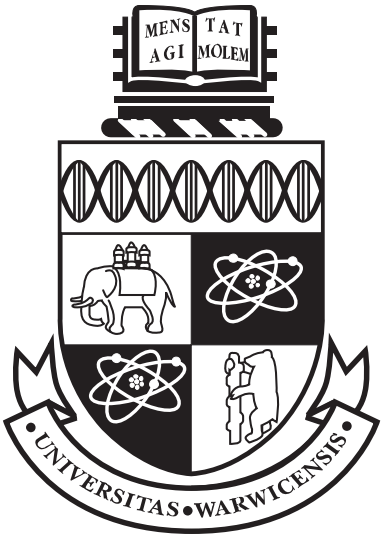


Fabrication and Characterisation of Novel Ge MOSFETs

by

Chris Beer

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Physics

November 2007

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

The author would like to acknowledge IMEC (particularly Gareth Nicholas) who supplied all Ge MOS devices, which make up the majority of results in this thesis. Warwick spin-off company, AdvanceSiS provided CVD growth and expertise, for which the author is grateful. Sample processing on behalf of the author was performed at Innos (Enrico Gilis, Riccardo Varrazza and Mir Mokhtari) and despite their suffering a catastrophic fire in late 2005, samples still arrived from Eindhoven, Netherlands in time for inclusion here. Useful supplementary measurements were organised by Lidia Lukasiak, at the Institute of Microelectronics and Optoelectronics in Warsaw, Poland.

I would like to thank fellow residents of office 522: Tim Naylor, Jon Parsons, Vishal Shah, Stephen Thomas, Dom Pearman and Andrew Dobbie; all of whom helped me through the last three years. A special thanks to recently appointed group leader, David Leadley, for help in writing this thesis. I acknowledge my supervisors, Evan Parker and Terry Whall, for their patience and advice.

Finally, I thank my parents, Karon and Derek Beer, for funding my study of physics through undergraduate and supporting me in my decision to undertake a PhD.

Declarations

This thesis is the result of research carried out by the author as a member of the Nano-Silicon Research Group, Department of Physics, University of Warwick; beginning Oct 2004 and ending Nov 2007. All practical work was performed solely by the author using equipment belonging to the Nano-Si Group and/or the Department of Physics, except where specifically stated in the text. Those samples reported in chapter 4 have been independently characterised by IMEC with the resulting publication:

P. Zimmerman, G. Nicholas, B. Jaeger, B. Kaczar, A. Stesmans, L. A. Ragnarsson, D. P. Brunco, F. E. Leys, M. Caymax, G. Winderickx, K. Opsomer, M. Meuris and M. M. Heyns. "High performance Ge p-MOS devices using a Si-compatible process flow" In *IEDM*, 2006,

to which the author makes no claim. Publications arising directly from this thesis are:

G. Nicholas, T. Grasby, D. Fulgoni, C. Beer and J. Parsons, "High mobility strained Ge p-MOSFETs with high-kappa/metal gate" *IEEE Electron Device Letters*, 28(9):825-827, 2007.

C. Beer, T. Whall, E. Parker, D. Leadley, B. De Jaeger, G. Nicholas, P. Zimmerman, M. Meuris, S. Szostak, G. Gluszko and L. Lukasiak, "Low temperature mobility in hafnium-oxide gated germanium p-channel metal-oxide-semiconductor field-effect transistors" *Applied Physics Letters*, 91:263512, 2007.

Abstract

As high-k dielectrics are introduced into commercial Si CMOS (Complimentary Metal Oxide Semiconductor) microelectronics, the 40 year channel/dielectric partnership of Si/SiO_2 is ended and the door opened for silicon to be replaced as the active channel material in MOSFETs (Metal Oxide Semiconductor Field Effect Transistor). Germanium is a good candidate as it has higher bulk carrier mobilities than silicon. In addition, Si and Ge form a thermodynamically stable SiGe alloy of any composition, allowing Ge to be implemented as a thin layer on the surface of a standard Si substrate. This thesis is a practical investigation on several aspects of Ge CMOS technology.

High-k dielectric Ge p-MOSFETs are electrically characterised. A large variation in interface state densities is demonstrated to be responsible for a threshold voltage shift and this is proportional to reciprocal peak mobility due to the Coulomb scattering of carriers by charged states. A theoretical mobility is fitted to that measured at 4.2 K and confirms that interface states are the main source of interface charged impurities. The model demonstrates a reduction in the interface charged impurity density in p-MOSFETs that underwent a PMA (Post Metallisation Anneal) in hydrogen atmosphere and that the anneal also reduces the RMS (Root Mean Square) dielectric/semiconductor interface roughness, from an average of 0.60 nm to 0.48 nm.

High-k strained Ge p-MOSFETs are electrically characterised and have peak mobilities at 300 K (470 $cm^2\ V^{-1}\ s^{-1}$) and 4.2 K (1780 $cm^2\ V^{-1}\ s^{-1}$) far in excess of those measured for the unstrained Ge p-MOSFETs (285 $cm^2\ V^{-1}\ s^{-1}$, 785 $cm^2\ V^{-1}\ s^{-1}$ respectively). Strained Ge n-MOSFETs perform significantly worse than standard Si n-MOSFETs primarily due to a high source/drain resistance.

A 10 nm thick SiGe-OI (On Insulator) layer with a Ge composition of 58% is obtained from a 55 nm $Si_{0.88}Ge_{0.12}$ initial layer on 100 nm Si-OI substrate via the germanium condensation technique. For the first time, germanium is demonstrated to diffuse through the BOX (Buried OXide) during Ge-condensation and into the underlying Si substrate. An order of magnitude increase in the calculated ITOX (Internal Thermal OXidation) rate of the BOX in the final stages of Ge-condensation is hypothesised to be responsible for stopping this diffusion.

Chapter 1

Introduction

1.1 CMOS Scaling

The exponentially increasing number of components in a single integrated circuit over time was first recognised, and predicted to continue, over forty years ago [Moore, 1965]. This phenomenon is now known as Moore's law and is still accurate. It is a law of economics more than anything else and the only way it can continue is through the shrinking of each component in an integrated circuit, thereby allowing a more complex circuit in the same area or an equally complex circuit in a smaller area. Either way, a reduction in semiconductor surface real-estate per component means a cheaper cost per component.

Logic circuits are predominantly complimentary metal-oxide-semiconductor (CMOS), which comprises of two types MOSFETs (metal oxide semiconductor field effect transistor). Thus to scale the integrated circuit exponentially with time we must scale this most basic component by a constant factor κ (> 1) at a fixed time interval. Table 1.1 lists various MOSFET/CMOS parameters and their scaling behaviour with κ. For the

	Parameters	Multiplicative factor
Scaling assumpsions	Device dimensions (t_{ox}, L, W, x_j)	$\frac{1}{\kappa}$
	Doping concentration (N_a, N_d)	κ
	Voltage (V)	$\frac{1}{\kappa}$
Derived device parameters	Electric field (E_{eff})	1
	Capacitance $(C = \frac{\epsilon WL}{t_{ox}})$	$\frac{1}{\kappa}$
	Current, drift (I)	$\frac{1}{\kappa}$
Derived circuit parameters	Circuit delay time $(= \frac{CV}{I})$	$\frac{1}{\kappa}$
	Power dissipation $= IV$	$\frac{1}{\kappa^2}$
	Power-delay $(= CV^2)$	$\frac{1}{\kappa^3}$
	Circuit density $(\propto \frac{1}{LW})$	κ^2
	Power density $(= \frac{IV}{LW})$	1

Table 1.1: Scaling parameters in CMOS circuit design, where, according to Moore's law, $\kappa \approx \sqrt{2}$ and is applied every two years. Adapted from Taur and Ning [1998].

purposes of this introduction, the MOSFET can be thought of as a capacitor on a semiconductor surface with a length L, width W, dielectric (called the gate dielectric) thickness t_{ox} of dielectric constant ϵ. Each has ion-implanted dopants of concentration N_a or N_d to a depth of x_j. When switched on a current I flows and a power supply voltage V is required for operation. These scaling parameters are derived from constant field scaling whereby the electric field E_{eff} at the semiconductor surface directly under the gate, is kept constant to control so-called short channel effects [Taur and Ning, 1998].

According to table 1.1, not only does scaling allow a higher circuit density but also increased circuit switching speed and lower power consumption per MOSFET. Specifically, Moore's law states that circuit density doubles every two years and hence $\kappa \approx \sqrt{2}$ and is applied every twenty four months when a new generation of integrated circuits is manufactured for the first time. Scaling MOSFETs to the next generation means increasing the dopant concentration and shrinking the dimensions; however, this becomes

less and less trivial with each subsequent iteration. So-called roadblocks appear, such as the requirement of an impossibly shallow ion-implant and they require an innovative step to overcome (halo-doping in this case). Historically, roadblocks have always been overcome by applying new techniques to the same semiconductor material.

1.2 Silicon

Silicon has formed the basis of the semiconductor industry, almost since its birth, for these primary reasons:

1. It is abundant on the earth's surface.
2. It has a stable, naturally forming insulator: silicon dioxide, SiO_2.
3. It has an appropriate band-gap of 1.12 eV.

These factors combine to provide cheap platforms (1) on which to manufacture reliable (2 and 3) MOSFETs. Not only is the gate dielectric formed by simply oxidising the silicon surface but the gate contact is highly-doped poly-silicon. The source and drain contacts can be formed by doping the silicon or depositing and then annealing a metal to form a silicide. Thus a functional MOSFET can be almost entirely comprised of silicon and little else; however, very recently new materials have had to be introduced into these almost pure silicon structures due to a fundamental limitation. The silicon dioxide dielectric has been scaled to such a small thickness that a large leakage current flows through it due to quantum mechanical tunnelling of electrons from the channel.

This limitation is a very good example of a fundamental roadblock, as one of the dimensions of the MOSFET reached a length comparable to a fundamental physical process (the tunnelling length of an electron). A hafnium based compound is now used

in place of silicon dioxide as the gate dielectric, which has a greater dielectric constant (often denoted κ or K — hence high-k dielectric) and hence offers the same capacitance for a greater thickness, thus reducing quantum tunnelling to an acceptable level. This allows scaling to continue but now the oxide thickness t_{ox} is an equivalent thickness of SiO_2 and not a real physical dimension.

As silicon technology is scaled further, many more roadblocks will undoubtedly appear and most will be solved by adapting the current material system. However, some will inevitably require new materials and this could even mean replacing the silicon that comprises the active channel region of the MOSFET.

1.3 Germanium

Germanium offers higher bulk carrier mobilities than silicon meaning a greater current flow for less lateral applied electric field, which translates to a Ge MOSFET potentially requiring less supply voltage and providing faster switching speeds than the equivalent Si MOSFET. However, this slightly naive reasoning ignores the problems associated with germanium:

1. It is not abundant on the earth's surface.
2. It has no stable, naturally forming insulator.
3. It has a small band-gap of 0.66 eV.

Despite these problems, much interest has been shown in germanium recently, with a dedicated symposium at EMRS 2006. The problems listed can, in principal, be overcome by (1) using silicon as a platform for a thin germanium layer on which to fabricate devices. Germanium is easily integrated with silicon as both are of group IV

in the periodic table and form a crystalline alloy of any Si/Ge composition. (2) SiO_2 is no longer used as the gate dielectric in silicon MOSFETs and hence there is no fundamental reason why the techniques developed to deposit an alternative dielectric cannot be adapted to germanium. (3) germanium is likely to be introduced after the geometry of MOSFETs has changed to so-called thin body architecture and in these structures, no extra leakage currents will occur due to a reduced band-gap.

1.4 Current Investigation

In general, current germanium research mainly focuses on fabricating standard Ge MOSFETs, which show an enhancement over the silicon equivalent. This requires a high quality germanium starting platform and a reliable gate dielectric; along with other technologies taken for granted in silicon processing: reliable ion-implantations, chemical etches, metallisation, etc.

This thesis is a report on an experimental investigation of germanium MOSFETs and substrates — the focus is primarily on the former, with only chapter 6 containing results regarding the latter. Specifically, the germanium condensation technique, which is a promising method of obtaining germanium-on-insulator substrates. These will most likely be the platform for future Ge MOSFETs.

Germanium MOSFETs with a high-k gate dielectric, fabricated on two different types of germanium platforms, are investigated. One platform is a thick layer of germanium on a silicon substrate, with the electrical properties of pure bulk germanium (chapter 4). The other is a thin layer of germanium under compressive strain, which alters the band structure of the semiconductor (chapter 5). The electrical characteristics of the MOSFETs are investigated, with a particular focus on low temperature measurements (down to 4.2 K).

A description of the practical techniques employed to physically and electrically characterise semiconductor material and devices is given in chapter 3.

Chapter 2 introduces some important general aspects of MOSFETs: device operation, carrier scattering and gate dielectrics. The chapter continues with a detailed discussion of germanium, which covers: integration with silicon, band-structure and Ge MOSFETs.

We begin by briefly introducing some basic physics of semiconductors.

Chapter 2

Background to Germanium MOSFETs

2.1 Fundamental Semiconductor Physics

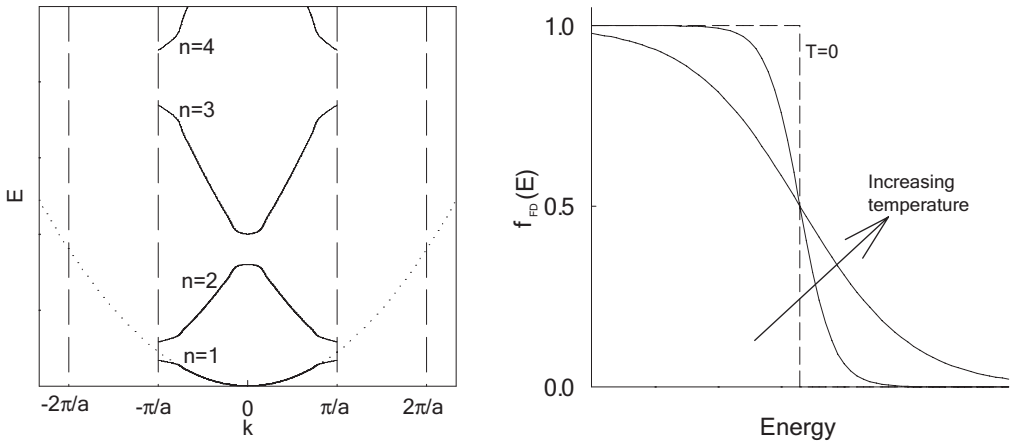
This description of basic semiconductor physics summarises that given in the first few chapters by Jaros [1989]. Treatment of a free electron in a bulk crystalline solid yields a series of equally separated, closely spaced, discrete allowed energy levels — each corresponding to a unique set of wave vectors $k_{x,y,z} = \frac{2\pi}{L} i_{x,y,z}$, where i are integers and L is a unit distance in real-space. By analogy to a classical particle:

$$E_k = \frac{k^2 \hbar^2}{2m} \tag{2.1}$$

where E_k is energy and m is mass. The density of states (number of electrons per energy level per volume of material) is:

$$\rho = \frac{\mathrm{d}N}{\mathrm{d}E} = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{\frac{3}{2}} E_k^{\frac{1}{2}} \tag{2.2}$$

Gaps are present in the continuum of allowed energy levels close to, and at, the



(a) (b)

Figure 2.1: The reduced representation of a semiconductor band-structure (a) with Brillouin zones (dashed lines) and standard E-k dispersion curve (dotted line). The Fermi-Dirac distribution function varying with temperature (b).

Brillouin zone boundaries (located at $k = \frac{n\pi}{a}$, where a is the length of a unit cell) due to interference of electrons interacting with the lattice. A reduced zone representation of the band structure is graphed by translating each segment of the E-k curve in nth Brillouin zone by an integer multiple of $\frac{2\pi}{a}$ into the first Brillouin zone. Bands are numbered from lowest to highest by n and hence any allowed energy is identified by two quantum numbers: $n = 1, 2, 3...$ and $-\frac{\pi}{a} < k < \frac{\pi}{a}$. A simplified reduced zone representation is shown in figure 2.1(a) but for real solids they are not this simple when all effects, such as crystal potential and symmetry, are considered.

We define the Fermi-energy E_F, corresponding to the Fermi-wave vectors of magnitude k_F, as the top of the distribution of energy states filled by electrons at $T = 0$. In a semiconductor this lies halfway between the valence band (the occupied band with the greatest energy) and the conduction band (the unoccupied band of the lowest energy). The valence band has no empty states at $T = 0$ and therefore electrons

are fixed and current unable to flow in the crystal.

At finite temperatures some electrons gain enough energy to occupy the conduction band, leaving behind empty states in the valence band (holes) that are treated as carriers of unit positive charge. Electrons in the conduction band can move freely as there are many available energy states and similarly, holes can move freely in the valence band. The occupancy of the conduction band by electrons and valence band by holes is obtained from the Fermi-Dirac distribution function (figure 2.1(b)):

$$f_{FD}(E) = \left(\exp\left(\frac{E - E_F}{k_B T}\right) + 1\right)^{-1} \tag{2.3}$$

This is approximated to the classical Boltzmann form $f_B(E)$ (it is assumed that $E_g >> k_B T$) and an integral performed over the density of states (equation 2.2) so that for electrons:

$$n = \int_{E_g}^{\infty} f_B^e(E)\rho^e(E)dE \tag{2.4}$$

This evaluates to

$$n = 2\left(\frac{m_e^* k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(\frac{E_F}{k_B T}\right) \exp\left(\frac{-E_g}{k_B T}\right) \tag{2.5}$$

A similar integral for holes gives:

$$p = 2\left(\frac{m_h^* k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(\frac{-E_F}{k_B T}\right) \tag{2.6}$$

The total density of carriers at a given temperature is given by the mass action law:

$$n_i^2 = np = 4\left(\frac{k_B T}{2\pi\hbar^2}\right)(m_e^* m_h^*)^{\frac{3}{2}} \exp\left(\frac{-E_g}{k_B T}\right) \tag{2.7}$$

Where the effective mass of the carriers m^* is obtained from:

$$m^* = \hbar^2\left(\frac{\partial^2 E_k}{\partial k^2}\right)^{-1} \tag{2.8}$$

Provided the E-k curve has a parabolic form, the effective mass of a carrier will be a constant and, observing the form of equation 2.1, can be expressed as a fraction of the electron mass in a vacuum. Band edges are normally approximated as parabolas for this reason; however, as one moves away from the conduction band minima and valence band maxima the deviation from parabolic form generally becomes larger. Consider the carriers in a semiconductor with constant effective masses at finite temperature. In a perfect semiconductor there should be no interactions of carriers with the lattice; however, in practice there are many mechanisms by which a carrier can interact and thus alter its momentum (discussed in section 2.4.2). The average time between these interactions is called the mean free time $\langle\langle\tau\rangle\rangle$.

If an electric field $\mathbf{F}$ is applied then carriers accelerate between the collisions. We now define the momentum relaxation time $\langle\langle\tau_k\rangle\rangle$ as the mean free time needed for a carrier to undergo enough collisions to lose all of its momentum. The velocity gained by the carrier between collisions, called the drift velocity, is (for an electron):

$$\mathbf{v}_e = -\left(\frac{q\,\langle\langle\tau_{\mathbf{k}}\rangle\rangle}{m_e^*}\right)\mathbf{F} \tag{2.9}$$

Note that we use q as the electron charge throughout this thesis, except in section 2.4.2 where we use e for the sake of convention.

We define electron mobiliy as:

$$\mu_e = \frac{q\,\langle\langle\tau_k\rangle\rangle}{m_e^*} \tag{2.10}$$

Similarly for holes:

$$\mathbf{v}_h = \mu_h\mathbf{F} \tag{2.11}$$

From these the drift currents are:

$$J_n^{Drift} = q\mu_e nF \tag{2.12}$$

$$J_p^{Drift} = q\mu_h pF \tag{2.13}$$

In an intrinsic semiconductor, electron concentration n and hole concentration p are exactly equal, as to excite an electron into the conduction band always leaves a hole in the valence band. Carrier concentration can be biased in favour of electrons or holes by introducing dopants into the semiconductor material. These either have one more electron (donors) than the semiconductor causing an excess of free electrons (n-type) or one fewer electron (acceptor) causing an excess of holes (p-type). We assume that dopants are introduced into semiconductors in concentrations low enough so that the band-structure of the semiconductor can be considered unaltered. Dopant levels are created just above valence band (p-type) or just below the conduction band (n-type), called shallow impurity levels.

At $T = 0$ the Fermi-level E_F is located at the additional shallow energy level but as the temperature increases the Fermi-level tends towards the centre of the band-gap where it is found in the case of an intrinsic semiconductor. This mid-gap value is now referred to as the intrinsic Fermi-level, E_i. The position of the Fermi-level E_F compared to the intrinsic Fermi-level E_i in a doped semiconductor is given by:

$$E_F - E_i = k_B T \ln\left(\frac{N_D}{n_i}\right) \tag{2.14}$$

in the case of n-type material with donor doping density N_D and

$$E_i - E_F = k_B T \ln\left(\frac{N_A}{n_i}\right) \tag{2.15}$$

in the case of p-type material with acceptor doping density N_A. We define n_i as the intrinsic carrier concentration, which decreases with decreasing temperature according to equation 2.7. This is given for several semiconductors including germanium in figure 2.2.

The interface between a region of n-type and p-type semiconductor, where the different Fermi-levels of the two regions align, is called a pn-junction. The spatially fixed

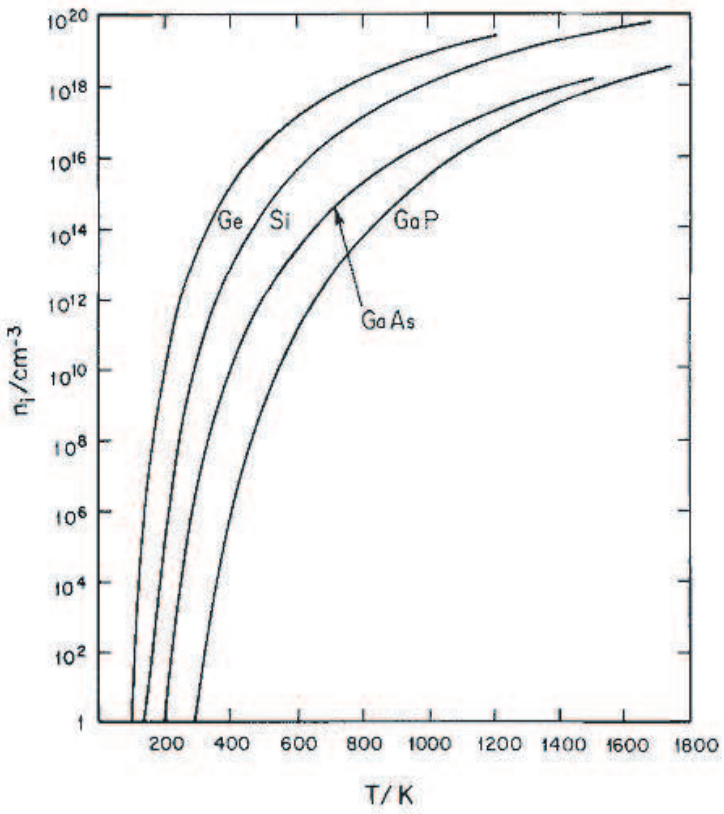


Figure 2.2: The intrinsic carrier concentration n_i as a function of temperature for several common semiconductor materials, reproduced from Thurmond [1975].

charges in the form of the donor and acceptor ions either side give rise to a potential across the junction — called the built-in potential. The electric field present due to the built-in potential causes a drift current across the junction without an external field being applied. Another current (the diffusion current) arises due to diffusion of electrons and holes down the respective concentration gradients. The total current across a pn-junction (in 1 dimension) is therefore $J_n + J_p$ where:

$$J_n = J_n^{Drift} + J_n^{Diff} = q\mu_e nF + qD_n \frac{\mathrm{d}n}{\mathrm{d}x} \tag{2.16}$$

$$J_p = J_p^{Drift} + J_p^{Diff} = q\mu_h pF - qD_p \frac{\mathrm{d}p}{\mathrm{d}x} \tag{2.17}$$

Where the x direction is defined as from the n-type to p-type material, which is in parallel to the applied field and measured current. In equilibrium the net hole and electron currents must both be 0 ($J_n, J_p = 0$). This allows the Einstein diffusion coefficients D_n and D_p to be obtained as:

$$D_n = \frac{k_B T}{q}\mu_e; D_p = \frac{k_B T}{q}\mu_h \tag{2.18}$$

2.2 The Metal-Oxide-Semiconductor Capacitor

Before attempting to understand MOSFET operation, it is best to begin with the metal oxide semiconductor (MOS) capacitor. Here we briefly introduce the MOS capacitor and define the depletion, inversion and accumulation surface conditions. The description given is after Taur and Ning [1998] but can be found in most applied solid state physics text books.

A MOS capacitor consists of a metal layer (the gate) atop a dielectric (the gate dielectric), which sits on a semiconductor surface the bulk of which can be doped n-type (n-MOS capacitor) or p-type (p-MOS capacitor). A voltage applied to the gate

will result in a potential drop across the dielectric and a change in the potential at the semiconductor surface.

$$V_g = -V_{ox} - \psi_s = -\frac{Q_s}{C_{ox}} - \psi_s \tag{2.19}$$

The potential across the dielectric V_{ox} is equal to the charge now present at the interface Q_s normalised by the capacitance of the dielectric C_{ox}. Further into the semiconductor, away from the semiconductor/dielectric interface, normal bulk potential is restored as the bands tend to their unperturbed state. For convenience, we define ψ_s as the perturbation of the intrinsic Fermi-level band potential at the surface relative to its bulk level. Note that the perturbations of the conduction band, valence band and intrinsic Fermi-level band are equal.

Before a voltage is applied to the gate, there will be some band bending at the surface already — necessary to align the Fermi-levels of the metal gate and the semiconductor. Therefore a voltage V_{fb} must be applied to the gate in order to negate this effect and restore the semiconductor surface to a state of no band bending, known as the flat band condition (figure 2.3). Flat-band voltage V_{fb} is also dependent on the non-ideal properties of real dielectrics, such as trapped charge, and this will be discussed later.

The majority of this thesis is concerned with p-MOSFETs, which are based on a n-type body, making the theory of n-MOS capacitor operation more relevant and hence the rest of this section is written in the context of n-MOS capacitors.

From the flat-band condition the semiconductor surface can be biased into depletion or accumulation (figure 2.4). Accumulation occurs when the band bending is downwards, giving rise to majority carriers at the surface (electrons) due to the conduction band tending towards the Fermi-level. Depletion occurs when band bending is upwards reducing the number of majority carriers at the interface as the conduction

band energy increases. As the electron density decreases, a fixed positive charge is left behind in the form of ionised donor nuclei and this is known as the depletion charge.

The band perturbation potential ψ as a function depth from the semiconductor surface z is approximated as a parabola with its vertex (the point where the bands are back to the bulk potential) at $z = W_d$ where $\psi(z = W_d) = 0$. At the surface ($z = 0$) the band perturbation potential must be equal to the surface potential $\psi(z = 0) = \psi_s$ and taking the inverse of this parabola gives the depletion layer width:

$$W_d = \sqrt{\frac{2\epsilon_0 \epsilon_{se} \psi_s}{qN_D}} \tag{2.20}$$

Where ϵ_{se} is the semiconductor relative permittivity and N_D is the ionised donor density, which is normally approximated to be the substrate doping density. The depletion charge sheet density is approximated by the depletion layer width multiplied by the donor ion density:

$$Q_d = qN_D W_d = \sqrt{2\epsilon_0 \epsilon_{se} qN_D \psi_s} \tag{2.21}$$

As the gate voltage is made increasingly more negative the depletion layer charge and width increase accordingly. The valence band begins to approach the Fermi-level and it becomes energetically favourable for holes to occupy the valence band at the surface. The process of holes populating the valence band at the surface in the n-type semiconductor is known as surface inversion. The thin sheet of charge comprised of holes at the surface is referred to as the inversion layer.

As the gate voltage is made further negative, a point will be reached where enough holes are present in the inversion layer to screen the semiconductor from any further band bending. This is called the strong inversion condition (figure 2.5) and any further applied gate voltage will generate more inversion charge, but not affect the surface potential nor the bands further into the solid. A common definition of

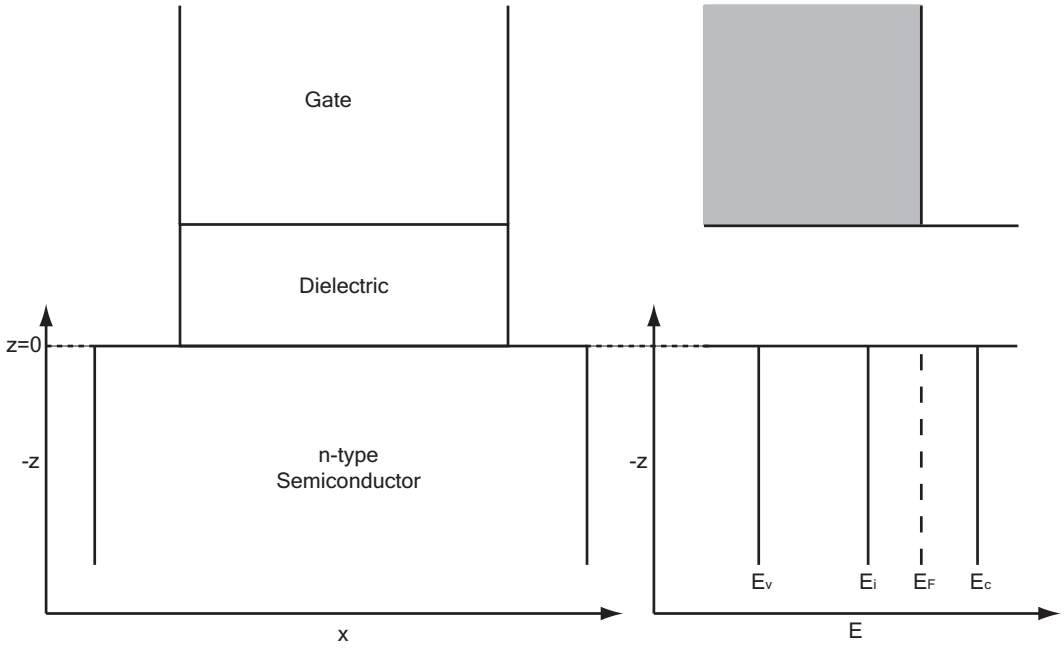


Figure 2.3: A n-MOS capacitor with corresponding band diagram at the flat-band condition with E_c, E_v, E_F, E_i denoting conduction band edge, valence band edge, Fermi-level and intrinsic Fermi-level respectively.

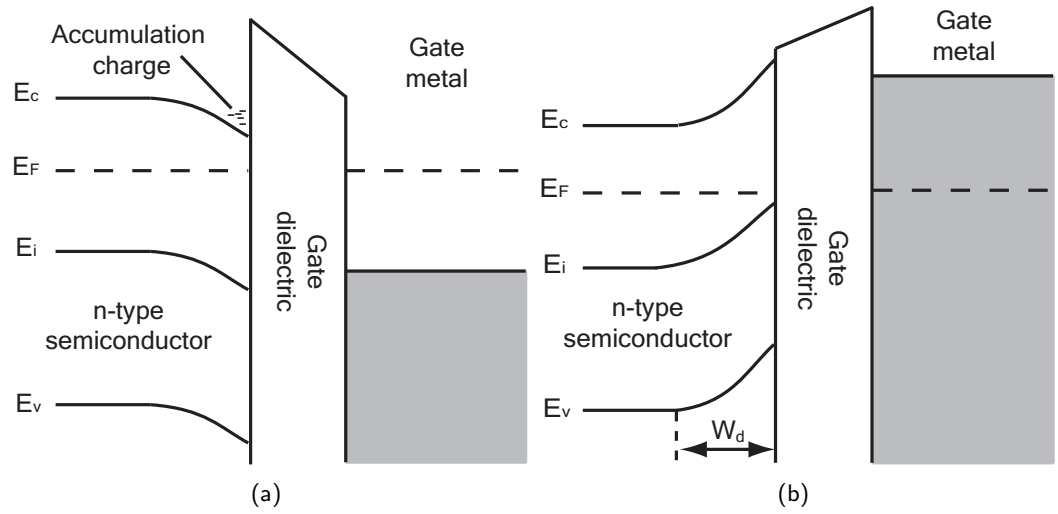


Figure 2.4: Accumulation (a) and depletion (b) conditions of a n-MOS capacitor.

the condition for strong inversion is when the inversion charge density is equal to the depletion charge density. This occurs when the surface potential is twice the bulk potential:

$$\psi_s = 2\phi_b = 2\frac{k_B T}{q}\ln\left(\frac{N_D}{n_i}\right) \tag{2.22}$$

Where the bulk potential $\phi_b = \frac{E_F - E_i}{q}$.

As the bands are no longer affected by a further negative increase in gate voltage, the depletion layer width and therefore depletion sheet charge density are fixed at maximum values:

$$W_{dm} = \sqrt{\frac{4\epsilon_0\epsilon_{se}k_B T\ln\left(\frac{N_D}{n_i}\right)}{q^2 N_D}} \tag{2.23}$$

$$Q_{dm} = \sqrt{4N_D\epsilon_0\epsilon_{se}k_B T\ln\left(\frac{N_D}{n_i}\right)} \tag{2.24}$$

2.3 The Metal-Oxide-Semiconductor-Field-Effect-Transistor

A MOSFET can be thought of as a modified MOS capacitor structure and hence its basic operation can be understood from the MOS capacitor surface conditions already given above. The description of MOSFET operation here does not consider the microscopic behaviour of carriers in the MOSFET channel and follows the description given by Taur and Ning [1998] to apply classical approximations in order to derive basic MOSFET operation.

A standard semiconductor MOSFET consists of a MOS capacitor with a highly doped source and drain region either side (figure 2.6), each with a respective metal contact. These relatively simple devices can be p-channel (n-doped body with highly p-doped source and drain) or n-channel (p-doped body with highly n-doped source and drain). This is a grossly simplified description as MOSFET design has evolved over the

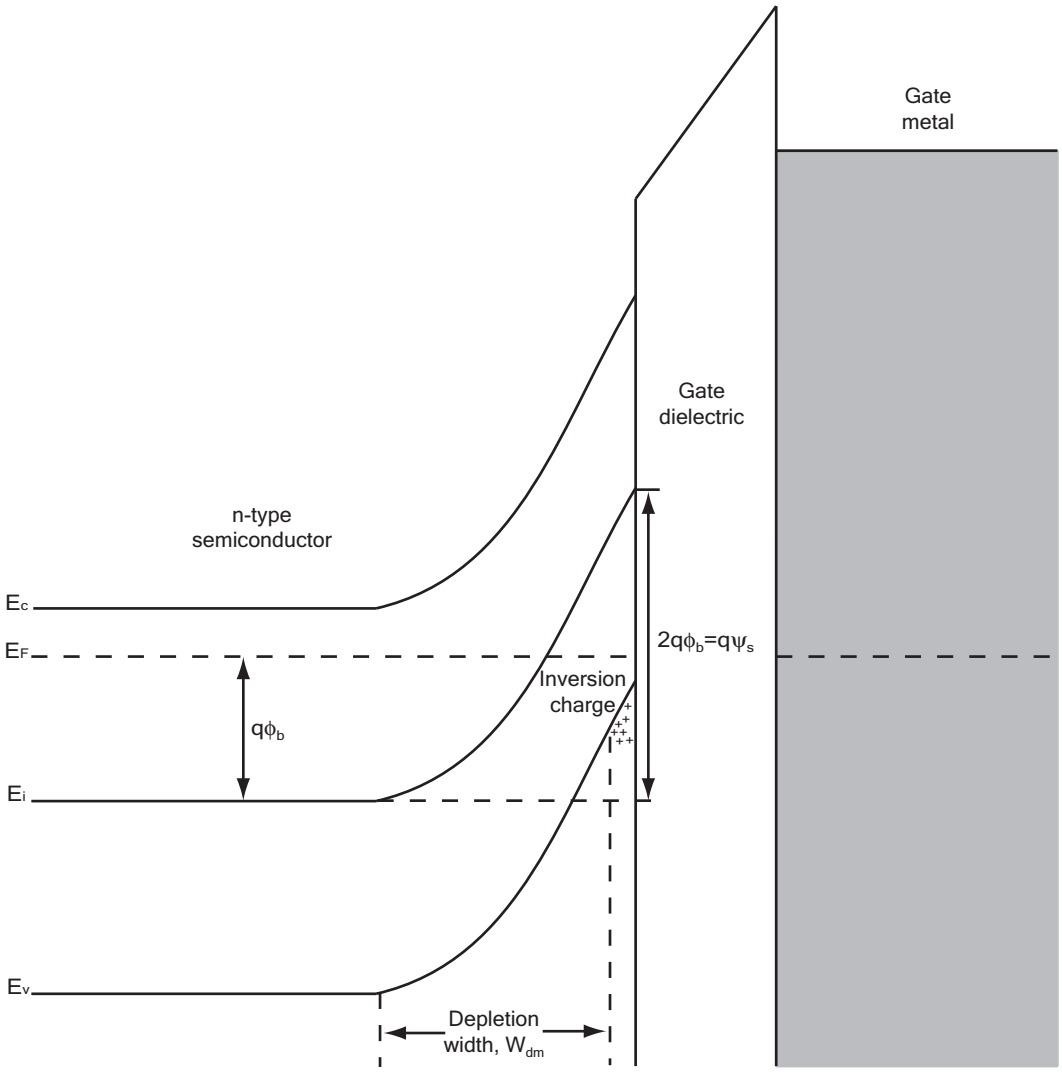


Figure 2.5: A n-MOS capacitor at the strong inversion condition, where at the surface bands are perturbed by $\psi_s = 2\phi_b$. Any further increase in gate voltage results in more inversion charge at the interface but no further band-bending.

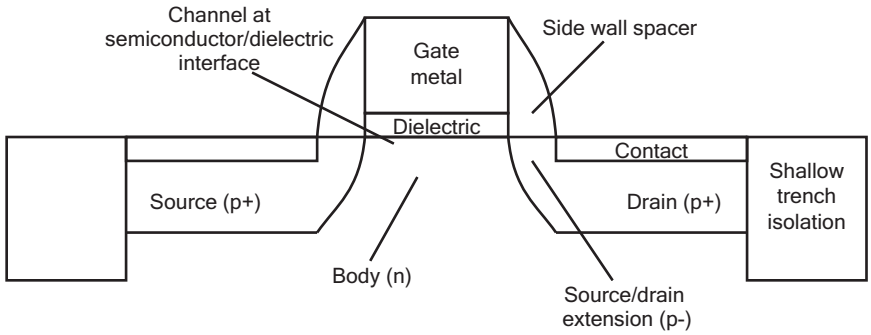


Figure 2.6: A p-MOSFET with key features labelled.

years to cope with higher levels of integration. One such modification of relevance is the source/drain extensions. These are lightly doped extensions of the source/drain regions to reduce dopant levels close to the channel region. This helps to control short channel effects [Taur and Ning, 1998], which are not discussed here.

The highly doped source and drain regions act as reservoirs for carriers. The p-doped drain is biased negatively for a p-channel MOSFET and the p-doped source neutral. This means that if they are able, holes will flow from source to drain, attracted by the negative bias. However, no holes can flow unless a sufficient number are generated in an inversion layer at the channel surface by applying the appropriate gate voltage, which will clearly be negative for a p-channel MOSFET. The band bending theory for a MOS capacitor structure given previously can now be applied to a p-channel MOSFET device. Throughout we assume that the voltage applied to the source contact V_s is zero.

2.3.1 The On-State

A MOSFET is said to be turned on once the strong inversion condition (equation 2.22) has been satisfied. Then mobile holes further populate the inversion layer with an increasingly negative applied gate voltage, allowing current to be mediated by holes from source to drain, referred to as the drain current I_d. The gate voltage at which

this criterion is met is called the threshold voltage V_t and is given by substituting the maximum depletion charge (equation 2.24) and $\psi_s = 2\phi_b$ condition into equation 2.19. Here we use the flat-band voltage to provide a known surface potential and ignore the non-ideal specifics of the gate stack.

$$V_t = V_{fb} - 2\phi_b - \frac{\sqrt{4\epsilon_0\epsilon_{se}qN_D\phi_b}}{C_{ox}} \tag{2.25}$$

As described above, the surface potential, and therefore the surface inversion charge density, is strictly a function of position along the length of the channel x and depth from the surface z. The MOSFET is assumed uniform across its width, which means, in general, that only a detailed analysis of the charge distribution into the semiconductor z and along the MOSFET channel x is required to formally describe the MOSFET operation. However, the charge sheet approximation invokes the assumption that the inversion charge is confined to a 2d sheet at the semiconductor surface and any charge distribution into the depth of the semiconductor z is ignored. A second assumption is that the surface potential ψ_s is equal all along the length of the channel.

The inversion charge Q_{inv} as a function along the length of channel is now given by:

$$Q_{inv}(x) = -C_{ox}(V_g - V_t - V_c(x)) \tag{2.26}$$

Where $V_c(x)$ is the potential at the surface of the channel region.

This layer of inversion charge is made up of mobile holes in the valence band and their velocity is related to the lateral electric field F by mobility μ which is assumed to be a constant:

$$v = -\mu\frac{\mathrm{d}V_c}{\mathrm{d}x} \tag{2.27}$$

Clearly if the number of mobile carriers and the velocity at which they traverse the length of the channel are known then the drain current can be found by multiplying

the two:

$$I_d = -C_{ox}\mu W\frac{\mathrm{d}V_c}{\mathrm{d}x}(-V_g + V_t - V_c(x)) \tag{2.28}$$

This can be integrated along the channel length. At the source $V_c(x = 0) = 0$ and at the drain $V_c(x = L) = V_d$. Current must be conserved along the channel length in the simplest case and thus all terms move outside of the integration, except the channel potential.

$$I_d = \mu\frac{W}{L}C_{ox}\left((V_g - V_t)V_d - \frac{V_d^2}{2}\right) \tag{2.29}$$

This equation captures the basic on-state behaviour of a MOSFET, which starts at $V_g \leq V_t$ and provided $0 > V_d > V_g - V_t$ the drain current increases linearly with gate voltage. This is called the linear region:

$$I_{dlin} = \mu\frac{W}{L}C_{ox}\left((V_g - V_t)V_d\right) \tag{2.30}$$

Note that when $V_d < V_g - V_t < 0$ pinch-off occurs — a portion of the channel towards the drain contains no inversion charge. Any further increase in drain voltage results in no increase in drain current but an increase in length of the highly resistive inversion charge free region. The MOSFET is said to be in saturation and the drain current, which is independent of drain voltage, is now given by:

$$I_{dsat} = \mu\frac{W}{2L}C_{ox}(V_g - V_t)^2 \tag{2.31}$$

Note that all electrical characterisation of MOSFET devices reported in this thesis was performed at low drain bias (50 mV).

2.3.2 The Off-State and Leakage Currents

Subthreshold Slope

When the applied gate voltage is more positive than the threshold voltage ($V_g > V_t$) the p-MOSFET is in the sub-threshold region where it is considered to be in the off-state.

The drain current is now limited by the emission of carriers over the barrier at the drain pn-junction to give:

$$I_d = \mu C_{ox} \frac{W}{L} \left(\left(1 + \frac{C_{dm}}{C_{ox}} \right) - 1 \right) \left(\frac{k_B T}{q} \right)^2 \exp \left(\frac{q(V_g - V_t)}{\left(1 + \frac{C_{dm}}{C_{ox}} \right) k_B T} \right) \tag{2.32}$$

where the maximum depletion layer capacitance C_{dm} can be approximated by:

$$C_{dm} \approx \frac{\epsilon_0 \epsilon_{se}}{W_{dm}} \tag{2.33}$$

The parameter S follows directly from this equation — the subthreshold slope (or subthreshold swing), which is a measure of how quickly a device turns off, specifically the change in gate voltage required to reduce the drain current by one decade:

$$S = \left(\frac{\mathrm{d}(\log_{10}(I_d))}{\mathrm{d}V_g} \right)^{-1} = 2.3 \frac{k_B T}{q} \left(1 + \frac{C_{dm}}{C_{ox}} \right) \tag{2.34}$$

Clearly, the lower the subthreshold slope, the faster the device is considered to switch off with respect to applied gate voltage. Note that according to the equation, subthreshold slope is dependent on temperature, depletion capacitance and little else; however, a capacitance due to interface traps (discussed in section 3.1.7) is in parallel to depletion capacitance and hence the subthreshold slope can be increased by these traps.

Leakage

Any current flowing through the device in the off-state is considered a leakage current and is undesirable. There are many components that add to the off-state current and that described by equation 2.32 is usually referred to as the subthreshold current. Generally this dominates in the immediate subthreshold region but once the current has been diminished sufficiently by a gate voltage further above V_t, it may be limited by various other leakage current components (figure 2.7) which are described below.

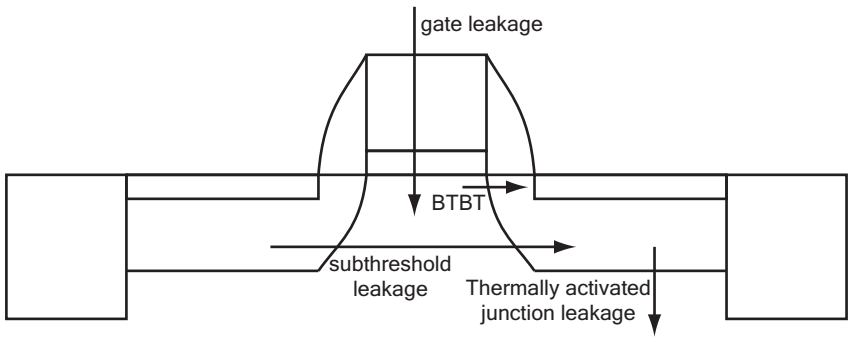


Figure 2.7: A MOSFET with leakage currents labelled.

A leakage current arises across the MOSFET drain pn-junction when an electron-hole pair is generated in the region by an electron tunnelling from the valence band to the conduction band (which leaves a hole in the valence band) [Chan et al., 1987]. In a p-MOSFET the hole is attracted to the negative drain bias and the electron to the neutral substrate. A carrier can tunnel from one band to the other by gaining enough energy to do so from random thermal motion (indirect tunnelling) or tunnelling across a pn-junction barrier with no change in energy required (direct tunnelling).

In general, indirect tunnelling has an associated activation energy and hence it is expected to vary exponentially with temperature:

$$I_{indirect} \propto T^2 \exp\left(-\frac{E_a}{k_B T}\right) \tag{2.35}$$

A Richardson plot ($\ln\left(\frac{I}{T^2}\right)$ vs T^{-1}) can be used to obtain the activation energy E_a which should be equal to the semiconductor band-gap E_g in the ideal case. However, energy states within the band-gap can reduce the activation energy. These deep impurity levels can take the form of dislocations, impurities or interface states (at the dielectric/semiconductor interface).

Direct tunnelling (band to band tunnelling or BTBT) currents generally only occur when the pn-junction has a large potential difference across it so that the con-

duction band on one side (the channel in a p-MOSFET) aligns with the valence band on the other (drain). This can occur due to highly doped source/drain regions (which is not relevant to this thesis) or arise with a large value of $(V_g - V_d)$ in the off-state, which is known as gate induced drain leakage (GIDL) [Chan et al., 1987]. This leakage current has been demonstrated to vary little with temperature [Rais et al., 1994]. Ge MOSFETs are expected to have increased leakage currents of the types described above due to germanium's smaller band-gap compared to silicon.

Gate leakage is also an important current occurring in both the on- and off-state. Much like the junction leakage mechanisms described, carriers can quantum-tunnel directly through the gate if it is thin enough. It is this problem that has bought about the introduction of high-k dielectrics, which are discussed in section 2.5.2. Carriers can also travel through the gate material via its conduction and valence bands (which should be at least 1 eV from the band edges of the semiconductor channel material to prevent this) or through those energy states in the dielectric that lie near the semiconductor conduction or valence band.

A well-designed, high-quality, relatively thick dielectric is therefore required to prevent significant gate leakage current. In the presence of significant gate leakage in the on-state it can be hard to assess MOSFET channel characteristics although corrections can be applied [Zeitzoff et al., 2003].

2.4 Carriers in the MOSFET Inversion Layer

Having described MOSFET operation on the device level, where inversion charge was approximated as a 2d-sheet at the interface and carrier mobility was treated as a constant, we now describe the carriers in the MOSFET channel without these approximations. This is necessary to investigate the physics of such a carrier, which can be broken into

two main parts: confinement at the gate dielectric interface and the scattering mechanisms arising from this.

Good descriptions of carrier confinement in the MOSFET channel are given by Jaros [1989] and Dobrovolsky and Litovchenko [1990]. The latter also describes scattering processes in the channel, but a more definitive description of this is given by Ando et al. [1982]. Although scattering in the MOSFET channel is a fairly specific case, a knowledge of general scattering processes in bulk solids, for example that given by Lundstrom [2000] and Butcher et al. [1986] may aid understanding.

2.4.1 Confinement

Both the confinement of carriers in the z direction (perpendicular to the channel) and the freedom of the carriers in the plane of the channel are considered separately and the respective allowed energy values calculated. These energies are then summed to give the total allowed energy of a carrier in the inversion layer E_T.

$$E_T = E_B + E_\perp + E_\parallel \tag{2.36}$$

Where E_B is the energy of band edge associated with the carriers. $E_\perp$ is the confinement contribution and $E_\parallel$ is the energy of a free carrier in a 2d semiconductor:

$$E_\parallel = \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right) = \frac{\hbar^2}{2m^*}k_\parallel^2 \tag{2.37}$$

Note that this is in the general case of a 2d-semiconductor (analogous to the 3d-case, equation 2.1) but the notation used is consistent with the inversion layer, where we are interested in the component of $\mathbf{k}$ parallel to the channel: $k_\parallel$. The density of states for a 2d-semiconductor is given as:

$$\rho_{2d} = \frac{\mathrm{d}N}{\mathrm{d}E} = \left(\frac{2m^*}{\hbar^2}\right)^{\frac{1}{2}} \frac{E^{-\frac{1}{2}}}{2}\frac{k_\parallel}{\pi} = \frac{m^*}{\pi\hbar^2} \tag{2.38}$$

In contrast to the bulk case (equation 2.2), the density of states in a 2d-semiconductor does not depend on the carrier energy.

In a MOSFET, confinement is normal to the channel (z direction) and gives allowed energy levels $E_\perp$. An appropriate shape for the potential well can be selected and the 1d-Schrodinger equation solved for a carrier in the well, in order to obtain the energy eigenvalues. A simple approximation for the confining potential at the interface is an infinitely deep triangular well. The assumption is that the potential decreases linearly from the interface:

$$V(z) = -ezE_{eff} \tag{2.39}$$

Where E_{eff} is the vertical effective field. Note that we use e as the electron charge here, as opposed q elsewhere.

We assume that the potential barrier at the gate dielectric/semiconductor interface is infinite and hence the potential is infinite for $z < 0$. The linear approximation of the bands allows the 1d-Schrodinger equation to be written:

$$\left(-\frac{\hbar^2}{2m^*}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + V(z)\right)\varphi(z) = E\varphi(z) \tag{2.40}$$

The solutions are Airy functions with Eigenvalues:

$$E_i = \left(\frac{\hbar^2}{2m^*}\right)^{\frac{1}{3}}\left(\frac{3\pi eE_{eff}}{2}\left(i+\frac{3}{4}\right)\right)^{\frac{2}{3}} \tag{2.41}$$

Where $i = 0, 1, 2....$ Note that the Airy functions are only valid for large i but can be corrected for the three lowest values by substituting the values of $(i + \frac{3}{4})$ for 0.7587, 1.7540 and 2.7575 for $i = 0, 1, 2$ respectively. This gives a series of discrete energy levels (figure 2.8(a)), the separation of which increases with increasing vertical effective field E_{eff} and decreasing carrier effective mass m^*. Combining this with $E_\parallel$ using equation 2.36 we obtain the energy as a function of k_x, k_y and i — a series of parabolas

with their minima increasing with i and all centred at $k_{\|} = 0$. These are known as sub-bands and at $T = 0$ carriers will occupy the lowest sub-band, filling it to larger values of energy until the minimum of the next sub-band has a lower value of energy than the remaining states in the lowest sub-band. Both bands will then fill in tandem until the third minimum is reached and so on. The density of states ρ_{2d} is constant with energy in each band and therefore the density of states function is just a series of steps (figure 2.8(b)), each increment marking the point where the next parabola begins to fill.

The most basic treatment of scattering is discussed in the next subsection where we confine all carriers into the lowest sub-band and ignore the higher energy states (the electrical quantum limit). This greatly simplifies calculations as all carriers are described by the same wave-function. In a MOSFET channel where high carrier densities are present, this condition is generally satisfied by cooling to low temperatures where typically the surface region is degenerate:

$$E_F >> k_B T \tag{2.42}$$

Clearly higher sub-bands will be occupied even at $T = 0$ if the carrier density is great enough, hence we must insist:

$$E_1 - E_F > 0 \tag{2.43}$$

Also note that in the fully degenerate electrical quantum limit the inversion charge sheet carrier density n_s will relate directly to the Fermi-wave vector and Fermi-energy. Integrating equation 2.38 with respect to energy and then rearranging:

$$E_F = \frac{\pi \hbar^2}{m^*} n_s \tag{2.44}$$

and substituting k_F for E_F:

$$k_F = (2\pi n_s)^{\frac{1}{2}} \tag{2.45}$$

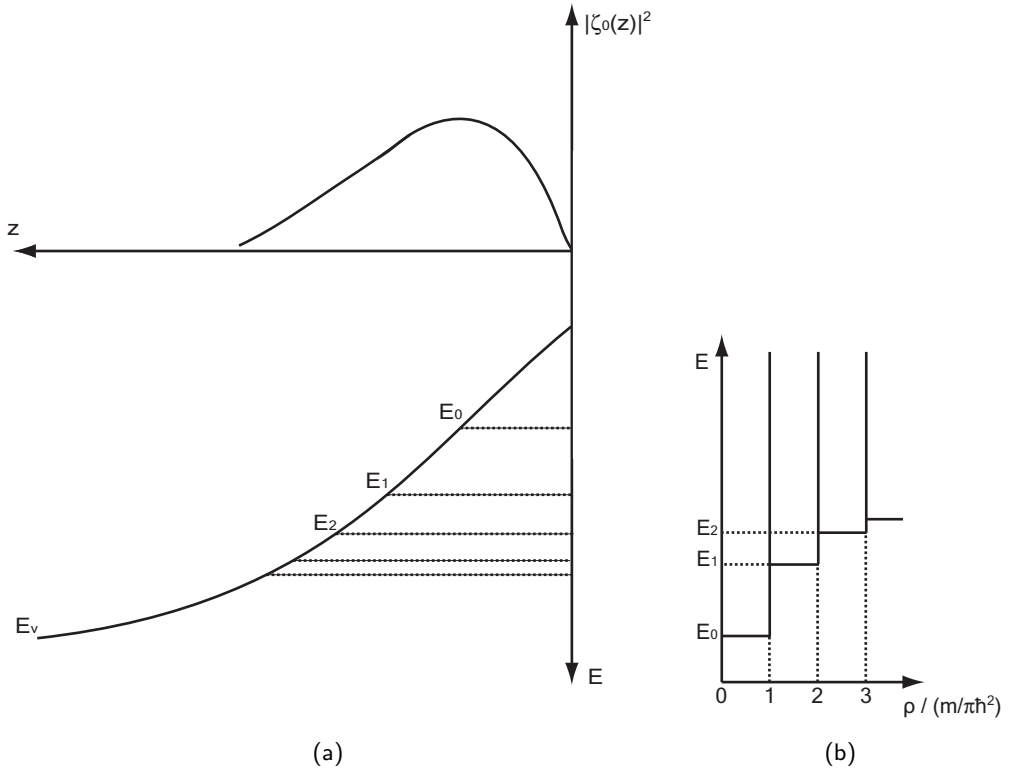


Figure 2.8: The distribution of inversion charge in the lowest sub-band (E_0) from the semiconductor/gate-dielectric interface, according to the Stern-Howard wave function ζ_0; along with energy levels of the lowest few sub-bands due to confinement (a). The 2d-density of states ρ increasing in steps as higher sub-bands become occupied (b). Adapted from Jaros [1989] and Dobrovolsky and Litovchenko [1990] respectively.

2.4.2 Scattering

The treatment of scattering here is adapted from that given by Horrell [2001], which is a detailed derivation of the expressions obtained in Gold and Dolgopolov [1986]. The latter is generally considered one of the better descriptions of scattering processes in the MOSFET inversion layer.

To simplify the algebra during scattering calculations, the Stern-Howard wave function is used in place of the $i = 0$ Airy function, given as:

$$\zeta_0(z) = \left(\frac{b^3}{2}\right)^{\frac{1}{2}} z\exp\left(-\frac{bz}{2}\right) \tag{2.46}$$

Parameter b is chosen so that the energy per carrier is minimised:

$$b = \left(\frac{48\pi m^* e^2}{\hbar^2 \epsilon_0 \epsilon_{se}}\right)^{\frac{1}{3}} \left(N_d + \frac{11}{32} n_s\right) \tag{2.47}$$

This is a self-consistent solution, where the triangular well is modified by the charge enclosed.

The scattering rate of a particle from wave vector $\mathbf{k}$ to $\mathbf{k}'$ is given by Fermi's golden rule [Lundstrom, 2000]:

$$P(\mathbf{k}, \mathbf{k}') = \left(\frac{2\pi}{\hbar}\right) \left|M(\mathbf{k}, \mathbf{k}')\right|^2 \delta(E_{\mathbf{k}} - E_{\mathbf{k}'}) \tag{2.48}$$

The delta function ensures that all transitions considered are elastic and $M(\mathbf{k}, \mathbf{k}')$ is the matrix element for the transition.

The scattering rate for a particle with wave-vector $\mathbf{k}$ is given by summing $P(\mathbf{k}, \mathbf{k}')$ over all $\mathbf{k}'$. We are interested in the momentum or transport relaxation time $\tau_{\mathbf{k}}$, hence each collision must be weighted by momentum change.

$$\frac{1}{\tau_{\mathbf{k}}} = \sum_{\mathbf{k}'} P(\mathbf{k}, \mathbf{k}')[1 - \cos(\theta_{\mathbf{k}\mathbf{k}'})] \tag{2.49}$$

Where $\theta_{\mathbf{k}\mathbf{k}'}$ is the angle between wave-vectors before and after the scattering event.

Note that the relaxation time appearing in equation 2.10 is averaged over energy:

$$\langle\langle \tau_{\mathbf{k}} \rangle\rangle = \frac{\langle E\tau_{\mathbf{k}} \rangle}{\langle E \rangle} \tag{2.50}$$

The averages of $(\tau_{\mathbf{k}}E)$ and E are calculated by integrating over all values of E in both cases; however, consider that scattering only ever occurs from an occupied state to an unoccupied state. In the degenerate case we cannot assume an infinite number of empty, nor filled states and hence we introduce a factor of $f_{FD}(E)(1 - f_{FD}(E))$ in the integrals to account for this. It can be shown that $f_{FD}(E)(1 - f_{FD}(E)) \propto \frac{\partial f_{FD}}{\partial E}$ [Butcher et al., 1986], hence:

$$\langle\langle \tau_{\mathbf{k}} \rangle\rangle = \frac{\int_0^\infty \tau_{\mathbf{k}} E \frac{\partial f_{FD}}{\partial E}}{\int_0^\infty E \frac{\partial f_{FD}}{\partial E}} \tag{2.51}$$

Clearly as $T \to 0$ the differential of the Fermi-Dirac function tends to a delta spike and all scattering takes place at an energy $E = E_F$, so that:

$$\langle\langle \tau_{\mathbf{k}} \rangle\rangle = \tau_{\mathbf{k}}(E_F) \tag{2.52}$$

Not only can we often achieve a degenerate semiconductor surface in the electrical quantum limit at low temperatures but also ensure that the majority of scattering occurs at the Fermi-energy.

Treating equation 2.49 as an integral by assuming a parabolic band and substituting wave-vector for energy using the dispersion relation:

$$\frac{1}{\tau_{\mathbf{k}}} = \frac{2\pi}{\hbar} \frac{A}{(2\pi)^2} 2 \int_0^\pi \mathrm{d}\theta \int_\infty^0 k' \mathrm{d}k' \frac{|M_{\mathbf{kk'}}|^2}{\epsilon_{\mathbf{kk'}}^2} [1 - \cos(\theta_{\mathbf{kk'}})] \delta \left(\frac{\hbar^2 k^2}{2m^*} - \frac{\hbar^2 k'^2}{2m^*} \right) \tag{2.53}$$

Where we introduce the dielectric screening function $\epsilon_{\mathbf{kk'}}$ which will be discussed later. Also note that $2\frac{A}{(2\pi)^2}$ is the number of states per unit area of k-space, introduced to normalise the density of scattering events.

This evaluates to:

$$\frac{1}{\tau_{\mathbf{k}}} = \frac{2\pi}{\hbar}\frac{A}{(2\pi)^2}2\frac{m^*}{\hbar^2}\int_0^{\pi} \mathrm{d}\theta \frac{|M_{\mathbf{kk'}}|^2}{\epsilon^2_{\mathbf{kk'}}}[1-\cos(\theta_{\mathbf{kk'}})] \tag{2.54}$$

To obtain the matrix transition element $|M_{\mathbf{kk'}}|$ we need a scattering potential $V(r,z)$ and appropriate wave function φ:

$$\varphi = \frac{1}{\sqrt{A}}\exp\left(i\mathbf{k}.\mathbf{r}\right)\zeta_0(z) \tag{2.55}$$

The scattering potential is dependent on specific scattering mechanism, which are considered later. The matrix transition element is given as:

$$M_{\mathbf{kk'}} = \langle\varphi_{\mathbf{k'}}\left|V(\mathbf{r},z)\right|\varphi_{\mathbf{k}}\rangle = \int \varphi^*_{\mathbf{k'}}V(\mathbf{r},z)\varphi_{\mathbf{k}}\mathrm{d}\mathbf{r}\mathrm{d}z \tag{2.56}$$

substituting equation 2.55 into equation 2.56 we obtain:

$$M_{\mathbf{kk'}} = \int \frac{1}{\sqrt{A}}\exp\left(-i\mathbf{k}.\mathbf{r}\right)\zeta_0(z)V(\mathbf{r},z)\frac{1}{\sqrt{A}}\exp\left(i\mathbf{k}.\mathbf{r}\right)\zeta_0(z)\mathrm{d}\mathbf{r}\mathrm{d}z \tag{2.57}$$

If $\mathbf{k'} = \mathbf{k}+\mathbf{q}$ then:

$$M_{\mathbf{kk'}} = \frac{1}{A}\int \exp\left(-i\mathbf{q}.\mathbf{r}\right)\zeta_0^2(z)V(\mathbf{r},z)\mathrm{d}\mathbf{r}\mathrm{d}z \tag{2.58}$$

Hence the matrix element $M_{\mathbf{kk'}}$ does not depend on the initial and final wave-vectors but only on the difference between them $\mathbf{q}$ and we substitute the matrix transition element for $M_{\mathbf{q}}$. Provided momentum is conserved (figure 2.9), i.e. $k = k'$:

$$q^2 = k^2 + k'^2 - 2kk'\cos(\theta_{\mathbf{kk'}}) \Rightarrow \frac{q^2}{2k^2} = 1-\cos(\theta_{\mathbf{kk'}}) \tag{2.59}$$

q has been used to denote the electron charge in previous sections; however, to be consistent with literature regarding scattering, we use it here to denote the magnitude of $\mathbf{q}$.

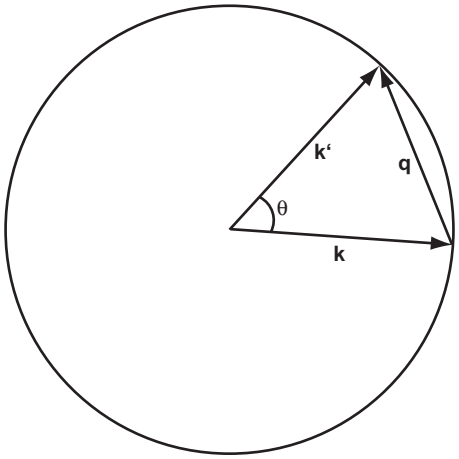


Figure 2.9: The geometry of a scattering event from $\mathbf{k}$ to $\mathbf{k}'$ through angle θ in k-space, reproduced from Horrell [2001].

Note that momentum is conserved in the case of surface roughness scattering and Coulomb scattering. Differentiating equation 2.59 gives:

$$\mathrm{d}q = k\cos(\frac{\theta_{\mathbf{kk}'}}{2})\mathrm{d}\theta = k\sqrt{1-\frac{q^2}{4k^2}}\mathrm{d}\theta \tag{2.60}$$

We now define the scattering function:

$$U_q^2 = A\left|M_q\right|^2 \tag{2.61}$$

which must be defined for each scattering mechanism.

Substituting for $M(\mathbf{k},\mathbf{k}')$ and θ terms in equation 2.54 using equation 2.61; equation 2.59 and equation 2.60 respectively:

$$\frac{1}{\tau_{\mathbf{k}}} = \frac{1}{2\pi\hbar E_k}\int_0^{2k}\frac{q^2\mathrm{d}q}{\sqrt{4k^2-q^2}}\frac{U_q^2}{\epsilon^2} \tag{2.62}$$

The dielectric function is given as:

$$\epsilon_q = 1+\frac{q_s}{q}F(q)\left[1-G(q)\right] \tag{2.63}$$

Where q_s is the Thomas-Fermi screening wave-number:

$$q_s = \frac{m^* e^2}{2\pi\epsilon_0\epsilon_{se}\hbar^2} \tag{2.64}$$

And where $F(q)$ is the form factor to take account of the finite extension of the electron wave function from the interface:

$$F(q) = \frac{\frac{1}{2}\left(1+\frac{\epsilon_{ox}}{\epsilon_{se}}\right)\left(1+\frac{9}{8}\frac{q}{b}+\frac{3}{8}\frac{q^2}{b^2}\right)}{\left(1+\frac{q}{b}\right)^3} + \frac{\frac{1}{2}\left(1-\frac{\epsilon_{ox}}{\epsilon_{se}}\right)}{\left(1+\frac{q}{b}\right)^6} \tag{2.65}$$

And where $G(q)$ is the local field correction in Hubbard's approximation:

$$G(q) = \frac{q}{\left(2q^2+k_F^2\right)^{1/2}} \tag{2.66}$$

All that remains is to select an appropriate form of U_q^2 for each scattering mechanism so that the relaxation time and hence carrier mobility can be obtained for each scattering mechanism. As phonon scattering is ignored in this thesis, on the grounds that phonons are frozen out at low temperatures, only Coulomb scattering and surface roughness scattering are considered here.

Coulomb Scattering

For a fixed charge magnitude e very close to, or in, the inversion layer the scattering potential will be proportional to e^2 and the inverse of the distance r [Stern and Howard, 1967]:

$$V(r,z) \propto \frac{e^2}{r} \tag{2.67}$$

Not considered in detail here is screening, where the high density of carriers in the inversion layer reduce the scattering potentials. Gold and Dolgopolov [1986] take this into account and obtain:

$$U_q^2 = n_{im}\left[\frac{2\pi e^2}{\epsilon q}\right]^2 F_i(q)^2 \tag{2.68}$$

Where n_{im} is the impurity charge sheet density and where $F_i(q)$ is a form factor:

$$F_i(q) = \frac{1}{(1+\frac{q}{b})^3} \tag{2.69}$$

The maximum of the integral in equation 2.62 occurs at $q = 2k_F$ and hence substituting U_q^2 for that given in equation 2.68 and assuming a single parabolic valence band (constant effective mass), Coulomb scattering limited mobility is given as [Emeleus et al., 1992]:

$$\mu_c = \frac{e\hbar n_s}{\pi m^* E_F n_{im}} \left(\frac{F(2k_F)[1-G(2k_F)] + (2k_F/q_s)}{F_i(2k_F)} \right)^2 \tag{2.70}$$

Surface Roughness Scattering

The interface between semiconductor and dielectric can be modelled as a gaussian function with perturbation height $\Delta(r)$ and correlation length Λ. The perturbing potential is proportional to the asperity height multiplied by the vertical effective field E_{eff} [Matsumoto and Uemura, 1974]:

$$V(\mathbf{r}, z) \propto E_{eff}\Delta(\mathbf{r}) \tag{2.71}$$

Transforming this to k-space and considering screening [Gold and Dolgopolov, 1986]:

$$U_q^2 = \pi\Delta^2\Lambda^2 q_s^2 E_F^2 \left(1 + \frac{2N_D}{n_s}\right)^2 \exp\left(\frac{-q^2\Lambda^2}{4}\right) \tag{2.72}$$

This is at a maximum when $q\Lambda = 2$ due to the exponential term and noting that the main contribution of the integral in equation 2.62 is still at $q = 2k_F$ (corresponding to $\theta = \pi$ according to equation 2.60), we get maximum scattering at $k_F\Lambda = 1$, i.e. where the Fermi-wave vector is equal to the inverse correlation length. In the $T = 0$ approximation Emeleus et al. [1992] gives:

$$\mu_{SR} = 0.48\frac{e\hbar}{m^*\Delta^2 k_F^2 E_F}\left[1 + \frac{2N_D}{n_s}\right]^{-2}\left[\frac{F_i(2k_F)}{F(2k_F)[1-G(2k_F)] + 2k_F/q_s}\right]^2 \tag{2.73}$$

2.5 MOS Dielectrics

Silicon can be oxidised to form SiO_2 with near perfect properties as a respective semiconductor/dielectric system for MOSFET applications. This is the result of years of research and development, which has revealed many types of defects associated with the material system. Here we introduce charged defects that affect device characteristics and briefly discuss high-k dielectrics.

2.5.1 Trapped Charges

A good description of the various charges discussed here is given by Taur and Ning [1998] and Schroder [2006] with the latter providing more detail. An extremely in-depth discussion of almost every aspect of Si/SiO_2 MOS capacitors and MOSFETs is given by Nicollian and Brews [1982].

In general, non-mobile charges are present in the gate dielectric in three forms: oxide fixed charge, oxide trapped charge and interface states. Comparing a device with these charges to one without would give an apparent applied gate voltage difference between the two:

$$\Delta V_g = \frac{Q_{ox}}{C_{ox}} \tag{2.74}$$

Where Q_{ox} is the sum of all contributions to charge trapped in the oxide.

Therefore, a MOSFET with a dielectric containing oxide charge would have parameters such as flat-band voltage and threshold voltage altered by an amount $-\Delta V_g$. However, this most basic observation does not account for the interaction of trapped charge with carriers in the channel and it is this interaction that defines the three oxide charge types discussed below.

Oxide Fixed Charge and Oxide Trapped Charge

Oxide trapped charge is present in the bulk of the dielectric, far enough from the channel to have no interaction with carriers in the inversion layer. A comparison of devices with and without oxide trapped charge would result in an observation of the gate voltage shift described above, but no other effects. This charge is still problematic as devices are usually designed to give, for example, a specific threshold voltage.

Oxide fixed charge is similar to oxide trapped charge except that its proximity to the channel means that it can scatter carriers via the Coulomb scattering mechanism. As a result, not only is a gate voltage shift observed but a reduction in carrier mobility μ also.

Interface States

Interface states are energy states distributed over the semiconductor band-gap, present at the dielectric/semiconductor interface. The density of interface states is usually denoted D_{it} with units cm^{-2} eV^{-1}. An electron will occupy any state below the Fermi-level and as the semiconductor surface bands are biased from accumulation, through depletion to strong inversion, the distribution of interface states sweeps the Fermi-level. Hence, the electron occupancy of the interface states is a function of surface potential and therefore applied gate voltage.

Depending on the nature of a state, when occupied by an electron it will either be neutral or have unit negative charge and when unoccupied it will either be neutral or have unit positive charge. We draw an analogy with dopants and define a state that is neutral when unoccupied by an electron as acceptor-like and one that is neutral when occupied by an electron as donor-like. Note that we discuss the occupancy of the state purely with respect to electrons although one can think of the states in terms of hole

occupancy also.

The accepted model for a silicon dioxide/silicon interface is that all states in the bottom half of the band-gap (below mid-gap) are donor-like and all states in the top half of the band-gap (above mid-gap) are acceptor-like, with the intrinsic Fermi-level of silicon approximately marking the boundary between the two types of state. If the bands are biased so that the intrinsic Fermi-level is aligned with the Fermi-level (the mid-gap condition — met at $V_g = V_{midgap}$) then all interface states will be neutral. In the case of a p-MOSFET, if we bias the gate from this condition towards strong inversion ($V_g < V_{midgap}$) the donor-like states in the lower half of the band-gap empty of electrons as they pass the Fermi-level and become positive (figure 2.10). Conversely, if we bias towards flat-band condition ($V_g > V_{midgap}$) then acceptor-like states gain an electron and become negatively charged.

The number of charged interface states per unit area $N_{it}(V_g)$ is the integral of $D_{it}(V_g)$ from mid-gap to V_g:

$$N_{it}(V_g) = \int_{V_g=V_{mg}} D_{it}(V_g)\mathrm{d}V_g \tag{2.75}$$

Unlike the static oxide charge types discussed above, the effect on device parameters is non-linear and depends on not only the number of interface states but also the distribution over the band-gap. Charged interface states are a source of scattering sites for carriers in the inversion layer due to their close proximity. With respect to V_{mg}, a larger gate voltage must be applied to obtain the same surface potential in a device with a higher density of interface states. A very high density of interface states can lead to Fermi-level pinning — where a maximum value of surface potential is reached before the expected value at strong inversion of $\psi_s = 2\phi_b$. Any further increase of applied gate voltage is balanced by interface states gaining charge instead of population of the inversion layer/increased surface band potential. Note that if the strong inversion

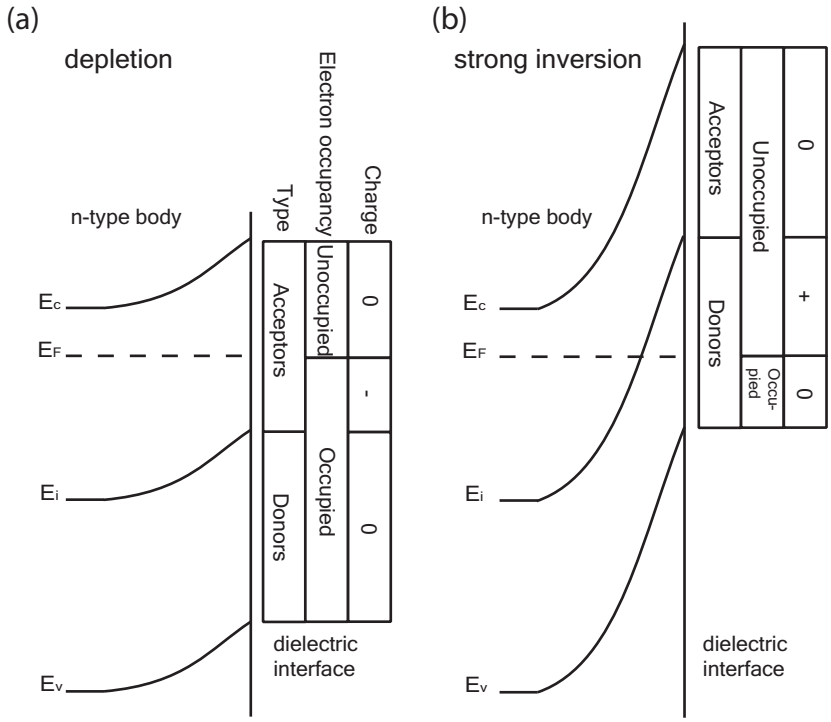


Figure 2.10: The donor/acceptor model for interface states in a p-MOSFET for depletion (a) and inversion (b) conditions. Adapted from Schroder [2006]

condition is successfully reached, we do not expect more interface states to be occupied as no further band bending occurs.

2.5.2 High-k Dielectrics

A good review of high-k dielectrics for application to silicon MOSFETs is given by Robertson [2006], which we briefly summarise here. Note that in this thesis ϵ with an appropriate subscript is generally used to denote relative permittivity. The "k" in "high-k dielectric" comes from an alternative notations of the relative permittivity K, k or κ.

According to table 1.1, the oxide thickness t_{ox} is scaled in each generation of MOSFET. After forty years of Moore's Law, this has led to a 1.2 nm thick silicon dioxide gate dielectric at the 90 nm node. At this thickness, direct quantum tunnelling of carriers through the oxide contribute significantly to the gate leakage current. This current exponentially increases with decreasing oxide thickness and an oxide thickness of 1.2 nm is the practical limit where more than 25% of circuit power is consumed by gate leakage [Derbyshire, 2007].

The solution has long been known and requires the substitution of silicon dioxide for a material with a greater dielectric constant, thereby allowing the same capacitance to the channel with a thicker gate dielectric and hence reduced tunnelling of carriers through the gate dielectric. One can see that if the relative permittivity of the dielectric is increased, then the thickness can proportionally increase to give the same capacitance per unit area.

$$C_{ox} = \frac{\epsilon_{ox}\epsilon_0}{t_{ox}} \tag{2.76}$$

It is normal to express the capacitance of a high-k dielectric as the equivalent thickness of silicon dioxide offering the same gate capacitance. The slightly naive version

is the equivalent oxide thickness t_{EOT}:

$$t_{EOT} = \frac{\epsilon_{SiO2}}{\epsilon_{hk}} t_{hk} \tag{2.77}$$

Where ϵ_{SiO2} is the relative permittivity of silicon dioxide and t_{hk} is the physical thickness of the high-k dielectric with relative permittivity ϵ_{hk}. However, consider that one must account for the fact that the inversion layer is not confined exactly at the interface, giving rise to a capacitance contribution in series from the semiconductor at the surface. We normalise this to an equivalent thickness of SiO_2 t_{se} so that it may be added directly to the equivalent oxide thickness to give the capacitive equivalent thickness:

$$t_{CET} = t_{EOT} + t_{se} \tag{2.78}$$

t_{CET} is the equivalent thickness of silicon dioxide one expects to directly calculate from a measurement of the oxide capacitance C_{ox} of a high-k gate dielectric MOSFET/MOS capacitor.

Often the gate metal comprises of highly-doped poly-silicon, which can deplete during normal MOSFET operation and adds another contribution to t_{CET}. Poly-Si deposition on high-k dielectrics results in a poor interface between the two and leads to Fermi-level pinning of the poly-Si. This results in a significant, unwanted threshold voltage shift of the MOSFET. In addition, the highly polarisable bonds in high-k dielectrics resonate with charge in the poly-Si gate, giving rise to optical phonons that scatter carriers in the channel. If a real metal gate is employed then this resonance does not occur due to the higher charge sheet density at a metal surface $\approx 10^{20}$ cm^{-2} compared to $\approx 10^{18}$ cm^{-2} for degenerately doped poly-Si. Note that all devices reported in this thesis have a real metal contact as opposed to poly-Si and hence these effects should not be present.

IBM and Intel almost simultaneously announced hafnium-based high-k gate dielectrics entering production at the 45 nm node in January 2007 [Singer, 2007]. A hafnium compound was chosen because it has a dielectric constant large enough to allow years of further scaling, is chemically stable in contact with silicon, is able to withstand the temperatures associated with silicon processing, acts as an insulator by having band offsets > 1 eV with the silicon band edges, forms a good interface with silicon and contains few electrically active defects. Note that the exact chemical formula of the gate dielectric and the associated fabrication techniques are closely guarded industry secrets but hafnium-oxide is well documented to fit these criteria, with a relative permittivity of $\approx$25 [Robertson, 2006].

2.6 Germanium

This section describes the physical and electrical properties of germanium and explains why it is considered a potential replacement for silicon as the channel material in future generations of MOSFET devices.

2.6.1 Substrates

Germanium substrates that are comparable to defect free silicon wafers are required in order to demonstrate germanium MOSFETs that exceed the electrical performance of their silicon counterparts. We begin by looking at crystalline layers made of a combination of silicon and germanium.

SiGe Heterostructures

Silicon and germanium are completely miscible — forming a stable alloy at any compositional fraction $(1 - x)$ and x $(0 < x < 1)$ respectively, usually written $Si_{1-x}Ge_x$.

A large body of publications regarding the electrical and physical properties of these alloy structures is available, including many reviews: for example that given by Schaffler [1997], Fischetti and Laux [1996] and Paul [2004].

A Si/Ge alloy is obtained by growth of SiGe on a standard (100) silicon wafer, either using Molecular Beam Epitaxy (MBE) or the more industry applicable chemical vapour deposition (CVD). The former allows a far greater parameter space to be explored in terms of temperature, growth rate and composition; the latter can mass produce wafers, but growth parameters are not independent due to the reliance on a controlled chemical reaction at the wafer surface, which depends on temperature, etc.

A bulk germanium crystal has a lattice constant of 5.61 Å, compared to 5.43 Å of silicon and the lattice constant of $Si_{1-x}Ge_x$ increases approximately linearly with x between the extremes of $x = 0$ and $x = 1$. A SiGe layer lattice matched to the silicon substrate on which it is grown will therefore be compressively strained in the plane of the wafer — known as a biaxially strained pseudomorphic layer. As the layer thickness increases, its strain energy will also increase, until it is energetically favourable for dislocations to form, which plastically relax the layer. These are called misfit dislocations and reside along the interface between the SiGe and the Si substrate. A dislocation cannot terminate in the bulk of a crystal and hence these are connected to the free top surface of the SiGe layer by threading arms, which do not themselves relieve strain.

The higher the Ge fraction of a SiGe layer is, the greater its lattice mismatch with the silicon substrate and hence the greater the strain energy for the given SiGe layer thickness. This reduces the thickness at which misfit dislocations form — called the critical thickness t_c. This has been calculated as a function of x (figure 2.11) by considering the strain force driving a threading dislocation balanced with the force required for it to glide [Matthews and Blakeslee, 1974]. As a pair of threading dislocations glide, the

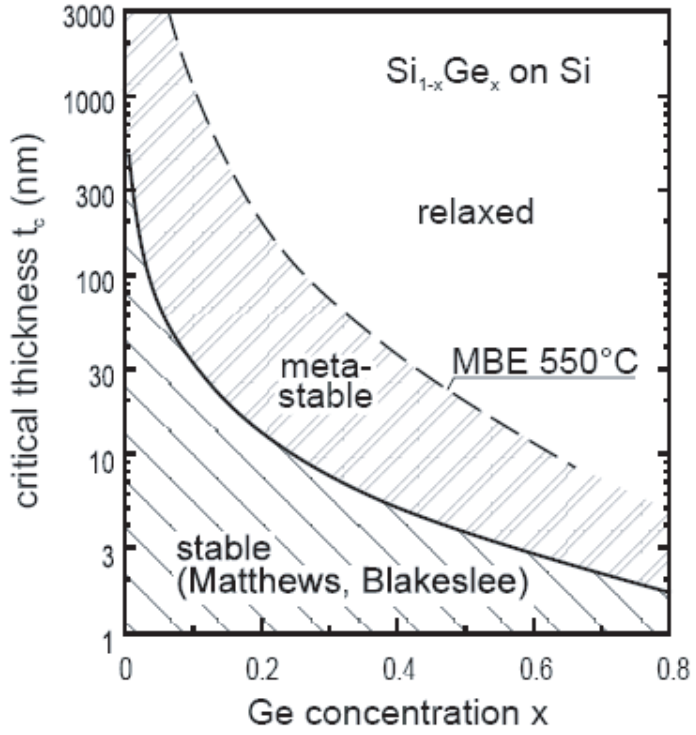


Figure 2.11: The critical thickness of a SiGe layer grown on bulk silicon (100) including the metastable region, reproduced from Schaffler [1997].

misfit dislocation connecting them along the Si/SiGe interface increases in length, thus further relaxing the SiGe layer. This particular derivation of t_c assumes that threading dislocations are already present in the SiGe layer, for example nucleated by defects during growth.

A larger value for the critical thickness is obtained by considering the activation energy required to nucleate a pair of threading dislocation from the surface, which ultimately reach the SiGe/Si interface and glide apart. Ensuring that the initial wafer surface and growth chamber remain free of particles allows a layer free of threading dislocations to be grown. A low growth temperature prevents threading dislocations from nucleating at the surface during growth and hence a higher critical thickness can

be obtained — called the metastable limit [People and Bean, 1985]. Any subsequent anneals sufficiently high in temperature will nucleate a significant number of threading dislocations and the layer will relax.

Although one can fabricate MOSFETs on pseudomorphic SiGe layers, they are more widely used as a platform for a silicon layer on which MOSFETs are fabricated. Allowing the SiGe to relax before the top Si layer is grown means that it will be biaxially tensile strained as it will lattice match to the underlying relaxed SiGe. Strain is maintained in the top Si-layer by ensuring that it does not exceed its own associated critical thickness [Samavedam et al., 1999]. The relaxed SiGe platform is often referred to as a virtual substrate and ideally is completely relaxed with few threading dislocations. Optimising virtual substrates to meet these requirements is the subject of much research.

Bulk Germanium

Due to limited world supplies of germanium, it is uneconomic to mass produce pure germanium substrates in the same vast quantities as silicon substrates. In addition, despite germanium's similarity to silicon, Ge substrates are far more difficult to handle, mainly due to their higher density and decreased mechanical strength [Vanhellemont and Simoen, 2007]. Combined with a lower melting temperature, these properties make the successful fabrication of devices on a Ge-substrate difficult. Therefore, if germanium is to become an industry acceptable semiconductor material, it will need to be implemented as a thin layer on a silicon substrate. Fortunately, silicon and germanium integrate well and this is certainly one reason for the renewed interest in germanium as an advanced channel material as opposed to other high mobility semiconductors.

An obvious method of obtaining a bulk-like germanium layer is to simply grow

germanium directly onto a silicon substrate and exceed the critical thickness by several microns [Nayfeh et al., 2004]. This results in a fully relaxed germanium layer far thicker than the vertical dimensions of a typical MOSFET and hence the device should behave no differently than if it were fabricated on a pure germanium wafer. Indeed high-performance Ge-devices have been demonstrated on substrates of this sort, for example those reported by Zimmerman et al. [2006] and Neyfeh et al. [2005]. However, although in the latter an effort is made to reduce the number of threading dislocations at the surface, generally leakage current caused by these is a major problem, as demonstrated by Nicholas et al. [2007b]. It is not entirely clear whether there is a significant contribution from leakage currents due to the small band-gap of germanium or whether high leakage is entirely due to high dislocation densities in the Ge layers. Threading density in the active Ge layer can be reduced by the use of a virtual substrate and has the additional advantage of allowing a strained Ge (s-Ge) layer [Nicholas et al., 2007a].

Germanium-on-Insulator

To enable MOSFET scaling into the deep sub-100 nm nodes, it is predicted that silicon-on-insulator (SOI) substrates will replace bulk silicon substrates [Current et al., 2000]. These consist of a top silicon layer, with an insulating layer directly under this, typically silicon dioxide and hence referred to as the buried oxide layer or BOX. This can range in thickness from 10's of nm to 10's of μm. The whole structure sits on a standard silicon substrate.

It is ultra-thin SOI, where the top silicon layer is 10 nm–50 nm thick, that enables the greatest MOSFET performance advantage over those on bulk-Si. As the junction depths of MOSFETs are scaled more shallow with each generation (see table 1.1), it becomes more difficult to activate the source/drain dopants with an anneal whilst

simultaneously preventing dopants from diffusing, resulting in an unwanted increase in junction depth. The main advantage of ultra-thin SOI is that the junction depths are automatically limited by the BOX (the resulting device is called a fully depleted SOI MOSFET as there is no part of the channel material that is not depleted when the device is turned on). The buried oxide also enables the potential implementation of a second gate stack below the MOSFET channel giving twice the gate capacitance for no increase in the MOSFET area. Thinner SOI top layers (5 nm–10 nm) will enable further scaling to so-called thin body MOSFET architecture, which offer further enhancements. SOI also has the advantage of better electrically isolating individual MOSFETs from each other and the substrate, reducing leakage current.

In light of the above, germanium is particularly suited to this type of substrate as it has the following issues:

- Dopants generally diffuse more in germanium than in silicon.
- Leakage currents are potentially greater due to germanium's smaller band-gap.
- Germanium needs to be integrated onto a silicon-substrate.
- Germanium will be implemented (if at all) in the deep sub-100 nm regime where thin-body architecture is required.

Germanium-on-insulator [Taraschi et al., 2004] can be obtained in two ways: wafer bonding or the germanium condensation technique. The former involves oxidising a standard silicon wafer and bonding this to a germanium layer grown on a silicon substrate (the donor substrate). The bonding process itself involves applying pressure to the wafers along with an anneal. Once bonding is complete, the large amount of excess material above the germanium layer is removed by various etches and proprietary techniques such as smart-cut [Bruel, 1995], which allow a smooth surface to be obtained.

The smart-cut process involves the implantation of hydrogen ions just below the active layer in the donor substrate. An anneal after bonding causes the H^+ ions to form H_2 gas bubbles, which break the wafer in the plane to leave only a very small amount of material above the active layer. This enables a smooth surface to be obtained by a selective chemical etch in addition to the donor-substrate being recovered for re-use, hence thin on-insulator substrates can be obtained from a single starting substrate.

Silicon-on-insulator is routinely obtained via bonding and as a sequence of SiGe layers is required on the donor substrate for selective chemical etching, the technique can, in principle, be easily adapted to any top surface layer, including pure germanium that could be strained or relaxed. Although thin SOI substrates are commercially available, problems of applying the technique to other top layers, such as germanium, arise from increased surface roughness and reduced thermal stability [Taraschi et al., 2004; Bedell et al., 2006].

The germanium condensation technique side-steps these problems by using thin SOI as a starting substrate, on which a low composition, relatively thick SiGe layer is grown (figure 2.12). An oxidation is then performed, where oxygen preferentially bonds to silicon rather than germanium to form SiO_2 [Paine et al., 1991]. As silicon is removed from the SiGe layer by an increasing thickness of oxide, germanium is rejected at the oxidising interface into the substrate and the overall Ge composition of the SiGe layer gradually increases. The raised temperature during oxidation allows germanium to diffuse down towards the BOX while simultaneously silicon diffuses up to the oxidising interface. One can stop the oxidation at any point to obtain a specific composition of SiGe-on-insulator (SGOI), but after a sufficiently long oxidation all of the silicon will be removed and the resulting substrate is pure germanium-on-insulator [Nakaharai et al., 2003].

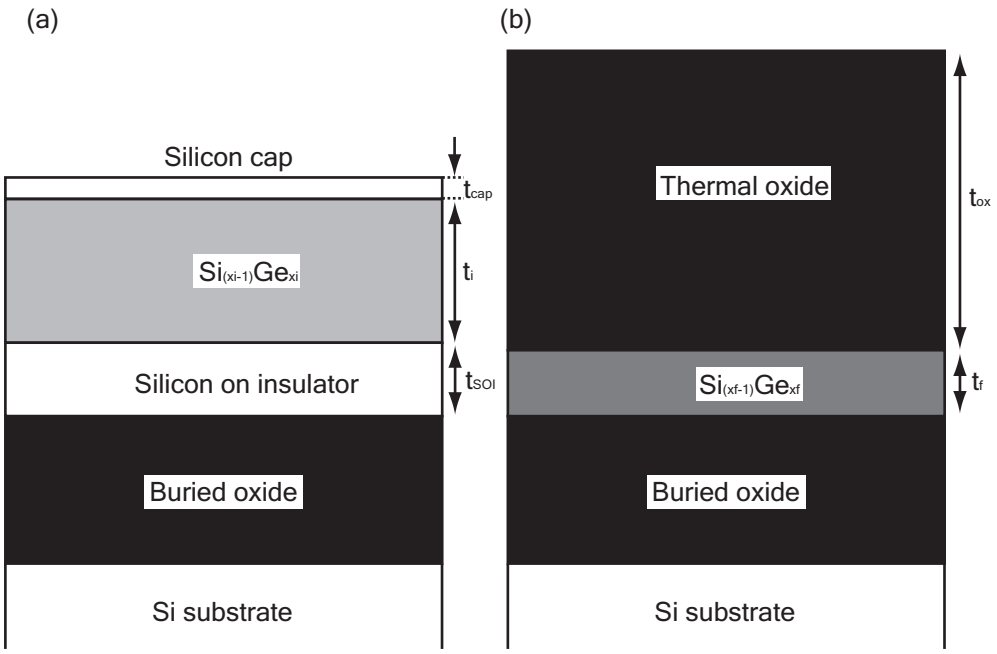


Figure 2.12: The initial (a) and final (b) germanium condensation layers.

There is a large parameter space to explore with regards to the germanium condensation technique: the initial layer thicknesses, the initial layer composition(s), the final layer thicknesses, the final SiGe layer composition, the oxidation duration(s) and oxidation temperature(s). These are related by the following equations:

$$t_i x_i = t_f x_f = t'_i x'_i \tag{2.79}$$

Where t_i and x_i are the initial thickness and composition of the SiGe layer and the same parameters with subscript f denote final parameters. We define the primed values for convenience, which is the total thickness of all the initial semiconductor layers t'_i and germanium composition normalised over this thickness x'_i:

$$t'_i = t_{SOI} + t_{cap} + t_i \tag{2.80}$$

$$x'_i = \frac{x_i t_i}{t'_i} \tag{2.81}$$

Where in addition to the silicon-on-insulator thickness t_{SOI}, a thin silicon cap of thickness t_{cap} is usually present. The total oxide thickness at the end of the Ge-condensation process t_{ox} is given by:

$$t_{ox} = \frac{1}{0.44}(t_i' - t_f) \tag{2.82}$$

Each single unit thickness of oxide consumes 0.44 unit thicknesses of silicon [Sze, 1985]. The oxidation time and temperature are related to the oxide thickness by the Deal-Grove model [Deal and Grove, 1965], which can be used as an estimate of the oxidation time/temperature required.

$$t_{ox} = \frac{A}{2}\left(\sqrt{1 + \frac{4B}{A^2}\tau} - 1\right) \tag{2.83}$$

Where τ is the oxidation time $\frac{B}{A}$ and B are the linear and parabolic rate constants of the form $D_0 \exp(\frac{-E_a}{kT})$. D_0 has a value of 7.8×10^6 $\mu\mathrm{m\ hr}^{-1}$ and 665×10^6 $\mu\mathrm{m\ hr}^{-1}$ for the linear and parabolic constants respectively and E_a has values of 2.01 eV and 1.21 eV respectively. Due to variation between different oxidation furnaces and the small tolerance of the thicknesses required for the Ge-condensation process, it is not normally practical to assume oxidation times will match that predicted by the Deal-Grove model, although it can provide a useful initial estimate of the times required. The oxide thickness is usually obtained by calibrating a furnace with a number of silicon test samples.

Having related the initial and final state variables, we now look at the process dynamics and associated design constraints. The most important of these is that the oxidation temperature should not exceed the melting temperature of any of the layers, which decreases with increasing germanium composition, as shown in figure 2.13. Any oxidations performed should be at a temperature below the lower line in this figure (the solidus line). An anneal temperature in the liquidus-solidus region results in inhomoge-

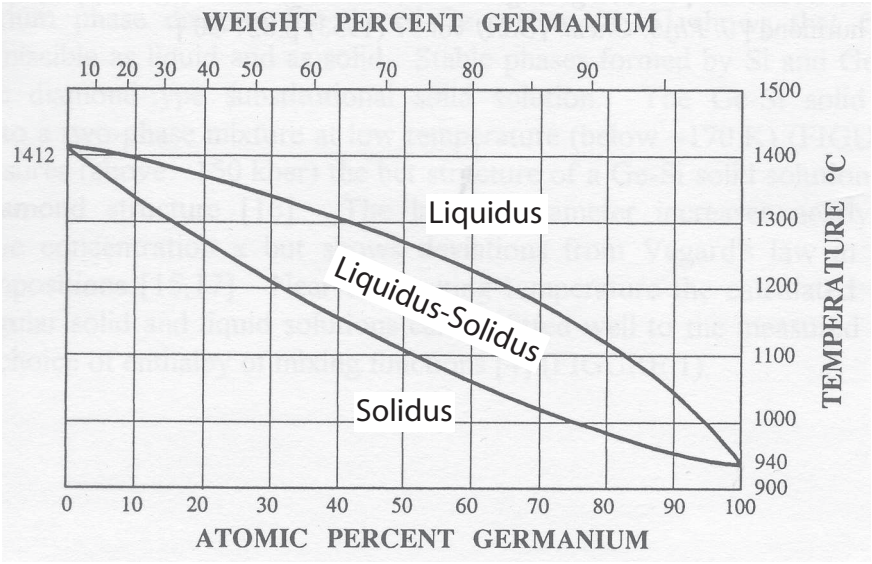


Figure 2.13: The liquidus-solidus phase diagram for SiGe alloy, reproduced from Kasper [1995].

neous layers and in the liquidus region results in large surface roughness [Usami et al., 2003].

Dry-oxidation of Si/SiGe progresses more slowly than wet-oxidation due to water catalysing the latter reaction and hence that the rate constants for dry oxidation are lower. It is this slower oxidation rate that allows a sufficient supply of silicon to the oxidising interface. Oxygen preferentially bonds with silicon rather than germanium to form SiO_2 and hence germanium is rejected from the oxide into the semiconductor. Note that Ge cannot occupy the oxide interstitially but only substitute a silicon atom [Minke and Jackson, 2004] and hence if all oxygen bonds to silicon, germanium must be rejected from the oxide. It has been demonstrated that performing a slow wet-oxidation, by reducing oxygen concentration in the furnace atmosphere, allows germanium to be rejected from the oxide in contrast to a normal wet-oxidation [LeGous et al., 1988]. Similarly increasing the dry-oxidation rate by pressurising the oxidising atmosphere causes

germanium to be incorporated into the oxide [Paine et al., 1991]. This supports the argument that it is the difference in oxidation rates between the wet- and dry-oxidations, rather than a difference in the chemistry of the reactions, that results in germanium incorporation/rejection into/from the oxide.

Once incorporated into an SiO_2 matrix the germanium will diffuse freely [Minke and Jackson, 2004] and under certain conditions, segregate into Ge-islands [Nobutoshi et al., 2007]. Germanium can be introduced to an SiO_2 matrix by ion implantation [Nobutoshi et al., 2007] or oxidation of certain SiGe layer structures [Stoica and Sutter, 2006] including over-oxidation in the Ge-condensation process to consume all Ge and Si into the oxide [Lai and Li, 2007].

One often overlooked oxidation process that occurs during Ge-condensation is the internal thermal oxidation (ITOX) of the buried oxide. As well as reacting with silicon at the top oxidising interface to form SiO_2, oxygen will diffuse into the underlying semiconductor material [Haas, 1960]. When oxidising an SOI wafer, oxygen will reach the BOX interface, where it will react to form SiO_2 [Takahashi et al., 1994]. This results in the BOX layer increasing in thickness and generally improves the quality of the BOX/Si interface [Matsumura et al., 2003]. It is the investigation of this process on low Ge composition SGOI wafers that lead to the invention of the germanium condensation technique [Mizuno et al., 2002].

A model has been proposed [Nakashima et al., 1996] for ITOX on pure silicon-on-insulator substrates, which considers the concentration gradient of oxygen through the top thermal oxide and SOI layers giving the ITOX layer thickness Y:

$$Y = K' \ln\left(L + \frac{2D_{Si}/k_2}{t_{SOI} - 0.44t_{ox} + (2D_{Si}/k_2)}\right) \tag{2.84}$$

Where t_{SOI} is the initial SOI thickness L is the thickness of an initial oxide cap and t_{ox}

is the thickness of oxide grown. K' is a constant given by:

$$K' = \frac{(k_1 + 2k_3)(D_{Si}/0.88k_2)}{k_1 + k_3} \tag{2.85}$$

and D_{Si} is the diffusion coefficient of oxygen in silicon [Haas, 1960] with units of $\text{cm}^2\ \text{s}^{-1}$:

$$D_{Si} = 0.21\text{exp}\left(\frac{-2.55}{k_B T}\right) \tag{2.86}$$

k_1, k_2 and k_3 are reaction rate constants [Nakashima et al., 1996] [Ghez and van der Meulen, 1972] with units of μm/min:

$$k_1 = 2.47 \times 10^{10}\text{exp}\left(\frac{-1.91}{k_B T}\right) \tag{2.87}$$

$$k_2 = 9.6 \times 10^{6}\text{exp}\left(\frac{-1.44}{k_B T}\right) \tag{2.88}$$

$$k_3 = 1.15 \times 10^{5}\text{exp}\left(\frac{-0.58}{k_B T}\right) \tag{2.89}$$

each corresponding to the chemical reactions at the oxidising interfaces:

$$(\text{Si} - \text{Si}) + \text{O}_2 \rightarrow (\text{Si} - \text{O} - \text{Si}) + \text{O} \tag{2.90}$$

$$(\text{Si} - \text{Si}) + \text{O} \rightarrow (\text{Si} - \text{O} - \text{Si}) \tag{2.91}$$

$$\text{O}_2 \leftrightarrow 2\text{O} \tag{2.92}$$

Other secondary considerations are important, when designing a Ge-condensation process, to maximise the quality of the final material layer. The diffusion of germanium from the oxidising interface into the substrate should be maximised and hence the temperature of the oxidation should be as high as possible, but not so high as to melt the SiGe layer. This promotes a uniform germanium composition throughout the semiconductor layers, reducing the strain gradient and subsequently reducing dislocation density [Sugiyama et al., 2004] and stacking fault density [Bedell et al., 2006]. If there

is a large difference in the initial and final germanium concentration of the layers, then a stepped temperature profile can be employed so that diffusion is maximised in the early stages of oxidation without melting the layers towards the end of the process [Mukherjee-Roy et al., 2005].

As with SiGe layers grown on a standard silicon substrate, SiGe layers associated with Ge-condensation are now understood to relax by misfit dislocations [Rehder et al., 2003], which has supplanted a suggested mechanism involving the compliance of the BOX layer [Di et al., 2005a]. A good agreement with the Matthew and Blakeslee critical thickness is demonstrated by Bedell et al. [2006] for many samples down to 40 nm.

2.6.2 Band Structure and Electrical Properties

The reduced zone representation of the band structure of germanium is displayed in figure 2.14. The energy levels of the top of the valence band and bottom of the conduction band (the respective band edges) define the energy gap as $E_G = 0.66$ eV. Note that the valence band edge is at the Brillouin zone centre and the conduction band edge is on the [111] directions.

The valence band edge is two-fold degenerate, being comprised of the so-called heavy-hole (HH) band and light-hole (LH) band. In the parabolic approximation the band edge hole masses are (equation 2.8): $m_{HH} = 0.3m_0$ and $m_{LH} = 0.04m_0$. Using these values in equation 2.6, 96% of holes per unit volume of germanium material are heavy holes.

The conduction band edge comprises of eight ellipsoidal surfaces, each centred at the zone edge on one of the eight [111] directions and are known as the L-ellipsoids (figure 2.15(a)). This gives four-fold conduction band degeneracy, as each Brillouin zone contains eight half-ellipsoids, modelled as four complete ellipsoidal surfaces. The

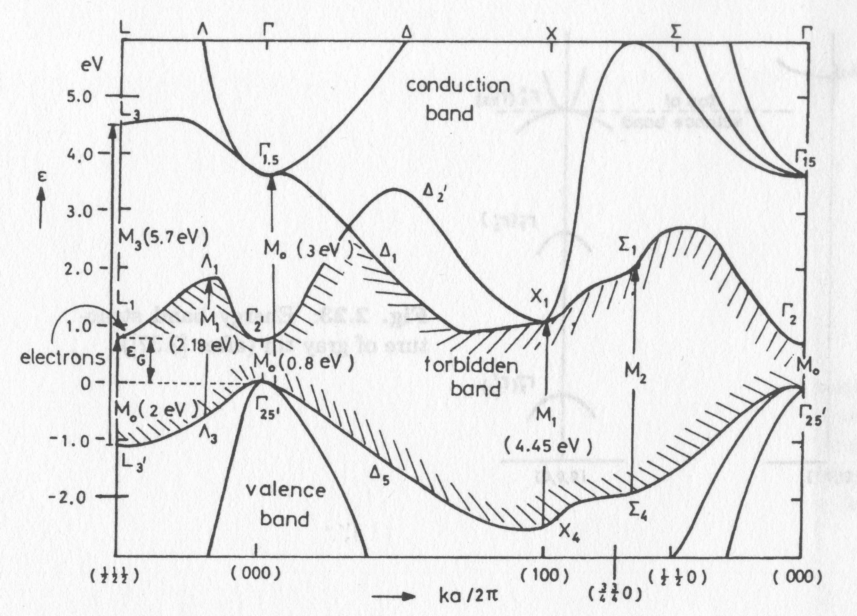


Figure 2.14: The band structure of germanium reproduced from Seeger [2004].

symmetry axes of the ellipsoids is down the [111] directions and hence in these directions is the longitudinal electron mass $m_l = 1.59m_0$. In the planes normal to these axes is the transverse electron mass $m_t = 0.08m_0$.

The effective hole masses in germanium are lower than the equivalent in silicon where $m_{HH}^{(Si)} = 0.53m_0$ and $m_{LH}^{(Si)} = 0.15m_0$. The electron masses for silicon are $m_t^{(Si)} = 0.19m_0$ and $m_l^{(Si)} = 0.91m_0$. Whereas the valence band edge in both semiconductors are almost identical in terms of band structure, the conduction band edge in silicon is comprised of six ellipsoids on the [100] directions (figure 2.15(b)), known as the Δ-ellipsoids.

Applying strain to a germanium or silicon layer changes the respective energies of the heavy hole and light hole band edges, along with distorting both bands to alter the respective effective masses. Here we assume that the strain is derived from a standard SiGe virtual substrate, i.e. it is biaxial in the substrate plane: compressive for

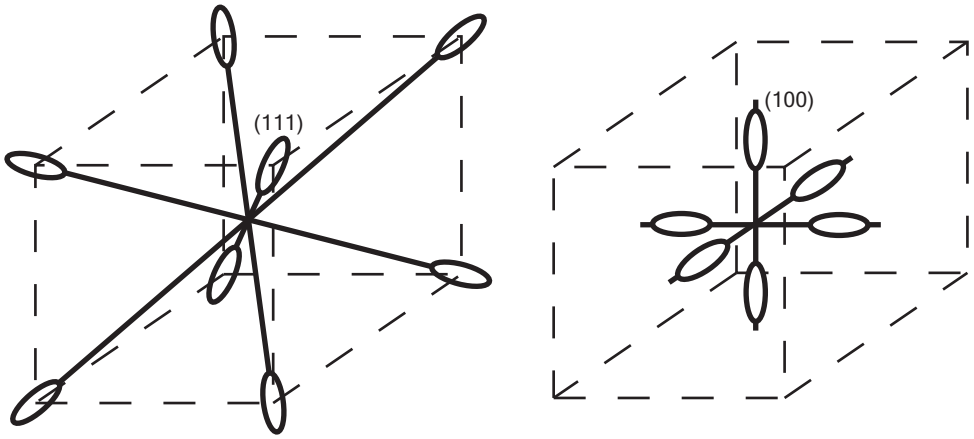


Figure 2.15: The Ge (a) and Si (b) conduction band edges approximated to ellipsoidal minimum energy surfaces.

a germanium layer and tensile for a silicon layer. This can be applied more generally to a relaxed $Si_{y-1}Ge_y$ virtual substrate with $Si_{x-1}Ge_x$ surface layer.

Under compressive strain ($y < x$) the heavy hole band edge increases in energy and light hole band edge decreases in energy (figure 2.16). The heavy hole band edge becomes sharper to give a decreased m_{HH} and the light hole band edge becomes shallower to increase m_{LH}. It is possible to have $m_{HH} < m_{LH}$, referred to as mass inversion. Tensile strain ($y > x$) has the opposite effects: the the light-hole band edge increases in energy and its effective mass increases also.

The most common wafer surface orientation is (100), which means bi-axial strain in the surface plane will break the symmetry of the Δ-ellipsoidal minima, comprising the conduction band in silicon, but not the L-ellipsoids in germanium. The energies of the various bands in silicon and germanium as a function of y are given by Schaffler [1997] and are reproduced in figure 2.17. Note that the split-off (SO) valence band has not been mentioned as it is a minimum of 250 meV below the HH/LH bands of (strained)

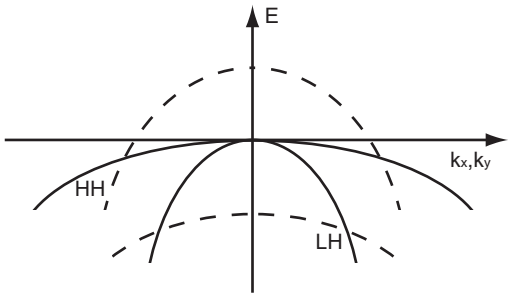


Figure 2.16: The heavy hole (HH) and light hole (LH) valence bands in bulk germanium (solid lines) and compressively strained germanium (dashed lines).

germanium; however, it does affect hole transport in silicon where its separation is comparable to k_BT at room temperature.

The room temperature bulk electron and hole mobilities of germanium, 3900 cm^2 V^{-1} s^{-1} and 1800 cm^2 V^{-1} s^{-1} respectively, greatly exceed those of silicon: 1450 cm^2 V^{-1} s^{-1} and 505 cm^2 V^{-1} s^{-1}. This can be attributed to the lower effective carrier masses and lower band degeneracies. The hole mobility is the highest of any semiconductor, which is particularly important as the performance of silicon p-MOSFETs is significantly worse than that of Si n-MOSFETs [Takagi et al., 1994a]. Room temperature hole mobilities of 2700 cm^2 V^{-1} s^{-1} in compressively strained germanium have been reported by Irisawa et al. [2002]. The improved mobility over bulk-Ge is thought to be due to decreased effective hole mass and the split degeneracy of the heavy hole and light hole bands [Sawano et al., 2005].

2.6.3 MOSFETs

It is the high bulk-mobility of electrons and holes in germanium, compared to those in silicon, that makes it a MOSFET channel material that could replace silicon in the future [Shang et al., 2006]. However, for this to happen the higher bulk mobility must

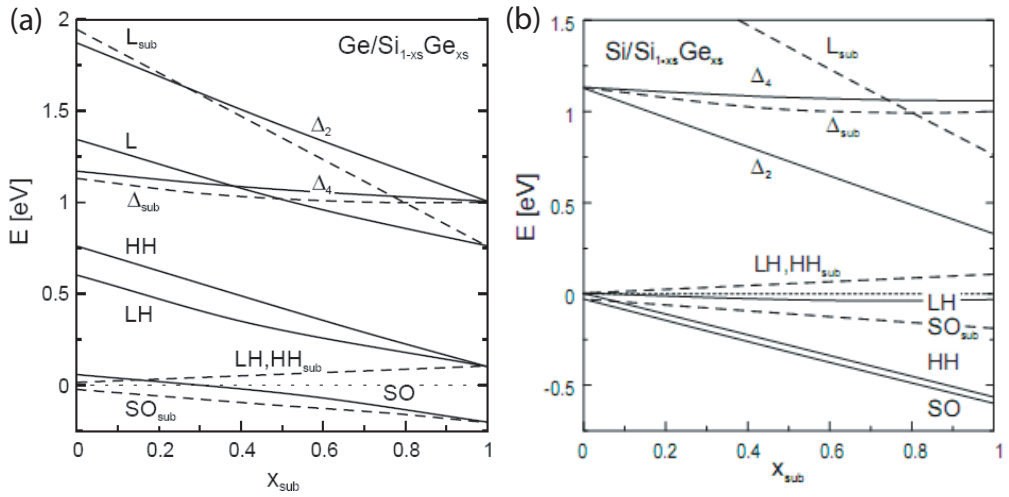


Figure 2.17: The various band edge alignments of strained germanium (a) and strained silicon (b) on a $Si_{1-xsub}Ge_{xsub}$ virtual substrate of (100) orientation. Note that although the energy level of the L-orbital conduction minima is altered with strain in germanium, its degeneracy is not split in contrast to the Si-like Δ minima. Reproduced from Schaffler [1997]

translate into an enhanced effective mobility in the MOSFET inversion layer. Any alternative semiconductor/dielectric material system is generally compared to silicon/silicon dioxide by comparing the effective mobility as a function of vertical effective field to the silicon universal mobility curves [Takagi et al., 1994a,b]. Germanium p-MOSFETs (with a channel length of about 100 μm) with an enhancement in channel mobility of 40% over the silicon universal curve have been demonstrated with a basic Ge-oxynitride/Ge-oxide gate dielectric [Shang et al., 2003]. After these promising early results, the technologies required to fabricate germanium MOSFETs of shorter channel lengths, have begun to emerge. 10 μm Ge p-MOSFETs, with a hafnium-oxide high-k dielectric, have $3\times$ mobility enhancement over the silicon universal mobility [Zimmerman et al., 2006]. Shorter devices with the same gate stack have recently been demonstrated and outperform silicon p-MOSFETs at 125 nm [Nicholas et al., 2007b] — although these devices suffer from large drain leakage. This problem could be solved by switching to GOI substrates:

working devices have been demonstrated on these [Park et al., 2006] with lower leakage and better on-state performance than the SOI equivalent.

Further mobility enhancements have been demonstrated by fabricating p-MOSFETs with strained Ge channels, using virtual substrates [Nicholas et al., 2007a] ($4\times$ silicon universal) and locally induced strain [Takagi et al., 2006] (up to $10\times$ silicon universal). The latter locally applies the Ge-condensation technique to the MOSFET channel region only.

Despite the promising enhancements observed for Ge p-MOSFETs; Ge n-MOSFETs have yet to show any enhancement over silicon n-MOSFETs, despite germanium's higher bulk electron mobility. Part of the problem lies with poor n-dopant activation in the source/drain regions due to the high anneal temperature required ($\geq$600 °C), which can damage the gate stack and result in too much diffusion of the junction dopants. A laser anneal has been applied, which allows the gate stack to remain cooler as the dopants are annealed at the appropriate temperature. This lead to reduced source/drain resistance and a corresponding increase in mobility but still not above the silicon universal [Zhang et al., 2006]. Maikap et al. [2007] demonstrates n-MOSFETs on strained Ge on (111) oriented substrates, which allows degeneracy splitting of the Ge-conduction band and results in a 40% mobility enhancement over the silicon (111) n-MOSFET universal curve but again not higher than the silicon (100) n-MOSFET universal curve. The advantage of the local Ge-condensation technique, is that it allows another type of channel material for n-MOSFETs (strained silicon for example) alongside Ge p-MOSFETs [Takagi et al., 2006; Shang et al., 2006].

Chapter 3

Electrical and Physical Characterisation Techniques

Here, the practical techniques used to electrically and physically characterise semiconductor devices and materials are described. This chapter is based on Schroder [2006] — a definitive source for nearly all modern semiconductor characterisation techniques.

3.1 Electrical Characterisation of MOS Devices

All measurements were performed using a probe station — either a standard one or the low temperature probe station detailed in section 4.1. They are operated in essentially the same way and have a basic set of similar components. These include an electrical shield along with high quality connections and coaxial cable throughout to ensure that the signal is protected from interference right to the probe needles. There is an electrically connected substrate chuck as many device batches have the substrate contact as metallisation on the back of the wafer. A good electrical contact between wafer and chuck is required and is normally achieved by many small perforations in the chuck

with a small pump sucking air through them. In the case of the low temperature probe station, the sample merely rests under its own weight. To improve the electrical contact a eutectic is smeared on the back of the sample before it is placed on a thin piece of copper and then the whole assembly onto the sample chuck. The needles themselves are manipulated with micrometers by the operator, whilst monitoring the tips through a microscope. When contacting the probe needles it is important to be consistent with the pressure applied and the position on the contact pads. In large or specialist institutions it is normal to have an automatic prober which, using a custom needle set to match a specific mask, will, once aligned, automatically perform consistent electrical measurements over an entire wafer with precision.

The room temperature probe station has a movable substrate chuck, which enables the sample to be moved relative to the probe needles. As devices on a sample usually have identical contact pad configurations, movement of the chuck, rather than individual probe needles, greatly reduces the time taken to probe many devices. Once contact to a device is made, the measurement chamber is closed or covered to provide a dark, electrically shielded environment.

The most common measurement to perform on a MOSFET is an I_d-V_g sweep: a gate voltage is applied to achieve the off-state and incremented to the on state (positive to negative bias for p-MOSFETs), a constant bias to the drain is applied and the resulting drain current is measured. Normally the currents in the source, gate and substrate are measured in addition to the drain (four-terminal measurement) to check that the MOSFET is operating normally without excessive leakage. The substrate and source are generally held at 0 V; however, some techniques require a substrate or source bias.

All I-V measurements presented in this thesis were made with a Agilent 4156c

parameter analyser. This is extremely sensitive (minimum detectable current of the order of a femto-Ampere), industry standard piece of equipment and allows for measurements to be made efficiently with a programmable interface. It contains in-built integration to minimise noise effects — although as with all sampling based noise reduction this increases the measurement time significantly. Quasi-static C-V is also performed with this equipment, which measures the capacitance (usually) between substrate and gate whist incrementing a DC gate voltage.

LCR (inductance, capacitance, resistance) measurements were made on an Agilent 4284A precision LCR meter. This operates by oscillating a small AC voltage signal about a varying DC voltage level at a selected frequency (20 Hz to 1 MHz) on the gate contact and measuring the resulting AC current through the source, drain or substrate contact. Note that the capacitance measured is $\frac{\mathrm{d}Q}{\mathrm{d}V}$, i.e. the charge which responds to the AC signal. All charges respond to the DC signal provided that the time delay between DC voltage steps is large enough.

By default the components between the terminals are modelled as a capacitance in parallel with a conductance, which are obtained as a function of DC voltage. Once the signal frequency, DC voltage range, time delays and signal averaging parameters have been specified this is all performed automatically by the 4284A. Various corrections must be performed to take into account the position and length of the cables, which themselves exhibit a capacitance and conductance. There is a built in function which determines correction parameters automatically.

To further simplify its operation a computer program was written in TestPoint. This program provides a user-friendly interface to specify measurement parameters and allows a sequence of measurements over a range of frequencies to be performed. The program graphs the capacitance and conductance measured as a function of DC voltage

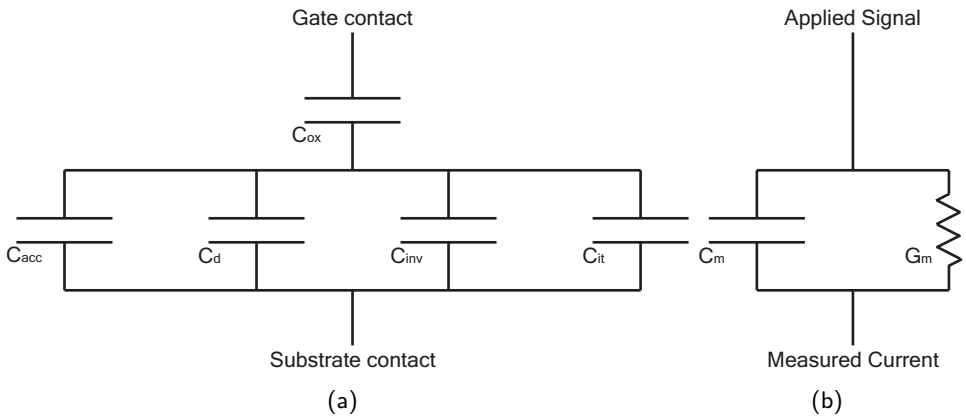


Figure 3.1: (a) Equivalent circuit for a MOS capacitor with C_{ox},C_{acc},C_d,C_{inv} and C_{it} as the oxide, accumulation, depletion, inversion and interface trap capacitances respectively, adapted from Schroder [2006]. (b) Assumed circuit by 4284A precion LCR meter.

to allow a quick analysis by the user as to whether the measurement was successful. This program was completed by the author in conjunction with Andrew Dobbie of the Nano-Silicon Research Group, Department of Physics, University of Warwick.

The rest of this section describes the interpretation of current, capacitance and conductance data measured for a MOS device with the equipment described above.

3.1.1 Equivalent Circuits

A MOS capacitor being characterised with the LCR meter is treated as equivalent circuit, shown in figure 3.1(a). Depending on the measurement frequency used and the condition of the surface potential (determined by the DC applied gate voltage) the charge associated with a particular capacitance is either not present, acts as a short circuit or cannot respond fast enough to the AC signal and hence the circuit simplifies. This allows the assumed equivalent circuit by the LCR meter to be applied (figure 3.1(b)) and MOS capacitor parameters extracted based on the measured capacitance.

In accumulation the large number of free majority carriers at the surface (electrons in a n-MOS capacitor) electrically short the dielectric interface to the substrate and the only capacitance measured is C_{ox} (figure 3.2(a)). Biasing the gate so that the MOS capacitor reaches the depletion condition gives the depletion region capacitance C_d in series with the oxide capacitance (figure 3.2(b)). Interface states with capacitance C_{it} can also respond, depending on the measurement frequency and act in parallel to C_d.

When the MOS capacitor is in inversion the fixed depletion layer width W_{dm} gives rise to a response time for the minority carriers that make up the inversion layer. The generation-recombination current density for minority carriers is:

$$J_r = \frac{qn_iW_d}{t} \tag{3.1}$$

Where t is the minority carrier lifetime. The time required to generate an inversion charge density equal to the depletion charge is therefore:

$$\frac{Q_d}{J_r} = \left(\frac{N_a}{n_i}\right) t \tag{3.2}$$

Hence, at high frequencies minority carriers do not respond and the measured capacitance remains at C_d (figure 3.2(c)), which will be fixed at or approaching its smallest value due to the depletion layer width being at its largest value W_{dm}. A low frequency signal allows the inversion charge to respond and once a high inversion charge density is achieved, it shorts the surface to the substrate and hence the measured capacitance tends to C_{ox} (figure 3.2(d)).

Note that for each surface condition, the equivalent circuit simplifies to one that corresponds almost directly to that assumed by the 4284A. An example of high and low frequency C-V curves obtained from a n-MOS capacitor are shown in figure 3.3(a).

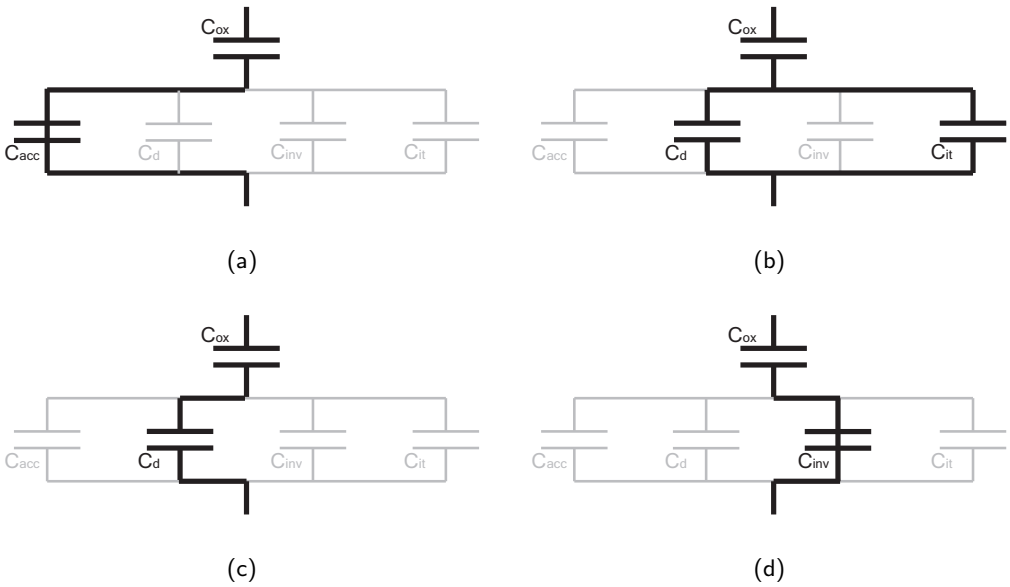


Figure 3.2: Equivalent circuits for a MOS capacitor in (a) accumulation, (b) depletion, inversion at (c) high frequency and (d) low frequency. Solid lines through a capacitor represent free charge shorting the surface to the bulk of the semiconductor. Adapted from Schroder [2006]

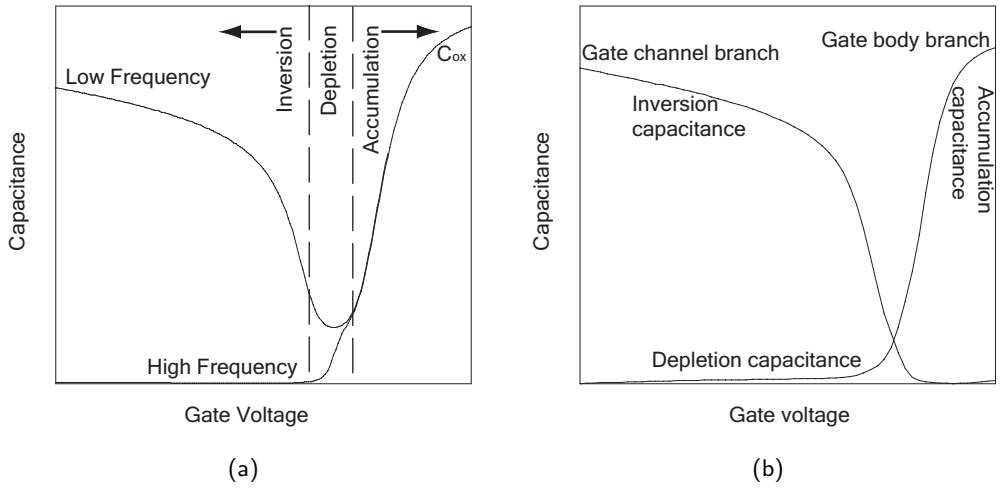


Figure 3.3: Typical C-V curve for a n-MOS capacitor (a) and split C-V curves for a p-MOSFET (b).

A MOSFET has the advantage of two extra contacts, which allow a more sophisticated LCR characterisation. Clearly if the source and drain are left floating and the capacitance measured through the substrate contact, the MOSFET should behave as a MOS capacitor (perhaps with some difference due to source/drain giving rise to a parasitic capacitance). However, if this measurement is performed with the source and drain contacts earthed then the depletion capacitance is obtained. The inversion layer charge is earthed and hence does not contribute to capacitance measured. Swapping the connections so that the substrate is earthed and the source and drain are connected as the measurement terminals, allows the measurement of the inversion capacitance independently of the depletion capacitance. These two terminal configurations are used in the split C-V technique to determine depletion charge and inversion charge independently — referred to as the gate body branch (GB) and gate channel branch (GC) respectively. The C-V curves obtained by performing split C-V on a p-MOSFET are shown in figure 3.3(b) and the connection configurations for this technique are shown in figure 3.4.

3.1.2 Flat-band Voltage

The gate voltage at which there is no band bending at the surface ($\psi_s = 0$) is called the flat-band voltage $V_g = V_{fb}$. This is an important parameter when analysing capacitance curves as it marks the boundary between accumulation and depletion. Schroder [2006] gives an experimental technique to obtain the flat-band voltage as plotting $\frac{1}{C_{hf}^2}$ against V_g, where the lower knee in the curve will correspond to V_{fb}. A high AC signal frequency is used to prevent acceptor-like interface traps from responding (n-MOS capacitor). However, traps will still respond to the DC signal, which reveals why an experimental technique is more appropriate than a theoretical calculation of flat-band voltage.

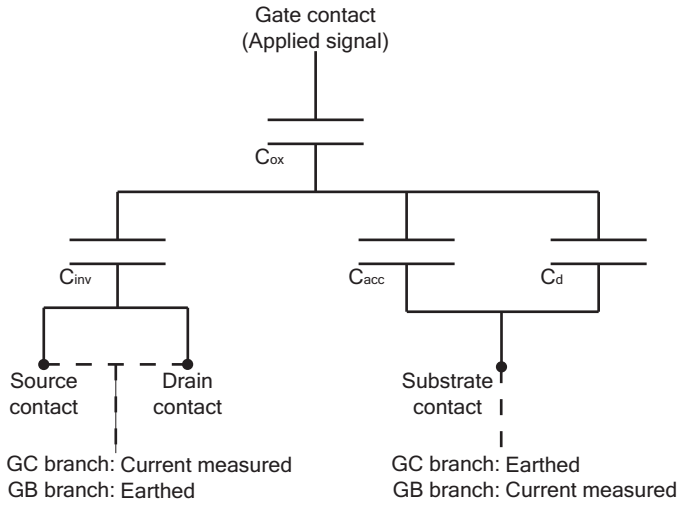


Figure 3.4: Equivalent circuit for the split C-V technique with dashed lines representing external cables.

Judging by eye where the knee in the $\frac{1}{C_{hf}^2}$ curve is located is difficult, but it will coincide with the steepest differential. Therefore, performing two differentiations gives a peak, the maximum of which coincides with V_{fb}. Unfortunately, differentiating amplifies any noise in the signal. Using small DC voltage increments ($\leq$0.01 V) or a lot of signal averaging usually reduces the noise to acceptable levels.

3.1.3 Inversion Charge, Depletion Charge and Vertical Field

The split C-V technique is applied to obtain inversion capacitance $C_{inv}(V_g)$ and depletion capacitance $C_d(V_g)$, from which inversion charge density $Q_{inv}(V_g)$ and depletion charge density $Q_d(V_g)$ are obtained by integrating from the flat-band voltage:

$$Q_{inv}(V_g) = \int_{V_{fb}}^{-\infty} C_{inv}(V_g)\mathrm{d}V_g \tag{3.3}$$

$$Q_d(V_g) \int_{V_{fb}}^{-\infty} C_d(V_g)\mathrm{d}V_g \tag{3.4}$$

For a p-MOSFET.

The vertical electric field is given as:

$$E_{eff} = \frac{Q_d + \frac{1}{3}Q_{inv}}{\epsilon_0 \epsilon_{se}} \tag{3.5}$$

This expression is obtained by treating the inversion and depletion charge as infinite sheets and applying the appropriate Gaussian surface (pillbox). The empirical factor of $\frac{1}{3}$ accounts for the distribution of holes ($\frac{1}{2}$ for electrons in n-MOSFETs) in the inversion layer. Note that this factor was given as $\frac{11}{32}$ for the scattering theory in section 2.4.2 according to equation 2.47. As mentioned E_{eff} allows a direct comparison of mobility to the universal silicon curve [Takagi et al., 1994a].

3.1.4 Threshold Voltage

There are many techniques to extract threshold voltage and in general they do not agree on a single value. Unlike the idealised MOSFET described in section 2.3, a real MOSFET is subject to leakage currents, source/drain resistance and a non-constant channel mobility. This means that drain current will not be absolutely linear in the linear region, nor will there be zero current in the off-state nor will drain current keep increasing with gate voltage indefinitely.

We assume that threshold voltage corresponds to the condition $\psi_s = 2\phi_b$, which should also be the point at which current begins to flow from source to drain. However, in a real MOSFET, the gate voltage at which significant current begins to flow is open to interpretation and it may or may not correspond precisely to the strong inversion condition. This gives rise to some confusion over the definition of the threshold voltage in a real device. From the point of view of creating a working CMOS circuit, the threshold voltages need to be defined as the point at which enough current flows to consider the devices as turned on. When studying the physics of MOSFET devices, it is

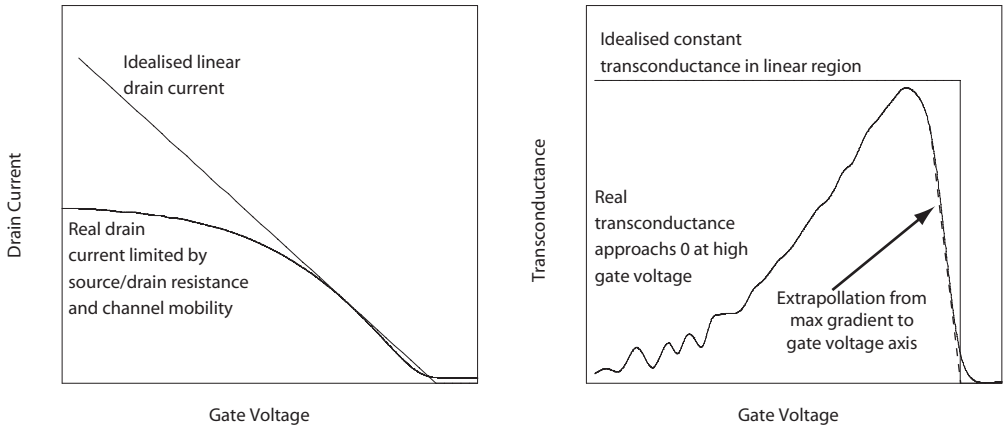


Figure 3.5: A comparison of the drain current (a) and transconductance (b) for an ideal p-MOSFET and a real p-MOSFET in the linear region. In the latter, the transconductance method, used in this thesis, is demonstrated.

more convenient to use the strong inversion condition than a practical definition of the threshold voltage.

Figure 3.5 compares the I_d-V_g and transconductance characteristics of a real and idealised MOSFET (low drain bias). Four common techniques to extract threshold voltage are: finding a gate voltage corresponding to a particular pre-selected drain current (constant drain current method); linearly extrapolating to the gate voltage axis from the steepest gradient in the drain current curve (linear extrapolation method); linearly extrapolating to a gate voltage from the steepest gradient in the transconductance curve (transconductance method); taking the gate voltage at which the peak in differentiated transconductance curve occurs (transconductance derivative method). In the case of the idealised MOSFET, all techniques always agree on a single value of threshold voltage (except the first technique, which only does if the chosen value of I_d is very small) whereas for a real MOSFET they, in general, do not.

In chapter 4 of this thesis we use the transconductance method to measure threshold voltage, which was first proposed by Tsuno et al. [1999]. A theoretical consideration, given by the authors, reveals that it is assumed that the transconductance, and therefore the channel mobility, is proportional to the density of charged impurities at low carrier densities. Coulomb scattering is conclusively demonstrated to dominate mobility at peak transconductance in the Ge MOSFETs reported in chapter 4 and hence the assumption is valid for these devices. Furthermore, there is a large variation in the areal density of charged impurities in these devices and given that a relationship is inferred between threshold voltage and mobility, any modification of the threshold voltage by these charges must be taken into account, lest a systematic error in threshold voltage determination lead to a erroneous relationship. The technique conveniently satisfies this criterion and has the advantage of being insensitive to gate length; hence, if the study is extended to Ge MOSFETs of shorter gate lengths, then the transconductance method will still be applicable.

3.1.5 Source/Drain Resistance

Between the probe needle tip on a source/drain contact pad and the MOSFET channel, there is electrical resistance called the source/drain resistance R_{sd}. Contributions to this resistance arise from the metallisation and doped source/drain regions of semiconductor. It is an important parameter as it affects all of the measurements made using these contacts on a MOSFET. The most common method used to extract the resistance is to treat the MOSFET as an equivalent circuit of three resistors in series (figure 3.6): the source (R_s), drain (R_d) and channel (R_{ch}). Note that $R_{sd} = R_s + R_d$.

Assuming that R_{sd} is a constant and that R_{ch} for a given gate overdrive ($V_{gt} = V_g - V_t$) is proportional to gate length (i.e. it has a constant resistivity per unit length

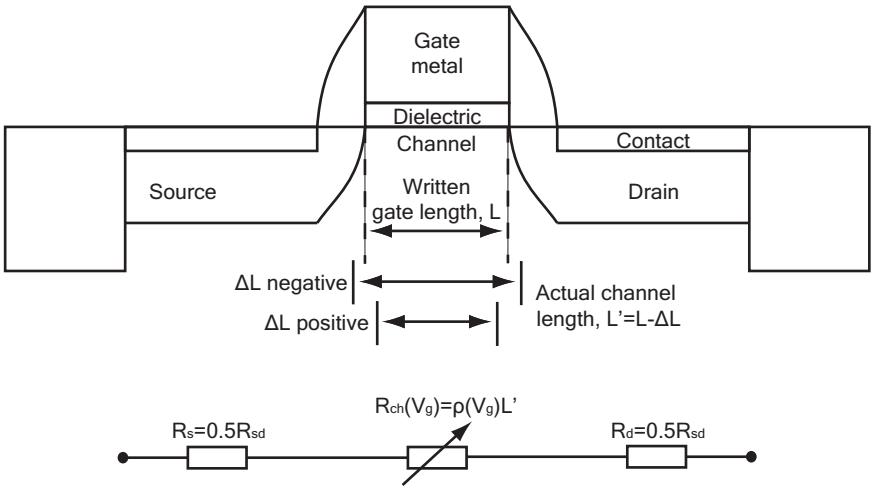


Figure 3.6: A MOSFET with equivalent circuit used for R_{sd} and ΔL extraction.

$\rho(V_{gt})$) then one can calculate R_{sd} from an array of MOSFETs of different gate lengths L. $\frac{V_d}{I_d}$ against L is plotted for several values of V_{gt}, which are typically small (see figure 4.2(a) for an example). Note that V_t must be known for each device and that the method used to measure it must not be systematically affected by gate length.

Straight lines of best fit are then drawn through each series of points associated with each gate overdrive. These lines of best fit should intersect at a single point with coordinates $(\Delta L, R_{sd})$. ΔL accounts for the fact that the written gate length L (as defined by the mask during fabrication) is not necessarily the real gate length L'. We define L':

$$L' = L - \Delta L \tag{3.6}$$

Hence, a positive value of ΔL means that the channel is shorter than the written gate length. As with R_{sd}, this has the greatest effect on devices with a short gate length where R_{ch} and L are comparable to R_{sd} and ΔL respectively. Once measured, the source/drain resistance and channel length difference are used to correct the applied

drain bias and gate voltage:

$$V_d' = V_d - I_d R_{sd} \tag{3.7}$$

$$V_g' = V_g - I_d R_{sd} \tag{3.8}$$

More accurate values of ΔL and R_{sd} may be obtained by plotting the intercept against the gradient of each line of best fit (see figure 4.2(b) for an example). This allows a better assessment of how good the original lines of best fit agree on a single value of R_{sd} and ΔL, and removes the need for judging by eye where the lines intersect.

3.1.6 Channel Mobility

The mobility of a MOSFET channel μ is calculated by considering the drift and diffusion drain current components of both electrons and holes. In a p-MOSFET we ignore electrons as the hole density is far greater in both the channel and p-doped source/drain regions. This leaves the hole drift current and hole diffusion current to contribute to the drain current. Normally the diffusion current is ignored and an expression for the mobility based on the drift current alone, is used. Using equation 2.13 and substituting for qp with the inversion charge sheet density Q_{inv}, lateral field with $\frac{V_d}{L}$ and noting that current sheet density is given as $\frac{I_d}{W}$:

$$\mu(V_g) = \frac{L}{W} \frac{I_d(V_g)}{Q_{inv}(V_g) V_d} \tag{3.9}$$

Hence, determining MOSFET channel mobility requires the measurement of the inversion charge and the drain current. Typically a voltage of -50 mV is applied to the drain (for p-MOSFETs) as we are interested in the low lateral field mobility. Inversion charge density can be accurately obtained from the split C-V technique as a function of gate voltage. Note that the application of this technique also allows the vertical effective field at the channel surface E_{eff} to be obtained as a function of V_g. Hence, with a

combination of I_d-V_g and split C-V, mobility can be measured as a function of vertical effective field and compared to the silicon universal mobility [Takagi et al., 1994a].

Equation 3.9 is the one most commonly used to obtain effective mobility but is not necessarily the most accurate. We reintroduce the hole diffusion current and hence from equation 2.17:

$$\frac{I_d}{W} = Q_{inv}(V_g, x)\mu E_x - \frac{k_B T \mu}{q}\frac{\mathrm{d}Q_{inv}}{\mathrm{d}x} \tag{3.10}$$

We must recognise that with a finite applied drain voltage, the inversion charge is not uniformly distributed over the channel length — the charge density decreases towards the drain, hence Q_{inv} is a function of position along the channel x and applied gate voltage V_g. The lateral field E_x will also be dependent on this charge distribution. Therefore, obtaining mobility from this expression is not as simple as the drift-current only case. Sodini et al. [1982] defines a function $0 \leq F(V_g) \leq 1$:

$$F(V_g) = \frac{\frac{\mathrm{d}Q_{inv}}{\mathrm{d}\psi_s}}{C_{ox} + \frac{\mathrm{d}Q_s}{\mathrm{d}\psi_s}} \tag{3.11}$$

Where Q_s is the total charge at the semiconductor surface. Provided drain bias is low:

$$C_{inv}(V_g) = F(V_g)C_{ox} \tag{3.12}$$

Where C_{inv} is the measured inversion charge capacitance and, ignoring interface states, is simply the measured gate channel branch capacitance in the split C-V technique. It is also shown that:

$$E_x = F(V_g)\frac{V_d}{L} \tag{3.13}$$

and that

$$\frac{\mathrm{d}Q_{inv}}{\mathrm{d}x} = -\frac{C_{ox}V_d F(V_g)}{L} \tag{3.14}$$

Substituting these three expressions into equation 3.10 yields:

$$\mu(V_g) = \frac{L}{W}\frac{I_d(V_g)}{V_d}\left[C_{inv}(V_g)\left(\frac{Q_{inv}(V_g)}{C_{ox}} + \frac{k_B T}{q}\right)\right]^{-1} \tag{3.15}$$

Which is a more accurate expression for mobility. Note that no extra measurements are required than when mobility is calculated with equation 3.9. One point not raised in Sodini et al. [1982] is that MOSFET inversion layers are often degenerate at large gate voltages, where the Fermi-energy must be substituted for k_BT. When the two energies are comparable, no simple expression for mobility is forthcoming.

Note that equation 3.15 will, in general, give a lower mobility than equation 3.9. It is worth pointing out that the diffusion corrected mobility expression has been misinterpreted in the past [Palmer, 2001; Nicholas, 2004] so that the diffusion current contribution actually *increases* the channel mobility. This is, of course, incorrect as if we measure a drain current and accept that some contribution has arisen due to diffusion, then we must see that a drift-only current would be lower and thus the diffusion correction must decrease channel mobility. Hence, the diffusion contribution increases the denominator in the diffusion-corrected expression for mobility (equation 3.15).

A convenient method of obtaining the channel mobility from a single $I_d - V_g$ measurement (provided C_{ox} is known) is now described. In general, the mobility obtained with this technique does not agree with the effective mobility (described above) and to avoid confusion, is called the field effect mobility μ_{FE}. We make the approximation that $Q_{inv}(V_g) = C_{ox}(V_g - V_t)$ and initially assume that mobility is constant with gate voltage, so that differentiating equation 2.30 gives:

$$\mu_{FE} = \frac{L}{W}\frac{1}{V_d C_{ox}}\frac{\mathrm{d}I_d}{\mathrm{d}V_g} = \frac{L}{W}\frac{1}{V_d C_{ox}} g_m \tag{3.16}$$

Where g_m is the transconductance.

However, mobility is actually a function of vertical field, which in turn is a function of gate voltage. Hence, we repeat the differentiation with $\mu(V_g)$:

$$\mu_{FE}(V_g) = \frac{L g_m}{W C_{ox} V_D}\left(1 + \frac{V_g - V_t}{\mu(V_g)}\frac{\mathrm{d}\mu}{\mathrm{d}V_g}\right) \tag{3.17}$$

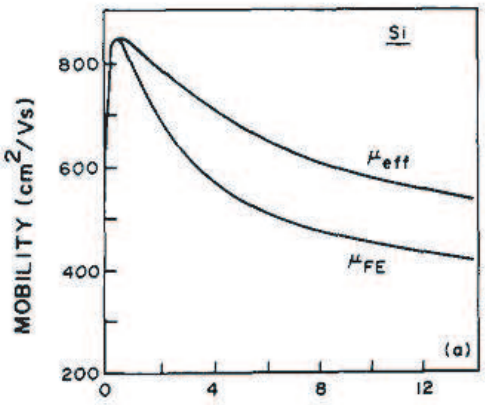


Figure 3.7: A comparison of the field effect mobility μ_{FE} and effective mobility μ_{eff} measured as a function of gate voltage and demonstrating an agreement at the peak values. Reproduced from Kang et al. [1989].

Taking the peak field effect mobility $\mu_{FE}^{(peak)}(V_g)$ with respect to V_g ensures that the differential term is zero and we observe that, at the maximum, equation 3.17 is equivalent to equation 3.16. Provided the approximation $Q_{inv}(V_g) = C_{ox}(V_g - V_t)$ is reasonable, peak field effect mobility will be equal to peak effective mobility $\mu_{FE}^{(peak)} = \mu^{(peak)}$. This has been demonstrated experimentally [Kang et al., 1989], shown in figure 3.7.

3.1.7 Interface State Density

Interface states give rise to an extra capacitance in series with C_{ox}, as shown in figure 3.1(a). Traps respond to DC and AC voltages applied to the gate, but only the latter if the signal frequency is not greater than the associated response time of a trap. Traps respond in depletion and weak inversion, as the surface potential of the semiconductor changes in these regions. Several techniques are now described, which allow D_{it} or N_{it} to be measured.

High-Low Frequency C-V Technique

This technique is applied to a MOS capacitor structure and exploits the finite response time of an interface trap. Consider a low frequency or quasi-static capacitance measurement:

$$C_{lf} = \left(\frac{1}{C_{ox}} + \frac{1}{C_d + C_{it}} \right)^{-1} \tag{3.18}$$

This can be rearranged to give C_{it}, from which D_{it} can be calculated:

$$D_{it} = \frac{Cit}{q^2} \tag{3.19}$$

The depletion capacitance is eliminated by making a high frequency measurement:

$$C_d = \frac{C_{ox} C_{hf}}{C_{ox} - C_{hf}} \tag{3.20}$$

This gives a final expression for D_{it}, which requires only a single low frequency and a single high frequency capacitance measurement to be performed:

$$D_{it} = \frac{C_{ox}}{q^2} \left(\frac{C_{lf}/C_{ox}}{1 - C_{lf}/C_{ox}} - \frac{C_{hf}/C_{ox}}{1 - C_{hf}/C_{ox}} \right) \tag{3.21}$$

C_{lf} and C_{hf} are a function of gate voltage and hence D_{it} is obtained as such. In accumulation the curves should tend to a common level C_{ox}. In inversion the curves deviate as $C_{lf} \rightarrow C_{ox}$ and $C_{hf} \rightarrow C_{dm}$. Therefore the band-gap is probed from the onset of depletion to the onset of inversion. D_{it} is obtained as a function of surface potential (zero corresponding to the flat-band condition) by:

$$q\psi_s = \int_{V_{g1}}^{V_{g2}} (1 - C_{lf}/C_{hf}) \mathrm{d}V_g + \Delta \tag{3.22}$$

Where Δ is an integration constant equal to $q\psi_s$ at $V_g = V_{g1}$ and hence setting $V_{g1} = V_{fb}$ gives $\Delta = 0$. Therefore, the integral should be split into two parts, one integrating from flat-band to strong inversion and one integrating from flat-band to accumulation. If

performed correctly and the MOS capacitor is relatively uniform, with no significant gate leakage, then $q\psi_s$ will generally correspond to the valence band edge in accumulation (for an n-MOS capacitor). Hence D_{it} can be plotted as a function of energy within the semiconductor band-gap.

Conductance Technique

The conductance technique is one of the most sensitive methods of determining D_{it} [Schroder, 2006]. It relies on the process of filling and emptying interface traps being "lossy", so that the equivalent circuit in figure 3.2(b) is modified with a resistance R_{it} in series with C_{it} (figure 3.8(a)). The response time for an interface trap τ_{it} will now correspond to the RC value: $\tau_{it} = C_{it}R_{it}$. The 4284A precision LCR meter simulates circuits as a parallel capacitance and conductance, and so far we have not looked at any technique that requires the measurement of the latter. However, here we require both capacitance and conductance for our equivalent circuit, which is a parallel conductance G_{pa} and capacitance C_{pa}, both in series with C_{ox} (figure 3.8(b)).

Assuming that only interface states with an energy equal to $q\psi_s$ respond, it is given that [Schroder, 2006]:

$$C_{pa} = C_d + \frac{C_{it}}{1 + (\omega\tau_{it})^2} \tag{3.23}$$

$$\frac{G_{pa}}{\omega} = \frac{q\omega\tau_{it}D_{it}}{1 + (\omega\tau_{it})^2} \tag{3.24}$$

Where C_{it} is substituted for $D_{it}q$.

However, generally traps k_BT above and below $q\psi_s$ respond and hence equation 3.24 is modified to account for this:

$$\frac{G_{pa}}{\omega} = \frac{qD_{it}}{2\omega\tau_{it}}\ln\left[1 + (\omega\tau_{it})^2\right] \tag{3.25}$$

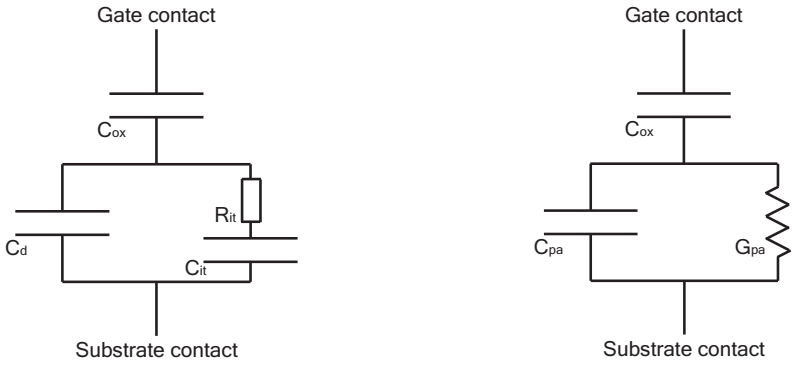


Figure 3.8: The modified equivalent circuit for a MOS capacitor in depletion as used for the conductance technique (a) and the same circuit modelled as a parallel conductance G_{pa} and capacitance C_{pa} in series with the oxide capacitance C_{ox} (b). Reproduced from Schroder [2006]

which has a maximum at $\omega \approx \frac{2}{\tau_{it}}$, hence:

$$D_{it} \approx \frac{2.5}{q} \left(\frac{G_{pa}}{\omega} \right)_{max} \tag{3.26}$$

We now relate the capacitance and conductance of the equivalent circuit to those measured (C_m and G_m respectively):

$$\frac{G_{pa}}{\omega} = \frac{\omega G_m C_{ox}^2}{G_m^2 + \omega^2 (C_{ox} - C_m)^2} \tag{3.27}$$

Measurements are usually performed by a sequence of DC voltage sweeps, each with a superimposed AC signal of a different frequency. To ensure that the maximum $\frac{G_{pa}}{\omega}$ is observed, the frequency range needs to be large (≈100 Hz to ≈1 MHz) as does the number of frequencies used. $\frac{G_{pa}}{\omega}$ is plotted as a function of ω and the peak corresponding to D_{it} found. This is repeated at each applied DC gate voltage to obtain D_{it} as a function of gate voltage, which can be converted to D_{it} as a function of band-gap energy using equation 3.22.

Subthreshold Slope Technique

A MOSFET turns off through weak inversion to depletion and the rate at which the current decreases (the subthreshold slope) is strongly dependant on the interface trap capacitance, which is in parallel with the depletion capacitance. We modify equation 2.34 to take this into account:

$$S = \frac{\ln(10)k_BT}{q}\left(1 + \frac{C_d + C_{it}}{C_{ox}}\right) \tag{3.28}$$

The depletion capacitance can be measured using the gate body branch of the split C-V technique or estimated from equation 2.33 by assuming $C_d = C_{dm}$, provided the substrate doping density is known.

Hence, using $D_{it} = C_{it}q^2$, the interface state density can be obtained from the subthreshold slope of a MOSFET:

$$D_{it} = \frac{C_{ox}}{q^2}\left(\frac{qS}{\ln(10)k_BT} - 1\right) - \frac{C_d}{q^2} \tag{3.29}$$

Threshold Voltage Technique

At the strong inversion condition, any interface traps between mid-gap and the surface energy $2q\psi_s$ will be charged (figure 2.10). In a p-MOSFET, this will be donor-like states with a positive charge. Given that threshold voltage corresponds to the strong inversion condition, we will observe a negative threshold voltage shift ΔV_t in a p-MOSFET after inducing interface states. This is normally done by applying a high-voltage to the gate for an extended length of time. Assuming that before the pulse there are no states and after there are N_{it} charged states (per area) at strong inversion then:

$$\Delta V_t = -\frac{qN_{it}}{C_{ox}} \tag{3.30}$$

Note that only the donor-like states (p-MOSFET) contribute to the threshold voltage shift as at $V_g = V_{mg}$ no interface traps are charged, hence $\Delta V_{mg} = 0$. An

average value of D_{it} in the band-gap between mid-gap and strong inversion can be obtained:

$$\langle D_{it} \rangle \, \Delta E = N_{it} \tag{3.31}$$

Where ΔE is the energy range from $E(V_{mg})$ to $E(V_t)$:

$$\Delta E = q\phi_b \tag{3.32}$$

Effect on Split C-V

It has already been shown that MOS capacitor capacitance measurements are affected by the presence of interface traps. Split C-V measurements performed on a MOSFET are also affected — the traps add an unwanted capacitance in parallel to the depletion charge capacitance and inversion charge capacitance in the gate-body and gate-channel branches respectively. This leads to an over-measurement of both charges, which in turn gives an under-measurement of effective mobility and an over-measurement of vertical effective field. The effect of interface traps is particularly noticeable in the gate-channel branch, where a shoulder will appear at low inversion charge density [Schroder, 2006].

A reliable split C-V measurement can be performed by selecting a frequency high enough so that the signal period is greater than the interface trap response time. However, high frequency measurements are more attenuated by resistances (source/drain, channel, gate and substrate) and hence the chosen measurement frequency should not allow traps to respond but also not be significantly attenuated by resistances [Zhu et al., 2004].

Charge Pumping

Charge pumping is a useful technique as it can be used to extract D_{it} as a function of surface potential in a MOSFET of any size. This is in contrast to the other techniques

described above, which only apply to large MOS capacitors; or give an average interface state density or D_{it} at a single energy level, in a MOSFET.

The source and drain are biased to attract inversion charge. A short voltage pulse that is applied to the gate, biases the MOSFET from inversion to accumulation and back again. At the start of the pulse, inversion charge carriers flow out through the source and drain. Interface traps of an energy close to the band edge also empty and the liberated carrier flows into the source or drain. However, those traps lying further from the band edge have a longer response time and hence keep their charge. Note that this takes place before the surface potential reaches flat-band and at the onset of accumulation, majority carriers arrive at the surface and recombine with charges in the interface states. At the end of the gate pulse the surface is biased back to inversion and those majority carriers that have not recombined with traps, leave the surface. Thus, by comparing the number of majority carriers that enter and leave the MOSFET in a single pulse, one obtains D_{it}. The number of carriers is normally inferred from a measurement of the substrate current. By varying the height of the pulse, one probes the band-gap at different energies.

3.2 Physical Characterisation of Materials

3.2.1 Transmission Electron Microscopy

Before a sample can be examined with a transmission electron microscope, it must be thin enough to allow a significant number of electrons to pass through, i.e. electron transparent (a few 100 nm). This is achieved by first mechanically grinding and polishing the sample and then using ion guns to erode the sample further to create the final, very thin viewable region. In the case of a silicon wafer, generally the region of interest is

at the surface and can be viewed from above (plan-view TEM or PVTEM) or in cross-section (XTEM). The former simply requires a small piece of substrate to be ground and polished from the back until it is about 30 μm thick, whereas the latter requires glueing to silicon support blocks and then grinding and polishing both sides to achieve the same thickness.

For grinding and polishing, samples are stuck with wax onto a flat metal block or glass slide, the latter having the advantage of allowing one to observe light passing through the sample when it is thin enough. Once sufficiently thin, a small copper ring with ≈ 1 mm aperture is glued on to the surface over the region of interest, and excess material is removed. The is wax melted to allow the, now supported, sample to be lifted off of the block/slide.

After a clean in boiling Isopropanol, a small region of the sample is made electron transparent by the use of a precision ion polisher (PIPS). Ions (in this case argon) are accelerated at the sample surface at a shallow angle (3 $^\circ$ - 5 $^\circ$) and on impact, erode material. The sample is constantly rotated, thus resulting in a conical dimple with a pinhole at the centre. The 100 μm^2 of material surrounding this hole is electron transparent, generally the thinnest material at the hole edge offers the best TEM images.

The sample is then loaded into the electron microscope, which consists of an evacuated column through which electrons, emitted from a filament, travel. They are focused by a series of magnetic lenses before transmitting through the sample and finally onto a fluorescent screen, where the image is viewed. Electrons are accelerated with a typical voltage of $\approx$100 kV resulting in a wavelength of less than 0.01 nm. This gives a theoretical image magnification of 100,000 and resolution of significantly less than 1 nm [Schroder, 2006]. However, in practice it is the aberrations of the magnetic lenses, which focus the beam, that limit the resolution and hence the 200 kV electron microscope used

to characterise samples in this thesis (Jeol 2000 fx) cannot resolve individual atoms.

Contrast in the image arises from the different electron absorption coefficients of the various material layers (for example SiGe layers of higher Ge-composition will appear darker) and the diffraction of electrons, called diffraction contrast. In general, the latter results in contrast due to the displacement of atoms from normal lattice sites. One form of this displacement is from dislocations within the material (diffraction contrast arising from this is specifically referred to as dislocation contrast), for example threading and misfit dislocations are common in SiGe virtual substrates. Also, if a layer of material is under global strain, for example a thin layer of germanium grown on a silicon substrate, then atoms in a particular plane are displaced from their normal lattice sites, which results in diffraction contrast, specifically referred to as strain contrast. A force is exerted by a strained layer on a relaxed layer at their interface, and hence the lattice distortion extends into the latter. Therefore, at these interfaces we expect strain contrast some way into the relaxed layer.

Both dislocation contrast and strain contrast in the SiGe heterostructure system can be maximised by tilting the sample to the appropriate 2-beam diffraction conditions, which are (220) and (004) respectively. XTEM reported in this thesis is fairly basic and used primarily to measure layer thicknesses; however, one should be aware of diffraction contrast when examining any TEM image and more a technical discussion of this, in SiGe heterostructures in particular, can be found in Nash [2005].

3.2.2 Secondary Ion Mass Spectroscopy

Secondary Ion Mass Spectroscopy (SIMS) is one of the most powerful techniques in analysing the composition of materials [Schroder, 2006]. An ion beam (the primary beam) is incident on the sample, which sputters material from the surface. The sputtered

material consists of uncharged particles, ions and electrons; the charged particles are focused by an electric field into a mass spectrometer. This measures the flux of a particular element. However, note that the signal ratio of several different elements does not necessarily correspond directly to the ratio of these elements within the material.

The primary beam sweeps an area of the order of 100 μm^2, gradually drilling into the sample surface and hence SIMS profiles can generally be obtained as a function of depth into the sample. A depth calibration is performed by sputtering a material for different amounts of time and then measuring the depth of the pit created by running a stylus over it. Note that different material layers will have different sputtering rates and therefore calibrating depth can be difficult. For this reason SIMS is often employed in conjunction with XTEM, allowing the SIMS profile to be matched to the TEM image.

Material is sputtered from the side of the pit also and hence it is standard procedure to use electronic (quadrupole) or optical (magnetic sector) gating so that the ion flux is not measured when the beam is sputtering at the extremes of the target area. Charge build-up on a sample surface can deflect the primary beam off target, sputtering material from the crater sides, or sample surface when it is not aimed there. Therefore, this effect must be avoided to obtain an accurate SIMS profile.

Chapter 4

Germanium p-MOSFETs with High-k Dielectric

4.1 Low Temperature Probe Station Development

Many low temperature measurements (down to 4.2 K) were performed on the devices reported in this thesis. A low temperature probe station is ideally suited to the task as it allows many devices to be probed in a single cooling cycle. This is in contrast to a cryostat, which generally requires devices to be mounted onto a chip packet and then gold wire bonded to the contact pads, hence limiting the number that can be measured in a single cycle. Also, the devices reported here have TiN contact pads — ill-suited to bonding.

A Desert Cryogenics low temperature probe station (figure 4.1) was used to perform electrical measurements. The probe station is a completely closed chamber as it needs to be under vacuum when cooling for thermal insulation and removal of moisture from the chamber environment. A vacuum of $\approx 10^{-6}$ mbar is achieved with

a two-stage pumping system consisting of a rotary roughing pump and a small turbo pump.

There are four probe arms operated with three micrometers each, allowing movement on all axes. The range of motion in each direction is large enough to allow any point on the substrate chuck (3 cm $\times$ 4 cm) to be reached by any of the four probe needles. Probe arms are able to move while maintaining vacuum with bellows, which expand and contract thus keeping the chamber sealed.

The region around the sample is kept free of electrical noise by a shield, which the probe arms enter through metal-braided holes that maintain a closed conducting surface around the sample area. Probe needles are thermally connected to both the shield and the substrate chuck with thick copper braids to ensure that the probes cool somewhat before having contact with the sample. A glass window is present above the sample, in the shield and the vacuum chamber, to allow the probe needles to be aligned with a microscope situated above. The shield window is lead plated so not to compromise the electrical shielding.

Temperature sensors are present both at the shield and substrate chuck, which are monitored electronically by a Lakeshore 330 temperature controller. This also controls two heaters, one under the substrate chuck and one under the shield, automatically using a negative feedback algorithm to achieve the user-specified temperature. Cryogenic liquid (nitrogen or helium) is contained in a closed dewar. The internal pressure of the dewar pushes liquid up a transfer tube, where it evaporates and flows through into the low temperature probe station. It travels under the substrate chuck and back out to vent.

Although a commercial piece of equipment, successful measurements require that it is well maintained and installed correctly. This was not the case when initial

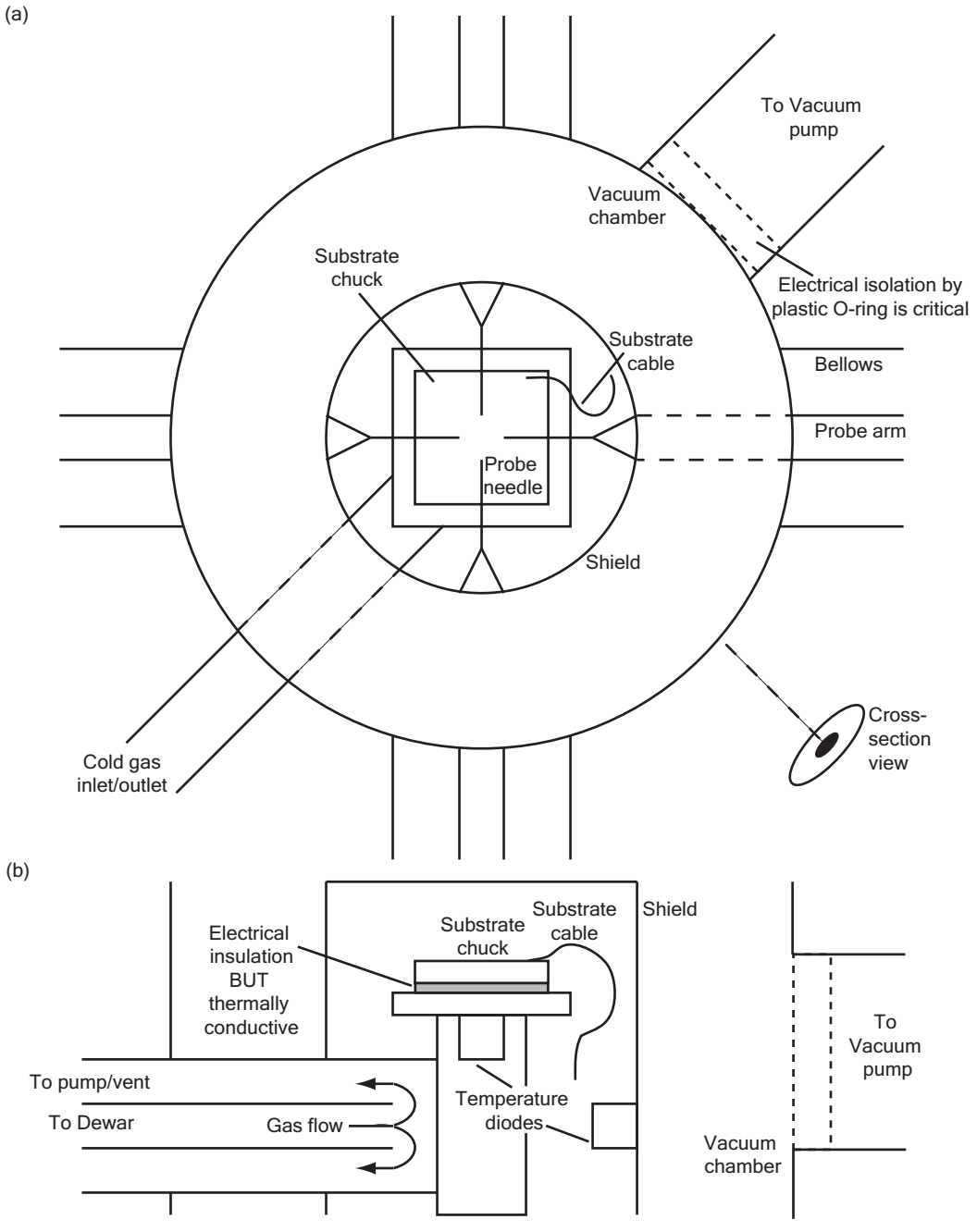


Figure 4.1: Schematics of the low temperature probe station from above (a) and in cross-section (b).

low temperature measurements were carried out on the devices reported in this chapter and hence considerable time and effort was put in by the author to rectify this. Simple, albeit non-trivial, problems included insufficient electrical isolation between noisy vacuum pumps and probe station resulting in large noise in electrical measurements. The solution was merely the application of an electrically insulating vacuum O-ring in the appropriate place (between pump and probe station). Making matters worse, was a temporary solution to the noise problem - implemented before the author's involvement - consisting of a glass slide placed between sample and substrate chuck. Although this prevented electrical noise from interfering with measurements, the glass slide was also thermally insulating, meaning that it was very unlikely that the sample was reaching the desired low temperature and indeed measurements were not repeatable with this set-up. Once the electrical noise had been eliminated at it's source, the glass slide was removed and the operator could be sure that the sample reached the chuck temperature. Although the chuck was now earthed cleanly, some measurement techniques require a signal to be applied or measured at the substrate contact. Fortunately, the manufacturer supplies a chuck with a thin, electrically insulating layer that conducts heat and once implemented all electrical measurements could be performed at low temperature.

Once this work was complete, measurements at 77 K (that of liquid nitrogen) were repeatable; however, those at other temperatures, including 4.2 K, were not. After much experimentation it was found that, for any measurement to be successful, it is important that the probe needles are at the same temperature as the sample and as they are thermally connected to the shield, care must be taken to ensure that the shield temperature is as close to the chuck temperature as possible. Measurements at 77 K with liquid nitrogen are the most simple, as the shield and the substrate chuck both reach the required temperature. However, as temperatures above this are achieved by use of

Step	Action
1	Start cryogenic flow
2	Enter temperature set point if required
3	Reduce flow so heater is on as lower power output as possible
4	Allow shield to stabilise (drift$<$1 °C/minute)
5	Contact probe tips to unused sample area for 20–30 minutes
6	Reconnect probes to sample area to be measured for 10 minutes
7	Reconnect and measure
8	Repeat measurement after 5 minutes — check for connection failure or data indicating temperature shift
9	Repeat step 8 until measurement consistent
10	Move probes one at a time to contact next structure
11	Measure and repeat (from step 8)
12	Once all measurements at the current temperature are complete, raise probes and warm/cool to next temperature

Table 4.1: The measurement procedure, developed for the reliable low temperature electrical characterisation of Ge MOSFETs, using a Desert Cryogenics low temperature probe station.

the heaters, in general the shield temperature is greater than the chuck temperature. This is also the case for 4.2 K measurements with helium, where the shield will only reach about 12 K.

To minimise the temperature difference, the flow rate of the cold vapour must be adjusted so that the chuck remains at the required temperature while the heater power can be as low as possible. For this reason a computer program was developed by the author, which plots the temperature of the chuck and shield over time so that the user can better adjust the flow rate to that required for minimising the shield temperature. In the case of liquid helium measurements, the flow rate must be as high as possible. The flow rate can be increased by pumping the vented gas through with a rotary pump; however, this can have the effect of an unstable flow rate which causes temperature fluctuations.

Probe needles deform with changes in temperature and hence cannot be in contact with the sample whilst cooling, which means they are generally warmer than the sample before a measurement. Only when the needles have been in contact with the sample surface for some time will they be at the desired temperature, but they can warm up again between measurements. For this reason the measurement procedure outlined in table 4.1 was developed by the author.

4.2 Overview

The devices reported here are germanium p-MOSFETs (with accompanying n-MOS capacitors) fabricated on the Pilot Line at IMEC — a large microelectronics research facility in Leuven, Belgium. The Pilot Line is used to fabricate non-silicon devices on 200 mm diameter wafers. The device batch is part of the Ge/III-V CMOS project with the ultimate goal of high-performance germanium p-MOSFET and III-V semiconductor (e.g. GaAs) n-MOSFET devices, fabricated simultaneously on a single wafer that would have separate regions of both material types. The wafers, which the devices reported here are fabricated on, are standard (100) silicon substrates with a several micron thick fully-relaxed germanium layer on the surface. This layer was grown by chemical vapour deposition by ASM international — a large corporation specialising in wafer processing technologies.

Three samples were provided to the author with the principal aim of a low-temperature electrical characterisation. All samples were cleaved from a single wafer at IMEC — two from the centre (sample C and sample H) and one from the edge (sample E). Sample H underwent an additional 400 °C anneal after fabrication (known as a post-metallisation anneal or PMA) in a pure hydrogen gas atmosphere. The mask set, used to define the electrically active areas during fabrication, contains n-MOS capacitors, p-

MOSFETs from 10 μm down to 125 nm gate lengths and MOSFETs with no gate stack (null-FETS). MOSFETs are arranged in lines (referred to as arrays) so that source/drain contact pads are shared between two neighbouring devices in order to save space on the wafer surface. The MOSFET gate lengths decrease down the length of each array. The germanium layer is n-doped by ion-implantation to give a substrate doping density of 7×10^{15} cm^{-3}.

The gate-stack is formed by partially oxidising a thin silicon cap (≈6 Å) on the Ge-surface by a dip in ozonated water. The hafnium-oxide high-k dielectric (≈4 nm thick) is immediately deposited via atomic layer deposition, directly onto the silicon dioxide to prevent any further oxidation. This results in a final gate dielectric stack from the Ge-surface of: Si/SiO_2/HfO_2. The reported capacitive equivalent thickness is 16 Å [Zimmerman et al., 2006].

Metallisation that electrically connects the devices to contact pads, is still in the development stage with the aim of obtaining as low as resistance as possible. Three different source/drain geometries are available with different areas of metal overlapping the doped source/drain regions. No detailed investigation into the different source/drain configurations was performed; but there was however, a configuration that had a lower source/drain resistance than the others. For simplicity we shall refer to these as "optimised source/drain contacts" and the others, with a higher resistance, as "unoptimised".

The source/drain resistance is reported to be ≈55 Ω for the MOSFETs with optimised source/drain contacts, with a large contribution from the contact pad itself. By changing the position of a probe needle on the source/drain contact pad it was shown to vary peak transconductance by 25% in a 125 nm device [Zimmerman et al., 2006]. This makes reliable R_{sd} extraction challenging as probe needle placement clearly has a large effect on this parameter. Also contributing to the resistance, is the lightly

doped drain (LDD) extension region, which is incorporated into the MOSFET design to reduce short channel effects by lowering the doping density close to the channel region.

Already published [Zimmerman et al., 2006] are the room temperature device characteristics. This includes the demonstration of an effective mobility $3\times$ greater than the silicon p-MOSFET universal mobility curve [Takagi et al., 1994a]. This was extracted from a MOSFET on a sample that had undergone hydrogen anneal as samples that had not, showed large variation in transconductance and threshold voltage and had generally lower mobilities. This was qualitatively demonstrated to be due to interface states with charge pumping measurements — the hydrogen anneal resulted in a reduction in interface state density and an improvement in device performance and uniformity.

We begin with a source/drain resistance extraction in section 4.3 before a room temperature electrical characterisation of MOS capacitors (section 4.4) and MOSFETs (section 4.5). A full low-temperature electrical characterisation is detailed in section 4.6. With the data obtained, a detailed, quantitative investigation of the interface states (section 4.7) and carrier mobilities (section 4.8) of these devices is reported.

4.3 Source/Drain Resistance

The source/drain resistance was measured at 300 K, 77 K and 4.2 K using the low temperature probe station. As will be shown in later sections, there is a large random variation in device performance between MOSFETs on a given sample, and a more systematic variation between samples. Sample H, which had been annealed in hydrogen, was chosen as it is reported to contain the most uniform MOSFETs and it is assumed that the hydrogen anneal has no effect on the source/drain resistance. First, a suitable array of working devices was selected, which was considered to be that which contained the best performing devices on initial inspection. That was defined as those with the

highest transconductance, lowest gate leakage (generally not an issue anyway) and good rectifying behaviour. This initial assessment was made from I_d-V_g measurements with V_d=-50 mV (not shown).

A single array of devices was chosen, rather than individually top performing devices of different gate lengths scattered over the sample, to minimise the distance between devices and thus time between contacting on the low temperature probe station. This helps to ensure that the probe needles remain cool between measurements at low temperatures, which is of particular importance at 4.2 K. However, this resulted in many device lengths on the mask not being included in the R_{sd} extraction, as they either performed less well than other devices in the same array, did not function at all, or the contact pads became damaged before sufficient data was collected.

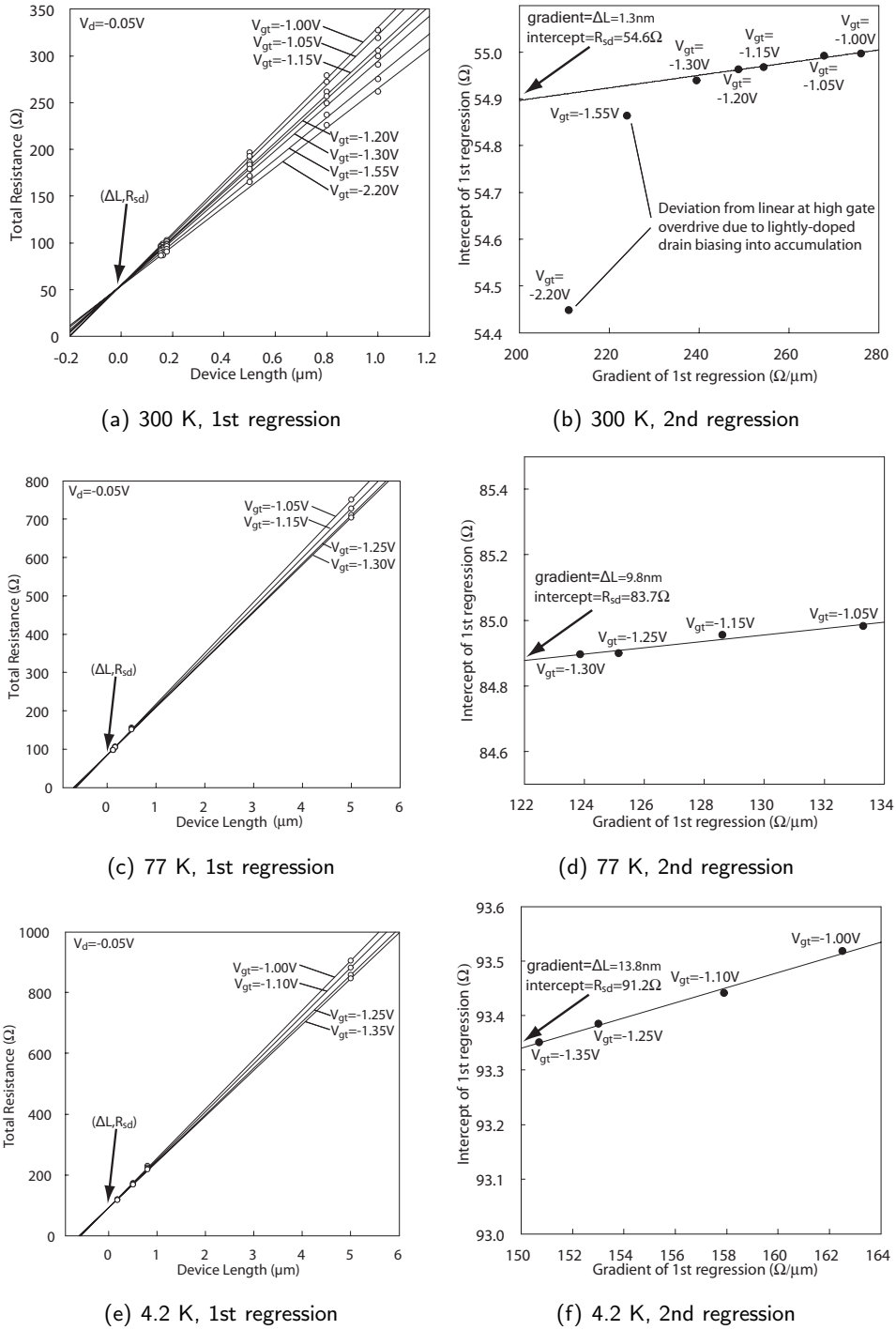
As mentioned, there are different types of source/drain contacts on the mask set. The R_{sd} extraction was performed on an array with the optimised source/drain configuration, i.e. that which has the lowest resistance. I_d-V_g measurements with V_d=-50 mV (not shown) were performed on the MOSFETs of different gate lengths. From the data obtained, a plot of $\frac{V_d}{I_d} = R_{tot}$ for different gate overdrives against the written gate length L (that measured from the photo lithography mask) was made and is generally referred to as a 1st regression plot (see section 3.1.5). Clearly, gate overdrive requires the threshold voltage to be known. This was measured from the same I_d-V_g data using the transconductance method (see section 3.1.4), which is particularly suited as it is reliable to short gate lengths [Tsuno et al., 1999]. This was performed at room temperature and repeated at 77 K and 4.2 K. A second regression plot, which is of the intercept against the gradient of the lines of best fit for each gate overdrive in the 1st regression plot, is displayed next to each of the corresponding first regression plots in figure 4.2. The values at which the straight line of best fit of the 2nd regression plot

intercepts the vertical and horizontal axes, is the R_{sd} and ΔL values respectively of these devices.

Note that only the lowest gate overdrives are used to obtain R_{sd} and ΔL in the second regression plot at room temperature, as higher gate overdrives result in a decrease in source/drain resistance, as can be seen in figure 4.2(b). This can be explained by the lightly p-doped drain extension (lightly doped drain or LDD) that overlaps the gate stack. It is thought that, as a greater gate overdrive is applied, the LDD is biased into accumulation, thus decreasing in resistance and in turn decreasing the overall R_{sd} value.

To gain further insight into the effect of the LDD, an array of null-FETs (MOSFETs with no gate stack) was electrically characterised at 300 K and 77 K. These devices vary in width along a given array and due to the lack of gate stack, only have source and drain contacts. The drain voltage is ramped between a small positive bias and small negative bias, which results in a proportional current and allows the calculation of resistance between the contacts. Due to the lack of channel region, the resistance measured for each should simply be equal to the R_{sd} value extracted for a standard MOSFET at a given device width. We expect this resistance to be inversely proportional to device width and hence the resistance is plotted against $1/W$ to give a straight line fit and allow R_{sd} to be inferred, by extrapolation, for a null-FET of any width (figure 4.3). However, this value of R_{sd} will not include a contribution from the lightly doped drain, as the highly doped source and drain regions of the null-FET will dominate due to their close proximity.

According to figure 4.3, the resistance of a 10 μm wide null-FET has decreased from 47 Ω at room temperature to 36 Ω at 77 K. This is in contrast to a standard MOSFET, that has increased source/drain resistance with decreased temperature, as evident in figure 4.4(b). Furthermore, the source/drain resistance of a MOSFET is al-



(a) 300 K, 1st regression

(b) 300 K, 2nd regression

(c) 77 K, 1st regression

(d) 77 K, 2nd regression

(e) 4.2 K, 1st regression

(f) 4.2 K, 2nd regression

Figure 4.2: First and second regression source/drain resistance plots for hydrogen annealed p-MOSFETs with optimised source/drain contacts. These were selected for their uniformity and it is assumed that the anneal has no effect on the source/drain resistance.

ways significantly greater than that of a null-FET with the same width, at the same temperature. Therefore, as the only difference in the source/drain regions of the two devices is the lack of lightly doped drain in the case of the latter, we must conclude that:

1. The lightly doped drain contributes significantly to the source/drain resistance of the MOSFETs.
2. The LDD increases in resistance with decreasing temperature.
3. There would be a decrease in the source/drain resistance of a MOSFET with decreasing temperature if the lightly doped drain was not present.

An explanation is that the dopants in the lightly doped drain of a MOSFET freeze-out when the temperature is decreased. The lightly doped drain makes a large contribution to the source/drain resistance at room-temperature and at low temperatures this contribution increases greatly.

The values of R_{sd} and ΔL plotted in figure 4.4 as a function of temperature are used to correct all relevant electrical characterisation data from here in. The reported problem [Zimmerman et al., 2006] of probe needle placement having a large enough effect on the source/drain resistance to alter the peak transconductance of a 125 nm gate length MOSFET by 25%, makes the accurate electrical characterisation of short channel length MOSFETs difficult. Indeed this was an issue when performing the above R_{sd} measurements on shorter devices. Therefore shorter device lengths are ignored and the focus of the rest of this chapter is on the characterisation of 10 μm$\times$10 μm MOSFETs.

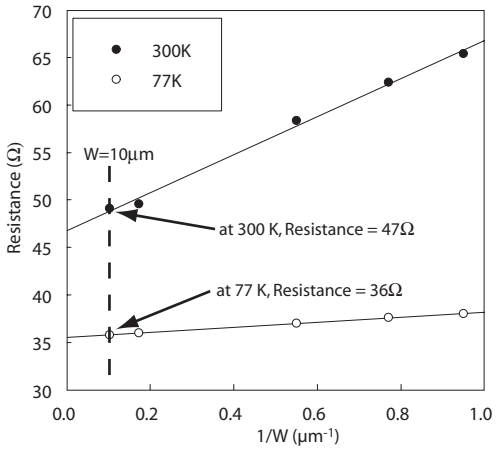
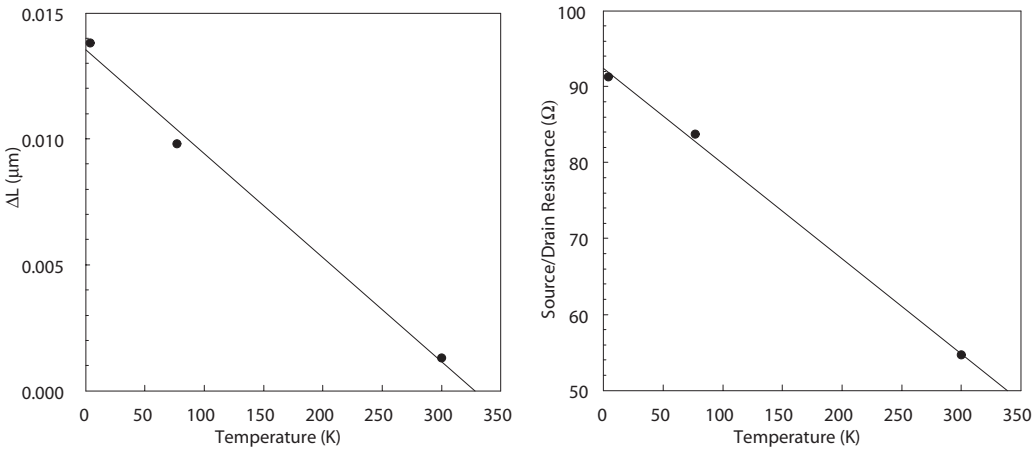


Figure 4.3: Resistance of 0-gate length MOSFETs at 300 K and 77 K measured by ramping the drain voltage from -1 V to 1 V and taking the gradient of the resulting V-I plots (not shown). In contrast to finite gate length MOSFETs, here we observe a decrease in source/drain resistance with temperature due to the lack of lightly-doped drain.



(a) (b)

Figure 4.4: Measured values of ΔL (a) and R_{sd} (b) as a function of temperature for Ge p-MOSFETs with optimised source/drain contacts. Both parameters appear to decrease linearly with increasing temperature; however, due to lack of data and device non-uniformity, this relationship is far from certain.

4.4 MOS Capacitor Characteristics

Room temperature C-V curves, measured at various frequencies using the 4284A precision LCR meter, for the n-MOS capacitors on the wafer centre (sample C) and hydrogen annealed (sample H) are shown in figure 4.5. Immediately apparent is the deeper 500 Hz curve through depletion in sample H. Note that minority carriers (holes) respond up to frequencies of 50 kHz. Interface states show a clear response at 10 kHz but at frequencies higher than 50 kHz, neither minority carriers nor interface states respond. At 1 MHz, attenuation in accumulation is apparent. This is caused by the resistance of the substrate contact and/or gate contact affecting high frequency measurements. For this reason, 100 kHz was chosen to measure the oxide capacitance (figure 4.6): 0.022 F m$^{-2} \pm 0.001$ F m^{-2} giving a capacitive equivalent thickness of 0.17 nm, consistent with the value of 16 Å already reported [Zimmerman et al., 2006].

There is little variation between capacitors located on a particular sample, although it should be noted that each sample does not contain many. Capacitors located at the wafer edge (sample E) have slightly deeper depletion regions at low frequency (not shown) than those on sample C, suggesting a lower interface state density. Those on sample H show by far the deepest valley at 500 Hz, probably due to the reported reduction in interface state density after a hydrogen anneal [Zimmerman et al., 2006].

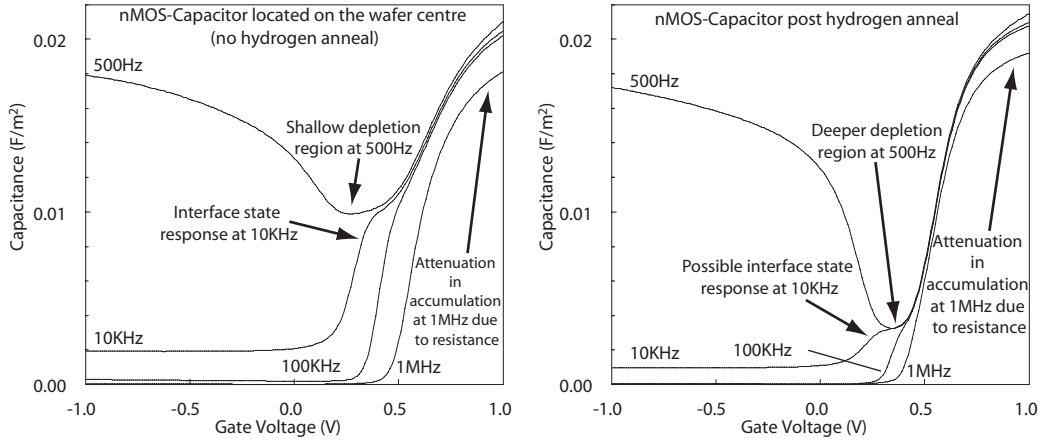


Figure 4.5: A family of C-V curves measured for n-MOS capacitors with an area of 10161 μm^2 located on a sample C, from the wafer centre (a) and sample H, that underwent hydrogen anneal (b). Note the improved low-frequency C-V curve after hydrogen anneal with a deeper depletion region, which indicates a reduction in interface state density.

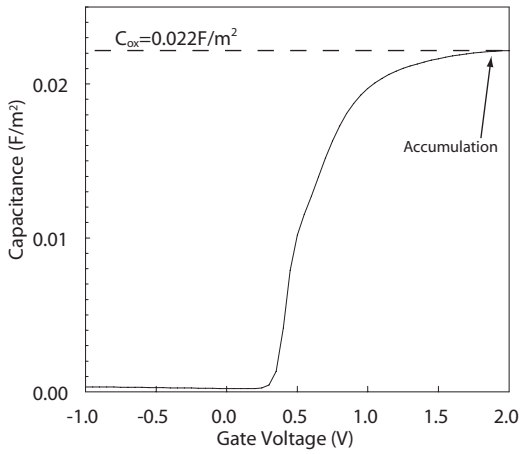


Figure 4.6: A 100 kHz C-V curve from a 10161 μm^2 hydrogen-annealed n-MOS capacitor used to determine $C_{ox} = 0.022$ F m^{-2} from the maximum capacitance in the accumulation region.

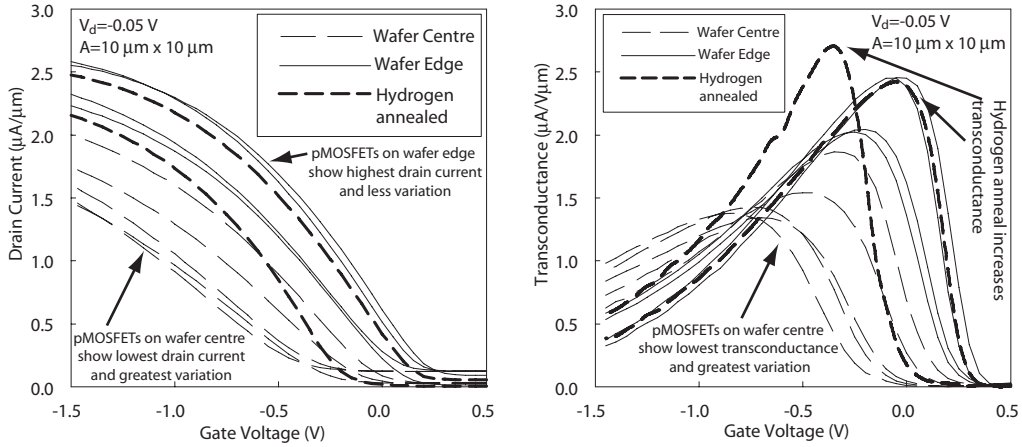


Figure 4.7: I_d-V_g (a) and corresponding transconductance curves from the derivatives (b) measured at $V_d = -0.05$ V for 10 μm$\times$10 μm p-MOSFETs located on the wafer centre (sample C) and wafer edge (sample E), and those that underwent a hydrogen anneal (sample H). In general, devices located on the wafer edge are more uniform and have on-currents and transconductances than those in the centre. A hydrogen anneal can lead to a greater transconductance.

4.5 MOSFET Characteristics

I_d-V_g was performed with a low drain bias (-50 mV) on over one hundred 10 μm$\times$10 μm MOSFETs at room temperature with the aim of investigating variation in device characteristics and the effect of the post metallisation anneal. The standard room temperature probe station was used because of the large number of devices that were probed; however, a few measurements were repeated in the low temperature probe station at 300 K to ensure the different probe needles and contact pressure did not affect the measurements.

I_d-V_g characteristics and the corresponding transconductance curves are shown for many devices, from all samples, in figure 4.7. This demonstrates a large variation in device performance over each sample and between samples. Sample E (wafer edge) is the

most uniform, whereas sample H (hydrogen annealed) contains devices with the highest transconductance and on-current. This is in contrast to sample C (wafer centre), which is the least uniform and contains the devices with the worst on-state characteristics.

Threshold voltage was extracted for each MOSFET by linear extrapolation from the point of steepest gradient in the transconductance curves to the gate voltage axis (transconductance method, details given in section 3.1.4). The extracted threshold voltage correlates strongly to the peak transconductance (shown alongside low temperature data in figure 4.8) of a given MOSFET. The threshold voltages of MOSFETs located on samples C and E vary from ≈-0.3 V to ≈0.3 V, giving a range of ≈0.6 V. MOSFETs on sample C generally have more negative threshold voltages, whereas those on sample E have more positive threshold voltages with less variance. Sample H contains devices which fit the peak transconductance-threshold voltage trend observed for the other two samples; and some that form a separate trend. Due to the more limited number of devices available on sample H it is more difficult to infer a clear trend for these MOSFETs.

The off-state current device characteristics (figure 4.9(a)) also show variation although, unlike measured threshold voltages and transconductances, this is not systematic with sample and devices that have undergone a post-metallisation anneal (sample H) show no improvement in this respect. Gate induced drain leakage (GIDL) is apparent at high positive gate biases, indicated by a large increase in off-current.

The minimum subthreshold slope S was extracted from the I_d-V_g data for each device with care taken to remove noise that would lead to an erroneous value. The subthreshold slope varies between devices, from about 100 mV/decade to 400 mV/decade, but is not related to the minimum off-current of a given device. One must be careful that this is the case as in general, a device with a higher off-current can exhibit a higher

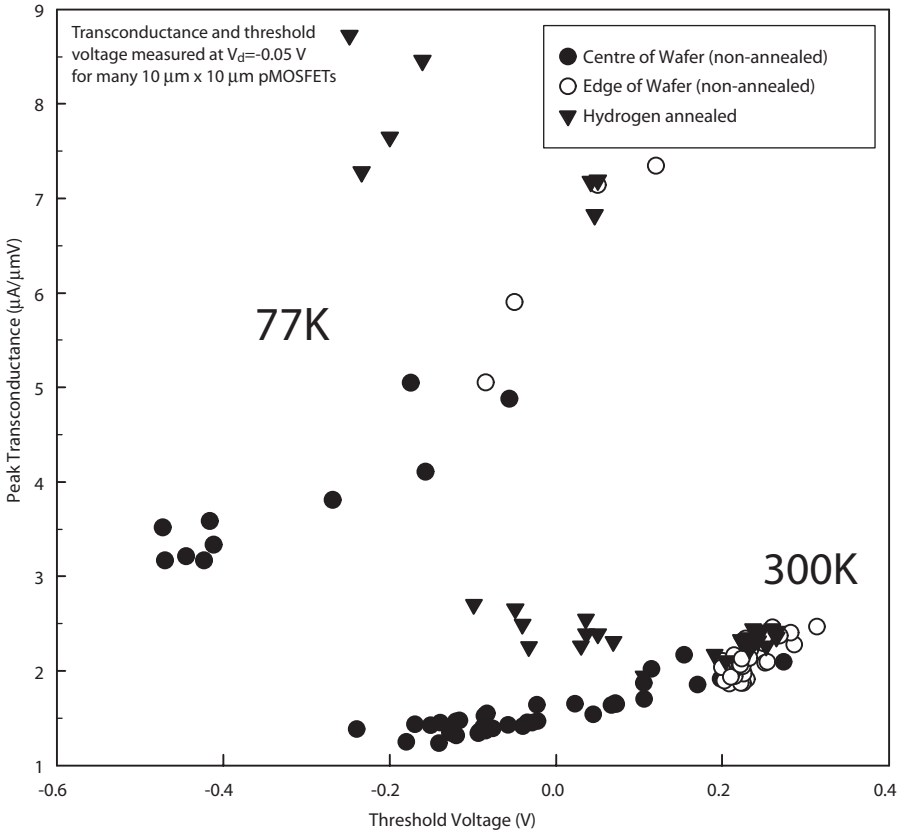
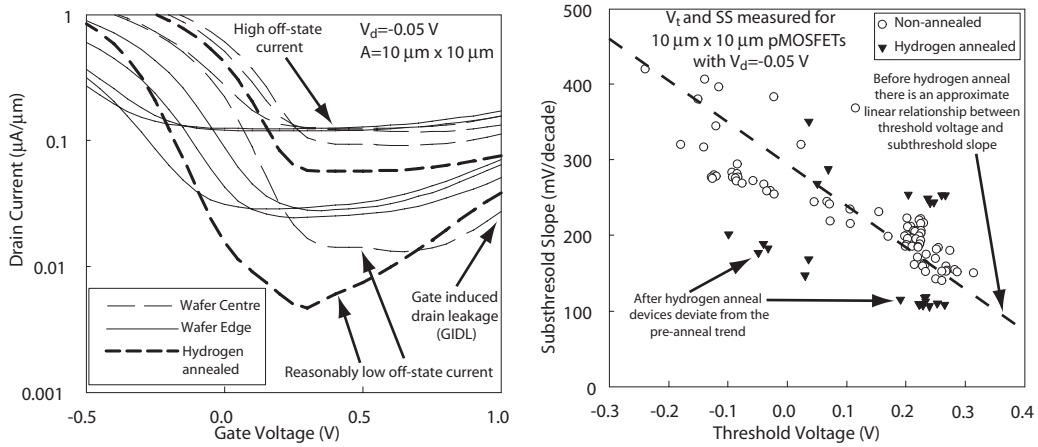
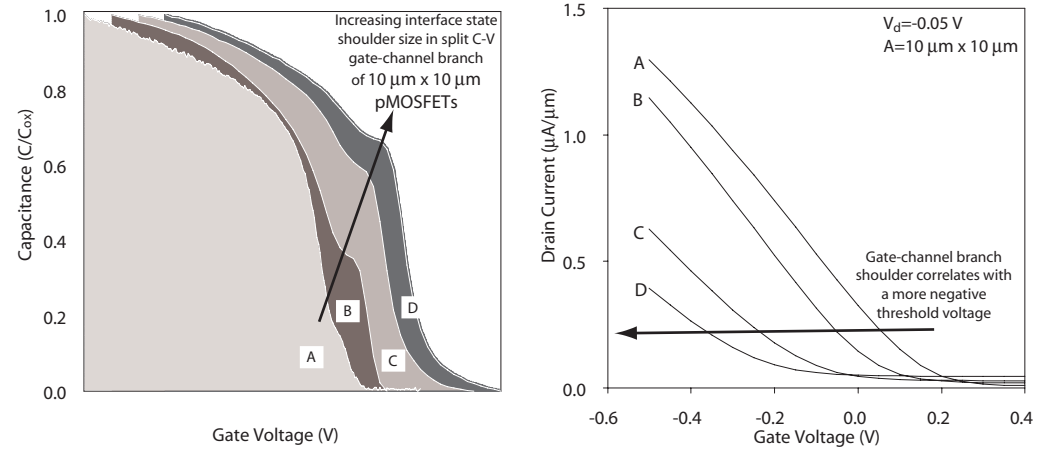


Figure 4.8: Peak transconductance against threshold voltage measured at $V_d = -0.05$ V for many 10 μm$\times$10 μm non-annealed (samples C and E) and hydrogen annealed (sample H) p-MOSFETs, demonstrating a strong correlation between the two parameters at both 300 K and 77 K.



(a) (b)

Figure 4.9: A sample of 10 μm×10 μm p-MOSFET characteristics in the off-state at $V_d = -0.05$ V (a) demonstrating a large variation in off-state currents which show no correlation between wafer position or hydrogen anneal. Subthreshold slope plotted against threshold voltage for many p-MOSFETs both non-annealed and hydrogen annealed (b).



(a) (b)

Figure 4.10: Gate-channel branch split C-V curves measured at 10 kHz (a) and I_d-V_g (b) for 10 μm×10 μm p-MOSFETs A, B, C and D. A higher interface state density, as qualitatively indicated by the interface state shoulder size in the GC-branch, corresponds to a negative threshold voltage shift, as demonstrated by the I_d-V_g curves.

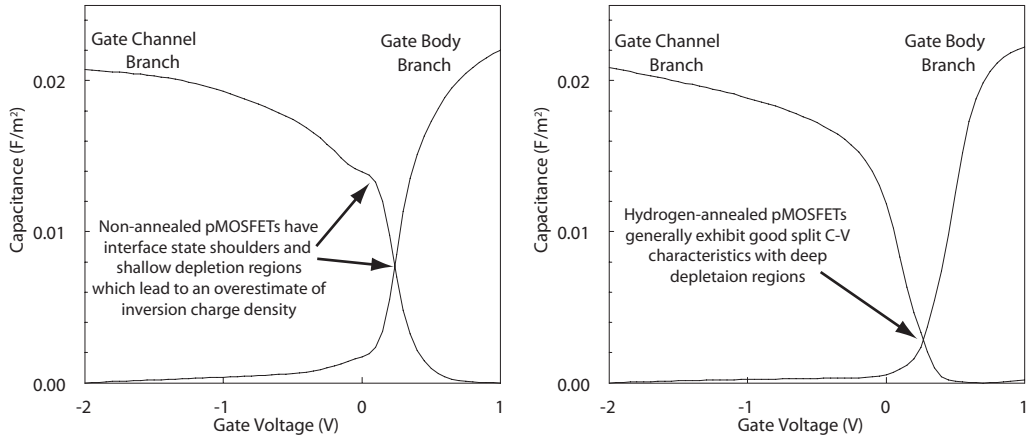


Figure 4.11: 100 kHz split C-V for a 10 μm$\times$10 μm p-MOSFET located at the wafer centre (a) and post hydrogen anneal (b). Note the shallow valley over the depletion region in the former, in contrast to the deep, well-defined valley present in the latter due to lack of interface state shoulder.

subthreshold slope due to the smaller difference between I_{on} and I_{off}. As there is no correlation between the off-current and the subthreshold slope measured here, it is safe to assume that higher off-currents do not increase the measured subthreshold slope. As with transconductance, subthreshold slope does correlate with threshold voltage (figure 4.9(b)).

Split C-V was performed on several devices from each sample at 10 kHz and there is a prominent feature in most of the gate-channel branch curves in the form of a shoulder. For a given MOSFET, the larger this shoulder is, the more negative the threshold voltage. This is demonstrated by a simple comparison of the GC C-V curve to the drain current for a few MOSFETs (figure 4.10). It is difficult to imply a quantitative relationship as the size of the shoulder is not well defined. This shoulder remains even at 100 kHz (figure 4.11(a)) despite interface states apparently not significantly affecting MOS capacitor characteristics at this frequency. Attenuation of both branches of the

split C-V occurs at frequencies greater than this (not shown — but demonstrated in 1 MHz MOS capacitor curves in figure 4.5). Hence, reliable split C-V cannot be performed on MOSFETS with a high density of interface states as no intermediate frequency unaffected by interface states and parasitic resistances exists [Zhu et al., 2004]. However, sample H contains some devices that do have very good split C-V characteristics at 10 kHz/100 kHz (figure 4.11(b)) with no interface state shoulder and a slightly lower depletion capacitance in the GB branch.

4.6 Low Temperature Measurements

As interesting, generally unexpected, strong correlations between certain MOSFET parameters (threshold voltage, peak transconductance, subthreshold slope and split C-V GC-shoulder) are observed at room temperature, we aim to see if these are still present at low temperature. As the low temperature probe station has a limited sample area only one or two chips could be placed on the chuck giving a total of twelve 10 μm$\times$10 μm devices per cycle — assuming 100% device yield. 77 K is an appropriate temperature as liquid nitrogen is relatively cheap and readily available meaning that many cycles can be performed. Also, both the shield and chuck reach 77 K and hence, once cool the probe needles remain at this temperature, which further simplifies the contacting procedure. In addition to the 77 K measurements, a more limited number of MOSFETs were characterised at 4.2 K using liquid helium.

I_d-V_g performed at 77 K and 4.2 K reveals that the variation in device performance observed at 300 K is still present at lower temperatures. Figures 4.12(a)-(d) show the drain current and transconductance as a function of gate voltage for a device located on sample C (wafer centre) and a top-performing device located on sample H (hydrogen annealed). As expected, all devices improve in performance with decreasing

temperature, showing greater drain current and transconductance at 77 K and 4.2 K than at room temperature. The split C-V characteristics at 10 kHz (figures 4.12(e),(f)) demonstrate very sharp transitions between accumulation, depletion and inversion. Note that there is now no interface state shoulder, which could be explained by an increased response time of the interface traps.

There is a strong correlation between peak transconductance and threshold voltage (figure 4.8) at 77 K, not dissimilar to that observed at room temperature (shown in the same figure). The peak transconductance increases with decreasing temperature and the threshold voltage for any given device becomes more negative. This change in threshold voltage is expected as the Fermi-level in the channel material (E_F) tends towards the conduction band edge due to the decrease in intrinsic carrier concentration (figure 2.2). Hence, the strong inversion condition $\psi_s = 2\phi_b$ corresponds to a more negative threshold voltage than at room temperature. However, we observe that a device with a more negative threshold voltage at room temperature experiences a greater negative threshold voltage shift with decreasing temperature, compared to a MOSFET with a more positive threshold voltage at room temperature (figure 4.13). One would expect all MOSFETs to undergo the same threshold voltage shift and the plot of $V_t(77\ K)$ against $V_t(300\ K)$ to give a gradient of unity with a negative intercept. This is assuming that the substrate dopant concentration (implantation and activation anneals were highly controlled) and intrinsic carrier concentration (a material property) is equal over the whole substrate from which the samples were cleaved.

Whereas at 300 K the off-state drain current at low drain bias varied greatly between MOSFETs, at 77 K devices fall into two groups. Some exhibit minimum off-currents of a single femto-amp (1×10^{-7} μA μm^{-1} when normalised by the device width of 10 μm) — the smallest measurable current with the 4156c parameter analyser.

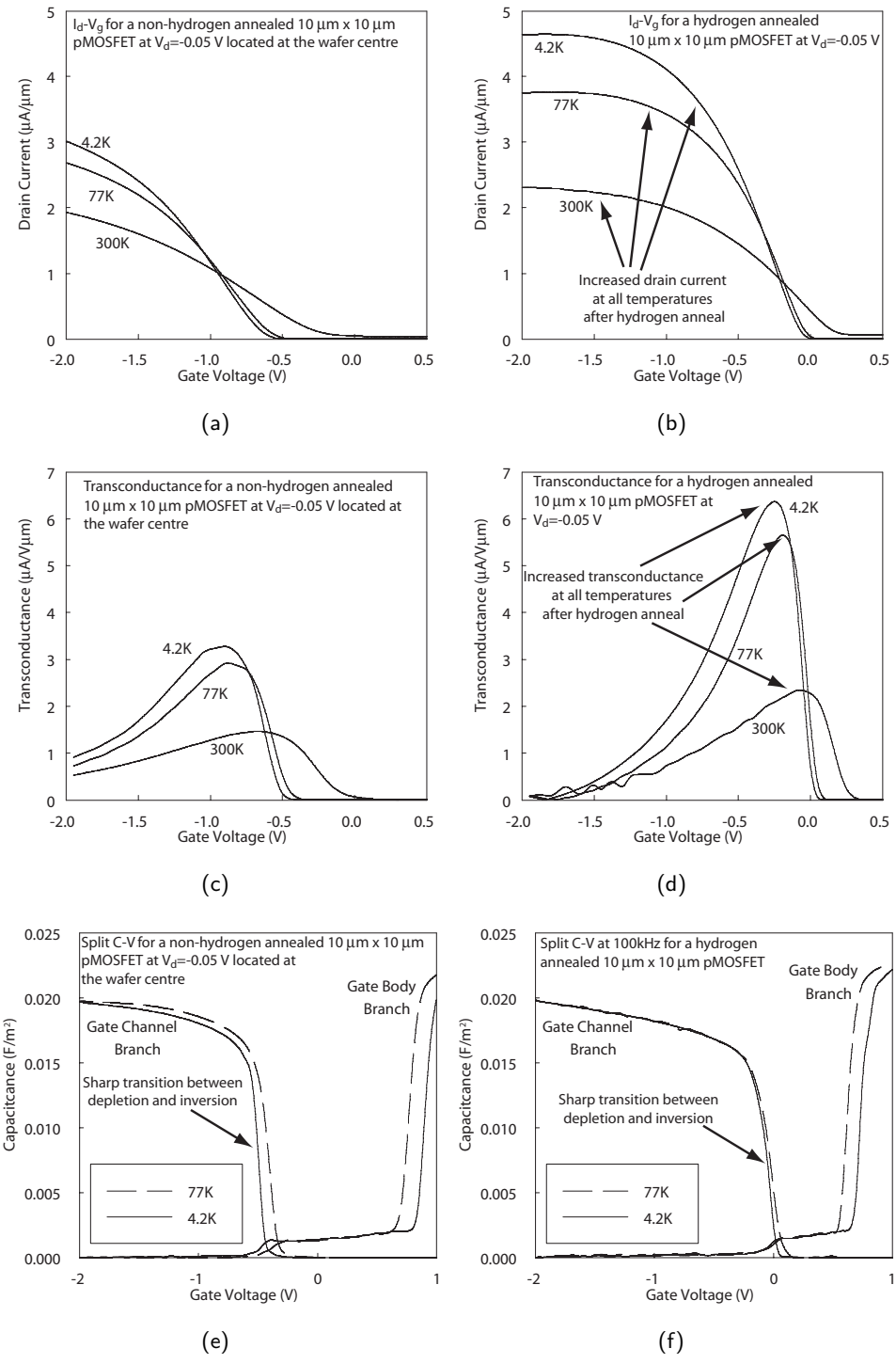


Figure 4.12: I_d-V_g, transconductance and split C-V at temperatures of 300 K, 77 K and 4.2 K on a p-MOSFET located at the wafer centre (a), (c), (e) and post hydrogen anneal (b), (d), (f).

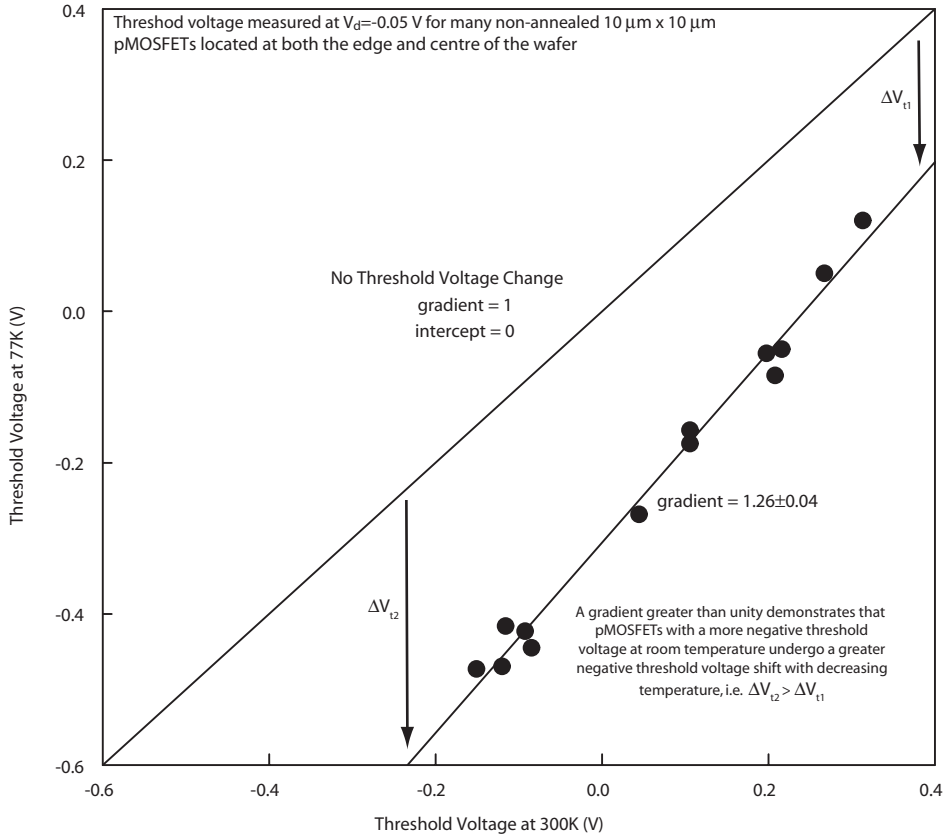
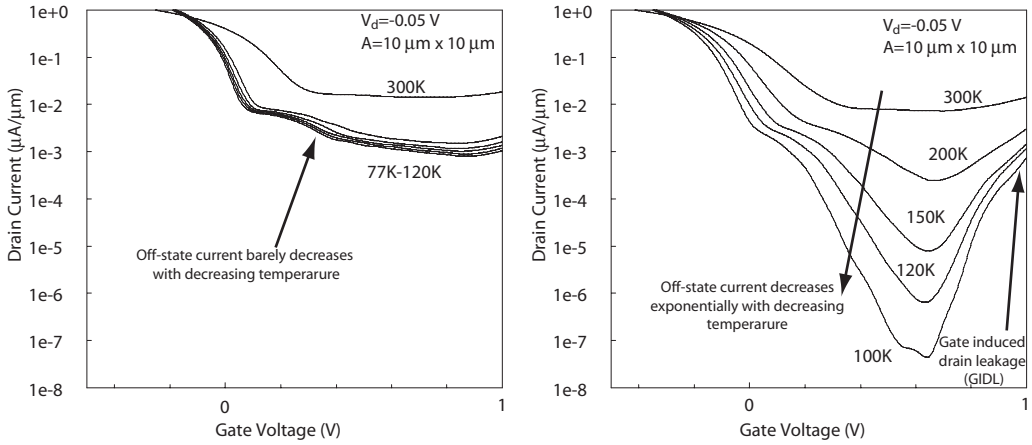


Figure 4.13: Threshold voltages measured with $V_d = -0.05$ V for many 10 μm$\times$10 μm non-annealed p-MOSFETs at 77 K against the same at 300 K. p-MOSFETs with a more negative threshold voltage at 300 K experience a larger negative threshold voltage shift with decreasing temperature, as indicated by a gradient greater than unity.



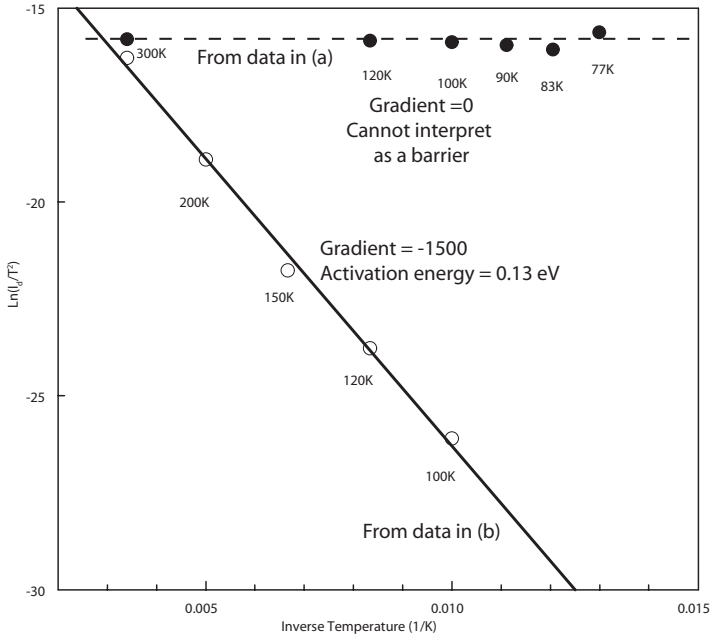


Figure 4.14: The off-state drain current with $V_d = -0.05$ V for two 10 μm$\times$10 μm p-MOSFETs (a,b) that are representative of the two different behaviours observed over all devices measured. Whereas the drain current in the former decreases exponentially with decreasing temperature, the latter sees almost no reduction in drain current. Corresponding Richardson plots (c) gives an activation energy of 0.13 eV for the latter and effectively no barrier at the source/drain for the former.

The others have a large minimum off-current of about 10 nA (1×10^{-3} μA μm^{-1}). As was the case at room temperature, the off-state drain currents do not correlate with sample, which implies that the hydrogen anneal has no effect. Several devices with high off-currents and several with low off-currents were selected and the off-currents of each measured at a range of temperatures. An example of each of the device behaviours is shown in figure 4.14(a) and 4.14(b), along with the corresponding Richardson plots (figure 4.14(c)).

MOSFETs with the lower off-state currents at 77 K give an activation energy of 0.13 eV. Although this could correspond to a single deep impurity level, it is more likely caused by the continuum of interface states in the gate/drain overlap region [Touhami and Bouhdada, 2005]. The devices with a large off-state current at low temperatures have a very small, or even no, activation energy and this cannot be interpreted as thermionic emission over a barrier.

4.7 Interface States

As it was proposed by Zimmerman et al. [2006] that N_{it} (determined from charge pumping), transconductance and threshold voltage are related, we now investigate the interface state density of the MOS devices in an attempt to explain the variation in performance observed in the previous sections.

4.7.1 MOS Capacitors

The capacitance data displayed in figure 4.5 is used to extract the interface state density in the band-gap, utilising the high-low frequency capacitance technique. Ideally the low frequency measurement should be quasi-static, where the capacitance is measured by incrementing a DC voltage applied to the gate. Even a small gate leakage current is

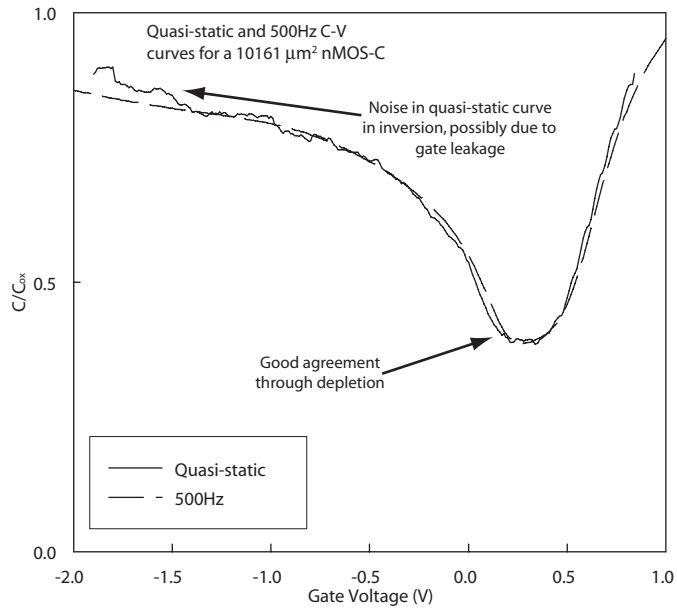


Figure 4.15: A comparison of a 500 Hz C-V curve with a Quasi-static C-V curve, measured for a 10161 μm^2 n-MOS capacitor. The good agreement through depletion demonstrates that all interface states respond to a 500 Hz signal.

problematic for these measurements as it causes a large amount of noise in the signal. A quasi-static measurement was performed using the 4156c with a gate voltage increment of -0.01 V and high signal averaging, in an attempt to eliminate noise. This is compared to a C-V measurement on the same capacitor with the 4284A precision LCR meter in figure 4.15. The latter was performed at 500 Hz with the same DC voltage increment but far lower signal averaging.

This shows that the two techniques agree and as the 500 Hz measurement has less noise, is quicker to perform, records the measured conductance and can be part of a sequence of measurements at different frequencies without changing connections; it is used in place of the quasi-static.

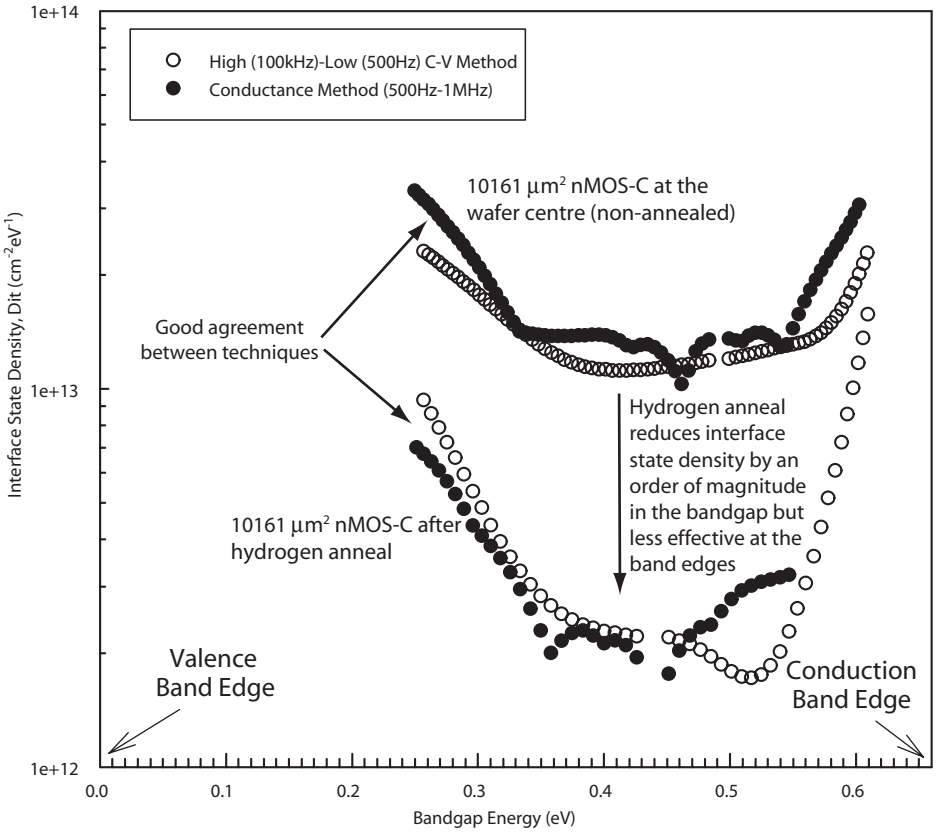


Figure 4.16: The density of interface states through the bandgap measured using the conductance technique and high-low capacitance technique on a non-annealed 10161 μm^2 n-MOS capacitor at the wafer centre and the same after hydrogen anneal. Of note is the agreement between the techniques and the order-of-magnitude reduction of interface sates in the bandgap after hydrogen anneal. At the band edges however, the hydrogen anneal appears to have less effect.

The high frequency was selected as 100 kHz, as it showed minimum attenuation due to substrate/gate contact resistance and no evidence of interface trap response. The interface state density as a function of band-gap energy is shown in figure 4.16. Flat-band voltage was needed for the calculation of band-gap energy using equation 3.22 and was extracted for each capacitor individually, using the technique described in section 3.1.2.

Clearly there is a large difference in the interface state density between samples C and H, with the hydrogen anneal (sample H) reducing it by one order of magnitude. Note that there is a high density of interface states throughout the band-gap and that this density increases towards both the valence band edge and the conduction band edge. Sample H shows a sharper increase in interface state density at the band edges than sample C, suggesting that the hydrogen anneal has reduced the interface state density significantly throughout the band-gap, but not as much towards the band edges.

The measured conductance over the range of frequencies was also used to calculate the interface state density as a function of energy using the method described in section 3.1.7. This agrees well with the capacitance technique (displayed in the same figure).

4.7.2 MOSFETs

In general, a shoulder present in the gate-channel branch of a split C-V curve indicate the presence of interface states (see section 3.1.7). As demonstrated for these Ge p-MOSFETs in figure 4.10, this shoulder increases in size when a more negative threshold voltage is associated with a particular MOSFET. We therefore assume that the negative threshold voltage shift is caused by interface states and we can thus calculate ΔN_{it} of any given device, where we define ΔN_{it} as the relative value of the charged interface state density per unit area of device 1 $N_{it}^{(1)}$ compared to that in device 2 $N_{it}^{(2)}$ where

$N_{it}^{(1)} > N_{it}^{(2)}$:

$$\Delta N_{it} = N_{it}^{(1)} - N_{it}^{(2)} \tag{4.1}$$

and hence according to the threshold voltage technique (section 3.1.7):

$$\Delta N_{it} = \frac{(V_t^{(1)} - V_t^{(2)})C_{ox}}{q} \tag{4.2}$$

Where V_t is the threshold voltage associated with each device. Note that this technique has been subtly modified as it is normally applied to a single MOSFET before and after a stress applied to the gate in order to induce interface states. Here, we exploit the large range of threshold voltages associated with individual MOSFETs due to interface traps.

Using 0.6 V as the range of threshold voltages measured gives an initial estimate of the maximum value of $\Delta N_{it} \approx 10^{13}$ cm^{-2}. This is interpreted as the difference in the charged interface state density at strong inversion between the device with the most positive threshold voltage and the one with the most negative. Note that this is considered as a large charged interface state density and explains the extreme variation in device performance observed.

The interface state density can also be extracted from the subthreshold slope method described in section 3.1.7, which will give $D_{it}(V_g = V_{ss})$ where V_{ss} is the gate voltage at which the minimum subthreshold slope occurs. From the specified substrate doping density of 7×10^{15} cm^{-3} we estimate the maximum depletion layer capacitance as 5×10^{-7} F cm^{-2} from equation 2.24. A value determined from the GB-branch in figure 4.11 gives a lower value of 1×10^{-7} F cm^{-2}.

As the subthreshold slopes are relatively large, even exercising caution and using the greater value of depletion capacitance in equation 3.29, when calculating D_{it}, makes negligible difference in the value obtained. The interface state capacitance ap-

pears to completely dominate the subthreshold slope and we obtain a range of D_{it} of 1×10^{13} cm^{-2} eV^{-1} to 6×10^{13} cm^{-2} eV^{-1} in non-annealed MOSFETs from samples C and E. Sample H (hydrogen annealed) contains devices with a subthreshold slope corresponding to interface state densities as low as $D_{it} = 5 \times 10^{12}$ cm^{-2} eV^{-1}.

To compare ΔN_{it} extracted using relative threshold voltages with D_{it} extracted using the subthreshold slopes, we convert ΔN_{it} values for a given MOSFET from a relative value to an absolute value N_{it} and then divide by the energy range which the interface states are spread over to obtain an average value of interface states per unit energy $\langle D_{it} \rangle$.

The conversion to an absolute value is accomplished from the charge pumping data published in Zimmerman et al. [2006], which shows that a non-annealed device (samples C and E) with a threshold voltage of ≈ 0.3 V has $N_{it} \approx 10^{12}$. As the distribution of threshold voltages at room temperature shown in figure 4.8, begins at $V_t \approx 0.3$ V the relative value of interface state density can be converted to an absolute value by:

$$N_{it} \approx \Delta N_{it} + 10^{12} \text{ cm}^{-2} \tag{4.3}$$

To convert the N_{it} values to $\langle D_{it} \rangle$ we assume that the donor/acceptor model applies and hence use equation 3.31. Given the interface state shoulder in the gate-channel split C-V branch, we can assume that the interface states are physically close to the channel and therefore most likely present at the SiO_2/Si interface. With the donor/acceptor model, we expect ΔE to be from mid-gap to strong inversion and assuming that threshold voltage corresponds to $\psi_s = 2\phi_b$ then $\Delta E = E_F - E_i = q\phi_b$, which is easily estimated from equation 2.14. With a substrate doping density of 7×10^{15} cm^{-3} and looking up the intrinsic carrier concentration for germanium at room temperature in figure 2.2 we get $\Delta E = E_F - E_i \approx$0.15 eV.

This enables the values of N_{it} to be divided by $\Delta E = q\phi_b$ to give the average interface state density $\langle D_{it} \rangle$ between strong inversion and mid-gap. Now that a value of D_{it} at a specific point in the band-gap is known from the subthreshold slope of each device in addition to an average value $\langle D_{it} \rangle$ between mid-gap and strong inversion, a comparison is made between the two in figure 4.17.

There is good agreement between the two values when applied to non-annealed MOSFETs (located on samples C and E) with the distribution overlapping and almost parallel to the unity line. Some hydrogen annealed MOSFETs (sample H) appear to lie in this distribution and some form a separate trend with a greater value of $\langle D_{it} \rangle$ compared to D_{it} determined from subthreshold slope. This could be because the density of interface states increases more sharply towards the valence band-edge in sample H, compared to samples E and C (figure 4.16). Although the subthreshold slope is reduced by reducing the density of interface states around mid-gap, a large threshold voltage shift is still possible due to a high density of states near the strong inversion condition.

Interface states also explain the larger negative threshold voltage shift observed at 77 K in devices with a more negative threshold voltage (and thus higher $\langle D_{it} \rangle$) at 300 K as shown in figure 4.13. As a device is cooled, the Fermi-level tends towards the conduction band edge and hence a greater band bending is required to reach the strong inversion condition (i.e. ϕ_b is larger). But consider that for a device with interface states in the lower half of the band-gap, as the band bending required to reach strong inversion increases, so does the number of interface states crossing the Fermi-level and gaining a positive charge. Hence, the negative shift in threshold voltage observed with decreasing temperature has two components, one due to the increased band bending required to reach strong inversion and the other due to the charging of more interface states. The second component is larger for a device with a higher density of interface

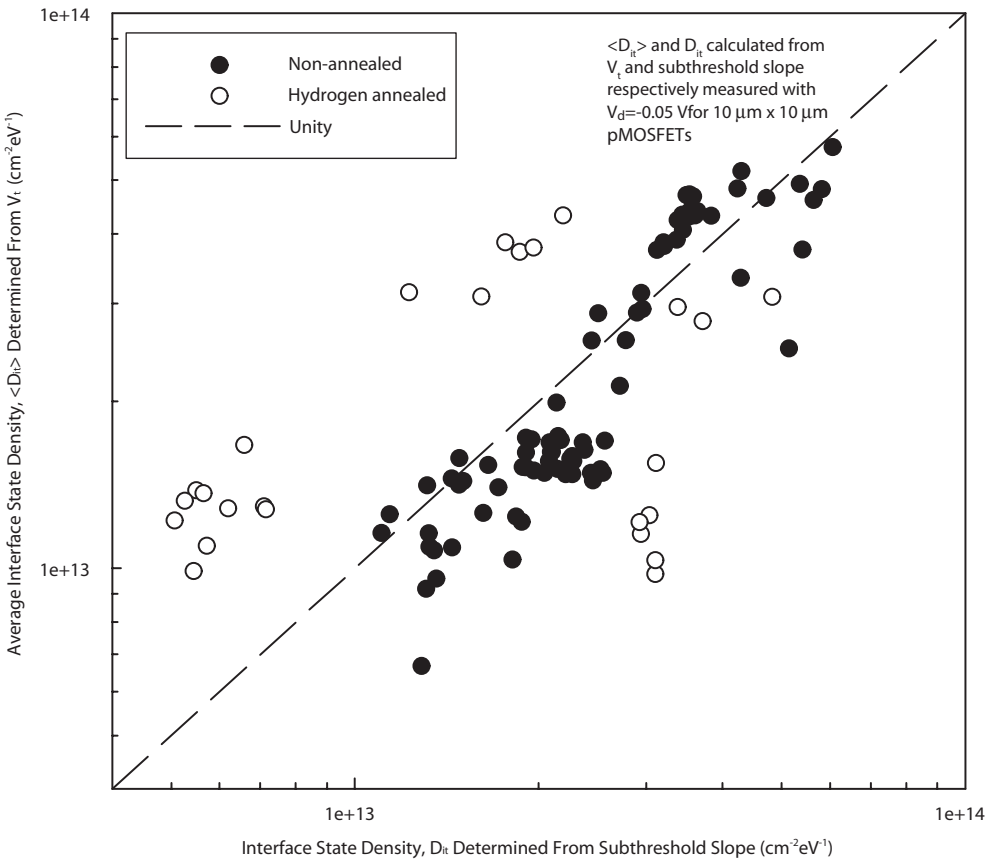


Figure 4.17: A comparison of $\langle D_{it} \rangle$ with D_{it} calculated from V_t and SS respectively measured with $V_d = -0.05$ V for many 10 μm$\times$10 μm non-annealed and hydrogen annealed p-MOSFETs.

states (and therefore a more negative threshold voltage at room temperature) and a larger threshold voltage shift with temperature decrease is observed.

The threshold voltage and subthreshold slope of a given MOSFET have been related by the presence of interface states. In turn these parameters offer a means of estimating the interface state density of a particular MOSFET. Every point on the graph in figure 4.17 references a particular device and each device is therefore easily assigned a value of D_{it} and $\langle D_{it} \rangle$.

As only limited charge pumping data is published Zimmerman et al. [2006], the author arranged for further measurements by Lidia Lukasiak at the Institute of Microelectronics and Optoelectronics in Warsaw, Poland. According to these measurements and those published, the range of D_{it} is 1×10^{12} cm^{-2} eV^{-1} to 5×10^{13} cm^{-2} eV^{-1}. This agrees well with that determined from the techniques applied to MOSFETs here.

Note that there is significantly less variation between interface state density extracted from capacitors (not shown) than that from MOSFETs. This could be because there are far fewer capacitors to characterise and that they have a significantly larger surface area: $\approx$10000 μm^2 compared to 100 μm^2 for the MOSFETs — effectively averaging out variation.

4.8 Mobility Analysis

It has been quantitatively demonstrated that interface states are responsible for the threshold voltage shift and range of subthreshold slopes observed in this device batch. However, we also observe a large decrease in peak transconductance with more negative threshold voltages (figure 4.8). It is a reasonable assumption that the decrease in peak transconductance is caused by the presence of interface states as it correlates strongly with threshold voltage, subthreshold slope and increasing interface state shoulder in the

gate-channel split C-V branch. Transconductance is directly related to the field effect mobility as described in section 3.1.6. Remember that only peak field effect mobility can be trusted as a reasonable measure of the effective mobility in the channel. We therefore begin with a simple analysis of peak field effect mobility before moving on to a more detailed analysis of the effective mobility as a function of vertical effective field.

4.8.1 Peak Field Effect Mobility

We have assumed that the interface states in the bottom half of the band-gap are donor-like and gain a positive charge as the device tends towards strong inversion. Given their close proximity to the channel (as indicated by the interface state shoulder in the gate-channel split C-V branch), these will then act as scattering sites for the holes that make up the inversion charge and increase the Coulomb scattering rate [Zhu et al., 2004].

The scattering rate will be proportional to the areal density of charged impurities n_{im}. Using Mathiessen's rule we see that the reciprocal of the total measured mobility should be proportional to the Coulomb scattering limited mobility:

$$\frac{1}{\mu} \propto \frac{1}{\mu_c} \tag{4.4}$$

In turn Coulomb scattering is inversely proportional to the number of charge scattering sites n_{im}:

$$\mu_c \propto \frac{1}{n_{im}} \tag{4.5}$$

It is now assumed that the main source of scattering sites is the charged interface states in inversion. Alternative sources could be other types of trapped oxide charge near the interface or impurities such as dopants in the channel. There is no evidence that these vary greatly or at all between devices and as we are performing a comparative study here, these constant factors are ignored. We therefore approximate $n_{im} \approx N_{it}$. It

has already been shown that $N_{it} \propto V_t$ so now:

$$\frac{1}{\mu} \propto \frac{1}{\mu_c} \propto N_{it} \propto V_t \tag{4.6}$$

Therefore, we expect that the reciprocal mobility of a carrier in the MOSFET channel will be proportional to the threshold voltage of that MOSFET. Peak transconductance has already been extracted for a large number of devices at 300 K and 77 K (figure 4.8) giving an easy method of calculating the peak field effect mobility μ_{FE} which can be compared to the corresponding threshold voltage of each device. A plot of μ_{FE}^{-1} against V_t demonstrates an extremely good linear relationship over the entire range of non-annealed devices (those from samples C and E) at 300 K and 77 K (figure 4.18). Note that the points at room temperature segregate into two groups, one above the line of best fit and one below. The MOSFETs with the higher peak mobility have optimised source/drain contacts and as expected have a lower R_{sd} than the other, unoptimised ones, lying on the lower peak mobility trend. All measurements are corrected for R_{sd} but only a single value — that extracted for the optimised devices — is used. It is expected that if the source/drain resistance was accurately measured for the unoptimised devices and used to correct the data, all devices would fit on the same trend.

To the author's knowledge, this linear trend of reciprocal peak field effect mobility against threshold voltage has not been observed before and therefore warrants some discussion. Zhu et al. [2004] demonstrates a decrease in peak effective mobility at low fields due to Coulomb scattering by interface traps, but with five data points compared to the one hundred or so here. A threshold voltage shift due to interface traps is well documented [Schroder, 2006] and therefore the relationship between mobility and interface traps demonstrated here is expected. A lack of previous publications could be for several reasons, the most obvious of which, is that it is rare to have such a variation in interface state density across a processed wafer. Normally when fabricating MOSFETs

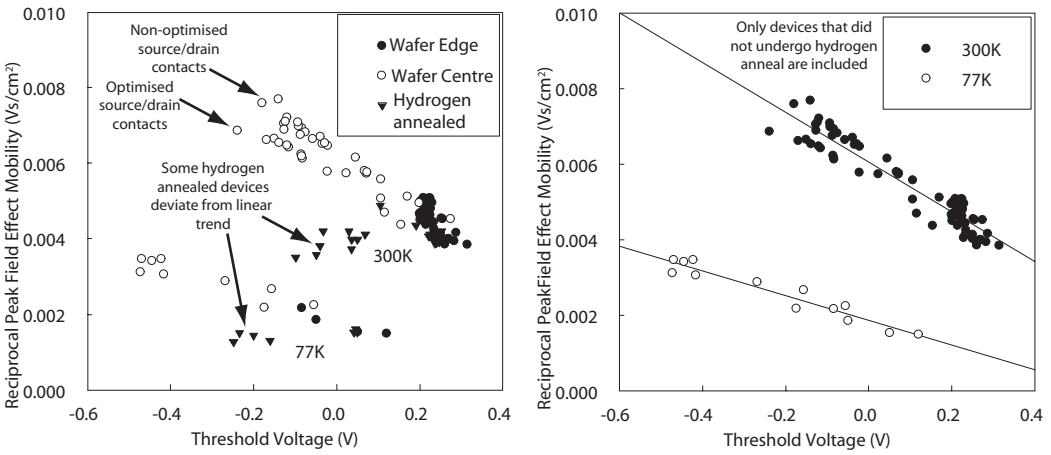


Figure 4.18: Reciprocal field effect mobility against threshold voltage extracted for many 10 μm×10 μm p-MOSFETs with $V_d = -0.05$ V at 77 K and 300 K. (a) includes hydrogen-annealed devices and (b) omits them to better demonstrate a linear relationship between the two parameters for those devices that did not undergo anneal.

for research or commercial purposes one hopes to obtain a uniform wafer with a low interface state density. Here we have quite the opposite and hence the application of a hydrogen anneal to achieve a more uniform, better performing set of devices (note that all subsequent germanium device batches fabricated at IMEC undergo hydrogen anneal as part of the standard fabrication process).

Once annealed in a hydrogen atmosphere, the devices appear to either fit the same linear relationship but lye at the high peak mobility end of the distribution, or to form another, apparently linear, relationship but increasing in mobility as threshold voltage becomes more negative (figure 4.18(b)). It is these devices that generally exhibit the highest peak field effect mobilities. It is difficult to assess the physical reason for this transition to a new trend. The initial state of the devices is unknown, only that they probably more closely resembled those found at the wafer centre (sample C) as opposed to the edge (sample E) as sample H did not come from the very edge of the wafer. This

means that before hydrogen anneal there was most likely a large range of interface state densities over sample H. Whereas a MOSFET with a high interface state density would be expected to move towards the more positive threshold voltage end of the trend with a hydrogen anneal; the post-anneal interface state density of a MOSFET with a low initial interface state density could be small enough that D_{it} no longer dominated its characteristics.

4.8.2 Effective Mobility

A more detailed investigation is now performed, examining the effective mobility as a function of vertical effective field or inversion charge density. The split C-V measurements at room temperature (figure 4.11) show a large dependence on interface state density and therefore can only be used to reliably calculate the inversion charge and depletion charge as a function of gate voltage if the interface state density is low. This is the case in some of the hydrogen annealed MOSFETs (sample H) and hence the mobility as a function of vertical field is calculated for a MOSFET on sample H at 300 K and compared to the strained-Ge devices in the next chapter (figure 5.13(a)). Here we are more interested in the effective mobility at 4.2 K.

Split C-V curves from measurements at low temperatures (77 K and 4.2 K) are not distorted by the presence of interface states. The threshold voltage shift associated with interface states at 300 K is still observed at low temperatures and therefore we assume that the interface states respond to DC voltages but not to the AC signal used in C-V measurements. Therefore, inversion charge and depletion charge can be accurately calculated from split C-V measurements at low temperatures for any device irrespective of its interface state density.

Mobility as a function of inversion charge density at 4.2 K was calculated for

nine MOSFETs from the wafer centre (sample C — labelled C1–C9) and two MOSFETs that underwent hydrogen anneal (sample H — labelled H1 and H2). Curves are fitted to this data based on carrier transport in the inversion layer in the electrical quantum limit described in section 2.4.2. First, we ensure that all carriers are in the lowest sub-band (i.e. the electrical quantum limit applies) by the condition specified in equations 2.42 and 2.43. Values of E_F and $E_1 - E_0$ calculated from equations 2.41 and 2.44 are compared in figure 4.19(a) over the entire range of vertical effective field applied to the channel. Note that the same figure demonstrates inversion layer degeneracy over the entire vertical field range.

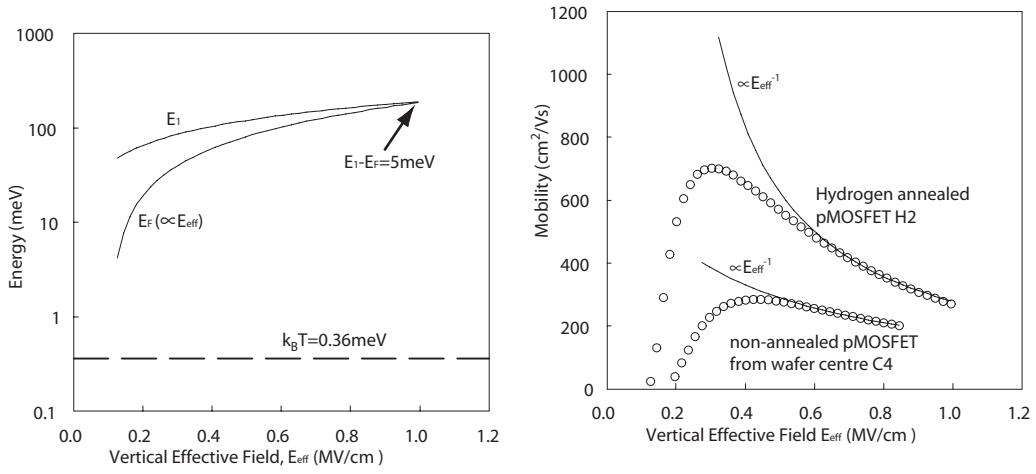


Figure 4.19: The Fermi-energy E_F and 1st sub-band energy E_1 plotted against vertical effective field E_{eff} (a) demonstrating that at 4.2 K the electrical quantum limit applies as $E_1 > E_F$ and that the inversion layer is degenerate as $E_F >> kT$. The mobility of p-MOSFETs measured at 4.2 K is proportional to E_{eff}^{-1} at high vertical effective fields (b) implying only local surface roughness scattering.

Only two scattering mechanisms are considered here: interface impurity scattering and surface roughness scattering. It is assumed that phonon scattering is negligible

at this low temperature [Takagi et al., 1994a]. Hafnium-oxide does have a very low optical phonon energy of 12.40 meV [Fischetti et al., 2001] but this is still far greater than the thermal energy of 0.36 meV at 4.2 K. Scattering from fixed charges away from the interface (remote impurity scattering), for example located in the hafnium gate dielectric, should also not have a significant effect [Emeleus et al., 1992]. This has been demonstrated for carrier densities greater than $\approx 1 \times 10^{12}$ cm^{-2} in s-Ge channels [Rossner et al., 2004].

Surface roughness scattering from interfaces other than the SiO_2/Ge interface, where the carriers are confined (remote interface roughness scattering), for example the hafnium-oxide/silicon dioxide interface, can also be ruled out. At high fields $\mu \propto E_{eff}^{-1}$ (figure 4.19(b)), which is exactly the relationship expected for holes at the SiO_2 interface in silicon based MOSFETs [Takagi et al., 1994a]. This dependence has been demonstrated to be due to direct surface roughness scattering using self-consistent simulations by Jallepalle et al. [1996].

The heavy-hole band constant effective mass of germanium ($m_h = 0.3m_0$) is used as the majority of holes will occupy this band. Non-parabolicity of the band is ignored. We also assume all scattering occurs at the Fermi-energy E_F corresponding to the Fermi-wave-vector k_F, i.e. the $T = 0$ relaxation time approximation is used.

The corrected effective mobility term (equation 3.15) is applied to all data with E_F in place of k_BT as the inversion layer is degenerate throughout the measurement. The R_{sd} correction is also applied and it is ensured that all MOSFETs have optimised source/drain contacts as this is the source/drain resistance that has been measured at 4.2 K (see section 4.3).

The effective mobility fit as a function of vertical effective field E_{eff} is obtained by adjusting the two parameters of interface impurity sheet density n_{im} and root mean

square (RMS) surface roughness Δ. The Coulomb mobility and surface roughness mobility are summed using Mathiessen's rule to give the total mobility as a function of vertical effective field. Parameters are adjusted until the total mobility gives the best agreement with the data for each device. Note that surface roughness generally limits the mobility at high vertical fields and interface impurity scattering limits mobility at low fields, hence the two contributions can be adjusted relatively independently from each other. A family of mobility curves for both mechanisms for various values of impurity density and RMS surface roughness is given in figure 4.20.

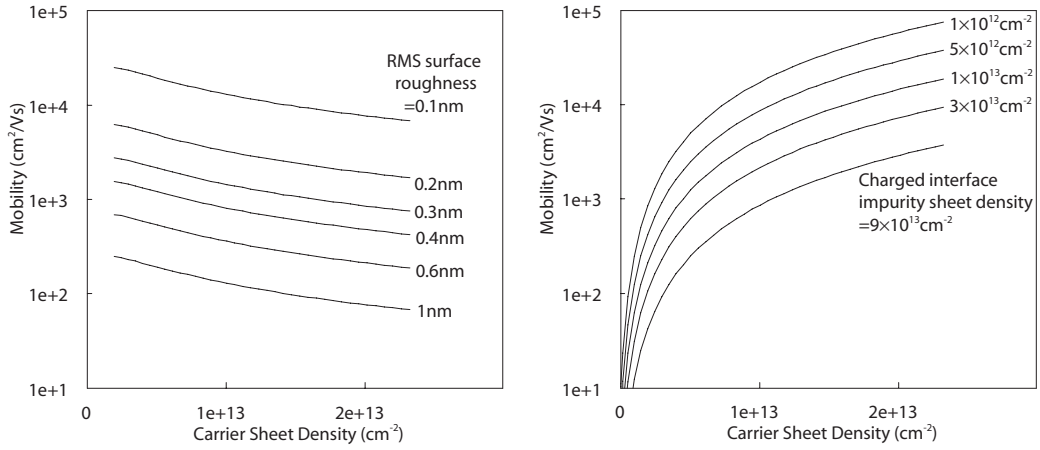


Figure 4.20: The surface roughness (a) and Coulomb scattering (b) limited mobility calculated for a Ge p-MOSFET with various RMS values and charged impurity sheet densities respectively. This is in the electrical quantum limit for a single sub-band of constant effective hole mass $0.3m_0$ and in the $T = 0$ approximation.

Mobility fits were successfully obtained for all eleven p-MOSFETs with two such mobility fits shown in figure 4.21 and figure 4.22. The interface impurity sheet density n_{im} and RMS surface roughness Δ used to fit mobility curves for each MOSFET, along with the interface state density extracted from the threshold voltage, are given in table 4.2. A plot comparing the interface state density to the interface impurity density

is given in figure 4.23(a). The RMS surface roughness is compared for each device in figure 4.23(b).

As expected, the hydrogen annealed MOSFETs (sample H) have a lower areal charged impurity density at the interface and note that the number of impurities deduced agrees reasonably with the $\langle D_{it} \rangle$ values measured for each MOSFET. However, the interface states are not spread over 1 eV but about 0.15 eV (at room temperature) between mid-gap and strong inversion, as calculated previously. This means that the number of interface impurities n_{im} used to obtain the mobility fit is actually significantly greater than the charged interface state density N_{it}.

Consider that interface state densities were extracted at room temperature where states were assumed to be present from mid-gap to the Fermi-level at strong inversion. At lower temperatures, the surface potential at strong inversion will tend towards the conduction band edge and hence the energy range over which interface states contribute to N_{it}, is larger. This could be closer to 0.33 eV as opposed to 0.15 eV estimated at room temperature, due to the decrease in intrinsic carrier concentration. We must also take into account the sharp increase in interface state density observed towards the conduction band edge in all samples (figure 4.16). Any change in the surface potential towards the conduction band edge at the strong inversion condition will give rise to a disproportionately large increase in the number of charged interface states. Therefore we might expect the density of charged interface states measured at strong inversion at room temperature to underestimate the 4.2 K density. As the number of charged impurities used to obtain the mobility fit correlates with the interface state density we conclude that the interface states are the main source of charged impurities at the interface.

An unexpected result is that surface roughness is reduced by the hydrogen anneal.

RMS roughness is reduced from an average of 0.60 nm in a MOSFET that has not been annealed in hydrogen (sample C) to an average of 0.48 nm on an hydrogen annealed MOSFET on sample H — a reduction of approximately 20%. Hydrogen anneals have been shown to reduce the RMS roughness of a free germanium surface [Nayfeh et al., 2004] but not the same at a dielectric interface before. As the dielectric interface in these MOSFETs is potentially quite complex as it has two interlayers of Si and SiO_2 and it is difficult to assess how the roughness is decreased. It could be a diffusion based process, for example, where Ge movement is aided by the presence of hydrogen [Nayfeh et al., 2004].

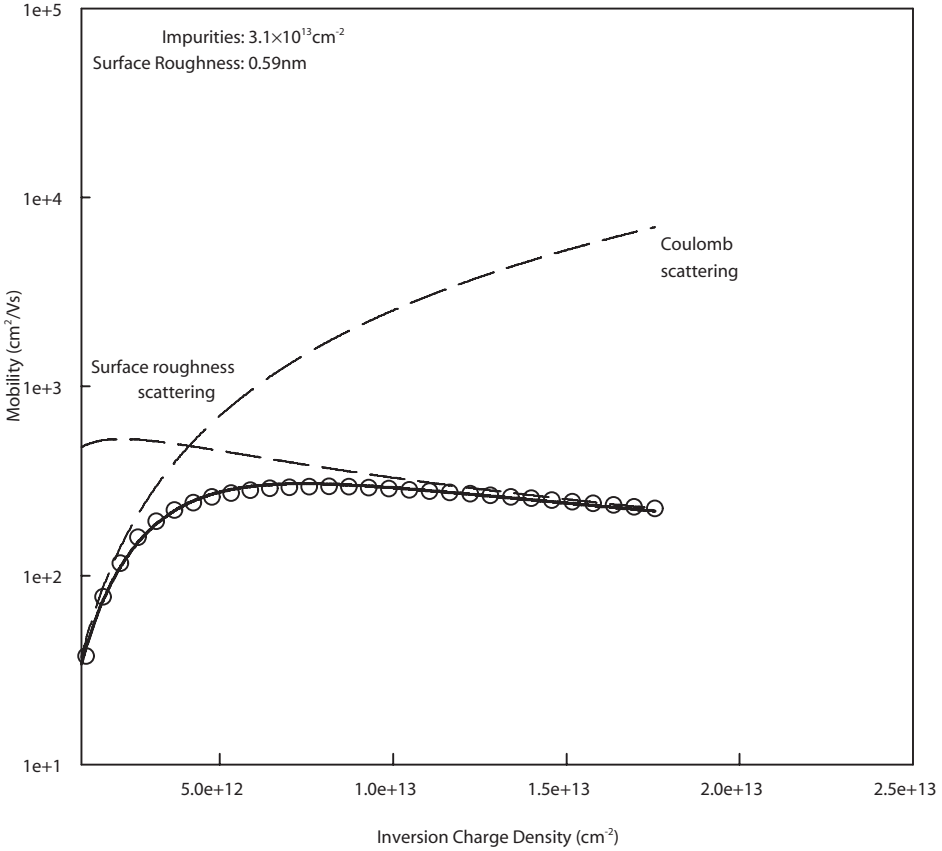


Figure 4.21: The measured mobility (points) of a 10 μm$\times$10 μm non-annealed p-MOSFET C4 (located at the wafer centre) with $V_d = -0.05$ V at 4.2 K and theoretical mobility (solid line) calculated using Mathiesen's rule to sum the modelled Coulomb and surface roughness scattering mobility components. Values of fitting parameters of RMS surface roughness Δ and charged impurity sheet density n_{im} are shown.

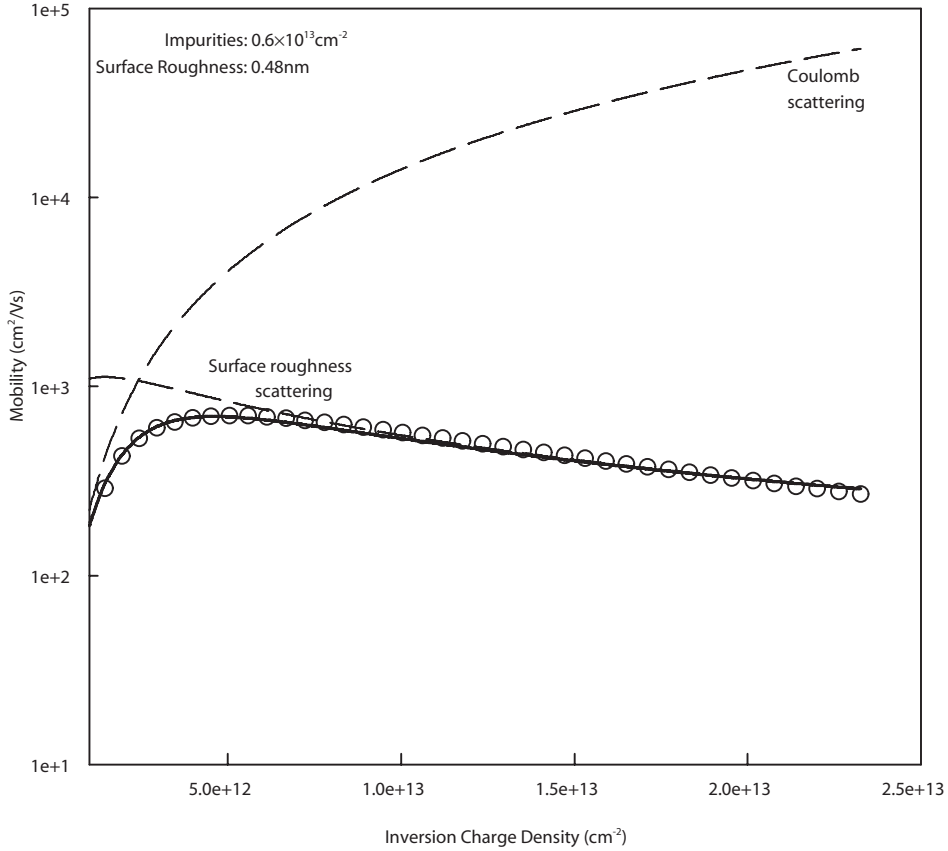
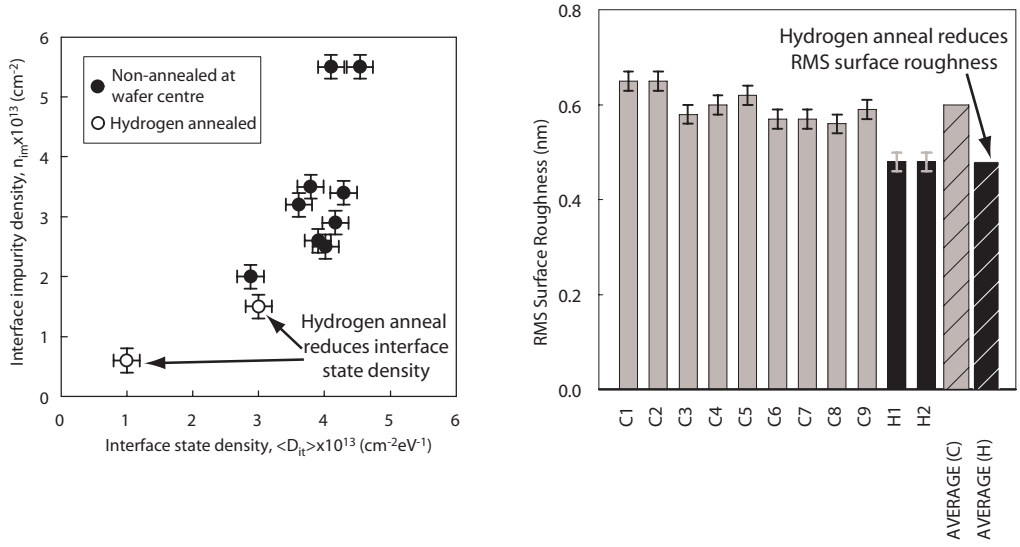


Figure 4.22: The measured mobility (points) of a 10 μm$\times$10 μm hydrogen annealed p-MOSFET H2 with $V_d = -0.05$ V at 4.2 K and theoretical mobility (solid line) calculated using Mathiesen's rule to sum the modelled Coulomb and surface roughness scattering mobility components. Values of fitting parameters of RMS surface roughness Δ and charged impurity sheet density n_{im} are shown.

Device	$\langle D_{it} \rangle$ cm^{-2} eV^{-1}	n_{im} cm^{-2}	Δnm
C1	4.1×10^{13}	5.5×10^{13}	0.65
C2	4.5×10^{13}	5.5×10^{13}	0.65
C3	3.8×10^{13}	3.5×10^{13}	0.58
C4	3.6×10^{13}	3.1×10^{13}	0.59
C5	3.9×10^{13}	2.6×10^{13}	0.62
C6	4.2×10^{13}	2.9×10^{13}	0.57
C7	4.0×10^{13}	2.5×10^{13}	0.57
C8	2.9×10^{13}	2.0×10^{13}	0.56
C9	4.3×10^{13}	3.4×10^{13}	0.59
H1	3.0×10^{13}	1.5×10^{13}	0.48
H2	1.0×10^{13}	0.6×10^{13}	0.48

Table 4.2: Measured interface state density $\langle D_{it} \rangle$ for 10 μm$\times$10 μm non-annealed (C) and hydrogen annealed (H) p-MOSFETs alongside charged impurity sheet density n_{im} and RMS surface roughness Δ values used to fit theoretical mobility curves to those measured at 4.2 K.



(a) (b)

Figure 4.23: Charged impurity sheet density n_{im} used to fit theoretical mobility curves against measured interface state density for p-MOSFETs (a). The RMS surface roughness Δ used to fit theoretical mobility curves is plotted for the same devices (b).

4.9 Summary

In this chapter we have investigated bulk-Ge p-MOSFETs with a high-k gate stack. These were shown to have a large range of interface state densities through threshold voltage shift and subthreshold slope. Although somewhat limited data is available, charge pumping measurements agree with the range of states measured. The high-low C-V and the conductance technique were used on n-MOS capacitors to measure the interface state density as a function of bandgap energy. All these techniques demonstrate a decrease in interface state density after a post-metallisation anneal in hydrogen atmosphere.

There are several clues as to the nature of the interface states. We see that they have a fast response time, appearing in the gate-channel branch split C-V curve as a shoulder at a measurement frequency of 100 kHz. This implies that they are physically close to the channel, most likely at the Si/SiO_2 interface. What's more, a hydrogen anneal is known to reduce the interface state density in silicon dioxide [Nicollian and Brews, 1982], further supporting the theory that they reside at this interface.

Not yet considered, is the physical process by which the interface states arose in these devices during fabrication. We now discuss this briefly with the assumption that the interface states are at the Si/SiO_2 interface. The silicon dioxide was not formed by standard thermal oxidation but by a dip in ozonated water, which is a proprietary process developed at IMEC. It is not clear why this would lead to a high density of interface states but the oxide certainly could have different properties to that of a standard thermal oxide. In addition, the germanium substrate consists of a relaxed, several micron thick Ge layer on a standard Si-substrate. This is not optimised and has a high threading density, which could potentially cause interface states to appear during oxidation of the Si-cap (figure 4.24(a)i,ii); however, this does not explain why more interface states are

present in devices at the wafer centre than at the edge.

Another aspect of these devices, not investigated in much detail here, was the off-state current leakage. Whereas some devices had the expected thermionic emission over a barrier, others did not have a barrier at all at the source/drain in the off-state. This could be due to interface states in the source/drain regions acting as a leakage path for carriers. As the gate stack is deposited over the entire wafer and then selectively etched, it could be this removal process that creates the states (figure 4.24(b)i,ii). Indeed, the chemistry of germanium device processing is in an experimental stage and far from perfected. Devices at the edge of the wafer had far lower interface state densities than those at the centre of the wafer, which could be to do with how it was handled when the gate etch was carried out. However, no correlation was found between the interface state density of a device and the off-state leakage current, although further investigation is needed.

The presence of interface sates had a large effect on the channel mobility of a given p-MOSFET. Peak field effect mobility (determined from peak transconductance) of a given device was shown, conclusively, to be inversely proportional to interface state density (determined from threshold voltage shift). It is proposed that the interface states are charged when the device is in the on-state and hence Coulomb scatter carriers in the channel. This scattering will be proportional to the areal density of these charged states. It was assumed that the SiO_2 donor/acceptor model of interface states applies (justifiable, given that the states appear to be at the SiO_2/Si interface) and therefore, the states will have a positive charge in the on-state.

However, although simple to measure, peak field effect mobility is far from accurate due to the inversion charge being approximated to $C_{ox}V_{gt}$ and hence to confirm the interface state/mobility relationship, a number of devices were cooled to 4.2 K and

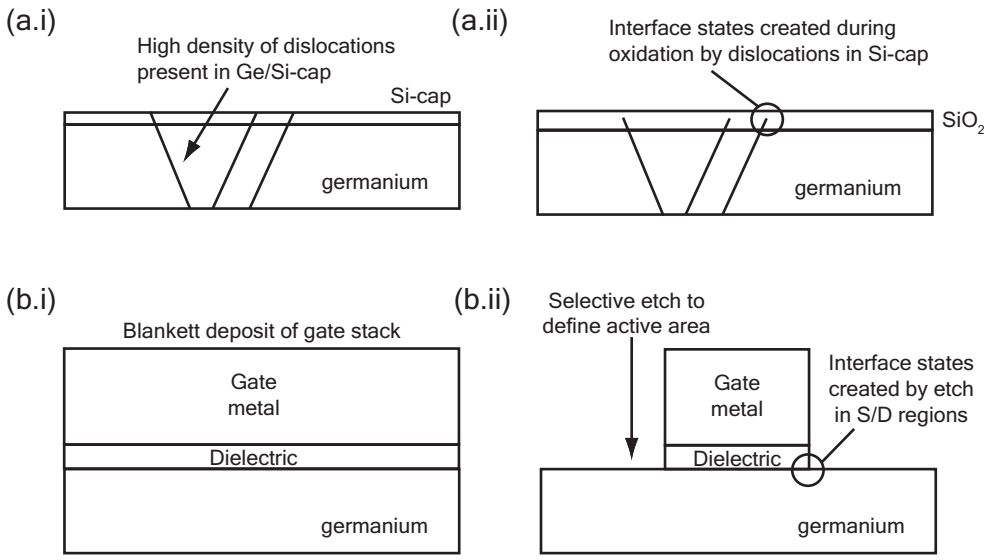


Figure 4.24: Dislocations present in the Si-cap before oxidation (a.i) result in interface states after oxidation (a.ii). Blanket deposition of gate stack (b.i) followed by a selective etch (b.ii) that introduces interface states in the source/drain regions where the etchant is active.

effective mobility accurately measured by split C-V. Applying an analytical model [Gold and Dolgopolov, 1986] to these mobility measurements quantitatively demonstrated that a higher interface state density resulted in a decreased channel mobility due to Coulomb scattering and that after a hydrogen anneal, the interface state density was reduced and the mobility increased.

A further, unexpected, effect of the hydrogen anneal was to reduce the RMS surface roughness, found through fitting the theoretical mobility to that measured at 4.2 K, by 20%. Given the simplicity of the analytical model applied, this result needs to be confirmed, perhaps with either a detailed computer simulation of these devices or experimentally by directly measuring the surface roughness. Despite this uncertainty, it is difficult to see what other scattering mechanism could be reduced by a hydrogen anneal and give an increase in mobility at high vertical fields/carrier densities.

Chapter 5

Strained Germanium MOSFETs with High-k Dielectric

5.1 Overview

The process used to fabricate the devices outlined in the previous chapter was repeated on a wafer with a strained germanium layer. This layer was obtained by first growing a relaxed $Si_{0.2}Ge_{0.8}$ buffer layer on six standard silicon (100) wafers by chemical vapour deposition. This was performed by AdvanceSiS — a spin-off company from the University of Warwick Department of Physics, specialising in epitaxial CVD Si/Ge material growth. Strained germanium layers were grown on the buffer layers with target thicknesses of 15 nm and 25 nm. A silicon cap was added for surface passivation and this also had two different thickness targets: 1 nm and 2 nm. Owing to time constraints, a detailed characterisation of the wafers before device fabrication was not performed, nor was the CVD recipe developed beyond an initial design based on similar structures grown previously. p-MOSFETs were fabricated on four wafers and n-MOSFETs on two.

Wafer	Device type	s-Ge thickness	Si-cap thickness
W1	n-MOSFET	15 nm	1 nm
W2	n-MOSFET	15 nm	2 nm
W3	p-MOSFET	15 nm	1 nm
W4	p-MOSFET	25 nm	1 nm
W5	p-MOSFET	15 nm	2 nm
W6	p-MOSFET	25 nm	2 nm

Table 5.1: Wafer specification for the s-Ge MOSFET device batch.

The wafers are referred to as W1–W6 and specified in table 5.1.

The wafers were 150 mm in diameter, which is incompatible with the Pilot Line processing facility at IMEC and therefore a simplified process was used to fabricate devices. The mask set used was also different than the one described in the previous chapter and, due to the simplified process flow, only a fraction of the structures worked. These are three MOSFETs of different gate lengths (defined in table 5.2) and capacitors of various sizes. Because of the simplified process flow, the MOSFETs are a primitive structure with a central dot as source/drain contact, a ring surrounding this as the gate stack and a further ring around the perimeter as the second source/drain contact (figure 5.1). Whereas the mask-set used in the previous chapter is old and well-tested, the mask used here is previously untested. Also, the simplified process flow means that there could be leakage current issues due to lack of proper device isolation. Note that all wafers here underwent post metallisation anneal in a pure hydrogen atmosphere at 400 °C for 20 minutes.

We begin with a physical characterisation of the wafers in section 5.2 before a detailed electrical characterisation of the p-MOSFETs on W3-6 in sections 5.3–5.6. The n-MOSFETs are electrically characterised in section 5.7.

MOSFET	Gate Length	Effective Width
D2	50 μm	453.2 μm
D3	100 μm	571.9 μm
D4	150 μm	679.9 μm

Table 5.2: Strained germanium (s-Ge) Ring MOSFET device dimensions.

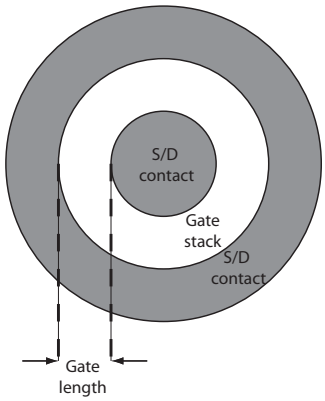


Figure 5.1: Plan-view schematic of a ring MOSFET.

5.2 Physical Characterisation

Cross-sectional transmission electron microscopy was performed on W1 and W2 after device fabrication to confirm the presence of the s-Ge layer and accurately measure the thickness of the different silicon caps. XTEM images are displayed in figure 5.2 and show that the silicon caps are thicker than the specified 1 nm and 2 nm for W1 and W2 respectively. Note that this is post-fabrication and some of the Si-cap will have been consumed during processing, for example in preparation for high-k deposition by the ozonated water dip. Therefore, the Si-caps as-grown will have been significantly greater than the specified thicknesses. From the XTEM images shown, the silicon cap thicknesses were measured as 1.3 nm$\pm$0.2 nm and 2.2 nm$\pm$0.2 nm for the 1 nm and 2 nm target thicknesses respectively. The s-Ge layer with a 15 nm target thickness was

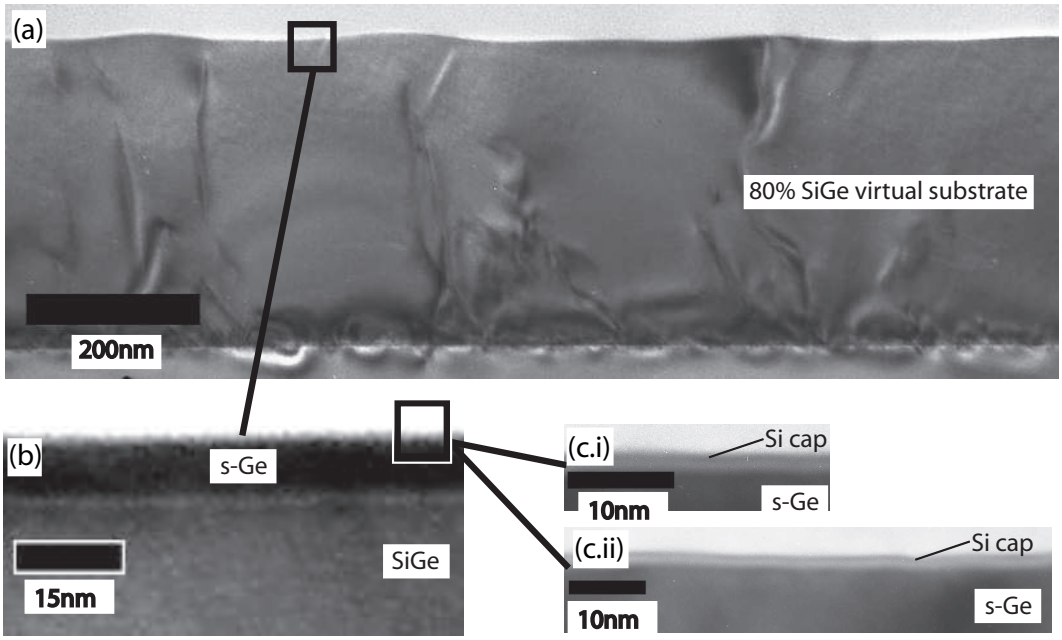


Figure 5.2: XTEM images of the relaxed 80% SiGe buffer (a), 15 nm target s-Ge layer (b) and Si-caps of two different thicknesses (c): 1 nm target (i) and 2 nm target (ii). Images are from the surface of wafers after device fabrication.

measured as 10 nm. No XTEM was performed on the thicker s-Ge target samples but it can be assumed that this layer is also thinner than specified.

At the request of the author, X-ray diffraction was performed on W4 by Jon Parsons, Nano-silicon Group, Department of Physics, University of Warwick with the aim of measuring the residual strain of the s-Ge layer [Koppensteiner et al., 1994]. A sample with a thicker s-Ge layer was chosen as a thicker layer will give a stronger signal; however, it could not be detected and only the SiGe buffer appeared on the X-ray scans (not shown). This was confirmed as fully relaxed.

5.3 p-MOSFET Off-state and Leakage Currents

I_d-V_g measurements were performed for p-MOSFETs on each wafer at 300 K, 77 K and 4.2 K (W3 and W4 only), and are shown in figure 5.3. Immediately apparent is the high drain current in the off-state at room temperature, which decreases significantly (six orders of magnitude) when the temperature is reduced and implies a thermally activated tunnelling mechanism. There are several possibilities for this high leakage current: dislocations, poor junction activation or poor device isolation. Each of these is considered in detail below.

5.3.1 Junction Activation

In order to minimise the risk of relaxing the top s-Ge layer, junction activation anneal temperatures were reduced from the usual 500 °C for germanium p-MOSFET processing to 450 °C. This temperature has proved effective in the past for boron activation in germanium; however, normally the Ge-layer is several microns thick. The thin s-Ge layers in this case mean that the junction depth (≈100 nm) greatly exceeds the thickness of the top s-Ge layer (10 nm–30 nm). If 450 °C is insufficient to activate the dopants in the underlying SiGe buffer layer then this is a viable path for leakage current. For this reason a series of anneals was performed to investigate whether the dopant activation could be improved. Improved activation should lead to a reduced off-state leakage current at room temperature.

The anneals performed were 450 °C, 500 °C, 550 °C, 600 °C, 700 °C, 750 °C for 5 minutes in addition to 700 °C for 10 minutes, all in nitrogen atmosphere. No reduction in the off-state current is forthcoming in any of the annealed devices (figure 5.4(a)), suggesting that poor dopant activation is not the cause of high leakage. Anneals of 450 °C – 550 °C for 5 minutes have no effect on device performance (figure 5.4(b))

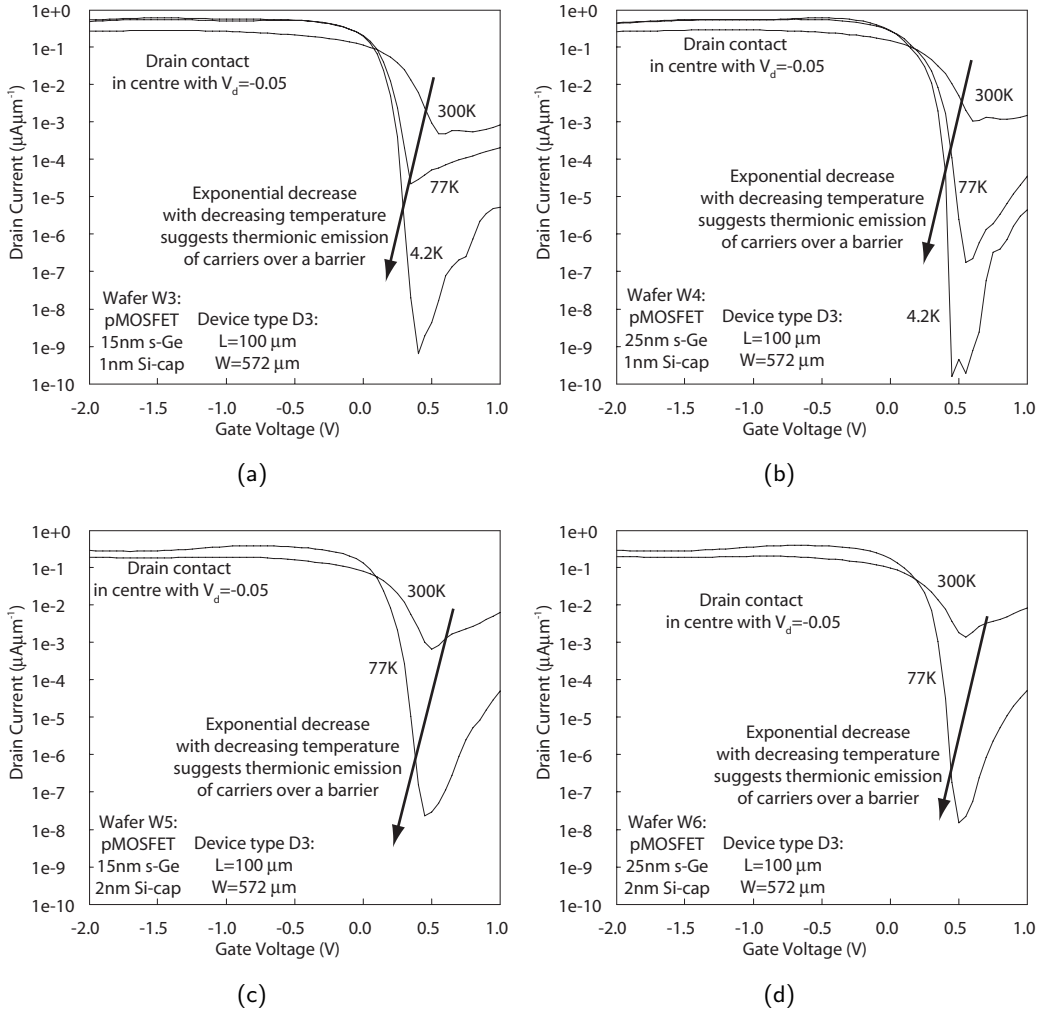


Figure 5.3: I_d-V_g characteristics at $V_d = -0.05$ V for best performing ring p-MOSFETs of size D3 (L=100 μm, W=572 μm) on wafers W3 (a), W4 (b), W5 (c) and W6 (d) at room temperature, 77 K and 4.2 K. The drain contact was at the central dot of the device and note that off-state leakage current, although high at room temperature, decreases exponentially with decreasing temperature suggesting that the dominant leakage mechanism is thermionic emission over a barrier.

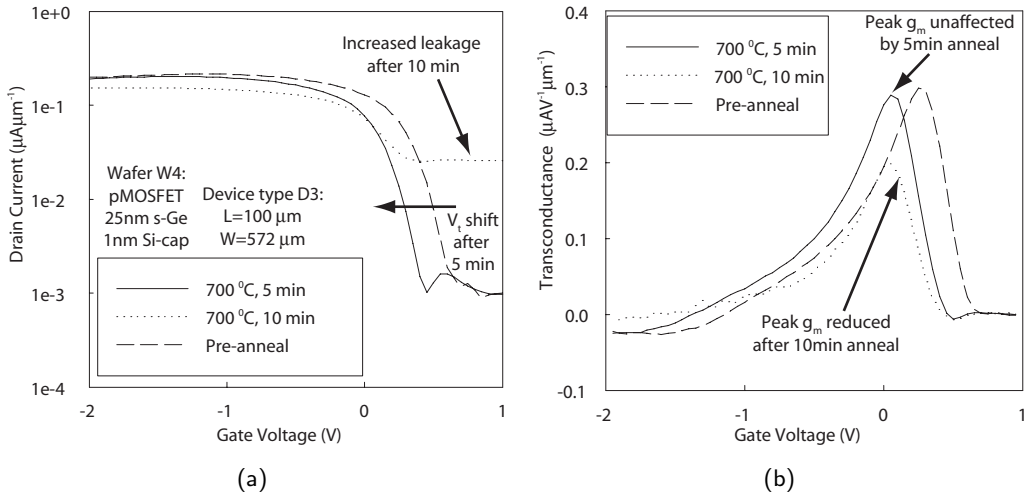


Figure 5.4: Drain current (a) and corresponding transconductance (b) measured at $V_d = -0.05$ V for p-MOSFETs of size D3 (L=100 μm, W=572 μm) on W4 pre-anneal and after 5 minutes and 10 minutes at 700 °C. A 5 minute anneal leads to a threshold voltage shift but no degradation in device performance, whereas 10 minutes increases the off-state current and reduces peak transconductance.

implying that they do not significantly affect the s-Ge layer nor gate stack. Anneals of 600 °C and 700 °C (only the latter shown) for 5 minutes give a threshold voltage shift of -0.1 V, which could be evidence of the gate stack being degraded as it is not normally subjected to temperatures this high. Anneals of 700 °C for 10 minutes and 750 °C (only the former shown) for 5 minutes have the same effect of severely degrading device performance, which is not surprising as the melting temperature of germanium (940 °C) is not much higher. The degradation is probably due to a combination of damage to the gate stack and s-Ge layer.

Low dopant activation can be ruled out as a cause of leakage, as no improvement in off-state current is observed for any of the anneals. It has been observed that anneals up to 550 °C (data not shown) do not affect the devices, which could be useful knowledge for future s-Ge device processing.

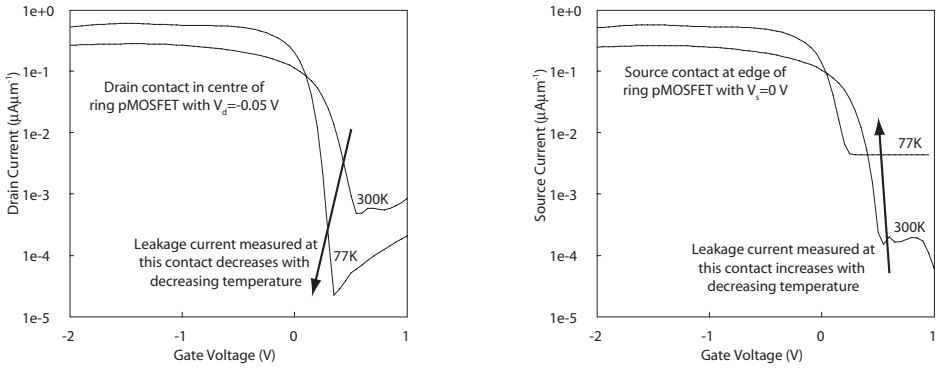


Figure 5.5: Comparison between drain current measured at the inner dot contact (a) and source current measured at the outer ring contact (b) of a size D3 (L=100 μm, W=572 μm) ring p-MOSFET with $V_d = -0.05$ V at 300 K and 77 K. In the former we observe a decrease in off-state current with decreasing temperature whereas in the latter an increase.

5.3.2 Device Isolation

Due to the simplified process there is no proper device isolation, and this potentially allows leakage current into a device from all over the wafer. The most vulnerable contact to this leakage is the outer ring which can be designated the source or drain contact during measurements. The drain contact is more susceptible to leakage in general, due to the voltage applied to it, hence the central dot contact was used as the drain. Figure 5.5 is a comparison of the source and drain currents measured at 300 K and 77 K with this contact configuration.

The drain current decreases significantly with a reduction in temperature but the source current increases. This suggests that as the temperature is reduced, the outer ring contact (source) becomes more electrically connected to the substrate than the MOSFET channel in the off-state. This could be due to dopants freezing out in the source/drain regions. The central dot contact (drain) therefore becomes more isolated from the source and hence the off-state current measured at the drain is significantly

reduced while that in the outer source contact is increased.

The current measured at the outer source contact in the off-state at room temperature is slightly lower than that at the central drain contact at room temperature. This is not consistent with the leakage currents measured at low temperature implying that the leakage mechanism described above does not dominate at 300 K. Therefore, although substrate leakage through the outer contact due to poor device isolation is clearly an issue at lower temperatures, the high leakage observed at room temperature in both contacts is caused by some other effect.

5.3.3 Threading Dislocations

Threading dislocations in the s-Ge layer are visible in the XTEM image of the virtual substrate presented in figure 5.2(a). This implies a high density of threading dislocations as the visible area of the XTEM image is small and having ruled out other leakage mechanisms, they are most likely responsible for high junction leakage at room temperature. The high density of threading dislocations provides a thermally activated tunnelling mechanism for current from either the source or drain contact into the substrate. If all leakage was due to this alone we would expect to see an exponential decrease in off-state current in both the inner dot and outer ring contacts; however, it has already been shown that this only occurs for the inner contact.

The unusual off-state currents measured as a function of temperature to the inner and outer contacts have been explained by a combination of isolation of the outer contact at lower temperatures due to dopant freeze-out and a high density of threading dislocations. Note that the on-state characteristics are not affected by contact configuration but we use the inner contact as the drain in subsequent measurements to avoid any contribution to the drain current by poor device isolation.

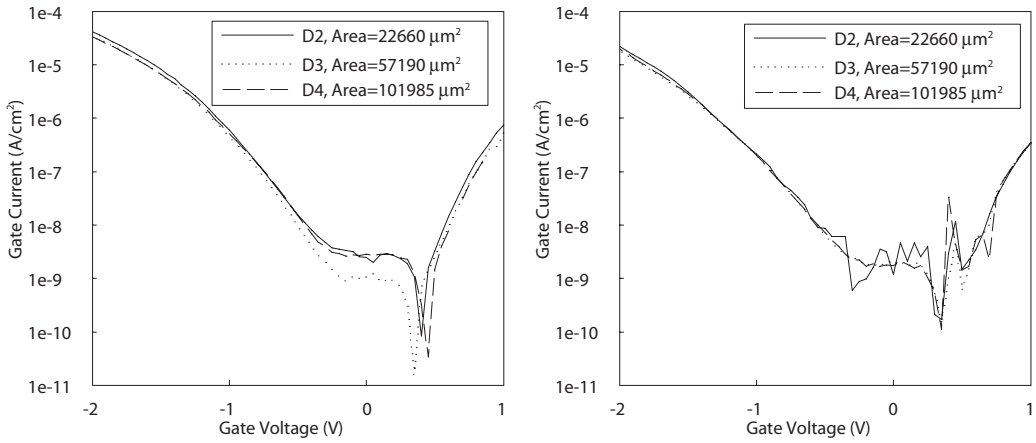


Figure 5.6: Gate leakage current normalised to gate stack area as a function of applied gate voltage for the ring p-MOSFETs of sizes D2, D3, D4 at 300 K (a) and 77 K (b). Note that the measured current density does not vary with gate area implying negligible leakage current around the perimeter of the gate stack. Gate leakage current changes little with temperature also.

5.3.4 Gate Leakage

Gate leakage is low (figure 5.6) and, even in the large area devices available for characterisation, not great enough to have significant impact on I_d. I_g scales with MOSFET area, meaning that no parasitic leakage current around the perimeter of the gate stack due to a processing error or poor mask design, for example. It is also relatively invariant with temperature.

5.4 p-MOSFET Characteristics

Drain current is plotted as a function of gate voltage at $V_d = -50$ mV for p-MOSFETs located on W3–W6 at 300 K, 77 K and 4.2 K in figure 5.7. At room temperature there is little variation between devices on different wafers. A feature of note is that the drain current decreases with more negative gate voltages at high gate overdrives. Reducing the

measurement temperature exaggerates this effect with very prominent double shoulders in the drain current curve appearing at data obtained at 77 K and 4.2 K. Note that throughout this chapter 4.2 K data is available for W3 and W4 only.

A thin silicon cap (W3 and W4) leads to a double peak in the drain current which is most prominent at 4.2 K. A thick cap (W5 and W6) gives a broader first peak with a far larger decrease in current with gate voltage. This behaviour is consistent with buried channel MOSFETs [Palmer, 2001] and is observed here due to the thicker than specified silicon caps allowing carrier confinement in the cap. As the Si-cap begins to populate with inversion charge at more negative gate voltages, we measure the drain current mediated by carriers confined at the Si/SiO_2 interface in addition to the drain current mediated by carriers confined at the s-Ge/Si interface. The transconductance plots in figure 5.8 derived from the drain currents in figure 5.7 highlight the two peaks in channel conductance. Note that the samples with a thinner cap (W3 and W4) offer a slightly higher peak transconductance at room temperature. At 77 K there is a far greater transconductance in W3 and W4 than in W5 and W6.

5.5 n-MOS Capacitor Characteristics

C-V was performed on capacitors on W3–W6 at 300 K, 77 K and 4.2 K at 10 kHz. Due to a fabrication error there is no metallisation on the back of the wafers, which is used as the substrate contact. This adds greatly to substrate resistance and hence causes severe attenuation at frequencies greater than 10 kHz in C-V measurements at room temperature (figure 5.9). One might expect dopant freeze-out in the substrate at lower temperatures to degrade the electrical contact between it and the substrate chuck further, thus introducing attenuation at lower frequencies; however, this was not the case. This could be due to the InGa eutectic applied to sample substrates as standard

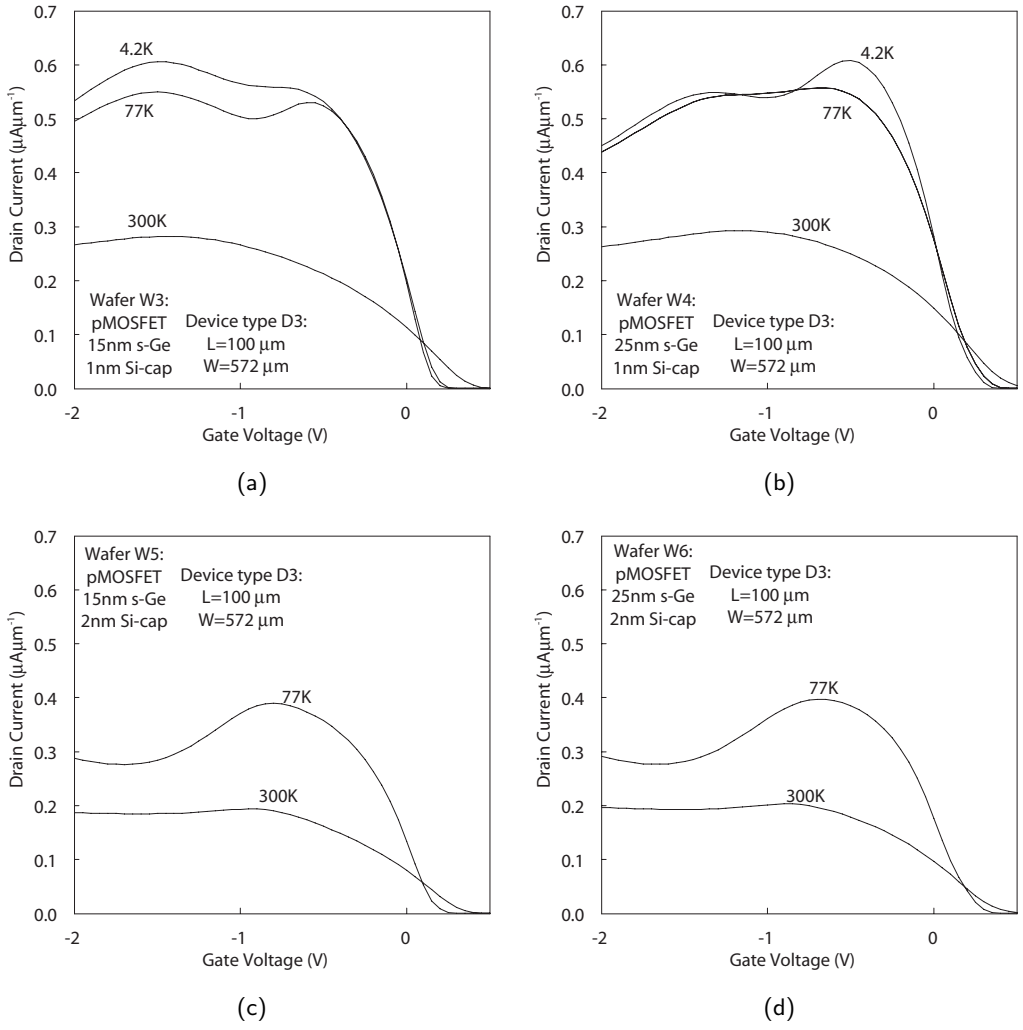


Figure 5.7: I_d-V_g characteristics with $V_d = -0.05$ V for the best performing ring p-MOSFETs of size D3 (L=100 μm, W=572 μm) on wafers W3 (a), W4 (b), W5 (c) and W6 (d) at 300 K, 77 K and 4.2 K. In each graph, it is clear that drain current decreases at high gate overdrive, which is inconsistent with normal MOSFET behaviour.

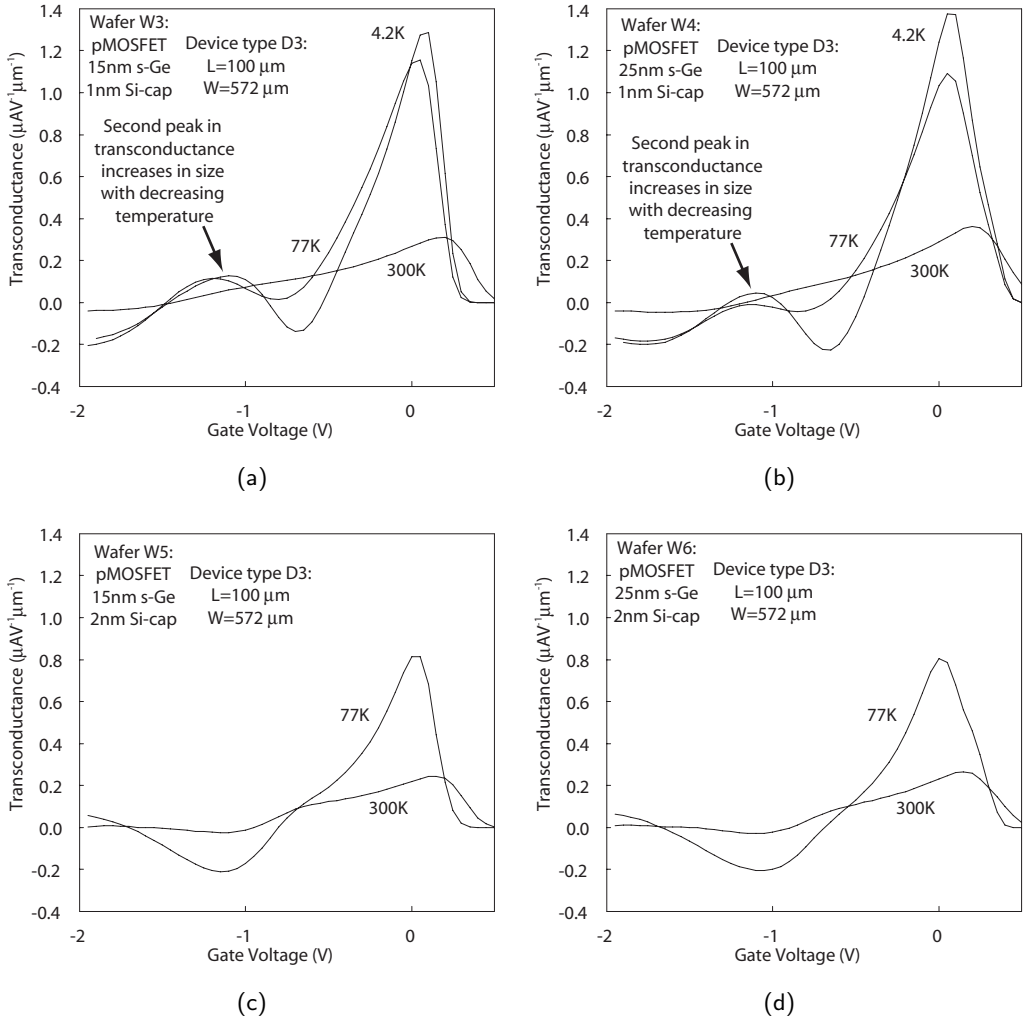


Figure 5.8: Transconductance measured with $V_d = -0.05$ V for the best performing ring p-MOSFETs of size D3 (L=100 μm, W=572 μm) on wafers W3 (a), W4 (b), W5 (c) and W6 (d) at 300 K, 77 K and 4.2 K. At lower temperatures for samples with a thinner Si-cap (W3, W4) there is a clear second peak in the transconductance. This is caused by holes populating the silicon cap at higher gate overdrives.

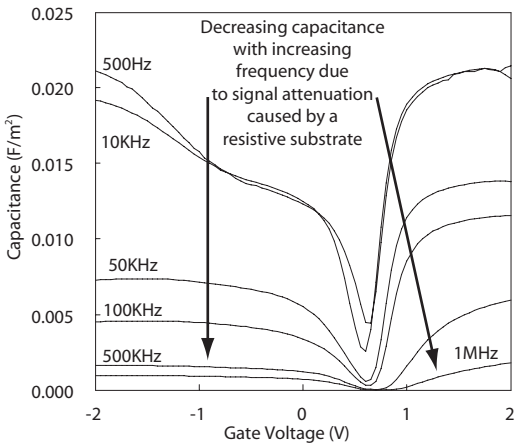


Figure 5.9: C-V curves measured at room temperature for a 250,000 μm^2 n-MOS capacitor on W4 (25 nm s-Ge, 1 nm Si-cap) at a range of frequencies. Note the severe attenuation in the measured capacitance as the frequency increases due to the resistive substrate. This was caused by a lack of metallisation on the back of the substrate due to a processing error.

procedure for low temperature measurements.

Present in the inversion region (negative gate voltage) of the C-V curves for all samples, is a shoulder (figure 5.10). This is thought to be due to the relatively thick silicon cap present in this device batch and we observe that the shoulder is lower for those samples with a thicker cap (W5 and W6 - the former shown in figure 5.10(b)). This is explained by holes preferentially occupying the Ge/Si-cap interface rather than the Si-cap/SiO_2 interface at lower gate overdrives. Until populated by carriers, the silicon cap acts as an extra capacitance in series with the oxide capacitance C_{ox} so that the shoulder capacitance C_{tot} is given as:

$$\frac{1}{C_{tot}} = \frac{1}{C_{ox}} + \frac{1}{C_{cap}} \tag{5.1}$$

Where C_{cap} is the silicon cap capacitance given by:

$$C_{cap} = \frac{\epsilon_0 \epsilon_{Si}}{t_{cap}} \tag{5.2}$$

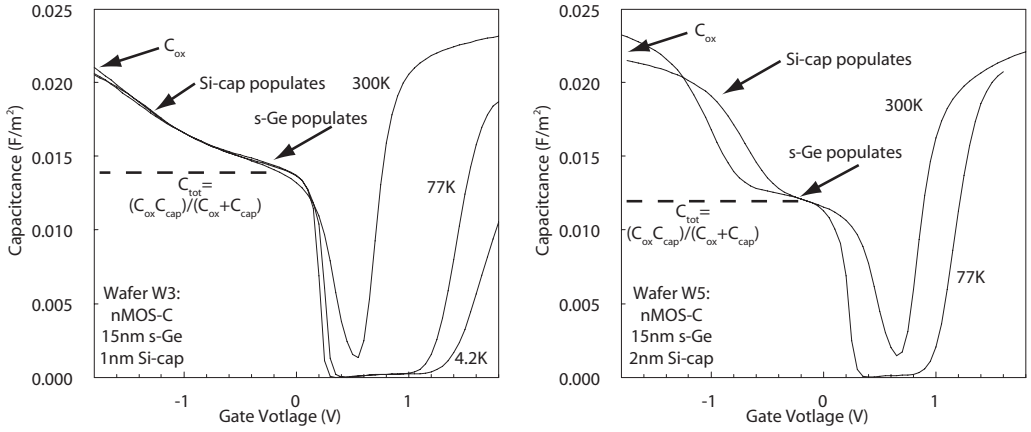


Figure 5.10: C-V curves measured at 500 Hz on 250,000 μm^2 n-MOS capacitors on wafer W3 (a) and wafer W5 (b). The former has a 1 nm Si-cap and the latter a 2 nm Si-cap, which has a significant effect on the shape of the inversion regions of the C-V curves. Before the cap populates with inversion charge it acts as an extra capacitance C_{cap} in series with the oxide capacitance C_{ox} and hence reduces the gate capacitance with holes in the s-Ge layer.

Where t_{cap} is the Si-cap thickness. Taking C_{ox}=0.023 F m^{-2} (determined from the accumulation maximum) and C_{tot}=0.015, 0.013 F m^{-2} as measured from the shoulders in the C-V profiles of samples with the thinner (W3, W4) and thicker caps (W5, W6) respectively; we calculate $t_{cap} = 2.5, 3.6$ nm respectively. These are significantly larger than the values measured from the XTEM images which are 1.3 nm and 2.2 nm respectively. As the inversion charge is not confined exactly at the interface but a distance determined by the maximum of the Stern-Howard wave function (figure 2.8(a)), this could contribute to lowering C_{tot} but not by the magnitude of the discrepancy between the XTEM images and electrical measurement. Especially when we consider that germanium has a higher relative permittivity than silicon.

An explanation is forthcoming when we consider that the silicon caps were not measured under a MOS gate stack in the XTEM, but on a piece of bare wafer surface

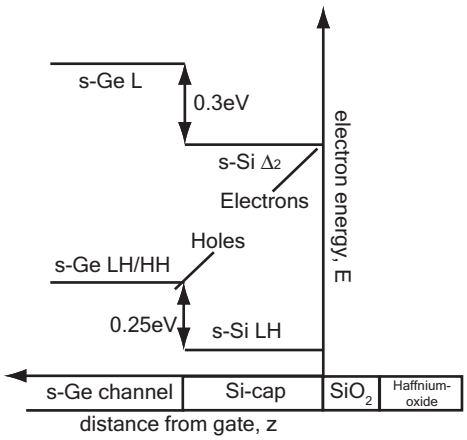


Figure 5.11: The band alignments of the s-Ge and s-Si cap lattice matched to a relaxed 80% Ge virtual substrate. At the flatband condition holes will occupy the s-Ge whereas electrons will occupy the Si-cap; however in inversion the bands will bend upwards for a p-MOSFET, resulting in holes occupying the Si-cap.

after device fabrication. High-k deposition is performed early on in the device fabrication process and therefore the silicon cap under the gate stack will have been protected from chemical etches performed later on in fabrication, whereas that on the wafer surface was exposed. This could well mean that the silicon cap in a device channel is greater than elsewhere on the wafer.

The accumulation region does not show any shoulder at all, suggesting that it is energetically favourable for electrons to occupy the silicon cap for any applied gate voltage. Assuming that the germanium and silicon cap are 100% strained (lattice matched to the underlying 80% SiGe buffer) we can use figure 2.17 to estimate the band offsets. These are shown in figure 5.11 and confirm that holes will preferentially occupy the s-Ge layer, whereas electrons will preferentially occupy the Si-cap.

5.6 p-MOSFET Mobility

I_d-V_g measurements were performed on many MOSFETs at 300 K on each wafer (W3–W6) and the peak field effect mobility extracted from the peak transconductance. The MOSFETs with the highest room temperature mobilities were cooled to 77 K and 4.2 K, where the peak field effect mobility was extracted again using the same method. Approximating the inversion charge to $C_{ox}(V_g - V_t)$ in the normal way will under estimate the peak field effect mobility, as the capacitance is significantly lower than C_{ox} due to the silicon cap adding a capacitance contribution in series. A better approximation for the inversion charge is to use the combined cap and oxide capacitance, $\frac{C_{cap}C_{ox}}{C_{cap}+C_{ox}}$, in place of C_{ox}. The peak field effect mobility calculated with and without the silicon cap correction are shown in figure 5.12 and compared to the best performing p-MOSFET on sample H reported in the previous chapter.

The mobilities of carriers in strained germanium p-MOSFET channels are far in excess of those on bulk-germanium. The highest peak mobility was measured at 4.2 K on sample W4 at 1780 cm^2 V^{-1} s^{-1} compared to the bulk-Ge mobility at the same temperature of 785 cm^2 V^{-1} s^{-1}. Note that without the silicon-cap correction (figure 5.12(a)) the peak mobility extracted from p-MOSFETs on W5 and W6 are far lower than those on W3 and W4, but correcting for the presence of the Si-cap (figure 5.12(b)) reduces the difference to be negligible at 300 K and 100 cm^2 V^{-1} s^{-1}-200 cm^2 V^{-1} s^{-1} ($\approx$10% reduction) at 77 K. It is not clear whether the thicker s-Ge-channel in samples W4 and W6 gives higher mobilities than the thinner channels in W3 and W5. At 300 K and 4.2 K the thicker channel appears to give a slight enhancement but at 77 K it does not.

Effective mobility was extracted as a function of vertical effective field by split C-V measurements at 300 K, 77 K and 4.2 K on the best-performing sample, W4.

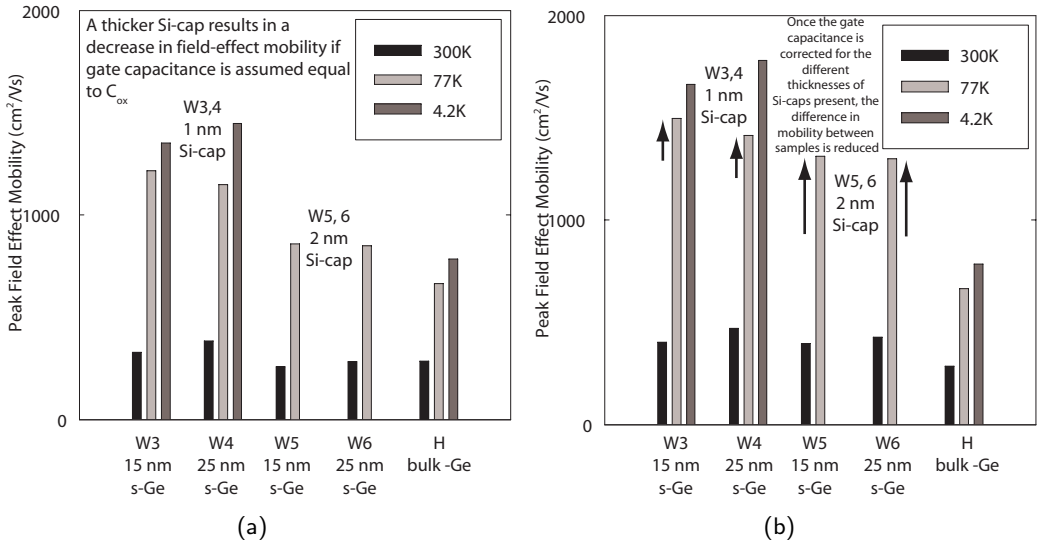


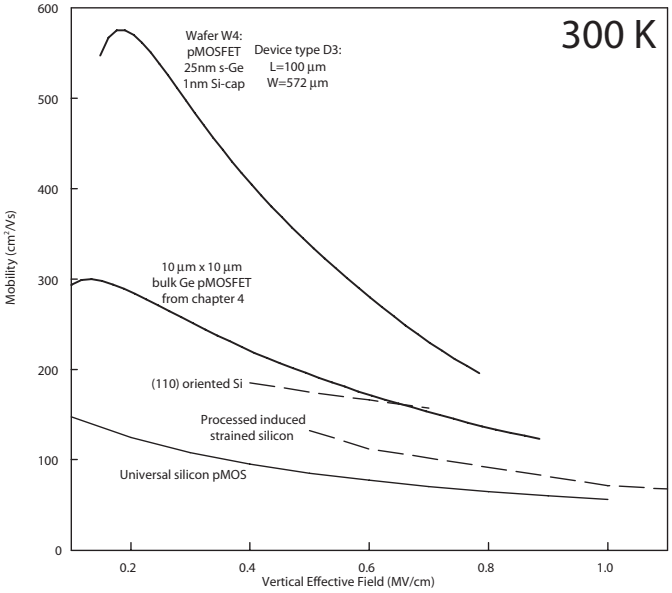
Figure 5.12: Peak field effect mobilities measured with $V_d = -0.05$ V for the best performing ring s-Ge p-MOSFETs (W3-W6) and bulk Ge p-MOSFET from the chapter 4 (sample H) at 300 K, 77 K and 4.2 K. (a) assumes that the gate capacitance is equal to C_{ox}=0.023 F m^{-2} whereas (b) corrects for the presence of the Si-caps substituting C_{ox} for $\frac{C_{cap}C_{ox}}{C_{cap}+C_{ox}}$. This should offer a more accurate mobility as otherwise the inversion charge is overestimated and hence the mobility is underestimated. The difference in apparent peak field effect mobility between samples is significantly reduced by taking the cap into account but it is not clear whether a thicker s-Ge layers has a significant effect on mobility.

This is compared to the silicon universal curve and that extracted from sample H in the previous chapter at 300 K, 77 K and 4.2 K in figure 5.13. In addition, at room temperature state-of-the-art silicon p-MOSFET mobilities are included for comparison.

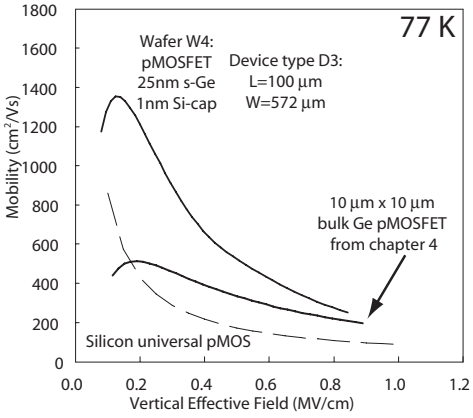
The effective mobility is far in excess of the silicon universal mobility over the whole vertical effective field range at all temperatures. The biggest enhancement is at low fields but towards high fields there is a rapid decrease towards a mobility comparable to that of the bulk-Ge p-MOSFET. This is due to significant population of the silicon cap by carriers and it expected that a larger mobility enhancement would be present at higher fields if the cap was thinner. Both the strained Ge and bulk Ge p-MOSFETs (chapter 4) exceed state-of-the-art silicon mobilities at room temperature, in particular the s-Ge devices. These mobilities are from recent publications by intel [Ghani et al., 2003] and IBM [Yang et al., 2003] and use novel techniques (process induced strain and (110) orientation) to enhance the silicon hole mobility.

5.7 s-Ge n-MOSFETs

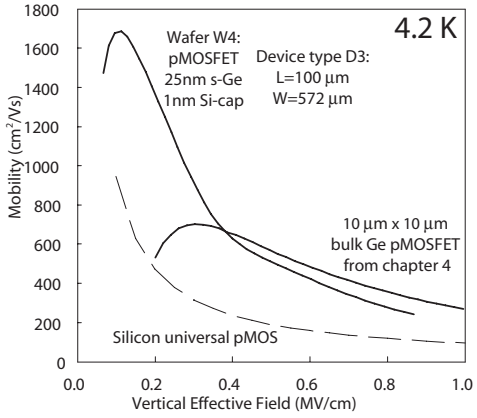
In contrast to the relatively good germanium p-MOSFETs presented in this chapter and the previous chapter, the s-Ge n-MOSFETs reported here perform poorly. These were fabricated alongside the s-Ge p-MOSFETs reported above but with n-dopants implanted into the source/drain regions and a slightly higher junction activation anneal of 500 °C compared to 450 °C for the p-MOSFETs. A further anneal performed on the p-MOSFETs at higher temperatures was shown to have no effect on device performance in the section 5.3.1. The n-MOSFETs on samples W1 and W2 have silicon cap targets of 1 nm and 2 nm respectively and identical s-Ge layers of 15 nm thickness. Figure 5.14 shows p-MOS capacitor C-V curves at 10 kHz of W1 and W2 that are very similar to that of the n-MOS capacitors on W3–W6. The holes (now the accumulation charge) still



(a)



(b)



(c)

Figure 5.13: (a) mobility against E_{eff} at 300 K of a s-Ge p-MOSFET and hydrogen annealed bulk Ge p-MOSFET from chapter 4. Mobility was measured with V_d=-0.05 V and exceeds state-of-the-art strained silicon [Ghani et al., 2003] and (110) oriented silicon [Yang et al., 2003] p-MOSFET mobilities (data taken from Dobbie et al. [2008]). Ge p-MOSFET mobilities are also compared to silicon universal p-MOS at 77 K (b) and 4.2 K (c).

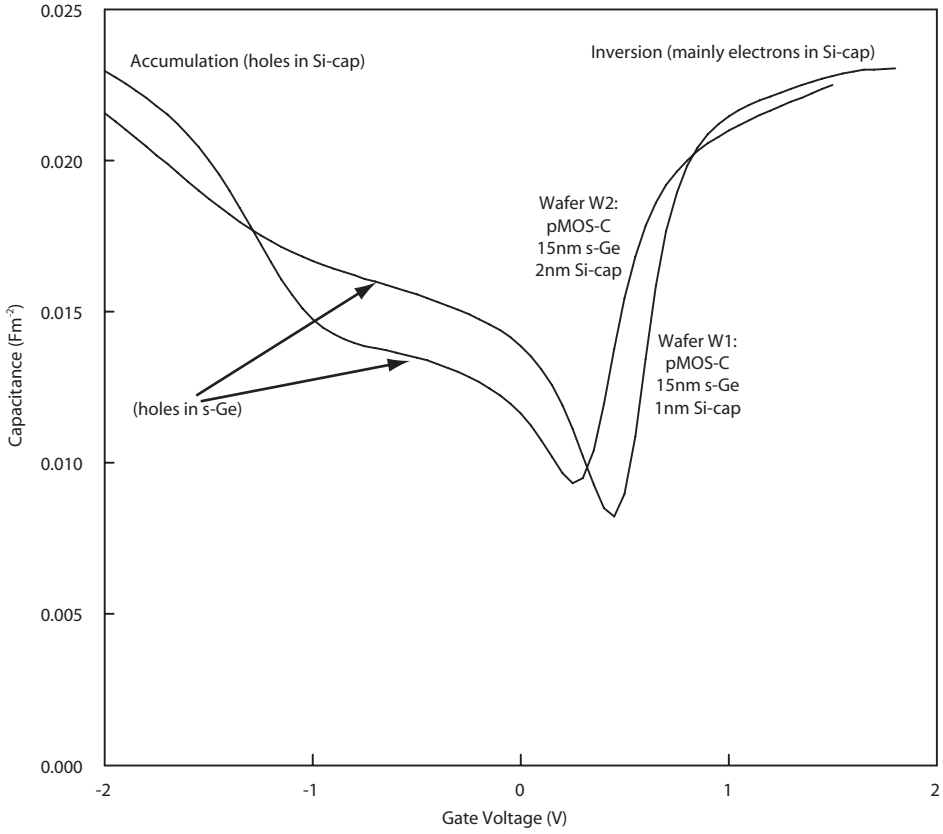


Figure 5.14: C-V curves measured for 250,000 μm^2 p-MOS capacitors located on W1 and W2 performed at 500 Hz.

occupy the s-Ge before being confined in the Si-cap at higher gate voltages, as indicated by the shoulders in the profile. Electrons (inversion charge) still show no confinement to the s-Ge channel and occupy the Si-cap, as expected for the band offsets between the s-Ge and Si-cap calculated previously (figure 5.11).

It could be argued that the inversion charge is entirely confined in the silicon-cap and therefore the s-Ge channel layer is superfluous. However, consider that the inversion layer should be thick enough to penetrate some way into the germanium if only in the case of the W1 with the thinner Si-cap. Drain current as a function of gate voltage at low

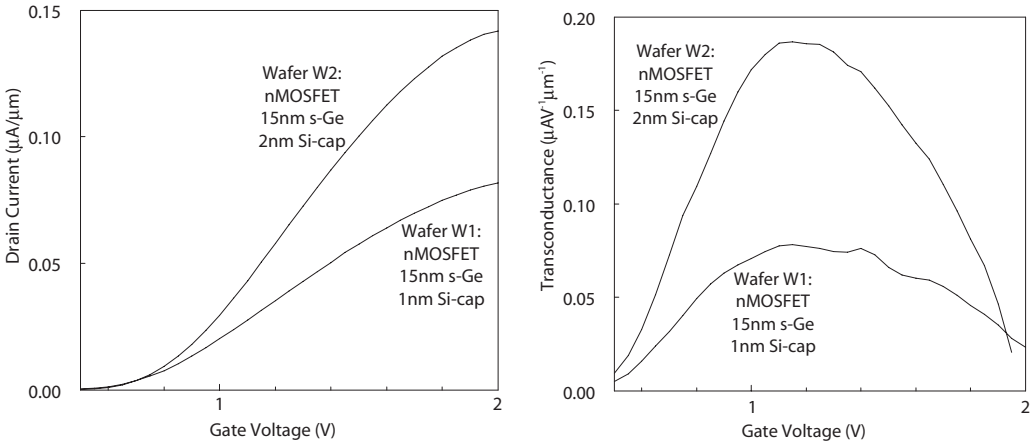
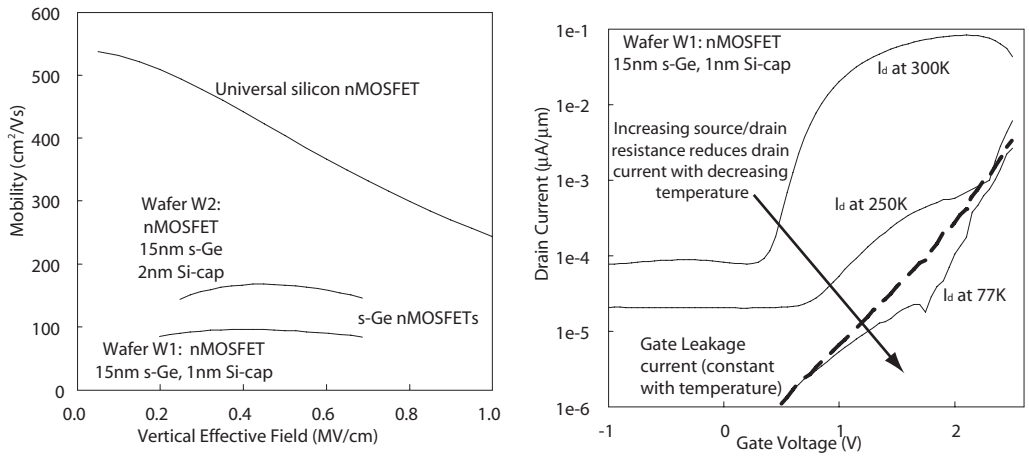


Figure 5.15: Drain current (a) and corresponding transconductance (b) measured at 300 K with $V_d = 0.05$ V for ring n-MOSFETs of size D3 (L=100 μm, W=572 μm) on wafers W1 (1 nm Si-cap) and W2 (2 nm Si-cap). The latter has roughly double the peak transconductance of the former, suggesting that a thicker Si-cap improves channel mobility. It is thought that more of the inversion charge in W1 occupies the s-Ge channel, which has a lower mobility than the Si-cap and thus decreases device performance overall. It is this sample therefore, that better represents a true s-Ge n-MOSFET with no Si-cap.

drain bias, along with transconductance curves, are shown in figure 5.15. Note that W1 performs significantly worse than W2 with a lower on-current and transconductance. If all carriers were confined to the Si-cap in both cases; we would expect to see no difference in device performance or if anything, W2 to perform worse as its thicker silicon cap has more chance of relaxing due to its increased thickness. This would introduce defects, thus reducing the mobility of carriers in this layer further [Olsen et al., 2005]. The fact that W1 shows poorer characteristics suggests that more carriers occupy the s-Ge layer in this sample due to the thinner Si-cap, which would appear to result in worse device characteristics.

Split C-V was performed on n-MOSFETs located on W1 and W2 to extract



(a) (b)

Figure 5.16: Effective room temperature mobility against vertical effective field measured with $V_d = 0.05$ V of D3 (L=100 μm, W=572 μm) s-Ge n-MOSFETs (a) compared to the universal silicon n-MOSFET mobility curve. Low temperature mobility extraction was not possible due to decreasing drain current with decreasing temperature as shown in (b). Below room temperature, gate leakage, which does not vary with temperature, completely dominates the on-state drain-current due to increasing source/drain resistance.

effective mobility as a function of vertical field and is compared to the universal silicon n-MOSFET curve at room temperature in figure 5.16(a). Unsurprisingly both W1 and W2 have mobilities far lower than this. No extraction could be performed at lower temperatures as drain current decreased greatly with decreasing temperature in both samples (figure 5.16(b)). It is assumed that this is due to increased source/drain resistance caused by dopant freeze-out, which would have a significant effect if the doping density were too low. As shown previously, gate leakage is invariant with temperature meaning that at temperatures of 200 K and under, all drain current measured was due to gate leakage.

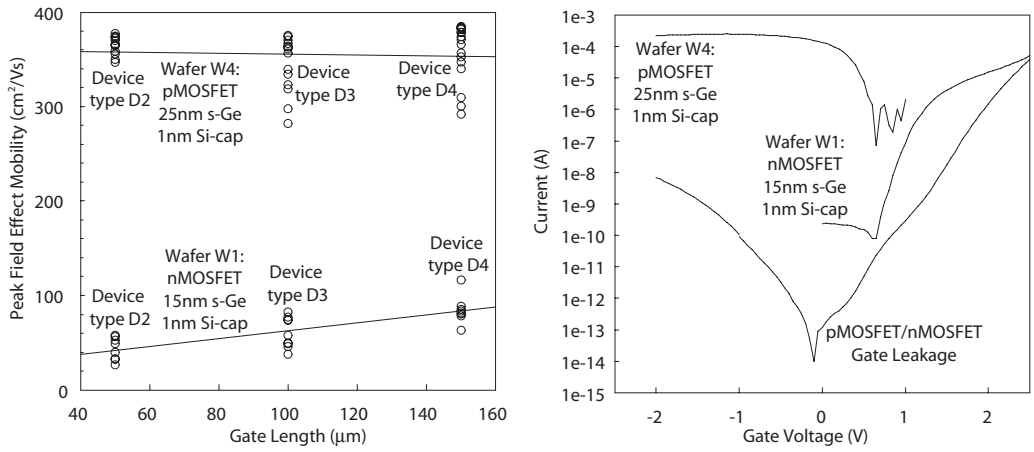
Peak field effect mobility was extracted from peak transconductance for many devices of various gate lengths on sample W1. The mobilities extracted are plotted

against gate length and compared to those for s-Ge p-MOSFETs on sample W4 in figure 5.17(a). Whereas there is no correlation of peak mobility with device length for sample W4, peak mobility decreases significantly with decreasing device length in the case of W1. The source/drain resistance must be exceedingly high to be a significant factor for the long gate lengths here. Finally we directly compare absolute drain and gate currents measured for an n-MOSFET on W1 and p-MOSFET on W4 at room temperature in figure 5.17(b).

The drain current measured in the s-Ge n-MOSFET device is one to two orders of magnitude lower than in s-Ge p-MOSFET, which is explained, in part, by the higher source/drain resistance. The gate leakage as a function of gate voltage is identical for both but is higher for positive gate voltages than the equivalent negative gate voltage. Therefore, gate leakage is greater for the on-state for n-MOSFETs than for the p-MOSFETs and combined with the reduced drain current, means that a large proportion of the device on-current is gate leakage. This increased gate leakage is qualitatively consistent with the band-offsets between the hafnium-oxide and semiconductor conduction bands. The hafnium-oxide conduction band offset is far lower than the corresponding valence band offset with both silicon [Robertson, 2006] and germanium [Afanas'ev et al., 2004]. Hence, holes have a greater energy barrier to overcome in order to enter the gate dielectric than electrons.

5.8 Summary

Compressively strained germanium MOSFETs have been electrically characterised. Whereas p-MOSFETs show a large enhancement over the silicon universal curve, n-MOSFETs perform poorly. The silicon cap, used for surface passivation before oxidation and high-k deposition, was thicker than specified, as determined from XTEM and capacitance



(a) (b)

Figure 5.17: Room temperature peak field effect mobility measured with $V_d = 0.05$ V as a function of s-Ge MOSFET channel length (a) showing that even with these long devices, source/drain resistance is still an issue for the n-MOSFETs as indicated by the decreasing mobility with decreasing channel length. A comparison of absolute drain currents and gate currents of s-Ge D3 (L=100 μm, W=572 μm) p-MOSFET and n-MOSFET reveals that the latter's downfall is it's aforementioned large source/drain resistance and high gate leakage.

measurements, and played a large role in the performance of the p-MOSFETs and n-MOSFETs. Whereas holes preferentially occupy the s-Ge at low gate overdrives in the former, electrons always occupy the Si-cap; hence, there is a large mobility enhancement for holes in the p-MOSFETs until conduction in the cap at high gate overdrives, in contrast to electrons in the n-MOSFETs where most conduction takes place through the Si-cap at all gate overdrives.

A thicker Si-cap results in an improved n-MOSFET mobility. It thought that this is because electron mobility in the Si-cap is higher than in the s-Ge channel and that those n-MOSFETs with the thinner Si-cap have an inversion layer that penetrates significantly into the s-Ge, thus lowering the overall measured mobility. Even in the relatively large n-MOSFETs (50–150 μm gate lengths) measured here, source/drain

resistance is significant and degrades the mobility measured. A reduced measurement temperature (250 K and lower) results in a greatly decreased drain current. It is thought that this is due to the doping density in the source/drain region being too low, and hence the drain current is reduced due to dopant freeze-out with a relatively small decrease in temperature. In turn, this gives rise to a large increase in source/drain resistance. At temperatures of less than 200 K, the drain current in the on-state is equal to the gate leakage current, and hence a low temperature electrical characterisation of these devices was not possible.

In the case of the s-Ge p-MOSFETs, the mobility measured far exceeds that of state-of-the-art strained and (110) oriented silicon p-MOSFETs. Although there is undoubtedly a large mobility enhancement due to strain, a portion of the overall mobility increase must be due to the fact that the inversion charge carriers are removed from the dielectric interface by the Si-cap. Indeed, Nicholas et al. [2007a] shows that for these particular devices, it is the thicker Si-cap that gives the highest peak mobility. This is consistent with the findings of the previous chapter, that demonstrates that surface roughness scattering and Coulomb scattering, both at the dielectric interface, to be the dominating factors in limiting the mobility at room temperature.

Chapter 6

Germanium Condensation

6.1 Overview

The germanium-condensation technique has received much attention from the semiconductor industry in the last few years [Bedell et al., 2004]. The large systematic studies performed require access to state-of-the-art processing facilities and a large number of ultrathin-SOI (30 nm) wafers — the latter, in general, not being commercially available.

The investigation here, is performed on a smaller scale and the wafers are the thinnest (100 nm) CVD-grade SOI that could be obtained commercially in a small quantity. This makes it interesting for two reasons: first, Ge-condensation has not been attempted on an initial substrate this thick and second, if successful it would mean that thin (S)GOI layers were obtained from relatively thick initial SOI wafers. The latter is of technological interest as currently, the Ge-condensation technique offers no advantage over wafer bonding in terms of processing [Bedell et al., 2006] as one must begin with an ultra-thin substrate. If a thicker, and therefore more simple to manufacture, initial SOI substrate could be substituted for the thinner, then Ge-condensation could be a

more commercially viable way of obtaining thin (S)GOI.

6.2 Process Design and Execution

There are three elements needed for the Ge-condensation process: SOI substrate(s), SiGe-layer growth and oxidation(s). All must be of the highest quality if the process is to succeed and hence each was outsourced to a specialist company and the various parts of the project coordinated by the author.

Seven SOI wafers were used in total, all with ≈230 nm buried oxide layer (BOX) and 100 nm silicon-on-insulator top layer (SOI). Although the SOI layer thickness is highly controlled, the buried oxide layer is approximate. It is relatively uniform over each wafer but can vary significantly between wafers. These were purchased from Soitec, a leading manufacturer of industry standard SOI wafers, and were manufactured by wafer bonding a donor silicon layer onto an oxidised Si-wafer.

On the author's request, a SiGe layer capped with a thin layer of silicon (≈5 nm) were grown on each wafer by AdvanceSiS using CVD. The SiGe layer had a target Ge-composition of 15% and target thickness of 60 nm. Although calibration growths of these layers, performed on bulk silicon wafers, came out almost exactly to specification (confirmed with XTEM by the author; ellipsometry and X-ray refraction by AdvanceSiS — none shown), the growth of the layers on the SOI substrates gave a thinner SiGe layer (55 nm) with a lower Ge-composition ($x = 0.12$) than specified (figure 6.1(a)). This is thought to be due to the lower wafer surface temperature during growth because of the insulating buried oxide layer.

One wafer was characterised by XTEM (sample 1) to measure the initial layer thicknesses and then the remaining six processed at Innos (a spin-off company of the University of Southampton based in Eindhoven, Netherlands; specialising in silicon-based

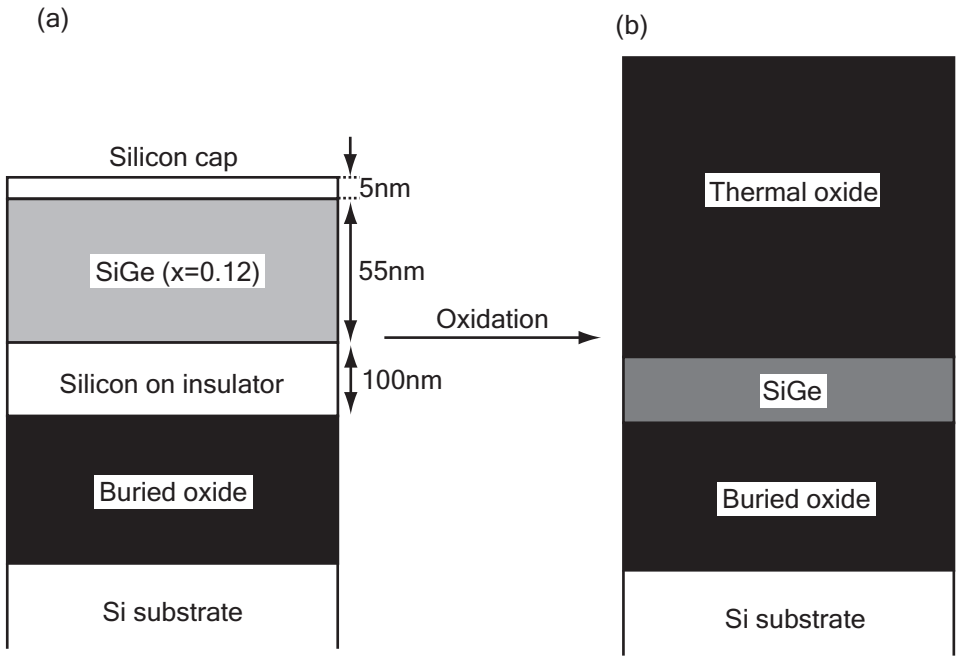


Figure 6.1: Initial (a) and post oxidation (b) Ge-condensation layers used here. The ultimate aim is to obtain a Ge-composition close to 1 but there will be many intermediate samples with low x.

device fabrication), where various oxidations were performed (figure 6.1(b)). To obtain a high composition of germanium in the final SiGe layer (>50%) its thickness will have to be less than or equal to 10 nm (equation 2.79), which requires about 350 nm of silicon dioxide (equation 2.82) to be grown. This is an extremely thick oxide compared to those grown in more standard silicon fabrication processes. A stepped oxidation temperature profile was employed to maximise diffusion of the germanium into the substrate whilst ensuring that the lower-melting temperature of the higher SiGe composition layers was not exceeded in the latter stages of the process [Mukherjee-Roy et al., 2005]. This had the added advantage of reducing the total oxidation time compared to the use of a constant, lower oxidation temperature.

To oxidise the silicon cap without germanium diffusing to the surface, a rapid

thermal oxidation (RTO) at 1150 °C was applied first. This novel step minimises diffusion as the sample is heated to the oxidation temperature in a few seconds, where the silicon cap is oxidised immediately, thus preventing Ge reaching the surface before an oxide is formed. This eliminates the need for an initial low temperature oxidation before full ramp-up as suggested by Sugiyama et al. [2007].

After the RTO, an oxidation at 1050 °C was performed in a standard oxidation furnace with a relatively long ramp-up time in a nitrogen atmosphere. To obtain a particular thickness of oxide, the oxidation process was repeated several times and between successive oxidations, the oxide thickness measured. Once the oxide thickness grown reached 200 nm, the oxidation temperature was reduced to 900 °C and more successive oxidations performed until a final target oxide thickness of 350 nm was obtained.

Between oxidations, oxide thicknesses were measured using a Nano-Metrics NanoSpec — a simple spectroscopic, single-layer thin-film thickness gauge, designed for non-destructive rapid characterisation and usually employed in fabrication facilities. Note that bulk-Si wafers (called drop-outs) are oxidised alongside the Ge-condensation samples and the oxide thickness grown is measured from these as the SGOI layer structure is too complicated for this type of characterisation. It was assumed that the oxide grown on a bulk-Si wafer is equal to that grown on the SiGe-structures [Sugiyama et al., 2004].

Note that the oxidation times for each stage are approximate as the moments at which oxidation begins and ends are generally not well defined. However, we are more interested in the oxide thicknesses grown than the times taken to grow them. Only sample 7 was subjected to the entire process described above. Samples 2–5 were aborted at various intermediate points so that they could be characterised. Table 6.1 shows how much of the entire Ge-condensation process, as described above, each sample went through.

Sample number	Oxidation temperature 1 (°C)	Oxidation time 1 (minutes)	Oxidation temperature 2 (°C)	Oxidation time 2 (minutes)	Oxidation temperature 3 (°C)	Oxidation time 3 (minutes)
1	–	–	–	–	–	–
2	1150	2.5 (RTO)	–	–	–	–
3	1150	2.5 (RTO)	1050	30	–	–
4	1150	2.5 (RTO)	1050	190	–	–
5	1150	2.5 (RTO)	1050	190	900	450
6	1150	2.5 (RTO)	1050	190	900	950
7	1150	2.5 (RTO)	1050	190	900	1500

Table 6.1: The oxidations comprising the Ge-condensation process that were performed on each sample before characterisation.

6.3 Characterisation

Each sample was characterised with XTEM and SIMS, which allowed the thickness and germanium composition of the various layers at the surface of each sample to be determined. These are listed in table 6.2.

SIMS measurements were performed by Richard Morris of the Analytical Science Projects Group, Department of Physics, University of Warwick; on the instruction of the author. The sample surfaces were sputtered with a 1 keV O_2^+ beam and the secondary ion fluxes for Si and Ge ejected from the surface was measured. This low energy beam was chosen to give a high depth resolution. At this low beam energy the recorded signal levels of each element correspond to the ratio of Ge and Si in the sample [Dowsett et al., 2003]. Positive charge build-up on the surface, which could deflect the primary beam, occurred due to the insulating BOX layer. This was counteracted by use of an electron beam, which required careful calibration.

The sputtering time was converted to a crater depth into the sample by sputtering sample 1 for a series of different amounts of time and measuring the corresponding crater depths. Note that calibration was only performed to depths not greater than the Si/SiGe layers and hence the sputter rate through silicon dioxide was not calibrated. Thus, any conversion of a sputtering time to a depth is only valid for regions of Si/SiGe and note that although the sputter rate is assumed constant through these regions, germanium does in fact increase the sputter rate.

The buried oxide was always sputtered through completely so that a constant silicon signal level was obtained corresponding to the underlying Si-substrate, on which the BOX sits. This signal was consistent over all of the samples and hence the back interface of the BOX layer could always be identified in the SIMS profiles. A comparison of two SIMS profiles as a function of sputtering time is shown in figure 6.2.

As the silicon dioxide sputter rate was not calibrated, the absolute depth of the SiGe layers is unknown due to different thicknesses of silicon dioxide on the sample surfaces. However, this oxide is of little interest and hence when converting the SIMS profiles from a function of sputtering time to depth, the top BOX/Si(Ge) interface is defined as zero depth. The SIMS profile as a function of depth corresponding to a XTEM image of sample 1 is shown in figure 6.3 alongside the calculated germanium fraction through the sample depth by comparing the signal ratio of Si to Ge.

This was repeated for all of the samples to give a SIMS profile corresponding to a XTEM image for each, shown in figure 6.4. The average germanium composition of each layer in table 6.2 was calculated by integrating over the SIMS profile of the SiGe layer and dividing by the layer thicknesses measured from each corresponding XTEM image. The expected Ge-composition of these final layers was calculated from the measured final thicknesses and the initial layer SiGe compositions/thicknesses using equation 2.79.

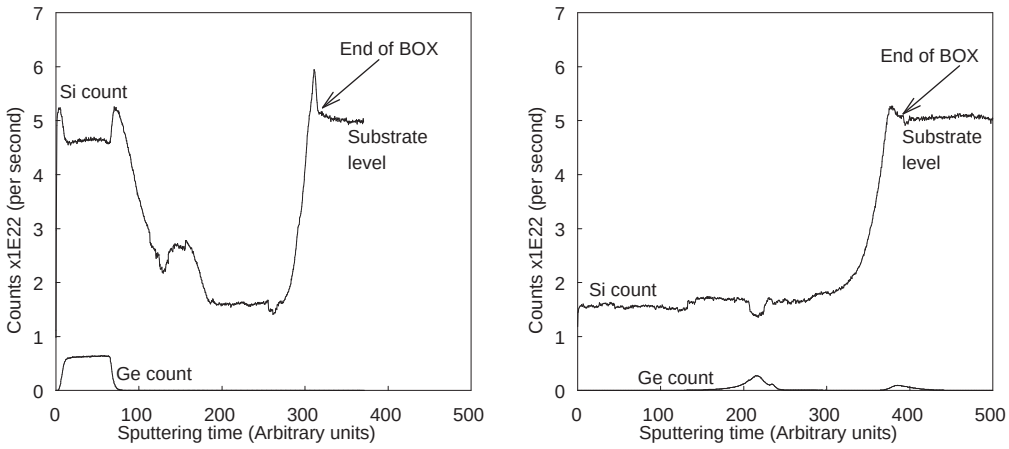


Figure 6.2: SIMS profile for the initial structure, sample 1 (a) compared to that for sample 3 (b), which has been oxidised.

The germanium-condensation process applied was a success, in that the final layers were crystalline (confirmed by observing the diffraction pattern of crystalline SiGe using XTEM) and of a greater average Ge-composition x' than the initial layers. The final layer in sample 7 (figure 6.4(f)) has a Ge-composition of 58% and is 10 nm thick. However, early on in the process (sample 3, figure 6.4(b)) dislocations in the SiGe layer and agglomerated germanium in the in thermal oxide are visible. The SIMS profile confirms the presence of Ge in the oxide.

XTEM and SIMS were used to measure the layer thicknesses and unfortunately both are prone to over estimating the thickness of the final 10 nm SGOI layer. Although not measured, we must assume that the final layer has significant residual compressive strain, as consistent with the wider Ge-condensation literature [Nakaharai et al., 2003; Bedell et al., 2006]. The XTEM image will therefore have a component of contrast resultant from the strain. This strain, and hence contrast, will extend some distance into the surrounding oxide.

Sample number	Measured SiGe thick-ness (nm)	SOI thick-ness (nm)	Peak Ge como-sition	Average Ge compo-sition	Oxide thick-ness (nm)	Expected average Ge compo-sition	Melting temperature of SiGe layer (°C)
1	55	100	0.12	0.12	0	–	1300
2	55	100	0.11	0.11	20	0.12	1300
3	130	0	0.17	0.04	70	0.05	1300
4	70	0	0.17	0.08	204	0.09	1300
5	50	0	0.29	0.11	250	0.13	1200
6	30	0	0.40	0.17	300	0.21	1150
7	10	0	0.77	0.58	340	0.63	1000

Table 6.2: The measured layer thicknesses and compositions of each sample determined with a combination of XTEM and SIMS measurements.

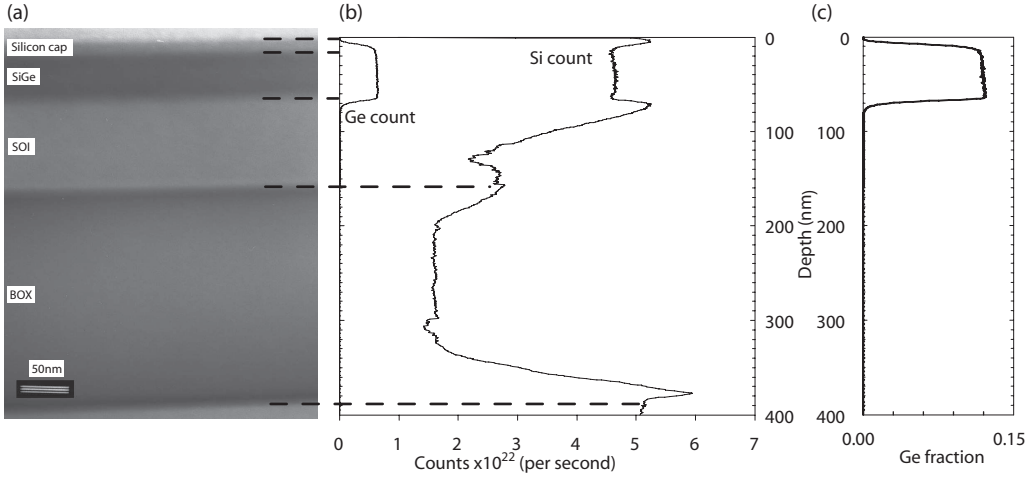


Figure 6.3: XTEM image of sample 1 (a), and the SIMS profile as a function of depth (b) with Ge-fraction as a function of the same calculated from the ratio of Si and Ge fluxes (c).

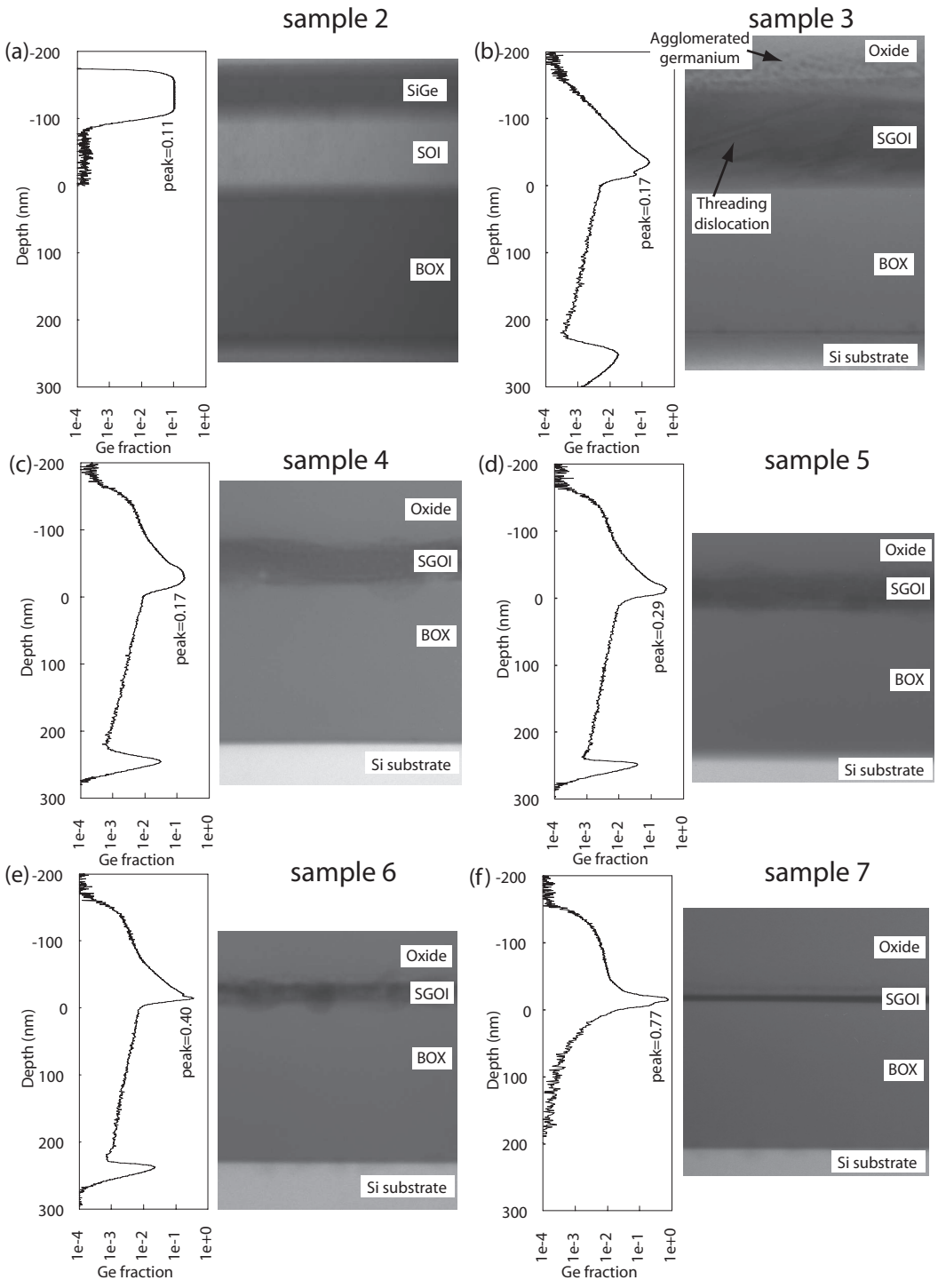


Figure 6.4: Ge-depth profiles determined from SIMS and corresponding XTEM images of samples 2–7 (a–f respectively).

SIMS is prone to broadening of features as a function of depth and this generally increases with primary beam energy. This, along with strain contrast in the XTEM, means that the final SGOI layer is most likely thinner than reported here. Furthermore, as the layer thickness is used to calculate the average Ge-composition of the SGOI layer, this will be underestimated. One solution is to perform SIMS at successively higher energies to increase the broadening effect. The apparent SGOI layer thickness could then be plotted as a function of primary beam energy and an extrapolation to the thickness at "0-energy" could be performed, which should correspond to the condition of no broadening and hence give the real layer thickness.

6.4 Germanium Diffusion in the BOX

Note that the germanium composition of the Ge-condensated layers (samples 3–7) are lower than expected in each case, which means that the germanium is not completely conserved by the diffusion barriers of the BOX and thermal oxide. The SIMS profiles (figure 6.4(b)–(e)) suggest that germanium is diffusing through the buried oxide layer during the oxidations. The Ge-diffusion profile through the BOX is consistent over all the samples except in samples 1 and 2 where no Ge is at the BOX/SOI interface and sample 7 where the germanium tends to a negligible level a short distance ($\approx$80 nm) into the BOX. In addition, we observe that where there is Ge-diffusion through the BOX, there is a high concentration of germanium at the BOX/Si-substrate interface. Note that the germanium fraction through the buried oxide is low, exponentially decreasing from 1% to 0.1% from the top to the bottom of the BOX layer. Most publications focus on the germanium profile through the SGOI layer and disregard the BOX entirely; however, we will look for evidence of germanium diffusion through the BOX in the published literature.

A SIMS profile in Sugiyama et al. [2004] shows germanium content in the BOX layer of about 1%–2%. A linear scale of germanium fraction is used and the profile remains constant to a depth of 50 nm into the BOX. However, the wafers used are not bonded but fabricated by the SIMOX process, which involves implanting oxygen into the substrate and annealing to form a BOX layer. Therefore, incorporation of Ge at this stage cannot be ruled out and hence, the validity of the comparison with samples reported here is questionable.

Di et al. [2005a] and Di et al. [2005b] show evidence for Ge diffusion into the BOX in high resolution XTEM images. Energy dispersive X-ray spectra (EDS) of the same sample [Di et al., 2005a] suggest a concentration of Ge of $\approx$0.04% about 5 nm into the buried oxide layer.

The use of linear intensity scales for Ge-composition through the BOX layer will conceal low Ge concentrations in most publications. Park et al. [2006] includes a before and after Ge-condensation SIMS profiles plotted on a logarithmic intensity scale. This shows germanium in the buried oxide, at the BOX/Si-substrate interface both before and after Ge-condensation. No mention is given to this large anomalous germanium peak and therefore the publication is of little use here.

Out of the context of Ge-condensation, germanium is well-known to diffuse substitutionally in silicon dioxide with the diffusion coefficient given by Minke and Jackson [2004]. The movement of Ge in SiO_2 is often exploited to obtain Ge-quantum-dots as it can agglomerate under certain conditions. This is demonstrated by Nobutoshi et al. [2007] with the low temperature annealing of ion-implanted germanium in a SiO_2 layer on a silicon substrate; however, more relevant is the reported behaviour (in the same research paper) of the germanium at higher anneal temperatures (900 °C and above) where germanium pile-up at the Si/SiO_2 interface is observed. This is exactly what we

observe, at the same temperature or higher temperature, in samples 3–6.

The reason given for this pile-up by Nobutoshi et al. [2007], is that the germanium diffuses at different rates in silicon and silicon dioxide. We would expect pile-up of Ge at this interface if it diffused more slowly in Si than in SiO_2, but the reverse is true: the diffusion rate of germanium is far greater in silicon than in silicon dioxide [Minke and Jackson, 2004; McVay and DuCharme, 1974]. However, consider that the diffusion of Ge in SiO_2 is for a block of silica and not thermally oxidised silicon; therefore, we cannot rule out that Ge could diffuse faster than suggested by Minke and Jackson [2004]. In fact, a principal conclusion of the paper is that Ge diffusion in silicon dioxide is defect mediated and if the BOX here, or silicon dioxide in Nobutoshi et al. [2007] obtained by the thermal oxidation of silicon had a higher density of defects, then we would expect to observe a higher Ge-diffusion rate. Hence, germanium could be diffusing faster through the oxide than the underlying Si-substrate, in which case pile-up would be expected at this interface. However, we cannot rule out that the observed Ge-peak is an artefact from the SIMS measurements, for example a discharge of positive ions from the surface when the insulating BOX is breached by the primary beam.

Sample 7 does not show significant germanium in the BOX. Note that it has been subjected to an anneal over 500 minutes longer at 900 °C than sample 6 and hence if the diffusion of Ge into the BOX stopped, then the Ge already there would most likely have time to dissipate into the Si-substrate. A small Ge-peak is observed at the BOX/Si-substrate interface in the SIMS profile for sample 7 (not shown), but it is barely above the background noise, peaking at 0.0001% Ge. It is therefore assumed that the Ge-diffusion into the BOX does stop at some point in the final stages of the oxidation process that only sample 7 was subjected to.

As mentioned in section 2.6.1, little or no consideration has been given to the

internal thermal oxidation of the buried oxide in previous publications regarding Ge-condensation. Here, we calculate the rate at which the buried oxide is growing due to ITOX in the Ge-condensation process performed. It is assumed that the reaction rate constants k_1, k_2 and k_3 (see equations 2.87–2.89) are the same as for the oxidation of pure silicon, as the dry-oxidation rate of silicon at the interface is not enhanced by the presence of germanium [Sugiyama et al., 2004]. The end product of the reaction is the same (SiO_2) as germanium is rejected into the semiconductor and therefore, reactions involving germanium are ignored.

Another important parameter is the diffusion rate of oxygen through silicon D_{Si}. This is applicable to pure SOI substrates, but here we have SGOI substrates of various Ge-compositions. There is no data available for the diffusion rate of oxygen in SiGe alloys but it can be assumed that D_{Si} will suffice for the low Ge composition alloys in the majority of samples here; however, for higher germanium concentrations (sample 7), the diffusion of oxygen through the SiGe layer may be more accurately described by the diffusion coefficient of oxygen through pure Ge (units of $cm^2\ s^{-1}$) [Haas, 1960]:

$$D_{Ge} = 0.17\exp\left(-\frac{2.02}{k_B T}\right) \tag{6.1}$$

The calculated internal oxidation rate of the BOX is plotted in figure 6.5. As the SiGe layer decreases in thickness, we see that the ITOX rate increases. Due to the greater diffusion rate of oxygen in germanium, compared to silicon, we expect a further increase in the ITOX rate towards the final stages of the Ge-condensation process.

When oxidising SiGe, if there is a sufficient supply of silicon to an oxidising interface, germanium is rejected from the oxide. The same should be true for the internal oxidation of the buried oxide. However, in the absence of significant internal oxidation, germanium could diffuse into the buried oxide and note that the ITOX rate is an order of magnitude lower in samples 5 and 6, compared to sample 7 (if one includes

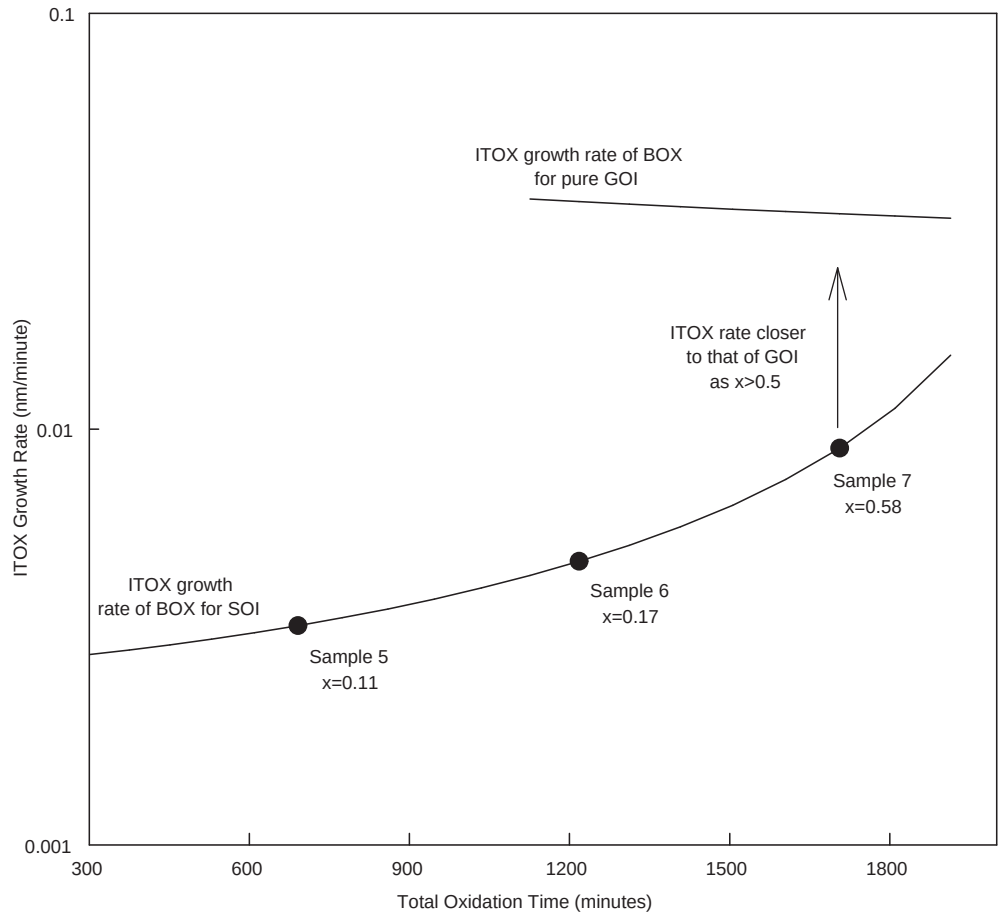


Figure 6.5: The calculated rate at which the internal thermal oxidation (ITOX) of the BOX layer is progressing in the final stages of the Ge-condensation process. The ITOX rate is calculated assuming oxygen diffusion through the SiGe layer is that for pure Si and pure Ge.

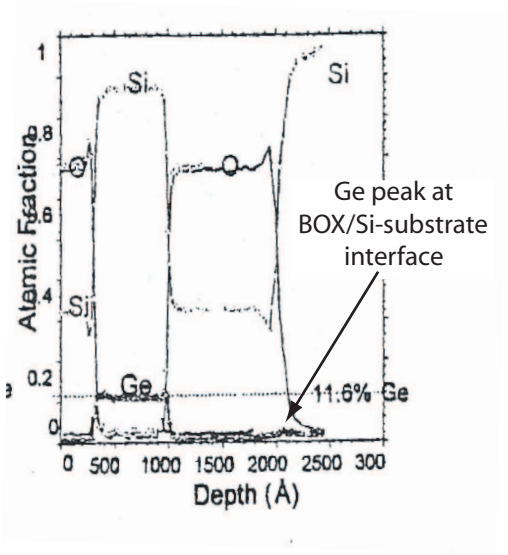


Figure 6.6: SIMS profile of a sample that has undergone Ge-condensation followed by a 2 hour anneal in argon at 1050 °C. Although plotted on a linear scale, a small Ge-peak can be seen at the BOX/Si-substrate interface. Reproduced from Sadaka et al. [2004].

the enhancement due to increased Ge composition). Therefore, it is proposed that the internal oxidation of the buried oxide layer is responsible for acting as an efficient diffusion barrier for germanium and not the static buried oxide layer alone, as generally assumed in the Ge-condensation literature.

Sadaka et al. [2004] includes a SIMS profile of Ge-condensation sample that has been annealed in argon for two hours at 1050 °C, reproduced here in figure 6.6. Although on a linear intensity scale, a Ge-peak at the BOX/Si-substrate interface is clearly visible. This could be the same Ge pile-up effect that is observed here and reported in Nobutoshi et al. [2007], and could suggest that there was significant Ge-diffusion into the BOX during the argon anneal. There is no internal oxidation of the buried oxide layer when annealing in argon and hence, if the ITOX hypothesis is correct, then we would expect to see germanium diffusing into the BOX in this case.

Balakumar et al. [2007] reports on the observation of SiGeO forming at both the top oxidising interface (as would be expected before Ge-rejection into the SiGe layer) and at the BOX/SiGe interface during a standard Ge-condensation process. Despite witnessing the formation of GeO at the BOX interface, they give no mention of the internal oxidation of the buried oxide, which could provide an alternative explanation to the one given (to do with the metastability of the SiGe at temperatures close to the L-S region).

6.5 Summary

An initial 55 nm SiGe layer with 5 nm Si-cap on 100 nm SOI substrate was oxidised for various times and temperatures to obtain a series of samples with various SGOI layer thicknesses and Ge-compositions. 1700 minutes of oxidation resulted in a final 10 nm thick SGOI layer with a Ge-composition of 58% and this is the thinnest, highest Ge-composition layer ever obtained, from thick (t_i'=160 nm) initial Si/SiGe layers of low Ge-composition (x'=4%). Furthermore, this layer could be thinner and, therefore, have a higher Ge-composition than reported here due to systematic errors occurring in both XTEM and SIMS measurements.

Ge was found to be present in the BOX after longer oxidations, except in sample 7, which underwent the longest final oxidation. Although the presence and subsequent absence of Ge in the BOX is explained with a hypothesis not inconsistent with the wider literature regarding Ge-condensation, it should be emphasised that results are based on a characterisation of a single sample set. Therefore, repeatability of this anomalous Ge-diffusion and the confirmation that ITOX does play a significant role in stopping this, has not been demonstrated conclusively. A peculiarity of these samples, for example inconsistent or non-uniform initial SOI wafers, cannot be ruled out.

Chapter 7

Conclusion and Further Work

A future CMOS technology with germanium as the active MOSFET channel material requires high-quality substrates and devices that offer a significant enhancement over the silicon equivalent. Both of these have been investigated, to varying degrees, in this work, which we now conclude.

State-of-the-art bulk and strained Ge p-MOSFETs with a high-k gate stack showed a large mobility enhancement over the universal silicon mobility and, particularly in the case of the latter, state-of-the-art silicon p-MOSFETs (strained [Ghani et al., 2003] and (110) [Yang et al., 2003] fabricated by intel and IBM respectively). The gate stack is far from perfect: a large variation of interface state densities in the bulk-Ge devices was observed. The states had a fast response time suggesting a close physical proximity to the channel; hence it was concluded that they reside at the Si/SiO_2 interface. Consistent with this, is that a hydrogen anneal greatly reduces the interface state density of a given device. Indeed, it was the hydrogen annealed devices that offered the highest mobilities.

The systematic variation of interface state densities allowed a detailed investigation into their effect on channel mobility. It was demonstrated, for the first time, that

peak field effect mobility is inversely proportional to the threshold voltage of a device if interface states dominate both parameters. The former is reduced by Coulomb scattering of carriers in the channel; whereas the latter is shifted by the filling of states, rather than the population of the channel with carriers, in the subthreshold regime.

Effective mobility measurements were made on a more limited number of devices at a temperature of 4.2 K. This required the development of a low temperature probe station and reliable measurement procedure, which are both documented in this thesis. To the resulting effective mobility curves (corrected for both source/drain resistance and diffusion) was fitted a theoretical mobility curve, which was the sum of a Coulomb scattering component and a surface roughness scattering component. The theoretical mobility was calculated using the model proposed by Gold and Dolgopolov [1986], which assumes that $T = 0$ and that the electrical quantum limit applies (this was demonstrated to be the case). Fitting parameters of areal charged impurity density and RMS surface roughness were used to adjust the theoretical mobility curve to obtain the best fit to the data.

The areal charged impurity density correlated to the measured interface state density for each device, confirming that Coulomb scattering by interface states is indeed responsible for the mobility variation. More interesting however, is the 20% decrease in RMS surface roughness, inferred by the model, after a device has been hydrogen annealed. Given the complexity of the gate stack, it is difficult to suggest an exact physical mechanism for this decrease but a diffusion based process is feasable. To the authour's knowledge, this has never been observed at a MOSFET gate-dielectric/channel interface before. Hence, this result does need to be confirmed, either by applying a more detailed model to the existing data or by a direct physical characteristion of this interface. Note that on the instruction of the author, computer simulations of these devices are

currently being performed at Universidad De Granada, Spain.

Strained Ge n-MOSFETs were also fabricated alongside the s-Ge p-MOSFETs at IMEC. They suffer from a particularly high source/drain resistance due to a low dopant density and, partly due to this, have mobilities far lower than the universal silicon n-MOSFET mobility. Their performance decreases rapidly with decreasing temperature and at 250 K the drain current is equal to the gate leakage current; hence no detailed low temperature characterisation was performed. In contrast to p-MOSFETs, Ge n-MOSFETs are well documented to perform badly [Shang et al., 2006] and hence current thinking is that if Ge p-MOSFETs are ever to be implemented, it will be alongside strained silicon or GaAs n-MOSFETs [Takagi et al., 2006].

Despite showing high mobilities, leakage currents of the bulk Ge p-MOSFETs and strained Ge p-MOSFETs are unacceptably high. A more detailed investigation of the latter demonstrated that it is most likely caused by a high density of threading dislocations. The substrates on which these devices have been fabricated are far from optimised and hence this is not necessarily a fundamental roadblock to Ge CMOS; it does however, highlight the need for good quality substrates. One potential method of obtaining these is the Ge-condensation technique and is the subject of the final results chapter.

The Ge-condensation technique was investigated using thicker SOI substrates (100 nm top Si-layer) than normally used ($\approx$30 nm) in this process. This resulted in very long oxidation times (1700 minutes in total) to achieve a final SGOI layer 10 nm thick and with an average Ge-composition of 58%. Furthermore, due to strain contrast in the XTEM image and broadening in the SIMS profile, this layer will most likely be thinner and of a higher Ge-composition than reported. It is proposed that a series of SIMS profiles is obtained at successively higher primary beam energies to allow an

extrapolation to "0-energy", which should give the layer thickness without broadening.

The thin, high Ge-composition layer obtained demonstrates that it is possible to obtain SGOI, or even GOI, substrates from relatively thick SOI starting substrates. These are easier to manufacture than thinner SOI substrates (100 nm were the thinnest commercially available from Soitec when the experimental work was carried out). However, a new phenomenon detrimental to the Ge-condensation process was observed: germanium diffused through the BOX during the oxidations. It has long been thought that silicon dioxide acts as a diffusion barrier to germanium and hence the presence of Ge in the BOX here, is somewhat surprising. An explanation was proposed, which is that usually the ITOX reaction at the SGOI/BOX interface that prevents Ge from diffusing into the BOX. The thicker SOI used here suppressed this reaction and hence allowed Ge to enter the BOX, except in the final stage of the process, where the SGOI layer was thin enough and contained sufficient germanium to increase the ITOX rate sufficiently enough that it stop germanium from entering the BOX.

General suggestions for further work are now proposed. The s-Ge p-MOSFETs investigated here had a thicker than specified silicon cap, which lead to buried channel behaviour. This resulted in artificially low mobilities at high vertical fields due to carrier population of this cap, but there is evidence to suggest that the peak mobility, which occurs at low vertical fields, was slightly enhanced by the presence of this Si-cap [Nicholas et al., 2007a]. Furthermore, due to the wafers being an inappropriate size, only very large s-Ge MOSFETs were fabricated. Therefore, reprocessing the batch on wafers with a thin silicon cap and of a size compatible with the main non-silicon processing facility at IMEC should be a priority. This would allow a more accurate assessment of the effect of strain in both long and short p-MOSFETs. Of particular relevance here, is that in shorter devices surface roughness has been identified as a critical parameter that

must be minimised [Du et al., 2005] and hence the hydrogen anneal could be even more important for these. Since the experimental work was carried out for this thesis, 125 nm Ge p-MOSFETs with halo implants have been fabricated at IMEC [Nicholas et al., 2007b] and hence applying this process to s-Ge substrates should be trivial.

s-Ge n-MOSFETs could be improved by changing the substrate orientation to (111), which would result in the degeneracy splitting of the conduction band. This would be a formidable challenge however, as (111) virtual substrates would have to be grown and the fabrication process adapted to these. Given the poor performance of n-MOSFETs here and in the wider literature, it is unlikely that such a device batch would be commissioned.

Finally, the ITOX process in Ge-condensation needs further investigation and has been completely ignored in the literature regarding the subject. It is proposed that a series of SOI wafers of different thicknesses are used in a Ge-condensation process. The ITOX rate of each should be calculated with the model used here and, in general, will be greater for samples that begin with a thinner SOI substrate. This experiment would perform several functions, not least of which is to confirm that the Ge-diffusion, that we observed here, was a real effect and not due to a peculiarity associated with this sample set, for example a damaged buried oxide. It would also test the proposed hypothesis — that ITOX is responsible for preventing Ge from entering the BOX — and allow the critical ITOX rate for this to occur to be ascertained. Finally, the SGOI/BOX interface of each sample could be studied using XTEM to see if ITOX affected it. The quality of this interface will be of significance if Ge MOSFETs are fabricated on thin GOI wafers obtained with the Ge-condensation technique and it is therefore critical that the dynamics of this interface during Ge-condensation are understood.

Bibliography

V. V. Afanas'ev, Y. G. Fedorenko, and A. Stesmans. "Electron states at the (1 0 0)Ge/HfO_2 Interface". *material science in semiconductor processing*, 7:191–196, 2004.

T. Ando, A. B. Fowler, and F. Stern. "Electronic properties of two-dimensional systems". *reviews of modern physics*, 54(2):438–621, 1982.

S. Balakumar, S. Peng, K. M. Hoe, A. Agarwal, G. Q. Lo, R. Kumar, N. Balasubramanian, D. L. Kwong, and S. Tripathy. "SiGeO layer formation mechanism at the SiGe/oxide interfaces during Ge condensation". *applied physics letters*, 90(3):032111, 2007.

S. W. Bedell, A. Reznicek, K. Fogel, J. Ott, and D. K. Sadana. "Strain and lattice engineering for Ge FET devices". *materials science in semiconductor processing*, 9: 423–436, 2006.

S.W. Bedell, K. Fogel, D. K. Sadana, and H. Chen. "Defects and strain relaxation in silicon-germanium-on-insulator formed by high-temperature oxidation". *applied physics letters*, 85(24):5869–5871, 2004.

M. Bruel. "Application of hydrogen ion beams to Silicon On Insulator material technology". *nuclear instruments and methods in physics research B*, 108:313–319, 1995.

P. Butcher, N. H. March, and M. P. Tosi. *Crystalline semiconducting materials and devices*. Plenum Press, New York, USA, 1986.

T. Y. Chan, J. Chen, P. K. Ko, and C. Hu. "The impact of Gate-Induced Drain Leakage on MOSFET scaling". In *IEDM proceedings*, pages 718–721, 1987.

M. I. Current, S. W. Bedell, I. J. Malik, L. M. Feng, and F. J. Henley. "Ultrashallow junctions or ultrathin SOI?". *solid state technology*, 43(9):66–77, 2000.

B. E. Deal and A. S. Grove. "General relationship for the thermal oxidation of silicon". *journal of applied physics*, 36(12):3770–3778, 1965.

K. Derbyshire. "Reconsidering germanium". *solid state technology*, 50(2):33–36, 2007.

Z. Di, A. Huang, P. K. Chu, M. Zhang, W. Liu, Z. Song, S. Luo, and C. Lin. "Strain relaxation mechanism in SiGe-on-insulator fabricated by Ge condensation". *journal of crystal growth*, 281(2–4):275–280, 2005a.

Z. Di, M. Zhang, W. Liu, S. Luo, Z. Song, C. Lin, Q. Lin, and P. K. Chu. "Interfacial characteristics of fully depleted SiGe-on-insulator (SGOI) substrate fabricated by modified Ge condensation". *semiconductor science and technology*, 20:L31–L35, 2005b.

A. Dobbie, B. De Jaeger, M. Meuris, T. E. Whall, E. H. C. Parker, and D. R. Leadley. "Channel backscattering characteristics of high performance germanium pMOSFETs". In *ULIS*, 2008.

V. N. Dobrovolsky and V. G. Litovchenko. *Surface electronic transport phenomena in semiconductors*. Oxford University Press, New York, USA, 1990.

M. G. Dowsett, R. Morris, P. Choub, S. F. Corcoranc, H. Kheyrandishd, G. A. Cooked, J. L. Maule, and S. B. Patele. "Charge compensation using optical conductivity enhancement and simple analytical protocols for SIMS of resistive $Si_{1-x}Ge_x$ alloy layers". *applied surface science*, 203–204:500–503, 2003.

G. Du, X. Y. Liu, Z. L. Xia, Y. K. Wang, D. Q. Hou, J. F. Kang, and R. Q. Han. "Evaluations of scaling properties of Ge on insulator MOSFETs in nano-scale". *japanese journal of applied physics*, 44(4B):2195–2197, 2005.

C. J. Emeleus, T. E. Whall, D. W. Smith, R. A. Kubiak, E. H. C. Parker, and M. J. Kearney. "Scattering mechanisms affecting hole transport in remote-doped Si/SiGe heterostructures". *journal of applied physics*, 73(8):3852–3856, 1992.

M. V. Fischetti and S. E. Laux. "Band structure, deformation potentials and carrier mobility in strained Si, Ge and SiGe alloys". *journal of applied physics*, 80(4):2234–2252, 1996.

M. V. Fischetti, D. A. Neumayer, and E. A. Cartier. "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering". *journal of applied physics*, 90(9):4587–4608, 2001.

T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, 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". In *IEDM technology digest*, page 978, 2003.

R. Ghez and Y. J. van der Meulen. "Kinetics of thermal growth of ultra-thin layers of SiO_2 on silicon". *journal of the electrochemical society*, 119(8):1100–1106, 1972.

A. Gold and T. Dolgopolov. "Temperature dependence of the conductivity for the two-dimensional electron gas: analytical results for low temperatures". *physical review B*, 33(2):1076–1084, 1986.

C. Haas. "The diffusion of oxygen in silicon and germanium". *journal of the physics and chemistry of solids*, 15:108–111, 1960.

A. I. Horrell. *A theoretical study of the hole mobility in silicon-germanium heterostructures*. PhD thesis, University of Loughborough, 2001.

T. Irisawa, S. Tokumitsu, T. Hattori, K. Nakagawa, S. Koh, and Y. Shiraki. "Ultra-high room-temperature hole Hall and effective mobility in $Si_{0.3}Ge_{0.7}/Ge/Si_{0.3}Ge_{0.7}$ heterostructures". *applied physics letters*, 81(5):847–849, 2002.

S. Jallepalle, W. K. Shih, J. D. Bude, M. R. Pinto, C. M. Maziar, and A. F. Tasch Jr. "Understanding the differences in the effective-field dependence of electron and hole inversion layer mobilities". In *IEDM proceedings*, pages 391–394, 1996.

M. Jaros. *Physics and applications of semiconductor microstructures*. Oxford University Press, New York, USA, 1989.

J. S. Kang, D. K. Schroder, and A. R. Alvarez. "Effective and field-effect mobilities in Si MOSFETs". *solid-state electronics*, 32(8):679–681, 1989.

E. Kasper. *Properties of strained and relaxed silicon germanium*. INSPEC, London, UK, 1995.

E. Koppensteiner, G. Bauer, H. Kibbel, and E. Kasper. "Investigation of strain-symmetrized and pseudomorphic Si_mGe_n superlattices by x-ray reciprocal space mapping". *journal of applied physics*, 76(6):3489–3501, 1994.

W.T. Lai and P. W. Li. "Growth kinetics and related physical/electrical properties of Ge quantum dots formed by thermal oxidation of SiGe-on-insulator". *nanotechnology*, 18:145402, 2007.

F. K. LeGous, R. Rosenberg, and B. S. Meyerson. "Kinetics and mechanism of oxidation of SiGe: dry versus wet oxidation". *applied physics letters*, 54:644–646, 1988.

M. Lundstrom. *Fundamentals of carrier transport*. Cambridge University Press, Cambridge, UK, 2nd edition, 2000.

S. Maikap, M. H. Lee, S. T. Chang, and C. W. Liu. "Characteristics of strained-germanium p- and n-channel field effect transistors on a Si (111) substrate". *semiconductor science and technology*, 22:342–347, 2007.

Y. Matsumoto and Y. Uemura. "Scattering mechanism and low temperature mobility of MOS inversion layers". *japanese journal of applied physics*, suppliment 2(part 2): 367–370, 1974.

A. Matsumura, I. Hamaguchi, K. Kawamura, T. Sasaki, S. Takayama, and Y. Nagatake. "Technological innovation in low-dose SIMOX wafers fabricated by an internal thermal oxidation (ITOX) process". *microelectronic engineering*, 66:400–414, 2003.

J. W. Matthews and A. E. Blakeslee. "Defects in epitaxial multilayers: I. misfit dislocations". *journal of crystal growth*, 47:118–125, 1974.

D. L. McVay and DuCharme. "Diffusion of Ge in SiGe alloys". *physical review B*, 9(2): 627–631, 1974.

M. V. Minke and K. A. Jackson. "Diffusion of germanium in silica glass". *journal of non-crystalline solids*, 351:2310–2316, 2004.

T. Mizuno, N. Sugiyama, T. Tezuka, and S. Takagi. "Relaxed SiGe-on-insulator substrates without thick SiGe buffer layers". *applied physics letters*, 80(4):601–603, 2002.

G. E. Moore. "Cramming more components onto integrated circuits". *electronics*, 38 (8), 1965.

M. Mukherjee-Roy, A. Agarwal, S. Balakumar, A. Y. Du, A. D. Trigg, R. Kumar, N. Balasubramanian, and D. L. Kwong. "A two step oxidation mediated condensation process for ultrathin high Ge content SiGe epitaxial films on insulator". *electrochemical and solid-state letters*, 8(7):G164–G166, 2005.

S. Nakaharai, T. Tezuka, N. Sugiyama, Y. Moriyama, and S. Takagi. "Characterization of 7-nm-thick strained Ge-on-insulator layer fabricated by Ge-condensation technique". *applied physics letters*, 83(17):3516–3518, 2003.

S. Nakashima, T. Katayama, Y. Miyamura, A. Matsuzaki, M. Kataoka, D. Ebi, M. Imai, K. Izumi, and N. Ohwada. "Investigations on the high-temperature thermal oxidation process at top and bottom interfaces of top silicon of SIMOX wafers". *journal of the electrochemical society*, 143(1):244–251, 1996.

L. Nash. *Growth and characterisation of terrace graded virtual substrates with* $Si_{1-x}Ge_x$ $0.15 \leq x \leq 1$. PhD thesis, University of Warwick, 2005.

A. Nayfeh, C. O. Chui, K. C. Saraswat, and T. Yonehara. "Effects of hydrogen annealing on heteroepitaxial-Ge layers on Si: surface roughness and electrical quality". *applied physics letters*, 85(14):2815–2817, 2004.

A. Neyfeh, C. O. Chui, T. Yonehara, and K. C. Saraswat. "Fabrication of high-quality p-MOSFET in Ge grown heteroepitaxially on Si". *IEEE electron device letters*, 26(5): 311–313, 2005.

G. Nicholas. *Investigation into the electrical properties of tensile strained silicon MOSFETs*. PhD thesis, University of Warwick, 2004.

G. Nicholas, T. J. Grasby, D. J. E. Fulgoni, C. S. Beer, and J. Parsons. "High mobility strained Ge pMOSFETs with high-kappa/metal gate". *IEEE electron device letters*, 28(9):825–827, 2007a.

G. Nicholas, B. Jaeger, D. P. Brunco, P. Zimmerman, G. Eneman, K. Martens, M. Meuris, and M. M. Heyns. "High-performance deep submicron Ge pMOSFETs with halo implants". *IEEE transactions on electron devices*, 54(9):2503–2511, 2007b.

E. H. Nicollian and J. R. Brews. *MOS (metal oxide semiconductor) physics and technology*. John Wiley and Sons, New York, USA, 1982.

A. Nobutoshi, T. Hiroshi, G. Naoyuki, M. Takashi, I. Toyotsugu, O. Tetsuya, A. Kouichirou, K. Hiroshi, G. Yasuhito, and I. Junzo. "Thermal diffusion behavior of implanted germanium atoms in silicon dioxide film measured by high-resolution RBS". *Surface and Coatings Technology*, 201:8312–8316, 2007.

S. H. Olsen, A. G. O'Neill, P. Dobrosz, S. J. Bull, L. S. Driscoll, S. Chattopadhyay, and K. S. K. Kwa. "Study of strain relaxation in Si/SiGe metal-oxide-semiconductor field-effect transistors". *journal of applied physics*, 97(11):114504, 2005.

D. C. Paine, C. Caragianis, and A. F. Schwartzman. "Oxidation of $Si_{1-x}Ge_x$ alloys at atmospheric and elevated pressure". *journal of applied physics*, 70(9):5076–5084, 1991.

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. Park, J. Bea, T. Fukushima, and M. Koyanagi. "Analysis of GOI-MOSFET with high-k gate dielectric and metal gate fabricated by Ge condensation technique". *surface and interface analysis*, 38:1720–1724, 2006.

D. J. Paul. "Si/SiGe heterostructures: from material and physics to devices and circuits". *reports on progress in modern physics*, 19(10):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". *applied physics letters*, 47(3):322–324, 1985.

K. Rais, F. Balestra, and G. Ghibaudo. "Temperature dependence of gate induced drain leakage current in silicon CMOS devices". *electronics letters*, 30(1):32–34, 1994.

E. M. Rehder, C. K. Inoki, T. S. Kuan, and T. F. Kuech. "SiGe relaxation on silicon-on-insulator substrates: an experimental and modelling study". *journal of applied physics*, 94:7892–7903, 2003.

J. Robertson. "High dielectric constant gate oxides for metal oxide Si". *reports on progress in modern physics*, 69:327–396, 2006.

B. Rossner, D. Chrastina, G. Isella, and H. von Kanel. "Scattering mechanisms in high-mobility strained Ge channels". *journal of applied physics*, 84(16):3058–3060, 2004.

M. Sadaka, M. Zavala, M. Rossow, A. Thean, A. Barr, T. White, V. Vartanian, and B. Y. Nguyen. "Enhancement of SiGe relaxation for fabrication of SGOI substrates using condensation". *electrochemical society proceedings*, 2004-7:281–288, 2004.

S. B. Samavedam, W. J. Taylor, J. M. Grant, J. A. Smith, P. J. Tobin, A. Dip, A. M. Phillips, and R. Liu. "Relaxation of strained Si layers grown on SiGe buffers". *journal of vacuum science and technology B*, 17(4):1424–1428, 1999.

K. Sawano, Y. Abe, H. Satoh, Y. Shiraki, and K. Nakagawa. "Compressive strain dependence of hole mobility in strained Ge channels". *applied physics letters*, 87: 192102–1–192102–3, 2005.

F. Schaffler. "High-mobility Si and Ge structures". *semiconductor science and technology*, 12:1515–1549, 1997.

D. K. Schroder. *Semiconductor material and device chacterization*. John Wiley and Sons, New Jersey, USA, 3rd edition, 2006.

K. Seeger. *Semiconductor physics*. Springer, Berlin, Germany, 2004.

H. Shang, H. Okorn-Schimdt, J. Ott, P. Kozlowski, S. Steen, E. C. Jones, H. S. P. Wong, and W. Hanesch. "Electrical characterization of germanium p-channel MOSFETs". *IEEE electron device letters*, 24(4):242–244, 2003.

H. Shang, M. M. Frank, E. P. Gusev, J. O. Chu, S. W. Bedell, K. W. Guarini, and M. Leong. "Germanium channel MOSFETs: opportunities and challenges". *IBM journal of research and development*, 50(4/5):377–386, 2006.

P. Singer. "High-k, metal gates a 'Go' for 45 nm". *Solid State Technology*, 50(3):28, 2007.

C. G. Sodini, T. W. Ekstedt, and J. L. Moll. "Charge accumulation and mobility in thin dielectric transistors". *solid-state electronics*, 25(9):833–841, 1982.

F. Stern and W. E. Howard. "Properties of semiconductor surface inversion layers in the electric quantum limit". *physical review*, 163(3):816–835, 1967.

T. Stoica and E. Sutter. "Ge dots embedded in SiO_2 obtained by oxidation of Si/Ge/Si nanostructures". *nanotechnology*, 17:4912–4916, 2006.

N. Sugiyama, T. Tezuka, T. Mizuno, M. Suzuki, Y. Ishikawaand, N. Shibata, and S.Takagi. "Temperature effects on Ge condensation by thermal oxidation of SiGe-on-insulator structures". *journal of applied physics*, 95:4007–4011, 2004.

N. Sugiyama, S. Nakaharai, N. Hirashita, T. Tezuka, K. Usuda, S. Takagi, and Y. Moriyama. "The formation of SGOI structures with low dislocation density by a two-step oxidation and condensation method". *semiconductor science and technology*, 22:S69–S62, 2007.

S. M. Sze. *Semiconductor devices physics and technology*. John Wiley and Sons, New York, USA, 1985.

S. Takagi, A. Toriumi, M. Iwase, and H. Tango. "On the universality of inversion layer mobility in Si MOSFET's: part 1—effects of substrate impurity concentration". *IEEE transactions on electron devices*, 41(12):2357–2362, 1994a.

S. Takagi, A. Toriumi, M. Iwase, and H. Tango. "On the universality of inversion layer mobility in Si MOSFET's: part 2—effects of surface orientation". *IEEE transactions on electron devices*, 41(12):2363–2368, 1994b.

S. Takagi, T. Tezuka, T. Irisawa, S. Nakaharai, T. Maeda, T. Numata, K. Ikeda, and N. Sugiyama. "Hole mobility enhancement of p-MOSFETs using global and local Ge-channel technologies". *materials science and engineering B*, 135:250–255, 2006.

Y. Takahashi, T. Ishiyama, and M. Tabe. "Counter-oxidation of superficial Si in single-crystalline Si on SiO_2 structure". *applied physics letters*, 65(23):2987–2989, 1994.

G. Taraschi, A. J. Pitera, and E. A. Fitzgerald. "Strained Si, SiGe and Ge on-insulator: review of wafer bonding fabrication techniques". *solid-state electronics*, 48:1297–1305, 2004.

Y. Taur and T. H. Ning. *Fundamentals of modern VLSI devices*. Cambridge University Press, Cambridge, UK, 1998.

C. D. Thurmond. "The standard thermodynamic functions for the formation of electrons and holes in Ge, Si, GaAs and GaP". *journal of the electrochemical society: solid-state science and technology*, 122(8):1133–1141, 1975.

A. Touhami and A. Bouhdada. "Impact of interface traps on gate-induced drain leakage current in n-type metal oxide semiconductor field effect transistor". *international journal of electronics*, 92(9):539–552, 2005.

M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Mattausch, and M. Hirose. "Physically-based threshold voltage determination for MOSFETs of all gate lengths". *IEEE transactions on electron devices*, 46(7):1429–1434, 1999.

N. Usami, K. Kutsukake, K. Fujiwara, T. Ujihara, G. Sazaki, S. Ito, B. P. Zhang, and K. Nakajima. "Impact of the annealing temperature on the homogeneity of SiGe-on-insulator". In *ICSI-3*, pages 200–202, 2003.

J. Vanhellemont and E. Simoen. "Brother silicon, sister germanium". *journal of the electrochemical society*, 154(7):H572–H583, 2007.

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". In *IEDM technology digest*, page 453, 2003.

P. M. Zeitzoff, C. D. Young, G. A. Brown, and Y. Kim. "Correcting effective mobility measurements for the presence of significant gate leakage current". *IEEE electron device letters*, 24(4):275–277, 2003.

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". *IEEE electron device letters*, 27(9):728–730, 2006.

W. Zhu, J. Han, and T. P. Ma. "Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics". *IEEE transactions on electron devices*, 51(1):98–105, 2004.

P. Zimmerman, G. Nicholas, B. Jaeger, B. Kaczar, A. Stesmans, L. A. 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". In *IEDM*, 2006.