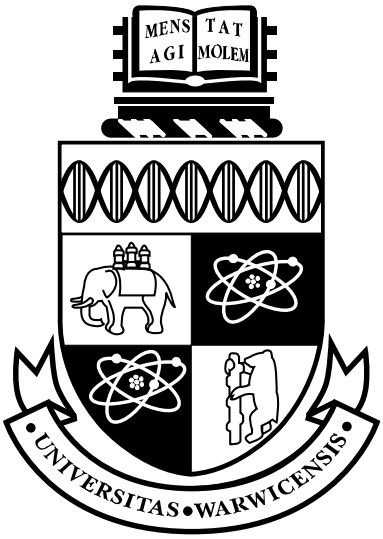


Investigation of the Electrical Properties of $Si_{1-x}Ge_x$ Channel pMOSFETs with High-κ Dielectrics

by

Andrew Dobbie

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Physics

September 2007

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to thank Dr. David Leadley, Prof. Evan Parker and Prof. Terry Whall for their support during this work. Terry deserves special praise for the number of hours spent proof-reading this thesis despite "retiring" a year ago! Secondly, I have to thank both past and present members of the Nano-Silicon Group for providing what was, on the whole, an enjoyable place to work for the last few years. In particular, Dr. Gareth Nicholas for help with both the Ge and high-field work, Chris Beer for his efforts in helping to set up the low temperature probe station, and Steve Thomas for additional electrical characterisation work deserve special mention. This work would not have been possible without the collaboratory efforts involving several institutions across Europe, and naturally I thank my working colleagues at IMEC, KTH, Chalmers University, AMO and LETI-CEA for providing the samples necessary for the work in this thesis. Finally, I wish to thank my friends and family who are too numerous to mention individually for all their support during the last few years. Finally, I thank my dad, Jim, and my mum, Yvonne for all their love and support! Thank-you!

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has not been submitted for a degree at any other University. Except where specifically stated, all of the work described in this thesis was carried out by the author or under his direction in the Department of Physics at the University of Warwick from September 2003 until September 2007.

Publications

The following publications and presentations contain material included in this thesis:

A. Dobbie, S. M. Thomas, D. J. Fulgoni, B. Raeissi, O. Engstrom, M. Schmidt, M. C. Lemme, P.-E. Hellstrom, E. H. C. Parker, D. R. Leadley and T. E. Whall, 'Electrical Characterisation of $Si_{1-x}Ge_x$ pMOSFETs using Sputtered HfO_2 Dielectric and W Metal Gate,' to be submitted.

A. Dobbie, G. Nicholas, M. Meuris, E. H. C. Parker, D. R. Leadley and T. E. Whall, 'Low Temperature Electrical Characterisation and Analysis on the Performance of Deep Submicron Germanium pMOSFETs,' to be submitted.

A. Dobbie, G. Nicholas, M. Meuris, E. H. C. Parker, and T. E. Whall, 'Low Temperature Performance of Deep Submicron Germanium pMOSFETs,' Proceedings of 2007 International Workshop on Electron Devices and Semiconductor Technology, Beijing, China, June 2007.

G. Nicholas, A. Dobbie, T. J. Grasby, T. E. Whall and E. H. C. Parker, 'Impact Ionisation in Strained SiGe pMOSFETs,' Electronics Lett., 41(16), pp925-927, 2005

Abstract

It is now apparent that the continued performance enhancements of silicon metal-oxide-semiconductor field effect transistors (MOSFETs) can no longer be met by scaling alone. High-mobility channel materials such as strained $Si_{1-x}Ge_x$ and Ge are now being seriously considered to maintain the performance requirements specified by the semiconductor industry. In addition, alternative gate dielectric, or high-κ dielectrics, will also be required to meet gate leakage requirements.

This work investigates the properties of using strained $Si_{1-x}Ge_x$ or Ge as alternative channel materials for pMOSFETs incorporating hafnium oxide (HfO_2) high-κ gate dielectric. Whilst the SiGe pMOSFETs ($x = 0.25$) exhibited an enhancement in hole mobility (300 K) over comparable silicon control pMOSFETs with sputtered HfO_2 dielectric, high Coulomb scattering and surface roughness scattering relating to the dielectric deposition process meant that the effective hole mobilities were degraded with respect to the silicon universal curve.

Germanium channel pMOSFETs with halo-doping and HfO_2 gate dielectric deposited by atomic layer deposition showed high hole mobilities of 230 $cm^2V^{-1}s^{-1}$ and 480 $cm^2V^{-1}s^{-1}$ at room temperature and 77 K, respectively. Analysis of the off-state current for the Ge pMOSFETs over a range of temperatures indicated that band-to-band tunnelling, gate-induced drain leakage and other defect-assisted leakage mechanisms could all be important.

Hole carrier velocity and impact ionisation were also studied in two batches of buried channel SiGe pMOSFETs with $x = 0.15$ and $x = 0.36$, respectively. SiGe channel pMOSFETs were found to exhibit reduced impact ionisation compared to silicon control devices, which has been attributed to a strain-induced reduction of the density of states in the SiGe conduction and valence bands. Analysis of the hole carrier velocity indicated that pseudomorphic SiGe offered no performance enhancements over Si below 100 nm, possibly due to higher ion implantation damage and strain relaxation of the strained SiGe channel. The results indicate that velocity overshoot effects might not provide the performance improvements at short channel lengths that was previously hoped for.

Abbreviations

A Area

ALD Atomic layer deposition

BTBT Band-to-band-tunnelling

C Capacitance

C_{dep} Depletion capacitance

C_{gb} Gate-body capacitance

C_{gc} Gate-channel capacitance

C_{hf} High-frequency capacitance

C_{it} Interface trap capacitance

C_{lf} Low-frequency capacitance

C_m Measured capacitance

C_{ox} Oxide capacitance

C_p Parallel equivalent circuit capacitance

C_s Series equivalent circuit capacitance

CET Capacitive equivalent thickness

CMOS Complimentary metal-oxide-semiconductor

CVD Chemical vapour deposition

D_p Einstein diffusion coefficient for holes

DIBL Drain-induced barrier lowering

D_{it} Interface trap density

DUT Device under test

E Energy

E_a Activation energy

E_c Conduction band edge energy

E_F Fermi level

E_i Intrinsic Fermi level

E_v Valence band edge energy

EOT Equivalent oxide thickness

ϵ_0 Permittivity of free space (= 8.85 x 10^{-14} F/cm)

ε Electric field

ε_{eff} Effective vertical field

f Frequency

FOX Field oxide

ϕ_B Barrier height

ϕ_{ms} Metal-semiconductor work function difference

g_m Transconductance

g_{mi} Intrinsic transconductance

G Conductance

G_m Measured conductance

G_p Parallel equivalent circuit conductance

GIDL Gate-induced drain leakage

HBT Heterojunction bipolar transistor

HRTEM High-resolution transmission electron microscopy

h Planck's constant (= 6.63 x 10^{-34} Js)

I Current

I_b Body current

I_{ds} Drain current

$I_{ds,sat}$ Saturation drain current

I_{off} Off-state current

I_{on} On-state current

I_s Source current

ITRS International Technology Roadmap for Semiconductors

J Current density

J_{DT} Current density due to direct tunnelling

k Wave vector

k_B Boltzmann's constant (=1.38 x 10^{-23})

κ Relative dielectric constant

L Gate length

L_D Debye length

L_{eff} Effective channel length

LPCVD Low pressure chemical vapour deposition

m_0 Free electron mass (= 9.1 x 10^{-31} kg)

m^* Effective mass

m^*_{ij} Effective mass in tensor form

M Avalanche multiplication factor

MOS Metal-oxide-semiconductor

MOSFET Metal-oxide-semiconductor field-effect-transistor

μ Mobility

μ_{eff} Effective mobility

n Electron density

N_D Doping density

N_{hys} Hysteresis charge density

N_s Inversion sheet density

PDA Post-oxide deposition anneal

PMA Post-metallisation anneal

q Electron charge (= 1.6 x 10^{-19} C)

Q_{dep} Depletion charge density

Q_{inv} Inversion charge density

Q_{ox} Oxide charge density

p Hole density

r_c Backscattering coefficient

R Resistance

$R_{channel}$ Channel resistance

R_d Drain series resistance

R_m Measured resistance

R_s Source series resistance

R_{sd} Source-drain series resistance

RTA Rapid thermal anneal

S Subthreshold slope

SCEs Short channel effects

T Temperature

t_{ox} Oxide thickness

τ Relaxation time

v Carrier velocity

v_{eff} Effective carrier velocity

v_T Thermal velocity

V Voltage

$V_c(x)$ Channel potential at a distance x from the source

V_{dd} Power-supply voltage

V_{ds} Drain-source voltage

V'_{ds} Intrinsic drain voltage

$V_{ds,sat}$ Saturation drain voltage

V_{fb} Flatband voltage

V_{gs} Gate-source voltage

V'_{gs} Intrinsic gate voltage

V_{int} Intrinsic voltage

V_t Threshold voltage

VLSI Very large scale integration

VTA Threshold voltage adjustment implant

W Channel width

ω Angular frequency

XPS X-ray photoelectron spectroscopy

Y Admittance

Y_p Parallel equivalent circuit admittance

ψ_B Bulk potential

ψ_S Surface potential

Z Impedance

Z_p Parallel equivalent circuit impedance

Chapter 1

Introduction

1.1 The Semiconductor Industry

The semiconductor industry the largest industry in the world with global sales totalling $213 billion in 2004 [SIA-Online, 2005]. At the core of this technology lies the metal-oxide-semiconductor field effect transistor (MOSFET) - effectively nothing more than a simple switch, but perhaps the most numerous man-made structure ever produced. The current generation of Intel Core 2 Duo microprocessors, fabricated at the 65nm technology node, contain a staggering 291 million transistors in an area of just 118 mm^2 [Intel, 2005].

For the last 30 years, silicon (Si) based technologies have dominated the industry, accounting for more than 97% of all microelectronic devices [Paul, 2004]. The dominance of silicon is due to a combination of both its physical properties and its high natural abundance and is therefore inexpensive. The oxide of silicon, SiO_2, forms an excellent insulator between the gate electrode and the channel in a MOSFET with a low defect density at the Si/SiO_2 interface, which is typically $\sim 10^{10}$ cm^{-2} [Nicollian and Brews, 1982]. In addition, its excellent insulating properties can also be exploited as a mask during device fabrication [Sze, 1983].

The growth of the semiconductor industry is based upon an observation by Gordon Moore, who in 1965, noted that the number of transistors on an integrated circuit doubles approximately every two years - an observation that has since become known as *Moore's Law* [Moore, 1965]. This remarkable feat has been achieved through a phenomenon known as *scaling* - a systematic reduction of the MOSFET dimensions. In the main this has been achieved by advancements in device fabrication techniques and in particular, advancements in lithographic techniques that allow ever-decreasing dimensions to be defined.

However, despite the historical success of scaling, it is now very clear that the continued performance enhancements in complimentary metal-oxide-semiconductor (CMOS) circuits that the industry has become accustomed to, cannot be achieved by scaling alone. Indeed, a major concern is the cost of new fabrication plants, predicted to exceed $10 billion by 2010 [Paul, 1999]. However, of even greater importance could be that MOSFET dimensions are rapidly approaching a regime where the key device features consist of just a few hundred atoms, such that quantum effects such as tunnelling become increasingly problematic, resulting in higher leakage currents and power consumption. Transistors with written gate lengths of 35 nm and 1.2 nm silicon oxynitride gate dielectric are already in high-volume production [Tyagi et al., 2005]. In addition, other effects such as random dopant fluctuations are likely to become critical in achieving uniform characteristics across a wafer [Asenov et al., 2003].

However, despite the industry pushing the boundaries closer and closer to its fundamental limit, there still remains great optimism that continued performance enhancements can be maintained using silicon-based technologies in the next few years. There are many materials with intrinsically superior electrical properties compared to silicon, which have found roles in niche market applications such as optical devices and the high frequency components of mobile communication devices. One of the most promising materials is *germanium* (Ge).

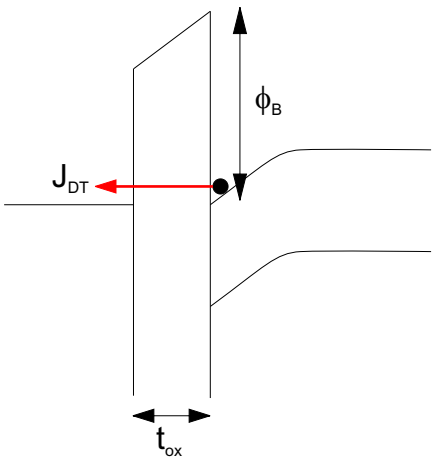


Figure 1.1: Schematic of direct tunnelling current, J_{DT}, through an oxide of thickness, t_{ox}, with a barrier height, Φ_B.

1.2 High-κ Dielectrics

The quality of the Si/SiO_2 interface is arguably the single-most important reason for the dominance of silicon in the microelectronics industry despite there being several known materials with superior electronic properties. However, the current scaling trends for CMOS devices, as specified by the International Technology Roadmap for Semiconductors (ITRS), dictates that the thickness of the gate dielectric should be reduced to below 1 nm by the end of the roadmap [ITRS, 2006 Update]. However, this results in a large increase in the gate leakage current due to quantum mechanical tunnelling, increasing exponentially with decreasing oxide thickness [Plummer and Griffin, 2001]. The gate leakage current due to direct tunnelling is given by:

$$J_{DT} = \frac{A}{t_{ox}^2} exp\left(-2t_{ox}\sqrt{\frac{2m^*q\phi_B}{\hbar^2}}\right) \tag{1.1}$$

where A is a constant, t_{ox} is the oxide thickness, m^* is the carrier effective mass and ϕ_B is the barrier height. The direct tunnelling gate leakage exhibits an exponential dependence on both the physical oxide thickness and the barrier height (Figure 1.1). Figure 1.2 shows the gate leakage limit and oxide thickness, as specified

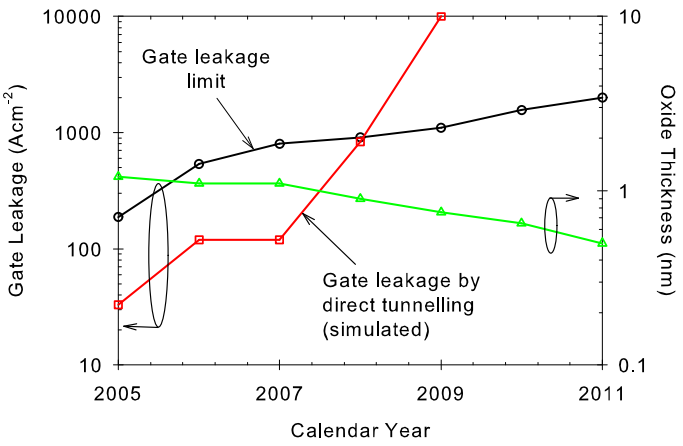


Figure 1.2: Gate leakage current and oxide thickness for silicon oxynitride as specified by the ITRS. The simulated gate leakage for a given oxide thickness is also shown. After ITRS [2006 Update].

by the ITRS [2006 Update], for high-performance devices in the forthcoming years. Also shown, are simulated results for the gate leakage due to *direct tunnelling* for bulk, planar silicon MOSFETs with a silicon oxynitride gate dielectric. The simulated results show that there exists a 'crossover point' such that beyond 2008, the gate leakage limit cannot be met using silicon dioxide (or oxynitride) because of direct tunnelling. This means that silicon dioxide has to be replaced as gate dielectric by a material with higher permittivity than SiO_2, a so-called *high-κ dielectric* in order to achieve the same capacitive coupling between the gate and the channel but for a thicker physical thickness to suppress tunnelling.

The gate capacitance, C, of a MOSFET can be considered as a parallel plate according to:

$$C = \frac{\kappa \epsilon_0 A}{t_{ox}} \tag{1.2}$$

where A is the area of the capacitor, κ is the dielectric constant, ϵ_0 is the permittivity of free space (8.85 x 10^{14} F/cm) and t_{ox} is the thickness of the gate oxide. For high-κ dielectrics, it is convenient to define an *equivalent oxide thickness* (EOT) as

being the theoretical physical thickness of SiO_2 required to achieve the same capacitance as that provided by an alternative dielectric, with a dielectric constant greater than that of SiO_2. Thus, the equivalent oxide thickness can be expressed in terms of an alternative dielectric with a dielectric constant, $\kappa_{high-\kappa}$, and physical thickness, $t_{high-\kappa}$, according to [Wilk et al., 2001]:

$$EOT = \frac{\kappa_{SiO_2}}{\kappa_{high-\kappa}} t_{high-\kappa} = \frac{3.9}{\kappa_{high-\kappa}} t_{high-\kappa} \tag{1.3}$$

This means, for example, that an alternative dielectric with a dielectric constant of 20 and a physical thickness of 5 nm could be used to replace SiO_2 to achieve an EOT of 1 nm, providing a physically thicker barrier to suppress the gate leakage due to direct tunnelling.

In selecting an alternative high-κ dielectric to replace silicon dioxide, there are several properties other than the dielectric constant itself, which must be given due consideration. Barrier height, thermodynamic stability, interface quality, film morphology, gate compatibility, process compatibility and reliability have all been shown to be important [Wallace and Wilk, 2005]. At the time of writing, the most promising candidates appear to be hafnium oxide (HfO_2) and zirconium oxide (ZrO_2).

Intel have recently announced that production of transistors at the 45 nm technology node will commence later this year and that these transistors will include high-κ gate dielectrics [Intel, 2007].

1.3 Germanium

Germanium channel transistors have long been an attractive alternative to silicon for CMOS circuits, mainly due to electron and hole bulk mobility values being two and four times that of silicon, respectively (Table 1.1). In addition, the electron and hole mobility values are closer than those of silicon, which would enable a more symmetrical

Table 1.1: Important properties of silicon and germanium relevant for CMOS device processing (after Sze [1981]).

Property	Silicon	Germanium
Electron bulk mobility (cm^2/Vs)	1450	3900
Hole bulk mobility (cm^2/Vs)	450	1900
Electron effective mass, m^*/m_0	$m_l^* = 0.98$ $m_t^* = 0.19$	$m_l^* = 1.64$ $m_t^* = 0.082$
Hole effective mass, m^*/m_0	$m_{lh}^* = 0.16$ $m_{hh}^* = 0.49$	$m_{lh}^* = 0.044$ $m_{hh}^* = 0.28$
Melting point (°C)	1415	937
Band gap (eV)	1.12	0.66

CMOS. Finally, the lower melting point of germanium compared to silicon, and the lower dopant activation energies in germanium, would allow the thermal budget of CMOS processing to be substantially reduced.

In fact, the first demonstration of the transistor, the bipolar transistor, back in 1948, used germanium as the active material [Bardeen and Brattain, 1948]. However, the absence of good quality native oxide has, to date, led to germanium only playing a bit-part role in the microelectronics industry. But with the inevitable replacement of SiO_2 as gate dielectric with an alternative high-κ dielectric, silicon's dominance in the microelectronics industry could be compromised. There have already been numerous publications on germanium pMOSFETs with hole mobilities exceeding those in silicon [Shang et al., 2003, Ritenour et al., 2006]. However, even at this early stage, results for germanium nMOSFETs appear less promising (see Chapter 2.6), and further research is still required in this area [Martens et al., 2006]. It might be such that in future generations, germanium is used in CMOS circuits for the p-channel devices, and a different high-mobility material such as gallium arsenide (GaAs) is used for n-channel devices. Ultimately, if germanium channel transistors are to become prevalent in CMOS circuits, the low natural abundance and high cost of germanium will mean that it will have to be incorporated in thin layers in an *on-insulator* technology [Feng et al., 2006,

Nakaharai et al., 2003].

Germanium can also be combined with silicon to form a random silicon-germanium (SiGe) alloy. In particular, this allows for the incorporation of *strain* into CMOS devices and higher carrier mobility (see Chapter 2). There are several ways in which strain can be introduced into CMOS devices, which can be divided into either *global* or *process-induced* (also, referred to as *local*) straining techniques.

Perhaps the easiest way to introduce global strain, is to grow a thin SiGe layer pseudomorphically on a silicon substrate, such that it assumes the lattice constant of the underlying silicon. Since the lattice constant of Ge is 4.2% bigger than silicon, the SiGe layer will be under biaxial compressive strain. Such structures are usually grown with a thin silicon cap on top of the SiGe layer for the purposes of cleaning and oxidation, which makes them unlikely to ever enter mainstream production due to the inherent problems of operating buried channel devices, although the use of a high-κ dielectric might overcome this problem.

An alternative global straining technique uses a SiGe *virtual substrate*. In such structures a thin strained silicon layer is grown on a relaxed SiGe buffer layer such that it is under biaxial tensile strain. Whilst this architecture permits surface channel devices, the high density of threading dislocations associated with the relaxation of the SiGe virtual substrate can lead to high leakage currents. Furthermore, the low thermal conductivity of the SiGe virtual substrate results in self-heating becoming an important issue [Nicholas et al., 2005a]. Whilst these problems mean that chips incorporating strained silicon on a virtual substrate are unlikely to enter commercial production, it is feasible that the virtual substrate platforms could be used to provide high-quality strained silicon layers for an on-insulator technology through layer transfer techniques [Langdo et al., 2004].

There are several methods by which process-induced straining techniques can be exploited for CMOS devices. For example, whilst germanium has yet to enter mainstream

production as a channel material for CMOS devices, Intel have recently shown that it can be incorporated into the source and drain regions of a silicon pMOSFET, resulting in a uniaxial compressively strained silicon channel, and these devices are now in high-volume production [Tyagi et al., 2005]. It is also possible to incorporate carbon into the source and drain regions to create a uniaxial tensile strained silicon channel to enhance electron mobility for nMOSFETs [Chui et al., 2007]. However, this work is still very much in development. Instead, state-of-the-art nMOSFETs use a silicon nitride thin film around the nMOSFET to induce a uniaxial tensile strain in the channel of the underlying MOSFET, to achieve higher electron mobilities [Tyagi et al., 2005].

As a final point, it should be mentioned that the narrow band gap of SiGe can be used in the base of a heterojunction bipolar transistor (HBT) to reduce the barrier to electron transport between the n-type emitter and collector contacts, to increase the current gain of the device. SiGe HBTs are already in commercial production.

1.4 Current Work

This thesis examines some of the important issues when either SiGe or Ge are used as conducting channel materials in pMOSFETs, in combination with high-κ dielectrics. This research has been conducted within several ongoing research development projects, including the Ge MOSFET Programme at IMEC (Belgium) and the New Channel Materials Work Package within the SINANO Network of Excellence.

Whilst there have been numerous publications regarding the performance of both SiGe and Ge channel MOSFETs with high-κ dielectrics, many aspects of carrier transport are not well understood. For example, the impact of the high-κ dielectric on the effective mobility of inversion layer carriers remains an important issue. Furthermore, it is still unknown as to whether band-to-band tunnelling will prove to be a fundamental problem in Ge channel MOSFETs due to the smaller Ge band gap.

Chapters 2 and 3 serve to provide the necessary background details required to understand the current investigation. Chapters 4 and 5 investigate the properties of SiGe channel capacitors and MOSFETs with high-κ dielectric and metal gate. Chapter 6 concerns work on the behaviour of long and short channel Ge pMOSFETs, with a particular emphasis on understanding the effective hole mobility and also the role of band-to-band tunnelling through measurements at cryogenic temperatures. Finally, with carriers experiencing very high lateral fields in modern state-of-the-art devices, the roles of both impact ionisation and velocity overshoot are discussed in Chapter 7 for two batches of SiGe channel pMOSFETs, one of which was processed by CEA-LETI (France) and the other as part of the UK-based HMOS project.

Chapter 2

Theoretical Background

2.1 Introduction

This chapter serves as an introduction to the field of $Si_{1-x}Ge_x (0 \leq x \leq 1)$ channel MOSFETs, including the integration of high-κ dielectrics and metal gates, which form the subject of this thesis. The chapter divides into four main parts.

In the first part, the basic operation of the simple bulk long-channel MOSFET is reviewed. The second part introduces the carrier mobility and the scattering mechanisms encountered by carriers in a MOSFET. This also includes a brief discussion on the role of the valence band structure, and in particular, how *strain* can be used to modify the valence band structure and the ensuing results. The third part provides a summary of *short-channel effects* (SCEs), now important in state-of-the-art CMOS devices that have resulted directly from the aggressive down-scaling of the MOSFET channel length. Finally, the chapter concludes with a section on process considerations specific to the fabrication of bulk Ge MOSFETs incorporating high-κ dielectrics and metal gates.

Whilst every effort has been made to include all of the important aspects relating to $Si_{1-x}Ge_x$ channel MOSFETs, this chapter merely serves as an introduction to the field, providing the salient points relevant to the current work. There are several

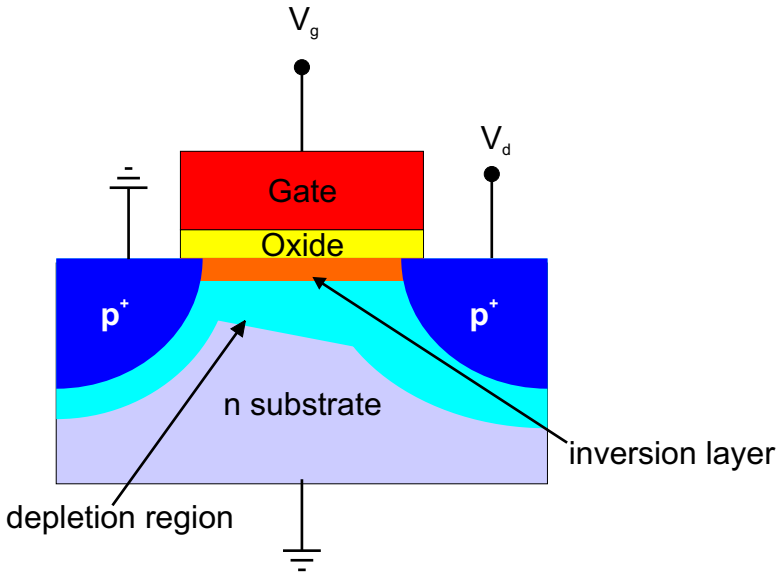


Figure 2.1: Schematic of a simple pMOSFET.

excellent books, to which the reader is referred for a more comprehensive discussion. For example, Taur and Ning [1998], Tsividis [1999], Sze [1981] and Shur [1990] provide an excellent foundation for device physics; Lundstrom [2000] is an excellent introduction to scattering; band structure of the silicon-germanium materials system is reviewed by Schäeffler, F. [1997]; and Huff and Gilmer [2005] provides one of the most thorough reviews of high-κ dielectrics available at the present time.

2.2 Basic MOSFET Operation

Figure 2.1 shows a schematic of a bulk silicon pMOSFET. The simple MOSFET is a four-terminal device, with a source, a drain, a gate and a substrate or body contact, each of which can be biased independently. For a pMOSFET, the source and drain are p^+ doped and the substrate is n doped. Traditionally, p^+ poly-silicon is used as gate electrode, although as will be seen in Section 2.6, a metal with suitable workfunction can be employed instead. The gate electrode is insulated from the bulk silicon substrate by a thin oxide, typically SiO_2, such that the gate and silicon substrate form a parallel

plate capacitor with the oxide layer serving as the dielectric. When the voltage applied to the gate is zero, majority carriers (electrons) in the n-substrate are accumulated under the gate, thus isolating the source and drain regions. This prevents current flow along the channel and the MOSFET is in the off-state. As the gate voltage is increased (i.e. made more negative for a pMOSFET), the electrons in the bulk silicon substrate are repelled away from the Si/SiO_2 interface, forming a depletion region under the gate, devoid of free carriers. If the gate voltage is increased further, it becomes energetically favourable for holes in the p^+ source and drain regions to populate the area under the gate forming an *inversion layer*, which serves as a conducting channel connecting the source and the drain, and the MOSFET is in the on-state. Since the drain is negatively biased with respect to the source, holes are able to flow along the channel from the source to the drain. Increasing the gate voltage even further increases the concentration of holes under the gate and also the drain current. Hence, the gate modulates the current flowing from the source to the drain. It is this drain current modulation that gives rise to the switching operation in fully integrated CMOS circuits.

Despite the simplicity of its operation, some of the effects observed in MOSFETs, and in particular short-channel MOSFETs, can be quite complicated. In order to explore some of these effects, a quantitative discussion of the operation of a long-channel MOSFET now follows, which will serve as the basis for the short-channel effects, introduced in Section 2.5.

The behaviour of the long-channel MOSFET can be described under the *gradual channel approximation*, in which it is assumed that the electric field along the channel is much smaller than the electric field perpendicular to it [Shur, 1990]. Furthermore, the analysis is further simplified under the *charge sheet approximation*, which assumes that the minority carriers that form the inversion layer are located in a sheet at the semiconductor surface [Taur and Ning, 1998]. This implies that no voltage is dropped across the inversion layer and Poisson's equation can be reduced to one-dimension.

The inversion charge density, Q_{inv}, at a point x along the channel can be expressed as:

$$Q_{inv} = C_{ox} \left(V_{gs} - V_t - V_c(x) \right) \tag{2.1}$$

where C_{ox} is the oxide capacitance, V_{gs} is the gate voltage with respect to the source, V_t is the threshold voltage of the device, which is assumed to be constant, and $V_c(x)$ is the potential at a distance x along the channel with respect to the source.

The carrier velocity at a position x in the channel, $v(x)$, is related to the longitudinal electric field, $\varepsilon(x)$, by the carrier mobility, μ, according to:

$$v(x) = \mu\varepsilon(x) = \mu \frac{dV_c(x)}{dx} \tag{2.2}$$

where it is assumed that the carrier mobility is constant along the channel. The current flowing from the source to drain, I_{ds}, is given by the product of the inversion charge, the carrier velocity and the device width:

$$I_{ds}(x) = W Q_{inv}(x) v(x) \tag{2.3}$$

Substituting Equations 2.1 and 2.2 into Equation 2.3, the drain current for the MOSFET can be written as:

$$I_{ds} = C_{ox} W \mu \left(V_{gs} - V_t - V_c(x) \right) \frac{dV_c x}{dx} \tag{2.4}$$

where it has been assumed that the drain current, I_{ds}, is constant along the channel.

Equation 2.4 can be integrated, with the appropriate boundary conditions such that $V_c(0) = 0$ and $V_c(L) = V_{ds}$, to obtain the expression for the drain current:

$$I_{ds} = C_{ox} \frac{W}{L} \mu \left[\left(V_{gs} - V_t \right) V_{ds} - \frac{V_{ds}^2}{2} \right] \tag{2.5}$$

The drain current has a maximum value when $V_{ds} = V_{gs}$ - V_t. This point corresponds to the MOSFET entering the *saturation regime*, and the *saturation current* can be expressed as:

$$I_{ds,sat} = C_{ox}\frac{W}{2L}\mu\left(V_{gs} - V_t\right)^2 \tag{2.6}$$

The onset of the saturation regime in a MOSFET can be understood as follows. As the drain voltage of the MOSFET increases, the drain current increases according to Equation 2.5. However, the inversion charge at the drain, $Q_{inv}(L)$, decreases according to Equation 2.1, such that as V_{ds} approaches V_{gs} - V_t, the inversion charge density approaches zero. This point is referred to as *pinch-off*, and results in a highly resistive region forming at the drain, since the inversion charge vanishes. Any further increase in V_{ds} will be result in the pinch-off point being moved towards the source, with the extra potential being dropped across this high resistive region near the drain. Hence, the drain current behaves as if $V_{ds} = V_{gs}$ - V_t, and saturates. At the saturation point, the gradual channel approximation breaks down, with carriers no longer being confined to the surface channel. Consequently, carriers are injecting from the pinch-off point into the drain depletion region and drift to the drain.

Whilst the assumptions made in the gradual channel approximation and charge sheet approximation are valid for describing the behaviour of long-channel MOSFETs, this simple model *cannot* be used to describe the behaviour of short channel MOSFETs, in which, short channel effects such as punchthrough, drain-induced barrier lowering (DIBL), velocity overshoot and quantum mechanical tunnelling must also be considered [Tsividis, 1999]. Hot carrier effects also become increasingly important due to the higher longitudinal field experienced by carriers along the channel, resulting directly from scaling. A summary of these effects is presented later in Section 2.5.

When considering the operation of a MOSFET, it is also important to consider

the effects of parasitic elements that effect its behaviour. One of the most important parasitic elements is the finite resistance of the source and drain regions, R_s and R_d, respectively [Schroder, 2006]. In a conventional planar MOSFET, the source and drain regions are formed by ion implantation of dopants, which are activated by a high temperature anneal. The voltages specified so far, refer to those voltages applied at the drain and gate electrodes, thus neglecting any voltage dropped across the source and drain regions. In order to account for these effects, it is necessary to replace the voltages, V_{ds} and V_{gs}, with their intrinsic equivalents, V'_{ds} and V'_{gs}:

$$V_{ds} \rightarrow V'_{ds} = V_{ds} - I_{ds}R_s - I_{ds}R_d \tag{2.7}$$

$$V_{gs} \rightarrow V'_{gs} = V_{gs} - I_{ds}R_s \tag{2.8}$$

2.3 Carrier Mobility and Scattering Mechanisms

The carrier mobility is perhaps the single-most important parameter with regards to MOSFET performance, particularly when trying to demonstrate the improved performance of alternative channel materials over silicon. Since the drain current of a MOSFET is proportional to the mobility, increasing the mobility should, in theory, improve the drive current and hence the CMOS circuit performance. Indeed, it is the higher electron and hole bulk mobility values in germanium [Sze, 1981] that provides the motivation for the incorporation of germanium into the channel region of a MOSFET.

As a carrier traverses the channel of a MOSFET from the source to the drain, its motion is likely to be impeded by one or more scattering events, which reduce the carrier mobility and hence the drive current. In the (unlikely) event that a carrier travels along the MOSFET without encountering such a scattering event, then it is said to be in the *ballistic regime*, which is predicted to occur for the shortest channel transistors,

where the gate length is shorter than the carrier mean free path [Natori, 1994].

The carrier mobility in a MOSFET is lower than the bulk mobility due to the presence of additional scattering mechanisms, associated with confining the inversion layer to a surface [Taur and Ning, 1998]. Consequently, this section discusses the important scattering mechanisms in a MOSFET, including ionised impurity (or Coulomb) scattering, phonon scattering, interface roughness scattering and alloy scattering.

In the treatment of carrier scattering, the carriers are frequently represented in the form of a Bloch wave, which is the product of a plane wave and a function with the periodicity of the lattice [Lundstrom, 2000]. Mathematically, this can be represented as:

$$\psi_k = u_k e^{ikz} \tag{2.9}$$

$$u_k(z) = u_k(z + a) \tag{2.10}$$

Bloch waves travel unimpeded through the periodic crystal structure but they can be scattered when they encounter *perturbations*, such as defects or lattice vibrations. When a carrier is scattered, it undergoes a transition from an initial state, $\mathbf{p}$, to a final state $\mathbf{p}'$. The scattering process can be described in terms of a scattering potential, which can be used to calculate the scattering rate for that particular scattering mechanism [Lundstrom, 2000]. It is often possible to use a scattering potential to describe the transition such that the scattering rate is proportional to the final density of states, such that the higher the number of available states there are for a carrier to scatter into, the more likely it is to occur.

2.3.1 Ionised Impurity Scattering

Ionised impurity (or Coulomb) scattering represents the simplest perturbation to the periodic lattice potential. Coulomb scattering is an example of scattering from a

static potential, meaning that carriers can respond in such a way that minimises the effect of the potential. This is known as *screening*. The effect of screening results in ionised impurities usually only limiting the mobility in weak inversion, where the mobile carrier concentration in the inversion layer is low. Scattering from ionised impurities is usually much lower for high energy carriers than for low energy carriers [Lundstrom, 2000].

Ionised impurities are often introduced into the vicinity of the channel during device fabrication steps such as ion implantation. In particular, the effects of ionised impurity scattering is likely to become increasingly important as the MOSFET gate length is scaled-down, due to the necessity of high substrate doping concentrations, including halo implants, which are required to control short-channel effects (Section 2.5). In addition, defects at both the semiconductor/gate dielectric interface and within the gate dielectric itself can also result in increased Coulomb scattering of mobile carriers. This can be particularly problematic in MOSFETs featuring high-κ dielectrics, which typically have higher interface trap densities than the Si/SiO_2 system (Section 2.6).

2.3.2 Phonon Scattering

Phonon scattering is the most important scattering mechanism in bulk semiconductors and also in metals. Phonon scattering occurs when the displacement of atoms caused by lattice vibrations changes the local band structure, which changes the local conduction and valence band energies. This produces a *grating* in the band edges, from which the Bloch waves can scatter [Lundstrom, 2000].

Both silicon and germanium have two atoms in the primitive basis. This results in the phonon dispersion relation splitting into two branches: *acoustical* and *optical* phonons [Kittel, 1996]. Acoustical phonon modes are like sound waves, in that adjacent atoms (or planes of atoms) are displaced in the same direction. In optical phonon modes, adjacent atoms are displaced in the opposite direction. Figure 2.2 distinguishes

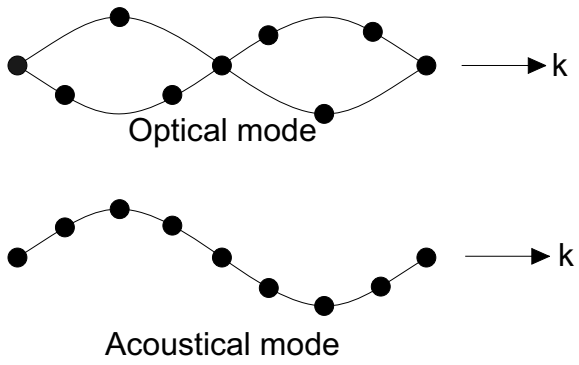


Figure 2.2: Simple schematic showing the difference between transverse optical and transverse phonon modes. After Kittel [1996].

between the two branches. Moreover, phonon modes can be either *longitudinal* or *transverse*, depending on whether atoms are displaced along or perpendicular to the phonon wavevector. Thus Figure 2.2 depicts the difference between transverse acoustic and tranverse optical phonon modes. For both silicon and germanium, there exist three acoustic phonon modes (one longitudinal and two transverse) and three optical phonon modes (one longitudinal and two transverse).

Acoustical phonons have negligible energy compared to that of the carriers, and so acoustic phonon scattering is effectively elastic. Optical phonons, on the other hand, have a much higher minimum energy, due to the greater distortion between adjacent atoms. Thus, optical phonon scattering cannot be considered as elastic, except for the highest energy carriers. It is possible for carriers to scatter by both phonon emission and phonon adsorption. However, there exists a threshold for optical phonon emission, $\hbar\omega_0$, above which the scattering rate increases significantly. This results in *velocity saturation* [Tsividis, 1999], which is approximately 7 x 10^6 cm/s for holes in silicon [Taur et al., 1993]. Unlike ionised impurity scattering, phonon scattering represents scattering from a non-static potential. This means, for example, that optical phonon scattering, that typically have high frequencies, will be largely unscreened.

2.3.3 Surface Roughness Scattering

When confining an inversion layer against a junction between two materials, the roughness of the interface will scatter carriers. The interface roughness is typically assumed to be distributed in a Gaussian distribution with an amplitude Δ and characteristic length Λ [Fischetti and Lauz, 1993].

Surface roughness scattering is another example of scattering from a static potential, such that it can be screened by mobile carriers. Interface roughness scattering is dominant in modern MOSFETs, which are operated at vertical fields greater than 1MV/cm.

2.3.4 Alloy Scattering

The introduction of germanium atoms into the silicon crystal disrupts the local band structure, which can result in additional short range scattering potentials [Kearney and Horrell, 1998]. The treatment of alloy scattering is one of the least well developed of the scattering processed. The scattering rate is highest for an alloy composition of 50%, although it is still unclear as to what impact it has on the carrier mobility and whether or not it is a screened mechanism, and is mentioned here for completeness.

2.4 Band Structure

The band gaps of both silicon and germanium are indirect with values of 1.12 eV and 0.66 eV at room temperature, respectively. In both of these semiconductors the conduction and valence bands consist of a number of *sub-bands*. In silicon, the conduction band minima lie along the <100> crystal direction and are six-fold degenerate. They are often referred to as the Δ_6 minima, since the Δ axis in the Brillouin zone is along the <100> direction and the subscript refers to the degeneracy. In germanium, the conduction band minima lie at the zone boundaries along the <111> direction of

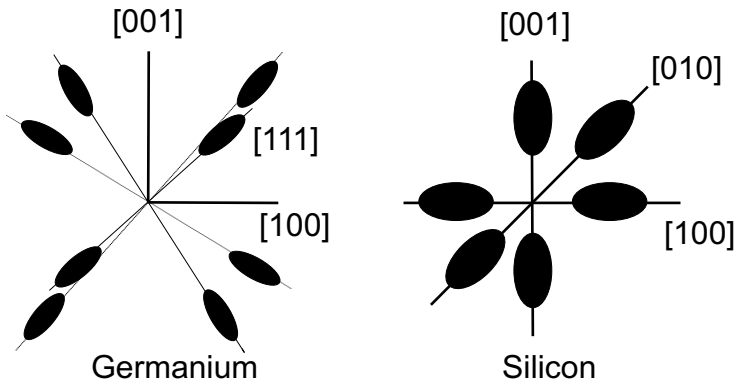


Figure 2.3: Schematic of the conduction band minima for bulk germanium and bulk silicon. After Paul [2004].

the Brillouin zone (L points) and are eight-fold degenerate [Sze, 1981] (Figure 2.3).

The valence band edge of both silicon and germanium occurs at the zone centre ($k = 0$). At this point, the valence band consists of two degenerate bands and a split-off band separated in energy from the two degenerate bands due to the spin-orbit interaction. The spin-orbit splitting energy is 44 meV in silicon and 296 meV in germanium. The two degenerate bands are defined as the light-hole band (the narrower band) and the heavy-hole band (the wider band). The effective mass, m^*, is given in tensor form to account for crystal anisotropy, by [Kittel, 1996]:

$$\frac{1}{m^*_{ij}} = \frac{1}{\hbar^2}\frac{\partial^2 E(k)}{\partial k_i \partial k_j} \tag{2.11}$$

The mobility is related to the effective mass according to:

$$\mu = \frac{q\tau}{m^*} \tag{2.12}$$

where τ is the relaxation time. The lower hole effective mass, in addition to the lower density of states in the valence band, and longer relaxation times in germanium compared to silicon is directly responsible for the higher bulk hole mobility observed in germanium, making it an attractive alternative channel material to silicon in pMOSFETs [Sze, 1981].

In addition to taking advantage of the intrinsically higher hole mobility offered by germanium, it is also possible to take advantage of the larger lattice constant of germanium and incorporate it into silicon to form a silicon-germanium ($Si_{1-x}Ge_x$) alloy to form a *strained* channel. The effects of strain on the $Si_{1-x}Ge_x$ band structure and carrier mobility are well documented Fischetti and Laux [1996], Paul [1999], Schäeffler, F. [1997], Whall and Parker [1998]. Consequently, the discussion presented will focus on how strain effects the valence band structure in a pseudomorphic $Si_{1-x}Ge_x$ layer structure, relevant to the hole mobility for pMOSFETs.

Silicon and germanium are both group IV elements that are entirely miscible such that it is possible to form a $Si_{1-x}Ge_x$ random alloy with properties that vary gradually from those that are Si-like to those which are Ge-like across the entire composition range. The unstrained alloy, like both silicon and germanium, crystallises in the diamond structure with a lattice constant that varies almost linearly with composition. The band structure of the *unstrained* $Si_{1-x}Ge_x$ is almost the same as that of silicon up to a germanium composition of 85%.

When a *thin* $Si_{1-x}Ge_x$ (with $x > 0$) layer is grown *pseudomorphically* on a silicon substrate such that the $Si_{1-x}Ge_x$ layer takes on the lattice constant of the underlying silicon substrate, the $Si_{1-x}Ge_x$ layer will be compressively strained in the plane and under tensile strain in the growth direction, resulting in a tetragonal distortion. This is depicted in Figure 2.4.

The effect of strain on the valence band is twofold and is shown schematically in Figure 2.5. Firstly, there is a splitting of the sub-bands. Secondly, there is a change in the effective mass of the holes that populate the sub-bands [Schäeffler, F., 1997].

The band-splitting results in the light and heavy holes becoming non-degenerate at the zone centre, with the heavy hole band being lowered in energy (remembering that the hole energy increases downwards in Figure 2.5), and the light hole band being raised in energy. This has the effect of reducing phonon scattering events between the two

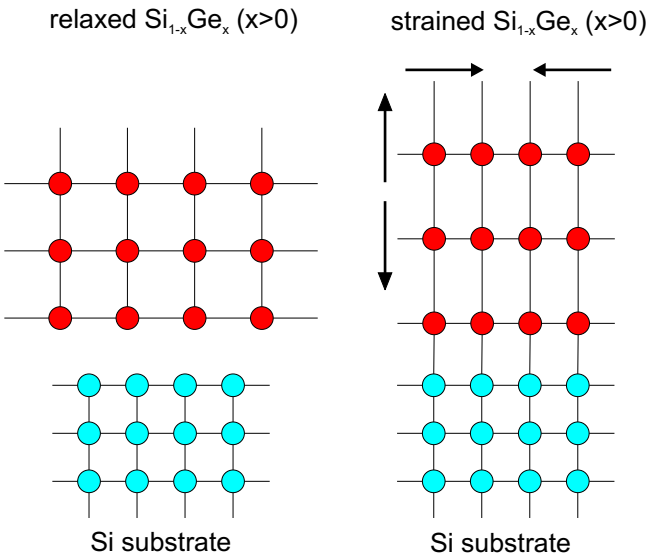


Figure 2.4: Schematic showing the pseudomorphic strained $Si_{1-x}Ge_x$ layer (below the critical thickness) on a silicon substrate.

sub-bands. The probability of a scattering event is strongly dependent on the number of available states for a carrier to scatter into. For the unstrained case, the degeneracy of the light and heavy hole bands at the zone centre yields a high number of available states with similar energies for a carrier to scatter into. The strain-splitting of the valence band significantly reduces the number of scattering events between the two sub-bands, thus helping to increase the effective hole mobility. Furthermore, the split-off band is raised even further in energy from the now non-degenerate light and heavy hole bands, further reducing the likelihood of holes scattering into this sub-band [Xie, 1999].

In addition to the band-splitting, the effective hole mass of both bands is changed. The effective mass of the heavy-hole band is lowered, whilst that of the light-hole band is increased. Depending upon the strain conditions, it is possible for the effective mass in the heavy hole band to be lower than that of the light hole band, thus resulting in *mass inversion*. The lower effective mass also reduces the density of states and increases the acceleration of holes between scattering events, which can lead to a higher hole mobility.

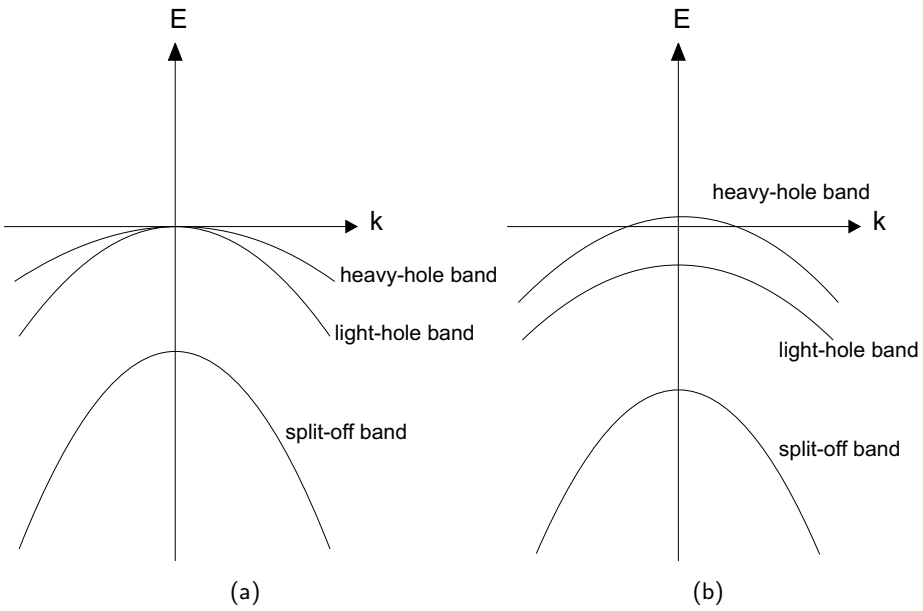


Figure 2.5: Schematic of the (a) unstrained and (b) biaxial compressively strained $Si_{1-x}Ge_x$ valence band showing the band splitting and change of effective mass.

The addition of germanium and the lowering of the heavy hole sub-band reduces the strained $Si_{1-x}Ge_x$ bandgap. The difference in the band gap between the strained $Si_{1-x}Ge_x$ and silicon substrate leads to discontinuities (or offsets) in both the conduction band and valence band [Schäeffler, F., 1997]. However, the conduction band offset is small, typically less than around 20 meV, and is often neglected, such that we can speak of a discontinuity in the valence band only. The valence band offset varies linearly with germanium composition and is given by [Galdin et al., 2000]:

$$\Delta E_v \approx 0.74x \tag{2.13}$$

The discontinuity in the valence band results in the confinement of holes in a quantum well in the high-mobility $Si_{1-x}Ge_x$ layer, with higher germanium compositions resulting in a larger valence band offset and a greater degree of confinement.

Unfortunately there is a trade-off between the germanium composition and the

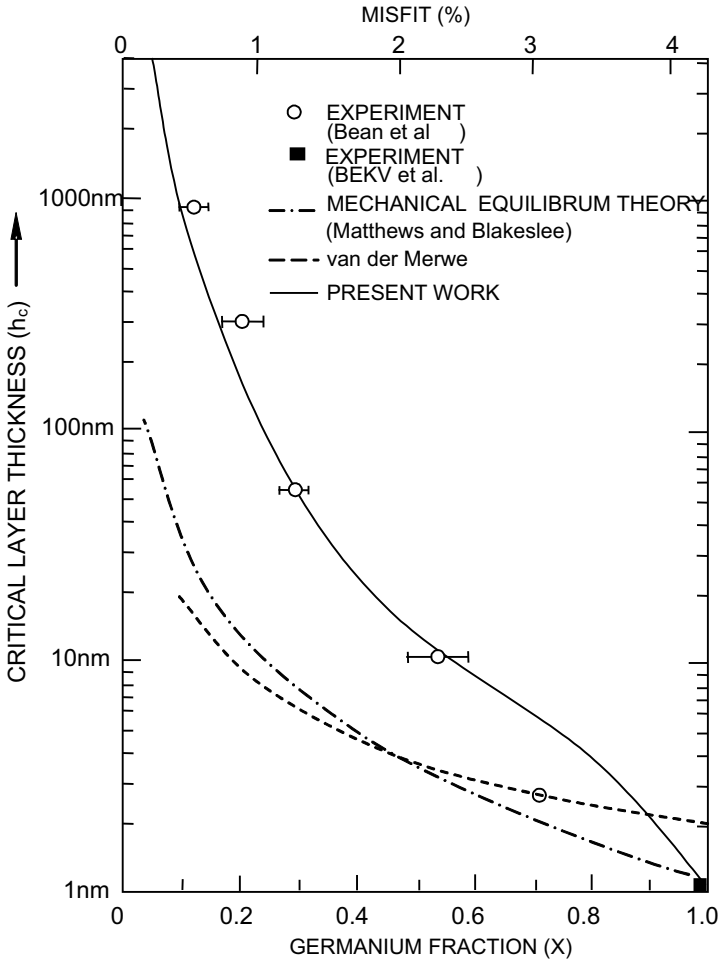


Figure 2.6: Critical thickness of $Si_{1-x}Ge_x$ layers grown on silicon substrate as a function of germanium composition. After People and Bean [1985].

maximum thickness of the silicon-germanium layer, such that it remains strained, the so-called *critical thickness* [People and Bean, 1985]. The critical thickness is shown as a function of germanium composition in Figure 2.6. Layers exceeding the critical thickness will relax by the formation of misfit dislocations. This emphasises that whilst a pure germanium layer might be best for quantum confinement of holes, only 3 monolayers of strained Ge can be grown on a silicon substrate before the layer starts to relax, which is of insufficient thickness to support a hole quantum well. It is possible to grow epitaxial layers that are thicker than the critical thickness by carefully control of the growth conditions. Such layers are said to be *metastable* and care must be taken to ensure that these layers do not relax during processing.

When designing a strained $Si_{1-x}Ge_x$ MOSFET, it is customary to grow a thin silicon cap on top of the strained $Si_{1-x}Ge_x$ channel. The silicon cap serves three main purposes [Palmer, 2001]. Firstly, oxidation of $Si_{1-x}Ge_x$ layer to form a gate dielectric tends to result in a snow-plough effect with a build-up of germanium at the interface and high densities of interface traps. Secondly, part of the silicon cap will be consumed during processing. Finally, it can provide extra stability against strain relaxation, particularly for layers that exceed the critical thickness.

2.5 Short Channel Effects

As the channel length of a MOSFET is reduced, simple models, such as those used in Section 2.2 can no longer be used to accurately describe MOSFET behaviour, which is said to suffer from *short channel effects* (SCEs) (See for example, Tsividis [1999]). Short channel effects are caused by the greater influence of the source and drain regions on the channel region. This is a direct result from the reduction of the channel length.

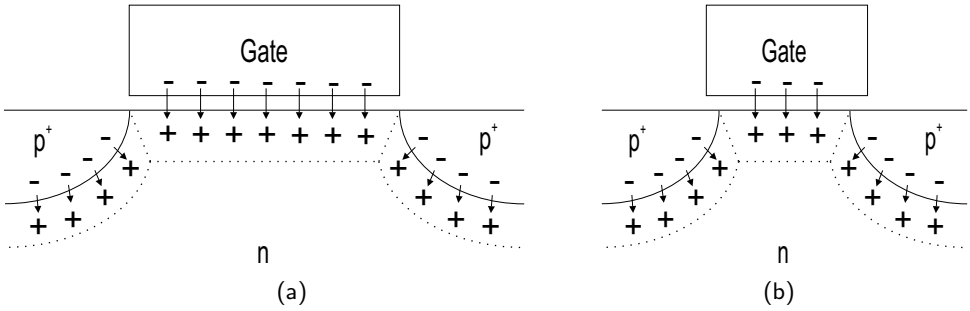


Figure 2.7: Schematic to show the charge sharing of the depletion region for (a) long and (b) short-channel pMOSFET.

2.5.1 Threshold Voltage Roll-off

Figure 2.7 shows a schematic of a MOSFET biased such that a depletion region is formed under the gate. Depletion regions are also formed around the source and drain junctions. The fixed charges in the depletion region are balanced by a combination of equal and opposite charges on the gate and within the depletion regions of the source and drain regions. This is referred to as the *charge sharing model* [Taur and Ning, 1998]. This allows a somewhat arbitrary division of the depletion region into three areas: one in which the charge is balanced by charges on the gate, and two others where the charge is balanced by charges within the depletion regions of the source and drain, respectively. The field lines in Figure 2.7 indicates the division and charge sharing of the depletion region.

In the long-channel MOSFET, the field lines terminating on the depletion charges originate almost entirely from the gate and to a very good approximation, the depletion charge is balanced entirely by charges in the gate. This is depicted in Figure 2.7(a), where the effects of the source and drain are shown to be negligible.

As the channel length is decreased, the width of the depletion regions around the source and drain begin to become comparable to the channel length, and the depletion regions around the source and drain form a greater proportion of the total depletion

region under the gate. This means that a larger proportion of the bulk depletion charge is balanced by charges residing within the depletion regions of the source and drains. Consequently, a smaller charge on the gate is required for the onset of inversion, which results in an accompanying reduction in the threshold voltage (Figure 2.7(b)).

The threshold voltage roll-off can be reduced by increasing the substrate doping and/or by reducing the junction depth. However, both techniques have their drawbacks. Reducing the junction depth suffers from an increased source-drain series resistance, which needs to be reduced by silicidation [Maex, 1993]. Increasing the substrate doping, reduces the carrier mobility and hence drive current, due to increases Coulomb scattering from the ionised dopants. Consequently, in deep-submicron MOSFETs, the substrate doping is only increased locally near the source and drain regions, using a technique known as *halo doping* [Roy et al., 2003, Yeh and Chou, 2001]. The use of halo-doping can also result in so-called *short channel effects*, since the average channel doping concentration increases, causing the threshold voltage to increase. However, as the gate length is further reduced, the regular SCE becomes more dominate and the threshold voltage decreases [Colinge and Colinge, 2006].

2.5.2 Channel Length Modulation

In section 2.2 it was shown that the drain current in a long-channel MOSFET saturates for a drain bias above $V_{ds} = V_g$ - V_t due to the channel pinching-off at the drain. As the drain bias is further increased, the saturation point moves away from the drain, extending into the channel, and the voltage at the pinch-off point remains at $V_{d,sat}$, with any additional increase in drain voltage dropped between the pinch-off point and the drain. Thus, the channel actually behaves as if it has an effective channel length of L - ΔL, where ΔL is the lateral distance between the pinch-off point and the drain. The distance ΔL is governed by the depletion width at the drain, which depends on the substrate doping and the drain voltage. For a long-channel MOSFET,

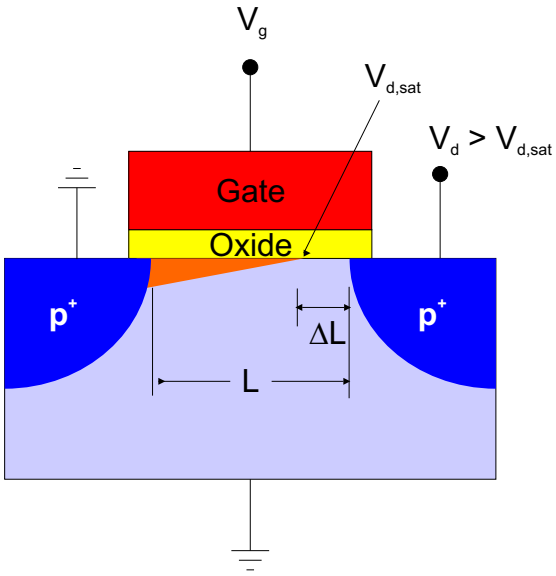


Figure 2.8: Channel length modulation in a MOSFET for a drain bias higher than the saturation drain voltage.

this effect is relatively small, since the channel length is much bigger than the depletion width formed at the drain. However, for a short-channel MOSFET, where the depletion width at the drain is comparable to the channel length, this results in a non-saturation of the drain current with increasing drain voltage, since ΔL increases with drain voltage. Thus, in the equation for the drain current of a MOSFET, L should be replaced by L - ΔL, according to:

$$I_{ds} = \frac{I_{dsat}}{1 - \left(\frac{\Delta L}{L}\right)} \tag{2.14}$$

2.5.3 Drain-Induced Barrier Lowering and Punchthrough

Figure 2.9 shows the surface potential as a function of lateral distance for a MOSFET. When the gate voltage is below threshold, only a limited number of minority carriers are injected over the potential barrier at the source into the channel, which results in the subthreshold current. In the long-channel MOSFET, the potential barrier is flat over most of the channel, and the lowering of the potential at the drain has a negligible

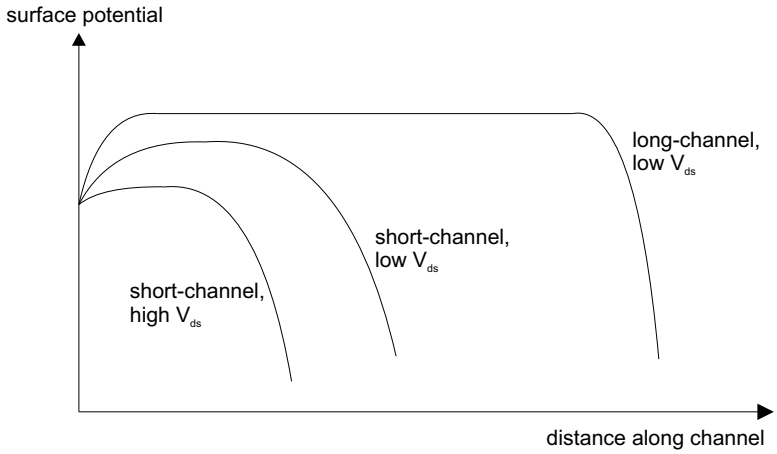


Figure 2.9: Schematic of the surface potential versus distance along a MOSFET channel showing how the drain bias in long and short-channel devices influences the barrier height at the source. After Taur and Ning [1998].

effect upon the barrier at the source. However, as the channel length decreases, the depletion region at the drain extends further into the MOSFET, lowering the potential barrier at the source, resulting in a *drain-induced barrier lowering (DIBL)*. This leads to both a reduction of the threshold voltage and an increase in the subthreshold current [Taur and Ning, 1998].

Punchthrough is essentially an extreme version of DIBL and occurs when the depletion regions of the source and drain regions overlap, without any depletion from the gate. This results in the potential barrier at the source being severely reduced or even vanishing entirely. Consequently, under the application of a drain voltage, V_{ds}, it is possible for a current to flow from the source to the drain, in the absence of a voltage being applied to the gate. There are two cases of punchthrough to consider: surface punchthrough, which is usually the case for a uniformly doped substrate, and bulk punchthrough, which usually occurs in devices with ion implanted source and drain regions that have a higher doping concentration at the surface. As is the case for the other short-channel effects, DIBL and punchthrough are best suppressed by increasing the substrate doping to reduce the width of the depletion regions around the source and

drain. The effects of DIBL and punchthrough on the MOSFET I-V characteristics are discussed in Section 3.3.3.

2.5.4 Velocity Overshoot

In very short channel transistors, the assumptions of the drift-diffusion models, in which carrier transport is treated to be in thermal equilibrium with the semiconductor lattice, breaks down. The presence of high fields and rapid spatially varying potentials can result in non-equilibrium transport effects such as *velocity overshoot*, where the carrier velocity can exceed the carrier saturation velocity. In a MOSFET the highest lateral fields are found in the vicinity of the drain region, and so if velocity overshoot effects are to occur, then it will most likely occur in this region. Whilst this might provide an opportunity to enhance the MOSFET drain current, Lundstrom [1996] argues that the effects of velocity overshoot are of secondary importance to the carrier backscattering at the source end of the channel. Lundstrom [1996] showed that the saturation drain current of a MOSFET can be expressed in terms of a carrier backscattering coefficient, r_c, according to:

$$I_{ds,sat} = C_{ox} W v_T \left(\frac{1 - r_c}{1 + r_c} \right) (V_{gs} - V_t) \tag{2.15}$$

in which the effective carrier velocity, v_{eff}, is given by:

$$v_{eff} = v_T \left(\frac{1 - r_c}{1 + r_c} \right) \tag{2.16}$$

Here v_T is the Richardson thermal velocity. In the ballistic regime, the carrier backscattering coefficient is equal to zero, and the maximum drain current is controlled by the carrier injection velocity at the source end of the channel. Whilst carriers cannot be injected into the channel higher than the Richardson thermal velocity, Lundstrom [1996] argues that the effects of velocity overshoot are still important since it has an

effect on the self-consistent field in the channel and hence upon the backscattering coefficient at the source. This has important implications for modern MOSFETs since if modern devices are already operating close to the thermal limit, then it may not be possible to realise further drive current enhancements from scaling. The value of the backscattering coefficient depends on several factors including temperature, gate length and operating voltage. A summary of backscattering coefficients at room temperature is provided by Barral et al. [2007] for silicon MOSFETs, with typical values being between 0.2 and 0.4, for gate lengths below 100 nm and gate voltages of around 1 V.

2.5.5 Further High Field Effects

The large electric fields that are increasingly prominent in short-channel devices can result in hot carriers that can degrade the performance of the device. One such degradation mechanism is that of *impact ionisation* [Yang, 1988]. An electron (or hole) in the conduction (valence) band gains energy from the electric field as it is accelerated towards the drain. The hot carrier collides with the crystal lattice ionising an atom, thus creating an electron-hole pair. If the field is high enough, these carriers can gain sufficient kinetic energy to cause further impact ionisation events. Thus, a carrier multiplication process occurs in the high-field region near the drain. The minority carrier joins the current flowing to the drain increasing the drain current, whilst the majority carriers are collected by the substrate, giving rise to a substrate current. The substrate current can produce a voltage drop across the channel-substrate junction from the spreading resistance in the bulk, which can charge the semiconductor bulk, and forward-bias the junction at the source. This lowers the threshold voltage and increases the drain current. Breakdown of the MOSFET occurs when the longitudinal field exceeds the breakdown field, typically $\sim 10^5$ Vcm^{-1}, and the carrier multiplication process that results from impact ionisation causes avalanche breakdown of the p-n junction [Taur and Ning, 1998]. The impact ionisation rate is a significant parameter in high-frequency power amplifiers

since it determines the breakdown voltage of the device, and hence the maximum power that a technology can give [Waldron et al., 2003].

Band-to-band tunnelling (BTBT) occurs when the electric field in a reverse-biased pn-junction is high enough such that an electron can tunnel from the valence band on the p-side into the conduction band on the n-side. Typically, the electric field has to exceed $\sim 10^6$ Vcm^{-1} for band-to-band tunnelling to occur [Taur and Ning, 1998]. Whilst not a true short-channel effect, since it can also occur in long-channel devices, band-to-band tunnelling is becoming increasingly important as MOSFET dimensions are reduced. This is due to the fact that short-channel MOSFETs require a higher substrate doping density, reducing the depletion layer width at the drain, resulting in a higher electric field.

The band-to-band-tunnelling current is a strong function of the semiconductor bandgap energy. This could potentially prove to be a fundamental problem for narrow band-gap materials, such as germanium. The contribution of the BTBT current to the leakage current of modern VLSI devices is becoming increasingly important as the industry shifts to smaller technology nodes.

The electric field around a p-n junction can also be increased by the gate bias. Increasing the gate voltage can increase the field crowding in and around the p-n junction, which can increase the junction leakage of the MOSFET at the drain. This is referred to as *gate-induced drain leakage* (GIDL) and can be an important leakage mechanism in modern MOSFETs with thin gate oxides [Bouhada et al., 1998].

2.6 Process Considerations for High-κ Gate Dielectric Ge Channel MOSFETs

Perhaps the most important consideration in the fabrication of Ge channel MOSFETs is the choice of gate dielectric. Unlike silicon, the native oxide of germanium is

volatile and water soluble, making it unsuitable for MOSFET applications. The absence of a good quality native oxide that has to date, led to germanium only playing a bit-part role in the microelectronics industry. It is now inevitable that silicon dioxide will be replaced by an alternative dielectric with a higher permittivity than SiO_2, a so-called *high-κ dielectric*, in order to suppress the direct tunnelling component of the gate leakage.

In selecting an alternative high-κ dielectric to replace SiO_2, there are several properties other than the dielectric constant itself, which must be given due consideration. Barrier height, thermodynamic stability, interface quality, film morphology, gate compatibility, process compatibility and reliability have all been shown to be important [Wilk et al., 2001]. Taking these considerations into account, the most promising high-κ dielectrics to replace SiO_2, at the time of writing, in silicon MOSFETs are hafnium oxide (HfO_2) and zirconium oxide(ZrO_2). Consequently, the majority of published research on Ge MOSFETs have incorporated either HfO_2 or ZrO_2 as gate dielectric (e.g. Chui et al. [2002], Ritenour et al. [2003]) . However, it is important to remember that compared to silicon, the research on germanium MOSFETs with high-κ dielectrics is still in its infancy and alternative dielectrics, such as rare earth oxide CeO_2 [Nicholas et al., 2007], La_2O_3 [Mavrou et al., 2007] and Gd_2O_3 [Evangelou et al., 2007] may ultimately prove more fruitful.

The quality of the dielectric/germanium interface is one of the most important aspects regarding the fabrication and operation of Ge channel MOSFETs. One of the key successes of the SiO_2/Si interface is that it produces a low interface state density $D_{it} \sim$ 1-2 x 10^{10} $cm^{-2}eV^{-1}$ [Nicollian and Brews, 1982]. Interface state densities for high-κ dielectrics on both silicon and germanium are typically $\sim 10^{11}$ - 10^{12} $cm^{-2}eV^{-1}$. High interface state densities are detrimental to device performance since they can act as trapping centres for inversion layer carriers, increasing both Coulomb scattering and threshold voltage instabilities and can result in higher gate leakage current due to trap-

assisted tunnelling mechanisms [Loh et al., 2002, Houssa et al., 2000]. Not only is the magnitude of the interface trap density important, but also the distribution of these traps within the band gap. This is particularly problematic for Ge nMOSFETs, which have frequently been reported to have effective electron mobilities lower than the effective hole mobilities [Wu et al., 2007]. At the time of writing, this is believed to be due to the density of interface states rising steeply towards the conduction band edge, leading to Fermi-level pinning, although further understanding of this phenomenon is still needed [Martens et al., 2006].

The interface state density can often be reduced through *passivation* of the germanium surface prior to high-κ gate dielectric deposition. Several different passivation techniques have already been shown to improve the carrier effective mobility in Ge/high-κ MOSFETs [Whang et al., 2004, Wu et al., 2004, Zimmerman et al., 2006]. In addition, post deposition annealing has also been shown to be effective in reducing the interface state density and increasing carrier effective mobility [Kamata et al., 2006]. Whilst surface passivation and post-deposition annealing can be effective in reducing the interface state density, they can often lead to the formation of a thin interfacial layer between the Ge and high-κ dielectric, which can reduce the overall capacitance of the gate stack. This can be particularly problematic for HfO_2 and ZrO_2, in which the oxygen diffusion rate is high [Robertson, 2006].

In addition to the high density of interface states, there are also numerous publications reporting the presence of fixed charge within the high-κ dielectric layer, and defects such as oxygen vacancies are well documented [Tse et al., 2007, Gavartin et al., 2006]. Recently, it has been shown that fluorine implantation can be used to passivate these defects in silicon MOSFETs to improve threshold voltage stability, increase drive current and reduce charge trapping [Tse and Robertson, 2006, Inoue et al., 2005, Seo et al., 2005].

The role of remote phonon scattering is likely to be critical to the performance

of high-κ dielectric MOSFETs. Theoretical studies have shown that low-energy soft optical phonon modes, a direct consequence from the highly polarisable metal-oxygen bonds in the high-κ dielectrics can result in severe mobility degradations in high-κ dielectric MOSFET [Fischetti et al., 2001, Ferrari et al., 2006]. However, Chau et al. [2004] suggest that metal gates could be effective in screening these phonon modes.

In addition to replacing SiO_2 as gate dielectric, it is also apparent that poly-silicon will have to be replaced as the gate electrode [Misra, 2005]. Fundamentally, this is because depletion of the poly-silicon gate at the gate dielectric interface can add around 0.4 nm to the EOT. The depletion width at the interface can, in theory, be reduced by increasing the doping density of the poly-silicon gate. However, this is limited by the solid solubility of the dopants. In addition, the higher doping density and the reduced thickness of the gate dielectric means that dopant diffusion from the gate to the channel is becoming an important concern. Furthermore, there are also process related compatibility issues with incorporating high-κ dielectrics and poly-silicon gates. For example, there have been reports that SiH_4 etches HfO_2 during poly-silicon gate formation [Robertson, 2006]. Furthermore, defects at the poly-silicon/HfO_2 interface are believed to be responsible for Fermi-level pinning that results in high threshold voltages for pMOSFETs [Hobbs et al., 2004a,b], although it has been proposed that oxygen vacancies in the high-κ dielectric could also be responsible [Robertson, 2006].

The introduction of metal gates into CMOS devices is just as complicated as the replacement of SiO_2 as gate dielectric, and only the key points are mentioned here. One of the main advantages of poly-silicon gate electrodes, is that the Fermi-level is controlled by using either n or p-type dopants, thus allowing the threshold voltage of the MOSFET to be adjusted. However, for metal gates, it is likely that two different metals will have to be used - one with a work function close to the conduction band for nMOSFETs and one with a work function close to the valence band for pMOSFETs [Misra, 2005]. From a device operation point of view, this is more favourable than the

alternative of using a single metal with a midgap workfunction, since this will add either 0.3 V or 0.5 V to the threshold voltage of germanium or silicon MOSFETs, respectively. Since the electron affinities of silicon and germanium are close, the choice of metal workfunction for silicon nMOSFETs should also be satisfactory for Ge nMOSFETs. For Ge pMOSFETs, the choice of metal gate is likely to be a midgap metal for silicon, such as TiN or TaN.

The final significant issue with regards to the fabrication of germanium MOSFETs relates to dopant control. The lower activation energies (and therefore annealing temperatures) required to activate dopants in germanium offers an advantage of Ge MOSFETs when using HfO_2 gate dielectric over silicon. Hafnium oxide is usually deposited in the amorphous state, but can crystallise upon annealing to become poly-crystalline, which can lead to high gate leakage. Despite the low crystallisation temperature of HfO_2, it has been shown that it can be increased by adding additional elements such as aluminium or tantalum [Wu et al., 2003, Lu et al., 2005]. However, for the lower dopant activation temperatures required for germanium [Chui et al., 2005], the HfO_2 is expected to remain in the amorphous state. The solid solubility of n-type dopants remains a concern for Ge nMOSFETs, which are already appear to be hampered by an asymmetric distribution of interface states. The low solid solubility of n-type dopants in Ge results in a high series resistance, and reduces the drive current. Laser annealing has already been demonstrated as a means of increasing the drive current [Zhang et al., 2006], but it might be such that alternative solutions such as Schottky barrier Ge nMOSFETs or alternative high-mobility channel materials such as GaAs are required for n-channel devices.

Chapter 3

Experimental Techniques

3.1 Introduction

This section describes the experimental and analysis techniques used to obtain the results presented in this thesis. Electrical characterisation was conducted using a combination of current-voltage (I-V) and capacitance-voltage (C-V) measurements between room temperature and 4 K. The chapter splits conveniently into two parts: the first describes the equipment and measurement techniques; the second describes the analysis techniques necessary to determine the key attributes of MOS devices such as the effective carrier mobility, threshold voltage and interface state density. A comprehensive review of semiconductor characterisation techniques is provided by Schroder [2006] and the references within.

3.2 Electrical Characterisation Techniques

3.2.1 Equipment and Measurement Techniques

Electrical characterisation was carried out using a range of current-voltage (I-V) and capacitance-voltage (C-V) measurements. I-V measurements were performed using

an Agilent 4156C semiconductor parameter analyser. An Agilent 4284A LCR meter was used to perform C-V measurements over a range of frequencies from 20 Hz to 1 MHz.

All room temperature electrical measurements were carried out using a Karl Suss probe station, in which measurements were conducted inside an earthed Faraday cage to eliminate electromagnetic interference, including visible light, which can create electron-hole pairs in the active regions of a device. The device-under-test (DUT) was typically on a wafer, which was mounted on a metal plate inside the Faraday cage. The metal plate could be raised or lowered and a small vacuum pump was used to prevent the wafer moving relative to the plate during measurement.

MOS devices were characterised using a four terminal measurement technique. The four terminals being the insulating gate contact, the source and drain contacts and the substrate or back contact. Tungsten needle probes with fine positioning control were used to contact to the device contact pads. In general, the metal plate on which the wafer was positioned, was used as the substrate contact, even when a top substrate contact pad was present. The needle probes were connected to the test equipment by either co-axial or tri-axial cables.

A major concern in the characterisation of MOSFET devices, particularly with thin gate oxides, is that they can be easily destroyed by electrical discharge. To minimise these risks, the wafers were stored in anti-static containers and anti-static wristbands were used when handling the wafers. The devices were connected to the test equipment by connecting the substrate first, followed by the source and drain, with the gate being the final connection. Since the metal plate upon which the wafer sat could be raised or lowered, the microscope lamp could be switched off before connecting the device. After the measurement, the device was disconnected in reverse order.

3.2.2 Current-Voltage (I-V) Measurements

Current-voltage (I-V) measurements were performed using an Agilent 4156C semiconductor parameter analyser. There are two main types of I-V measurement for a MOSFET.

In the first, and arguably the most useful, of these techniques, the source and substrate are grounded, and a constant potential, V_{ds}, is applied to the drain. The drain current, I_{ds} is then measured in response to a varying bias, V_{gs}, applied to the gate. These device characteristics are usually referred to as the *transfer characteristics* of the device and can be used to determine many useful device properties.

The transfer characteristics can be used to examine the subthreshold behaviour of the MOSFET, since the subthreshold slope is determined from a logarithmic plot of I_{ds} versus V_{gs} and is a measure of how well the device turns on and off. The transfer characteristics also show the I_{on}/I_{off} ratio, an important figure-of-merit in device performance. The transconductance, g_m, follows from differentiating the linear drain current with respect to the gate voltage, from which, one can also determine the device threshold voltage, V_t.

The transfer characteristics are typically determined at two different drain biases, one being a low drain bias and one being high. The low drain bias measurement, typically for $V_{ds} = \pm$50 mV, can be combined with a split-CV measurement to determine the low-field effective carrier mobility, μ_{eff}, perhaps the single-most important device parameter. The transfer characteristics at high drain bias are used to examine how the device would perform in a CMOS circuit, and the applied drain bias, would depend upon the technology node that the device was designed for, typically being between $\pm$ $1V$ and $\pm$ $2.5V$ for the devices in this work. A direct comparison between the transfer characteristics measured at low and high drain bias can be used to determine the effects of drain-induced-barrier-lowering (DIBL), an important factor in short-channel devices.

The *output (or drain) characteristics* resemble a more classical FET measurement. In this measurement, the source and substrate are grounded once again and a constant voltage, V_{gs}, is applied to the gate. The drain current, I_{ds}, is then measured in response to a varying drain bias, V_{ds}, and the measurement can be used to compare the drive currents of different devices. However, for this to be a fair comparison, it is important to compare devices at the same gate overdrive, $V_{gt} = V_{gs} - V_t$ and not the same gate voltage. Thus, these measurements were performed after the threshold voltage had been determined from an $I_{ds} - V_{gs}$ measurement. In addition, the drain characteristics can also show whether a device suffers from short-channel effects or self-heating.

3.2.3 Capacitance-Voltage (C-V) Measurements

Capacitance-voltage (C-V) measurements were performed using an Agilent 4284A LCR meter, which was controlled using a PC and a TestPoint application written by the author. During a C-V measurement, the LCR meter applies an oscillating ac signal superimposed on a DC bias at one terminal of the DUT, and the response is measured through a second terminal of the DUT. The amplitude of the ac signal was small, typically 25 mV, and the frequency was varied between 20 Hz and 1 MHz.

The conventional C-V measurement consists of a two-terminal measurement on a capacitor, using the gate and substrate contacts. In the MOS capacitor, the gate forms the top plate of the parallel capacitor and the charge in the semiconductor substrate forms the bottom plate. Since the distance of the charge in the semiconductor substrate from the gate oxide interface depends on the applied bias, the capacitance also changes with the applied bias.

Figure 3.1 shows the simplified form of the equivalent circuit for the MOS capacitor measured by the LCR meter. The MOS capacitor can be represented by either a parallel $G_p - C_p$ equivalent circuit, where G_p is the conductance and C_p is the capac-

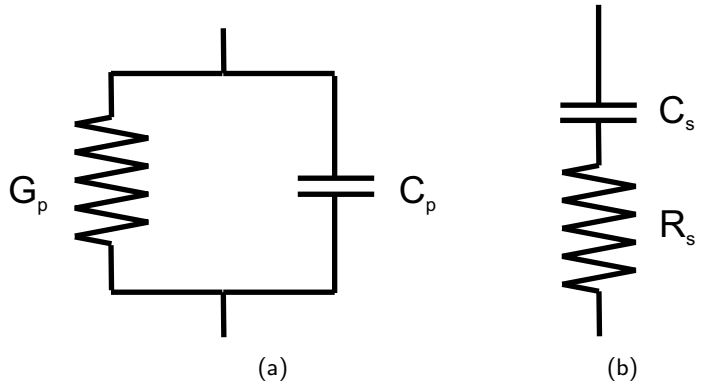


Figure 3.1: Simplified equivalent MOS circuits measured in (a) parallel and (b) series mode.

itance, or a series $R_s - C_s$ equivalent circuit, where R_s is the series resistance of the MOS capacitor and C_s its capacitance. In this work, the MOS capacitance was measured using the parallel equivalent circuit mode, since this provided a useful measure of the gate leakage current. However, in cases where the series resistance is particularly high, then the series equivalent circuit mode may be more appropriate.

In the parallel circuit mode, the admittance, Y_p is given in terms of the impedance, Z_p, and the angular measurement frequency, ω, by:

$$Y_p = \frac{1}{Z_p} = G_p + j\omega C_p \tag{3.1}$$

Rearranging Equation 3.1, the impedance can then be written as:

$$Z_p = \frac{G_p}{G_p^2 + (\omega C_p)^2} - \frac{j\omega C_p}{G_p^2 + (\omega C_p)^2} \tag{3.2}$$

Thus, the LCR determines the conductance and the capacitance of the sample by applying a sinusoidal signal to the MOS capacitor and measuring the phase shift of the output signal, where the zero phase shift component of the output signal gives the conductance and the 90 $^\circ$ phase shift gives the capacitance [Schroder, 2006].

3.2.4 Cryogenic Measurements

Low temperature electrical characterisation was carried out using a Desert Cryogenics TT-Probe Station. Low temperature electrical characterisation was carried out using both liquid nitrogen and liquid helium in this work.

The cryogenic fluid travels from a self-pressurised storage dewar along a transfer tube, with a flow-controlling needle valve, and enters the cold finger of the probe station, which is thermally connected to the sample stage. A small diaphragm pump connected to the exhaust is used to suck the, now gaseous, cryogen out of the cold finger. The base temperature for nitrogen and helium was typically around 73 K and 3.8 K, although a lower base temperature could be achieved by using a stronger exhaust pump.

The temperature of the probe station was controlled by changing the flow rate of the liquid cryogen through the needle valve and also by applying power to a resistive heater. Silicon diodes were used to measure the temperature of both the sample stage and the radiation shield.

The DUT was mounted on a two-inch square metal plate, which allowed devices to be characterised at wafer level, albeit if the wafers had to be cleaved into smaller pieces first, thus eliminating the need for bonding devices into chip packages. Unlike the Karl Suss probe station for room temperature characterisation, it was not possible for the sample to be held in place by vacuum, so care had to be taken not to move the sample when making the connections between the probe needles and the contact pads. In the absence of a top substrate contact pad, the sample stage was used to provide the substrate contact. It was found that due to the wafer not being held under vacuum to the metal plate, that the electrical connection was not as good compared to that achieved on the Karl Suss probe station. Consequently, an InGa eutectic solution was applied to the back of the wafer, which was then mounted onto a thin copper sheet, which was in turn mounted onto the sample stage. Comparable results between the two

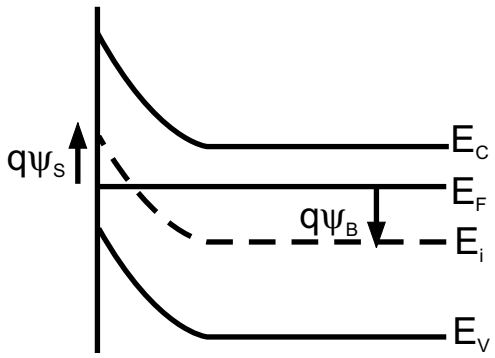


Figure 3.2: Definition of threshold voltage for a pMOSFET. The condition for threshold is when $q\psi_B = q\psi_S$.

probe stations was subsequently achieved.

3.3 MOS Analysis Techniques

3.3.1 Threshold Voltage

The threshold voltage of a MOSFET is loosely defined as the voltage at which the device turns on. It is important to know the threshold voltage in order to make a fair comparison in performance between devices, and also to extract other important device parameters including, for example, series resistance and effective channel length.

The most commonly accepted definition of the threshold voltage is the point where the potential difference between the Fermi level, E_F, and the intrinsic level, E_i, at the surface is equal in magnitude and opposite in sign to that in the bulk [Schroder, 2006]. This is shown in Figure 3.2, which depicts the threshold voltage for a pMOSFET.

Unfortunately, this condition is very difficult to determine experimentally. As a result, researchers have come up with several alternative definitions of threshold voltage that are somewhat easier to determine. A recent review of threshold voltage extraction methods is provided by Ortiz-Conde et al. [2002], and thus, is not repeated here.

The method used in this work, is that proposed by Tsuno et al. [1999] and is based on the linear extrapolation of the transconductance (Figure 3.3). The method is

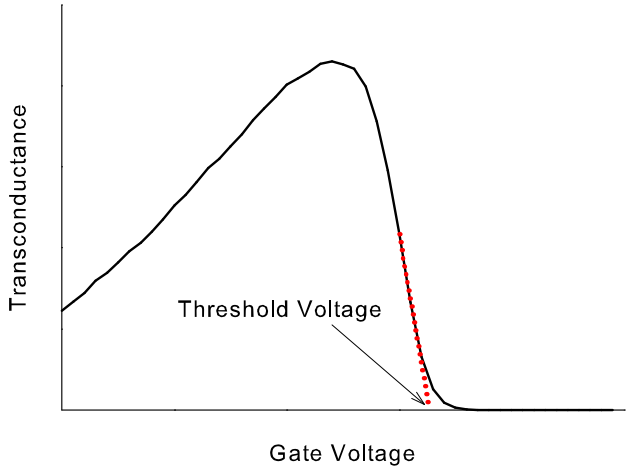


Figure 3.3: Extraction of threshold voltage by linear extrapolation of the transconductance.

used because not only does it meet the physical requirements of the threshold condition but it is valid for both long and short channel MOSFETs, relevant to this work. Furthermore, it has the added advantage that it does not make the assumption that the mobility is independent of V_{gs}, unlike, for example, the method based on the linear extrapolation of the drain current.

The transconductance is defined as:

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}} \tag{3.3}$$

The threshold voltage by this technique is found by extrapolation of the transconductance from the point of its maximum slope back to zero. Since this point is usually found in the tail of the $I_{ds} - V_{gs}$ curve, the threshold voltage is closer to the point where the drain current is negligibly small. Care must be taken, since the technique is sensitive to both noise and the resolution of the equipment.

The technique assumes that in the threshold region, the mobility is dominated by Coulomb scattering, and is thus, proportional to the inversion carrier sheet density, N_s:

$$\mu_{eff} = \alpha N_s \tag{3.4}$$

The inversion carrier sheet density is given by the approximation:

$$N_s = \frac{C_{ox}(V_{gs} - V_t)}{q} \tag{3.5}$$

Substituting Equations 3.4 and 3.5 into Equation 2.5, and differentiating with respect to V_{gs}, gives the transconductance as:

$$g_m = \frac{W}{L}\frac{2\alpha}{q}C_{ox}^2(V_{gs} - V_t)V_{ds} \tag{3.6}$$

where the small V_{ds}^2 term has been neglected. As expected, the transconductance exhibits a linear dependence on the gate voltage, and when extrapolated back to zero, gives $V_{gs} = V_t$.

3.3.2 Subthreshold Slope

The subthreshold slope, S, sometimes referred to as the subthreshold swing, is measured directly from the gradient of the $log_{10} I_{ds} - V_{gs}$ graph, and for a long channel devices is given, in units of mV/dec, according to [Wolf, 1995]:

$$S = \left(\frac{d(log_{10} I_{ds})}{dV_{gs}}\right)^{-1} = 2.3\frac{k_B T}{q}\left(1 + \frac{C_{dep}}{C_{ox}}\right) \tag{3.7}$$

where C_{dep} is the depletion capacitance per unit area.

Typically, the subthreshold slope will be in the region of 70 - 100 mV/decade, and it is not possible to have a subthreshold slope lower than approximately 60 mV/dec at room temperature. The subthreshold slope is an important parameter for MOSFETs since it determines how quickly the device can turn on/off. Importantly, the subthreshold slope can be dramatically increased in the presence of a high interface trap density, since the capacitance associated with interface states, C_{it}, is in parallel with the depletion

capacitance, C_{dep}, and more of the charge on the gate will be used to image the interface states and not to form the inversion layer.

3.3.3 Drain-Induced Barrier Lowering

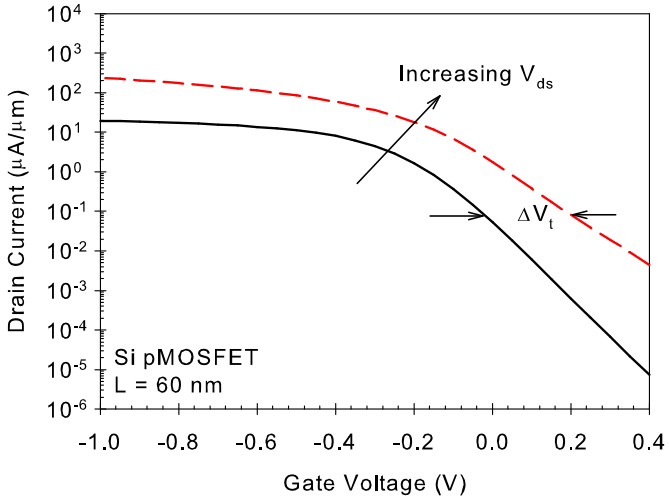
Drain-induced barrier lowering, or DIBL, is defined as:

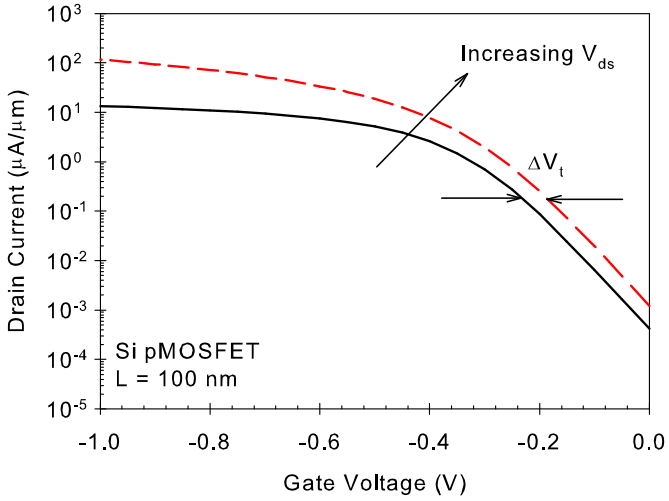
$$DIBL = \frac{\Delta V_t}{\Delta V_{ds}} \tag{3.8}$$

In order to accurately measure DIBL, it is necessary to accurately determine the threshold voltage. The technique used to determine threshold voltage has to be accurate for both low and high drain bias and also for long and short devices. A simpler method to determine DIBL, is to measure the voltage shift directly from a plot of two $log_{10} I_{ds} - V_{gs}$ measured at two different values of V_{ds}. It has been proposed that DIBL should be measured at a given value of I_{ds}, typically 0.1 μA/μA [Wolf, 1995]. However, provided that DIBL is measured in the linear region of the subthreshold plot, the choice of the subthreshold current value makes little difference. Care must also be taken when the subthreshold slope is degraded due to punchthrough. This is shown in Figure 3.4 for two silicon pMOSFETs with gate lengths of 60 nm and 100 nm investigated in the current work. The degradation of the subthreshold slope with increasing drain bias indicates that the presence of punchthrough in this particular device, and so, for this device, and other such cases the DIBL value was extracted at the threshold voltage.

3.3.4 Channel Length and Series Resistance

Many of the models used to describe the drain current for a MOSFET are somewhat simplified. For example, two assumptions that are frequently made are that there is no deviation from the mask-designed gate length and that the parasitic resistances associated with the source and drain contact regions can be neglected. Of course, in





(b)

Figure 3.4: Example of DIBL extraction for a Si pMOSFET (a) with and (b) without the effects of punchthrough.

practice, neither is true, and it is important that both the change from the mask-designed channel length, ΔL, and the series resistance, R_{sd} are determined, such that the effective carrier mobility can be compared accurately between devices. Such corrections are particularly important for gate lengths below a few microns, when these effects become increasingly significant.

There are several reasons why the effective channel length, L_{eff}, of a MOSFET can differ from the mask-defined gate length, L. Dopants from the source/drain regions can extend underneath the gate stack, either from the initial ion implantation step to form the source/drain regions, or by diffusion during subsequent annealing steps in the process. In addition, fringing effects of the electric field at the end of the gate electrode result in non-vertical electric fields at the end of the channel. Further differences can arise from lithographic errors and the ambiguity in defining *exactly* where the channel starts and ends can. Finally, the effective channel length can also depend on the bias conditions of the MOSFET. The effective channel length is related to the gate length by:

$$L_{eff} = L - \Delta L \tag{3.9}$$

In a real MOSFET, since the resistance of the source and drain contact regions is not negligible, some of the applied voltage, V_{ds}, is dropped across these resistances. This so-called *series resistance* can then lead to a lower than expected value of the drain current.

There are several contributions to the series resistance of a MOSFET [Schroder, 2006]. The most important contribution to the series resistance is the resistance of the heavily doped source and drain regions underneath the metal contacts. This is strongly dependent upon the doping dose and activation anneal temperature, with the resistance usually decreasing as both of these factors are increased. Where an extension implant

is used, this will add significantly to the total resistance. The second most important contribution is the 'spreading resistance' where carriers travel from the heavily doped semiconductor regions, which are typically tens of nanometres thick, into an inversion layer, just a few nanometres thick. In addition, there will also be a resistance associated with the source and drain contact pads, although since these are usually made of metal, this component is negligibly small. However, the Schottky barrier between the metal contact pad and the heavily doped semiconductor surface will have a resistance, but this is likely to be small since the depletion layer formed will be very narrow, and carriers will tunnel through easily.

As a result, the potential difference along the channel is usually modified to include the effects of these resistances according to:

$$V'_{ds} = V_{ds} - I_{ds}R_{sd} \tag{3.10}$$

where R_{sd} is the combined series resistance of the source and drain regions.

3.3.4.1 Linear Regression of Resistance versus Channel Length

The drain current, I_{ds}, of a MOSFET operating in the linear regime, taking into account the effects of series resistance and deviations to the channel length, can be expressed as:

$$I_{ds} = Q_{inv}\mu_{eff}V'_{ds}\frac{W}{L_{eff}} \tag{3.11}$$

The measured resistance of a device, R_m, is given by [Chern et al., 1980]:

$$R_m = \frac{V_{ds}}{I_{ds}} = R_{channel} + R_{sd} \tag{3.12}$$

where the channel resistance, $R_{channel}$, is given by:

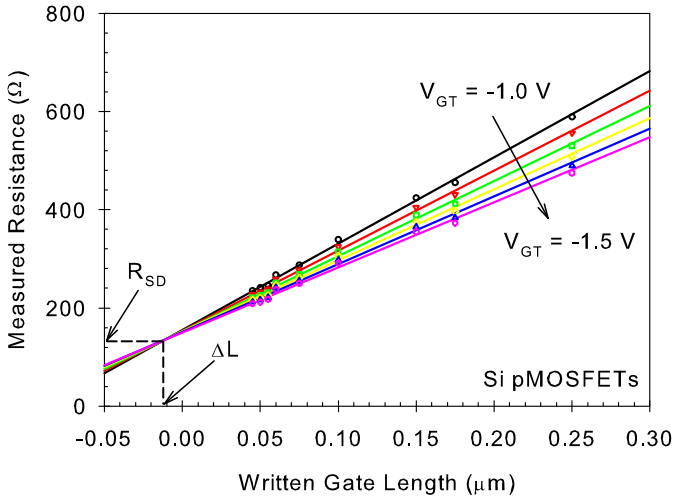


Figure 3.5: Example of series resistance and effective length extraction using the 1st regression method for bulk silicon MOSFETs.

$$R_{channel} = \frac{V'_{ds}}{I_{ds}} \tag{3.13}$$

Combining equations 3.11, 3.12 and 3.13, the measured resistance of the device is given by:

$$R_m = \frac{L - \Delta L}{W \mu_{eff} Q_{inv}} + R_{sd} \tag{3.14}$$

where the effective gate length, L_{eff} has been written as $L - \Delta L$.

Equation 3.14 suggests that plotting R_m versus L for different values of gate overdrive should produce a series of straight lines that should meet at the point where $L = \Delta L$ and $R_m = R_{sd}$. An example of this is shown in Figure 3.5 for Si pMOSFETs investigated in the present work..

Despite the technique being favoured for its simplicity, it is possible to extract inaccurate values of R_{sd} and ΔL. The main error is due to the difficulty of accurately defining a point where all of the lines meet. In such a case, this usually indicates that R_{sd} and ΔL vary with gate overdrive, when the technique assumes they do not. Secondly, since the resistance of all devices must be measured at the same gate overdrive, the

threshold voltage must be extracted using a technique that accurately compares both long and short devices. Finally, there is also some ambiguity in choosing which devices to use in the method. Large devices can have a large leverage on the gradient of the regression line and can introduce large errors in ΔL. However, it has been reported by some authors that the measured resistance is lower than expected for the shortest devices, causing the straight lines to deviate from linearity. Usually, the technique works best for devices with $L \leq$ 1 - 2 μm.

3.3.4.2 Double Regression Method

Terada and Muta [1979] proposed an alternative method to extract R_{sd} and ΔL based on the linearity between R_m and L.

Writing equation 3.14 in the form:

$$R_m = AL + R_{sd} - A\Delta L; A = \frac{1}{W\mu_{eff}qN_s} \tag{3.15}$$

For a given value of V_{gt}, a plot of R_m versus L will have a gradient of A and an intercept of $R_{sd} - A\Delta L$. If the slope and the intercept are determined from the linear regression and plotted against each other for different values of V_{gt}, then this should produce another straight line with a gradient of $-\Delta L$ and an intercept of R_{sd}, which can be determined from a second regression. Whilst this technique overcomes the problem of finding a suitable intercept using the first regression method, it suffers from many of the same problems. For example, if R_{sd} and ΔL are functions of V_{gt}, then the second regression may not produce a straight line.

3.3.5 Capacitance Measurements

Figure 3.6 shows the equivalent circuit model of a two-terminal MOS capacitor. The capacitance of an MOS capacitor is measured as a function of voltage, to elicit

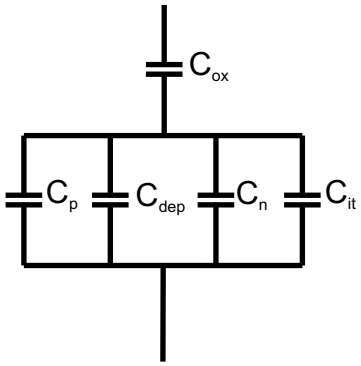


Figure 3.6: Equivalent circuit model showing the capacitances in the MOS capacitor.

information about the various charges within the structure. Capacitance-voltage measurements can be used to determine the oxide capacitance, C_{ox}, used to determine the oxide (or equivalent oxide) thickness. They can also be useful in determining the doping profile in the semiconductor bulk, the inversion charge density, the interface trap density and the oxide charge density. In theory, one often refers to the change in the MOS capacitance as a function of gate bias. However, in the two-terminal configuration, measuring the change in the MOS capacitance in response to an applied substrate bias, proved to be 'less noisy' [Palmer, 2001].

The MOS capacitance is given by:

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_p + C_n + C_{dep} + C_{it}} \tag{3.16}$$

where C_{ox} is the oxide capacitance, given by:

$$C_{ox} = \frac{\epsilon_0 \kappa A}{t_{ox}} \tag{3.17}$$

where ϵ_0 is the permittivity of free space; κ is the relative permittivity of the dielectric; A is the area of the capacitor; and t_{ox} is the dielectric thickness.

For an n-type capacitor, C_n is the accumulation layer capacitance; C_p is the inversion layer capacitance; C_{dep} is the depletion layer, or bulk, capacitance; and C_{it} is

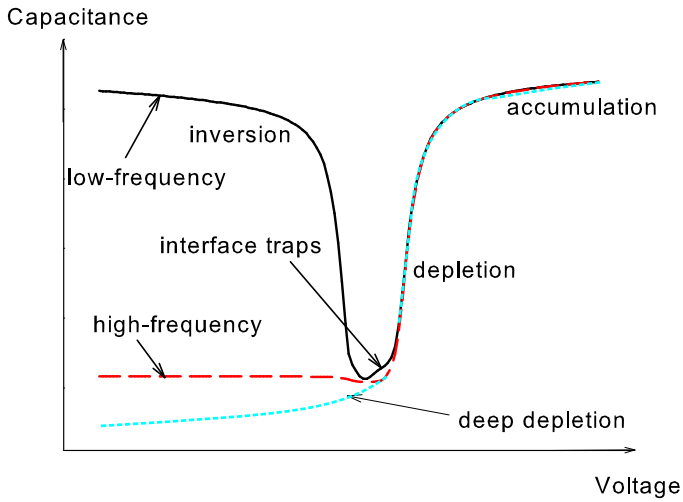


Figure 3.7: Example of low and high frequency C-V characteristics of an n-type MOS capacitor.

the capacitance due to interface traps. For a p-type capacitor, C_n and C_p are reversed, such that C_p is the accumulation capacitance and C_n is the inversion capacitance.

Figure 3.7 shows an example of a C-V measurement on an n-type MOS capacitor measured at low and high frequency. The variation of the capacitance with gate voltage can be split up into the three regions: accumulation, depletion and inversion. When a large positive bias is applied to the gate, the majority carriers (electrons) in the substrate, are attracted to the semiconductor-oxide interface, and the MOS capacitance is high. As the gate voltage is made increasingly negative, the electrons in the substrate are repelled away from the interface, forming a depletion region under the gate. If the gate voltage is made increasingly negative, then the band-bending at the semiconductor surface makes it energetically favourable for minority carriers (holes) to form an inversion layer at the surface. In the MOS capacitor, holes are formed by generation processes in the semiconductor bulk and depletion regions, and then drift to the semiconductor surface, which has a certain timescale. Thus, to measure an inversion capacitance in an MOS capacitor, it is important that the frequency is low enough such that the minority carriers are able to respond. If the frequency is too high for minority carriers to respond, the

inversion capacitance will be approximately equal to the depletion capacitance, as shown in Figure 3.7. For capacitance measurements made on MOSFETs (see Section 3.3.6), this constraint is more relaxed, since the minority carriers are supplied from the heavily doped source and drain regions. Figure 3.7 also shows the deep depletion capacitance, a non-equilibrium capacitance measurement, which results when the gate voltage is swept quickly during a high frequency C-V measurement.

3.3.5.1 Doping Profile

The capacitance of an MOS capacitor in depletion can be used to obtain information about the doping profile in the semiconductor body. The doping concentration can be measured as a function of depth from the interface, and is given by [Schroder, 2006]:

$$N_D(W) = \frac{2}{q\kappa_s\epsilon_0 A^2 d(1/C^2)/dV} \tag{3.18}$$

The depth from the interface, W, is given by:

$$W = \kappa_s\epsilon_0 A\left(\frac{1}{C} - \frac{1}{C_{ox}}\right) \tag{3.19}$$

In fact, the technique does not exactly measure the doping profile. Instead, it is the effective carrier concentration that is measured. This is because in deriving Equation 3.18, the contribution of minority carriers is neglected and the depletion region is assumed to be totally depleted of majority carriers to a depth W. The model assumes that the ac signal is used to vary the number of donor ions at the edge of the space-charge region, which gives rise to the differential capacitance. However, it is the majority carrier electrons in the substrate which move in response to the signal. The effective carrier density is approximately equal to the majority carrier density, and so $N_D(W)$ in Equation 3.18 should be replaced by $n(W)$.

Since the majority carrier density is measured instead of the doping profile, an error is introduced in the determination of the doping profile. This measurement error is governed by the Debye length, L_D. The Debye length is the length scale for which majority carriers are able to respond to an electric field, and is given by:

$$L_D = \sqrt{\frac{k_B T \kappa_s \epsilon_0}{q^2(p+n)}} \tag{3.20}$$

3.3.5.2 Flatband Voltage

The flatband voltage is defined as the value of V_g at which the conduction and valence bands edges are flat at the surface. That is, the surface potential, ψ_s, is equal to zero. The flatband voltage depends upon the difference between the metal and semiconductor work function and also on the density of the different types of charges within the oxide, Q_{ox} according to:

$$V_{fb} = \phi_{ms} - \frac{Q_{ox}}{C_{ox}} \tag{3.21}$$

The flatband voltage can be determined from the flatband capacitance, C_{fb} which is given by Equation 3.22:

$$C_{fb} = \frac{\kappa_s}{L_D} \tag{3.22}$$

where κ_s is the relative permittivity and L_D is the Debye length.

However, this method of determining the flatband voltage requires accurate knowledge of the average doping density. Alternatively, the flatband voltage can be determined experimentally from a high-frequency C-V curve by plotting $(1/C_{hf})^2$ versus V_g. The lower knee of this curve occurs at $V_g = V_{fb}$ [Schroder, 2006]. Since it was difficult to determine the exact position of the knee, differentiating the $(1/C_{hf})^2$ curve yields a second curve for which the maximum value of the slope occurs at V_{fb}, which

can thus be determined by a second differentiation.

3.3.5.3 Interface Traps

There are several techniques for determining the interface trapped charge in either an MOS capacitor or a MOSFET, which include the high-low frequency C-V method, the Terman method, the conductance method and charge pumping, to name but a few. A good review of the varying techniques, including a discussion of their relative strengths and weaknesses, can be found in Schroder [2006].

For an MOS capacitor, the most popular methods are the high-low frequency method and the conductance method. Charge-pumping is the preferred method when determining the interface trapped charge using a MOSFET. It is important to remember that no single method exists that gives a *perfect* measurement of the interface trapped charge. Care must be taken when interpreting the results from all of the methods and the limitations of each technique have to be understood.

The low and high frequency C-V method is based on two C-V measurements of an MOS capacitor: one at a low frequency, such that both interface traps and minority carriers are able to respond to the ac signal, and one at a high frequency, where they cannot. In depletion, such that C_p and C_n are assumed to be zero, the low frequency capacitance is determined from the equivalent circuit in Figure 3.6 as:

$$\frac{1}{C_{lf}} = \frac{1}{C_{ox}} + \frac{1}{C_{dep} + C_{it}} \tag{3.23}$$

In depletion, the high frequency capacitance, where it is assumed that C_{it} is negligible, is given by:

$$\frac{1}{C_{hf}} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}} \tag{3.24}$$

The interface trap capacitance is defined in terms of the interface trap density,

D_{it}:

$$D_{it} = \frac{C_{it}}{q^2} \tag{3.25}$$

The factor of q^2 appears in Equation 3.25 rather than the more commonly used q in order to make the units on both sides of the equation consistent [Schroder, 2006].

Eliminating C_{dep} from Equations 3.23 and 3.24, the interface trap density can be written in terms of the measured high and low frequency according to:

$$D_{it} = \frac{C_{ox}}{q^2}\left(\frac{C_{lf}/C_{ox}}{1 - C_{lf}/C_{ox}} - \frac{C_{hf}/C_{ox}}{1 - C_{hf}/C_{ox}}\right) \tag{3.26}$$

The appeal of the technique is based in the simplicity of the measurement. Traditionally a quasi-static C-V measurement has been used for the low frequency and 1 MHz used for the high frequency. However, when considering dielectrics with high leakage currents and high interface trap densities, considerable error can be introduced. Firstly, it is not often possible to perform a quasi-static C-V measurement on large area capacitors with high-κ dielectrics due to the intolerable leakage current. It is possible to perform a low-frequency measurement at a frequency of 100 Hz, for example, but even at this low frequency, not all interface traps are able to respond to the ac signal, thus underestimating the contribution to C_{it}. Furthermore, for an MOS capacitor with a high interface trap density, it is still possible for a significant number of interface traps to respond to the ac signal, even at 1 MHz. Hence, the low and high frequency C-V method, would, at best, provide a lower limit on the interface trap density.

The interface trap density can also be measured from the parallel conductance. Peaks in the $G - V$ curve are indicative of the 'loss' processes associated with the presence of interface traps. An approximate expression for the average interface trap density is given by [Schroder, 2006]:

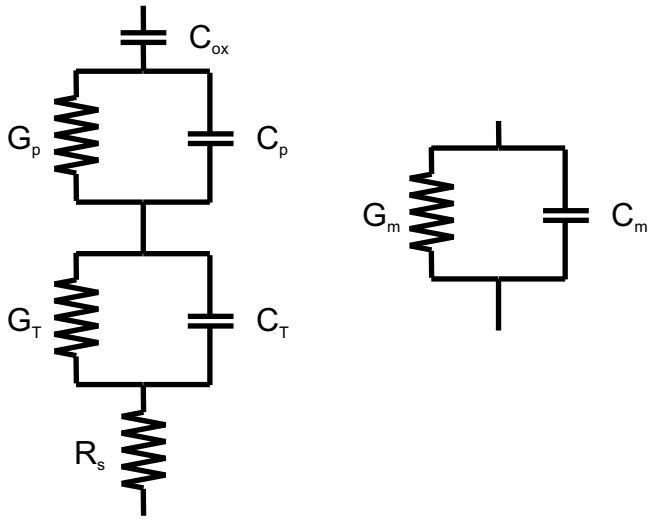


Figure 3.8: (a) Equivalent MOS circuit model to account for an interfacial dielectric layer and (b) the equivalent circuit measured by the LCR meter.

$$D_{it} \approx \frac{2.5}{q}\left(\frac{G_p}{\omega}\right)_{max} \tag{3.27}$$

Whilst the technique can not be used to study the interface trap density distribution within the bandgap, its simplicity is particularly useful in wafer mapping.

3.3.5.4 Series Resistance

When the series resistance of an MOS capacitor is high, the accumulation capacitance can become frequency dependent. Whilst every effort should be made to reduce the effects of series resistance, for example, by forming good Ohmic contacts to the capacitor, Kwa et al. [2003] demonstrated that series resistance effects can be observed in samples with a poor quality interfacial layer dielectric sandwiched between the semiconductor surface and the gate dielectric.

They modified the equivalent circuit model to include the effects of this so-called 'lossy' interfacial dielectric. The modified equivalent diagram is shown in Figure 3.8, in which the interfacial layer is represented by the parallel circuit consisting of C_T and R_T ($=1/G_T$).

Following analysis of the circuit diagrams, the corrected capacitance, C_c, and conductance, G_c can be expressed in terms of the measured quantities, C_m and G_m, according to:

$$C_c = \frac{(\omega^2 C_m C_E - G_m^2 - \omega^2 C_m^2)(G_m^2 + \omega^2 C_m^2) C_E}{(\omega^2 C_E^2)[G_m(1 - R_s' G_m) - \omega^2 R_s' C_m^2]^2 + (G_m^2 + \omega^2 C_m^2 - \omega^2 C_m C_E)^2} \tag{3.28}$$

$$G_c = \frac{[G_m(1 - R_s' G_m) - \omega^2 R_s' C_m^2](G_m^2 + \omega^2 C_m^2) C_E}{(\omega^2 C_E^2)[G_m(1 - R_s' G_m) - \omega^2 R_s' C_m^2]^2 + (G_m^2 + \omega^2 C_m^2 - \omega^2 C_m C_E)^2} \tag{3.29}$$

where the following quantities are defined as:

$$C_E = \frac{G_T^2 + \omega^2 C_T^2}{\omega^2 C_T} = \frac{-C_{ox}(G_{ma}^2 + \omega^2 C_{ma}^2)}{\omega^2 (C_{ma}^2 - C_{ma} C_{ox}) + G_{ma}^2} \tag{3.30}$$

$$R_s' = \frac{G_{ma}}{G_{ma}^2 + \omega^2 C_{ma}^2} \tag{3.31}$$

The corrected capacitance-voltage curves can then be used for a reliable extraction of the MOS capacitor parameters such as interface trap density, flatband voltage and doping density.

3.3.6 Split C-V

The split C-V method was determined by Koomen [1973] as a way of measuring the inversion capacitance and the depletion/accumulation capacitance separately. The technique requires a MOSFET as the test structure, and thus the proposed method of applying the voltage to the substrate to reduce the noise in the C-V measurement, as mentioned for the MOS capacitor, is no longer applicable.

There are two configurations for the split C-V experiment. In the first configuration, the substrate is held at Earth, and the capacitive response to the voltage applied at the gate is measured at the tied source and drain contacts. This is the *gate-channel*

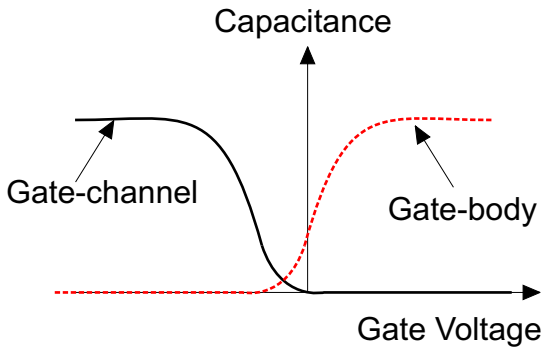


Figure 3.9: Typical split C-V characteristics for a silicon pMOSFET.

branch of the split C-V curve. In the second configuration, the tied source and drain contacts are held at Earth and the response is measured through the substrate contact. This is the *gate-body* branch of the split C-V curve (Figure 3.9).

In the gate-channel branch of the split C-V, the inversion capacitance is very close to the oxide capacitance, for a surface channel device, assuming a low density of interface traps, since the source and drain regions are able to supply minority carriers for an inversion layer to form. As the bias is made more positive, the gate-channel capacitance decreases, until it reaches zero in accumulation. In the gate-body branch of the split C-V measurement, the capacitance in accumulation is very close to the oxide capacitance. As the gate voltage is made increasingly negative, the capacitance decreases as majority carriers are repelled further from the surface. The capacitance finally decreases to zero when the MOSFET is biased into inversion, and the depletion charge is screened from the gate by the inversion charge.

The split C-V technique is used to obtain both the inversion and depletion charge densities, by integration of the gate-channel and gate-body branches of the split C-V measurement. Accordingly, the inversion charge is given by:

$$Q_{inv} = \int_{\infty}^{V_{gs}} C_{gc}(V_{gs}) dV_{gs} \tag{3.32}$$

The depletion charge is given by:

$$Q_{dep} = \int_{V_{fb}}^{V_{gs}} C_{gb}(V_{gs}) dV_{gs} \tag{3.33}$$

The effective vertical field is then given by a linear combination of the inversion and depletion charges, according to:

$$\varepsilon_{eff} = \frac{Q_{dep} + \eta Q_{inv}}{\kappa_s \epsilon_0} \tag{3.34}$$

where η is an empirical parameter, equal to $1/2$ for electrons and $1/3$ for holes.

3.3.7 Mobility

Perhaps, the single most important device parameter, is the carrier mobility. The carrier mobility is critical to the drain current of the MOSFET. In semiconductor physics, there are several ways of defining the mobility. For example, one can measure the conductivity mobility, the Hall mobility, the magnetoresistance mobility or the effective mobility. A good discussion of the different types of mobility can be found in Schroder [2006]. In a MOSFET, the relevant mobility is the effective mobility, since this allows one to probe the carrier mobility within a narrow region at the semiconductor/oxide surface, where additional scattering mechanisms, such as Coulomb scattering from oxide charges and interface states and surface roughness scattering may be dominant. These additional scattering mechanisms typically result in the effective mobility being significantly lower than the bulk mobility values.

The effective mobility is usually determined from a combination of I-V and C-V measurements on a long-channel device, in order to minimise short-channel and edge effects. In the linear region, the drain current of a MOSFET is given by:

$$I_{ds} = \frac{W}{L} \mu_{eff} Q_{inv} V'_{ds} \tag{3.35}$$

where Q_{inv} is the inversion charge density and V'_{ds} is the intrinsic drain voltage to account for the source/drain series resistance.

There are two common ways of determining the inversion charge density. In the first method, the inversion charge density is approximated by:

$$Q_{inv} = C_{ox} (V_{gs} - V_t) \tag{3.36}$$

In this case, Equation 3.35 is equivalent to Equation 2.5 with the small V_{ds}^2 term neglected. However, this approximation is rather limited. Firstly, the method requires an accurate extraction of the device threshold voltage. Furthermore, the assumption that the inversion charge depends linearly on the gate voltage is particularly inaccurate near threshold. Secondly, the inversion layer capacitance is not necessarily equal to the oxide capacitance, C_{ox}, which does not include the effects of poly-Si gate depletion, quantum confinement of the inversion layer, or any parasitic channels.

A more accurate way of determining the inversion charge density is by employing the split C-V method discussed in Section 3.3.6, where the inversion charge was given by Equation 3.32.

When measuring the inversion charge density from split C-V measurements, care must be taken in selecting the measurement frequency, which must satisfy the condition [Schroder, 2006]:

$$f \ll \frac{1}{2\pi\tau_{gc}} = \frac{4R_{sh,ch}}{2\pi C_{gc} L_{eff}^2} \tag{3.37}$$

where τ_{gc} is the channel time constant, $R_{sh,ch}$ is the channel sheet resistance, C_{gc} is the gate-channel capacitance and L_{eff} is the channel length.

The equation for the effective mobility still assumes that the diffusion contribution to the current is negligible and that there is no dependence on the drain voltage. This is generally not true near threshold. Sodini et al. [1982] introduced a correction

such that the MOSFET drain current is given by:

$$I_{ds} = W\mu_{eff}Q_{inv}\varepsilon(x) - W\mu_{eff}\frac{k_BT}{q}\frac{dQ_{inv}}{dx} \tag{3.38}$$

where $\varepsilon(x)$ is the longitudinal electric field and the Einstein relation for the diffusion current was used:

$$D_p = \frac{k_BT}{q}\mu_{eff} \tag{3.39}$$

When the correction for the diffusion current is made, the effective mobility can be expressed as:

$$\mu_{eff} = \frac{L}{W}\frac{I_{ds}}{V'_{ds}}\frac{1}{Q_{inv} - \frac{k_BT}{q}C_{gc}(V_{gs})} \tag{3.40}$$

3.3.8 Carrier Velocity

The ultimate limit to the performance of a MOSFET is thought to be determined by the Richardson thermal injection velocity, v_T, from the source into the channel. Therefore, to determine how close to the thermal limit a MOSFET is operating, it is important to be able to determine the velocity of carriers in a MOSFET. Furthermore, determination of the carrier velocity will also elucidate whether or not velocity overshoot is occurring.

The average carrier velocity can be determined from the intrinsic transconductance, g_{mi} according to:

$$v_{eff} = \frac{g_{mi}}{WC_{ox}} \tag{3.41}$$

where C_{ox} is the oxide capacitance per unit area.

If velocity overshoot is occurring in a device, it is most likely to occur in the high-field regions of the drain. However, whilst this would not be particularly beneficial

to the potential performance of the device, velocity overshoot at the source, would be beneficial. Thus, Lochtefeld et al. [2002] proposed an alternative method to determine the carrier velocity, to extract the carrier velocity at a position closer to the source, x_0, than the transconductance method.

They propose the following relation for the effective carrier velocity at the source:

$$v_{eff} = \frac{I_{on}}{WQ_{inv}(x_0)} \tag{3.42}$$

where $Q_{inv}(x_0)$ is the inversion carrier density near the source. This quantity is very difficult to determine in short devices due to relatively large overlap and fringing capacitances, uncertainties in the effective channel length, and nonuniform charge distribution along the channel. However, in strong inversion and in the gradual channel approximation, $Q_{inv}(x_0)$, for a short device should correspond closely to that for a long-channel device, which can be accurately measured and is given by:

$$Q_{inv}(x_0)_{long} = \int_{\infty}^{V_{gs}} C_{sd}(V_{gs})dV_{gs} \tag{3.43}$$

for a p-channel device. Thus, the carrier velocity near the source in a short-channel MOSFET can be determined from a split C-V measurement on a long-channel device. To accurately determine the inversion carrier sheet density in the short device, corrections have to be made to the upper integration limit, in order to account for the differences in threshold voltage between two devices due to DIBL and V_t roll off, and also due to the voltage drop on the source resistance. Thus, the inversion carrier density near the source for a pMOSFET, can be determined according to:

$$Q_{inv}(x_0)_{short} = \int_{\infty}^{V_{gs}-\Delta V_t - I_{on}R_s} C_{sd}(V_{gs})dV_{gs} \tag{3.44}$$

Chapter 4

Development of $Si_{1-x}Ge_x$ Capacitors with High-κ Dielectric and Metal Gate

4.1 Introduction

Described in this chapter is a study of the electrical properties of $Si_{1-x}Ge_x$/high-κ dielectric/metal gate capacitors, with a view to developing surface-channel $Si_{1-x}Ge_x$ MOSFETs employing a high-κ/metal gate stack. Sputtered hafnium oxide was used as gate dielectric and both titanium nitride and tungsten were investigated as metal gates. The work was carried out as part of a joint collaboration involving Warwick University, Chalmers University (Sweden) and AMO (Germany) within the framework of the SINANO Network of Excellence under the guidance of the author.

4.2 Process Outline and Specification

Five wafers with pseudomorphic $Si_{0.75}Ge_{0.25}$ layers were grown by low pressure chemical vapour deposition (LPCVD) at Warwick University on (100) n^- Si substrates. Each wafer was grown with a thin silicon cap with varying thicknesses of 1 nm, 2 nm, 3 nm, 4 nm and 5 nm, since it was unknown as to how much of this cap would be consumed during capacitor processing. The $Si_{0.75}Ge_{0.25}$ layers were all 7 nm thick, in order to stay below the critical thickness for strain relaxation [People and Bean, 1985]. An epitaxial silicon control wafer was also grown as part of the wafer batch.

Prior to hafnium oxide deposition, the wafers underwent an RCA standard pre-deposition cleaning step [Kern, 1993], in order to remove any impurities and undesired native oxide on the surface of the wafers, according to the following recipe:

1. 10 minutes at 70 °C in 1:1:5 of NH_4OH:H_2O_2:H_2O.
2. 2% HF dip for 10 seconds.
3. 10 minutes at 70 °C in 1:1:5 of HCl:H_2O_2:H_2O.
4. 2% HF dip for 10 seconds.

Hafnium oxide was then deposited by sputtering at Chalmers University, with a nominal physical thickness of around 10 nm, expected to result in a capacitive equivalent thickness (CET) of between 3 nm and 4 nm. The reason for using a sputtered hafnium oxide gate dielectric was twofold. Firstly, this particular oxide had previously been demonstrated to exhibit excellent gate leakage characteristics in preliminary experiments conducted within the SINANO Network [Engstrom, 2007]. Secondly, it was hoped that it might alleviate the need for elaborate surface passivation techniques, usually required before the deposition of CVD-based high-κ dielectrics to achieve a high quality dielectric [Van Elshocht et al., 2004]. Following hafnium oxide deposition, all sputtered wafers

were given a post-oxide deposition anneal (PDA), at 800 °C for 10 minutes in N_2, in order to repair some of the damage at the $Si_{1-x}Ge_x/HfO_2$ interface caused by the sputtering process.

Capacitor dots were then fabricated at AMO on the $Si_{1-x}Ge_x/HfO_2$ layers using one of three different metal gates, since it was unknown which metal gates would provide the appropriate workfunction values and stability on the hafnium oxide dielectric. Prior to capacitor fabrication, each wafer was cleaved into $\sim$2 cm^2 pieces, to permit a split in the final anneal condition. The fabrication process can be describes by 3 main process steps:

1. Active Area Definition - Circular capacitor dots were defined using a two-stage lithography process to define capacitor dots with diameters of 100 μm, 200 μm, 300 μm and 400 μm. The first stage defines the active areas, whilst the second step was used to achieve a good undercut, required for the metal lift-off process.

2. Metal Gate Deposition and Lift-Off - Three different metal gates were used in this investigation:

   (a) Thick TiN - 100 nm TiN

   (b) Thin TiN - 20 nm TiN capped with 100 nm in-situ doped (5 x 10^{18} cm^{-3}) n^+ poly-silicon

   (c) 50 nm W

   All metal gates were deposited by sputtering. Following metal gate deposition, a lift-off process using acetone, propanol and de-ionised water was used to remove undesired metal between the capacitor dots.

3. Annealing - The final stage in the capacitor fabrication was annealing. In this investigation, four different anneal conditions were used:

(a) 'As-deposited' - These samples were unannealed

(b) 'Post-metallisation anneal' or PMA - These samples were annealed at 425 °C for 30 minutes in N_2/H_2 and represented the optimum anneal conditions for the metal gate, following previous studies using metal gates on SiO_2 [Schmidt et al., 2005].

(c) 'Rapid-thermal anneal 1' or RTA1 - These samples were annealed at 800 °C for 30 seconds in Ar and would represent the anneal temperature that could be used in a reduced thermal budget CMOS process incorporating silicided source and drain regions.

(d) 'Rapid-thermal anneal 2' or RTA2 - These samples were annealed at 950 °C for 30 seconds in Ar and would represent the temperature used to activate dopants in the source and drain regions of a typical CMOS process [Hallstedt et al., 2006].

The capacitors were processed as two separate batches. The first batch consisted of the thick TiN gate capacitors, whilst the second batch consisted of capacitors with either thin TiN or W gates. It should also be noted that there were no PMA samples with thin TiN metal gate, since these anneal conditions were insufficient to activate the doping in the poly-silicon layer, thus effectively leaving a high resistance poly-silicon layer on top of the thin TiN layer. Furthermore, the 'as-deposited' thin TiN samples actually have a thick TiN gate, also due to the activation problem, but in addition, this would serve as a comparison between the processing conditions of the thick TiN and thin TiN, fabricated as separate batches.

In order to realise $Si_{1-x}Ge_x$ channel devices with high-κ dielectric and metal gate, it is imperative that the gate stack can withstand the high thermal budgets of RTA1 and RTA2. The electrical characterisation results will now be presented and discussed with a view to fabricating surface-channel $Si_{1-x}Ge_x$ transistors with high-κ

Table 4.1: Capacitor specification for $Si_{1-x}Ge_x/HfO_2$/Thick TiN capacitors.

Wafer Description	PDA	As-dep	PMA	RTA1	RTA2
Si Control	Y	Y	Y	Y	Y
SiGe 1 nm Cap	Y	Y	Y	Y	Y
SiGe 2 nm Cap	Y	Y	Y	Y	Y
SiGe 3 nm Cap	Y	Y	Y	Y	Y
SiGe 4 nm Cap	N	Y	Y	Y	Y
SiGe 5 nm Cap	Y	Y	Y	Y	Y

Table 4.2: Capacitor specification for $Si_{1-x}Ge_x/HfO_2$/Thin TiN capacitors.

Wafer Description	PDA	As-dep	RTA1	RTA2
Si Control	Y	Y	Y	Y
SiGe 1 nm Cap	Y	Y	Y	Y
SiGe 2 nm Cap	Y	Y	Y	Y
SiGe 3 nm Cap	Y	Y	Y	Y

Table 4.3: Capacitor specification for $Si_{1-x}Ge_x/HfO_2$/W capacitors.

Wafer Description	PDA	As-dep	PMA	RTA1	RTA2
Si Control	Y	Y	Y	Y	Y
SiGe 1 nm Cap	Y	Y	Y	Y	Y
SiGe 2 nm Cap	Y	Y	Y	Y	Y
SiGe 3 nm Cap	Y	Y	Y	Y	Y

dielectric/metal gate stack.

The full capacitor batch specification is summarised in Tables 4.1, 4.2 and 4.3.

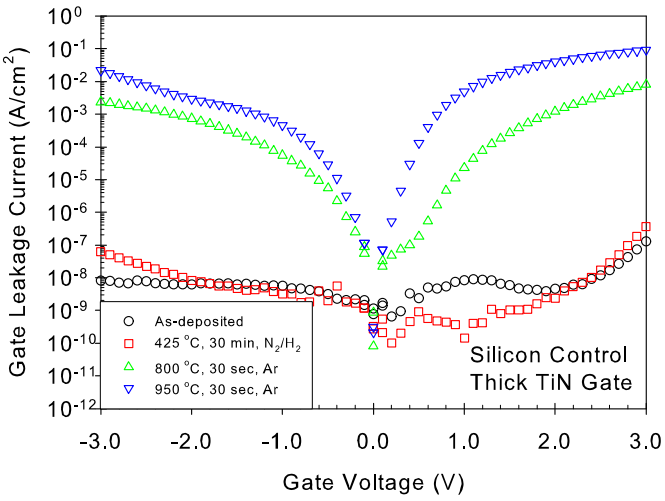


Figure 4.1: Gate leakage characteristics for Silicon control capacitors with thick TiN gate with different anneal conditions.

4.3 Electrical Characterisation Results

4.3.1 Gate Leakage

4.3.1.1 Effect of Anneal Conditions

The gate leakage behaviour of the $Si_{1-x}Ge_x/HfO_2$/metal gate capacitors was found to be strongly dependent on the final anneal condition for each of the metal gates investigated. Figure 4.1 shows typical gate leakage characteristics for silicon control capacitors with thick TiN gate following each of the anneal conditions.

Excellent gate leakage behaviour was generally observed for all as-deposited samples and also for samples following the PMA at 425 °C for 30 minutes in N_2/H_2. In fact, gate leakage current following the low temperature anneal at 425 °C, was frequently observed to be lower than that for the as-deposited samples, particularly in the low voltage regime. This could be attributed to two factors. Firstly, the interface trap density is reduced following the anneal, which would reduce the component of the gate leakage current due to trap-assisted tunnelling. Secondly, during the annealing stage, regrowth of an $HfSi_xO_y$ interfacial layer can occur at the HfO_2/Si interface, which increases the

oxide thickness, thus providing a thicker barrier for tunnelling.

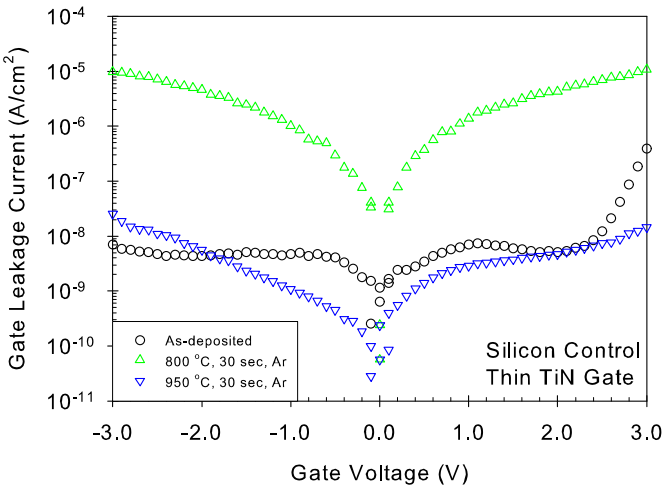
Following RTA at either 800 °C or 950 °C, the gate leakage current increases by several orders of magnitude across the entire bias range, with the gate leakage on samples annealed at 950 °C having the highest leakage. This could be due to the thermal instability of the TiN/HfO_2 interface, leading to a reaction at this interface, reducing the effective thickness of the dielectric and increasing the leakage current. A similar effect with regards to the thermal stability of TiN on SiO_2 has already been reported [Lemme et al., 2006]. Similar results were observed on all SiGe capacitors with thick TiN gates, irrespective of thin silicon cap thickness.

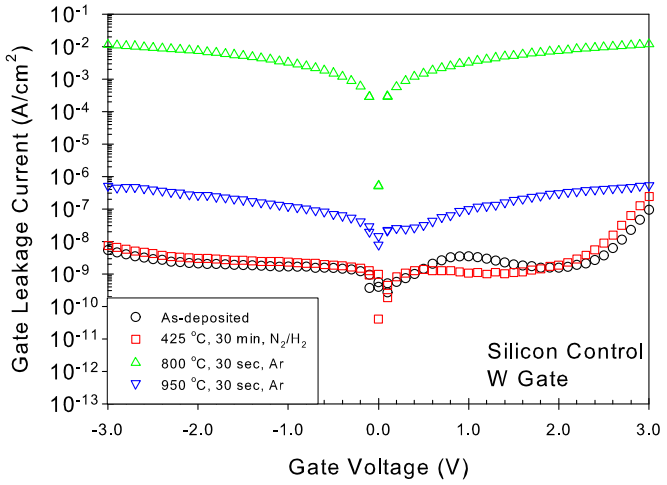
The gate leakage characteristics of silicon capacitors with both thin TiN and W gates are shown in Figure 4.2. In general, gate leakage currents lower than 10^{-4} Acm^{-2} were frequently observed on both Si and SiGe capacitors across the entire range of measured gate biases, thus confirming the improved thermal stability of both thin TiN gates and W gates, with respect to thick TiN, following high temperature anneals, and demonstrating their potential as metal gates that can withstand the high CMOS thermal budgets for device processing.

Interestingly, for these samples, both the thin TiN and W gate capacitors show higher gate leakage current following an 800 °C anneal, compared to the 950 °C anneal. No immediate explanation for this was apparent. Consequently, the experiment was briefly repeated for Si/HfO_2/W capacitors that were annealed at 700 °C, 800 °C and 950 °C. This experiment confirmed that the gate leakage increased with temperature, as expected (not shown). The earlier anomaly is likely to be due non-uniform thickness of the hafnium oxide across the wafer.

4.3.1.2 Effect of Gate Material

The gate leakage characteristics of as-deposited silicon control capacitors with different metal gates are shown in Figure 4.3. Low gate leakage currents, typically less





(b)

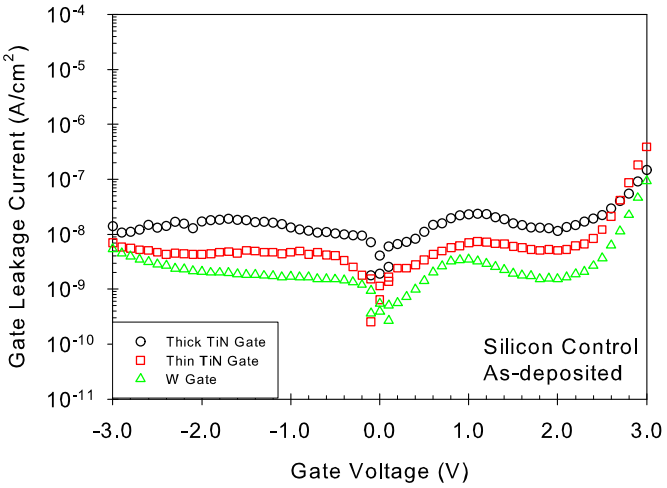
Figure 4.2: Gate leakage charactereristics of silicon capacitors with (a) thin TiN and (b) W metal gates following different anneal conditions.

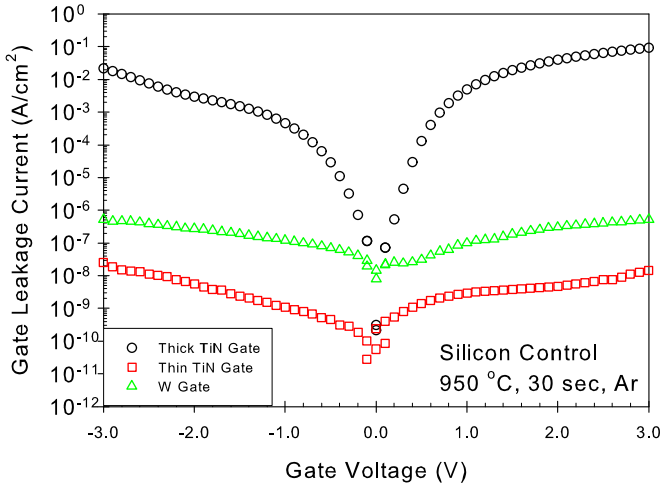
than $10^{-7} - 10^{-8}\ Acm^{-2}$ are observed for gate voltages below 3 V. Above 3 V, the gate leakage increases by several orders of magnitude. In general, the gate leakage current is lower for negative gate voltages than positive gate voltages, due to the depletion region that is formed in the MOS capacitor for negative gate voltages. It should also be noted that, in general, the gate leakage was higher for the 'thick TiN' capacitors compared to the 'thin TiN' samples, despite them both having supposedly identical gate stacks. This difference is likely to be due to slight changes in process conditions between different batches, since the 'thin TiN' batch was processed at a later date and thus highlights the difficulties associated with achieving repeated results between successive batches, especially when multiple laboratories are used.

Interestingly, the gate leakage behaviour of the as-deposited capacitors, showed a small leakage hump at low gate voltages, for all metal gates investigated. This effect was typically observed for positive gate voltages around 1.5 V. Occasionally, the effect was also observed for negative gate voltages, but the hump was significantly less pronounced. The effect could arise from a transient charge-trapping effect, similar to that recently reported in W/GeO_xN_y/Ge capacitors [Chui et al., 2006].

Similar effects were also observed on SiGe capacitors and the gate leakage characteristics of SiGe capacitors with a 2 nm silicon cap are shown in Figure 4.4. Once again, low gate leakage currents are observed for gate voltages below 2.5 V. Low gate leakage currents are observed for thin TiN and W gates on both the silicon control and SiGe capacitors with a 2 nm silicon cap. Good capacitors on these samples were frequently found with gate leakage currents below $10^{-3}\ Acm^{-2}$ for gate voltages up to 3.5 V.

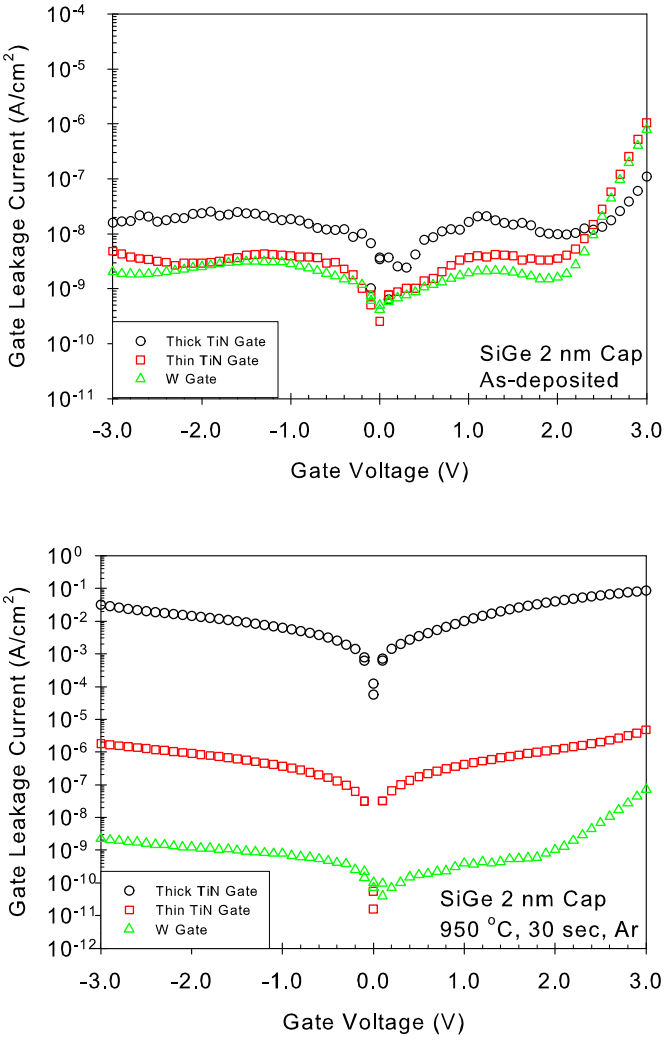
Gate leakage currents were typically observed to be several orders of magnitude higher on all capacitors with a thick TiN gate following the high temperature anneal. As previously discussed, this could be due to a thermal instability at the TiN/HfO_2 interface. The thermal stability of both W and the thin TiN layers on HfO_2 appears to





(b)

Figure 4.3: Gate leakage charactereristics for silicon capacitors with different metal gates that are (a) as-deposited and (b) following RTA anneal at 950 °C for 30 seconds in Ar.



(b)

Figure 4.4: Gate leakage charactereristics for SiGe capacitors with 2 nm Si cap and different metal gates that are (a) as-deposited and (b) following RTA anneal at 950 °C for 30 seconds in Ar.

be much greater than that of the thick TiN layer. The differences in thermal stability between the thick TiN films and the thin TiN films needs to be fully understood. For example, is the thickness of the TiN layer the dominant factor, or is the role of the poly-silicon capping layer that is responsible for increasing the thermal stability.

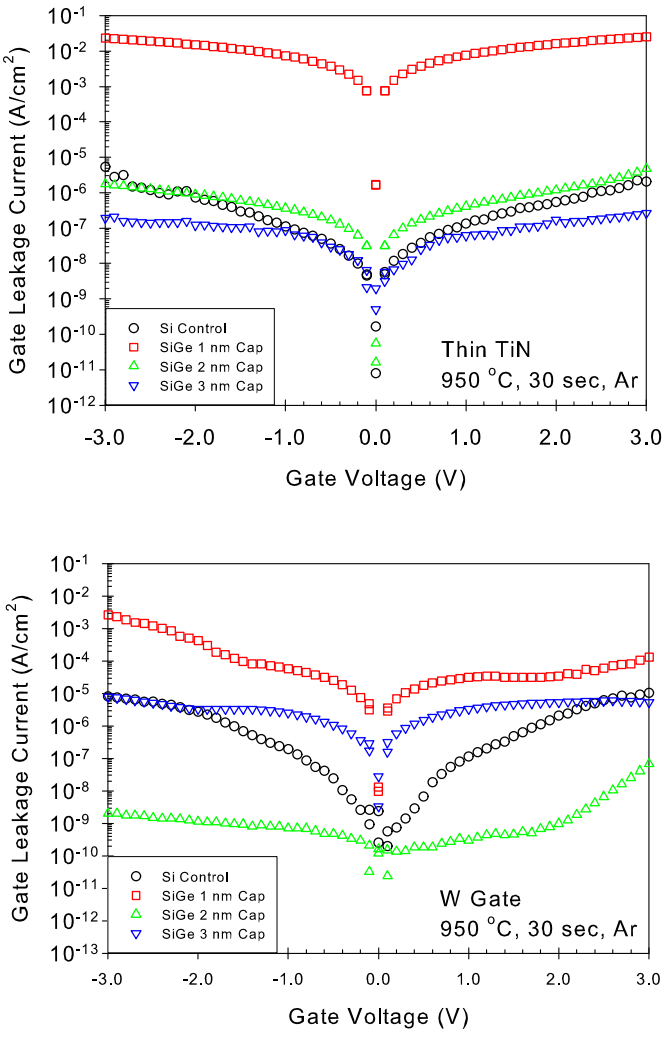
4.3.1.3 Effect of Silicon Cap Thickness

The role of the silicon cap thickness and the presence of germanium at the interface was also investigated. Figure 4.5 shows the gate leakage characteristics for SiGe capacitors after RTA2 at 950 °C with different silicon cap thicknesses for thin TiN and W metal gates.

Due to the large parameter space investigated, it is difficult to draw systematic conclusions. However, the gate leakage appears to be higher for the SiGe capacitors with 1 nm Si cap for both gate materials. Similar observations were also found for SiGe capacitors with thick TiN metal gate (not shown). This is likely to be due to the presence of germanium atoms near the interface, which might lead to the formation of volatile GeO_x and generate a higher density of interface traps [Park et al., 2005]. For the other SiGe capacitors, it is likely that a thin silicon cap remains following the capacitor fabrication and the formation of GeO_x is suppressed.

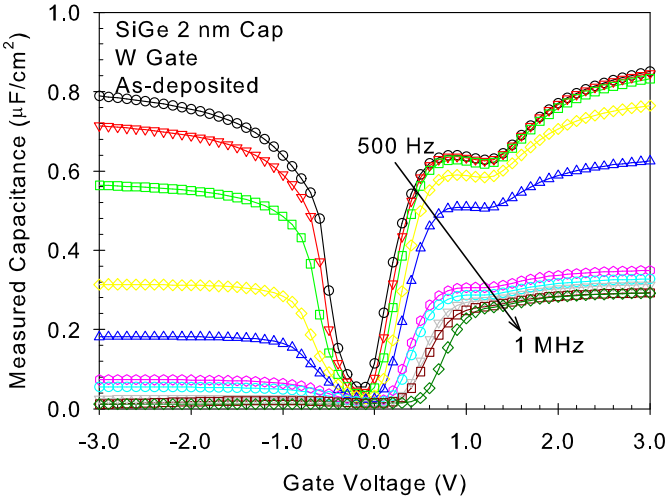
4.3.2 Capacitance-Voltage Characteristics

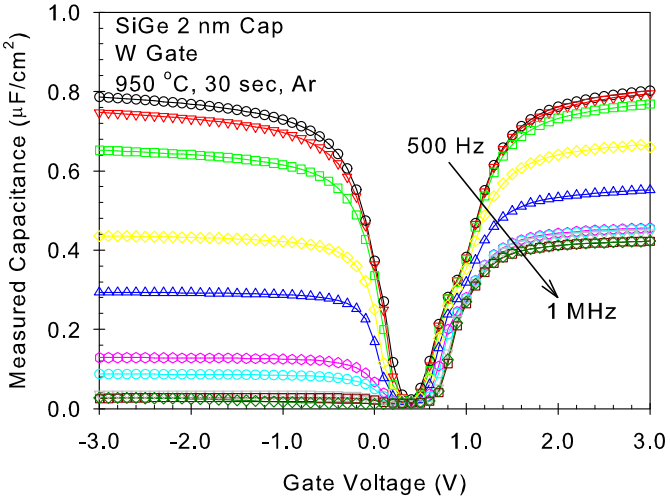
Typical capacitance-voltage (C-V) characteristics for SiGe capacitors with 2 nm Si cap and W gate are shown in Figure 4.6. Figure 4.6 shows the characteristics of both as-deposited capacitors and those following RTA2 anneal at 950 °C for 30 sec in Ar for different measurement frequencies. Similar C-V characteristics were observed for all $Si_{1-x}Ge_x$ capacitors. Neither the as-deposited or annealed capacitors show particularly well-behaved C-V characteristics, with a strong dependence on the measurement frequency. The frequency dependence in inversion and depletion is to be expected and



(b)

Figure 4.5: Gate leakage charactereristics for SiGe capacitors showing the effects of the silicon cap for (a) thin TiN and (b) W metal gates after RTA2.





(b)

Figure 4.6: Measured C-V characteristics for (a) as-deposited and (b) 950 °C annealed SiGe capacitors with 2 nm cap and W gate.

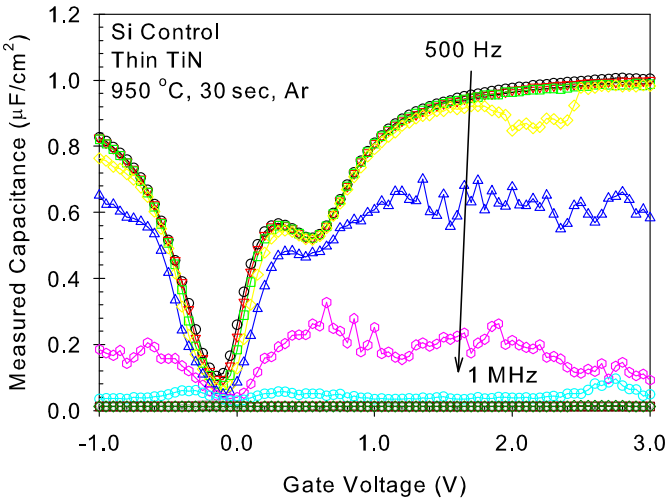


Figure 4.7: Capacitance-voltage characteristics for Si control capacitors with thin TiN gate following RTA2.

is understood by the the frequency-dependent response of minority carriers and interface traps, respectively. The frequency dependence in accumulation is likely to be due to a series resistance effect associated with the presence of an interfacial SiO_2 layer and is discussed in Section 4.3.2.1. The as-deposited C-V characteristics show a clear distortion in the region of the C-V curve between accumulation and flatband. This is due to the presence of a high density of interface traps (see Section 4.3.4). The interface trap density is significantly reduced by annealing, evident by the distortion almost disappearing in the C-V characteristics of annealed samples.

Unfortunately the C-V characteristics for SiGe capacitors with both thick and thin TiN metal gates were not well-behaved. The C-V characteristics for thick TiN gates following high temperature anneal were dominated by excessive gate leakage (not shown) making it impossible to extract reliable parameters such as interface trap density, from the C-V curves. The behaviour of the thin TiN gates was rather unusual. It was earlier shown that the gate leakage for these capacitors was similar to that for W gate capacitors. Figure 4.7 shows the C-V characteristics for silicon control capacitors with thin TiN gate following RTA at 950 °C. There appears to be a strong suppression of the

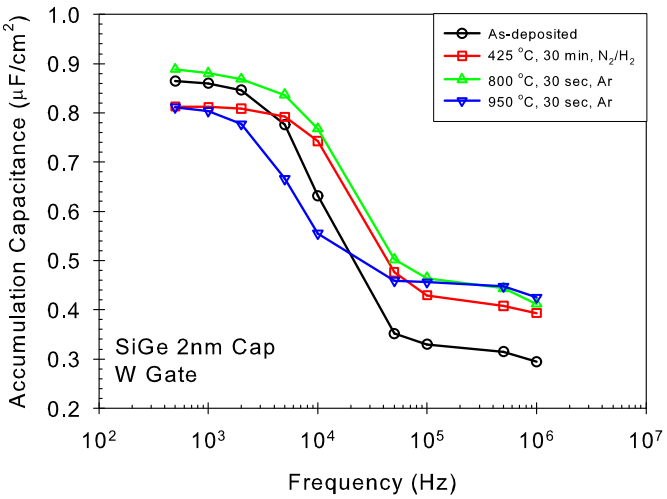


Figure 4.8: Accumulation capacitance for SiGe capacitors with 2 nm cap and W gate as a function of measurement frequency and anneal condition.

high-frequency C-V characteristics. The reason for this remains unknown and requires further understanding. It is possible that the doping density in the poly-Si capping layer (either the nominal doping density was insufficient or the doping has not been well activated). This might result in a high series or contact resistance, which might suppress the high frequency C-V characteristics. Consequently, since it was not possible to extract reliable parameters for either the thick TiN or thin TiN gates, the analysis from the C-V characteristics shall be limited to the W gate capacitors.

4.3.2.1 Series Resistance and Corrected C-V Characteristics

Typical C-V characteristics were shown in Figure 4.6(a) for as-deposited SiGe capacitors with 2 nm Si cap and W gate. The oxide capacitance, C_{ox} is, to a good approximation, given by the capacitance in strong accumulation. However, it is evident that the accumulation capacitance is a strong function of the measurement frequency. Figure 4.8 shows the measured accumulation capacitance, C_{ma} as a function of frequency, measured at a gate voltage, $V_g = +3.5$ V. The accumulation capacitance approaches a constant value for measurement frequencies below 2 kHz. This behaviour

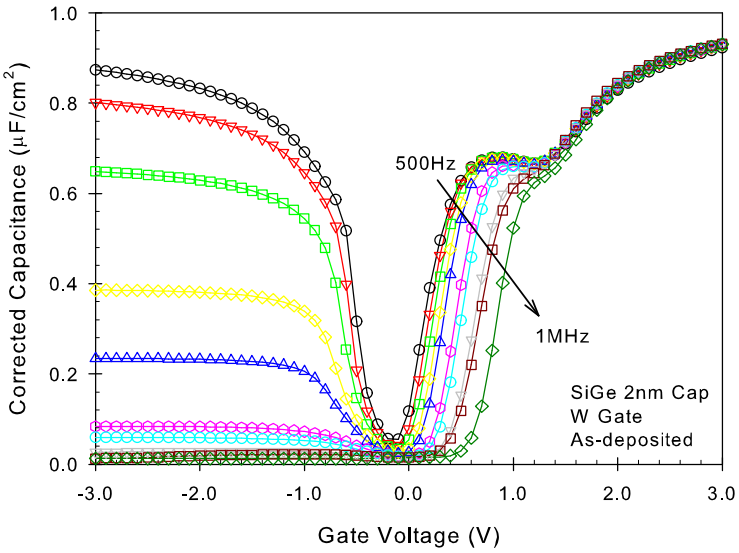


Figure 4.9: Corrected C-V characteristics for as-deposited SiGe capacitors with 2 nm cap and W gate.

was observed for all SiGe capacitors for all anneal conditions. Typically, the accumulation capacitance should not depend strongly upon the measurement frequency, since the majority carriers in the substrate should be able to respond to the ac signal. However, it has previously been reported, for both high-κ dielectrics and silicon dioxide that a frequency dispersion in the accumulation capacitance can arise from a series resistance effect due to the presence of a thin interfacial layer at the gate/dielectric interface [Kwa et al., 2003]. The model proposed by Kwa et al. [2003] was applied to the measured W gate capacitors in Figure 4.6(a) to recover the 'true' C-V characteristics. The results of this correction are shown in Figure 4.9. Typical values for the series resistance ranged from $\sim 10^4 \Omega$ at low frequencies (e.g. 500 Hz) to $\sim 10^2 \Omega$ at 1 MHz. Figure 4.9 shows that following the correction, no information is 'lost' from the C-V curves, the only significant difference being that the different frequency C-V curves now coincide in accumulation. The corrected C-V curves were used to extract various parameters for the W gate capacitors, including interface trap density and flatband voltage.

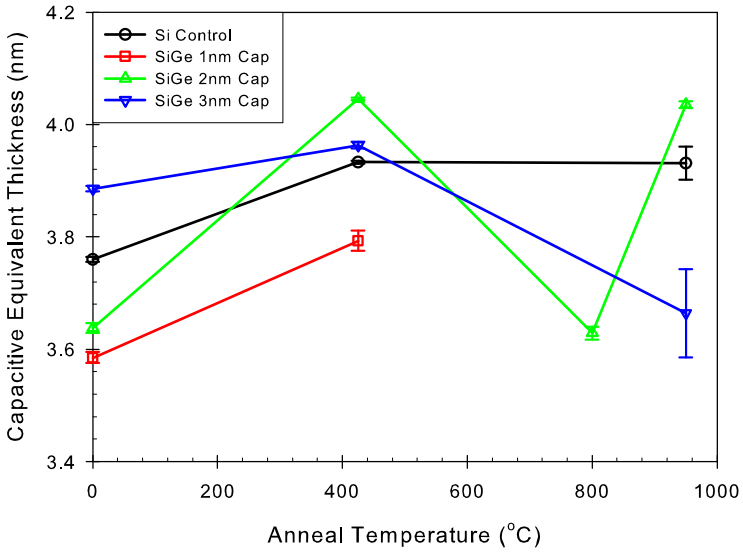


Figure 4.10: Capacitive equivalent thickness (CET) for $Si_{1-x}Ge_x$ capacitors as a function of anneal temperature.

4.3.3 Oxide Capacitance and Thickness

The oxide capacitance was extracted from the accumulation capacitance according to the technique outlined by Ghibaudo et al. [2000]. The lowest frequency C-V curve (typically 500 Hz) was used to determine the oxide capacitance in order to alleviate the series resistance effects reported in Section 4.3.2.1. The capacitive equivalent thickness (CET) follows by treating the capacitor as a parallel plate with a dielectric constant of 3.9. Figure 4.10 shows the CET values measured on both Si and SiGe capacitors as a function of anneal temperature. The error bars indicate the statistical error associated with extracting the CET values for several capacitors on a single sample. It should be noted that CET values were not extracted for all samples. In general, the C-V characteristics for these samples were dominated by high gate leakage and reliable extraction of the oxide capacitance was not possible.

Figure 4.10 shows that the CET values were typically between 3.6 nm and 4.1 nm. Given that the physical thickness of the hafnium oxide was intended to be approximately 10 nm, and the dielectric constant of hafnium oxide is typically reported to be around

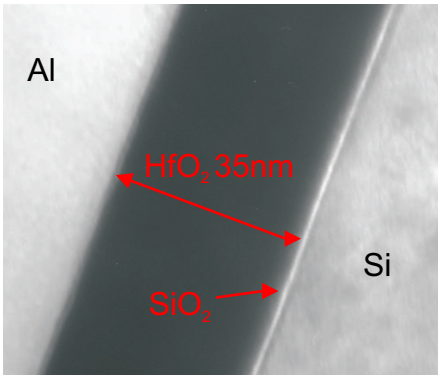


Figure 4.11: TEM image of sputtered HfO_2 gate dielectric on silicon with Al gate showing the presence of an SiO_2 interfacial layer. Reproduced with permission from Raeissi [2007].

20, the high CET values measured on the W gate capacitors suggest the formation of an interfacial layer during the sputter deposition and post deposition annealing stage. Indeed, an SiO_2 interfacial layer was observed by transmission electron microscopy (TEM) during preliminary development work of the sputtered hafnium oxide gate dielectric (Figure 4.11).

Following PMA at 425 °C for 30 min in Ar, the CET was found to increase with respect to the as-deposited samples on all wafers. This is likely to be due to the regrowth of a silicon and germanium containing thick interfacial layer at the hafnium oxide/$Si_{1-x}Ge_x$ interface [Curreem et al., 2006]. On the other hand, the CET values for the RTA samples (at both 800 °C and 950 °C) show a less clearer trend. For both Si control and SiGe with 2 nm cap, the CET values following RTA at 950 °C increase in a manner similar to those following PMA at 425 °C, whereas the CET values for the SiGe with 3 nm cap show a stark decrease following RTA at 950 °C compared to the as-deposited samples. The latter result could indicate a problem with the hafnium oxide thickness uniformity. Unfortunately, this is an inherent problem of this experiment, since the wafers were cleaved into smaller pieces to permit a greater split of anneal conditions, meaning different samples originate from different points on the full wafer.

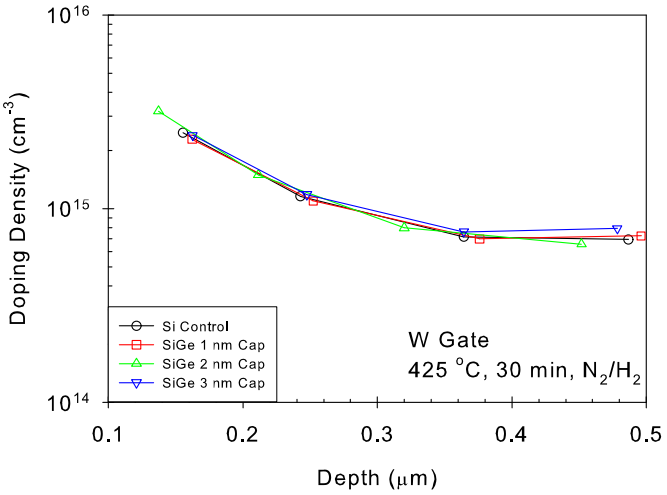


Figure 4.12: Typical doping density profiles for both silicon and SiGe capacitors following PMA anneal.

4.3.4 Doping Profile and Interface Trap Density

The doping density profile was determined from the corrected 1 MHz C-V curves. Typical doping density profiles are shown in Figure 4.12. The doping density was determined to be $\sim 10^{15}$ cm^{-3}, consistent with the initial wafer specification of 10 Ωcm [Sze, 1981]. The low doping values are to be expected since the wafers were not intentionally doped at any stage during either growth or fabrication. The only doping present in the capacitors is likely to originate from any doping present in the starting substrate and any background doping present in the CVD chamber during the growth stage, although this is only likely to be around 10^{14} cm^{-3}. The average doping density was not found to depend on either the Si cap thickness in SiGe capacitors or the anneal conditions, confirming excellent doping uniformity across all wafers.

The interface trap density distribution was calculated according to the high-low frequency method outlined in Section 3.3.5.3. A 500 Hz frequency curve was used for the low frequency C-V curve and the 1 MHz C-V curve was taken for the high frequency curve. The technique is useful because of its simplicity, yet the point should be stressed that this technique will only provide a lower estimate for the interface trap density. The

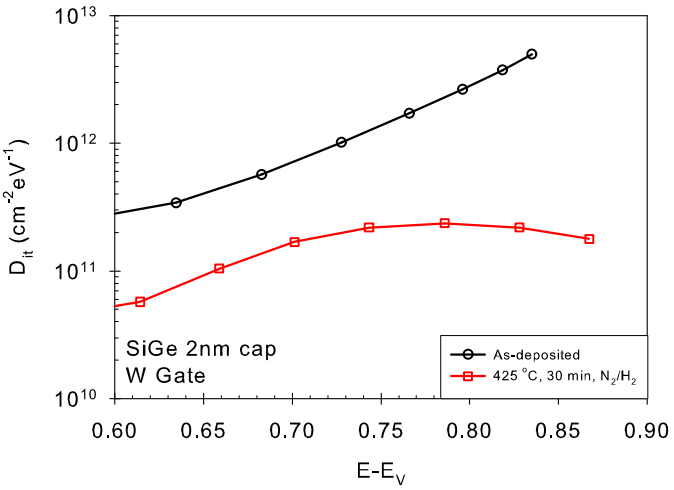


Figure 4.13: Interface trap density distribution for SiGe capacitors with 2 nm Si cap.

reason is because the technique works on the assumption that all interface traps are able to respond to the ac signal during the low frequency measurement but they are unable to follow the high frequency ac signal. However, in the case of high-κ dielectrics where the interface trap density is typically much higher than the Si/SiO_2 system, interface traps *do* respond to the high frequency ac signal, as indicated by the shoulder in the C-V curves (e.g. in Figure 4.9). Figure 4.13 shows the distribution of interface traps within the bandgap for a SiGe capacitor with 2 nm Si cap for both as-deposited and PMA anneal conditions. The effect of the PMA anneal can clearly be seen to reduce the the interface trap density by approximately 1 order of magnitude. Due to the position of the flatband voltage and using n-type substrates only, it was only possible to determine the interface density distribution for the upper half of the bandgap. The interface trap energy, E, is measured with respect to the valence band edge, E_V.

An estimate for the average interface trap density, D_{it} can also be obtained from the peak value of the conductance (see Section 3.3.5.3). Figure 4.14 shows the average interface trap density for $Si_{1-x}Ge_x$ capacitors as a function of anneal temperature. The as-deposited samples have the highest D_{it} values, typically around 2 x 10^{12} cm^{-2}. The lowest values for the interface trap density were obtained following PMA at 425 °C for

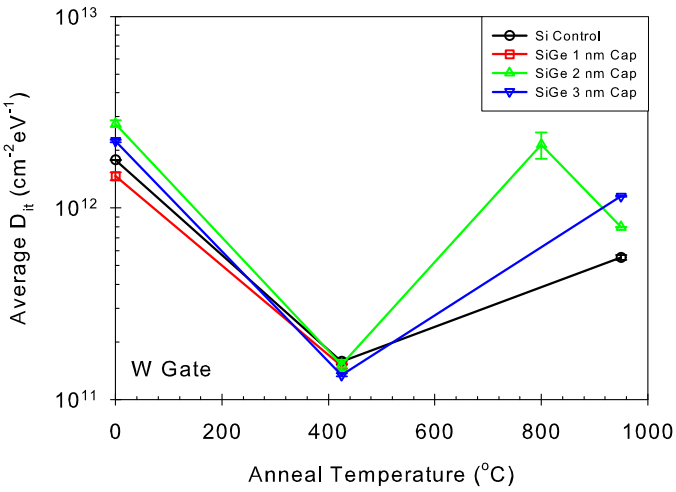


Figure 4.14: Average interface trap densities for $Si_{1-x}Ge_x$ capacitors as a function of anneal temperature.

30 min in N_2/H_2. High temperature annealing at either 800 °C or 950 °C has also reduced the interface trap density with respect to the as-deposited condition, but less effectively.

4.3.5 Flatband Voltage

The flatband voltage, V_{fb} was calculated from the method described in Section 3.3.5.2. The flatband voltage was calculated from the 1 MHz C-V curves swept from inversion to accumulation, to minimise the effects of interface traps on the flatband voltage extraction. The flatband voltage is shown in Figure 4.15 as a function of silicon cap thickness for each of the anneal conditions.

The flatband voltage for the as-deposited capacitors is around 0.4 V, independent of silicon cap thickness. A positive flatband shift, typically around 0.3 V, with respect to the as-deposited capacitors, was measured on all of the annealed samples. This flatband voltage shift corresponds to a reduction in positive charge of approximately 1 x 10^{12} cm^{-2} and is consistent with the reduction of interface states discussed in Section 4.3.4.

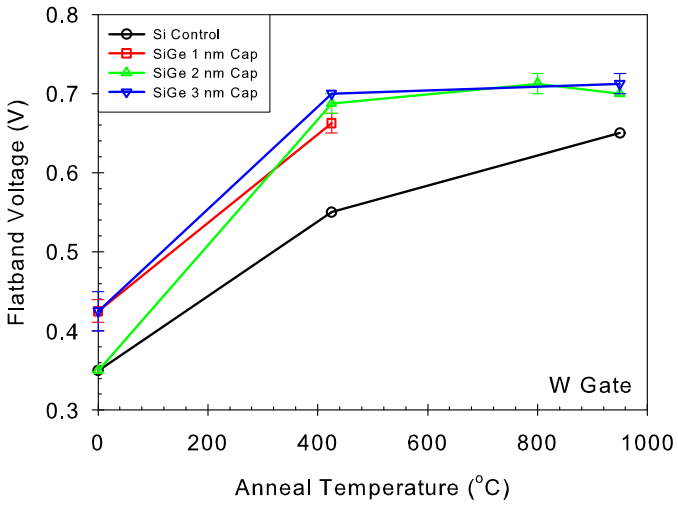


Figure 4.15: Flatband voltage for silicon and SiGe capacitors as a function of cap thickness and anneal condition.

4.3.6 Hysteresis

The hysteresis behaviour of the Si_{1-x}/HfO_2/W gate capacitors was investigated by sweeping the capacitance from inversion to accumulation and then back from accumulation to inversion at a measurement frequency of 1 MHz to minimise the effects of interface traps. Figure 4.16 shows the hysteresis for as-deposited SiGe capacitors with 3 nm Si cap, showing the shift of the C-V characteristics between the two sweep directions. The width of the hysteresis, ΔV_{fb} was typically found to be between 50 mV and 100 mV for all as-deposited samples. Figure 4.16 shows the hysteresis to be in the clockwise direction for as-deposited capacitors.

The hysteresis effect is caused by charge trapping in the oxide during the experiment, and the magnitude and direction of the hysteresis loop can be used to gain information about the nature of the charge trapping. The density of charges, N_{hys}, can be calculated according to Equation 4.1 [Kukli et al., 2006]:

$$N_{hys} = \frac{C_{ox}\Delta V_{fb}}{q} \tag{4.1}$$

The clockwise hysteresis for the as-deposited capacitors can be understood as

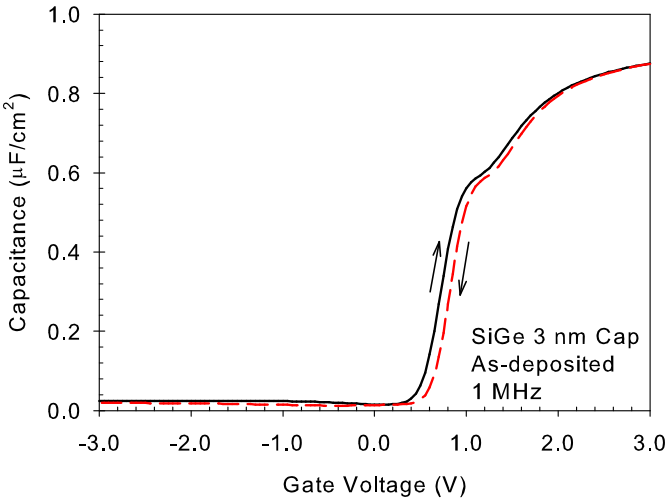


Figure 4.16: High frequency C-V characteristics showing the clockwise hysteresis behaviour for as-deposited SiGe capacitors with 3 nm Si cap.

follows. When the gate voltage is swept from inversion to accumulation, electrons from the substrate are injected into the oxide increasing the net density of negative charge in the oxide. When the gate voltage is swept back from accumulation to inversion, the trapped negative charge results in a positive flatband voltage shift, giving rise to the observable clockwise hysteresis shown in Figure 4.16. The hysteresis shown in Figure 4.16 corresponds to an oxide trapped charge density of 6.5 x 10^{11} C/cm^2.

The hysteresis can be significantly reduced following PMA at 425 °C for 30 min in N_2/H_2 to approximately 1 mV or less, still in the clockwise direction, indicating that the post metallisation anneal has reduced the number of traps into which electrons can tunnel into by 1-2 magnitudes.

Whilst the hysteresis appears to be in the clockwise direction for all as-deposited and PMA samples, the hysteresis following RTA at either 800 °C or 950 °C is in the *anticlockwise* direction (Figure 4.17), indicating a different type of charge-trapping. It is possible that this could be due to charge-trapping at the HfO_2/W interface, where electrons injected into the W gate during the forward sweep, would result in a negative flatband voltage shift for the reverse sweep, and hence, an anticlockwise hysteresis.

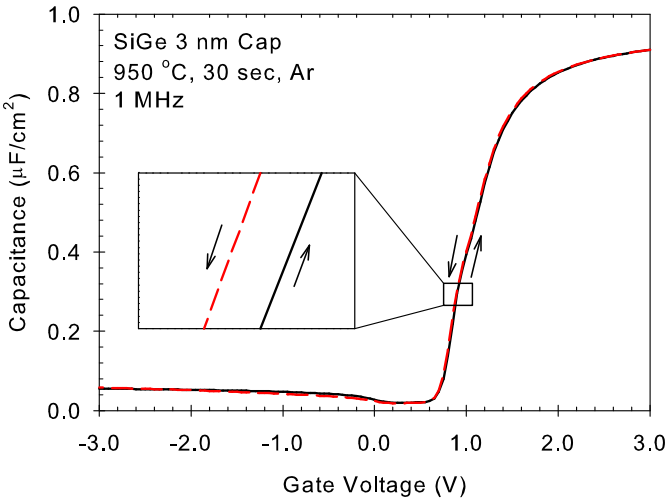


Figure 4.17: High frequency C-V characteristics showing the clockwise hysteresis behaviour for SiGe capacitors with 3 nm Si cap following RTA2.

Table 4.4: Magnitude of oxide charges responsible for hysteresis behaviour in $Si_{1-x}Ge_x/HfO_2/W$ gate capacitors.

	As-deposited (cm^{-2})	PMA (cm^{-2})	RTA1 (cm^{-2})	RTA2 (cm^{-2})
Si Control	3 x 10^{11}	5 x 10^{9}	-	4 x 10^{10}
SiGe 1 nm Cap	3 x 10^{11}	5 x 10^{9}	-	-
SiGe 2 nm Cap	9 x 10^{11}	5 x 10^{9}	1 x 10^{10}	5 x 10^{10}
SiGe 3 nm Cap	7 x 10^{11}	1 x 10^{10}	-	7 x 10^{10}

The magnitude of the hysteresis following RTA typically fell between the measured values for the as-deposited and PMA capacitors. Furthermore, the width of the hysteresis was found to be higher following RTA at 950 °C compared to RTA at 800 °C. The hysteresis of all samples is summarised in Figure 4.18 as a function of anneal condition. In Figure 4.18, open symbols indicate clockwise hysteresis behaviour and closed symbols represent anti-clockwise hysteresis. The magnitude of the corresponding charge densities, calculated from Equation 4.1 are shown in Table 4.4.

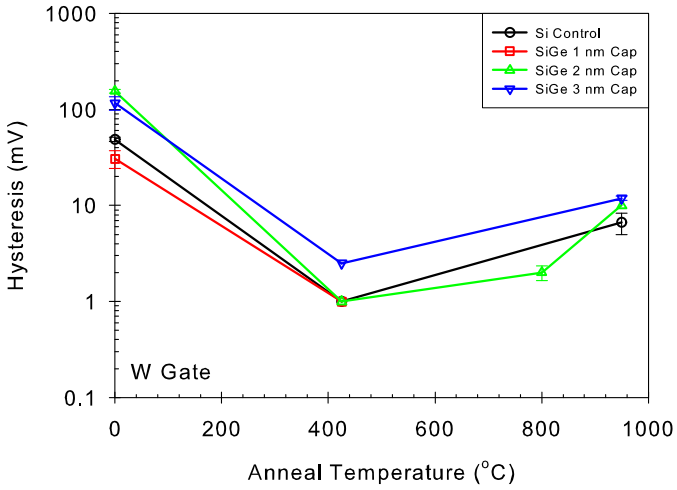


Figure 4.18: Magnitude of hysteresis for $Si_{1-x}Ge_x$ capacitors with W gate as a function of anneal temperature.

4.4 Conclusion

The electrical properties of $Si_{1-x}Ge_x$ capacitors with sputtered hafnium oxide gate dielectric and three different metal gate have been discussed in this chapter. Unfortunately, the time constraint of the SINANO project resulted in a large parameter space for this particular experiment, making it difficult to draw systematic conclusions. However, of the three metal gates studied, only W demonstrated both tolerable gate leakage and reasonable C-V characteristics. High gate leakage on $Si_{1-x}Ge_x/HfO_2$ capacitors using thick TiN metal gate was attributed to a thermal instability of TiN on HfO_2, an observation previously seen on Si/SiO_2 capacitors with TiN gate, following high temperature annealing. However, the use of a thin TiN metal gate capped with poly-silicon was found to be thermally stable on HfO_2 following high temperature annealing with low gate leakage and this warrants further investigation. The C-V characteristics of the thin TiN gate capacitors, were found to be strongly suppressed at high measurement frequencies. Although this is not fully understood, high series and contact resistance due to either insufficient doping or dopant activation was offered to explain the behaviour.

$Si_{1-x}Ge_x/HfO_2$ capacitors using W gate demonstrated both excellent gate leak-

age and reasonable C-V characteristics. The capacitors attributed CET values of around 4 nm, suggesting the presence of an SiO_2-like interfacial layer, and interface state densities were reduced to below $\sim 10^{12}$ $cm^{-2}eV^{-1}$ by annealing. Two types of charge-trapping in the HfO_2/W gate capacitors was prominent during the C-V measurements, indicated by the observation of both clockwise and anti-clockwise hysteresis in the C-V curves. As-deposited and PMA capacitors showed clockwise hysteresis, which can be explained by electron injection from the substrate during the C-V measurement. Electron injection into the gate could explain the anti-clockwise hysteresis behaviour observed on all capacitors following high temperature annealing at 800 °C and 950 °C. As a result of this experiment, the HfO_2/W gate stack was chosen to develop a batch of $Si_{1-x}Ge_x$ transistors.

Chapter 5

Electrical Characterisation Results of $Si_{1-x}Ge_x$ Channel pMOSFETs with High-κ Dielectric and Metal Gate

5.1 Introduction

This chapter describes the electrical characterisation of a batch of pseudomorphic $Si_{1-x}Ge_x$ channel pMOSFETs with hafnium oxide/W gate stack developed in Chapter 4. The majority of the transistor processing steps were carried out at KTH, Sweden. As for the capacitor development work in Chapter 4, this work was carried out within the framework of the SINANO Network of Excellence under the direction of the author.

Table 5.1: Growth specification for $Si_{1-x}Ge_x/HfO_2$/W gate pMOSFETs

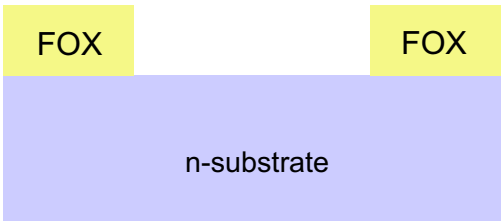
Wafer Description	Ge Composition (%)	SiGe Thickness (nm)	Si Cap Thickness (nm)
Si Control	0	-	-
SiGe 1 nm Cap	25	7	1
SiGe 2 nm Cap	25	7	2
SiGe 3 nm Cap	25	7	3
SiGe 4 nm Cap	25	7	4
SiGe 5 nm Cap	25	7	5

5.2 Device Specification and Processing

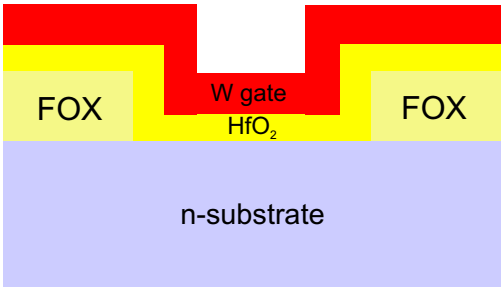
The batch of pseudomorphic $Si_{1-x}Ge_x/HfO_2$/W gate MOSFETs consisted of six wafers grown by LPCVD at the University of Warwick. The growth specification is summarised in Table 5.1. Each of the SiGe wafers featured a 7 nm SiGe channel with a Ge composition of 25% and a thin silicon cap, varying in thickness from 1 nm to 5 nm. An epitaxial silicon control wafer was also grown.

A schematic of the transistor process sequence is shown in Figure 5.1. Following active area formation and well implants, hafnium oxide and tungsten were deposited to form the gate stack as outlined in Section 4.2, with each wafer receiving both a post-oxide deposition anneal (PDA) at 800 °C anneal for 10 min in N_2 following hafnium oxide deposition and a post-metallisation anneal (PMA) at 425 °C for 30 min in N_2/H_2 after W gate deposition. Following metal gate etch, deep implantations for the source and drain regions were made and source/drain activation anneal was performed at 950 °C for 30 seconds in Ar before contact metallisation. Unfortunately, it was not possible to process a silicon control with SiO_2 gate dielectric as part of this batch.

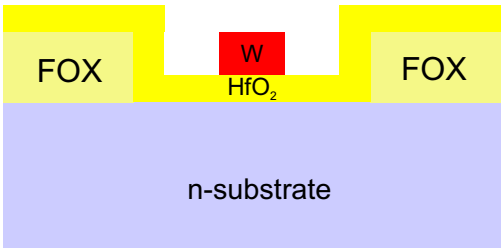
Several technical difficulties were encountered during the planning and fabrication of the device batch. Firstly, the difficulties arising from the W metal etch meant that it was not possible to process devices with gate lengths shorter than around 1 μm. Secondly, the W gate etch stopped on the HfO_2. This meant that HfO_2 remained in the



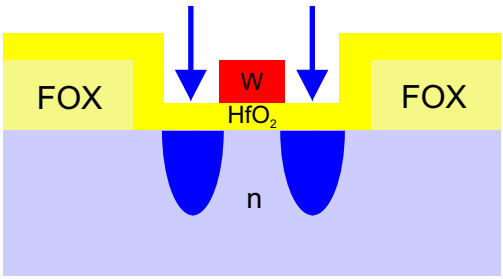
(a) Active area formation



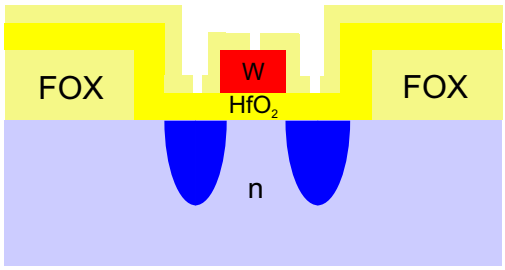
(b) Oxide and gate deposition



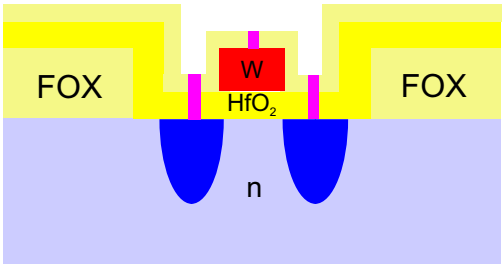
(c) Gate etch



(d) Source/drain implantation



(e) Contact hole formation



(f) Contact metallisation

Figure 5.1: Schematic of the process flow for $Si_{1-x}Ge_x/HfO_2/W$ transistors.

source and drain regions after the gate etch (see Figure 5.1). This had to be removed without invoking a lateral etch of the W gate, which would be exposed due to the absence of sidewall spacers. Consequently, the HfO_2 in the source and drain regions had to be removed by sputter etch, despite it being unclear as to whether this would result in any further damage to the transistors that might degrade their performance. Thirdly, with the focus being on long-channel transistors and given the relatively short timescale available for processing, the source and drain regions were processed without silicidation. It was not known how detrimental to the series resistance this would be.

5.3 Electrical Characterisation of $Si_{1-x}Ge_x$ Channel pMOSFETs

5.3.1 Uniformity

It quickly became apparent that the electrical characteristics of the $Si_{1-x}Ge_x$ transistors would be severely hampered by gross non-uniformities in the transistor characteristics across each wafer. Figure 5.2 shows the transfer characteristics for several 50 μm x 50 μm SiGe transistors with 1 nm Si cap located at random positions across the wafer. This behaviour was observed on all wafers and for different size transistors.

The source of the variation is likely to be due to a high and varying hafnium oxide thickness across each of the wafer. It is postulated that the thicker hafnium oxide dielectric could have resulted in some hafnium oxide (maybe 1 - 2 nm) remaining after the sputter-etch formation of the source/drain contact holes, resulting in a high series resistance. Nevertheless, systematic trends could still be obtained between wafers. In addition, comparisons were also made between the *best-performing* transistors on each wafer, and reference is made when this has been done.

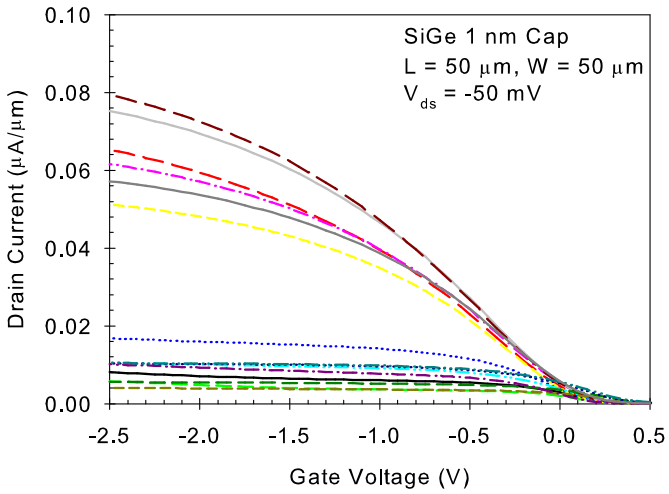
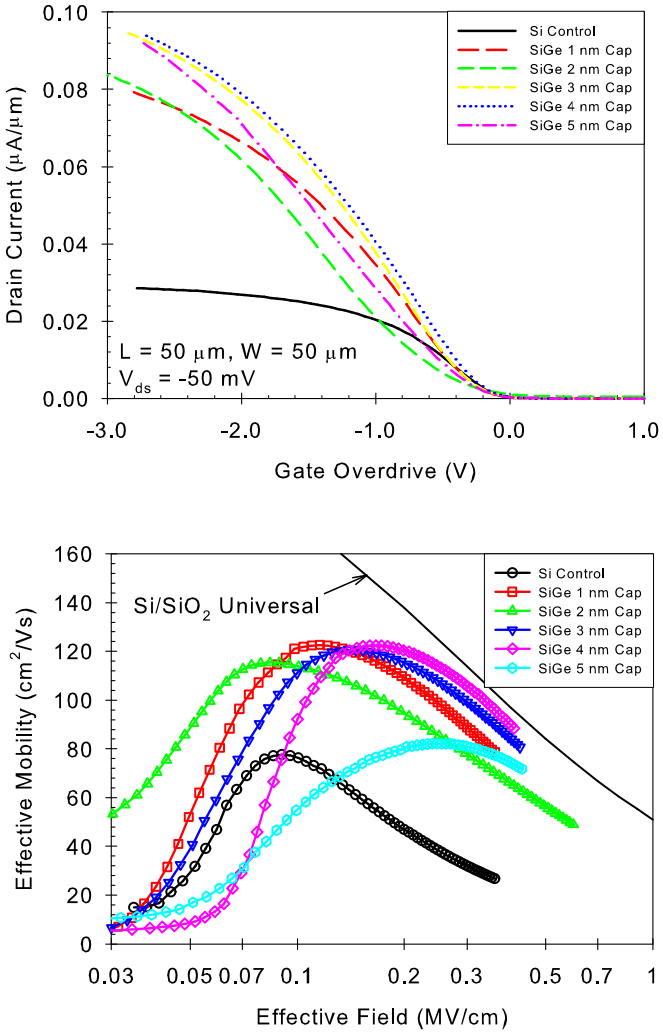


Figure 5.2: Linear transfer characteristics for 50 μm x 50 μm SiGe transistors with 1 nm Si cap highlighting the significant spread in device characteristics across a wafer.

5.3.2 Effective Mobility and Split C-V

The effective hole mobility was extracted for 50 μm x 50 μm transistors using the split C-V technique outlined in Section 3.3.7. The drain current in the linear regime for the *best-performing* devices is shown in Figure 5.3(a). The best-performing SiGe channel devices show a clear performance enhancement over the best Si control devices, showing, on average, three times higher drain current at a gate overdrive of -2.5 V. The corresponding effective hole mobilities are shown as a function of effective field in Figure 5.3(b). The inversion charge was determined from split C-V measurements taken at 100 kHz to minimise the contribution of interface traps. The universal mobility curve for holes in silicon is shown for comparison [Takagi et al., 1994]. The effective mobility curves in Figure 5.3(b) remain uncorrected for series resistance.

The first thing to note is that the effective mobility extracted from the best transistors across all of the wafers are disappointing, in that they are lower than both the silicon universal mobility for holes and also the corresponding mobility values previously reported in the literature. For example, Wu et al. [2005] show hole mobility enhancements over the silicon universal curve for strained SiGe pMOSFETs with a composite



(b)

Figure 5.3: (a) Linear drain current and (b) effective hole mobility as a function of effective field for the best-performing 50 μm x 50 μm $Si_{1-x}Ge_x/HfO_2$/W gate pMOSFETs.

$Al_2O_3/HfO_2/Al_2O_3$ gate dielectric, whereas Weber et al. [2006] and Jin et al. [2004] demonstrate SiGe pMOSFETs with HfO_2/TiN gate stack having a 60% improvement over the silicon, although in all these cases, the hafnium oxide layer was deposited by atomic layer deposition (ALD) and not by sputtering.

However, despite the disappointing values for the mobility, the best-performing SiGe devices exhibit a significant hole mobility enhancement over the best-performing Si control device, showing nearly two times enhancement in the peak mobility and around three times enhancement at a vertical field of 0.3 MV/cm. This hole mobility enhancement can be attributed to the effects of the strained SiGe band structure.

The reason for the observed mobility degradation is likely to be due to a combination of several factors. At low vertical fields, the carrier mobility is likely to be degraded due to high Coulomb scattering owing to a high density of interface traps created during the hafnium oxide sputter deposition. It is also possible that a sputter-induced surface roughening has occurred during the hafnium oxide deposition, which has resulted in high surface roughness scattering at higher vertical fields. Furthermore, it is speculated that the presence of an SiO_2-like interfacial layer between the $Si_{1-x}Ge_x$ and HfO_2 could result in an additional 'remote' surface scattering. Remote optical phonon scattering from the hafnium oxide might also play a role.

High series and contact resistance, attributed to the non-uniformity in device characteristics, will also impact the extracted mobility values in a negative way. It was hoped that some of the mobility degradation could be recovered by correcting for the effects of series resistance. However, a reliable extraction of the series resistance was not possible due to the afore-mentioned uniformity issue.

Figure 5.3(b) hints at a mobility dependence on the silicon cap thickness for the SiGe devices, with higher hole mobilities observed for thicker silicon caps. The presence of the silicon cap is likely to improve the mobility by removing the inversion charge away from the hafnium oxide/silicon interface and reducing the effects of Coulomb scattering

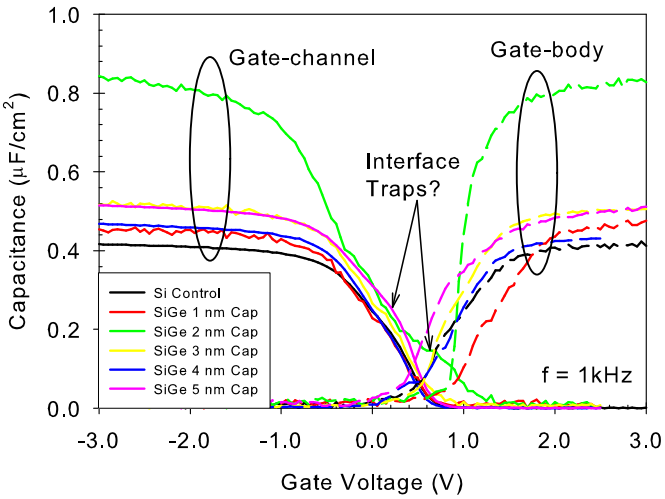


Figure 5.4: Typical split C-V characteristics for $Si_{1-x}Ge_x/HfO_2/W$ gate pMOSFETs taken at 1 kHz.

by interface and oxide charges and the effects of surface roughness scattering. This is indeed true for the SiGe devices with 1 nm, 3 nm and 4 nm Si caps. However, in respect of this,the mobility behaviour of the SiGe devices with 2 nm and 5 nm caps are slightly anomalous.

Figure 5.4 shows the split C-V characteristics that were used to calculate the effective mobility for the best-performing transistors measured at a frequency of 1 kHz. The split C-V characteristics for the SiGe wafer with 2 nm Si cap are very strange. Firstly, the capacitive equivalent thickness (CET) appears to be about half that of the other wafers. Secondly, there is an observable 'kink' in the gate-channel branch of the split C-V curve in weak inversion, indicating a possible 'early turn-on' in these devices and could be due to the presence of interface traps. Further investigation with regards to these effects are required and whether it can be related to the degraded mobility shown in Figure 5.3(b) remains to be understood. The effective mobility for the SiGe device with 5 nm cap was strongly suppressed at low vertical fields, indicating that the mobility is limited by Coulomb scattering in this regime. The corresponding split C-V curve shows an observable distortion, or shoulder, in the gate-channel branch in weak

Table 5.2: Capacitive equivalent thickness (CET) values for $Si_{1-x}Ge_x/HfO_2/W$ gate pMOSFETs

Wafer Description	CET (nm)
Si Control	8.3
SiGe 1 nm Cap	7.3
SiGe 2 nm Cap	4.2
SiGe 3 nm Cap	6.8
SiGe 4 nm Cap	8.0
SiGe 5 nm Cap	6.7

inversion. This additional capacitance is likely to be due to the presence of a high density of interface traps, which would be consistent with the observed mobility degradation at low vertical fields.

The capacitive equivalent thickness (CET) values were taken from the accumulation capacitance of the 1 kHz gate-body branch of the split C-V curves and are summarised in Table 5.2. The CET values are significantly higher than those reported in Section 4 for the capacitor development work, despite the two experiments supposedly featuring identical gate stacks. The high CET values for the transistors are likely to be due to a higher physical thickness of hafnium oxide than deposited on the capacitors. This would be consistent with the problems associated with gross non-uniformities in device characteristics across each wafer, since this additional hafnium oxide could impede the sputter etch in the source/drain regions.

The split C-V characteristics showed significantly less variation than the I-V characteristics, reported in Section 5.3.1. This is shown in Figure 5.5 which shows good uniformity in the split C-V characteristics of 50 μm x 50 μm SiGe devices with 1 nm Si cap from random locations across the wafer, despite the significant variations in the drain current (see Figure 5.2). Whilst this is not entirely understood, it allowed an average oxide capacitance to be determined, which could be used to determine the field effect mobility from the transconductance, and perform a statistical analysis on devices

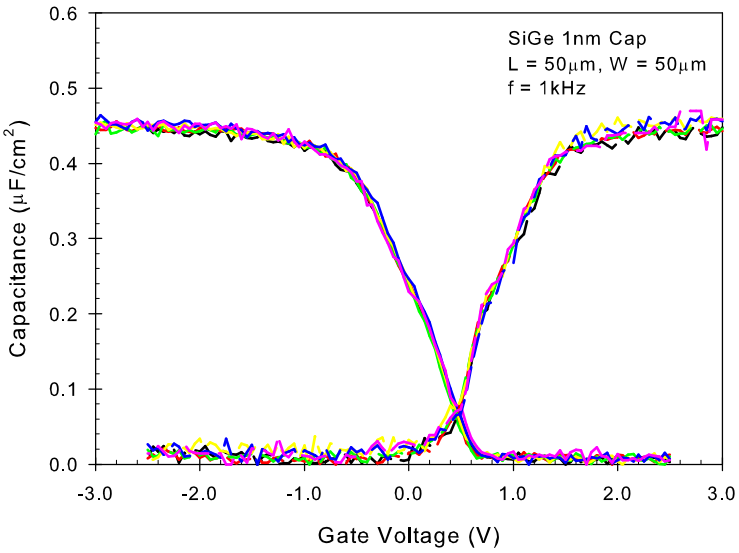
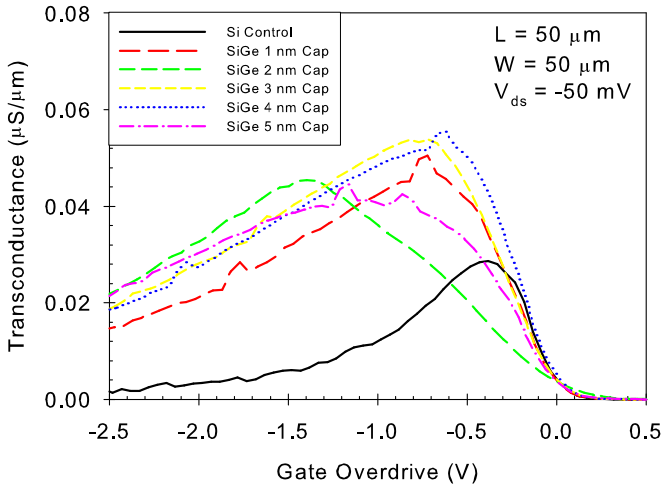


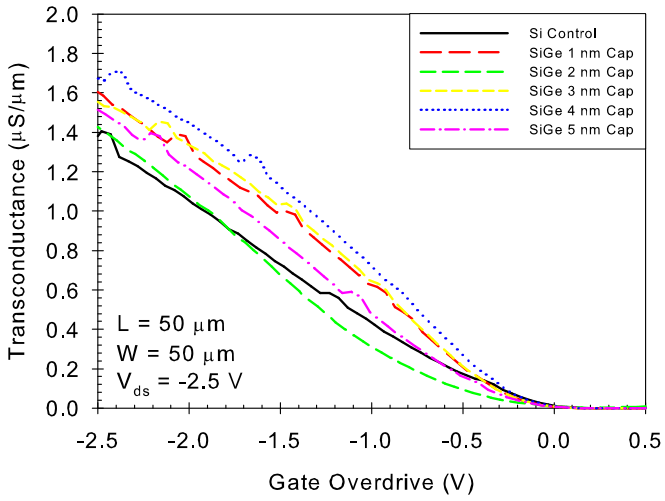
Figure 5.5: Split C-V characteristics measured at 1 kHz for several 50 μm x 50 μm SiGe pMOSFETs with 1 nm Si cap.

for which split C-V measurements were unreliable. It was hoped that this data could be used to ascertain whether or not the mobility enhancements observed on the best devices are, indeed, indicative of the entire batch. The results of this statistical analysis are presented in Section 5.3.3, when the transconductance of the $Si_{1-x}Ge_x$ devices is discussed.

5.3.3 Transconductance, Threshold Voltage and Subthreshold Slope

Figure 5.6 shows the linear and saturation transconductance for the best-performing 50 μm x 50 μm $Si_{1-x}Ge_x$ transistors. As expected from Section 5.3.2, the best-performing SiGe channel devices exhibit higher transconductance over silicon in both the linear and saturation regimes. However, due to the large variation in the device characteristics across a wafer, it was necessary to compare the transconductance characteristics for a large number of $Si_{1-x}Ge_x$ channel devices, located at random positions across the wafer. Typically, between 10 and 20 transistors (of a given dimension) were compared. In order to make a fair comparison between the $Si_{1-x}Ge_x$ wafers, the transconductance





(b)

Figure 5.6: (a) Linear and (b) saturation transconductance for $Si_{1-x}Ge_x/HfO_2/W$ gate transistors.

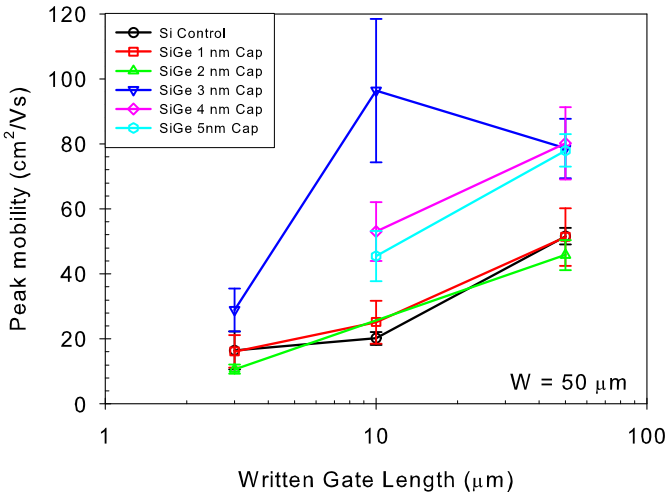


Figure 5.7: Average field effect mobility for $Si_{1-x}Ge_x/HfO_2/W$ gate transistors as a function of written gate length.

was scaled by the oxide capacitance and the transistor dimensions to determine the field effect mobility.

Figure 5.7 compares the *average* field effect mobility for Si control devices with a width of 50 μm. The error bars represent the error associated with measuring a large number of transistors and gives a measure of the uniformity. In general, the SiGe channel devices *do* show a mobility enhancement over the silicon control. This enhancement is particularly evident for the longer gate length transistors with thicker silicon caps (i.e. greater than 3 nm). The SiGe transistors with silicon caps of 1 nm and 2 nm show a much less obvious improvement over silicon. In fact, the average mobility for the SiGe devices with a 2 nm cap falls below that of silicon for all gate lengths. However, since this average also includes 'non-working' transistors, this could imply that the postulated problems with the sputter etch has reduced the yield on this wafer, leading to a lower average mobility. This highlights just how difficult it is to draw firm conclusions for the performance of the $Si_{1-x}Ge_x/HfO_2/W$ gate devices featured in this batch, when the characteristics are dominated by non-uniformities associated with both the gate dielectric.

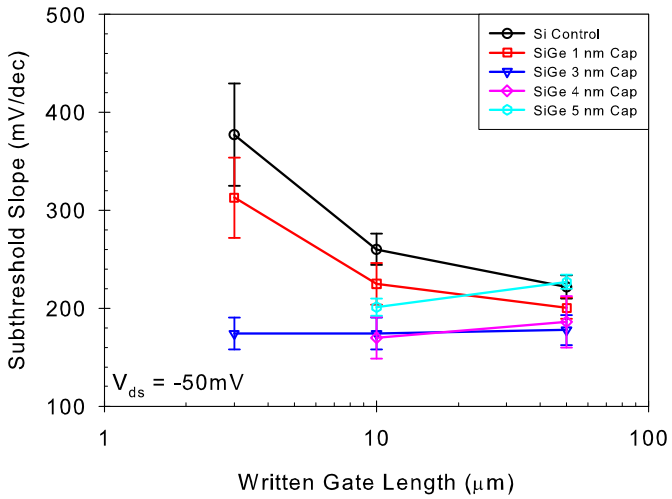


Figure 5.8: Average values for the subthreshold slope measured in the linear regime for $Si_{1-x}Ge_x/HfO_2/W$ gate transistors as a function of written gate length.

The peak mobility decreases with decreasing channel length. This implies that the results are affected by the effects of series resistance, even at these relatively long gate lengths. It is also speculated that be damage at the source/drain regions from the hafnium oxide sputter etch in these regions, which could become more important for shorter gate lengths.

Figure 5.8 shows the average subthreshold slope values for a drain bias of V_{ds} = -50 mV as a function of written gate length for all wafers. Typical values for the subthreshold slope are around 200 mV/dec, indicating a high density of interface traps. In general, the subthreshold slope for the SiGe wafers is lower than that of the silicon control, except for the wafer with 2 nm cap, where subthreshold slopes in excess of 500 mV/dec were observed (not shown). The lower subthreshold slope values for the SiGe transistors suggests a lower interface trap density compared to silicon control. Why this should be true is unclear. Typically, the presence of Ge at the interface degrades the subthreshold slope of SiGe devices compared to silicon control devices. However, it is difficult to conclude whether or not this is a 'true' effect or is related to the problems associated with a uniform deposition of the hafnium oxide dielectric on different wafers.

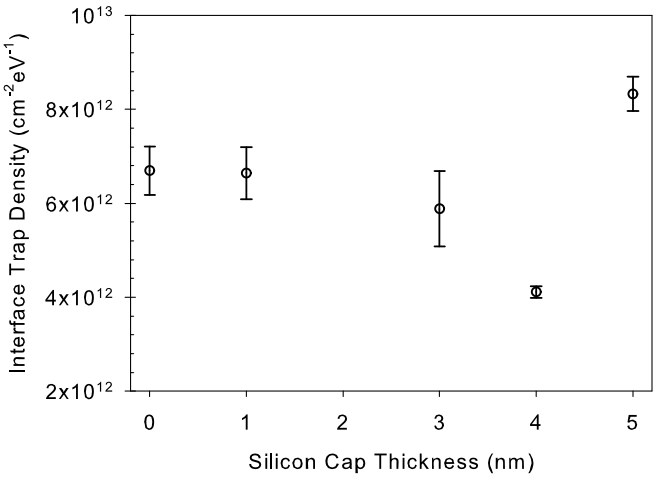


Figure 5.9: Interface trap density measured from the subthreshold slope on $Si_{1-x}Ge_x/HfO_2/W$ pMOSFETs as a function of Si cap thickness, where a cap thickness of '0' has been used to represent the Si control.

The interface trap density, $D_{it} = C_{it}/q$, was calculated from the subthreshold slope, S, according to [Sze, 1981]:

$$S = 2.3\frac{k_B T}{q}\left(1 + \frac{C_{it} + C_{dep}}{C_{ox}}\right) \tag{5.1}$$

where C_{dep} and C_{ox} are the depletion and oxide capacitances, respectively. Figure 5.9 shows the measured interface trap density as a function of silicon cap thickness, where the silicon control is represented by a cap thickness of '0'. Typical values of the interface trap density ranged from ~4 x 10^{12} $cm^{-2}eV^{-1}$ to 8 x 10^{12} $cm^{-2}eV^{-1}$ and much higher than the values measured on the capacitors in Chapter 4. Part of this could be due to the differences in processing between the capacitor and transistor batches. However, it should also be mentioned that the discrepancies could be related to the methodology, where the interface trap densities reported in Chapter 4 were calculated following the correction of the C-V data for series resistance effects, whereas those for the transistor batch were not.

The threshold voltages for 50 μm x 50 μm devices is plotted against the peak transconductance in Figure 5.10. Typical values for the threshold voltage are between

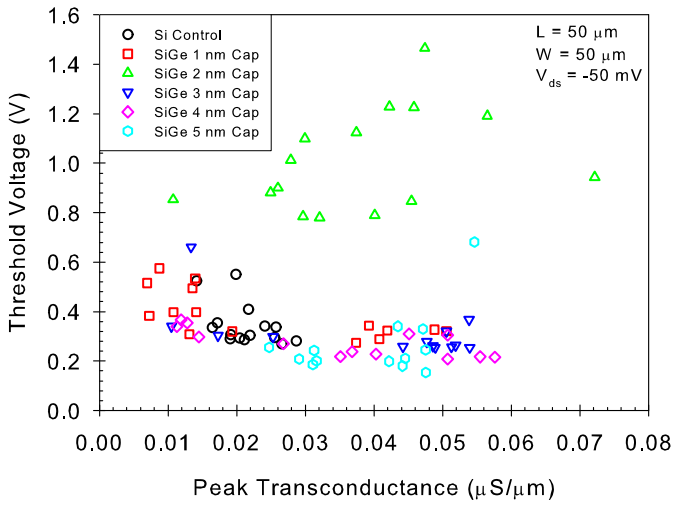
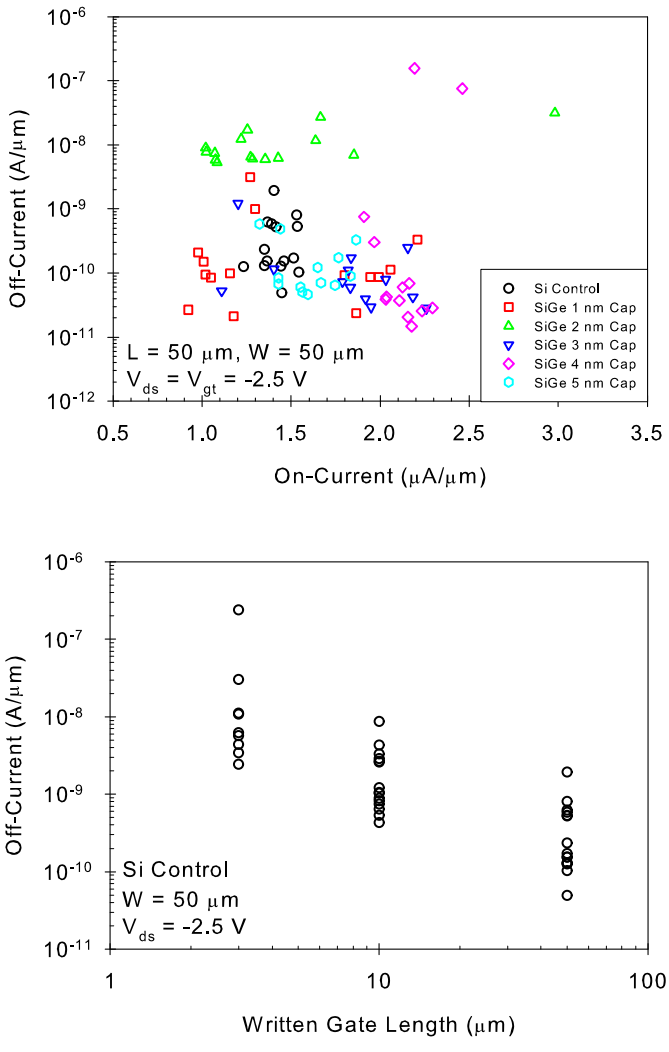


Figure 5.10: Scatter plot of threshold voltage vs peak linear transconductance for 50 μm x 50 μm transistors.

+0.2 V and +0.4 V. Whilst this is off-target for pMOSFETs, optimisation of the threshold voltage was beyond the scope of the work. The SiGe devices with 2 nm Si cap again show anomalous behaviour, with threshold voltages between +0.8 V and +1.5 V. This is consistent with the other anomalous behaviour for devices on this wafer, such as high subthreshold slope and low CET, and can probably be related to an anomaly in the hafnium oxide deposition for this wafer. From Figure 5.10, there appears to be little correlation between the threshold voltage and the transconductance. It has been postulated that a correlation between these two variable might be related to the nature of interface traps. However, the data for this device batch reinforces the significant variations in device characteristics across all of the wafers.

5.3.4 Saturation Drain Current and Off-Current

Figure 5.11(a) shows the saturation drain current (or the on-current, I_{on}) versus the off-current, I_{off} for 50 μm x 50 μm $Si_{1-x}Ge_x$ devices. In general, the SiGe devices with Si caps of 3 nm, 4 nm and 5 nm show higher saturation drain currents than the silicon control wafer. The SiGe devices with a 1 nm Si cap, also show an improvement



(b)

Figure 5.11: (a) Saturation drain current versus off-current for $Si_{1-x}Ge_x/HfO_2/W$ gate transistors and (b) off-current versus gate length for silicon control devices.

in saturation drain current over the silicon control, although, there appears to be a number of devices exhibiting lower current. The drain current in these devices is likely to be severely degraded due to the series resistance problem. All wafers show similar trends in the off-current except the SiGe devices with a 2 nm cap, which typically shows around one magnitude higher off-current. The off-current on all wafers was found to be a strong function of gate length, with shorter gate lengths showing significantly higher subthreshold leakage (see Figure 5.11(b)). The long gate lengths in this work mean that short channel effects are not likely to be responsible. It is possible that defects at the hafnium oxide/$Si_{1-x}Ge_x$ interface, generated either in the early stages of the hafnium oxide deposition or during the sputter etch for contact hole formation could be responsible.

5.4 Conclusion

Following the development of a high-κ/metal gate stack on $Si_{1-x}Ge_x$ (see Chapter 4), a batch of pseudomorphic $Si_{1-x}Ge_x/HfO_2$/W gate pMOSFETs have been fabricated and characterised. Whilst there were some indications of improved hole mobility in the SiGe channel MOSFETs, which could be attributed to the effects of the strained SiGe band structure, the hole mobility values were largely disappointing, with the mobility values for the best-performing transistors on all wafers being lower than the silicon universal mobility curve for holes. The observed mobility degradation is likely to be due to a combination of high Coulomb and surface roughness scattering directly related to the hafnium oxide sputter deposition process. In addition, for this particular transistor batch, the electrical characterisation was severely compromised by gross non-uniformities in the transistor characteristics across all wafers, although this might be expected in a development transistor batch using a non-standard fabrication process. This meant that a reliable extraction of some parameters, most notably series resistance, meaning that

the effective mobility values were not corrected for this important effect. Furthermore, the complications in uniformity made it difficult to draw firm conclusions on the performance of these transistors. The effective mobility values were also found to be worse than those previously reported in the literature [Jin et al., 2004, Weber et al., 2006, Wu et al., 2005], although it should be noted that in all these cases the hafnium oxide was deposited by atomic layer deposition (ALD) and not sputtering. It remains to be seen as to whether or not the sputter-deposited dielectric can be sufficiently optimised to produce a dielectric comparable in quality to those deposited by ALD, which would subsequently permit a detailed study of hole transport in these transistor structures. In particular, physical characterisation techniques, such as high-resolution transmission electron microscopy (HRTEM) and x-ray photoelectron spectroscopy (XPS), not used in the current work, could prove crucial in gaining a fundamental understanding of the properties of the $Si_{1-x}Ge_x/HfO_2$ interface.

Chapter 6

Electrical Characterisation Results and Discussion on Germanium Channel pMOSFETs

6.1 Introduction

The absence of a good quality native oxide has, to date, led to germanium only playing a bit-part role in the microelectronics industry since it was used to demonstrate the first transistor [Bardeen and Brattain, 1948]. However, it has become apparent that SiO_2, arguably the principal reason for silicon's dominance in this industry, can no longer support gate leakage currents that result from scaling the gate dielectric thickness to targets below 1 nm. This has led to the pursuit of alternative gate dielectric materials with a dielectric constant higher than SiO_2, a so-called 'high-κ' dielectric. However, there have been numerous reports citing mobility degradations in silicon transistors incorporating these dielectrics compared to those with a traditional SiO_2 dielectric due at least to increased Coulomb and remote optical phonon scattering. Consequently, higher mobility channel materials such as strained Si, strained SiGe, Ge and strained Ge

are attractive to recover, or more than compensate for the 'lost' mobility and continue to provide improved transistor performance at future technological nodes.

In recent years, there have been several reports on long-channel Ge pMOSFETs with high hole mobility values exceeding the silicon universal hole mobility for a variety of different gate stacks and passivation techniques [Chui et al., 2002, Ritenour et al., 2003, Shang et al., 2002]. However, only very recently have high drive currents been realised in deep submicron Ge pMOSFETs [Nicholas et al., 2007]. Furthermore, the recently reported deep submicron Ge pMOSFETs were fabricated in a Si-like process line, demonstrating that deep submicron Ge pMOSFETs are compatible with current Si processing technology.

Measurements of the electrical characteristics of deep sub-micron germanium pMOSFETs, as a possible alternative-channel based devices to silicon, are reported in this chapter, as part of the Ge MOS technology development programme at IMEC. The devices have been characterised over a range of temperatures between 4 K and room temperature and their electrical properties evaluated. Following an initial discussion of the basic device characteristics including threshold voltage, subthreshold swing and transconductance, the author will go on to discuss the effective mobility in these devices, and in particular to investigate the dominant scattering factors that limit the hole mobility in germanium channel pMOSFETs. The final section focuses on the off-state leakage currents in germanium devices. With a bandgap of 0.66 eV compared to 1.12 eV in silicon, it is, at present, unknown as to whether or not band-to-band tunnelling will limit the lowest possible achievable off-state current.

6.2 Device Specification and Processing Details

Germanium pMOSFETs were fabricated with written gate lengths down to 125 nm in a Si pilot line at IMEC, Belgium. A schematic of the processed device is shown in

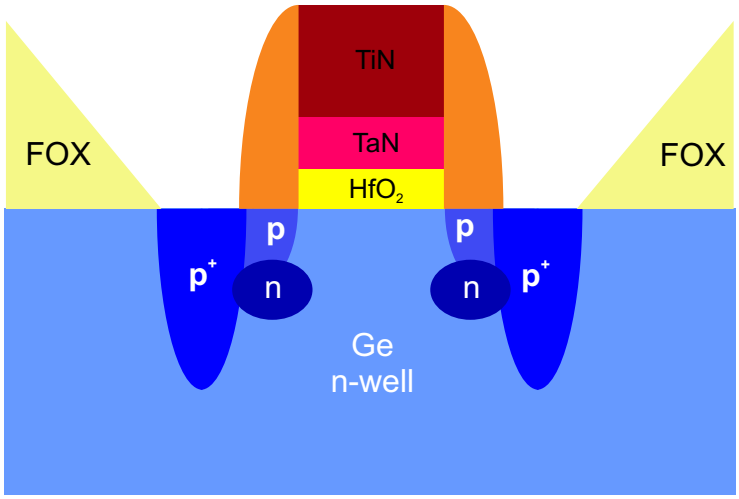


Figure 6.1: Schematic of a Ge/HfO_2/TaN/TiN pMOSFET.

Figure 6.1. The germanium devices were processed on commercially available 200 mm (100) germanium-on-silicon wafers from ASM. The pure germanium layer is $\sim$2 μm thick, and is fully relaxed with a threading dislocation density of approximately 10^8 cm^{-2}.

The well implants are formed first by implantation of ^{31}P through 30 nm deposited SiO_2. The devices in this investigation featured a deep well implant at 570 keV, a shallow well implant at 180 keV and a threshold voltage adjustment (VTA) implant with a dose of 4 x 10^{12} cm^{-2} at 90 keV. Well activation is performed at 600 °C for 5 min in N_2.

Prior to gate dielectric deposition, the germanium surface is passivated with an ultrathin ($\sim$ 4 monolayers) epitaxial Si layer. Previous work has shown that this passivation technique yields a smooth germanium/gate dielectric interface, with low interface state densities that result in high mobility germanium devices. The epitaxial Si layer is partially oxidised that results in thin interfacial layers, consisting of $\sim$0.6 nm Si and $\sim$0.4 nm SiO_2 [De Jaeger et al., 2005].

The devices used 4 nm hafnium oxide (HfO_2) as gate dielectric, deposited by atomic-layer deposition (ALD). The HfO_2 layer is formed at 300 °C through sequential pulses of $HfCl_4$ and H_2O reactants that result in self-saturating reactions at the germa-

nium surface. N_2 gas is used to purge the chamber after one pulse of each reactant. The HfO_2 layer was deposited immediately after the epitaxial silicon passivation, without exposure to air, thus avoiding any uncontrolled oxidation of the interface. Following gate dielectric deposition, a metal gate of 10 nm TaN capped with 70 nm TiN was deposited by physical vapour deposition.

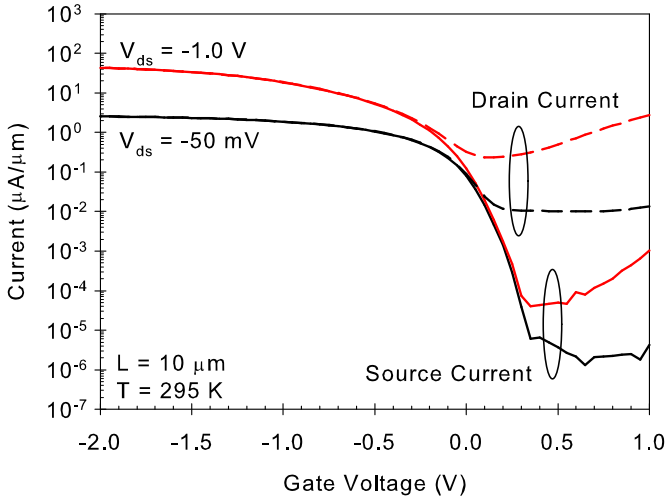
Halo implants were then made before extension implant and nitride sidewall spacer formation. Highly doped drain implants were performed and subsequent activation anneal at 500 °C for 5 min in N_2, before the NiGe source/drain regions were formed [De Jaeger et al., 2007]. Following contact metal formation, a final anneal at 350 °C for 20 min in H_2 was performed, since this has been found to reduce the density of interface states [Zimmerman et al., 2006].

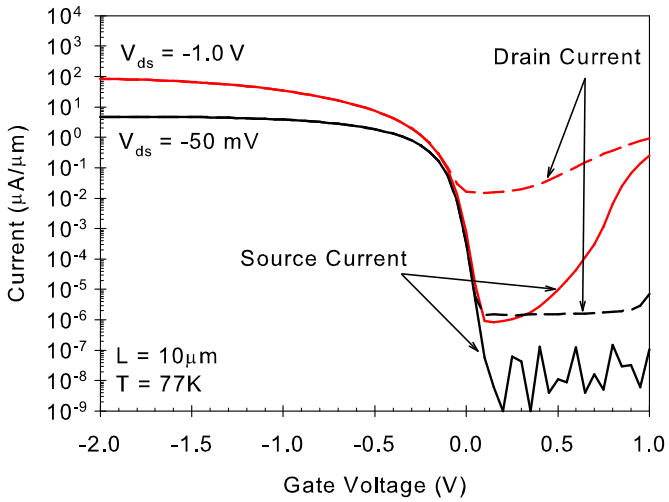
6.3 Device Characteristics

6.3.1 I-V Characteristics

Figure 6.2(a) shows an example of the typical room temperature transfer characteristics of a long-channel device (L = 10 μm) measured at V_{ds} = -50 mV and -1.0 V, as a function of gate voltage. Whilst the on-state characteristics are the same as measured at the source and the drain, there is a significant difference in the off-state leakage behaviour. The off-state current at the drain is typically observed to be several orders of magnitude higher than that measured at the source, for both V_{ds} = -50 mV and -1.0 V. This is a big drawback and is discussed in more detail in Section 6.3.7.

The transfer characteristics for the same L = 10 μm device measured at 77 K are shown in Figure 6.2(b). As for the room temperature characteristics, the on-state currents measured at the source and drain are equivalent, but the off-state currents differ by several magnitudes. Compared to room temperature, the device measured at 77 K, exhibits an increase in on-current and a significant decrease in both the off-state





(b)

Figure 6.2: Transfer characteristics for L = 10 μm Ge pMOSFET measured at (a) room temperature and (b) 77 K.

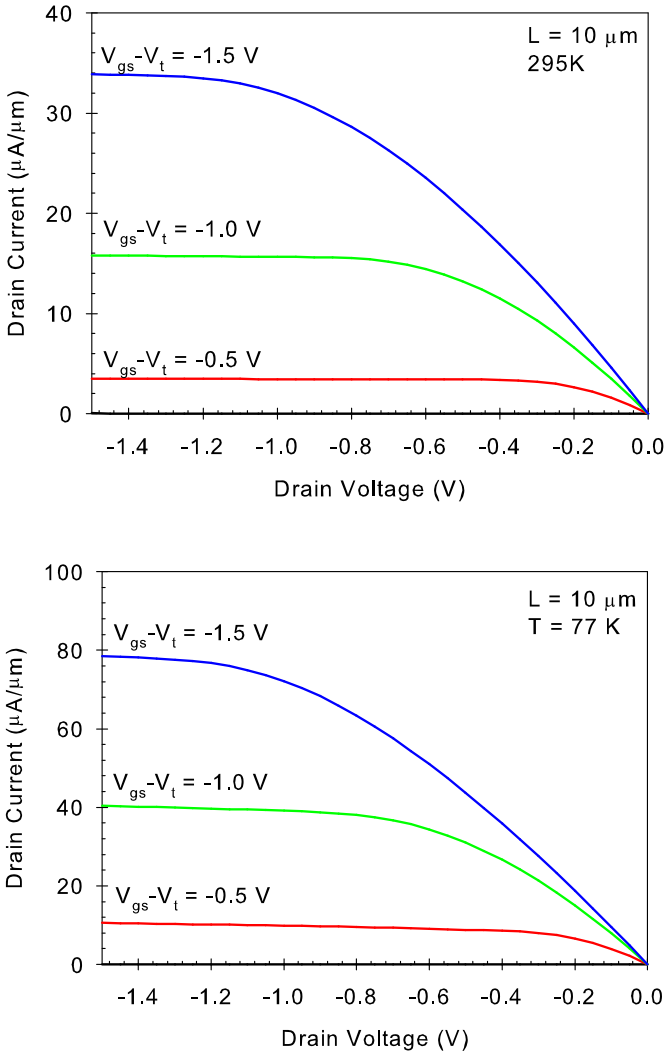
current and the subthreshold slope.

Figure 6.3 shows the output characteristics for the L = 10 μm device as a function of gate overdrive measured at both (a) room temperature and (b) 77 K. Figure 6.3 shows that the drive current at 77 K is nearly three times higher than than at room temperature, which, as will be shown later, is mainly due to the higher hole mobility that results from the reduction of phonon scattering.

Figure 6.4 shows the transfer characteristics of a short channel device, with a written gate length of 125 nm, measured at (a) room temperature and (b) 77 K, as a function of gate voltage. As for the long-channel device, there is a significant difference in the off-state currents, as measured at the source and drain respectively, with the device measured at 77 K showing higher on-current, lower off-current and steeper subthreshold slope.

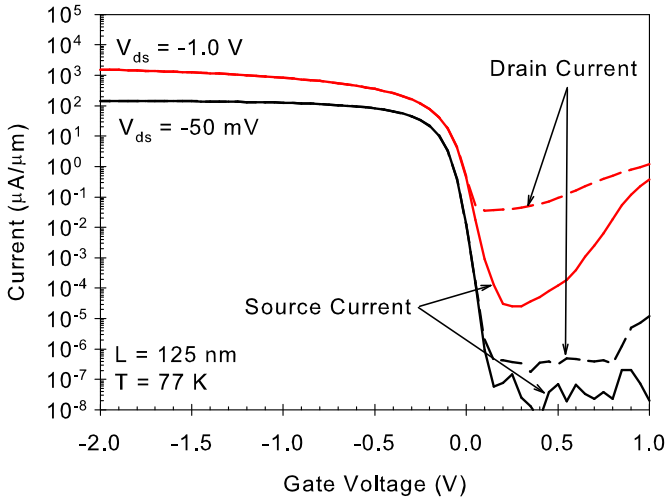
The output characteristics for the 125 nm device are shown in Figure 6.5 as a function of gate overdrive, at both (a) room temperature and (b) 77 K. This particular device shows very large drive currents of 980 μA/μm and 1366 μA/μm at V_{gs} - V_t = V_{ds} = -1.5 V, at room temperature and 77 K, respectively.

In the absence of a silicon control for this investigation, the performance of the Ge pMOSFETs was assessed by comparing to the drive currents specified by the International Technology Roadmap for Semiconductors (ITRS) for silicon MOSFETs [ITRS, 2000 Update]. A written gate length of 125 nm is close to that specified for the 0.18 μnm technology node, for which the corresponding power supply voltage, V_{dd} is given as 1.5 V. The drive current for a silicon pMOSFETs is specified as 230 μA/μm, meaning that the Ge pMOSFETs in this work show four times the drive current compared to their silicon counterparts at this particular technology node. However, the results of the comparison should be treated with some caution, since the threshold voltage of the Ge pMOSFETs has not yet been considered. On a second note, it was quickly found that this high supply voltage of 1.5 V results in high leakage current at the drain due to the

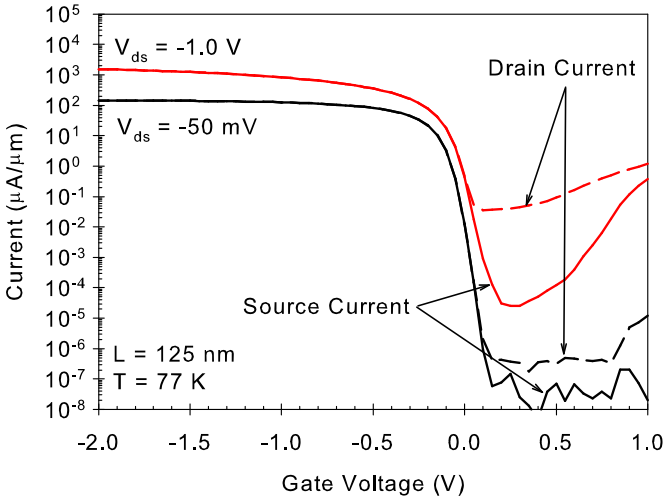


(b)

Figure 6.3: Output characteristics of L = 10 μm Ge pMOSFET measured at (a) room temperature and (b) 77 K.

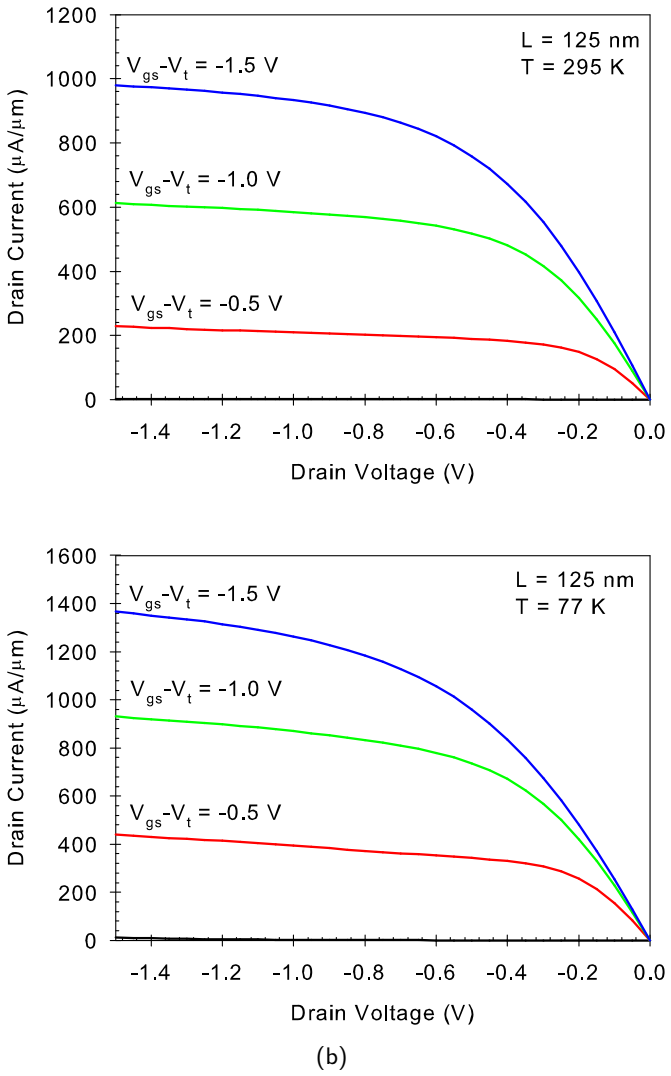


(a)



(b)

Figure 6.4: Transfer characteristics of L = 125 nm Ge pMOSFET measured at (a) room temperature and (b) 77 K.



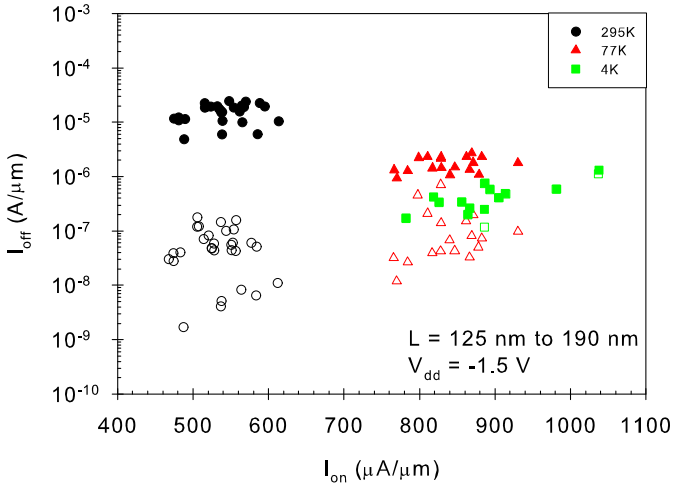
(b)

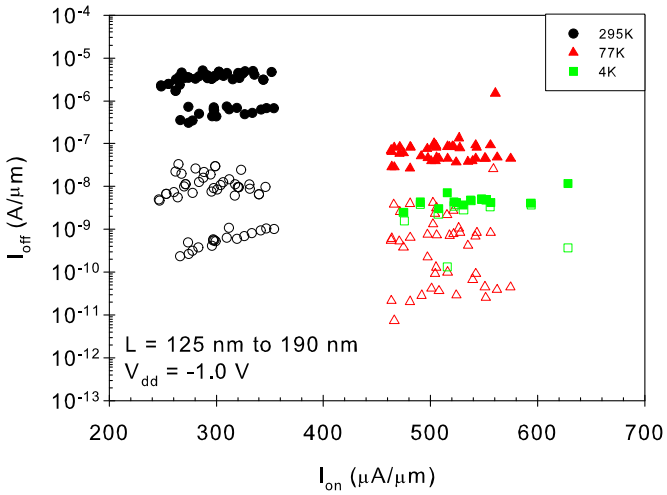
Figure 6.5: Output characteristics of L = 125 nm Ge pMOSFET measured at (a) room temperature and (b) 77 K.

potential across the junction.Consequently, the germanium transistors were evaluated at power supply voltages of V_{dd} = -1.0 V and -1.5 V.

Traditionally I_{on} and I_{off} are evaluated at gate voltages equal to V_{dd} and 0 V, respectively. However, as will be found in Section 6.3.2, the threshold voltage for the germanium devices is not currently optimised, with the devices actually *on* for a gate voltage of 0 V. Consequently, it makes little sense to evaluate the on and off-state currents in the usual way. Instead, the on-current is defined as the current flowing when V_{gs} - V_t = -2/3|V_{dd}| and the off-current is defined as the current flowing when V_{gs} - V_t = 1/3|V_{dd}|, with the minus sign for the on-current required for operation of a pMOS device [Nicholas et al., 2007]. This preserves the historical trend for Si transistors with optimised threshold voltage, where V_t is roughly 1/3 of the power supply voltage, thus leaving 2/3 of the gate voltage swing above V_t for the drive current and 1/3 of the swing below threshold to evaluate the off-current.

Figure 6.6 shows I_{on} versus I_{off} for germanium pMOSFETs with gate lengths between 190 nm and 125 nm, measured at room temperature, 77 K and 4 K, for power supply voltages of (a) -1.5 V and (b) -1.0 V. At room temperature, the highest I_{on} value, for a 125 nm device, was found to be 613 μA/μm and 341 μA/μm, for supply voltages of -1.5 V and -1.0 V, respectively, which is more realistic. At 77 K, these values were improved by more than 50% to 930 μA/μm and 551 μA/μm, respectively. As expected, the I_{on} values measured at the source are equal to those measured at the drain. Even accounting for the non-optimised threshold voltage for the Ge pMOSFETs, the drive current represents nearly a threefold increase over silicon MOSFETs for the 0.18 μm technology node. Further improvements in drive current are realised at 4 K, where the highest drive currents were measured as 1037 μA/μm and 629 μA/μm for supply voltages of -1.5 V and -1.0 V, respectively. It should be mentioned that the variation in the measured I_{off} values (nearly 1 decade) at a given temperature, are are due to subtle variations in the processing of the source and drain regions between





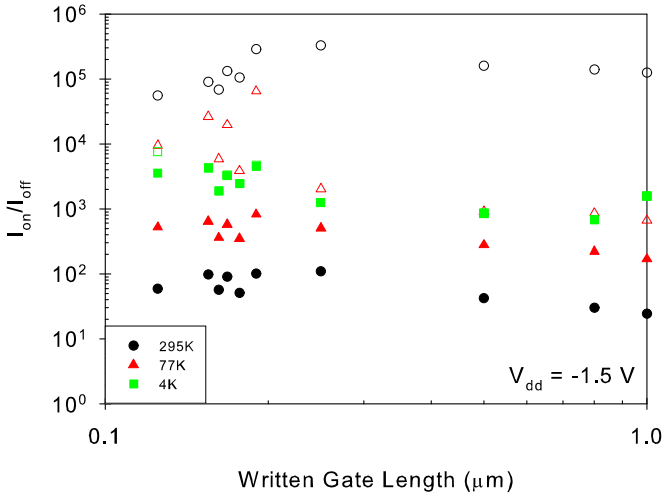
(b)

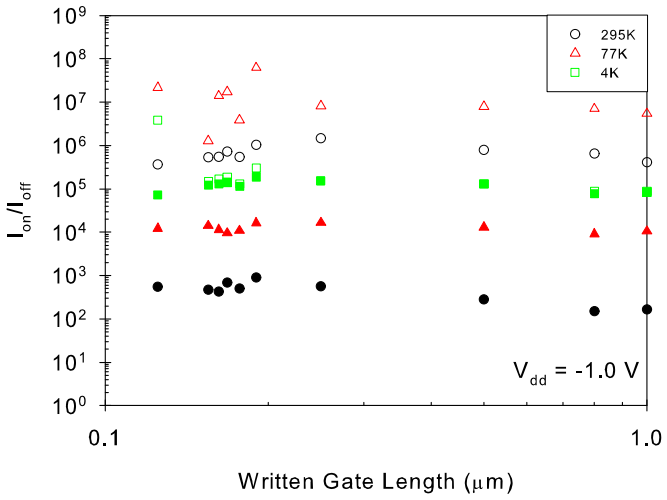
Figure 6.6: I_{off} versus I_{on} for Ge pMOSFETs measured at both the source and the drain at room temperature, 77 K and 4 K for power supply voltages of (a) -1.5 V and (b) -1.0 V. The open symbols represent the current measured at the source; the closed symbols the current at the drain.

different arrays of transistors with gate lengths ranging from 190 nm to 125 nm, and not due to unintentional process variations between transistors of the same gate length like those observed in Chapter 5.

The off-current measured at the drain is typically found to decrease by between 1 and 2 magnitudes when measured at 77 K compared to room temperature, and is several orders of magnitude higher than that measured at the source. However, the off-current measured at the source is found to increase at 77 K compared to room temperature for the higher supply voltage of -1.5 V. The reason for this could be due to a substrate-charging effect, due to a combination of the high drain junction leakage and an increased substrate resistivity, that could induce a slight forward bias between the source and the substrate, giving rise to a higher leakage current across this junction. At 4 K, the off-current measured at the source is equal to the off-current at the drain for both power supply voltages. Whilst the off-current at the drain has been further reduced from those measured at 77 K, the off-current at the source is higher than the values measured at 77 K for both supply voltages, likely to be due to the effects of substrate charging.

The I_{on}/I_{off} ratio is shown in Figure 6.7 as a function of gate length at room temperature, 77 K and 4 K for power supply voltages of (a) -1.5 V and (b) -1.0 V. An I_{on}/I_{off} ratio in excess of 10^5 is observed at room temperature, for the shortest gate length devices, when measured at the source, for a power supply voltage of -1.0 V. This is improved to in excess of 10^7, when measured at 77 K. The I_{on}/I_{off} measured at the drain improves from around $\sim 10^3$ to $\sim 10^4$ for the shortest gate length devices, when the devices are cooled from room temperature to 77 K. The reduced I_{on}/I_{off} ratio is entirely due to the higher leakage observed at the drain. In the same way, the I_{on}/I_{off} ratio is lower for a power supply voltage of -1.5 V compared to $V_{dd} =$ -1.0 V. This is due to the higher off-currents measured at the drain for the higher supply voltage. At 4 K, the I_{on}/I_{off} ratio measured at the source decreases, whilst that at the drain increases.





(b)

Figure 6.7: I_{on}/I_{off} ratio versus gate length for Ge pMOSFETs measured at both the source and the drain at room temperature, 77 K and 4 K for power supply voltages of (a) -1.5 V and (b) -1.0 V. The open symbols represent the current measured at the source; the closed symbols the current at the drain.

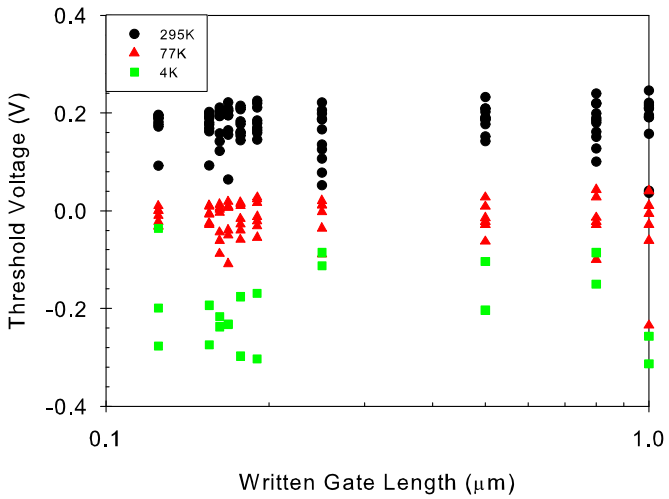


Figure 6.8: Threshold voltage values for germanium pMOSFETs as a function of gate length measured at room temperature, 77 K and 4 K.

This is due to the measured off-current increasing at the source and decreasing at the drain.

6.3.2 Threshold Voltage

Typical values for the threshold voltages of the present germanium transistors are shown as functions of both temperature and gate length in Figure 6.8. At room temperature, the threshold voltage was found to be approximately +0.2 V, which is currently not optimised for pMOS devices. Originally, it was believed that the positive threshold voltage at room temperature was due to a combination of the difference in work function between Ge and the TaN metal gate and also the possibility of negative charge in the HfO_2 layer. However, recent evidence seems to show that this could be related to the use of a Si passivation layer prior to high-κ gate dielectric deposition due to the reduction of the substrate work function by the formation of a dipole layer at the Si/Ge interface [Pourtois et al., 2007].

Figure 6.8 also shows that there is at least 100 mV variation in threshold voltage on different devices with the same gate lengths, residing on the same chip and also on

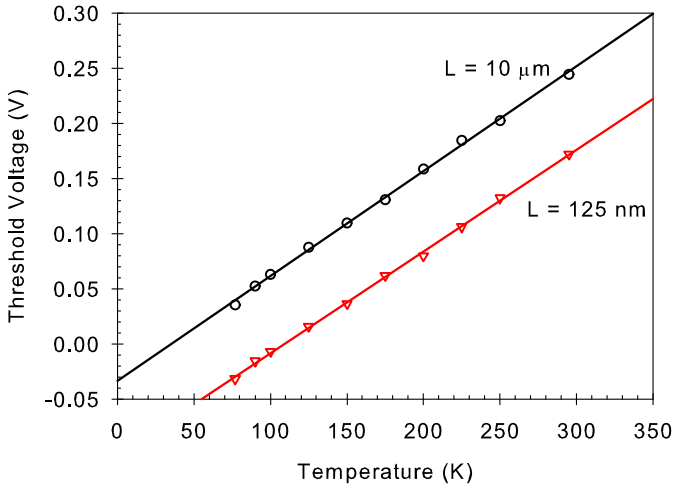


Figure 6.9: Threshold voltage vs temperature for L = 10 μm and L = 125 nm Ge devices.

different chips. This is likely to be due to a combination of interface traps and oxide charges. This variation in threshold voltage needs to be reduced since CMOS circuits will only tolerate a variation in threshold voltage of up to 50 mV across a wafer.

As the temperature is decreased, the threshold voltage becomes more negative (i.e. higher) for all gate lengths. The threshold voltage for devices with gate lengths of 10 μm and 125 nm is plotted as a function of temperature in Figure 6.9. The shift in threshold voltage with temperature was found to be approximately 1 mV/K for all devices. This shift is due to a combination of an increase in bandgap and also the position of the bulk Fermi level, which resides closer to the conduction band edge.

As the channel length is decreased, no significant threshold voltage roll-off is observed indicating that the halo implants are effective in controlling short channel effects in this particular batch of Ge pMOSFETs.

6.3.3 Subthreshold Slope, DIBL and Transconductance

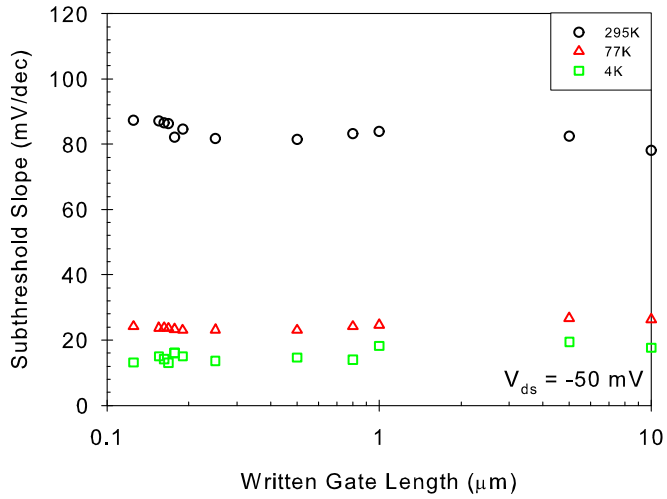
The value of the subthreshold slope is one of the fundamental criteria for benchmarking new technologies for logic applications, providing a direct insight into the scal-

ability of a technology [Chau et al., 2005]. Fundamentally, the subthreshold slope determines the ability of a MOSFET to turn on and off. The subthreshold slope, measured from the source current, is shown as a function of gate length in Figure 6.10, measured at a drain bias of (a) V_{ds} = -50 mV and (b) V_{ds} = -1.0 V, at room temperature, 77 K and 4 K. The subthreshold slope was evaluated at the source rather than the drain since the higher leakage current at the drain means that it is not possible to extract the subthreshold slope from the steepest part of the transfer characteristics [Nicholas et al., 2007].

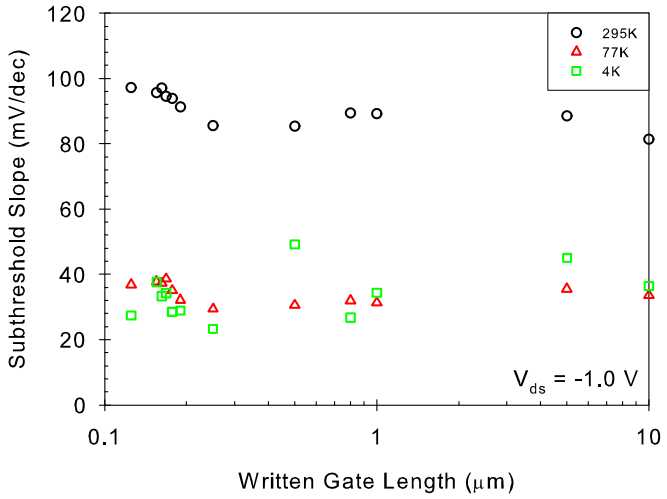
For a low drain bias of V_{ds} = -50 mV, typical values for the subthreshold slope at room temperature are around 80 mV/dec and are approximately constant with channel length. These values are reduced to around 25 mV/dec and 15 mV/dec at 77 K and 4 K, respectively.

A subthreshold slope of 97 mV/dec were measured at gate lengths of 125 nm at a drain bias of V_{ds} = -1.0 V. This is somewhat higher than the 70 mV/dec reported in planar Si transistors with comparable gate lengths [Chau et al., 2005]. Even for long-channel devices, the subthreshold slope only reaches 81 mV/dec. Whilst this compares favourably to previously reported results in the literature [Ritenour et al., 2003, Shang et al., 2002], this indicates that further improvements can still be made in the quality of the germanium/high-κ interface. At 77 K, the subthreshold slope is typically reduced to between 30 and 40 mV/dec at a drain bias of V_{ds} = -1.0 V. Whilst there is a slight improvement of the subthreshold slope for the shortest channel devices at 4 K, some of the long channel devices actually exhibit higher subthreshold slope values than observed at 77 K. The reason for this is unclear, but the extraction may be adversely affected by the higher off-current at the source.

Values for the subthreshold slope for a device with a gate length of 125 nm are shown for V_{ds} = -50 mV and -1.0 V in Figure 6.11 as a function of temperature. The device shows a smooth, continuous reduction in the subthreshold slope for both values



(a)



(b)

Figure 6.10: Subthreshold slope versus gate length for devices measured at room temperature and 77 K, with V_{ds} = (a) -50 mV and (b) -1.0 V.

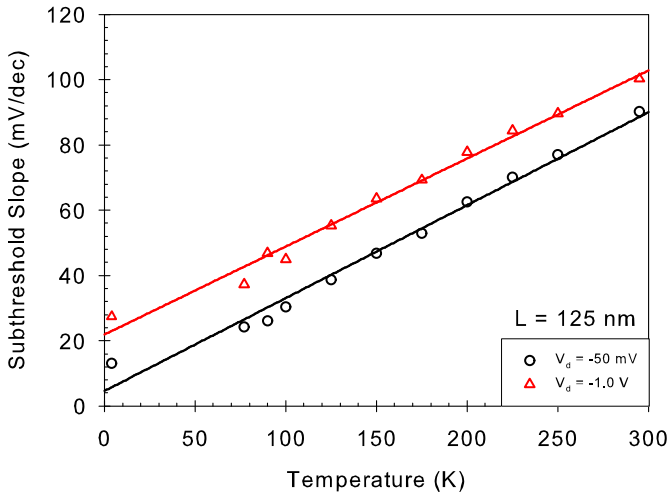


Figure 6.11: Subthreshold slope versus temperature for a device with a gate length of 125 nm with V_{ds} = -50 mV and -1.0 V.

of V_{ds}. Recalling Equation 6.1, the subthreshold slope is given by:

$$S = 2.3\frac{k_B T}{q}\left(1 + \frac{C_{it} + C_{dep}}{C_{ox}}\right) \tag{6.1}$$

Consequently, a plot of the subthreshold slope vs temperature should produce a straight with a gradient equal to $1 + (C_{dep} + C_{it})/C_{ox}$, assuming that the depletion, interface trap and oxide capacitances are independent of temperature. Using this method a value for D_{it} was extracted as $\sim$ 4 x 10^{12} $cm^{-2}eV^{-1}$, higher than those previously reported by room temperature charge-pumping measurements on these transistors [Kaczer et al., 2007]. The large discrepancies are concerning. Recent work by Batude et al. [2007] showed that capacitances associated with bulk traps in Ge (not considered in this simple theory) cannot be ignored as they are in silicon, suggesting that the traditional techniques used to extract D_{it} values in silicon will have to be readdressed, and adapted, as necessary, for the specific case of germanium.

Figure 6.12 shows the values of DIBL as a function of written gate length, extracted from the source current. Tolerable values of DIBL of approximately 50 mV/V were observed at gate lengths of 125 nm. As the temperature is decreased to 77 K and

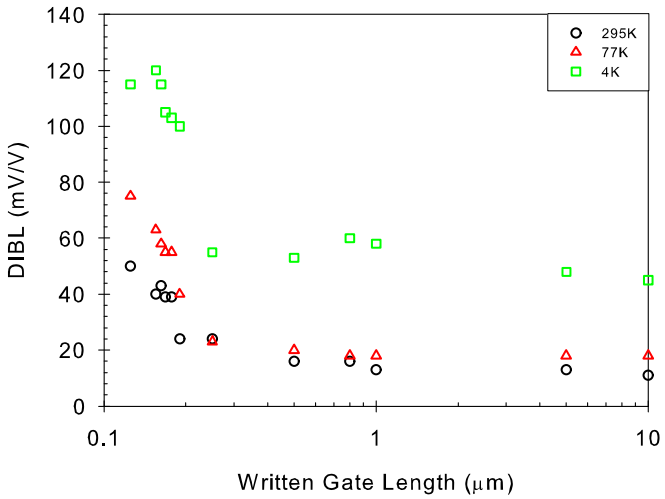
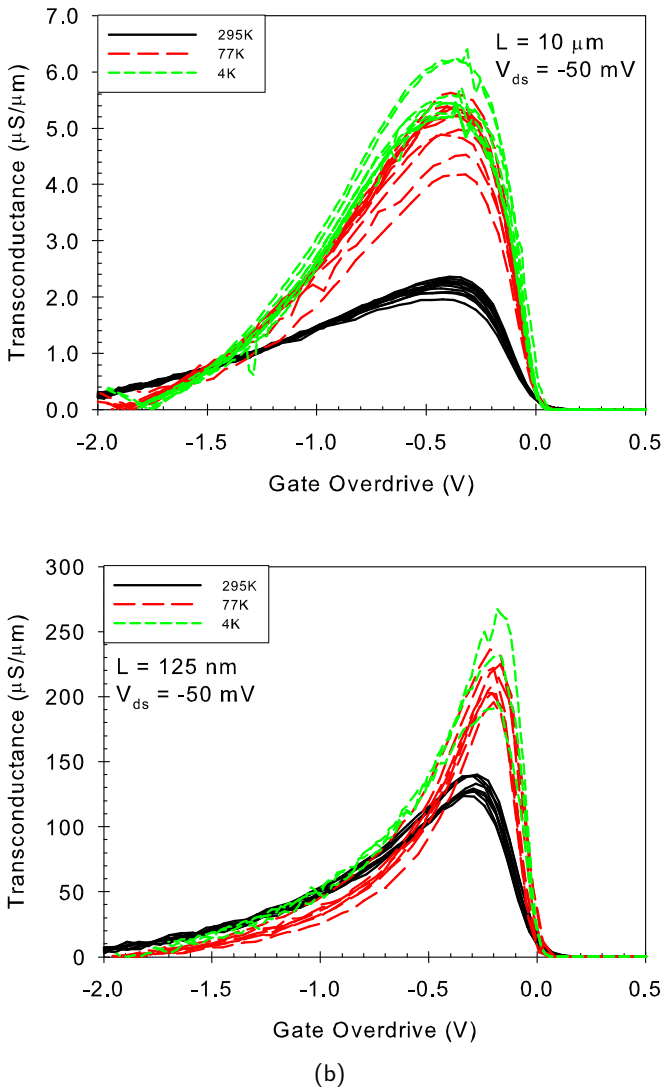


Figure 6.12: DIBL as a function of gate length measured at room temperature, 77 K and 4 K.

4 K, the DIBL values were found to increase. This could be due to partial freeze out of the halo dopants, thus reducing their effectiveness. In addition, the substrate charging effect might also be important, particularly at 4 K, where the off-current measured at the source was found to increase.

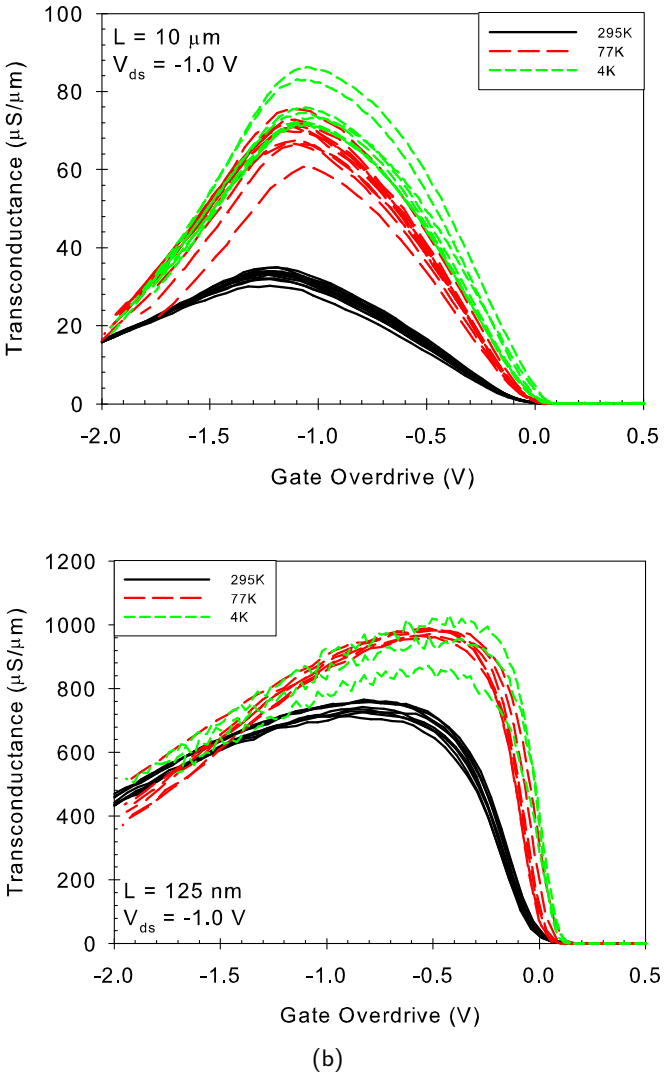
The transconductance is shown in Figures 6.13 and 6.14 for a selection of devices with gate lengths of 10 μm and 125 nm, measured at both room temperature, 77 K and 4 K.

The long-channel (L = 10 μm) transconductance measurements taken in the linear regime, with low drain bias, show between 2 and 2.5 times enhancement in the maximum value when measured at 77 K compared to room temperature. The short-channel devices, with a gate length of 125 nm, exhibit a more modest improvement in the peak transconductance, with a typical enhancement of around 60%. at either 77 K or 4 K. Whilst one or two devices show a higher peak transconductance at 4 K, in the main there is no significant improvement in transconductance between 77 K and 4 K. For the long-channel devices, the improvement in the peak transconductance for a device operated at either 77 K or 4 K diminishes for values of gate overdrive, V_{gs} -



(b)

Figure 6.13: Linear transconductance for a selection of devices with gate lengths of (a) L = 10 μm and (b) 125 nm at room temperature, 77 K and 4 K.



(b)

Figure 6.14: Saturation transconductance for a selection of devices with gate lengths of (a) L = 10 μm and (b) 125 nm at room temperature, 77 K and 4 K.

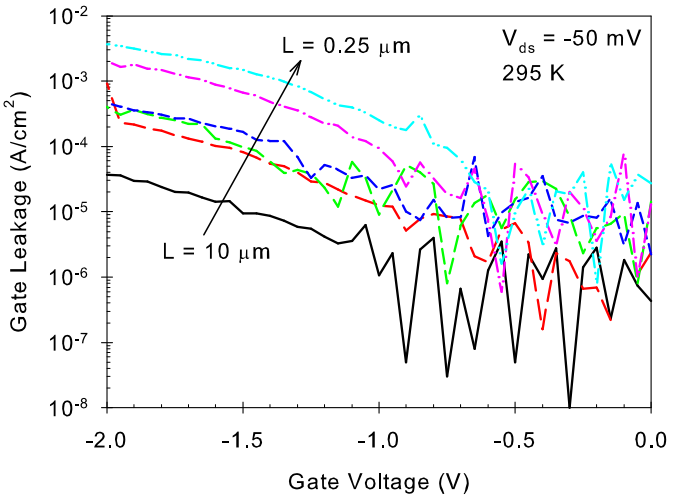
V_t = -1.5 V. As will be discussed in further detail in Section 6.3.6, this is the region where the mobility is limited by surface roughness scattering, which should be relatively temperature independent. The transconductance improvements with decreasing temperature are only realised in the short-channel devices for values of gate overdrive below -0.5 V.

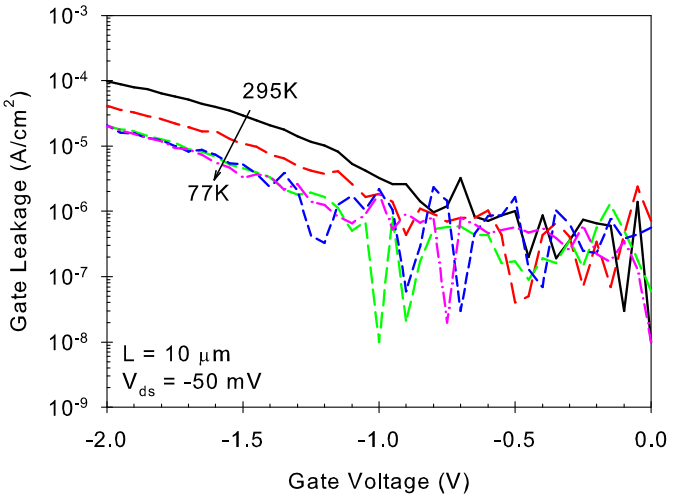
The corresponding saturation transconductance measurements are shown in Figure 6.14 for an operating voltage of $V_{dd} = V_{ds}$ = -1.0 V. This is more realistic in terms of how a device would usually be operated. In the absence of any mobility degradation, the saturation transconductance should be linearly dependent on the applied gate bias for long-channel devices. However, in practice, as the gate voltage is increased, the mobility decreases due to interface roughness scattering. Clearly, this is evident in the germanium devices, which clearly show a degradation in the transconductance with increasing gate bias. Furthermore, mobility degradation is more severe in devices operated at cryogenic temperatures, and therefore interface roughness scattering is more dominant there.

6.3.4 Gate Leakage

The gate leakage current for germanium pMOSFETs with gate lengths from 10 μm down to 0.25 μm is shown in Figure 6.15(a) for low drain bias at room temperature in the inversion regime. Low gate leakage is observed at all gate lengths, and is much lower than ITRS roadmap requirements for the specified EOT of 1.3 nm. Figure 6.15(a) also indicate a gate leakage dependence on the channel length of the transistor, which may be due to gate leakage in the gate overlap regions of the source and drain.

Figure 6.15(b) shows the gate leakage current on 10 μm x 10 μm pMOSFETs as a function of temperature. The gate leakage is observed to decrease with temperature. This could be due to a suppression of the trap-assisted component of the gate leakage at lower temperatures and possibly a change in the tunnelling barrier height.





(b)

Figure 6.15: (a) Gate leakage at room temperature as a function of gate length (b) as a function of temperature for an L = 10 μm device.

6.3.5 Split C-V Characteristics

Split C-V measurements were conducted on 10 μm x 10 μm devices to calculate the effective mobility at a range of temperatures between 4 K and room temperature. The inversion charge density was calculated using a measurement frequency of 1 MHz to minimise the contributions of interface traps to the mobile inversion charge, that could lead to an overestimate over the inversion charge and hence an underestimate of the effective mobility. Figure 6.16 shows the split C-V characteristics on a 10 μm x 10 μm pMOSFET as a function of the ac signal frequency. The gate-channel branch of the split C-V curve shows a small bump in weak inversion, not present at a frequency of 1 MHz. This additional capacitance is likely to be due to interface states. No additional frequency dispersion was observed in inversion, and so extracting the inversion charge density at a frequency of 1 MHz is justified. The gate-body branch of the split C-V characteristics does exhibit a frequency dependence. Firstly, in strong accumulation, the 1 MHz C-V curve is suppressed compared to the 10 kHz and 100 kHz curves, which is likely to be the result of series resistance. Secondly, a bump in the gate-body branch of the split C-V curves is observed in depletion at around 0.5 V, which is likely to be due to some defect states at the interface. The bump is still visible at 1 MHz, albeit smaller, indicating that these defects are fast-responding states close to the band edge. The fact that these defects appear in the gate-body branch of the split C-V characteristics indicates that they are in the upper half of the Ge bandgap. It should be noted that the defect bump in the gate-body branch of the split C-V characteristics is not typical behaviour of other germanium transistor lots processed by IMEC and appears related to the chips from this particular batch. The origin is unknown and is likely to be due to defects that have been introduced during a process step. Despite the suppressed accumulation capacitance, the 1 MHz curve was also used to determine the depletion charge, required to calculate the effective field, since the depletion charge only requires

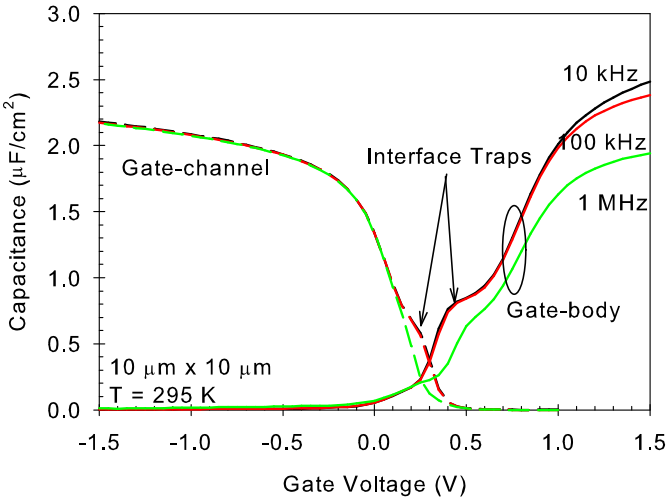


Figure 6.16: Split C-V characteristics of a 10 μm x 10 μm device as a function of frequency at room temperature.

the tail of the gate-body curve, and hence it was assumed that the error associated with the suppression of the 1 MHz curve in accumulation would be smaller than the additional defect-related capacitance in depletion.

Figure 6.17 shows split C-V characteristics on a 10 μm x 10 μm device at a range of temperatures between 4 K and 295 K. As the temperature is decreased, the width of the depletion capacitance region increases, as the device undergoes abrupt transition from accumulation to depletion and from depletion to inversion. This is due to a positive shift in the flatband voltage and the threshold voltage becoming increasingly negative. Interestingly, as the temperature decreases, the bumps in both branches of the C-V characteristics are diminished, such that at a frequency of 100 kHz they are no longer visible. Both the depletion and inversion capacitance are observed to decrease with decreasing temperature, due to partial freezing out of substrate dopants.

The temperature dependence on the flatband voltage is also evident in Figure 6.18 . As the temperature is decreased, a positive flatband voltage is observed, indicating an increase in negative charge. This is speculated to be due to acceptor-like defects at the Ge/HfO_2 interface [Gray and Brown, 1966]. Acceptor-like defects occu-

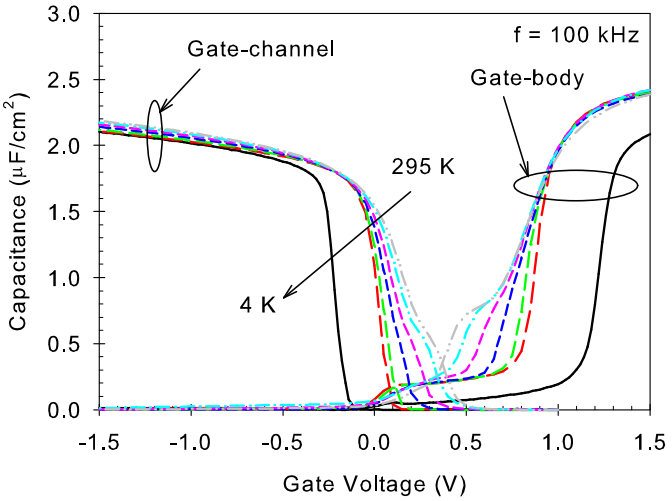


Figure 6.17: Split C-V characteristics of a 10 μm x 10 μm device as a function of temperature.

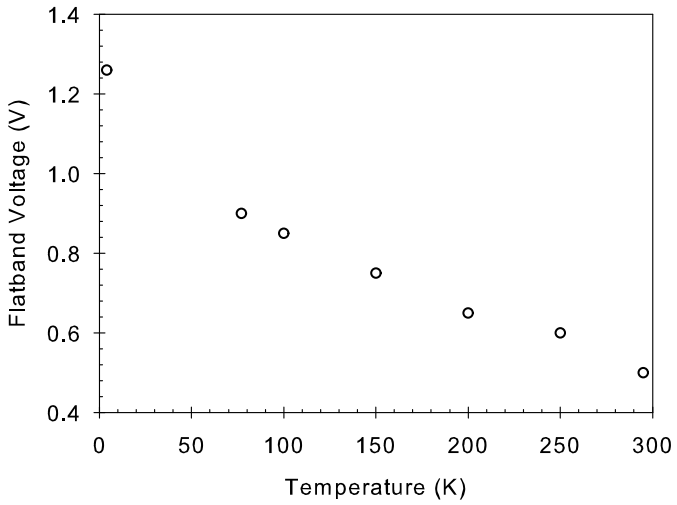


Figure 6.18: Flatband voltage as a function of temperature for a 10 μm x 10 μm.

pied below the Fermi-level are negative, and those above the Fermi-level are unoccupied and neutral.

Ideally, the flatband voltage should be zero. However, in reality the flatband voltage will be non-zero, due to the metal-semiconductor work function and the presence of charges in the oxide. Consequently, the flatband voltage is the voltage that must be applied in order to restore the flatband condition. As the temperature is decreased from room temperature to 4 K, the Fermi-level is pushed closer to the conduction band edge. In the absence of interface states, this would not cause a shift in flatband voltage, since the Fermi-level at the surface will be equal to that at the bulk. However, if there are defects at the interface such that they become negatively charged when occupied below the Fermi-level, then these negative charges will act to bend the bands upwards. Hence, in order to restore the flatband condition, it will be necessary to apply a positive voltage on the gate.

6.3.6 Effective Mobility

Traditionally, the hole mobility has been considered to be the most important parameter, especially for long-channel devices. In particular, comparison with the silicon universal mobility curve is useful means of gauging the performance benefits of the new technology over silicon. Additionally, when the effective mobility is plotted against the effective field, valuable insight is gained into the dominant scattering mechanisms in the device, which is useful in highlighting areas in which further improvements in design and processing can be made.

The effective mobility is plotted against vertical field in Figure 6.19 for a selection of 10 μm x 10 μm pMOSFETs at room temperature, 77 K and 4 K. The effective field was calculated using the relation:

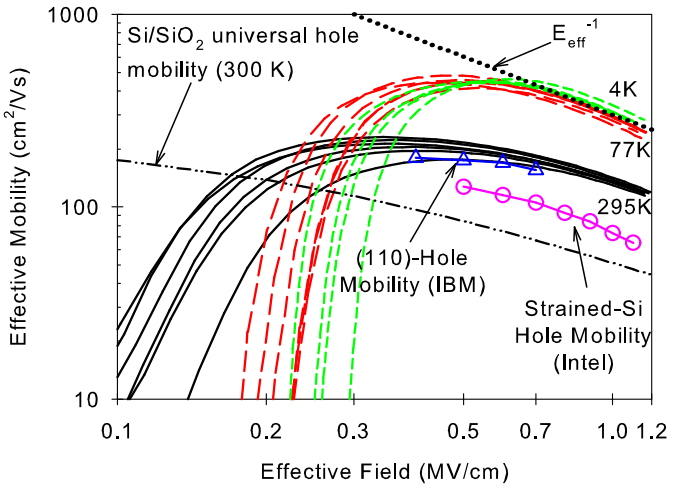


Figure 6.19: Effective hole mobility at room temperature, 77 K and 4 K for 10 μm x 10 μm germanium pMOSFETs. Data also includes hole mobility values for (100)-silicon [Yang et al., 2003] and for uniaxial strained silicon with SiGe source/drain regions ([Ghani et al., 2003]).

$$\varepsilon_{eff} = \frac{Q_{bulk} + \eta Q_{inv}}{\kappa_{Ge}\epsilon_0} \tag{6.2}$$

There is some question about the value of η. Takagi et al. [1994] showed that a value of $1/3$ is valid for holes at both room temperature and 77 K for silicon MOSFETs. This was the value of η such that the effective mobility could be described by a single universal curve. However, at present, the universal behaviour of germanium MOSFETs has not yet been explored. A universal law in germanium MOSFETs could be potentially be quite complicated due to the vast combination of gate stacks to choose from, and the possibility that they might not all behave in the same way. Consequently, to the best of the author's knowledge, all of the published literature on the effective mobility of germanium MOSFETs is based on the assumption that they behave in the same way as silicon MOSFETs and assume the same value for the parameter η. In this work, a value of $1/3$ was assumed for the value of η for all temperatures. Perhaps, this assumption is not so unrealistic, given that the germanium pMOSFETs used in this work are passivated with Si/SiO_2, and so is likely to be as close to resembling a standard Si MOSFET as is

possible for a Ge device.

The Ge pMOSFETs exhibit a maximum room temperature hole mobility of approximately 230 $cm^2V^{-1}s^{-1}$. The silicon universal curve is shown by means of comparison. The germanium devices maintain a hole mobility enhancement of 2.5 times that of the universal curve, even at fields of 1 $MVcm^{-1}$, typically where the devices would be operated. Figure 6.19 also indicates that there is a spread in mobility with the worst device having a peak mobility of 150 $cm^2V^{-1}s^{-1}$ at room temperature. Nonetheless, even the worst device in this case exhibits a hole mobility enhancement of two times that of the silicon universal curve at an effective field of 1 $MVcm^{-1}$ at room temperature. These devices represent some of the highest reported mobilities in germanium MOSFETs with high-κ dielectrics at vertical fields where sub-micron devices would typically be operated.

Figure 6.19 also shows the effective hole mobility for (110)-oriented silicon substrates [Yang et al., 2003] and also for uniaxially strained silicon pMOSFETs with SiGe source/drain regions [Ghani et al., 2003]. The germanium pMOSFETs exhibit a significant enhancement (nearly two times) in hole mobility over the strained-Si pMOSFETs, whilst the improvement in mobility over the (110)-oriented silicon pMOSFETs is much more modest. This is particularly encouraging since neither of these devices employed high-κ gate dielectric. It is envisaged that further mobility enhancements can still be made in germanium pMOSFETs through strain and changing the crystal orientation. Additionally, it is also possible that a reduction of the gate dielectric thickness might also be expected to bring out a further increase in the hole mobility due to reduced charge trapping in the thinner oxide, providing that this does not result in any significant increase in gate leakage [Ragnarsson et al., 2006].

Figure 6.19 also shows the effective hole mobility of Ge pMOSFETs at both 77 K and 4 K. The effective hole mobility can be increased by approximately a factor 2 at moderate to high vertical fields when operated at 77 K, with a peak mobility of around

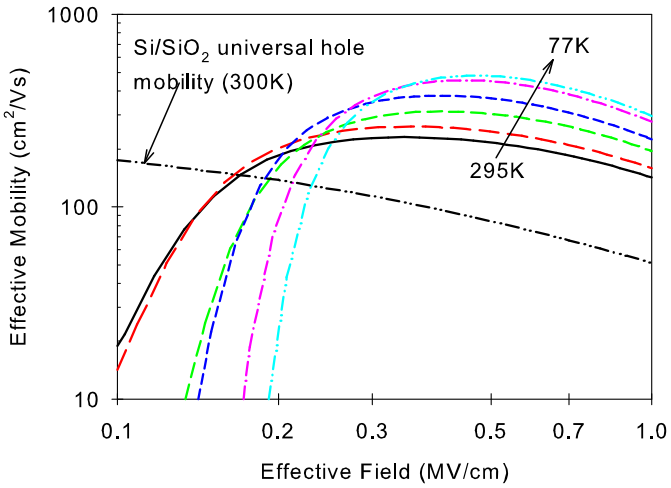


Figure 6.20: Effective hole mobility vs effective field for a 10 μm x 10 μm germanium pMOSFET as a function of temperature.

480 $cm^2V^{-1}s^{-1}$. This increase in hole mobility is almost entirely due to the reduction of phonon scattering in the channel. However, no significant increases in hole mobility were observed when cooling from 77 K to 4 K. This can be interpreted that below 77 K, phonon scattering is not the limiting factor to the effective hole mobility. Rather, that the hole mobility is limited by Coulomb scattering at low vertical fields and surface roughness scattering at high vertical fields. This is further illustrated by Figure 6.20, which shows the effective hole mobility as a function of vertical field measured on the same transistor over a range of temperatures. At low vertical fields, the hole mobility is found to exhibit a steeper roll-off with decreasing temperature. The Coulomb scattering results from the high substrate doping of these transistors, including halo implants, and is found to become increasingly important with decreasing temperature due to the reduction of the hole carrier energy. In addition, it is also possible that additional Coulomb scattering could result from the increased occupancy of acceptor-like defects at the Ge/HfO_2 interface as the temperature is decreased (see Section 6.3.5). At high vertical fields, the hole mobility increases until 77 K, below which, the mobility is almost temperature-independent, indicating that the mobility in this temperature regime is limited entirely

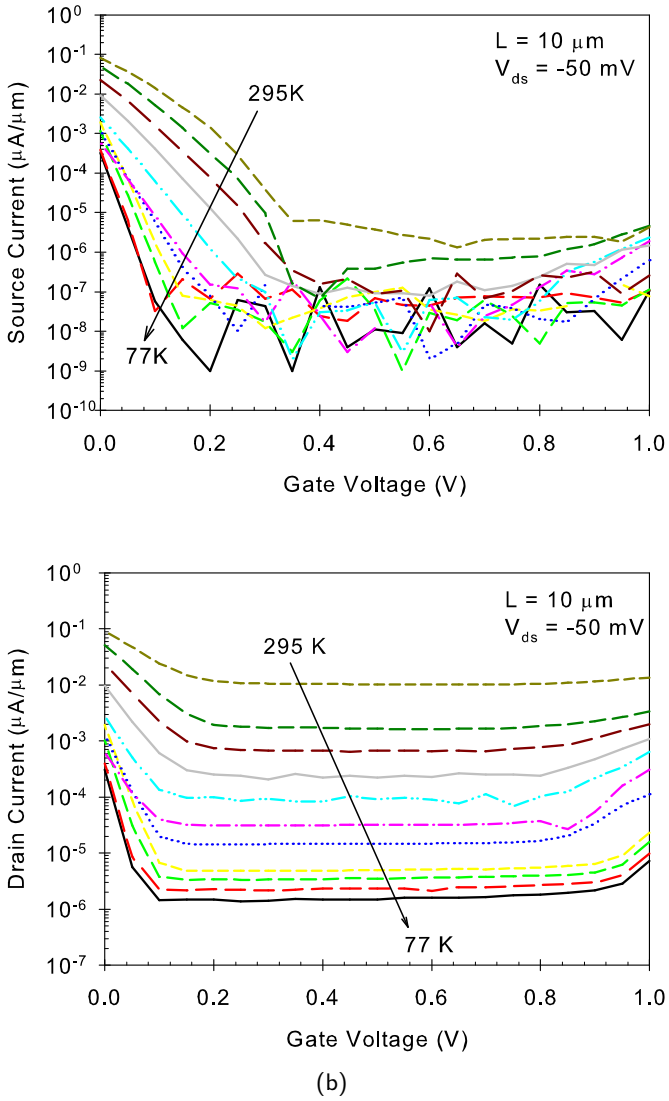
by surface scattering. In this regime, the hole mobility is nearly proportional to ε_{eff}^{-1}, which is the same dependence as reported in silicon [Takagi et al., 1994].

6.3.7 Off-State Leakage

Band-to-band tunnelling is becoming increasingly important to the leakage currents in modern VLSI devices [Taur and Ning, 1998]. However, it is perceived that band-to-band tunnelling could be even more important in germanium channel devices due to the lower bandgap. The off-state leakage current was investigated as a function of temperature on 10 μm x 10 μm devices. Figures 6.21 and 6.22 shows the off-state characteristics measured at both the source and drain as a function of temperature for both low and high drain biases of V_{ds} = -50 mV and -1.0 V.

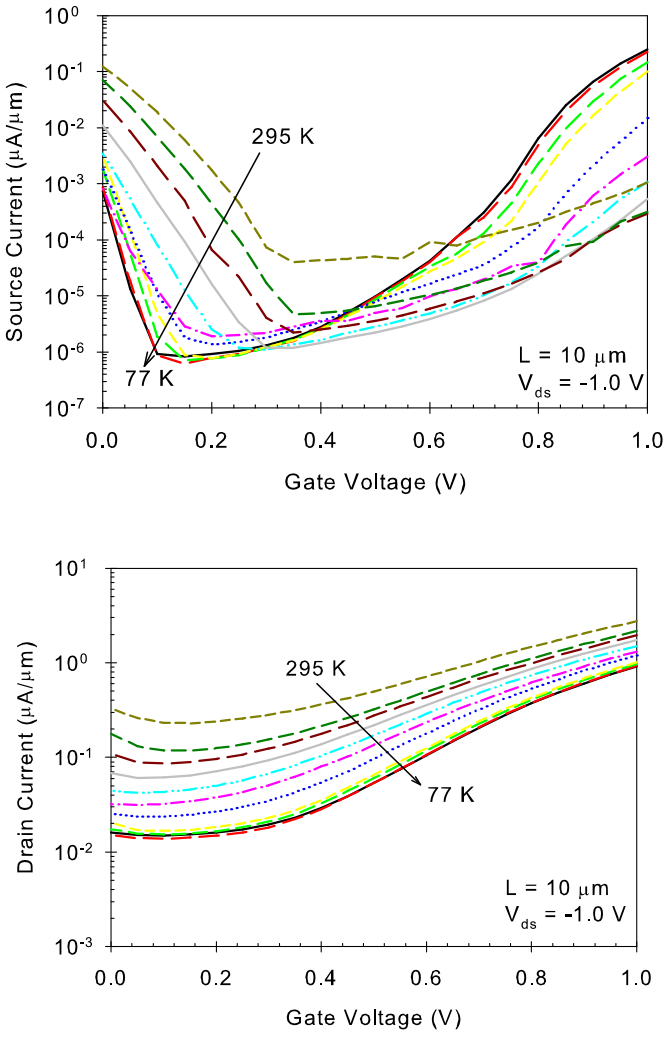
When measured at room temperature, the leakage current in the linear regime decreases by approximately 4 decades as the device is cooled from room temperature to 77 K, as shown in Figure 6.21(b), when measured at the drain. There is a slight increase in the off-state leakage current when the gate voltage exceeds around 0.8 V, which could be due to the effects of gate-induced drain leakage (GIDL). Conversely, the leakage current for the device measured in the saturation regime, decreased by less than 2 decades between room temperature and 77 K when measured at the drain, suggesting that different leakage mechanisms are involved in the germanium transistors. GIDL also appears to play an important role when the device is measured in the high drain bias regime, with the leakage current increasing for gate voltages above 0.4 V. This behaviour was typical of devices of all gate lengths down to 125 nm, as shown in Figure 6.23, which shows the leakage currents on a 125 nm x 10 μm device at temperatures between room temperature and 77 K.

At the source, the leakage current is typically several orders of magnitude lower than that at the drain, under the application of a low drain bias, and is almost independent of temperature for gate voltages above 0.4 V. The increased noise at the source is



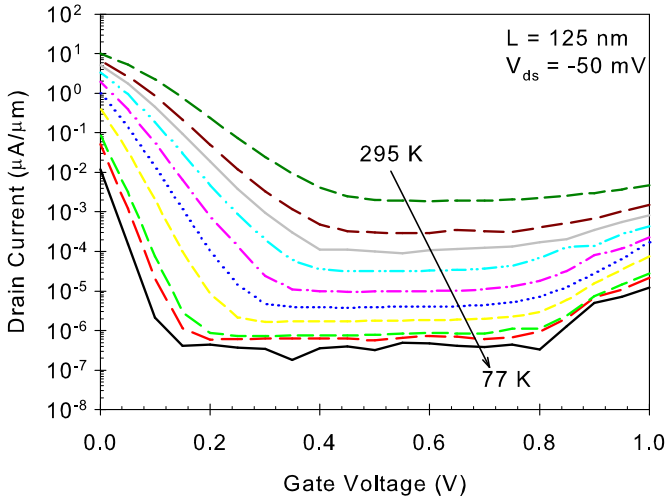
(b)

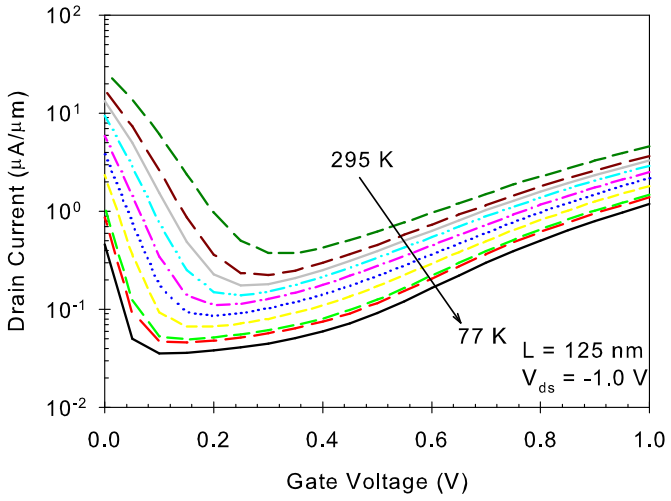
Figure 6.21: Leakage current measured from (a) the source current and (b) the drain current on 10 μm x 10 μm devices at a low drain bias of V_{ds} = -50 mV as a function of temperature.



(b)

Figure 6.22: Leakage current measured from (a) the source current and (b) the drain current on 10 μm x 10 μm devices at a high drain bias of V_{ds} = -1.0 V as a function of temperature.





(b)

Figure 6.23: Leakage current measured on 125 nm x 10 μm devices with (a) V_{ds} = -50 mV and (b) V_{ds} = -1.0 V, as a function of temperature.

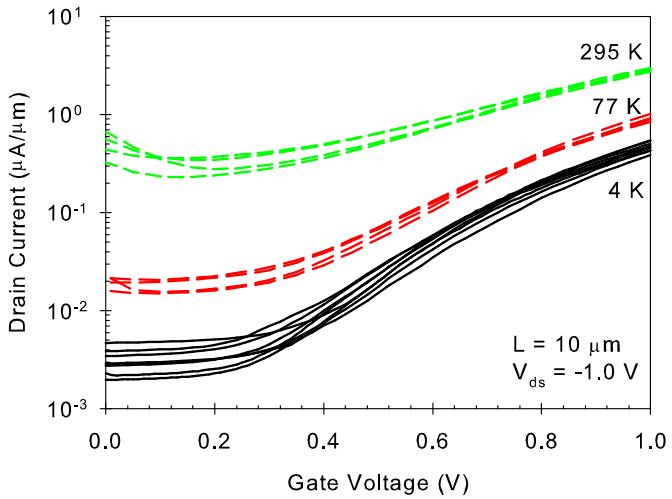


Figure 6.24: Leakage current measured on a selection of 10 μm x 10 μm devices with V_{ds} = -1.0 V, at room temperature, 77 K and 4 K.

due to the current being close to the resolution of the parameter analyser. However, at high drain bias, there appears to exist a 'crossover', since for low gate voltages, the leakage shows a decrease of between 1 and 2 decades between room temperature and 77 K, but for gate voltages higher than around 0.4 V, the leakage increases with decreasing temperature. This is likely to be due to a substrate-charging effect, which results in the forward bias of the source-substrate junction, thus increasing the leakage at the source. The remainder of this section shall focus on the the leakage currents at the drain under a high drain bias, since this is the most relevant to band-to-band-tunnelling.

Figure 6.24 shows the leakage current measured at the drain for a selection of 10 μm x 10 μm transistors measured at room temperature, 77 K and 4 K with a drain bias of -1.0 V. For a gate voltage of 0 V, there is approximately a two-decade reduction in the off-state current. However, as the gate voltage increases, the effects of GIDL become important and for a gate voltage of +1.0 V, there is only 1 decade difference between room temperature and 4 K.

In order to try and understand the mechanisms behind the leakage mechanisms, an Arrhenius plot was attempted to try and determine an activation energy for the

Table 6.1: Summary of the activation energy values for the off-state leakage in germanium pMOSFETs.

	Slope 1	Slope 2
V_{ds} = -50 mV	$E_a \sim$ 0.2 eV	$E_a \sim$ 0.04 eV
V_{ds} = -1.0 V	$E_a \sim$ 0.05 eV	$E_a \sim$ 0.01 eV

leakage mechanism, assuming a leakage current should be of the form:

$$I_{off} = Aexp\left(-E_a/k_BT\right) \tag{6.3}$$

where E_a is the activation energy of the leakage current and A is a pre-exponential function, which in this work has been assumed to be independent of temperature. For this purpose, the off-current was taken as the minimum value of the drain current to avoid the problem of the unoptimised threshold voltage and minimise the effects of GIDL.

Arrhenius plots of I_{off} vs $1/k_BT$ are shown in Figure 6.25 for drain voltages of (a) -50 mV and (b) -1.0 V, respectively, for devices with gate lengths from 10 μm to 0.25 μm. In both cases, the Arrhenius plots appear to exhibit two slopes, with a transition at around 200 K, giving two activation energies for both low and high drain bias, which are summarised in Table 6.1. The activation energies are found to be independent of gate length.

The low activation energies, especially when the devices are biased at V_{ds} = -1.0 V, suggest a leakage mechanism that is not thermally activated. Therefore, the leakage mechanisms are likely to be related to band-to-band tunnelling mechanisms, including GIDL. Indeed, it might be possible to study the independent role of the gate and drain voltages on the off-state leakage, possibly with additional measurements on p^+-n Ge diodes, but this was beyond the scope of this work. The higher activation energy of 0.2 eV for a drain bias of -50 mV at lower temperatures implies a different leakage mechanism. The fact that the energy lies within the germanium bandgap suggests that

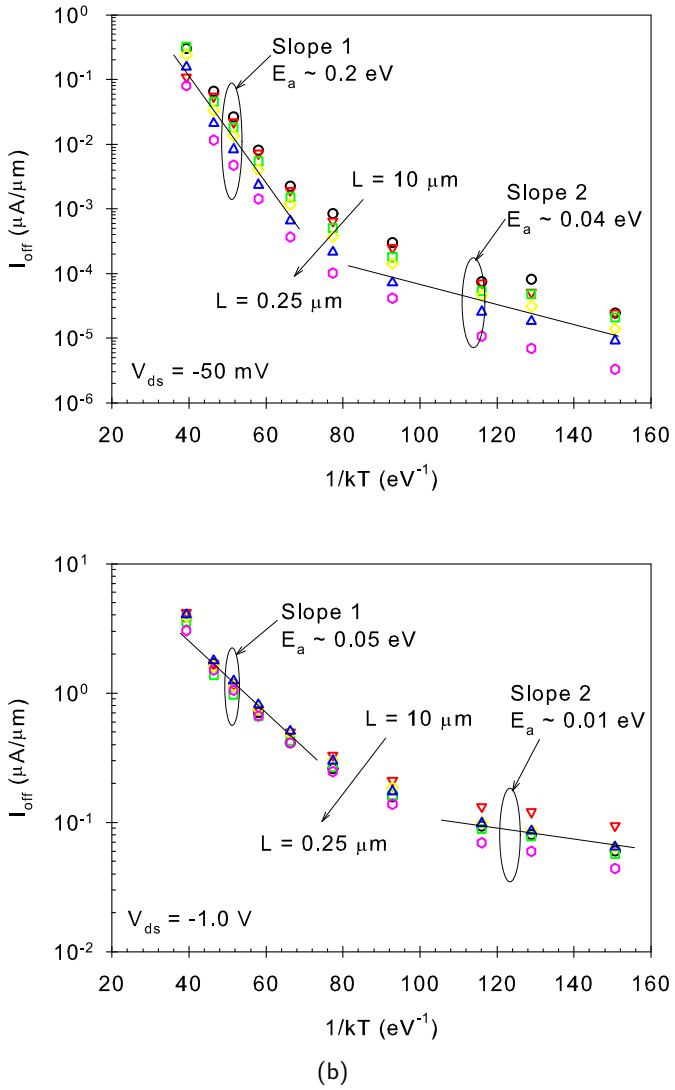


Figure 6.25: Arrhenius plots of I_{off} vs $1/k_BT$ for devices with gate lengths from 10 μm to 0.25 μm for (a) V_{ds} = -50 mV and (b) V_{ds} = -1.0 V.

the defects may be responsible. These defects could either be bulk defects such as threading dislocations, due to the relaxation of the germanium layer, or interface traps in the gate-drain overlap region.

Attempts have also been made to fit the experimental data to GIDL models in the literature, such as those proposed by Chen et al. [1987]. So far, this has proved unsuccessful and remains a work in progress to be reported at a later date.

6.4 Conclusion

Germanium pMOSFETs have been fabricated in a Si-like process with written gate lengths down to 125 nm, and their electrical characteristics have been studied as a function of temperature down to 4 K. The germanium pMOSFETs have high hole mobility, with a peak value of around 230 $cm^2V^{-1}s^{-1}$ at room temperature, and are higher than those reported in the literature for (110)-silicon pMOSFETs and uniaxial strained Si pMOSFETs. Whilst higher peak mobility values have been reported, the germanium pMOSFETs investigated in this thesis maintain their hole mobility enhancement over the silicon universal curve at high effective fields (over 1 $MVcm^{-1}$), typically where a device would be operated. The hole mobility was found to increase to around 480 $cm^2V^{-1}s^{-1}$ when measured at 77 K, due to the reduction of phonon scattering. However, no further significant improvements in the hole mobility were observed at 4 K, indicating that below 77 K, the mobility is limited by Coulomb scattering and surface roughness scattering at low and high vertical fields, respectively.

The high hole mobility values extracted in long-channel devices were found to translate to high drive currents in short-channel devices, with 125 nm gate length devices, easily surpassing the ITRS requirements for the 180 nm technology node, with the exception of the off-state leakage current. The off-state leakage current in germanium pMOSFETs remains problematic. It may be that band-to-band-tunnelling mechanisms,

including GIDL, will be a limiting factor in the off-current for germanium devices. Ultimately, if germanium is going to succeed as a channel material in CMOS circuits, it is highly likely that it will have to be on an insulator platform, which should substantially reduce the leakage currents. Additional reductions in leakage current may also be achieved through reduction of the threading dislocation density of the germanium layers [Nayfeh et al., 2005] and also by using Ge-channel heterostructure transistor designs [Saraswat et al., 2006]. It might also be possible to investigate Schottky barriers as a way of reducing the off-state leakage without significant degradation to the drive current [Li et al., 2006].

Chapter 7

High-Field Transport Effects in Buried $Si_{1-x}Ge_x$ Channel pMOSFETs

7.1 Introduction

In recent years, there have been many published works showing the improved performance of long-channel silicon-germanium alloy channel devices compared to standard bulk silicon MOSFETs (see, for example [Parker and Whall, 1999, Lee et al., 2005, Xie, 1999, Leadley et al., 2003]). In particular, this area of research has focussed on improving the hole mobility, since this is four times lower than the electron mobility in bulk silicon, and ultimately results in pMOSFETs being approximately twice as wide as the nMOSFETs in a CMOS circuit, effectively reducing the number of devices on a chip [Taur and Ning, 1998]. The majority of the research on silicon-germanium channel devices has demonstrated improved hole mobility and current drives in long channel devices due to the reduction of the hole effective mass and reduced intervalley scattering that results from the splitting of the heavy and light hole subbands [Schäeffler, F., 1997].

However, there have been relatively few in-depth studies on short channel devices, and with current state-of-the-art silicon devices already having gate lengths below 50 nm, there is a need to understand whether or not this improvement in silicon-germanium channel devices is maintained at technologically important gate lengths. There are several technological problems associated with the integration of silicon-germanium channel devices. In particular, the issue of finding a suitable high-κ dielectric was addressed in Chapters 4 and 5. Although, perhaps a more fundamental problem for silicon-germanium channel devices could be the lower hole saturation velocity compared to silicon [Bufler and Meinerzhagen, 1998].

In Section 2.5.4 the concept of velocity overshoot was discussed in that for short channel devices the high lateral fields can result in non-equilibrium transport such that it can become possible for the carrier velocity to temporarily exceed the saturation carrier velocity. Due to the high field experienced by carriers near the drain, velocity overshoot is most likely to occur in this region and several reports citing carrier velocities exceeding the saturation velocity have already been published. However, in the model proposed by Lundstrom, the effects of velocity overshoot are secondary to the backscattering coefficient at the source [Lundstrom, 1996]. Furthermore, Lundstrom argued that carriers cannot be injected into the channel region with a velocity higher than the Richardson thermal velocity. Consequently, there exists significant interest in measuring the carrier velocity in deep submicron devices, to firstly determine whether or not velocity overshoot effects are occurring and also their importance and secondly to assess how close modern devices are to the fundamental thermal limit and the impact on future transistor scaling.

The chapter begins with a brief review of the long channel behaviour of $Si_{1-x}Ge_x$ pMOSFETs ($x = 0$, 0.15) provided by CEA- LETI, France. The short channel characteristics of these devices, with written gate lengths down to 40 nm, will then be presented and their performance compared to the silicon control devices, with a particular emphasis placed upon the extraction of the *effective carrier velocity* near the source-end of the

channel, as proposed by Lochtefeld et al. [2002], to ascertain whether or not the effects of velocity overshoot predicted, for example, in the work by Palmer et al. [2001] and Kaya et al. [2000], can actually be observed in deep submicron SiGe devices. This work was carried out at both room temperature and at 77 K in order to benefit from higher effective hole mobilities.

The final section of the chapter will compare impact ionisation between silicon and $Si_{0.64}Ge_{0.36}$ channel devices. The impact ionisation rate is a significant parameter in high-frequency power amplifiers since it determines the breakdown voltage of the device, and hence the maximum power that a technology can give [Waldron et al., 2003]. Impact ionisation is also important in flash memory devices, in which hot carriers charge a floating gate. It has also recently been suggested that impact ionisation could be exploited as a means of reducing the subthreshold slope below the room temperature limit of 60 mV/dec and be used to overcome the problem of increasing off-state leakage currents [Gopalakrishnan et al., 2005a,b]. Whilst there has been considerable investigation on impact ionisation and breakdown in silicon MOSFETs, there have been relatively few reports on the impact ionisation characteristics of SiGe devices.

Strained SiGe has a smaller bandgap than both silicon and the unstrained alloy, which one would expect to lead to a higher impact ionisation rate. Monte Carlo simulations have shown that the impact ionisation rate in the unstrained alloy might be suppressed by the effects of alloy scattering [Yeom et al., 1996]. Meanwhile, Lee et al. [1995] found that the hole impact ionisation coefficient increased with increasing germanium composition from experiments on relaxed $Si_{1-x}Ge_x$/Si diodes. Hot carrier effects have been studied in vertical devices with strained SiGe in the drain region [Date and Plummer, 2001]. However, to the best of the author's knowledge, the impact ionisation rate for MOSFET's with a strained SiGe channel has not been investigated. Consequently, the final section of this chapter sets out to compare the impact ionisation rates in strained $Si_{1-x}G_{1-x}$ pMOSFETs.

Neither set of devices used to investigate velocity overshoot and impact ionisation in this chapter featured high-κ dielectrics or metal gates, since it was felt that the SiGe/high-κ dielectric interface has yet to be properly optimised, which would add further complications to what can be considered a more fundamental study. It might be such that, when the interface is improved, then this study could be repeated to see what, if any, effect the gate dielectric has upon the carrier velocity and impact ionisation of strained SiGe-channel MOSFETs.

7.2 Investigation of Hole Carrier Velocity in $Si_{1-x}Ge_x$ Channel pMOSFETs

7.2.1 Device Specification and Processing Details

The batch of devices used to study carrier velocity were processed in a standard LETI 50 nm CMOS process [Andrieu et al., 2003]. Figure 7.1 shows a schematic of the devices, including the layer structure. A 15 nm thick $Si_{0.85}Ge_{0.15}$ layer was grown on a 10 nm epitaxial Si buffer on (100) bulk Si substrate by low pressure chemical vapour deposition (LPCVD). The $Si_{0.85}Ge_{0.15}$ layer was capped with a thin 2 nm Si cap, which was optimised to provide a good SiO_2/Si interface without creating a parasitic hole channel at high effective fields. The devices had a 2 nm SiO_2 gate dielectric, that was grown at 700 °C, with a p^+ poly-crystalline silicon gate and nitride spaces. $Si_{1-x}Ge_x$ pMOSFETs were processed with written gate lengths down to 40 nm.

7.2.2 Mobility and Long-Channel Behaviour

Much of the room temperature electrical characterisation of this batch of devices has been performed by Andrieu et al. [2003] as part of a different investigation. That investigation included a comparison of effective hole mobility, interface trap density and

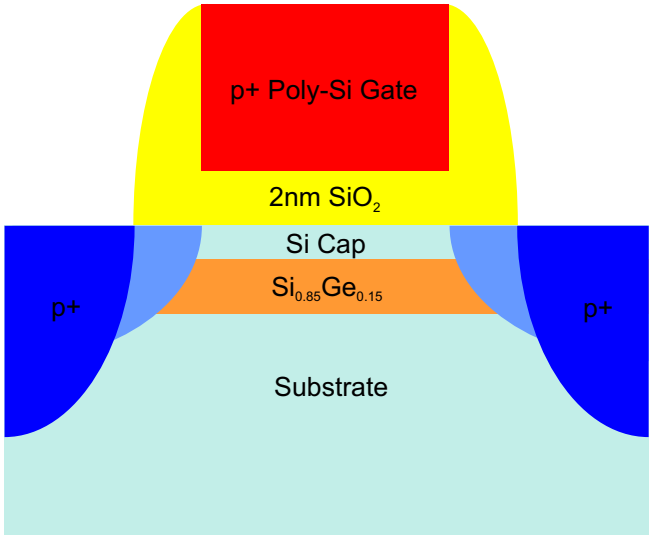


Figure 7.1: Device schematic of $Si_{0.85}Ge_{0.15}$ MOSFET used to study hole carrier velocity.

saturation drain current. However, due to its importance and relevance in this study, the mobility dependence on the electric field at room temperature was performed again, here.

The effective hole mobility is shown in Figure 7.2 as a function of vertical effective field for 10 μm x 10 μm SiGe and Si devices, measured at both room temperature and 77 K. The silicon control agrees well with the universal mobility curve at room temperature for vertical fields higher than 0.6 MVcm^{-1}, with a peak mobility of 75 cm^{2}V^{-1}s^{-1} at 0.5 MVcm^{-1}. At room temperature, the SiGe device maintains a mobility enhancement of approximately 50% over the silicon control at all fields up to 1 MVcm^{-1}, with a peak mobility of 110 cm^{2}V^{-1}s^{-1} at 0.5 MVcm^{-1}. The room temperature mobility agrees closely with that previously determined by Andrieu et al. [2003].

As expected, Figure 7.2 shows that the hole mobility at 77 K is significantly higher than at room temperature due to the reduction of phonon scattering for both the SiGe and the silicon control devices, with peak mobilities of 135 cm^{2}V^{-1}s^{-1} and 235 cm^{2}V^{-1}s^{-1} at 0.5 MVcm^2, respectively. A hole mobility enhancement of nearly two times that of silicon is maintained at all fields in the SiGe device at 77 K. For vertical

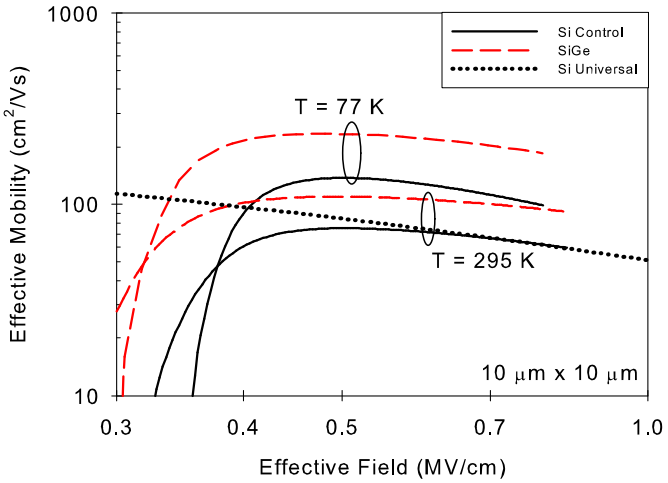
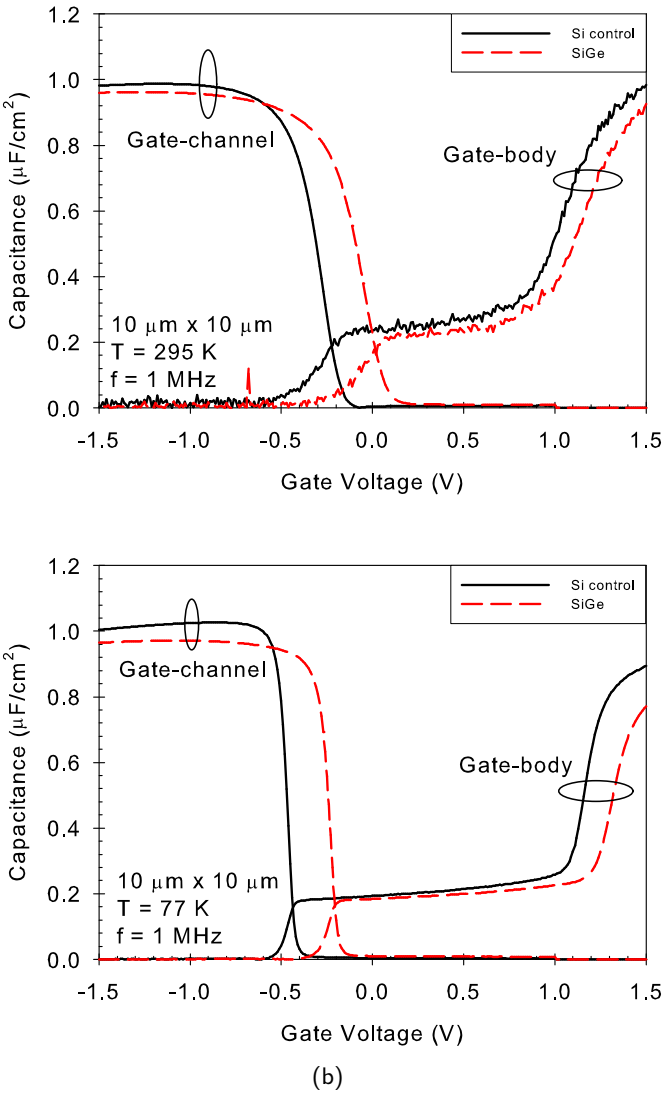


Figure 7.2: Comparison of hole effective mobility for 10 μm x 10 μm SiGe and Si control pMOSFETs at room temperature and 77 K. The universal curve for silicon is based on the work of Takagi et al. [1994].

fields greater than 0.5 MVcm^{-1}, the dependence of the hole mobility on the vertical field is stronger at 77 K, than at room temperature, as indicated by the steeper slope of the mobility curve, due to the mobility being increasingly limited by surface roughness as the temperature is lowered. Another point of note is that both devices exhibit a steeper mobility roll-off in the low-field regime, such that the hole mobility at 77 K is lower than that at room temperature, meaning that at low vertical fields, the effective mobility becomes increasingly limited by Coulomb scattering.

The split CV characteristics of 10 μm x 10 μm SiGe and Si devices used to determine the effective mobility, measured at room temperature and 77 K, are shown in Figures 7.3(a) and 7.3(b), respectively. The measurement frequency was taken as 1 MHz to minimise the effect of interface states contributing to the inversion charge. No frequency dispersion in accumulation due to any series resistance effects were observed for C-V measurements taken at frequencies above 1 kHz [Kwa et al., 2003]. Interestingly, both the SiGe and the silicon control, in particular, show a higher inversion capacitance at 77 K than at room temperature. It is speculated that this is due to a lower channel resistance. In addition, the depletion capacitance, given by the plateau in the gate-body



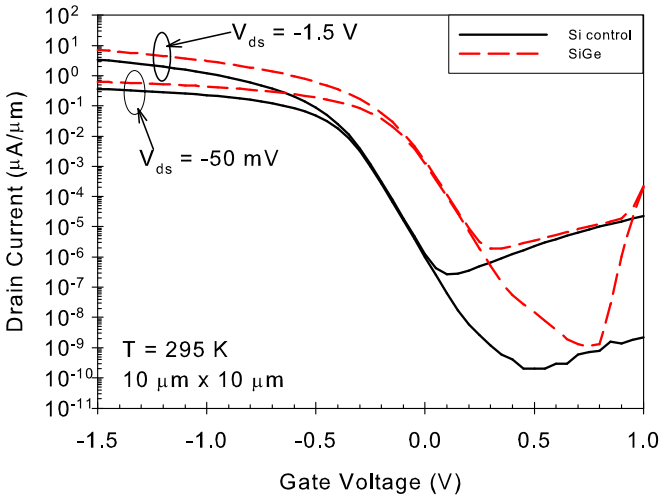
(b)

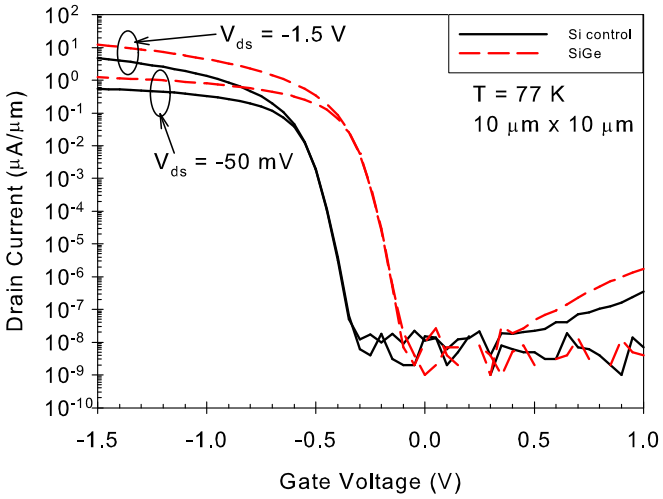
Figure 7.3: Split CV characteristics of 10 μm x 10 μm SiGe and Si pMOSFETs measured at (a) room temperature and (b) 77 K.

branch of the split C-V curve, is slightly reduced in both sets of devices, due to some freezing out of dopant atoms. The flatband voltage dependence is also clearly evident from Figure 7.3, with both SiGe and silicon devices showing a positive flatband voltage shift of approximately 0.2 V at 77 K. The positive flatband voltage shift with decreasing temperature can be attributed to the increased occupancy of acceptor-like states at the Si/SiO_2 interface, which could also be responsible for the increased Coulomb scattering at 77 K observed in Figure 7.2 [Gray and Brown, 1966].

The transfer characteristics of 10 μm x 10 μm SiGe and Si pMOSFETs with V_{ds} = -50 mV and -1.5 V, measured at (a) room temperature and (b) 77 K, are shown in Figure 7.4. At room temperature, both the SiGe and Si control 10 μm devices exhibit more than eight decades difference between I_{on} and I_{off} at low V_{ds} and between 6 and 7 decades at high V_{ds}. At 77 K, both devices exhibit around eight decades difference between I_{on} and I_{off} at low V_{ds}, comparable to that at room temperature, whereas the the difference between I_{on} and I_{off} at high V_{ds} is nine decades, a significant improvement compared to room temperature. This is due to a two decade reduction in the off-current. The sharp increase in the off-current for the SiGe device, measured at V_{ds} = -50 mV, at room temperature is due to gate leakage, possibly due to an early breakdown of the gate oxide. Both SiGe and Si control devices exhibit a large decrease in the subthreshold slope from 89 mV/decade and 83 mV/decade at room temperature to 33 mV/decade and 30 mV/decade at 77 K, respectively. The threshold voltage is also higher in both devices at 77 K due to a combination of the bulk Fermi-level lying closer to the conduction band edge and a larger bandgap. In addition, the occupancy of interface states could also be important.

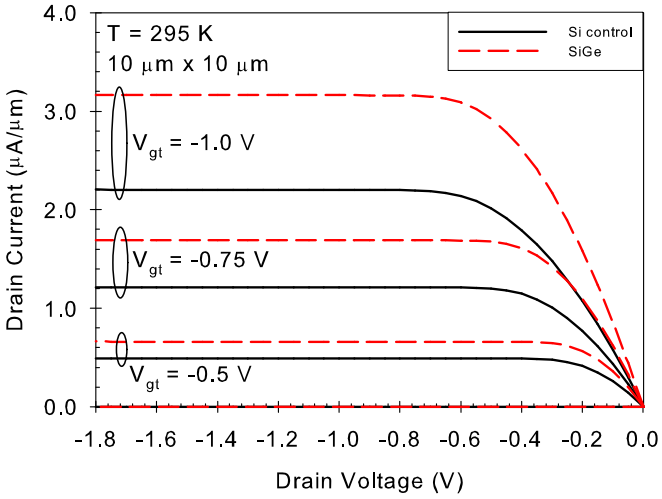
The output characteristics are shown in Figure 7.5 for both 10 μm x 10 μm SiGe and silicon control devices, measured at (a) room temperature and (b) 77 K. As expected, the SiGe device exhibits approximately 40% improvement in current drive over the silicon control at room temperature. The improvement in drain current is further

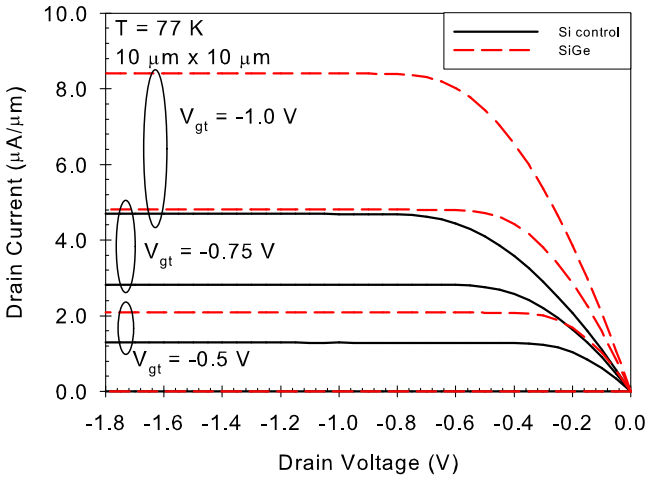




(b)

Figure 7.4: Transfer characteristics for 10 μm x 10 μm SiGe and Si pMOSFETs measured at (a) room temperature and (b) 77 K.





(b)

Figure 7.5: Output characteristics for 10 μm x 10 μm SiGe and Si pMOSFETs measured at (a) room temperature and (b) 77 K as a function of gate overdrive.

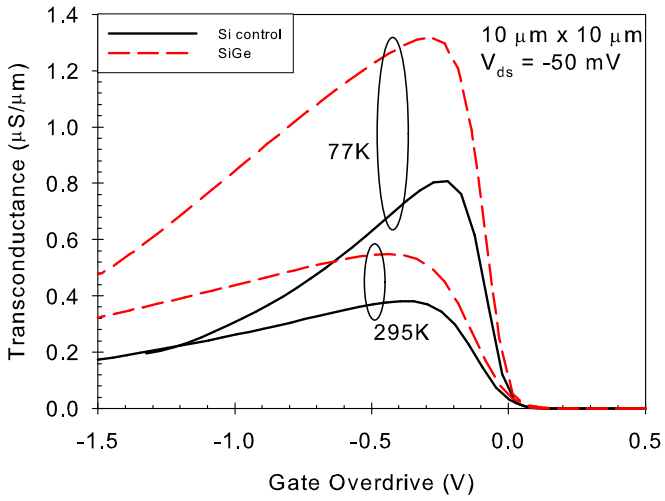
enhanced at 77 K, where it is found to be 80% higher than silicon at a gate overdrive of V_{gt} = -1 V.

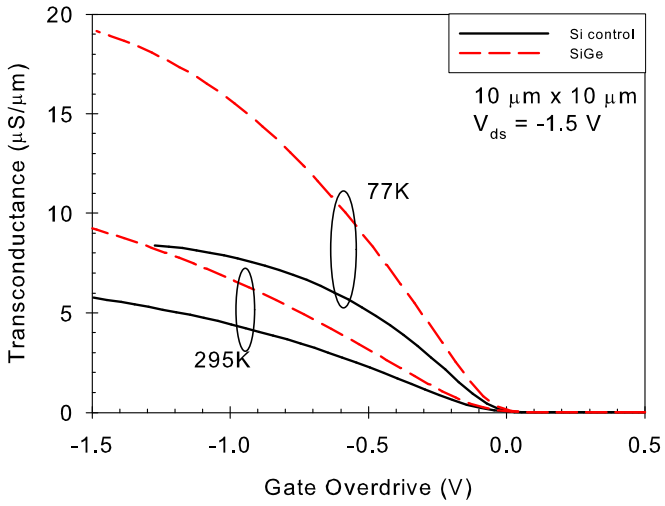
The transconductance at room temperature and 77 K for 10 μm x 10 μm SiGe and silicon control devices, is shown in Figure 7.6 for (a) V_{ds} = -50 mV and (b) -1.5 V. The SiGe devices exhibit more than 60% higher transconductance in the linear regime and more than two times higher saturation transconductance at both room temperature and 77 K, resulting from the higher hole mobility in SiGe.

Whilst long-channel silicon-germanium devices have frequently been observed to outperform their silicon counterparts, only a handful of studies have been performed on short-channel devices. For example, strained-Si CMOS devices with gate lengths below 50 nm are already in production [Tyagi et al., 2005], and hence it is vital that the performance enhancements already demonstrated in long-channel devices can be scaled to current, and future, technological dimensions. The next section sets out to investigate the short channel behaviour of these devices, and whilst there have been a handful of reports on deep submicron SiGe devices, to the best of the author's knowledge, this is one of the first studies on sub 100 nm gate-length strained SiGe channel pMOSFETs at cryogenic temperatures.

7.2.3 Short Channel Behaviour and Carrier Velocity

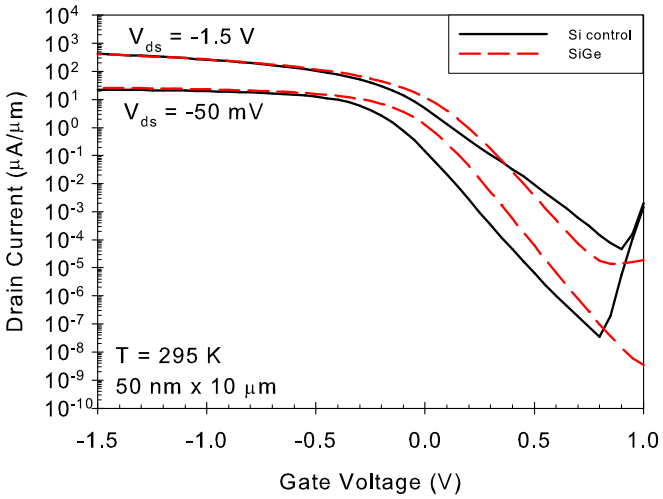
Figure 7.7 shows the transfer characteristics for 50 nm x 10 μm SiGe and silicon control devices measured at (a) room temperature and (b) 77 K, with V_{ds} = -50 mV and -1.5 V. At room temperature, both devices exhibit excellent I_{on}/I_{off} ratios of 8 decades and 6/7 decades in the linear and saturation regimes, respectively. The SiGe devices are less susceptible to short channel effects, exhibiting lower DIBL, lower threshold voltage roll-off and a smaller degradation of the subthreshold slope at higher drain bias, indicative of punch-through. This could be due to quantum confinement of holes in the SiGe channel devices. At 77 K, both devices exhibit around 10 decades difference

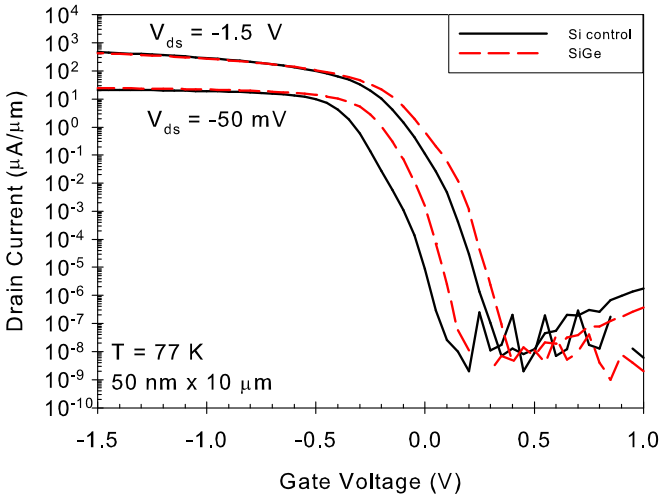




(b)

Figure 7.6: (a) Linear and (b) saturation transconductance for 10 μm x 10 μm SiGe and Si pMOSFETs at room temperature and 77 K





(b)

Figure 7.7: Transfer characteristics for 50 nm x 10 μm SiGe and Si pMOSFETs measured at (a) room temperature and (b) 77 K.

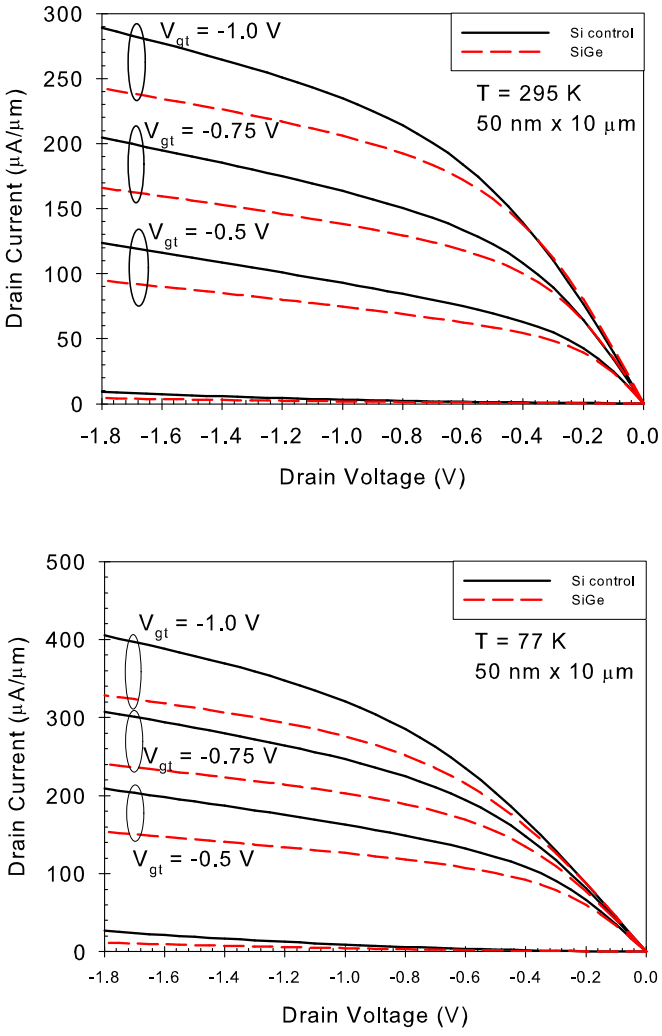
between I_{on} and I_{off} at high drain bias. Interestingly, the short channel effects of the silicon devices are suppressed at 77 K, with a reduction in both DIBL from 189 mV/V to 155 mV/V, and negligible degradation of the subthreshold slope with increasing drain bias.

The saturation drain current is shown in Figure 7.8 as a function of gate overdrive for 50 nm x 10 μm SiGe and Si control devices, measured at (a) room temperature and (b) 77 K. Interestingly, at both room temperature and 77 K, the silicon control exhibits higher drain current than SiGe for all gate overdrive values. This indicates that the long-channel performance benefits resulting from a high-mobility strained SiGe channel may not be realised in the deep submicron regime, which would be of particular concern if strained SiGe channels were to be considered as a way of improving performance in sub-100 nm pMOS devices.

This concern is emphasised again in Figure 7.9, which shows both (a) the linear and (b) the saturation transconductance of SiGe and silicon control devices, with a 50 nm gate length, at room temperature and 77 K. Accordingly, both silicon and SiGe exhibit higher transconductance at 77 K, particularly in the linear regime. Yet, the linear transconductance is comparable between the two devices, but the saturation transconductance is actually worse in the SiGe 50 nm device.

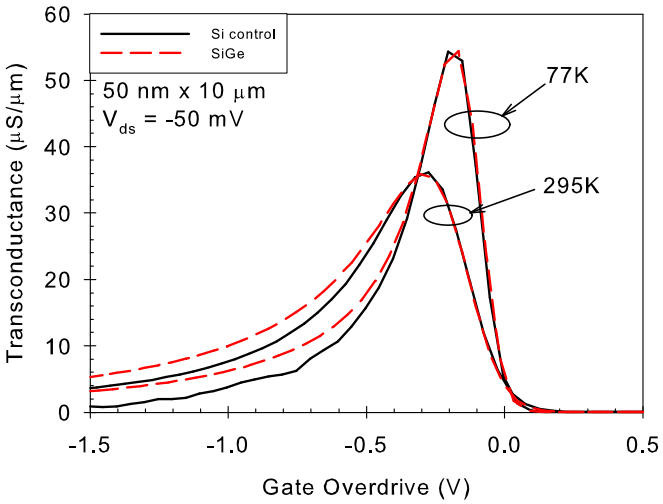
The on-current, I_{on} for these devices was measured at V_{ds} = -1.5 V and $V_{gs} - V_t$ = -1.0 V. A gate overdrive of -1.0 V was preferred to -1.5 V to avoid any unnecessary stressing of the oxide. The off-current was taken at the minimum of the $I_{ds} - V_{gs}$ curve as opposed to at V_{gs} = 0 V to reduce the effects of GIDL and non-optimised threshold voltage, particularly for the SiGe devices. Figure 7.10 shows the I_{on} values for both SiGe and Si control devices, measured at both room temperature and 77 K as a function of the off-current.

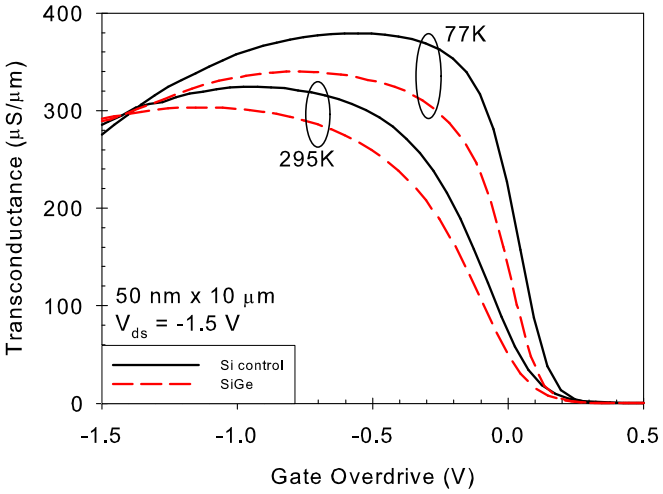
In order to further examine the short-channel behaviour of these devices, the intrinsic carrier velocity was extracted and compared at room temperature and 77 K



(b)

Figure 7.8: Output characteristics for 50 nm x 10 μm SiGe and Si pMOSFETs measured at (a) room temperature and (b) 77 K.





(b)

Figure 7.9: (a) Linear and (b) saturation transconductance for 50 nm x 10 μm SiGe and Si pMOSFETs measured at room temperature and 77 K.

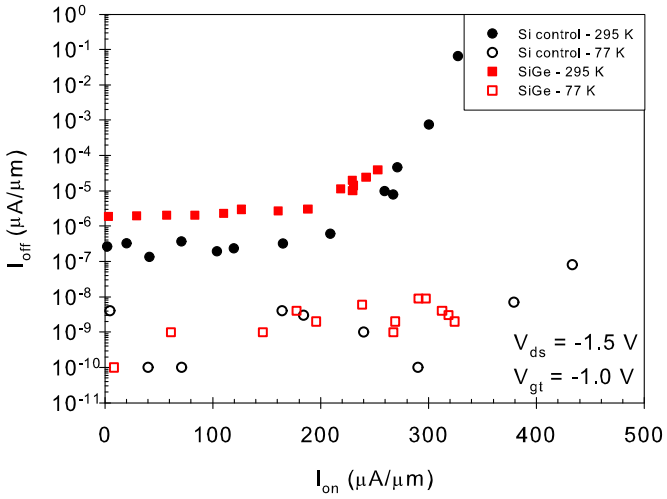
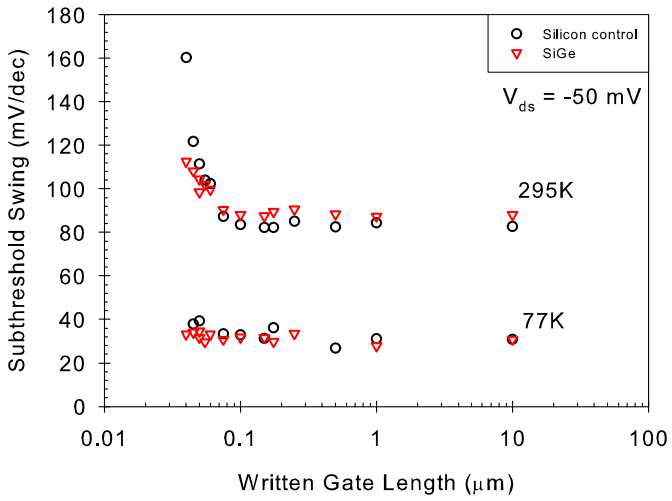
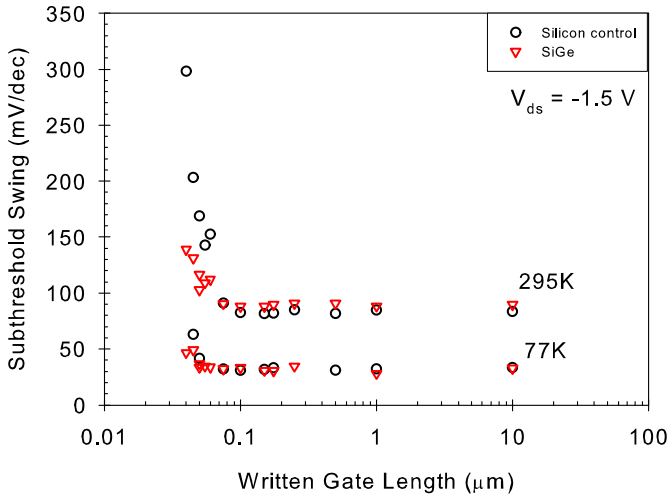


Figure 7.10: I_{on} versus I_{off} for Si and SiGe pMOSFETs at room temperature and 77 K.

as a function of gate length. As a result of the lower carrier saturation velocity, the effects of velocity overshoot could prove decisive if SiGe is to be used as an alternative channel material to silicon in deep submicron pMOSFETs. Before this extraction was conducted, it was first necessary to confirm the electrostatic integrity of the devices, and in particular to assess whether short-channel effects were dominating either set of devices.

The subthreshold slope is shown in Figure 7.11 as a function of gate length, and was measured in both the linear and saturation regime at room temperature and 77 K. At room temperature, both the silicon control and SiGe devices have similar subthreshold slopes for written gate lengths down to 75 nm. Below 75 nm, the subthreshold slope for the silicon control devices is typically higher than those found in the SiGe devices, particularly for high drain bias. This degradation of the silicon control subthreshold slope with decreasing gate length indicates that the silicon devices are more adversely affected by punchthrough in the shortest-channel devices than SiGe. As expected, a stark reduction in the subthreshold slope is observed in both silicon and SiGe devices at 77 K, with typical values of 30 mV/dec realised in the saturation regime for gate lengths down to 75 nm. Furthermore, the short channel effects appear to be suppressed





(b)

Figure 7.11: Subthreshold slope measured in the (a) linear and (b) saturation regime for SiGe and Si control pMOSFETs as a function of written gate length at room temperature and 77 K.

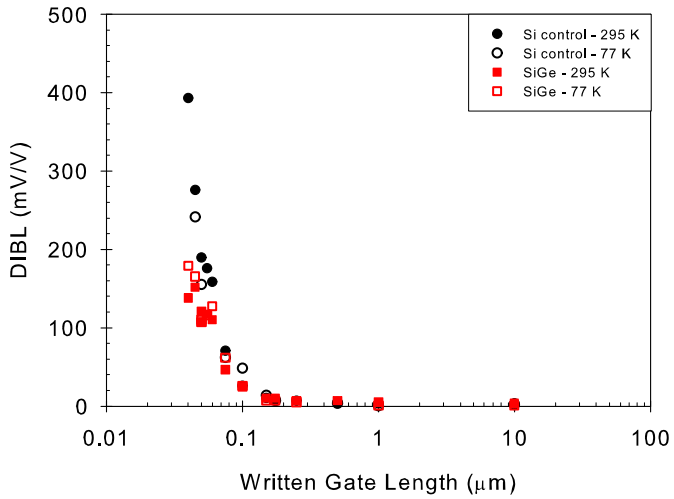


Figure 7.12: DIBL for SiGe and Si pMOSFETs at room temperature and 77 K as a function of written gate length.

in the silicon devices at 77 K, since comparable values for the subthreshold slope at V_{ds} = -1.5 V are obtained for gate lengths down to 50 nm.

In addition to the subthreshold slope, it is imperative to consider the effects of DIBL upon the experiment. Importantly, the silicon control and SiGe devices should have comparable DIBL values for all gate lengths to be compared. This ensures that the effects of DIBL do not contribute to the drain current and hence result in erroneous values of the carrier velocity. Figure 7.12 shows the DIBL values for both SiGe and silicon control devices as a function of gate length at room temperature and 77 K. Comparable DIBL values are found for gate lengths down to 100 nm and 75 nm at room temperature and 77 K, respectively. Below these gate lengths, the DIBL increases significantly more for the silicon devices than for SiGe devices, again confirming that the silicon devices are more susceptible to suffering from short channel effects. The DIBL values should be as low as possible, and whilst DIBL values between 100 mV/V and 200 mV/V might be considered a little high, they are nonetheless tolerable.

In order to determine the effective carrier velocity at the source, v_{eff}, according to the method proposed by Lochtefeld et al. [2002], it is also necessary to determine the

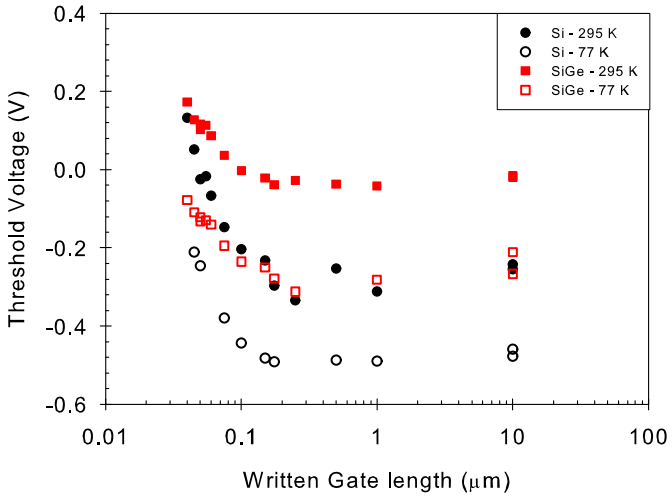


Figure 7.13: Threshold voltage for SiGe and Si pMOSFETs at room temperature and 77 K as a function of written gate length.

threshold voltage roll-off, the device on-current and the source-drain series resistance, such that the inversion charge in a short device can be accurately found from a split CV measurement on a long-channel device.

The threshold voltages for both silicon and SiGe devices, extracted at room temperature and 77 K, are shown in Figure 7.13 as a function of written gate length. Accordingly, both sets of devices show a more negative (i.e. higher) threshold voltage at 77 K, due to the increase in bandgap, the bulk Fermi level lying closer to the conduction band edge and possibly changes in occupancy of interface states. The threshold voltage roll-off for the silicon control devices is worse than that for the SiGe devices ($\sim$0.3 V compared to $\sim$0.1 V), again indicating that the silicon control devices are more susceptible to short channel effects.

The source-drain series resistance, R_{sd}, and the deviation between the written and effective channel length, ΔL are summarised in Table 7.1. The values were determined using the second regression technique described in Section 3.3.4.2. At room temperature, the values for R_{sd} were found to be 1300 $\Omega\mu$m and 1250 $\Omega\mu$m for silicon and SiGe, respectively, and are in agreement with those reported by Romanjek et al.

[2004] on similar devices. The extraction indicates that the series resistance in both the silicon control and SiGe channel devices are very comparable. This might be due to the higher solid solubility of boron in SiGe compared to silicon [Salm et al., 1997]. However, as described in Section 3.3.4.2, it is possible for uncertainties to arise when using the regression techniques and comparisons between the extracted values (for both R_{sd} and ΔL should be made with care.

Whilst the room temperature values for R_{sd} seem reasonable, the values obtained for ΔL, in particular for SiGe, appear to be somewhat erroneous. The reason for this probably lies in the assumptions of the extraction procedure, in which it is implicitly assumes that the change in the effective channel length is a constant function of gate length. In SiGe devices, this is known to not necessarily be true [Ghibaudo, 2006]. Andrieu et al. [2003] extracted the effective channel length using split CV measurements and found the effective channel length in Si MOSFETs to be approximately 10 nm longer than the written gate length, which is comparable to the ΔL value of 11.2 nm in this work. Furthermore, Andrieu et al. [2003] found that the effective channel length for both the silicon control and SiGe devices was comparable in this particular batch, justifying the comparisons made between the Si and SiGe devices as a function of written gate length in the present work.

At 77 K, both sets of devices show an increase in the source-drain series resistance, to around 1800 $\Omega\mu$m and 1500 $\Omega\mu$m for silicon and SiGe, respectively. It is speculated that the increase in series resistance could be due to dopant freeze-out, particularly in the lower-doped extension regions, giving rise to the observed increase in the series resistance at 77 K. Both sets of devices show an increase in the effective channel length at 77 K, although the value obtained for the SiGe devices is not realistic and should be ignored. It is possible that the slight increase in the effective channel length of the silicon devices can also be related to freezing out of dopants in the extension regions.

Table 7.1: Summary of series resistance and effective channel length for silicon and SiGe pMOSFETs at room temperature and 77 K.

	Silicon		SiGe	
	295 K	77 K	295 K	77 K
R_{sd} ($\Omega\mu$m)	1300	1800	1250	1500
ΔL (nm)	-11.2	-11.9	-29.6	-62.4

The effective carrier velocity near the source, v_{eff}, is shown in Figure 7.14, as a function of written gate length, for both SiGe and silicon control devices at room temperature and 77 K. The carrier velocity was evaluated at V_{ds} = -1.5 V and $V_{gs} - V_t$ = -1.0 V. Whereas the long channel SiGe devices show an enhancement of approximately 50% in the hole carrier velocity over silicon, this enhancement is not realised for the shortest devices. For devices with gate lengths below 100 nm, the hole carrier velocity in the siicon devices exceeds those in the SiGe devices.

Figure 7.14(b) shows the ratio of hole carrier velocity for these devices. From this graph, it is clear that the enhancement in carrier velocity, and ultimately in drive current, diminishes as the gate length is reduced below around 100 nm. Figure 7.14(b) also indicates that the extracted hole carrier velocities are below their saturation values, thus indicating that velocity overshoot effects may not be occurring. The extracted hole velocities are lower than those reported by Kar et al. [2002], who reported a hole velocity of 1.1 x 10^7 cms^{-1} in $Si_{0.793}Ge_{0.2}C_{0.007}$ pMOSFETs, which they attributed to the onset of velocity overshoot. However, no account of short channel effects, and in particular, DIBL was reported. This work also contradicts the theoretical predictions by Palmer et al. [2001]. In the work by Palmer et al. [2001], it was predicted that whilst the ratio of the saturation transconductance of SiGe to Si would diminish as the gate length is reduced, the onset of velocity overshoot in the SiGe devices, indicated by a large increase in the SiGe saturation velocity, will lead to the 'recovery' of the performance

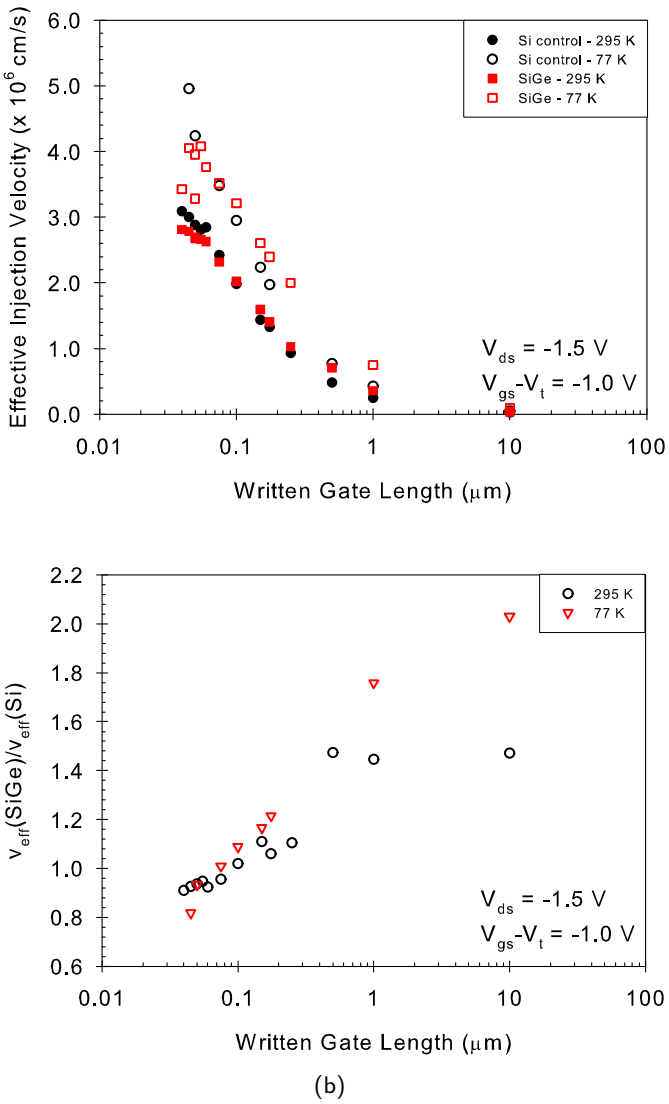


Figure 7.14: (a) Hole carrier velocity (after Lochtefeld et al. [2002]) and (b) ratio of the hole carrier velocity for SiGe and silicon control pMOSFETs measured at room temperature and 77 K.

enhancement in the SiGe devices over silicon, and result in saturation transconductance ratio increasing again. An enhancement of between 1.4 and 1.9 times over silicon was predicted for a gate length of 70 nm, a result not observed in the current work.

Alternatively, in terms of the velocity overshoot model, it may be that velocity overshoot *does* occur in deep sub-micron SiGe channel devices, but is not observed in the present case. One possibility is that of local Coulomb scattering near the source and drain regions of the device, that results from damage caused to the crystal structure by ion implantation, being higher in SiGe than in silicon. The damage caused by ion implantation could leave a high density of charge at the source-channel interface, which could lead to a higher backscattering coefficient, and also induce some relaxation of the strained SiGe layer. In long channel devices, this effect is not so important, since the channel length is much longer than the region over which ion implantation damage is likely to occur, and the higher mobility of the strained SiGe channel provides the performance enhancements that are routinely reported. As the channel length is decreased, the relative size of the damaged area becomes more comparable to the channel length, and thus the effect becomes increasingly important, dominating the performance of the shortest SiGe devices. If this hypothesis is correct, then it will have severe implications for SiGe channel devices as a viable technological replacement for silicon. However, the hypothesis is very difficult to confirm experimentally, and would probably require elaborate simulation work, or the fabrication of Schottky barrier SiGe devices, both of which lie beyond the scope of this particular investigation. This result should not be used to completely rule out the possibility of SiGe channel devices as a future CMOS technological replacement for silicon, and further work is essential in understanding the behaviour of these very short devices. A second possibility is that being that velocity overshoot does lead to an enhancement in drive current, but for devices with gate lengths shorter than 50 nm or with germanium concentrations higher than 15%, which were not available for this investigation.

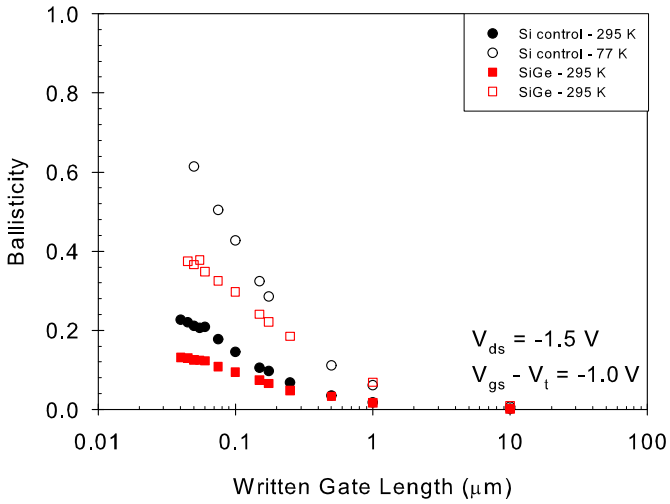


Figure 7.15: Ballisticity as a function of written gate length for SiGe and Si pMOSFETs at room temperature and 77 K.

Comparing the effective velocity with the theoretical value for the thermal injection velocity, it is possible to calculate the ballisticity (v_{eff}/v_T) of the short-channel $Si_{1-x}Ge_x$ pMOSFETs (see Section 2.5.4). The ballisticity values are shown as a function of written gate length in Figure 7.15. The fact that the SiGe devices appear to be operating at a lower ballisticity is not surprising given that the effective velocities for written gate lengths of $\sim$ 100 nm are comparable, and the thermal injection velocity is higher than silicon due to the lower effective hole mass. The lower ballisticity values for the SiGe pMOSFETs indicate that the backscattering coefficient at the source is higher in these devices, which would be consistent with the suggestion of ion implantation damage mentioned previously.

Table 7.2: Growth specification of $Si_{0.64}Ge_{0.36}$ wafers to investigate impact ionisation. After Palmer [2001]

Wafer ID	Substrate Type	Ge Composition (%)	SiGe Thickness (nm)	Si Cap Thickness (nm)	Oxide Thickness (nm)
2	n^+	36	10	5	10
6	n^+	0	-	-	10

7.3 Impact Ionisation Investigation

7.3.1 Device Specification and Processing

Impact ionisation was studied in a batch of pseudomorphic long-channel $Si_{1-x}Ge_x$ pMOSFETs. The specification and fabrication details of these devices have been extensively reported elsewhere and so the only the main details are summarised here. For a complete description of the growth and processing conditions, reference is made to Palmer [2001].

The layers were grown by solid source molecular beam epitaxy (SS-MBE) at the University of Warwick and the devices were fabricated at Southampton University. The devices featured a 10 nm-thick compressively strained $Si_{0.64}Ge_{0.36}$ channel with a 5 nm Si cap following processing, a 10 nm gate oxide grown by thermal oxidation and a polycrystalline silicon gate. The source/drain implants were made by ion implantation of 5 x 10^{15} cm^{-2} BF_2 at 50 keV and were activated by annealing at 850 °C for 60 seconds.

The devices used in this investigation were long channel devices with written gate lengths of 10 μm to minimise complications arising from short channel effects.

7.3.2 Impact Ionisation

Impact ionisation was evaluated from the relation [Taur and Ning, 1998]:

$$M - 1 = \frac{I_b}{I_s} \tag{7.1}$$

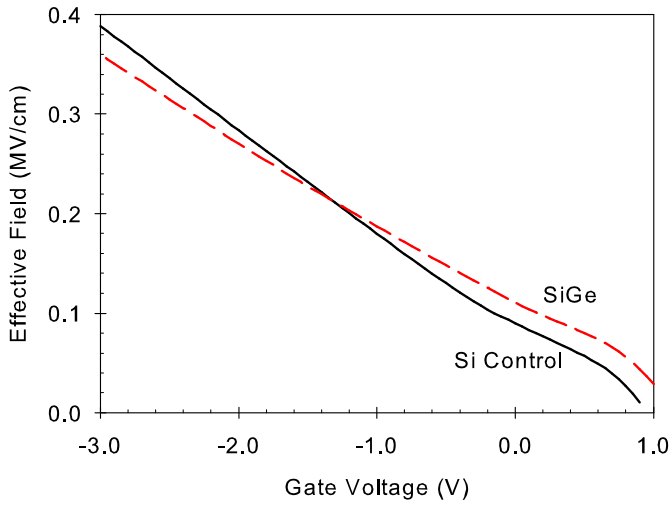


Figure 7.16: Vertical effective field extracted from split C-V measurements against gate voltage for Si control and strained SiGe pMOSFETs.

where I_b and I_s, are the body and source currents, respectively. During impact ionisation measurements, the SiGe and Si control pMOSFETs were evaluated at the same gate voltage, rather than the same gate overdrive, despite a significant difference in threshold voltage of approximately 0.3 V. This is because a change in gate voltage affects the shape of the energy bands at the semiconductor surface, which consequently affects the electric field strength at the drain [Nicholas et al., 2005b]. Because impact ionisation is evaluated from the ratio of the body current to the source current, the comparison is still valid despite the unavoidable difference in carrier sheet density. The use of different gate overdrives would be expected to lead to a difference in effective field. This would be expected to affect the mobility and thus render any comparison invalid. To overcome this complication, the effective field, extracted from split C-V measurements, is plotted as a function of gate voltage in Figure 7.16 for both SiGe and Si control devices. A gate voltage of V_{gs} = -1.5 V was selected for impact ionisation measurements, such that the two devices were subjected to almost identical electric fields and, importantly, the gate voltage was small enough such that there was no significant population of the thin silicon cap in the SiGe device.

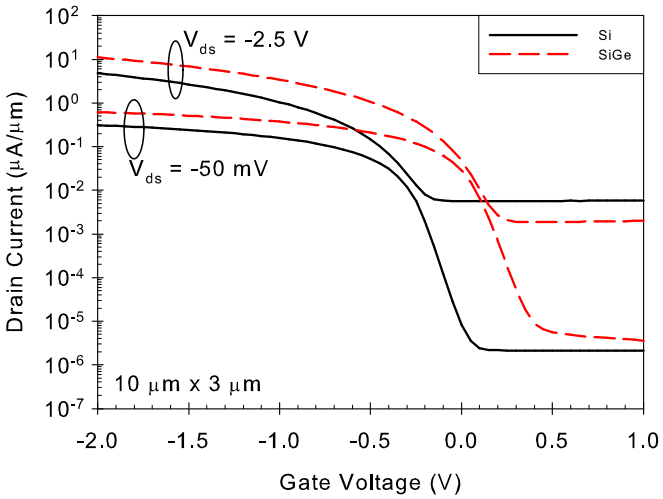


Figure 7.17: Comparison of the linear and saturation transfer characteristics of 10 μm x 3 μm Si and SiGe pMOSFETs.

Figure 7.17 shows the linear and saturation transfer characteristics for the 10 μm x 3 μm silicon and SiGe pMOSFETs. Both devices exhibited over five decades of variation between the on and off state in the linear regime. In the saturation regime, the devices only exhibit around three decades between the on and off state, which is mainly due to the increased off-current observed for both devices.

The impact ionisation coefficient, $M-1$, is plotted against the intrinsic voltage, V_{int}, in Figure 7.18 for 10 μm x 3 μm silicon and SiGe pMOSFETs. Here, V_{int} has been calculated from the relation:

$$V_{int} = V_{ds} - I_s R_{sd} \tag{7.2}$$

where V_{ds} is the drain voltage with respect to the source and R_{sd} is the parasitic resistance of the source and drain. The values of R_{sd} were obtained from Palmer [2001] from the double regression method as being 5000 $\Omega\mu$m and 4720 $\Omega\mu$m for Si and SiGe, respectively. Figure 7.18 indicates that the impact ionisation for holes may be suppressed in SiGe devices compared to silicon.

The low field hole mobility is shown as a function of effective field for SiGe and

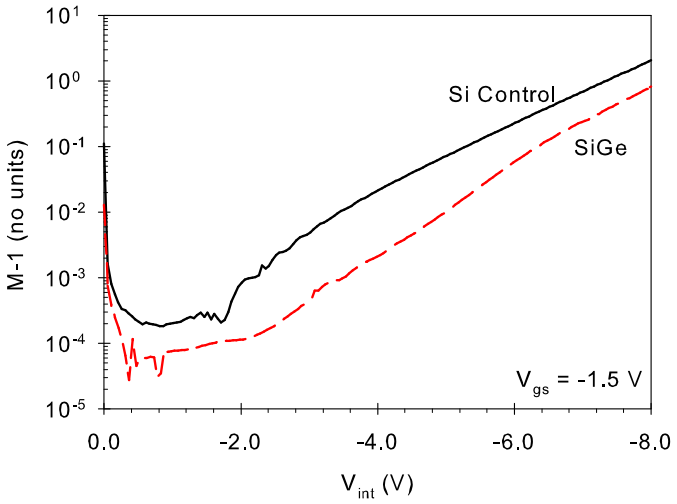


Figure 7.18: Impact ionisation multiplication coefficient against intrinsic voltage for Si control and strained SiGe pMOSFETs.

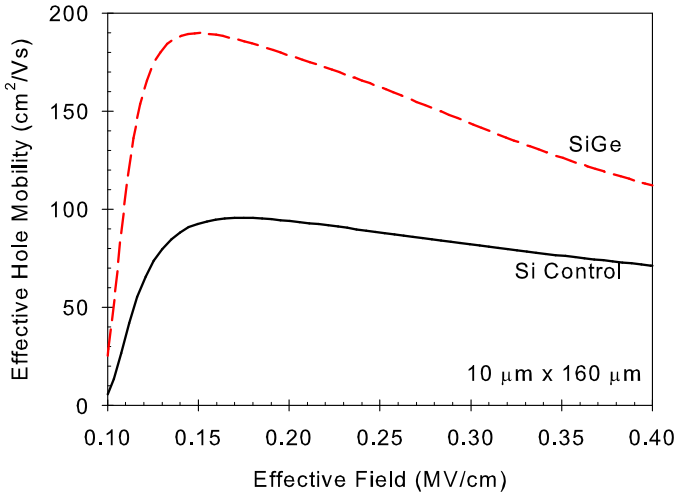


Figure 7.19: Low field hole mobility against effective vertical field for Si and strained SiGe pMOSFETs.

Si control devices in Figure 7.19. Impact ionisation rates were measured at a vertical field of approximately 0.2 MVcm^{-1}. This corresponds to a 90% enhancement in hole mobility for the SiGe channel device compared to the Si control device. It is therefore somewhat surprising that, despite the increased mobility and smaller bandgap, the SiGe devices exhibit reduced impact ionisation. For compressively strained $Si_{0.64}Ge_{0.36}$, the density of states in the valence band is thought to be almost six times lower than in bulk Si, whilst the density of states in the conduction band is also considerably reduced [Yang et al., 2004]. This lack of availability of free states is likely to severely impede impact ionisation, which involves the creation of a relatively low energy electron and hole pair. This is a similar result to the findings of Richard et al. [2005], in which a reduced density of states was cited for the reduction in impact ionisation rate in strained Ge compared to bulk Ge.

It is also possible that, while the low field hole mobility is greatly increased, the reduced saturation velocity in strained SiGe should mean less impact ionisation. Another effect that must be considered is that due to the reduced diffusion of boron in SiGe [Rajendran and Scoenmaker, 2001]. One would expect a more sharply defined p^+-n junction in the strained SiGe devices, which would lead to an increased peak electric field near the drain, which could enhance the impact ionisation rate in the strained SiGe devices compared to the Si control. However, since this was not observed, this effect does not undermine the validity of the result. Finally, consideration must be given to the conduction band offset between silicon and strained SiGe. For most practical purposes, the conduction band offset is considered to be negligibly small. However, it is an important consideration, since any confinement of the electrons created after an impact ionisation event could reduce the measured body current and hence impact ionisation. There exists some controversy regarding whether the SiGe/Si band alignment is type I or type II [Baier et al., 1994, Houghton et al., 1995]. However, photoluminescence results now appear to confirm a type II band alignment in SiGe/Si quantum wells such

that the SiGe conduction band edge lies above that of silicon [Thewalt et al., 1997]. Therefore, electrons created in the impact ionisation event will not be confined to the SiGe layer, thus, not suppressing the measured impact ionisation.

7.4 Conclusion

The role of hole carrier velocity in compressively strained buried channel pseudomorphic SiGe pMOSFETs has been investigated by electrical characterisation at both room temperature and 77 K. The SiGe pMOSFETs were found to exhibit both hole mobility and velocity enhancements over silicon in long-channel length transistors. However, these performance enhancements were found to diminish at both room temperature and 77 K as the gate length decreased, and for gate lengths below 100 nm, higher hole velocities were exhibited in the silicon control pMOSFETs, although the silicon control transistors were found to be more susceptible to short channel effects, including slightly higher DIBL, for gate lengths below 75 nm. The reasons for the 'loss' in performance enhancement for the high-mobility SiGe channel transistors is not well understood. It is speculated that the epitaxial SiGe layer suffers higher ion implantation damage in the vicinity of the source/drain regions, which could induce a high density of defects at the source/channel interface, and maybe result in a higher backscattering coefficient of hole carriers. In addition, it is speculated that if the strain in the SiGe layer is not successfully maintained in these short channel transistors, then this could also reduce the performance enhancements.

Impact ionisation was studied in a batch of long-channel SiGe pMOSFETs with a germanium composition of 36%. The strained SiGe channel pMOSFETs were found to exhibit reduced impact ionisation compared to the silicon control devices, despite having higher hole mobility and smaller bandgap. This was attributed to a strain-induced reduction of the density of states in both the SiGe conduction and valence bands,

suppressing impact ionisation events. The validity of the impact ionisation result could still depend upon the importance of the conduction band offset between Si and SiGe. In particular, if the band alignment is type I, electrons created in an impact ionisation event could be confined in a quantum well, lowering the measured substrate current and the observed impact ionisation. It might be an idea to model the experimental data using the conduction band offset as a parameter to confirm the experimental findings.

Chapter 8

Conclusion

8.1 Summary

A batch of long-channel strained SiGe pMOSFETs incorporating high-κ dielectric and metal gate has been developed and characterised within the framework of the European SINANO Network of Excellence project under the guidance of the author. The transistor batch featured hafnium oxide gate dielectric and employed a tungsten metal gate, since this combination was found to tolerate the high thermal budgets required for transistor processing as demonstrated through a series of preliminary short-loop capacitor experiments. Whilst the presence of a pseudomorphic SiGe channel was found to yield a hole mobility enhancement over the silicon control, the effective hole mobilities on all wafers were largely disappointing. The effective hole mobilities were found to be degraded with respect to the silicon universal mobility curve for holes and other literature values. This was believed to be due to a combination of high interface trap densities (typically in the mid-high 10^{12} $cm^{-2}eV^{-1}$ range), resulting in increased Coulomb scattering, and possible sputter-induced surface roughening during the early stages of the hafnium oxide deposition. The electrical characterisation of the batch was hampered by poor uniformity in the transistor characteristics across all wafers.

The reason for the poor uniformity was believed to be due to high contact and series resistance due to the incomplete removal of hafnium oxide in the source and drain regions by sputter etching prior to contact hole formation.

The electrical properties of a batch of Ge pMOSFETs with written gate lengths down to 125 nm have been studied through extensive characterisation over a range of temperatures between room temperature and 4 K. The Ge pMOSFETs were fabricated in a Si-like process as part of the Ge MOSFET programme at IMEC, Belgium. The halo-doped transistors were found to exhibit a room temperature hole mobility enhancement of 2.5 times over the silicon universal mobility curves for holes – an enhancement maintained at high vertical effective fields – and are comparable to some of the highest hole mobility values reported in the literature for Ge pMOSFETs to date. The effective hole mobility reached a maximum value of 480 cm^2/Vs when measured at 77 K, an enhancement due almost entirely to the reduction of phonon scattering. However, no further significant improvements in hole mobility were observed when measured at 4 K, indicating that below 77 K, the effective mobility in these halo-doped Ge pMOSFETs is limited entirely by Coulomb scattering and surface roughness scattering at low and high vertical fields, respectively. The use of halo implants were found to successfully suppress short-channel effects and excellent drive currents were obtained for transistors with a written gate length of 125 nm, easily surpassing the performance targets specified by the ITRS for silicon pMOSFETs for a comparable technology node. The off-state leakage current in Ge pMOSFETs remains a concern. The off-state leakage in this particular batch of Ge pMOSFETs was found to differ by several orders of magnitude between that measured at the source and the drain. Indications suggest that band-to-band tunnelling mechanisms, including GIDL, are important, although defect-assisted leakage mechanisms might also play a role.

The role of hole carrier velocity was investigated by electrical characterisation at both room temperature and 77 K in a batch of buried channel pseudomorphic SiGe

pMOSFETs provided by CEA-LETI, France. The SiGe pMOSFETs exhibited hole mobility and carrier velocity enhancements over silicon for long-channel transistors. However, whilst the SiGe pMOSFETs were found to be less susceptible to short-channel effects such as DIBL and threshold voltage roll-off, measurements of the hole carrier velocity indicated that this enhancement is 'lost' in aggressively-scaled transistors, with the silicon control transistors exhibiting higher hole carrier velocities in transistors with gate lengths less than 100 nm. It was speculated that higher implantation damage or some strain relaxation in the SiGe devices could be responsible although this remains to be fully understood. This result was found to be in contradiction to simulation results for deep sub-micron SiGe pMOSFETs, where it was predicted that velocity overshoot effects would lead to a performance enhancement in SiGe transistors over silicon in the deep sub-micron regime.

Impact ionisation was studied in a batch of buried channel pseudomorphic SiGe pMOSFETs fabricated as part of the UK-based HMOS project. It was found that, despite an increased hole mobility and smaller bandgap, the pseudomorphic SiGe channel pMOSFETs exhibited reduced impact ionisation compared to the silicon control. This has been attributed to a strain-induced reduction in the density of states in both the conduction and valence bands of SiGe, impeding the probability of an impact ionisation event occurring.

8.2 Discussion and Suggestions for Further Work

Whilst significant mobility enhancements have been demonstrated in long-channel high-κ pseudomorphic SiGe pMOSFETs over the silicon control, the effective mobility values reported in this work were found to be degraded with respect to the silicon universal mobility curve for holes and also lower to values already reported in the literature [Weber et al., 2006, Wu et al., 2005, Jin et al., 2004], although these all employed HfO_2

deposited by ALD. Further optimisation of the transistor fabrication process is evidently required before an in-depth study on hole carrier transport in these structures can be made. In particular, further understanding of the initial stages of the hafnium oxide deposition sputter process is critical and *in-situ* surface analysis techniques could prove useful, with particular emphasis on understanding surface roughening and the formation of an interfacial layer between the silicon surface and the hafnium oxide dielectric, to which the high capacitive equivalent thickness (CET) values were attributed. Whilst the high CET values can be tolerated for long-channel mobility studies, extension of this work to shorter channel lengths will require the CET values to be reduced. Ultimately, it might be such that alternative hafnium oxide deposition techniques have to be explored. These could be alternative to sputter-based techniques such as oxidation of a thin Hf film deposited by evaporation [Chang, 2005], or CVD-based techniques, which could also require the development of a surface passivation process. As a final note, the pseudomorphic SiGe pMOSFETs were fabricated as part of a joint collaboration involving four different institutions for growth and transistor processing. It still remains to be seen whether such a collaborative network is conducive to producing high quality transistor batches or whether a single processing facility is required.

Germanium appears to offer considerable advantages over silicon for pMOS applications. Hole mobility enhancements of 2.5 times over the silicon universal mobility curve for holes were maintained for high vertical fields of 1 MV/cm, at which a transistor for this particular technology would be operated. An obvious extension of this work could be to investigate the effective mobility using alternative high mobility crystal orientations, such as (110) or (111) and also strained Ge channels.

It is still unknown as to whether increased band-to-band tunnelling due to the reduced bandgap will prove to be a fundamental obstacle in Ge MOSFETs, and further work in this area is clearly warranted. In addition, the Ge pMOSFETs were fabricated on Ge-on-silicon substrates, where the threading dislocation density of the Ge layers is

expected to be $\sim 10^7$ - 10^8 cm^{-2}, and it is not known how detrimental this could prove to be to the performance of the Ge MOSFETs. The effects of GIDL were found to be important in this particular batch of transistors due to the non-optimised threshold voltage requiring a positive gate voltage to turn the transistor off. Ultimately, germanium will probably have to be incorporated using thin layers in an on-insulator technology due to its high cost and low natural abundance, which might assist in reducing leakage currents. Recent work has also demonstrated that low off-currents can also be achieved using heterostructure device designs with thin-layer Ge channels [Saraswat et al., 2006] and Schottky barrier source and drain regions [Li et al., 2006].

Experimental determination of the effective hole carrier velocity in deep sub-micron SiGe pMOSFETs was found to be in contradiction to previous simulation studies, which predicted that velocity overshoot effects in SiGe would result in a performance enhancement of SiGe pMOSFETs over silicon for gate lengths approaching 100 nm [Palmer et al., 2001, Kaya et al., 2000]. In the current work, no such performance enhancements were observed, with silicon pMOSFETs exhibiting higher hole carrier velocities than those observed in SiGe for channel lengths less than 100 nm. Greater ion implantation damage to the SiGe channel was speculated as a possible reason, and the author speculates as to whether such damage would be visible using high-resolution transmission electron microscopy (HR-TEM). A review of the original simulation work could also provide further evidence of the experimental and theoretical discrepancies. This experimental result could have important technological implications for biaxially strained SiGe as a potential channel material for future generation CMOS circuits. Uni-axially strained SiGe-on-insulator pMOSFETs have recently been demonstrated to exhibit an 80% enhancement in drive current over SOI control devices for gate lengths of 40 nm using elevated source and drain regions, which could reduce the likelihood of ion implantation damage in the SiGe layer. This indicates that SiGe may still hold some promise as an alternative channel material to silicon, and that the transistor design and

the technique by which strain is introduced into the channel is a critical factor [Irisawa et al., 2006].

The final section of this thesis examined impact ionisation in SiGe pMOSFETs. Pseudomorphic SiGe pMOSFETs were found to exhibit reduced impact ionisation compared to the silicon control pMOSFETs, and this was attributed to the strain-induced reduction of the density of states in both the valence and conduction bands, reducing the probability of an impact ionisation event. The result remains valid subject to the confirmation of the conduction band offset. There is evidence in the literature to suggest that the conduction band offset between Si and SiGe could be either type I or type II [Baier et al., 1994, Houghton et al., 1995], although the more recent consensus supports a type II band alignment [Thewalt et al., 1997]. A type I band alignment could adversely affect the result if it led to electrons being confined in a SiGe quantum well, which would result in a lower substrate current being measured and hence an apparently lower impact ionisation. The author speculates as to whether modelling of the experimental data with the conduction band offset as a parameter could be done sufficiently accurately to verify this result.

As a final word, there still exists enormous interest in alternative high-mobility channel materials such as strained SiGe and Ge. In particular, the results observed for short-channel Ge pMOSFETs in this work, hold much promise for Ge channel pMOSFETs in future generation CMOS circuits. Despite not observing a performance improvement in pseudomorphic SiGe pMOSFETs for channel lengths less than 100 nm in the current work, the recent results by Irisawa et al. [2006] for 40 nm SiGe-on-insulator pMOSFETs are very encouraging and at this stage SiGe channel transistors still cannot be ruled out as possible silicon replacements. Indeed, there is still much research to be done in the field of SiGe and Ge-channel transistors, with exciting times still ahead.

Bibliography

F. Andrieu, T. Ernst, K. Romanjek, O. Weber, C. Renard, J. M. Hartmann, A. Toffoli, A. M. Papon, R. Truche, P. Holliger, L. Brevard, G. Ghibaudo, and S. Deleonibus. "SiGe Channel p-MOSFETs Scaling-Down". *Proc. ESSDERC*, pages 267–270, 2003.

A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva. "Simulation of Intrinsic Paramter Fluctuations in Decananometer and Nanometer-Scale MOSFETs". *I.E.E.E. Trans. Electron Device*, 50:1837–1852, 2003.

T. Baier, U. Mantz, K. Thonke, R. Sauer, Schäeffler, F., and H.-J. Herzog. "Type-II Band Alignment in $Si/Si_{1-x}Ge_x$ Quantum Wells from Photoluminescence Line Shifts Due to Optically Induced Band-Bending Effects: Experiment and Theory". *Phys. Rev. B*, 50:15191–15192, 1994.

J. Bardeen and W. H. Brattain. "The Transistor, A Semi-Conductor Triode". *Phys. Rev.*, 74:230–231, 1948.

V. Barral, T. Poiroux, M. Vinet, J. Widiez, B. Previtali, P. Grosgeorges, G. Le Carval, S. Barraud, J. L. Autran, D. Munteanu, and S. Deleonibus. "Experimental Determination of the Channel Backscattering Coefficient on 10 - 70 nm Metal-Gate Double-Gate Transistors". *Solid-State Electron.*, 51:537–542, 2007.

P. Batude, X. Garros, L. Clavelier, C. Le Royer, J. M. Hartmann, V. Loup, P. Besson, L. Vandroux, S. Deleonibus, and F. Boulanger. "In-Depth Investigation of the Mech-

anisms Impacting C-V/G-V Characteristics of Ge/GeON/HfO_2/TiN Stacks by Electrical Modeling". *Microneleclon. Eng.*, 84:2320–2323, 2007.

A. Bouhada, A. Touhami, and S. Bakkali. "New Model of Gate-Induced Drain Current Density in an NMOS Transistor". *Microelectronics Journal*, 29:813–816, 1998.

F. M. Bufler and B. Meinerzhagen. "Hole Transport in Strained $Si_{1-x}Ge_x$ Alloys on $Si_{1-y}Ge_y$ Substrates". *J. Appl. Phys.*, 84:5597–5602, 1998.

J. P. Chang. High-κ gate dielectric deposition technologies. In *High Dielectric Constant Materials: VLSI MOSFET Applications*. Springer, 2005.

R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, and M. Radosavljevic. "Benchmarking Nanotechnology for High-Performance and Low-Power Logic Transistor Applications". *I.E.E.E. Trans. Nanotechnol.*, 4:153–158, 2005.

R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz. "High-κ/Metal-Gate Stack and its MOSFET Characteristics". *I.E.E.E. Electron Device Lett.*, 25: 408–410, 2004.

J. Chen, T. Y. Chan, I. C. Chen, P. K. Ko, and C. Hu. "Subbreakdown Drain Leakage Current in MOSFET". *I.E.E.E. Electron. Device Lett.*, 5:515–517, 1987.

J. G. J. Chern, P. Chang, R. F. Motta, and N. Godinho. "A New Method to Determine MOSFET Channel Length". *I.E.E.E. Electron Device Lett.*, 1:170–173, 1980.

C. O. Chui, F. Ito, and K. C. Saraswat. "Nanoscale Germanium MOS Dielectrics — Part I: Germanium Oxynitrides". *I.E.E.E. Transactions on Electron Devices*, 53:1501–1508, 2006.

C. O. Chui, H. Kim, D. Chi, B. B. Triplett, P. C. McIntyre, and K. C. Saraswat. "A

Sub-400°C Germanium MOSFET Technology with High-κ Dielectric and Metal Gate". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 437–440, 2002.

C. O. Chui, L. Kulig, J. Moran, W. Tsai, and K. C. Saraswat. "Germanium n-Type Shallow Junction Activation Dependencies". *Appl. Phys. Lett.*, 87:091909, 2005.

K.-J. Chui, K.-W. Ang, N. Balasubramanian, M.-F. Li, G. S. Samudra, and Y.-C. Yeo. "n-MOSFET with Silicon-Carbon Source/Drain for Enhancement of Carrier Transport". *I.E.E.E. Trans. Electron Device*, 54:249–256, 2007.

J.-P. Colinge and C. A. Colinge. *Physics of Semiconductor Devices*. Springer, 2006.

K. K. S. Curreem, P. F. Lee, K. S. Wong, J. Y. Dai, M. J. Zhou, J. Wang, and Q. Li. "Comparison of Interfacial and Electrical Characteristics of HfO_2 and HfAlO High-k Dielectrics on Compressively Strained $Si_{1-x}Ge_x$". *Appl. Phys. Lett.*, 88:182905, 2006.

C. K. Date and J. D. Plummer. "Increased Hot-Carrier Effects using SiGe Layers in Vertical Surrounding-Gate MOSFETs". *I.E.E.E. Trans. Electron Device*, 48:2690–2694, 2001.

B. De Jaeger, R. Bonzom, F. Leys, O. Richard, J. Van Steenbergen, G. Winderickx, E. Van Moorhem, G. Raskin, F. Letertre, T. Billon, M. Meuris, and M. Heyns. "Optimisation of a Thin Epitaxial Si layer as Ge Passivation Layer to Demonstrate Deep Sub-micron n- and p-FETs on Ge-On-Insulator Substrates". *Microelectron. Eng.*, 80: 26–29, 2005.

B. De Jaeger, B. Kaczer, P. Zimmerman, K. Opsomer, G. Winderickx, J. Van Steenbergen, E. Van Moorhem, V. Terzieva, R. Bonzom, F. Leys, C. Arena, M. Bauer, C. Werkhoven, M. Caymax, M. Meuris, and M. Heyns. "Ge Deep Sub-micron HiK/MG pFETs with Superior Drive Compared to Si HiK/MG State-of-the-Art Reference". *Semicond. Sci. Tech.*, 22:S221–S226, 2007.

O. Engstrom. unpublished, 2007.

E. K. Evangelou, G. Mavrou, A. Dimoulas, and N. Konofaos. "Rare Earth Oxides as High-κ Dielectrics for Ge Based MOS Devices: An Electrical Study of Pt/Gd_2O_3/Ge Capacitors". *Solid-State Electron.*, 51:164–169, 2007.

J. Feng, Y. Liu, P. B. Griffin, and J. D. Plummer. "Integration of Germanium-on-Insulator and Silicon MOSFETs on a Silicon Substrate". *I.E.E.E. Electron Device Lett.*, 27:911–913, 2006.

G. Ferrari, J. R. Watling, S. Roy, J. R. Barker, P. Zeitoff, G. Bersuker, and A. Asenov. "Monte Carlo Study of Mobility in Si Devices with HfO_2-Based Oxides". *Mater. Sci. Semicond. Process.*, 9:995–999, 2006.

M. V. Fischetti and S. E. Laux. "Band Structure, Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiGe Alloys". *J. Appl. Phys.*, 80:2234–2252, 1996.

M. V. Fischetti and S. E. Lauz. "Monte Carlo Study of Electron Transport in Silicon Inversion Layers". *Phys. Rev. B*, 48:2244–2274, 1993.

M. V. Fischetti, D. Neumayer, and E. A. Cartier. "Effective Electron Mobility in Si Inversion Layers in Metal-Oxide-Semiconductor Systems with a High-κ Insulator: The Role of Remote Phonon Scattering". *J. Appl. Phys.*, 90:4587–4608, 2001.

S. Galdin, P. Dollfus, V. Aubry-Foruna, P. Hesto, and H. J. Osten. "Band Offset Predictions for Strained Group IV Alloys: $Si_{1-x-y}Ge_xC_y$ on Si(001) and $Si_{1-x}Ge_x$ on $Si_{1-z}Ge_z$(001)". *Semicond. Sci. Tech.*, 15:565–572, 2000.

J. L. Gavartin, D. M. Ramo, A. L. Shluger, G. Bersuker, and B. H. Lee. "Negative Oxygen Vacancies in HfO_2 as Charge Traps in High-κ Stacks". *Appl. Phys Lett.*, 89: 082908, 2006.

T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffman, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr. "A 90nm High Volume Manufacturing Logic Technology Featuring Novel 45nm Gate Length Strained Silicon CMOS Transistors". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 978–980, 2003.

G. Ghibaudo. "Electrical Characterisation of Single and Multi Gates Devices". *MIGAS Emerging Silicon Devices for the End of the Roadmap*, page 18, 2006.

G. Ghibaudo, S. Bruyère, T. Devoivre, B. DeSalvo, and E. Vincent. "Improved Method for the Oxide Thickness Extraction in MOS Structures with Ultrathin Gate Dielectrics". *I.E.E.E. Trans. on Semicon. Manufact.*, 13:152–154, 2000.

K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer. "Impact Ionisation MOS (I-MOS)–Part I: Device and Circuit Simulations". *I.E.E.E. Trans. Electron Device*, 52:69–76, 2005a.

K. Gopalakrishnan, R. Woo, C. Jungemann, P. B. Griffin, and J. D. Plummer. "Impact Ionisation MOS (I-MOS)–Part II: Experimental Results". *I.E.E.E. Trans. Electron Device*, 52:77–84, 2005b.

P. V. Gray and D. M. Brown. "Density of SiO_2-Si Interface States". *Appl. Phy. Lett.*, 8:31–33, 1966.

J. Hallstedt, M. von Hartmann, P.-E. Hellstrom, M. Ostling, and H. H. Radamsson. "Hole Mobility in Ultrathin Body SOI pMOSFETs with SiGe or SiGeC Channels". *I.E.E.E. Electron Device Lett.*, 27:466–468, 2006.

C. C. Hobbs, L. R. C. Fonseca, A. Knizhnik, V. Dhandapani, S. B. Samavedam, W. J. Taylor, J. M. Grant, L. G. Dip, D. H. Triyoso, R. I. Hegde, D. C. Gilmer, R. Garcia,

D. Roan, M. L. Lovejoy, R. S. Rai, E. A. Hebert, H.-H. Tseng, S. G. H. Anderson, B. E. White, and P. J. Tobin. "Fermi-Level Pinning at the Polysilicon/Metal Oxide Interface - Part I". *I.E.E.E. Trans. Electron Dev.*, 51:971–977, 2004a.

C. C. Hobbs, L. R. C. Fonseca, A. Knizhnik, V. Dhandapani, S. B. Samavedam, W. J. Taylor, J. M. Grant, L. G. Dip, D. H. Triyoso, R. I. Hegde, D. C. Gilmer, R. Garcia, D. Roan, M. L. Lovejoy, R. S. Rai, E. A. Hebert, H.-H. Tseng, S. G. H. Anderson, B. E. White, and P. J. Tobin. "Fermi-Level Pinning at the Polysilicon/Metal Oxide Interface - Part II". *I.E.E.E. Trans. Electron Dev.*, 51:978–984, 2004b.

D. C. Houghton, G. C. Aers, S.-R. Eric Yang, E. Wang, and N. L. Rowell. "Type I Band Alignment in $Si_{1-x}Ge_x$/Si(100) Quantum Wells: Photoluminescence Under Applied [110] and [100] Uniaxial Stress". *Phys. Rev. Lett.*, 75:866–869, 1995.

M. Houssa, M. Tuominen, M. Naili, V. Afanas'ev, A. Stesmans, S. Haukka, and M. M. Heyns. "Trap-Assisted Tunneling in High Permittivity Gate Dielectric Stacks". *J. Appl. Phys.*, 87:8615–8620, 2000.

H. R. Huff and D. C. Gilmer. *High Dielectric Constant Materials: VLSI MOSFET Applications*. Springer, 2005.

M. Inoue, S. Tsujikawa, M. Mizutani, K. Nomura, T. Hayashi, K. Shiga, J. Yugami, J. Tsuchimoto, Y. Ohno, and M. Yoneda. "Fluorine Incorporation into HfSiON Dielectric for V_th Control and its Impact on Reliability for Poly-Si Gate pFET". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 413–416, 2005.

Intel. Intel Unveils World's Best Processor, available online http://www.intel.com/ca/pressroom/2006/0727.htm, 2005.

Intel. Intel's Transistor Technology Breakthrough Represents

Biggest Change to Computer Chips in 40 Years, available online http://www.intel.com/ca/pressroom/archive/releases/20070128comp.htm, 2007.

T. Irisawa, T. Numata, T. Tezuka, K. Usuda, N. Hirashita, N. Sugiyama, E. Toyoda, and S. Takagi. "High Performance Uniaxially Strained SiGe-on-Insulator pMOSFETs Fabricated by Lateral-Strain-Relaxation Technique". *I.E.E.E. Trans. Electron Device*, 53:2809–2815, 2006.

ITRS. "Process Integration, Devices and Structures", 2000 Update.

ITRS. Process Integration, Devices and Structures, 2006 Update.

B. Jin, S. Datta, G. Dewey, M. Doczy, B. S. Doyle, K. Johnson, J. Kavalieros, M. Metz, N. Zelick, and R. Chau. "Mobility Enhancement in Compressively Strained SiGe Channel pMOSFET with HfO_2/TiN Gate Stack". *206th ECS Meeting*, page 1293, 2004.

B. Kaczer, B. De Jaeger, G. Nicholas, K. Martens, R. Degraeve, M. Houssa, G. Pourtois, F. Leys, M. Meuris, and G. Groeseneken. "Electron and Reliability Characterisation of Metal-Gate/HfO_2/Ge FETs with Si Passivation". *Microneleclon. Eng.*, 84:2067–2070, 2007.

Y. Kamata, Y. Kamimuta, T. Ino, R. Iijima, M. Koyama, and A. Nishiyama. "Influences of Annealing Temperature on Characteristics of Ge p-Channel Metal Oxide Semiconductor Field Effect Transistors with ZrO_2 Gate Dielectrics". *Jap. J. Appl. Phys.*, 45: 5651–5656, 2006.

G. S. Kar, S. Maikap, S. K. Banerjee, and S. K. Ray. "Hole Velocity Overshoot in Partially Strain Compensated $Si_{0.793}Ge_{0.2}C_{0.007}$ Inversion Layers". *Electron. Lett.*, 38:141–142, 2002.

S. Kaya, Y.-P. Zhao, J. R. Watling, A. Asenov, J. R. Barker, G. Ansaripour, G. Braithwaite, T. E. Whall, and E. H. C. Parker. "Indication of Velocity Overshoot in Strained $Si_{0.8}Ge_{0.2}$ p-Channel MOSFETs". *Semicond. Sci. Technol.*, 15:573–578, 2000.

M. J. Kearney and A. I. Horrell. "The Effect of Alloy Scattering on the Mobility of Holes in a $Si_{1-x}Ge_x$ Quantum Well". *Semicond. Sci. Tech.*, 13:174–180, 1998.

W. Kern. *Handbook of Semiconductor Wafer Cleaning Technology*. Noyes Publications, 1993.

C. Kittel. *Introduction to Solid State Physics Seventh Edition*. Wiley, 1996.

J. Koomen. "Investigation of the MOST Channel Conductance in Weak Inversion". *Solid-State Electonics*, 16:801–810, 1973.

K. Kukli, M. Ritala, S. Sajavaara, J. Keinonen, and Leskelä, M. "Comparison of Interfacial and Electrical Characteristics of HfO_2 and HfAlO High-k Dielectrics on Compressively Strained $Si_{1-x}Ge_x$". *Appl. Phys. Lett.*, 88:182905, 2006.

K. S. A. Kwa, S. Chattopadhyay, N. D. Jankovic, S. H. Olsen, L. S. Driscoll, and A. G. O'Neill. "A Model for Capacitance Reconstruction from Measured Lossy MOS Capacitance-Voltage Characteristics". *Semicond. Sci. Tech.*, 18:82–87, 2003.

T. A. Langdo, M. T. Currie, Z.-Y. Cheng, J. G. Fiorenza, M. Erdtmann, G. Braithwaite, C. W. Leitz, C. J. Vineis, J. A. Carlin, A. Lochtefeld, M. T. Bulsara, I. Lauer, D. A. Antoniadis, and M. Somerville. "Strained Si on Insulator Technology: From Materials to Devices". *Solid-State Electronics*, 48:1357–1367, 2004.

D. R. Leadley, H. Fischer, L. Risch, M. J. Kearney, A. I. Horrell, E. H. C. Parker, and T. E. Whall. "Performance Enhancement in $Si/Si_{0.5}Ge_{0.5}/Si$ Strained Alloy p-Channel Metal Oxide Semiconductor Field Effect Transistors". *Semicond. Sci. Tech.*, 18:945–949, 2003.

J. Lee, A. L. Gutierrez-Aitken, S. H. Li, and P. K. Bhattacharya. "Impact Ionisation Coefficients in $Si_{1-x}Ge_x$". *Appl. Phys. Lett.*, 66:204–205, 1995.

M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. Currie, and A. Lochtefeld. "Strained Si, SiGe, and Ge Channels for High-Mobility Metal-Oxide-Semiconductor Field-Effect Transistors". *J. Appl. Phys.*, 97:1–27, 2005.

M. C. Lemme, J. K. Efavi, T. Mollenhauer, M. Schmidt, H. D. B. Gottlob, T. Wahlbrink, and H. Kurz. "Nanoscale TiN Metal Gate Technology for CMOS Integration". *Microelectron. Eng.*, 83:1551–1554, 2006.

R. Li, J. S. Lee, H. B. Yao, D. Z. Chi, M. B. Yu, and D.-L. Kwong. "Pt-Germanide Schottky Source/Drain Germanium p-MOSFET with HfO_2 Gate Dielectric and TaN Gate Electrode". *I.E.E.E. Electron Device Lett.*, 27:476–477, 2006.

A. Lochtefeld, I. J. Djomehri, G. Samudra, and D. A. Antoniadis. "New Insights into Carrier Transport in n-MOSFETs". *IBM J. Res. and Dev.*, 46:347–357, 2002.

W. Y. Loh, B. J. Cho, and M. F. Li. "Correlation Between Interface Traps and Gate Oxide Leakage Current in the Direct Tunneling Regime". *Appl. Phys. Lett.*, 81: 379–381, 2002.

N. Lu, H.-J. Li, M. Gardner, S. Wickramanayaka, and D.-L. Kwong. "Electrical Properties of Amorphous High-κ HfTaTiO Gate Dielectric with Dielectric Constants of 40-60". *I.E.E.E. Electron Device Lett.*, 26:298–300, 2005.

M. Lundstrom. "Scattering Theory of the Short Channel MOSFET". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 387–390, 1996.

M. Lundstrom. *Fundamentals of Carrier Transport 2nd Edition*. Cambridge University Press, 2000.

K. Maex. "Silicides for Integrated Circuits: TiS_2 and $CoSi_2$". *Materials Science and Engineering*, 11:53–153, 1993.

K. Martens, B. De Jaeger, R. Bonzom, J. Van Steenbergen, M. Meuris, G. Groeseneken, and H. Maes. "New Interface State Density Extraction Method Applicable to Peaked and High-Density Distributions for Ge MOSFET Development". *I.E.E.E. Electron Device Lett.*, 27:405–407, 2006.

G. Mavrou, S. F. Galata, A. Sotiropoulos, P. Tsipas, Y. Panayiotatos, A. Dimoulas, E. K. Evangelou, J. W. Seo, and Ch. Dieker. "Germanium Metal-Insulator-Semiconductor Capacitors with Rare Earth La_2O_3 Gate Dielectric". *Microelectron. Eng.*, 84:2324–2327, 2007.

V. Misra. Issues in metal gate electrode selection for bulk cmos devices. In *High Dielectric Constant Materials: VLSI MOSFET Applications*. Springer, 2005.

G. E. Moore. "Cramming More Components onto Integrated Circuits". *Electronics.*, 38: 116–119, 1965.

S. Nakaharai, T. Tezuka, N. Sugiyama, Y. Moriyama, and S. Takagi. "Characterisation of 7-nm-Thick Strained Ge-on-Insulator Layer Fabricated by Ge-Condensation Technique". *Appl. Phys. Lett.*, 83:3516–3518, 2003.

K. Natori. "Ballistic Metal-Oxide-Semiconductor Field Effect Transistor". *J. Appl. Phys.*, 76:4879–4890, 1994.

A. Nayfeh, C. O. Chui, T. Yonehara, and K. C. Saraswat. "Fabrication of High-Quality p-MOSFET in Ge Grown Heteroepitaxially on Si". *I.E.E.E. Electron. Device Lett.*, 26:311–313, 2005.

G. Nicholas, T. J. Grasby, E. H. C. Parker, T. E. Whall, and T. Skotnicki. "Evidence of

Reduced Self-Heating in Strained Si MOSFETs". *I.E.E.E. Electron Device Lett.*, 26: 684–686, 2005a.

G. Nicholas, T. J. Grasby, E. H. C. Parker, T. E. Whall, and T. Skotnicki. "Impact Ionisation in Strained Si Devices". *J. Appl. Phys.*, 98:104502, 2005b.

G. N. Nicholas, D. P. Brunco, D. A. Dimoulas, J. Van Steenbergen, F. Bellenger, M. Houssa, M. Caymax, M. Meuris, Y. Panayiotatos, and A. Sotiropoulos. "Germanium MOSFETs with CeO_2/HfO_2/TiN Gate Stacks". *I.E.E.E. Trans. Electron Dev.*, 54:1425–1430, 2007.

E. H. Nicollian and J. H. Brews. *MOS (Metal Oxide Semiconductor Physics and Technology*. Wiley, 1982.

A. Ortiz-Conde, F. J. García Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue. "A Review of Recent MOSFET Threshold Voltage Extraction Methods". *Microelectron. Rel.*, 42:583–596, 2002.

M. J. Palmer. *"Investigation of High Mobility Pseudomorphic SiGe p-Channels in Si MOSFETs at Low and High Electric Fields"*. PhD thesis, University of Warwick, 2001.

M. J. Palmer, G. Braithwaite, M. J. Prest, E. H. C. Parker, T. E. Whall, Y. P. Zhao, S. Kaya, J. R. Watling, A. Asenov, J. R. Barker, A. M. Waite, and A. G. R. Evans. "Enhanced Velocity Overshoot and Transconductance in $Si/Si_{0.64}Ge_{0.36}/Si$ pMOSFETs - Predictions for Deep Submicron Devices". *Proc. ESSDERC*, pages 199–202, 2001.

T. J. Park, S. K. Kim, J. H. Kim, J. Park, M. Cho, S. W. Lee, S. H. Hong, and C. S. Hwang. "Electrical Properties of High-k HfO_2 Films on $Si_{1-x}Ge_x$ Substrates". *Microelectron. Eng.*, 80:222–225, 2005.

E. H. C. Parker and T. E. Whall. "SiGe Heterostructure CMOS Circuits and Applications". *Solid-State Electron.*, 43:1497–1506, 1999.

D. J. Paul. "Silicon-Germanium Strained Layer Materials in Microelectronics". *Adv. Mater.*, 11:191–204, 1999.

D. J. Paul. "Si/SiGe Heterostructures: From Material and Physics to Devices and Circuits". *Semicond. Sci. Tech.*, 19:R75–R108, 2004.

R. People and J. C. Bean. "Calculation of Critical Layer Thickness Versus Lattice Mismatch for Ge_xSi_{1-x}/Si Strained-Layer Heterostructures". *Appl. Phys. Lett.*, 47: 322–324, 1985.

J. D. Plummer and P. B. Griffin. "Material and Process Limits in Silicon VLSI Technology". *Proc. I.E.E.E.*, 89:240–258, 2001.

G. Pourtois, M. Houssa, B. De Jaeger, F. E. Leys, B. Kaczer, M. Caymax, M. Meuris, G. Groeseneken, and M. Heyns. "A Step Towards a Better Understanding of Silicon Passivated (100)-Ge p-Channel Devices". *211th ECS Meeting*, page 558, 2007.

B. Raeissi. Private Communication, 2007.

L.-Å. Ragnarsson, S. Severi, L. Trojman, K. D. Johnson, D. P. Brunco, M. Aoulaiche, M. Houssa, T. Kauerauf, R. Degraeve, A. Delabie, V. S. Kaushik, S. De Gendt, W. Tsai, G. Groeseneken, K. De Meyer, and M. Heyns. "Electrical Characteristics of 8-ÅEOT HfO_2/TaN Low Thermal-Budget n-Channel FETs With Solid-Phase Epitaxially Regrown Junctions". *I.E.E.E. Trans. Electron. Dev.*, 53:1657–1668, 2006.

K. Rajendran and W Scoenmaker. "Studies of Boron Diffusivity in Strained $Si_{1-x}Ge_x$ Epitaxial Layers". *J. Appl. Phys.*, 89:980–987, 2001.

S. Richard, N. Zerounian, N. Cavassilas, G. Hock, T. Hackbarth, H. J. Herzog, and

F. Aniel. "Electroluminescence of Ge/SiGe p-MODFETs". *Mater. Sci. Semicond. Process.*, 8:377–382, 2005.

A. Ritenour, A. Khakifirooz, D. A. Antoniadis, R. Z. Lei, W. Tsai, A. Dimoulas, G. Mavrou, and Y. Panayiotatos. "Subnanometer-Equivalent-Oxide-Thickness Germanium p-Metal-Oxide-Semiconductor Field Effect Transistors Fabricated Using Molecular-Beam-Deposited High-κ/Metal Gate". *Appl. Phys. Lett.*, 88:132107, 2006.

A. Ritenour, S. Yu, M. L. Lee, N. Lu, W. Bai, A. Pitera, E. A. Fitzgerald, D.-L. Kwong, and D. A. Antoniadis. "Epitaxial Strained Germanium p-MOSFET with HfO_2 Gate Dielectric and TaN Gate Electrode". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 433–436, 2003.

J. Robertson. "High-κ Dielectrics Constant Gate Oxides for Metal Oxide Si Transistors". *Rep. Prog. Phys.*, 69:327–396, 2006.

K. Romanjek, F. Andrieu, T. Ernst, and G. Ghibaudo. "Improved Split C-V Method for Effective Mobility Extraction in Sub-0.1-μm Si MOSFETs". *I.E.E.E. Electron Device Lett.*, 25:583–585, 2004.

K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits". *Proc. I.E.E.E.*, 91:305–327, 2003.

C. Salm, D. T. van Veen, D. J. Gravesteijn, J. Holleman, and P. H. Woerlee. "Diffusion and Electrical Properties of Boron and Arsenic Doped Poly-Si and Poly-Ge_xSi_{1-x} (x$\sim$0.3) as Gate Material for Sub-0.25μm Complementary Metal Oxide Semiconductor Applications". *J. Electrochem. Soc.*, 144:3665–3673, 1997.

K. C. Saraswat, C. O. Chui, T. Krishnamohan, D. Kim, A. Nayfeh, and A. Pethe. "High Performance Germanium MOSFETs". *Mat. Sci. Eng. B*, 135:242–249, 2006.

Schäeffler, F. "High-Mobility Si and Ge Structures". *Semicond. Sci. Tech.*, 12:1515–1549, 1997.

M. Schmidt, M. C. Lemme, H. Kurz, T. Witters, T. Schram, K. Cherkaoui, A. Negara, and P. K. Hurley. "Impact of H_2/N_2 Annealing on Interface Defect Densities in Si(100)/SiO_2/HfO_2/TiN Gate Stacks". *Microelectron. Eng.*, 80:70–73, 2005.

D. K. Schroder. *Semiconductor Material and Device Characterization Third Edition*. John Wiley & Sons, Inc, 2006.

K.-I. Seo, R. Sreenivasan, P. C. McIntyre, and K. C. Saraswat. "Improvement in High-κ (HfO_2/SiO_2) Reliability by Incorporation of Fluorine". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 417–420, 2005.

H. Shang, H. Okorn-Schmidt, K. K. Chan, M. Copel, J. A. Ott, P. M. Kozlowski, S. E. Steen, S. A. Cordes, H.-S. P. Wong, E. C. Jones, and W. E. Haensch. "High Mobility p-Channel Germanium MOSFETs with a Thin Ge Oxynitride Gate Dielectric". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 441–444, 2002.

H. Shang, H. Okorn-Schmidt, J. Ott, P. Kozlowski, S. Steen, E. C. Jones, H.-S. P. Wong, and W. Hanesch. "Electrical Characterisation of Germanium p-Channel MOSFETs". *I.E.E.E. Electron Device Lett.*, 24:242–244, 2003.

M. Shur. *Physics of Semiconductor Devices*. Prentice Hall, 1990.

SIA-Online. SIA Issues - Economy, available online http://www.sia-online.org/iss_economy.cfm, 2005.

C. G. Sodini, T. W. Ekstedt, and J. L. Moll. "Charge Accumulation and Mobility in Thin Dielectric MOS Transistors". *Solid-State Electonics*, 25:833–841, 1982.

S. M. Sze. *Physics of Semiconductor Devices 2nd Edition*. Wiley, 1981.

S. M. Sze. *VLSI Technology*. McGraw-Hill, 1983.

S. Takagi, A. Toriumi, M. Iwase, and H. Tango. "On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I – Effects of Substrate Impurity Concentration". *I.E.E.E. Trans. Electron Device*, 41:2357–2362, 1994.

Y. Taur, C. H. Hsu, B. Wu, R. Kiehl, B. Davari, and G. Shahidi. "Saturation Transconductance of Deep-Submicron-Channel MOSFETs". *Solid-State Electron.*, 36:1085–1087, 1993.

Y. Taur and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 1998.

K. Terada and H. Muta. "A New Method to Determine Effective MOSFET Channel Length". *Jap. J. of Appl. Phys.*, 18:953–959, 1979.

M. L. W. Thewalt, D. A. Harrison, C. F. Reinhart, J. A. Wolk, and H. Lafontaine. "Type II Band Alignment in $Si_{1-x}Ge_x$/Si(100) Quantum Wells: The Ubiquitous Type I Luminescence Results from Band Bending". *Phys. Rev. Lett.*, 79:269–272, 1997.

K. Tse, D. Liu, K. Xiong, and J. Robertson. "Oxygen Vacancies in High-k Oxides". *Microelectron. Eng.*, 84:2028–2031, 2007.

K. Tse and J. Robertson. "Defect Pasivation in HfO_2 Gate Oxide by Fluorine". *Appl. Phys. Lett.*, 89:142914, 2006.

Y. Tsividis. *Operation and Modeling of the MOS Transistor*. McGraw-Hill, 1999.

M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch, and M. Hirose. "Physically-Based Threshold Voltage Determination for MOSFETs of All Gate Lengths". *I.E.E.E. Trans. Electron Device*, 46:1429–1434, 1999.

S. Tyagi, C. Auth, P. Bai, G. Curello, H. Deshpande, S. Gannavaram, O. Golonzka, R. Heussner, R. James, C. Kenyon, S.-H. Lee, N. Lindert, M. Liu, R. Nagisetty, S. Natarajan, C. Parker, J. Sebastian, B. Sell, S. Sivakumar, A. St Amour, and K. Tone. "An Advanced Low Power, High Performance, Strained Channel 65nm Technology". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 245–247, 2005.

S. Van Elshocht, B. Brijs, M. Caymax, T. Conard, T. Chiarella, S. De Gendt, B. De Jaeger, S. Kubieck, M. Meuris, B. Onsia, O. Richard, I. Teerlinck, J. Van Steenbergen, C. Zhao, and M. Heyns. "Deposition of HfO_2 on Germnaium and the Impact of Surface Pretreatments". *Appl. Phys. Lett.*, 85:3824–3826, 2004.

N. S. Waldron, A. J. Pitera, M. L. Lee, E. A. Fitzgerald, and J. A. del Alamo. "Impact Ionisation in Strained-Si/SiGe Heterostructures". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 813–816, 2003.

R. M. Wallace and G. D. Wilk. Materials issues in high-κ gate dielectric selection and integration. In *High Dielectric Constant Materials: VLSI MOSFET Applications*. Springer, 2005.

O. Weber, J.-F. Damlencourt, F. Andrieu, F. Ducroquet, T. Ernst, J.-M. Hartmann, A.-M. Papon, O. Renault, B. Guillaumot, and S. Deleonibus. "Fabrication and Mobility Characteristics of SiGe Surface Channel pMOSFETs with a HfO_2/TiN Gate Stack". *I.E.E.E. Trans. Electron Device*, 53:449–456, 2006.

T. E. Whall and E. H. C. Parker. "SiGe Heterostructures for FET Applications". *J. Phys. D: Appl. Phys.*, 31:1397–1416, 1998.

S. J. Whang, S. J. Lee, F. Gao, N. Wu, C. X. Zhu, J. S. Pan, L. J. Tang, and D.-L. Kwong. "Germanium p- & n-MOSFETs Fabricated with Novel Surface Passivation (Plasma-PH_3 and Thin AlN) and TaN/HfO_2 Gate Stack". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 307–310, 2004.

G. D. Wilk, R. M. Wallace, and J. M. Anthony. "High-κ Dielectrics: Current Status and Materials Properties Considerations". *J. Appl. Phys.*, 89:5243–5275, 2001.

S. Wolf. *Silicon Processing for the VLSI Era Volume 3 - The Submicron MOSFET*. Lattice Press, 1995.

D. Wu, A.-C. Lindgren, S. Persson, G. Sjöblom, M. von Hartmann, J. Seger, P.-E. Hellström, J. Olsson, H.-O. Blom, S.-L. Zhang, M. Östling, E. Vainonen-Ahlgren, W.-M. Li, E. Tois, and M. Tuominen. "A Novel Strained $Si_{0.7}Ge_{0.3}$ Surface-Channel pMOSFET with an ALD $TiN/Al_2O_3/HfAlO_x/Al_2O_3$ Gate Stack". *I.E.E.E. Electron Device Lett.*, 24:171–173, 2003.

D. Wu, J. Lu, E. Vainonen-Ahlgren, E. Tois, M. Tuominen, Östling, M., and S.-L. Zhang. "Structural and Electrical Characterization of $Al_2O_3/HfO_2/Al_2O_3$ on Strained SiGe". *Solid-State Elec.*, 49:193–197, 2005.

N. Wu, Q. Zhang, N. Balasubramanian, D. S. H. Chan, and C. Zhu. "Characteristics of Self-Aligned Gate-First Ge p- and n-Channel MOSFETs Using CVD HfO_2 Gate Dielectric and Si Surface Passivation". *I.E.E.E. Trans. Electron. Dev.*, 54:733–741, 2007.

N. Wu, Q. Zhang, C. Zhu, D. S. H. Chan, A. Du, N. Balasubramanian, M. F. Li, A. Chin, J. K. O. Sin, and D.-L. Kwong. "A TaN-HfO_2-Ge pMOSFET with Novel SiH_4 Surface Passivation". *I.E.E.E. Electron. Device Lett.*, 25:631–633, 2004.

Y. H. Xie. "SiGe Field Effect Transistors". *Materials Science and Engineering*, 25: 89–121, 1999.

E. S. Yang. *Microelectronic Devices*. McGraw-Hill, 1988.

L. Yang, J. R. Watling, R. C. W. Wilkins, M. Borici, J. R. Barker, A. Asenov, and

S. Roy. "Si/SiGe Heterostructure Parameters for Device Simulations". *Semicond. Sci. Technol.*, 19:1174–1182, 2004.

M. Yang, M. Ieong, L. Shi, K. Chan, V. Chan, A. Chou, E. Gusev, K. Jenkins, D. Boyd, Y. Ninomiya, D. Pendleton, Y. Surpris, D. Heenan, J. Ott, K. Guarini, C. D'Emic, M. Cobb, P. Mooney, B. To, N. Rovedo, J. Benedict, R. Mo, and H. Ng. "High Performance CMOS Fabricated on Hybrid Substrate with Different Crystal Orientations". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 453–456, 2003.

W.-K. Yeh and J.-W. Chou. "Optimum Halo Structure for Sub-0.1 μm CMOSFETs". *I.E.E.E. Trans. Electron Dev.*, 48:2357–2362, 2001.

K. Yeom, J. M. Hinckley, and J. Singh. "Calculation of Electron and Hole Impact Ionisation Coefficients in SiGe Alloys". *J. Appl. Phys.*, 80:6773–6782, 1996.

Q. Zhang, J. Huang, N. Wu, G. Chen, M. Hong, L. K. Bera, and C. Zhu. "Drive Current Enhancement in Ge n-Channel MOSFET Using Laser Annealing for Source/Drain Activation". *I.E.E.E. Electron Device Lett.*, 27:728–730, 2006.

P. Zimmerman, G. Nicholas, B. De Jaeger, B. Kaczer, A. Stesmans, L.-Å. Ragnarsson, D. P. Brunco, F. E. Leys, M. Caymax, G. Winderickx, K. Opsomer, M. Meuris, and M. M. Heyns. "High Performance Ge pMOS Devices Using a Si-Compatible Process Flow". *I.E.E.E. I.E.D.M. Tech. Dig.*, pages 655–658, 2006.