed previously. For measurement, a set current is applied and the corresponding voltage is measured, however for this analysis it is assumed that a known voltage is applied to a standard 200 μm length Hall bar and the resultant current is measured. Although this is physically equivalent, this allows for easier determination of potential barrier heights. Standard two-carrier fitting is performed as described previously [106], and a 2D carrier density and electron mobility are determined (resulting in the values given in Table 3.3). For the sample analysed in the transport model previously (sample SF1056, Figure 4.18), the relevant parameters are given in Table 5.1. This table also contains the corresponding values for sample SF1055 for use as a check of the model.

Table 5.1 Values and symbols for parameters included in the following Landauer and Monte Carlo models for the two samples used for varification of the model. These samples were those analysed in section 4.5.

Parameter	Symbol	SF1056 Value	SF1055 Value
Number of barriers	N	82	82
Carrier density	n_{2D}	$2.14 \times 10^{11}\ cm^{-2}$	$2.89 \times 10^{11}\ cm^{-2}$
Mean surface feature size	l_e	2.43 μm	2.43 μm
Experimental current	I	1 μA	1 μA
Experimental voltage	V_0	0.75 mV	0.4 mV

The method of determining the range of possible barrier heights is summarised in the flowchart in Figure 5.1 for a series of uniform barriers. Initially, for an arbitrary barrier width, a barrier height is chosen, and the transmission probability as a function of energy for this single barrier is determined using the method described subsequently in section 5.3. Following this, the current as a function of voltage is determined using the Landauer formalism and compared to the values given in Table 5.1. If the current value is too low or too high, the barrier height is adjusted accordingly until the correct current is achieved. Following this, a MC model is used to determine the corresponding mobility (section 5.4), with this again compared to Table 5.1. The effective transmission across the multi-barrier

system ($T^*(E)$) is also determined from this MC model, with this used to calculate an updated current value. If this new current and mobility match the values in Table 5.1, the process is stopped and the barrier height is determined, otherwise the barrier height is adjusted and the process repeated. This process is then repeated for a set of barrier widths to determine a range of possible barrier width-height combinations.
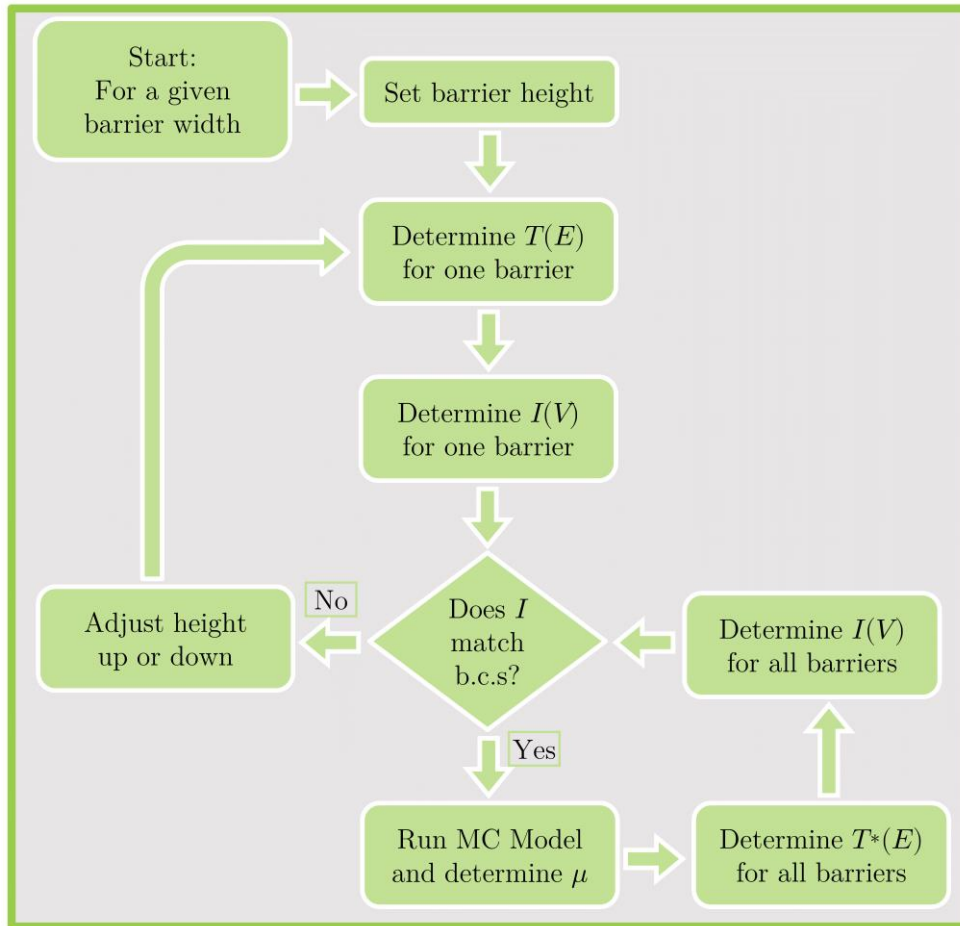
Figure 5.1 Flowchart describing the method of determining the range of possible barrier heights via determining a Landauer tunnelling current and MC modelled mobility for a given potential barrier width. For a single barrier with a given width, an initial height is set, and the transmission probability and the tunnelling current calculated. This is compared to the boundary condition (b.c.) value given in Table 5.1, and if the values do not agree, the barrier height is adjusted and the process repeated. Following this, a MC model is used to determine a corresponding mobility through all of the barriers, again compared to Table 5.1. An effective transmission ($T^*(E)$) is also determined from this MC model, with this used to calculate an updated current value. If the mobility and current match the b.c.s, the process is repeated for a new width, otherwise the height is adjusted and the process iterated.

5.3 Landauer Tunnelling Current

To determine the net tunnelling current through a potential barrier at the boundary of a feature, the difference between the electrons tunnelling in each direction across the barrier must be determined and integrated over energy. As conduction is assumed ballistic between barriers, then the voltage drop across the length of the sample occurs only across the barriers, where this gives rise to a chemical potential difference across each barrier, and subsequently, a current. This idea of finite probability for an electron to tunnel through a barrier, resulting in a net flow across a barrier, is the basis of Landauer theory.

The tunnelling current density, J, is then determined by integrating over the Fermi distributions either side of a barrier and multiplying by the energy dependent transmission probability $T(E)$. For a 2D sheet of charge (xy) tunnelling in the perpendicular z-direction, this is given by (amongst others) Davies [63] and Tsu and Esaki [134]. However, in this case, the 2D sheet of charge is considered to be tunnelling in the plane (in the y-direction). In this orientation, the tunnelling current for an applied voltage, V, at a temperature, T, is given by

$$J = \frac{2e}{h} \int_0^\infty [n_{1D}(E_F - E_y) - n_{1D}(E_F - E_y - eV)]T(E_y)dE_y, \qquad (5.1)$$

$$n_{1D}(E_x) = \frac{\sqrt{2m^*}}{h} \int_0^\infty (\sqrt{E_x})^{-1} \left[1 + \exp\left(\frac{E_x - E_F}{k_B T}\right)\right]^{-1} dE_x. \qquad (5.2)$$

Here, E_F is the Fermi energy and m^* the effective mass, both calculated from n_{2D} using the non-parabolic Kane model as stated previously. To obtain the current, I, the current density must be multiplied by the device width (40 μm).

The tunnelling current given by equation (5.1) is then dependant on the transmission probability, where this is dependent on the choice of barrier. For a few select barrier shapes the transmission probability can be determined analytically, however, for the majority of barrier shapes, this must be determined numerically. This can be achieved following the method of Ando and Itoh [135], whereby a barrier is broken up into a number (n) of smaller rectangular slices. In each slice, the potential is regarded as constant and there is a finite probability of transmission into the next slice. This reduces the problem of determining the overall transmission probability through a single barrier to determining the result of $n + 1$ matrix multiplications. In the limit of an infinite number of slices, the continuously varying potential is recovered, and the true transmission determined. In practice however, when n becomes too large (i.e. if attempting to calculate the energy

dependant transmission directly across the full N-barrier system), this system requires a very fine energy resolution, and the results can become unstable. In this scenario, a small change in an input value (i.e. energy, barrier height, etc.) can lead to a very large (unphysical) change in the associated transmission probability. Therefore, this method has only been applied to determine $T(E)$ for individual barriers, and the MC model described subsequently has then been used to determine the effective $T^*(E)$ across the full N-barrier system.

Following Tsu and Esaki [134], to determine the current due to a voltage drop applied over several barriers, the integral in equation (5.1) is performed across the whole system, with an effective transmission probability for the full N-barrier system. As a first approximation to determine the tunnelling current, it is initially assumed that the voltage applied is dropped across a single barrier, and $T(E_y)$ is determined for this barrier, with this representative of the N-barrier system. This is then used to determine a range of possible barrier shapes that match the boundary condition of the experimental current. These barrier shapes are then input into the Monte Carlo model (described subsequently), where the output from this MC model can be used to give a new effective transmission probability $T^*(E_y)$. Substituting $T^*(E_y)$ into equation (5.1) then gives an updated current corresponding to the voltage dropped across N-barriers, however due to the large numbers of barriers and small voltages considered, $T(E_y) \approx T^*(E_y)$, and the change in predicted current is small.

5.3.1 Rectangular barriers

Initially, the barriers considered for these models consisted of regularly spaced simple rectangular barriers (a schematic diagram of these barriers is given later in Figure 5.6). Barrier heights were then determined for barrier widths from 1 to 50 monolayers (MLs) by calculating $T(E_y)$ and J using equation (5.1), and matching this to the value in Table 5.1. The results for the barrier heights for each width matching this criteria are shown in Figure 5.2 for the two samples considered.
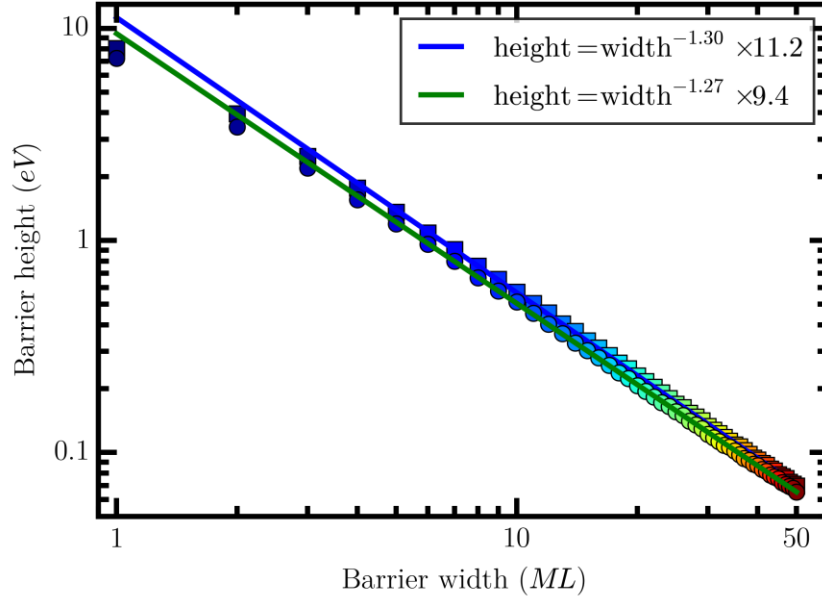
Figure 5.2 Logarithmic plot of barrier height required to achieve the correct tunnelling current for the applied voltage given in Table 5.1, for range of barrier widths given in monolayers, for samples SF1056 ($n_{2D} = 2.14 \times 10^{11}$ cm^{-2}, squares) and SF1055 ($n_{2D} = 2.89 \times 10^{11}$ cm^{-2}, circles). Linear least squares fits are shown for sample SF1056 (blue) and SF1055 (green), with the best fit equation for each given in the legend.

Figure 5.2 shows a clear correlation between the barrier widths and corresponding height for each sample, as expected. Using a least squares linear fit to the data shown in Figure 5.2, the barrier heights (in eV) for both samples were found to approximately follow the form height $\approx$ width$^{-1.3} \times 10$, where width is measured in MLs. These barrier heights and widths are then included in a Monte Carlo model to determine the corresponding mobility, with this included in the previous transport model to determine the overall limiting scattering rate.

5.4 Monte Carlo Model Mobility (Rectangular Barriers)

Whilst the tunnelling current was determined via a mixture of numerical and analytical techniques, a different model is needed to determine the mobility through this multi-barrier system. This is achieved using a mixture of a many particle Monte Carlo (MC) model and a simple Drude model, with these discussed subsequently.

In this Monte Carlo model, the electrons are treated as non-interacting, and travel at a constant velocity v. This "billiard ball" model only treats the scattering due to the barriers at the feature boundaries, treating the electron transport within a feature as

ballistic. This scattering rate will then subsequently be combined with the other scattering mechanisms described previously to give a modified transport model.

It is assumed for the MC model that electrons are travelling from left to right and that the left contact is a perfect emitter or source (and perfect reflector for any particles impinging on the contact), whilst the right contact is a perfect acceptor (sink) [136]. The velocity of each electron is dependent on its energy, E, such that

$$v = \sqrt{\frac{2E}{m^*}},$$ (5.3)

and the number of electrons at each energy is given by the Fermi distribution. Initially variation between barriers is neglected (i.e. each barrier is treated as identical), and the voltage drop across each barrier is the total voltage applied divided by the number of barriers (i.e. V_0/N). As the voltage applied is small, the effect on barrier height is negligible. The spacing between each barrier is given by the average determined from Nomarski imaging ($l_e = 2.43 \ \mu m$).

5.4.1 Mobility

To determine the mobility, a large number of particles (electrons) are input into the system (with energies following the Fermi distribution) and the system is evolved. When an electron reaches a barrier, a random number between 0 and 1 is generated, and if this number is larger than $T(E_y)$ at the energy of the electron, the electron is reflected, if it is smaller it is transmitted. If an electron is transmitted, no collision occurs and there is no change in energy, whereas a reflection is treated as a scattering event. When an electron is scattered it becomes re-thermalised, attaining the energy it entered the system with, giving a uniform thermal electron distribution across the sample.

The mobility can then be calculated using a simple Drude model, where the average force on an electron is determined by the difference between the electric field force and the average change in momentum (due to the change of velocity induced by a scattering event), divided by the average time between scattering events. At steady state, when the net force is zero and the momentum gained by an electron before scattering is equal to the momentum lost in a collision, the drift velocity, v, can be calculated. The mobility is then defined as the ratio of the average drift velocity to the electric field. Therefore, assuming each collision imparts an average momentum change m^*v, the mobility in the MC model at a given energy can then be calculated as an average over particles, such that

$$\mu = \frac{e\tau}{m^*} = \frac{e}{m^*}\overline{\left(\frac{t}{N_c}\right)} = \frac{e}{m^*}\frac{l_e}{v} \times \overline{\left(\frac{N_s}{N_c}\right)}. \qquad (5.4)$$

In equation (5.4), t is the total time taken for a particle to travel from contact to contact whilst τ is the average time between collisions, N_c and N_s are the number of collisions and steps respectively. The average time between collisions for an individual particle (an individual MC trial) is given by the total time taken to traverse from left to right, divided by the corresponding number of collisions, N_c. To determine the overall mobility, this is repeated for many particles (trials) and averaged, where the number of trials is typically on the order of several thousand or higher. As electrons travel ballistically between barriers, the time taken to travel this distance (l_e) is given by l_e/v. The total time taken (t) is then given by l_e/v multiplied by the total number of times this distance was traversed (i.e. the number of steps, N_s).

At finite temperatures, conduction doesn't occur at a given individual energy but over a spread of energies proportional to the rate of change of the Fermi distribution, f. Therefore, an energy average of mobility must be performed, accounting for a temperature dependent band gap, Fermi energy, and effective mass at each energy [137].

This is the basis of the Monte Carlo and Drude models used to calculate the mobility for a given rectangular barrier shape (height and width).

5.4.2 Transport model

The resultant mobilities from this MC model can then be combined with the standard scattering mechanisms described previously (remote ionised impurities, interface roughness, etc.), to give an overall modelled mobility. The limiting mobility for rectangular barriers of varying widths between 1 and 50 *MLs* for sample SF1056 is shown in Figure 5.3. This is a repeat of Figure 4.18, replacing the scattering from equation (4.48) with that from the MC model, and using the reasonable values for the other scattering mechanisms given there.
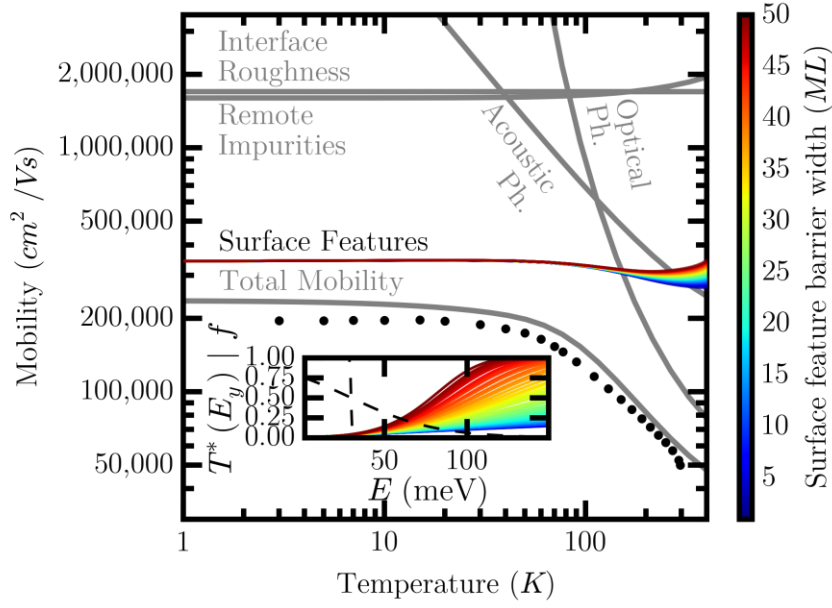
Figure 5.3 Reproduction of Figure 4.18 ($n_{2D} = 2.14 \times 10^{11}$ cm^{-2}), replacing the simple Drude model mobility given by equation (4.48) with the MC modelled mobility due to scattering from features with potential barriers from 1 ML (bottom) to 50 MLs (top). Barrier width is represented by colour. Inset: Transmission $T^*(E_y)$ vs energy (solid lines) for barriers from 1 ML (bottom) to 50 MLs (top), as well as Fermi distribution f vs energy for 3 K and 300 K (black, dashed). As previously, when surface feature related scattering from 2.43 µm features is included in the transport model, the predicted mobility is in excellent agreement with that measured.

As before, the total mobility closely matches the experimentally measured values, with the surface feature related scattering acting as the limiting scattering mechanism. Treating the boundaries between features as potential barriers allows the electron conduction to be treated as a series of tunnelling events, where reflections act as scattering events limiting the mobility.

The results of repeating this analysis for the higher carrier density sample SF1055, verifying this scattering across a range of carrier densities, are shown in Figure 5.4. Again there is a very close match to the measured mobility.
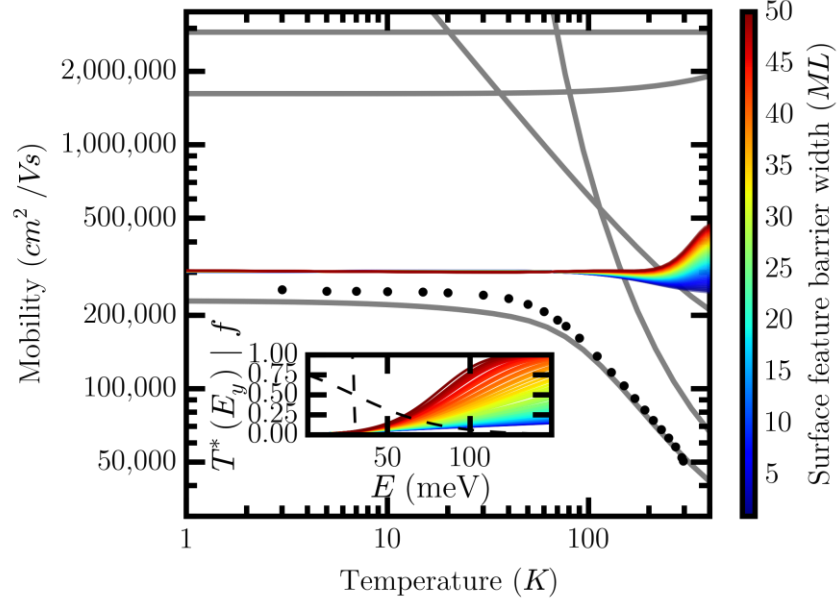
Figure 5.4 As Figure 5.3, reproduction of Figure 4.19 ($n_{2D} = 2.89 \times 10^{11}\ cm^{-2}$), replacing the simple Drude model mobility given by equation (4.48) with the MC modelled mobility due to scattering from features with potential barriers from 1 ML (bottom) to 50 MLs (top). Barrier width is represented by colour. Inset: Transmission $T^*(E_y)$ vs energy (solid lines) for barriers from 1 ML (bottom) to 50 MLs (top), as well as Fermi distribution f vs energy for 3 K and 300 K (black, dashed). As previously, when surface feature related scattering from 2.43 µm features is included in the transport model, the predicted mobility is in excellent agreement with that measured.

To investigate the dependence on voltage, the model was repeated for both samples with variations in applied voltage. As predicted, the voltage drop per barrier is negligible, and so the mobility is independent of voltage, however the tunnelling current does directly depend on voltage, as expected. The transmission probability is also extremely low, giving rise to very few successive tunnelling events, and ensuring any difference in energy between electrons that have successfully tunnelled and those that haven't is negligible.

Analysing the mobility as a function of both temperature and barrier width, both Figure 5.3 and Figure 5.4 show that the low temperature mobility is consistent for all barrier widths (due to the approximately constant value of $T^*(E_y)$ at $E = E_{\mathrm{F}}$). At higher temperatures the effect of barrier shape is more apparent, where the mobility trends for each barrier width begin to diverge due to the relation between $T^*(E_y)$ and the Fermi distribution. This divergence does not occur until the temperature is greater than ~70 K, however above this temperature, the transport model shows that phonon scattering is the limiting transport mechanism, and it is therefore not possible to precisely determine the shape of the potential barriers due to the features present on these samples.

These figures do show however that the low temperature limiting scattering can be modelled as a series of potential barriers, with a range of widths, separated by an average value obtained through Nomarski imaging. Conduction occurs ballistically within a feature (only scattering due to the other standard scattering mechanisms described previously), whilst at feature boundaries, electrons have a finite chance of tunnelling or scattering, limiting the mobility. This model gives reasonable thicknesses of between 1 ML and 50 MLs, consistent with possible screw-like growth around a threading dislocation. Thicknesses above this value give a mobility that begins to grow exponentially at higher temperatures, giving 50 ML as the reasonable upper barrier thickness range.

5.4.3 Variation (barrier height, width, and separation)

The above Monte Carlo modelling, whilst matching reasonably to experimental data, was limited in its scope. Only barriers of uniform width, height and separation were considered, however as is visible from Nomarski imaging (as seen in Figure 4.13), there is a large spread in feature sizes, and so therefore barrier separations. It is also therefore reasonable to assume there will be an equally large spread in the size of the potential barriers between features. This has subsequently been included in the MC model, initially varying only one parameter whilst keeping the others constant, to investigate the sensitivity of the mobility to each variation. This was then extended to included variations from multiple parameters.

Initially the separation between barriers was varied, with values following a Gaussian distribution with mean of 2.43 μm and standard deviations of up to 50% of the mean value, representing levels of variation. These randomly spaced barriers give no appreciable difference in calculated current or mobility, with the mobility dependent only on the mean separation. The mobility is linearly dependent on this mean separation, where for the 2.14×10^{11} cm^{-2} sample, this relationship is $\mu \approx 141{,}000 \times l_e$, where μ is measured in cm^2/Vs and l_e is measured in μm.

Subsequently, barrier heights were varied following a similar Gaussian spread about the values determined previously, i.e. for a given barrier width (kept constant for all barriers), the corresponding height was varied according to a Gaussian distribution. The current and mobility were calculated in a similar way to Figure 5.1, where for a given barrier width, the average barrier height across all barriers was that determined previously. However, when determining the mobility and effective transmission $T^*(E)$ across all barriers via the MC model, the corresponding transmission for each specific barrier width-height combination was used. The effective transmission from the MC model was then used to calculate an updated predicted current value with this matched to Table 5.1. This

process was then repeated for the set of barrier widths analysed previously (i.e. from 1 to 50 MLs). Again, this does not give any appreciable difference in predicted current values or mobilities as compared to the uniform case at low temperature, with only a marginal difference at higher temperatures.

A similar treatmeant was also performed for random barrier widths, however barrier widths are now Poisson distributed rather than Gaussian. In this model, the barrier width must always be an integer multiple number of monolayers, each barrier is assumed independent, and there cannot be a 0 monolayer thickness barrier.

Performing the same analysis as previously (fixing barrier height and separation), variations in barrier widths again give minimal change in resultant mobility. This largest effect on mobility variation is achieved for smaller barrier widths. Here, due to the restriction of a positive integer thickness barrier, the distribution of sizes is more strongly peaked on the average than at larger thicknesses, though this also gives rise to a skewed distribution. There are therefore more above average thickness barriers than thinner, with this having a detrimental effect on the mobility, though this effect is again minimal. At larger thicknesses, the distribution is more uniform and this effect is reduced. For all barrier thicknesses this mobility variation is small, and again can be considered to depend solely on the average value, with the corresponding height determined previously.

Finally, the effect on mobility of combining the variations in barrier heights, widths and separations was also investigated. A schematic example of the resulting barrier distribution from varying height and spacing by 20%, and using an averge barrier thickness of 25 MLs is shown in Figure 5.5 (the seperation between barriers has been reduced for clarity). The coresponding barrier transmission probability at the Fermi energy is also shown. Performing this analysis as previously shows that even when considering large variations of multiple parameters, the effect on mobility is small, only dependant on the mean value for each parameter.

This modelling has shown that variation of barrier heights, widths and separations can be introduced into the MC model described previously, and that this variation does not significantly affect the predicted mobility. Individual variations in heights, widths and separations of up to 50% produce no appreciable difference in the low temperature regime concerned, the mobility depending solely on the mean value for each parameter. This is also true when introducing variation from multiple parameters, whilst for variations in separation, the mobility is dependent on the average separation with the relationship $\mu \approx 141{,}000 \times l_e$ for the $n_{2D} = 2.14 \times 10^{11}\ cm^{-2}$ sample. This shows the validity of using the Monte Carlo model propsed previously, and expanded here, in describing the mobility due to surface features observed through Nomarski imaging.
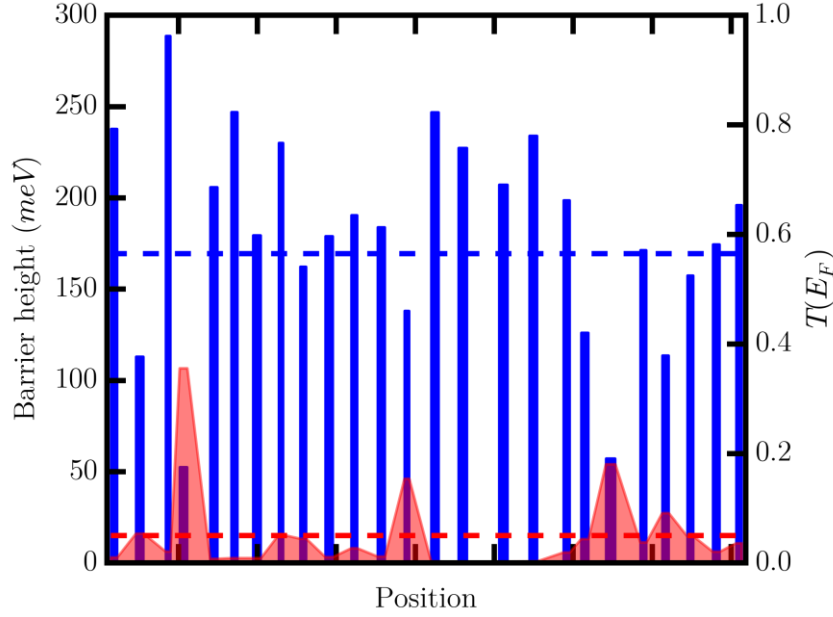
Figure 5.5 Example of randomly distributed barriers of varying widths and heights (blue). Barrier height is measured on the left axis, with the transmission at E_F ($T(E_F)$, red) measured on the right axis. Spacing between barriers is reduced by a factor of 100 for clarity. The mean barrier height and transmission are marked by the blue and red dashed lines respectively.

5.5 Refined Potential Barrier Model

To improve the MC model, various refinements have been implemented, including to the barrier shapes, incorporation of the 2D nature of system, and the derivation of an analytical mobility. The following sections describe these improvements, whilst Appendix B discusses the speed improvements and associated run-time reduction of the python code through the use of the Cython programming language.

5.5.1 Barrier shape (back-to-back Schottky-like barriers)

The previous model assumed surface features were separated by rectangular potential barriers, however due to the change in crystal structure at a feature boundary and possible trapped charge, it is more likely the interface will behave as a back-to-back Schottky-like barrier (similar to those seen at surfaces, e.g. Figure 4.3) [93, 131, 132]. It is also more likely that the barrier would, similar to the surface, be pinned in the band gap of the InSb, with the barrier height as measured from the conduction band edge in the quantum well given by $\Phi = E_g/2 + E_F$ if mid-gap pinned. Taking account of non-parabolic effects, this is typically on the order of ~150 meV. The width of the barrier is then governed by the

charge density of the 2DEG, with a width measured at the Fermi energy of l, given by [138, 139]

$$l = (\Phi - E_F)\frac{2\varepsilon}{e^2\pi n_{2D}}.$$

(5.5)

This is typically on the order of 120 nm. A schematic diagram of the barrier shapes used previously, as well as the refined Schottky-like barriers described here, is given in Figure 5.6.
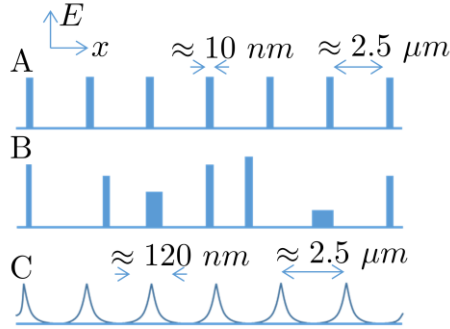
Figure 5.6 Schematic barriers for regular rectangular barriers (A), randomly distributed rectangular barriers (B) and back-to-back Schottky-like barriers (C), with typical barrier widths and separations given.

5.5.2 Incorporation of 2D nature into 1D model

The previous model equated the current flow through the material to many electrons following independent 1D paths, with an average length matching that of the Hall bar. This is in effect a 1D chain of resistors, as an individual electron would experience, however this neglects the overall 2D nature of the current flow measured in experiment. To calculate current flow between features in the model, it was assumed that all N barriers had an equal voltage drop, and therefore the same individual resistance of $1/N$ of the resistance of the sample. However, in an infinite 2D grid of identical resistors, the resistance measured between adjacent nodes (the equivalent of adjacent features) is half that of each individual resistor [140]. Therefore, as the model considers only the single resistor case, to equate the modelled current to the experimental value, the modelled current for a single resistor must be half the experimentally measured value.

As an example, if a voltage of 1 mV is applied to a sample, and a current of 1 μA measured, a resistance of 1 $k\Omega$ is determined. This is for the 2D grid of resistors, so each individual resistor in this grid will have a resistance of 2 $k\Omega$. Applying the same 1 mV to

this individual resistor would then give half the current, 0.5 μA. Therefore, to match the modelled current/voltage values to those measured, the modelled current must be half the experimental value. For all subsequent modelling, this half value has been used.

5.5.3 Analytic mobility derivation

At very low temperatures, an analytic approximation for mobility can be obtained by consideration of the relationship between N_s, N_c, and the transmission probability $T(E_F)$, simplified to T_{E_F}. This allows for the comparison of the MC model to theory, as well as to the simple Drude model used in the previous chapter, validating the assumptions made for this model. Verification of this analytical model would also allow for less computationally intense calculations to determine a predicted mobility given an average barrier shape and separation.

In any random walk, the relative proportion of steps where direction is reversed is given by the ratio of the number of reflection events N_R to the total (transmitted N_T, and reflected N_R),

$$\frac{N_R}{N_T + N_R} = \frac{(1 - T_{E_F})}{T_{E_F} + (1 - T_{E_F})} = (1 - T_{E_F}). \tag{5.6}$$

The total number of events, $N_T + N_R$, is simply the total number of steps N_s, whilst the number of reflection events is the number of collisions, N_c, giving the average number of collisions as a proportion of the total number of steps as

$$N_c = N_s(1 - T_{E_F}). \tag{5.7}$$

Substituting this into equation (5.4) gives the mobility at low temperatures as

$$\mu = \frac{e}{m^*} \frac{l_e}{v_F} \frac{1}{1 - T_{E_F}}, \tag{5.8}$$

where v_F is the Fermi velocity.

A convenient check for the validity of this expression can be performed by considering the average path length between scattering events $\lambda = l_e(N_s/N_c)$. This can be calculated by considering the probability of successfully tunnelling through n consecutive barriers, where $0 < n \leq N$, and N is the total number of barriers. Each successful tunnelling event

advances the particle a distance equal to the average spacing of l_e, giving the average path length as

$$\lambda = \left(\sum_{n=1}^{N} T_{E_F}^n \right) l_e + l_e, \qquad (5.9)$$

where the $+l_e$ accounts for the greater number of spaces between barriers compared to the number of barriers due to the left and right contacts. As T_{E_F} is sufficiently small, and N is sufficiently large, this can be approximated by an infinite sum as [67]

$$\lambda = l_e \frac{T_{E_F}}{1 - T_{E_F}} + l_e = \frac{l_e}{1 - T_{E_F}}. \qquad (5.10)$$

Substitution into equation (5.4) shows $\mu = e\lambda/m^*v_F$, and confirms the result reached in equation (5.8). A graphical comparison of this analytic derivation compared to the MC model solution is shown later in Figure 5.11.

At lower carrier densities, E_F and v_F are both reduced compared to higher carrier densities, with this giving a lower corresponding transmission probability. However, as T_{E_F} is already small, this changing value has minimal effect on the mobility, and equation (5.8) predicts that, at these lower carrier densities, the change in v_F has more of an effect on the mobility. This equation predicts that at lower carrier densities, the longer time taken to travel between surface feature boundaries will reduce the time between scattering events, and therefore increase the mobility. At higher carrier densities, where E_F, v_F and T_{E_F} are all increased, the increasing transmission has a larger effect, acting to increase the mobility, though the increase in Fermi velocity acts to oppose this, decreasing the scattering time. The overall mobility will be a combination of these two competing effects.

5.6 Detailed Magnetoresistance Measurements

5.6.1 Longitudinal magnetoresistance measurements

To attempt to verify the proposed Monte Carlo model, as well as investigate the scattering, detailed longitudinal magnetoresistance measurements were performed, analysing the variation in R_{xx} as a function of B-field. These measurements were performed using a Stanford Research Systems SR830 digital lock-in amplifier to measure the longitudinal V_{xx} as a function of magnetic field and temperature, using $0.02\ T$ magnetic field steps. A plot of the measured low temperature ($5\ K$ to $25\ K$) longitudinal R_{xx} resistance with B-field is shown in Figure 5.7, where colour represents temperature.

At lower temperatures and high magnetic fields, Shubnikov-de-Haas oscillations are observed.
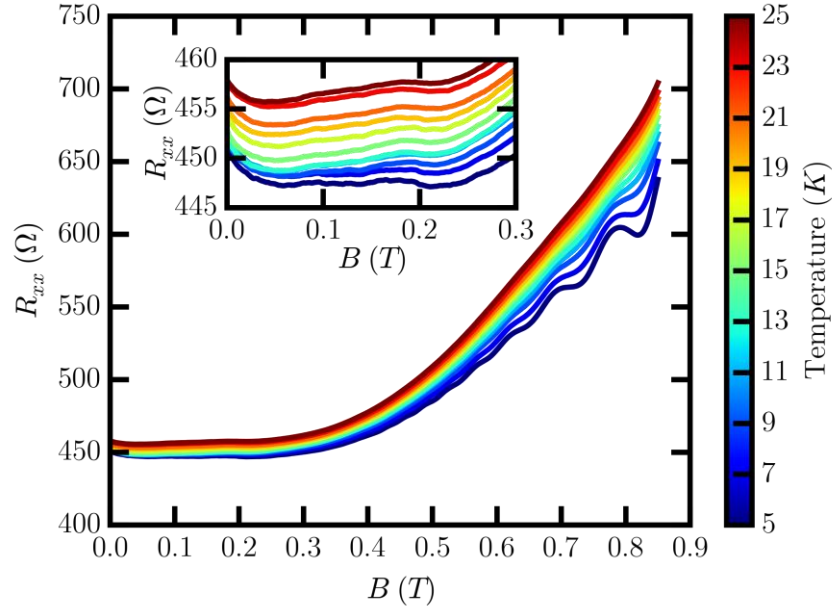
Figure 5.7 Longitudinal (R_{xx}) magnetoresistance measurements as a function of magnetic field (B) and temperature. The colour of each line represents temperature. Inset: Low field R_{xx} measurements showing local minima and oscillations analysed in this section.

Following Gilbertson et al. [133], scattering mechanisms with a scattering length comparable to the electronic cyclotron orbit size can manifest as local minima in the percentage change in low field voltage from the $B = 0$ field value. To analyse these, a plot of the smoothed percentage change ($\Delta V(B) = (V_B - V_{B=0})/V_{B=0}$) against cyclotron orbit radius ($R_c = \hbar k_F/eB$) is shown in Figure 5.8 for temperatures between $5\ K$ and $25\ K$. Smoothing is achieved using the Savitzky-Golay algorithm [141], where a polynomial is fitted to a moving window of data to achieve a smooth interpolation. The Savitzky–Golay filter is a type of low-pass filter, particularly suited to smoothing noisy data, removing high frequency noise. This approach smooths each point by performing a least-squares fit with a polynomial of high order over an odd-sized window centred on that point. This method has the advantage of preserving the original shape and features present in the data better than other types of filtering techniques (e.g. moving average techniques) [142].

Also shown in Figure 5.8 are the extracted local minima shown by square symbols. As an aid to size scales involved, vertical lines are plotted to show the typical surface feature diameter (l_e, 2.43 μm, A) and radius (1.2 μm, B), the average spacing between threading

dislocations (TDs) for a density of $0.75 \times 10^9 \ cm^{-2}$ ($0.37 \ \mu m$, C) as given in Table 4.2, and the average spacing between background impurities for a density of $N_{bkg} = 5 \times 10^{14} \ cm^{-3}$ ($0.13 \ \mu m$, D). The deBroglie wavelength for these electrons is $\lambda \sim 0.05 \ \mu m$.
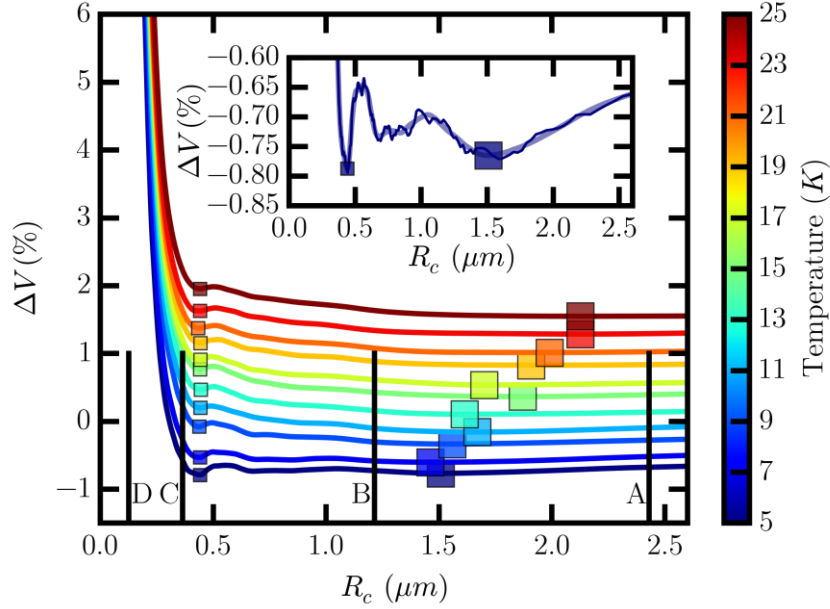
Figure 5.8 Low field magnetoresistance measurements shown in the inset of Figure 5.7 plotted as percentage voltage change from the $B = 0$ field value ($\Delta V = (V_B - V_{B=0})/V_{B=0}$) as a function of cyclotron radius ($R_c = \hbar k_F / eB$). The colour of each line represents temperature. Square symbols show extracted local minima, with the two sizes corresponding to the two distinct sets of minima persistent across temperature. Vertical lines correspond to: surface feature diameter (A), radius (B), threading dislocations (C) and background impurities at $N_{bkg} = 5 \times 10^{14} \ cm^{-3}$ (D). Inset: Raw data for the 5 K ΔV measurement (thin line), smoothed data (thick line) and extracted local minima (squares).

Figure 5.8 clearly shows two sets of well-defined minima in ΔV persisting across temperature, with strong alignments to values expected from independent measurements. The minima observed between ~1.5 μm and ~2 μm (large squares) clearly correspond to scattering related to the surface features, whilst the minima at ~0.5 μm (small squares) could either correspond to scattering from threading dislocations, or from a background dopant density $N_{bkg} < 5 \times 10^{14} \ cm^{-3}$.

To extract the characteristic scattering sizes corresponding to these minima, the R_c value was extrapolated to $T = 0 \ K$, allowing the determination of an average surface feature diameter and average background impurity density or threading dislocation density. For the surface features (large squares in Figure 5.8), the minima will occur when

the cyclotron radius is equal to the average feature radius. At this value, an electron can complete a full orbit within a feature without scattering from the surface feature boundary. Using this, the magnetoresistance measurement gives an extracted surface feature size of 2.48 μm. This is in excellent agreement with the 2.43 μm value extracted from optical Nomarski imaging. This gives further evidence these features are acting as the limiting scattering mechanism, as shown by transport lifetime modelling. In Figure 5.8, there is a clear trend with temperature for the R_c values at the surface feature related local minima. As yet this increase in R_c value is not well understood, with an explanation of this trend a possible avenue of future exploration.

Repeating the above analysis for the smaller squares in Figure 5.8, corresponding to either the background dopant density, or the threading dislocation density, the minima in the magnetoresistance gives a value of $R_c \approx 0.4 \ \mu m$. Considering the threading dislocations that propagate through the plane of the QW, these can be assumed to be evenly distributed, forming a regular square grid with spacing s (density N_{td}). If however the charged background dopant is considered, this can similarly be assumed to be arranged in a regular cubic grid (so having a square cross-sectional grid in the 2DEG), again with spacing s. When an electron scatters off an impurity ion or a dislocation line, it will begin a circular orbit with radius R_c from that point. It will continue in this orbit until it scatters from another ion or dislocation, where the shortest distance between scattering events for these orbits occurs when $R_c = s$ and $R_c = s/2$.

For a given s, the corresponding densities will be given by

$$N_{bkg} = \left(\frac{1}{s}\right)^3 \ \& \ N_{td} = \left(\frac{1}{s}\right)^2, \qquad (5.11)$$

and as the larger $R_c = s$ radius occurs at a lower B field, this condition will be met first.

Using the extracted value of $R_c \approx 0.4 \ \mu m$ therefore corresponds to an average threading dislocation density of $N_{td} \approx 5.1 \times 10^8 \ cm^{-2}$. If however the deBroglie wavelength is included so as to cause the electron wavefunction to extend further, then R_c can be equated to $s + \lambda$, giving a density of $N_{td} \approx 6.5 \times 10^8 \ cm^{-2}$. This is in strong agreement with the average value of $N_{td} \approx 7.5 \times 10^8 \ cm^{-2}$ given in Table 4.2.

Now, considering the minima to be related to background impurity scattering, the value of $R_c \approx 0.44 \ \mu m$ corresponds to an average background density of $N_{bkg} \approx 1.2 \times 10^{13} \ cm^{-3}$. Again, if the deBroglie wavelength is included so as to cause the electron wavefunction to extend further, then R_c can be equated to $s + \lambda$, giving a density of $N_{bkg} \approx 1.7 \times 10^{13} \ cm^{-3}$. This is again in agreement with independent measurements

performed on bulk AlInSb which gave a value of $N_{bkg} < 5 \times 10^{14}$ cm^{-3}, and is consistent with the Schrödinger-Poisson modelling performed in section 4.3.

This analysis suggests that scattering from threading dislocations, which was previously neglected from the transport model, should now be considered. Scattering from dislocations is often considered to be due to charged trapped along the line of the dislocation. The scattering rate from these dislocations have been considered by various authors, including Jena and Mishra [143], and Jena, Gossard and Mishra [144], with these works giving scattering rates of

$$\frac{1}{\tau_{Dis}} = \frac{N_{dis}m^*e^2}{\hbar^3\varepsilon^2c^2}\left(\frac{1}{4\pi k_F^4}\right)\left(\frac{2k_F}{q_{TF}}\right)^2\int_0^1 \frac{du}{\left(1+\left(\frac{2k_F}{q_{TF}}\right)^2 u^2\right)^2\sqrt{1-u^2}} \qquad (5.12)$$

and

$$\frac{1}{\tau_{Dis}} = \frac{N_{dis}m^*e^2\rho_L^2}{\hbar^3\varepsilon^2}\left(\frac{1}{16\pi k_F^4}\right)\int_0^1 \frac{du}{\left(u+\frac{q_{TF}}{2k_F}\right)^2\sqrt{1-u^2}} \qquad (5.13)$$

respectively, where $u = q/2k_F = \sin(\theta/2)$. In equations (5.12) and (5.13), N_{dis} is the number of threading dislocations per unit area penetrating the plane of the quantum well (i.e. the threading dislocation density, N_{td}), and $\rho_L = 1/c$ is the charge density along each dislocation line (i.e. there is a separation of length c between individual charges along the dislocation). It is often assumed that the dislocation is made up of a series of dangling bonds, with the value of c then given by the monolayer spacing. Using this value in equations (5.12) and (5.13), along with the dislocation density given previously, limiting mobilities of $\mu_{Dis} \approx 740,000$ cm^2/Vs and $\mu_{Dis} \approx 34,000$ cm^2/Vs respectively are predicted. These two scattering rate equations give large disagreements in predicted mobilities, with the origin of this discrepancy unclear. Due to this, and as both scattering rates are derived for the physical situation studied here (with reference [144] citing [143] in their derivation of scattering rate) the results of both equations are considered, giving a representative range in which the true scattering rate and limiting mobility are likely to lie. Incorporation of these values into the transport model for the sample with a 3 K mobility of $\mu \approx 200,000$ cm^2/Vs, including surface feature related scattering, would then give limiting mobilities of $\mu \approx 170,000$ cm^2/Vs and $\mu \approx 30,000$ cm^2/Vs. If the surface feature scattering were neglected, these values would be $\mu \approx 370,000$ cm^2/Vs and $\mu \approx 33,000$ cm^2/Vs respectively.

Clearly there is a large discrepancy between observed mobilities and those predicted from scattering due to a line of charge dislocations, even when considering the range of scattering times predicted by Jena and Mishra, and Jenna, Gossard and Mishra. If the background charge density measured of $N_{bkg} < 5 \times 10^{14}\ cm^{-3}$ is considered associated with these threading dislocations, then the density of charge along the line dislocation, and the spacing between charges, is given by

$$\rho_L = \frac{N_{bkg}}{N_{dis}} = \frac{5 \times 10^{14}\ cm^{-3}}{7.5 \times 10^8\ cm^{-2}} = 6.7 \times 10^5\ cm^{-1} = 67\ \mu m^{-1}, \qquad (5.14)$$

$$c = \frac{1}{\rho_L} = 15\ nm \approx 46 ML. \qquad (5.15)$$

Using this new value in equations (5.12) and (5.13) gives new limiting mobilities of $\mu_{Dis} \approx 1.6 \times 10^9\ cm^2/Vs$ and $\mu_{Dis} \approx 7.2 \times 10^7\ cm^2/Vs$ respectively. These new limiting values are now far higher than the measured mobility, showing that this scattering has a minimal effect on the observed mobility, and justifying neglecting this scattering from the transport model previously.

With this scattering having minimal effect on the mobility, it is likely that the minima in Figure 5.8 are associated with the charged background impurities, giving a density of $N_{bkg} \approx 2 \times 10^{13}\ cm^{-3}$. If this charge is associated with the threading dislocations penetrating the QW, this gives a line charge density and spacing of

$$\rho_L = \frac{N_{bkg}}{N_{dis}} = \frac{5 \times 10^{14}\ cm^{-3}}{7.5 \times 10^8\ cm^{-2}} = 2.7 \times 10^4\ cm^{-1} = 2.7\ \mu m^{-1}, \qquad (5.16)$$

$$c = \frac{1}{\rho_L} = 0.37\ \mu m \qquad (5.17)$$

respectively.

This analysis implies the low value of background charged impurities measured $(N_{bkg} < 5 \times 10^{14}\ cm^{-3})$ is likely associated with threading dislocations through the structure. This low value has been verified by Schrödinger-Poisson, and refined using detailed magnetoresistance measurements, giving a value of $N_{bkg} \approx 2 \times 10^{13}\ cm^{-3}$.

5.7 Combined MC and MR

This detailed magnetoresistance measurement allows for the refinement of the possible barrier shapes used in the MC modelling, where tunnelling currents, mobilities and thermionic emission currents can all be examined. This allows the model to be refined, and made more physically realistic, allowing for the verification and application of the model across all the single carrier samples measured (regions 1 and 2 in Figure 4.1). With this, the limiting scattering mechanisms across this range of samples can be determined, with this knowledge applied to improve future sample design iterations, and improving the subsequent mobilities achieved.

The first stage of this refinement is to implement the back-to-back Schottky-like barriers described previously, with the results compared to the simple Hall and magnetoresistance measurements shown above.

5.7.1 Thermionic emission current

As there is little temperature dependence in the voltage at the extracted minima, tunnelling through the potential barriers dominates, with very little thermal activation. Therefore, the height of the potential barriers between surface features must be sufficient to limit the thermionic emission over the barrier. The thermionic emission current density from the 2DEG over the barrier is then be given by [145]

$$J = A_{2D}^* T^2 \exp\left(\frac{q\Phi}{k_B T}\right) \tanh\left(\frac{qV}{2k_B T}\right),$$

$$A_{2D}^* = \frac{q(8\pi k_B^3 m^*)^{\frac{1}{2}}}{h^2}.$$

(5.18)

A plot of the thermionic emission current as a function of temperature for a typical mid-gap pinned barrier height ($\Phi = 150~meV$) and experimental voltage ($V = V_0 = 0.4~mV$) is shown in Figure 5.9. It can clearly be seen that the thermal current is significantly smaller than the experimental current measured in the low temperature region and would give no noticeable thermal activation. This matches the conclusions from the magnetoresistance measurements, and confirms that a mid-gap pinned Schottky-like barrier is a reasonable candidate.

Extending this analysis in the low temperature $3~K$ limit using this barrier, the energy dependent tunnelling probability can be determined as described previously. This

tunnelling probability can then be used to determine a tunnelling current and corresponding Monte Carlo mobility. Uncertainty in the pinning of the barrier height can be introduced through a variable height barrier (Φ'), and varying this to match the observed currents and mobilities as previously.
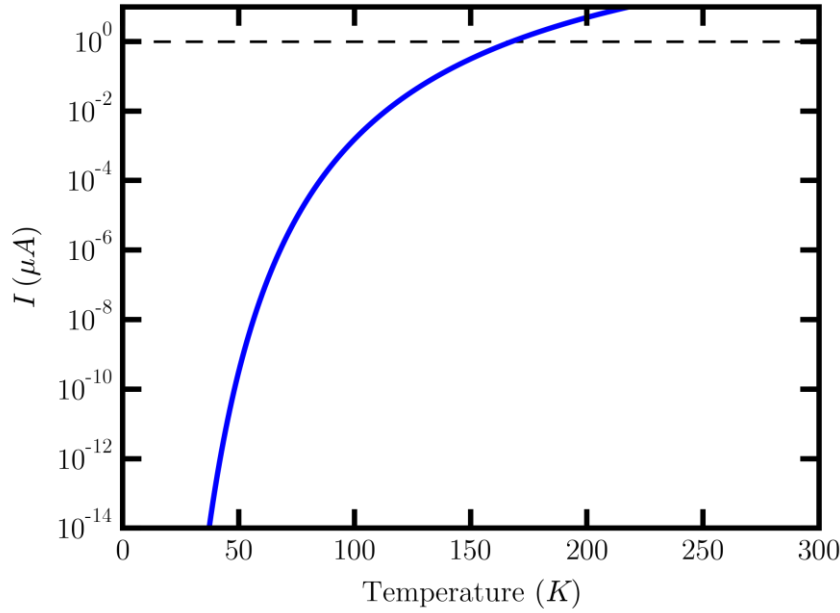
Figure 5.9 Thermionic emission current (blue) as a function of temperature for a back-to-back Schottky-like barrier as given by equation (5.18). The value of experimental current of 1 μA is marked by the horizontal dashed line (black), showing that for this height barrier, the thermionic emission current over the barrier is negligible at low temperatures.

The tunnelling currents and Monte Carlo mobilities calculated for a range of barrier heights $0.25\Phi \leq \Phi' \leq \Phi$, for both samples analysed previously, are shown in Figure 5.10. This figure shows the tunnelling current (red) and MC mobility (blue) for both the lower carrier density sample (SF1056, dashed) and higher carrier density sample (SF1055, solid). It can clearly be seen from Figure 5.10 that a reduced height barrier ($\Phi' < \Phi$) is needed to obtain the required 0.5 μA current matching the experimental value (accounting for the 2D nature as described in section 5.5.2). It also shows that the change of barrier height has a more significant effect on the current than on the predicted mobility. This mobility value is approximately equal to that determined previously with the rectangular barrier shape, which when included in the transport model, matched the experimentally determined value.

To achieve the correct value of tunnelling current and mobility, for the higher carrier density sample a barrier height of $\Phi' = 0.77\Phi \approx 115\ meV$ is required, where this value will be used subsequently. This is a small deviation from the assumed mid-gap value but gives excellent agreement between the Monte Carlo model and the magnetoresistance measurements.
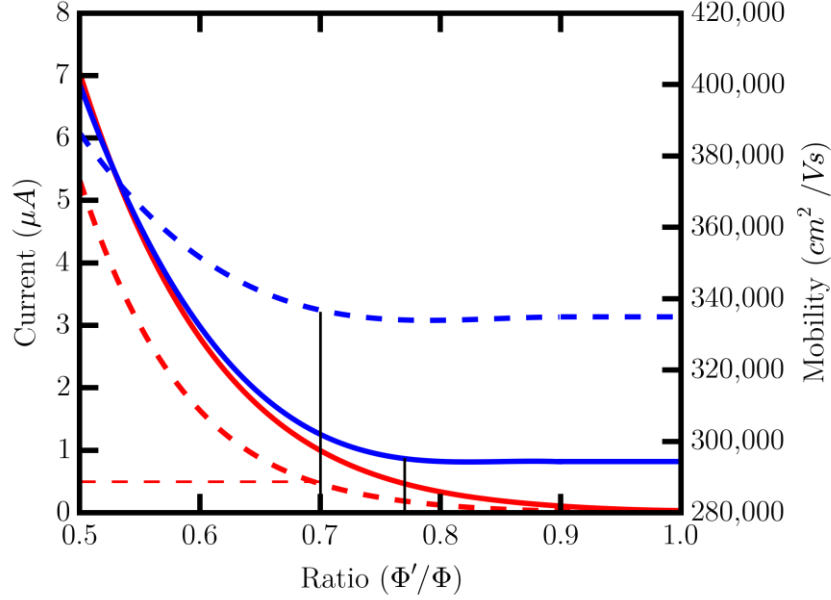
Figure 5.10 Calculated current (red) and MC modelled mobility (blue) as a function of the ratio of barrier height Φ′/Φ for the low carrier density sample (SF1056, dashed) and high carrier density sample (SF1055, solid). Vertical black lines mark the ratios where the calculated current matches the experimental value. This figure shows that the current is very sensitive to the barrier height, whilst the mobility has a much weaker dependence, with barrier height ratios of ~0.7 and ~0.77 giving the correct current values.

5.7.2 MC model for current and mobility as a function of carrier density

With the conclusion of the previous section that this potential barrier model is valid, a natural extension of the Monte Carlo modelled mobility is as a comparison to the originally postulated Drude model scattering rate (equation (4.48)), as well as the theoretical analytic expression (equation (5.8)). This has been performed as a function of carrier density at $3\ K$, with the results shown in Figure 5.11. Subsequent to this, the MC modelled mobility was included in the transport model described previously, and applied across the range of single carrier samples to match the trends shown in Figure 4.1 (i.e. for

regions 1 and 2). This determines the limiting scattering mechanisms across a range of carrier densities, with the results of this analysis shown in Figure 5.12.
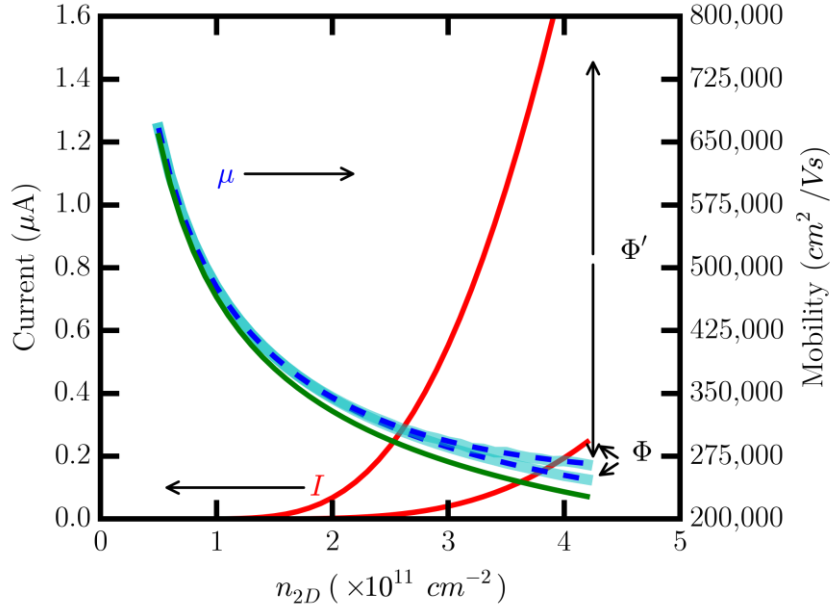
Figure 5.11 Calculated current (red, solid) and MC modelled mobility (cyan, solid) for both Φ and Φ' barrier heights as a function of carrier density n_{2D}. Also shown is the analytical mobility given by equation (5.8) (blue, dashed) matching exactly the MC mobility, and the simple Drude model mobility given by equation (4.48) (green, solid) giving lower values than the MC model and equation (5.8).

Figure 5.11 shows the calculated currents (left axis, red) and MC mobilities (right axis, cyan), as well as a comparison to the theoretical mobility (blue dashed) and Drude model mobility (green) for both barrier heights. This figure shows that the current is very sensitive to the barrier height used, with the 77% mid-gap Φ' barrier giving the correct current ($0.5\ \mu A$ at $n_{2D} = 2.89 \times 10^{11}\ cm^{-2}$) matching the experimental $1\ \mu A$ when accounting for the 2D nature. This figure also shows that the mobility is less sensitive to this change in barrier height, with the MC modelled values in excellent agreement with the theoretical predictions. At lower carrier densities there is little difference between the two barriers and the simple Drude model (equation (4.48)), however at increased carrier density, there is a larger discrepancy between the MC model and the Drude model, with the MC model predicting a marginally higher mobility. For all subsequent modelling, only the 77% mid-gap Φ' barrier is considered.

Extending this MC model further, the model's results were incorporated in to the previous transport model, with the aim of matching the trends shown in Figure 4.1 for regions 1 and 2. This requires accurate knowledge of the variation with carrier density of

various parameters included in the transport limiting scattering mechanisms. These parameters are primarily the non-parabolic effective mass (important for all scattering mechanisms), the separation between the 2DEG and the δ-dopant plane, d, and the total amount of dopant required to produce the correct carrier density in the well. These final two values primarily effect the remote ionised impurity scattering rate, with these relationships calculated through S.P. modelling in section 4.3, taking account of dopant dragging in the top cap.

With these parameters calculated as a function of carrier density, and using the lower background dopant density of $N_{bkg} = 2 \times 10^{13}\ cm^{-3}$ determined from magnetoresistance measurements, the transport model was used in conjunction with a least squares fitting routine to determine the best fit to the data. The fitting parameters varied were then the average surface feature diameter, l_e, using a range of $3\delta_n = 0.39\ \mu m$, and the interface roughness parameters I_Δ and I_Λ. The results of this fitting are shown in Figure 5.12.

Figure 5.12 shows that at lower carrier densities (i.e. region 1), when screening is reduced, interface roughness is the dominant scattering mechanism limiting the mobility, with a typical roughness scale of $I_\Delta \approx 1.4\ nm$ and $I_\Lambda \approx 17\ nm$ giving the best fit to the data. As the carrier density is increased, and so screening increases, the upper mobility limit also increases. Simultaneously, the scattering rate due to the surface feature related scattering also increases. This is due to the increasing Fermi velocity, v_F, decreasing the time taken for electrons to cross each feature, with this scattering giving an overall plateau in the mobility (i.e. region 2). The best fit to the data is given by surface features with an average size of $l_e \approx 2.8\ \mu m$. This value is remarkably close to the independent optical and magnetotransport measurements of $l_e = 2.43\ \mu m$ and $l_e = 2.48\ \mu m$ respectively, and a basic sensitivity analysis of these parameters shows this fitting is relatively insensitive to small changes from these optimal values.

In region 3, intersubband scattering becomes significant, however this is not considered in this work.
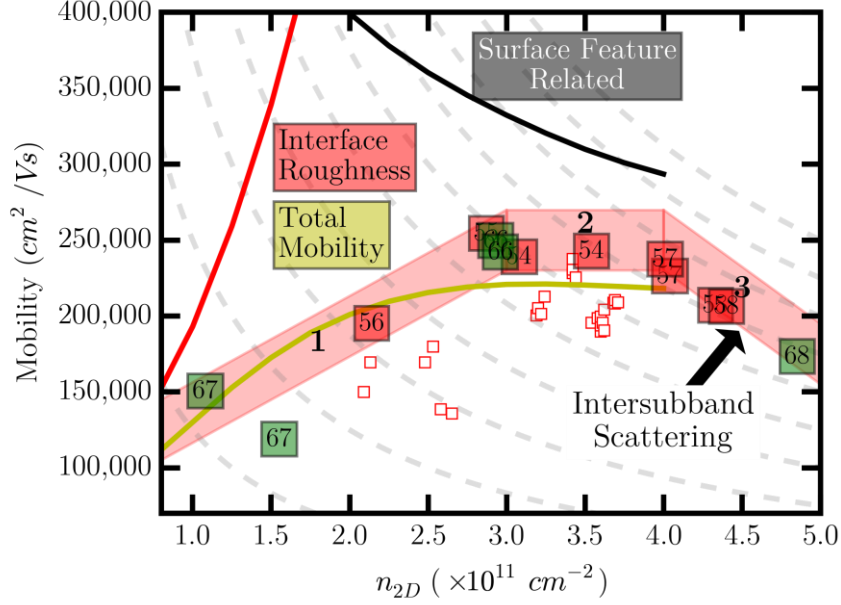
Figure 5.12 As Figure 4.1, measured 3 K mobility as a function of carrier density for batch IV samples (large, filled squares) and historical samples (small, unfilled squares). Solid lines represent a least squares fit to the total mobility (yellow), including scattering due to the interface roughness (red) and MC modelled surface feature related scattering (black). The best fit parameters are given by $I_\Delta \approx 1.4\ nm$, $I_\Lambda \approx 17\ nm$ and $l_e \approx 2.8\ \mu m$ respectively.

5.7.3 Possible mobility limits given future buffer redesign

With the optimal parameters determined previously, predictions of possible mobility improvements can be made due to improvements in average surface feature size alone. Figure 5.13 shows the range of possible mobilities for surface features ranging from $l_e = 1\ \mu m$ to $l_e = 5\ \mu m$, with a range of $3\delta_n = 0.39\ \mu m$ for each surface feature size considered.

Figure 5.13 shows that significant mobility improvements are possible through reduction of surface feature size, however these returns are limited. To achieve further improvements, structure design changes must also occur.
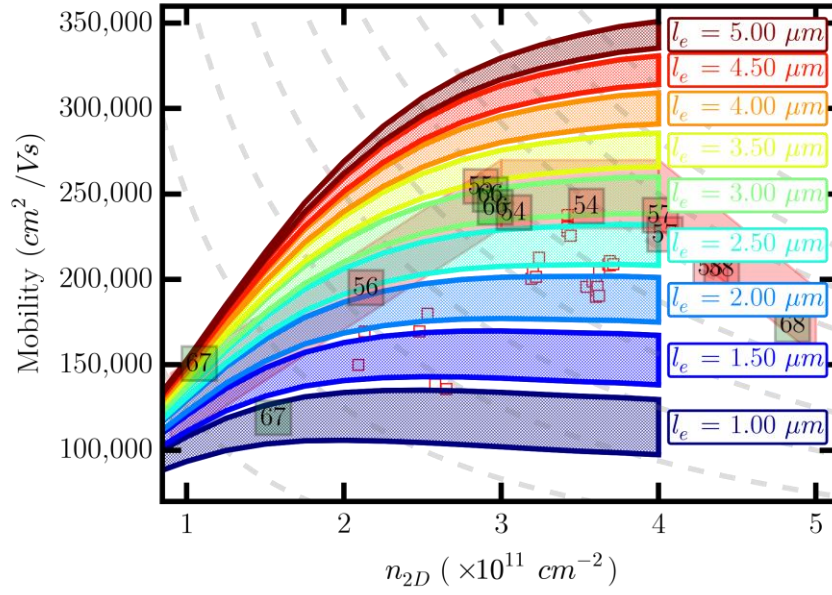
Figure 5.13 Predicted total mobility for a range of surface features sizes (SF), ranging from SF = 1 μm to SF = 5 μm, with a range for each interval of $3\delta_N \approx 0.4$ μm. The values for interface roughness scattering have been fixed at the best fit values determined previously.

This modelling and analysis has shown that a randomly distributed approximately mid-gap pinned back-to-back Schottky-like barrier, when combined with a Landauer tunnelling current calculation, a Monte Carlo and Drude model mobility calculation, and thermionic emission current calculations, gives excellent agreement to experimental data. Experimentally measured currents and mobilities can be matched using this potential barrier model, as can the lack of thermal activation of current over the barriers. This gives justification to the statement that scattering related to surface features, observed through Nomarski imaging, are the principle mobility limiting scattering mechanism. This means higher mobilities will not be achievable without a buffer redesign to reduce this surface feature density (and so increase the size of these features).

5.8 Conclusions

With scattering related to surface features acting as the limiting scattering mechanism, Landauer theory was used alongside Monte Carlo modelling to determine the nature of this scattering.

Initially, the boundaries between surface features were treated as rectangular barriers, having a regular height and width, and separated by the average surface feature size, l_e. Then, using Landauer theory to determine a tunnelling current for a given barrier height

and applied voltage, a range of barrier heights were determined for two samples for barrier widths ranging between 1 and 50 monolayers. These barrier heights were found to follow the approximate relationship height $\approx$ width$^{-1.3} \times 10$, where height is measured in eV and witdth in MLs.

Following this, these barriers were input into a Monte Carlo model, where the region within a surface feature was treated as ballistic, with scattering only occurring at the boundaries. An electron at an energy E then has a finite probability $T(E)$ of tunnelling through a barrier, with reflections acting as scattering events. With this assumption, the model was run, and a corresponding mobility for each barrier height and width determined. This was shown to again be the limiting scattering mechanism for multiple samples, with differences in mobilities for different barrier heights only arising at higher temperatures. At these temperatures phonon scattering is dominant and so it is not possible to determine a single barrier height consistent with experimental measurements.

Expanding on this, variations were introduced between barriers in the simulations, where variations in barrier heights, widths and separations were examined. It was shown that only the average values of these parameters affect the mobility. Refining the model further, more realistic Schottky-like barriers were introduced, and an analytic derivation for the mobility was performed, where this was subsequently shown to give an excellent match to the mobility as predicted by the MC model.

Detailed magnetoresistance measurements were performed, where it was shown that local minima in the low field resistance values correspond to characteristic scattering sizes. Extracting these sizes confirms scattering at a scale ~2.48 μm, in excellent agreement with the 2.43 μm value extracted through Nomarski imaging. Minima are also observed at a size giving a background dopant density of $N_{bkg} = 2 \times 10^{13}\ cm^{-3}$, again consistent with previous S.P. modelling and measurements.

Combining these magnetoresistance measurements with the MC model, it was shown that a back-to-back Schottky-like barrier pinned at ~77% of the mid-gap value is consistent with experimentally determined tunnelling currents and extracted mobilities. Extending this model to cover the range of single carrier samples measured, it was shown that observed trends could be described using a combination of interface roughness and surface feature related scattering. At lower carrier densities, the interface roughness scattering dominates, and as the carrier density increases, increased screening occurs, and the upper mobility limit increases. At further increased carrier densities, surface feature related scattering begins to dominate, with higher carrier densities giving higher Fermi velocities

and decreased lifetimes. For carrier densities above $\sim 4 \times 10^{11}\ cm^{-2}$, significant 2 carrier behaviour is observed, where mobilities are then limited by intersubband scattering.

With these surface features acting as the limiting scattering mechanism, mobility improvements will not be achieved without corresponding improvements in surface feature sizes. These larger feature sizes may be realised through buffer redesigns, where the corresponding upper mobility limits for a range of surface feature sizes were shown across a range of carrier densities.

6 : GATED STRUCTURE MODELLING

6.1 Introduction

The previous chapters have focussed on characterising the InSb QW samples studied in this work, determining the relationships between carrier density and mobility, and analysing the limits of these relationships. This chapter moves away from characterising the material, and instead focusses on a potential application of these InSb QW samples, namely in advanced "quantum" structures. Specifically, the devices studied here consist of nanoscale surface bar gated samples, with the aim of producing quantised conductance via a Zener tunnelling current (i.e. tunnelling laterally from the conduction band, through the forbidden energy band gap region, into to the valence band). Using the large Landé g-factor of InSb [11, 14], it is in principle possible to exploit the large Zeeman splitting to create a spin polarised, tunnelling current, where this could have many potential uses in future quantum computation applications [146, 147].

The following sections outline the design and fabrication of these devices, the modelling of the potential created by applying voltage to these devices, and the dependence of the currents achieved on device and material parameters. These simulation results are then analysed to determine the predicted quantised conductance due to a bar gated device.

6.2 Sample Design and Fabrication

The devices studied in this chapter consist of small area, triple bar gated structures. These consist of Ohmic current contacts either end of a short mesa, where this mesa is crossed by three parallel bar gates, with these connected to larger area pads for external fine-wire bonding. These gates consist of Schottky contacts, where ideally, any voltage applied would not give rise to a current through the device. In reality, any contact has a finite leakage current when a voltage is applied, and so to minimise the effect of this, the bar gates on the mesa are air-bridged to the bonding pads, ensuring only a small area connection to the active mesa area, and so a minimised leakage current. A scanning electron microscope (SEM) image of a typical device is given in Figure 6.1, whilst Figure 6.2 shows an increased magnification image of the triple bar gates in the active area of the device. Three bar gates are used so that the central gate can raise the

conduction and valence band edges to create a potential barrier and allow Zener tunnelling, whilst the outer two gates can in theory help to pin the potential away from the central gate, narrowing the potential barrier created.
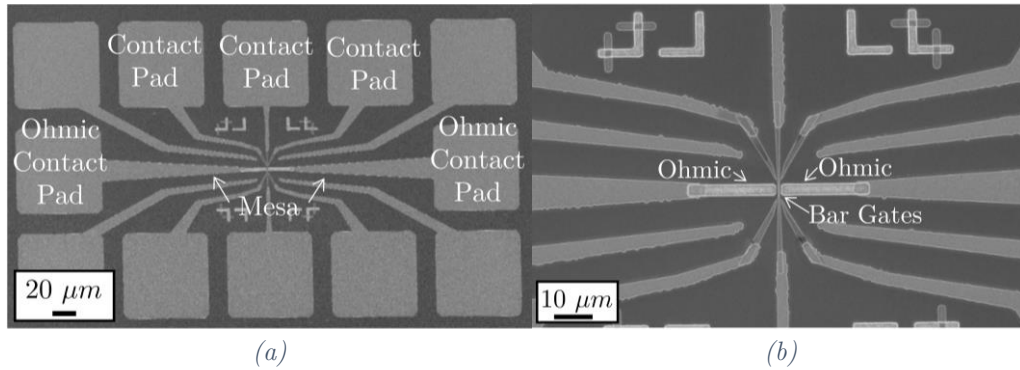
Figure 6.1 SEM images of device patterns for triple bar gated structures, showing the whole device (a), with large contact pads, contacting to a central mesa with 3 bar gates in the centre, and (b) central area showing 3 parallel bar gates crossing the central mesa between two Ohmic contacts
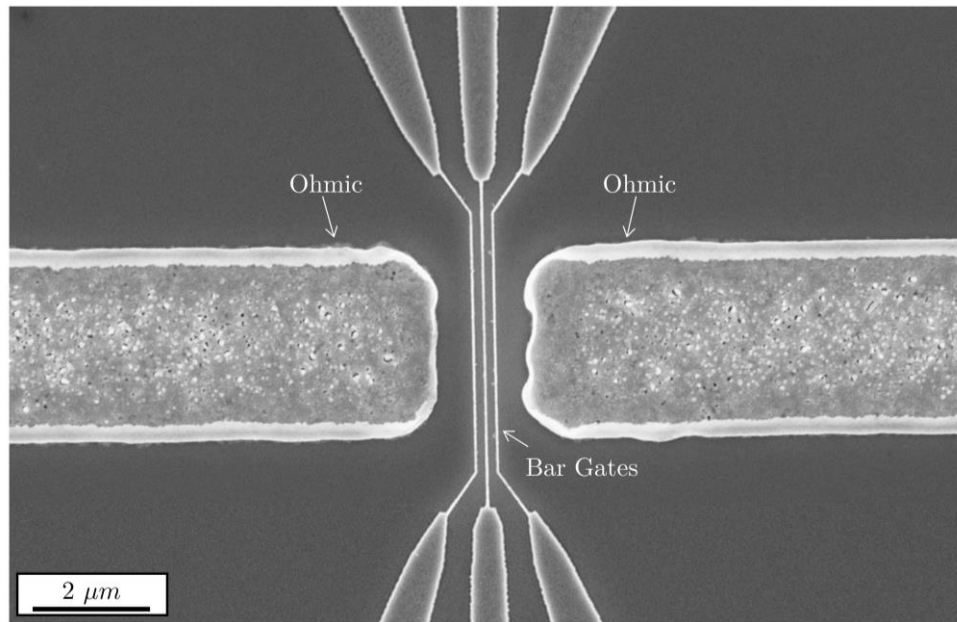
Figure 6.2 Scanning electron microscope (SEM) image of a triple bar gated structure between two Ohmic contacts.

To form the necessary Schottky contacts, new processing steps and a revised order of steps are needed compared to those given in section 3.4 for Hall bar devices. In theory, to form a Schottky barrier, a metal with the correct work function is necessary to create a potential barrier at the surface, causing a depletion layer within the semiconductor. However, in practice, in InSb the work function of metals does not directly correlate with the barrier height achieved, and a more empirical choice must be

made [16, 93, 148, 149, 150]. For the contacts used in these devices, a layer of 10 nm of nickel is used, followed by a 90 nm layer of gold, which in combination with the small area of the bar gates, has been found to give reasonably low leakage Schottky contacts. A feed metal recipe of 50 nm nickel, 300 nm gold is used for the link between the bar gates and the bonding pads.

The process order used to create a Hall bar consists of a mesa lithography and etch, followed by an Ohmic lithography, metal deposition and lift off. For these bar gated devices, the order is reversed, consisting of an optical lithography, metal deposition and lift off for the Ohmic contacts, an e-beam lithography, metal deposition and lift-off for the bar gates, and an optical lithography, metal deposition and lift off for the feed metal. Finally the mesa lithography and etch are performed.

6.2.1 Measurement

To measure these gated structures, a different program is used as compared to the Hall experiment control software described previously. These measurements use an Agilent E5270B Precision Measurement Mainframe with 4 Agilent E5281B Series Source Measurement Units. These can both supply current and apply voltage to the device under test, whilst also measuring the current or voltage sourced. This allows for monitoring of leakage currents whilst performing measurements.

6.3 Modelling the Potential Profile

6.3.1 Background and assumptions

Modelling of the bar gated structures has been performed to analyse the choice of sample design, the effect of applied voltage, and to make predictions about experimental results.

The modelling of the bar gated structures predominantly follows that of John H. Davies, Ivan A. Larkin and E. V. Sukhorukov in their work on "*Modeling the patterned two-dimensional electron gas: Electrostatics*" [151]. This work focuses on calculating analytically the potential in a heterostructure 2DEG, located a fixed distance, d, below the surface, resulting from biased patterned polygonal gates on the surface. It treats two main cases; the potential arising solely from electrostatics and governed by the depth of the 2DEG, and the screened potential (the potential profile when charge in the 2DEG is considered).

An example of the band structures studied in reference [151], for which the original derivations were performed, is shown in Figure 6.3. This consists of a 2DEG formed by a simple III-V modulation doped heterostructure, where the 2DEG is confined to a triangular well. The structure of the device consists of a gated surface on top of an n-doped barrier layer, such as AlGaAs, an undoped spacer layer of the same material, followed by an undoped channel layer, such as GaAs, where the 2DEG resides.
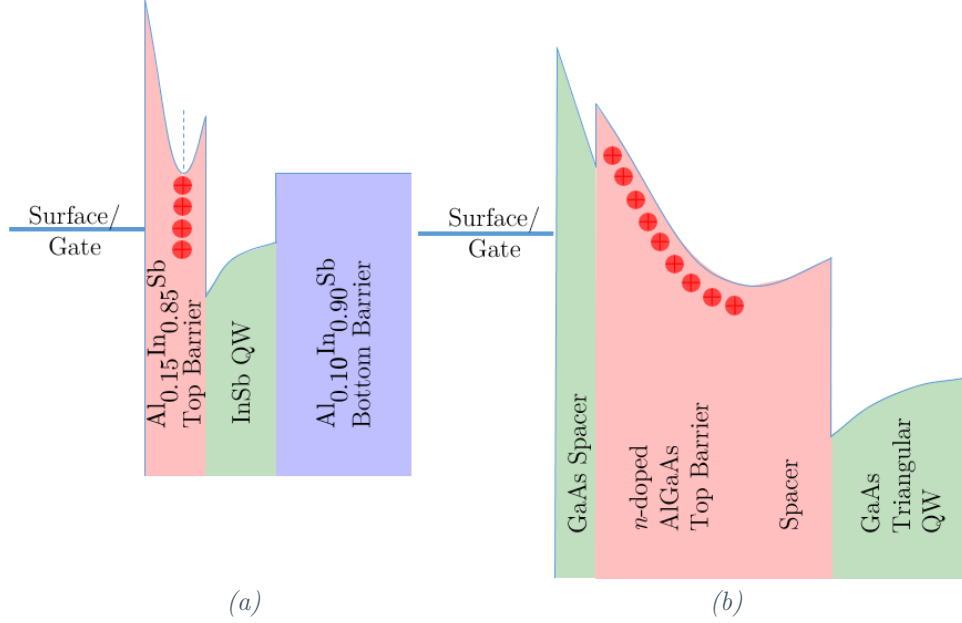
Figure 6.3 Schematic diagrams for comparison of structures studied in this work (a) and those studied in reference [151] (b). The doping is denoted by red $\oplus$ symbols, where in (a), the δ-doped layer is indicated by a dashed line. In (b), the doping is homogenous and a triangular well is formed.

These structures are similar in many regards to those studied here, e.g. with the conducting channel of electrons spatially separated from the positive ion cores of the dopant atoms, and with the number of charge carriers (approximately) equal to the number of positive ion cores. However there are also important differences. In the 2DEG formed in the triangular well, the positive ion cores are (approximately) evenly spread over the range from the top of the spacer above the 2DEG to the surface, giving a homogeneous charge density in the top cap. In the InSb quantum wells studied here, the doping is not uniform in the top cap, instead forming a δ-doped layer, giving a non-uniform charge density. This differing arrangement of charge will affect the electric-fields formed from applied gate voltages, and therefore slightly affect the numerical accuracy of some of the approximations and results considered subsequently.

To allow for simple calculation of the potential in the 2DEG, it is assumed that the positive charge donors are fixed in response to an applied gate voltage, i.e. there are no

free charge carriers, and all donors are either fully ionised, or electrons are trapped in deep states. As temperature increases, this approximation becomes less valid, however at lower temperatures, this approximation has been shown to be reasonable [152]. Any free carriers remaining in the donor layer will act to cause higher experimental voltages being required as compared to those predicted by theory.

Before calculation of the potential profile in the structure, the potential at the surface must be considered, where, as previously, it is assumed that the Fermi energy at the free surfaces of III-V structures is pinned at the mid-gap. When no voltage is applied, this is constant throughout the structure and for convenience this is taken as the zero of the scale of energy and potential. When a voltage is applied to a gate, the Fermi energy is shifted, where at the surface this shift is considered to occur only in the gate, with the free surface remaining at 0 V.

These assumptions now allow calculation of the potential profile at the surface due to a patterned gate, as well as at a depth into the structure corresponding to that of the 2DEG. The following sections derive the important equations necessary to achieve this.

6.3.2 Defining equations

Following the approach of Larkin and Davies [153] and Larkin and Sukhorukov [154], the potential profile $(\Phi(\boldsymbol{R}) = \Phi(\boldsymbol{r}, z))$ for a depth $z > 0$ can be calculated by solution of Poisson's equation with two charge densities, that of the positive donor layer and that of the 2DEG charge carriers, giving

$$\nabla^2 \Phi = \frac{\rho_p + \rho_e}{\varepsilon}, \tag{6.1}$$

where ρ_p and ρ_e are the volume charge density of the donors and the electrons in the quantum well respectively. For a 2DEG at a depth d, $\rho_e = -en_{2D}\delta(z - d)$. The boundary conditions on the potential are then given by

$$\Phi|_{z=0} = 0 \text{ or } \Phi|_{z=0} = V_g, \tag{6.2}$$

$$\left.\frac{d\Phi}{d_z}\right|_{z\to\infty} = 0, \tag{6.3}$$

where equation (6.2) represents the exposed areas of the surface, or those covered by gates with an applied voltage V_g respectively.

As ρ_p is assumed constant and uniform, the contribution due to the positive donors can be removed by substituting for a new potential, ϕ, and altering the boundary conditions accordingly. Now, as the total charge of the donors is approximately equal to that in the 2DEG, the potential on the unperturbed 2DEG is given by

$$\phi_{z=d} = V_t = -\frac{en_{2D}d}{\varepsilon}, \qquad (6.4)$$

where V_t is the threshold voltage, the voltage required to just remove all electrons from underneath a large area gate. The threshold voltage is determined by considering the system as a parallel plate capacitor, with one plate as the gate and one as the 2DEG, with this being the dominating energy scale of the system.

To account for the bound state of the quantum well, the depth d is extended from the interface of the QW to the peak of the wavefunction density. The kinetic Fermi energy in this system is given (approximately) by

$$E_F = \frac{n_{2D}}{g_{2D}}, \qquad (6.5)$$

where $g_{2D} = m^*/\pi\hbar^2$ is the 2D density of states. Any change in kinetic Fermi energy due to a potential on the gate can then be neglected as $|V_t| \approx 0.12\ V$ and $E_F \approx 0.03\ eV$, giving the ratio $|E_F/eV_t| \approx 0.25$. With this, it is possible to assume the potential at $z = d$ is constant everywhere the 2DEG is present, with this allowing Poisson's equation (equation (6.1)) to be replaced with Laplace's, such that

$$\nabla^2\phi = 0. \qquad (6.6)$$

This potential ϕ then represents the "bare" (unscreened) potential and is a function solely of depth and applied voltage. To account for charge present in the 2DEG and the associated screening of the potential, the induced change in the 2DEG charge density due to the applied potential must be determined, and the resulting change in energy calculated. As the electrostatic potential energy is larger than the kinetic Fermi energy, any local variation in the Fermi energy can be neglected, and the 2DEG can be treated as an equipotential, where it is also assumed that the 2DEG has a linear response due to a potential applied to the gate. This assumption is valid so long as the applied voltage is small enough that electrons remain in the 2DEG.

The charge induced at a point is proportional to the vertical electric field at that point (i.e. in the 2DEG at a depth d, $E_z = -d\phi/dz|_{z=d}$), multiplied by the permittivity ε.

Dividing this value of charge by the density of states, g_{2D}, then gives the corresponding change in energy, and so therefore the change in potential as

$$\phi_{scr}(\boldsymbol{r}, d) = \frac{\varepsilon}{g_{2D}} \frac{E_z(\boldsymbol{r}, d)}{e^2} = \frac{\pi \hbar^2 \varepsilon}{m^*} \frac{E_z(\boldsymbol{r}, d)}{e^2} = -\frac{a_0^*}{4} \frac{d\phi}{d_z}\bigg|_{z=d}. \qquad (6.7)$$

Here, a_0^* is the Bohr radius scaled by the relative permittivity and effective mass of the electron. This is the equivalent of a screening radius.

In this derivation it was assumed the donor layer formed an evenly distributed charge density between the 2DEG and the surface, though this is not the case for the structures studied here. A uniform charge density would lead to a uniform electric field and voltage drop across the top cap, that would respond linearly to gate voltage changes. In our structures the donor layer is instead focused in an (approximately) single δ-doped plane. This acts to change the way the voltage is dropped and how that changes with changing gate voltage. Now, the δ-doped layer would act as a form of pivot, where a large increase in gate voltage creates a large change between the surface and the δ-doping layer, and a relatively smaller change between the doping layer and the 2DEG. This means that estimates for quantities such as the threshold voltage will be underestimates, and potential profiles calculated for a specific voltage will correspond to experimentally higher voltages.

This offset can be determined by comparing the theoretical cut off voltage (when the potential causes the conduction band edge to cross the Fermi energy and conduction ceases) to the actual pinch off voltage experimentally measured. When the channel has been pinched off, all further trends should approximately match those predicted by the above theory, except for a constant voltage offset with the voltage offset being the difference between the two pinch off values.

6.3.3 Potential from patterned gates

This approach follows that of Davies et al. [151], where the aim is to find a general form, $\phi(x, y, z)$, as a solution to equation (6.6) for the given the boundary conditions. This is then simplified to give analytic equations for simple polygonal gates, where more complex structures can be expanded from there. To find a general formula for ϕ as a function of position, a 2D Fourier transform is performed on the surface potential, $\phi(\boldsymbol{r}, 0)$, transforming this to $\tilde{\phi}(\boldsymbol{q}, 0)$. Now $\tilde{\phi}(\boldsymbol{q}, z)$ must have the form of an exponential decay to satisfy equations (6.3) and (6.6), giving

$$\tilde{\phi}(\boldsymbol{q}, z) = \tilde{\phi}(\boldsymbol{q}, z) \exp(-|qz|). \qquad (6.8)$$

Performing the inverse Fourier transform, the potential in real space is then given by

$$\phi(\boldsymbol{r}, z) = \int \frac{|z|}{2\pi(z^2 + |\boldsymbol{r} - \boldsymbol{r}'|^2)^{3/2}} \phi(\boldsymbol{r}', 0) d\boldsymbol{r}' \qquad (6.9)$$

This is a complicated equation to solve for generic shaped gates, so again following Davies et al. [151], the potential is first considered under an infinite triangular gate, as shown in Figure 6.4.

The infinite triangle considered here is defined as having one vertex at the origin, with one edge aligned along the positive x-axis, and another edge at an angle $\theta = 2A$ to the x-axis. The voltage applied to the gate is V_g.
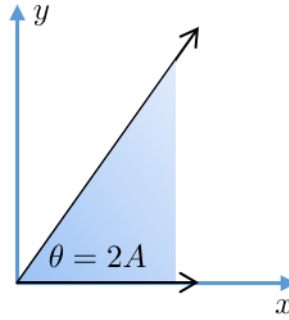
Figure 6.4 Infinite triangular gate in the xy-plane, enclosing an angle $\theta = 2A$ as measured anticlockwise from the positive x-axis. The triangular gate is represented by the shaded area.

Performing the integral in equation (6.9) in polar coordinates ($\boldsymbol{r} = (r, \theta)$), the bare potential profile in the 2DEG due to a bias on the gate is given by

$$\frac{\phi(\boldsymbol{r}, d)}{V_g} = -\frac{d}{2\pi} \int_0^{2A} d\theta' \int_0^\infty dr' \frac{r'}{[d^2 + r^2 + r'^2 - 2rr' \cos(\theta - \theta')]^{3/2}}. \qquad (6.10)$$

The integral over r' can be evaluated using the standard integral (2.264.6 from [155])

$$\int \frac{xdx}{\sqrt{T^3}} = -\frac{2(2a + bx)}{\Delta\sqrt{T}}, \qquad (6.11)$$

where $T = a + bx + cx^2$ and $\Delta = 4ac - b^2$, with a, b, and c given by $a = d^2 + r^2$, $b = -2r\cos(\theta - \theta')$ and $c = 1$ respectively. Now Δ and T are given by

$$\Delta = 4(d^2 + r^2) - (-2r\cos(\theta - \theta'))^2,$$
$$T = d^2 + r^2 - 2rr'\cos(\theta - \theta') + r'^2. \tag{6.12}$$

Using the standard result from equation (6.11), and expanding $\Delta\sqrt{T}$ about 0 and ∞ to account for the limits of integration, the potential is then given by

$$\frac{\phi(\boldsymbol{r}, d)}{V_g} = -\frac{d}{2\pi} \int_0^{2A} \frac{d\theta'}{R - r\cos(\theta - \theta')}, \tag{6.13}$$

where R is the 3D distance from the vertex of the gate to a position (x, y, d), and r is this projection on the (x, y) plane. This can again be evaluated using a standard integral (2.553.3 from [155]), giving the potential at depth d for a triangular gate enclosing an angle of $2A$ as

$$\frac{\phi(\boldsymbol{r}, d)}{V_g} = -\frac{1}{\pi}\arctan\left(\frac{d\sin(A)}{(R - x)\cos(A) - y\sin(A)}\right). \tag{6.14}$$

This potential is plotted in Figure 6.5 for a unit positive potential. In equation (6.14) it is important to use the correct branch of the arctangent function, where the value depends on the individual signs of the argument denominator and numerator (i.e. x and y for $\arctan(y/x)$), and not simply their ratio. This is determined using the NumPy function "arctan2(y, x)", where the values are summarised in Table 6.1 for the range of possible x and y input values.

The "bare" potential in the 2DEG at a typical depth of $d \approx 65$ nm for a triangle gate with an angle $\theta = 2A = \pi/4$, and applied voltage $V_g = +1$ V is shown in Figure 6.5.

To determine the potential from more complex gate geometries, the triangular gate example can be extended to cover larger angles (e.g. to cover the upper half plane), and the range of inputs can be limited or rotated (e.g. to give a stripe gate). Using these results, superpositions of gates can be used to determine the potentials from any arbitrary gate geometry. An example of a short wire formed by a split gate structure is shown in Figure 6.6 for a unit negative potential. This structure can be used to restrict a 2-dimensional system to a 1D wire, giving rise to quantised conductance [50, 156, 157].

Table 6.1 Output from the Numpy "arctan2(y, x)" function for various x and y.

Input range	arctan2(y, x)
$x > 0$	$\arctan(y/x) + 0$
$x < 0, y \geq 0$	$\arctan(y/x) + \pi$
$x < 0, y < 0$	$\arctan(y/x) - \pi$
$x = 0, y > 0$	$+\pi/2$
$x = 0, y < 0$	$-\pi/2$
$x = 0, y = 0$	undefined
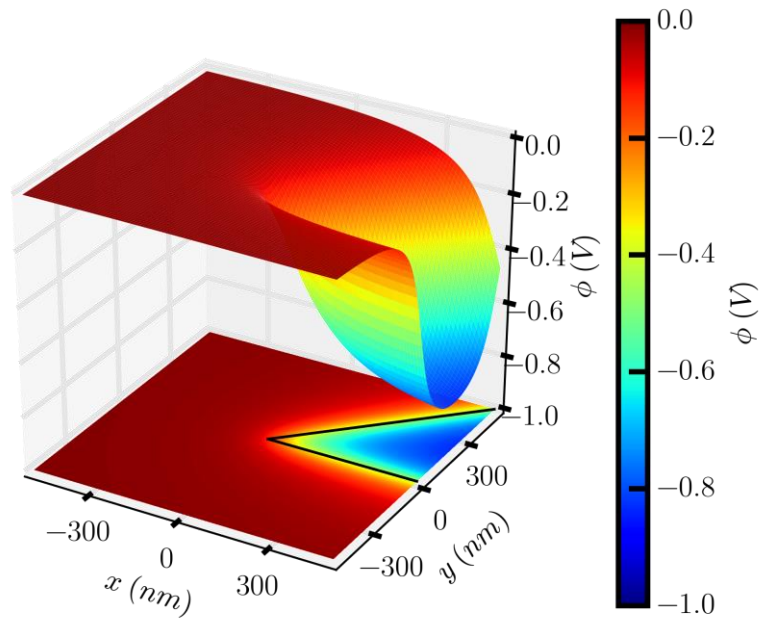
Figure 6.5 Potential at a depth of 65 nm from triangular gate with unit positive gate voltage applied.
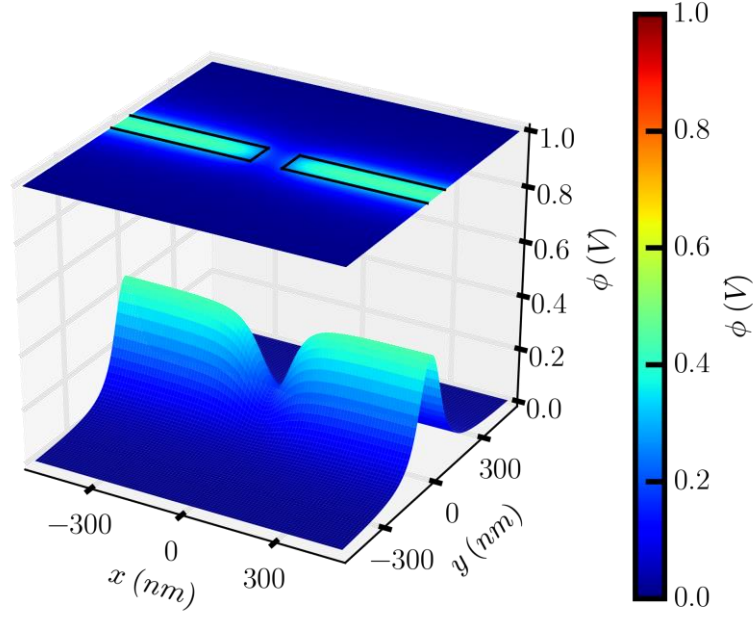
Figure 6.6 Potential at a depth of 65 nm from a split gate structure with unit negative voltage applied. A short 1D wire is formed which can be used, for example, to observe quantised conductance.

For the bar gated structures studied here, the equivalent model is of a stripe gate invariant in the y-direction, extending from $x = -a$ to $x = +a$ (i.e. a width $2a$), where the bare potential is then given by

$$\frac{\phi(\mathbf{r}, d)}{V_g} = \frac{1}{\pi} \left(\arctan\left(\frac{a+x}{d}\right) + \arctan\left(\frac{a-x}{d}\right) \right). \tag{6.15}$$

To extend equation (6.15) to give the potential from a triple bar gated structure, the separation between gates must be given, where this edge to edge separation is given by $2b$. Similarly, the voltage on the inner and outer gates must be denoted separately, where the voltage on the centre gate is given by V_{g1}, and the outer gates by V_{g2}. An example of the bare potential from a triple bar gated structure is shown in Figure 6.7.

To determine the equivalent screened potential, a conformal map must be performed, where this gives ϕ_{scr} as

$$\frac{\phi_{scr}(\mathbf{r}, d)}{V_g} = \left(\frac{a_0^*}{4d}\right) \left(\frac{\tanh\left(\frac{\pi a}{2d}\right)}{1 + \text{sech}^2\left(\frac{\pi a}{2d}\right)\sinh^2\left(\frac{\pi a}{2d}x\right)} \right). \tag{6.16}$$

A cross-section comparison of the bare and screened potentials is shown in Figure 6.8 for a triple bar gated structure with gate widths of $2a = 100\ nm$, sperations $2b = 100\ nm$, and applied voltages $V_{g1} = -V_{g2} = -1\ V$.
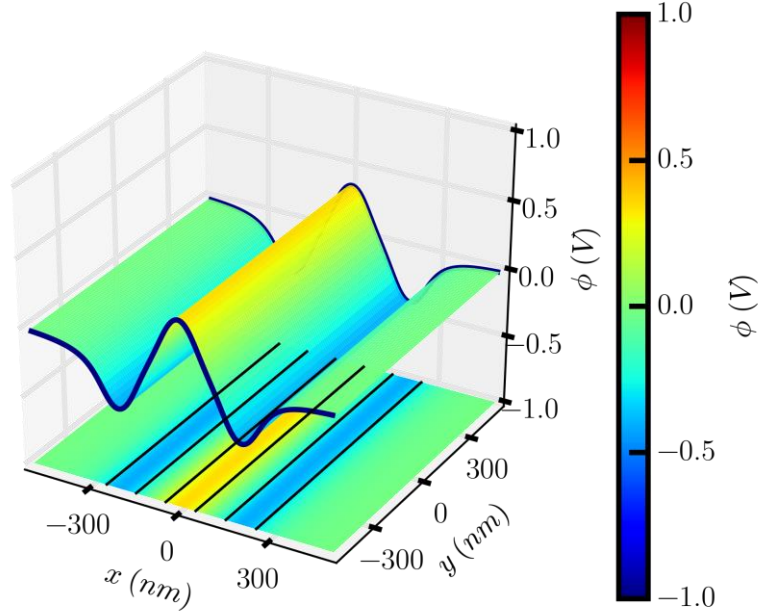
Figure 6.7 Potential at a depth of 65 nm from a repeated stripe (bar) gated structure. For all gates, the widths and separations are $2a = 2b = 100$ nm. For the central gate, the applied voltage is $V_{g1} = -1\ V$, whilst for the outer gates, $V_{g2} = +1V$.
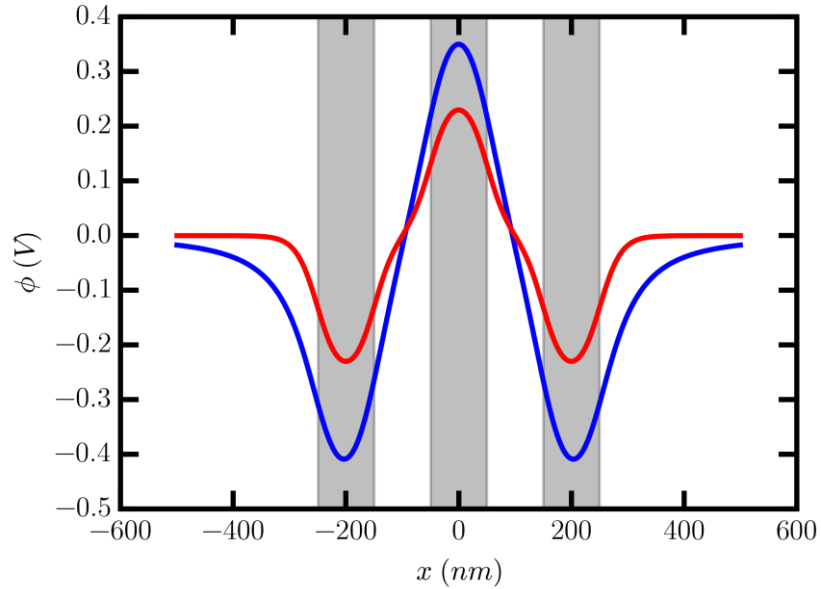
Figure 6.8 Comparison between bare potential (blue) and screened potential (red) at a depth of 65 nm due to a triple bar gated structure with $2a = 2b = 100$ nm and $V_{g1} = -V_{g2} = -1\ V$. The area covered by the gates is shown by the grey shading.

6.3.4 Pinch off voltage

To calculate the pinch off voltage required to just deplete the area under the middle of a narrow gate, stopping conduction, the Fermi energy under the gate must be raised above the threshold voltage (equation (6.4)). The relationship between Fermi energy and threshold voltage is given by

$$E_F = e \frac{a_0^*}{4d} V_t,$$
(6.17)

and so the screened potential in the 2DEG beneath the centre of the gate must be raised to this value, i.e.

$$\phi_{scr}(0, d) = \frac{a_0^*}{4d} V_t.$$
(6.18)

By simple rearrangement of equation (6.16), this corresponds to a gate voltage of

$$V_g = V_t \coth \left(\frac{\pi}{2} \frac{a}{d} \right).$$
(6.19)

For a 2DEG at a depth of $d \approx 65 \ nm$, V_g ranges between gate voltages of ~$2V_t$ to ~$1V_t$ (~$0.3 \ V$ to ~$0.15 \ V$) for a ranging from $a = 20 \ nm$ to $a = 50 \ nm$ respectively. The region occupied by electrons is then given by the contour equal to E_F/e marking the boundary between occupied and unoccupied regions.

6.4 Band Structure Variation due to Gated Structures

The potential profiles calculated above for various applied gate voltages can now be considered as representing the conduction band profile in the QW in the vicinity of the gate, where a value of $1 \ V$ corresponds to an energy of $1 \ eV$. Figure 6.9 shows an example of the conduction and valence band profiles due to various values of V_{g1} applied to the central bar gate of the triple bar gated structures, with $V_{g2} = 0 \ V$ or $V_{g2} = 1 \ V$ applied to the outer gates. This figure shows the clear potential for Zener tunnelling between the conduction and valence bands at low voltages due to the narrow band gap of InSb.
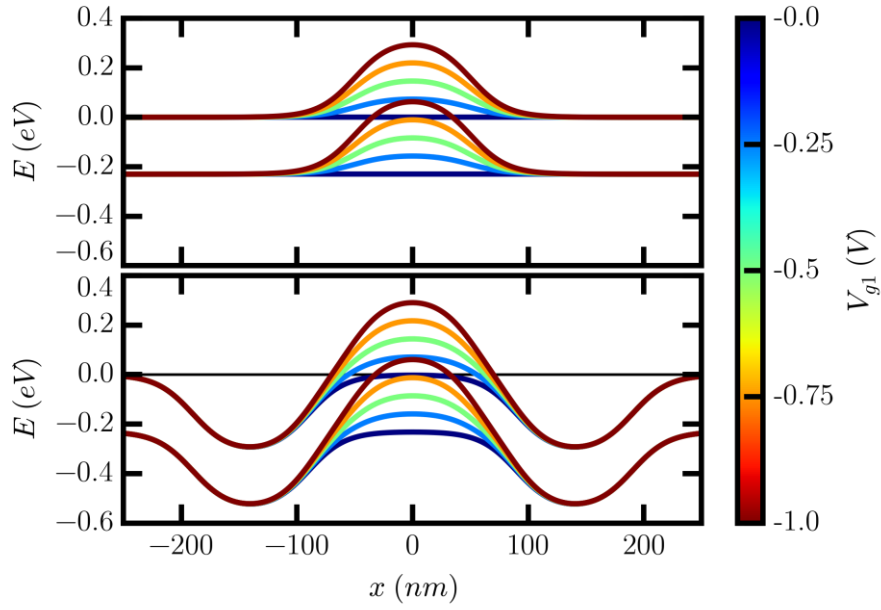
Figure 6.9 Conduction and valence band profiles in the region of bar gates with various applied central gate voltages, V_{g1} (the voltage is represented by colour). The applied outer gate voltage is $V_{g2} = 0$ V (top) and $V_{g2} = 1$ V (bottom).

In designing and experimentally measuring these bar gated structures, several parameters can be varied, including inner gate width, outer gate width, separation of gates, and the voltages applied to all gates. Variations in any of these will change the band structure obtained and affect the physical observations made. There are however physical limits placed on the ranges of these parameters, such as the resolution of the e-beam lithography, and the maximum voltage the small gates can withstand. These limitations must be accounted for when modelling these devices with the aim of influencing future design, fabrication and experimental testing.

The main figures of merit used in subsequent modelling are the voltage required to achieve Zener tunnelling (i.e. for the top of the valence band to cross the energy of the unperturbed conduction band edge), the corresponding width of the potential barrier between the conduction and valence bands, and the probability of tunnelling across this potential barrier.

6.4.1 Voltage to achieve Zener tunnelling

As the primary gate used to achieve Zener tunnelling is the central gate, the value of V_{g1} required to achieve Zener tunnelling can be determined for a range of physical device parameters (the gate width $2a$, and the gate separation $2b$). This voltage can be used to assess whether a device of the given dimensions would be feasible, where a device is feasible

if the voltage to achieve Zener tunnelling is sufficiently small that gate leakages will be minimal, and the gate would not break down. This value of V_{g1} to achieve Zener tunnelling for no voltage applied to the outer gates ($V_{g2} = 0$) is shown in Figure 6.10, whilst the corresponding value for various applied V_{g2} values is shown in Figure 6.11. For simplicity, the outer gate width is fixed as the same width as the inner gate, and the outer gate voltages examined range from $V_{g2} = 0$ V to $V_{g2} = 5$ V in 1 V steps.

Figure 6.10 and Figure 6.11 together show that for most gate dimensions, the voltage required to achieve Zener tunnelling is around $V_{g1} \approx -1$ V to $V_{g1} \approx -2$ V, with a weak dependence on b and V_{g2}. However, for smaller gate widths with smaller separations, this voltage rises sharply, though the magnitude remains $\lesssim 5$ V. This means that for all gate widths above $2a \approx 40$ nm, a reasonable lower width limit for e-beam fabrication, the voltages required for all dimensions are reasonable. However, as stated previously, these voltages will be underestimates due to the stated assumptions, though this value is still indicative of the approximate value for the scale of experimental voltage required.
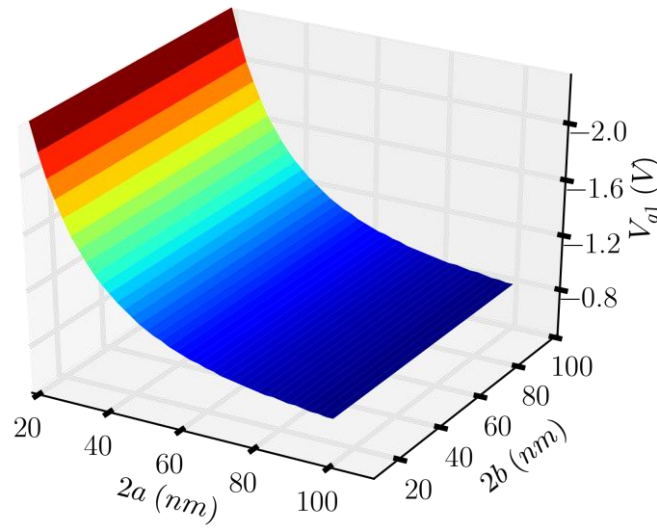
Figure 6.10 Central gate voltage (V_{g1}) required to enter the Zener regime (i.e. to raise the valence band above the energy of the unperturbed conduction band edge) as a function of gate width (2a) and separation (2b) for $V_{g2} = 0$ V. For gate widths > 40 nm the magnitude of voltage required is less than 1.5 V, though this value rises sharply for thinner gates, being $|V_{g1}| > 2$ V for 20 nm gates.
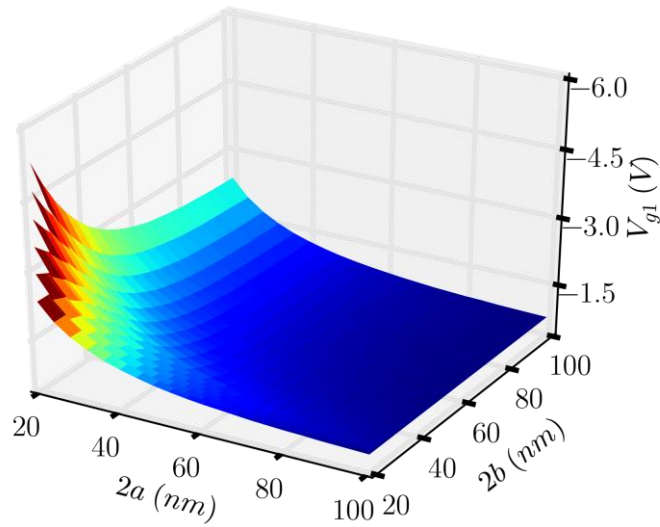
Figure 6.11 Voltage (V_{g1}) required enter the Zener regime for $V_{g2} = 0$ V (bottom) to $V_{g2} = 5$ V (top) in 1 V steps. For the majority of gate widths and separations, the magnitude of voltage required is small (less than 1.5 V), though again this rises sharply for narrow, closely spaced gates.

6.4.2 Investigations into the effect of gate voltage and barrier separation

Knowing the voltages required to enter the Zener regime, the specifics of the band structure in this regime can be examined for a range of device dimensions and applied voltages. As the width of the gates and their separations are altered, or the applied voltages are altered, the band structure in the area of interest will change, either narrowing or increasing the width of the gap between the conduction and valence bands. This width is important in determining the tunnelling probability for an electron to tunnel from the conduction to the valence band, and so is important for the resultant current.

This gap between the bands can be treated as an effective right angled triangular potential barrier [158] (as shown in Figure 6.12), with a width dependant on the specific device dimensions and voltages, and a height E_g. The width of this barrier is determined between the point where the conduction band crosses the Fermi level and the equivalent point in the valence band.
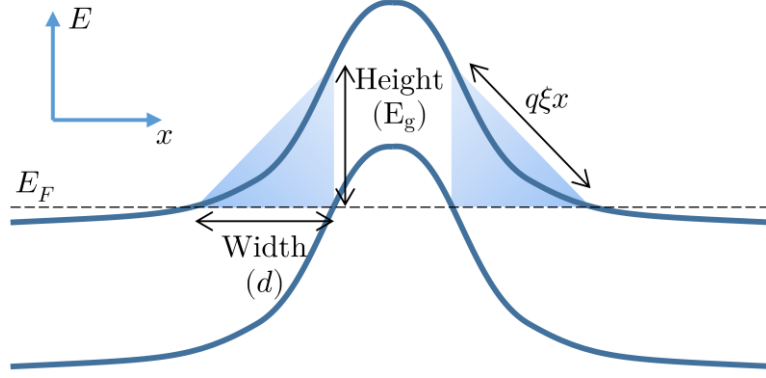
Figure 6.12 Triangular potential barrier approximation for the barrier to Zener tunnelling. The height of the barrier is given by the band gap, E_g, the width as measured at the Fermi energy is given by d, and the barrier height as a function of position is given by qξx, where ξ is the maximum electric field of the barrier.

In the schematic barrier shown in Figure 6.12, ξ is the maximum electric field of the barrier, and q is the electronic charge. To determine an approximation for the tunnelling probability for an electron at an energy E to cross such a barrier, the Wentzel-Kramers-Brillouin (WKB) approximation can be used [158]. In the WKB approximation, the solution to the Schrödinger equation is treated as a set of running waves, given by

$$\psi(x) = \psi_0 \exp\left(\int_0^x ik(x)dx\right), \qquad (6.20)$$

where

$$k(x) = \sqrt{\frac{2m_e^*(V(x) - E)}{\hbar^2}} \qquad (6.21)$$

and

$$V(x) = E + q\xi x. \qquad (6.22)$$

Then, using the WKB approximation, a general equation for the transmission probability (T_{WKB}) across a barrier of width d (where $d = E_g/q\xi$) is given by

$$T_{WKB} = \exp\left(-2\int_0^d |k(x)|dx\right), \qquad (6.23)$$

giving

$$T_{WKB} = \exp\left(-\frac{4\sqrt{2m_e^*}E_g^{\frac{3}{2}}}{3q\hbar\xi}\right) = \exp\left(-\frac{4\sqrt{2m_e^*}E_g^{\frac{1}{2}}}{3\hbar}d\right). \tag{6.24}$$

Equation(6.24) shows that the transmission probability is exponentially dependant on the barrier width, determined by the separation of the conduction and valence bands at $E = E_F$.

Using equation (6.24), the barrier width and transmission probability can be examined as a function of the central gate voltage for a range of V_{g2} values (Figure 6.13) and gate separations (Figure 6.14). For both of these figures, the central gate width was fixed at a value $2a = 100$ nm, whilst for Figure 6.13 the gate separation was fixed at $2b = 40$ nm, and for Figure 6.14 the outer gate voltage was fixed at $V_{g2} = 2$ V.
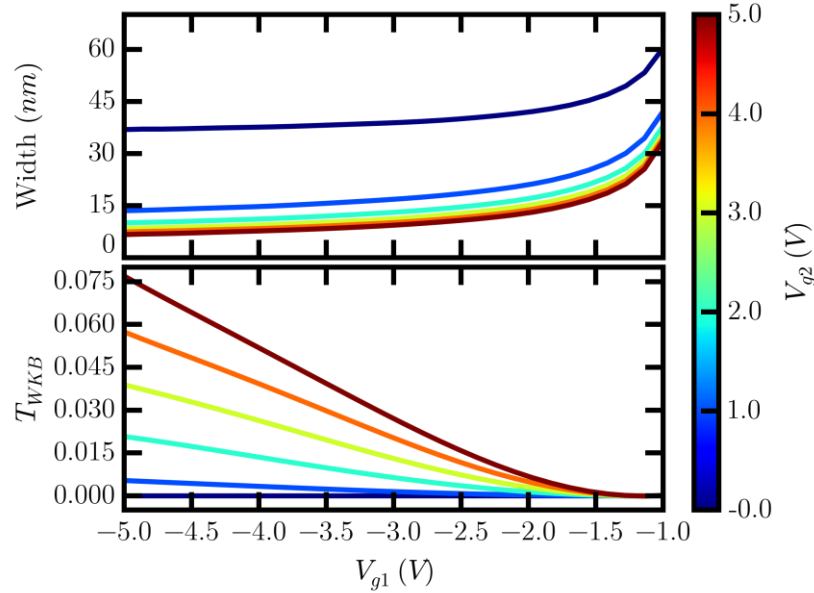
Figure 6.13 Triangular barrier width (top) and corresponding transmission probability, T_{WKB} (bottom), as a function of central gate voltage (V_{g1}) for several outer gate voltages (V_{g2}) between 0 V and 5 V (colour represents the value of V_{g2}). The width of all gates is fixed at $2a = 100$ nm, whilst the separation is fixed at $2b = 40$ nm to give an increased sensitivity to V_{g2}. For increasing magnitude of V_{g1} the barrier width reduces, with this effect further increased with increasing V_{g2}, however for all V_{g1} and V_{g2} the probability of tunnelling is minimal.

Figure 6.13 shows that for increasing V_{g1} values, the width of the effective potential barrier decreases, and the T_{WKB} value increases, however this value is still very small. Applying moderate outer gate voltages decreases the width of the barrier, and increases the transmission probability, though again this value is very small.

A similar trend to Figure 6.13 is seen in Figure 6.14, where at large gate separations, the outer gates have little effect on the barrier width. Reducing this separation decreases the widths of the effective potential barriers and increases the transmission probability, however, as previously, this value is very small.
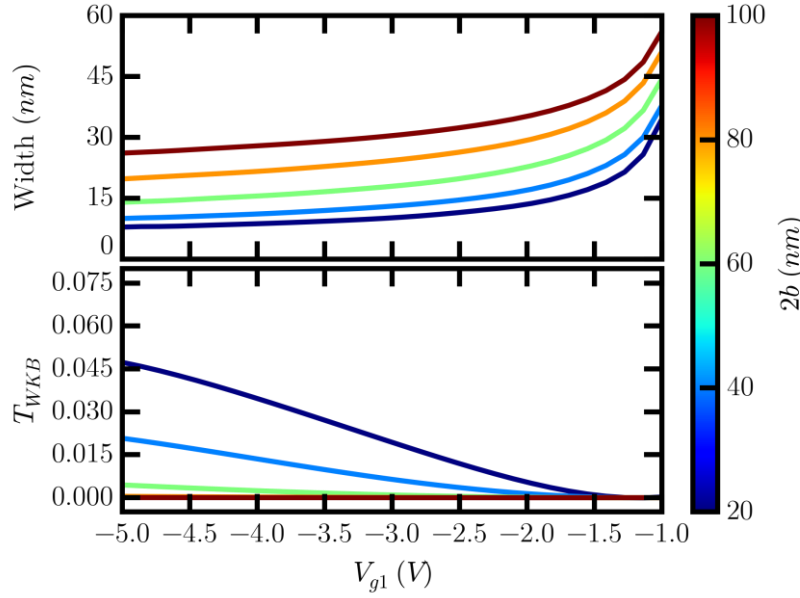
Figure 6.14 Triangular barrier width (top) and corresponding transmission probability, T_{WKB} (bottom), as a function of central gate voltage (V_{g1}) for several gate separations (2b) between 20 nm and 100 nm (colour represents the value of 2b). The width of all gates is fixed at 2a = 100 nm, whilst the outer gate voltage is fixed at $V_{g2} = 2$ V. For increasing magnitude of V_{g1} the barrier width reduces, with this effect further increased with decreasing 2b, however for all V_{g1} and 2b, the probability of tunnelling is minimal.

It is also worth noting that to tunnel into the valence band, a state must be available for the incoming electron. Due to the dimensions of the confining potential, the states in this region are further confined from the 2-dimensional states of the QW to 1-dimensional states. This quantisation gives a separation in energy levels, and so restricting the energies at which electrons may tunnel. Therefore, tunnelling will only occur when a confined state is aligned to the energy of the incoming electron.

It is also worth noting that, for symmetrical barriers, if the incident energy is aligned to a state between the barriers, the system is then at resonance (as in a double barrier resonant tunnelling structure). When a system of double barriers is symmetric, and at resonance, the total transmission probability through this system raises, in principle, to unity [135, 159, 160, 161, 162, 163, 164]. Applying a longitudinal voltage to this system will shift the system off resonance, though for small longitudinal voltages, the transmission

will still be greatly elevated compared to the off resonance states as given by the T_{WKB} values shown above. It is therefore still beneficial to realise the smallest barrier widths and largest T_{WKB} values practically possible, with this giving an increased transmission when longitudinal voltages are applied.

6.4.3 Direct tunnelling across the conduction band

The alternative to tunnelling into a confined valence band state on resonance is tunnelling directly across the conduction band barrier (i.e. bypassing the intermediate valence band state). This can be calculated using the matrix methods of Ando and Itoh as described previously [135].

Figure 6.15 shows this probability as a function of energy for a potential barrier created by a 40 nm central gate with a bias of $V_{g1} = -2\ V$. Figure 6.15 also shows the confined light hole state energies and wavefunctions for these states in the confining valence band potential.

As Figure 6.15 shows, the tunnelling probability is vanishingly small for energies close to the conduction band edge, only rising to appreciable values near the conduction band maximum. This energy is far above the Fermi energy of the system, confirming that electronic conduction can only occur via Zener tunnelling into confined valence band states.

6.4.4 Light hole & heavy hole bound states

To calculate the energies at which conduction will occur, the energies of the bound hole states must be determined. This was achieved via use of a Schrödinger solver, using the shooting method and the Newton-Raphson method, to calculate the energy and wavefunction solutions given a confining potential (given by the modelling performed above). A description of this Schrödinger solver is given in Appendix A.

In the valence band, both heavy and light hole states will be present. At $k = 0$ these states are degenerate, however due to the different curvature of each dispersion relation, the light and heavy holes have significantly different masses ($m_{lh}^* = 0.015\ m_e$ and $m_{hh}^* = 0.43\ m_e$ respectively). This large discrepancy in effective mass will give rise to significant differences in state energies and separations.

For light hole states, the spacing between state energy levels will be large, whilst for heavy holes the separation will be reduced. There will also be a difference in the ground state energies (as measured from the peak of the valence band), with this ground state

energy offset and energy level separation having significant consequences for the quantised conductance discussed later.

These state energies will also be directly influenced by the dimensions of the device, with narrow gates giving rise to a narrower, steeper well, and giving larger state energy separations. Applying a voltage to the outer gates will also change the band structure, making the potential well steeper again, and therefore further increasing the energies of each state. However, as can be seen from Figure 6.9, this change is small and the difference in state energies as compared to the $V_{g2} = 0$ case will be minimal.

Figure 6.15 shows the transmission probability across the barrier as a function of energy for a typical valence band well, with the light hole wavefunction solutions superimposed. This shows that, when not aligned with a state, the probability of electrons near the Fermi energy tunnelling directly across the conduction band barrier is ≈ 0. However, there are also numerous confined states within this potential, with clear energy separations. At these specific state energies, the system will be at or near resonance, and as stated previously, the transmission will be increased to a value ≈ 1. The effect of a non-unity transmission will be to reduce the measured current compared to that predicted.
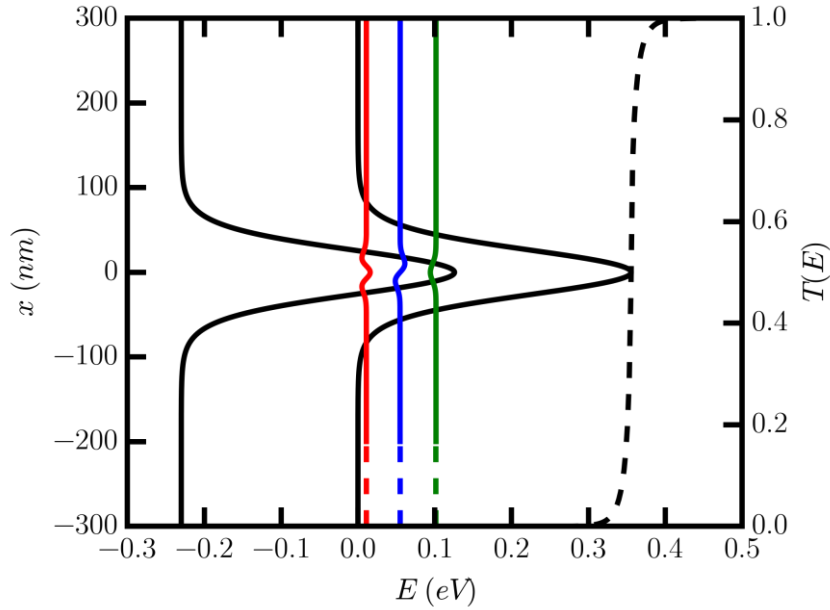
Figure 6.15 Band structure profile under a 40 nm gate with an applied voltage $V_{g1} = -2\ V$ (black, solid), with the first three light hole confined states (coloured). The direct transmission probability across the conduction band barrier is given by the dashed black line, where this is ≈ 0 for energies below the conduction band peak. At the energies of the confined states, the system is at resonance and the probability of transmission increases to ≈ 1.

6.4.5 Energy state separation and ground state energies

Examining the state separations shown in Figure 6.15, it can be seen that these are approximately evenly spaced. This is due to the approximately parabolic nature of the potential in the centre of the gate (i.e. the peak of the valence band). In a parabolic well, the energy is given by $E = (n + 1/2)\hbar\omega$, where for the ground state, $n = 0$. With a large separation between states ($\gg k_B T$), each individual state becomes distinct, and when considering quantised conductance in section 6.5, this will lead to distinct conductance levels. In these levels, the conductance value is proportional to the number of states around the Fermi energy. For a gate width of $2a = 40\ nm$, and no outer gate voltage, this separation is $\approx 5\ meV$ ($\approx 20\ k_B T$ at 3 K) for heavy holes, and $\approx 30\ meV$ for light holes. When outer gate voltages are applied, the potential well steepness is increased, increasing these separations. As these separations are significantly greater than $k_B T$ at low temperatures, there will be a clear demarcation between filled and empty states, giving sharp conductance steps.

The ground state energy can also be considered, with a difference in energy between the light hole and heavy hole states giving rise to Zener tunnelling into the heavy hole state at a lower gate voltage than the light hole state. For a narrow central gate width ($2a = 40\ nm$) and a V_{g1} voltage of $-2\ V$ (neglecting any outer gate voltages), the ground state energies for light and heavy holes are approximately 100 meV and 120 meV respectively. At 3 K, this gives an energy separation of $\approx 80\ k_B T$, significantly greater than the spread of carrier energies given by the Fermi distribution, and giving distinct separation between states. As the magnitude of V_{g1} is increased to push the regime into the Zener regime, initially electrons will tunnel into only heavy hole states. Increasing V_{g1} further, electrons will Zener tunnel will through ~3 to 4 heavy hole states (and giving regular conductance steps) before also tunnelling through the light hole ground state. This mixture of states will then give rise to irregularly spaced conductance steps.

6.5 Quantised Conductance

Applying a voltage to a bar gate constricts the 2-dimensional system. Initially the channel is completely blocked, stopping conduction. Then, as the valence band confined states become available, conduction can again occur, in a now 1-dimensional system. This transport is similar to that seen in quantum point contact split gates [50, 156, 157], where applying a voltage to two split gates separated by a small gap creates a 1-dimensional restricted channel (as shown in Figure 6.6). When a constriction is formed by these split gates, 1D transverse modes are formed in this constriction, and, depending on the number

of these modes available, quantised conductance is observed. A similar effect is expected in these bar gated structures with the quantised states formed in the valence band constriction.

Quantised conductance arises due to treatment of the system as ballistic, as described by the Landauer-Büttiker formalism. The following description primarily follows that of Chapter 6.1 of reference [50], and is an expansion on the description given in section 5.3. Now, the system concerned consists of two electron reservoirs (the 2DEG far from the gate) separated by a ballistic region (the region under the gates). An incident electron has a finite probability of transmission through the ballistic region, $T(E)$, whilst the transmission out of the ballistic region is considered to be unity. Finite resistance of the constriction occurs due to the large discrepancy between the number of states in the 2DEG reservoirs and the number of confined states in the ballistic region, resulting in a large reflection probability.

To calculate the conductance of a single confined state (energy E) at $0\ K$, the current must first be calculated, where this is given by

$$I = n_{1D}ev = e\int_0^\infty vg_{1D}[F_L(E, E_L) - F_R(E, E_R)]T(E)dE, \qquad (6.25)$$

where v is the velocity, g_{1D} is the 1D density of states and F_L and F_R are the Fermi-Dirac distributions for the two reservoirs, with Fermi levels E_{FL} and E_{FR} respectively.

Assuming parabolic dispersion, $v = \sqrt{2E/m^*}$ and $g_{1D} = (1/\pi\hbar)\sqrt{m^*/2E}$, meaning their product,

$$v \times g_{1D} = \frac{1}{\pi\hbar} = \frac{2}{h}, \qquad (6.26)$$

is independent of energy.

At $0\ K$, the Fermi distributions become step functions, with the difference in Fermi levels being proportional to applied voltage, and therefore the integral in equation (6.25) can be replaced such that

$$I = \frac{2e^2}{h}VT(E). \qquad (6.27)$$

The conductance, G, is then simply the change in current with applied voltage, giving

$$G = \frac{\partial I}{\partial V} = \frac{2e^2}{h} T(E). \qquad (6.28)$$

For a ballistic conductor, $T(E) = 1$, and all voltage dropped is dropped in the contacts. As stated previously, at the resonance condition the transmission probability raises to unity (a ballistic conductor) however, in principle, any asymmetry in the system will reduce the transmission value from this peak. However, for small applied longitudinal voltages this effect will be minimised, and so for simplicity will be neglected here. The effect of this assumption will be to reduce any measured current compared to that predicted, however this effect is expected to be small.

Equation (6.28) applies to a single conducting mode through the constriction, however if there are multiple modes below the Fermi energy (at $0\ K$), each mode will contribute $2e^2/h$ to the conductance, with the total conductance given by a sum over modes. At temperatures above $0\ K$, the effect of altered occupation of states must be considered via the Fermi-Dirac distribution. Now the total conductance is given by

$$G = \frac{2e^2}{h} \sum_N \frac{1}{\exp\left(-\frac{(E_F - E_N)}{k_B T}\right) + 1}, \qquad (6.29)$$

where E_N is the energy of state N. At $0\ K$ the Fermi-Dirac distribution is a step function, then as the applied gate voltage is increased, altering the number of states contributing to the conductance, sharp conductance steps are observed. As temperature is increased, these steps begin to smooth out until at high temperature, no quantisation is observed (the conductance becomes Ohmic). An example of the quantised conductance due to a modelled split gate structure as a function of gate voltage is shown in Figure 6.16.
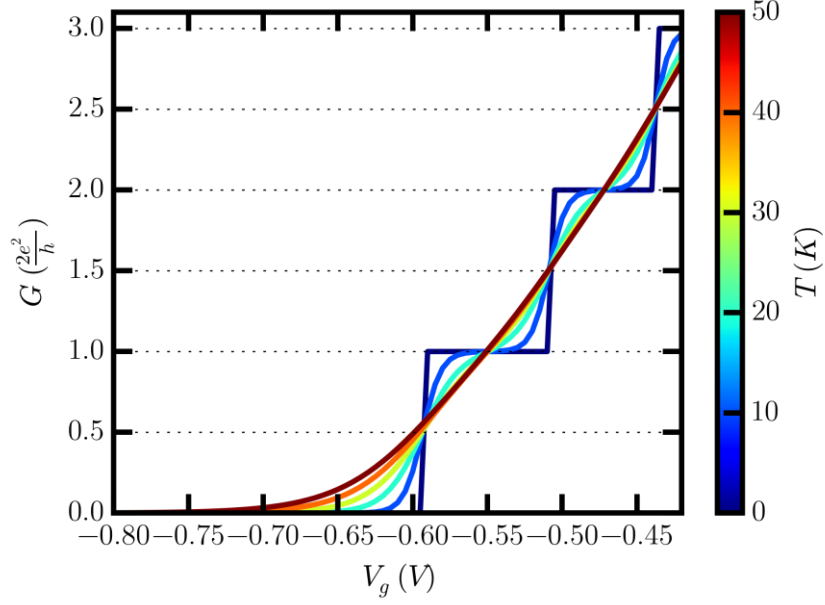
Figure 6.16 Example quantised conductance for a traditional split gate device as shown in Figure 6.6 (and as experimentally measured in e.g. reference [50]). As the magnitude of the gate voltage (V_g) is increased, the energy minima in the potential constriction is increased and the number of states below E_F able to conduct reduces. For low temperatures this decrease is stepwise due to the abrupt change in the Fermi distribution, whereas at higher temperatures these steps are smoothed out.

For quantised conductance in bar gated structures, the case is more complex. Now there are a finite number of states available in the energy range between the peak of the valence band (giving an upper state energy limit), the Fermi energy (giving the upper range of occupied electron states), and the conduction band edge (limiting $E > 0$). An example of the confining potentials at 3 distinct voltages, and the states that can contribute to conduction (at 0 K), are shown in Figure 6.17.
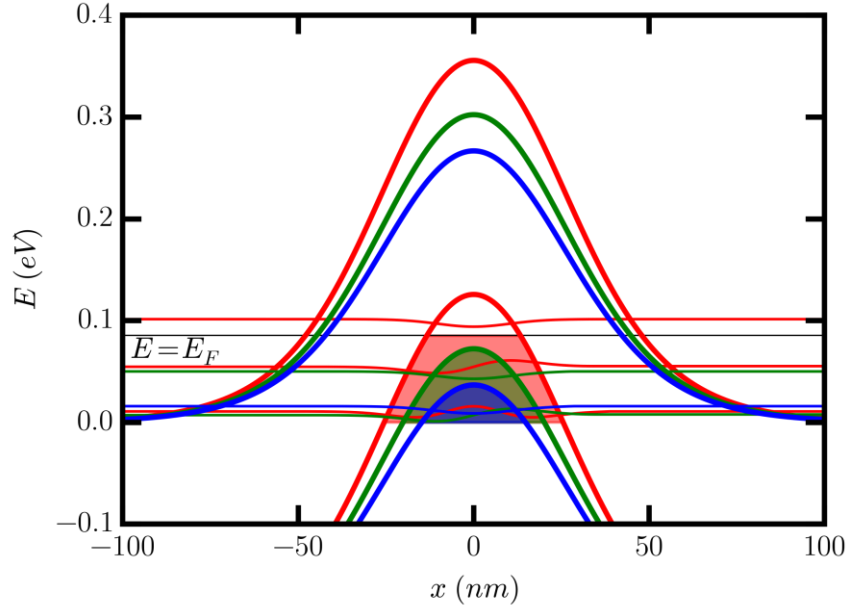
Figure 6.17 Example conduction band profile in the region of the central gate for 3 different V_{g1} values (thick lines), with confined light hole states (thin lines). The energy region in which Zener tunneling can occur for each V_{g1} at 0 K ($0 < E < E_F$) is shaded the corresponding colour. For the smallest V_{g1} (blue), only one state is accessable. For the next V_{g1} value (green), two states are now accessable. For the largest V_{g1} (red), still only two states are available as the ground state is above E_F.

As shown in Figure 6.17, as V_{g1} is increased and the system enters the Zener regime, the first states enter this conducting region and the first conductance steps are observed. Initially these states are all heavy hole states, giving rise to regularly spaced steps in energy and conductance. As the potential is increased further, light hole states become conducting, the result being irregularly spaced steps in the conductance. Eventually the potential is raised further and the highest states rise above the Fermi energy, where these can now only contribute via thermal excitation of incoming electrons. Now the confined states cover the full (finite) range of the energy scale where it is possible to conduct, and no more states are added with increasing voltage. As the voltage is increased and more states are raised above the Fermi energy, these are replaced by confined states passing above $E = 0$. This leads to a plateauing of conductance.

As well as increased voltage raising the potential, it simultaneously makes the well steeper, giving more widely spaced energy levels. It is therefore possible that the spacing of states can be increased such that a reduced number of states lie in the conducting region. This means that the number of states contributing to the conductance can actually decrease with increasing potential. This plateauing and fluctuation is shown in Figure 6.18 for the light and heavy hole states for 5 temperatures.
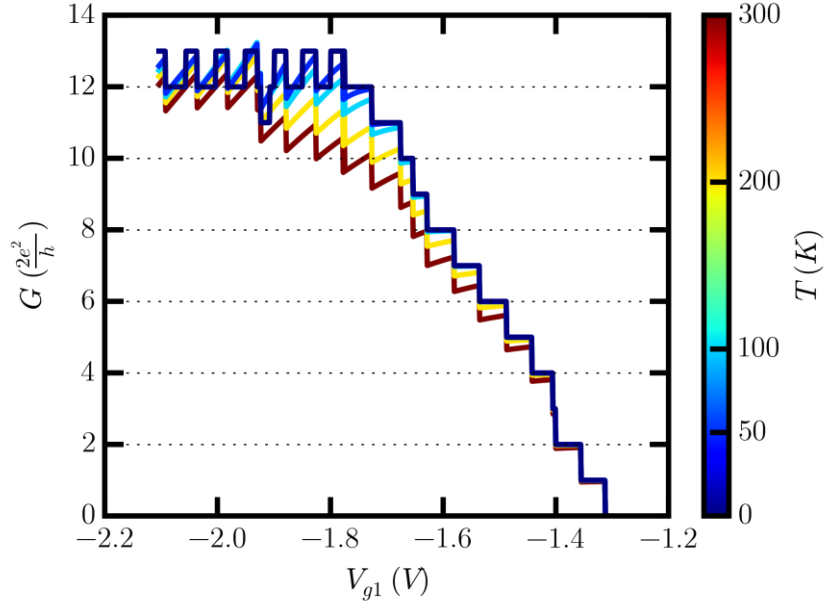
Figure 6.18 Conductance as a function of central gate voltage for a gate with width $2a = 40\ nm$, for various temperatures between $0\ K$ and $300\ K$ (where colour represents temperature). Initially regular steps are observed due to heavy hole states, with an irregular step at $V_g1 \approx -1.4\ V$ due to the first light hole state. At voltages of $\sim -1.8\ V$, plateauing of the conductance occurs due to the finite energy range over which tunnelling can occur ($0 < E < E_F$), where beyond this, oscilations in conductance occur due to the oscilating number of states in this range. For larger temperatures, decreased conductance is observed.

Figure 6.18 clearly shows the conductance features described previously, including regular and irregular steps, plateauing, and fluctuations in conductance. Examining the effect of raised temperature on the conductance, at low voltages the temperature dependence is weak for the states that are well below the Fermi energy. This is because the Fermi-Dirac distribution is approximately equal for all temperatures at these energies. The dependence only becomes significant for sate energies around E_F, where the conductance begins to plateaux. Around the Fermi energy, the Fermi-Dirac distribution changes from 1 to 0 over a range of $\sim 4\ k_B T$, where this is $\sim 1\ meV$ at $3\ K$ and $\sim 30\ meV$ at $100\ K$. This range gives rise to the more significant temperature dependence at lower voltages in Figure 6.18 for higher temperature curves.

In split gate quantised conductance, the Fermi-Dirac function acts to smooth out the conductance steps as occupation increases in higher states and decreases in lower ones (Figure 6.16). This is also true to an extent in bar gated structures, however, due to the limited number of states above E_F, at large temperatures, there are a significant proportion of electrons far enough above the highest state that they cannot tunnel and

conduct. This acts to reduce conductance. Eventually, with increasing voltage, the potential is raised high enough that the highest (ground) state energy increases far enough to allow these high energy electrons to reach more states, and so conductance is again raised. For further voltage increases, the temperature now only acts to smooth out the steps.

All of these factors combine to give a very interesting and unique conductance behaviour expected from these bar gated structures. Introducing a perpendicular magnetic field would then act to spin split these quantised states (Zeeman splitting), with this increasing the complexity of the quantised conductance further.

6.6 Conclusions

Modelling of triple bar gated structures has been performed, showing that it should be possible to use a triple bar gated device, exploiting the large Zeeman splitting of InSb, to achieve a spin dependant quantised conductance.

Modelling of the potential due to surface gated structures has been performed following the work of Davies et al. [151] to model the potential due to these triple bar gated structures. Using this, investigations were performed into the device dimensions, analysing the effect of gate widths, separations, and applied voltages on the operation of such devices. It was shown that the voltage required to enter the Zener regime is small ($\sim -1\ V$) due to the narrow band gap of InSb, with this value having only a slight dependence on the device dimensions. Following this, the effective barrier width and transmission probability were examined as a function of outer gate voltage and separation, where again, any increases in transmission probability were negligible, with a minimal probability of tunnelling across this barrier. To achieve tunnelling, a resonance condition is required between the states in the 2DEG and confined states in the valence band, where the large discrepancy between the effective masses of the light and heavy holes gives rise to a large energy separation between these confined states.

Due to the 1D nature of the valence band constriction, quantised conductance through these states is possible, where the conductance increases in steps of $2e^2/h$ for increasing gate voltages. Because of the difference in the ground state energies of the light and heavy hole states, initially conduction will only occur through the heavy hole states. Following this, light hole states will also conduct, giving rise to irregularly spaced steps. At larger gate voltages, confined state energies rise above E_F and stop conducting, giving rise to a plateau in conductance. With increased temperature, where the spread of the Fermi distribution gives carriers at higher energies, these carriers are at energies above the highest

energy states in the valence band constriction and so cannot conduct. This gives rise to a decreasing conductance with temperature.

7 : CONCLUSIONS AND FUTURE WORK

Transport in InSb/AlInSb QW heterostructures has been examined with the aim of understanding a range of phenomena observed.

For low carrier density samples, interface roughness scattering is dominant, whereas for high carrier densities, intersubband scattering becomes important. To explain the plateauing upper mobility limit observed at intermediate carrier densities, it has been shown that a previously unaccounted for scattering mechanism must be incorporated into the transport lifetime modelling performed. This surface feature related scattering is related to structural quality of the material, and is the cause of the upper mobility measured of $\approx 250,000 \ cm^2/Vs$. This scattering can be quantified through Nomarski optical image analysis, where surface features (hillocks) are present across all material imaged in this work, with an average diameter of $l_e = 2.43 \ \mu m$, giving the ratio of the average feature size to the largest mean free paths of $l_e/\lambda \approx 1$. Further, evidence of this scattering is present in detailed magnetoresistance measurements, where local minima give a feature size of $2.48 \ \mu m$. These measurements also give a background dopant density of $N_{bkg} = 2 \times 10^{13} \ cm^{-3}$, consistent with independent electrical measurements and self-consistent Schrödinger-Poisson modelling.

Scattering from these features has been modelled through use of Landauer-Büttiker theory and Monte Carlo modelling using a potential barrier model. Due to the screw-like growth around a central threading dislocation, the material within a feature is uniform, however at the boundaries potential barriers are formed. Treating these boundaries as back-to-back Schottky like barriers, a barrier pinned at $\approx 77\%$ of the mid-gap value is required ($\Phi' \approx 115 \ meV$) to match experimental values. Understanding this scattering mechanism, predictions can be made about possible mobility improvements given improvements in feature sizes. Also, predictions can be made as to the mobility of a sample through simple optical analysis, without requiring any processing, fabrication or measurement.

Following this, modelling of bar gated structures was performed, where these are designed to exploit the extreme parameters of InSb to attain quantised conductance with a spin polarised current. This is achieved through applying gate voltages to drive the system into the Zener regime, with conduction occurring via electrons tunnelling into

confined valence band states. In these devices, this quantised conduction should occur at gate voltages of $\approx -1.3\ V$, with irregular steps, plateauing, and decreasing conductance with increasing temperature predicted.

This work has been described in detail in chapters 4 to 6 respectively, with a full summary given in the following section. Following this, a brief description of preliminary measurements of bar gated structures is given, as well as proposals for future works, where these focus on possible increases in measured mobilities through improvements in material quality, as well as applications for this improved material, specifically in the bar gated structures.

7.1 Conclusions

Chapter 4 : Schrödinger-Poisson and Transport Modelling analysed the mobility trends observed in a batch of MBE grown QW samples with varying δ-doping levels. This was performed to determine the transport limiting scattering mechanisms in these InSb QWs as a function of sheet carrier density, with the overall aim of understanding how to increase the upper mobility limits observed. It was shown in this chapter that for low carrier density samples ($n_{2D} < 3 \times 10^{11}\ cm^2/Vs$), interface roughness scattering was the dominant scattering mechanism, with an increasing carrier density giving rise to increased screening, giving increased upper mobility limits. At moderately increased carrier densities ($3 \times 10^{11}\ cm^2/Vs < n_{2D} < 4 \times 10^{11}\ cm^2/Vs$), scattering related to surface features observed through Nomarski imaging was found to dominate, with this giving the overall upper mobility limit observed of $\approx 250,000\ cm^2/Vs$. At even higher carrier densities, multisubband filling and intersubband scattering dominate. These surface features are present on all material examined in this work, and have been observed previously on similar InSb based structures.

To determine these stated limiting factors, and so to understand the trends observed, the level of dopant for a given carrier density was investigated through self-consistent Schrödinger-Poisson modelling, accounting for dopant dragging in the top cap as determined through SIMS measurements. This modelling resulted in the determination of the relationship between dopant and free carriers, whilst also indicating a p-type background dopant density of $N_{bkg} \leq 5 \times 10^{14}\ cm^3$ is required to match the experimentally measured onset of multicarrier behaviour. This value for N_{bkg} is consistent with independent electrical measurements of AlInSb layers from the same growth run, indicating the relatively clean nature of these samples due to the low level background charge.

7 : Conclusions and Future Work

With the relationship between dopant levels and corresponding carrier density determined, a transport lifetime model was implemented to understand the measured mobilities as a function of temperature. This model included standard scattering mechanisms as have been examined previously, including acoustic and optical phonons, remote (dopant) and background charged impurities, and interface roughness. However, whilst similar to previously implemented models, this model used refined values (such as from the S.P. modelling) as compared to the extreme values used previously, giving more realistic mobility predictions. Applying this model to a sample with a $3\ K$ carrier density of $n_{2D} = 2.14 \times 10^{11}\ cm^{-2}$ and mobility $\mu \approx 200{,}000\ cm^2/Vs$, the predicted mobilities were significantly higher than those experimentally measured. Extreme values are required to compensate for the difference, with this inconsistent with previous observations. This modelling then clearly indicates that the mobility limiting mechanisms are not understood, and a previously unaccounted for scattering mechanism is therefore required.

To investigate this large discrepancy between the predicted and measured mobility, optical analysis of the unprocessed wafer surfaces was performed using differential interference (Nomarski) microscopy. This imaging showed a clear surface roughness present on all material imaged, where these surface features (hillocks) consist of approximately circular features with well-defined boundaries. Through use of image analysis techniques, average surface feature sizes were extracted, where these had an average diameter of $l_e = 2.43 \pm 0.13\ \mu m$, remarkably comparable with the largest electrical mean free paths measured, with a ratio of $l_e/\lambda \approx 1$. Using a simple Drude relationship, this diameter was turned into an average lifetime, where incorporation into the previous transport lifetime model, and using reasonable values, gave an excellent agreement between the predicted and measured mobilities for the full range of temperatures from $3\ K$ to $300\ K$.

These features (hillocks) limiting the mobility are related to the structural quality of the sample, where screw like growth occurs around threading dislocations that form at the substrate/epilayer boundaries, and propagate through the sample to the surface. To achieve increases in mobility, decreases in hillock densities must be achieved, where to do this would require a buffer redesign. To achieve this aim, investigations in similar structures have already been performed, where it has been shown that reduction in defect densities can be achieved using interlayers or superlattice buffers. This work however has not examined the electrical effect of such buffers, with this being an obvious area for future exploration.

With scattering related to these surface features shown to be the limiting scattering mechanism in the highest mobility samples, Chapter 5 : Monte Carlo Modelling and Magnetoresistance, extended this analysis to determine the nature of scattering from these surface feature boundaries. Following this, the upper mobility limits possible across the full range of sample carrier densities was considered.

Due to the nature of the formation of these surface features as screw-like growth around a threading dislocation, the material within a feature can be considered to be uniform, and so ballistic with regards to the surface feature related scattering. The same cannot be said however for the boundaries between features, where the regularity of the crystal no longer holds. This irregularity is treated as creating a potential barrier for electrons propagating through the QW.

To investigate the range of potential barrier shapes possible, a combination of Landauer-Büttiker theory and Monte Carlo modelling was used to determine a tunnelling current, and corresponding mobility, with these compared to experimental values. A simple treatment of these barriers investigated rectangular shaped barriers, with an average separation given by l_e. The transmission probability as a function of energy ($T(E)$) was calculated numerically, with this used to calculate a tunnelling current for a given applied voltage, whilst also being used in the MC model to determine an average mobility. These values were determined for two samples for barrier widths ranging between 1 and 50 monolayers, where barrier heights were found to follow the approximate relationship height $(eV) \approx$ width $(ML)^{-1.3} \times 10$. When included in the previous transport lifetime modelling, this surface feature potential barrier model again gave excellent agreement between the measured and predicted values, indicating the role of scattering related to these features in limiting the highest mobilities achieved.

Expanding the model to include variations between barriers, in the form of varying heights, widths and separations gave no significant change in the mobility, with only the average values determining the mobility achieved. Further refinements to this model included the implementation of more realistic back-to-back Schottky-like barriers, as well as an analytic derivation for the simulated MC mobility, with excellent agreement achieved between the two.

To further examine the scattering mechanisms present in these InSb QW samples, detailed magnetoresistance measurements were performed. These measurements showed clear local minima in the low field resistance values present across a range of temperatures, where the location of these minima correspond to characteristic scattering sizes. Extrapolation of these minima to 0 K gives a surface feature size of 2.48 μm and a

background dopant density of $2 \times 10^{13}~cm^{-3}$. Both of these values are in excellent agreement with the independently measured values of $2.43~\mu m$ from Nomarski imaging, and $N_{bkg} \leq 5 \times 10^{14}~cm^{-3}$ from S.P. modelling and electrical flat layer measurements.

To determine the height of the back-to-back Schottky-like barriers between surface features, a mid-gap pinned value was assumed, as well as a reduced height barrier, and the MC model simulations compared to experimental results. Examining the dependence on barrier height, it was shown that the value of current predicted from the model is sensitive to the barrier height input, whilst the mobility has a weaker dependence. The barrier height required to achieve results consistent with the experimental values was a back-to-back Schottky-like barrier pinned at ~77% of the mid-gap value ($\Phi' \approx 115~meV$). Applying this model to determine the limiting scattering mechanisms across the range of single carriers, it was shown that at lower carrier densities, the interface roughness scattering is dominant, with increased screening with increased carrier density leading to increased mobilities. As carrier densities increase further, scattering from the potential barriers surrounding the screw-like growth that is visible at the surface as hillocks begins to dominate. This gives the overall upper mobility observed of $\approx 250,000~cm^2/Vs$. For carrier densities above ~$4 \times 10^{11}~cm^{-2}$, significant 2 carrier behaviour is observed, where mobilities are then limited by intersubband scattering.

To improve on these upper mobility limits, increases in these feature sizes must be realised through buffer redesigns, with corresponding large increases in mobility possible.

To examine the advanced devices possible with this high quality material, triple bar gated structures were modelled, fabricated and measured, as described in Chapter 6 : Gated Structure Modelling. These devices are designed to operate through use of Zener tunnelling to create a quantised current, where application of a transverse magnetic field can further spin polarise this current through Zeeman splitting due to the large Landé g-factor of InSb.

These devices consist of three parallel Schottky surface gates, with the potential profile due to bias of the gates modelled to examine the operation of such devices. It was shown that only moderate voltages ($\sim -1~V$) are required to enter the Zener regime, though resultant barrier widths are large and corresponding transmission probability is low. Due to this, a current can only be achieved when tunnelling occurs via intermediate aligned confined 1D hole states, where the system is then at resonance, and where the conductance is quantised in steps of $2e^2/h$ for increasing gate voltages.

These confined hole states consist of both light and heavy hole states, with a large effective mass discrepancy between the two. This large discrepancy gives rise to large

differences between ground state energies, and energy level separations, with this giving interesting consequences for the quantised conductance, such as irregularly spaced steps. Further, as conduction can only occur between the energies of the conduction band edge and the Fermi energy (at $0\ K$), at larger gate voltages, confined state energies rise above E_F and stop conducting, giving rise to a plateau in conductance. At increased temperatures, where the Fermi energy is no longer a distinct step function, carriers at higher energies can conduct, acting to smooth out the distinct conductance steps. However, at low gate voltages these higher energy electrons are above all the confined valence band states and so cannot conduct, reducing the observed conductance.

These bar gated devices are therefore predicted to have interesting and unique conductance behaviour, with the introduction of a perpendicular magnetic field further complicating the conductance. This magnetic field would act to spin split the confined states, potentially giving rise to a spin polarised tunnelling current.

7.2 Future Work

Preliminary work has been performed examining the fabrication of bar gated devices through the use of e-beam lithography. This was primarily aimed at examining the minimum gate widths and separations possible through varying the dosage applied for the inner and outer gates separately. An image showing a test pattern for various dosages and separations is shown in Figure 7.1.

Examining the gate patterns produced, only gate separations of 175 nm and 200 nm resulted in three individual gates that did not merge. The gate widths in these patterns are approximately 40 nm, the minimum realistic single pixel line width achievable using this e-beam lithography system that can be metallised as robust gates. An example of these fabricated gates is shown in Figure 7.2.
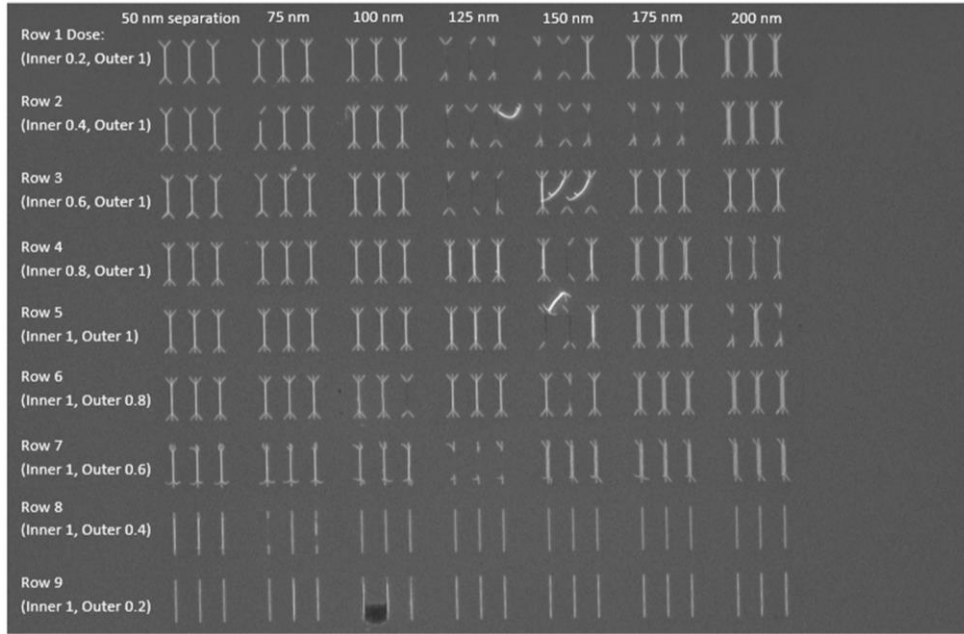
Figure 7.1 Triple gate e-beam lithography test patterns, with varying gate separations (columns) and e-beam dosages (rows). The various patterns are used to find the optimum parameters to ensure reliable gate fabrication with the minimum separations possible, ensuring gates do not merge.
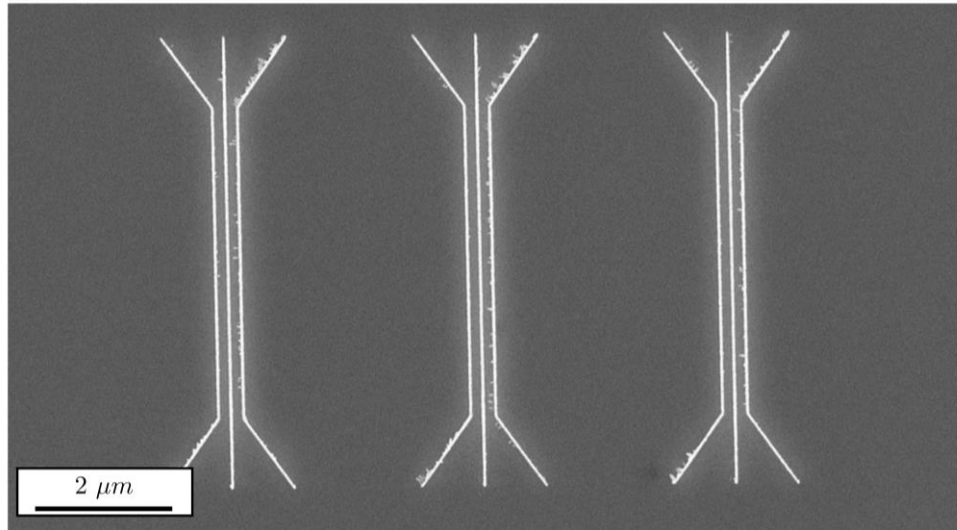
Figure 7.2 Magnified triple bar gate test pattern showing well defined patterns with clear separations (i.e. not merged). The gate widths are ~40 nm whilst the minimum separations are ~200 nm.

Measurements were then performed (described in section 6.2.1) where a range of longitudinal voltages were applied and the corresponding current measured, with the conductance determined via a linear least squares fit. This was repeated for increasing gate voltages, with the leakage current also monitored. A selection of typical results are shown in Figure 7.3 for a temperature of 3 K.
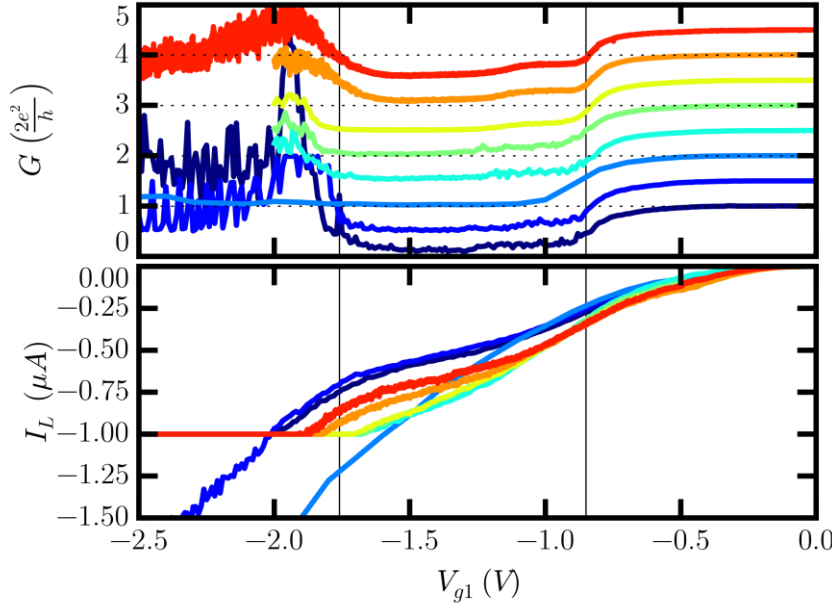
Figure 7.3 Conductance (G, top) and leakage current (I_L, bottom) as a function of gate voltage (V_{g1}) for several bar gated devices. The conductance has been normalised to give a 0 V value of 1, with each measurement offset vertically by a value of 0.5 for clarity. Vertical lines denote the approximate region of channel pich off where (almost) no conduction occurs. At higher gate voltages, conductance begins to increase, as predicted, however no quantisation is observed. Here, the noise also increases rapidly, whilt simultaneously the leakage currents become large, and comparable to the longitudinal current. Therefore, the origin of this inreasing conductance is inconclusive.

For the results shown in Figure 7.3, clear channel modulation and pinch off is observed, where this occurs at gate voltages of $V_{g1} \approx -0.85\ V$. This is reasonably comparable to the value of $\sim -0.3\ V$ predicted previously. Increases in conductance then occur again at $V_{g1} \approx -1.75\ V$, comparable to the $-1.3\ V$ predicted for the onset of quantised conductance as shown in Figure 6.18. These two experimental values are both separated from the predicted values by ~0.5 V, consistent with the assumption that the differences in charge distribution between these structures and those the model was originally derived for would result in an approximately constant voltage offset.

Quantised conductance is not observed however, as whilst the conductance increases, the leakage current through the gates simultaneously increases. This acts to reduce the effective modulation of the gate and increase noise on the measurement. Further increases in gate voltage then only act to exponentially increase current flow through the gate, leading to the eventual breakdown of the gate. As the leakage currents become comparable

to the longitudinal current, and due to the increased noise, the origin of this inreasing conductance is inconclusive.

These initial results show promise for the possibility of observing quantised conductance in a bar gated devices, however difficulties remain, namely the fabrication of good quality Schottky contacts. These are required to reduce the leakage current for a given bias on a gate, and so reducing the noise created in the longitudinal resistance measurements. Given improvements in these Schottky characteristics, and increased reliability through the fabrication process, increased quality results should be possible. Recent work has already been undertaken with this aim, investigating the characteristics of AlInSb/InSb QW Schottky diodes [148].

With these improvements, investigations into applied transverse magnetic fields can be undertaken, with the aim of examining the Zeeman splitting, achieving spin dependant tunnelling and a spin polarised current. Similarly, the effect of Rashba splitting on such a device could be examined, where increased knowledge of the Rashba coefficients in InSb could be used in other, spin dependant devices.

Also, whilst the operation of these devices is not strictly dependant on the mobility, improvements in material quality will be beneficial. Namely, reductions in defect densities should aid in device reliability, and increased scattering lifetimes will lead to less scattering in what was considered the ballistic region for the quantised conductance derivation. Simultaneously, this will also give rise to a reduced chance of scattering after an electron has tunnelled into a hole state. Scattering of an electron that has tunnelled into a hole state would act to destroy the resonance condition, drastically reducing the probability of transmission. Improved mobilities would give rise to longer times between these scattering events, allowing for a longer range of time for an electron to tunnel out of this state.

Given these improvements in sample and gate qualities, the operation of a bar gated device such as that described here, observing quantised and spin dependant conductance, is entirely feasible.

Further, increased material quality and the associated higher mobility is highly advantageous for many potential device applications, including low power high frequency electronics, as well as the possibility of the realisation of Majorana Fermions, generated in a solid state material system where advanced planar fabrication can be utilized. InSb has already attracted a lot of attention in these areas due to its light effective mass, with the large Landé g-factor also attracting attention for potential exploitation in spintronics and quantum information control. Further improvements in the material system via buffer redesigns, including the possibility of graded, interlayer, or superlattice buffers giving

increased mobilities, will make this material system even more attractive for these purposes. Work has already been performed with the aim of achieving increased material quality, though the effect of these buffers on the electrical quality has yet to be confirmed [99, 100, 101].

8 : REFERENCES

[1]   W. F. Brinkman, D. E. Haggan and W. W. Troutman, *"A History of the Invention of the Transistor and Where It Will Lead Us,"* IEEE Journal of Solid-State Circuits, *vol. 32, no. 12, pp. 1858-1865, 1997.*

[2]   G. E. Moore, *"Cramming more components onto integrated circuits,"* Electronics, *vol. 38, no. 8, 1965.*

[3]   Y. Kim, C. Pham and J. P. Chang, *"Potentials and challenges of integration for complex metal oxides in CMOS devices and beyond,"* Journal of Physics D: Applied Physics, *vol. 48, no. 6, 2015.*

[4]   L. B. Kish, *"End of Moore's law: thermal (noise) death of integration in micro and nano electronics,"* Physics Letters A, *vol. 305, pp. 144-149, 2002.*

[5]   T. N. Theis and H.-S. P. Wong, *"The End of Moore's Law: A New Beginning for Information Technology,"* Computing in Science & Engineering, *vol. 19, no. 2, pp. 41-50, 2017.*

[6]   H. N. Khan, D. A. Hounshell and E. R. H. Fuchs, *"Science and research policy at the end of Moore's law,"* Nature Electronics, *vol. 1, pp. 14-21, 2018.*

[7]   M. Schulz, *"The end of the road for silicon?,"* Nature, *vol. 399, pp. 729-730, 1999.*

[8]   M. Cooke, *"Digital III-Vs review,"* III-Vs Review, *vol. 19, no. 3, pp. 18-22, 2006.*

[9]   Nobel Media, *"The Nobel Prize in Physics 2000,"* *[Online]. Available: https://www.nobelprize.org/prizes/physics/2000/press-release/.*

[10]   IOFFE Physico-Technical Institute, St. Petersburg, Rusian Federation, *"Physical Properties of Semiconductors, New Semiconductor Materials. Characteristics and Properties,"* *[Online]. Available: http://www.ioffe.ru/SVA/NSM/Semicond/.*

[11]   O. Madelung, U. Rossler and M. Schulz, Landolt and Bornstein Group III Condensed matter: Group IV Elements, (IV-IV) and (III-V) Compounds. Part b Electronic, Transport, Optical and Other, Berlin: Springer, 2006.*

[12] S. Adachi, *Properties of Group-IV, III–V and II–VI Semiconductors, John Wiley & Sons, Ltd, 2205.*

[13] C. Kittel, *Introduction to Solid State Physics, John Wiley & Sons, Inc, 2005.*

[14] B. Nediniyom, R. J. Nicholas, M. T. Emeny, L. Buckle, A. M. Gilbertson, P. D. Buckle and T. Ashley, *"Giant enhanced g-factors in an InSb two-dimensional gas,"* Physical Review B, *vol. 80, p. 125328, 2009.*

[15] W. Yi, A. A. Kiselev, J. Thorp, R. Noah, B.-M. Nguyen, S. Bui, R. D. Rajavel, T. Hussain, M. F. Gyure, P. Kratz, Q. Qian, M. J. Manfra, V. S. Pribiag, L. P. Kouwenhoven, C. M. Marcus and M. Sokolich, *"Gate-tunable high mobility remote-doped InSb/In1-xAlxSb quantum well heterostructures,"* Applied Physics Letters, *vol. 106, p. 142103, 2015.*

[16] O. J. Pooley, A. M. Gilbertson, P. D. Buckle, R. S. Hall, M. T. Emeny, M. Fearn, M. P. Halsall, L. F. Cohen and T. Ashley, *"Quantum well mobility and the effect of gate dielectrics in remote doped InSb/AlxIn1-xSb heterostructures,"* Semiconductor Science and Technology, *vol. 25, no. 12, 2010.*

[17] J. M. S. Orr, A. M. Gilbertson, M. Fearn, O. W. Croad, C. J. Storey, L. Buckle, M. T. Emeny, P. D. Buckle and T. Ashley, *"Electronic transport in modulation-doped InSb quantum well heterostructures,"* Physical Review B, *vol. 77, p. 165334, 2008.*

[18] F. Qu, J. v. Veen, F. K. de Vries, A. J. A. Beukman, M. Wimmer, W. Yi, A. A. Kiselev, B.-M. Nguyen, M. Sokoloich, M. J. Manfra, F. Nichele, C. M. Marcus and L. P. Kouwenhoven, *"Quantized Conductance and Large g-Factor Anisotropy in InSb Quantum Point Contacts,"* Nano Letters, *vol. 2016, no. 16, p. 7509–7513, 2016.*

[19] J. W. G. van den Berg, S. Nadj-Perge, V. S. Pribiag, S. R. Plissard, E. P. A. M. Bakkers, S. M. Frolov and L. P. Kouwenhoven, *"Fast Spin-Orbit Qubit in an Indium Antimonide Nanowire,"* Physical Review Letters, *vol. 110, p. 066806, 2013.*

[20] R. Li, J. Q. You, C. P. Sun and F. Nori, *"Controlling a nanowire spin-orbit qubit via electric-dipole spin resonance,"* Physical Review Letters, *vol. 111, no. 8, p. 086805, 2013.*

[21] D. D. Awschalom, L. C. Bassett, A. A. Dzurak, E. L. Hu and J. R. Petta, *"Quantum Spintronics: Engineering and Manipulating Atom-Like Spins in Semiconductors,"* Science, *vol. 339, no. 6124, pp. 1174-1179, 2013.*

[22] *T. W. Kim, B. S. Yoo, M. A. McKee and J. Y. Lee, "Structural and electrical properties of a strained InSb/GaAs heterostructure,"* Physica Status Solidi (a), *vol. 142, no. 1, pp. K23-K27, 1994.*

[23] *M. A. Ball, J. C. Keay, S. J. Chung, M. B. Santos and M. B. Johnson, "Mobility anisotropy in InSb/AlxIn1-xSb single quantum wells,"* Applied Physics Letters, *vol. 80, p. 2138, 2002.*

[24] *T. D. Mishima, J. C. Keay, N. Goel, M. A. Ball, S. J. Chung, M. B. Johnson and M. B. Santos, "Anisotropic structural and electronic properties of InSb/AlxIn1-xSbquantum wells grown on GaAs (001) substrates,"* Journal of Crystal Growth, *vol. 251, pp. 551-555, 2003.*

[25] *T. D. Mishima, J. C. Keay, N. Goel, M. A. Ball, S. J. Chung, M. B. Johnson and M. B. Santos, "Effect of structural defects onIn InSb/AlxIn1−xSb quantum wells grown on GaAs (001) substrates,"* Physica E, *vol. 200, pp. 260-263, 2004.*

[26] *T. D. Mishima, J. C. Keay, N. Goel, M. A. Ball, S. J. Chung, M. B. Johnson and M. B. Santos, "Effect of micro-twin defects on InSb quantum wells,"* Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, *vol. 23, p. 1171, 2005.*

[27] *M. Edirisooriya, T. D. Mishima and M. B. Santos, "Effect of Al composition on filtering of threading dislocations by AlxIn1-xSb/AlyIn1-ySb heterostructures grown on GaAs (001),"* Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, *vol. 25, p. 1063, 2007.*

[28] *T. D. Mishima, M. Edirisooriya and M. B. Santos, "Reduction of microtwin defects for high-electron-mobility InSb quantum wells,"* Applied Physics Letters, *vol. 91, p. 062106, 2007.*

[29] *D. G. Hayes, C. P. Allford, G. V. Smith, C. McIndo, L. A. Hanks, A. M. Gilbertson, L. F. Cohen, S. Zhang, E. M. Clarker and P. D. Buckle, "Electron transport lifetimes in InSb/Al1-x In x Sb quantum well 2DEGs,"* Semiconductor Science and Technology, *vol. 32, p. 085002, 2017.*

[30] *C. J. McIndo, D. G. Hayes, A. Papageorgiou, L. A. Hanks, G. V. Smith, C. P. Allford, S. Zhang, E. M. Clarke and P. D. Buckle, "Determination of the transport lifetime limiting scattering rate in InSb/AlxIn1−x Sb quantum wells using optical surface microscopy,"* Physica E, *vol. 91, pp. 169-172, 2017.*

[31] C. J. McIndo, D. G. Hayes, A. Papageorgiou, L. A. Hanks, G. V. Smith, C. P. Allford, S. Zhang, E. M. Clarke and P. D. Buckle, *"Optical Microscopy as a probe of the rate limiting transport lifetime in InSb/Al1-xInxSb quantum wells,"* Journal of Physics: Conference Series, *vol. 964, p. 012005, 2018.*

[32] C. Zener and R. H. Fowler, *"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 Charachter, *vol. 145, no. 885, 1934.*

[33] Q. Zhang, W. Zhao and A. Seabaugh, *"Low-subthreshold-swing tunnel transistors,"* IEEE Electron Device Letters, *vol. 27, no. 4, pp. 297-300, 2006.*

[34] P.-F. Wang, T. Nirschl, D. Scmitt-Landsiedel and W. Hansch, *"Simulation of the Esaki-tunneling FET,"* Solid-State Electronics, *vol. 47, no. 7, pp. 1187-1192, 2003.*

[35] P.-F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel and W. Hansch, *"Complementary tunneling transistor for low power application,"* Solid-State Electronics, *vol. 48, no. 12, pp. 2281-2286, 2004.*

[36] T. S. Plaskett, J. M. Woodall and A. Segmuller, *"The Effect of Growth Orientation on the Crystal Perfection of Horizontal Bridgman Grown GaAs,"* Journal of The Electrochemical Society, *vol. 118, no. 1, pp. 115-117, 1971.*

[37] J. R. Arthur Jr., *"Interaction of Ga and As2 Molecular Beams with GaAs Surfaces,"* Journal of Applied Physics, *vol. 39, p. 4032, 1968.*

[38] W. P. McCray, *"MBE deserves a place in the history books,"* Nature Nanotechnology, *vol. 2, pp. 259-261, 2007.*

[39] L. Pfeiffer and K. W. West, *"The role of MBE in recent quantum Hall effect physics discoveries,"* Physica E, *vol. 20, pp. 57-64, 2003.*

[40] G. C. Gardner, S. Fallahi, J. D. Watson and M. J. Manfra, *"Modified MBE hardware and techniques and role of gallium purity for attainment of two dimensional electron gas mobility >35×106 cm2/Vs in AlGaAs/GaAs quantum wells grown by MBE,"* Journal of Crystal Growth, *vol. 441, pp. 71-77, 2016.*

[41] E. H. Putley, The Hall Effect and Semi-Conductor Physics, Dover, 1968.

[42] E. H. Putley, The Hall Effect and Related Phenomena, Butterworth, 1960.

[43] T. J. Phillips, *"High performance thermal imaging technology,"* III-Vs Review, The Advanced Semiconductor Magazine, *vol. 15, no. 7, 2002.*

8 : References

[44] T. Ashley, T. M. Burke, M. T. Emeny, N. T. Gordon, D. J. Hall, D. J. Lees, J. C. Little and D. Milner, "Epitaxial InSb for elevated temperature operation of large IR focal plane arrays," in Proceedings Volume 5074, Infrared Technology and Applications XXIX, 2003.

[45] J. G. Crowder, H. R. Hardaway and C. T. Elliott, "Mid-infrared gas detection using optically immersed, room-temperature, semiconductor devices," Measurement Science and Technology, vol. 13, no. 6, p. 882–884, 2002.

[46] K. L. Litvinenko, B. N. Murdin, J. Allam, C. R. Pidgeon, M. Bird, K. Morris, W. Branford and S. K. Clowes, "Spin relaxation in n-InSb/AlInSb quantum wells," New Journal of Physics, vol. 8, p. 49, 2006.

[47] J. J. Harris, T. Zhang, W. R. Branford, S. K. Clowes, M. Debnath, A. Bennett, C. Roberts and L. F. Cohen, "The role of impurity band conduction in the low temperature characteristics of thin InSb film," Semiconductor Science and Technology, vol. 19, pp. 1406-1410, 2004.

[48] J. M. S. Orr, "Low-Dimensional Transport Measurements in InSb Heterostructure Devices," PhD Thesis, 2007.

[49] A. M. Gilbertson, "Spin and Magnetotransport Properties of Narrow Gap Semiconductors," PhD Thesis, 2009.

[50] O. J. Pooley, "Electron Transport in InSb/AlInSb Semiconductor Heterostructures," PhD Thesis, 2011.

[51] S. J. Chung, M. A. Ball, S. C. Lindstrom, M. B. Johnson and M. B. Santos, "Improving the surface morphology of InSb quantum-well structures on GaAs," Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, vol. 18, p. 1583, 2000.

[52] T. D. Mishima, M. Edirisooriya, N. Goel and M. B. Santos, "Dislocation filtering by AlxIn1-xSb/AlyIn1-ySb interfaces for InSb-based devices grown on GaAs (001) substrates," Applied Physics Letters, vol. 88, p. 191908, 2006.

[53] T. D. Mishima, M. Edirisooriya and M. B. Santos, "Micro-twin induced structural misalignment in InSb quantum wells," Physica Status Solidi C, vol. 5, no. 9, pp. 2775-2777, 2008.

[54] Y. Fu and M. Willander, "Elemental and compound semiconductors," in Physical Models of Semiconductor Quantum Devices. Electronic Materials Series, vol 5, Boston, Springer, 1999.

[55]   F. Shimura, *"Single-Crystal Silicon: Growth and Properties," in* Springer Handbook of Electronic and Photonic Materials*, Springer, Cham, 2017.*

[56]   R. W. Keyes, *"Physical limits of silicon transistors and circuits,"* Reports on Progress in Physics, *vol. 68, no. 12, pp. 2701-2746, 2005.*

[57]   *P. Hoffman, Solid State Physics, An Introduction, Wiley-VCH, 2008.*

[58]   *N. W. Ashcroft and N. D. Mermin, Solid State Physics, Brooks Cole, 1976.*

[59]   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, no. 2, p. 556, 1976.*

[60]   *J. Singleton, Band Theory and Electronic Properties of Solids, Oxford University Press, 2001.*

[61]   E. O. Kane, *"Band structure of narrow gap semiconductors," in* Narrow Gap Semiconductors Physics and Applications*, vol. 133, Springer-Verlag Berlin Heidelberg, 1980, pp. 13-31.*

[62]   *E. F. Schubert, Light-Emitting Diodes, Cambridge University Press, 2003.*

[63]   *J. H. Davies, The Physics of Low-Dimensional Semiconductors: An Introduction, Cambridge University Press, 1998.*

[64]   E. O. Kane, *"Band structure of indium antimonide,"* Journal of Physics and Chemistry of Solids, *vol. 1, no. 4, pp. 249-261, 1957.*

[65]   P. Pfeffer and W. Zawadzki, *"Conduction electrons in GaAs: Five-level k.p theory and polaron effects,"* Physical Review B, *vol. 41, p. 1561, 1990.*

[66]   J. P. Stanley, *"Theory and Modelling of Semiconductor Spintronic Heterostructures,"* PhD Thesis, *2004.*

[67]   *M. Abramowitz and I. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, New York: Dover Publications, Inc., 1964.*

[68]   L. Vegard, *"The constitution of the mixed crystals and the space filling of the atoms,"* Physics, *vol. 5, no. 1, pp. 17-26, 1921.*

[69]   I. Vurgaftman and J. R. Meyer, *"Band parameters for III–V compound semiconductors and their alloys,"* Journal of Applied Physics, *vol. 89, p. 5815, 2001.*

[70] N. Dai, F. Brown, R. E. Doezema, S. J. Chung, K. J. Goldammer and M. B. Santos, "Determination of the concentration and temperature dependence of the fundamental energy gap in AlxIn1-xSb," Applied Physics Letters, vol. 73, p. 3132, 1998.

[71] R. L. Anderson, "Germanium-Gallium Arsenide Heterojunctions [Letter to the Editor]," IBM Journal of Research and Development, vol. 4, no. 3, pp. 283 - 287, 1960.

[72] N. Dai, G. A. Khodaparest, F. Brown, R. E. Doezema, S. J. Chung and M. B. Santos, "Band offset determination in the strained-layer InSb/AlxIn1-xSb system," Applied Physics Letters, vol. 76, p. 3905, 2000.

[73] Y. Hinuma, A. Gruneis, G. Kresse and F. Oba, "Band alignment of semiconductors from density-functional theory and many-body perturbation theory," Physical Review B, vol. 90, p. 155405, 2014.

[74] I.-H. Tan, G. L. Snider, L. D. Chang and E. L. Hu, "A self-consistent solution of Schrödinger–Poisson equations using a nonuniform mesh," Journal of Applied Physics, vol. 68, p. 4071, 1990.

[75] O. J. Pooley, A. M. Gilbertson, P. D. Buckle, R. S. Hall, L. Buckle, M. T. Emeny, M. Fearn, L. F. Cohen and T. Ashley, "Transport effects in remote-doped InSb/AlxIn1-xSb heterostructures," New Journal of Physics, vol. 12, p. 053022, 2010.

[76] A. M. Gilbertson, W. R. Branford, M. Fearn, L. Buckle, P. D. Buckle, T. Ashley and L. F. Cohen, "Zero-field spin splitting and spin-dependent broadening in high-mobility InSb/In1-xAlxSb asymmetric quantum well heterostructures," Physical Review B, vol. 79, p. 235333, 2009.

[77] R. A. Stradling and P. C. Klipstein, Growth and Characterisation of Semiconductors, Adam Hilger imprint by IOP Publishing Ltd, 1990.

[78] D. L. Partin, M. Pelczynski, P. Cooke, L. Green, J. Heremans and C. M. Thrush, "The influence of stoichiometry on the growth of tellurium-doped indium antimonide for magnetic field sensors," Journal of Crystal Growth, vol. 195, no. 1-4, pp. 378-384, 1998.

[79] T. Zhang, J. J. Harris, S. K. Clowes, M. Debnath, A. Bennett, L. F. Cohen, T. Lyford and P. F. Fewster, "Evidence for dislocation-related amphoteric behaviour of Si dopant in high-mobility InSb thin films," Semiconductor Science and Technology, vol. 20, no. 12, p. 1153–1156, 2005.

[80] J. J. Harris, "Delta-doping of semiconductors," Journal of Materials Science: Materials in Electronics, *vol. 4, no. 2, pp. 93-105, 1993.*

[81] E. F. Schubert, "Delta doping of III–V compound semiconductors: Fundamentals and device applications," Journal of Vacuum Science & Technology A, *vol. 8, p. 2980, 1990.*

[82] J. J. Harris, "Review: Delta-doping of semiconductors," Journal of Materials Science: Materials in Electronics, *vol. 4, pp. 93-105, 1993.*

[83] A. H. Etgens, M. Sauvage-Simkin, R. Pinchaux, J. Massies, N. Jedrecy, A. Waldhauer, S. Tatarenko and P. H. Jouneau, "ZnTe/GaAs(001): Growth mode and strain evolution during the early stages of molecular-beam-epitaxy heteroepitaxial growth," Physical Review B, *vol. 47, no. 16, 1993.*

[84] C. A. Lehner, T. Tschirky, T. Ihn, W. Dietsche, J. Keller, S. Falt and W. Wegscheider, "Limiting scattering processes in high-mobility InSb quantum wells grown on GaSb buffer systems," Physical Review Materials, *vol. 2, p. 054601, 2018.*

[85] L. J. Nash, "Growth and Characterisation of Terrace Graded Virtual Substrates with Si1-xGex 0.15 ≤ x ≤ 1.," PhD Thesis, *2005.*

[86] W. T. Read Jr., Dislocations in Crystals, McGraw-Hill Book Company, Inc., *1953.*

[87] S. M. Hu, "Misfit dislocations and critical thickness of heteroepitaxy," Journal of Applied Physics, *vol. 69, p. 7901, 1991.*

[88] P. M. J. Marée, J. C. Barbour, J. F. van der Veen, K. L. Kavanagh, C. W. T. Bulle-Lieuwma and M. P. A. Viegers, "Generation of misfit dislocations in semiconductors," Journal of Applied Physics, *vol. 62, p. 4413, 1987.*

[89] J. Parsons, "Relaxation of strained silicon on virtual substrates," PhD Thesis, *2008.*

[90] D. Hull and D. J. Bacon, "Chapter 5 - Dislocations in Face-centered Cubic Metals," in Introduction to Dislocations, *Butterworth-Heinemann, 2011, pp. 85-107.*

[91] O. Skibitzki, "Material Science for high performance SiGe HBTs: Solid-Phase Epitaxy and III-V/SiGe hybrid approaches," Phd Thesis, *1982.*

[92] M. Li, Y. Qiu, G. Liu, Y. Wang, B. Zhang and L. Zhao, "Distribution of dislocations in GaSb and InSb epilayers grown on GaAs (001) vicinal substrates," Journal of Applied Physics, *vol. 105, p. 094903, 2009.*

[93] C. R. M. Grovenor, Microelectronic Materials, Taylor & Francis, *1998.*

[94] J. W. Matthews and A. E. Blakeslee, "Defects in epitaxial multilayers: I. Misfit dislocations," Journal of Crystal Growth, *vol. 27, pp. 118-125, 1974.*

[95] A. Maros, N. Fallev, R. R. King and C. B. Honsberg, "Critical thickness investigation of MBE-grown GaInAs/GaAs and GaAsSb/GaAs heterostructures," Journal of Vacuum Science & Technology B, *vol. 34, p. 02L113, 2016.*

[96] Q. Meng, X. Zhang and L. Zhang, "Analysis of Structural Stress in InSb Array Detector without Underfill," Journal of Computers, *vol. 7, no. 8, 2012.*

[97] R. People and J. C. Bean, "Calculation of critical layer thickness versus lattice mismatch for GexSi1-x/Si strainedlayer heterostructures," Applied Physics Letters, *vol. 47, p. 322, 1985.*

[98] J. B. Vílchez and D. G. G. Andrade, "Wikimedia Commons File:Vector de Burgers.PNG," 2008. [Online]. Available: https://commons.wikimedia.org/wiki/File:Vector_de_Burgers.PNG.

[99] Y. Shi, D. Gosselink, V. Y. Umansky, J. L. Weyher and Z. R. Wasilewski, "Threading dislocations in MBE grown AlInSb metamorphic buffers: Revealed and counted," Journal of Vacuum Science & Technology B, *vol. 35, p. 02B112, 2017.*

[100] Y. Shi, D. Gosselink, K. Gharavi, J. Baugh and Z. R. Wasilewski, "Optimization of metamorphic buffers for MBE growth of high quality AlInSb/InSb quantum structures: Suppression of hillock formation," Journal of Crystal Growth, *vol. 477, pp. 7-11, 2017.*

[101] X. Zhao, Y. Zhang, M. Guan, L. Cui, B. Wang, Z. Zhu and Y. Zeng, "Effect of InSb/In0.9Al0.1Sb superlattice buffer layer on the structural and electronic properties of InSb films," Journal of Crystal Growth, *vol. 470, no. 15, pp. 1-7, 2017.*

[102] Keithley, Keithley Model 6220 and 6221 Information Sheet.

[103] National Institute of Standards and Technology (NIST), "Hall Effect Measurements Introduction," [Online]. Available: http://www.nist.gov/pml/div683/hall_intro.cfm.

[104] S. M. Sze and K. K. Ng, Physics of Semiconductor Devices, John Wiley & Sons, Inc., 2007.

[105] I. Isenberg, B. R. Russell and R. F. Greene, "Improved Method for Measurig Hall Coefficients," Review of Scientific Instruments, *vol. 19, p. 685, 1948.*

[106] M. Reed, W. Kirk and P. Kobiela, "Investigation of parallel conduction in GaAs/AlxGa1−xAs modulation-doped structures in the quantum limit," IEEE Journal of Quantum Electronics, vol. 22, no. 9, pp. 1753 - 1759, 1986.

[107] Y. Shao, S. A. Solin, L. R. Ram-Mohan and K.-H. Yoo, "Optimizing the physical contribution to the sensitivity and signal to noise ratio of," Journal of Applied Physics, vol. 101, p. 123704, 2007.

[108] G. M. Dutta, D. Chattopadhyay and B. R. Nag, "Drift and Hall mobilities of electrons in InSb at 30 and 77 K," Journal of Physics C: Solid State Physics, vol. 7, p. 1854, 1974.

[109] F. A. Riddoch and B. K. Ridley, "On the scattering of electrons by polar optical phonons in quasi-2D quantum wells," Journal of Physics C: Solid State Physics, vol. 16, p. 6971, 1983.

[110] V. K. Arora and A. Naeem, "Phonon-scattering-limited mobility in a quantum-well heterostructure," Physical Review B, vol. 31, no. 6, pp. 3887-3892, 1985.

[111] B. K. Ridley, "The electron-phonon interaction in quasi-two-dimensional semiconductor quantum-well structures," Journal of Physics C: Solid State Physics, vol. 15, p. 5899, 1982.

[112] W. Walukiewicz, H. E. Ruda, J. Lagowski and H. C. Gatos, "Electron mobility in modulation-doped heterostructures," Physical Review B, vol. 30, p. 4571, 184.

[113] J. Bardeen and W. Shockley, "Deformation Potentials and Mobilities in Non-Polar Crystals," Physical Review, vol. 80, p. 72, 1950.

[114] B. J. F. Lin and D. C. Tsui, "Mobility of the two-dimensional electron gas in GaAs-AlxGa1-xAs heterostructures," Applied Physics Letters, vol. 45, p. 695, 1984.

[115] P. J. Price, "Two-dimensional electron transport in semiconductor layers. I. Phonon scattering," Annals of Physics, vol. 133, no. 2, pp. 217-239, 1981.

[116] J. Singh, Physics of Semiconductors and their Heterostructures, McGraw-Hill, 1993.

[117] U. Penner, H. Rucker and I. N. Yassievich, "Theory of interface roughness scattering in quantum wells," Semiconductor Science and Technology, vol. 13, no. 7, 1998.

[118] D. K. Ferry and S. M. Goodnick, Transport in Nanostructures, Cambridge University Press, 1997.

[119] W.-P. Hong, J. Singh and P. K. Bhattacharya, "Interface roughness scattering in normal and inverted In0.53Ga0.47As-In0.52Al0.48As modulation-doped heterostructures," IEEE Electron Device Letters, vol. 7, no. 8, pp. 480-482, 1986.

[120] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Lilientl, D. Fathy and O. L. Krivanek, "Surface roughness at the Ai(100)-SiO2 interface," Physical Review B, vol. 32, p. 8171, 1985.

[121] P. T. Coleridge, "Small-angle scattering in two-dimensional electron gases," Physical Review B, vol. 44, p. 3793, 1991.

[122] P. T. Coleridge, P. Zawadzki and A. S. Sachrajda, "Peak values of resistivity in high-mobility quantum-Hall-effect samples," Physical Review B, vol. 49, p. 10798(R), 1994.

[123] S. Syed, M. J. Manfra, Y. J. Wang, R. J. Molnar and H. L. Stormer, "Electron scattering in AlGaN/GaN structures," Applied Physics Letters, vol. 84, p. 1507, 2004.

[124] S. Das Sarm and F. Stern, "Single-particle relaxation time versus scattering time in an impure electron gas," Physical Review B, vol. 32, p. 8442(R), 1985.

[125] P. T. Coleridge, "Inter-subband scattering in a 2D electron gas," Semiconductor Science and Technology, vol. 5, no. 9, 1990.

[126] M. Shinohara, T. Ito, K. Wada and Y. Imamura, "Electrical Properties of Oval Defects in GaAs Grown by MBE," Japanese Journal of Applied Physics, vol. 23, no. 6, pp. L371-L373, 1984.

[127] T. Ohachi, M. Inada, K. Asai and J. M. Feng, "Arsenic pressure dependence of hillock morphology on GaAs (n11)A substrates grown using MBE," Journal of Crystal Growth, Vols. 227-228, pp. 67-71, 2001.

[128] T. Ohashi, G. W. Wicks, S. Mukherjee and L. F. Eastman, "Sb Induced Nucleation of InSb on (Ill) InSb Substrates by Molecular Beam Epitaxy," Journal of Electronic Materials, vol. 14, no. 4, 1985.

[129] S. Fujikawa, T. Taketsuru, D. Tsuji, T. Maeda and H. I. Fujishiro, "Improved electron transport properties of InSb quantum well structure using stepped buffer layer for strain reduction," Journal of Crystal Growth, vol. 425, pp. 64-69, 2015.

[130] W. K. Burton, N. Cabrera, F. C. Frank and N. F. Mott, "The growth of crystals and the equilibrium structure of their surfaces," Philisophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, vol. 243, no. 866, 1951.

[131] J. M. S. Orr, P. D. Buckle, M. Fearn, G. Giavaras, P. J. Wilding, C. J. Bartlett, M. T. Emeny, L. Buckle, J. H. Jefferson and T. Ashley, "Low temperature impact ionization in indium antimonide high performance quantum well field effect transistors," Journal of Applied Physics, vol. 99, p. 083703, 2006.

[132] A. M. Gilbertson, J. M. S. Orr, P. D. Buckle, S. Clowes, M. Fearn, C. J. Storey, L. Buckle, L. F. Cohen and T. Ashley, "Low-temperature Schottky barrier tunneling in InSb/InxAl1-xSb quantum well heterostructures," Physicl Review B, vol. 76, p. 085306, 2007.

[133] A. M. Gilbertson, M. Fearn, A. Kormányos, D. E. Read, C. J. Lambert, M. T. Emeny, T. Ashley, S. A. Solin and L. F. Cohen, "Ballistic transport and boundary scattering in InSb/In1-xAlxSb mesoscopic devices," Physical Review B, vol. 83, p. 075304, 2011.

[134] R. Tsu and L. Esaki, "Tunneling in a finite superlattice," Aplied Physics Letters, vol. 22, p. 562, 1973.

[135] Y. Ando and T. Itoh, "Calculation of transmission tunneling current across arbitrary potential barriers," Journal of Applied Physics, vol. 61, p. 1497, 1987.

[136] P. F. Bagwell and T. P. Orlando, "Landauer's conductance formula and its generalization to finite voltages," Physical Review B, vol. 40, p. 1456, 1989.

[137] A. Asgari, S. Badanejad and L. Faraone, "Electron mobility, Hall scattering factor, and sheet conductivity in AlGaN/AlN/GaN heterostructures," ournal of Applied Physics, vol. 110, p. 113713 , 2011.

[138] L. I. Glazman and I. A. Larkin, "Lateral position control of an electron channel in a split-gate device," Semiconductor Science and Technology, vol. 6, p. 32, 1991.

[139] D. B. Chklovskii, B. I. Shklovskii and L. I. Glazman, "Electrostatics of edge channels," Physical Review B, vol. 46, p. 4026, 1992.

[140] G. Venezian, "On the resistance between two points on a grid," American Journal of Physics, vol. 62, p. 1000, 1994.

[141] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry, vol. 36, no. 8, 1964.

[142] SciPy Cookbook, "Savitzky Golay Filtering," 2007. [Online]. Available: https://scipy-cookbook.readthedocs.io/items/SavitzkyGolay.html.

[143] D. Jena and U. K. Mishra, "Quantum and classical scattering times due to charged dislocations in an impure electron gas," Physical Review B., vol. 66, no. 241307(R), 2002.

[144] D. Jena, A. C. Gossard and U. K. Mishra, "Dislocation scattering in a two-dimensional electron gas," Applied Physics Letters, vol. 76, no. 1707, 2000.

[145] A. J. Chiquito, C. A. Amorim, O. M. Berengue, L. S. Araujo, E. P. Bernardo and E. R. Leite, "Back-to-back Schottky diodes: the generalization of the diode theory in analysis and extraction of electrical parameters of nanodevices," Journal of Physics: Condensed Matter, vol. 24, p. 225303, 2012.

[146] S. A. Wolf, D. D. Awwschalom, R. A. Buhrman, J. M. Daughton, S. von Molnar, M. L. Roukes, A. Y. Chtchelkanova and D. M. Treger, "Spintronics: A Spin-Based Electronics Vision for the Future," Science, vol. 294, no. 5546, pp. 1488-1495, 2001.

[147] S. Murakami, N. Nagaosa and S.-C. Zhang, "Dissipationless Quantum Spin Current at Room Temperature," Science, vol. 301, no. 5638, pp. 1348-1351, 2003.

[148] F. Alshaeer, "A study of current transport in Schottky diodes based on AlInSb/InSb-QW heterostructures," PhD Thesis, 2018.

[149] H. K. Henisch, Semiconductor Contacts, Oxford University Press, 1984.

[150] I. P. Batra, Metalization and Metal-Semiconductor Intferaces, New York: Plenum Press, 1989.

[151] J. H. Davies, I. A. Larkin and E. V. Sukhorukov, "Modeling the patterned two-dimensional electron gas: Electrostatics," Journal of Applied Physics, vol. 77, p. 4504, 1995.

[152] A. R. Long, J. H. Davies, M. Kinsler, S. Vallis and M. C. Holland, "A simple model for the characteristics of GaAs/AlGaAs modulation-doped devices," Semiconductor Science and Technology, vol. 8, no. 8, p. 1581, 1993.

[153] I. A. Larkin and J. H. Davies, "Edge of the two-dimensional electron gas in a gated heterostructure," Surface Science, vol. 361/362, pp. 517-520, 1996.

[154] A. I. Larkin and E. V. Sukhorukov, *"Method to investigate the random potential in a quantum point contact,"* Physical Review B, *vol. 49, no. 8, pp. 5498-5507, 1994.*

[155] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products, Eighth Edition, Academic Press, 2014.*

[156] J. M. S. Orr, P. D. Buckle, M. Fearn, C. J. Storey, L. Buckle and T. Ashley, *"A surface-gated InSb quantum well single electron transistor,"* New Journal of Physics, *vol. 9, p. 261, 2007.*

[157] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. ven der Marel and C. T. Foxon, *"Quantized Conductance of Point Contacts in a Two-Dimensional Electron Gas,"* Physical Review Letters, *vol. 60, no. 9, 1988.*

[158] Q. Zhang, *"Interband Tunnel Transistors,"* PhD Thesis, *2009.*

[159] R. Tsu and L. Esaki, *"Tunneling in a finite superlattice,"* Applied Physics Letters, *vol. 22, p. 562, 1973.*

[160] K. Miyamoto and H. Yamamoto, *"Resonant tunneling in asymmetrical double barrier structures under an applied electric field,"* Journal of Applied Physics, *vol. 84, p. 311, 1998.*

[161] H. Yamamoto, *"Resonant Tunneling Condition and Transmission Coefficient in a Symmetrical One-Dimensional Rectangular Double-Barrier System,"* Applied Physics A, *vol. 42, no. 3, p. 225, 1987.*

[162] H. H. Lee, *"Barrier structure for near unity transmission in double-barrier resonance tunneling,"* Journal of Applied Physics, *vol. 69, p. 6725, 1991.*

[163] J. Nanda, P. K. Mahapatra and C. L. Roy, *"Transmission coefficient, resonant tunneling lifetime and traversal time in multibarrier semiconductor heterostructure,"* Physica B, *vol. 383, p. 232, 2006.*

[164] C. Pacher, W. Boxleitner and E. Gornik, *"Rigorous derivation of coherent resonant tunneling time and velocity in finite periodic systems,"* Physical Review B, *vol. 71, p. 125317, 2005.*

[165] P. Harrison, *Quantum Wells, Wires and Dots, Wiley, 2005.*

[166] D. S. Seljebotn, *"Fast numerical computations with Cython," in* Proceedings of the 8th Python in Science Conference (SciPy 2009)*, 2009.*

[167] S. Brehnel, R. Bradshaw, L. Dalcin, M. Florisson, V. Makarov and D. S. Seljebotn, "Cython C-Extensions for Python," [Online]. Available: https://cython.org/.

[168] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn and K. Smith, "Cython: The Best of Both Worlds," Computing in Science & Engineering, vol. 13, no. 2, pp. 31-39, 2011.

[169] S. Behnel, R. Bradshaw, D. S. Seljebotn, G. Ewing, W. Stein and G. Gellner, "Cython for NumPy users," Cython 3.0a0, 2018. [Online]. Available: https://cython.readthedocs.io/en/latest/src/userguide/numpy_tutorial.html.

[170] I. M. Wilbers, H. P. Langtangen and A. Ødegård, "Using Cython to Speed up Numerical Python Programs," in Fifth National Conference on Computational Mechanics, 2009.

[171] S. Behnel, R. W. Bradshaw and D. S. Seljebotn, "Cython tutorial," in Proceedings of the 8th Python in Science Conference (Scipy 2009), 2009.

APPENDIX A

This appendix details the numerical calculations used to determine solutions to the Schrödinger equation given a confining potential, namely the shooting method, the finite difference method, and the Newton-Raphson method.

When considering potential wells, such as in the QW structure, or in the valence band peak created due to surface gating, confined states will exist. These states will be electron like states for the QW, or heavy and light hole states for the valence band potential, where the specific energies and wavefunctions of these states are of interest. To determine these, a numerical 1D Schrödinger Solver program written in Python by Laura Hanks was used, being adapted for the needs required here. The following sections give an overview of the program's operation, following that of Paul Harrison in his book "*Quantum Wells, Wires and Dots*" [167].

The energies (E) and wave functions (ψ) of the bound states can be found by finding the eigenfunction and eigenvalue solutions to the Schrödinger equation for a 1D potential ($V(z)$)

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2}\psi(z) + V(z)\psi(z) = E\psi(z). \tag{A.1}$$

This assumes a constant effective mass and only considers the envelope function of the bound states. The boundary conditions are such that the states are stationary, continuous, and decay to 0 at infinity, i.e.

$$\psi(z) \to 0 \text{ and } \frac{\partial}{\partial z}\psi(z) \to 0 \text{ as } z \to \pm\infty. \tag{A.2}$$

The first step in numerically solving equation (A.1) is to turn the differential into a finite difference equation, dependant on a finite step size δz (the finite difference method). This is done in two steps, giving

$$\frac{df}{dz} \approx \frac{\Delta f}{\Delta z} = \frac{f(z + \delta z) - f(z - \delta z)}{2\delta z}, \tag{A.3}$$

and

Appendix A

$$\frac{d^2 f}{dz^2} \approx \frac{f(z + 2\delta z) - 2f(z) + f(z - \delta z)}{(\delta z)^2}. \qquad (A.4)$$

By substitution into equation (A.1) and rearranging, ψ at a position $z + \delta z$ can be obtained as a function of $\psi(z)$, such that

$$\psi(z + \delta z) = \left[\frac{2m^*}{\hbar^2} (\delta z)^2 (V(z) - E) + 2 \right] \psi(z) - \psi(z - \delta z) \qquad (A.5)$$

Therefore if ψ is known at 2 consecutive points, the wave function can be deduced for all points, given a small enough step size δz. The shooting method states that general starting conditions of

$$\psi(z - \delta z) = 0 \text{ and } \psi(z) = 1 \qquad (A.6)$$

can be used to give a first guess solution to the wavefunction, which can then be refined.

For a given energy, substituting equation (A.6) into the finite difference version of the Schrödinger equation (equation (A.5)) will return an un-normalised wave function, however multiplying an eigenfunction by a constant does not affect the eigenvalue. The wavefunction can then be normalised through integration, such that

$$\psi(z) \rightarrow \frac{\psi(z)}{\sqrt{\int_{all \ space} \psi^*(z)\psi(z)dz}}. \qquad (A.7)$$

The problem now is to find the correct energy (the energy eigenvalue) to pass to equation (A.5), where this is achieved through monitoring the final point of the wave function generated. For the exact eigenvalue energy, the wavefunction at large z will approach 0, however for an arbitrary energy, the end tail of the wave function will begin to diverge to $\pm\infty$. As this energy is swept in value, it will cross an eigenenergy at which point the tail will flip sign, and diverge to $\mp\infty$. This flip indicates that the solution must have passed through 0, and so there is an eigenenergy between two successive energy values. An example of this divergence and flip of the wavefunction tail is shown in Figure A.1.

This energy range can then be narrowed using an iterative process such as the Newton-Raphson method to analyse the final point of this wavefunction solution, and so approaching the true eigenvalue. In this method, a new estimate eigenenergy (E_N) is

calculated from the current energy and wavefunction final point guess solutions (E_1 and ψ_1) by calculating two new solutions (E_0, ψ_0 and E_2, ψ_2) at energies ΔE below and above E_1 respectively. Then, E_N is given by

$$E_N = E_1 - \frac{\psi_1}{\Delta\psi},$$

$$(A.8)$$

where

$$\Delta\psi = \frac{\psi_2 - \psi_0}{2\Delta E}.$$

$$(A.9)$$

Once the energy has converged below an acceptable level, the remaining exponential tail can be neglected, leaving the true wavefunction. This method was used to determine the bound state solutions as analysed in Chapter 6 : Gated Structure Modelling.
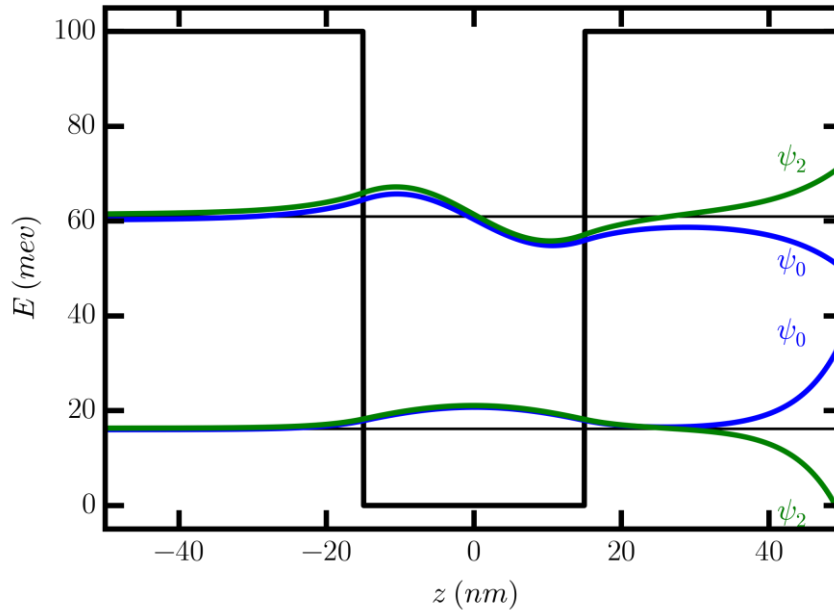

Figure A.1 Example numerical wavefunction solutions of the Schrödinger equation for a 100 meV quantum well. The eigenenergies of the first two states are shown as thin black lines, where the numerical solutions for energies ΔE below (ψ_0, blue) and above (ψ_2, green) this energy show the characteristic diversion to ∞ at large z, altering sign on crossing the eigenenergy solution.

APPENDIX B

The initial MC simulation was implemented in Python, however due to the large number of trials required, and the large number of steps in each trial, this is inefficient. Pure Python is a dynamically typed, interpreted language, with a global interpreter lock (GIL)[15]. These properties make Python easy to write and understand, though each property causes Python code to execute slower than alternatives such as C.

Despite the inherent GIL in Python, it is possible to exploit the use of parallelisation through the Python multiprocessing module. This allows the Python code to be executed across multiple cores in the computer's CPU, allowing a speed up proportional to the number of cores. Many speed improvements can also be achieved using the numerical python package NumPy, which gives support for array objects using faster C and C++ code. NumPy is written in primarily C, and releases the GIL, explaining the significant speed improvements possible [168].

However, due to the nature and variability of the simulation (i.e. the unknown number of steps necessary in each trial), a full implementation through NumPy is not possible and numerical loops in pure python code are still required. To further aid in speeding up execution of the Python code, the Cython language can be exploited. Cython is a "*superset of the Python language that additionally supports calling C functions and declaring C types on variables*" allowing "*the compiler to generate very efficient C code*", speeding up the execution of Python code [169].

Using Cython, variables can be properly typed as in C, and the code is compiled before execution, both of which allow significant speed improvements. Cython is designed to compile Python code directly to C, and working with NumPy, allows the embedding of C-speed numerical loops directly in Python [170]. The speed improvements compared to pure Python can be several thousand, and even dozens of times faster than optimized NumPy [171, 172, 173]. These features, as well as the reduced learning curve compared to C, are why Cython was used to implement the revised MC simulation. This Cython version gives a several hundred times speed improvement over the original implementation in Python, allowing the use of a much larger number of trials within an acceptable runtime.

---

[15] The GIL in Python ensures only a single thread is run at a time when executing pure Python.