A theoretical study of the hole mobility in silicon–germanium heterostructures

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:


Metadata Record: [https://dspace.lboro.ac.uk/2134/34855](https://dspace.lboro.ac.uk/2134/34855)

Publisher: © Adrian Ifor Horrell

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
Please note that fines are charged on ALL overdue items.
A theoretical study of the hole mobility in Silicon-Germanium heterostructures

by

Adrian Ifor Horrell

A doctoral thesis submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University

©Adrian Ifor Horrell
Department of Electronic & Electrical Engineering,
Loughborough University
December 2001
Abstract

The incorporation of Si$_{1-x}$Ge$_x$ alloy heterostructures into conventional Si processes has been proposed as a means of improving the operating frequency and overall performance of Si field effect transistors. One parameter expected to benefit from this approach is the hole mobility, which would have important implications for high speed CMOS applications. Measured values of the hole mobility, however, have failed to live up to early expectations, and much ongoing research is directed at understanding whether this is an intrinsic limitation (e.g. due to alloy disorder scattering), or due to imperfections arising in the growth and fabrication process.

In this thesis, a detailed theoretical study is presented of the hole mobility in single sub-band Si$_{1-x}$Ge$_x$ heterostructures. The possible sources of scattering allowed for are alloy disorder, interface impurities, background impurities, strain fluctuations, interface roughness, acoustic phonons and non-polar optic phonons. The theory is developed not just within a lowest order scattering framework, but also allowing for multiple scattering, and a full and proper account is taken of screening. Detailed comparisons are made with experimental data obtained from a variety of structures both at low temperature (4K) and room temperature (300K). Very accurate fits to the mobility data as a function of sheet carrier concentration have been obtained. The overall conclusion is that alloy disorder alone cannot explain the observed behaviour of the mobility. Whilst alloy scattering is undoubtedly important, the mobilities presently seen in real devices are more than likely being limited by other scattering mechanisms (such as interface roughness) to an equal or even greater extent.
Contents

List of Figures .................................................. xi

List of Tables ................................................... xii

Acknowledgements .............................................. xiii

Published Work .................................................. xiv

1 Introduction .................................................. 1

1.1 Background ................................................ 1

1.2 The pseudomorphic Si_{1-x}Ge_x pMOSFET ............... 4

1.3 This thesis ............................................... 7

2 Review of previous results .................................. 8

2.1 Strain and bandstructure ................................. 8
3.6.1 Alloy Disorder ........................................ 35
3.6.2 Charged impurity scattering .......................... 40
3.6.3 Interface (or surface) roughness ....................... 42
3.6.4 Strain fluctuations ........................................ 43
3.6.5 Acoustic phonons ......................................... 44
3.6.6 Non-polar optic phonons ................................. 45
3.7 Screening and the dielectric function .................... 46

4 Analysis of experimental results .......................... 49

4.1 Low temperature hole mobility in normal devices ...... 50
  4.1.1 Physical description of the devices ................... 50
  4.1.2 Hall mobility measurements .......................... 52
  4.1.3 Lowest order mobility modelling ...................... 53
  4.1.4 Multiple scattering mobility modelling ............... 61

4.2 Room temperature hole mobility in a 50% Ge 'normal' device 65
  4.2.1 Physical description of the device .................... 66
4.2.2 Hole mobility ........................................ 66

4.3 The 'inverted' device and suppression of interface roughness scattering ........................................ 72

5 Theoretical analysis and prediction ......................... 78

5.1 Alloy disorder scattering and the effects of screening .......... 79

5.2 The transport to state lifetime ratio for alloy scattering and ionized impurity scattering ......................... 85

5.2.1 Alloy disorder scattering only ............................ 87

5.2.2 Alloy disorder and background impurity scattering ...... 90

5.2.3 Alloy disorder and planar impurity scattering ............. 92

5.2.4 Observations ............................................ 94

6 Conclusions and future work ................................ 96

6.1 What now for \( \text{Si}_{1-x}\text{Ge}_x \) pMOS? .............................. 100

6.2 Suggested future work .................................... 101

Bibliography .................................................. 105
List of Figures

1.1 Schematic view of typical Si$_{1-x}$Ge$_x$ p-channel MOSFET. . . . . . . . . . . . 5

1.2 Band alignment in the Si$_{1-x}$Ge$_x$ p-channel MOSFET (schematic). . . . . . . . 6

2.1 Effect of strain on valence band edges (after [Briggs et al., 1998]).
   The split-off spin-orbit band (SO) is also shown. . . . . . . . . . . . . . . . . . . . . . 10

2.2 Conduction and valence band discontinuities in 20% Ge alloy grown
   on bulk Si (after [Paul, 1999]). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

3.1 k-space geometry of an elastic scattering event. . . . . . . . . . . . . . . . . . . . . 24

4.1 Device structure (schematic only) for section 4.1. . . . . . . . . . . . . . . . . . . . . 51

4.2 Mobility measurements for 'normal' (non-inverted) devices at 4K. . . . . 53
4.3 Fit for sample 35/47. Circles: measured, solid line: theory. Dashed lines a: alloy disorder, b: strain fluctuations, c: interface impurities, d: surface roughness. .......................... 57


4.5 Fits for sample 33/33. Squares: experimental data, solid lines: theory. (a) \( \Delta = 0.58 \text{nm}, \Lambda = 3.5 \text{nm} \), (b) \( \Delta = 0.58 \text{nm}, \Lambda = 20 \text{nm} \) and (c) \( \Delta = 0.1 \text{nm}, \Lambda = 3.5 \text{nm} \). .......................... 60

4.6 Multiple scattering and lowest order fits for sample 35/47 mobility. Circles: experimental data, solid curve: multiple scattering \( (n_i = 1.15 \times 10^{11} \text{cm}^{-2}) \), dashed curve: lowest order \( (n_i = 1.15 \times 10^{11} \text{cm}^{-2}) \) and dotted curve: lowest order \( (n_i = 1.5 \times 10^{11} \text{cm}^{-2}) \). For all, \( \Delta=0.58 \text{nm} \) and \( \Lambda=3.5 \text{nm} \). .......................... 63

4.7 Multiple scattering and lowest order fits for sample 33/33 mobility. Triangles: experimental data, solid curve: multiple scattering \( (n_i = 1.0 \times 10^{11} \text{cm}^{-2}) \), dashed curve: lowest order \( (n_i = 1.0 \times 10^{11} \text{cm}^{-2}) \) and dotted curve: lowest order \( (n_i = 1.2 \times 10^{11} \text{cm}^{-2}) \). For all, \( \Delta=0.1 \text{nm} \) and \( \Lambda=3.5 \text{nm} \). .......................... 64

4.8 Structure of 50% Ge device (schematic only). .......................... 67

4.9 Hole mobility at 300K in the 50% Ge device (circles) and in the Si control device (triangles). The solid curves are included as a guide to the eye. .......................... 68
4.10 50% Ge device theoretical mobility calculations and experimental data (circles) at 300K. 70

4.11 4K mobility for Si$_{1-x}$Ge$_x$ device (circles) and Si control device (triangles). Theoretical fit for the alloy device mobility is shown by the solid curve, while the dashed curves show the contributions of individual scattering mechanisms as labelled. 71

4.12 Structure of inverted device and schematic band structure. The 2DHG is formed in the region marked '+++'. 73

4.13 Measured (circles) and theoretical mobilities in the inverted device. Individual contributions of scattering mechanisms also shown. 74

4.14 Variation of the state 'width' (taken, somewhat arbitrarily, to be $6/b$ where $b$ is the Fang-Howard variational parameter) with $n_d$. 75

4.15 Comparison between the effects of roughness scattering in the inverted device (solid curves) with a normal device (dashed curves). 76

5.1 Alloy scattering mobility vs. temperature ($n_d = 10^{12}$cm$^{-2}$, $x = 0.2$). a: no screening, b: screened with no local field correction, c: screening with local field correction. 81

5.2 Alloy scattering mobility vs. $n_d$ at 4K ($x = 0.2$). a: no screening, b: screened with no local field correction, c: screened with local field correction. 82
5.3 Alloy scattering mobility vs. $n_s$ at 300K ($x = 0.2$). a: no screening, b: screened with no local field correction, c: screened with local field correction. 83

5.4 Alloy scattering mobility vs. Ge fraction, $x$ ($T = 4K, n_s = 10^{12} \text{cm}^{-2}$). a: no screening, b: screened with no local field correction, c: screened with local field correction. 84

5.5 Alloy scattering mobility vs. Ge fraction, $x$ ($T = 300K, n_s = 10^{12} \text{cm}^{-2}$). a: no screening, b: screened with no local field correction, c: screened with local field correction. 85

5.6 $\tau/\tau_s$ versus $n_s$ for alloy disorder scattering in $\text{In}_{1-x}\text{Ga}_x\text{As}$. The solid curve shows the multiple scattering result and the dashed curve shows the lowest order result. Inset shows mobility versus $n_s$ on the basis of multiple scattering theory. 88

5.7 $\tau/\tau_s$ versus $n_s$ for alloy disorder scattering in $\text{Si}_{1-x}\text{Ge}_x$. The solid curve shows the multiple scattering result and the dashed curve shows the lowest order result. Inset shows mobility versus $n_s$ on the basis of multiple scattering theory. 89

5.8 $\tau/\tau_s$ for background impurity and alloy disorder scattering in $\text{In}_{1-x}\text{Ga}_x\text{As}$. Solid lines show multiple scattering $\tau/\tau_s$, dashed lines show $\tau_s$ (lowest order)/ $\tau_s$ (multiple). $N_B = (a)10^{15} \text{cm}^{-3}$, (b) $10^{16} \text{cm}^{-3}$ and (c) $10^{17} \text{cm}^{-3}$. Inset shows multiple scattering result for background impurities only. 90
5.9 \( \tau_f / \tau_s \) for background impurity and alloy disorder scattering in Si\(_{1-x}\)Ge\(_x\).
Solid lines show multiple scattering \( \tau_f / \tau_s \), dashed lines show \( \tau_f \) (lowest order)/\( \tau_s \) (multiple). \( N_B = (a)10^{15}\text{cm}^{-3}, (b)10^{16}\text{cm}^{-3} \) and (c)\( 10^{17}\text{cm}^{-3}\). 
Inset shows multiple scattering result for background impurities only. 91

5.10 \( \tau_f / \tau_s \) versus \( n_s \) in In\(_{1-x}\)Ga\(_x\)As for alloy scattering together with a plane of impurities of concentration \( n_i = 10^{11}\text{cm}^{-2} \) placed at distances from the centre of the quantum well of (a) 20nm, (b) 10nm and (c) 5nm. Solid curves show the results on the basis of multiple scattering theory and dashed curves show the lowest order results.
The inset shows \( \tau_f / \tau_s \) versus \( n_s \) for planar impurities only. 93

5.11 \( \tau_f / \tau_s \) versus \( n_s \) in Si\(_{1-x}\)Ge\(_x\) for alloy scattering together with a plane of impurities of concentration \( n_i = 10^{11}\text{cm}^{-2} \) placed at distances from the centre of the quantum well of (a) 17.5nm, (b) 7.5nm and (c) 2.5nm. Solid curves show the results on the basis of multiple scattering theory and dashed curves show the lowest order results.
The inset shows \( \tau_f / \tau_s \) versus \( n_s \) for planar impurities only. 94

6.1 Predicted mobility improvement at 4K in device 35/47 through reducing \( n_i \) to \( 5 \times 10^{10}\text{cm}^{-2} \) and changing Gaussian roughness parameters to (a) \( \Delta = 0.25nm, \Lambda = 4nm \) or (b) \( \Delta = 0.1nm, \Lambda = 4nm \). 99
List of Tables

4.1 'Normal' (non-inverted) device growth parameters. 52
4.2 'Normal' (non-inverted) device fixed parameters. 56
4.3 'Normal' (non-inverted) device lowest order fitting parameters. 60
4.4 50% Ge device fixed parameters. 69
4.5 'Inverted' device fitting parameters. 73

5.1 Material parameters for $\tau_l/\tau_g$ calculation. The parameters for Si$_{1-x}$Ge$_x$ are typical of the work in this thesis, while those for In$_{1-x}$Ga$_x$As are taken from [Gold, 1988]. 87
Acknowledgements

This work was supported in part by a grant (GR/L54011) of the Engineering and Physical Sciences Research Council of the UK.

Though one name must appear on the front page of a doctoral thesis, science is a collaborative business. This is especially true of one’s first venture into research where one relies on the advice of those who have gone before. In this respect, I have been particularly fortunate in finding in Professor Michael Kearney a supervisor and teacher of the very best sort. His patient guidance has brought this work into being.

I also owe much to Professors Terry Whall and Evan Parker and the other members of the Advanced Semiconductors Research Group at the Department of Physics, Warwick University, for access to experimental results and many enlightening (if sometimes hard to keep up with!) conversations. Former members of that group, Rob Lander and M. Ali Sadeghzadeh, provided valuable experimental data, as did H. Fischer and L. Risch of Infineon Technologies. I must also record my debt to some others not directly connected with my work: I have been lucky to count on the friendship of many people during my time as a doctoral student. Foremost must be the 89 Forest Road household - Eric, Mark, Jayne and the satellite figure of Alistair May. My family have contributed materially to my being able to follow this career path and, more importantly, have provided much needed encouragement and support.

Lastly, thank you to Sharon. I can still hardly believe you waited for me.
Published work

A large portion of the work reported in this thesis has already been published or submitted for publication. These publications are listed here in chronological order.


6. H Fischer, L Risch, M J Kearney, A I Horrell, E H C Parker and T E Whall. 'Improved room temperature mobility in a coherently strained Si/Si0.5Ge0.5/Si p-channel field effect transistor.' Submitted to IEEE Transactions on Electron Devices
Chapter 1

Introduction

1.1 Background

In the half century since the invention of the transistor the solid state electronics industry has grown to be one of the largest enterprises ever developed by mankind. The annual volume of the semiconductor components market is now several hundred billion US$. Within this market silicon dominates completely: in 1995 the estimated market share held by silicon technologies was 98%, with Si CMOS alone accounting for 84.5% [Paul, 1999]. The remaining section of the market represents those applications where it is impossible, or presently impractical, to use Si - for example optoelectronics or radiofrequency above a few GHz. In these areas III-V materials such as gallium-arsenide are used, but at great expense. As a point of comparison, at 1995 prices the cost of Si CMOS was 0.01 US$ per mm², while that of epitaxially grown GaAs was 2.00 US$ per mm². The reasons for this difference
include the ease of growing high volumes of good quality Si and the excellent insulating properties of the oxide and nitride of silicon. When one considers, also, the huge capital already sunk into Si fabrication facilities and the industry's years of accumulated experience in handling Si, it would clearly be advantageous to try to squeeze higher performance out of silicon technology in order to meet the demands of new applications rather than have to invest in other technologies.

In the past, improvements in the speed of integrated circuits have been gained by shrinking individual devices - indeed, since the first MOSFETs were developed in the 1960s, gate lengths have fallen exponentially with an approximate doubling of device packing density every eighteen months (the celebrated 'Moore's Law'). Now the limit for ultraviolet lithography is being approached and, though x-ray lithography may allow further progress, the reduction cannot continue indefinitely since, below a certain size, the physical character of the devices will be radically altered as quantum effects become more important. Therefore other means of gaining performance improvement are being considered and exploited, for example the use of copper interconnects.

Another method of achieving this improvement is through the selective incorporation of the alloy $\text{Si}_1-x\text{Ge}_x$ into conventional Si devices. This approach has proved successful with Si bipolar technology, leading to the development of the $\text{Si}_1-x\text{Ge}_x$ heterojunction bipolar transistor (HBT), wherein the base of the transistor is composed of the $\text{Si}_1-x\text{Ge}_x$ alloy while the emitter and collector regions remain in Si. The advantage in this approach comes from the difference in bandgap, $\Delta E_g$, between Si and $\text{Si}_1-x\text{Ge}_x$. The current gain, $\beta$, depends upon this and other parameters as follows [Kasper & Luy, 1991]
\[ \beta \sim \frac{N_E w_E D_E}{N_B w_B D_B} \exp\left(\frac{\Delta E_g}{kT}\right) \]  

(1.1)

where \( N_E \) (\( N_B \)), \( w_E \) (\( w_B \)) and \( D_E \) (\( D_B \)) are, respectively, the doping concentration, layer width and diffusivity in the emitter (base). The salient feature is the factor \( \exp\left(\frac{\Delta E_g}{kT}\right) \) which, for example, with a Ge concentration of 25% has the value 1550 [Kasper & Luy, 1991]. This gives greater current gain for given design parameters and allows greater flexibility in thickness and doping levels. Such devices are already on the market, e.g. IBM manufacture a 2GHz low noise amplifier using Si\(_{1-x}\)Ge\(_x\) HBT technology. The volume of the Si\(_{1-x}\)Ge\(_x\) HBT market in 1999 was estimated to be about a billion US$.

Whether or not this success can be duplicated in Si MOS is an open question. Two relative weaknesses of Si MOS are the comparatively low electron and hole mobilities and the disparity between the mobilities of electrons and holes. Typically, in high purity bulk Si at room temperature, the electron mobility \( \mu_n \sim 1500 \text{cm}^2 \text{V}^{-1} \text{s}^{-1} \) and the hole mobility \( \mu_p \sim 450 \text{cm}^2 \text{V}^{-1} \text{s}^{-1} \) [Davies, 1998] (these values will be substantially lowered in practical devices). The latter leads to problems in CMOS design since, in order to retain symmetric transfer characteristics, the channel width of the p-channel device must be made larger than that of the n-channel device by a factor \( \mu_n/\mu_p \), typically about 3, which in turn restricts device packing density and overall performance.

It is the goal of much ongoing research to increase the mobility of both electrons and holes in Si MOS via the use of Si\(_{1-x}\)Ge\(_x\) heterostructures. While considerable improvements have been made in both, it is interesting that the observed mobility of holes in Si\(_{1-x}\)Ge\(_x\) is lower than one might expect. For comparison, a low temperature (~4K) electron mobility in excess of
500,000 cm²V⁻¹s⁻¹ in a modulation doped quantum well formed by a Si channel within Si₁₋ₓGeₓ has been reported, while the hole mobility in a Si₁₋ₓGeₓ channel in Si has only reached values of about 20,000 cm²V⁻¹s⁻¹ [Schäffler, 1997]. An important question in current research is: is this discrepancy the result of fabrication related issues, or is it evidence of a fundamental limit imposed by the material - specifically through the action of alloy disorder scattering?

We concentrate our study on the p-channel device for two main reasons. Firstly, the benefits for CMOS design of improving the performance of p-channel MOSFETs to a level comparable with that of n-channel devices would be considerable. Secondly the sort of device we consider here is likely to be easier to incorporate into a standard CMOS process as compared with its n-channel counterpart. The essential features of the device will be described in the next section.

### 1.2 The pseudomorphic Si₁₋ₓGeₓ pMOSFET

The device shown schematically in figure 1.1 is typical of the structures studied in this thesis. The main difference from the conventional Si p-channel MOSFET is the inclusion of a thin layer of the alloy Si₁₋ₓGeₓ, typically a few nm thick. The alloy layer is grown epitaxially upon a nominally undoped Si buffer layer, which separates the channel from the n⁺ Si substrate, followed by an epitaxially grown Si cap, gate oxide and metallised or polysilicon contacts. Boron is implanted to form the p⁺ source and drain wells. In some devices, an additional thin p-doped layer is included outside the Si₁₋ₓGeₓ channel to adjust threshold characteristics, and, for low temperature investigations, a
Schottky gate is sometimes used.

In this structure the $\text{Si}_{1-x}\text{Ge}_x$ layer, if kept below a certain critical thickness (a few nm), adopts the lateral lattice spacing of pure Si and therefore a biaxial compressive strain is built in to the layer. Such a structure is referred to as pseudomorphic or coherently strained. The difference in bandgap between Si and $\text{Si}_{1-x}\text{Ge}_x$ is typically a few hundred meV (this depends on alloy composition). In this structure, the difference is mainly accommodated in the valence band (i.e. the structure exhibits type-II band alignment [Paul, 1999]). With a negative gate bias, the band structure in the device resembles that shown
Device structures such as this have been fabricated and studied, with several variations in structure and mode of operation, by many groups, for example [V-Vandeboek et al., 1994], [Kesan et al., 1991] and [Tsutomu et al., 1998]. Generally improvements in hole mobility and transconductance have been observed. The mobility improvement in these devices arises from (i) the intrinsically lower hole effective mass in Ge and (ii) the effects of strain on the bandstructure. Quantum confinement has further beneficial effects on bandstructure and separates carriers from certain scattering sources, for example the Si / SiO₂ interface.
1.3 This thesis

The work described in this thesis is intended to shed some light upon the question of whether the performance of Si$_{1-x}$Ge$_x$ (in terms of low-field hole mobility) is limited by intrinsic material properties or by deficiencies in growth or fabrication. We have, through theoretical calculation, interpreted measured hole mobilities in a particular class of Si$_{1-x}$Ge$_x$ pMOS device (one having a pseudomorphically grown, buried Si$_{1-x}$Ge$_x$ channel - as described in section 1.2) in terms of the scattering mechanisms that limit these mobilities. We show, amongst other things, that while alloy disorder scattering certainly does place an upper limit on hole mobility, there is a wide range of carrier mobility values over which other factors are more important and that, in this range, there is considerable scope for improvement. A long established means of estimating mobility from basic principles is to calculate the relaxation time for a carrier undergoing a single scattering event due to a potential or combination of potentials associated with impurities etc. in the semiconductor crystal. We call this the lowest order theory, and it is widely used in the literature. An improvement to this basic framework attempts to treat multiple scattering events in a self-consistent manner. We call this the multiple scattering theory. Both are used in this thesis, and we will describe some of the features of each theory as we use them to analyse real device structures. A comprehensive review of 2-dimensional transport theory may be found in [Ando et al., 1982].
Chapter 2

Review of previous results

Here we present a survey of some relevant existing work. It is intended to be a broad overview, setting the scene for what follows, rather than an in depth examination of a specific area of work. Detailed technical information will be brought in as required later in the thesis.

2.1 Strain and bandstructure

Si and Ge are the only miscible group IV elements, forming a random substitutional alloy with properties which vary with composition. The alloy crystallises in the diamond lattice with spacing $a$, which varies almost linearly with Ge fraction, $x$ [Schäffler, 1997]. Pure Ge has a lattice spacing of 0.565nm and pure Si one of 0.543nm - the mismatch being about 4.2%. The alloy can be grown epitaxially on a pure Si substrate where, because of the
differing lattice constants, an in-plane compressive strain is built in to the layer. The layer can be grown up to a critical thickness beyond which the strain is relaxed through the formation of defects [Hull, 1994].

The critical thickness for growth in equilibrium is commonly predicted via the model of Matthews and Blakeslee (see [Hull, 1994] for a simple description). A metastable critical thickness greater than the equilibrium critical thickness exists when the layer is grown at low temperatures. For a Ge fraction of 20%, the Matthews and Blakeslee model predicts a critical thickness of about 10nm, whereas, for the same alloy composition, a metastable layer more than 100nm thick may be grown [Paul, 1999]. An alloy layer thinner than the critical thickness which is forced to assume the lattice spacing of the substrate is called a pseudomorphic layer.

The symmetry breaking effect of strain on the alloy is to shift the valence band edges for Heavy and Light Holes lifting the degeneracy as shown schematically in figure 2.1. In the case of compressive strain (b) in Si$_{1-x}$Ge$_x$, the heavy hole mass is substantially lowered and the HH band lies lowest in hole energy. Inter- and intra-band scattering cross sections are reduced by the energy gap $\Delta_m$ [Whall & Parker, 2000]. The variation of the strain induced by varying alloy composition allows for tuning of certain material parameters.

2.2 Low dimensional structures

A benefit of heterojunction devices is that carriers may be confined in a quantum well formed by discontinuities in the conduction band (for electrons)
or the valence band (for holes). A simple case is the quantum well formed in the valence band when a layer of Si$_{1-x}$Ge$_x$ is grown pseudomorphically on a Si substrate. The valence band discontinuity is then given approximately by $\Delta E_v \approx 0.74x$ eV [Paul, 1999], [Schäffler, 1997]. The conduction and valence band discontinuities are shown in figure 2.2.

The devices we consider in this thesis are an example of a broad class of structures described as low dimensional, meaning simply that the carriers in the structure are confined to move in fewer than 3 dimensions [Davies, 1998]. The confinement of carriers in a 2-dimensional quantum well confers several advantages. Carriers may be kept spatially separate from certain sources.
Figure 2.2: Conduction and valence band discontinuities in 20% Ge alloy grown on bulk Si (after [Paul, 1999]).

of scattering (e.g the generically poor quality interface between Si and its oxide). The quantum confinement also leads to further splitting of the bands into a ladder of subbands, further reducing scattering cross sections. Often single subband occupancy can be guaranteed or used as a suitable starting point for analysis. For a comprehensive review of the electronic properties of 2-dimensional systems, see [Ando et al., 1982].

2.3 Effective mass

Effective mass is a key parameter - much of the benefit of strained Si$_{1-x}$Ge$_x$ lies in its reduction by the Ge content, $x$, and the strain. A wide range of experimental values for the hole effective mass in strained Si$_{1-x}$Ge$_x$ is reported
in the literature. Of particular interest in 2-dimensional hole gas (2DHG) transport is the in-plane effective mass. Two important experimental means of measuring $m^*$ are cyclotron resonance and the Shubnikov de Haas effect.

A cyclotron resonance study of structures with $\text{Si}_{1-x}\text{Ge}_x$ layers grown on Si by chemical vapour deposition (CVD) reported in [Cheng et al., 1994] describes a fall in the in-plane effective mass from $0.40m_0$ to $0.29m_0$ as Ge fraction in the alloy is increased from $x = 0.13$ to $x = 0.37$. The authors also point out a discrepancy between measured and calculated effective masses, citing non-parabolicity effects as the cause.

A lower value for $m^*$ than those above was presented by [Whall et al., 1994]. Here, the Shubnikov de Haas effect was used to extract the effective mass. Measurements were made on a $\text{Si} / \text{Si}_{0.87}\text{Ge}_{0.13} / \text{Si}$ structure, with $m^* = 0.23m_0$ reported. The exact value of $m^*$ is subject to uncertainty, but we have enough information to make progress as regards transport theory.

For comparison, the Heavy Hole effective mass in bulk Si is $0.54m_0$ and in bulk Ge it is $0.28m_0$ [Davies, 1998]. A value of $< 0.1m_0$ was reported for strained Ge in [Xie et al., 1993].

### 2.4 Hole mobility in pseudomorphic p-channel devices

Many groups are working on improving hole mobility in the Si/Si$_{1-x}$Ge$_x$ MOS devices, with a wide variety of device structures and growth techniques and
a range of peak hole mobilities at low and room temperature being reported. In reviewing the literature in this field one is almost never comparing like with like - authors report results at different temperatures, using different assumptions to extract hole mobility from their measurements and so on, leaving a seemingly endless list of variations. Nonetheless, it is possible to get a feel for what has been attained from a few representative results.

[Garone et al., 1992] reported on a device much like those described in this thesis, having a buried pseudomorphic Si$_{1-x}$Ge$_x$ layer 10nm thick and 7.5-10nm below the gate oxide. The Si$_{1-x}$Ge$_x$ layers were grown by CVD at 600 - 625°C and in a range of alloy compositions from $x = 0.2$ to $x = 0.4$. A peak effective mobility of 290 cm$^2$V$^{-1}$s$^{-1}$ was reported for the 30% Ge device at a temperature of 300K, an improvement of 50% over a Si only control device. The same device exhibited a 90K peak mobility of over 970 cm$^2$V$^{-1}$s$^{-1}$, an improvement of over 100% on the Si control. The authors also report that the performance of the 40% Ge device was markedly poorer than that of devices with lower Ge fraction, citing defects in the alloy layer, which was thicker than the equilibrium critical thickness, as a possible cause.

[V-Vandebroek et al., 1994] examined a series of structures with graded channels. That is, the alloy composition was varied with distance from the gate. Alloy layers were grown by CVD at 550°C, 15-20nm thick and having composition graded from 25% to 15% Ge and 0 to 25% (a so called retrograded profile). The structures were modulation doped - with the source of hole being a boron doped Si layer, separated from the alloy channel. The authors reported mobilities of 220 cm$^2$V$^{-1}$s$^{-1}$ at 300K and 980 cm$^2$V$^{-1}$s$^{-1}$ at 82K, and observed that the graded channel allowed a larger valence band discontinuity for a given total Ge content giving better carrier confinement in the alloy.
A more recent study [Kaya et al., 2000] reports a peak hole mobility at 300K of about 550cm²V⁻¹s⁻¹ in a device having a 20% Ge layer, nominally 20nm thick, grown by molecular beam epitaxy (MBE) upon a Si buffer layer and situated 7nm below the gate oxide. The mobility performance here was approximately 3 times better than an equivalent Si only device. In [Lander et al., 2000] a comparable peak 295K mobility of 460cm²V⁻¹s⁻¹ is reported in a device having a pseudomorphic Si₀.₆₅Ge₀.₃₅ layer grown by a mass production-like CVD process upon a Si substrate.

2.5 Pure Ge channels and graded buffer layers

A very different structure was reported by [Madhavi et al., 2001]. A pure Ge channel was grown by MBE upon a relaxed Si₀.₃Ge₀.₇ buffer layer. A 295K hole mobility of 1700cm²V⁻¹s⁻¹ was attained, which the authors claim as the highest achieved to date in a Si based p-channel heterostructure. The authors are among several groups who have investigated pure Ge channels in the belief that mobility is almost completely degraded by alloy scattering in Si₁₋ₓGeₓ - a view which is strongly argued for in [Schöffler, 1997]. This commonly held view is challenged in this thesis.

Another pure Ge channel device was fabricated by [Xie et al., 1993]. Again, a strained Ge layer was grown on a relaxed Si₁₋ₓGeₓ buffer by MBE. The
very high hole mobility value of 55,000 cm$^2$V$^{-1}$s$^{-1}$ was recorded at 4.2K. The authors identify interface roughness effects as limiting performance in this device and make the point that a substantial improvement in performance over their own earlier work was achieved by fabricating devices in which carriers were confined away from the roughest interface.

The above two results may be seen as a strong indication that alloy scattering is perhaps a major limitation to performance. Many groups, however, have published results which indicate that the issue is not so clear cut. For example, table 4.1 and figure 4.2 in this thesis (published in [Lander et al., 1997]) show a marked improvement in mobility with increasing Ge fraction in an alloy device. Furthermore, the use of a pure Ge channel requires fairly elaborate processing, whereas the simple pseudomorphic buried alloy channel is, in theory at least, a simple addition to existing CMOS process. Indeed [Lander et al., 2000] explicitly pointed out that the growth technique used was compatible with production processing.

2.6 Mobility in strained Si n-channel devices

For comparison, we will list here some results in n-channel Si$_{1-x}$Ge$_x$ heterostructure research. While n-channel structures certainly provide greater improvements in mobility in discrete devices, they are expected to be harder to integrate than p-channel structures.

A high room temperature electron mobility was reported by Nelson et al. [Nelson et al., 1993]. A series of layers of graded composition from pure
Si to Si$_{0.7}$Ge$_{0.3}$ were grown by CVD on a Si substrate, ending with a thick (about 1.5\( \mu m \)) layer of Si$_{0.7}$Ge$_{0.3}$. This last relaxed layer is known as a virtual substrate. A layer of Si grown on this virtual substrate will be subject to tensile strain; the effects on the conduction band structure lead to a lower effective mass for electrons. This study reports peak electron mobility values of over 2600cm$^2$V$^{-1}$s$^{-1}$ at 294K - this being nearly double the peak mobility observed in bulk Si.

[Welser et al., 1994] compared mobilities in surface and buried strained Si layers grown on a virtual substrate, of similar composition to that described above, with mobility in a pure Si control structure. The buried channel device exhibited a peak mobility of 1620cm$^2$V$^{-1}$s$^{-1}$, as compared with the Si control mobility of 560cm$^2$V$^{-1}$s$^{-1}$ at 290K. The surface layer device showed a smaller improvement in mobility (1010cm$^2$V$^{-1}$s$^{-1}$), but the improvement did not degrade so much with increasing temperature and the structure was easier to fabricate.

At low temperatures (~4K), very high electron mobilities can be attained in strained Si. A review by [Maiti et al., 1998] list several results by different groups including 17,000cm$^2$V$^{-1}$s$^{-1}$ for strained Si on a uniform-composition buffer, 200,000cm$^2$V$^{-1}$s$^{-1}$ on a virtual substrate and even over 500,000cm$^2$V$^{-1}$s$^{-1}$ at 0.4K when front and back gating were used to control the carrier wavefunction. (Compare with a high of about 9300cm$^2$V$^{-1}$s$^{-1}$ for holes in a Si$_{0.87}$Ge$_{0.13}$ device [Whall et al., 1993]).

While these results are certainly impressive, it should be borne in mind that the processing required to fabricate these devices, particularly those involving graded buffer layers and a virtual substrate, is vastly more involved.
than that required for simple pseudomorphic buried $\text{Si}_{1-x}\text{Ge}_x$ channel device. Given the disparity between n-channel and p-channel performance in CMOS, and the greater need for improvement in p-channel performance, this is a motivation for much work in $\text{Si}_{1-x}\text{Ge}_x$ pMOS.

2.7 Commercial $\text{Si}_{1-x}\text{Ge}_x$ technology

Silicon Germanium HBT technology is now firmly established in the market-place. A typical example is Maxim's MAX2648 low noise amplifier (LNA), which is claimed to outperform a similar GaAs LNA at operating frequencies around 5GHz (source: Maxim Integrated Products, www.maxim-ic.com).

IBM offer a custom $\text{Si}_{1-x}\text{Ge}_x$ BiCMOS process supported by design services, claiming a maximum operating frequency of 70GHz and excellent noise performance (www.ibm.com). The company SiGe Semiconductors was founded expressly to exploit $\text{Si}_{1-x}\text{Ge}_x$ technology, and offers several products for the low-noise, low-power, high-frequency environment of broadband wireless systems (www.sige.com).

So far, $\text{Si}_{1-x}\text{Ge}_x$ MOS technology has not been successfully commercialised, although several products are reportedly close to market. Given the dominant position of Si MOS (over Si bipolar) technology, utilisation of $\text{Si}_{1-x}\text{Ge}_x$ in this way may be financially more significant. One may ask: at what point in the Si CMOS roadmap might we expect $\text{Si}_{1-x}\text{Ge}_x$ to become viable? Currently, it is assumed that this will be at gate-lengths of 0.1 $\mu$m or less. Here, we are approaching a regime where the basic physics of operation of
the devices begins to change. Given, also, the fact that Si MOS technology remains dominant and continues to improve despite its oft-predicted demise, this question is hard to answer definitively. Many issues beyond the scope of this thesis must be addressed before such questions may be resolved.

2.8 Theory

The theory of two-dimensional charge transport is presented in a comprehensive review by [Ando et al., 1982]. This reference lays out the framework for both lowest order and multiple scattering analyses of mobility. The lowest order analysis is also described by [Stern & Howard, 1967]. The treatment of multiple scattering in this thesis is drawn largely from a series of papers by Gold and Götze [Gold & Götze, 1981], [Gold & Götze, 1986] and [Gold, 1988]. Further detail will be presented as the need arises - particularly in chapter 3. Note that, although the general framework of transport theory is well established, there are many contentious issues in its application (for example treatment of screening and alloy scattering etc.). Some of these issues are addressed in this thesis.
Chapter 3

Theory of hole transport

In this chapter, we review the theory of hole transport as it applies to the device structures considered in this thesis. We develop expressions for the transport relaxation time and quantum (or state) lifetime for holes due to single and multiple scattering events. The single scattering or lowest order results are given in a temperature dependent form, while the multiple scattering results as shown here are strictly applicable only at zero temperature. Several distinct sources of scattering will be described and screening will be taken into account. Screening is often neglected in analysis of experimental data. We will show that this is not tenable, given the importance of screening at low temperature and the fact that low temperature measurements are commonly used to extract material parameters.

A full analysis of a real device with a realistic band structure is a very difficult problem. Given, also, the many uncertainties in basic theory and material parameters, it is not clear that an all-inclusive (i.e. beyond that presented
here) modelling exercise based upon current understanding would be of great merit in any case (see also section 6.2). In what follows we assume that holes occupy a single two dimensional heavy-hole subband which is parabolic and isotropic in the plane, and characterised by an effective mass $m^*$. We assume, also, that the confining potential of the quantum well is infinite. None of these assumptions is expected to be drastic and each could, in principle, be relaxed.

3.1 Basic concepts

The relationship between the velocity, $v$, of the carriers and the electric field strength, $E$, is known as the velocity-field characteristic. The velocity saturates at high electric field strength. Below this point, at low electric field strengths, the relationship between velocity and electric field strength is approximately linear with the constant of proportionality being the mobility, $\mu$, which is conventionally expressed in units of cm$^2$V$^{-1}$s$^{-1}$:

$$v = \mu E.$$  \hfill (3.1)

The mobility for a carrier with unit charge $e$ and constant effective mass $m^*$ depends upon the transport relaxation time, $\tau_t$ as follows [Ando et al., 1982].

$$\mu = \frac{e}{m^*} \langle \tau_t \rangle.$$  \hfill (3.2)
where $E_k$ and $\tau_k$ are, respectively, the energy and transport relaxation time for a carrier with wavevector $k$ and $f(E)$ is the Fermi-Dirac distribution function. At zero temperature (3.3) reduces to

$$\mu = e\tau(E_F)/m^*,$$

(3.4)

where $E_F$ is the Fermi energy. In the next section we will review the main steps in deriving the lowest order (single scattering) expression $\tau_t$ within the framework of linear transport theory.

### 3.2 Lowest order transport relaxation time

The theory of linear carrier transport is well established (see, for example, [Nag, 1980] - chapter 7, [Stern & Howard, 1967], [Ashcroft & Mermin, 1976] - particularly chapter 16, or [Ando et al., 1982]). The theory is linear in the sense that the semiclassical equation of carrier motion is expanded to linear order, with a term added to correct for collisions, to produce the Boltzmann transport equation. Here we review the main steps in obtaining an expression for $\tau_t$.

Finding the lowest order transport time amounts to solving the linearised
Boltzmann transport equation within the relaxation time approximation. This approach may be exactly applied when all scattering is elastic. As a convenient starting point we shall take the following [Ando et al., 1982]

\[
\frac{1}{\tau_k} = \frac{2\pi}{\hbar} \sum_{k'} \frac{|M_{kk'}|^2}{\varepsilon_{kk'}^2} (1 - \cos \theta_{kk'}) \delta(E_k - E_{k'}). \tag{3.5}
\]

This gives the characteristic transport time over which a carrier in a state with wavevector \( k \) is scattered to some other state with wavevector \( k' \) by a potential associated with an impurity (we will write 'impurity' for clarity - in reality, the scattering potential may arise from other sources e.g. surface roughness or strain fluctuations). Generally, there will be some concentration of impurities and \( \langle |M_{kk'}|^2 \rangle \) is an ensemble averaged square matrix element (defined below) where the averaging is performed over the impurity distribution. The \( \delta \)-function ensures that energy is conserved. \( \varepsilon_{kk'} \) is the dielectric screening function which will be described in section 3.7.

The basic matrix element for scattering by a potential \( V(r, z) \) is of the form

\[
M_{kk'} = \langle \psi_{k'} | V(r, z) | \psi_k \rangle = \int \psi_{k'}^* V(r, z) \psi_k drdz. \tag{3.6}
\]

We consider the movement of carriers to be confined to a plane by virtue of the narrow quantum well, so that \( r, k \) and \( k' \) are now explicitly 2 dimensional vectors and \( z \) a distance perpendicular to the plane. \( V(r, z) \) is the potential due to the scattering impurity and the wavefunctions, normalised within area \( A \), have the form
\[ \psi_k = \frac{1}{\sqrt{A}} e^{i k r} \chi(z), \quad (3.7) \]

where \( \chi(z) \) is the normalised envelope function perpendicular to the plane. The sum (3.5) may be converted to an integral in the plane, assuming a parabolic band and substituting for the energy using the dispersion relation \( E = \hbar^2 k^2 / 2m^* \), to give

\[ \frac{1}{\tau_k} = \frac{2\pi}{h} \frac{A}{(2\pi)^2} \int_0^\pi \int_0^{k'k'} \frac{|M_{k'k'}|^2}{c^2_{k'k'}} (1 - \cos \theta_{k'k'}) \delta \left( \frac{\hbar^2 k^2}{2m^*} - \frac{\hbar^2 k'^2}{2m^*} \right). \quad (3.8) \]

Noting that \( \int f(x) \delta(x - a) dx = f(a) \), and making the substitution \( dX = \hbar^2 k'dk'/m^* \), we obtain

\[ \frac{1}{\tau_k} = \frac{2\pi}{h} \frac{A}{(2\pi)^2} \frac{2m^*}{\hbar^2} \int_0^\pi \frac{|M_{k'k'}|^2}{c^2_{k'k'}} (1 - \cos \theta_{k'k'}) d\theta. \quad (3.9) \]

Now a few subsidiary pieces of information are needed. First consider the matrix element, \( M_{k'k'} \). Substituting appropriate wavefunctions of the form (3.7) into (3.6),

\[ M_{k'k'} = \int \frac{1}{\sqrt{A}} e^{-ik'r} \chi(z)V(r,z) \frac{1}{\sqrt{A}} e^{ikr} \chi(z) dr dz. \quad (3.10) \]

Let \( k' = k + q \), then

23
Figure 3.1: $k$-space geometry of an elastic scattering event.

\[
M_{kk'} = \frac{1}{A} \int e^{-iq\cdot x} \chi^2(z)V(r,z)\, dr\, dz. \tag{3.11}
\]

Clearly, the matrix element depends upon the momentum transfer, $|q|$, and not upon the initial and final values of the momentum, so we write $M_q$ in place of $M_{kk'}$ (the screening function also depends only upon $|q|$). Secondly, consider the geometry of a scattering event in $k$-space as shown in figure 3.1.

From elementary trigonometry ($\theta = \theta_{kk'}$):

\[
q^2 = k^2 + k'^2 - 2kk' \cos \theta. \tag{3.12}
\]

Since, for impurities and roughness scattering etc., the scattering is elastic (this does not apply for non-polar optic phonons but is approximately valid for acoustic phonons), $k = k'$, so
\[ \frac{q^2}{2k^2} = 1 - \cos \theta. \]  

(3.13)

Further manipulation leads to

\[ dq = k \cos \frac{\theta}{2} d\theta = k \sqrt{1 - \frac{q^2}{4k^2}} d\theta. \]  

(3.14)

Now we can return to equation (3.9) and, substituting from (3.14) and using 

\[ E_k = \hbar^2 k^2 / 2m^* \]  

for the carrier energy, we obtain

\[ \frac{1}{\tau_k} = \frac{A}{2\pi \hbar E_k} \int_0^{2\pi} \frac{q^2 dq}{\sqrt{4k^2 - q^2}} \frac{< |M_q|^2 >}{\epsilon_q^2} \]  

(3.15)

Finally, we define the scattering function, \( U_q^2 \), which must be specified for each scattering mechanism considered, as

\[ U_q^2 = A < |M_q|^2 >, \]  

(3.16)

so

\[ \frac{1}{\tau_k} = \frac{1}{2\pi \hbar E_k} \int_0^{2\pi} \frac{q^2 dq}{\sqrt{4k^2 - q^2}} \frac{U_q^2}{\epsilon_q^2} \]  

(3.17)
3.2.1 Mobility

Having established an expression for $\tau_k$, the mobility can be obtained from (3.3), or at low temperature from (3.4) - in either case, the dispersion relationship is used to relate carrier energy to the wavevector. Assuming the quantum well to be infinitely deep so that carriers are never scattered out of it, at finite temperature, $T$, (3.3) can be expressed in the form [Laikhtman & Kiehl, 1993]

\[
\mu = \frac{e}{4\pi h^2 n_s k_B T} \int_0^\infty \frac{E\tau(E) dE}{\cosh^2 \left( \frac{E-E_0}{2k_BT} \right)},
\]  

where $k_B$ is Boltzmann’s constant, $n_s$ is the sheet carrier density and $\xi$ is the chemical potential, which is related to $n_s$ by

\[
\xi = k_B T \ln \left( e^{\frac{\pi h^2 n_s}{k_B T m^*}} - 1 \right).
\]

Where several distinct scattering mechanisms act together but do not interfere with one another, the relaxation times can be summed according to Mathiessen’s Rule

\[
\tau(E) = \left( \sum_i \tau_i(E)^{-1} \right)^{-1},
\]

where $\tau_i(E)$ is the relaxation time due to the $i$th mechanism. This is equivalent to summing the squared scattering functions so that $U_q^2 = U_1^2 + U_2^2 + ...$. 

26
3.3 Lowest order quantum lifetime

The lowest order state or quantum lifetime (as opposed to the transport time) for carriers undergoing impurity or other scattering may be found by directly applying first order time-dependent perturbation theory in the form of Fermi's Golden Rule. If the impurities are sufficiently dilute that carriers may be considered to interact with only one at a time, we can write

\[
\frac{1}{\tau_s} = \frac{1}{2\pi\hbar E_F} \int_0^{2k_F} \frac{2k_F^2 dq}{\sqrt{4k_F^2 - q^2}} \frac{U_q^2}{e_q^2}
\]

(3.21)

Written in this form, equation (3.21) applies only at zero temperature, and is not valid in certain cases e.g. that of uniform background impurity scattering in a square quantum well, which will be considered later (see also [Gold, 1988]). The ratio of transport to quantum lifetime, \( \tau_t/\tau_s \), provides information about the nature of the scattering processes present in a particular structure. For example \( \tau_t/\tau_s \gg 1 \) indicates the dominance of a long-range scattering potential [Gold, 1988] and gives information about the angular dependence of scattering: processes favouring small angle scattering do not greatly affect momentum and so have a smaller degrading effect on current.

The quantum lifetime may be obtained experimentally from Shubnikov de Haas measurements [Davies, 1998], [Whall et al., 1994].

27
3.4 Multiple scattering transport relaxation time

The lowest order expressions for $\tau_1$ and $\tau_s$ given above really apply only when the scattering is weak. Nonetheless, due to their relative simplicity, they are often used to interpret experiments where the scattering (or disorder) is known to be strong.

The problem of how to modify the lowest order expressions to treat multiple scattering exactly is unsolved. For weak disorder, the corrections to the the lowest order results are fairly well understood, but for the strong disorder regime the theory is not well developed at all. In the intermediate regime very little is known.

Over a series of papers, Gold and Götze developed a theory which treats multiple scattering in a self-consistent manner ([Gold, 1988],[Gold & Götze, 1981] and [Gold & Götze, 1986]). The following is based upon their theory which is summarised below. This is not an exact treatment and has some limitations, but we expect that is will capture most of the features of interest.

In general, the zero temperature conductivity is given by the Drude relationship

$$\sigma(z) = \frac{e^2 n_s}{m^*} \frac{1}{z + M(z)}$$

(3.22)

where $z = \omega' + i\omega''$ is a complex frequency, and $M(z)$ is the current relaxation
kernel, given by

\[ M(z) = \frac{1}{4\pi n_e m^*} \int_0^\infty q^2 U_q^2 \phi(q, z) dq. \]  

(3.23)

\( \phi(q, z) \) is the density propagation function which itself depends upon \( M(z) \) - this is the self-consistent aspect of the theory. Within certain assumptions and approximations ([Gold, 1988], [Gold & Götte, 1981] & [Gold & Götte, 1986]).

\[ \phi(q, z) = \frac{\phi_c(q, z + M(z))}{1 + M(z) \phi_c(q, z + M(z))/\chi_c(q, i0^+)} \]  

(3.24)

where \( \phi_c \) is the propagation function in the absence of a random (scattering potential) which is defined in terms of the dressed polarizability \( \chi_c(q, z) \) and its zero frequency limit \( \chi_c(q, i0^+) \):

\[ \phi_c(q, z) = \frac{\chi_c(q, z) - \chi_c(q, i0^+)}{z}. \]  

(3.25)

(N.B. Gold and Götte express the theory in terms of the carrier gas compressibility, but this is equivalent). The dressed polarizability is in turn expressed in terms of the bare polarizability \( \chi_0(q, z) \) and the interaction potential \( V(q) \):

\[ \chi_c(q, z) = \frac{\chi_0(q, z)}{1 + V(q)\chi_0(q, z)}, \]  

(3.26)
\[ \chi_0 = \frac{m^*}{\pi \hbar^2} + \frac{m^*}{2\pi \hbar^2 Q^2} \left[ \sqrt{(Z - Q(Q + 1)\sqrt{Z - Q(Q - 1)}} - \sqrt{Z + Q(Q + 1)\sqrt{Z + Q(Q - 1)}} \right] \]  

(3.27)

where \( Q = q/2k_F \) and \( Z = \hbar z/4E_F \).

The zero-frequency limit of these equations may be found in several ways. We choose to take it along the imaginary axis in the upper half of the complex frequency plane, i.e. \( z = i\omega'' \) as \( \omega'' \to 0 \), which makes later simplification easier. In general, for a complex function \( K(q, z = i\omega'') \), \( \omega'' \geq 0 \), we write \( K(q, i\omega'') \equiv K'(q, i\omega'') + iK''(q, \omega'') \) where \( K' \) and \( K'' \) are real functions.

Then the following results hold:

\[ \chi''_0(q, i\omega'') = 0, \]

(3.28)

\[ \chi''_c(q, i\omega'') = 0, \]

(3.29)

\[ \phi''_c(q, i\omega'') = 0. \]

(3.30)

A self-consistent solution can be found by putting \( M'(i0^+) = 0 \), then from (3.22) we have that the transport time \( \tau_t \equiv i/M''(i0^+) \), and from (3.23)

\[ \frac{1}{\tau_t} = \frac{1}{4\pi n_e m^*} \int_0^\infty q^3 U^2 \phi''(q, i0^+) dq. \]

(3.31)
The imaginary part of the zero-frequency density propagator is given by

\[
\phi''(q, i0^+) = \tau_t \left[ \frac{\chi'_0(q, i0^+) - \chi'_0(q, i/\tau_t)}{\chi'_0(q, i/\tau_t)} \right] \frac{\chi'_0(q, i0^+)}{\varepsilon_q^2} \tag{3.32}
\]

where \(\varepsilon_q = 1 + V(q)\chi'_0(q, i0^+)\) is the static dielectric function. The real part of the bare polarizability for zero frequency is given by

\[
\chi'_0(q, i0^+) = \frac{m^*}{\pi \hbar^2} \left[ 1 - \sqrt{1 - \frac{1}{Q^2}} \theta(Q^2 - 1) \right] \tag{3.33}
\]

where \(\theta(\bullet)\) is the Heaviside unit step function. For non-zero frequency \(\chi'_0(q, i0^+)\) becomes

\[
\chi'_0(q, i/\tau_t) = \frac{m^*}{\pi \hbar^2 Q^2} \left[ F[-Q(Q + 1)]F[-Q(Q - 1)] \right. \\
\left. - F[Q(Q + 1)]F[Q(Q - 1)] \right] \tag{3.34}
\]

where

\[
F(x) = \sqrt{x + \sqrt{x^2 + \frac{\hbar^2}{16\varepsilon_q^2 \tau_t^2}}}. \tag{3.35}
\]

The mobility can be evaluated by finding a self-consistent solution for equations (3.31) to (3.35), which can be achieved by numerical iteration.
For weak disorder where \( h \ll 4E_p \tau_t \), it can be shown that, for \( |Q(Q \pm 1)| \gg 4E_p \tau_t/h \),

\[
\phi''(q, i0^+) \to \frac{2m^* \theta(4k_F^2 - q^2) \frac{1}{\pi h^3 q \sqrt{4k_F^2 - q^2} \epsilon_q^2}} {2.36}
\]

and (3.31) reduces to the lowest order result (3.17). In the strong disorder limit, i.e. \( h \gg 4E_p \tau_t \), the situation is more complicated. For \( |Q(Q \pm 1)| \ll 4E_p \tau_t/h \) we obtain

\[
\phi''(q, i0^+) \to \frac{\pi h^2 \chi_q(q, i0^+)} {E_p \tau_t q^2 \epsilon_q^2}.
\]

An important consequence of this is that (3.31) can only have a non-zero solution if the quantity

\[
A = \frac{1}{4 \pi n^2} \int_0^\infty q \chi_q(q, i0^+) \frac{U^2}{\epsilon_q^2} dq < 1.
\]

In other words, for sufficiently strong disorder, the mobility is strictly zero. This is not rigorous, but it captures the basic features of the problem in that the condition \( A = 1 \) may be viewed as defining a 'mobility edge' which separates strongly localised states from extended states giving rise to non-zero (i.e. measurable) mobility.
3.5 Multiple scattering quantum lifetime

The state lifetime for multiple scattering can be estimated from the single particle Green's function evaluated within the mass shell approximation as demonstrated in [Gold, 1988]. The Green's function is written as

\[ G(k, E) = \frac{1}{E - \hbar^2 k^2 / 2m^* - i\hbar / 2\tau_s}. \] (3.39)

For the self energy, \( i\hbar / 2\tau_s \), only the hole-disorder interaction is considered and in the mass shell approximation we calculate for \( k = k_F \) and \( E = E_F \):

\[ \frac{i\hbar}{2\tau_s} = \sum_q \frac{U_q^2}{\xi_q^2} G(q + k_F, E_F), \] (3.40)

which may be simplified to

\[ \frac{1}{\tau_s} = \frac{2m^*}{\pi^2 \hbar^3} \int_0^\infty \frac{U_q^2}{\xi_q^2} I(q, \tau_s) dq. \] (3.41)

The function \( I(q, \tau_s) \) is defined as

\[ I(q, \tau_s) = \int_{-\pi/2}^{\pi/2} \frac{(m^*/\tau_s)}{(q^2 + 2k_F q \sin \phi)^2 + (m^*/\tau_s)^2} d\phi \] (3.42)

which can be evaluated exactly using contour integration techniques - a step not taken by Gold - to give
\[ I(q, \tau_s) = \frac{\pi}{2q} \sqrt{\frac{\tau_s}{m^* \cosh \theta(q)}} \frac{e^{\theta(q)/2}}{m^* \cosh \theta(q)} \]  

(3.43)

where

\[ \theta(q) = \sinh^{-1}\left[ \frac{(m^*/\tau_s)^2 - q^4 + 4k_F^2q^2}{2(m^*/\tau_s)q^2} \right]. \]  

(3.44)

As \( \tau_s \to \infty \), (3.41) reduces to the lowest order result (3.21), except in the case of uniform background impurities in a square quantum well, where the integral in the lowest order expression diverges and the full self-consistent form must, of necessity, be used.

3.6 Scattering mechanisms

In this thesis we have considered the following distinct sources of scattering: alloy disorder, interface impurities, background impurities, interface roughness, strain fluctuations, acoustic phonons and non-polar optic phonons. Since an important part of this work concerns alloy disorder scattering, we shall examine the scattering function for this mechanism in some detail, paying particular attention to the definition of the alloy potential which is a source of great confusion in the literature. Other scattering mechanisms will be described in outline, with appropriate scattering functions taken from the extensive literature, with errors corrected in some cases.

To cover various situations, we consider each mechanism acting within either
a square quantum well profile or within a triangular one. For the out-of-plane envelope function in the case of an infinite square quantum well of width $L$ we take

$$\chi(z) = \sqrt{\frac{2}{L}} \cos \frac{\pi z}{L}. \quad (3.45)$$

For the infinite triangular quantum well, we use the well known Fang-Howard form [Ando et al., 1982]

$$\chi(z) = \sqrt{\frac{b^2}{2}} z \exp \left( -\frac{b z}{2} \right). \quad (3.46)$$

The variational parameter, $b$, is given (for the non-inverted device) by

$$b = \left[ \frac{12m_z e^2}{\hbar^2 \varepsilon_L} \left( N_d + \frac{11}{32} n_s \right) \right]^{1/3}, \quad (3.47)$$

where $m_z$ is the out-of-plane effective mass and $N_d$ is the depletion charge density arising from the Si buffer layer.

### 3.6.1 Alloy Disorder

The random nature of the $\text{Si}_{1-x}\text{Ge}_x$ alloy and the differing atomic potentials associated with Si and Ge lead to a non-periodic potential in the lattice which acts to scatter carriers.
Consider first 3-dimensional scattering in a bulk sample of the alloy $\text{Si}_{1-x}\text{Ge}_x$, later moving to the 2-dimensional case. The alloy has the diamond structure—a face-centred cubic Bravais lattice with a 2-point basis [Ashcroft & Mermin, 1976].

We begin by labelling the set of lattice sites occupied by Si atoms $\{r_s\}$ and the set occupied by Ge atoms $\{r_G\}$. The set of all sites is then $\{r_L\} = \{r_s\} \cup \{r_G\}$. The potential in the system can then be written as

$$U(r) = \sum_{r_s} U_S(r - r_s) + \sum_{r_G} U_G(r - r_G)$$  \hspace{1cm} (3.48)

or, equivalently, as

$$U(r) = \sum_{r_L} [(1 - x)U_S(r - r_L) + xU_G(r - r_L)] + \sum_{r_G} (1 - x)\Delta U(r - r_G) + \sum_{r_s} (-x)\Delta U(r - r_s),$$ \hspace{1cm} (3.49)

$$\Delta U(r) = U_G(r) - U_S(r),$$ \hspace{1cm} (3.50)

where $U_S(r)$ is the Si potential and $U_G(r)$ is the Ge potential. The first term in the RHS of equation (3.49) is the purely periodic virtual crystal term (having the same periodicity as the Bravais lattice) and the other terms represent the scattering perturbation which is assumed to be small. The wavefunctions in the unperturbed virtual crystal may be written as

$$\psi_k(r) = \frac{1}{\sqrt{V}} \exp(ik \cdot r)u_k(r).$$ \hspace{1cm} (3.51)
where $V$ is the total volume of the crystal and the Bloch function $u_k(r)$ has the same periodicity as the virtual crystal. (3.51) is normalised as

$$\int_{\Omega_c} u_k^*(r)u_k(r)dr = \Omega_c$$

(3.52)

where $\Omega_c$ is the volume of the primitive cell of the Bravais lattice. For lattice spacing $l$, $\Omega_c = l^3/4$. The volume occupied by a single atom is $\Omega_0 = l^3/8$, so that $\Omega_c = 2\Omega_0$.

We make two approximations to evaluate the matrix element $M_q$. Firstly, the $k$ dependence of $u_k(r)$ is small near the $\Gamma$ point and is ignored, so we take $u_k(r) \approx u_0(r)$. Secondly $ql \ll 1$, where $q$ is the momentum transfer and the lattice spacing, $l$, characterises the range of the localised potential $\Delta U(r)$. To further simplify matters, we assume that $u_k(r)$ has the full periodicity of the Bravais lattice, i.e. $u_k(r + r_L) = u_k(r)$. Having made these assumptions, the matrix element for an intra-band scattering process becomes

$$M_q \approx \frac{U_{AL}\Omega}{V} \left[ \sum_x (1-x)e^{-iqr_0} - \sum_x xe^{-iqr_s} \right]$$

(3.53)

$$U_{AL} \equiv \frac{1}{\Omega} \int_V \Delta U(r)u_0^*(r)u_0(r)dr$$

(3.54)

where $U_{AL}$ is, by definition, the alloy potential and $\Omega$ is some volume introduced to give $U_{AL}$ the dimensions of energy. Notice that (3.53) is independent of the particular choice of $\Omega$. We can introduce an effective scattering
potential $U_{eff}(r)$ acting on the envelope functions $1/\sqrt{V}e^{ikr}$,

$$U_{eff}(r) = U_{AL}\Omega \left[ \sum_{r_G} (1 - x)\delta (r - r_G) + \sum_{r_S} (-x)\delta (r - r_S) \right]. \quad (3.55)$$

We can now consider the 2 dimensional case where the envelope functions are of the form (3.7). From here on, all vectors are 2 dimensional. The new matrix element is

$$M_q \approx \frac{U_{AL}\Omega}{A} \left[ \sum_{r_G, r_G} (1 - x)e^{-i\mathbf{q}\cdot \mathbf{r}_G} \chi^2 (z_G) - \sum_{r_S, r_S} xe^{-i\mathbf{q}\cdot \mathbf{r}_S} \chi^2 (z_S) \right]. \quad (3.56)$$

An expression for $U_q^2$ can be derived from (3.56). For configuration averaging, note that there are $x/\Omega_0$ Ge atoms per unit volume and $(1 - x)/\Omega_0$ Si atoms per unit volume.

$$U_q^2 \equiv A(|M_q|^2) = \frac{x(1 - x)U_{AL}^2\Omega^2}{\Omega_0} \int \chi^A (z) dz. \quad (3.57)$$

Finally, the assumption that $u_0(r)$ has the full periodicity of the crystal is relaxed. By considering the lattice as two interlocking face-centred cubic lattices, and considering each sub-lattice separately, the above expressions can be modified. The result is that (3.57) still holds, but $U_{AL}$ is now
where the two-point basis is specified by the vectors $0$ and $a$. It is possible, in principle, to evaluate $U_{AL}$ but it is usually left as a fitting parameter in interpreting experimental data. A critical point is that we are free to choose any volume for $\Omega$. Different authors make different choices [Fischetti & Laux, 1996], so one cannot simply compare reported values of $U_{AL}$ on a like-by-like basis. Often the definition is not made clear, which is a great source of confusion. For example, the alloy potential 0.6eV quoted in [Venkataraman et al., 1993] is 'equivalent' to the value 0.3eV quoted in [Laikhtman & Kiehl, 1993], which goes some way toward explaining the wide range of values for $U_{AL}$ found in the literature: 0.2eV [Hinckley & Singh, 1990a] [Chun & Wang, 1992], 0.27eV [Manku & Nathan, 1991] [Manku et al., 1993], 0.3eV [Laikhtman & Kiehl, 1993], 0.6eV [Li et al., 1993] [Venkataraman et al., 1993], 0.9eV [Fischetti & Laux, 1996] and 1.0eV [Liou et al., 1994]. We will return to this point in section 4.1.3. In what follows we set $\Omega = \Omega_0$, as done in [Venkataraman et al., 1993]. The analysis given by [Laikhtman & Kiehl, 1993] is equivalent to the choice $\Omega = \Omega_e$. Using the Fang-Howard wavefunction (3.46) we obtain

\begin{equation}
U_{q}^2 = x(1-x)U_{AL}^2\Omega_0\frac{3b}{16}, \tag{3.59}
\end{equation}

and for a square quantum well of width $L$,
3.6.2 Charged impurity scattering

We will consider Coulomb scattering from (i) impurities concentrated in a \( \delta \)-layer positioned at a Si/Si_{1-x}Ge_x interface and (ii) impurities distributed uniformly throughout the channel. These cases are chosen because of their technological importance e.g. we expect that impurities will be concentrated at heterointerfaces. The matrix element for Coulomb scattering is (see, for example, [Laikhtman & Kiehl, 1993])

\[
U_q^2 = \frac{3x(1-x)U_{ML}^2\Omega_0}{2L}.
\]  

\text{(3.60)}

\[\begin{align*}
U_q^2 &= \left[ \frac{e^2}{2\varepsilon_L q} \right]^2 \int N_I(z)[F_I(q,z)]^2dz, \\
\end{align*}\]

\text{(3.61)}

where \( N_I \) is the impurity concentration, \( \varepsilon_L \) is the dielectric constant and \( F_I(q,z) \) is the form factor, which is derived from electrostatic considerations.

\textbf{Interface impurities}

For a sheet of charged impurities at the Si/Si_{1-x}Ge_x interface [Gold & Dolgopolov, 1986]

\[
U_q^2 = n_i \left[ \frac{e^2}{2\varepsilon_L q} \right]^2 F_R^2(q, z_i),
\]

\text{(3.62)}
where $n_i$ is the area density of impurities and $F_R^2(q, z_i)$ is the form factor, which for a triangular quantum well profile (i.e. using the Fang-Howard wavefunction with variational parameter $b$) is

$$F_R^2(q, z_i) = \frac{1}{(1 + \frac{q}{b})^8}. \quad (3.63)$$

In the case of a square quantum well, of width $L$, the form factor is [Laikhtman & Kiehl, 1993]

$$F_R^2(q, z_i) = \frac{8\pi^2}{qL} \frac{1}{4\pi^2 + q^2L^2} \left[ \frac{1}{2} e^{\frac{qL}{2}} e^{-qL}(1 - e^{-qL}) \right] \quad (3.64)$$

**Uniform background impurities**

For uniform background impurities at a concentration of $N_I$ per unit volume in a square quantum well of width $L$, the scattering function takes the form (a corrected version of the expression appearing in [Laikhtman & Kiehl, 1993])

$$U^2_q = N_I L \left[ \frac{e^2}{2\pi q^2} \right]^2 F_B(q), \quad (3.65)$$

where

$$F_B(q) = \frac{1}{qL} \left( \frac{4\pi^2}{4\pi^2 + q^2L^2} \right)^2 \left[ \frac{4}{qL} \left( 1 - \frac{3}{2qL}(1 - e^{-qL}) + \frac{1}{2} e^{-qL} \right) \right].$$
Deriving the correct form factor for the triangular quantum well is straightforward though laborious, leading to a cumbersome result. Thus, in what follows, a numerical approach has been used in this case.

3.6.3 Interface (or surface) roughness

At least two factors lead to the formation of an uneven interface between Si and Si$_{1-x}$Ge$_x$: The random distribution of Si and Ge atoms means that a few monolayers at the interface should be considered as a transition region from Si to Si$_{1-x}$Ge$_x$, and non-optimal growth temperature and deposition rate will lead to the formation of islands, rather than planar growth [Emeleus et al., 1993]. These factors lead to small variations in the width of the quantum well which perturb the eigenstates for carriers in the well, producing an effective scattering potential for mobile carriers [Ando et al., 1982].

For a single interface and a triangular well

\[
U_q^2 = \frac{e^4}{\hbar^2 L} \left( N_d + \frac{n_s}{2} \right)^2 \Delta_q^2, \tag{3.67}
\]

and for a single interface and square well

\[
U_q^2 = \frac{\pi^4 h^4}{m_e^2 L^6} \Delta_q^2, \tag{3.68}
\]
We consider two roughness distributions: one is the classic Gaussian distribution characterised by a height $\Delta$ and an in-plane correlation length $\Lambda$ [Ando et al., 1982], for which

$$\Delta_q^2 = \pi \Delta^2 \Lambda^2 \exp \left( -\frac{q^2 \Lambda^2}{4} \right). \quad (3.69)$$

The other is a power law distribution characterised by a height $\Delta$, a length $\Lambda$ and a dimensionless exponent, $n$ (typically $\sim 1$ to 3), which governs the roll-off of the distribution at high values of $q$, for which [Feenstra & Lutz, 1995]

$$\Delta_q^2 = \pi \Delta^2 \Lambda^2 \left( 1 + \frac{q^2 \Lambda^2}{4n} \right)^{-(n+1)}. \quad (3.70)$$

There is no substantial qualitative or quantitative difference between these roughness models once the degree of roughness has been characterised.

3.6.4 Strain fluctuations

A further consequence of the roughening of the Si/Si$_{1-x}$Ge$_x$ interfaces is the creation of local strain fluctuations deep in the channel. These fluctuations may act as a distinct source of scattering, through a deformation potential $\Xi_u$. [Feenstra & Lutz, 1995] have developed a theory for the effect and provided some evidence of its importance for electrons. We have modified this theory to make it applicable to the present situation.

For a single rough interface and a triangular quantum well
and for a single rough interface and square quantum well

\[ U_q^2 = \frac{f^2 \pi^2}{4} \left( \frac{1 + \nu}{1 - \nu} \right)^2 \frac{q^2}{(1 + q/b)^6} \Delta_q^2, \quad (3.71) \]

Here, \( f \) is the lattice mismatch factor, which we evaluate by simple linear interpolation, and \( \nu \) is Poisson’s ratio. The factor \( \Delta_q^2 \) may be specified for Gaussian or power law roughness distributions, as described for interface roughness scattering, above.

### 3.6.5 Acoustic phonons

We take a fairly simplistic approach to acoustic phonon scattering, following that used by [Laikhtman & Kiehl, 1993]. The principal assumption made is the standard one that acoustic phonon scattering can be treated as quasi-elastic [Nag, 1980]. Our expression for the triangular quantum well is a simple modification of that given in [Laikhtman & Kiehl, 1993], where the well width is replaced by the mean well width weighted by the out-of-plane envelope function \( \chi^2(z) \), i.e. \( L \approx \langle z \rangle = 3/b \), to give

\[ U_q^2 = \frac{k_B T \pi^2 b}{2 \rho \nu^2 L}. \quad (3.73) \]

The expression for the square well is
For both, $\rho$ is the material density, $v_L$ is the longitudinal speed of sound and $\Xi_u$ is the deformation potential. The exact value of $\Xi_u$ is a matter of some controversy. [Laikhtman & Kiehl, 1993] choose a value of about 10eV on the basis of linearly interpolating between values in bulk Si and Ge. A value of 4.5eV has been reported [Braithwaite et al., 1997]. There is, however, some recent evidence from hot carrier measurements that $\Xi_u$ may be as low as 3eV [Whall, private communication].

3.6.6 Non-polar optic phonons

We include non-polar optic phonons in a simplified manner following the treatment of [Laikhtman & Kiehl, 1993]. Firstly, phonon scattering is inherently inelastic, so the relaxation time approach to a solution of the linearised Boltzmann transport equation is already a significant compromise. Secondly, the complicated band structure means that the matrix elements governing various allowed and disallowed transitions will themselves be complicated (see e.g. [Hinckley & Singh, 1990a], [Hinckley & Singh, 1990b]). We therefore consider it prudent to err on the pessimistic side by ignoring selection rules altogether; in any case, errors introduced by so doing may well be absorbed by uncertainties in the deformation potentials. Lastly, we consider a non-polar material, assuming that the transverse and longitudinal phonon branches have the same frequency. With these approximations, for an infinite square quantum well of width $L$, we write

\[
U^2 = \frac{3\Xi^2 \kappa B T}{2\rho v_T^2 L}.
\] (3.74)
\[
\frac{1}{\tau_{\text{opt}}} = \frac{3d_0^2 m^*}{\rho \hbar^2 \omega L^2} \left( 1 + \frac{\theta(E - \hbar \omega) e^{\frac{\hbar \omega}{k_B T}}}{e^{\frac{\hbar \omega}{k_B T}} - 1} \right),
\]

(3.75)

where \(d_0\) is an effective optical phonon deformation potential, \(\hbar \omega\) the phonon energy, \(\rho\) is the material density, \(L\) is the lattice constant and \(\theta(\bullet)\) is the Heaviside step function (which embodies the fact that the carrier energy must be greater than the phonon energy for a phonon to be emitted). The parameter \(d_0\) is not well known so, following [Laikhtman & Kiehl, 1993], we choose a value by linear interpolation between the values for Si (29.3eV) and Ge (40eV) in order to make progress. Finally, with the above approximations, for an infinite triangular confining potential, we modify (3.75) by simply replacing \(L\) with a weighted mean well width derived from the confined wavefunction \(\chi(z)\) (3.46).

### 3.7 Screening and the dielectric function

The effects of screening are considered for both long and short range scattering potentials, though not for phonons since they are considered only at high temperature when the screening is expected to be weak anyway. In contrast, screening is expected to be important at low temperature, which is significant since measurements intended to allow extraction of material parameters are often performed at low temperature. We take for the dielectric function

\[
\epsilon_q(T) = 1 + \frac{q_z}{q} F(q) \Pi(q, T)(1 - G_H(q)).
\]

(3.76)
\( q_s = \frac{m^* e^2}{2\pi \varepsilon_{TL} \hbar^2} \) is the Thomas-Fermi screening wavenumber, \( F(q) \) is a form factor due to the confining potential and \( G_H(q) \) is Hubbard's local field correction factor [Gold & Dolgopolov, 1986]

\[
G_H(q) = \frac{q}{2\sqrt{q^2 + k_F^2}}. \tag{3.77}
\]

The static temperature dependent polarizability is given within the random phase approximation by [Laikhtman & Kiehl, 1993]

\[
\Pi(q, T) = \int_0^1 \frac{dx}{\exp\left[\frac{E_q(1-x^2)}{4k_BT} - \frac{\xi}{k_BT}\right] + 1} \tag{3.78}
\]

where \( E_q = \hbar^2 q^2 / 2m^* \).

Neglecting image effects, which is justifiable given the similar dielectric constants in Si and Ge, the form factor for the triangular well is given by [Gold & Dolgopolov, 1986], [Ando et al., 1982]

\[
F(q) = \left[ 1 + \frac{9}{8} \left( \frac{q}{b} \right) + \frac{3}{8} \left( \frac{q}{b} \right)^2 \right] \left( 1 + \frac{q}{b} \right)^{-3}. \tag{3.79}
\]

For the square well, the form factor is

\[
F(q) = \frac{1}{4\pi^2 + q^2 L^2} \left[ 3qL + \frac{8\pi^2}{qL} - \frac{32\pi^4}{q^2 L^2} \right] \frac{1 - e^{-qL}}{4\pi^2 + q^2 L^2}. \tag{3.80}
\]

Generally we find that screening is important in modelling scattering and its
neglect, as is common in the literature, is not justifiable. In particular, the alloy scattering potential is often taken as unscreened on the basis that it is a short range potential. We will show that this is a serious deficiency and that screening has a substantial influence on the effect of alloy scattering. There is also controversy over the inclusion of the local field correction - we will discuss this later, as and when appropriate.
Chapter 4

Analysis of experimental results

In this chapter, we apply the theoretical framework developed in chapter 3 to analyse a variety of experimental data. This will be divided into three sections: one dealing with low temperature mobility in devices similar to those described in chapter 1; one dealing with room temperature and low temperature mobility in a proper device having a 50% Ge alloy layer and, finally, a section investigating a variant on these device structures which leads to a reduction in the strength of interface roughness scattering. Through fitting the experimental data, we aim to characterise the relative importance of various scattering mechanisms and to demonstrate the range of applicability of the lowest order and multiple scattering theories. As this work progressed, understanding of the theory and material parameters was refined, leading to some changes in approach. Relevant details will be explained as and when they arise.
4.1 Low temperature hole mobility in normal devices

This section deals with low temperature Hall mobility data obtained from three samples fabricated by a team coordinated by R. Lander of the Advanced Semiconductors Research group at the Warwick University Department of Physics. The devices are said to be 'normal' to distinguish them from the 'inverted' device which is essentially the same but has a doping slab included below the alloy channel. Conduction takes place at the upper (closest to the gate) Si/\text{Si}_{1-x}\text{Ge}_x interface in the 'normal' device and at the lower interface in the 'inverted' device.

4.1.1 Physical description of the devices

The three devices are very similar in structure to those described in the introduction to this thesis. Figure 4.1 shows the structure of the devices in schematic form. The band structure is taken to be similar to that shown in figure 1.2, and for the purposes of mobility modelling, we shall assume a triangular quantum well profile. This assumption is borne out by a self-consistent solution of the Schrödinger-Poisson equations.

Each sample was grown at Warwick University by solid-source molecular beam epitaxy on an n⁺ (100) substrate and consisted of a 300nm Si buffer layer followed by the pseudomorphically grown strained Si$_{1-x}$Ge$_x$ layer and an undoped Si cap. The charge in the inversion layer was controlled by a
Figure 4.1: Device structure (schematic only) for section 4.1.

self-aligned polysilicon gate. In this way, mobility could be studied as a function of carrier concentration, \( n_+ \), in a single device - an improvement on earlier work (e.g. [Emeleus et al., 1993]). The Si\(_{1-x}\)Ge\(_x\) layers were grown to a nominal composition of 20% Ge, though the actual Ge concentrations as measured by x-ray diffractometry differ slightly and are listed in table 4.1. Two different gate oxides were used, one being a high-quality plasma anodic oxide grown at temperatures below 300°C, and the other produced by low-pressure chemical vapour deposition (LPCVD) at 400°C (see table 4.1). The gate and channel contacts were made conductive by implanting BF\(_2\) at 20keV to a dose of \( \sim 1 \times 10^{16}\text{cm}^{-2} \). Implant activation was performed at 750°C for 30 minutes. We expect the channel to be coherently strained in all three devices [Hull, 1994].
## 4.1.2 Hall mobility measurements

Hall measurements were performed at 4K on each sample listed in table 4.1. A two-carrier model was assumed for the Hall coefficient, $R_H$, in order to extract mobilities and carrier concentrations:

$$\frac{1}{eR_H} = \frac{(N_1\mu_1 + N_2\mu_2)^2}{N_1\mu_1^2 + N_2\mu_2^2}, \quad (4.1)$$

where $N_1$ and $N_2$ are the carrier concentrations and $\mu_1$ and $\mu_2$ the mobilities in two channels - one the $\text{Si}_{1-x}\text{Ge}_x$ channel and the other a parasitic channel at the $\text{Si}/\text{SiO}_2$ interface (see figure 1.2). It is assumed that at low temperatures the Hall scattering factor will be close to unity, so the calculated drift mobility can be compared with the measured Hall mobility, and that the carrier concentration in the second (oxide) channel becomes significant only beyond a certain gate voltage. Both assumptions are borne out by independent measurements [Whall & Parker, 2000]. The results are shown in figure

<table>
<thead>
<tr>
<th></th>
<th>33/33</th>
<th>35/47</th>
<th>39/48</th>
</tr>
</thead>
<tbody>
<tr>
<td>channel Ge conc., %</td>
<td>18</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>channel thickness, nm</td>
<td>24</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>growth temperature, °C</td>
<td>640</td>
<td>640</td>
<td>600</td>
</tr>
<tr>
<td>Si cap thickness, nm</td>
<td>31</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>gate oxide</td>
<td>50nm anodic + 100nm LPCVD</td>
<td>69nm anodic + 100nm LPCVD</td>
<td>100nm LPCVD only</td>
</tr>
</tbody>
</table>

Table 4.1: 'Normal' (non-inverted) device growth parameters.
4.2, where mobility is plotted against carrier concentration in the Si$_{1-x}$Ge$_x$ channel for each of the three samples. Also shown, for comparison, is the inferred mobility in the oxide channel of one particular device, namely 35/47.

### 4.1.3 Lowest order mobility modelling

First, we will attempt to fit the measured data using the lowest order results for mobility. We use a simple numerical integration technique to evaluate the expressions for the wavevector dependent relaxation time and the full temperature dependent mobility - although, since the measurements were made at 4K, using the simple zero temperature expression should not introduce a
large error. Screening effects are included, allowing for the Hubbard local field correction as outlined in section 3.7.

Based upon our understanding of the technologically important performance limiting factors, four scattering mechanisms are considered for this first group of samples: charged interface impurities, alloy disorder, surface roughness and strain fluctuations. Even with this limited selection of scattering mechanisms, the number of adjustable parameters is large, leading, on the face of it, to potential issues in the interpretation of the result of the fitting exercise. However, we can reduce this problem by noting that many of the parameters should be to a large extent independent of the details of the device while others can be approximated by interpolating values between pure Si and pure Ge (these fixed values are summarised in table 4.2). Also, the alloy concentration for each device is well known from x-ray diffractometry. This leaves only four fitting parameters: the interface impurity density, \( n_i \), the roughness scales \( \Delta \) and \( \Lambda \), and the alloy potential, \( U_{AL} \).

We include the alloy potential as a fitting parameter due to the uncertainty in its exact value. The range of values encountered in the literature is huge (see section 3.6.1): 0.2eV [Hinckley & Singh, 1990a] [Chun & Wang, 1992] up to 1.0eV [Liou et al., 1994]. Taken at face value, this represents a factor of 25 difference in the strength of scattering. In section 3.6.1 we described one reason for this spread of values, namely ambiguity in the actual definition of \( U_{AL} \); different authors are really reporting different quantities, so the spread is not as bad as it first may seem. In what follows, we will show that there are at least two other factors at work. First, screening is often ignored in the interpretation of experimental data obtained at low temperature, despite the fact that it is demonstrably important. Second, the presence of other scat-
tering mechanisms (e.g. interface impurities and surface roughness at low temperature and phonons at high temperature) makes it difficult to extract the alloy component of the overall mobility. Our analysis cannot answer the question of the 'correct' value of $U_{AL}$. Based upon information in the literature and with our definition (see section 3.6.1), however, we believe that $U_{AL}$ lies in the range 0.43eV to 0.6eV, with the latter value representing a 'worst-case' view of the alloy scattering. Later, in section 5.1, we will work with the worst-case scenario to establish the importance of alloy scattering in relation to other mechanisms. To begin, though, we take the lower value which, at the time this work was done, was widely believed to be appropriate. This value corresponds to the band-gap offset between Si and Ge - this approach is accepted in the wider alloy scattering literature, but we note that the form of equation 3.58 suggests there is little theoretical basis for estimates of $U_{AL}$ based only on band-gap differences or offsets. One may also expect $U_{AL}$ to have a weak dependence on alloy concentration and strain since these factors affect $u_0(r)$, but an evaluation is beyond the scope of this work. We stress that a detailed sensitivity analysis of how the results in this section are affected by varying $U_{AL}$ shows clearly that the general conclusions are unchanged, and that the quantitative changes can easily be compensated for by small changes in other parameters within acceptable limits of uncertainty. Therefore the actual value chosen within the range suggested above is not critical.
**Depletion charge, \( N_d \)**  10^{11}\text{cm}^{-2}  Schrödinger-Poisson calculation

**In-plane effective mass, \( m^* \)**  0.28\( m_0 \)  [Whall et al., 1994], [Whall et al., 1995]

**Out-of-plane effective mass, \( m_z \)**  0.28\( m_0 \)  [Laikhtman & Kiehl, 1993]

**Dielectric constant, \( \varepsilon_L \)**  12.8  interpolated

**Atomic spacing, \( a \)**  \( \sim 0.547\text{nm} \)  interpolated

**Lattice mismatch, \( f \)**  \( \sim 0.01 \)  interpolated

**Poisson’s ratio, \( \nu \)**  \( \sim 0.28 \)

**Temperature, \( T \)**  4\text{K}  fixed

**Deformation potential, \( \Xi_d \)**  4.5eV  [Braithwaite et al., 1997], [Mironov et al., 1997]

**Alloy potential, \( U_{AL} \)**  0.43eV  see discussion in main text

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depletion charge, ( N_d )</td>
<td>10^{11}\text{cm}^{-2}</td>
<td>Schrödinger-Poisson calculation</td>
</tr>
<tr>
<td>In-plane effective mass, ( m^* )</td>
<td>0.28( m_0 )</td>
<td>[Whall et al., 1994], [Whall et al., 1995]</td>
</tr>
<tr>
<td>Out-of-plane effective mass, ( m_z )</td>
<td>0.28( m_0 )</td>
<td>[Laikhtman &amp; Kiehl, 1993]</td>
</tr>
<tr>
<td>Dielectric constant, ( \varepsilon_L )</td>
<td>12.8</td>
<td>interpolated</td>
</tr>
<tr>
<td>Atomic spacing, ( a )</td>
<td>( \sim 0.547\text{nm} )</td>
<td>interpolated</td>
</tr>
<tr>
<td>Lattice mismatch, ( f )</td>
<td>( \sim 0.01 )</td>
<td>interpolated</td>
</tr>
<tr>
<td>Poisson’s ratio, ( \nu )</td>
<td>( \sim 0.28 )</td>
<td></td>
</tr>
<tr>
<td>Temperature, ( T )</td>
<td>4\text{K}</td>
<td>fixed</td>
</tr>
<tr>
<td>Deformation potential, ( \Xi_d )</td>
<td>4.5eV</td>
<td>[Braithwaite et al., 1997], [Mironov et al., 1997]</td>
</tr>
<tr>
<td>Alloy potential, ( U_{AL} )</td>
<td>0.43eV</td>
<td>see discussion in main text</td>
</tr>
</tbody>
</table>

Table 4.2: ‘Normal’ (non-inverted) device fixed parameters.

**Sample 35/47**

This sample, although not exhibiting the highest mobility, proved to be the easiest to fit using lowest order theory. The basic fit is shown in figure 4.3, and it can be seen that the agreement between theory and experiment is excellent over almost the entire range of measurement. The agreement is not perfect at very low carrier concentration, which is perhaps indicative of the inability of the lowest order theory to model localisation effects (this will be discussed further in section 4.1.4). The fitting parameters used were \( n_i = 1.5 \times 10^{11}\text{cm}^2 \), \( \Delta=0.58\text{nm} \) and \( \Lambda=3.5\text{nm} \), and these values are consistent with other measurements on similar structures [Emeleus et al., 1993], [Braithwaite et al., 1997]. The large interface impurity density is also seen in other devices considered in this thesis.
For this sample, we show also the individual contribution of each scattering mechanism to the overall mobility in figure 4.3. Clearly interface impurity scattering dominates at low carrier concentration, while at high carrier concentration surface roughness appears to be the limiting mechanism - this is generally true of the 'normal' (as distinct from 'inverted') structures studied in this thesis. Notice the relative weakness of screened alloy scattering. It is common to neglect screening in analyses of this kind. To do so here would result in an alloy scattering limited mobility of $\sim 17000 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$, largely independent of carrier concentration.
Figure 4.4: Fit for sample 39/48. Squares: experimental data, solid line: theory.

Sample 39/48

Figure 4.4 shows the lowest order fit to the mobility in sample 39/48. The fit is not quite as good as that obtained for the previous sample, but is still reasonable over a wide range of sheet carrier concentration. The fitting parameters used here are $n_t = 4.0 \times 10^{11} \text{cm}^{-2}$, $\Delta = 0.58 \text{nm}$ and $\Lambda = 3.5 \text{nm}$. These are the same roughness parameters as used for sample 35/47, but the interface impurity charge is greater, and we will comment further on this shortly.
Finally we examine sample 33/33, which proved to be the most awkward to fit. This is only to be expected given that the comparatively low maximum carrier concentration means that localisation or multiple scattering events are very much in evidence, leading to the steep fall off of mobility at the low end of the curve. Nonetheless, it is clear that one cannot retain similar roughness parameters as for samples 35/47 and 39/48. The effect of surface roughness must be reduced to fit the observed mobility at high carrier concentrations. This may be achieved by reducing $\Delta$ or by increasing $\Lambda$ and, as can be seen from curves b and c in figure 4.5, the required change in the roughness is large. For all fits in figure 4.5, $n_i = 1.2 \times 10^{11}\text{cm}^{-2}$. In the next section, we will re-examine this sample in the context of multiple scattering theory where a better fit to this data will be demonstrated, while retaining similar values of the fitting parameters.

Review

We can, on the basis of this modelling exercise, draw several conclusions about the relative differences between samples 33/33, 35/47 and 39/48. For easy comparison, the fitting parameters for each sample are summarized in table 4.3.

(i) We can see that the reduced mobility of 39/48 as compared with 35/47 is due to increased interface impurity charge. The devices were grown under similar conditions and have the same cap layer thickness so it is not sup-
Figure 4.5: Fits for sample 33/33. Squares: experimental data, solid lines: theory. (a) $\Delta = 0.58\text{nm}$, $\Lambda = 3.5\text{nm}$, (b) $\Delta = 0.58\text{nm}$, $\Lambda = 20\text{nm}$ and (c) $\Delta = 0.1\text{nm}$, $\Lambda = 3.5\text{nm}$.

Table 4.3: 'Normal' (non-inverted) device lowest order fitting parameters.

<table>
<thead>
<tr>
<th></th>
<th>35/47</th>
<th>39/48</th>
<th>33/33</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_i(\times 10^{11}\text{cm}^{-2})$</td>
<td>1.5</td>
<td>4.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$\Delta (\text{nm})$</td>
<td>0.58</td>
<td>0.58</td>
<td>0.10 - 0.58</td>
</tr>
<tr>
<td>$\Lambda (\text{nm})$</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5 - 20</td>
</tr>
</tbody>
</table>

posed that the difference occurred during growth. The oxidation processes, however, were different. It is not known how extra impurity incorporation may occur in oxide formation, though the LPCVD oxide seems preferable.
Degradation of electron mobility due to thermal oxide growth has been commented upon before, e.g. in [Prasad et al., 1995]. The roughness parameters are the same and are in good agreement with those obtained for modulation doped structures where no oxide was used [Emeleus et al., 1993], suggesting that in this case oxide formation has had little effect on roughness.

(ii) Reduced roughness is responsible for the enhanced mobility of sample 33/33 as compared with 35/47. Both used the same oxide process and have similar interface impurity densities (lending weight to (i) above). There are two possible reasons for the smoother interface. Firstly, the thicker capping layer may buffer the interface from distortion during oxidation, though (i) above may to some extent preclude this. Even if this were the case, the reduction in maximum attainable carrier concentration is drastic. Secondly, 33/33 employs a slightly lower Ge concentration which, at the growth temperatures used here, may lead to a smoother interface being formed (see table 4.1).

4.1.4 Multiple scattering mobility modelling

Here we apply the self-consistent multiple scattering theory to reexamine the normal devices. We pick out two cases - one in which the observed mobility was readily explained by lowest order theory (sample 35/47) and one in which the lowest order theory did not cope so well (sample 33/33).
Figure 4.6 shows the result of fitting the mobility data using the multiple scattering theory (solid curve). The parameters were as for the lowest order fit except for $n_t$ which is reduced to a value of $1.15 \times 10^{11}\text{cm}^{-2}$ - a change of about 23% (for comparison, the results of the lowest order modelling are shown for $n_t = 1.15 \times 10^{11}\text{cm}^{-2}$ (dashed line) and $n_t = 1.5 \times 10^{11}\text{cm}^{-2}$ (dotted line)). The multiple scattering fit is just as good at high values of the carrier concentration, but significantly better at lower values. This is because the multiple scattering theory captures some of the localisation behaviour expected in the system, which will be increased for increased disorder or, equivalently, for lower carrier concentration (since adjusting the chemical potential for fixed disorder has a similar effect as adjusting the disorder for fixed chemical potential, and the chemical potential depends upon the sheet carrier density). Note that this theory assumes strictly zero temperature, but as is evident from figure 4.6 this does not create a great disparity with the finite temperature lowest-order model.

Sample 33/33

It was shown in section 4.1.3 that observed mobility in sample 33/33 was difficult to explain with the lowest order theory due to the steep fall off at low values of $n_t$. It was mentioned that this fall off was due to localisation effects, and that the multiple scattering theory at least approximates these effects. It is pleasing to note, therefore, that, as shown in figure 4.7 (solid curve), the multiple scattering theory does indeed allow a much closer fit to the
Figure 4.6: Multiple scattering and lowest order fits for sample 35/47 mobility. Circles: experimental data, solid curve: multiple scattering \((n_i = 1.15 \times 10^{11} \text{cm}^{-2})\), dashed curve: lowest order \((n_i = 1.15 \times 10^{11} \text{cm}^{-2})\) and dotted curve: lowest order \((n_i = 1.5 \times 10^{11} \text{cm}^{-2})\). For all, \(\Delta=0.58\text{nm}\) and \(\Lambda=3.5\text{nm}\).

It can be seen that, in general, multiple scattering theory should be used, as experimental data than that obtained in the previous analysis. In generating this curve the interface impurity density, \(n_i\), was reduced to \(1.0 \times 10^{11} \text{cm}^{-2}\) - a change of about 17% from its previous value. As for sample 35/47 we also show the 'best' lowest order fit (dotted line) and the result of performing the lowest order calculation with the value of \(n_i\) extracted from the multiple scattering model (dashed line). For all three theoretical curves, the roughness parameters were \(\Delta=0.1\text{nm}\) and \(\Lambda=3.5\text{nm}\).
Figure 4.7: Multiple scattering and lowest order fits for sample 33/33 mobility. Triangles: experimental data, solid curve: multiple scattering \((n_i = 1.0 \times 10^{11} \text{cm}^{-2})\), dashed curve: lowest order \((n_i = 1.0 \times 10^{11} \text{cm}^{-2})\) and dotted curve: lowest order \((n_i = 1.2 \times 10^{11} \text{cm}^{-2})\). For all, \(\Delta=0.1\text{nm}\) and \(\Lambda=3.5\text{nm}\).

it gives a better fit to experimental data; it certainly should be used at lower carrier concentration. At higher carrier concentration, however, the lowest order theory gives fairly good results. The discrepancy in extracted material parameters in this latter regime may, in any case, be within the limits of experimental uncertainty.
4.2 Room temperature hole mobility in a 50% Ge ‘normal’ device

In this section we examine a coherently strained p-channel FET having a doped Si$_{0.5}$Ge$_{0.5}$ channel. This device was fabricated by H. Fischer and L. Risch of Infineon Technologies, Munich, and is closer to a realistic working device structure than those structures so far considered. The results of their measurements, together with the theoretical work presented here have been submitted for publication [Fischer et al., 2000].

The increase in Ge fraction may engender conflicting expectations of the device’s performance. On the one hand, the increase in Ge content may be expected to lower the effective mass leading to increased mobility and, on the other, it may be expected to strengthen the effect of alloy disorder scattering and reduce mobility. Comparison with a Si only control device clearly indicates substantial mobility improvement in the alloy channel device, and through theoretical calculation of the mobility in much the same way as in the previous section, we can to show that the observed mobility is chiefly limited by interface roughness effects.

Since the multiple scattering theory developed in Chapter 3 is applicable only at zero temperature, we are forced to use the lowest order theory to calculate mobility at 300K and, for consistency, at low temperature. We have seen above that the results produced by the two theories are not qualitatively very different except at very low sheet carrier concentration. Also it is not expected that either will emphasise or reduce the effect of any one scattering mechanism with respect to others.
4.2.1 Physical description of the device

The basic structure of the device is shown in figure 4.8. The devices were grown by reduced pressure chemical vapour deposition and the gate oxide was formed by thermal oxidation. The device is uniformly doped n-type, ensuring enhancement mode operation with the threshold voltage being $\sim -1.0V$. The background doping concentration was found by SIMS and MOS capacitance-voltage measurements to be $\sim 1.5 - 2 \times 10^{17} \text{cm}^{-3}$ and the depletion charge after inversion was found to be $\sim 4 \times 10^{12} \text{cm}^{-2}$. Poisson-Schrödinger calculations show that the confining quantum well profile is triangular and that conduction occurs predominantly at the upper Si/Si$_{0.5}$Ge$_{0.5}$ interface up to a carrier concentration of about $3 \times 10^{12} \text{cm}^{-2}$. This carrier concentration is substantially greater than those seen earlier in this thesis. This reflects the better carrier confinement obtained in the deeper quantum well due to the higher Ge content.

4.2.2 Hole mobility

In figure 4.9, the room temperature hole drift mobility (in this case the drift mobility, as opposed to the Hall mobility, was measured directly) is plotted against carrier concentration for the 50% Ge device and for a Si control device having identical structure except that the channel is composed of pure Si rather than the alloy. Clearly the alloy device exhibits higher mobility. The improvement is by a factor of 2 at a sheet carrier concentration of about $10^{12} \text{cm}^{-2}$. A marked increase in mobility is still evident when the carrier concentration has reached about $10^{13} \text{cm}^{-2}$, even though by this time there
is significant parallel conduction at the oxide interface.

To produce a theoretical fit for this mobility data, we consider the following scattering mechanisms: surface roughness, strain fluctuations, alloy disorder, charged interface impurities, background doping, acoustic phonons and non-polar optic phonons. In the calculation all scattering potentials are screened apart from those relating to phonons (which are important only at high temperature where screening is weak). Surface roughness was modelled by a power-law distribution, fixing the dimensionless exponent $n$ to be 2 (see chapter 3). Similar results may be obtained from a Gaussian roughness model. As far as possible we have attempted to reduce the number of adjustable parameters in fitting the data. This, of course, requires that we provide values for many parameters at the outset. These are drawn from diverse sources and are summarized in table 4.4.
Figure 4.9: Hole mobility at 300K in the 50% Ge device (circles) and in the Si control device (triangles). The solid curves are included as a guide to the eye.

The appropriate value for the effective mass in this situation i.e. at large carrier concentration, is not well known. For example, experimental values of $0.29m_0$ for $n_s \sim 2.3 \times 10^{12}$ cm$^{-2}$ in a 37% Ge alloy [Cheng et al., 1994] and $0.26m_0$ at $n_s \sim 10^{12}$ cm$^{-2}$ in a 40% Ge alloy [Kiatgamolchay, 2000] have been reported. The theory of Zhang and Singh suggests a value above $0.2m_0$ for $n_s > 10^{12}$ cm$^{-2}$ at 300K [Zhang & Singh, 1998]. In the present work, good fits to experimental data have been obtained taking $m^* = 0.27m_0$ at 300K, which appears consistent with the above.

We are now left with the parameters $n_i$, $\Delta$ and $\Lambda$ to fit the data. Selecting values of $n_i = 0.7 \times 10^{11}$ cm$^{-2}$, $\Delta = 0.86$nm and $\Lambda = 0.87$nm, the result
shown in figure 4.10 is obtained where the theory is in good agreement with the experimental data. The importance of surface roughness scattering is clearly seen when the mobility limited by that mechanism (and the related mechanism of strain fluctuations) is plotted separately from the mobility limited by all other mechanisms - as also shown in figure 4.10. The roughness parameters quoted here are not dissimilar to those found elsewhere for the same material system (see e.g. [Hull et al., 1989], [Powell et al., 1992], [Penner et al., 1998]). On the basis of this information it would seem that,
were it possible to reduce significantly the effect of surface roughness, then room temperature mobilities approaching $1000\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ might be achieved in devices like the one considered here. The technological implications of such an improvement would be considerable.

By way of improving confidence in the above modelling exercise, we show in figure 4.11 that a very good fit can be obtained for the mobility of the same device measured at low temperature (4K), although certain parameters must be adjusted by a small amount: $\Delta = 0.72\text{nm}$, $\Lambda = 0.7\text{nm}$ and $m^* = 0.22m_0$ [Zhang & Singh, 1998]. These changes are plausible given the differing thermal distribution of the carrier energy at the two temperatures.
Figure 4.11: 4K mobility for Si$_{1-x}$Ge$_x$ device (circles) and Si control device (triangles). Theoretical fit for the alloy device mobility is shown by the solid curve, while the dashed curves show the contributions of individual scattering mechanisms as labelled.

and our simplifying assumption of a parabolic subband. Shown also is the mobility measured in the Si control device and it can be seen that the improved mobility in the Si$_{0.5}$Ge$_{0.5}$ device is maintained at low temperatures. It can be seen also that charged interface impurities are the dominant scattering mechanism at low carrier concentration and surface roughness takes over at higher carrier concentration, while the effect of the screened alloy disorder scattering is comparatively weak - as observed in the low temperature work presented in section 4.1.
4.3 The ‘inverted’ device and suppression of interface roughness scattering

An alternative structure for Si$_{1-x}$Ge$_x$ p-channel MOS is shown in figure 4.12. The main difference from the other structures described in this thesis is the inclusion of a boron doped layer beneath the alloy channel. This structure minimises parallel conduction in the cap layer and also leads to a reduction in vertical effective field strength with increasing carrier concentration - this corresponds to the width of the carrier wavefunction increasing with $n_x$ as opposed to the narrowing seen in the other devices. Furthermore, conduction now takes place at the lower Si/Si$_{1-x}$Ge$_x$ interface as can be inferred from the schematic band structure shown in figure 4.12.

The device was grown by solid source MBE, with the nominal layer thicknesses as indicated in figure 4.12 and the Hall mobility measured at 4K is shown in figure 4.13 (circles). The multiple scattering theory was used to calculate the hole mobility, taking into account 4 sources of scattering i.e. interface impurities, surface roughness, strain fluctuations and alloy disorder. Assuming a triangular well profile (again, borne out by self consistent Schrödinger-Poisson calculations), including screening and taking the parameters listed in table 4.5 we arrive at the mobility curve shown as a solid line in figure 4.13. Note that the value of $U_{AL}$ is larger than that used previously. This reflects our belief that this is a more realistic value, or at least represents an upper bound on $U_{AL}$ [Kearney & Horrell, 1998], and leads to a more pessimistic result for alloy scattering. Also shown are the individual contributions of the scattering mechanisms.
Figure 4.12: Structure of inverted device and schematic band structure. The 2DHG is formed in the region marked '+++'.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_t$</td>
<td>$0.9 \times 10^{11} \text{cm}^{-2}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.95nm</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.9nm</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
</tr>
<tr>
<td>$U_{AL}$</td>
<td>0.6eV</td>
</tr>
</tbody>
</table>

Table 4.5: 'Inverted' device fitting parameters.
The effects of alloy disorder and interface impurity scattering are in line with what we have shown elsewhere. A noteworthy feature of figure 4.13, however, is the eventual upward turn in the mobilities limited by the roughness-associated scattering mechanisms as the carrier concentration is increased. If attainable in a realistic device structure, then this effect is clearly desirable for Si$_{1-x}$Ge$_x$ devices, particularly in the light of our contention that the main factor limiting the room temperature mobility in the 50% Ge device described in section 4.2 was indeed interface roughness. The effect may be explained by the sensitivity of the roughness limited mobility to change in the carrier state 'width' which varies in the opposite sense to that in the conventional devices - this variation is shown in figure 4.14. Further, the
Figure 4.14: Variation of the state 'width' (taken, somewhat arbitrarily, to be $6/b$ where $b$ is the Fang-Howard variational parameter) with $n_x$.

lower Si/Si$_{1-x}$Ge$_x$ interface where the alloy is grown epitaxially upon Si may be of better quality than the upper one.

In figure 4.15 we compare the roughness limited mobilities for a normal device (i.e. one without the B-doped slab below the channel) and the inverted device. The roughness parameters in both cases are as quoted above for the inverted device. The suppression of the roughness scattering with increasing carrier concentration can clearly be seen. This result, of course, does not represent a full calculation for a realistic device and the only difference between calculations for the normal and inverted mobilities shown here is in the evaluation of the variational parameter $b$, which in the case of the inverted
device is extracted from self-consistent Schrödinger-Poisson modeling of the band structure. Nonetheless, the effect is fairly pronounced and one may hope that any departure from the behaviour illustrated here in a real device will be quantitative rather than qualitative. Once again, we see that the dominant scattering mechanism is not alloy disorder scattering, but surface roughness scattering.

In this chapter, we have presented a detailed comparison of theory with experimental results; more detailed than most encountered in the literature. A clear conclusion is that alloy scattering is not dominant in devices presently being studied—surface roughness seems to dominate. This is a cause for
optimism, since in principle we have some technological control over the factors currently limiting device performance. Whether such improvements can be realised remains to be seen. Eventually, of course, the mobility will be limited by alloy (and phonon) scattering, but this is expected to be at levels well in excess of those currently observed (see also chapter 5).
Chapter 5

Theoretical analysis and prediction

In this chapter, we apply the lowest order and multiple scattering theories of hole mobility, transport lifetime and quantum lifetime to investigate areas in which we do not have experimental data available. The preceding chapter has demonstrated the validity and range of application of these theoretical approaches in the interpretation of experiment and part of the purpose of this chapter is to provide more information about the conditions in which these theories may be expected to give realistic results.

This chapter is broken into two separate sections. In the first we present some results on the effect of alloy scattering on hole mobility in a Si$_{1-x}$Ge$_x$ quantum well within the lowest order theory, examining in some detail the effects of screening and temperature variation. This provides a concrete illustration of the treatment of alloy scattering described in chapter 3. In the
second section, we make a detailed comparison between the effects of alloy scattering and scattering from ionised impurities upon the transport lifetime, quantum lifetime and mobility in $\text{Si}_{1-x}\text{Ge}_x$ and $\text{In}_{1-x}\text{Ga}_x\text{As}$ quantum wells, employing lowest order and multiple scattering theory.

5.1 Alloy disorder scattering and the effects of screening

In order to lend further weight to our contention that alloy disorder scattering is not the greatest limiting factor currently restricting device performance, here we consider the behaviour of alloy disorder scattering in the absence of all other sources of scattering, and examine also the impact of screening on the results. This work was carried out using the lowest order theory of mobility, but later we will show that the results are not qualitatively different under the multiple scattering theory.

In the fitting of theoretical calculations to experimental data, the question of how to treat screening is of paramount importance. Throughout this work, we consider three possibilities: Firstly, that screening is neglected altogether i.e. we set the dielectric function equal to 1; secondly that screening is allowed for (using the dielectric function described in chapter 3) but no account is taken of local field effects; and finally that screening is included as above and a local field correction is made (we use Hubbard’s approximation). It seems possible that inconsistency in the treatment of screening (together, of course, with the unclear definition of the parameter) may have contributed to the
wide range of values reported for $U_{AL}$, and led to the deleterious effect of alloy scattering being overestimated by some authors. Taking these concerns and several sources of data into account, a sensible choice for $U_{AL}$ as defined in section 3.6.1 would seem to be $\sim 0.6$eV. This is probably towards the upper end of the range of possible values, and in choosing it we are likely to describe a worst-case situation as regards alloy scattering.

In what follows we have assumed a triangular quantum well profile, and a dependence of effective mass upon Ge fraction, $x$, of $m^*/m_0 = 0.44 - 0.42x$ [Venkataraman et al., 1993]. To begin, in figure 5.1 we show, with the three screening possibilities outlined above, the alloy scattering limited mobility versus temperature for an alloy composition $x = 0.2$ and fixed $n_s = 10^{12}$cm$^{-2}$, which is typical of devices studied in this thesis. The importance of screening at low temperature is instantly visible. While the inclusion of the local field correction does tend to reduce the mobility back toward its unscreened value, it is clearly a poor approximation to neglect screening altogether in the interpretation of low temperature data (it should be pointed out that the value of the local field correction $G_H$ may, of course, be larger than predicted by Hubbard's approximation here which would ultimately tend to compensate for screening). Low temperature measurements are the norm in material diagnostics because phonon scattering is suppressed and therefore the question of whether to allow for screening in extracting material parameters remains an important one.

As far as device applications are concerned, the variation of mobility with sheet carrier concentration is of greater interest than the temperature dependence. A typical range of carrier concentration for devices presently being studied is $n_s = 2 \times 10^{11}$cm$^{-2}$ to $3 \times 10^{12}$cm$^{-2}$. The best choice of $x$ is unclear.
increasing $x$ up to 0.5 increases the size of the alloy scattering matrix element while reducing the effective mass, so there is a trade off to be made at low $x$. Again, we take $x = 0.2$ as typical and with this choice of parameters we obtain the results shown in figure 5.2 (at $T=4K$) and figure 5.3 (at $T=300K$). It can be seen that at low temperature the effects of screening become more pronounced as the carrier concentration is reduced. This is a well known effect in two dimensional transport (see, for example, [Ando et al., 1982]). At room temperature the trend is in the opposite sense, though the size of the effect is comparatively small.
Figure 5.2: Alloy scattering mobility vs. \( n_s \) at 4K \((x = 0.2)\). a: no screening, b: screened with no local field correction, c: screened with local field correction.

Returning to the question of the optimal choice of \( x \), we show in figure 5.4 and 5.5 the variation of mobility with alloy composition at low and room temperature. Again, the effect of screening is manifest at low temperature, though as \( x \) increases \( m^* \) falls and the effect is lessened. Due to the above mentioned trade-off in parameters related to \( x \) (mainly \( m^* \) and the magnitude of the matrix element) a mobility minimum exists - in this case at \( x \simeq 0.25 \). One can also see that it would be useful to move to values of \( x \) greater than 0.5 though this presents obstacles to fabrication in that it would probably require the use of a relaxed buffer layer and, further, the expression for \( m^* \) used here is suspect for very high \( x \).
Figure 5.3: Alloy scattering mobility vs. $n_s$ at 300K ($x = 0.2$). a: no screening, b: screened with no local field correction, c: screened with local field correction.

Screening is commonly neglected in analyses presented in the literature, e.g. [Hinckley & Singh, 1990a], [Venkataraman et al., 1993]. The above illustrates how this can lead to erroneous results.
Figure 5.4: Alloy scattering mobility vs. Ge fraction, $x$ ($T = 4K, n_s = 10^{12} \text{cm}^{-2}$).

a: no screening, b: screened with no local field correction, c: screened with local field correction.
Figure 5.5: Alloy scattering mobility vs. Ge fraction, $x$ ($T=300\,\text{K}, n_s = 10^{12}\,\text{cm}^{-2}$).

- a: no screening,
- b: screened with no local field correction,
- c: screened with local field correction.

5.2 The transport to state lifetime ratio for alloy scattering and ionized impurity scattering

We have shown how the hole mobility may be calculated via evaluation of the transport lifetime, $\tau_t$, which implies that it is possible to extract a value for $\tau_t$ from measurement of mobility. Also, we have defined the quantum or state lifetime, $\tau_s$, which can be obtained experimentally by measurements of
the Shubnikov de Haas effect. These times are not equivalent and can be markedly different. In fact, the ratio $\tau_l/\tau_s$ contains information about the nature of the scattering processes operating within a sample. For example, long range scattering processes (such as scattering from remote ionized impurities) lead to values of $\tau_l/\tau_s \gg 1$, while short range scattering (such as alloy disorder scattering) gives $\tau_l/\tau_s \sim 1$ [Gold, 1988].

When alloy disorder scattering and ionised impurity scattering are simultaneously present, the behaviour of the ratio $\tau_l/\tau_s$ may be quite complex in its dependence upon parameters such as the sheet carrier concentration, alloy potential and ionized impurity concentration. In this section we examine this behaviour in some detail, employing both the lowest order and multiple scattering approaches to calculate $\tau_l$ and $\tau_s$.

We consider electron transport in $\text{In}_{1-x}\text{Ga}_x\text{As}$ as well as hole transport in $\text{Si}_{1-x}\text{Ge}_x$, in each case assuming an infinite square confining potential of width $L$, and three separate scattering mechanisms: alloy disorder scattering, uniform background ionized impurities and ionized impurities concentrated in a $\delta$-layer at various distances outside the quantum well and at the heterointerface. We have chosen to examine the $\text{In}_{1-x}\text{Ga}_x\text{As}$ case to allow comparison with results in the literature, many of which are for electrons rather than holes. Furthermore, we choose a Ga content of 0.47 which is representative of the classic lattice matched (to InP) $\text{In}_{1-x}\text{Ga}_x\text{As}$ system. Taking the alloy layer to extend a distance $L/2$ either side of $z = 0$, for the $\delta$-layer, the form factor (3.64) is evaluated for varying values of $z_i$, with $z_i = L/2$ corresponding to interface impurities.

We take the parameters shown in table 5.1 - where the $\text{Si}_{1-x}\text{Ge}_x$ parameters,
Table 5.1: Material parameters for $\tau_l/\tau_s$ calculation. The parameters for $\text{Si}_{1-x}\text{Ge}_x$ are typical of the work in this thesis, while those for $\text{In}_{1-x}\text{Ga}_x\text{As}$ are taken from [Gold, 1988].

In particular, are representative of pMOS devices presently studied around the world, and the $\text{In}_{1-x}\text{Ga}_x\text{As}$ parameters are drawn from [Gold, 1988].

5.2.1 Alloy disorder scattering only

If ionized impurities and surface roughness can be eliminated, then the mobility limiting mechanism at low temperature is indeed alloy scattering. In figures 5.6 and 5.7, we plot the ratio $\tau_l/\tau_s$ for $\text{In}_{1-x}\text{Ga}_x\text{As}$ and $\text{Si}_{1-x}\text{Ge}_x$ respectively evaluated via the lowest order (dashed line) and multiple scattering (solid line) theories. One can show analytically that within the lowest order theory as $n_s \to 0$, $\tau_l/\tau_s \to 2/3$ [Gold, 1988]. This is confirmed here by the full calculation. As expected, $\tau_l/\tau_s \sim 1$, indicating a short range scattering potential. Also, as we have described elsewhere, it is clear that multiple scattering effects become more important as $n_s$ is reduced.
Figure 5.6: $\tau_1/\tau_s$ versus $n_s$ for alloy disorder scattering in In$_{1-x}$Ga$_x$As. The solid curve shows the multiple scattering result and the dashed curve shows the lowest order result. Inset shows mobility versus $n_s$ on the basis of multiple scattering theory.

Compared with what is to follow, the lowest order results here are in good agreement with the multiple scattering results over a wide range of carrier concentrations. This is simply because, in both cases, the critical value of carrier concentration, $n_s^*$ (obtained from equation (3.38)), below which the transport time is zero according to the self consistent theory being used, is well outside the range of values plotted. For In$_{1-x}$Ga$_x$As $n_s^* = 2.2 \times 10^7$ cm$^{-2}$ and for Si$_{1-x}$Ge$_x$ $n_s^* = 1.1 \times 10^9$ cm$^{-2}$. Shown as insets to the two figures are the mobilities resulting from applying the relation $\mu = e\tau_1/m^*$ to the multiple
Figure 5.7: $\tau_t/\tau_s$ versus $n_e$ for alloy disorder scattering in $\text{Si}_{1-x}\text{Ge}_x$. The solid curve shows the multiple scattering result and the dashed curve shows the lowest order result. Inset shows mobility versus $n_e$ on the basis of multiple scattering theory.

Scattering values of $\tau_t$, which are comparable with our previous lowest order results for the alloy scattering limited mobility for $\text{Si}_{1-x}\text{Ge}_x$ and with the results of Gold for $\text{In}_{1-x}\text{Ga}_x\text{As}$. Overall, the results are very similar in form, as may be expected from the theory, but quite different in scale - this reflects the widely different values of $m^*$ in the two material systems.
5.2.2 Alloy disorder and background impurity scattering

The behaviour is complicated by the addition of ionized impurity scattering. In figures 5.8 and 5.9, we show $\tau_l/\tau_s$ against $n_s$ for various values of the background impurity concentration $N_B$. The solid lines show the multiple
Figure 5.9: $\tau_1/\tau_s$ for background impurity and alloy disorder scattering in Si$_{1-x}$Ge$_x$. Solid lines show multiple scattering $\tau_1/\tau_s$, dashed lines show $\tau_1$(lowest order)/ $\tau_s$(multiple). $N_B = (a)10^{15}$cm$^{-3}$, (b)$10^{16}$cm$^{-3}$ and (c)$10^{17}$cm$^{-3}$. Inset shows multiple scattering result for background impurities only.

scattering ratio $\tau_1/\tau_s$ while the dashed lines show the ratio of lowest order $\tau_1$ to multiple scattering $\tau_s$, since the lowest order state time cannot be calculated for background impurities within the lowest order approximation. Also shown, in the inset to each figure, is $\tau_1/\tau_s$ for background impurity scattering alone.

It is readily apparent that the ratio is reduced substantially by the inclusion of
alloy scattering, and that the lowest order theory significantly overestimates the ratio at lower values of carrier concentration. This inaccuracy is such that the lowest order theory increasingly suspect for \( n_s \) below about \( 10^{11}\text{cm}^{-2} \) for electrons in \( \text{In}_{1-x}\text{Ga}_x\text{As} \) and about \( 10^{12}\text{cm}^{-2} \) for holes in \( \text{Si}_{1-x}\text{Ge}_x \). The critical carrier concentrations \( n_s^* \) are much higher than for alloy scattering alone and can easily be read off from 5.7 and 5.9. Near these critical concentrations, the ionised impurity scattering dominates, with alloy scattering becoming increasingly significant as the carrier concentration rises - hence the roll-over in the curves. It is interesting to note that because of this roll-over, at certain carrier concentrations, the larger values of \( N_B \) correspond with the largest values of \( \tau_I/\tau_s \) - this is not the case when alloy scattering is absent, when increasing \( N_B \) always reduces \( \tau_I/\tau_s \).

5.2.3 Alloy disorder and planar impurity scattering

In some devices (e.g. HEMTs) a remote plane of dopants is deliberately introduced as a source of carriers. In 5.10 and 5.11 we show the variation of \( \tau_I/\tau_s \) with \( n_s \) for a \( \delta \)-layer of impurities (or ionised dopants) located at two positions outside the quantum well ((a) and (b)) and at one of the heterointerfaces (c). We set the sheet impurity concentration at \( n_s = 10^{11}\text{cm}^{-2} \) (the background impurity concentrations studied above are 'equivalent' to sheet concentrations in the range \( n_s = 5 \times 10^8\text{cm}^{-2} \) to \( n_s = 10^{11}\text{cm}^{-2} \)). For comparison, we show in the insets to these figures \( \tau_I/\tau_s \) versus \( n_s \) for scattering due to the impurity layer alone. Evidently, the ratio \( \tau_I/\tau_s \) can be vastly improved by remote doping - the more so for larger doping offsets. But, as the figures also show, the inclusion of alloy scattering drastically reduces
Figure 5.10: $\tau_l/\tau_s$ versus $n_s$ in In$_{1-x}$Ga$_x$As for alloy scattering together with a plane of impurities of concentration $n_d = 10^{11}$ cm$^{-2}$ placed at distances from the centre of the quantum well of (a) 20nm, (b) 10nm and (c) 5nm. Solid curves show the results on the basis of multiple scattering theory and dashed curves show the lowest order results. The inset shows $\tau_l/\tau_s$ versus $n_s$ for planar impurities only.

This improvement back toward unity. Surprisingly, perhaps, at high carrier concentration the largest value of $\tau_l/\tau_s$ may be seen for impurities within the quantum well. Consequently, one should be wary of attempting to gain too much information about the impurity distribution within a device from measurement of the scattering time ratio alone. It is often assumed, however, that just such a treatment is possible.
Figure 5.11: $\tau_t/\tau_s$ versus $n_s$ in Si$_{1-x}$Ge$_x$ for alloy scattering together with a plane of impurities of concentration $n_d = 10^{11}$ cm$^{-2}$ placed at distances from the centre of the quantum well of (a) 17.5nm, (b) 7.5nm and (c) 2.5nm. Solid curves show the results on the basis of multiple scattering theory and dashed curves show the lowest order results. The inset shows $\tau_t/\tau_s$ versus $n_s$ for planar impurities only.

5.2.4 Observations

On the basis of these results, it appears that lowest order scattering theory should only be used for carrier concentrations above about $10^{11}$ cm$^{-2}$ (electrons) in In$_{1-x}$Ga$_x$As and $10^{12}$ cm$^{-2}$ (holes) in Si$_{1-x}$Ge$_x$. The inclusion of alloy scattering results in a reduction in the scattering time ratio from
the case when only ionized impurity scattering is present, and to a more complex variation of the ratio with \(n_s\) - which is of practical importance as low temperature transport measurements are often made to determine the presence and strength of particular scattering mechanisms in devices. These results show that understanding observations in alloy materials requires a more careful analysis than may be tolerable for a non-alloy system. Extending this analysis to other scattering mechanisms is straightforward, as one need only supply the correct form of the scattering function \(U^2_q\).
Chapter 6

Conclusions and future work

We have undertaken a variety of analyses of experimental data, attempting to explain observed mobility behaviour in terms of the various scattering mechanisms present in real devices. We have included the effects of screening, local field corrections and carrier localisation - all constitute a closer approach to a full physical analysis than is often encountered in the literature.

The direct result of this modelling has been a series of very good theoretical fits to experimental mobility data. We have seen in all cases, a quite rich interaction between different scattering mechanisms, with marked regimes of dominance by certain of these: at low temperature we have seen charged impurity scattering to be strong at low carrier concentrations, giving way to interface roughness scattering at higher carrier concentrations (e.g. figure 4.3 and figure 4.11)

Throughout, the effect of alloy disorder has been comparatively small. Fur-
thermore, the weaker dependence of alloy disorder scattering on carrier concentration indicates that it cannot be considered dominant in explaining the form of the mobility vs. carrier concentration curves that we observe.

Having singled out surface roughness scattering as a strong mobility limiting factor, we have seen in section 4.3 a means of suppressing the effects of this scattering mechanism in the 'inverted' device structure. It should be noted that the main benefit of this device is as an investigative tool to examine interface roughness scattering. Conversion to a practical device presents difficult technological problems.

While it is certainly true that there are a great many parameters which may be varied to get our lowest order and multiple scattering models to fit the observed data, we have attempted to confine the variable space by picking out some fairly well known parameters and keeping them constant while adjusting those parameters which characterise the various scattering mechanisms. In general, the values of parameters arrived at (e.g. surface roughness length scales) seem comparable with experimentally determined values. We are confident that the principal conclusions of this work are substantially unaffected by small changes in these parameters.

It is true, also, that carrier mobility is not the ultimate measure of device performance - many other factors contribute and may make the difference between an excellent device and a poor one. Nonetheless, mobility is an excellent diagnostic of material viability, and through fitting exercises like those conducted in this work, quite detailed information about the material may be extracted.
In addition to the interpretation of experimental results, we have included two pieces of purely theoretical analysis. In the first, in order to lend some further weight to our contention that alloy disorder is not so strong a limiter of mobility as is often supposed, we have calculated the variation in mobility as determined solely by this scattering mechanism. We have taken typical material parameters and varied alloy composition, temperature and carrier concentration. Screening and local field corrections have been accounted for, with the result that mobility is greatly in excess of typical measured values reported here and elsewhere in the literature - another pointer to the fact that another explanation for poor performance must be sought. This last point is critical - with present technology we have little control over the nanoscale structure of the alloy, and therefore little control over alloy scattering, but we can control other aspects of the material system, e.g. impurity incorporation and interface roughness, with substantial scope for improvement. Figure 6.1 shows what may be achieved, albeit at low temperature, taking sample 35/47 from section 4.1 and reducing the effects of interface impurity scattering and surface roughness scattering.

We have also examined in detail the ratio of transport to state lifetime within the multiple scattering theory for holes in $\text{Si}_{1-x}\text{Ge}_x$ and electrons in $\text{In}_{1-x}\text{Ga}_x\text{As}$. A complex interplay between alloy disorder scattering and other mechanisms was revealed, serving to reiterate the point that a simplistic explanation of mobility in terms of alloy disorder scattering alone is deeply suspect. This work also served to indicate the limits of applicability of the lowest order scattering theory most commonly used in the literature.
Figure 6.1: Predicted mobility improvement at 4K in device 35/47 through reducing \( n_s \) to \( 5 \times 10^{10} \text{cm}^{-2} \) and changing Gaussian roughness parameters to (a) \( \Delta = 0.25 \text{nm}, \Lambda = 4 \text{nm} \) or (b) \( \Delta = 0.1 \text{nm}, \Lambda = 4 \text{nm} \).
6.1 What now for Si$_{1-x}$Ge$_x$ pMOS?

Today, a ready market for devices combining the attributes of high operating frequency, low power consumption and low noise exists in the area of broadband wireless communications.

Two important standards operating at a radio frequency of 2.4GHz are IEEE 802.11b and Bluetooth. Bluetooth, for example is predicted to be added to 80% of mobile telephones and over 1 million personal computers in 2002, and to reach a market volume of $7 billion by 2005 (Business Wire, Oct. 2000). Presently, a variety of technologies are in use by providers of Bluetooth hardware, including a 0.18μm Si CMOS only system by Cambridge Silicon Radio. Where extra amplification is required, Si$_{1-x}$Ge$_x$ HBT LNAs are a common choice. With a target price of $5 for a Bluetooth system, the lower cost of Si$_{1-x}$Ge$_x$ LNAs compared with GaAs LNAs is advantageous. A drawback however, is the fact that harnessing the benefits of Si$_{1-x}$Ge$_x$ presently requires a separate HBT device or an expensive BiCMOS process - hampering efforts at reducing cost through greater integration. Ultimately, one may hope to produce a complete system in Si$_{1-x}$Ge$_x$ MOS.

Higher operating frequencies are demanded by other standards, for example ETSI HiperLAN2 and IEEE 802.11a - both operating at a radio frequency of 5GHz and placing heavy demands on the power consumption and noise performance of active devices.

The calls for high performance and low cost create an excellent market opportunity for Si$_{1-x}$Ge$_x$ MOS technology. A useful interim step might be the
use of pseudomorphic p-channel $\text{Si}_{1-x}\text{Ge}_x$ devices of the type discussed in this thesis to boost the performance of $\text{Si}_{1-x}\text{Ge}_x$ pMOS. A substantial step towards reducing the cost of GaAs technology was recently announced by Motorola, with a process offering GaAs integrated on a Si wafer (EE Times, Sept. 2001): $\text{Si}_{1-x}\text{Ge}_x$ MOS research must move quickly toward commercial production to capitalise on this opportunity, especially in the face of competition from the giant Si MOS industry, which keeps delivering performance improvement and cost reduction in defiance of predictions that it will one day saturate.

6.2 Suggested future work

There is extensive scope for further application of and improving the detail in the modelling presented in this thesis. Some possibilities and their implications are given here.

The treatment of band structure in this thesis is somewhat simplistic: we have assumed an infinite confining potential and the bands to be parabolic and isotropic. We have also assumed throughout that carriers remain in the lowest energy subband arising from the quantum confinement. This last assumption is well approximated in the samples studied here and could, in principle, be relaxed by means of defining an approximate relaxation time for each subband, and coupling these together via inter-subband scattering. As for the infinite confining potential, a finite potential would result in a shift of the subband minimum energy - without affecting transport properties. Energetic carriers may be scattered out of the quantum well. This may perhaps
be important at high temperature in samples with a low Ge content (and thus shallow quantum well). This effect could be approximately accounted for by cutting off the appropriate integrals over energy (e.g. equation (3.18)) at the energy corresponding to the top of the potential barrier. In practically encountered cases, the error resulting from assuming an infinitely deep quantum well is small. In any case, unless the carriers are largely confined to the alloy layer, the benefits of the heterostructure will not be observed.

The assumption of parabolic and isotropic bands, common in much transport modelling, is not so easily justified. For low carrier energies, there is evidence that it is reasonable [Hinckley & Singh, 1990a], [Hinckley & Singh, 1990b]. These approximations will fail at higher energies. The full band structure is complicated [Hinckley & Singh, 1990a], [Hinckley & Singh, 1990b] and would perhaps best be treated by a Monte-Carlo simulation. The means of modifying the theory presented here to take proper account of the band structure is not clear, particularly in the case of multiple scattering. At the least, an analytical approximation to the band structure which captures enough detail to make the exercise worthwhile would be needed. Such an approximation does not yet exist and, in its absence, we are probably safe in assuming that the effects of deviation from the parabolic, isotropic case can be incorporated to some degree through the use of a phenomenological value for the effective mass. This approach is at least consistent, in that the value of effective mass is usually extracted from transport measurements.

The expressions used in this work and their underlying assumptions are best suited to low temperature situations. At room temperature, an obvious target for improvement is the treatment of phonon scattering. A large literature concerning linear transport theory with full account taken of phonon scat-
tering exists (e.g. [Nag, 1980]). It would be straightforward to incorporate appropriate elements of these treatments in the present case, but to gain any great benefit by so doing would require better information about certain material parameters, e.g. deformation potentials, than is currently available. Given uncertainties in such parameters and in issues such as the applicability of selection rules for certain transitions, it is not clear that one should have any great faith in the results of more detailed modelling - at least until the experimental evidence has improved.

Another matter of concern lies in the thinness (perhaps 2nm or less) of the capping layer separating the alloy from the gate oxide in structures being considered for deep sub-micron devices. The proximity of carriers to the oxide layer may introduce significant extra sources of scattering. In particular, roughness and trapped impurities will doubtless be present at the Si/\(\text{SiO}_2\) interface as well as at the Si/Si\(_{1-x}\text{Ge}_x\) heterointerface. As a first approximation, one could treat the former in the same way as the latter. Difficulty may arise in the understanding of how roughness at the two interfaces is correlated: measurement of the characteristics of these interfaces would be an important experimental contribution.

As yet, no suitable experimental results with which to compare the calculations of transport time to state lifetime ratio made in chapter 5 exist. Such measurements would provide more detailed information about scattering processes than does measurement of mobility alone. It would be interesting to investigate more carefully the applicability of the multiple scattering theory in samples with low carrier concentrations, though one may expect to encounter the greatest uncertainties here, in just that regime where the theory should be most relevant. Further, while we have been mostly concerned with
modelling the relation between mobility and carrier concentration at fixed temperature, modelling the variation of mobility with temperature for fixed carrier concentration (and comparing with experimental results) may give some insight into the accuracy of the treatment of screening. Lastly, there is considerable scope for investigation of alloy scattering in Si$_{1-x}$Ge$_x$, perhaps using structures expressly fabricated for the purpose: particularly, the exact value of the alloy scattering potential, $U_{AL}$, remains hard to pin down.
Bibliography


108


