Monte Carlo simulation of silicon-germanium transistors

Yangthaisong, Anucha

Monte Carlo Simulation of Silicon-Germanium Transistors

Anucha Yangthaisong

A thesis submitted for the
degree of Doctor of Philosophy
at the University of Durham,
Department of Physics

October 2002

Declaration

The work presented in this thesis has been carried out by the candidate (except where otherwise acknowledged) and has not been previously submitted for any degree.

Ph.D. Candidate

Ph.D. Supervisor

Acknowledgements

I am indebted to the numerous friends and colleagues who have made this work possible and have contributed towards my enjoyment of the time spent at Durham. Special thanks go to Prof. Dick Abram for his supervision and encouragement during the course of my study, and for his comments on the first draft of this thesis. I am also very grateful to Gavin Crow for all his assistance. I would also like to thank Lydia Heck, Stuart Clark, Stuart Brand, Hadi Arabshahi, Daniel Harrison, Rick Coles, David Dugdale, Ian Bolland, Vicente Salinero and other members of Condensed Matter Theory Group (CMT) for discussion and practical help.

I gratefully acknowledge financial support for the period of my study from the Royal Thai Government (MOSTE).

I would also like to thank my family for their encouragement. I appreciate Ketesuda Panpho for her support, especially during the time spent writing this thesis. Last but not least, I would also like to thank Su Chen, Pattamaporn Chompoonutprapa, Pornsawan Amornsakchai, Mika Toronen, Ruchaboom Bunrawd, Auttakit Chattaraputi and others for helping me to take a regular break from my work to discuss on various topics.

"Hark, O Goddess! Some people in this world strive to get results for their endeavours even if they don't succeed. Hark, O Goddess! You do see clearly the results of actions, don't you? All the others have drowned in the ocean; we alone, are still swimming and have seen you hovering near us. As for us, we are going to endeavour further to the utmost of our ability; we are going to strive like a man should to reach the shores of the ocean."

His Majesty King Bhumibol Aduljadej

(The Story of Mahajanaka)

Abstract

Self-consistent Monte Carlo simulation studies of n-channel Si/SiGe modulation doped field effect transistors (MODFETs) and silicon-on-insulator lateral bipolar junction transistors (SOI-LBJTs) are reported in this thesis.

As a preliminary to the device studies Monte Carlo simulations of electron transport in bulk Si strained as if grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ substrates have been carried out at 300 K, for field strengths varied from 10^4 to 2×10^7 Vm^{-1}. The calculations indicate an enhancement of the average electron drift velocity when Si is tensilely strained in the growth plane. The enhancement of electron velocity is more marked at low and intermediate electric fields, while at very high fields the velocity saturates at about the same value as unstrained Si. In addition the ensemble Monte Carlo method has been used to study the transient response to a stepped electric field of electrons in strained and unstrained Si. The calculations suggest that significant velocity overshoots occurs in strained material.

Simulations of n-channel $Si/Si_{1-x}Ge_x$ MODFETs with Ge fractions of 0.23, 0.25, and 0.45 have been performed. Five depletion mode devices with $x = 0.23$ and 0.25 were studied. The simulations provide information on the microscopic details of carrier behaviour, including carrier velocity, kinetic energy and carrier density, as a function of position in the device. Detailed time-dependent voltage signal analysis has been carried out to test device response and derive the frequency bandwidth. The simulations predict a current gain cut-off frequency of 60 ± 10 GHz for a device with a gate length of 0.07 μm and a channel length of 0.25 μm. Similar studies of depletion and enhancement mode n-channel $Si/Si_{0.55}Ge_{0.45}$ MODFETs with a gate length of 0.18 μm have been carried out. Cut-off frequencies of 60 ± 10 GHz and 90 ± 10 GHz are predicted for the depletion and enhancement mode devices respectively.

A Monte Carlo model has also been devised and used to simulate steady state and transient electron and hole transport in SOI-LBJTs. Four devices have been studied and the effects of junction depth and silicon layer thickness have been investigated. The advantage of the silicon-on-insulator technology SOI device is apparent in terms of higher collector current, current gain, and cut-off frequency obtained in comparison with an all-silicon structure. The simulations suggest that the common-emitter current gain of the most promising SOI-LBJT structure considered could have a cut-off frequency approaching 35 ± 5 GHz.

Contents

1 Introduction **1**

1.1 Structure of the thesis 3

2 Si/Ge Materials **6**

2.1 Introduction 6

2.2 Properties of $Si/Si_{1-x}Ge_x$ structures 7

2.2.1 Pseudomorphic epitaxial layers 8

2.2.2 Critical Thickness 9

2.2.3 Effect of strain on band structure 12

2.2.4 Effect of strain on effective masses 18

2.2.5 Effect of strain on mobilities 18

2.2.6 Band offset and band alignment 19

2.3 Field Effect Transistor 22

2.4 Concept of Modulation doping 24

2.5 n-$Si/Si_{1-x}Ge_x$ MODFET 28

2.6 npn-Si SOI LBJT 30

2.7 Summary 33

3 Monte Carlo Simulation **36**

3.1 Introduction 36

3.2 The Boltzmann transport equation 37

3.3 Monte Carlo simulation 39
3.4 Band Structure 40
3.5 Scattering Mechanisms 46
3.5.1 Lattice Scattering 47
3.5.2 Ionised impurity scattering 52
3.5.3 Alloy scattering 53
3.6 Types of Monte Carlo simulation 54
3.6.1 One-particle simulation 54
3.6.2 Ensemble simulation 54
3.6.3 Self-consistent simulation 55
3.7 Device Simulation Model 55
3.8 Transport Model 65
3.8.1 Band structure 65

4 MODFET Simulations **69**

4.1 Introduction 69
4.2 Simulation of electron transport in bulk Si and strained Si 71
4.2.1 Steady state simulation results 71
4.2.2 Transient response simulation results 75
4.3 Device simulation 77
4.4 Results for moderate tensilely strained Si n-channel MODFETs 81
4.4.1 Layer design 81
4.4.2 Device Characteristics and Analysis 86
4.5 Results for high tensilely strained Si n-channel MODFETs 114
4.5.1 Layer design 114
4.5.2 Device Characteristics and Analysis 117
4.6 Summary and Conclusion 127

5 SOI-LBJT Simulations **130**

5.1 Introduction 130

5.2 Device Simulation 132

5.3 Device Characteristics and Analysis 137

5.3.1 DC characteristics 137

5.3.2 Signal Analysis 140

5.3.3 Microscopic Analysis 142

5.4 Summary and Conclusion 155

6 Summary, Conclusions and Suggestions for Further Works **157**

6.1 Summary & conclusions 157

6.2 Suggestions for Further work 160

References **162**

Chapter 1

Introduction

Despite the superior high-speed performance of transistors based on III-V materials, silicon is still the dominant semiconductor in modern microelectronics, with applications that can be found in both discrete devices and monolithic integrated circuits (ICs). A metal-oxide-semiconductor field-effect transistor (MOSFET) is a building block of VLSI circuits in microprocessors and dynamic memories. The n-type device involves controlling the current flow through a channel whilst a thin layer of SiO_2 is used to separate the gate electrode and p-silicon. Complementary MOSFET (CMOS) technology, which combines both n-channel and p-channel MOSFETs, provides very low power consumption combined with high speed. The performance and packing density of VLSI circuits is still improving as a result of the continuing miniaturisation of MOSFETs and advances in circuit fabrication techniques. However, enhancement of submicron device performance by reducing device size is facing more and more difficulties [1,2]. As a result, several new device concepts and new technologies have been developed in recent years in order to overcome or circumvent the problems. Two devices are investigated in this thesis, which might play an important role when the down-scaling of critical device dimensions is no longer an affordable option for bulk production.

Strain has the effect of changing the band structure and other parameters of a material so that it has different electronic properties and this provides a method of tailoring

device performance. It is possible to fabricate an n-channel MODFET structure, in which a thin silicon channel is under tensile strain as a result of growing it in a thin layer on a silicon-germanium alloy. The tensile strained Si provides a confining channel for the electrons. In addition the strain-induced splitting of the Si conduction band valleys means that the conductivity effective mass is effectively reduced and intervalley scattering is suppressed, resulting in an enhanced electron mobility. It is also possible to achieve spatial separation of the electrons from the ionised donors which supply them by only doping the structure away from the channel. This has the advantage of reducing ionised impurity scattering of the electrons. Due to these effects, modulation doped and strained $Si/Si_{1-x}Ge_x$ ($x < 0.5$) structures have been shown to exhibit high-electron mobility, and hence have considerable potential for the development of fast Si-based MODFETs.

Field-effect transistors are used in the majority of VLSI circuits, but bipolar transistors are valuable in applications requiring high speed, high current and lower integration levels than state of the art circuits. Devices that take advantage of both the high current-drive capability of high speed bipolar transistors and the mass production techniques of CMOS technology are particularly attractive. Bipolar junction transistors with an unconventional lateral geometry can be incorporated into integrated circuits with only minor modifications to CMOS processing. In particular the use of silicon-on-insulator for lateral bipolar devices permits simple circuit integration and results in very small junction capacitances, ease of isolation, and a CMOS-like device area. These features are particularly attractive for dense high-performance CMOS-type applications.

Carrier transport in submicron semiconductor devices is complicated by the rapid spatial and temporal variations in electric fields and associated carrier dynamics that can occur. The modelling of such devices requires physics-based simulators that can provide a realistic description of non-equilibrium carrier transport. The advantage of

accurate device simulation based on firm physical foundations is that it can be considered as a form of experiment, which can be carried out much faster and with less expense than real experimental investigation. Furthermore, computer simulation provides a wealth of microscopic detail and the potential for a deep physical interpretation of the results that can point the way to enhancements in device performance.

This thesis is concerned with the modelling of electron transport in n-channel strained $Si/Si_{1-x}Ge_x$ MODFETs with Ge contents of 0.23, 0.25, and 0.45 and electron and hole transport in silicon-on-insulator lateral bipolar transistors (SOI-LBJTs). For both types of device the self-consistent ensemble Monte Carlo method is used. The thesis is organised as follows:

1.1 Structure of the thesis

Chapter 2

Chapter 2 describes the elastic and electronic properties of Si and SiGe alloys, with particular emphasis on the effects of strain on the electronic band structure, and the transport properties of carriers, and how the effects can be utilised in high performance devices. The two types of transistors which are the subject of study, the modulation doped field-effect SiGe transistor (MODFET) and the lateral bipolar junction transistor (LBJT), are also introduced.

Chapter 3

In this chapter, the Boltzmann transport equation (BTE) is introduced as a semiclas-

sical description of charge carrier transport in semiconductors. The Monte Carlo (MC) method is also introduced as a numerical method of solving the BTE by tracking the trajectories of particles representative of electrons and holes which move through a device under the influence of the local electric field and the stochastic scattering by phonons, impurities and alloy disorder. The chapter concludes with a description of the use of the Monte Carlo model for the device simulations which produce the results presented in Chapter 4 and 5.

Chapter 4

Chapter 4 presents the results of simulations of n-channel strained $Si/Si_{1-x}Ge_x$ MODFETs using the self-consistent ensemble Monte Carlo method. The devices studied can be categorised as being in two groups according to their Ge content. The moderately strained devices have a Ge fraction $x = 0.23$ or 0.25 and the highly strained devices have $x = 0.45$. Macroscopic device characteristics and parameters, such as the drain current-drain voltage curve and, the transconductance (g_m) are obtained from the simulations. The simulations also provide information on the microscopic details of carrier behaviour, including carrier velocity, kinetic energy and density as a function of position in the device, facilitating a thorough analysis of the mechanisms determining device performance. The device response to a time-dependent voltage signal has also been simulated to derive the frequency bandwidth for various geometries.

Chapter 5

In this chapter, we report how a Monte Carlo simulation has been devised and used to model steady state and transient electron and hole transport in SOI-LBJTs. Four devices are studied in order to investigate the effects of junction depth and silicon layer

thickness. Simulations have also been carried out to investigate the high frequency performance of the devices.

Chapter 6

Conclusions drawn from this work are presented in Chapter 6 and suggestions are made for further study.

Chapter 2

Si/Ge Materials

2.1 Introduction

To overcome the physical limits on the speed of the field effect transistor while maintaining compatibility with conventional integrated circuit fabrication processes, several innovations have been developed in recent years. In particular Si/SiGe heterostructures have been introduced as a promising method for improving device speed without reduction of device size while solving a number of other problems encountered in MOSFETs [1,2]. Much recent research in $Si/Si_{1-x}Ge_x$ heterostructures has had the aim of obtaining higher hole and electron mobilities through strain-induced band splitting, and enhanced mobilities in both n-channel and p-channel devices have been reported in recent years [3,4]. Fabrication processes for Si/SiGe devices are quite compatible with those routinely used in Si ICs, in contrast to the situation with III-V compound devices [5]. Hence, mobility-enhanced field-effect transistors based on SiGe could not only fit well into the area of mainstream microelectronics, but might also provide the performance advantage that may ultimately be decisive when the down-scaling of critical device dimensions is no longer an affordable option for bulk IC production.

The physics of Si/SiGe materials that is relevant to the operation of the devices and to the simulation model that we use to investigate them is presented in this chapter.

In Section 2.2 the basic properties of Si, Ge and the alloy $Si_{1-x}Ge_x$ in the bulk are briefly introduced and then the effect of strain on the band structure, effective masses, and mobilities of Si and $Si_{1-x}Ge_x$ are described. A knowledge of the band alignment of a $Si/Si_{1-x}Ge_x$ heterojunction is also required to describe in quantitative terms the control of carrier transport in the heterostructures, and information on this is also given in Section 2.2. The basic concept of the field-effect transistor and the technique of modulation doping are introduced in Sections 2.3 and 2.4 respectively. Also in Section 2.4, we describe how the changes due to the strain-induced band structure can be used to tailor device performance. Section 2.5 deals with the use of heterojunctions and strain in specific n-$Si/Si_{1-x}Ge_x$ MODFETs. The basic operation of the silicon-on-insulator bipolar transistor (SOI LBJT) is described in Section 2.6.

2.2 Properties of $Si/Si_{1-x}Ge_x$ structures

Si and Ge are indirect bandgap materials and the fundamental bandgap of $Si_{1-x}Ge_x$ alloys is also indirect for all compositions. The conduction band edge minima are sixfold degenerate in Si, where they are located along the [100] directions near the X-points of the Brillouin zone and are referred to as Δ-minima. In Ge, the conduction band edge minima are located right at the Brillouin-zone edge in the [111] directions (L-minima). In Si, the six equivalent minima of the conduction band are ellipsoids of revolution and therefore are characterised by different effective masses in the various directions. There is a longitudinal mass $m_l = 0.91m_0$, which describes the curvature in the major axis direction, and a transverse mass $m_t = 0.19m_0$, which relates to the two minor axes. For Ge, there are eight half ellipsoids equivalent to four complete ellipsoids in the periodic zone scheme, which have a transverse mass of $m_t = 0.82m_0$ and a longitudinal mass of $m_l = 1.64m_0$. According to the work of Braunstein *et al.* [6], the lowest lying conduction bands cross over in the alloy from the Δ Si-like states to the L Ge-like states for a Ge content of 85 %. Consequently, the conduction band structure of bulk

SiGe alloy is Si-like for a wide range of alloy compositions.

According to Vegard's law, the lattice constant of unstrained $Si_{1-x}Ge_x$ is given by

$$a(x)_{Si_{1-x}Ge_x} = (1-x)a_{Si} + xa_{Ge} \tag{2.1}$$

That is, the lattice constant of $Si_{1-x}Ge_x$ can then be varied in a continuous fashion from that of Si to Ge. Note that the lattice constant $a(x)$ of the $Si_{1-x}Ge_x$ alloy is bigger than that of Si since $a_{Si} = 5.430$ Å and $a_{Ge} = 5.650$ Å.

2.2.1 Pseudomorphic epitaxial layers

Consider the effect of attempting to grow an epilayer of a material with cubic crystal structure and lattice constant a_{epi} on a substrate with the same type of crystal structure but a different lattice constant a_{sub}. Our attention is restricted to the growth along the (100) axis of the crystal. The lattice mismatch is described by a misfit parameter f_m as:

$$f_m = \frac{a_{epi} - a_{sub}}{a_{sub}} \tag{2.2}$$

where a_{epi} is the lattice constant of the epitaxial layer (or epilayer) and a_{sub} is the lattice constant of the substrate. A match between the two crystalline structures can be achieved only if one or both crystals are elastically strained. In general, the substrate is much thicker than the growing epitaxial film, and then it is the atoms of the epitaxial layer that must be displaced. If $f_m < 0$ there is an in-plane (biaxial) tensile strain in the epilayer and also the lattice constant normal to the plane is reduced. Alternatively, if $f_m > 0$ there is an in-plane (biaxial) compression of the epilayer and tensile strain normal to the plane. In both cases there is a tetragonal distortion of the cubic unit cell of the epilayer crystal structure as schematically illustrated in Figure 2.1. When the epilayer grows with this type of simple elastic distortion of the crystal structure, without the appearance of dislocations, the growth is said to be pseudomorphic.

2.2.2 Critical Thickness

The basic understanding of the growth of pseudomorphic layers was first provided by Frank and van der Merwe [7] but a more complete analysis of the growth of lattice-mismatched layers was performed by Matthews and Blakeslee [8]. They pointed out that the growth of a sufficiently thin epitaxial layer, whose lattice constant is close to but not equal to the lattice constant of the substrate, may give rise to a coherently strained layer as opposed to a polycrystalline or amorphous layer. In particular, there exists a critical thickness (h_c) of the epilayer below which the film is thermodynamically stable and no misfit dislocations exist in the interface. For layers with thickness above h_c, misfit dislocations become energetically favourable and provide partial strain relaxation of the film, the degree of which increases with increasing layer thickness. According to the theory of Matthews and Blakeslee (MB) [8], the critical thickness can be calculated from [9]:

$$h_c = \frac{b^2(1-\nu\cos^2\beta)}{8\pi f_m(1+\nu)b_1}\ln\frac{\rho_c h_c}{q} \tag{2.3}$$

where b is the Burgers vector, ν is Poisson's ratio, β is the angle between the dislocation line and the Burgers vector, $b_1 = b\cos\beta$, q is the core cut-off parameter, which is taken to be equal to b_1, and ρ_c is the core energy parameter. Note that the formula does not take account of the interactions between dislocations and a more rigorous calculation of h_c has been performed by Jain *et al.* [9]. Nonetheless, accurate predictions of h_c differ only slightly from the values obtained from Equation 2.3.

Thus, when a layer of $Si_{1-x}Ge_x$ is grown on a Si substrate, it becomes compressively strained with the lattice symmetry changing from cubic to tetragonal if the growth is pseudomorphic. On the other hand, if the layer thickness is above the critical thickness, there are misfit dislocations and other defects and the carrier mobility in the material

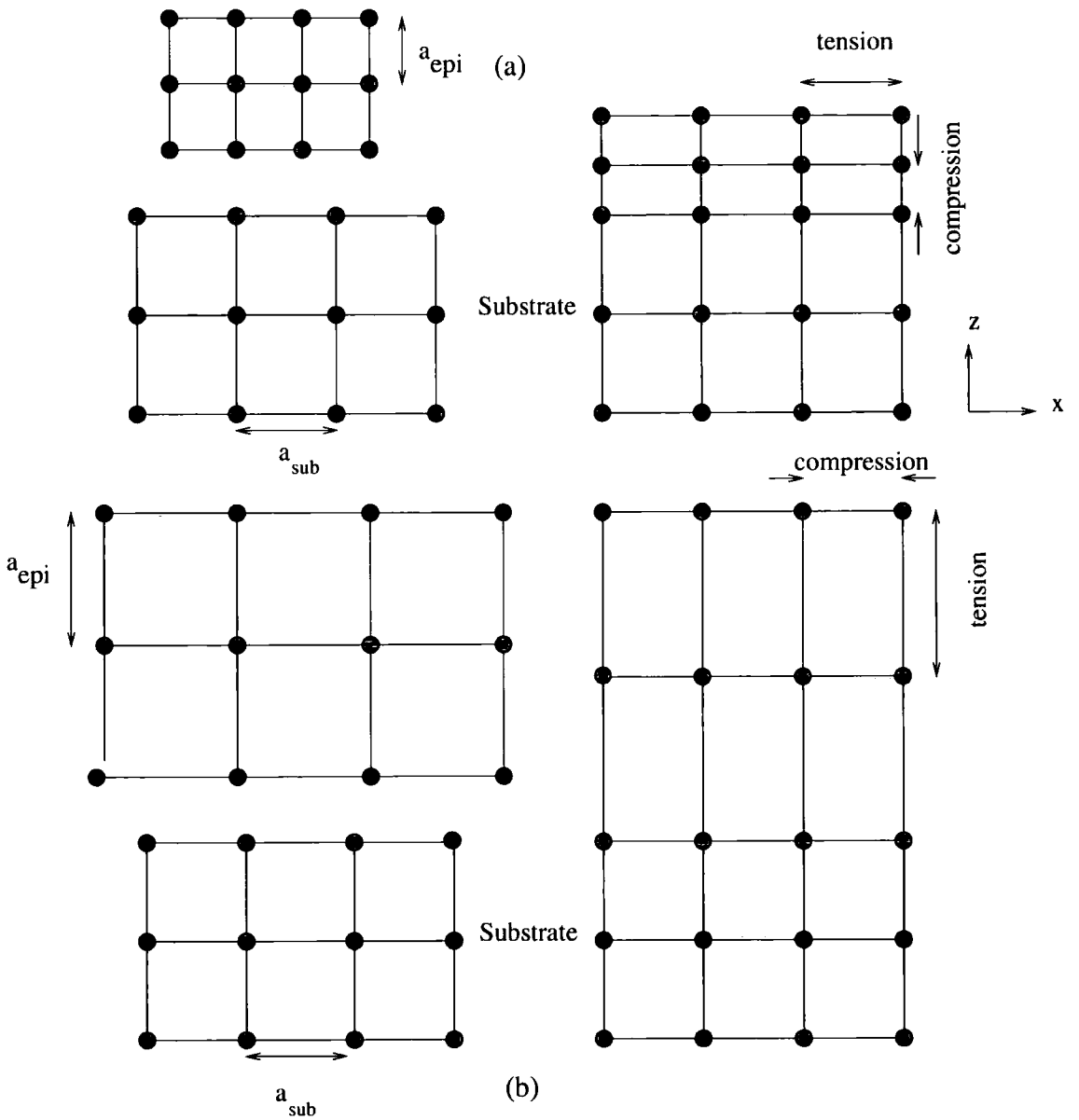
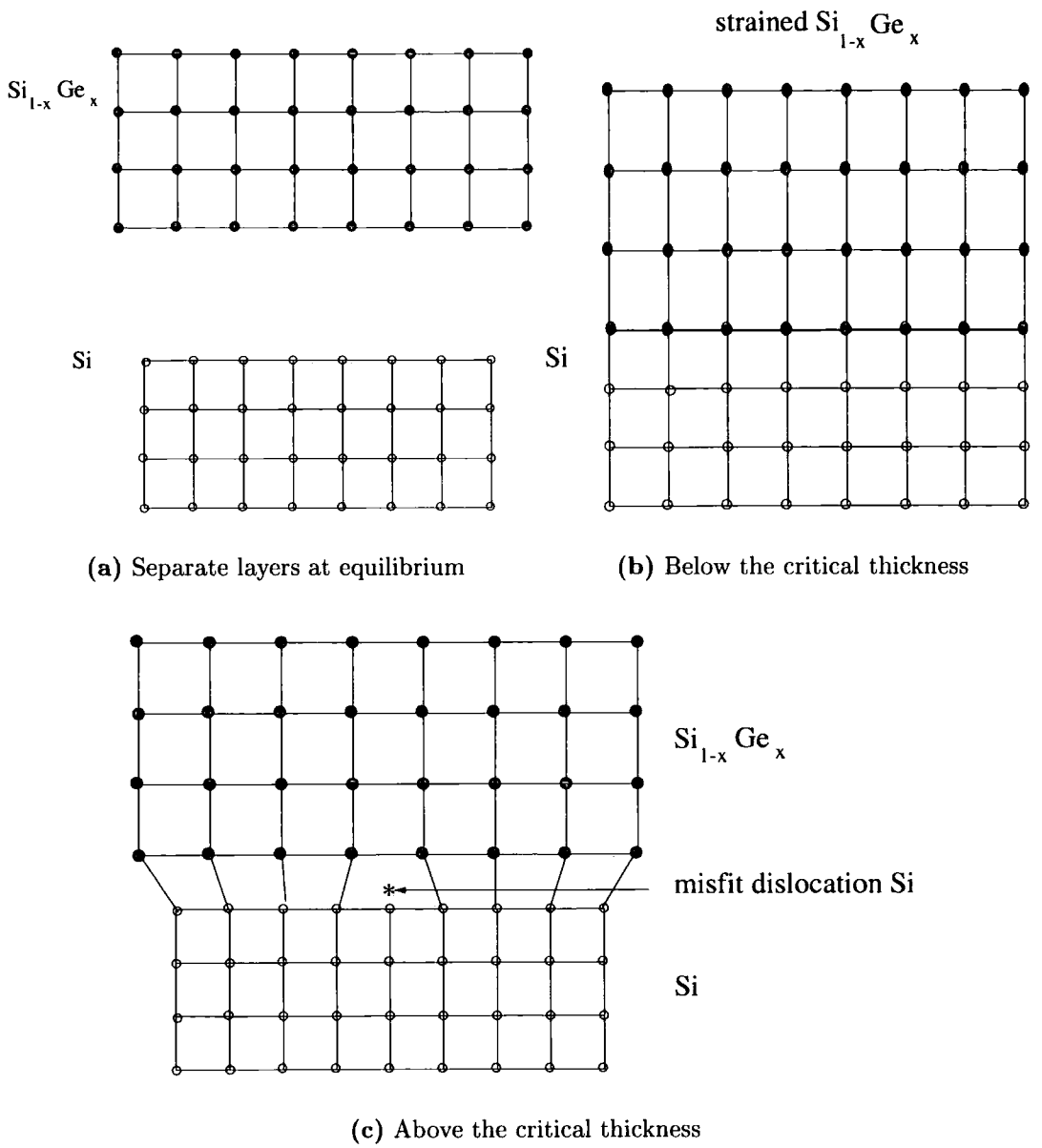


Figure 2.1: A material with a lattice constant a_{epi} grown on a substrate with a lattice constant a_{sub}: a) $a < a_{sub}$ b) $a > a_{sub}$



(a) Separate layers at equilibrium

(b) Below the critical thickness

(c) Above the critical thickness

Figure 2.2: Schematic illustration of **(a)** Separate layers at equilibrium. There is a 4.2% difference in the lattice constant of Si and Ge. Thus when a layer of $Si_{1-x}Ge_x$ is grown on top of Si, it has a bulk relaxed lattice constant which is larger than Si. **(b)** If a layer grown is below the critical thickness it becomes strained with the lattice symmetry changing from cubic to tetragonal. **(c)** If the layer thickness is above the critical thickness, misfit dislocations appear, which act to relieve the strain in the epitaxial film.

is reduced by the high defect density, making it unsuitable for device applications. These two cases are illustrated in Figure 2.2. Figure 2.3 illustrates the opposite to Figure 2.2(b) in the sense that an epilayer of Si is grown on a SiGe substrate resulting in biaxial tensile strain.

2.2.3 Effect of strain on band structure

In this section, we examine the effect that strain in an epitaxial layer has on the electronic properties of the semiconductor. To this end, it is important to know the strain tensor for the epitaxial layer. Once the strain tensor is known, the effect of strain on the band structure can be obtained from deformation potential theory. In the present work, the epitaxial semiconductor layer is biaxially strained in the plane of the substrate by an amount ϵ, and uniaxially strained in the normal direction by an amount $\epsilon_\perp$.

The definition of stress, strain and their relationship are discussed in many textbooks, e.g. Kittel [10], and only the essential points will be presented here. A strain within the elastic limit is defined by the components of a symmetric second-rank tensor ε:

$$\varepsilon_{ij} = \frac{1}{2}(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}), \quad \text{where} \quad i,j = 1,2,3 \tag{2.4}$$

Note that the diagonal part ε_{ii} is the specific increment of length in direction i and the dilation θ (the specific increment of the volume V) is simply the trace of ε

$$\theta = \frac{\Delta V}{V} = \varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} = Tr(\varepsilon) \tag{2.5}$$

The nondiagonal elements of ε_{ij} $(i \neq j)$ correspond to shearing deformations. The stress-tensor σ is defined by its components σ_{ij}, where $i,j = 1,2,3$. According to

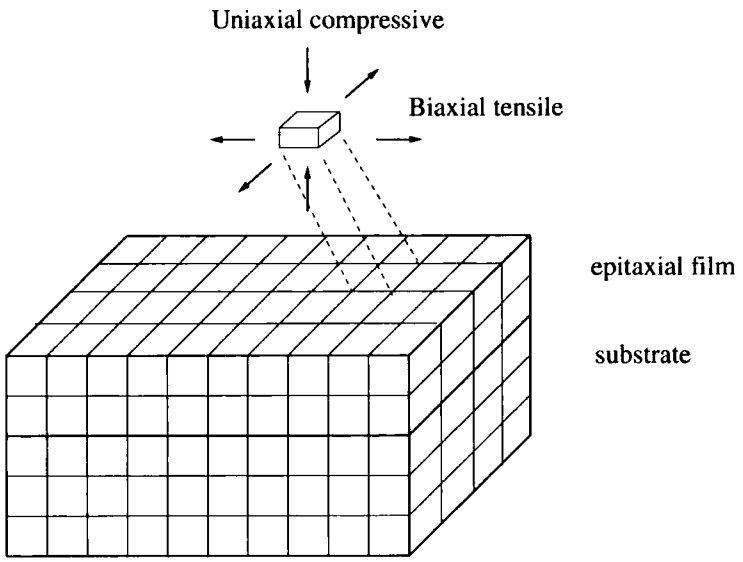


Figure 2.3: Schematic diagram of an epitaxial film of Si grown on $Si_{1-x}Ge_x$ under tetragonal distortion with in-plane biaxial tensile strain and out-of-plane uniaxial compressive strain.

Hooke's law, the stress-strain relations are

$$\varepsilon_\mu = \sum_{\eta=1}^{6} s_{\mu\eta}\sigma_\eta \tag{2.6}$$

where in the conventional reduced notation the values of μ and η have the meaning $1 \equiv xx, 2 \equiv yy, 3 \equiv zz, 4 \equiv yz, 5 \equiv zx, 6 \equiv xy$ and $s_{\mu\eta}$ are the elastic compliance constants. The stress components can be written as a function of the strain components:

$$\sigma_\mu = \sum_{\eta=1}^{6} c_{\mu\eta}\varepsilon_\eta \tag{2.7}$$

where $c_{\mu\eta}$ are the elastic stiffness constants. The energy is a quadratic function of the strain, and can be written as

$$U = \frac{1}{2}\sum_{\lambda=1}^{6}\sum_{\mu=1}^{6} \tilde{c}_{\lambda\mu}e_\lambda e_\mu \tag{2.8}$$

where $\tilde{c}_{\lambda\mu}$'s are related to the c's. For cubic systems, the elastic energy U simplifies to

$$\begin{aligned} U &= \frac{1}{2}c_{11}(e_{xx}^2 + e_{yy}^2 + e_{zz}^2) + \frac{1}{2}c_{44}(e_{yz}^2 + e_{zx}^2 + e_{xy}^2) \\ &+ c_{12}(e_{yy}e_{zz} + e_{zz}e_{xx} + e_{xx}e_{yy}) \end{aligned} \tag{2.9}$$

The non-vanishing components of the stress tensor ($\sigma(epi)$) of the epitaxial layer are $\sigma_{xx}(epi)$ and $\sigma_{yy}(epi)$ with $\sigma_{xx}(epi) = \sigma_{yy}(epi)$. The independent stress-strain relations for the epitaxial layer are

$$\begin{aligned} \sigma_{xx}(epi) &= c_{11}(epi)\varepsilon_{xx}(epi) + c_{12}(epi)[\varepsilon_{yy}(epi) + \varepsilon_{zz}(epi)], \\ \sigma_{zz}(epi) &= c_{11}(epi)\varepsilon_{zz}(epi) + c_{12}(epi)[\varepsilon_{xx}(epi) + \varepsilon_{yy}(epi)] \end{aligned} \tag{2.10}$$

with $c_{11}(epi)$ and $c_{12}(epi)$ being the elastic stiffness constants of the epitaxial layer. The in-plane strain $\varepsilon_{||}$ $(= \varepsilon_{xx} = \varepsilon_{yy})$ is defined for the epitaxial layer

$$\varepsilon_{||} = \frac{a_{sub}}{a_{epi}} - 1 \tag{2.11}$$

Since there is no stress along the perpendicular direction, the perpendicular strain $\varepsilon_{\perp}$ (ε_{zz}) can be written as

$$\begin{aligned} \varepsilon_{\perp} &= -2\frac{c_{12}}{c_{11}}\varepsilon_{||} \\ &= \frac{-2\sigma}{1-\sigma}\varepsilon_{||} \end{aligned} \tag{2.12}$$

where σ is Poisson's ratio.

Once the strain tensor is known, deformation potential theory can be used to calculate the effect of the strain on the electronic band structure. Deformation potential theory describes the effect of strain in terms of deformation potentials which are obtained for each band by experimental measurements of the band structure with applied stress. The strain perturbation Hamiltonian is defined and its effects are calculated in first order perturbation theory. Here, we will summarise the relevant results of that formalism.

There are three basic elements significantly affecting the band structure in strained films [11–14];

1) changes in bandgaps due to hydrostatic strain,

2) removal of the degeneracy of states due to uniaxial strain,

3) shift in the spin orbit split off (Δ_{so}) band.

Hydrostatic strain shifts the average position of the conduction band, $\Delta E_{c,av}$, according to the formula

$$\Delta E_{c,av} = a_c \frac{\Delta V}{V} \tag{2.13}$$

where the net change in volume of the unit cell $\Delta V/V = Tr(\vec{\varepsilon}) = 2\varepsilon_{||} + \varepsilon_{\perp}$ and a_c is the hydrostatic deformation potential for the conduction band. Hydrostatic strain also shifts the average position of valence band, which can be expressed in a similar form to Equation 2.13. The general effect of hydrostatic strain is to increase the band gap with compressive strain and to decrease the gap with tensile strain.

However, the strain in a pseudomorphic epitaxial layer has both hydrostatic and uniaxial components. For example, if silicon is grown pseudomorphically on $Si_{1-x}Ge_x$, the in-plane lattice constant changes by Δa while the lattice constant in the growth direction changes by $\Delta a_{\perp}$ where $\Delta a_{\perp} = -\Delta a_{||}/\sigma$. The deformation does not retain the symmetry of the material as in the hydrostatic case, and the strain can be considered to be the sum of hydrostatic and uniaxial components. The effect of uniaxial strain on the six-fold Δ valleys of the degenerate conduction band of Si (and SiGe with a Si-like lower conduction band) is that the valley degeneracy is partly removed. The valleys split into two sets, one set being a doublet and the other a quadruplet as shown in Figure 2.4. Denoting the four equivalent in-plane axes by Δ_4 and the two valleys on the axis parallel to the growth direction by Δ_2, the energy shifts are

$$\Delta E_c^{\Delta} = \begin{cases} +\frac{2}{3}\Xi_u^{\Delta}(\varepsilon_{\perp} - \varepsilon_{||}) & \text{for the } \Delta_2 \text{ valleys} \\ -\frac{1}{3}\Xi_u^{\Delta}(\varepsilon_{||} - \varepsilon_{\perp}) & \text{for the } \Delta_4 \text{ valleys} \end{cases} \tag{2.14}$$

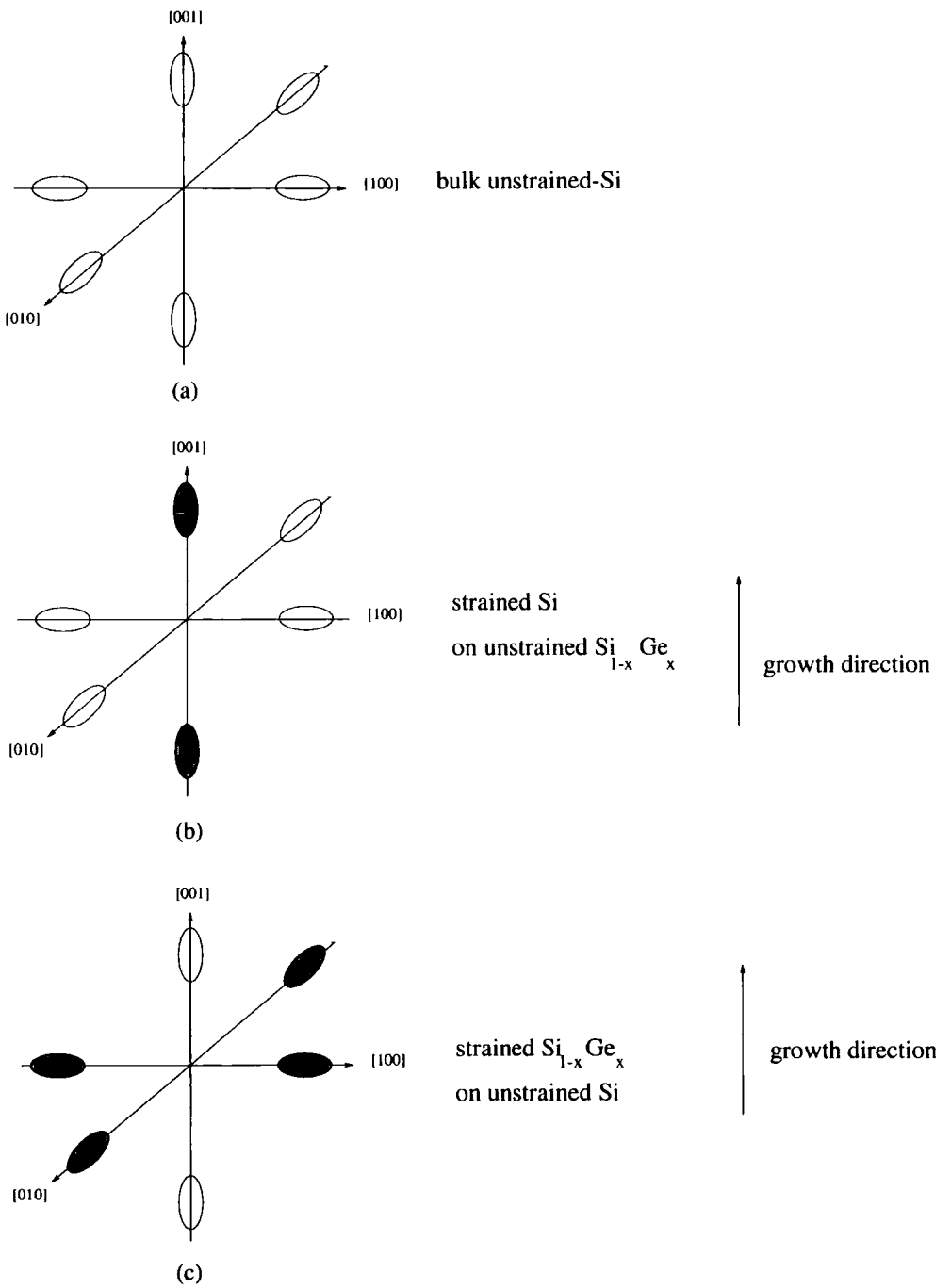


Figure 2.4: **(a)** The constant energy surfaces of the conduction bands of silicon. There are six equivalent conduction bands located along the [100], [010], and [001] directions. **(b)** The constant energy surfaces of strained Si grown on a [001] $Si_{1-x}Ge_x$ virtual substrate. The [001] conduction bands have shifted downwards in energy relative to the [010] and [100] bands. **(c)** The constant energy surfaces of strained $Si_{1-x}Ge_x$ grown on a [001] Si substrate. The [010] and [100] bands have shifted downwards in energy relative to the [001] bands.

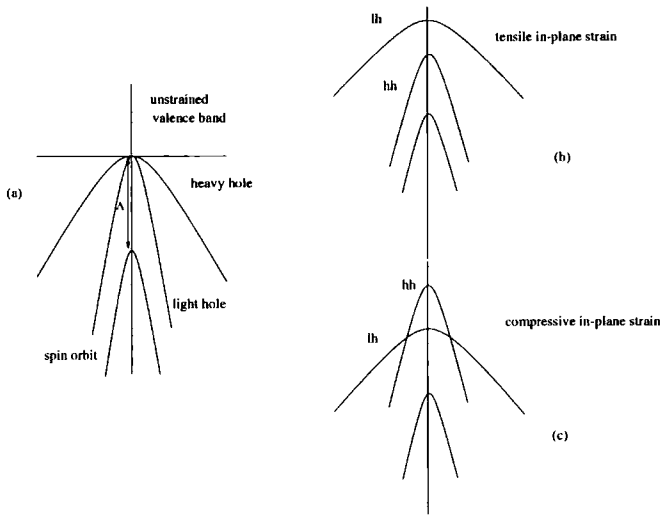


Figure 2.5: **(a)** Valence band structure in the absence of strain, **(b)** effect of tensile in-plane strain, **(c)** as **(b)** but for compressive in-plane strained band structure

where Ξ_u^Δ is the Δ-valley deformation potential. If Si is grown on $Si_{1-x}Ge_x$, the Si is under biaxial tensile and uniaxial compressive strain and the two valleys in black in Figure 2.4(b) have the lower energy. Note that the Ge content dependence of Ξ_u^Δ (in eV) for $Si_{1-x}Ge_x$ is parameterised as [14]

$$\Xi_u^\Delta = 9.29 + 0.2188x + 0.6912x^2 \tag{2.15}$$

The conduction band offset ΔE_c is about 0.12 – 0.3 eV for a Ge fraction of 0.2 – 0.5. Figure 2.4(c) shows the opposite case of the alloy epilayer on a Si substrate.

The effect of the strain in an epilayer on the valence band is to remove the heavy-hole and light-hole band degeneracy as illustrated in Figure 2.5. If the strain in the plane of the layer is compressive, the heavy-hole band is raised relative to the light-hole band; if tensile, the reverse occurs. For the former case, the in-plane effective mass of the heavy-hole band is relatively light [15], resulting in an improved hole mobility. In addition the interband scattering is reduced due to the lifting of the heavy hole-light

hole band degeneracy. On the other hand, it is expected that the tensile in-plane strain will also increase the low-field hole mobility since most holes reside in the heavy-hole band in the bulk and interband scattering is suppressed when the bands are split by strain.

2.2.4 Effect of strain on effective masses

There are no experimental data available on electron masses of SiGe alloys [16]. The theoretical investigation using non-local pseudopotential calculations performed by Rieger *et al.* [14] shows that there are only small changes of the X-valley mass parameters due to strain and Ge content. Hence, the effect of strain on effective mass is not included in our Monte Carlo transport model.

2.2.5 Effect of strain on mobilities

The carrier mobility is one of the most important parameters of a semiconductor material since the carrier velocity is proportional to the mobility for low electric fields. All other things being equal a transistor made from a higher mobility material will have a higher frequency response since carriers take less time to travel through the device. In addition, higher mobility materials imply lower resistances and lower RC time constants, again resulting in a higher frequency response.

We have already noted that the strain in Si grown on SiGe removes some of the degeneracy of the Δ valleys in unstrained Si. A typical value for the splitting is 0.18 eV which is greater than $6k_BT$ at room temperature. As a result, electrons preferentially occupy the two lower valleys (shown in black in Figure 2.4(b)) which have low in-plane masses. As the in-plane mass is much lower than the conductivity mass of bulk Si we might expect a substantial enhancement in mobility. In addition, the lower density of

states of the two valleys compared to the original six reduces the intervalley scattering and also acts to increase the electron mobility.

The initial experimental attempts to realise the predicted enhancement of electron mobility due to the effects of strain on the band structure were severely impeded by the quality of the strain-adjusting buffer layers used in an effort to produce a SiGe substrate by growing relaxed material on a Si substrate. The best mobilities achieved as a result of these initial efforts reached a value close to 17 000 $cm^2V^{-1}s^{-1}$ [17]. With better growth conditions and a breakthrough in reducing the defect densities of the $Si_{1-x}Ge_x$ buffer layer, electron mobilities now exceed 500 000 $cm^2V^{-1}s^{-1}$ at low temperature [18]. The mobilities at room temperature have also been increased to almost 3000 $cm^2V^{-1}s^{-1}$ [3], which is more than a factor of 3 higher than that in conventional Si MOSFETs [19,20], and a factor of 2 higher than the (3D) mobility of intrinsic bulk Si.

2.2.6 Band offset and band alignment

A knowledge of band alignments is essential for informed device design. As well as the normal space charge effects that occur as a result of doping we must consider the band offsets that result from the use of different materials. In the lattice-matched case, it is necessary to know just the valence-band and conduction-band offsets, ΔE_v and ΔE_c respectively. When there is a lattice mismatch, as for Si-$Si_{1-x}Ge_x$ heterostructures, we also need to know the shifts in bands that occur as a result of strain, including the strain splitting of the valence and conduction bands. In fact this information is often required for a range of different strain conditions, since there is generally some flexibility in the alloy compositions used in a particular device.

To determine the band alignment for a given $Si_{1-y}Ge_y$-Si heterointerface it is necessary to know the band gaps of the strained layer constituents and either ΔE_c or

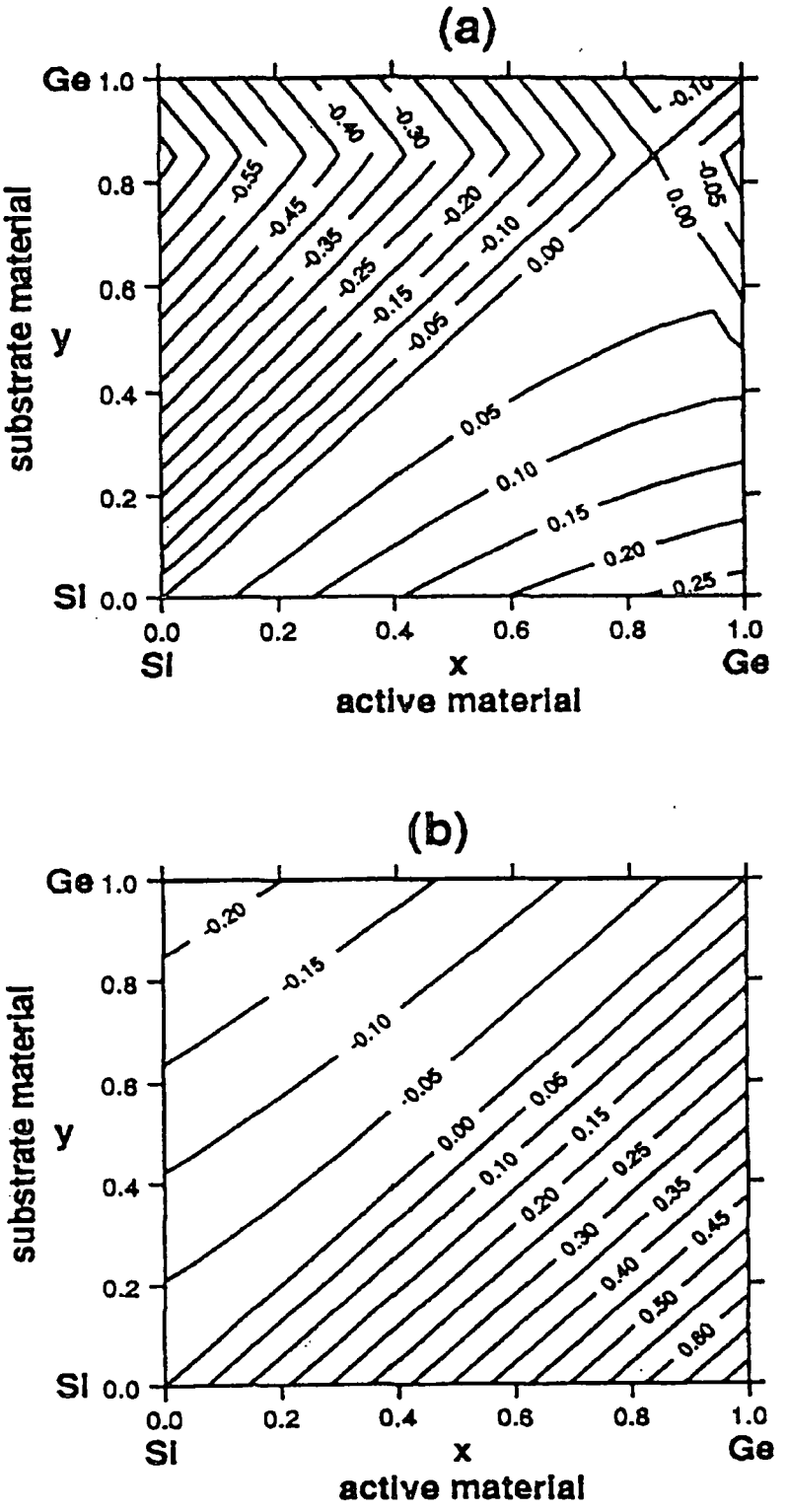


Figure 2.6: Band offsets in (a) conduction band and (b) in valence band between a substrate of $Si_{1-y}Ge_y$ and an active layer of $Si_{1-x}Ge_x$ that is strained to match the substrate [14].

ΔE_v [21]. The energy gap of strained $Si_{1-x}Ge_x$ is given by [21]:

$$E_g(Si_{1-x}Ge_x)_{strained} = E_g(Si_{1-x}Ge_x)_{unstrained} \quad (2.16)$$

$$-\frac{2}{3}(c_{11}+2c_{12}) \times (1-(c_{12}/c_{11}))\frac{dE_g}{dP} \times \varepsilon_{xx} \quad (2.17)$$

where c_{11}, c_{12} are elastic constants of unstrained $Si_{1-x}Ge_x$ and dE_g/dP is the hydrostatic pressure coefficient of the energy gap of unstrained $Si_{1-x}Ge_x$. According to the calculation of Rieger and Vogl [14], the offset between the average energy of the valence-band edge in a strained $Si_{1-x}Ge_x$ layer and an unstrained $Si_{1-y}Ge_y$ layer can be expressed as:

$$\Delta E_{v,av} = (0.47 - 0.067y)(x-y) \quad (2.18)$$

The energy gap of the alloy $Si_{1-y}Ge_y$ can be obtained by interpolating the energy gap of Si and Ge as:

$$E_g(Si_{1-y}Ge_y) = yE_g(Ge) + (1-y)E_g(Si) + y(1-y)b(SiGe) \quad (2.19)$$

where $E_g(Si) = 1.12$ eV, $E_g(Ge) = 0.975$ eV and b is the bowing parameter for the ternary alloy, which is taken to have the value -0.19 eV [22]. Once we know the valence band offset, we can use our knowledge of the bandgap in the unstrained alloy, combined with deformation potential theory, to obtain the conduction-band offset according to

$$\Delta E_c = E_g(Si_{1-y}Ge_y) + \Delta E_v(Si_{1-x}Ge_x/Si) - E_g(Si_{1-x}Ge_x)_{strained} \quad (2.20)$$

for the strained $Si_{1-x}Ge_x$ grown on unstrained $Si_{1-y}Ge_y$. Figure 2.6 shows the calculated band offsets for a layer of $Si_{1-x}Ge_x$ on a $Si_{1-y}Ge_y$ substrate. Note there are sharp changes in the way the conduction band offset ΔE_c varies with composition, which are due to the changes in the nature of the lowest minima of the conduction band in the two materials. Figure 2.7(a) relates to a strained $Si_{0.77}Ge_{0.23}$ layer on a silicon substrate. Note most of the offset is in the valence band. The type I band alignment shown is favourable for hole confinement and can be exploited in several

novel heterostructure devices. For example, pseudomorphic $Si_{1-x}Ge_x$ grown on (001) Si is used in the base region of n-p-n heterojunction bipolar transistors (HBTs), which show good high-frequency performance. In fact Si/SiGe HBTs, with cut-off frequencies and maximum oscillation frequencies well beyond 100 GHz have been reported [23]. On the other hand, Si forms the epilayer on a $Si_{0.77}Ge_{0.23}$ substrate in Figure 2.7(b). The Si is now under biaxial tension, the band alignment is type II, and there is a bigger discontinuity in the conduction band. This arrangement facilitates electron and hole confinement in different layers. Thus, it is useful for both n- and p-type devices for strained Si/SiGe-based CMOS technology.

Figure 2.8 shows the conduction (ΔE_c) and valence band (ΔE_v) offsets for strained Si-$Si_{0.77}Ge_{0.23}$ and strained Si-$Si_{0.55}Ge_{0.45}$ heterojunctions. Note that a ΔE_c of 0.139 eV is obtained for the former and 0.285 eV for the latter.

2.3 Field Effect Transistor

The concept of a field-effect transistor (FET) was first proposed by Lilienfeld and Heil in the 1930s [24,25], but only in the 1950s had semiconductor-material processing technology progressed enough that Dacay and Ross [26] were able to demonstrate working devices. Currently FET technology plays the dominant role in microelectronics, and FET devices and integrated circuits are made in a variety of designs and with many different semiconductor materials. The tremendous versatility of the device combined with high manufacturing yield and operational reliability have allowed it to become the industry workhorse for a variety of applications.

The device operates as a conducting semiconductor channel between source and drain ohmic contacts, in which the number of charge carriers is controlled by a third contact called the gate. Carriers flow from the source to the drain while the width of the channel is modulated by a potential applied to the gate as depicted in Figure 2.9(a).

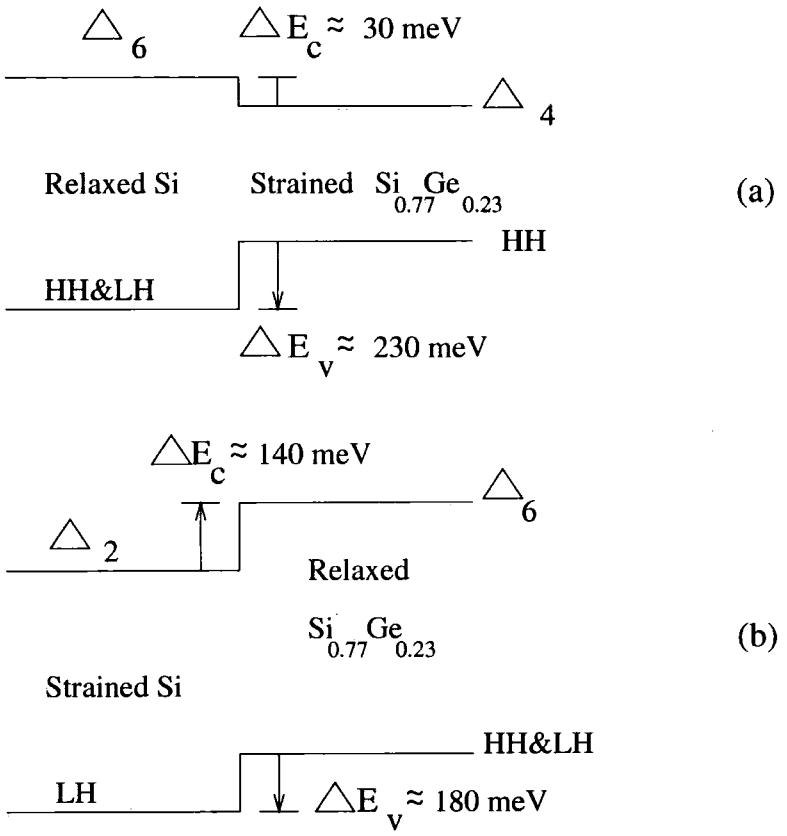


Figure 2.7: Band alignments between Si and $Si_{0.77}Ge_{0.23}$ on two substrates (a) Si and (b) $Si_{0.77}Ge_{0.23}$

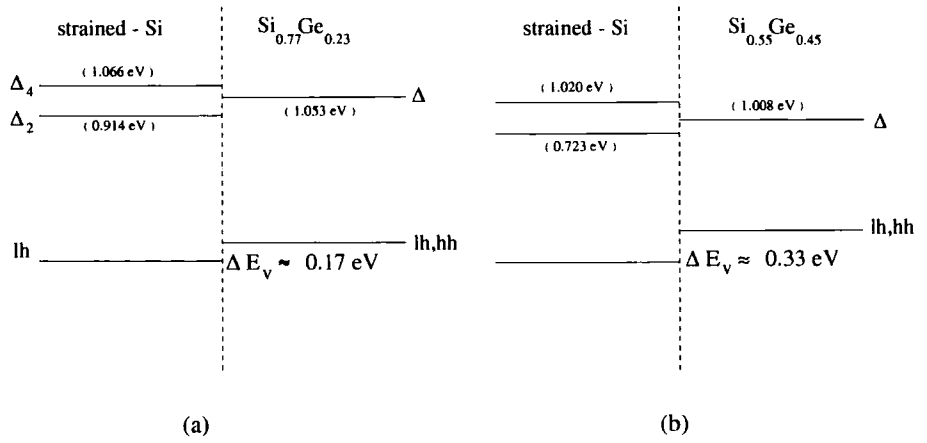
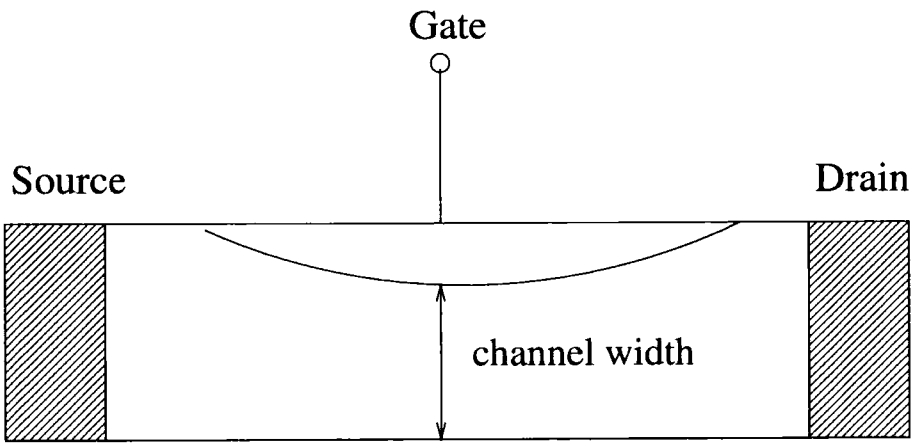


Figure 2.8: Band alignments between Si and $Si_{0.77}Ge_{0.23}$ (a), Si and $Si_{0.55}Ge_{0.45}$ (b) as grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates, respectively.

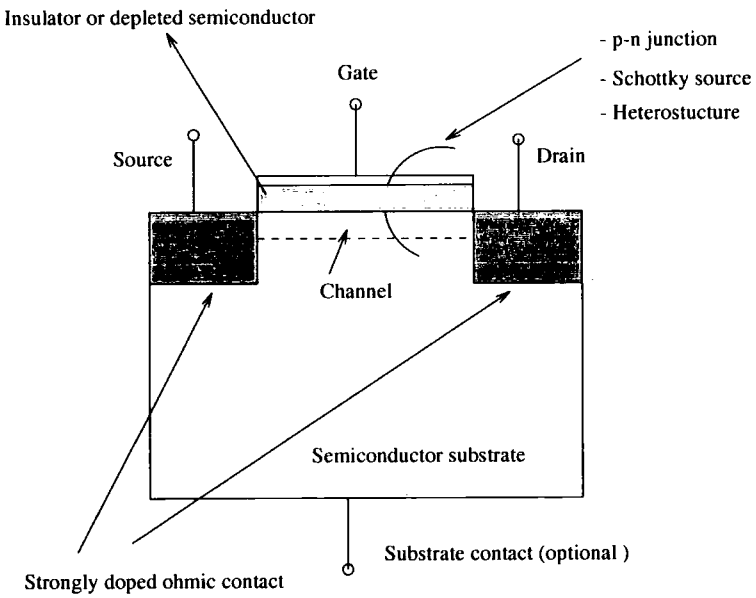
The basic principle has been implemented in a variety of different devices, such as the Metal Oxide Semiconductor FET (MOSFET), or Metal Insulator Semiconductor FET (MISFET), Metal Semiconductor FET (MESFET) and Heterostructure FET (HFET) as illustrated in Figure 2.9(b).

2.4 Concept of Modulation doping

The purpose of semiconductor(s) doping is generally to change the free carrier density in a controllable fashion. This requires that the dopant centres be ionised, leaving positively charged donor centres in the case of n-doping. These fixed charged centres have a screened Coulomb interaction with the free electrons and the resultant ionised impurity scattering is usually an important scattering mechanism of low energy carriers in semiconductors. In a conventional transistor, the impurities and carriers exist in the same regions of the device and the detrimental impurity scattering cannot be avoided. However, the quality of carrier transport in the device can be enhanced by



(a) Basic principle of FET



(b) Conventional FET structure

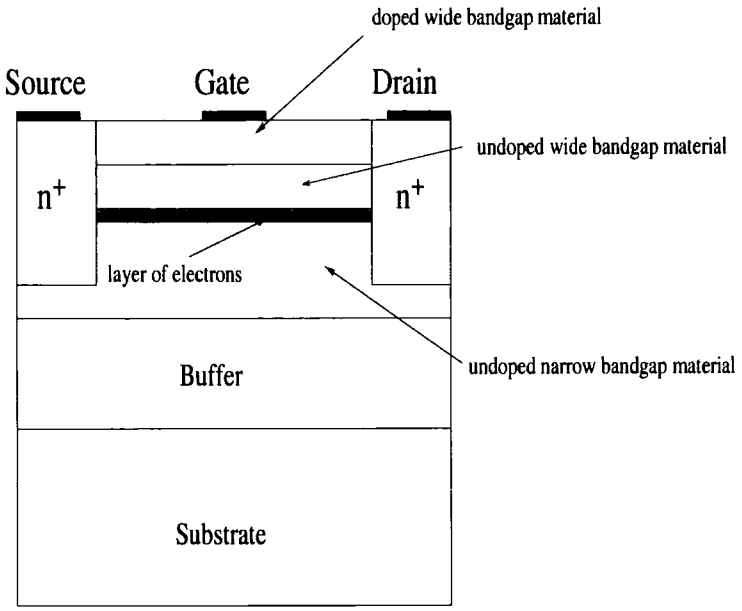
Figure 2.9: **(a)** Schematic FET structure. Device operation involves the use of a gate to modulate the charge in the channel and hence in the current flow between source and drain. **(b)** Conventional FET structure. The gate isolation is achieved in a variety of ways (p-n junction, Schottky gate, heterostructure) resulting in a number of different versions of the device.

removing the dopants from the regions where the critical carrier transport processes occur. This is the modulation doping concept, which can be realised by the use of spatially-dependent doping and the use of heterojunctions to control the location of the carriers.

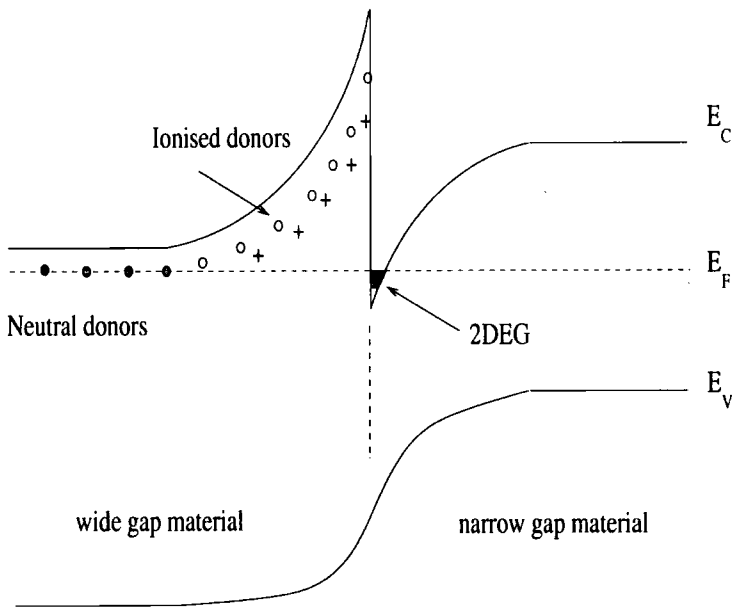
Stern and Howard first formulated a theory of two-dimensional electron transport in Si inversion layers in 1967 [27]. In 1969 Esaki and Tsu [28] proposed that the high electron concentration and the spatial separation of donors diminishes the effects of impurity scattering in the two-dimensional electron gas, thereby enhancing the electron mobility. The effect of mobility enhancement was first experimentally observed in an AlGaAs/GaAs heterostructure by Dingle *et al.* [29].

The basic idea behind modulation doping is illustrated in Figure 2.10 for a heterojunction between a wide bandgap material (e.g AlGaAs) and narrow bandgap material (e.g GaAs) in a FET-type structure. Note that a layer of the wide gap material immediately adjacent to the interface is left undoped, and is referred to as the spacer layer. As a result of the conduction band offset there is a discontinuity in the band profile through the device and a layer of electrons is formed at the heterointerface. Note that the spatial separation of free carriers from their parent-dopant centres could not be realised in bulk material since there is a large electrostatic restoring force on the electrons after the separation. In the heterostructure the band discontinuity prevents the electric field from returning the electrons to the vicinity of the donors. In fact the field can only squeeze the electrons against the interface, where they are trapped in a roughly triangular potential well. The net result of the modulation doping is that carriers in the undoped heterointerface region are spatially separated from the doping centres and can have a substantially enhanced mobility due to the reduced impurity scattering. This is despite the fact that interface roughness scattering at the heterointerface(s) appears as an additional scattering mechanism.

The first transistor based on the modulation doping concept was a GaAs/AlGaAs



(a) General structure of a MODFET



(b) Band diagram of n-MODFET

Figure 2.10: (a) General structure of a MODFET. The wide-bandgap material is doped and the carriers diffuse to the undoped narrow-bandgap material. **(b)** The band profile of an n-MODFET showing band bending leading to a roughly triangular quantum well at the heterostructure interface. Mobile electrons generated by the donors diffuse into the small band gap material, and are prevented from returning by the potential barrier at the interface.

device demonstrated by Mimura *et al.* [30] in 1980, which showed a dramatic enhancement in the drain current and transconductance. Since then attention has focused on the use of modulation-doped heterostructures for the realisation of high speed field-effect transistors. A field-effect transistor that takes advantage of the superior carrier transport properties of modulation-doped heterostructures is called a modulation-doped field-effect transistor (MODFET). MODFETs based on the GaAlAs/GaAs heterostructures have very successfully driven microelectronics to higher speeds for specialist applications. However, the drawback of III-V compound devices is their incompatibility with the well established Si-based integrated circuit technology. As a result there is interest in devices which exploit the properties of the strained SiGe/Si system, and can be thought of as a natural extension of the Si homosystem.

2.5 n-Si/$Si_{1-x}Ge_x$ MODFET

Figure 2.11 illustrates an n-channel Si/$Si_{1-x}Ge_x$ MODFET, in which a channel for the electrons is formed by the tensile strained silicon layer with a $Si_{1-x}Ge_x$ alloy barrier. A $Si_{1-x}Ge_x$ alloy spacer layer separates the doped region of the layer from the silicon channel to form a modulated doped structure.

The device is based on the ability to grow a thick relaxed SiGe layer on a silicon substrate. The upper part of the SiGe layer has the lattice constant of bulk material and can be used as a substrate to grow a strained pseudomorphic Si layer and produce a heterostructure with the band alignments shown in Figure 2.7(b). The most widely used technique is the initial growth of a SiGe buffer layer on the Si substrate, with a Ge content which is graded from zero up to the final concentration of typically 20-30 %, followed by the growth of a constant composition buffer and then the active layers of the device [31]. This graded SiGe buffer concept provides what are called virtual substrates with low defect densities at arbritrary Ge content [31,32].

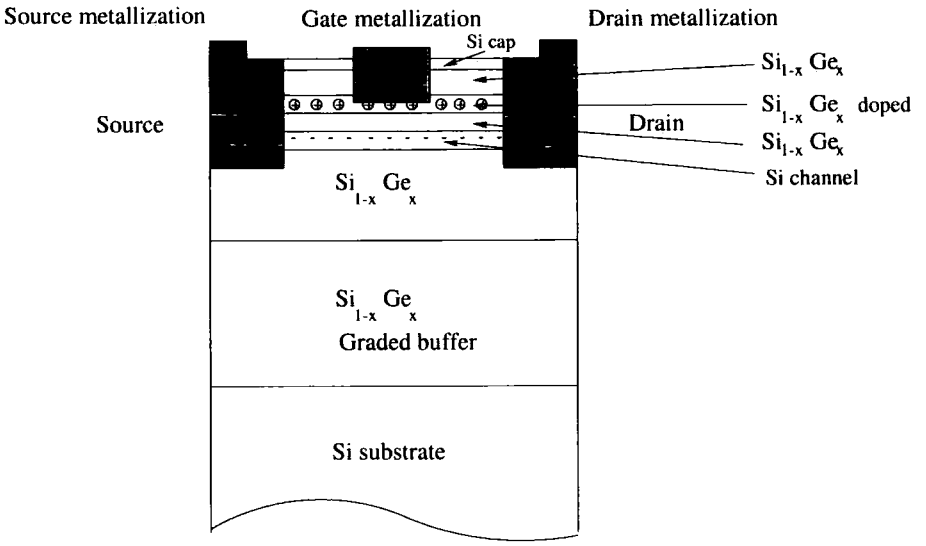


Figure 2.11: n-channel $Si/Si_{1-x}Ge_x$ MODFET.

In recent years the electron mobility in the strained Si channel of n-type modulation doped quantum well structures (MODQW) has reached a high level. For example, a room temperature mobility in excess of 2500 $cm^2V^{-1}s^{-1}$ was reported by Nelson *et al.* [3], which is almost twice that of bulk Si and three times higher than in conventional MOSFETs. (Note that the peak electron mobility in intrinsic bulk Si is about 1400 $cm^2V^{-1}s^{-1}$ [33]). The first n-channel modulation-doped SiGe/Si heterostructure field-effect transistors based on this type of structure were then fabricated by Daenbkes *et al.* [34] by the use of molecular-beam epitaxial (MBE) growth. The first transistor they reported exhibited an extrinsic transconductance of 40 mS mm^{-1} for a gate length of 1.6 μm, which is higher than that of comparable conventional Si MOSFETs.

Ismail *et al.* [35] have grown devices in two successive UHV-CVD processes and reported a carrier density of 2.5×10^{12} cm^{-2} as well as a high mobility of 1500 $cm^2V^{-1}s^{-1}$ at room temperature. Also, MBE-grown n-MODFETs, with Hall mobilities of 2200 $cm^2V^{-1}s^{-1}$ at room temperature, have been reported by König *et al.* [36]. A current gain cut off frequency f_T as high as 62 GHz for a n-$Si/Si_{0.75}Ge_{0.25}$ MODFET has been

recently reported by the IBM group [37]. Similar high values of f_T have been achieved in high Ge content (45 %) n-MODFETs that have been investigated experimentally by the Daimler-Chrysler group [36].

P-channel MODFETs can be produced by growing a pseudomorphic $Si_{1-x}Ge_x$ layer on Si. In this case nearly all the band offset occurs in the valence band and holes are confined to the $Si/Si_{1-x}Ge_x$ interface. In addition, the strain lifts the degeneracy of the light- and heavy-hole valence band; a $Si_{1-x}Ge_x$ layer on Si is in biaxial compression and the heavy-hole band is raised relative to the light hole band as shown in Figure 2.5(c). Since the heavy hole band has the smaller in-plane effective mass, the strained $Si_{1-x}Ge_x$ film can be expected to exhibit a higher hole mobility than bulk Si. The reduced interband scattering that results from the lifting of the heavy hole-light hole band degeneracy also acts to increase hole mobility. A novel method for the fabrication of high hole mobility structures is the growth of a Ge channel on a relaxed graded $Si_{1-x}Ge_x$ buffer. The Ge channel, being under an in-plane biaxial compressive strain, provides a hole effective mass lower than $0.1m_0$ in strained Ge. (Note that the hole density-of-states effective mass in bulk Ge is much higher with a value of $0.29m_0$). As a consequence high hole mobilities at 4.2 K of 55000 $cm^2V^{-1}s^{-1}$ have been observed [38].

2.6 npn-Si SOI LBJT

Devices that combine the field effect and bipolar effect are very attractive for some applications since they can offer the advantages of high current driving capability and high current gain together with low power consumption. By combining both technologies on the same chip, one can optimise the tradeoff between the speed and power dissipation, and attain digital/analog systems with a performance which may exceed that of circuits based on either technology alone.

Bipolar transistors are normally fabricated in a vertical geometry; that is with

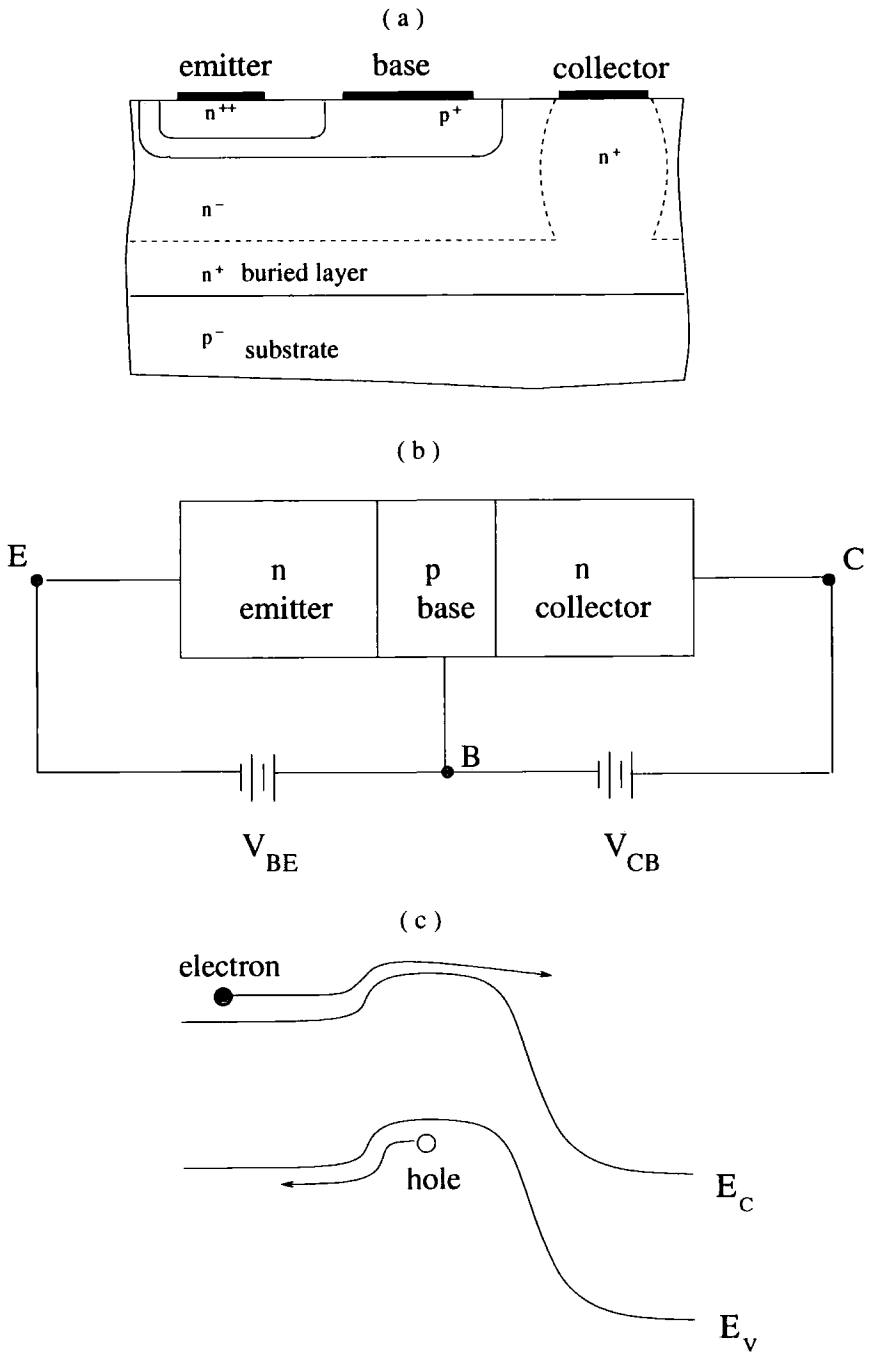


Figure 2.12: **(a)** Two-dimensional representation of the structure of a vertical npn BJT. **(b)** Schematic diagram illustrating the applied voltages in normal operation, **(c)** Corresponding band-diagram with carrier flows indicated.

emitter, base and collector formed by adjacent layers of the structure as illustrated in Figure 2.12(a). However, it is possible to produce lateral transistors where different parts of the device are fabricated using lithographic methods. Since the vertical dimensions can be made smaller than the lateral dimensions, vertical transistors are preferable for high-speed applications. However, lateral bipolar junction transistors (LBJTs) can be incorporated into a MOSFET-based integrated circuit with minor modifications to processing methods. A Si lateral bipolar junction transistor can be derived from a standard Si n-channel MOSFET by using an ohmic gate electrode to make contact to the p-Si channel which is used to form a base. In addition the use of thin-film silicon-on-insulator (SOI) technology for LBJTs permits simple circuit integration. An LBJT using SOI technology is an attractive device because of its low parasitic capacitances and simple fabrication process. In principle, it provides a good alternative to the MOSFET since, for extremely thin gate oxides, the gate leakage current of a MOSFET become comparable to the base current of an LBJT. At the same time, the LBJT could offer a highly scalable approach for CMOS circuits.

There are two types of bipolar devices; the n-p-n type, which has a p-type base and n-type emitter and collector, and the p-n-p type, which has an n-type base and p-type emitter and collector. We will concentrate on the n-p-n transistor operating in the active mode as illustrated in Figure 2.12(b). The emitter-base junction is forward biased, resulting in holes being injected from the base into the emitter, and electrons being injected from the emitter into the base. However, since the emitter is more heavily doped than the base, the electron current dominates and depends exponentially on the forward bias of the junction. The base-collector is reverse biased and so in the absence of any other effects there would be only a small leakage current flow at the base-collector junction. However, essentially all the electrons injected from the emitter into the base reach the collector, and give rise to a collector current, since the base width (typically 0.1 μm or less) is much less than the electron diffusion length. The

holes injected from the base into the emitter give rise to a base current. The current gain of a bipolar transistor is defined as the ratio of its collector current to its base current. Thus one basic objective in bipolar transistor design is to achieve a collector current significantly larger than the base current and hence obtain current amplification.

Figure 2.13(a) shows a schematic cross-section of a lateral npn transistor, available in a typical CMOS process. Figure 2.13(b) illustrates the modelled lateral SOI bipolar transistor whose simulation results are reported and analysed in Chapter 5. The emitter, intrinsic base, and collector region are laterally formed in thin-film SOI. Note from Figure 2.13(b) that the heavy p^+ doping at the top of the base pedestal provides a low-resistance ohmic contact and a built-in electric field which repels electrons in the base away from the metal base contact [39].

2.7 Summary

In this chapter, we have described the relevant properties of Si and SiGe alloys. The effects of strain on the band structure and the electron transport properties have been explained. In particular, we have pointed out that the strain present in a $Si/Si_{1-x}Ge_x$ heterostructure breaks the six-fold degeneracy of the Si Δ valleys, resulting in an improved conduction band offset, a low in-plane effective mass at the conduction band edge, and reduced intervalley scattering. Thus, the strain provides an additional variable parameter, which can be used to tailor the device parameters discussed in Section 2.4. Specifically, a combination of bandgap engineering and strain can be used to produce an n-channel MODFET structure, in which a thin silicon channel is placed under tensile strain as discussed in Section 2.5. In Chapter 4 we shall present simulation results of n-strained Si channel MODFETs. In addition, the npn-SOI LBJT, which could be particularly suitable in dense high performance CMOS technology, has been introduced and the basic operation described. Chapter 5 of this thesis is devoted

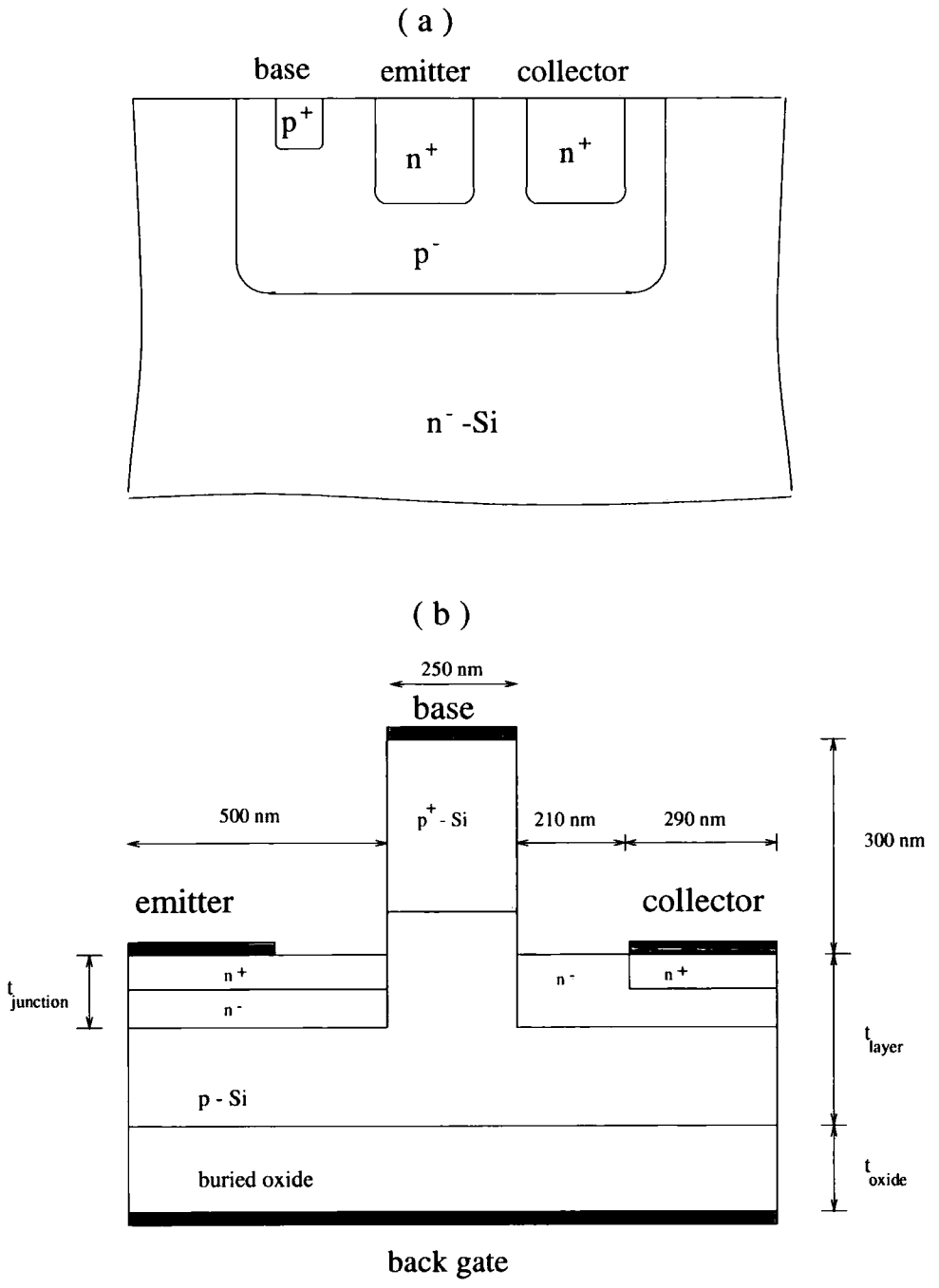


Figure 2.13: **(a)** Lateral bipolar junction transistor (LBJT) [40]. **(b)** Two-dimensional representation of the structure of a npn SOI-LBJT whose simulation results are reported in Chapter 5.

to simulations of npn-SOI LBJTs.

Chapter 3

Monte Carlo Simulation

3.1 Introduction

In order to describe the behaviour of carriers in a semiconductor device at a microscopic level, we have to solve the Boltzmann transport equation (BTE) for the structure. The BTE describes the time evolution of the carrier distribution in real and reciprocal space through the motion of particles, the forces acting on them and their scatterings. Since the Boltzmann equation is a nonlinear integro-differential equation, it can only be solved analytically in certain simple cases, and practical problems require the use of approximations or numerical methods. Monte Carlo (MC) simulation is one such numerical method and is the approach used in the research reported in this thesis. In Monte Carlo simulation [22,41–44], a computer is used to predict the trajectories in real and reciprocal space of typically tens of thousands of particles which are representative of electrons and holes in the real device. As a result, microscopic and macroscopic quantities (such as the carrier distribution functions, carrier mean velocity, device current, cut-off frequency, etc.) can be predicted. Monte Carlo simulation is extremely powerful because it is possible to incorporate detailed band structure information, band and wavevector-dependent carrier scattering events and realistic device geometries in a device model. The method is well suited to the analysis of non-stationary carrier

transport in submicron semiconductor devices, in which the electric field varies appreciably over distances comparable with the electron mean free path or on time scales which are comparable to mean free flight times. Hence it is possible to include many new and interesting non-equilibrium carrier dynamics such as ballistic transport and velocity overshoot which become important as device dimensions shrink to the deep sub-micrometer scale. In many cases, Monte Carlo simulation is in fact the most accurate technique available for analysing transport in devices [45].

This chapter is devoted to the description of the ensemble Monte Carlo method used in the device simulations described in Chapters 4 and 5. First, the Boltzmann transport equation is introduced and some common methods of solution are mentioned. The general methodology of the Monte Carlo simulation of carrier transport in bulk material is described in Section 3.3. Section 3.4 presents the band structure model used in the simulations. The carrier scattering mechanisms relevant to Si and SiGe alloys are described in Section 3.5. The application of the Monte Carlo method to the analysis of semiconductor devices is discussed in Section 3.6.

3.2 The Boltzmann transport equation

The quantum mechanical state of an electron in a crystal is specified by a band index and a wavevector $\mathbf{k}$ in the first Brillouin zone. For each band there is a definite relationship between the electron energy E and the wavevector. In the semiclassical model the electron can also be ascribed a position $\mathbf{r}$ in real space. (Note this is not possible in a fully quantum mechanical theory since the specification of both $\mathbf{r}$ and $\mathbf{k}$ is a contradiction of the Heisenberg Uncertainty Principle). Hence the dynamics of each electron can be represented by a moving point in a six-dimensional phase space ($\mathbf{r}$, $\mathbf{k}$).

The equation of motion for the particle is given by

$$\hbar\dot{\mathbf{k}} = -e\mathbf{F} \tag{3.1}$$

$$\dot{\mathbf{r}} = \mathbf{v} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}) \tag{3.2}$$

where e is the magnitude of electronic charge, and $\mathbf{F}$ is the electric field. Equations 3.1 and 3.2 are the equation of motion of free electrons in a perfect crystal. In reality, this so called ballistic motion is interrupted by collisions with impurities, phonons, defects, etc. If a collision is instantaneous, it may be visualised as the electron disappearing and instantaneously reappearing at a different value of $\mathbf{k}$ of the phase space.

We can define a distribution $f(\mathbf{r}, \mathbf{k}, t)$ that represents the probability of finding an electron at position $\mathbf{r}$, with vector $\mathbf{k}$ and at time t. $f(\mathbf{r}, \mathbf{k}, t)$ describes the distribution of carriers in both the position and the momentum and can be used to obtain various quantities of interest such as the mean carrier density, kinetic energy, and so forth. The equation for the distribution function f is called the Boltzmann transport equation:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t}\right)_{coll} \tag{3.3}$$

where $\partial f/\partial t_{coll}$ represents the change in f due to electron scattering, and can be expressed as:

$$\left(\frac{\partial f}{\partial t}\right)_{coll} = \int [S(\mathbf{k}', \mathbf{k}) f_{\mathbf{k}'}(1 - f_{\mathbf{k}}) - S(\mathbf{k}, \mathbf{k}') f_{\mathbf{k}}(1 - f_{\mathbf{k}'})] dV_{\mathbf{k}'} \tag{3.4}$$

where the integration is over the first Brillouin zone. $S(\mathbf{k}, \mathbf{k}')dV_{\mathbf{k}'}$ is the conditional rate of transitions from the state $\mathbf{k}$ to the states $\mathbf{k}'$ in $dV_{\mathbf{k}'}$ given that there is initially an electron in state $\mathbf{k}$ and state $\mathbf{k}'$ is empty.

The Boltzmann equation is generally difficult to solve. However, some analytical approaches to achieve approximate closed-form solutions have been developed for the case of carrier transport in bulk material. For example, in low electric fields a solution may be based on the fact that the distribution function will not be very different

from the equilibrium Fermi-Dirac distribution function. For elastic scattering (such as impurity, acoustic phonon, or piezoelectric scattering) one may use a relaxation time approximation (RTA) where $\partial f/\partial t_{coll}$ is taken equal to $-(f - f_0)/\tau$, where f_0 is the equilibrium distribution function, and τ is the relaxation time constant. However, for most semiconductors, all the scattering processes are not elastic and the relaxation time approximation usually fails to provide adequate accuracy for calculating the transport properties. Rode's iterative procedure [46] is a numerical method to calculate the distribution function, but is straightforward to implement and is not demanding on computer resources. However, this method is limited to cases where the carriers remain in one band and to low electric fields where linear response is valid.

The Monte Carlo method is a more general numerical but computationally intensive approach which is based on the simulation of the carriers dynamics in a semiconductor. It provides numerical solutions to the Boltzmann transport equation by stochastically tracking the movements of a set of representative particles in phase space under the action of the electric field and the various carrier scattering processes. When coupled with a self-consistent numerical solver for Poisson's equation, the Monte Carlo method permits the exploration of the properties of realistic two- and three-dimensional models of semiconductor devices and the investigation of the microscopic detail underlying the device behaviour. It is Monte Carlo simulation which is used for device modelling in the research reported here, and the rest of this chapter is devoted to an outline of the simulation method.

3.3 Monte Carlo simulation

The Monte Carlo (MC) method solves the Boltzmann transport equation by simulating the motion of electrons and holes in the bulk semiconductor medium or device structure. The main feature of the MC method is that a carrier is treated as a point-like

semiclassical particle moving under the action of the local electric field and the scattering processes. The individual trajectories of carriers are computed by the following steps:

1. **Each carrier is moved in real and reciprocal space in the "free flight"** between collisions under the action of the electric field. This is done by solving Equations 3.1 and 3.2 in the crystal.

2. **The carriers scatter** conditionally at the end of their free flights. The duration of a free flight and the new state after scattering are chosen stochastically in accordance with quantum mechanical theory.

Such an approach requires a knowledge and appropriate description of the band structure of the semiconductor.

3.4 Band Structure

In the absence of applied electric field, carriers reside close to the edges of the bands on either side of the fundamental gap. The relationship between the momentum $\mathbf{k}$ and the energy $E(\mathbf{k})$ of an electron close to the minimum of the conduction band can be described by various models at different levels of complexity.

(a) Parabolic Bands

The simplest model of a valley in the conduction band is to assume a parabolic and isotropic relation between energy and wavevector. Specifically

$$E(\mathbf{k}) = \frac{\hbar^2|\mathbf{k}|^2}{2m_o^*} \tag{3.5}$$

where m_o^* is the effective mass at the conduction band minimum. The particle velocity is evaluated as

$$\mathbf{v} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}) = \frac{\hbar\mathbf{k}}{m_o^*} \tag{3.6}$$

and gives the simple relation for the electron momentum

$$\mathbf{p} = \hbar\mathbf{k} = m_o^*\mathbf{v} \tag{3.7}$$

(b) **Non-parabolic Bands**

In general conduction band valleys are not parabolic but for sufficiently small $\mathbf{k}$ may often be described by the approximate formula

$$E(\mathbf{k})(1 + \alpha E(\mathbf{k})) = \frac{\hbar^2|\mathbf{k}|^2}{2m_o^*} \tag{3.8}$$

where α is the coefficient of nonparabolicity and has the dimensions of an inverse energy [47]. The solution of equation (3.8) for the energy in terms of wavevector is

$$E(\mathbf{k}) = \frac{1}{2\alpha}\left(\sqrt{1 + \frac{4\alpha\hbar^2|\mathbf{k}|^2}{2m_o^*}} - 1\right) \tag{3.9}$$

The particle velocity is

$$\mathbf{v} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}) = \frac{\hbar\mathbf{k}}{m_o^*}\left(1 + 4\alpha\frac{\hbar^2|\mathbf{k}|^2}{2m_o^*}\right)^{-1/2} \tag{3.10}$$

which, using (3.9), can be written as

$$\mathbf{v} = \frac{\hbar\mathbf{k}}{m_o^*[1 + 2\alpha E(\mathbf{k})]} \tag{3.11}$$

So far, a spherically symmetric band (or isotropic effective mass) has been assumed. This is normally a good approximation for the Γ valley of GaAs, which is at the centre of the Brillouin zone. In the cases of Si and Ge, a number of degenerate valleys

away from the zone centre form the conduction band edge and the constant energy surfaces for these valleys are ellipsoids rather than a sphere. Ge has eight valleys with their minima at the edge of the Brillouin zone in the [111] directions, called the L-points. Si has six valleys with minima at approximately $\mathbf{k} = (2\pi/a)(0.85, 0, 0)$, and the equivalent points, which are close to the X-points at the edge of the Brillouin zone. In an ellipsoidal valley the mass is not isotropic, and is instead described by a tensor with components

$$m^*_{ij} = \hbar^2 \left(\frac{\partial^2 E}{\partial k_i \partial k_j} \right)^{-1} \tag{3.12}$$

where i, j denote the Cartesian axes. The $E - \mathbf{k}$ relation for an ellipsoidal constant energy surface is

$$E = \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right) \tag{3.13}$$

where k_x, k_y, k_z are the components of $\mathbf{k}$ along the principal axes of the ellipsoid measured from the position of the minimum, and m_x, m_y, m_z are the corresponding effective masses. For Si, the surfaces of constant energy are ellipsoids of revolution and the $E-\mathbf{k}$ relation in the conduction band can be expressed as

$$E(\mathbf{k}) = \frac{\hbar^2 k_z^2}{2m_l} + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_t} \tag{3.14}$$

where l and t refer to the so called longitudinal and transverse directions. For Si, the value of m_l is $0.91m_o$ and m_t is $0.19m_o$. By using the Herring-Vogt transformation of $\mathbf{k}$-space [48], Equation 3.13 can be written as

$$\begin{aligned} E &= \frac{\hbar^2}{2m_o}(k_x^{*2} + k_y^{*2} + k_z^{*2}) \\ &= \frac{\hbar^2 k^{*2}}{2m_o} \end{aligned} \tag{3.15}$$

where m_o is the free electron mass. To obtain the above result, the transformation

$\mathbf{k}^* = T_m^{1/2} \cdot k$ has been used, where $T_m^{1/2}$ is defined as

$$T_m^{1/2} = \begin{pmatrix} (m_o/m_l)^{1/2} & 0 & 0 \\ 0 & (m_o/m_t)^{1/2} & 0 \\ 0 & 0 & (m_o/m_t)^{1/2} \end{pmatrix} \tag{3.16}$$

With this transformation, Equation 3.8 becomes

$$E(1+\alpha E) = \frac{\hbar^2}{2m_D^*}\mathbf{k}^{*2} \tag{3.17}$$

where $m_D^* = (m_t^{*2} m_l^*)^{1/3}$ is the density of states effective mass. The electron velocity as a function of $\mathbf{k}^*$ is then given by

$$v_i = \frac{\hbar}{m_o(1+2\alpha E)} T_m^{1/2} k_j^* \tag{3.18}$$

Equations 3.13-3.14 describe a band with ellipsoidal constant energy surfaces whose principal directions are along the crystallographic axes. After the Herring-Vogt transformation the constant-energy surfaces are spherical in k^* space, which has the advantage that the integrals involved in the various scattering probabilities may be simplified. However, the equation of motions are vector equations, and the transformation in Equation 3.16 has to be applied to other vector quantities such as the electric field and the phonon wavevectors. For example, the equation of motion for an electron under the influence of a field $\mathbf{F}$ becomes

$$\frac{d(\hbar\mathbf{k}^*)}{dt} = -e\mathbf{F}^* \tag{3.19}$$

where F^* is an electric field in the starred coordinate system. The valence bands of a semiconductor are made up of the light and heavy hole bands and the spin split-off band. At the top of the valence band, the light and heavy hole states are degenerate and the constant energy surfaces are warped. In Si, the spin split-off band is only 44 meV below the degenerate bands at $\mathbf{k} = 0$. The spin splitting is larger in Ge with a value of 0.29 eV [42]. $\mathbf{k} \cdot \mathbf{p}$ theory gives, for the light and heavy hole valence bands, an

$E - \mathbf{k}$ relation of the following form [45,49,50]:

$$\begin{aligned} E_v &= -\frac{\hbar^2}{2m_o}Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_x^2k_z^2)]^{1/2} \\ &= -\frac{\hbar^2}{2m_o}\gamma_1 k^2 \pm [4\gamma_2^2k^4 + 12(\gamma_3^2 - \gamma_2^2)(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2)]^{1/2} \end{aligned} \tag{3.20}$$

where k_x, k_y, k_z are the components of $\mathbf{k}$ along the crystal axes, and $\gamma_1, \gamma_2, \gamma_3$ are Luttinger parameters. The $\pm$ symbols refer to the light and heavy hole bands, respectively. The values of γ_1, γ_2, γ_3 for Ge and Si are given in Table 3.1. The spin split-off valence band has a simple spherical form given by [51]

$$E_v = -\Delta + Ak^2 \tag{3.21}$$

where A is -4.22 for Si and -13.35 for Ge.

Material	m_e/m_o	γ_1	γ_2	γ_3	Δ (eV)	A
Ge	1.58/0.082	13.25	4.20	5.56	0.044	-4.22
Si	0.916/0.191	4.26	0.34	1.45	0.29	-13.35

Table 3.1: Band structure parameters for Ge and Si [42]

(c) **Full band Structure**

If we need to specify $E(\mathbf{k})$ for various bands throughout the Brillouin zone, algebraic approximations are no longer possible. Then the function $E(\mathbf{k})$ must be described numerically using the results of detailed band structure calculations, such as those based on the pseudopotential method, or using data from experiments.

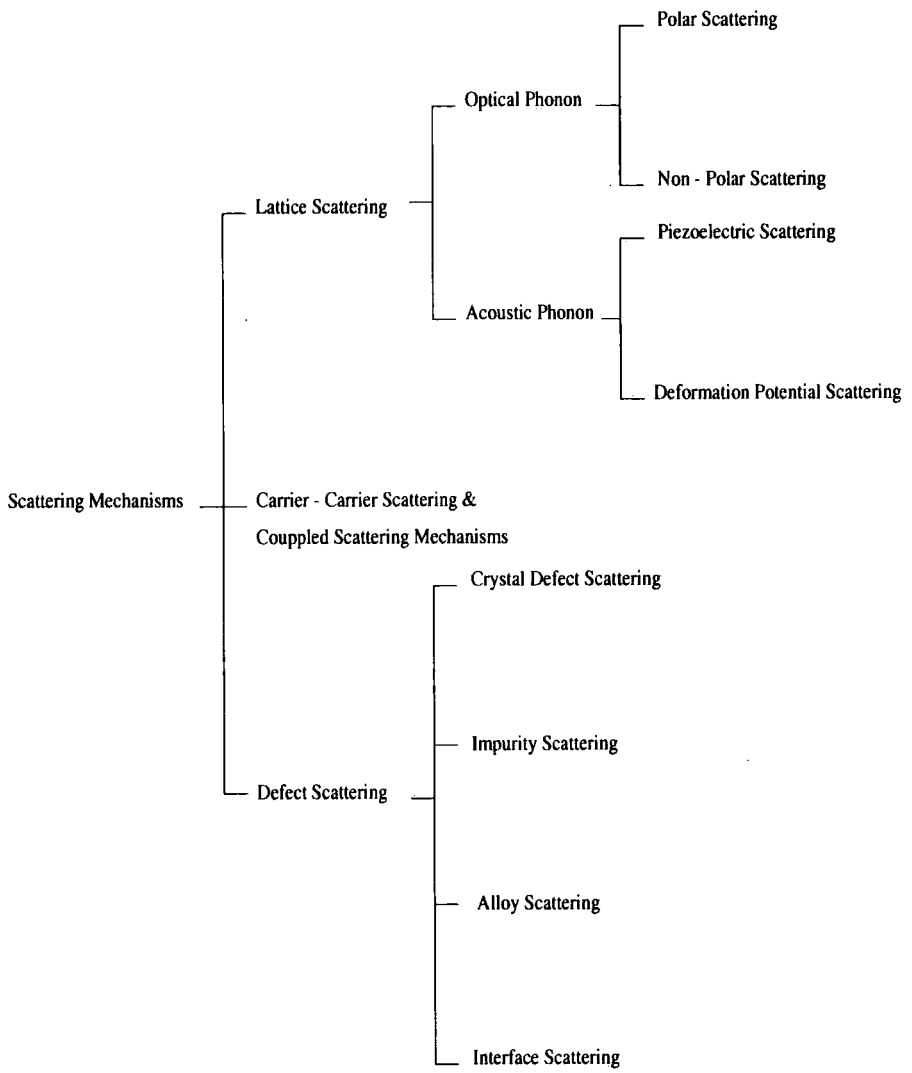


Figure 3.1: A schematic classification of the various scattering mechanisms in semiconductors based on their origin.

3.5 Scattering Mechanisms

Scattering mechanisms can be classified into three main categories : lattice or phonon scattering, carrier-carrier scattering, and defect scattering as shown in Figure 3.1. Carrier-carrier scattering is only normally important in very heavily doped or highly excited semiconductors and is ignored in our simulations. The scattering of electrons and holes in bulk silicon has been treated extensively in the past (see, e.g. Jacoboni *et al.* [42,43], Fischetti *et al.* [22], Moglestue [44]). The main scattering mechanisms to be considered in the Si-Ge material system are lattice scattering (acoustic and nonpolar optical), ionised impurity scattering, and alloy scattering (for $Si_{1-x}Ge_x$). There is also scattering due to the roughness of the heterointerface in a Si/SiGe structure.

Normally carrier scattering is described quantum mechanically by time-dependent perturbation theory. For example in the case of lattice scattering, it is necessary to find the rate at which an electron is scattered out of one state $\mathbf{k}$ into another state $\mathbf{k}'$ while either absorbing or emitting a phonon of wavevector $\mathbf{q}$. This is given by Fermi's golden rule as

$$S(\mathbf{k},\mathbf{k}') = \frac{2\pi}{\hbar}|M_{\mathbf{k}',\mathbf{k}}|^2 G(\mathbf{k},\mathbf{k}')\delta(\mathbf{k}'-\mathbf{k}\mp\mathbf{q}-\mathbf{G})\delta(\mathbf{E}(\mathbf{k}')-\mathbf{E}(\mathbf{k})\pm\hbar\omega_{\mathbf{q}}) \qquad (3.22)$$

where $\hbar\omega_{\mathbf{q}}$ is the phonon energy and $\mathbf{G}$ is any reciprocal lattice vector. The upper sign corresponds to the absorption of a phonon and the lower sign to emission. $G(\mathbf{k},\mathbf{k}')$ is the overlap factor defined as

$$G(\mathbf{k}',\mathbf{k}) = |\int_{cell} u^*_{k'}(r)\cdot u_k(\mathbf{r})d^3\mathbf{r}|^2 \qquad (3.23)$$

between the periodic parts $u_{\mathbf{k}}(\mathbf{r})$ and $u_{\mathbf{k}'}(\mathbf{r})$ of the Bloch functions of the initial and final states. The overlap integral is equal to unity for parabolic conduction bands (pure s-state wave functions). More generally for a scattering between an initial state $\mathbf{k}$ and

a final state $\mathbf{k}'$ in the same valley $G(\mathbf{k}', \mathbf{k})$ is often taken to be as [41]:

$$G(\mathbf{k}', \mathbf{k}) = \frac{[(1+\alpha E)^{1/2}(1+\alpha E')^{1/2} + (\alpha^2 E E')^{1/2}\cos\beta]^2}{(1+2\alpha E)(1+2\alpha E')} \tag{3.24}$$

where β is the angle between the state $\mathbf{k}$ and $\mathbf{k}'$. For the hole transitions within the heavy and light hole bands $G(\mathbf{k}', \mathbf{k})$ can be taken as [52]

$$G(\mathbf{k}', \mathbf{k}) = \frac{1}{4}(1 + 3\cos^2\theta) \tag{3.25}$$

For transitions between the light and the heavy hole bands,

$$G(\mathbf{k}', \mathbf{k}) = \frac{3}{4}\sin^2\theta = \frac{3}{4}(1 - \cos^2\theta) \tag{3.26}$$

It turns out that the overlap integral is always less than one if the nonparabolicity of the bands is taken into account. The total scattering rate for any process is obtained by integrating Equation 3.22 over all final states. Next, we will examine carrier scattering mechanisms and evaluate the transition rate, $S(\mathbf{k}, \mathbf{k}')$ and total scattering rate $S(\mathbf{k})$ for the various scattering processes.

3.5.1 Lattice Scattering

Carriers can interact with various types of lattice vibrations, by absorbing or emitting a phonon in each process. There are two main types of phonon scattering in group IV semiconductors as described below.

3.5.1.1 Acoustic Phonon Scattering

Generally acoustic phonons may cause scattering in two different and independent ways, through deformation potential scattering and piezoelectric scattering. Piezoelectric scattering occurs in semiconductors with two or more atoms per unit cell, where there is no crystal inversion symmetry. This scattering results from the interaction of

the electron with the induced polarisation of the lattice. Thus this type of scattering can be important in semiconductors like GaAs, GaN, etc. However, crystals of silicon and germanium are formed of identical atoms, and do not exhibit piezoelectric scattering. Hence, only acoustic deformation scattering need be considered here.

An acoustic wave induces a change in the spacing of neighbouring atoms in a semiconductor which causes a variation of the local energy of the bands and hence a source of electron scattering. The deformation potential is defined as the change of energy of the band edge per unit strain due to acoustic phonons. This type of scattering is usually the most important cause of scattering of electrons in undoped silicon and germanium crystals at room temperatures. Energy and momentum conservation restricts intravalley scattering by acoustic phonons to that associated with long-wavelength lattice modes. For longitudinal acoustic mode scattering the emission of a phonon, the matrix element is

$$|M_{\mathbf{k},\mathbf{k}'}|^2 = \frac{\hbar \Xi_1^2 q^2}{2MN\omega_q}(N_q + 1) \tag{3.27}$$

where Ξ_1 is the deformation potential, $\mathbf{q}$ is the phonon wavevector, $M = \rho V/N$, ρ is the mass density, V is the volume and N_q is the phonon distribution function. The quantity $(N_q + 1)$ is replaced by N_q for phonon absorption. Normally the phonon energy will be of order of a few millivolts or less. Therefore, when considering phonon scattering at room temperature the phonon energy is much lower than the thermal energy and the so called equipartition approximation to the Bose-Einstein distribution can be used for N_q, which takes the form:

$$N_q = \frac{1}{\exp(\hbar\omega_q/k_BT) - 1} \approx \frac{k_BT}{\hbar\omega_q} >> 1 \tag{3.28}$$

Since the typical electron energy is much larger than the acoustic phonon energy, the scattering can be treated as elastic and the term $\hbar\omega_q$ ignored in the energy delta function of Equation 3.22. Now considering the scattering rate due to both phonon emission and absorption and making use of the equipartition approximation above (which is ac-

tually valid for temperatures above about 50 K in Si and Ge), the matrix element relating to the combined effect of emission and absorption is

$$|M_{\mathbf{k},\mathbf{k}'}|^2 = \frac{\Xi_1^2 q^2 k_B T}{MN\omega_q^2} = \frac{\Xi_1^2 k_B T}{\rho\, V v_s^2} \tag{3.29}$$

where the dispersion relation for the acoustic phonons is simply represented as $\omega_q = qv_s$. With these approximations, the acoustic phonon scattering rate is

$$S(\mathbf{k}) = \frac{2D_{ac}^2 k_B T}{8\pi^2 \hbar \rho v_s^2} \int d^3k' \delta(\mathbf{k} - \mathbf{k}')\delta(E_{\mathbf{k}'} - E_{\mathbf{k}'}) \tag{3.30}$$

Carrying out the integration over the final wavevector, the acoustic phonon scattering rate is [49]:

$$S(\mathbf{k}) = \frac{2\pi D_{ac}^2 k_B T N(E_{\mathbf{k}})}{\hbar \rho v_s^2} \tag{3.31}$$

$N(E_{\mathbf{k}})$ is the density of states at energy $E_{\mathbf{k}}$, and for a parabolic band

$$N(E_{\mathbf{k}}) = \frac{(2m^*)^{3/2} E_{\mathbf{k}}^{1/2}}{4\pi^2 \hbar^3} \tag{3.32}$$

where m^* is the electron effective mass. Inspection of Equation 3.31 shows that the acoustic phonon scattering rate is proportional to the temperature. For a non-parabolic band, the total scattering rate is given by [53]

$$S(\mathbf{k}) = \frac{(2m^*)^{3/2} k_B T D_{ac}^2}{2\pi \rho v_s^2 \hbar^4} \gamma^{1/2}(E)(1 + 2\alpha E) \times \frac{(1 + \alpha E)^2 + \frac{1}{3}(\alpha E)^2}{(1 + 2\alpha E)^2} \tag{3.33}$$

where α is the coefficient of nonparabolicity described in Section 3.4 and $\gamma(E) = E(1 + \alpha E)$.

3.5.1.2 Non-polar Optical Phonon Scattering

Optical phonon scattering generally becomes the predominant electron scattering mechanism in semiconductors at high temperatures or at high electric fields and is an important influence on electron transport in Si and Ge above room temperature. The scattering can originate from either polar or nonpolar effects but polar scattering does

not occur in Si and Ge for the same reasons as piezoelectric scattering is absent as described in Section 3.5.1.1. The scattering of electrons by nonpolar optical phonons may be treated as a type of deformation potential scattering process. The transition rate for optical phonon intravalley scattering is [54]

$$S(\mathbf{k},\mathbf{k}') = \frac{\pi D_o^2}{\rho\omega_o\Omega}(N_o + \frac{1}{2} \mp \frac{1}{2})\delta(\frac{\hbar^2 q^2}{2m^*} \pm \frac{\hbar^2 kq\cos\theta'}{m^*} \mp \hbar\omega_o) \tag{3.34}$$

where D_o is an optical deformation potential constant, and Ω is the cell volume. Because the optical phonon energy is nearly constant as a function of $\mathbf{q}$, we can approximate $\omega_\mathbf{q}$ and N_q by the constants ω_o and N_o.

For scattering in an isotropic parabolic band the minimum and maximum phonon wavevectors involved in the scattering are

$$\begin{aligned} q_{min} &= k|1 - (1 \pm \frac{\hbar\omega_o}{E_k})^{1/2}| \\ q_{max} &= k[1 + (1 \pm \frac{\hbar\omega_o}{E_k})^{1/2}] \end{aligned} \tag{3.35}$$

Integrating Equation 3.34 over $\mathbf{q}$ and using limits of phonon wavevectors indicated above one can obtain [54]:

$$S(\mathbf{k}) = \frac{\pi D_o^2}{\rho\omega_o}(N_o + \frac{1}{2} \mp \frac{1}{2})N(E_\mathbf{k} \pm \hbar\omega_o) \tag{3.36}$$

Note that the optical phonon energy $\hbar\omega_o$ is comparable to the average thermal energy of carriers at room temperature, and therefore this type of scattering must be treated as inelastic. For non-parabolic bands, the non-polar optical phonon scattering rate is [43]

$$S(\mathbf{k}) = \frac{D_o^2(m^*)^{3/2}}{\sqrt{2}\pi\hbar^3\omega_o}(1 + 2\alpha E')\gamma^{1/2}[N_o \text{ or } N_o + 1] \tag{3.37}$$

where E' is the electron energy at the state $\mathbf{k}'$.

3.5.1.3 Equivalent Intervalley Phonon Scattering

The discussion of phonon scattering so far has concentrated on intravalley scattering. However, acoustic and optical phonons of large wavevector can cause electron transitions between different conduction band minima. Intervalley scattering is usually accompanied by the absorption or emission of a longitudinal-mode optical phonon and its rate defined by an appropriate deformation potential. Since the phonon energy is comparable to the typical electron energy, intervalley scattering is generally regarded as inelastic. In the case of multiple parabolic valleys, the conditions of conservation of momentum and energy have the form

$$\mathbf{k}' = \mathbf{k}_{ij} + \mathbf{k} \pm \mathbf{q} \tag{3.38}$$

$$\frac{\hbar^2 k'^2}{2m^*} = \frac{\hbar^2 k^2}{2m^*} \pm \hbar\omega_{\mathbf{q}} \tag{3.39}$$

where $\mathbf{k_{ij}}$ is the vector between the minima of two different valleys, and $\mathbf{k}$ and $\mathbf{k}'$ are measured from the different valley minima. The matrix element for intervalley phonon scattering is [50]

$$|M_{\mathbf{k},\mathbf{k}'}| = \left(\frac{\hbar}{2NM\omega_{\mathbf{q}}}\right)^{1/2} D_{ij}[\sqrt{N_{\mathbf{q}}} \text{ or } \sqrt{(N_{\mathbf{q}}+1)}] \tag{3.40}$$

where D_{ij} is defined as the intervalley optical deformation potential constant for scattering between valleys i and j. In Si, scattering between the minima on a given axis (e.g. $< \bar{1}00 >$ to $< 100 >$) is called g-type scattering and is different from the so called $f-$type scattering between minima on different axes (e.g. $< \bar{1}00 >$ to $< 010 >$) with regard to the magnitude of the wavevector of the phonon involved. Hence the two processes are often considered separately and different matrix elements used for them. Denoting the number of equivalent valleys for the final state by Z_i the intervalley scattering rate is given by [55]:

$$S_{inter}(E) = \frac{D_{op}^2}{8\pi^2\rho\omega_{iv}} Z_i[N_o + \frac{1}{2} \pm \frac{1}{2}]N(E \pm \hbar\omega_{iv} - \Delta_{iv}) \tag{3.41}$$

where $N(E)$ is the density of states of a single valley. D_{op} is the coupling constant (deformation potential) for the relevant type of intervalley scattering, $\hbar\omega_{iv}$ is the intervalley phonon energy for the relevant phonon branch, and Δ_{iv} is the energy separation between the initial and final valley minima (if applicable). For bulk Si there are six degenerate Δ-minima, and two different types of intervalley scattering. For g-type scattering $Z_i = 1$, and $Z_i = 4$ for f-type scattering. The corresponding phonon energies are 61.2 and 59 meV [43] respectively. The density of states in Equation 3.41 is given by

$$N(E) = \frac{(m^*)^{3/2}}{2^{1/2}\hbar^3\pi^2}(1 + 2\alpha E)[E(1 + \alpha E)]^{1/2}, E > 0 \tag{3.42}$$

3.5.2 Ionised impurity scattering

The substitution of some of the host atoms by dopant atoms perturbs the periodic potential of a semiconductor crystal and causes electron scattering. Ionised impurity scattering is a significant process for low energy electrons in Si and Ge at 300 K when doping densities exceed 10^{23} m^{-3}, becoming comparable to low energy phonon scattering rates. At room temperature shallow impurities are normally assumed to be ionised in uncompensated material. An ionised donor produces a Coulomb potential which is screened by the free carriers

$$V_{ap}(r) = -\frac{e}{4\pi\epsilon_o\epsilon_r r}e^{-\kappa r} \tag{3.43}$$

where $\kappa^2 = ne^2/\varepsilon_o\epsilon_r k_B T$, n is the density of electrons in the conduction band, and ϵ_r is the relative dielectric constant of the material. The matrix element of the screened impurity potential is

$$M_{\mathbf{k},\mathbf{k}'} = \frac{e^2}{\epsilon_o\epsilon_r(|\mathbf{k}' - \mathbf{k}|^2 + \kappa^2)} \tag{3.44}$$

and, if N_I is the impurity concentration, the ionised scattering rate follows as

$$S(\mathbf{k}, \mathbf{k}') = \frac{1}{4\pi^2\hbar}\int \frac{e^4 N_I \delta(E(\mathbf{k}) - E(\mathbf{k}'))}{(\epsilon_o\epsilon_r)^2(4k^2\sin^2(\theta/2) + \kappa^2)^2}d^3k' \tag{3.45}$$

Inspection of the integrand in Equation 3.45 reveals that the scattering is anisotropic and favours forward scattering ($\theta = 0$). Note also that the argument of the delta function indicates elastic scattering, which is a consequence of the negligible recoil of the relatively massive impurity centre in the electron scattering process. For electrons with energy E in a valley with nonparabolicity coefficient α, the ionised impurity scattering rate is thus [56]:

$$\begin{aligned} S &= \frac{N_I}{8\sqrt{2m^*}\pi\epsilon_r^2\epsilon_o^2(1+2\alpha E)[E(1+\alpha E)]^{\frac{3}{2}}} \\ &\times \left[\alpha^2E^2 + \frac{[1+\alpha E(1+\xi)]^2}{\xi^2-1} + \alpha E[1+\alpha E(1+\xi)]\ln\frac{(\xi-1)}{(\xi+1)}\right] \end{aligned} \tag{3.46}$$

where $\xi = 1 + \kappa^{-2}/4m^*E(1+\alpha E)$. The approach described here is normally referred to as the Brooks and Herring model [57].

3.5.3 Alloy scattering

Alloy scattering arises from the random positioning of the different atomic constituents of the material and Harrison and Hauser [58] have described the theory of this process. The scattering is elastic and the matrix element is given by

$$M_{\mathbf{k},\mathbf{k}'} = \Delta U\left[\frac{x(1-x)}{N_c}\right]^{1/2}\delta_{\mathbf{k}\pm\mathbf{q},\mathbf{k}'} \tag{3.47}$$

where ΔU is the spherical scattering potential and N_c is the number of cation sites, defined as

$$V(r) = \begin{cases} \Delta U & \text{for } r \le r_0 \\ 0, & \text{for } r > r_0 \end{cases} \tag{3.48}$$

where r_0 defines the spatial extent of the scattering potential. The scattering rate for non-parabolic bands is given by

$$S(\mathbf{k}) = \frac{4\sqrt{2}\pi m^{*3/2}r_0^6}{9\hbar^4}\frac{x(1-x)(\Delta U)^2}{\Omega^2}\gamma^{1/2}(E)(1+2\alpha E) \tag{3.49}$$

where Ω is the volume of the primitive cell.

A very simple model for the scattering of electrons off rough heterointerfaces has been included in the Monte Carlo simulation. When an electron is incident on a heterointerface at which it would suffer a classical reflection, it is assumed to be reflected in a random direction relative to its initial path rather than suffer a specular reflection.

3.6 Types of Monte Carlo simulation

3.6.1 One-particle simulation

In this scheme, the trajectory of a single particle in phase space $(\mathbf{r}, \mathbf{k})$ is followed in time in order to obtain the steady state physical properties of an ensemble of similar non-interacting particles. According to the ergodic theorem, one can use the time averaged properties of the single particle as representative of the ensemble averaged properties at a given time for all the particles. This approach is most useful to study steady state problems such as mean carrier drift velocity as a function of field in a bulk semiconductor or device structure when carrier-carrier interactions or space charge effects are not significant.

3.6.2 Ensemble simulation

The algorithm is similar to that for one-particle simulation but the dynamics of a number of particles are simulated simultaneously, as illustrated in Figure 3.2. The advantage of this approach is that an ensemble distribution exists and its evolution with time as a result of time-dependent external influences can be studied if appropriate.

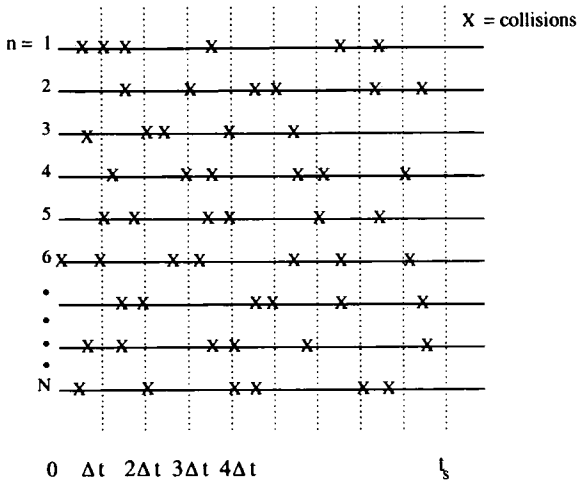


Figure 3.2: Illustration of distribution of scatterings (×) relative to the timesteps in ensemble simulation of the dynamics of N particles.

3.6.3 Self-consistent simulation

This method employs the ensemble Monte Carlo procedure in combination with the solution of Poisson's equation to take account of carrier space charge effects, and therefore is suitable for full device simulation.

3.7 Device Simulation Model

A self-consistent ensemble Monte Carlo device simulator called SLURPS (Software Library for Universal Random Particle Simulation) has been used in the research reported in this thesis. SLURPS has been developed in recent years at the University of Durham [59–61] and is composed of a library of routines, each of which is designed to perform a specific task in the simulation, such as drifting the particles, specifying charge distribution, solving Poisson's equation, etc. A flow chart of the SLURPS program is shown in Figure 3.3.

Since the number of electrons in a real device, N, is normally extremely large, it is impossible to simulate the motion of all of them by the Monte Carlo method. Hence, in the course of the ensemble Monte Carlo simulation, a set of N_{sp} superparticles which are representative of the electrons in the device is considered. For the superparticle dynamics to be representative of the electrons, the response of each superparticle to the electric field and scattering interactions must be the same as that of an electron. The use of superparticles means that the ensemble size is smaller than the actual number of particles in the physical system, and the mean values of physical quantities are only estimators of those in the true system with relative error that decreases as the number of superparticles is increased.

A) Setup geometry and discretisation scheme.

The first step in the device simulation is the definition of a simplified and tractable model of the actual device. The model needs to contain the essential physics of the device and its constituent materials but it must also be amenable to Monte Carlo simulation with the computational resources available. For example three main simplifications are made in modelling field effect transistors:

a) A two-dimensional real space model of the device is used. This means that there is no change of any physical quantity along the third (z) axis (known as the device width). Hence motion of particles in that direction has no effect on the behaviour of the model device. In effect, the superparticles may be considered to be charge rods extending infinitely in the z direction. The effective charge per unit length q_{sp} of the charge rod is taken as

$$q_{sp} = -en/N_{sp}W \tag{3.50}$$

where n is the number of electrons in the real device and W is the device width. This means that the total electronic charge within the real device is equal to the total su-

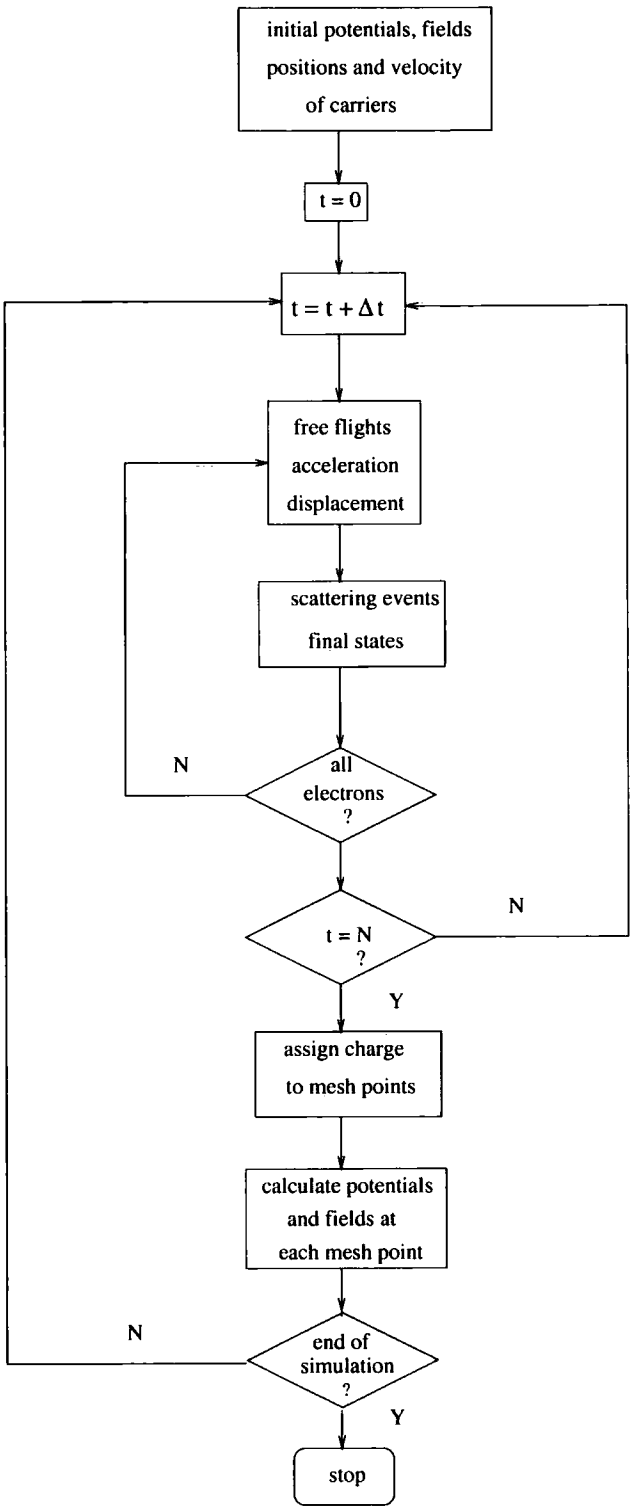


Figure 3.3: Flowchart of a self consistent MC program for device simulation.

perparticle charge in a range of z which is equal to the device width. This quantity is important for the calculation of the electric field within the device and the flow of current through the electrodes.

b) Source and drain are treated as ideal, zero-resistance ohmic contacts, and the gate is a Schottky barrier which is perfectly absorbing for electrons with energy in excess of the conduction band barrier.

c) Only a small part of the source and drain contacts is included in the model device because of their large extent in the real device.

It is worthwhile emphasising that the MC program is inherently three dimensional (3D) and the carrier dynamics in the 3D band structure is fully represented. A 2D real space model is used mainly because of the large computational demands of solving Poisson's equation in 3D, and the larger number of superparticles that would be required to describe effects in the third dimension.

The potential is represented by values on a regular array of mesh points. The charge density is given by the sum of the mobile charge, which can be obtained from the simulation via the particle-mesh calculation, and the fixed impurity profile which is specified at the start of the simulation. In SLURPS, model devices are built up as a series of joined rectangular regions with the mesh cell sizes matched along the joins between each region. Each region can consist of multiple layers of different alloy composition and doping/compensation density. For example, Figure 3.4 shows the type of finite difference grids used for a strained $Si/Si_{0.77}Ge_{0.23}$ n-channel MODFET whose simulation results will be described in Chapter 4. This model MODFET is made up of three regions that are beneath the source, drain and gate. Typically, the field cell dimensions used for the central region and the more highly doped source and drain implants are a few nanometres .

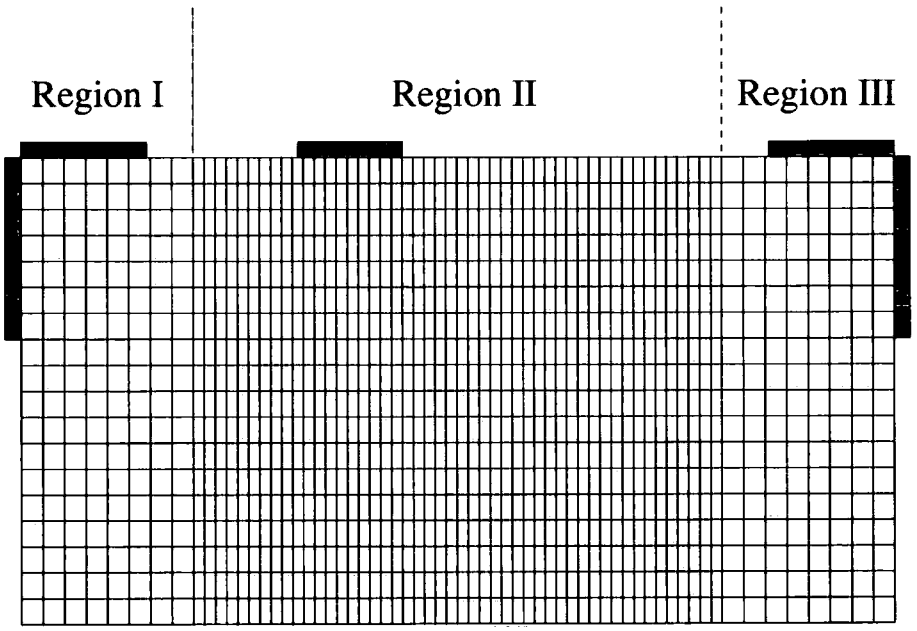


Figure 3.4: Illustration of a finite difference scheme used in the simulation of a strained $Si/Si_{0.77}Ge_{0.23}$ MODFET presented in Chapter 4.

B) Potential solution and boundary conditions

The solution of Poisson's equation requires a knowledge of the charge distribution in the device, which is available from the Monte Carlo simulation and can be represented by the charge assigned to each grid point. A fast solution of Poisson's equation in each rectangular region is provided by a combined Fourier analysis-cyclic reduction technique [61]. The use of a capacity matrix approach facilitates the solution of Poisson's equation in the joined rectangular regions described in the previous section with specified conditions at the device boundary. Ohmic contacts are modelled by specifying the value for the electric potential. The gate contact is a Schottky barrier which is represented by a contact with potential equal to the applied potential minus the barrier height. Elsewhere on the boundary, a zero value of the electric field normal to the boundary is prescribed. It is also necessary to specify the behaviour of particles when

they reach the surface of the device. It is required that there is no current flowing through the device boundary except at the electrodes, which is implemented in the model by reflecting any particle that hits the boundary. Ohmic contacts absorb all particles which are incident on them, as do Schottky contacts for electrons with energy in excess of the conduction band barrier. In addition, to maintain charge neutrality in the immediate vicinity of ohmic contacts, particles are added or removed as appropriate; in effect the source and the drain act as carrier reservoirs.

C) Charge assignment

The specification of the charge density profile is based on assigning the mobile superparticle and fixed dopant charge to the grid point by finding the total charge in the cell surrounding each grid point, as shown in Figure 3.5. The charge density ρ_{ij} at a grid point ij is taken to be the total charge per unit length in the cell surrounding the grid point divided by the cell area. The electrostatic potential values are calculated at the same grid points as indicated in Figure 3.5(b). In general the electrostatic potential ϕ is related to the charge density $\rho(x, y)$ by Poisson's equation

$$\nabla^2\phi(x,y) = -\frac{\rho(x,y)}{\epsilon_0\epsilon_r} \tag{3.51}$$

which can be discretised by employing the method of finite difference. Suppose that a rectangular region of the device is split into an n_x by n_y array of rectangular cells with each cell having dimensions Δx by Δy as in Figure 3.5(b). The cells are labelled with the indices $i = 0, ..., n_x - 1$ and $j = 0, ..., n_y - 1$. The electric field $F = -\nabla\phi$ together with the electric potential and charge density are discretised as F_{ij}, ϕ_{ij}, and ρ_{ij} respectively. If the value of electric field F^x_{ij} $[F^y_{ij}]$ refers to the interface of the cells ij and $(i-1)j$ $[i(j-1)]$ as shown as solid line in Figure 3.5(a), integrating Poisson's

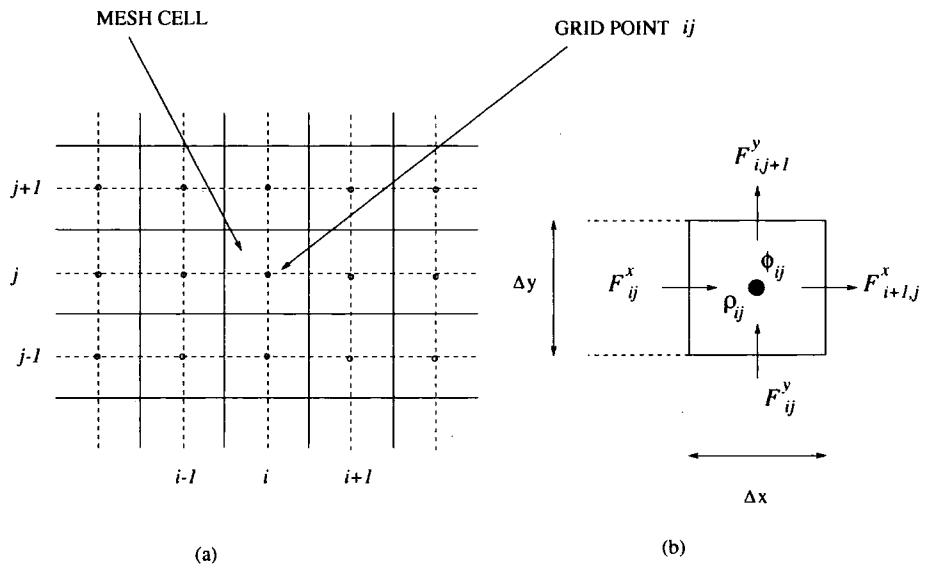


Figure 3.5: **(a)** Two-dimensional illustration of the nearest grid point scheme. **(b)** A cell for use with the discretised Poisson equation showing discretised electric field F_{ij}, electric potential ϕ_{ij} and charge density ρ_{ij}

equation over the area of an individual cell yields

$$-(F^x_{i+1,j} - F^x_{ij})\Delta y - (F^y_{i,j+1} - F^y_{ij})\Delta x = -\frac{\rho_{ij}\Delta x \Delta y}{\epsilon_0 \epsilon_r} \tag{3.52}$$

(which is actually an expression of Gauss's law) where

$$F^x_{ij} = -\frac{(\phi_{ij} - \phi_{i-1,j})}{\Delta x} \tag{3.53}$$

and

$$F^y_{ij} = -\frac{(\phi_{ij} - \phi_{i,j-1})}{\Delta y} \tag{3.54}$$

D) Carrier dynamics and scattering

Each superparticle undergoes the standard Monte Carlo sequence of scatterings and free flights in the local field previously determined from the solution of Poisson's

equation. The MC sequence is stopped after each time step (typically 1 femtosecond), when the field is adjusted by following the steps described above. The time evolution of the wavevector of a superparticle during a free flight starting at $t = 0$ is

$$\mathbf{k(t)} = \mathbf{k(0)} \pm \frac{e\mathbf{F}(t)}{\hbar} \tag{3.55}$$

where the plus sign applies for holes, and the minus for electrons. The differential scattering probability $S(\mathbf{k}, \mathbf{k}')$ that an electron undergoes a transition from an initial state $\mathbf{k}$ to a final state $\mathbf{k}'$ is calculated by Fermi's Golden Rule as described for the various scattering mechanisms in Sections 3.5.1-3.5.3. At each instant there exists a definite total scattering rate $S(\mathbf{k})$ for the superparticle, which is dependent on the carrier state, and can be obtained by adding up the scattering rates $S_i(\mathbf{k})$ for all possible scattering processes (labelled by i). The probability for a carrier to be scattered in the small time interval $[t, t + \delta t]$ is $S(\mathbf{k}(t))\delta t$. Thus the probability that the carrier will not scatter in the same interval is $(1 - S(\mathbf{k}(t))\delta t)$. The probability for a carrier to travel without scattering for a time t becomes simply

$$P_f(t) = \prod_i (1 - S(\mathbf{k}(t))\delta t_i) \tag{3.56}$$

where $t = \sum_i \delta t_i$. To handle Equation 3.56, it is easier to take the logarithm of both sides and we have

$$\begin{aligned} \ln(P_f(t)) &= \sum_i \ln(1 - S(\mathbf{k}(t))\delta t_i) \\ &\approx -\sum_i S(\mathbf{k}(t))\delta t_i \end{aligned} \tag{3.57}$$

assuming $S(\mathbf{k}(t))\delta t_i \ll 1$. The summation can be replaced by an integral, and hence

$$P_f(t) = \exp(-\int_o^t S(\mathbf{k}(t'))dt') \tag{3.58}$$

The probability therefore that an electron will suffer its next collision during an inter-

val dt around t is

$$\begin{aligned} p(t)dt &= P_f(t)S(\mathbf{k}(t))dt \\ \text{and hence } p(t) &= S(\mathbf{k(t)})e^{-\int dt' S(\mathbf{k(t')})} \end{aligned} \tag{3.59}$$

$p(t)$ represents a nonuniform distribution of free flight times over a semi-infinite interval. Consider selecting a time t_r by generating a random number r between 0 and 1 and solving the equation

$$\int_0^{t_r} p(t)dt = r \tag{3.60}$$

for t_r. The probability that the random number lies in the interval $r, r + dr$ is dr. Therefore, if we define dt_r by

$$\int_0^{t_r+dt_r} p(t)dt = r + dr \tag{3.61}$$

the probability $P(t_r)dt_r$ that the upper limit of the integral in Equation 3.59 lies in the interval $t_r, t_r + dt_r$ is equal to dr. Now if

$$P(t_r)dt_r = dr \tag{3.62}$$

it follows that

$$P(t_r) = \frac{dr}{dt_r} = p(t_r) \tag{3.63}$$

where the second equality comes from Equation 3.60. That is the probability of obtaining a particular t_r by solving Equation 3.60 for t_r when r is a random number is equivalent to the distribution of free flight times $p(t_r)$.

In practice, it is difficult to obtain the probability distribution of the scattering events by this method since the total scattering rate $S(\mathbf{k})$ is a complicated function of $\mathbf{k(t)}$. In the circumstances it is convenient to introduce an additional scattering process S_{ss}, called self-scattering [62], which actually has no effect on the state of the

carriers. The rate of self scattering is chosen so that

$$S(\mathbf{k}(t)) + S_{ss} = \Gamma \tag{3.64}$$

where Γ is the new total scattering rate and is a constant. Now free flight times can be simply calculated from Equations 3.59 and 3.60 as

$$t_r = -\frac{\ln(1-r)}{\Gamma} \tag{3.65}$$

Obviously, the self-scattering is just a mathematical trick, with no physical meaning. It does not alter the statistical distribution of the real scattering events but does make the selection of free flight times much easier. However, there is an added computational burden due to the increased number of scattering events. In practice, Γ is fixed as the largest scattering rate possible in the simulation, to be sure that S_{ss} is never negative.

At the end of the free flight, random numbers are required to select the type of scattering mechanism and to determine the electron state after scattering. The final state after scattering must have an energy $E(\mathbf{k}')$ given by

$$E(\mathbf{k}') = E(\mathbf{k}) \pm \hbar\omega_q \tag{3.66}$$

where $E(\mathbf{k})$ is the energy at the end of the free flight and, $\hbar\omega_q$ is the phonon energy in the case of lattice scattering. The particle wavevector after scattering is selected using random numbers and the results of the quantum mechanical theory for the process. To simplify the procedure for ellipsoidal valleys it is convenient to carry out a Herring-Vogt transformation of $\mathbf{k}$-space so that the constant energy surfaces are spherical as described in Section 3.4. Once the final state is known, the sequence of free flight/scattering is simply repeated with a new free flight.

3.8 Transport Model

Having described the general methodology of self-consistent ensemble Monte Carlo simulation, we now present the specific features of the models which are used in the device simulations reported in Chapters 4 and 5.

3.8.1 Band structure

The model of the conduction band structure used in the Monte Carlo simulations of bulk Si has six ellipsoidal non-parabolic Δ valleys located along the [100] directions at 85% of the way to the Brillouin zone edge from the zone centre [43]. The energy-wavevector relationship is taken to be:

$$E(1+\alpha E) = \frac{\hbar^2}{2}\left(\frac{k_l^2}{m_l} + \frac{k_t^2}{m_t}\right)$$

where k_l and k_t are the longitudinal and transverse components of the wavevector with respect to the $< 100 >$ crystallographic directions, $1/m_l$ and $1/m_t$ are the longitudinal and transverse components of the inverse effective-mass tensor, and α is the coefficient of nonparabolicity, as given in Table 3.2.

The effect of strain in Si layers is to split the Δ-valleys as described in Chapter 2. The energy difference between the twofold degenerate valleys and the fourfold degenerate valleys is determined by the deformation potential and the amount of strain. The strain induced modifications are parameterised as a function of Ge content according to the results of Rieger *et al.* [14]. The effective masses of the valleys are assumed to be unchanged. Changes of the nonparabolicity and scattering rates with strain have also been neglected.

To model the electron transport in bulk $Si_{1-x}Ge_x$(for $x \leq 0.8$), we assume that the conduction-band structure remains Si-like with six Δ-valleys whose effective masses remain unchanged. The additional mechanism of alloy scattering occurs in the alloy

parameters	Si	Ge
Lowest minima	Indirect	Indirect
Degeneracy	6	4
E_g (eV)	1.12	0.644
m_l/m	0.916	1.64
m_t/m	0.191	0.082
m_{hh}/m	0.50	0.44
m_{lh}/m	0.16	0.28
Δ_{so} (eV)	0.044	0.29
α (X valley eV^{-1})	0.5	0.65

Table 3.2: Energy band parameters for Si and Ge [42]

and is described by the method described in Section 3.5.3 using the potential in Table 3.3. The acoustic intravalley phonon scattering is treated as an elastic process. The deformation potential D_{ac} for $Si_{1-x}Ge_x$ is calculated by linear interpolation. The intervalley $\Delta - \Delta$ transitions are treated by considering both f and g processes with different deformation potentials of 9.3×10^{10} and 6.3×10^{10} eV m^{-1} respectively. In addition to electron transport, it is necessary to consider the transport of holes in the npn-SOI LBJT simulations. The maxima of the valence bands of Si are at Γ. The heavy, light and spin-split-off bands are represented by simple spherical parabolic approximations. Band effective masses and other material parameters are listed in Table 3.3.

It is worthwhile discussing the expected accuracy of the bandstructure model employed in comparison to the more sophisticated full band model. Figure 3.6 shows the first few conduction bands for Si obtained from an empirical-pseudopotential calculation [22]. In addition to the six equivalent ellipsoidal constant energy surfaces which form the conduction band edge, there is a second conduction band which is only 0.1 eV higher. Thus electrons with kinetic energy $\geq$ 0.1 eV may reside in either of the two conduction bands.

A comparison of the density of electronic states of the silicon conduction band

parameters	Si	Ge	$Si_{0.77}Ge_{0.23}$	$Si_{0.55}Ge_{0.45}$
density $\rho(\mathrm{kgm}^{-3})$	2330	5481	3018	3676
sound velocity $v_s(\mathrm{ms}^{-1})$	9044	5369	7767	6877
nonparabolicity(X valley) (eV^{-1})	0.5	0.65	0.385	0.275
effective mass (m_l^*/m_o)	0.9160	0.1791	1.117	1.31
effective mass (m_t^*/m_o)	0.1906	0.2040	0.1937	0.197
$c_{11}(\times 10^{10}\mathrm{N/m}^2)$	16.75	13.15	15.92	15.13
$c_{12}(\times 10^{10}\mathrm{N/m}^2)$	6.5	4.94	6.141	5.79
low-frequency dielectric constant (ϵ_s)	11.7	16.2	12.73	13.73
high-frequency dielectric constant (ϵ_∞)	10.82	16.2	12.06	13.24
acoustic deformation potential (D_{ac} (eV))	6.0	7.996	7.190	7.4
f, g deformation potential $(10^{10}$ eV $\mathrm{m}^{-1})$	9.3, 6.3		9.2, 9.2	9.2,9.2
lattice constant a (Å)	5.430	5.650	5.481	5.53
alloy scattering potential ΔU (eV)	-	-	0.8	0.7
bowing parameter	-	-	−0.19	−0.19
Energy band gap (eV)	1.12	0.975	1.053	1.008

Table 3.3: Material parameters at 300 K for Si, Ge, and $Si_{1-x}Ge_x$ used in the simulations [63,64]

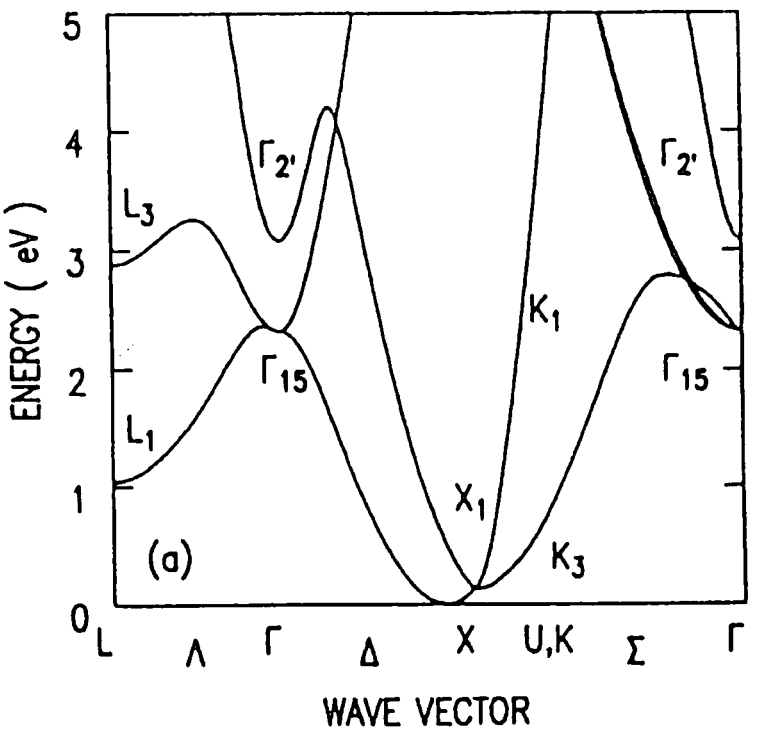


Figure 3.6: Band structure for Si obtained from the empirical-pseudopotential calculation [22].

using the parabolic and nonparabolic valley models and the full bandstructures is reported in the work of Kunikiyo *et al.* [65]. The study reveals that the nonparabolic band model provides a reasonable approximation to the realistic density of states up to about 2 eV above the band edge. Nevertheless, the full-band Monte Carlo investigation of electron transport in unstrained Si by Bufler *et al.* [66] reveals that the analytical bandstructure model results in an overestimate of the electron drift velocity at energies in a range above 130 meV. However, in the high energy region the drift velocity is strongly underestimated by the analytical bandstructure model.

Comparisons of the results of using parabolic, nonparabolic ellipsoidal and full bandstructure models for high energy electron transport in silicon have been reported by Abramo *et al.* [67]. The study deals with calculating the steady state energy distribution of electrons in homogeneous, intrinsic silicon at room temperature. The calculations suggest that the nonparabolic ellipsoidal bandstructure model is a good approximation for electron energies up to about 1 eV. However, for higher energies, detailed full-band calculations are essential for accurate results.

Obviously, self-consistent ensemble full-band Monte Carlo simulation of semiconductor devices is the ideal for obtaining insight into the full details of hot electron transport in semiconductor devices [22,65]. However, full-band Monte Carlo simulators involve large processing time even with major computational resources, which in any case are not generally available for device technology development. Hence simulations based on nonparabolic elliposiodal, nonparabolic spherical or even simple parabolic bandstructure models are still widely used and have had substantial success in device simulations [42–44]. For example, Formicone *et al.* used a nonparabolic ellipsoidal ensemble Monte Carlo simulation to model hole and electron transport in strained Si [68] and submicron $Si_{1-x}Ge_x$ based MOSFETs [69]. Monte Carlo simulations of $Si_{1-x}Ge_x$ MODFETs by use of nonparabolic ellipsoidal bandstructure model of $Si_{1-x}Ge_x$ MODFETs were carried out by Dollfus [70,71].

Chapter 4

MODFET Simulations

4.1 Introduction

In an attempt to overcome some of the limiting physical effects due to aggressive scaling of MOSFETs, while maintaining compatibility with conventional processes, several new technologies have been developed as described in Chapter 2. In particular, during the past few years, modulation doped Si/SiGe heterostructures grown on relaxed SiGe virtual substrates have been shown to produce enhanced electron mobility, and have great potential for the development of fast Si-based MODFETs [19,37,72,73]. The improved electron transport in strained $Si/Si_{1-x}Ge_x$ MODFETs and the natural compatibility with existing Si-CMOS technology makes the devices an attractive prospect for improving the circuit performance and packing density of ultra large scale integrated circuits (ULSICs). Of interest in this chapter are the impact of strain on high field transport and the implications for the performance of deep submicron devices.

In the deep submicron regime, non-equilibrium carrier transport is expected to have a significant effect on the performance of field-effect transistors. Optimum transistor design is ultimately based on a full understanding and accurate modelling of carrier transport in the device and Monte Carlo simulation is now well established as a powerful tool for studying the microscopic behaviour of carriers in semiconductor devices

where semiclassical models of carrier transport are appropriate [44,45]. In this chapter, we describe studies of electron transport in n-channel $Si/Si_{0.77}Ge_{0.23}$ and $Si/Si_{0.55}Ge_{0.45}$ MODFETs using 2D self-consistent ensemble Monte Carlo simulation. The details of the transport model were described in Chapter 3. The simulation provides information on the microscopic details of the carrier behaviour, including carrier velocity, kinetic energy and carrier density, as a function of position in the device, and provides a physical insight into device behaviour. Analysis of the microscopic and macroscopic data can provide valuable understanding of the device properties, and in addition, indicate possible ways to improve device performance.

Section 4.2, describes a detailed analysis of the in-plane electron velocity-field characteristics, of strained Si grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates, and also of unstrained Si. Both steady state and transient bulk properties have been studied. For the steady state bulk simulations, the drift velocity of the electrons plotted against the homogeneous electric field is obtained. In small devices, it is expected that velocity overshoot will play an important role in aiding high frequency performance. To this end, a study of the transient response of electrons to a stepped application of electric field has been performed.

The simulations of n-channel Si/SiGe MODFETs that have been performed are described in Section 4.3. The simulated devices can be classified into two groups; moderate and high tensilely strained MODFETs. Four depletion mode devices are considered in the first group; three of them make use of a 0.23 Ge fraction whilst a 0.25 fraction is used for the fourth. The simulations have been carried out in order to investigate the high frequency performance. The effect of varying gate length L_G and source-gate separation L_{SG} are studied. For the higher Ge fraction of 0.45, two devices working in the depletion and enhancement modes are studied. The model devices are similar to those experimentally investigated by Glück *et al.* [73]. For each group of devices the layer design and the geometries of the simulated devices are described first

and then the simulation results are presented and discussed. The simulations provide microscopic information on the carrier dynamics and also the output characteristics (drain current versus drain voltage), from which the transconductance (g_m) can be obtained. The frequency response of the MODFETs is investigated by modelling the response of the device when a voltage pulse is applied to the gate.

4.2 Simulation of electron transport in bulk Si and strained Si

4.2.1 Steady state simulation results

To investigate the performance of small, high-speed Si/$Si_{1-x}Ge_x$ MODFETs, it is instructive to know the behaviour of the drift velocity of electrons versus electric field for steady state and transient conditions in bulk material. To this end we have performed a detailed analysis of the in-plane velocity characteristics at 300 K of unstrained bulk Si and of bulk Si strained as it would be if it were grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates at 300 K. The present section is devoted to the steady-state case. The calculated steady state velocity-field curves are shown in Figure 4.1. The impurity concentration is taken to be $1\times10^{21}m^{-3}$. The electric field is varied from 10^4 to 2×10^7 Vm^{-1} and the orientation of the electric field is chosen to be parallel to the [100] direction, which is the case in the active layer of most Si/SiGe MODFETs.

Figure 4.1 shows that the velocity increases monotonically with increasing applied field, but eventually saturates at high field for both the unstrained Si and strained Si cases. It is apparent that strain causes a significant enhancement of the in-plane drift velocity-field for both $x = 0.23$ and 0.45 substrates, at least for electric fields up to about 15 MVm^{-1}. As explained in Chapter 2, the strain in Si splits the six-fold degen-

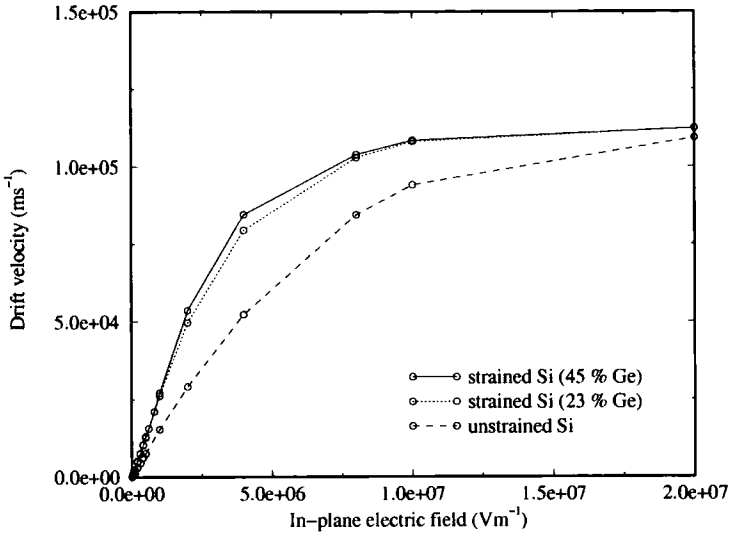


Figure 4.1: Calculated steady state drift velocity of electrons in bulk Si as if grown on $Si_{0.77}Ge_{0.23}$, $Si_{0.55}Ge_{0.45}$ virtual substrates and bulk Si. The impurity concentration is assumed to be 10^{21} m^{-3}.

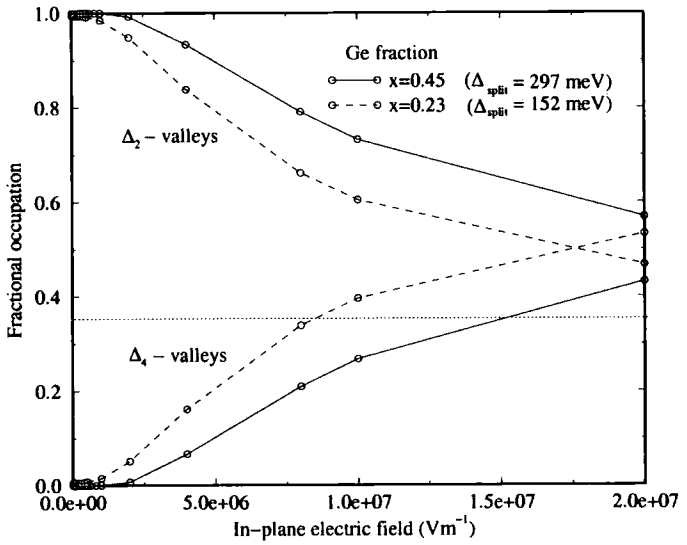


Figure 4.2: Fractional valley occupation in bulk Si as if grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates and bulk Si respectively. The impurity concentration is 10^{21} m^{-3}. The dotted line represents the fractional occupation of electrons in the Δ_2 valleys for bulk Si.

erate Δ valleys into two groups. As a result the four valleys (Δ_4) on the Δ axes parallel to the plane of the heterostructure lie higher in energy than the two valleys (Δ_2) on the Δ axes normal to the plane, which have low mass parallel to the field. The velocity enhancement in strained Si at 300 K is partly due to the smaller in-plane effective mass experienced by the electrons in the two lower energy valleys. There is also reduced f-type intervalley scattering. Note that the drift velocity for moderate tensile strain shows similar general behaviour to the high tensile strain case. Inspection of Figure 4.1 reveals that the velocity enhancement is more efficient at low and intermediate electric fields; a consequence of the fact that the higher electric field strengths increase the population of the fourfold higher valleys with their larger mean effective mass in the direction of transport. Under a low in-plane electric field ($F_{||}$ <10^6 Vm^{-1}), and a Ge fraction of 0.23 in the virtual substrate, the valley splitting energy $\Delta E_{split} \sim 150$ meV is sufficient for most of the electrons to occupy the Δ_2 valleys as shown in Figure 4.2. Under higher electric fields the electrons tend to distribute among the Δ_2 and Δ_4 valleys and more electrons reside in the Δ_4 valleys at 20 MVm^{-1}. For the higher Ge content of 0.45, the larger valleys splitting ($\Delta E_{split} \sim 300$ meV) means that more than 50% of electrons still occupy the Δ_2 valleys at 20 MVm^{-1}. However, the saturation velocities are practically identical in both cases, at about 10^5 ms^{-1}, which is consistent with experiment and other recent calculations [66,70,74–78]. For example the experiments of [74] show the saturation velocity of electrons in strained Si to be about 0.9×10^5 ms^{-1} at 295 K. The full band Monte Carlo simulations of Fischer *et al.* [78] predict that the calculated electron drift velocity saturates at about 0.95×10^5 ms^{-1} in strained Si at 300 K, whilst the value for unstrained Si is slightly lower. Another full band Monte Carlo study by Bufler *et al.* [66] shows a similar result in which the saturation velocity of electrons in strained Si grown on $Si_{0.7}Ge_{0.3}$ is about 1.05×10^5 ms^{-1} whilst that for unstrained Si is 10^5 ms^{-1}.

In the very high field regime, the electrons are distributed among the Δ valleys in

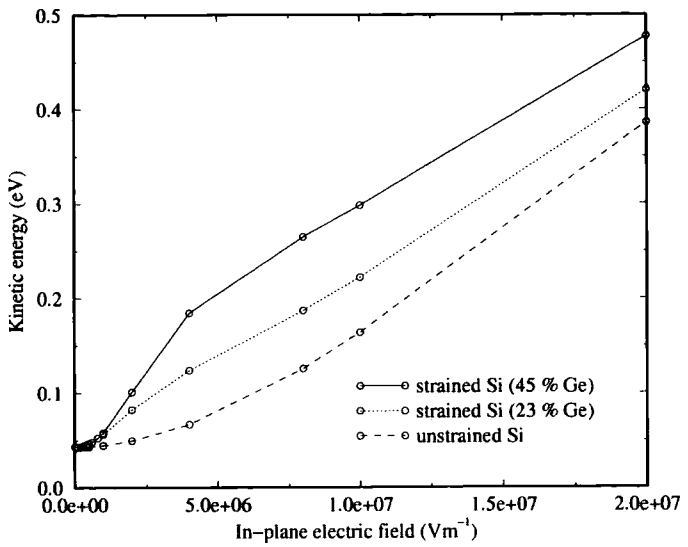


Figure 4.3: Average kinetic energy of electrons in bulk Si as if grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates and bulk Si, labelled as strained Si and unstrained Si respectively. The impurity concentration is assumed to be 10^{21} m^{-3}.

much the same way as in unstrained Si, and the drift velocity is similar. The velocity saturation occurs in both strained and unstrained Si because the electrons lose energy by optical phonon emission.

Figure 4.3 demonstrates the higher average kinetic energy of electrons in Si grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates compared with that of unstrained Si for a given applied field along a [100] direction. The kinetic energy at thermal equilibrium is $3k_BT/2 = 39$ meV at 300 K and increases monotonically with the field. The average kinetic energy increases with substrate Ge content for fixed field since the electrons in the Δ_2 valleys must achieve a higher energy before they can transfer to the Δ_4 valleys.

As seen in Figure 4.3 the maximum average kinetic energy about 0.5 eV is obtained for the applied field of 2×10^7 Vm^{-1}. Referring to Figure 3.6, it is expected that substantial number of electrons can occupy the second conduction band, affecting

the electron transport. In fact, the highest electric fields encountered in Monte Carlo simulation reported in this thesis is about 3×10^7 Vm^{-1} as will be seen in Figure 4.32. The corresponding maximum average kinetic energy is about 0.5 eV as illustrated in Figure 4.33. Thus, the results such as average kinetic energy and velocity obtained by use of an analytic bandstructure model as employed here may differ significantly from those obtained by use of full band model. However, the overall output characteristics such as drain current, transconductance and cut-off frequency may not differ from those obtained by use of full band model.

4.2.2 Transient response simulation results

In this section, we describe the use of ensemble Monte Carlo simulation to study the transient response to a stepped electric field of electrons in strained and unstrained Si. Particular emphasis is placed on those short-time-short-distance phenomena that are expected to be important in small and very small devices. High electric fields are common in short-channel devices, and thus the transient or saturation velocities rather than the low field velocity (as determined by the low field mobility) may ultimately limit the performance of small devices.

Transient transport of electrons in unstrained and strained Si is simulated for electric fields of 2×10^6, 6×10^6 and 2×10^7 Vm^{-1} applied along the [100] direction. Before the field is switched on, the superparticle ensemble is set up with a Maxwellian energy distribution and associated random velocity distribution occupying the states of the degenerate Δ-valleys of unstrained Si and the split Δ-valleys in the strained case. These are the conditions that would exist if the ensemble had been set up in some arbitrary distribution and a simulation had been carried out at zero field until equilibrium was achieved.

Figure 4.4 shows the transient response of the electron drift velocity at 300 K to

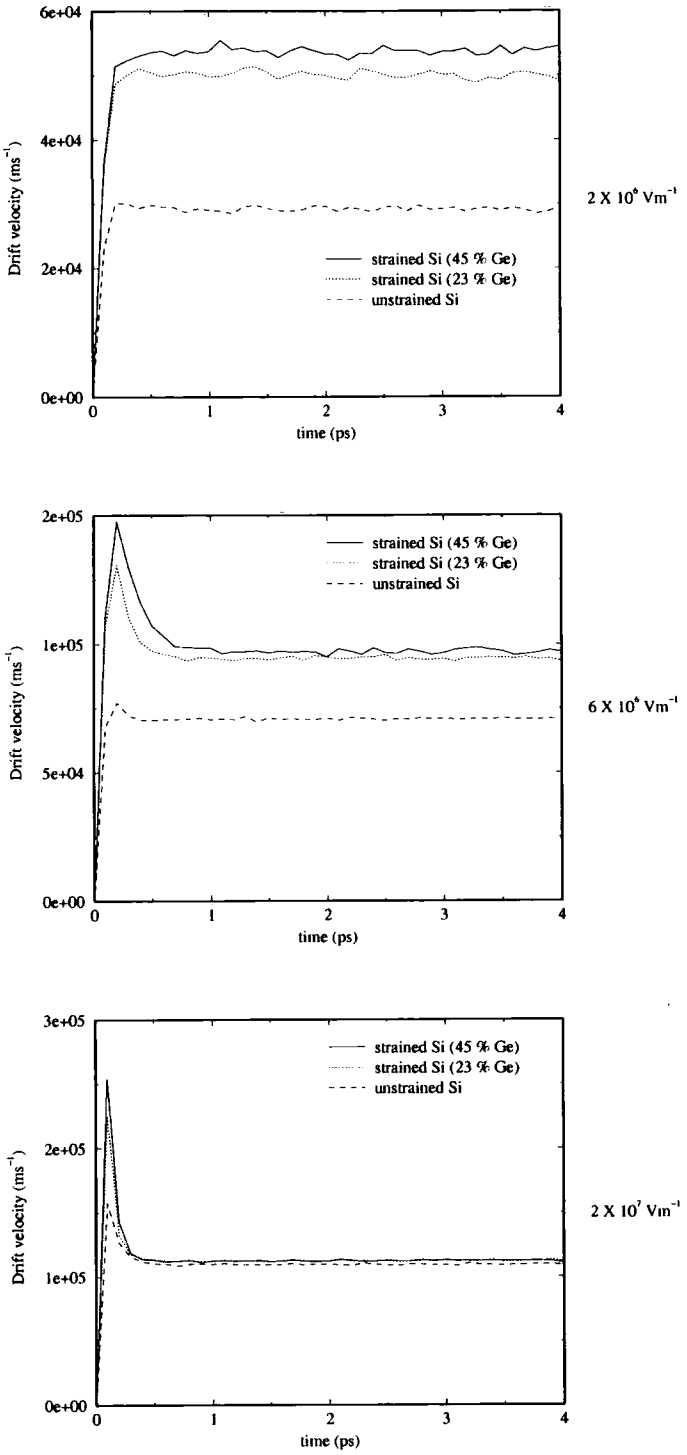


Figure 4.4: From top to bottom, average drift velocity of electrons in bulk Si as if grown on $Si_{0.55}Ge_{0.45}$, $Si_{0.77}Ge_{0.23}$ virtual substrates are plotted as a function of time. The relevant data for unstrained bulk Si has been plotted as a reference. The applied electric field intensities are 2×10^6, 6×10^6, and 2×10^7 Vm^{-1} respectively. All applied electric fields are oriented along the [100] direction.

applied fields of 2×10^6, 6×10^6, and 2×10^7 Vm^{-1}. Velocity overshoot is not found for the lowest field although the results do confirm the enhanced steady state drift velocity for strained Si. At the higher field of 6×10^6 Vm^{-1}, significant velocity overshoot occurs in strained Si reaching the peak of 2×10^5 ms^{-1} for the higher tensile strain case, but the effect in unstrained Si is almost negligible. For the highest field, the peak velocity is about 2.5×10^5 ms^{-1} for the higher tensile strain and a smaller but significant effect occurs in the unstrained material. In all cases, the velocity reaches its saturation value in less than $\sim$ 1 ps. Consequently, it is possible that the transient overshoot could be used to partly overcome some limitations due to the velocity saturation for devices in the deep submicron regime. Velocity overshoot has a significant effect when the carrier transit time along the channel is comparable to the energy relaxation time and the changes in the Si band structure due to strain may have a significant impact on the effective channel velocity. As the momentum relaxation time is smaller, the electron distribution is shifted in momentum space but it is somewhat later that energy relaxation becomes effective so that the distribution function spreads out, and the drift velocity decreases. This is particularly the case if some scattering has an energy threshold as results from the energy splitting of the Δ valleys in strained Si.

Figure 4.5, shows the Δ_2 and Δ_4 valley occupancies as a function of time for applied fields of 2×10^6, 6×10^6, and 2×10^7 Vm^{-1} at 300 K. The top picture corresponds to the moderate tensile strain case whilst the bottom is for the high tensile strain case. The plots show that at the highest field about 60% of electrons occupy the Δ_2 valleys for the higher tensile strain in agreement with the steady state results of Figure 4.2.

4.3 Device simulation

A substantial improvement of the in-plane electron transport properties of tensilely strained Si layers is suggested by the results on bulk material in the previous sections.

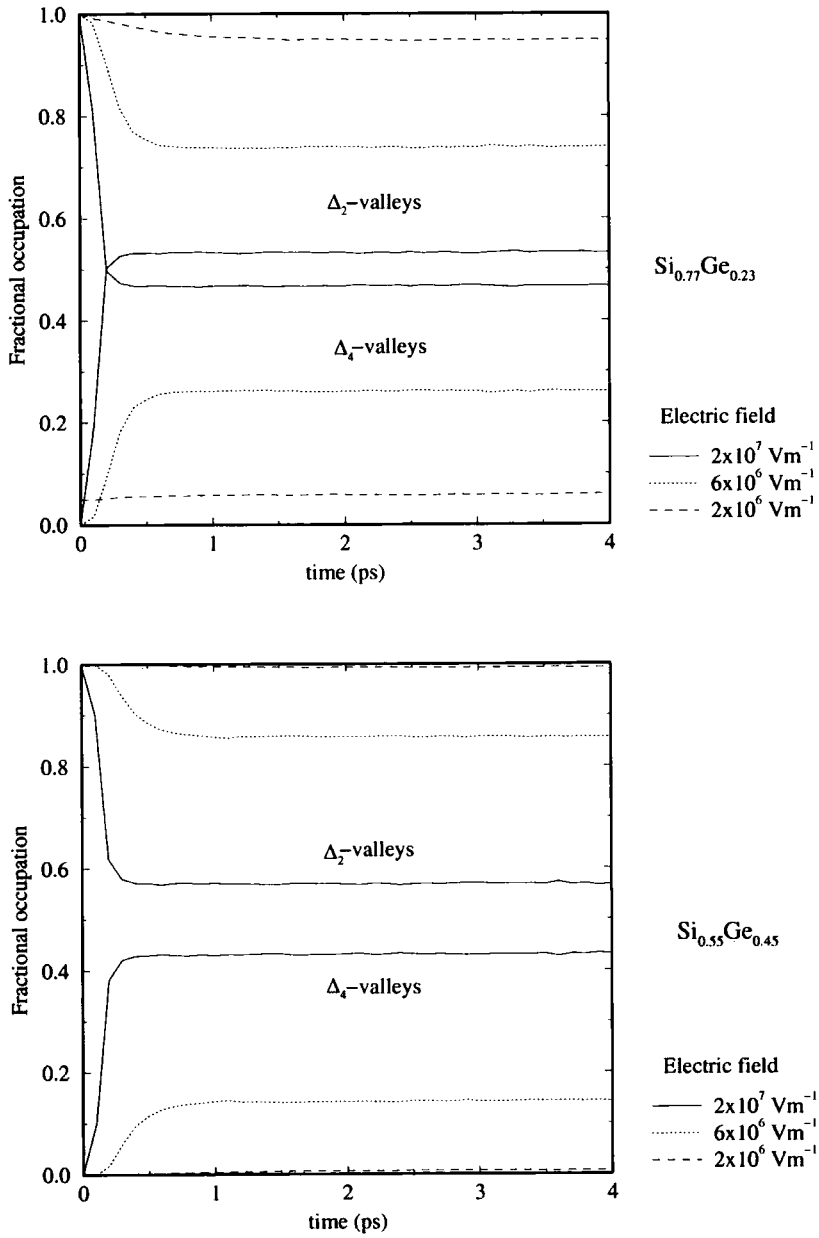


Figure 4.5: Proportion of electrons in Δ_2 and Δ_4 valleys in bulk Si as if grown on $Si_{0.77}Ge_{0.23}$ (top picture) and $Si_{0.55}Ge_{0.45}$ (bottom picture) virtual substrates as a function of time for various applied electric field strength 2×10^6, 6×10^6, and 2×10^7 Vm^{-1} respectively. The electric field is oriented along the [100] direction.

We found an enhanced drift velocity up to fields of approximately 2×10^7 Vm^{-1} which can be expected to reduce the electron transit time in the high field regions of a MODFET. In addition the low field mobility is much increased, which is beneficial in the source and drain contact regions. Further, we have demonstrated overshoot effects resulting in a peak velocity of 2.5×10^5 ms^{-1}. Such effects in a short channel MODFET, could reduce electron transit times further and consequently boost the dynamic performance of the transistor.

To investigate the improvement in device performance that might be achieved due to the enhancement of carrier transport properties in strained Si layers, ensemble self-consistent Monte Carlo simulations have been performed for n-channel $Si/Si_{1-x}Ge_x$ MODFETs. The simulated devices can be classified as moderate tensile strain ($x = 0.23$ and 0.25 Ge fraction) and high tensile strain ($x = 0.45$ Ge fraction). In an attempt to optimise n-MODFET performance, the following factors have been considered [79,80].

1) SiGe buffer

Obviously, a larger Ge content in the buffer provides stronger electron confinement in the channel due to the larger strain induced splitting of the Δ_2-Δ_4 valleys. A large conduction band discontinuity is desirable to confine the electrons to the Si channel. However, in practice the production of Ge-rich relaxed virtual substrates requires the growth of deep graded buffer layers which are several microns in depth, and which is comparatively costly.

2) Strained Si channel

The thickness of the strained Si layer has to be kept below a critical thickness. Otherwise, the homogenous strain energy, which increases with increasing layer thickness, becomes so large that it becomes energetically favourable for misfit dislocations to be introduced. Normally the thickness of the channel is around 5-20 nm [81]. The

Si channel is the desired path for source to drain electron transport but in some cases electrons cannot transfer completely from the supply layer to the channel. The supply layer is then a second, parallel path with a low mobility as a result of the presence of donor impurities and alloy scattering.

3) SiGe spacer layer

The spacer layer is an important element of modulation-doped structures since it further increases the separation of electrons from the ionised donor impurities, which can improve the electron mobility. Nonetheless, a thick spacer can also decrease the electron concentration. A trade-off between the electron mobility and the electron concentration in the channel has to be made.

4) Concentration of the 2D electron gas in the strained Si channel

If we assume that all donor levels in the $Si_{1-x}Ge_x$ supply layer are activated and the electrons are transferred to the Si-channel layer, the area density is $N_D d$, where N_D is the doping density in the supply layer and d is the thickness of the supply layer. To obtain a high electron density in the channel it is necessary to have a high donor concentration in the supply layer and a thin cap layer and gate-to-channel separation to ensure effective charge transfer from the supply layer to the channel. Reducing the gate-to-channel separation also increases the transconductance but if taken too far it can result in a decrease of the electron mobility due to interface roughness effects.

Device	Ge content	L_G (μm)	L_{SG} (μm)	L_{SD} (μm)	h_{CH} (nm)	h_{GC} (nm)	L_G/h_{GC}
1	0.23	0.09	0.09	0.5	10	35	2.57
2	0.23	0.1	0.1	0.3	10	35	2.86
3	0.23	0.07	0.05	0.25	10	35	2.00
4	0.25	0.2	0.15	0.5	5	25	8.00

Table 4.1: Device parameters used in simulations of n-channel SiGe MODFETs with moderate tensile strain. For devices 1-3, Ge fraction of 0.23 is used. Device 4 employs a 0.25 Ge fraction. (L_G : gate length, L_{SG} : source-gate separation, L_{SD} : source-drain separation, h_{CH} : channel depth, h_{GC} : gate-to-channel separation, L_G/h_{GC} : aspect ratio)

4.4 Results for moderate tensilely strained Si n-channel MODFETs

4.4.1 Layer design

To achieve a reasonable conduction-band offset between Si and $Si_{1-x}Ge_x$, a high Ge content in the $Si_{1-x}Ge_x$ layer is required. Figure 2.8(a) shows the conduction band offset between strained Si and $Si_{0.77}Ge_{0.23}$ to be about 139 meV. In addition, the splitting of the Δ_2-Δ_4 valleys is about 152 meV for a strained Si layer on $Si_{0.77}Ge_{0.23}$, and scattering into the higher valleys in the silicon layer can only happen if the electrons actually have sufficient energy to escape the channel well. From the technological point of view, smaller Ge fraction layers are preferred because of the lower cost of fabrication. In this regard, experimental results for strained Si n-MODFETs show that promising performance is possible with a 0.2-0.3 Ge fraction [37,72]. For example, a cut-off frequency f_T as high as 62 GHz for an n-Si/$Si_{0.75}Ge_{0.25}$ MODFET at a low drain bias of 1 V has been reported recently by the IBM group [37].

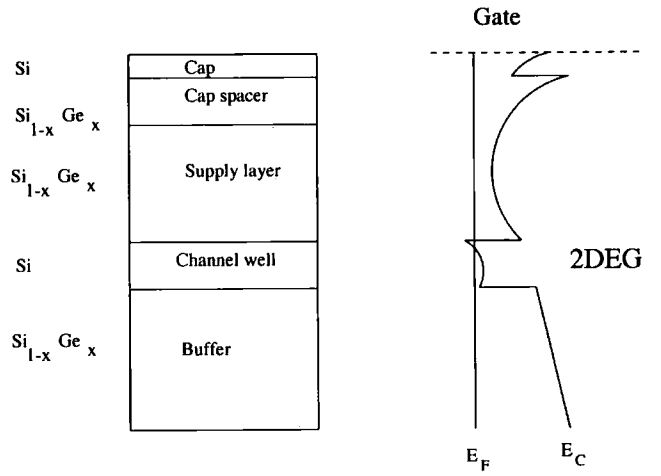
We have studied three depletion[a] mode devices with a modulation-doped het-

[a]Field effect transistors can be designed to work in enhancement or depletion mode, depending on the required applications. Enhancement or depletion mode operation can be achieved by the

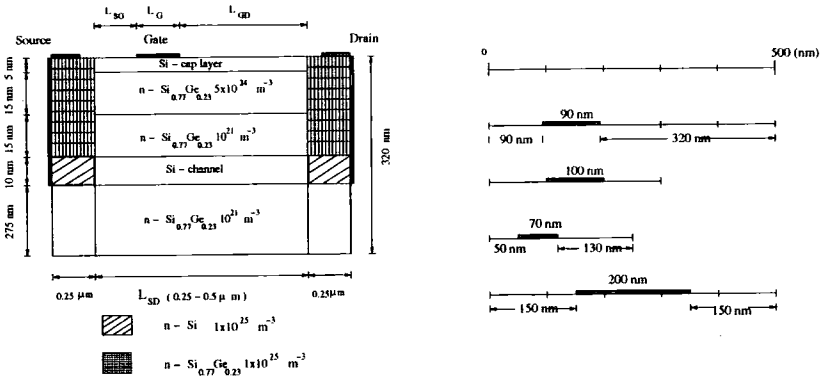
erostructure of $Si_{0.77}Ge_{0.23}/Si/Si_{0.77}Ge_{0.23}$ and gate lengths of 0.09, 0.1 and 0.07 μm (labelled as device 1, 2, and 3 respectively and illustrated in Figure 4.6). The devices are similar to those being investigated experimentally by a number of groups, such as Daimler-Chrysler [73], and IBM [82]. The effect of varying the gate length L_G and source-gate separation L_{SG} has been studied. The 10 nm thick tensile strained Si channel is undoped and separated from the supply SiGe layer by a 15 nm thick $Si_{0.77}Ge_{0.23}$ spacer layer. The doping of the supply layer $Si_{0.77}Ge_{0.23}$ is taken to be 5.0×10^{24} m^{-3}, and the background doping of the substrate is 10^{21} m^{-3}. The device parameters are summarised in Table 4.1. The gate is deposited on a 5 nm thick Si cap-layer doped to 10^{21} m^{-3}. The gate Schottky barrier height Φ_B is assumed to be 0.8 eV. The gate-to-channel separation is 35 nm. The source and drain implants have a doping density of 10^{25} m^{-3}. The temperature for the simulations has been set at 300 K.

In addition to the three devices described above, another depletion mode device (labelled as device 4) has been considered whose structure is similar to the $Si/Si_{0.75}Ge_{0.25}$ MODFET demonstrated by the IBM group with $f_T = 62$ GHz at 1 V drain bias [37]. The structure of devices 1-3 has been modified to get something as close as possible to the device investigated experimentally by Koester *et al.* [37]. Device 4 has a modulation-doped heterostructure of $Si_{0.75}Ge_{0.25}/Si/Si_{0.75}Ge_{0.25}$. The 10 nm thick tensile strained Si channel is undoped and separated from the donor supply layer by a thin (5 nm) spacer layer $Si_{0.75}Ge_{0.25}$. The doping of the 15 nm thick supply layer is 4×10^{24} m^{-3}, the background doping of the substrate is the same as for devices 1-3 but the source and drain implants are now 4×10^{24} m^{-3}. The gate is centred between the source and drain contacts and deposited on a 5 nm Si cap-layer doped to 10^{21} m^{-3}. Note that the gate-to-channel distance is now reduced to 25 nm. In order to get output characteristics as close as possible to those reported experimentally, the gate Schottky barrier height Φ_B has been varied from 0.8 – 1 V. The simulations have been done for

appropriate choice of doping level or by recessing the gate so that it is closer to the channel.



(a) device layers and conduction band energy



(b) simulated structures and device parameters

Figure 4.6: **(a)** Layer design of Si/SiGe n-MODFET with moderate tensile strain ($x = 0.23$ or 0.25) and the corresponding conduction band profile in equilibrium. At these Ge contents the band offset at the heterojunction is about 150 meV. **(b)** The simulated device structure. A 10 nm tensile strained Si quantum well is grown on a relaxed $Si_{0.77}Ge_{0.23}$ virtual substrate, and is modulation doped via the top supply layer (15 nm, 5×10^{24} m^{-3}). The confined electron gas is separated from the supply layer by a 15 nm spacer (background doping density 10^{21} m^{-3}). Source and drain implants are taken to be 1×10^{25} m^{-3}. The gate is deposited on the top of the Si-cap layer (5 nm) and a Schottky barrier height of 0.8 eV has been used to represent the contact potential at the gate electrode.

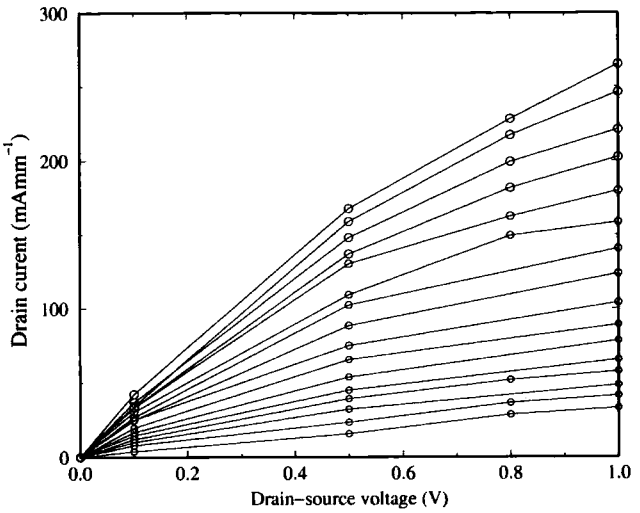


Figure 4.7: Simulated drain current plotted as a function of drain voltage at different gate voltages for the simulated $Si/Si_{0.77}Ge_{0.23}$ MODFET: device 1 illustrated in Figure 4.6 with a gate length 0.09 μm. The gate voltage ranges from -1 V to $+0.5$ V in 0.1 V intervals.

a temperature of 300 K.

Initially, the cell size used for the mesh in the central region (see Figure 3.4) in the simulations of device 3 was 5.21×5 nm^2 (horizontal×vertical) and 7.81×5 nm^2 for the more highly doped source and drain implants. However, due to the small channel and gate lengths employed in this structure, the effect of reducing the cell dimension in the longitudinal (source-drain) direction was investigated. The number of mesh cells was 48×64 in the central region. Further simulations in which the number of mesh cells in the central region were 64×64 and 96×64 respectively have been performed. The corresponding mesh dimensions are 3.91×5 nm^2 and 2.60×5 nm^2. Figure 4.8 shows the Δ_2-valley population, average kinetic energy, drift velocity of electrons, and longitudinal electric field recorded along the channel of device 3. The bias condition is $V_{DS} = 1$ V and $V_{GS} = 0.3$ V. It can be seen that the results obtained using the smaller and smallest cell dimensions have somewhat higher values compared to those using the initial cell dimensions. However, the results for the smaller and smallest cell

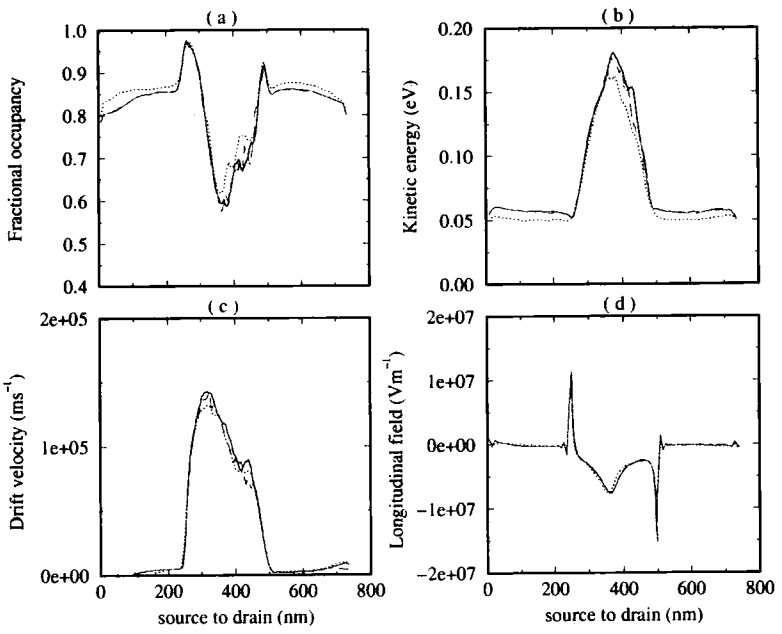


Figure 4.8: **(a)** Δ_2-valley population recorded along the channel, **(b)** Average kinetic energy of Δ_2 and Δ_4 electrons, **(c)** Average longitudinal velocity of Δ_2 and Δ_4 electrons, **(d)** Longitudinal electric field. Note the purple dashed line, black line, and red dashed line represent the data by use the number of mesh cells in the central region of device 3 of 48 × 64, 64 × 64, and 96 × 64 respectively.

dimensions are very similar, which suggests that cells with dimensions of 3.91 × 5 nm^2 can be used to ensure results of acceptable accuracy without too large a penalty in terms of computing time. After following a similar procedure for the other devices, the cell dimensions in the central region used for devices 1-4 were chosen as 5.21 × 5 nm^2, 4.69 × 5 nm^2, 3.91 × 5 nm^2, and 7.81 × 5 nm^2, respectively.

The cells have a vertical dimension of 5 nm which is half the channel thickness. It would be nice to obtain results with smaller cells but 5 nm is the minimum value that can be achieved in the MODFET devices studied here. Further reduction of the cell size in the vertical (gate-buffer) direction is not possible due to the memory requirement in the computations. The problem lies in solving Poisson's equation by combining Fourier transforms in the horizontal direction (source-gate) with Buneman cyclic reduction in

the vertical direction (gate-buffer) [61]. The flexibility of the fast Fourier transform method permits practically any number of cells for the horizontal extent of a region, but the number of cell in the vertical direction is restricted to a power of two. Because of the need for the cells to match at joins between regions, doubling the number of cells in the central region also require doubling the number of cells in the source and drain regions. It might seem that the problem could be circumvented by splitting the central region and having a small region which with includes the channel. However, then the total number of joins exceeds a predefined limit in SLURPS.

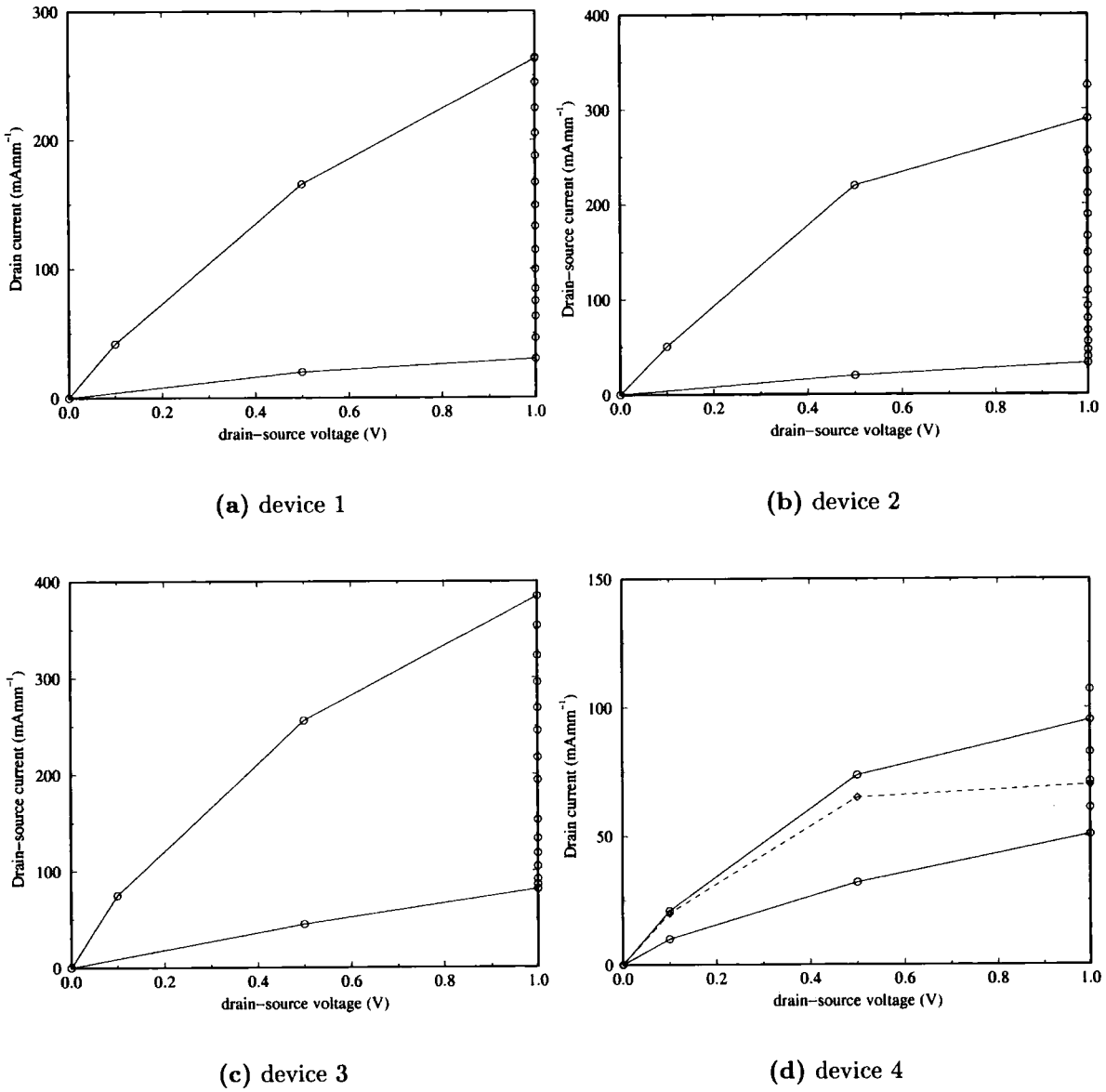
4.4.2 Device Characteristics and Analysis

1) Static (DC) characteristics

Simulations of the dc current characteristics have been carried out using 30,000 electron superparticles. Figure 4.7 shows the simulated output characteristics of device 1 for various applied gate voltages. The device works in depletion mode which means that a negative bias at the gate is required to turn the conducting channel off. The calculated intrinsic drain current is about 270 mA mm^{-1} at V_{DS} = 1 V and V_{GS} = 0.5 V. Because of the large amount of computing time required for the simulations, a complete set of output characteristics has only been calculated for device 1 and then only up to V_{DS} = 1 V. For the remaining devices, Figure 4.9 shows the output characteristics at selected biases. One of the most important figures of merit of a transistor is the transconductance (g_m), which relates the small-signal drain current to the small-signal gate voltage at fixed drain voltage [33]:

$$g_m = \left.\frac{\Delta I_D}{\Delta V_{GS}}\right|_{V_{DS}=constant} \tag{4.1}$$

Simulations at V_{DS} = 1 V and with gate bias varied from −1.0 V to 0.5 V in 0.1 V



(a) device 1

(b) device 2

(c) device 3

(d) device 4

Figure 4.9: Simulated drain current plotted as a function of drain voltage at maximum and minimum gate voltages for simulated $Si/Si_{0.77}Ge_{0.23}$ MODFETs (device 1-3) and $Si/Si_{0.75}Ge_{0.25}$ MODFET (device 4). The maximum drain bias is 1 V. For devices 1 and 3 the applied gate voltage ranges from -1 V to $+0.5$ V. The applied gate voltage ranges from -1 V to $+0.6$ V for device 2 and 0 V to 0.5 V for device 4. The experimental result (dashed line) taken from [37] is shown for comparison for device 4 at the gate voltage of 0.4 V.

steps have been performed in order to derive the transconductance of each device. The transconductance as a function of gate voltage for device 1 is shown in Figure 4.10(a). The intrinsic drain current (I_D) is also plotted in the same figure. The results show a general pattern in which the transconductance increases with the gate voltage at low gate bias, but, as the gate bias is raised further, the transconductance levels off and then starts to decrease. For device 1, the calculated maximum intrinsic transconductance is about 200 mS mm^{-1} at a gate voltage of 0.3 V, which is comparable to the devices investigated in [37]. This moderate value of the transconductance is due to relatively large gate-to-channel spacing of 35 nm, and the source-channel access resistance.

In principle, ideal MODFET operation requires electron transport under the gate to occur only in the channel layer, with the supply donor layer remaining depleted. Otherwise, degradation in device current, transconductance, and switching speed is expected to occur, since all of the properties depend on the average electron velocity, which is considerably lower in the supply layer. The figures of merit for MODFETs therefore depend on the fraction of the electrons induced in the device by the gate voltage that are transported in the channel, and on the average electron velocities in the channel and the supply layer. The drain current is limited at gate voltages $V_{GS} \geq \Phi_B - 0.2$ V, by the appearance of a parasitic transport path in the SiGe supply layer and by gate leakage current. In particular, for a gate voltage exceeding 0.3 V for device 1 with drain bias 1 V, strong charge accumulation occurs in the supply layer.

The effect is illustrated by the superparticle distribution in device 1 shown (solid circle : Δ_2-electrons, open-square : Δ_4-electrons) in Figure 4.11. It is apparent from the Figure 4.11(a) that no parallel conduction channel forms in the upper supply layer under the gate in the case of negative gate voltage bias. However, parallel conduction does occur in the upper supply layer in the absence of gate bias. Increasing the gate voltage to 0.5 V causes a stronger parallel conduction channel.

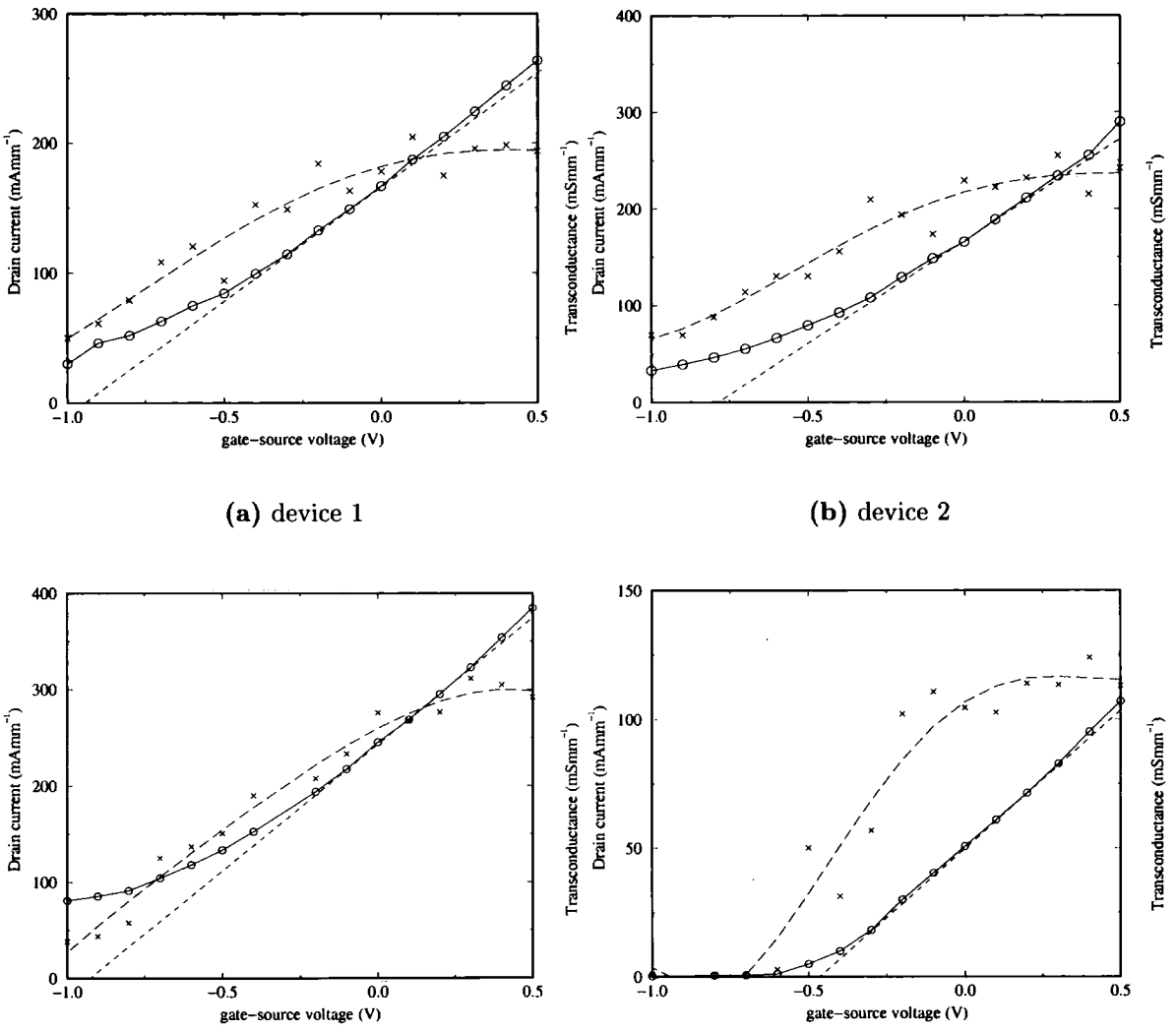


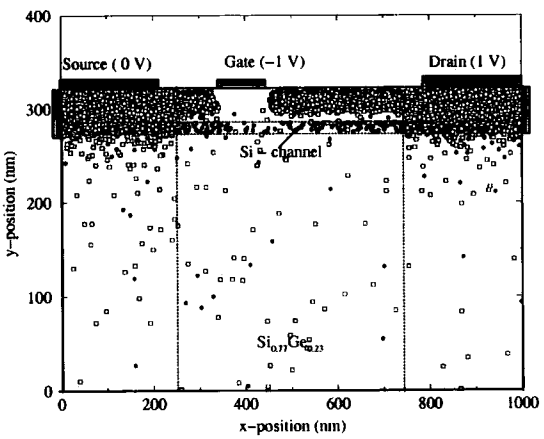
Figure 4.10: **(a)-(d)** The transfer characteristics of devices 1-4 for 1 V drain bias with the drain currents given by the left hand side vertical scale. The transconductance g_m(slope of the transfer characteristic) is also shown as a long dashed line for each device with g_m given by the right hand side vertical scale (same numerical values as left hand scale). The maximum intrinsic transconductances g_m of devices 1-4 are about 200, 230, 300, and 120 mS mm^{-1} at gate voltages of 0.3, 0.3, 0.3, and 0.4 V respectively. The threshold voltage (V_T) is determined by extrapolating the linear part of the transfer characteristic to zero as indicated by the dashed line for each device. V_T for devices 1-4 is about $-0.9, -0.75, -0.9$, and-0.45 V.

Normally we expect the transconductance of a MODFET to be enhanced as the channel length L_{ch} and/or gate length L_G is reduced or the drain-source bias is increased. However, the expected dependence is complicated for deep-sub-micrometer values of L_{ch} due to short channel effects and velocity overshoot. Other simulations, that we have carried out but not shown here, suggest that g_m increases from 200 to 250 mS mm^{-1} as the drain voltage increases from 1 V to 1.5 V. Not surprisingly, the transconductances for shorter channel devices at a drain bias of 1 V are higher than 200 mS mm^{-1}. Figure 4.9(b) shows that the maximum intrinsic transconductance derived from the transfer characteristic of device 2 is about 220 mS mm^{-1} at V_{GS} = 0.5 V while the corresponding value for device 3 (Figure 4.9(c)) is about 240 mS mm^{-1} (at V_{GS} = 0.3 V). Hence, decreasing the channel length by about 40-50% results in enhancement of the transconductance by about 20-25%. Note that the maximum intrinsic transconductance of device 4, with a much longer channel, is about 120 mS mm^{-1}. Experimentally, devices of this type exhibit a peak extrinsic transconductance of approximately 135 mS mm^{-1} [37].

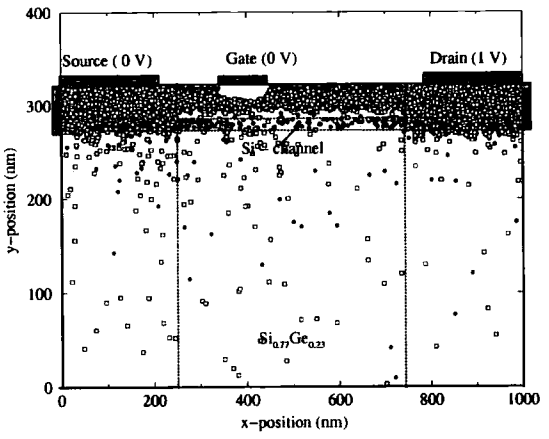
For RF applications, an important figure of merit for the speed of response of a transistor is the frequency at which the current gain in a common-source configuration becomes unity for short-circuit conditions at the output [33]. This is referred to as the unity current-gain frequency or cut-off frequency (f_T). In the simplest model, this can be determined from the transconductance and total gate capacitance C_g using the expression [33]:

$$f_T = \frac{g_m}{2\pi C_g} \tag{4.2}$$

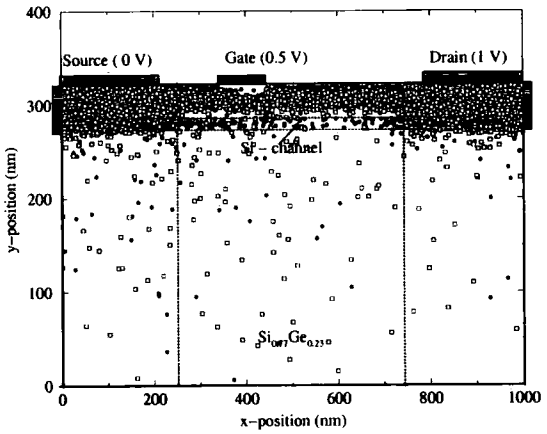
It follows that if C_g is taken as constant for all the simulated devices, we expect that device 3 will show the highest cut-off frequency while device 1 will show the lowest. More generally Equation 4.2 shows that the ratio of transconductance to gate capacitance provides a useful guide to the effect of geometry on high-frequency FET operation. The transconductance (g_m) can be easily obtained from the transfer characteristic curve as



(a) $V_{DS} = 1$ V, $V_{GS} = -1$ V



(b) $V_{DS} = 1$ V, $V_{GS} = 0$ V



(c) $V_{DS} = 1$ V, $V_{GS} = 0.5$ V

Figure 4.11: Superparticle distribution in device 1 (solid circle : Δ_2 electrons, open square : Δ_4 electrons) for various drain (V_{DS}) and gate (V_{GS}) voltages.

explained previously. The total gate capacitance (C_g) can be obtained directly from the Monte Carlo simulation since the charge throughout the device is available at each time step. However, this calculation will underestimate the capacitance of the real device because it does not take into account the parasitic capacitances [83]. We also calculate f_T in the absence of parasitic effects and refer to it as the intrinsic cut-off frequency f_T^i. An alternative method for obtaining f_T^i which we have adopted, is to calculate the current gain directly from the Monte Carlo simulations.

Another important parameter for FETs is the threshold voltage (V_T), which is the gate-source voltage necessary to just allow the passage of drain current. The threshold voltage can be determined directly from the linear part of drain current versus gate-source voltage curve as illustrated in Figure 4.10. The threshold voltages of devices 1 and 2 are about the same, at -0.8 V, whilst V_T is about -0.9 V for device 3 and -0.6 V for device 4.

2) Intrinsic Response of MODFET

We start the discussion of the time-dependent behaviour of the MODFETs by considering the response of device 3 to a step change in gate voltage. The initial condition is the steady state with $V_{DS} = 1$ V and $V_{GS} = 0$ V. At time $t = 0^+$s the gate bias is instantaneously switched to $V_{GS} = 0.5$ V. The response of the drain current I_D (solid line), gate current I_G (dotted line), and source current I_S (dashed line) is shown in Figure 4.12. The convention is that an inward-flowing terminal current is positive. The currents include both the particle and displacement components and Figure 4.12 shows the transient phase (phase I) and the subsequent steady state (phase II). The transient phase is characterised by the flow of substantial currents through all the terminals. During this phase the gate current decreases with time while the drain and source currents increase. In the steady state, the gate current is close to zero while the source and drain currents are equal in magnitude but opposite in sign. The sum of the

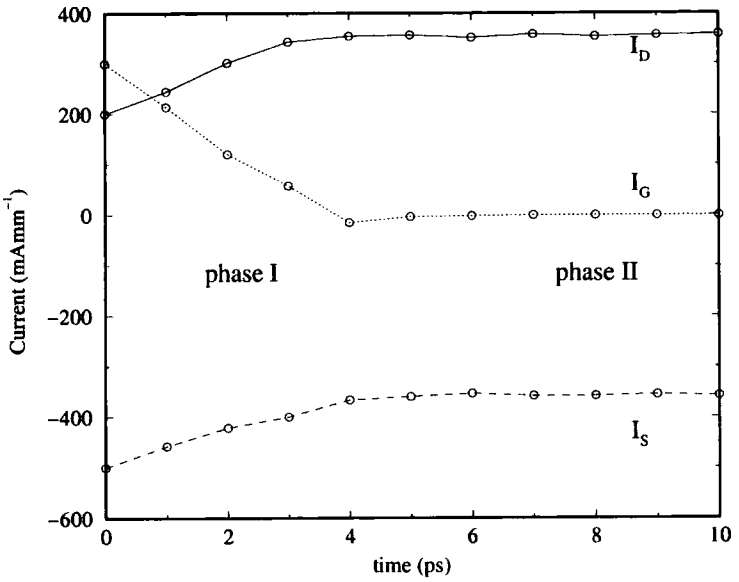


Figure 4.12: The source, drain and gate currents I_S, I_D, and I_G as a function of time for the gate step response of device 3. The drain-source bias is 1 V. The gate voltage is pulsed from 0 V to 0.5 V at time $t = 0$.

currents is zero to within statistical error at all times. The transient current through the gate is characteristic of the discharging of a capacitor in response to a change in applied voltage. Here the discharge process is accompanied by a reduction in the extent of the gate depletion region in the channel. This process influences the drain current and therefore it is to be expected that the current should have a response time that is similar or greater. An additional delay might be expected because the drain current response is carried by the electrons in the channel and it takes a finite amount of time for the electrons to reach the drain from the source but it is apparent from the figure that this effect does not add significantly to the response time.

3) Dynamic (AC) characteristics

The DC characteristics in Section 4.4.2 show that approximately $40 - 50\%$ improvement of transconductance (g_m) is obtained when the channel length is decreased from

0.5 μm (device 1) to 0.25 μm (device 3). Equation 4.2 suggests that an associated increase in device speed might be expected. However, instead of using Equation 4.2, we have performed simulations of device response to obtain a value for the intrinsic cut-off frequency f_T^i directly. Of particular interest is the effect of source-gate separation (L_{SG}), gate length (L_G), and gate-drain separation (L_{GD}) on the cut-off frequency. Note that these parameters also affect the parasitic capacitances of the devices, which are not included in our model but could have a significant influence on device performance.

The frequency response has been investigated by applying a series of truncated sinc voltage pulses to the gate contact with the device in the common source configuration as illustrated in Figure 4.13(a) [84]. The time dependence of the voltage signal is

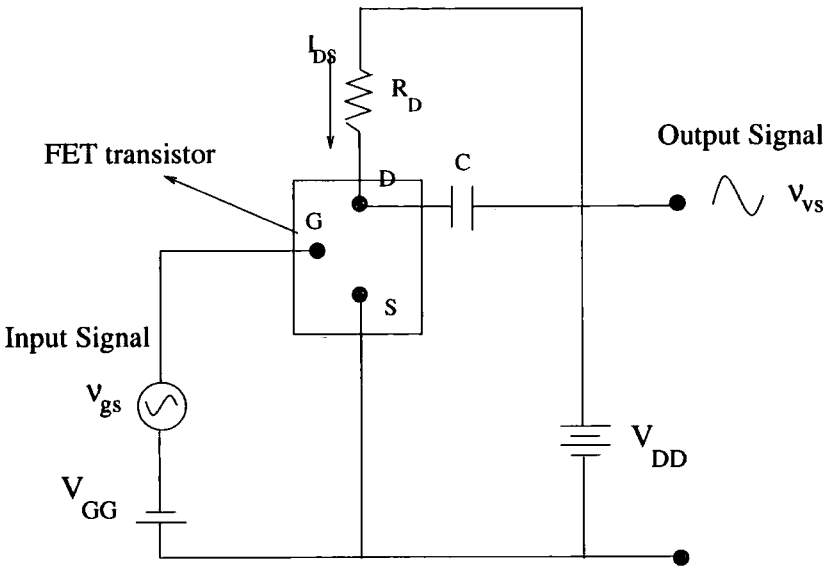
$$V_{gate}(t) = V_0 + V_{amplitude}\left(\frac{\sin\omega(t-T/2)}{\omega(t-T/2)}\right) \tag{4.3}$$

where V_0 is the steady state gate bias upon which the modulation signal at angular frequency ω is superimposed, $V_{amplitude}$ is the peak amplitude of the signal, T is its duration. A 100 ps duration sinc-form pulse is used which provides a flat frequency spectrum up to 100 GHz. The number of superparticles used in the simulations was increased to about 50,000 in order to suppress noise to an acceptable level.

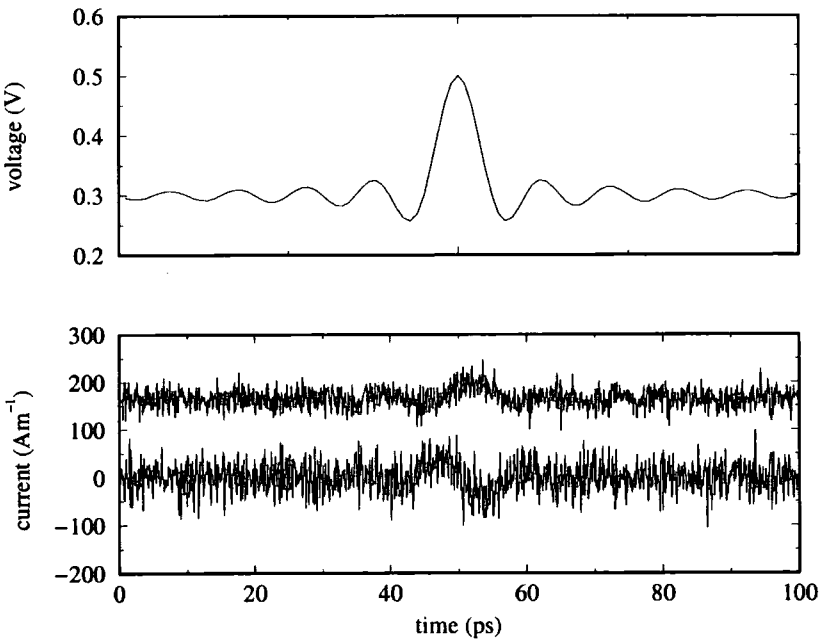
For device 1, a signal with $V_{amplitude} = 0.2$ V is superimposed on a steady gate bias $V_0 = 0.3$ V. The drain bias is 1 V. Figure 4.13(b) illustrates the voltage as a function of time applied to the gate, the output drain current response I_{drain} and the corresponding gate current I_{gate}. Note that the gate current is about $\pi/2$ out of phase with the drain current. The phase delay can be understood in terms of the equation for the total current flow through contact i [44],

$$I_i(t) = \frac{dQ_i(t)}{dt} + \epsilon\epsilon_o\mathbf{S}\cdot\frac{\partial\mathbf{F}}{\partial t} \tag{4.4}$$

where Q_i is the charge flow in the form of individual charged particles and the second



(a) Common source configuration for ac amplification



(b) Frequency response

Figure 4.13: **(a)** MODFET as a current amplifier. **(b)** The simulated frequency response of a $Si/Si_{0.77}Ge_{0.23}$ n-channel MODFET (device 1) to a sinc gate voltage pulse. From top to bottom, the gate voltage (V_{gate}) as a function of time for a truncated pulse of duration 100 ps, the drain current response (I_{drain}), and the gate current (I_{gate}).

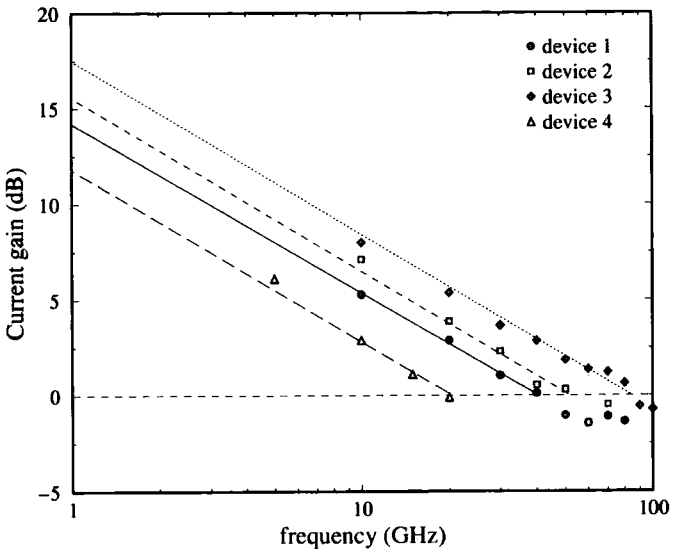


Figure 4.14: Logarithmic plot of the intrinsic current gain calculated for n-channel MODFETs with various gate lengths. The intrinsic current gain cut-off frequencies f_T^i for the four devices are 40 ± 10, 50 ± 10, 80 ± 10 GHz, and 20 ± 10 GHz, respectively.

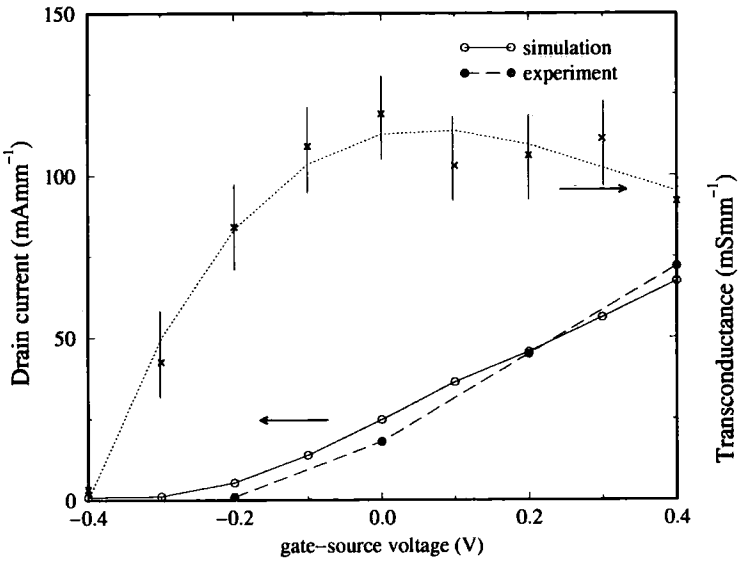
term on the right hand side represents the displacement current caused by temporal changes in the applied electric field **F** at the contact (whose area is represented by the vector **S**). The gate current is mainly displacement current since the gate is a Schottky barrier contact. Some displacement current also flows in the source and drain but the magnitude is much smaller than for the gate.

Similar frequency responses have also been obtained for the other devices. However, a 200 ps duration sinc-form pulse, providing a flat frequency spectrum up to 50 GHz in the frequency domain is used for device 4 because of its inferior high frequency performance. The current gain has been derived as a function of frequency by taking fast Fourier transforms of the simulated drain and gate current signals, and then calculating the ratio of the drain and gate current transforms.

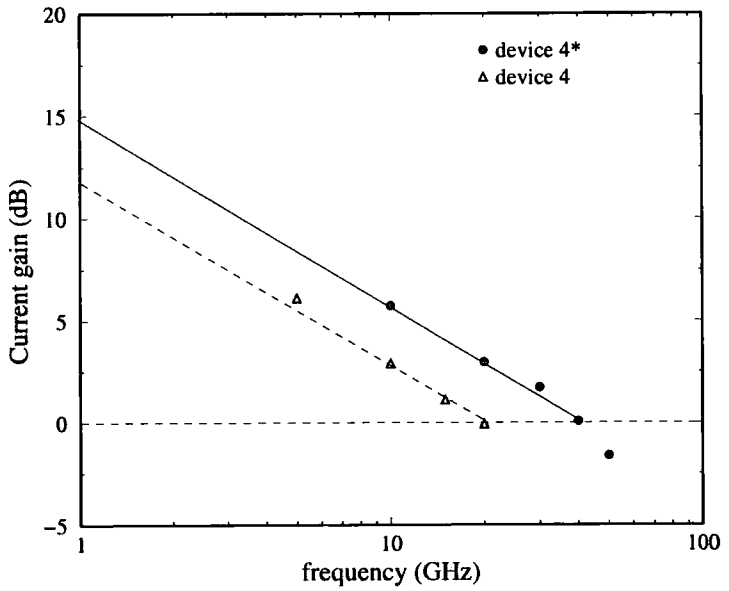
The unity current gain cut-off frequency is expected to increase as the channel length is reduced since the electron transit time is also reduced. Therefore it is expected

that f_T will increase as the channel length is reduced from the 0.5 μm of device 1 to the 0.3 μm of device 2 and the 0.25 μm of device 3. This is verified in Figure 4.14. The intrinsic current gain cut-off frequency reaches 80 $\pm$ 10 GHz for the smallest device simulated here, which is device 3 (L_{CH} = 0.25 μm, L_G = 0.07 μm). In contrast device 4 (L_{CH} = 0.5 μm, L_G = 0.2 μm) has a cut off-frequency of only 20 $\pm$ 10 GHz, which is about half that found in device 1(L_{CH} = 0.5 μm, L_G = 0.09 μm). Apart from the longer gate length (L_G = 0.2 μm) and centred gate configuration employed in device 4, the reason for the lower performance relative to device 1 is the lower doping in the supply layer and in the contact implants. The device investigated experimentally by the IBM group has a cut-off frequency of 62 GHz at the drain voltage of 1 V [37]. By increasing the doping density of source and drain implants to 10^{25} m^{-3}, to enhance electron injection, and at the same time decreasing the doping of the supply layer from 4×10^{24} m^{-3} to 3×10^{24} m^{-3}, to keep good agreement between the simulated and experimental output characteristics, it is possible to raise f_T^i to 40 $\pm$ 10 GHz, as illustrated in Figure 4.15(b). We will refer to this device as device 4*. Figure 4.15(a) shows that the highest g_m of device 4* is about 120 mS mm^{-1} at zero gate voltage. Note that the g_m of device 4 reaches its peak value of about 150 mS mm^{-1} at a gate voltage of 0.4 V (Figure 4.10(d)). The experimental device investigated in [37] has a g_m which peaks at about 135 mS mm^{-1} at a drain bias of 1 V and zero gate voltage. Figure 4.15(a) shows that the simulated drain current agrees well with the experimental results denoted by the dashed-line.

The reduction of current gain at higher frequency can be explained by inspecting the small-signal equivalent circuit illustrated in Figure 4.16. Since the output is taken from the drain while the input is applied to the gate, the gate-to-drain capacitance C_{DG} can be considered as a feedback capacitor from the output back to the input. The output is then out of phase with respect to the input signal because of the feedback signal through the gate-to-drain capacitance. As the frequency of the input signal increases,



(a) Transfer characteristics



(b) Cut-off frequency

Figure 4.15: **(a)** The transfer characteristics for 1 V drain bias of device 4*. The maximum intrinsic transconductance g_m (dotted-line) is about 120 mS mm^{-1} at zero gate voltage. Note the error bars on the simulated g_m values take account of the fact that noise on the simulated drain current is at least ±10 mA mm^{-1}, when 30,000 electron particles are used for the simulation. Experimentally the g_m of the device investigated in [37] is about 135 mS mm^{-1} at the same bias arrangement. **(b)** Logarithmic plot of the maximum intrinsic current gain calculated for device 4 and 4*. The intrinsic current gain cut-off frequencies f_T^i for devices 4 and 4* are 20 ± 10 and 40 ± 10 GHz, respectively.

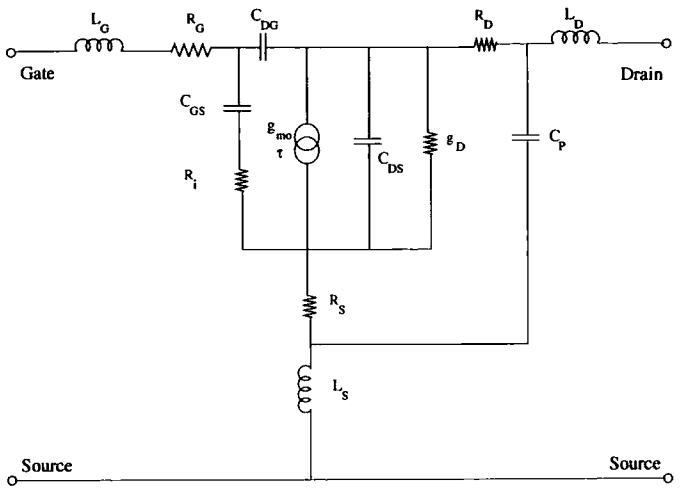


Figure 4.16: Equivalent circuit model of MODFET.

the effect of the negative feedback capacitance also increases and the observed current gain begins to decrease rapidly.

Another figure of merit in characterising transistors for RF application is f_{max} (maximum frequency of oscillation). The maximum frequency of oscillation is the maximum frequency at which the device can still be used as an oscillator [33]. Actually f_{max}, which is a function of f_T, indicates the ability of a given transistor to deliver the intrinsic performance to an external circuit and is the ultimate figure of merit for system performance. The determination of f_{max} is not only dependent on the intrinsic device but also on the external circuit. This requires the inclusion of the parasitic-resistance and capacitance effects into the Monte Carlo model, and is not considered here.

4) Microscopic Behaviour

To understand the device performance of deep submicron devices, it is necessary to obtain a good understanding of the underlying physics that governs their operation. A self-consistent ensemble Monte Carlo device simulation is well suited for this purpose since at every time step the coordinates and electronic states of all the electrons are

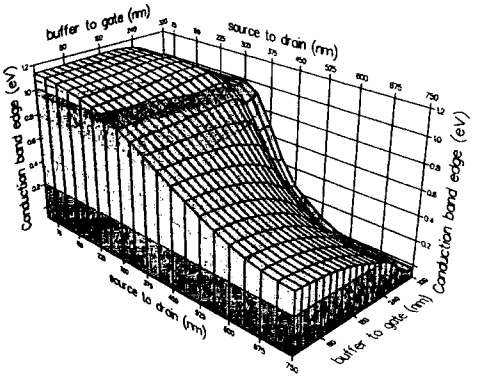


Figure 4.17: The conduction band edge (Δ_2) of device 3 at 1 V drain bias and 0.3 V gate voltage.

available together with the distribution of electrostatic potential or field throughout the device.

Figure 4.17 shows the energy of the Δ_2 conduction band edge for device 3 as a function of position at a drain-source bias of 1 V and a gate voltage of 0.3 V. The device has its peak transconductance under these conditions. On the basis of the figure we might expect that Δ_2 electrons injected at the source are heated along the channel under the gate due to the longitudinal electric field, and some of them will reach the threshold kinetic energy for non-polar (intervalley) phonon scattering into the Δ_4 satellite valleys.

The conduction band edge energy is plotted as a function of position along the channel for devices 1-4 in Figure 4.18(a). The conduction band in the source and drain regions is quite flat compared with the channel region. The region of high longitudinal field is more extensive in devices 2-3. As a consequence, the average drift velocities of electrons along the channels of devices 2-3 are expected to be higher, provided the electron velocity does not saturate. Figure 4.18(b) shows how the conduction band edge varies through the thickness of the channel from the buffer to the gate of device 3,

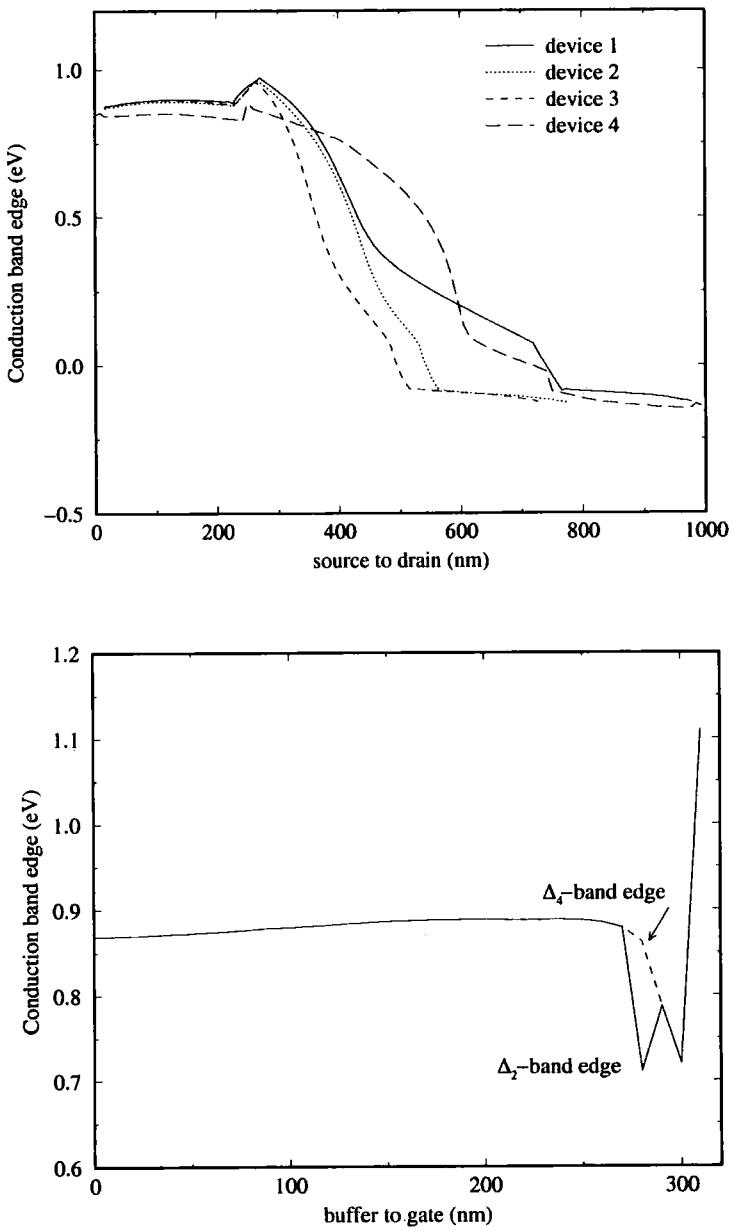
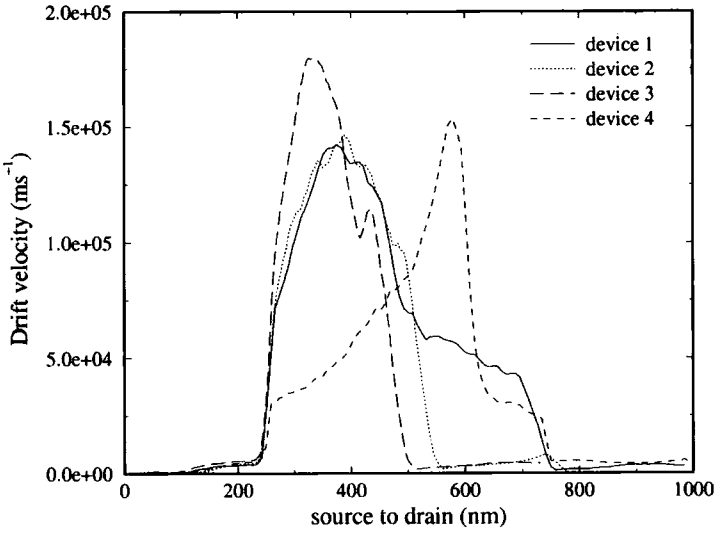


Figure 4.18: **(a)** Conduction band edge as a function of position along the channel for devices 1-4. The drain bias is 1 V. The gate voltages are 0.3, 0.5, 0.3, and 0.4 V for devices 1-4, respectively. **(b)** Conduction band edge through the device 3 under the gate. The bias arrangement is the same as in **(a)**. The dashed line corresponds to the Δ_4 valleys conduction band edge.

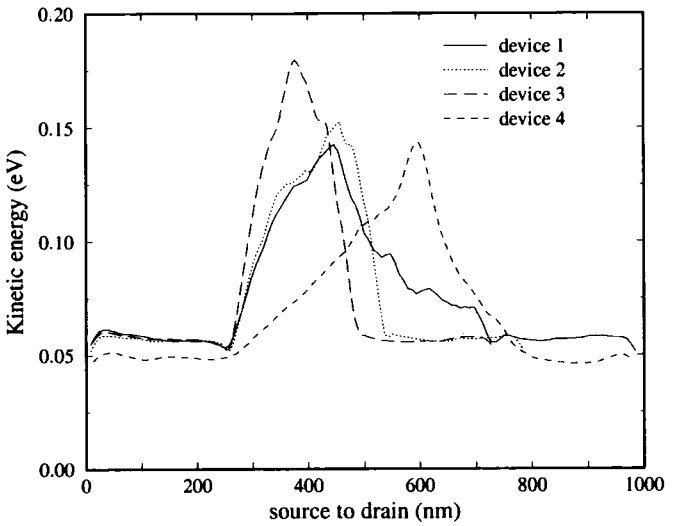
which is the smallest device simulated in this study. Immediately, under the Schottky gate, there is the expected conduction band barrier. The channel with a barrier height of 0.15 eV is also apparent in the figure.

In order to understand the carrier dynamics in the devices, it is instructive to obtain the carrier energy, velocity and population profiles as a function of distance along the device. Longitudinal Δ_2 electron velocities along the channel of the four devices for $V_{DS} = 1.0$ V, with V_G chosen to provide maximum transconductance for each device, are shown in Figure 4.19(a). The average electron velocity increases steeply near the source edge of the gate and then increases more gradually thereafter. In devices 2 and 3 the maximum velocity v_m of Δ_2 electrons is significantly higher than the saturation velocity v_{sat} (which is about 1×10^5 ms^{-1}) whilst v_m is close to the saturation value for device 1. The average electron velocity in device 4 is also depicted with the maximum velocity of Δ_2 electrons predicted to be 1.5×10^5 ms^{-1}. However, the larger part of the channel of device 4 has a much lower electron velocity, resulting in a lower velocity averaged along the channel than in the other devices, and hence a lower cut-off frequency f_T^i of 20 $\pm$ 10 GHz.

It was pointed out in Section 4.2 that the average electron kinetic energy increases monotonically with increasing electric field in bulk material. We also reported an enhancement of the kinetic energy of electrons in strained Si. However, in a MODFET the electric field is a function of position, which complicates the interpretation of the simulated results. Figure 4.19(b) shows the profile of the corresponding average electron energy as a function of position along the channel of the devices. The bias conditions are the same as for Figure 4.19(a). As the electrons flow from the source to the drain they gain kinetic energy as they are accelerated by the electric field. Recalling that the strain induced energy splitting of the Δ valleys for 23-25% Ge fraction is about 0.15 eV, electrons in the Δ valleys whose energies are higher than the splitting can scatter from Δ_2 to Δ_4 valleys via non-polar optical phonon intervalley scattering. The



(a) Average velocity of Δ_2 electrons



(b) Average kinetic energy of Δ_2 and Δ_4 electrons

Figure 4.19: (a) Average longitudinal velocity of Δ_2 valleys electrons along the channel of devices 1-4 for the bias which gives the maximum transconductance for each device at (V_{GS} = 0.3, 0.3, 0.3, and 0.4 V for devices 1-4. and V_{DS} = 1 V). **(b)** Average kinetic energy of $\Delta_2 + \Delta_4$ valleys electrons at the same bias arrangements as in **(a)**.

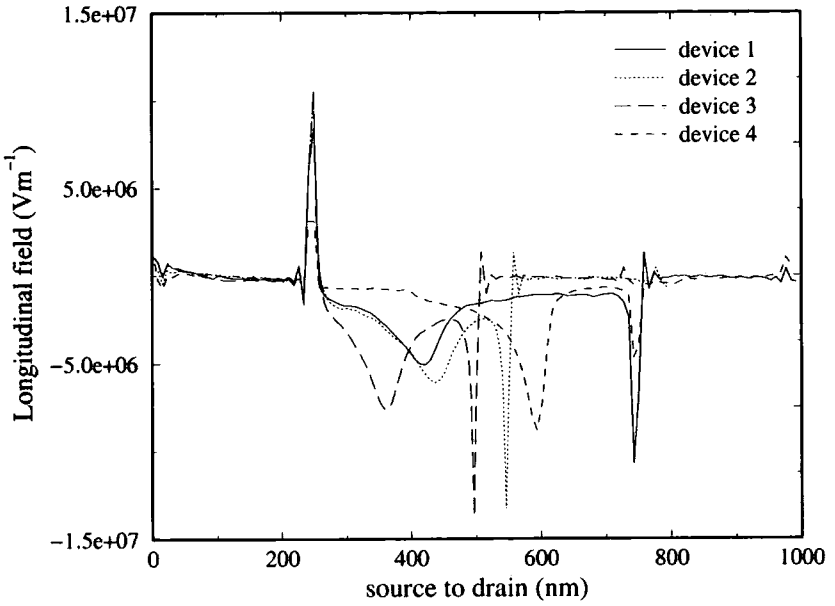


Figure 4.20: Longitudinal electric field recorded along the channel of devices 1-4. The drain-source bias is 1 V. The gate voltage is selected to provide an intrinsic maximum transconductance for each device.

intervalley transfer results in a reduction of the average kinetic energy of electrons.

To consider the electron heating in more detail it is instructive to look at the longitudinal electric field along the channel of each simulated device as shown in Figure 4.20. There are field spikes at the source region-channel and drain region-channel interfaces with the lower doping of the source and drain implants in device 4 resulting in smaller spikes at the interfaces than for the other devices. In the simplest 1-dimensional model, the sequence of source-channel-drain in the MODFET can be thought of as an n^+-i-n^+ structure. In such a structure, the out-diffusion of electrons from the highly doped n^+ regions creates space-charge regions between the high-low junction (n^+-i interface) and the low-high junction (i-n^+ interface). This space-charge layer produces an electric field, which causes a drift current opposite to the electron diffusion. The field distribution reach its maximum value at the position where the space charge changes sign. For an n^+-i doping sequence the field spike is positive while the reverse sequence produces

a negative spike.

A similar strong spill-over of carriers from the n^+ to the n-region is also reported by Starikov *et al.* [85] in a Si structure. In [85], hydrodynamic and Monte Carlo simulations of n^+-n-n^+ Si structure with doping levels $n = 2\times10^{21}$ m^{-3} and $n^+ = 5\times10^{23}$ m^{-3} was investigated. The electric field spike at the n^+-n homojunction was found to have a peak value around 4×10^6 Vm^{-1} when a bias of 1.5 V was applied. The hydrodynamic calculation of an n-i-n structure, with an i-region thickness of 0.4μm reported by Ramaswamy *et al.* [86] also shows similar features with an electric field spike around 3×10^6 Vm^{-1}. Apart from the spikes the field strength in the source and the drain regions is very small but it is much larger around the gate where it reaches between about 5 and 9 MVm^{-1} depending on the particular device.

The electron energy is related to the field profile and therefore the fractional occupancies of the Δ_2 and Δ_4 valleys are functions of position in the devices. Figure 4.21 shows the fractional occupation of electrons in the Δ_2 and Δ_4 valleys for the four simulated devices. In the source and the drain regions of devices 1-3, about 80% of electrons occupy the Δ_2 valleys whereas the relevant value for device 4 is about 90%. In all the devices, it is only in the channel region that the valley populations change significantly from these values. For each device, over a large part of the channel most electrons remain in the Δ_2 valleys, to the benefit of the average electron velocity. However, electrons are strongly heated near the gate due to the high local electric field, with a resultant fall in the Δ_2 electron population. In device 3, the Δ_2 valleys occupation decreases from about 95% in source region to about 65% on the drain side of the gate.

It was shown in Section 4.2.2 that a significant velocity overshoot occurs in strained Si grown on a $Si_{0.77}Ge_{0.23}$ virtual substrate, reaching a peak value of 1.5×10^5 ms^{-1} for a stepped application of an electric field of 6×10^6 Vm^{-1}. Hence, it might be expected that the overshoot velocity effect would be found in our simulated devices, and especially in devices 2-4. Figure 4.22(a)-(d) depicts the average drift velocity of electrons

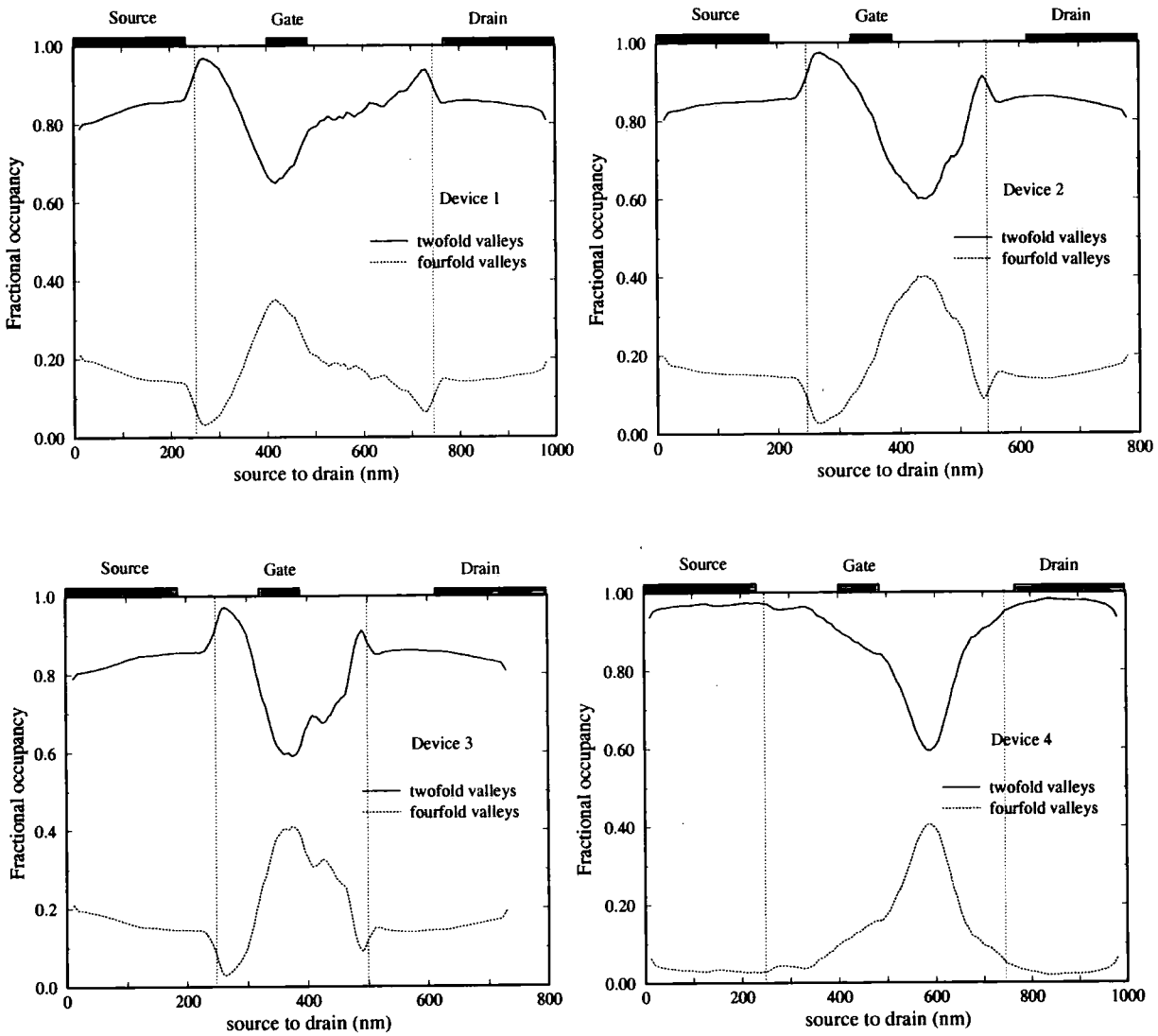
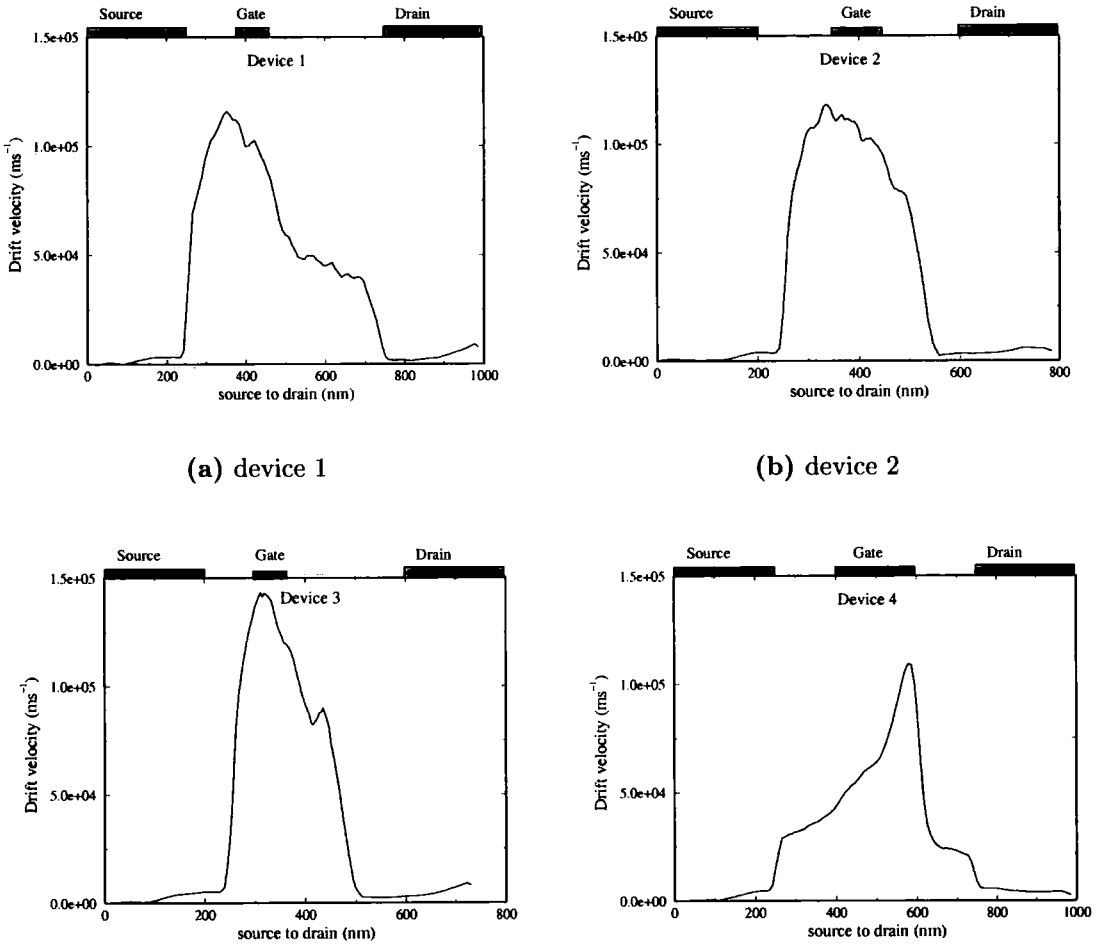


Figure 4.21: Fractional occupation of the Δ_2 and Δ_4 valleys recorded along the channel of device 1-4. The drain bias is 1 V. The gate voltage applied for each device provides the maximum intrinsic transconductance.



(a) device 1

(b) device 2

(c) device 3

(d) device 4

Figure 4.22: (a)-(d) Average longitudinal velocity of electrons ($\Delta_2 + \Delta_4$) and electron density along the channel of devices 1-4 for the bias which gives the maximum transconductance for each device at ($V_{DS} = 1$ V and $V_{GS} = 0.3$, 0.3, 0.3, and 0.4 V for device 1, 2, 3, and 4 respectively).

recorded along the channel for each device. The drain bias is 1 V and the gate-source voltages are 0.3, 0.4, 0.3, and 0.4 V for devices 1, 2, 3, and 4 respectively. However, Figure 4.22(a) shows that the average drift velocity is below the saturation velocity of electrons in bulk Si ($\sim 10^5$ ms^{-1}) in device 1 while it is similar to the saturation velocity in devices 2 and 4. On the other hand, the overshoot velocity effect is apparent in the shortest device (device 3), which exhibits a peak velocity of about 1.4×10^5 ms^{-1}. This implies that after injection the electrons are accelerated in a high field region that is shorter than the distance required for the distribution function to equilibrate to the steady state. The electrons encounter a step in the electric field which is a situation similar to the case where the electric field is just switched on and the temporal scale of the velocity response found in Subsection 4.2.2 translates into a distance scale here. Note that in devices 1 and 3, the highest field is in the source-gate region while it is in the gate-drain region for devices 2 and 4.

Since the average drift velocity along the channel is known from the simulations, it is possible to estimate the transit time (t_{tr}) for electrons to pass through any given part of the channel. The transit time can be calculated by the simple formula:

$$t_{tr} = \int \frac{dx}{< v_x(x) >} \tag{4.5}$$

where $< v_x(x) >$ is the average velocity of the electrons at position x, and the integral has been evaluated for the region of interest. The average velocity as a function of position along the channel of each simulated device shown in Figure 4.22(a)-(d) can be used to calculate the integral. If the section of channel of interest is taken to be that immediately under the gate of each device, the calculated transit times for devices 1-4 are 1.46, 1.04, 0.88, and 2.86 ps, respectively. An estimate of the cut off frequency can be obtained using the formula [33]

$$f_T = \frac{1}{2\pi t_{tr}} \tag{4.6}$$

Equations 4.5 and 4.6 suggest that a shorter gate will result in a higher f_T (even if the electron velocity is not increased). However, as the gate length is reduced, one must also proportionally reduce the vertical dimensions of the FET in order to maintain gate control of electron transport in the channel. The associated device parameter is the aspect ratio, which is the ratio of the gate length (L_G) to the gate-to-channel separation (h_{GC}). Note from table 4.1 that device 4 has the highest aspect ratio whilst device 3 has the lowest value. Reducing the gate-to-channel separation generally increases the transconductance and Equation 4.2 suggests this will also have a beneficial effect on f_T.

Using the calculated transit times, the f_T of devices 1-4 are estimated from Equation 4.6 to be 109, 152, 181, and 56 GHz, respectively. If the average electron drift velocity under the gate in each device were the saturation velocity for bulk Si, the value of f_T would be about 177, 159, 227, and 80 GHz, respectively. On the other hand, if the region of interest were the channel length for each device, the value of f_T would be about 16, 44, 56, and 16 GHz, respectively. The intrinsic cut-off frequencies f_T^i of devices 1-4 derived from the sinc pulse response are 40 ± 10, 50 ± 10, 80 ± 10, and 20 ± 10 GHz respectively.

The sinc pulse response suggests that device 4* has a cut-off frequency of 40 ± 10 GHz. To understand the better performance of this device compared to device 4, it is necessary to look at the microscopic details of the electron transport in the channel. Figure 4.23 shows the average kinetic energy and drift velocity of electrons in the channel of devices 4 and 4*. It is apparent from Figure 4.23(a) that there is an enhancement of average kinetic energy and drift velocity of the electrons in device 4* compared to those in device 4. The drift velocity of the electrons reaches 2×10^5 ms^{-1} in device 4*, as illustrated in Figure 4.23(b), while the peak velocity is about 1.2×10^5 ms^{-1} in device 4. For device 4* the corresponding transit time of electrons traversing the gate is 2.36 ps, providing an estimate of f_T of 67 GHz. The device similar to 4 and 4* that was

investigated by the IBM group [37] has $f_T = 62$ GHz for the the same bias. For device 3(Figure 4.22) the drift velocity peaks at about 1.43×10^5 ms^{-1}, which is a lower value than for device 4*, but nevertheless it has a higher f_T. In this connection it should be noted that a large part of the channel in device 4* corresponds to a lower electron velocity compared to that in device 3, resulting in a lower drift velocity averaged over the channel.

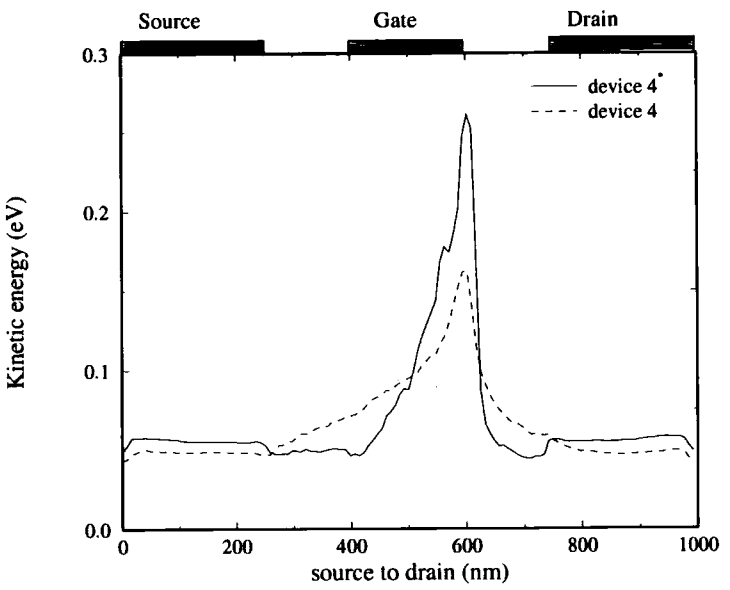
It is worthwhile discussing the expected accuracy of f_T values calculated using either Equations 4.2 or 4.6, or by considering the current gain of the device in the frequency domain. The intrinsic limits on the high frequency response of a MODFET can be considered to be determined by various capacitance charging times and carrier transit times. It is common to describe the static and high-frequency response characteristics of semiconductor devices by using an equivalent circuit such as that shown in Figure 4.16, where the electrical behaviour of the device is represented by a circuit consisting of lumped two-terminal elements with ideal resistance, capacitance, or inductance. From Figure 4.16, the current generator in the output is $g_m V_G$. The cut-off frequency f_T is defined as the frequency at which current through C_{GS} is equal to the current of the generator, giving Equation 4.2. While g_m can be extracted from current-voltage characteristics, in principle, the capacitances of the intrinsic transistor can be calculated by recording the charge stored in each region throughout the device at each timestep. Note that a device model that does not include parasitic capacitances will tend to underestimate C_G and overestimate f_T.

An alternative formula for the cut-off frequency is

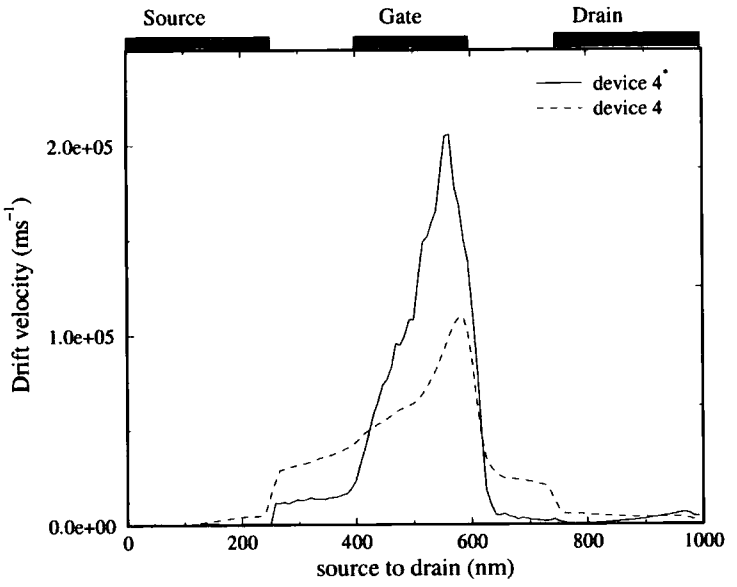
$$f_T = \frac{1}{2\pi t_T} \tag{4.7}$$

where t_T is the intrinsic total delay given by

$$t_T = t_{channel} + t_{transit} + t_{drain} \tag{4.8}$$



(a) Average kinetic energy of Δ_2 and $\Delta_4$electrons



(b) Average drift velocity of Δ_2 and $\Delta_4$electrons

Figure 4.23: **(a)** Average kinetic energy of $\Delta_2+\Delta_4$ valley electrons along the channel of devices 4 and 4* for the bias which gives the maximum transconductance for each device (V_{DS} = 1 V, V_{GS} = 0.4 and 0 V for devices 4 and 4* respectively). **(b)** Average drift velocity of $\Delta_2 + \Delta_4$ valley electrons at the same bias arrangement as **(a)**.

and $t_{channel}$ is the channel charging delay, $t_{transit}$ the transit delay and t_{drain} the drain delay. $t_{channel}$ is associated with RC delays and is proportional to channel resistance. t_{drain} is the time required by the electron to traverse the depletion region between the gate and the drain. If the transit delay is the dominant delay time and is identified with the carrier transit time under the gate, the cut-off frequency f_T can be estimated from Equation 4.6. f_T calculated using Equation 4.6 will be an upper estimate because the approximations described above underestimate the delay time t_T. The transit time for the entire channel and not just under the gate would lower the value of f_T but it is no longer possible to claim that it produces a bound on the actual value.

In conclusion, although the transit time in any region of interest can be obtained accurately within statistical error from Monte Carlo simulation, it is expected that f_T estimated via Equation 4.6 will be less accurate than that obtained via Equation 4.2. However, the equivalent circuit that is used to derive Equation 4.2 is only a simplified model of carrier response in the idealised device of the simulation model and therefore Equation 4.2 cannot be regarded as entirely reliable.

This suggests that a more straightforward way of obtaining an accurate picture of the frequency response is by applying time-dependent voltage signal to the gate in the Monte Carlo simulations. The cut-off frequency f_T is then determined directly from the simulation by transforming the current gain into the frequency domain. In this scheme both capacitive and transit time factors are naturally taken into account.

Although a direct comparison of f_T between our simulated devices and other devices investigated experimentally or theoretically is often difficult due to the wide range of device parameters, geometries, transport models etc., and the uncertainties in experimental structures, the results of our calculations can be compared with the Monte Carlo simulations performed by Dollfus [70,71]. In [70] a maximum f_T value of 55(66) GHz is predicted for a 0.18 μm gate $Si_{1-x}Ge_x$ MODFET with $x = 0.2$ (0.3), $L_{CH} = 0.38$ μm and a $Si_{1-x}Ge_x$ supply layer doped to $N_D = 3 \times 10^{24}$ m^{-3}. In comparison device

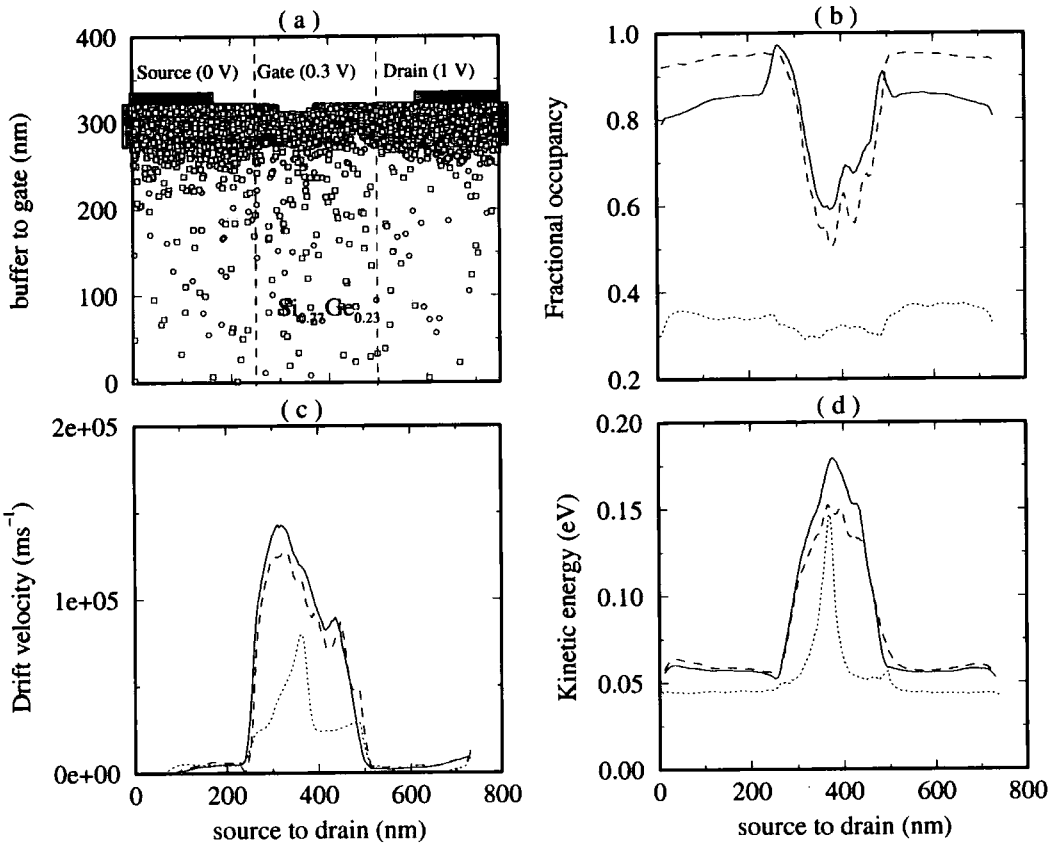


Figure 4.24: **(a)** Superparticle distribution (solid circle : Δ_2 electrons, opened square : Δ_4 electrons) at a drain bias of 1 V of device 3. The gate voltage is 0.3 V. This bias arrangement provides the peak transconductance, **(b)** Δ_2-valley population recorded along the channel, in the Si-channel, donor supply layer, and the top of buffer, **(c)** The corresponding average velocity, **(d)** The corresponding kinetic energy. Solid line : channel, dotted line : supply layer, and dashed-line : sub-channel.

2 ($L_G = 0.1$ μm, $L_{CH} = 0.3$ μm) is found to have an f_T^i of 50 $\pm$ 10 GHz using the signal response method. Also, device 3 ($L_G = 0.07$ μm, $L_{CH} = 0.25$ μm) has an f_T^i of 80 $\pm$ 10 GHz which is about half the 135 GHz of the device with $L_G = 0.08$ μm, $L_{CH} = 0.28$ μm calculated by Dollfus [71]. f_T in [70,71] is calculated by the use of Equation 4.2 and at least some of the discrepancy with the result obtained here can be explained by our use of the signal response method, which is expected to be more reliable.

Since device 3 provides the highest predicted cut-off frequency f_T^i we look in more

detail at its properties. Figure 4.24 shows electron distribution, fractional occupation of Δ_2 electrons, average drift velocity, and kinetic energy at various positions parallel to the channel, namely in the channel itself, in the supply layer and in the sub-channel at 1 V drain bias and 0.3 V gate-source voltage. Note that the sub-channel layer lies 270 nm from the bottom of the $Si_{0.77}Ge_{0.23}$ buffer. The advantages of confining the electrons in the channel well are manifest in view of the much reduced velocity in the proximity of the donors in the supply layers. However, inspection of Figure 4.24(a) reveals that there is a parallel conduction path in the donor $Si_{0.77}Ge_{0.23}$ supply layer because electrons in the channel gain enough kinetic energy to escape. The presence of electrons in the supply layer has a parasitic effect, since the gate modulates the charge in the layer but has a much reduced effect on the drain current. The valley population in the Δ_2 valleys is illustrated in Figure 4.24(b). Since the density of states of the Δ_4 electrons is double that of the Δ_2 electrons, it can be seen that about 67% of electrons occupy the Δ_4 valleys in the donor layer.

4.5 Results for high tensilely strained Si n-channel MODFETs

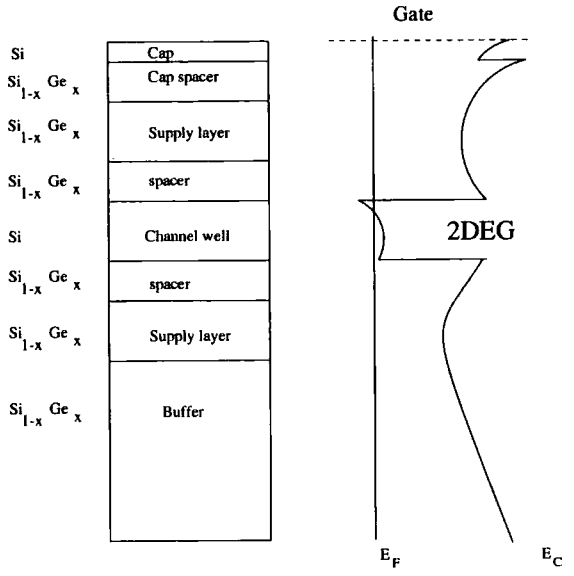
4.5.1 Layer design

The calculations performed in Section 4.2 (Figure 4.2) for strained bulk Si suggest that for a Si layer on a 0.45 Ge content virtual substrate, the majority of carriers occupy the Δ_2 valleys for fields up to at least 2×10^7 Vm^{-1}. It was also found that the electron drift velocity can reach 2.5×10^5 ms^{-1} in the transient response to a stepped electric field of 2×10^7 Vm^{-1}. These observations imply that small MODFETs with high Ge fraction and internal electric fields comparable to 10 MVm^{-1} should perform well. Experiments

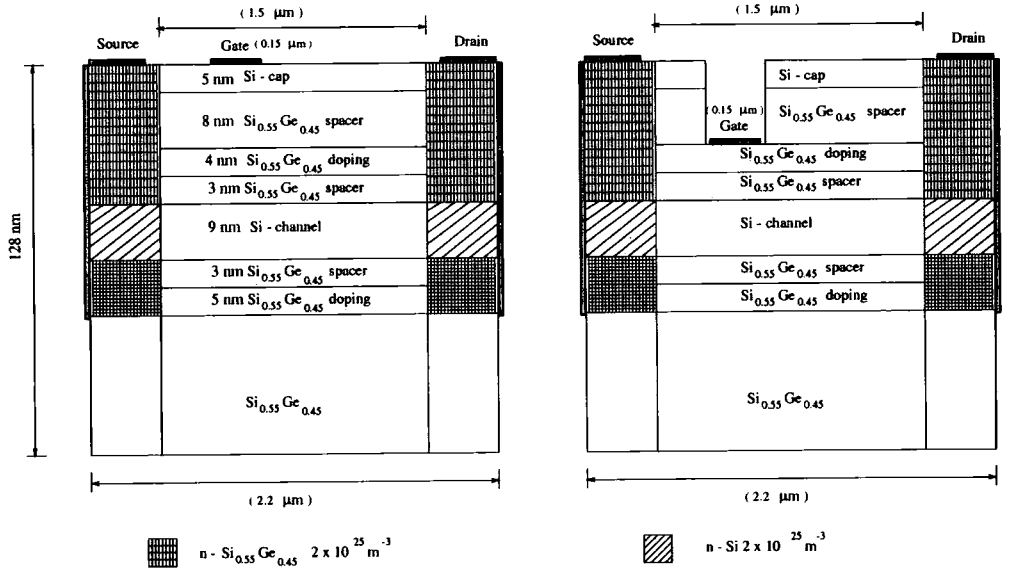
on n-channel SiGe MODFETs with 45% Ge have been recently reported by Glück *et al.* [73] of the Daimler-Chrysler group. In that work, n-channel MODFETs have been demonstrated with recessed and surface Schottky gates, operating as enhancement and depletion mode devices respectively. High transconductances of 476 (205) mS mm^{-1} and the extrinsic cut-off frequencies of 30 (43) GHz have been obtained for enhancement (depletion) mode devices.

We have simulated two structures similar to those of Glück *et al.* which are shown in Figure 4.25. In principle, enhancement mode MODFETs can be achieved either by reducing the doping concentration in the supply layer [81,87] or by a recessed gate configuration [73]. The enhancement mode device in [73] is derived from a device that would otherwise work in the depletion mode by recessing the gate. Accordingly, our simulated MODFET geometries are identical except that the gate of the enhancement mode device is recessed.

The layer design and simulated structures for the two devices are illustrated in Figure 4.25. A 9 nm tensile strained Si channel is sandwiched between relaxed $Si_{0.55}Ge_{0.45}$ layers. Electrons are supplied by doping in the layer below and the layer above the channel. The idea is that a careful choice of doping in the lower layer can reduce charge storage and parasitic channel formation [88]. The upper layer doping is taken to be 1.5×10^{25} m^{-3} while doping of 4×10^{24} m^{-3} is used for the lower supply layer. A 3 nm thick undoped $Si_{0.55}Ge_{0.45}$ layer separates each supply layer and the channel. The source and drain implants are taken to be doped to 2×10^{25} m^{-3} while the background doping is 1.0×10^{21} m^{-3} in the $Si_{0.55}Ge_{0.45}$ virtual substrate. The heterostructure is capped with a 5 nm thick Si layer. A Schottky barrier height of 0.8 eV has been used to represent the contact potential at the Au/Pt electrodes. The channel length is 1.5 μm, the source-gate separation (d_{SG}) is 0.2 μm and the gate length is 0.15 μm for both devices. The gate-to-channel separations (d_{GC}) are 20 and 12 nm for the depletion and enhancement mode devices respectively.



(a) Layer design



(b) depletion mode

(c) enhancement mode

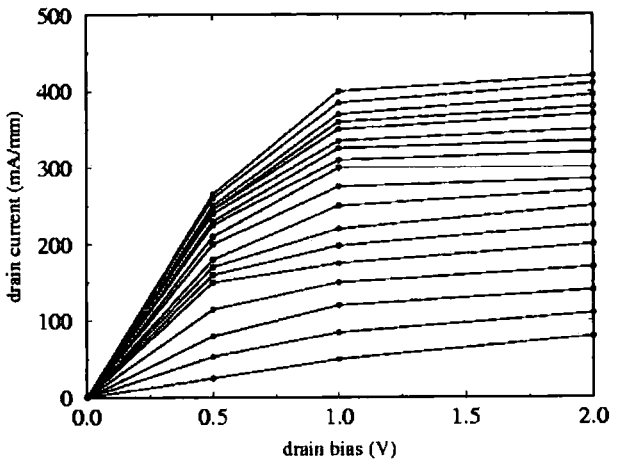
Figure 4.25: **(a)** Layer design of Si/SiGe n-MODFETs with high tensile strain and corresponding conduction band profile in equilibrium. **(b)** The simulated device structure. A 9 nm tensile strained Si quantum well is grown on a relaxed $Si_{0.55}Ge_{0.45}$ virtual substrates, and is modulation doped via both the substrate side (5 nm, 2×10^{24} m^{-3} Sb) and top side (4 nm, 1.5×10^{25} m^{-3} Sb). The confined electron gas is separated from the supply layer by a 3 nm spacer (background doping density 10^{21} m^{-3}). Source and drain implants are taken to be 2×10^{25} m^{-3}. The gate is deposited on the top of a Si-cap layer (5 nm) and a Schottky barrier height of 0.8 eV has been used to represent the contact potential at the gate electrode. The device on the left works in depletion mode, whilst the enhancement mode device on the right is derived from it by recessing the gate.

4.5.2 Device Characteristics and Analysis

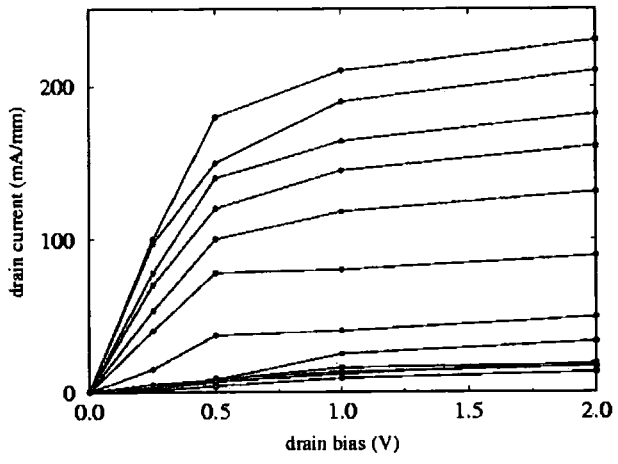
1) Static (DC) characteristics

The simulations reported in this section were done in collaboration with Dr G.C. Crow and have been published with him [89]. The simulations of steady current characteristics were carried out using 30,000 superparticles for the two device structures described in Section 4.5.1. Figure 4.26(a) and 4.26(b) show the simulated output characteristics of the depletion and enhancement mode devices. For the depletion mode device, the drain current is calculated to reach 420 ± 30 mA mm^{-1} at 2 V drain voltage and +0.6 V gate bias, with a knee voltage of about 0.8 V. The experimental drain current is rather lower at 320 ± 30 mA mm^{-1} for the same bias conditions as shown in Figure 4.27(a) [73]. For the enhancement mode device, the calculated drain current is 220 ± 30 mA mm^{-1} for the same bias, and is close to the experimental result at 2 V drain and +0.7 V gate bias as shown in Figure 4.27(b) [73]. The knee voltage is at about 0.5 V. It is not surprising that the simulations give a current higher than experiment in the depletion mode devices since the simulations are of the intrinsic device, with ideal ohmic contacts and no external parasitics included. However, if that is the reason for the discrepancy in the depletion mode device, it is necessary to explain the much better agreement between simulation and experiment in the enhancement mode device. It is possible that the experimental device structures are not quite as assumed or that there are different parasitics brought about, for example, by the existence or otherwise of a recess.

The simulated drain current plotted as a function of gate voltage at a fixed drain bias of 1 V is shown in Figure 4.28. The soft pinch-off behaviour is seen for both devices but especially for the enhancement mode device. By selecting the linear part of each

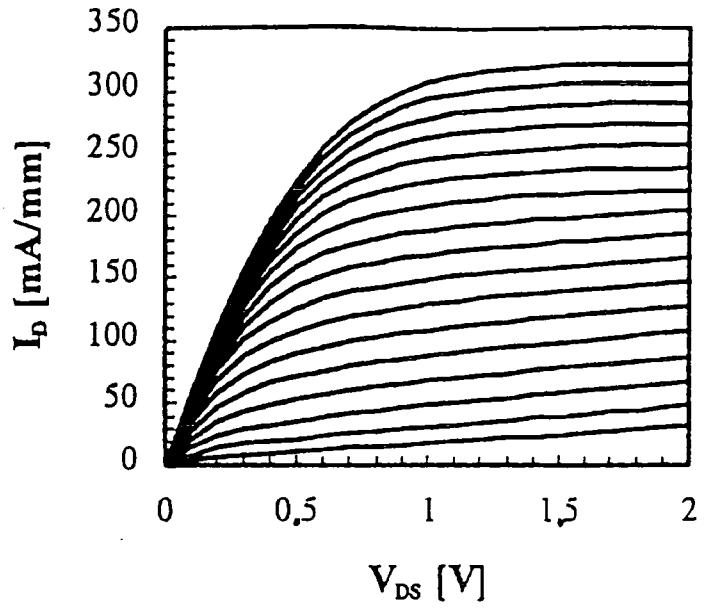


(a) depletion mode

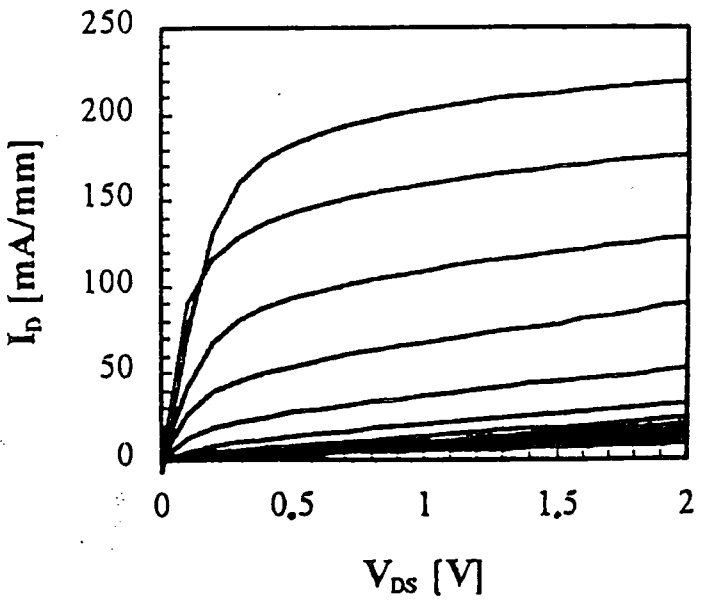


(b) enhancement mode

Figure 4.26: (a) Simulated drain current plotted as a function of drain voltage at different gate voltages for the simulated $Si/Si_{0.55}Ge_{0.45}$ n-MODFET illustrated in Figure 4.25(a), for which the gate length is 0.15 μm. The gate voltage ranges from -1.1 to $+0.6$ V, in 0.1 V intervals. The knee voltage is around 0.8 V.(b) Simulated current-voltage characteristic for the enhancement mode device. The gate voltage has the values$-1, -0.6, -0.4, -0.2, -0.1$ to $+0.6$ V in 0.1 intervals. The knee voltage is now about 0.5 V. Extracted from [89].



(a) depletion mode



(b) enhancement mode

Figure 4.27: DC output characteristics of $Si/Si_{0.55}Ge_{0.45}$ n-type MODFETs investigated experimentally by the Daimler-Chrysler group [73]. **(a)** Device working in depletion mode with a maximum extrinsic transconductance g_{me} = 205 mS mm^{-1}, (gate voltage weep: $V_G = -1$ V to 0.6 V, in $\Delta\ V_G = 0.1$ V steps). **(b)** Enhancement mode device with a 20 seconds recess process and a high maximum g_{me} = 476 mS mm^{-1}. (gate voltage sweep: $V_G = -1$ V to 0.7 V, in $\Delta\ V_G = 0.1$ V steps.) Extracted from [73].

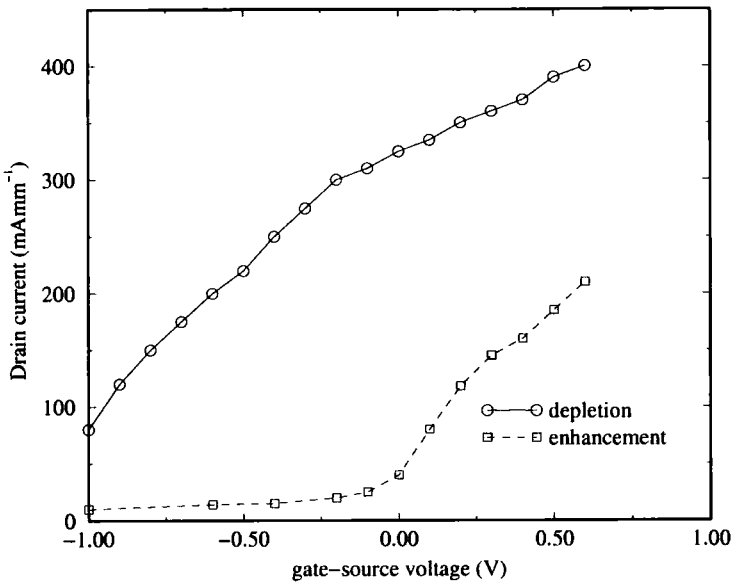
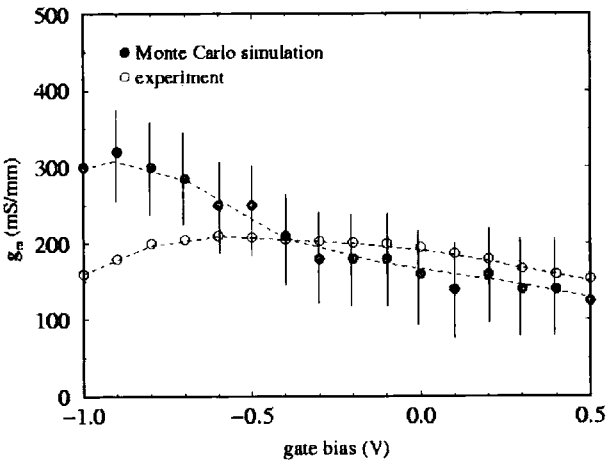


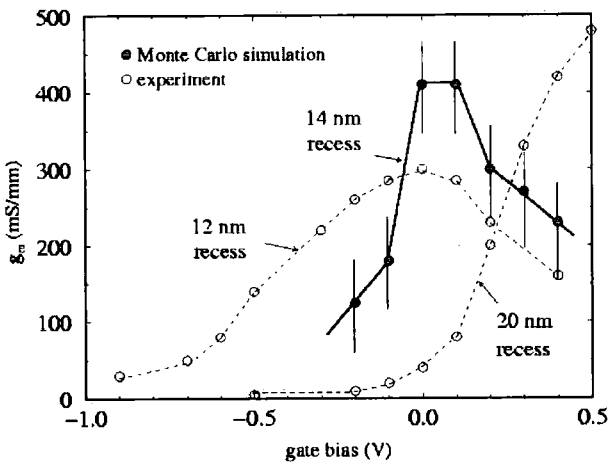
Figure 4.28: Simulated drain current plotted as a function of gate voltage at a fixed drain bias of 1 V.

curve and extrapolating, threshold voltages of -1.6 and -0.2 V can be estimated for the depletion and enhancement cases respectively.

Figure 4.29 shows the transconductance derived from the current characteristics at 2 V drain bias for the depletion (surface-gate) and enhancement (recessed-gate) modelled devices illustrated in Figures 4.25. Due to the statistical nature of the Monte Carlo method, the noise on the simulated contact currents is at least $\pm$ 30 mA mm^{-1}, and the error placed on $\Delta I_{drain}/\Delta V_{gate}$ is $\pm$ 60 mS mm^{-1}. The peak of the transconductance of the depletion mode device is about 300 mS mm^{-1} at a gate bias of -0.9 V and a drain bias of 2 V. Note that the experimental transconductance profile [73], which is also shown in Figure 4.29(a), is quite flat with a value of about 200 mS mm^{-1}. The calculated result agrees quite well for the positive gate voltages and small negative gate voltages up to -0.4 V. Nonetheless, the discrepancy is noticeably larger for higher negative gate voltage. Since the depletion device is already on, even when there is no voltage applied to the gate, the positive gate voltages do not produce a substantial increase in current. Instead, the depletion device requires a negative gate voltage in order



(a) depletion mode



(b) enhancement mode

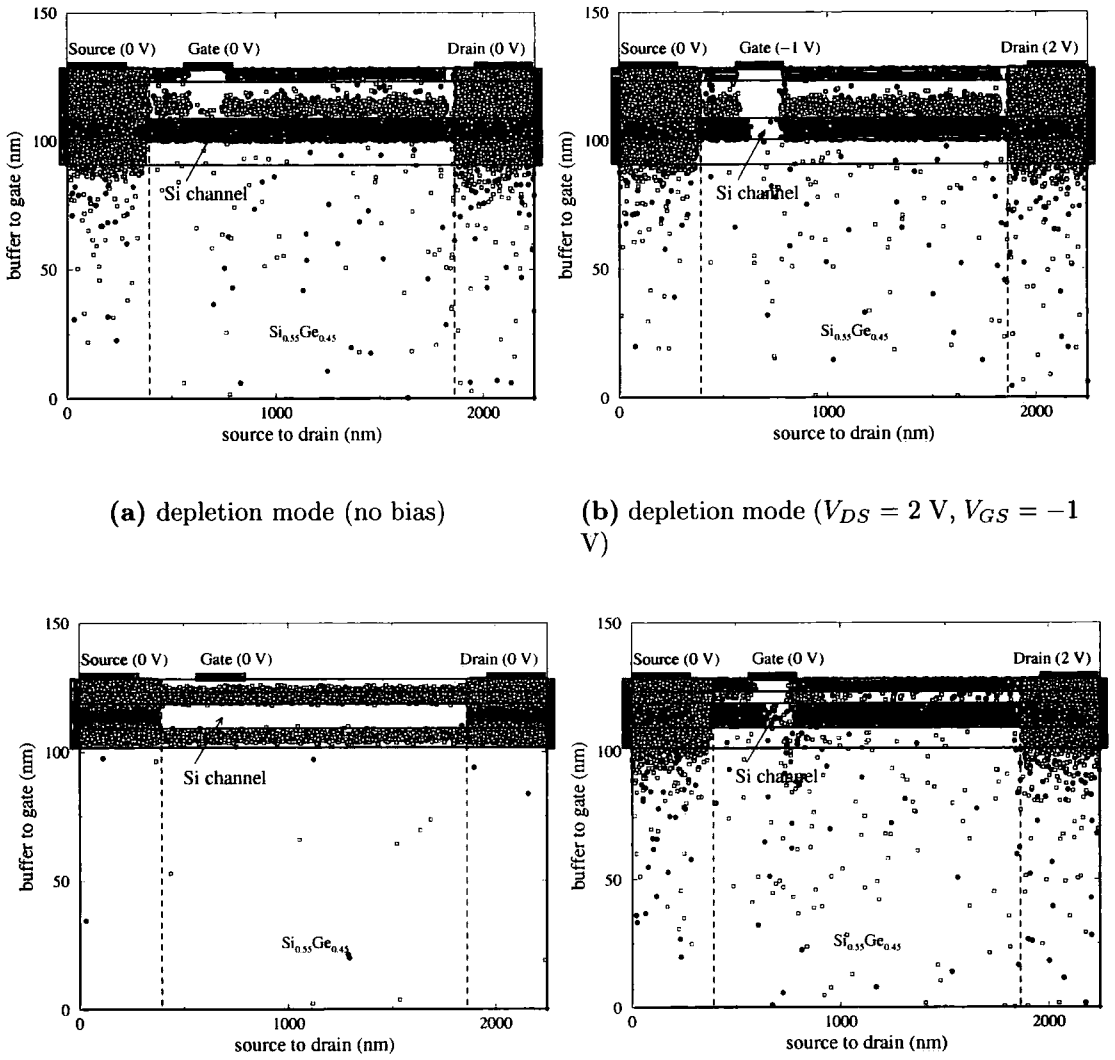
Figure 4.29: **(a)** The calculated transconductance g_m (filled circles) from the transfer characteristic and the measured transconductance g_m (open circles) of the depletion mode device. The drain bias is 2 V. Note that the error bars on the simulated g_m values take account of the fact that noise on the simulated drain current is at least $\pm$ 30 mA mm^{-1}, when 30,000 electron particles are used for the simulation. **(b)** A comparison of predicted transfer characteristic for the enhancement mode device at the same bias arrangement as above figure. Experiment data from Gluck *et al.* [73]. Extracted from Crow *et al.* [89].

to cut the conducting channel. The negative gate voltage causes a significant change in the conducting channel, and hence a large change in drain current. Thus a high transconductance can occur at a low drain current. The flat profile of the reported transconductance indicates the absence of any significant parallel conduction which would take place in the low-mobility SiGe doped layer. On the other hand, due to the smaller gate-to-channel separation, the transconductance obtained from simulations of the enhancement mode device (Figure 4.29(b) reaches the higher value of about 400 mS mm^{-1}). The peak is found between gate voltages of 0 and 0.1 V and this agrees well with the experimental results shown for a device with a 14 nm gate recess.

Figure 4.30 shows the spatial electron distribution in the depletion and enhancement devices. For the depletion device, it is apparent from Figure 4.30(b) that a -1 V can effectively cut the conducting channel. Note the substantial electron transfer from the back supply layer to the channel. Figure 4.30(c) shows the electron distribution of the enhancement device in the absence of bias. The channel well is essentially unoccupied by electrons as expected. At a drain bias of 2 V, with zero voltage on the gate, more (Δ_2) electrons occupy the channel and there is a significant drain current. Note that at this bias condition, there is no parallel conduction path under the gate in the top supply layer in either device. However, there is parasitic charge layer in the source-gate and gate-drain regions, which is a consequence of the incomplete electron transfer from the top supply layer to the Si-channel.

2) Dynamic (AC) characteristics

Simulations of the response of the devices to modulation of the gate voltage have also been performed. The methodology is essentially identical to that described in Section 4.4. A 100 ps duration sinc-form pulse containing ten mini-cycles provides a flat frequency spectrum up to 100 GHz, with 10 GHz resolution. Figure 4.31(a) shows the



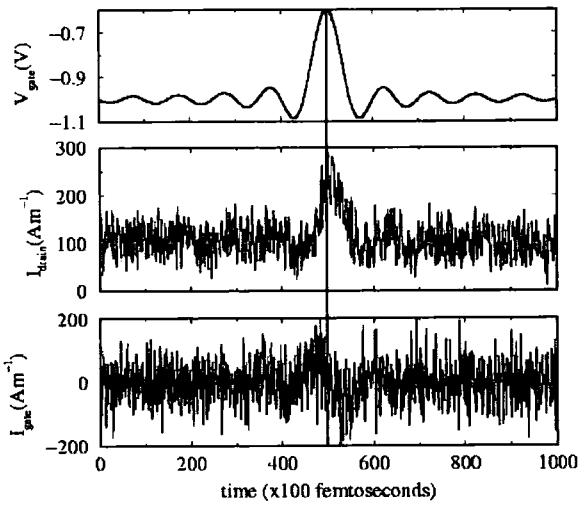
(a) depletion mode (no bias)

(b) depletion mode ($V_{DS} = 2$ V, $V_{GS} = -1$ V)

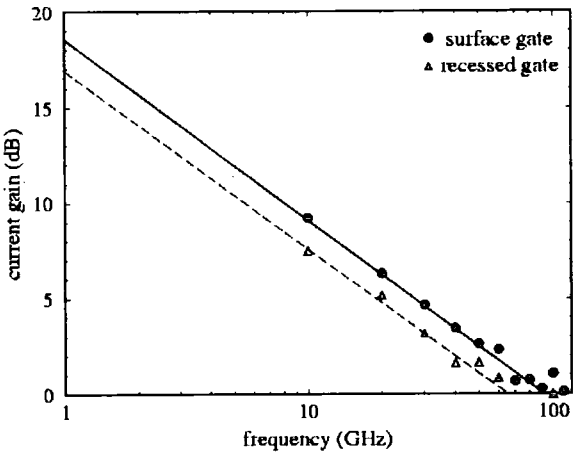
(c) enhancement mode (no bias, $t = 0$)

(d) enhancement mode ($V_{DS} = 2$ V, $V_{GS} = 0$ V)

Figure 4.30: Superparticle distribution in the depletion and enhancement mode devices (closed circles are Δ_2-electrons, opened squares are Δ_4-electrons). **(a)** The electrons distribution at the unbiased condition for the depletion mode device. Note that the channel is already populated. **(b)** The electron distribution at 2 V drain bias. A gate voltage of -1 V is needed to cut the conducting channel. **(c)** The initial ($t = 0$) electrons distribution of the enhancement mode device at no bias arrangement prior to charge transfer to the Si-channel. The positions of the top and back supply layers are clearly seen. **(d)** The electron distribution at 2 V drain bias. The gate is unbiased. This bias arrangement provides the maximum of transconductance illustrated in Figure 4.29.



(a) frequency response



(b) cut-off frequency

Figure 4.31: (a) The simulated frequency response of a $Si/Si_{0.55}Ge_{0.45}$ n-channel MODFET to a 100 ps sinc gate voltage pulse. From top to bottom, this figure shows the gate voltage (V_{gate}) as a function of time, the drain current response (I_{drain}), and the gate current (I_{gate}). **(b)** Logarithmic plot showing the maximum intrinsic current gain calculated for the depletion device (solid circles) and enhancement device (solid triangle). The cut-off frequencies f_T^i for the surface and recessed gate devices are 90 ± 10 GHz and 60 ± 10 GHz, respectively. Extract from Crow *et al.* [89].

voltage signal applied to the gate contact, the output drain current, and the corresponding gate current of a depletion device at 2 V drain bias. The gate bias is -1 V and the amplitude of the applied signal is 0.2 V. The total number of particles employed in the simulations is about 50,000. The Schottky gate is a non-emitting contact and the current measured is mainly displacement current. Consequently the gate current is $\pi/2$ out of phase with the drain current. An identical voltage pulse was applied to the enhancement n-MODFET, in which the gate signal was superimposed on a steady gate bias of 0.1 V. The calculated response was similar to Figure 4.31(a).

Figure 4.31(b) shows that the intrinsic current gain cut-off frequencies f_T^i for the enhancement and depletion mode devices are 60 $\pm$ 10 and 90 $\pm$ 10 GHz. Not surprisingly, our calculated cut-off frequencies for both devices are higher than the reported experimental results [73], in which the extrinsic cut-off frequency f_T is 30(43) GHz. The discrepancy can be explained by the fact that parasitic capacitances and resistances are not included in the Monte Carlo model. The reason for the higher cut-off frequency of the depletion mode device relative to the enhancement mode device is discussed in the next section.

3) Microscopic Behaviour

In Section 4.2 it was explained that the advantages of high tensile strain were the stronger confinement of electrons and the greater response of the drift velocity to a step of electric field as a function of time. The contribution of velocity overshoot to the effective or average channel velocity depends on the energy relaxation time, and therefore altering the energy band structure by strain may have a significant impact on the effective channel velocity. Figure 4.32 shows the longitudinal electric fields along the channel of the two simulated devices. Note that there are field spikes at the source region-channel and drain region-channel interfaces of the two devices in

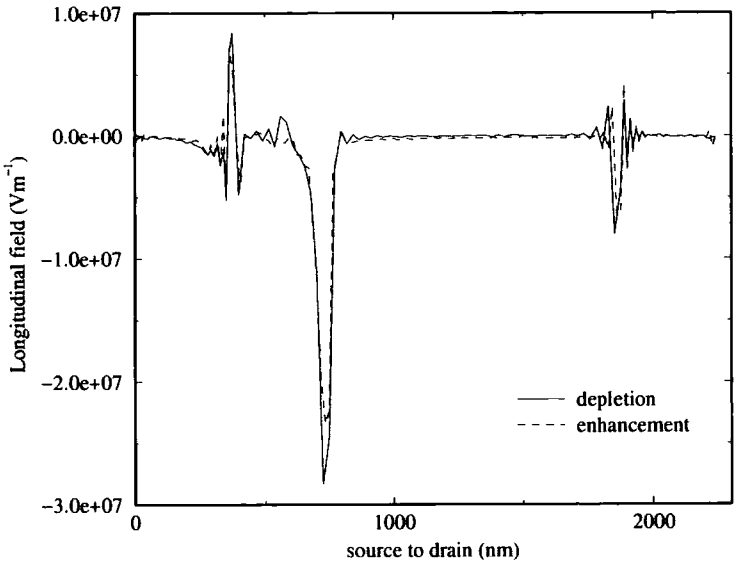


Figure 4.32: Longitudinal electric field along the channel for the 0.15 μm gate length devices working in depletion mode (solid line) and enhancement mode (dashed line). The drain bias is 2 V. The gate voltage is -1 V for depletion mode device and 0 V for the enhancement mode device.

Figure 4.32. The origin of these is the same as described in Section 4.4. A very high electric field of about 3×10^7 Vm^{-1} is found under the gate for the depletion mode device. Our bulk simulations of strained bulk material showed that the drift velocity of electrons can reach 2.5×10^5 ms^{-1} during the transient response to an applied field. Similar values are found for electrons in the channel of the depletion mode device as illustrated in Figure 4.33. The very high field under the gate region results in strong velocity overshoot. The velocity overshoot of Δ_2 electrons and the subsequent Δ_2-Δ_4 intervalley scattering results in a spatial separation of the peak of average drift velocity and the peak of kinetic energy. The fractional occupation of the Δ_4 valleys is 50% at the end of the gate section. The electrons subsequently emit phonons and cool, and the Δ_2 valley electrons drift along the low-field gate-drain section of the Si channel with a net velocity of 2×10^4 ms^{-1}. Since there is a lower electric field in the channel of the enhancement mode device, the electrons are not heated so much as in depletion mode device, and the population of the Δ_2 valleys is about 80% immediately after passing

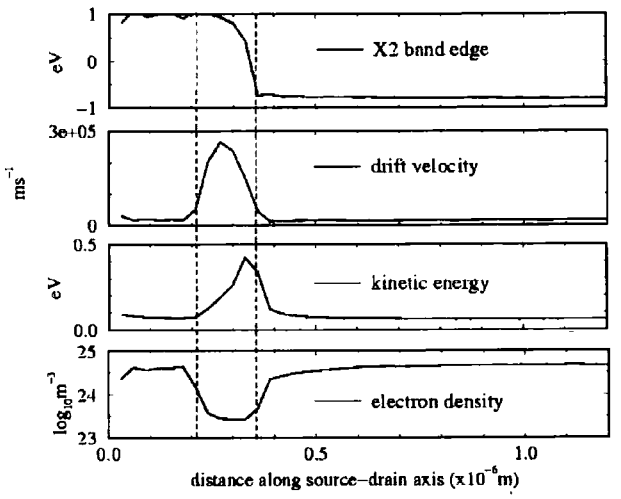


Figure 4.33: The microscopic data of electron transport as a function of position through the source-drain section of the Si channel of the depletion device. The steady drain-source bias is 2 V, and the gate bias is −1 V. From top to bottom, the Δ_2 valley band profile, the average (Δ_2 valley + Δ_4-valley) drift velocity and kinetic energy and total electron density. The dashed lines indicate the section of the channel under the gate. Extracted from Crow *et al.* [89].

the high field region.

4.6 Summary and Conclusion

A detailed analysis of the in-plane electron drift velocity characteristics of bulk Si strained as if grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates has been presented. Both steady state and transient simulations were carried out and the results compared to unstrained Si. For the steady state, the simulations reveal that the average electron drift velocity and kinetic energy are enhanced as a function of electric field by the effects of strain. The enhancement of kinetic energy is especially pronounced at low fields. At very high fields, electrons in strained Si behave in much the same way as in unstrained Si. However, the transient response simulations suggest that the electron drift velocity can reach 2.5×10^5 ms^{-1} in strained Si as a result of velocity overshoot

compared to a saturation velocity of about 10^5 ms^{-1} in unstrained Si.

Ensemble self-consistent Monte Carlo device simulations have been used to model steady state and transient electron transport in n-channel $Si_{1-x}Ge_x$ MODFETs. The simulated devices studied can be categorised into two groups: moderate and high strain n-MODFETs.

Three depletion-mode n-$Si/Si_{0.77}Ge_{0.23}$ MODFETs were simulated. The effects of strain on the high field transport and device performance were studied. The effect of varying gate length L_G and source-gate separation L_{SG} have also been investigated. Macroscopic and microscopic device properties were obtained from the simulations and the results analysed. Detailed time-dependent signal analysis was carried out to study the device response and derive the frequency bandwidth. For each device, a sinc voltage pulse was applied to the gate and the resulting drain and gate currents used to calculate the current gain as a function of frequency. The calculations showed that the highest predicted cut-off frequency was 80 $\pm$ 10 GHz for the smallest device with a gate length of 0.07 μm.

Simulations of a depletion-mode n-$Si/Si_{0.75}Ge_{0.25}$ MODFET (device 4) were also carried out. The modelled transistor has some similarities to that recently investigated by the IBM group [37], was found to have which a cut-off frequency of 62 GHz. Device 4 is predicted to have a cut-off frequency of 20 $\pm$ 10 GHz. However, device 4*, which is similar to device 4 but has higher doping in the source and drain implant regions and is probably closer to the IBM device, has a calculated cut-off frequency of 40 ± 10 GHz.

Regarding device design and optimisation of the moderate strain n-MODFETs, it is apparent from our investigations that a general reduction of the channel length and the gate length results in an enhancement of the cut-off frequency f_T. The shorter gate length results in a smaller gate transit time when electrons travel at their saturation velocity but velocity overshoot effects can also occur, resulting in a further enhancement of device speed. However, there are some other factors that can affect the high

frequency performance. For example, the gate-source separation and the doping levels in the source and drain implant regions. The smaller gate-source separation results in a higher electric field near the source, which promotes velocity overshoot in the source-gate region and hence reduces the gate transit time, as seen in devices 1-3 in comparison to devices 4 and 4*. High doping in the source and drain implant regions is also important as it facilitates the transfer of electrons between the contacts and the channel, as seen in the comparison of devices 4 and 4*.

In addition, simulations of depletion and enhancement-mode n-Si/$Si_{0.55}Ge_{0.45}$ MODFETs were performed. The device structures are close to those experimentally investigated by the Daimler-Chrysler group. Calculated cut-off frequencies for the enhancement and depletion mode devices are 60 $\pm$ 10 and 90 $\pm$ 10 GHz respectively, which should be compared with the experimental values of 30(43) GHz.

Chapter 5

SOI-LBJT Simulations

5.1 Introduction

Although most microelectronics products are made of MOS field effect transistors for reasons described earlier, bipolar transistors have the advantage of high current capability, inherently high speed for digital circuit applications and superior characteristics for analog circuit applications [90]. Therefore there are considerable attractions in devices which combine the advantages of bipolar transistors with well established CMOS technology, including reliable mass production methods. Also, silicon-on-insulator (SOI) [91,92] is an emerging technology for integrated circuits due to its low leakage currents, reduced parasitic capacitances, and potential advantages in terms of speed and power relative to conventional bulk and epitaxial wafers. SOI is expected to extend bulk CMOS performance limits beyond 0.1 μm and provide the ultimate speed/low power performance achievable with scaled CMOS [93]. The above observations have been the motivation for recent research into silicon-on-insulator lateral bipolar transistors (SOI-LBJT) [94].

Initially lateral bipolar transistors were realised as a parasitic effect in MOSFETs, but suffered from low current gain [95,96]. Nevertheless, several applications, such as current mirrors, voltage references, and comparators have been demonstrated [97,98].

SOI has the advantage of reducing leakage currents and also of drastically reducing the emitter-substrate and collector-substrate parasitic capacitances, which in turn improves device speed [91]. Since bipolar transistors exhibit higher transconductance for a given area and bias current than MOSFETs [99], SOI bipolar devices may find application either by themselves or in combination with MOSFETs in the same circuit (SOI CMOS).

Lateral bipolar transistors fabricated using an SOI CMOS process have been demonstrated with current gains up to 70 (and 40) for npn (and pnp) devices with a base width of 0.5 μm [100]. In these CMOS-compatible devices, the gate contact is preserved but the base contact is made on one side of the base (see Figure 2.12), resulting in high base resistance. To overcome this problem, the device reported by Sturm *et al.* [39] utilises a polysilicon contact all around the base with nominal base widths from 2 to 10 μm, providing an acceptably low base resistance. In this case the current gain of the devices appeared to be limited by recombination in the SOI film. Most recent devices use base contacts above the active part of the base, reducing the base resistance [101]. Recently, the simulation results of an npn LBJT built on a 0.1 μm thick SOI substrate and based on a minor modification of a submicron MOSFET without gate oxide suggest that it should be possible to achieve a cut-off frequency f_T of about 35 GHz and common emitter current gain of about 60 for a base doping p^+ $N_B = 5{\times}10^{23}$ m^{-3} [94]. Since there is no thin gate oxide employed in SOI-LBJTs, the devices could be an alternative to MOSFETs when down-scaling of critical device dimensions will no longer be an affordable option for bulk production.

In this chapter we report how a Monte Carlo simulation has been devised and used to model steady state and transient electron and hole transport in SOI-LBJTs. Four devices are studied and the effects of junction depth and silicon layer thickness are investigated [102]. The device geometries that have been modelled and the transport model used are described in Section 5.2. Section 5.3 is devoted to a presentation of the

device characteristic and an analysis of the simulation results. Conclusions from the results of the work are drawn in Section 5.4.

5.2 Device Simulation

Figure 5.1 shows the structure of the SOI-LBJT that has been modelled. It has lateral dimensions and doping densities which are close to those of the experimental device of Gómez *et al.* reported in [94]. The background p-type doping throughout the transistor is 2×10^{21} m^{-3}. The ohmic emitter and collector contacts have shallow n^+ implants of doping density 10^{25} m^{-3} and the sub-collector and region beneath the emitter are formed by a n^- doping implants of 2×10^{23} m^{-3}. The base contact is placed at the top of the device in order to reduce the base resistance [91]. The 0.25 μm $\times$ 0.3 μm base pedestal is p^+ doped to a level of 2×10^{24} m^{-3}, with the doping set back from the emitter and collector junctions by 75 nm. Note that the doping profile of our model structure is different from a normal homojunction bipolar transistor, in which the emitter has the highest doping, the collector the lowest and the base doping density at some intermediate level.

The four devices described in Table 5.1 and Figure 5.1 are studied and the effects of junction depth and silicon layer thickness are investigated. Device 4 has minimal modifications to a conventional MOSFET, and the intrinsic base width is equivalent to the channel length of the MOSFET. The structures of devices 1-3 are similar to that of device 4 except that the LBJTs are fabricated on silicon-on-insulator (SOI) substrates. The motive behind this development is two fold: i) it removes the need for n-type well-stops and hence the presence of parasitic pnp transistors within the structure, ii) current leakage via the substrate is removed.

The transistors here have been modelled using the SLURPS software. Both electrons and holes are included in the simulations using the bandstructure and carrier

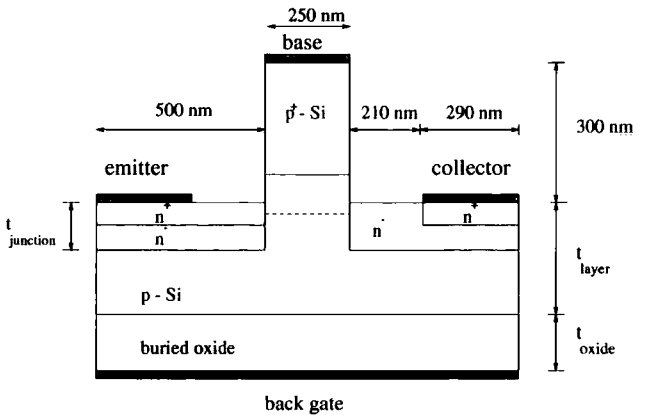


Figure 5.1: The simulated SOI-LBJT device structure. The emitter and collector n^+ implants are 10^{25} m^{-3}, and the n^- doping for the sub-collector and the region around the emitter is 2×10^{23} m^{-3}. The background p-doping throughout the device is 2×10^{21} m^{-3}. The p^+ base doping is 2×10^{24} m^{-3}. The dashed line beneath the base pedestal indicates the depth in the channel at which the electric field and other microscopic details are recorded for subsequent figures in Chapter 5.

transport models described in Section 3.8.1. Electron-hole recombination is difficult to incorporate into Monte Carlo simulation because the timescale on which the process occurs is generally much longer than that for carrier scattering, and it is not included here. Its omission is not expected to have a major effect on the carrier dynamics in the device because the base width is much less than the electron diffusion length.

As discussed in Section 3.6, model devices in SLURPS are built up as a series of joined rectangular regions, with the electric field cell sizes matched along the join connecting each region. Eight connected regions are used in order to model the emitter, base, sub-collector, collector and back gate oxide. Note that the buried oxide layer is broken into three regions corresponding to the main features of the device structure above it. Initially the number of mesh cells (horizontal×vertical) for the buried oxide layer was 16×4 whilst 16×16 was used for other regions. Further simulations in which the number of mesh cells in the horizontal direction of every region was increased to 32 have also been performed.

Device	t_{oxide} (nm)	t_{layer} (nm)	$t_{junction}$ (nm)
1	100	300	112.5
2	100	300	75
3	100	200	50
4	0	400	75

Table 5.1: Device parameters used in the simulations.

Figure 5.2 shows the average kinetic energy of electrons, longitudinal electric field, electron and hole densities, and drift velocities of electrons and holes recorded along the channel of device 1. The bias condition is $V_{BE} = 1.1$ V and $V_{CE} = 3$ V. It can be seen that the results obtained using 16 (solid line) and 32 (dashed line) cells in the horizontal direction for each region are very similar, which suggests that 16 cells in the horizontal direction can be used to ensure results of acceptable accuracy without too large a penalty in terms of computing time.

The number of cells in the vertical direction quoted above is the maximum that can be achieved in the LBJT structures studied here without encountering computational problems similar to those described in Section 4.4.

As explained in Chapter 3, in the 2D simulation each superparticle can be considered to be a charge rod with charge per unit length given by Equation 3.51 (and its equivalent for holes). In Equation 3.51 the number of electrons n is taken equal to the magnitude of the net impurity charge (in units of e) in the n-type regions of the device. For hole superparticles, n is replaced by the number p of holes, which is equal to the net impurity charge in the p-type regions of the device.

In the interests of reducing fluctuations in charge density within the device and noise in the currents through the contacts, it is advantageous to have as many superparticles as possible. However, the computation time increases with superparticle number and it is normal practice to use the minimum number for which reliable results can be obtained - here between 2×10^4 and 4×10^4. There is then the question of how

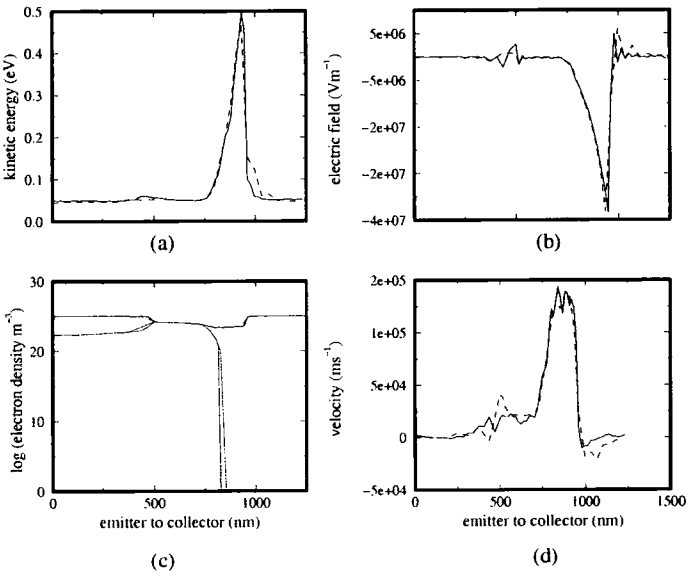


Figure 5.2: A comparison of the transport properties of device 1 by use of 16 (solid line) and 32 (dashed line) cells in the horizontal direction in each region. **(a)** Average kinetic energy of electrons recorded along the channel, **(b)** Longitudinal electric field, **(c)** Electron and hole densities (red line for 16 cells and green line for 32 cells). The electron density for 32 cells is not shown since it is essentially identical to the 16 cells result. **(d)** Average longitudinal drift velocity of electrons.

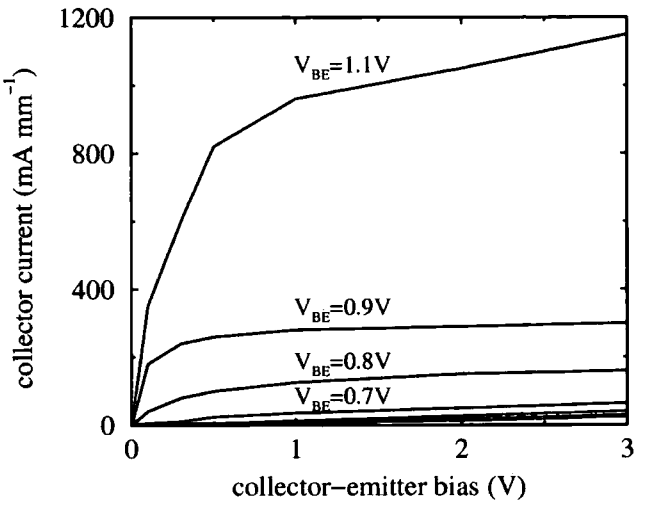


Figure 5.3: Calculated common emitter characteristics of device 2 with a base width of 0.25 μm. The base voltage has values from 0.1, 0.3, 0.5, 0.7, 0.9 and 1.1 V.

many superparticles to use for the electrons and for the holes in the device. In bipolar devices there can be large differences in the total donor and acceptor charges and, if the same numbers of electron and hole superparticles are used, it can lead to one type of superparticle having a much larger charge, resulting in problems with charge fluctuations. Here the ratio of donor to acceptor charge is only about 5 but we have found it beneficial to the suppression of fluctuations in the simulation results to increase the number of electron superparticles relative to the number of hole superparticles so that the charge ratio is only 1.78.

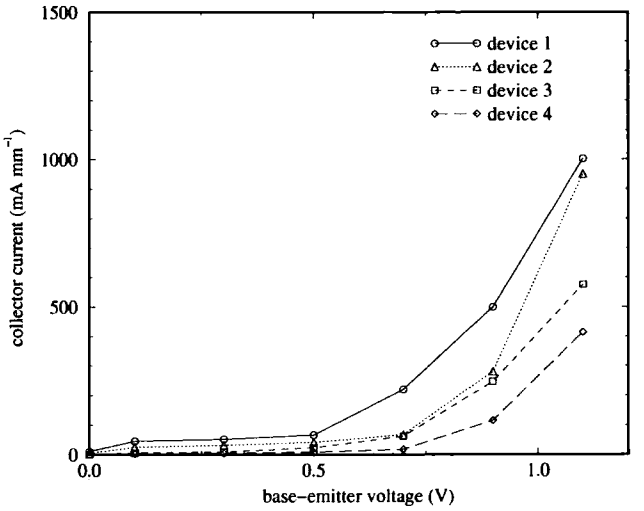


Figure 5.4: Calculated collector current (I_C) versus base-emitter voltage (V_{BE}) for the four devices. The collector bias is 3 V.

5.3 Device Characteristics and Analysis

5.3.1 DC characteristics

Figure 5.3 shows the simulated collector current-collector voltage characteristic for device 2 at 300 K, for base voltages V_{BE} of 0.1, 0.3, 0.5, 0.7, 0.8, 0.9, and 1.1 V. The emitter and the back gate beneath the buried oxide are grounded. For collector bias $V_{CE} < V_{BE}$, the collector-base and base-emitter junctions are forward biased and the transistor is in saturation mode. In this bias arrangement, the collector current is mainly the difference of the electron current injected from the emitter into the base and the electron current injected from the collector into the base. When V_{CE} is increased, the collector current grows as the transistor becomes less saturated. For $V_{CE} > V_{BE}$, the collector-base junction is reverse-biased and the transistor is now in the active mode of operation; most of the electrons injected from the emitter into the base are collected by the collector, but there is negligible electron injection from the collector into the base.

The calculated collector current (I_C) versus base-emitter voltage (V_{BE}) character-

istics for the four simulated devices are depicted in Figure 5.4. The applied collector-emitter bias is 3 V for all four devices, which are therefore operating in the active mode for the range of V_{BE} shown. The plot shows ideal behaviour in the sense that the collector current is proportional to $\exp(eV_{BE}/k_BT)$, a consequence of the exponential variation with base-emitter voltage of the electron concentration at the emitter edge of the base. It can be seen from this figure that the collector current of device 1 is highest at each value of base-emitter bias for which calculations have been made.

The (static) current amplification of a transistor in the common-emitter configuration is defined as $\beta = I_C/I_B$ where I_B is the base current. Figure 5.4 shows that at a collector-emitter voltage of 3 V the maximum common emitter current gain of device 1 is about 400 at the low base bias of 0.1 V whilst the maximum of device 2 is about 80 at the higher base bias 0.5 V. An important difference between these two devices is that the Si layer thickness of device 1 is 50 % larger than device 2. In simple terms, the thicker Si layer allows higher currents without high injection effects. Although the static current gain of device 1 peaks at about 400 at a base voltage of 0.1 V, the current gain at a voltage of 0.5 V is much lower due to the significantly higher base current.

The rapid decline of the current gain of device 1 with V_{BE} can be attributed to a substantial increase in base current as illustrated in Figure 5.6. Note from Figure 5.6 that the current gain is less than unity at base-emitter voltages of 0.7 V or higher. The current gain of device 3 peaks at a higher value than that of device 2 but does so at the low base-emitter bias of 0.3 V. Device 4 exhibits low current amplification at all V_{BE}, and suffers from current leakage through the substrate at the larger base voltages (V_{BE} = 0.7 V and larger) as shown in Figure 5.7. Note from Figure 5.5 that the current gain peaks at the same V_{BE} of 0.3 V in devices 3 and 4, although the collector current in device 3 is somewhat higher at the larger base voltages.

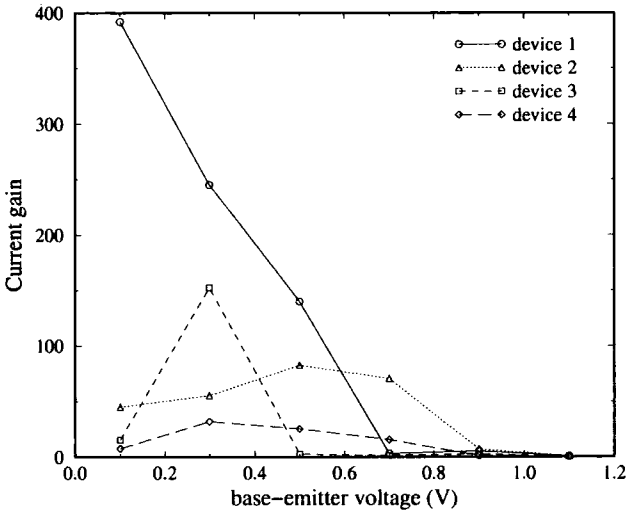


Figure 5.5: Calculated current gain versus base-emitter voltage (V_{BE}) for the four devices. The collector bias is 3 V.

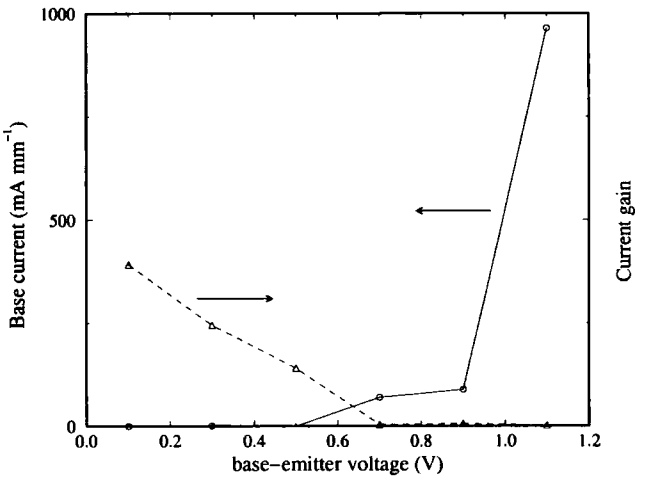


Figure 5.6: Calculated base current versus base-emitter voltage (V_{BE}) for device 1. The collector bias is 3 V. The current gain is also depicted as a dashed line.

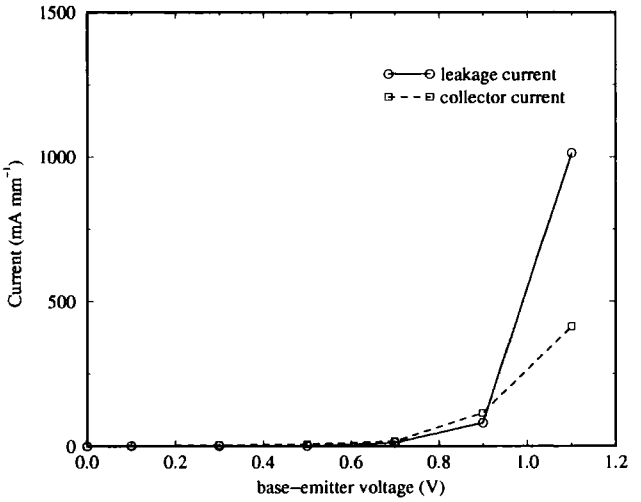


Figure 5.7: Calculated leakge current (solid line) through the substrate and collector current (I_C) (dashed line) versus base-emitter voltage (V_{BE}) for in device 4. The collector bias is 3 V.

5.3.2 Signal Analysis

Simulations of the effect of modulating the base bias have been carried out, in order to test the collector current response and derive the intrinsic cut-off frequency f_T^i. The frequency response for each device has been investigated by applying a truncated sinc voltage pulse to the base contact and taking Fourier transforms of the output base and collector currents in a procedure analogous to that described in Section 4.4.2 for MODFETs. The time dependence of the base voltage is

$$V_{base}(t) = V_o + V_{amplitude}\left(\frac{\sin\omega(t-T/2)}{\omega(t-T/2)}\right) \tag{5.1}$$

where T is the duration of the signal. A 200 ps duration sinc voltage pulse containing ten mini-cycles provides a flat frequency spectrum up to 50 GHz, with 5 GHz resolution. The base voltage signal was applied in the simulation once a device had reached a stable collector current at 3 V collector bias with the base bias level set at 0.5 V. These bias conditions were chosen to provide minimal base injection current at an acceptable level of collector current and hence the possibility of high current gain.

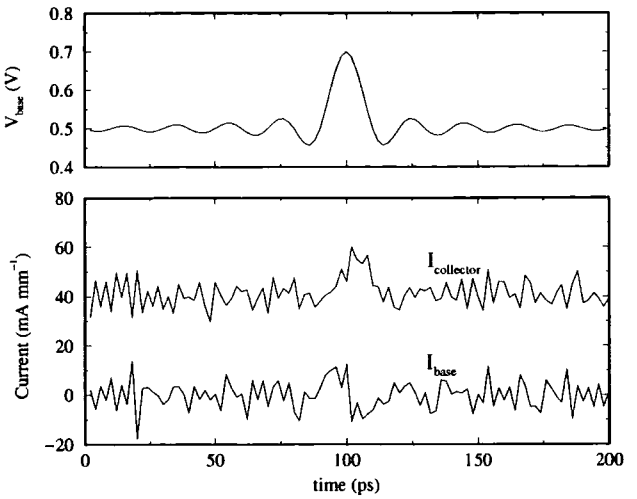


Figure 5.8: (Top) The sinc voltage pulse of peak amplitude of 0.2 V superimposed on steady state value of 0.5 V applied to the base of device 2. The collector-emitter voltage is 3 V. (Bottom) The collector and base current responses.

The response of device 2 is shown in Figure 5.8. The top diagram shows the sinc voltage pulse of peak amplitude 0.2 V applied to the base, whilst the bottom diagram shows the collector and base current responses. Note that the collector current resembles the shape of the applied voltage signal while the base current is largely composed of displacement current for this bias arrangement, and hence resembles the time derivative of the applied signal. The high frequency noise on the response currents is a consequence of the use of a limited number of superparticles (electrons and holes) with each superparticle carrying a substantial charge because of the relatively heavy doping in the emitter and collector.

The current gain of device 2 plotted as a function of frequency is illustrated in Figure 5.9. The predicted current gain has a cut-off frequency f_T^i approaching 35 $\pm$ 5 GHz. Note that f_T of the simulated SOI npn LBJT investigated by Gómez *et al.* is about 25 GHz for the same base doping of 2×10^{24} m^{-3} [94]. Figure 5.4 shows that device 1 provides the higher collector current for various base-emitter bias compared to that of device 2, and as a result it might be expected that device 1 would provide

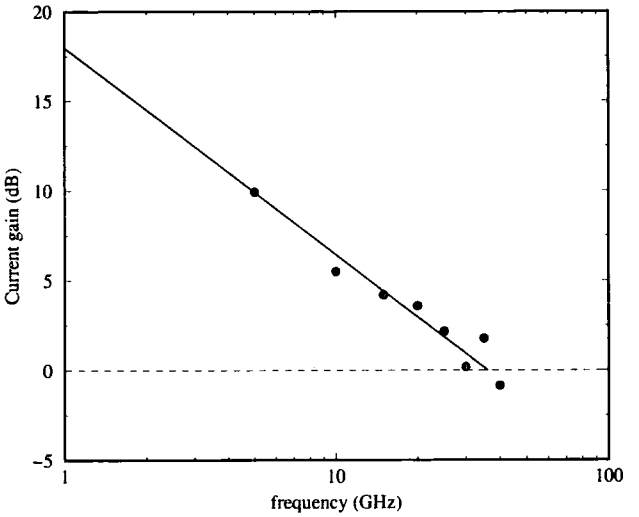


Figure 5.9: The calculated current gain plotted as a function of frequency for device 2. The predicted cut-off frequency f_T^i is 35 ± 5 GHz.

a higher f_T^i. However, as already noted, device 1 suffers from a large base current for base voltages in excess of 0.7 V and since the base signal takes the device into that bias regime, it is not altogether surprising that the cut-off frequency is lower at 15 ± 5 GHz. Device 3 with a thinner p-Si layer and junction depth is predicted to have f_T^i of 10 ± 5 GHz. Device 4, which is deposited on bulk p-type silicon, does not exhibit current amplification in the GHz frequency range. In addition, device 4 suffers from current leakage through the substrate at high base voltages, resulting in a low collector current.

5.3.3 Microscopic Analysis

This subsection is devoted to a microscopic analysis and understanding of the device behaviour that has been presented in previous sections. Note that the nominal base width of all the simulated devices is 0.25 μm, which is much shorter than the diffusion length of the minority carriers (holes). Hence electrons injected from the emitter diffuse across the base without recombination with the holes in the base. In addition, the

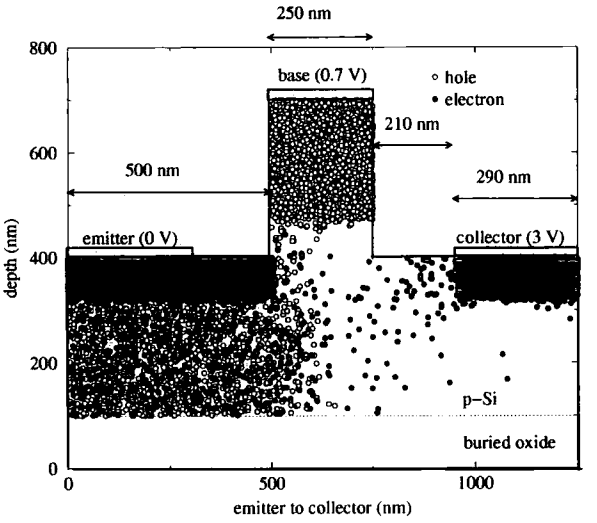


Figure 5.10: Cross-section through the model SOI-LBJT device 1 showing the instantaneous distribution of electrons (black dots) and holes (open circles) for a collector bias of 3 V and base bias of 0.7 V.

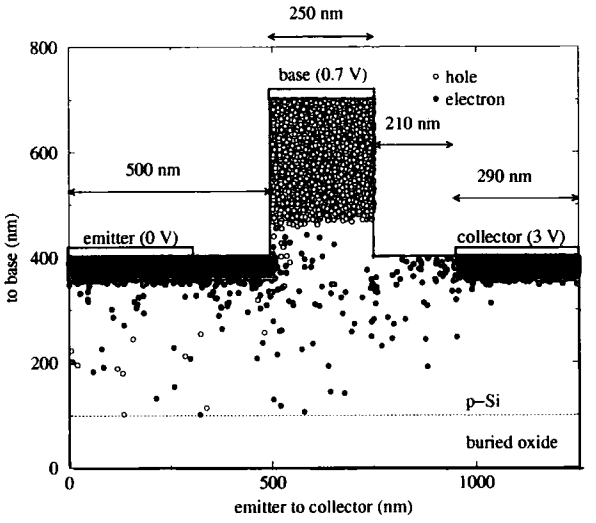


Figure 5.11: Cross-section through the model SOI-LBJT device 2 showing the instantaneous distribution of electrons (black dots) and holes (open circles) for a collector bias of 3 V. The base voltage is 0.7 V.

emitter-base and collector-base junctions can influence each other.

Figures 5.10 and 5.11 show the instantaneous distribution of electrons and holes in devices 1 and 2 respectively with an applied base bias of 0.7 V. The collector-emitter voltage is 3 V. Note at this base voltage the diffusion of holes from the base pedestal into the channel of the device 1, and also the accumulation of injected electrons and holes beneath the emitter. At this base voltage and larger, the emitter current is made up of electron injection and hole leakage. The base current has a significant contribution from electron leakage in addition to hole injection, dramatically reducing the common emitter current gain. It should be emphasised that no recombination processes have been included in our model, and the base current is the sum of the conduction current due to all the electrons and holes passing through the base contact. In contrast to device 1, the diffusion of holes from the base pedestal of device 2 is small. Thus the emitter current is made up mainly of electrons and the base current is still low, resulting in high current gain. Note that the distribution of electrons and holes in device 3, as shown in Figure 5.12, is similar to that in device 1. Figure 5.13 shows the distribution of electrons and holes in device 4 for a collector bias of 3 V and a base voltage of 0.9 V. The advantage of the SOI structure is apparent from the behaviour of the carriers in this all-silicon structure. The distribution of carriers is characteristic of an emitter current which passes into the substrate without contributing to transistor action. The current leakage through the p-Si substrate causes the lower collector current for this device seen in Figure 5.4.

It can be seen in Figure 5.10 that substantial electron and hole populations appear to co-exist throughout the region under the emitter contact in the case of LBJT device 1. The co-existence of a substantial density population of electrons and holes suggests that recombination process could be significant. However, the diffusion length of holes in silicon at this electron concentration is expected to be of the order of microns, and significantly larger than the length of emitter region of the devices simulated

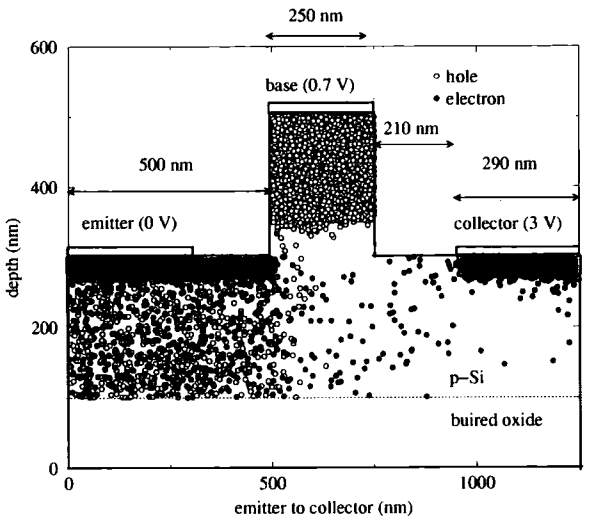


Figure 5.12: Cross-section through the model SOI-LBJT device 3 showing the instantaneous distribution of electrons (black dots) and holes (open circles) for a collector bias of 3 V. The base voltage is 0.7 V.

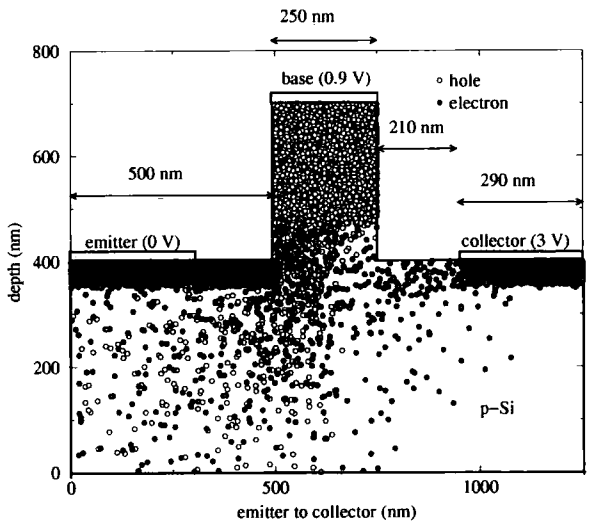


Figure 5.13: Cross-section through the model LBJT device 4 (a conventional all-silicon structure) showing the instantaneous distribution of electrons (black dots) and holes (open circles) for a collector bias of 3 V. The base voltage is 0.9 V.

here. Thus, it is unlikely that the inclusion of recombination processes in our transport model will result in a significant deviation of device behaviour from that presented in the thesis.

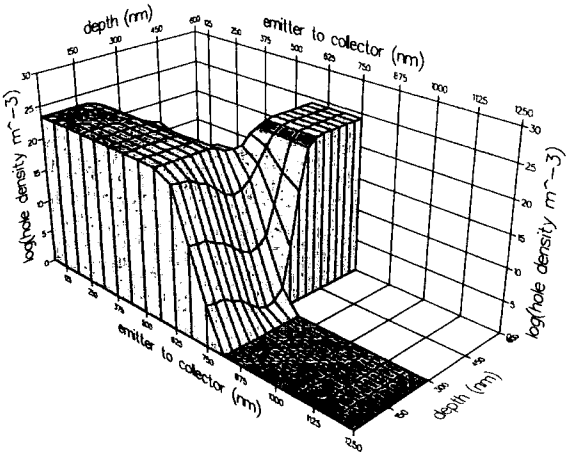
Figure 5.14 shows the hole density as a function of position in devices 1 and 2 at $V_{CE} = 3$ V, and $V_{BE} = 0.7$ V. The stronger diffusion of holes from the base pedestal into the emitter region in device 1 compared to that in device 2 is clear. The hole density in the emitter region of device 2 is about the same as the background p-doping with a value of 2×10^{21} m^{-3}. However, in the emitter region of device 1, the hole density is about 10^{24} m^{-3}. Electron-hole scattering can be expected to be significant in this latter case.

However, it is in the intrinsic base region that strong electron-hole scattering would have a significant influence on device performance and the hole density is much lower there. At higher values of V_{BE} the hole density increases and electron-hole scattering could become more important. To include it in the device simulation would involve a major increase in simultion time. Its effect is likely to be similar to an increased ionised impurity scattering rate. This would decrease currents and increase transit times relative to those calculated here.

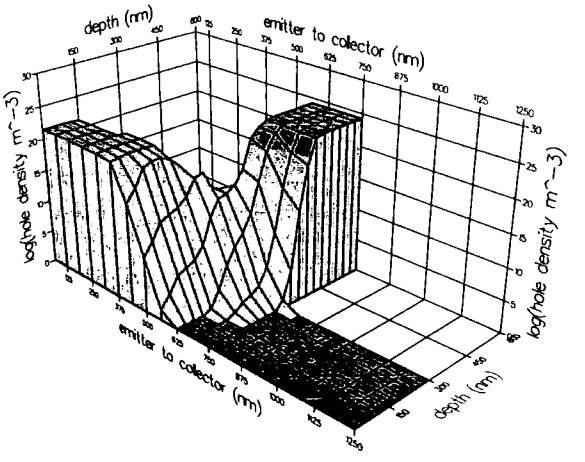
As mentioned earlier the current gain of device 1 declines rapidly with base-emitter voltage, and one of the reasons behind this phenomenon can be explained by considering Figures 5.16 and 5.17. Both figures relate to the bias conditions $V_{BE} = 0.7$ V and $V_{CE} = 3$ V for which the current gain has declined to a low value in device 1 (see Figure 5.5). From Figures 5.16 and 5.17 it is clear that the transverse (vertical in Figure 5.1) velocities[a] of holes and electrons in the vicinity of the base pedestal are much higher in device 1 than in device 2, resulting in a rather higher base current and a lower current gain.

Figure 5.18 shows the longitudinal electric field in the channel of device 2 (along

[a]For the purpose of the plots in Figure 5.16 and 5.17 electron velocity is taken as positive for vertically upward direction while hole velocity is positive for the vertically downward direction.



(a) Device 1



(b) Device 2

Figure 5.14: Hole density plotted as a function of position in devices 1 and 2. The collector bias is 3 V and the base bias is 0.7 V.

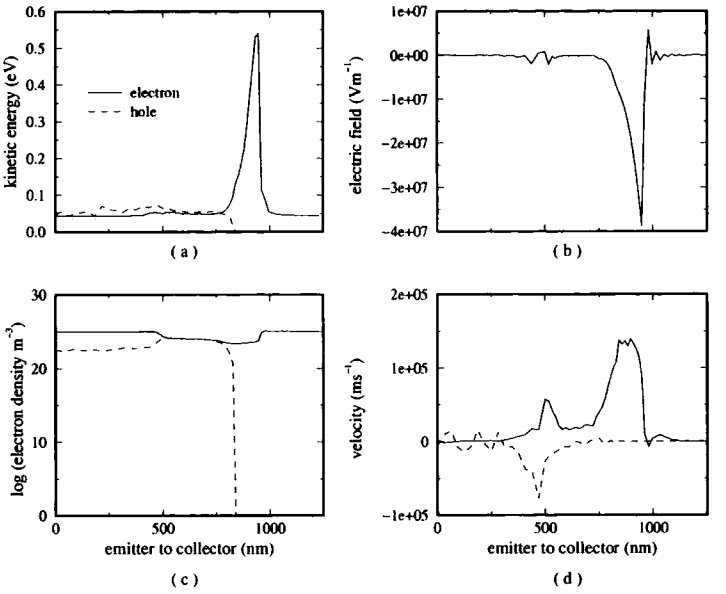
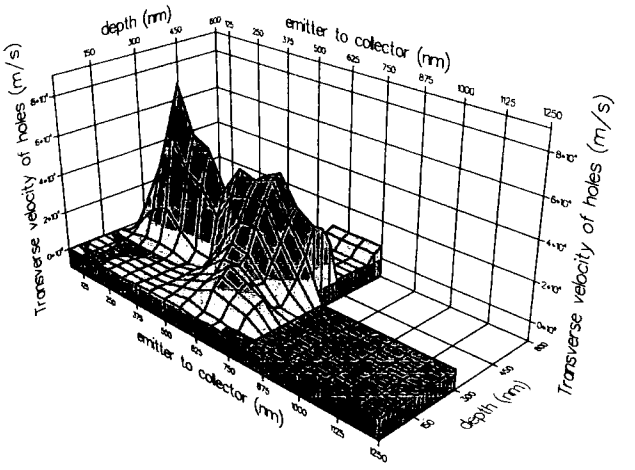


Figure 5.15: Microscopic details extracted from simulations of device 2 for V_{CE} = 3V, V_{BE} = 1.1 V. (a) kinetic energy (b) longitudinal electric field, (c) electron density, and (d) longitudinal velocity.

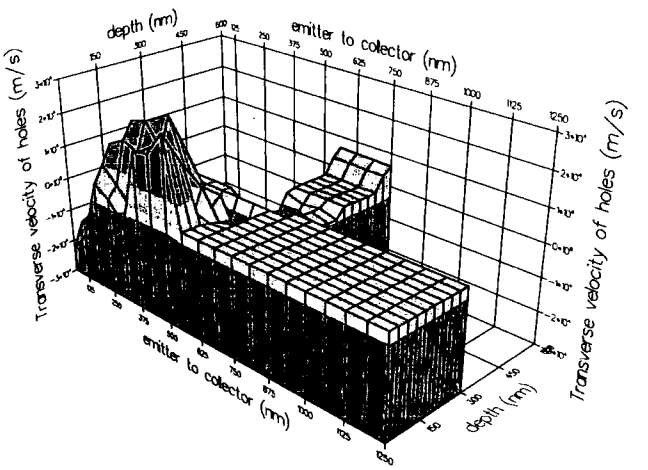
the dashed line shown in Figure 5.1) with zero bias and also with V_{BE} = 1.1 V and V_{CE} = 3 V. The electric fields due to the junction space charge regions are apparent and, as expected, around the forward biased emitter-base junction the electric field is smaller than in thermal equilibrium, whereas the reverse is the case for the reverse-biased base-collector junction.

The Δ valley band edge profile through device 2 in the unbiased state is illustrated in Figure 5.19. For normal operation the base-emitter junction is forward-biased and the barrier between the channel and the heavily doped part of the base pedestal is reduced. As a result electrons are injected into the base pedestal and holes are injected into the emitter. The plot of the Δ valley band profile through the SOI-LBJT (device 2) with V_{BE} = 0.9 V and V_{CE} = 1.5 V is illustrated in Figure 5.20.

In the following discussion the transistor is operated in the active mode with the emitter heavily forward biased, and the collector heavily reverse biased. Figure 5.21 shows microscopic details of electron transport extracted along the channel of device 2

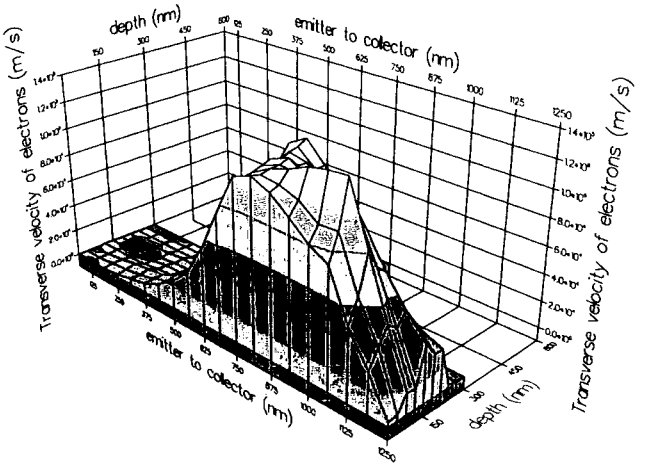


(a) Device 1

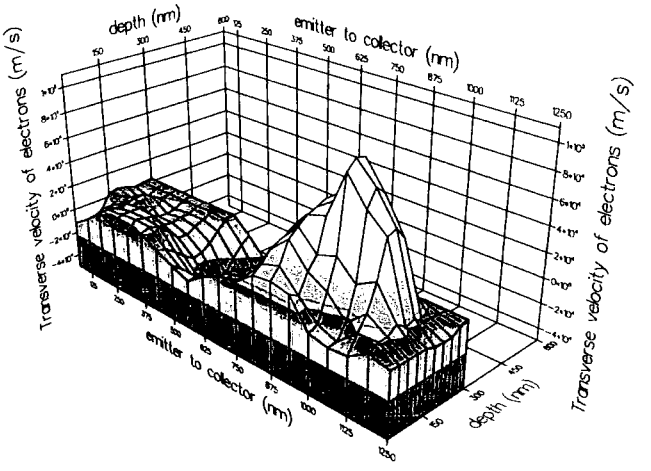


(b) Device 2

Figure 5.16: Hole transverse velocity as a function of position in device 1 (a) and device 2 (b). The collector bias is 3 V and the base bias is 0.7 V.



(a) Device 1



(b) Device 2

Figure 5.17: Electron transverse velocities as a function of position in device 1 (a) and device 2 (b). The collector bias is 3 V and the base bias is 0.7 V.

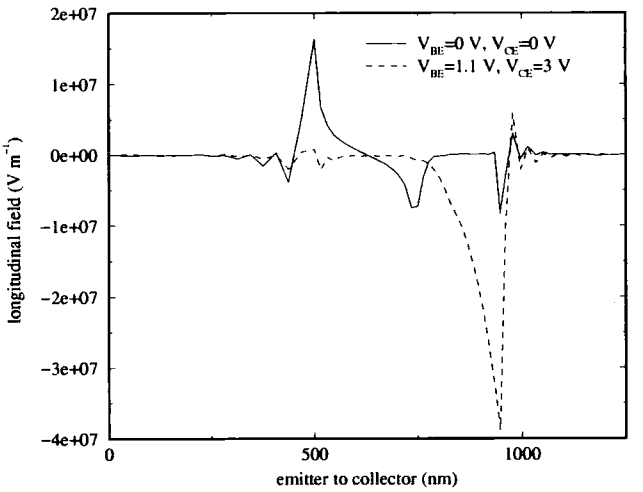


Figure 5.18: The electric field distribution along the channel (see dashed line in Figure 5.1) of device 2 at thermal equilibrium (solid line) and at $V_{BE} = 1.1$ V and $V_{CE} = 3$ V (dashed line).

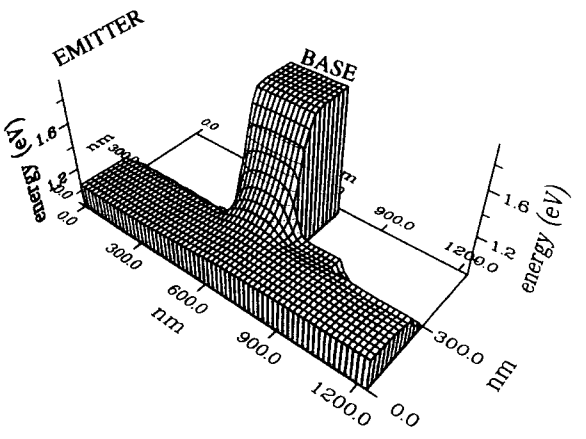


Figure 5.19: Contour plot showing the steady state Δ-valley band edge profile though the unbiased SOI-LBJT device 2. The polysilicon back gate beneath the buried oxide is grounded.

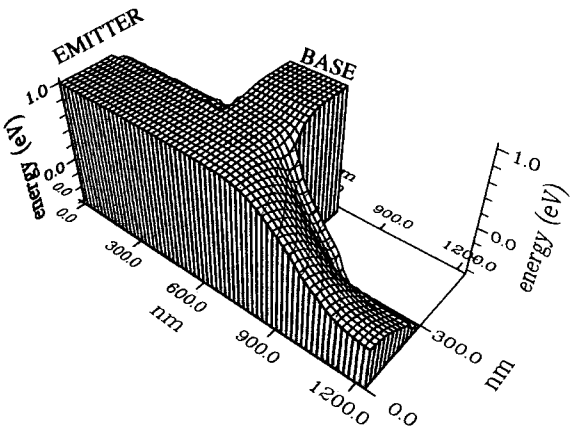


Figure 5.20: Contour plot showing the steady state Δ-valley band edge profile through the SOI-LBJT device 2 when the base-emitter and collector-emitter voltages are 0.9 V and 1.5 V, respectively

when the collector-emitter bias is 3 V and the base voltages are 0.5 (solid line) and 0.7 V (dashed line). Of interest here is how the transport properties change with varying bias. The kinetic energy of the electrons is depicted in Figure 5.21(a). Electrons that drift through part of the base (at 0.5 V bias) are heated to an average peak kinetic energy of 0.45 eV on reaching the field at the base-sub-collector junction as shown in Figure 5.21(b). It can be seen that increasing the base voltage to 0.7 V significantly reduces the kinetic energy of the electrons despite the relatively small reduction of the longitudinal electric field.

The electron density along the channel is illustrated in Figure 5.21(c). As expected, the electron density is essentially determined by the donor density in the emitter and collector regions. However, a significant electron density is found in the base and sub-collector regions. Due to the low injection of holes from the base pedestal to the channel at a base voltage of 0.5 V, the hole density in the channel part of the base is very low for this bias arrangement as it is shown in Figure 5.21(c). However, at the higher base voltage of 0.7 V, hole injection is stronger and the hole density is shown as the dotted line. Figure 5.21(d) shows the longitudinal velocity for electrons moving from the emitter to the collector. The peak average velocity at a base bias of 0.7 V is

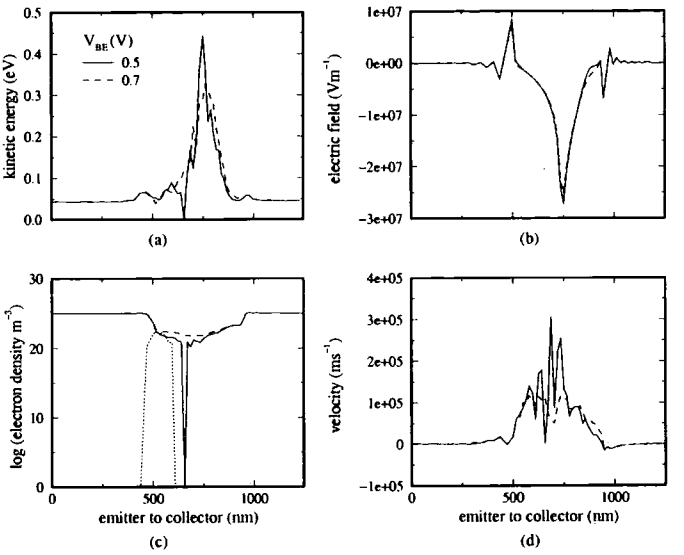


Figure 5.21: Microscopic details extracted from simulations for device 2. The data correspond to electron transport along the channel and are averaged over 28.75 nm thickness of the channel. Note that the recorded layer is 281.3 nm above the top of buried oxide layer. The collector-emitter bias is 3 V. The base voltages are 0.5 V (solid line) and 0.7 V (dashed line). (a) kinetic energy, (b) longitudinal electric field, (c) electron density and hole density (dotted line) at $V_{BE} = 0.7$ V, and (d) longitudinal velocity.

approximately the saturation velocity of electrons in bulk n-Si (10^5 ms^{-1}). At the lower base voltage the velocity reaches about 3×10^5 ms^{-1} in the base-sub-collector region, which corresponds to the position of the peak of the longitudinal electric field shown in Figure 5.21(b).

It was demonstrated in Section 5.3.2 that although device 1 provides a larger collector current than device 2 for the same bias arrangement, its cut-off frequency is rather lower. The base current of device 1 increases significantly for base voltages in excess of 0.5 V, substantially reducing the current gain and lowering the predicted cut-off frequency. Figure 5.22 shows microscopic details of the simulated electron transport along the channel of device 1. The base voltage is 0.5 (solid line) or 0.7 (dashed line) V and the collector-emitter bias is 3 V. From this figure, it is apparent that the transport properties of the device are similar to those of device 2 (see Figure 5.21). At the lower base bias 0.5 V, the hole density along the channel is negligible but a significant hole

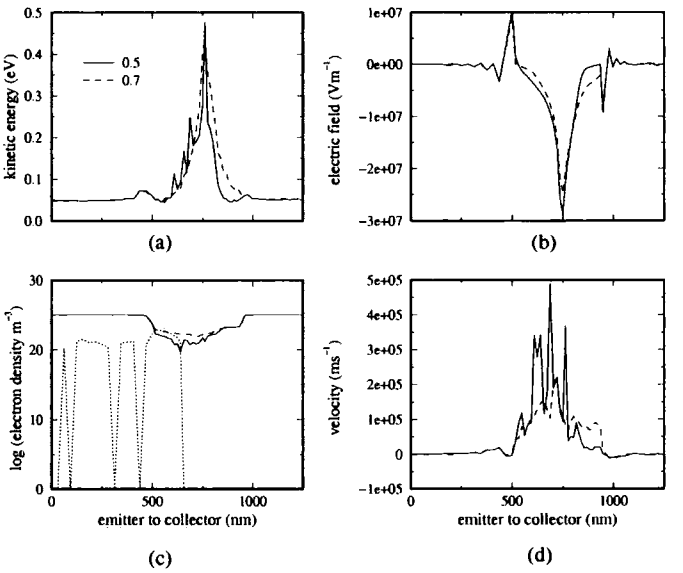


Figure 5.22: Microscopic details extracted from simulations. The data correspond to electron transport along the channel of device 1 with the base width of 0.25 μm. The collector-emitter bias is 3 V. The base voltages are 0.5 V (solid line) and 0.7 V (dashed line). (a) kinetic energy (b) longitudinal electric field, (c) electron density and hole density (dotted line) for V_{BE} = 0.7 V, and (d) longitudinal velocity.

density exists beneath the emitter and in the base part of the channel for a base bias of 0.7 V.

The microscopic details of electron transport recorded along the channel of device 3 are illustrated in Figure 5.23. The bias arrangement, chosen to obtain the maximum frequency response, is the same as for device 2. The carrier densities are shown in Figure 5.23(c). The electron density in much of the base part of the channel is negligible at a base bias of 0.5 V and holes exist only at the emitter-base junction. The absence of electrons means that zero values are recorded for their kinetic energy and drift velocity. After increasing the base bias to 0.7 V, stronger hole injection from the base pedestal is found and more holes reach the emitter region. Note that at this bias arrangement the electron density in the base region also increases significantly.

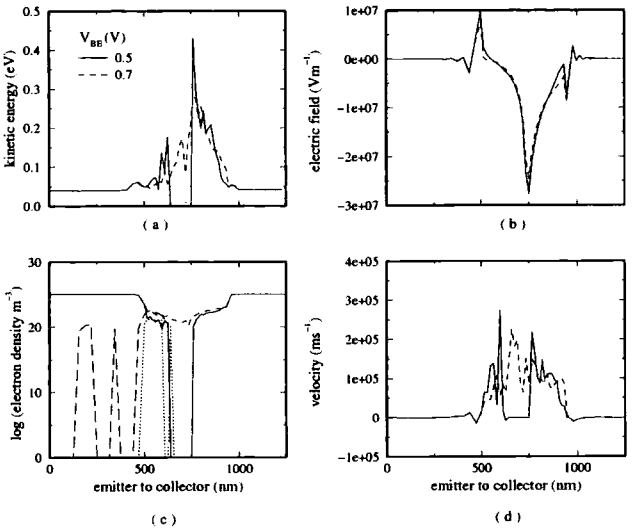


Figure 5.23: Microscopic details extracted from simulations. The data correspond to electron transport along the channel of device 3. The collector-emitter bias is 3 V. The base voltages are 0.5 V (solid line) and 0.7 V (dashed line). (a) kinetic energy, (b) longitudinal electric field, (c) electron density, and (d) longitudinal velocity. Note that in (c) the dotted line represents hole density at the base bias of 0.5 V whilst the long-dashed line corresponds to hole density at higher base voltage of 0.7 V.

5.4 Summary and Conclusion

In this chapter, we have described self-consistent ensemble Monte Carlo simulations that were used to model steady state and transient electron and hole transport in n-p-n silicon-on-insulator lateral bipolar junction transistors (n-p-n-SOI LBJTs). The simulated transistors are similar to those which have recently been the subject of experimental investigation. The effects of junction depth and silicon layer thickness were investigated. Macroscopic and microscopic device characteristics were extracted from the simulations, providing some insight into the carrier transport and transistor performance. Time-dependent signal analysis was carried out to test the response of some of the devices and derive their frequency bandwidths. For each device, a sinc voltage pulse was applied to the base and the resulting collector and base currents used to calculate the current gain as a function of frequency. The calculations predict

that the common-emitter current gain for the device 2 has a cut-off frequency of 35 $\pm$ 5 GHz, provided the shallow emitter and collector contact resistances are minimised.

It is apparent from our investigations that the buried oxide layer is critical for good LBJT performance as it prevents the current leakage through the p-Si substrate and associated low dc current amplification which is apparent in device 4. The thickness of Si layer is found to be an important factor. Device 1 has the highest collector current at all base-emitter biases considered because of its thicker Si layer and p-n-junction depth. It also has the highest common emitter current gain at low bias but this declines rapidly as the bias is increased. The decline can be attributed to a substantial increase in base current. Device 2 has a smaller junction depth and results in a dc gain which, although lower than device 1, is much less dependent on bias. Device 2 also exhibits the highest cut-off freqeuncy (35 ± 5 GHz) of all the devices considered. It is concluded that an important factor in limiting dc and high frequency device performances is the increase of base current with base-emitter bias and it would be worthwhile to consider ways of suppressing the current rise by the use of a heterostructure or other means.

Chapter 6

Summary, Conclusions and Suggestions for Further Works

6.1 Summary & conclusions

The main aim of the work presented in this thesis has been the study, using the self-consistent ensemble Monte Carlo method, of carrier transport in n-strained Si/SiGe MODFETs and npn-SOI LBJTs with the aim of gaining insight into the factors that determine device performance.

In Chapter 2 we explained the basic properties of Si and $Si_{1-x}Ge_x$ heterostructures. The effect of strain on the transport properties was discussed and the way to implement it in an effort to improve field-effect transistor performance was then described. In particular, we described how strain-induced band structure changes may lead to increased charge carrier mobility within the pseudomorphic layers of Si-SiGe field effect transistors. It was noted how mobility-enhanced field effect transistors based on Si-compatible SiGe technology can not only fit easily into the arena of mainstream microelectronics but can also provide a performance advantage that might ultimately be decisive, when down-scaling of critical device dimensions is no longer an affordable

option for bulk production of integrated circuits. The chapter was concluded with a brief description of the basic mode of operation of the LBJT.

To perform a thorough investigation of the transport properties of electrons (holes) inside a semiconductor device on a microscopic level, we have to be able to solve the Boltzmann transport equation (BTE), which provides a fundamental description of the semiclassical carrier transport. Our method of solution was by the ensemble Monte Carlo method, and its application to device simulation was briefly reviewed in Chapter 3.

In Chapter 4 we reported a detailed analysis by Monte Carlo simulation of the in-plane electron velocity characteristics of an unstrained layer of Si and of strained Si grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates. An enhancement of the average drift velocity was found when the Si was tensilely strained in the plane. The enhancement of velocity is significant at low and intermediate electric fields, but at very high fields the velocity saturates to about the same value as unstrained Si.

It is well known that the transit time of electrons passing under the gate is an important factor in determining the frequency response of field-effect transistors. Thus, to realise enhanced high-frequency operation, the gate length generally has to be reduced. However, for small gate lengths the electron transport does not reach a steady state and transient phenomena also become a significant factor in determining the frequency response. To investigate such effects, an ensemble Monte Carlo method was used to study the transient response to a stepped electric field of electrons in Si and strained Si grown on $Si_{0.77}Ge_{0.23}$ and $Si_{0.55}Ge_{0.45}$ virtual substrates. The calculations suggested that the mean electron velocity overshoots its saturated value. The peak velocity of electrons in strained Si grown on a $Si_{0.55}Ge_{0.45}$ virtual substrate reaches 2×10^5 ms^{-1} and 2.5×10^5 ms^{-1} at electric fields 6×10^6 Vm^{-1} and 2×10^7 Vm^{-1} respectively. The values for the same applied fields for strained Si grown on a $Si_{0.77}Ge_{0.23}$ virtual substrate are 1.6×10^5 ms^{-1} and 2.2×10^5 ms^{-1}.

Device simulations based on the self-consistent ensemble Monte Carlo method were also performed. The devices studied were strained $Si/Si_{1-x}Ge_x$ MODFETs with Ge fractions of 0.23, 0.25, and 0.45. For depletion mode n-channel $Si/Si_{0.77}Ge_{0.23}$ MODFETs, the effects of varying the gate length L_G and the source-gate distance L_{SG} were studied. Simulations of the effect of modulating the gate bias were also carried out to test the device response and derive the maximum intrinsic bandwidth. The highest intrinsic current gain cut-off frequency f_T^i of 80 $\pm$ 10 GHz was obtained for a gate length of 0.07 μm. Also, the simulations of depletion and enhancement n-channel $Si/Si_{0.55}Ge_{0.45}$ MODFETs with a gate length of 0.18 μm, whose geometries and doping were matched to those investigated experimentally by Glück *et al.* [73], showed fair agreement with the measured data. Predicted f_T^i of 60 $\pm$ 10 GHz and 90 $\pm$ 10 GHz were obtained.

Chapter 5 was devoted to the steady state and transient charge transport in a novel silicon-on-insulator lateral bipolar junction transistor. The modelled devices had lateral dimensions and doping densities close to those of the experimental device reported by Gómez *et al.* [94]. Four devices were studied in detail and the effects of junction depth and silicon layer thickness were investigated. Substantial advantages in growing the Si-bipolar junction transistor on a buried oxide layer were found in terms of higher collector current, higher current gain, and higher speed. In addition, a significant current leakage via the substrate, found in the conventional all-silicon structure and resulting in a low collector current, was eliminated. The common-emitter current gain of an SOI-LBJT with a 100 nm oxide layer, 300 nm Si layer, and 75 nm of junction depth, was predicted to have an intrinsic cut-off frequency f_T^i approaching 35 $\pm$ 5 GHz.

6.2 Suggestions for Further work

The work in this thesis has concentrated on applying ensemble self-consistent Monte Carlo device simulation to the study of carrier transport in n-channel strained Si/SiGe MODFETs and npn-SOI LBJTs. There are several ways in which further work could be done which would build on the progress to date.

Several new technologies and new device concepts have been developed in order to overcome the physical problems due to the miniaturisation of MOSFETs. As explained in Chapter 2 the strained Si MOSFET is a promising device for sub-0.1 μm high-speed and low-power CMOS, because of the high electron and hole mobilities that can be achieved. For n-channel devices, a relaxed $Si_{1-x}Ge_x$ buffer layer is required to ensure tensile strain and to provide the conduction band offset in Si/SiGe for the confinement of electrons. However dislocation defects are inevitably introduced in the epitaxial films, degrading device performance. In addition, the device structures normally are not completely compatible with the CMOS standard processes, requiring well formation and isolation at high temperature. An alternative is the use of silicon-on-insulator (SOI) devices. The advantage of using SOI is that parasitic capacitances can be significantly reduced, and isolation is straightforward. The unique properties of SOI allow low-power and low-voltage operation to be improved [103]. It is expected that SOI-MOSFETs using thin film SOI substrates with low channel dopant could extend the capabilities of high-speed, sub-100 nm CMOS. Devices that take advantage of carrier transport enhancement due to strain and of using SOI are very attractive as they might provide an alternative way to avoid the huge investment required for sub 0.1 μm photolithography system development.

Recently, a new strained Si MOSFET based on strained Si/SiGe-on insulator (strained-SOI) substrates has been reported by Mizuno *et al.* [104]. The demonstrated devices provide enhanced drain current compared with SOI-MOSFETs. The electron and hole mobilities of the strained-SOI MOSFETs are higher by about 62% and 5% than the

conventional bulk Si MOSFETs. The potential impact of new methods of achieving strain-induced device performance enhancement should be explored. In addition, high-field and transient transport properties are expected to dominate the characteristics of deep-submicron transistors. To this end, investigation by self-consistent ensemble Monte Carlo simulation is the appropriate tool to provide a detailed understanding of carrier transport in new device designs.

Apart from simulating new devices as mentioned above, it would also be interesting to include some additional physical mechanisms into the Monte Carlo transport model to provide a more complete description of carrier transport. For example, in the SOI LBJT device simulations, the base current includes the current caused by the injection of minority-carriers holes into the emitter. Although this is the dominant source of base current under most conditions of bias for the npn transistor, a more accurate model of the device would include the current due to the recombination of injected electrons in the base, and of injected electrons in the emitter-base depletion region which can be important at low emitter currents. A possible extension of the simulation would be to consider these recombination processes. In addition, the recombination process between electrons and holes at the emitter and SiO_2 interface could be important to determine the emitter current and this surface recombination might also be included in the simulation.

References

[1] Taur Y., Buchanan D. A., Chen W., Frank D. J., Ismail K. E., Lo S., Sai-Halasz G. A., Viswanathan R. G., Wan H. C., Wind S. J., and Wong H. S, "CMOS scaling into nanometer regime," *IEEE Proc.* **85(4)**, pp. 486–504, 1997.

[2] Williams S. C., Kim K. W., and Holton W. C., "Ensemble Monte Carlo study of channel quantisation in a 25-nm n-MOSFET," *IEEE Trans Electron Devices* **47(10)**, pp. 1864–1872, 2000.

[3] Nelson S. F., Ismail K., Saenger K. L., Chu J. O., and Myerson B. S., "Room-temperature electron-mobility in strained Si/SiGe heterostructures," *Appl. Phys. Lett* **63(3)**, pp. 367–369, 1993.

[4] Whall T. E., Smith D. W., Plews A. D., Kubiak R. A., Phillips P. J., and Parker E.H. C., "High hole mobilities in a p-type modulation-doped $Si/Si_{0.87}Ge_{0.13}/Si$ heterostructure," *Semicond. Sci. Technol.* **8(3)**, pp. 615–616, 1993.

[5] Lie D.Y.C and Wang K. L., "Si/SiGe heterostructures for Si-based nanoelectronics," in *Handbook of advanced electronic and photonic materials and devices: Vol. 2 (semiconductor devices)*, Nalwa H.S, Ed., chapter 1. Academic press, 2001.

[6] Braunstein R., Moore A. R., and Herman F., "Intrinsic otical absorption in germanium-silicon alloys," *Phys. Rev.* **109**, pp. 695–710, 1958.

[7] Frank F. C. and van der Mervwe J. H., "One dimensional dislocations II. Misfitting monolayers and oriented overgrowth," *Proc.R. Soc. A* **198**, pp. 216–219, 1949.

[8] Matthews J. W. and Blakeslee A. E., "Defects in epitaxial multilayers," *J. Crystal Growth* **27**, pp. 118–125, 1974.

[9] Jain S. C., Decoutre S., Willander M., and Maes H. E., "SiGe HBTs for application in BiCMOS technology: I. stability, reliability and material parameters," *Semicond. Sci. Technolo.* **16(6)**, pp. R51–R65, 2001.

[10] Kittel C., *Introduction of solid state physics*, Wiley, seventh edition, 1996.

[11] Bean J. C., Feldman L. C., Fiory A. T., Nakahara S., and Robinson I. K., "Ge_xSi_{1-x}/Si strained-layer superlattice grown by molecular-beam epitaxy," *J. Vac. Sci. Technol. A* **2(2)**, pp. 436–440, 1984.

[12] Van de Walle C. G. and Martin R. M., "Theoretical calculations of heterojunction discontinuities in the Si/Ge system," *Phys. Rev. B* **34(8)**, pp. 5621–5634, 1986.

[13] Van de Walle C. G., "Strain effects on the conduction-band structure of SiGe," in *Properties of strained and relaxed silicon germanium emis Data Review Series 12*, Kasper E, Ed., chapter 4.5, pp. 99–102. INSPEC, IEE, London, 1995.

[14] Rieger M. and Vogl P., "Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," *Phys. Rev. B* **48(19)**, pp. 14276–14287, 1993.

[15] Hasegawa H., "Theory of cyclotron resonance in strained silicon crystals," *Phys. Rev* **129(3)**, pp. 1029–1040, 1963.

[16] Nötzel J. F., Engelhardt C. M., and Abstreiter G., "Effective masses in SiGe," in *Properties of strained and relaxed silicon germanium emis Data Review Series 12*, Kasper E., Ed., pp. 103–109. INSPEC, IEE, London, 1995.

[17] Schubert G., Schäffler F., Besson M., Abstreiter G., and Gornik E., "High electron mobility in modulation-doped Si/SiGe quantum well structures," *Appl. Phys. Lett.* **59(25)**, pp. 3318–3320, 1991.

[18] Ismail K., Arafa M., Saenger K. L., Chu J. O., and Meyerson B. S., "Extremely high-mobility in Si/SiGe modulation-doped heterostructures," *Appl. Phys. Lett* **66(9)**, pp. 1077–1079, 1995.

[19] Schäffler F., "Review article : High-mobility Si and Ge structure," *Semicond. Sci. Technol.* **12(12)**, pp. 1515–1549, 1997.

[20] Schäffler F., "Si/$Si_{1-x}Ge_x$ and Si/$Si_{1-y}C_y$ heterostructures: materials for high speed field-effect transistors," *Thin solid films* **321**, pp. 1–10, 1998.

[21] People R. and Bean J. C., "Band alignments of coherently strained Ge_xSi_{1-x}/Si heterostructures on $< 001 >$ Ge_ySi_{1-y} substrates," *Appl. Phys. Letter* **48(8)**, pp. 538–540, 1986.

[22] Fischetti M. V. and Laux S. E., "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Phys. Rev. B.* **38(14)**, pp. 9721–9745, 1988.

[23] Crabbe E., Meyerson B., Harame D., Stork J., Megdanis A., Cotte J., Chu J., Gilber M., Stanis C., Comfort J., Patton G., and Subbanna S., "113-GHz f_t graded SiGe HBT's," *IEEE Electron Devices* **40(11)**, pp. 2100–2101, 1993.

[24] Lilienfeld J. E., "Method and apparatus for controlling electric currents," *US patent 1,745,175*, 1930.

[25] Lilienfeld J. E., "Device for controlling electric currents," *US patent 1,900,018*, 1933.

[26] Shur M. and Fjeldly T. A., "Compound-Semiconductor Field-Effect Transistor," in *Modern Semiconductor Device Physics*, Sze S.M., Ed., chapter 2. Addison-Wiley, 1998.

[27] Stern F. and Howard W. E., "Properties of semiconductor inversion layers in the electric quantum limit," *Phys. Rev.* **163**, pp. 816–835, 1967.

[28] Esaki L. and Tsu R., "Superlattice and negative differential conductivity in semiconductors," *IBM Journal of Research and Development* **14(1)**, pp. 61–65, 1970.

[29] Dingle K., Störmer H. L., Gossard A. C., and Wiegmann W., "Electron mobilities in modulation-doped semiconductor heterojunction superlattices," *Appl. Phys. Lett.* **33**, pp. 665–667, 1978.

[30] Mimura T., Hiyamizu S., Fujii, and Nambu K., "A new field effect transistor with selectively doped GaAs/n-$Al_xGa_{1-x}As$ heterostructures," *Jpn. J. Appl. Phys.* **19(5)**, pp. L225–L227, 1980.

[31] Meyerson B. S., "UHV/CVD growth of Si and Si:Ge alloys: chemistry, physics, and device applications," *Proc. IEEE* **80(10)**, pp. 1592–1608, 1992.

[32] Hackbarth G., Höck G., Herzog H. J., and Zeuner M., "Strain relieved SiGe buffer for Si-based heterostructure field effect transistors," *J. Cryst. Growth* **201**, pp. 734–738, 1999.

[33] Sze S. M., *Physics of Semiconductor Devices*, Wiley, New York, second edition, 1981.

[34] Daenbkes H., Herzog H. J., Jorke H., Kibbel H., and Kasper E., "The n-channel SiGe/Si modulation-doped field-effect transistor," *IEEE Electron device* **33(5)**, pp. 633–638, 1986.

[35] Ismail K., Meyerson B. S., Rishton S., Chu J., Nelson S., and Nocera J., "High-transconductance n-type Si/SiGe modulation-doped field-effect transistor," *IEEE Electron Device Lett.* **13(5)**, pp. 229–231, 1992.

[36] König U., Boers A. J., Schäffler F., and Kasper E., "Enhancement mode n-channel Si/SiGe MODFET with high intrinsic transconductance," *Electronic Lett.* **28(2)**, pp. 160–162, 1992.

[37] Koester S. J., Chu J. O., and Groves R. A., "High-f_t n-MODFETs fabricated on Si/SiGe heterostructures grown by UHV-CVD," *Electronics Letters* **35(1)**, pp. 86–87, 1999.

[38] Monroe D. Xie Y. H., Fitzgerald E. A., Silverman P. J., Thiel F. A., and Watson G. P., "Very high mobility two-dimensional hole gas in Si/Ge_xSi_{1-x}/Ge structures grown by molecular beam epitaxy," *Appl. Phys. Lett.* **63(16)**, pp. 2263–2264, 1993.

[39] Sturm J. C., McVittie J. P., Gibbons J. F., and Pfeiffer L., "A lateral silicon-on-insulator bipolar transistor with a self-aligned base contact," *IEEE Electron Device Lett.* **8(3)**, pp. 104–106, 1987.

[40] Ng K. K., *Complete Guide To Semiconductor Devices*, McGraw-Hill, 1995.

[41] Fawcett W., Boardman D. A, and Swain S., "Monte Carlo determination of electron transport properties in gallium arsenide," *J. Phys. Chem. Solids* **31**, pp. 1963–1990, 1970.

[42] Jacoboni C. and Reggiani L., "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.* **55(3)**, pp. 645–705, 1983.

[43] Jacoboni C. and Lugli P., *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, 1989.

[44] Moglestue C., *Monte Carlo Simulation of Semiconductor Devices*, Chapman & Hall, 1993.

[45] Lundstrom M., *Fundamentals of Carrier Transport*, Cambridge University Press, second edition, 2000.

[46] Rode D. L., "Low-Field Electron Transport," in *Semiconductors and Semimetals Vol. 10*, Willardson R.K. and Beer A.C., Eds., pp. 1–89. Academic Press, New York, 1975.

[47] Conwell E. M. and Vassel M. O., "High-field transport in n-type GaAs," *Phys. Rev.* **166(3)**, pp. 797–821, 1968.

[48] Herring C. and Vogt E., "Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering," *Phys. Rev* **101**, pp. 944–961, 1956.

[49] Ridley B. K., *Quantum Processes in Semiconductors*, Clarendon Press, 1982.

[50] Nag B., *Electron Transport in Compound Semiconductors*, Springer-Verlag, New York, 1980.

[51] Briggs P. J., Walker A. B., and Herbert D. C., "Calculation of hole mobilities in relaxed and strained SiGe by monte carlo simulation," *Semicond. Sci. Technol.* **13(7)**, pp. 680–691, 1998.

[52] Wiley J. D., "Mobility of holes in III-V compounds," in *Semiconductors and Semimetals Vol. 10*, pp. 91–174. Academic Press, New York, 1975.

[53] Singh J., *Physics of Semiconductors and Their Heterostructures*, chapter 8, McGraw-Hill, 1993.

[54] Tomizawa K., *Numerical Simulation of Submicron Devices*, chapter 2, Artech House, 1993.

[55] Bude J., "Scattering mechanisms for semiconductor transport calculations," in *Monte Carlo Device Simulation: Full Band and Beyond*, Hess K., Ed., pp. 27–66. Kluwer Academic Publisher, 1991.

[56] Crow G. C. and Abram R. A., "Performance predictions for a silicon velocity modulation transistor," *J. Appl. Phys.* **85(2)**, pp. 1196–1202, 1999.

[57] Brooks H., "Scattering by ionized impurities in semiconductors," *Phys. Rev.* **83(4)**, p. 879, 1951.

[58] Harrison J. W. and Hauser J. R., "Alloy scattering in ternary III-V compounds," *Phys. Rev. B* **13(12)**, pp. 5347–5350, 1976.

[59] Crow G. C. and Abram R. A., "Monte Carlo simulations of carrier transport in AlGaInP laser diodes," *IEEE Q.E.* **33(9)**, pp. 1551–1556, 1997.

[60] Hoare D. and Abram R. A., "Monte Carlo simulation of PHEMTs operating up to terahertz frequencies," *Int. J. Electronics*, pp. 429–439, 1997.

[61] Walmsley M. and Abram R. A., "A fast Poisson solver for realistic semiconductor device structures," *J. Comput. Math. Elec. Electron. Eng.* **15(2)**, pp. 31–52, 1996.

[62] Rees H. D., "Calculation of distribution functions by exploiting the stability of the steady state," *J. Phys. Chem. Solids.* **30**, pp. 643–655, 1969.

[63] Fischetti M. V. and Laux S. E., "Band structure, deformation potentials, and carrier mobility in strained Si, Ge and SiGe alloys," *J. Appl. Phys.* **80(4)**, pp. 2234–2252, 1996.

[64] Landolt-Börnstein, "Numerical Data and Functional Relationships in Science and Technology, **17**," in *Semiconductors subvol. d*, M. Schulz O. Madelung and H. Weiss, Eds. Springer, Berlin, 1984.

[65] Kunikiyo T., Takenaka M., Kamakura Y., Yamaji M., Mizuno H., and Morifuji M., "Monte Carlo simulation of anisotropic electron transport in silicon using full band structure and anisotropic impact-ionization model," *J. Appl. Phys.* **75**, pp. 297–312, 1994.

[66] Bufler F. M., Graf P., Keith S., and Meinerzhagen, "Full band Monte Carlo investigation of electron transport in strained Si grown on $Si_{1-x}Ge_x$ substrates," *Apply. Phys. Lett* **70(16)**, pp. 2144–2146, 1997.

[67] Abramo A., Baudry L., Brunetti R., Casagne R., Charef M., Dessenne C. F., Dollfus P., Dutton R., Engl W. L., Fauquembergue R., Fiegna C., Fischetti M. V., Galdin S., Goldsman N., Jungemann C., Kamakura Y., Kosina H., Kunikiyo T., Laux S. E., Lin H., Maziar C., Mizuno H., Peifer H. J., Ramaswamy S., Sano N., Scrobohaci P. G., Selberherr S., Takenaka M., Tang T. W., Taniguchi K., Thobel J. L., Thoma R., Tomizawa K., Tomizawa M., Vogelsang T., L. Wang S, Wang X., S Yao C, Yoder P. D., and Yoshi A., "A comparison of numerical solutions of the Boltzmann transport equation for high-energy electron transport silicon," *IEEE Trans. Electron. Dev.* **41(9)**, pp. 1646–1654, 1994.

[68] Formicone G. F., Vasileska D., and Ferry D. K., "2D Monte Carlo simulation of hole and electron transport in strained Si," *VLSI Design* **6(4)**, pp. 167–171, 1998.

[69] Formicone G. F., Vasileska D., and Ferry D. K., "Modeling of submicron $Si_{1-x}Ge_x$-based MOSFETs by self-consistent Monte Carlo," *Physica Stat. Solidi. B* **204(1)**, pp. 531–533, 1997.

[70] Dollfus P., "$Si/Si_{1-x}Ge_x$ heterostructures: electron transport and field-effect transistor operation using Monte Carlo simulation," *J. Appl. Phys.* **82(8)**, pp. 3911–3916, 1997.

[71] Dollfus P., "Study of non-stationary transport in $Si/Si_{0.7}Ge_{0.3}$ n-MODFET," *Physica Status Solidi B-Basic Research* **204(1)**, pp. 541–544, 1997.

[72] Churchill A. C., Robbins D. J., Wallis D. J., Griffin N., Paul D. J., and Pidduck A. J., "High-mobility two-dimensional electron gases in Si/SiGe heterostructures on relaxed SiGe layers grown at high temperature," *Semicond. Sci. Technol.* **12(8)**, pp. 943–946, 1997.

[73] Glück M., Hackbath T., Birk M., Haas A., Khon E., and König U., "Design and fabrication of Si/SiGe n-type MODFETs," *Physica E.* **2**, pp. 763–767, 1998.

[74] Sadek A. and Ismail K., "Si/SiGe CMOS possibilities," *Solid-State Electronics* **38(9)**, pp. 1731–1734, 1995.

[75] Vogelsang Th. and Hofmann K. R., "Electron transport in strained Si layers on $Si_{1-x}Ge_x$ substrates.," *Appl. Phys, Lett* **63(2)**, pp. 186–188, 1993.

[76] Miyata H., Yamada T., and Ferry D. K., "Electron transport properties of a strained Si layer on a relaxed $Si_{1-x}Ge_x$ substrate by Monte Carlo simulation," *Appl. Phys. Lett.* **62(21)**, pp. 2661–2663, 1993.

[77] Rashed M., Shih W. K., Jallepalli S., Zaman R., Kwan T.J.T, and Maziar C. M., "A Monte Carlo study of electron transport in strained Si/SiGe heterostructures," *VLSI Design* **6(1-4)**, pp. 213–216, 1998.

[78] Fischer B. and Hofmann K. R., "Full-band Monte Carlo model of electron and hole transport in strained Si including inelastic acoustic phonon scattering," *Appl. Phys. Lett* **74(15)**, pp. 2185–2187, 1999.

[79] Jain S. C., *Germanium - Silicon Strained Layers and Heterostructures*, chapter 8, Academic Press, 1994.

[80] Nayak D. K. and Shiraki Y., "Strain adjustment for n-MODFETs: SiGe unstrained, Si strained (1.5%)," in *Properties of strained and relaxed silicon germanium emis Data Review Series 12*, Kasper E., Ed., pp. 205–211. INSPEC, IEE, London, 1995.

[81] König U., Glück M., Gruhle A., Höck G., Kohn E., Bozon., Nuernbergk D., Ostermann T., and Hagelauer R., "Design rules for n-type SiGe hetero FETs," *Sol. Stat. Elec.* **41(10)**, pp. 1541–1547, 1997.

[82] Ismail K., Meyerson B. S., Chu J, Nelson S., and Nocera J., "High-transconductance n-type Si/SiGe modulation-doped field effect transistors," *IEEE Electron Dev. Lett.* **13**, pp. 229–231, 1992.

[83] Dollfus P., "Monte Carlo simulation of pseudomorphic InGaAs/GaAs high electron mobility transistors: physical limitations at ultrashort gate length," *J. Appl. Phys.* **73(2)**, pp. 804–809, 1993.

[84] Bergland G. D., "A guided tour of the fast Fourier transform," *IEEE Spectrum, Jul.*, pp. 41–52, 1969.

[85] Starikov E., Shiktorov P., Gruzinskis V., Gonzalez T., Martin M. J., Pardo D., Reggiani L., and Varani L., "Hydrodynamic and Monte Carlo simulation of steady-state transport and noise in submicrometer n^+nn^+ silicon structures," *Semicond. Sci. Technol* **11**, pp. 865–872, 1996.

[86] Ramaswamy S. and Tang T., "Comparison of semiconductor transport models using a Monte Carlo consistency test," *IEEE Tran. Electron Devices.* **401**, pp. 76–83, 1994.

[87] Ismail K., Rishton S., Chu J.O., Chan K., and Meyerson B. S., "Electron transport properties of Si/SiGe heterostructures: Measurement and device implications," *Appl. Phys. Lett.* **63(5)**, pp. 660–662, 1993.

[88] König U. and Glück M., "Si/SiGe field-effect transistors.," *J. Vac. Sci. Technol. B* **16(5)**, pp. 2609–2614, 1998.

[89] Crow G. C., Abram R. A., and Yangthaisong A., "Monte Carlo simulations of SiGe n-MODFETs with high tensile strained Si channels," *Semicond. Sci. Technol.* **15(7)**, pp. 770–775, 2000.

[90] Ashburn P. and Morgan D., "Heterojunction bipolar transistor," in *Physics and technology of heterojunction devices*, Morgan D. and Williams R., Eds., chapter 6. IEE, 1991.

[91] Coligne J. P., "A SOI voltage controlled bipolar-MOS device," *IEEE Trans. Electron Devices* **34(4)**, pp. 845–849, 1987.

[92] Cristoloveanu S. and Sheng S. Li, *Electrical characterization of silicon-on-insulator materials and devices*, Kluwer academic, 1995.

[93] Vasudev P. K., Mendicino M., and Seidel T. E., "Advanced materials for low power electronics," *Solid-State Electronics* **39(4)**, pp. 489–497, 1996.

[94] Gómez R., Bashir R., and Neudeck G. W., "On the design and fabrication of novel lateral bipolar transistor in a deep submicron technology," *Microelectronics J.* **(31)**, pp. 199–205, 2000.

[95] Tsaur B. Y., Silversmith D. J., Fan J.C. C., and Mountain R. W., "Fully isolated lateral bipolar-MOS transistors fabricated in zone-melting-recrystallized Si films on SiO_2," *IEEE Electron Device Lett.* **4(8)**, pp. 269–271, 1983.

[96] Rodder A. and Antioniadis D. A., "Silicon-on-insulator bipolar transistors," *IEEE Electron. Device Lett.* **4(6)**, pp. 193–195, 1985.

[97] Vittoz E., "MOS transistors operated in the lateral bipolar mode and their

application in CMOS technology," *IEEE J Solid-State Cir.* **18(3)**, pp. 273–279, 1983.

[98] Degrauwe M.G.R., Leuthold O. N., Vittoz E. A., Orguey H. J., and Descombes A., "Voltage references using lateral bipolar transistors," *IEEE J Solid-State Cir* **20(6)**, pp. 1151–1157, 1985.

[99] Warner Jr. R. M. and Schrimpf R., "BJT-MOSFET transconductance comparisons," *IEEE Electron Devices* **ED-34(5)**, pp. 1061–1065, 1987.

[100] Colinge J. P., "Half-micrometer-base lateral bipolar transistors made in thins silicon-on-insulator films," *Electronics Letter* **22(17)**, pp. 886–887, 1986.

[101] Huang W. M., "TFSOI complementay BiCMOS technology for low-power applications," *IEEE Trans Electrons Dev.* **42(3)**, pp. 506–512, 1995.

[102] Yangthaisong A., Crow G. C., and Abram R. A., "Monte Carlo simulations of submicron SOI-LBJTs," *IEEE 2001 Topical Meeting on Silicon Monolithic Integrated Circuits in RF System Dig., Ann Arbor, MI*, 2001.

[103] Coligne J. P., "SOI and three-dimensional structures," in *ULSI devices*, Chang C.Y. and Sze S.M., Eds., chapter 5. Wiley, 2000.

[104] Mizuno T., Takagi S., Sugiyama N., Satake H., Kurobe, and Toriumi A., "Electron and hole mobility enhancement in strained-Si MOSFET's on SiGe-on-insulator substrates fabricated by SIMOX technology," *IEEE Elec. Dev. Lett.* **21(5)**, pp. 230–232, 2000.