MSc. Thesis:

Electrical Characterisation of Strained Silicon devices

Andrew McGinn, August 2006

Andrew McGinn
Nano-Silicon Research Group
Dept of Physics
University of Warwick
Coventry, CV4 7AL
UK

Tel: +44 (0)2476 522448
E-mail: a.j.p.mcginn@warwick.ac.uk

Abstract

The rate of development in the standard of living in the western world is dependant on technology that is based on the planar transistor. Performance improvement of the transistor over the last six decades has followed Moore's law, but recently has slowed down due to fundamental limits with miniaturising the components involved. This has led to the industry looking into new novel techniques to maintain progress without drastically altering the current processing lines. One of the most promising short term novel techniques allows significant performance improvements by placing the channel region of a transistor under a small amount of strain.

This MSc looks at the use of two different methods for inducing strain on the channel and the benefits and drawbacks of implementing them.

The first method involves the use of a silicon germanium alloy as the foundation or virtual substrate upon which a layer of silicon is grown. This top layer of silicon forms the channel of the device. By incorporating Ge into the substrate the atomic spacing becomes larger and the silicon that is grown upon it will adopt this larger spacing. The channel will be biaxially tensile strained and has been shown to improve both the performance of nMOS and pMOS devices.

The second method is simpler and involves depositing a strained nitride layer on top of a device. As the layer attempts to revert back to a structure that is most energetically favourable it affects the device underneath. The nitride layer induces uniaxial strain and it has been shown that for performance increases tensile stress is required for nMOS devices and compressive stress for pMOS devices.

The main criteria used to evaluate modern day MOSFETs are the mobility, current drive, leakage current, threshold voltage shift, source/drain resistance and subthreshold slope. All of these have been looked at to determine if the use of a strained channel induced by a virtual substrate or a combined virtual substrate and nitride capping layer, can bring an overall performance enhancement compared to standard silicon MOSFETs. The main aim of this work was to maximise the increase in mobility and drive current without significant degradation of the other criteria. This is because these are considered second order effects which theoretically will be negatively impacted, but can be minimised through additional process steps.

Mobility enhancements of 1.2 to 2 times were observed for the nMOS biaxially strained devices tested and enhancements of 1.15 times for pMOS. Depositing a nitride layer over the gate stack and surrounding features also had a positive impact on mobility of 1.05 times for nMOS devices and 1.5 times for pMOS. While such enhancements appear modest they have allowed the '2005' batch of devices to meet targets set by the ITRS in 2005, for devices fabricated in that year. Other device parameters have been adversely affected and the level of degradation coincides with theoretical and experimental published work [Goo 2003] [Lim 2004] [Thompson 2004b] [Xiang 2003] [Sugii 2002]. As these issues can be resolved and the enhancements kept, strained silicon has a very bright future in the next 10 years or so with Intel already manufacturing commercial chips based around this novel technology.

Table of Contents

Glossary of Symbols

Å	Angstroms
C_{dep}	depletion capacitance
C_{gc}	gate to channel capacitance
C_{it}	interface capacitance
C_{ox}	oxide capacitance
d	spacer width
e	electron charge
E	energy
E_c	conduction band
E_{eff}	effective electric field
E_f	final energy level
E_F	Fermi level
E_{Fi}	intrinsic Fermi level
E_g	band gap
E_i	initial energy level
E_{lat}	lateral electric field
E_v	valence band
$g(\sigma)$	function of strain
g_m	transconductance
h	gate height
$\hbar$	Planck's constant
$\hbar\omega$	phonon energy
I_D	drain current
I_G	gate current
$I_{S/D}$	source/drain current
I_{SUB}	substrate current
k	momentum
k_B	Boltzmann constant
K_s	semiconductor dielectric constant
L	device length
L_{eff}	effective length
m	body effect coefficient
m*	effective mass
n	electron density
N_a	acceptor density
N_s	sheet density
N_v	density of states in the valence band
n^+	negative heavily doped region
Q_n	mobile channel charge density
R_{ch}	channel resistance
R_m	device total resistance
$R_{S/D}$	source/drain resistance
S	subthreshold slope
S_{int}	intrinsic strain of the nitride layer
T	temperature

T_c	critical thickness
t_N	nitride layer thickness
T_{ox}	oxide layer thickness
μ	mobility
μ_{eff}	effective mobility
v	carrier velocity
V_D	drain voltage
V_{fb}	flat band voltage
V_G	gate voltage
V_{po}	potential drop across the inversion region
V_{SUB}	substrate voltage
V_T	threshold voltage
v_{th}	carrier thermal velocity
W	device width
ε_0	permittivity of vacuum
ε_d	oxide dielectric constant
σ	strain
τ	relaxation time
Δ	rms height
λ	correlation length
ϕ_F	work function relating the Fermi level and intrinsic Fermi level of a semiconductor
ϕ_s	work function of a semiconductor

List of Figures

List of Tables

Acknowledgements

I would like to take this opportunity to thank my supervisor Dr. D. R. Leadley for his support and guidance throughout the year. Without his help this would still be far from complete. I would also like to thank Prof. E.H.C. Parker and Prof. T.E. Whall for the additional support they gave me.

I would like the thank all the members of the Nano-Silicon Research Group, Chris Beer, Lee Nash, Jonathan Parsons, Dominic Pearman, Vishal Shah, Steve Thomas, Tim Naylor and especially Gareth Nicholas and Andy Dobbie for not only their help but for making this year a pleasant and enjoyable one.

I would like to thank Fabrice Payet at STMicroelectronics for providing me with the latest batch of devices that were investigated for this thesis, and for his help on them.

Finally I would like to thank my Father, as well as the rest of my family and friends.

Declaration

This thesis is submitted to the University of Warwick in support of my application for a Master of Science degree. Except where specifically stated all of the work described in this thesis was carried out by the author or under his direction.

Andy M^{c}Ginn

Andrew James Phillip M^{c}Ginn (*31.08.06*)

1 Introduction

1.1 Semiconductor industry

The transistor was developed in the late 1940’s by scientists at Bell Telephone Laboratories in New Jersey, spawning the semiconductor industry. This industry has developed extremely rapidly over the past six decades, enabling western civilisation to reach its current high technological position. Rapid development of semiconductor capabilities are now expected, as standard, by the rest of the World’s technologies. In an attempt to keep up with this demand for progress a road map is laid out by the ITRS every few years illustrating where the industry should head over the next 15 years and highlighting where progress might end if certain technological breakthroughs are not achieved [ITRS 2001/05]. The industry strives to achieve the set goals and solve the identified technology challenges.

Intel co-founder Gordon Moore observed that the markets demand for functionality per chip doubles roughly every one and a half to two years (*Figure 1-1*) and it has been this rate of change, known as Moore’s Law, that the semiconductor companies are trying to maintain. Moore’s Law is related to the integration level of a chip (the number of components per chip).

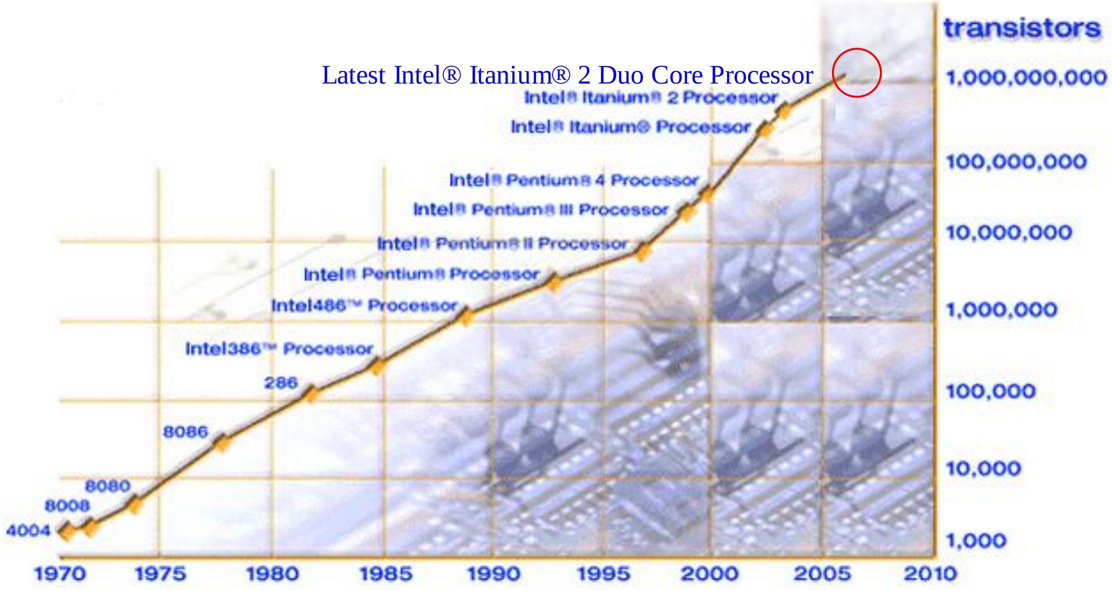

Figure 1-1: Intel® processors and how the number of transistors in each has changed over time [Intel 2005]

While Moore’s law has been the main scaling trend there are others that have become more dominant in recent years as the ability to shrink components at the same time as increasing performance has slowed. Other scaling trends are related to cost, speed, power consumption, compactness, and functionality.

Currently Complementary Metal Oxide Semiconductor (CMOS) devices are considered to be the industry workhorses and will continue to be until around 2020 where it is believed that post CMOS devices such as those in the field of spintronics are going to take over.

In the past it is has been the physical scaling of CMOS devices that have allowed them to keep up with Moore’s law as this increases the packing density achievable on a chip while improving speeds (frequency response α 1/Channel length) and the current drive. However, in recent years, CMOS devices scaling has had to be constrained in order to keep negative effects such as high leakage currents within reasonable bounds. The problem is that without gate oxide scaling, reducing the channel length has little impact on device performance. This has lead to the pursuit of other technologies (technology boosters) that can be used as part of CMOS technology to improve the current Si technology platform. The main four are:

(1) Strained channel regions
(2) High (k) dielectric constant materials
(3) Dual orientation devices
(4) New channel material (e.g. Germanium)

While viewed as ‘short term’ fixes, these are expected to allow the industry to continue moving forwards until the post CMOS era starts, and are therefore extremely important. The use of strained channel regions and high-k dielectrics to replace the gate oxide have been most heavily investigated in recent years as they appear to be the most beneficial. Strain technology has already been implemented by Intel into the industry’s first high-volume 90nm production line (they use uni-axially strained silicon (S-Si) layers using selective SiGe source and drain regions for pMOS and Si_3N_4 cap layers for nMOS) and now the new Itanium® 2 Duo Core Processor with 1.72 billion transistors.

1.2 Strained silicon

The notion of placing the channel of a CMOS device under strain has been in the scientific community since IBM announced in June of 2001 that it had developed a new form of silicon. Strained Silicon (S-Si) is achieved by either increasing (tensile) or reducing (compressive) the atomic spacing between the Si atoms (*Figure 1-2*). The type of strain is described relative to the x and y planes of the device channel. If stress is applied to:

- One axis → uniaxial strained Si
- Two axis → biaxial strained Si

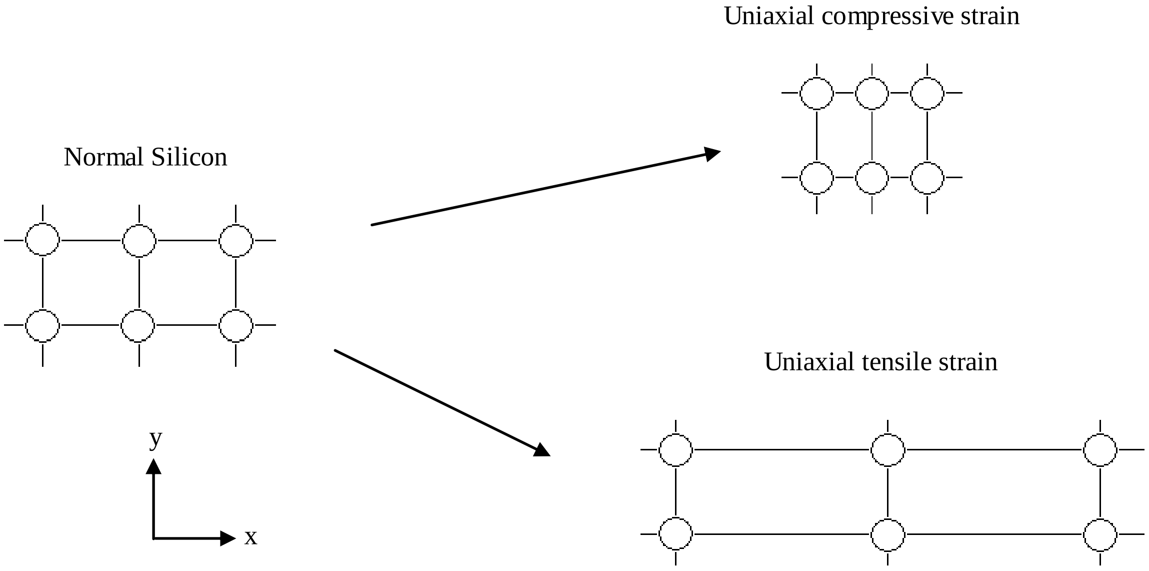

Figure 1-2: Top down view of a silicon structure under varying strain conditions

Altering the atomic spacing between the Si atoms causes a change in the energy band structure that leads to carrier mobility enhancements that are directly linked to performance.

1.3 Current investigation

There are three main ways of inducing strain in the channel region of a device:

1. Mechanical – standard devices placed under global physical strain (*Figure 1-3*)
2. Substrate – channel placed under strain by the underlying substrate (*Figure 1-4*)
3. Process – altered process steps resulting in channel strain (*Figure 1-5*)

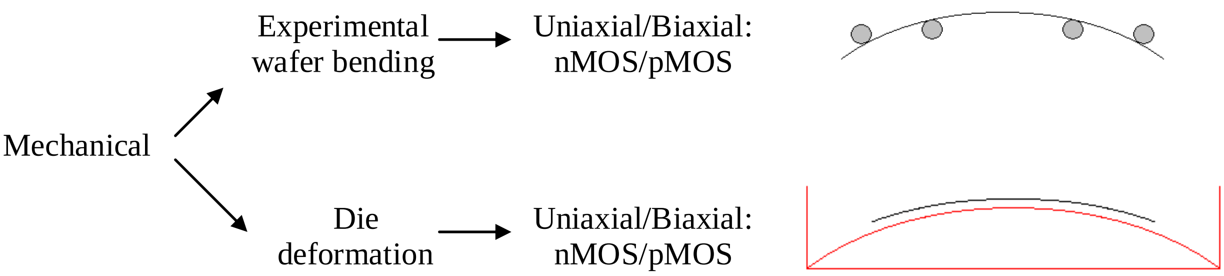
Experimental wafer bending
Uniaxial/Biaxial: nMOS/pMOS
Mechanical
Die deformation
Uniaxial/Biaxial: nMOS/pMOS

Figure 1-3: Types of mechanical induced channel strain

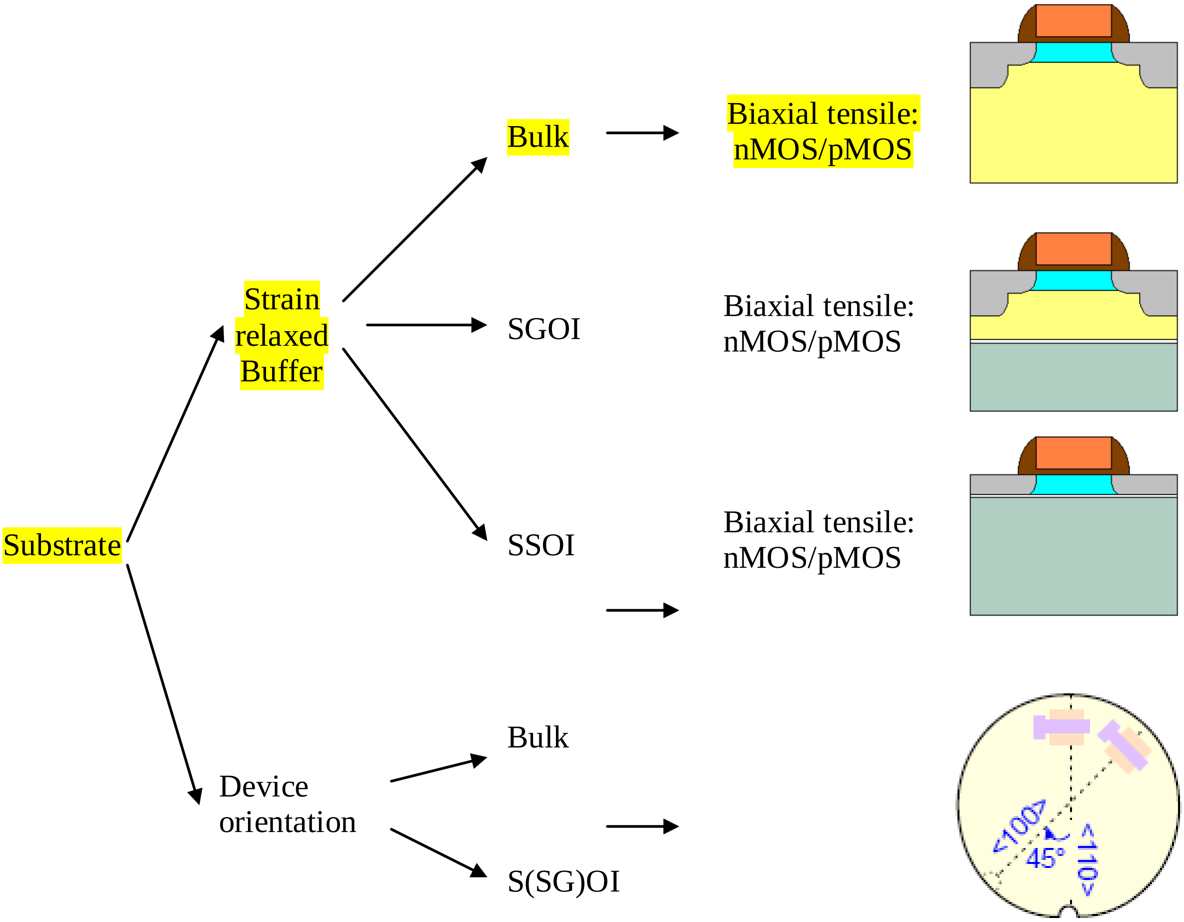

Figure 1-4: Types of substrate induced channel strain. Highlighted methods have been researched in this MSc.

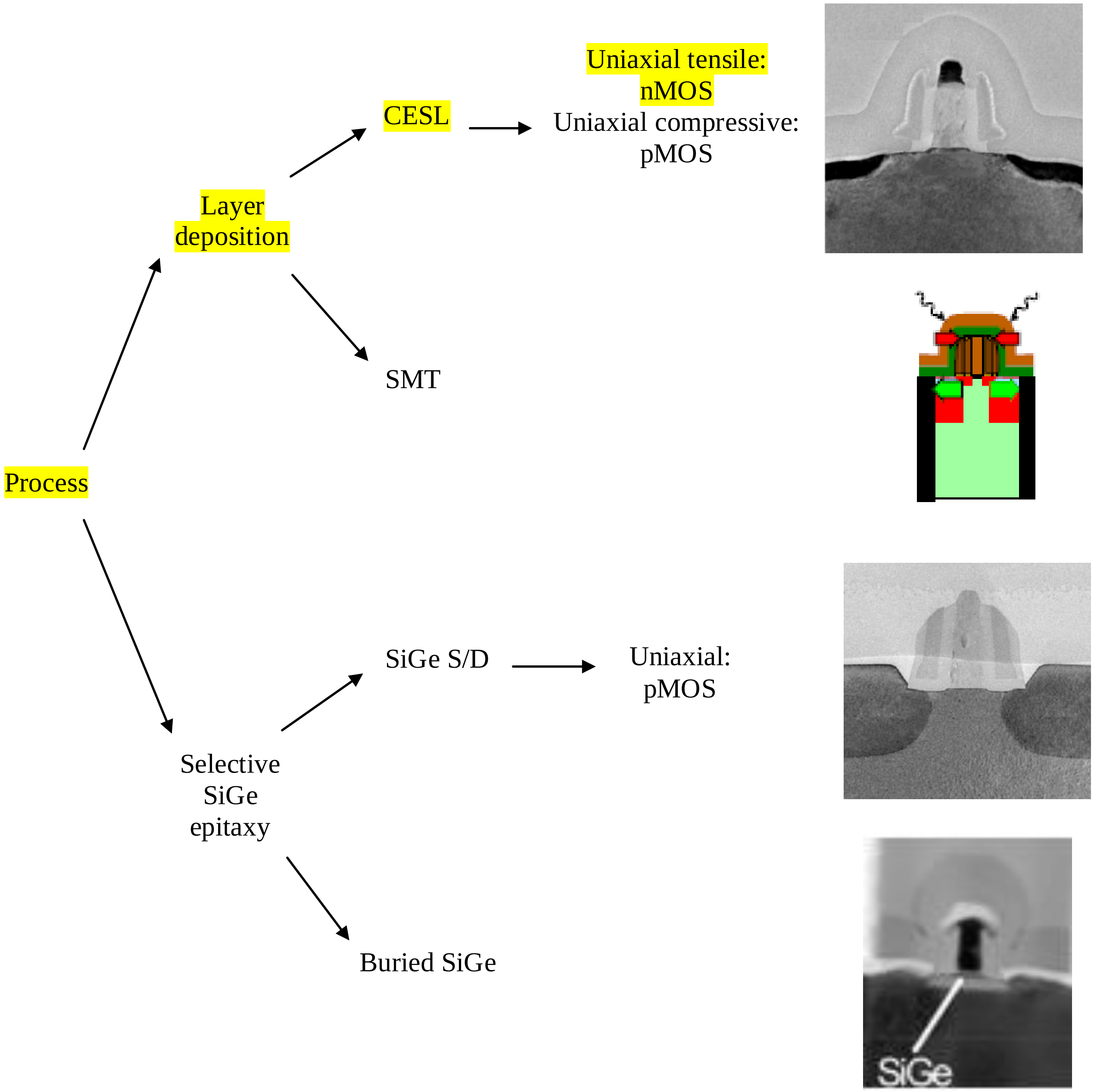

Figure 1-5: Types of process induced channel strain. Highlighted methods have been researched in this MSc.

This investigation looks at two methods for deploying strain technology, substrate induced (*Figure 1-3)* and process induced (*Figure 1-4*) along with a number of refinement steps such as Chemical Mechanical Polishing (CMP) and threshold voltage optimization, as a means of harnessing the benefits of the strain while minimising the drawbacks. Electrical characterisation has been performed under room temperature conditions on two wafer batches produced by ST Microelectronics (France).

While SiGe VS has been quite exhaustively tested the optimum processing refinements have yet to be found. To the best of my knowledge there are no papers or articles regarding the combining of VS and CESL strain technology.

The theoretical background for this field is presented in *Chapter 2* detailing basic MOSFET operation, and the two methods of applying stress to the channel region for performance gain. The experimental procedure and parameter extraction details are described in *Chapter 3*. *Chapter 4* reports on the fabrication of the device wafers and lists the specifications of the final devices that were tested. The data collected on the novel strain techniques are presented and analysed in detail in *Chapter 5* before conclusions are drawn in *Chapter 6*.

2 Theoretical Background

2.1 Introduction

This chapter serves as a summary of the important issues relating to strained silicon MOSFETs. A basic overview of MOSFETs and mobility is given before a more in depth review of the effect of stress applied to the channel region of a MOSFET.

Unless explicitly stated otherwise, all theory and diagrams discuss n-type MOSFETs where the minority carriers that allow for the flow of current are electrons.

2.2 MOSFET structure & operation

A MOSFET (Metal Oxide Semiconductor Field Effect Transistor) is an extension of the MOS Capacitor (for background reading see [Singh 1994]). The major difference is the addition of source and drain regions. A schematic of a simple MOSFET is shown in *Figure 2-1*.

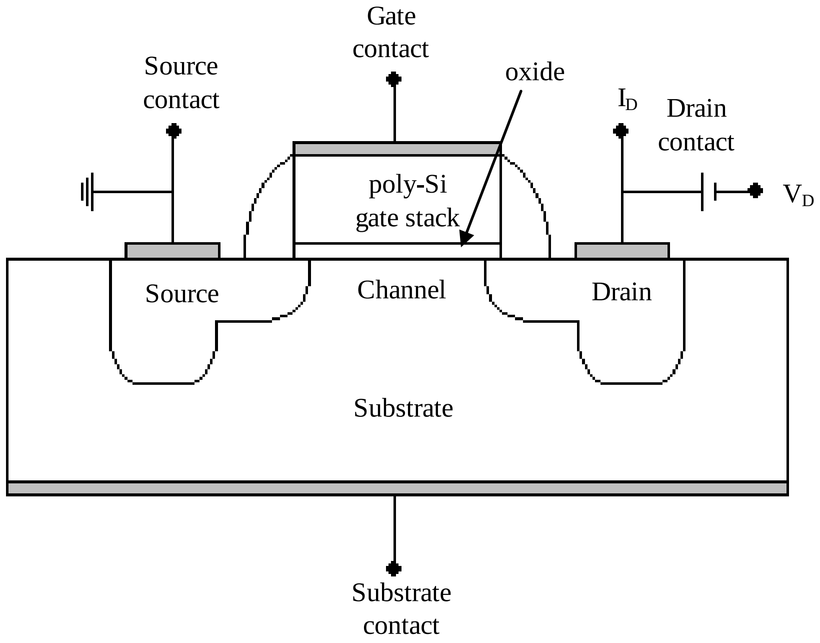

Figure 2-1: MOSFET schematic

By applying a bias to the gate contact the channel can be forced to enter a state of inversion. This is related to the surface band bending at the interface given by $e\Phi_s$. The condition for the onset of inversion is given by the criteria that the surface band bending equal twice that of the band bending $e\Phi_F$ in the semiconductor (*Figure 2-2*).

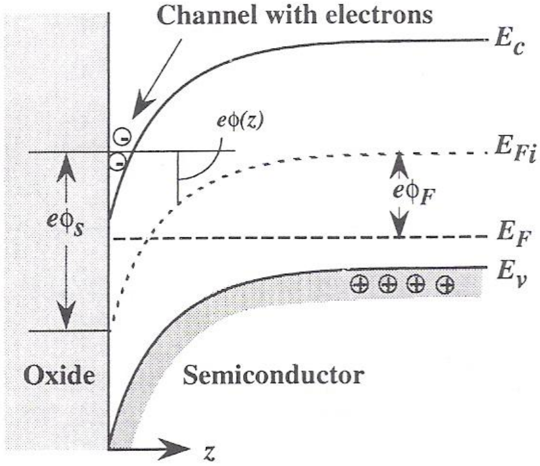

Figure 2-2: Band bending in the semiconductor at the onset of inversion [Singh 1994]

Here the concentration of minority carriers, which have come predominantly from the source and drain regions where they are plentiful, is equal to the bulk concentration of majority carriers and we have

$$e\phi_s = 2e\phi_F \qquad (2.1)$$

The gate voltage applied (V_G) at which this criteria is met, is labelled the threshold voltage (V_T) and we find it to be

$$V_T = V_{fb} + \phi_s + \frac{2\sqrt{\varepsilon_s e N_a \phi_F}}{C_{ox}} \qquad (2.2)$$

In deriving this equation we are assuming:

i. the full depletion approximation

ii. the inversion layer charge = 0 for $V_G \leq V_T$.

iii. the linear relationship $Q_i = C_{ox}(V_G - V_T)$ exists for $V_G > V_T$

As the inversion layer is grounded in a MOSFET due to the addition of extra contacts the condition of inversion is altered slightly to

$$e\phi_s = 2\phi_F - V_{sub} \tag{2.3}$$

Field

.

The charges in the channel region are now able to carry a current between the source and the drain. The amount of current is dependant on the number of free carriers in the channel and hence the gate is able to modulate the current flow.

A number of approximations are made in order to look at a simplified model for device operation. The ‘charge sheet approximation’ assumes that the inversion charges are at the semiconductor surface in a sheet and that no potential drop or band bending occurs across the inversion layer. The ‘Gradual channel approximation’ is also made where we assume that the horizontal field strength is much less than the vertical field ($E_{lat} << E_{eff}$) in the channel, which allows us to use Poisons 1D equation. As gate lengths reduce these approximations no longer apply as short channel effects (SCEs) occur which are not taken into account in these simple models.

Since the drain is positively biased with respect to the source, electrons flow from the source to the drain; however, conventional current flows in the other direction. Almost no current flows into the gate as it sits on an insulator.

Once the gate bias is such that a channel of free carriers is formed ($V_G > V_T$) we see two distinct regions of operation when drain current (I_D) is plotted as a function of the drain/source bias (V_D) (*Figure 2-3*).

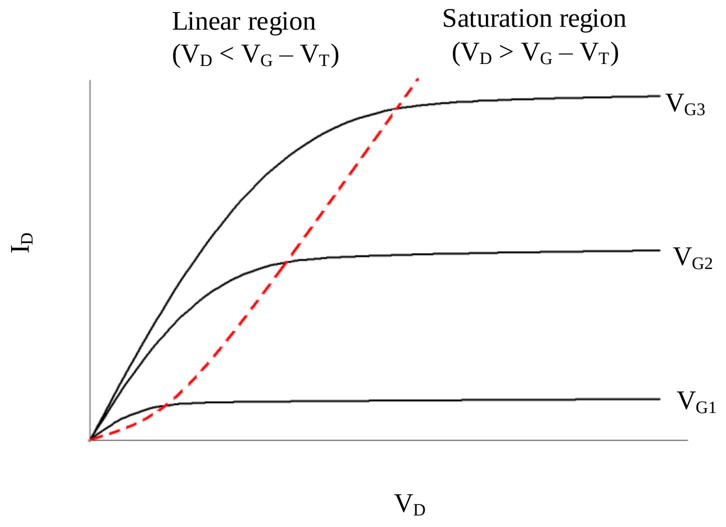

Figure 2-3: Drain current versus drain bias for typical MOSFET at different gate biases ($V_{G3}>V_{G2}>V_{G1}$)

When a small drain source voltage is applied the channel inversion region behaves like an ordinary resistive material giving an expected linear I-V characteristics (*Figure 2-4 (a)*). Given that the carrier velocity

$$v = \mu \frac{dV_c}{dx} \tag{2.4}$$

Field

We know the drain current at any point in the channel is

$$I_D = C_{ox}\mu \frac{dV_c}{dx}(V_G - V_T - V_C(x)) \tag{2.5}$$

Field

By integrating over the channel length from the source (x=0, $V_C(0)=0$) to the drain (x=L, $V_C(L)=V_D$) we obtain

$$I_D = \frac{1}{2}\mu C_{OX}\left(\frac{W}{L}\right)\left[2(V_G - V_T)V_D - V_D{}^2\right] \tag{2.6}$$

Field

And the second order term is ignored (as V_D is small) giving a linear relationship. From this we can obtain the drain conductance which illustrates the effect the drain bias has on the drain current:

$$g_D = \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G=\text{constant}} \qquad (2.7)$$

When the drain voltage is equal to the overdrive voltage ($V_D = V_G - V_T$), the voltage drop between the gate and the drain end of the channel will no longer be sufficient to maintain an inversion region along the entire channel length and we get the pinch off point (*Figure 2-4 (b)*). The source end of the channel remains unaffected as it is at zero potential.

As the drain voltage is further increased the pinch off point will move (*Figure 2-4 (c)*) creating a larger depletion region between the inversion region and the drain. This region is however usually very small and as such the electric field across it will be large allowing the carriers to be accelerated across it and into the drain. I_D is no longer dependant on V_D. Now

$$I_D = \frac{1}{2}\mu C_{OX}\left(\frac{W}{L}\right)(V_G - V_T)^2 \qquad (2.8)$$

Since the drain current is independent of V_D the drain conductance would be zero so instead we use the transconductance. This gives us an indication of the gates control over the drain current.

$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D=\text{constant}} \qquad (2.9)$$

The problem occurs with short channel devices as the current continues to increase in the saturation region. Starting with a basic equation for current we can obtain

$$I_D = \frac{enA\mu V_{po}}{L} \qquad (2.10)$$

V_{po} (the potential drop across the inversion region) remains constant as the pinch off point moves therefore having no effect on I_D however L decreases causing I_D to increase. For very long devices the ΔL due to the pinch off point moving will be negligible having no effect on the drain current, however for short devices (<1μm) this change in the channel length is important.

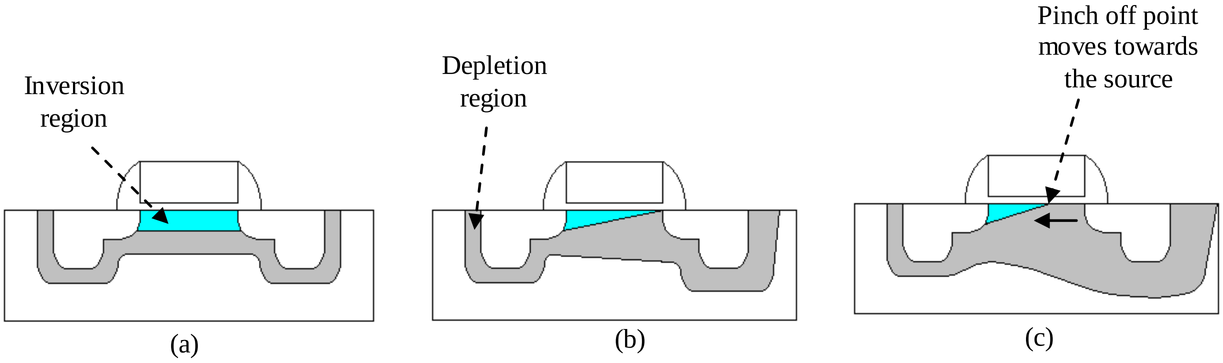

Figure 2-4: n-type MOSFET (a) linear region, (b) pinch off point, (c) saturation region [Parker 2004]

2.3 Scattering principles & Carrier mobility

If a charge carrier moved through a perfectly periodic crystal in the absence of external forces it would remain in a particular electronic energy state for an infinite period of time. External effects such as imperfections or impurities in the lattice ‘perturb’ the carrier. These perturbations cause the carrier to move from one energy state to another in a process known as ‘scattering’. The main causes for a carrier to scatter are:

1. ionized impurities
2. phonons
3. the lattice being an alloy
4. interface roughness

It is possible using Fermi’s Golden Rule to calculate the probability of such scattering events occurring at a quantum level. The scattering probability is dependant on the strength at which the perturbation couples the initial and the final states, and the number of ways it can happen (the density of final states) [Harrison 2002]. Scattering can alter momentum of a carrier as well as its energy. The average time a carrier can travel through the crystal lattice undisturbed before it is scattered, is known as the relaxation time (τ). This microscopic parameter of the system can be related to a macroscopic one that quantifies how easily a carrier can move through a material, which is called the mobility. The mobility, μ, is given by:

$$\mu = \frac{\tau e}{m^*} \tag{2.11}$$

Field

The mobility of electrons and holes in undoped bulk Si is 1450 cm^{2}v^{-1}s^{-1} and 505 cm^{2}v^{-1}s^{-1} respectively [Schaffler 1997]. This disparity is the reason that pMOS devices are made wider for CMOS to match drive currents. A high mobility is desirable as an increased mobility results in increased drive currents (*Equation 2.6*) and this is what the semiconductor is striving for. This can be achieved by reducing the amount of scattering that takes place. One way of doing this is to place the lattice in the channel region of a MOSFET under strain.

2.3.1 Ionized impurity scattering

When doping semiconductors to create structures such as the source and drain regions, a fraction of ionized dopants are often deposited unintentionally in the channel region. The ionised dopants alter the local periodicity of the lattice potential. As this is a static change in the potential it is possible for the charge carriers to screen its effect. When there are few carriers, or if the carriers don’t have much energy (low temperatures) this form of scattering can cause a problem otherwise its effect is minimal. The number of ionized dopants in the channel can be increased during high temperature stages of processing. In such conditions the dopants diffuse from the areas they were originally implanted. The industry tends to employ a Rapid Thermal Anneal (RTA) technique to reduce this effect where a high temperature (~1000^{o}C) is applied only for a few seconds (low thermal budget).

2.3.2 Phonon scattering

If a semiconductor is at a finite temperature above absolute zero, the atoms in the crystal lattice will vibrate about a central position. The independent motion of each atom causes local changes in band structure which in turn alters band energies. These lattice vibrations act as a perturbing potential for charge carriers and are called phonons. Phonons are bosons and are governed by Bose-Einstein statistics. They represent time dependant perturbations of the system since they illustrate the motion of the atoms that are the centres of electronic charge. The total energy of the system before a scattering event is equal to the energy of the charge carrier plus or minus the phonon (ħω) that is either absorbed or emitted. The energy of the charge carrier can be split into the potential (energy band minimum) and kinetic energy [Harrison 2002].

$$E_i^T = E_i + \frac{\hbar^2 k_i^{\ 2}}{2m^*} \pm \hbar\omega \tag{2.12}$$

Field

$$E_f^{\ T} = E_f + \frac{\hbar^2 k_f^{\ 2}}{2m^*} \tag{2.13}$$

Field

In optical phonons the distance between neighbouring atoms is greater than acoustic so have a much higher minimum energy causing large changes in carrier energy. Acoustic phonons have negligible energy compared to carriers so scattering is approximately elastic. The assumption of parabolic bands limits the use of this model for holes in valence bands. Solving mathematically there are two possible scattering events available to the system for either the absorption or emission process (*Figure 2-5*) and conservation laws are followed.

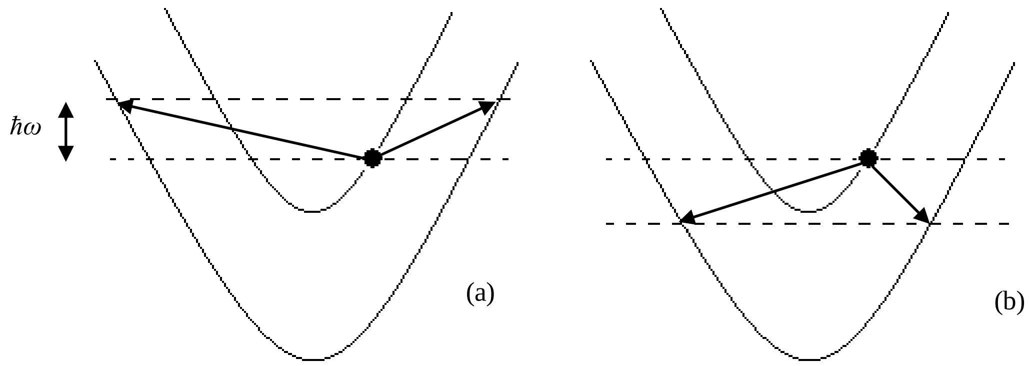

Figure 2-5: Energy band representation of phonon scattering by charge carriers. (a) Absorption of a phonon by a charge carrier, (b) emission of a phonon by a charge carrier [Harrison 2002]

Scattering depends on the number of active phonon modes which is a function of temperature so it is imperative to keep the channel temperature low. Optical phonons have a high energy but once the system passes this threshold emission greatly increases and carriers experience velocity saturation.

2.3.3 Alloy scattering

When Germanium atoms (Ge) are added to Si to form a Virtual Substrate (VS) it creates atomic disorder. The Ge atoms change the band structure of the crystal as a whole and also cause localised changes in band structure. The distribution of atoms in the alloy tends to be random with the scattering rate at a maximum when the Ge content is at 50%. It is possible for carriers to rearrange themselves to screen static potentials reducing its influence but it is not known whether screening occurs here.

If carriers are confined within Si channel on top of the SiGe VS, alloy scattering is believed not to cause a problem. However for pMOS devices we tend to get a parasitic channel in the SiGe VS so alloy scattering could have an effect.

2.3.4 Interface roughness scattering

Interface roughness at the semiconductor surface (*Figure 2-6*) causes ‘potential bumps’ in the way of carriers causing them to scatter

.

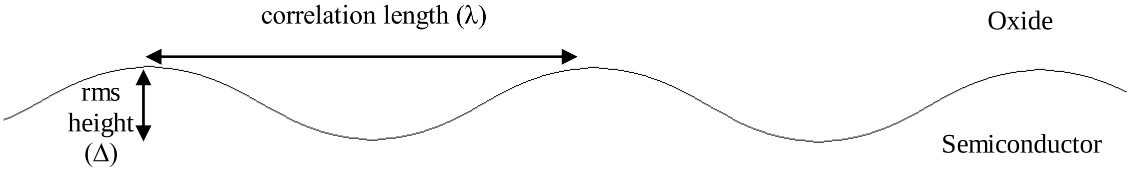

Figure 2-6: Semiconductor/oxide interface with the parameters that effect interface roughness scattering

This roughness occurs even with interfaces of high quality, and exists for one or two mono-layers due

- to non ideal growth conditions.
- Imprecise shutter control of SC species

The charge carriers form a 2D electron/hole gas and are confined against the semiconductor surface. Since the surface roughness acts as a static potential the carriers can rearrange in response and screen it in such a way that its influence is reduced. The problem is when E_{eff} is large, as under such conditions the carriers are forced against the channel surface and screening effects can not be set up. Since modern short channel devices all have electric field greater than 5MVcm^{-1} interface roughness scattering is becoming debilitating.

2.4 $Si_{1-x}Ge_x$ virtual substrate induced biaxial strain

Silicon that is under biaxial strain has been either stretched or compressed in two of the three orthogonal directions. The biaxial strain investigated in this MSc involved placing the Si channel region under tensile strain. The Si atoms are therefore further apart than they would be if no external forces were acting on them and this is illustrated in *Figure 2-7*.

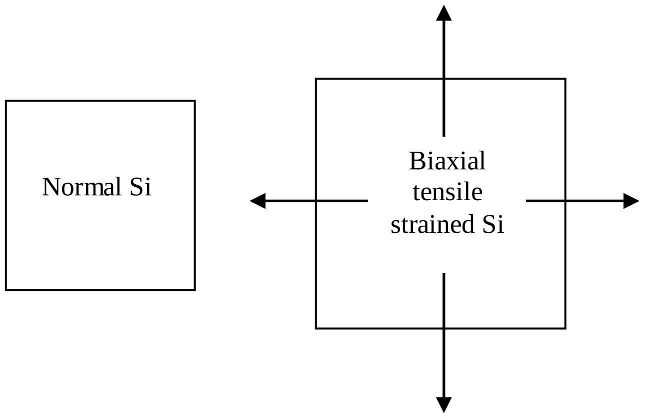

Figure 2-7: Stress applied to the Silicon in two directions resulting in this illustration in biaxial tensile strain

2.4.1 $Si_{1-x}Ge_x$ virtual substrate growth

In order to achieve the biaxial strain in the Si channel a $Si_{1-x}Ge_x$ VS can be used. Ge was chosen because of its compatibility with the Si technology and its marginally larger lattice parameter:

- Si – 5.431Å
- Ge – 5.657Å

Vegards law is used to calculate the approximate atomic spacing in an alloy. However it has been noted during theoretical research that semiconductor alloys do not follow this law precisely [Fong 1976]. This violation is most noticeable in pseudobinary alloys. The deviation from the law was subsequently backed up by experimental evidence for $Ga_{1-x}In_xAs$ [Mikkelsen 1982], $Al_xGa_{1-x}As$ [Gehrsitz 1999] and $Si_{1-x}Ge_x$/Si [Nikulin 1996] [Dismukes 1993] heterostructures. From experimental data on $Si_{1-x}Ge_x$ structures [Dismukes 1993], the following polynomial was derived [Payet 2005] that more accurately reflects the changing atomic constant with Ge fraction.

$$a_{Si_{1-x}Ge_x}(x) = 2.78192\cdot10^{-3}x^2 + 1.98821\cdot10^{-2} + 0.54313 \tag{2.14}$$

It is important to be able to accurately calculate the atomic spacing of the VS so that the effect on the Si channel deposited on top can be known.

When the $Si_{1-x}Ge_x$ is deposited by epitaxial growth on top of a bulk Si wafer (which forms the starting platform of the VS), the SiGe atoms will initially line up with the Si below and be under compressive strain and as the depth of the layer increases it will begin to relax. The most commonly used way of relaxing this $Si_{1-x}Ge_x$ layer is to grade the Ge content. If a final $Si_{0.8}Ge_{0.2}$ composition is required the percentage of Ge would be slowly increased up to 20% over a distance of ~4µm. After a critical thickness has formed it will become energetically favourable for the lattice to relax [Huang 2005] and where the atoms do not line up due to the difference in atomic spacing misfit dislocations are formed. While it is desirable to have a large number of defects in the graded layer to maximise relaxation [Vdovin 2002], the problem is that on interaction with each other the misfit dislocations form threading dislocations which can move to the surface. This causes major problems for device performance. In order to try and reduce the number of threading dislocations point defects (PD) in the form of vacancies (absence of atoms) or interstitials (additional atoms) are often intentionally introduced. These condense on {111} planes forming dislocation loops. High numbers of PDs should promote dislocation climbing therefore annihilating threading dislocations creating a smooth surface morphology and low defect density.

Generally dislocations are not mobile at room temperatures and so only become debilitating when a wafer undergoes a high temperature process, Dislocations then travel across the wafer destroying the device. Once a graded $Si_{1-x}Ge_x$ layer is deposited a uniform layer of $Si_{0.8}Ge_{0.2}$ would then be grown for ~1µm allowing a high Ge content to be achieved on the surface, with a high degree of misfit strain relaxation but without introducing a crippling number of threading dislocations. H or He implantation has been tried as a way of restricting dislocations to areas below the upper surface [Lyutovich 2004] but the most commonly used method these days in a Chemical Mechanical Polishing (CMP) process performed between the graded and uniform SiGe layers.

The properties that are desirable in a VS and the issues associated are summarised in *Table 2-1*.

Desirable VS properties	VS issues
i. High Ge content – lattice parameter proportional to the Ge atom fraction therefore more channel strain ii. High strain relaxation iii. Smooth surface and low defect density to reduce interface carrier scattering effects iv. Low thickness to prevent SRB (strain relaxed buffer) acting as a thermal barrier allowing localized heating in the channel and performance decreases.	i. Low yields (as desire <10^3cm^{-2} defects) ii. Processing challenges iii. pMOS performance not so good. iv. Expensive v. Degree of relaxation generally insufficient vi. Development of "cross hatch pattern" increases surface roughness vii. High density of threading dislocations

Table 2-1: VS properties and issues

Due to the number of years of research on Si substrates it is possible to have Si wafers that are completely free of defects, containing almost no impurities and at a low cost and these qualities are what are being strived for in the development of SiGe VSs.

2.4.2 Biaxial strained Si energy bands

The conduction band minima of Si occur along the 6 <100>/<001> crystal directions (Δ minima). The bands are associated with electrons moving in the six orthogonal directions (±x, ±y, ±z) and are 6 fold degenerate (*Figure 2-8*). Electrons are scattered between these bands via a process known as inter-band scattering.

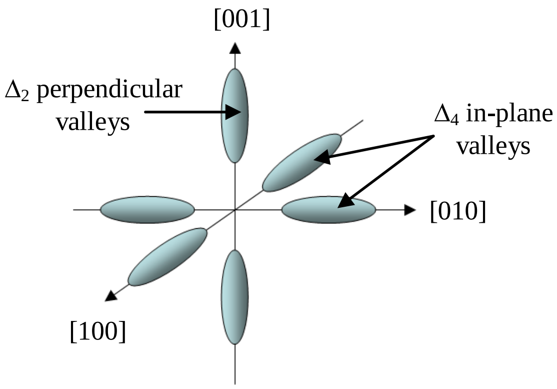

Figure 2-8: Energy orbitals of unstrained Silicon

The biaxial strain induced in the silicon causes the conduction bands to split dependant on the type of strain (*Figure 2-9*).

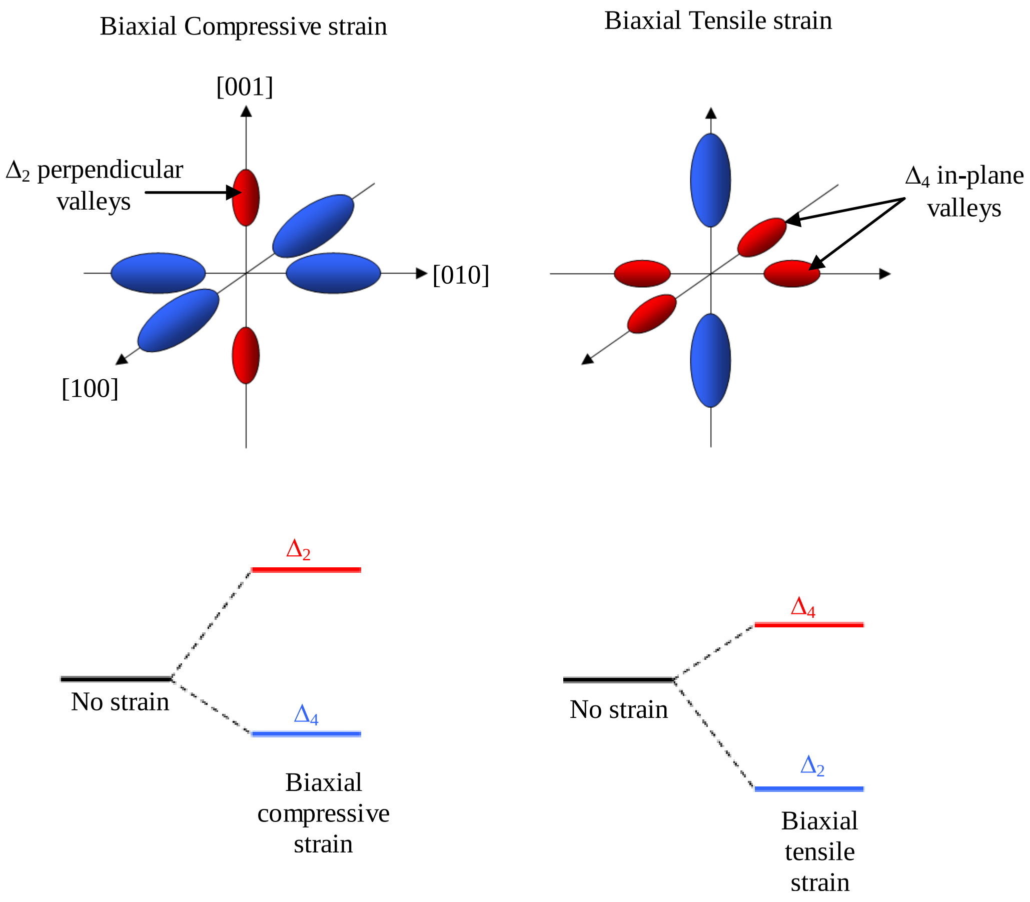

Figure 2-9: Energy orbitals under compressive and tensile strain, and the effect on the energy levels that represent the orbitals

For every 10% Ge in SiGe buffer layer, the Si energy bands split by 67meV. Electrons preferentially fill the lower energy bands, therefore tensile strain is more beneficial for nMOS devices as it leaves only the lower energy 2 fold degenerate levels for inter-band scattering to occur between by presenting fewer possible final states for the carriers to scatter into. As a result electrons can travel further through the lattice before scattering (τ increased). Strain also narrows the shape of the bands in an energy-momentum diagram indicating a reduction in effective mass. This allows the electrons to accelerate more for a given E field. From the equation for mobility (*Equation 2.11*) we have an increase in mobility from the strain.

The degeneracy of the holes in the valence bands is also altered (*Figure 2-10*).

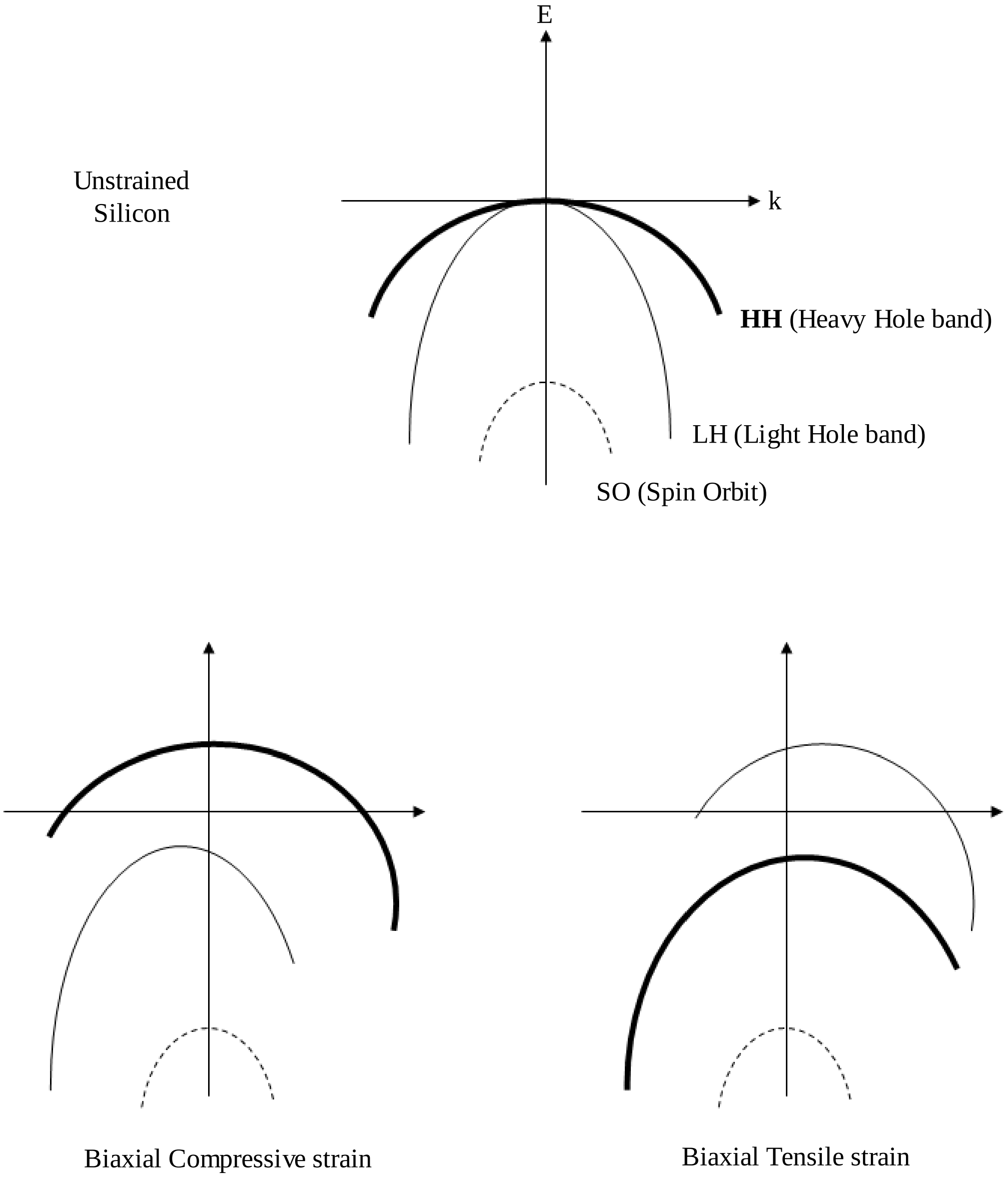

Figure 2-10: Light Hole and Heavy Hole band splitting of the Silicon valence bands under strain

Strain has less of an effect on holes than electrons with only a 38meV split for every 10% Ge in the VS. Scattering still occurs between the two bands but is greatly reduced. The effective mass is hardly affected by the strain [Thompson 2004b] and so the mobility improvements are due to the reduced scattering alone.

The biaxial strain improves mobility for nMOS and pMOS but requires low electric fields and high stress. At high E fields (>5MVcm^{-1}) where modern CMOS operates we obtain 2D surface confinement of the carriers. It has been shown recently [Thompson 2004b] that the surface confinement cancels the light to heavy hole band splitting that resulted from the biaxial strain placed upon the channel region. This means that the mobility enhancements are lost [Mikkelsen 1982].

2.4.3 Critical thickness

The critical thickness (t_c) relates to the amount of Si that can be grown on top of a SiGe VS that is dislocation free and thermodynamically stable [Matthews 1974]. Once the critical thickness is passed the S-Si will begin to relax and defects form.

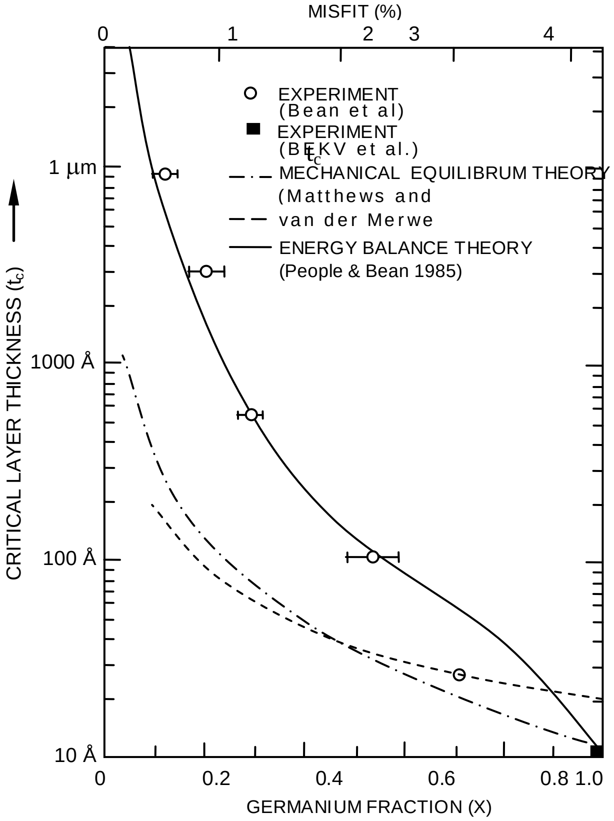

Figure 2-11: Critical thickness of Si than can be grown on top of a SiGe VS [People 1985]

There are a number of approaches for predicting the critical thickness of Si. The most widely quoted is the Mechanical Equilibrium Theory proposed by Matthews and Blakeslee in 1974. Using this, for a 20% Ge content the critical thickness is 16nm (*Figure 2-11*) which is already marginal as we have to take into account the processing steps (oxidation and cleaning) which will remove some of the layer before we can think about how much will actually be left to form the channel. Cleaning the deposited layer typical consumes ~3nm and the oxidation step also typically consumes ~3nm. If the channel region ends up too thin increased scattering at the S-Si/SiGe VS interface can lead to performance degradation.

For thicknesses less than t_c no misfit dislocations are observed using TEM, but when a Si layer is deposited thicker than the critical thickness misfit dislocations are generated as they become energetically favourable. These misfit dislocations are formed at the S-Si/SiGe VS interface via a gliding process due to the threading dislocations that exist in the VS. This creates areas of plastic relaxation in the Si layer.

It was originally thought that the poor performance in devices where the deposited layer of Si is thicker than t_c, was from the loss of strain. However it has been shown that relaxation is not complete due to dislocation blocking at the S-Si/SiGe VS interface [Fiorenza 2004]. Therefore the device keeps the mobility enhancements brought about by the strained Si. Using mobility degradation due to relaxation of the S-Si layer, as a reason for sticking below t_c is not valid.

Instead it has been seen that as you go above t_c the off current increases which has a debilitating result. This has been shown to be due to the misfit dislocations acting as dopant diffusion pipelines that cause a short circuit between the source and drain contacts [Fiorenza 2004].

Other explanations have been put forward such as the misfit dislocations causing a short circuit between the drain and the substrate or that the dislocations allow direct conduction without dopant diffusion. These explanations have been ruled out. Dopant diffusivity does indeed increase along misfit dislocations [Nabarro 1967] and has been experimentally shown that Arsenic diffuses 6 times more along a misfit dislocation at an interface [Braga 1994]. By measuring source and drain current simultaneously it has been shown that it is indeed leakage from the source to the drain [Fiorenza 2004]. During normal operation there is always light emitted due to the electron impact ionization and recombination however for layers greater than t_c light more intense is observed (using Photon Emission Spectroscopy) in discrete points along the channel width [Fiorenza 2004]. These would be at the sites of misfit dislocations. However at this point it could still be considered possible that direct conduction is taking place. This is ruled out by the fact that the off current was reduced as the gate length increased for the same S-Si layer thickness. If direct conduction was occurring this would not happen. Instead it is because the gate length becomes longer than the diffusion length of the dopant therefore stopping the short circuits.

2.4.4 Dopant diffusion issues

The three dopants most commonly used for semiconductors are Boron, Arsenic and Phosphorous. The most important structures they are used to form are the source and drain regions. It is therefore essential to know exactly how these dopants diffuse. The dopant diffusion rates in SiGe and S-Si for Boron and Arsenic are quite different while Phosphorous diffuses at approximately the same rate in both materials. Differences in the diffusion rates cause problems when forming the source and drain regions as they are no longer well defined. A device with a short channel can be prevented from turning off if the source and drain regions touch (*Figure 2-12(b)*).

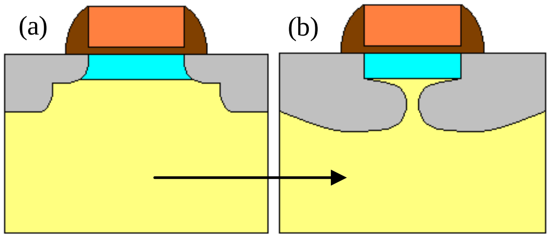

Figure 2-12: (a) A well defined source and drain in a MOSFET and (b) a poorly defined source and drain as a result of differing diffusion rates

2.4.5 Self heating

The thermal conductivity of $Si_{0.8}Ge_{0.2}$ is approximately 15 times lower than bulk Si [Jenkins 2002]. The random presence of Ge atoms in the alloy scatters phonons which prevent effective heat transfer. The heat generated during operation cannot dissipate as the channel is thermally isolated by the underlying VS. This effect is known as self heating.

Heat in semiconductors is mainly carried by acoustic phonons which are also scattered by imperfections and impurities. If threading dislocation count is high we also get scattering of acoustic phonons further decreasing thermal conductivity [Kotchetkov 2001].

For gallium arsenide and gallium nitride thermal conductivity reduces dramatically when the dislocation line density becomes as high as 10^{11} cm^{-2}. For low threading dislocations the thermal conductivity is mostly defined by the intrinsic crystal properties and point defects. SiGe is similar to GaAs and GaN so is likely to behave in the same manner.

It is during d.c. operation that self heating occurs. As the temperature in the channel increases it will cause a decrease in the current drive. Static d.c. operation therefore does not give an indication of the full enhancements in performance. Dynamic data can be obtained by applying a short pulse (~7ns) to the gate at a low repetition rate (<0.01% duty cycle) which ensures that heat build up doesn't take place [Jenkins 1995][Jenkins 2002].

2.4.6 pMOS parasitic channel

Due to the band structure in a S-Si on SiGe MOSFET electrons are confined at the surface of the S-Si channel. Holes however are not confined purely inside the S-Si layer but are able to form a parasitic channel in the $Si_{1-x}Ge_x$ layer because of the alignment of the valence band at this interface (*Figure 2-13*).

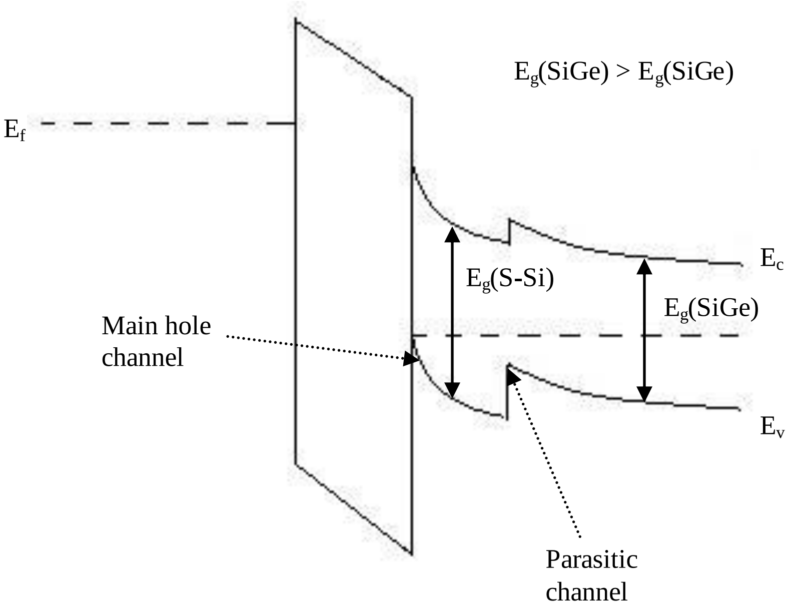

Figure 2-13: Energy band diagram for a pMOS device under inversion bias conditions [Fong 1976]

There are a number of ways to try and eliminate the parasitic channel. Increasing the thickness of the S-Si layer lowers the parasitic conduction in the $Si_{1-x}Ge_x$ buffer. It is also possible to smooth the valence band by depositing a grade back layer prior to depositing the Si. This involves reducing the Ge content down to zero at the top of the VS over a distance of a few nm. This then means that a thinner layer of S-Si must be grown otherwise it will be thermally unstable for further processing steps. Another option is to add dopants near the bottom of the S-Si layer which can eliminate the hole confinement in the VS.

2.4.7 Performance improvement

VS technology allows performance increases in both nMOS and pMOS devices. A Ge content of 40% however is required (*Figure 2-14*) for the maximum gain for both device types.

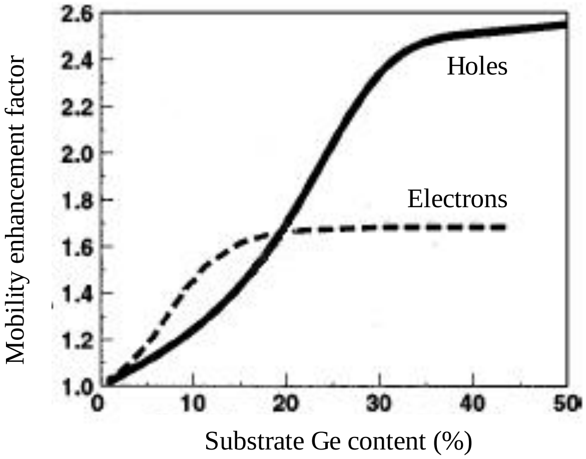

Figure 2-14: Theoretical mobility enhancements for electrons and holes as a function of the % Ge in a $Si_{1-x}Ge_x$ VS under low field conditions [Oberhuber 1998]

A Ge content of 40% would ideally be used; however, the amount of Si that can be grown on top is limited by the critical thickness (*Figure 2-11*), and the increased number of defects results in compromises having to be made.

2.5 Uniaxial Strain

An alternative to global biaxial strain for device enhancement is local uniaxial strain introduced in the direction of the device channel length.

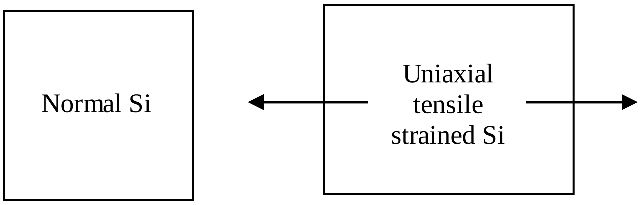

Figure 2-15: Stress applied to the Silicon in one direction in this case resulting in uniaxial tensile strain

This form of strain is achieved locally in the channel region through the use of a strained nitride capping layer (contact etch stop layer or CESL). Uniaxial strain of this kind is known as “process induced” as the nitride layer is deposited during the fabrication process of the device. The nitride (SiN) is deposited after the silicidation process directly on top of the device (*Figure 2-16*) and the thickness and intrinsic strain of the layer partially determines the strain applied to the underlying channel.

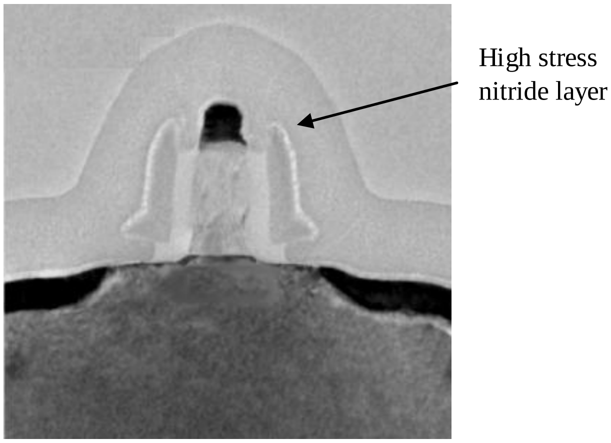

Figure 2-16: TEM of a 45nm nitride capped nMOS channel [Thompson 2004a]

The film itself is not the only factor in determining the resultant channel strain. It is shown later that a number of other factors are also involved:

- Gate stack dimensions
- Spacer dimensions
- Source Drain elevation

As a process technique, strain of this type is easier to implement into current fabrication lines that biaxial strain. Modest performance increases (drive currents higher by 6% with only a 100MPa increase in tensile strain in the channel [Ootsuka 2000]) show promise for this technology without the debilitating effects associated with the use of a VS.

2.5.1 Nitride growth

A Plasma Enhanced CDV (PECVD) reactor is used to deposit nitride layers. By varying the standard gas mixture (SiH_4, NH_3 and N_2), the deposition rate, and the temperature it is possible to deposit nitride layers with an intrinsic strain in the ±GPa range (positive values indicate tensile strain).

It has been shown necessary [Payet 2005] [Ootsuka 2000] [Ito 200] to use tensile stress for nMOS and compressive stress for pMOS in order to achieve performance enhancements. If a blanket covering of tensile strained SiN is deposited on the wafer it will have a positive effect on the nMOS devices and a negative effect on the pMOS. This causes a problem as it is not desirable to cause the performance of one device type to degrade. Two possible ways of getting round this are:

1. the nitride layer could be selectively deposited only over one particular device type.
2. a blanket layer is deposited and its strain is neutralised using a masking layer and Ge ion implant over a particular device type[Thompson 2004c], [Shimizu 2004].

The second method was used in the device fabrication for this MSc work (*Figure 2-17*).

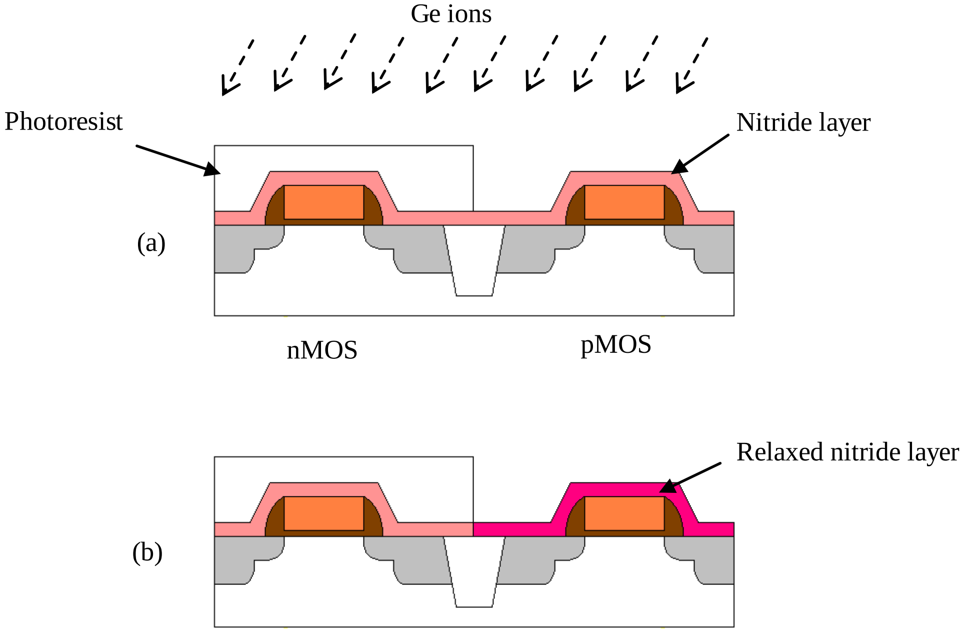

Figure 2-17: Schematic of MOSFET with a nitride layer undergoing (a) a Ge ion bombardment process and (b) the result relaxed nitride layer is only over the selected devices

The energy of the Ge implantation directly affects the level of strain relaxation in the nitride layer. The process destroys bonds in the SiN, and the ions are able to penetrate deeper if they have more energy. 80-100keV seems to be the optimum energy (*Figure 2-18*) as above this value the relaxation achievable saturates. There is also a critical dose to destroy bonds and going above this has minimal effect [Shimizu 2004].

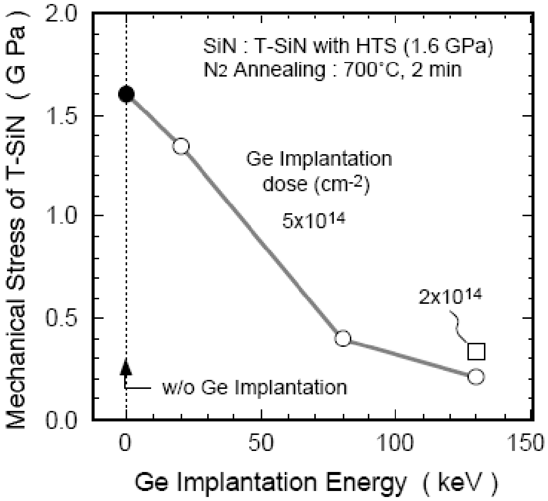

Figure 2-18: Effects of Ge implantation on mechanical stress of nitride layer [Shimizu 2004]

2.5.2 SiN qualities

Nitride layers have been used in the industry for a while as they form excellent barriers. While almost impervious to wet HF dip they etch readily in fluorine containing plasmas. This allows them to be selectively etched with respect to Si [Toda 2001].

It is only recently that nitride layers have been noticed as having the potential to be used as a method of inducing strain in the channel. To be used for such a purpose the nitride layer must be thermally stable after formation due to the high thermal budgets used in CMOS fabrication. The nitride layers do possess this quality (*Figure 2-19*).

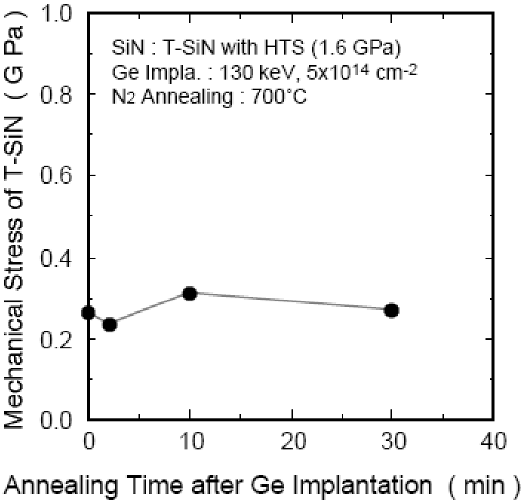

Figure 2-19: Thermal stability of deposited nitride layer [Shimizu 2004]

2.5.3 SiN$_x$ stress determination

Nitride layers can be deposited with a large range of intrinsic strains. The atomic structure, which is determined by the deposition conditions, allows for this range. In the deposited nitride layer you have N-H, Si-H and Si-N bonds. The Si-N bonds are dominant but it is the ratio of the other two types of bonds that determine the intrinsic strain of the film [Arghavani 2004].

Figure 2-20 shows the FTIR spectra for nitride layers with varying intrinsic strains. The peak position of the N-H bond increases with tensile strain and the peak positions of the Si-H and Si-N bonds decrease. The area under the peaks also changes as the strain does, but the combined areas under the N-H and Si-H peaks remains almost constant for all strains. This is because the total hydrogen content in the nitride layer does not change.

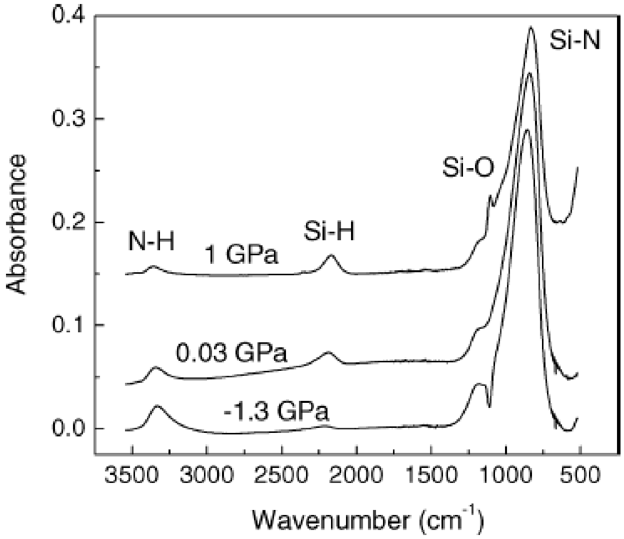

Figure 2-20: FTIR spectra showing the N-H, Si-H, Si-O, and Si-N peaks of three differently strained samples. For clarity each has been shifted on the vertical axis. [Arghavani 2004]

When Si-H bonds are formed they have a weakening effect on the Si-N bonds that are attached to the same Si atom. The distance between the Si and N atoms is made larger, because of the weakening of the bonds, and the nitride layer is tensile strained. In addition the weaker Si-N bonds can strengthen the remaining N-H bonds making them shorter. By observing the ratio of the bonds we can determine the stress of a nitride layer

N-H>>Si-H	gives compressive strain
N-H<<Si-H	gives tensile strain

2.5.4 Nitride action

Once a nitride layer has been deposited it will exert a stress on the atomic layers it is in contact with as it reverts back to its optimal lattice spacing. Due to its nature it will only have a finite range of influence. If the nitride layer is under tensile strain the nitride atoms will pull together. The resultant stress, if it has been deposited on a flat surface, on the atomic layers underneath will be compressive in the x and y-directions and tensile in the z-direction. This is not the case however, as the nitride layer is deposited over the device (*Figure 2-16*). The main parameters that have been shown to determine the resultant strain in the channel region are:

- Intrinsic strain of the nitride layer (S_{int})
- Nitride layer thickness (t_N)
- Gate length (L)
- Gate width (W)
- Gate height (h)
- Spacer width (d)

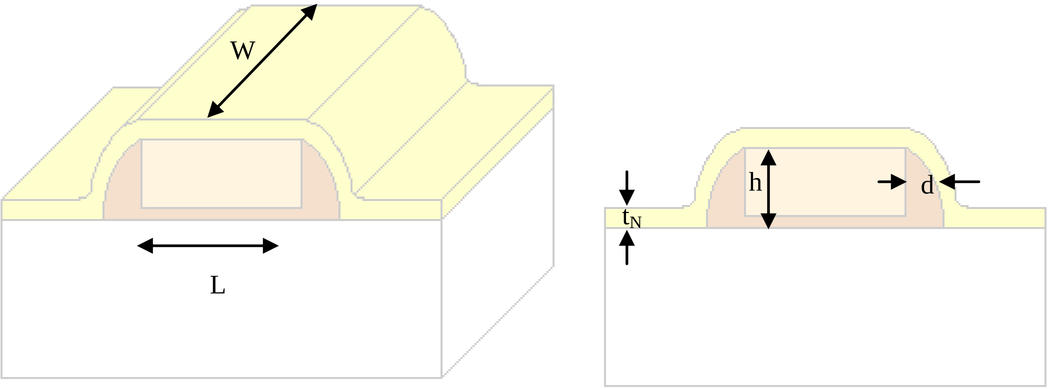

Figure 2-21: Schematic of MOSFET with additional nitride layer specifying important parameters for channel induced strain

Simulation has shown how these parameters affect the resultant strain in the channel region [Payet 2005].

2.5.4.1 Gate length and width

Figures 2-22 & 23 illustrate the strain for a long channel (L=2µm, W=10µm) device with a nitride layer 20nm thick. The channel is only under tensile strain in the x-direction at the source and drain ends. This is because the gate stack is too long for the nitride to be able to place the entire channel under tensile strain. Here we see why the process induced strain brings about only a uniaxial strain component to the channel. The width of the channel is so great that the nitride layer is unable to have an effect on the y-axis just as it is having difficulty with the channel length of the device. In the vertical direction the channel is only under a small amount of compressive strain again at the source and drain ends. This is because the nitride layer is trying to stretch out and pushes down on the gate stack.

The magnitude of strain along the channel length and in the vertical direction is important as both elements contribute to the band splitting as described later in *Chapter 2.5.5*.

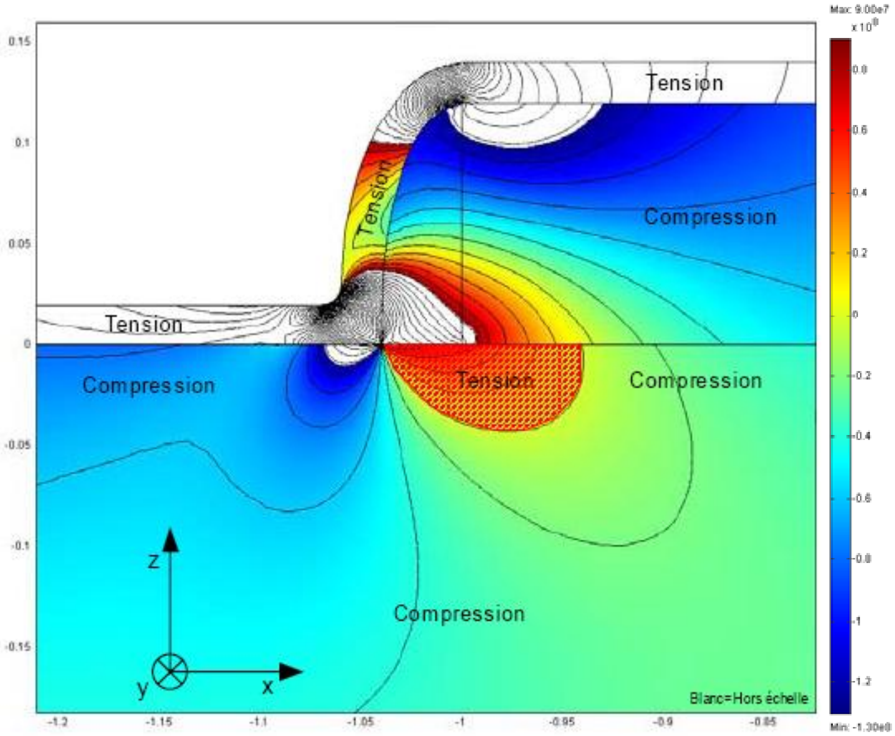

Figure 2-22: Simulated strain in the x-direction for a MOSFET with a tensile nitride layer. L=200nm, S_{int}=1.5GPa [Payet 2005]

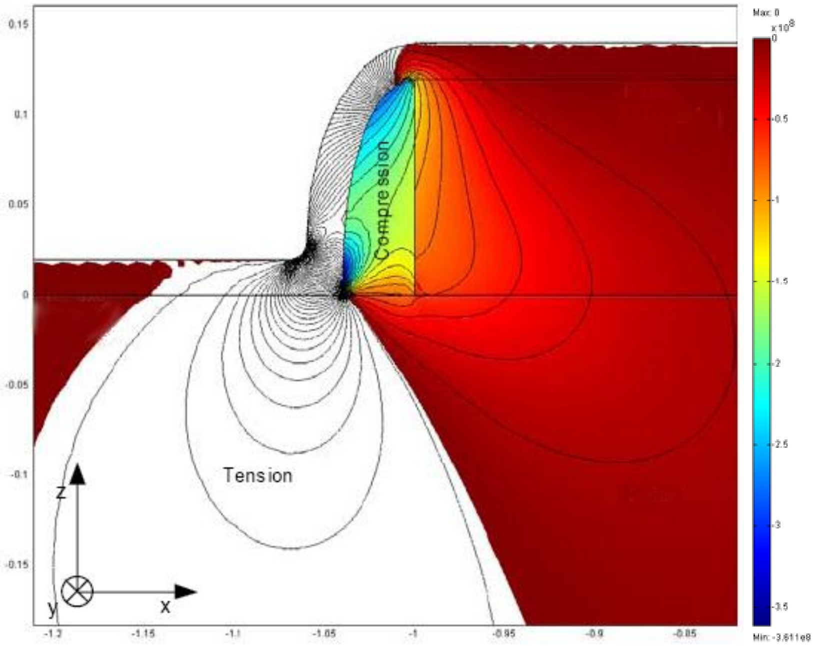

Figure 2-23: Simulated strain in the z-direction for a MOSFET with a tensile nitride layer. L=200nm, S_{int}=1.5GPa [Payet 2005]

When the channel length is reduced (L=60nm, W=10μm) which can be seen in *Figures 2-24 & 25* the areas of tensile strain at the source and drain ends of the channel have met. The entire channel is now under the desired strain. The component of strain in the vertical direction has also increased along the channel length.

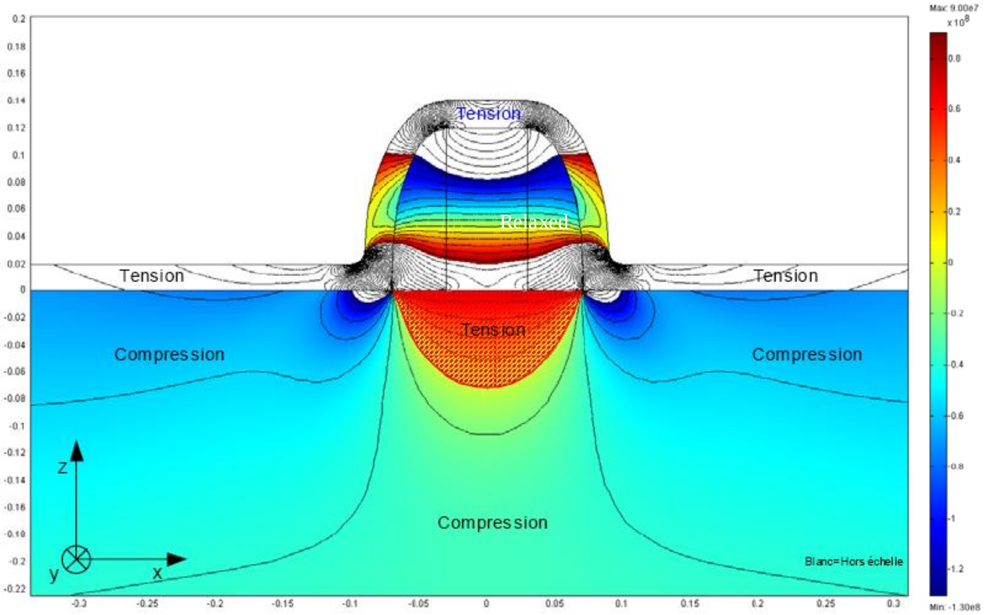

Figure 2-24: Simulated strain in the x-direction for a MOSFET with a tensile nitride layer. L=60nm, S_{int}=1.5GPa [Payet 2005]

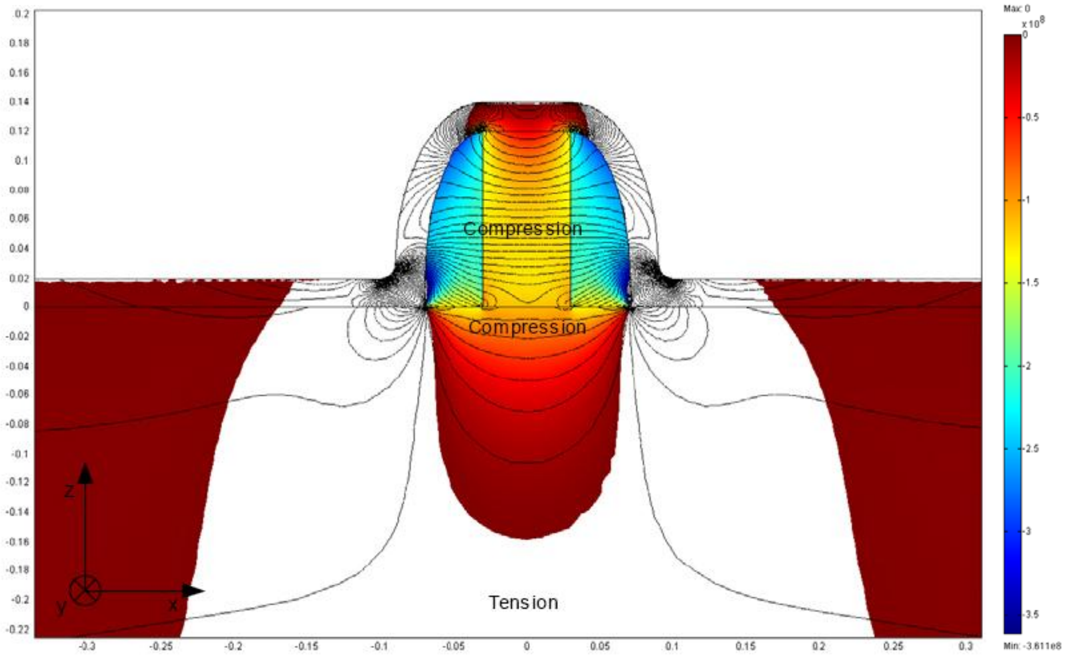

Figure 2-25: Simulated strain in the z-direction for a MOSFET with a tensile nitride layer. L=60nm, S_{int}=1.5GPa [Payet 2005]

It has been found [Payet 2005] that only devices with a gate length less than 200nm have the entire channel under the desired uniaxial strain. Experimental data [Toh 2005] [Toda 2001] has verified these simulations.

2.5.4.2 Intrinsic strain of the nitride layer

Figure 2-26 shows that as the intrinsic strain of the nitride layer is increased, the x direction component of strain in the device channel increases linearly and the z component decreases linearly.

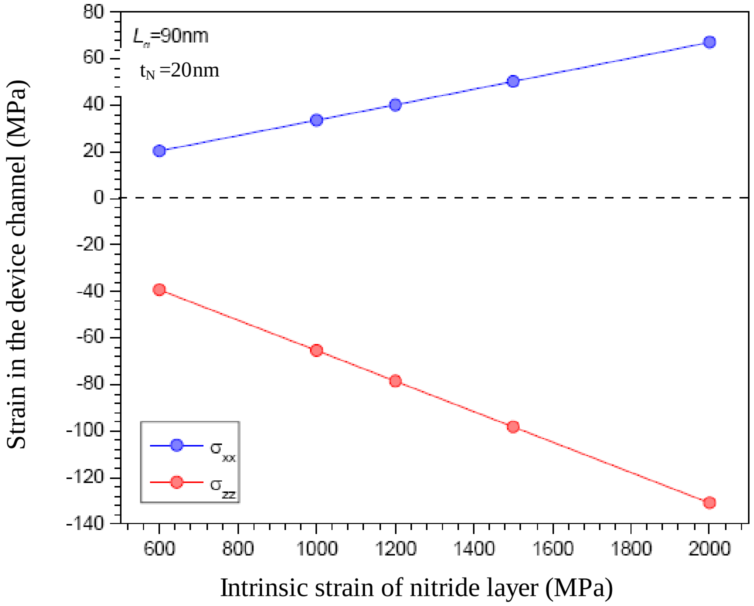

Figure 2-26: Strain in the device channel as a function of the intinsic stress applied by the nitride layer [Payet 2005]

2.5.4.3 Nitride layer thickness

Figure 2-27 relates the nitride layer thickness to the stress in the channel. As the thickness is increased for a tensile strained layer, the x and z components of the channel strain start to reduce.

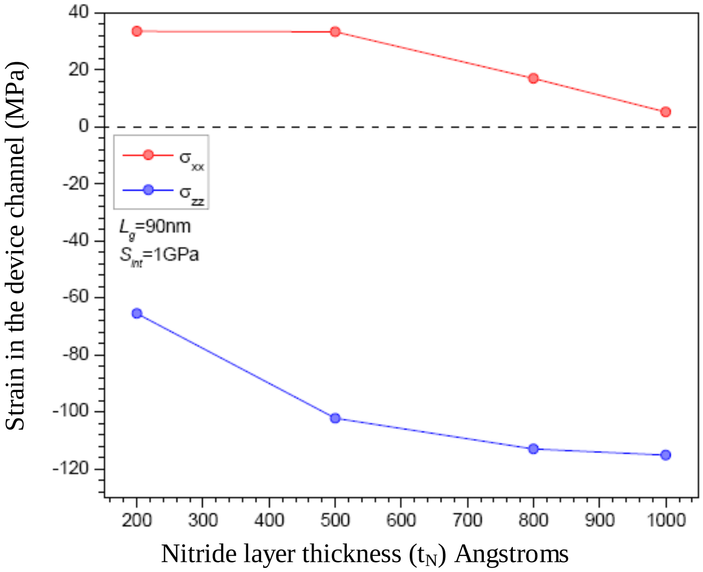

Figure 2-27: Strain in the device channel as a function of the nitride layer thickness [Payet 2005]

2.5.4.4 Spacer width

Figure 2-28 shows the relationship between the spacer width and the strain in the channel. As the spacer width increases the x direction strain is reduced and the z direction strain is increased.

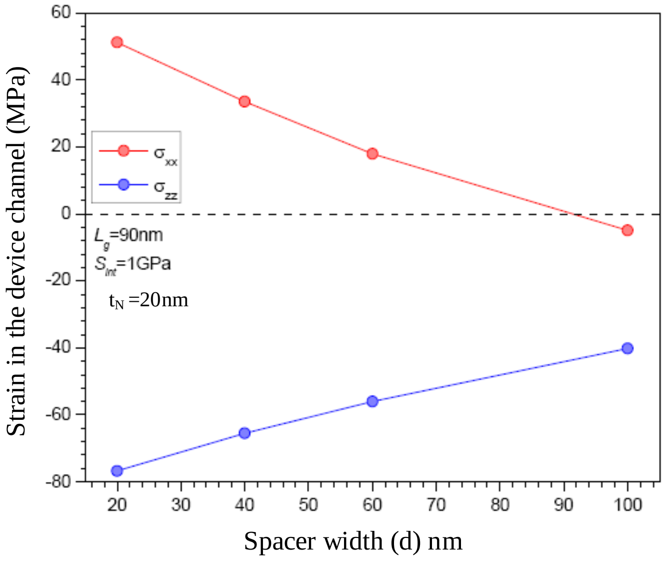

Figure 2-28: Stress in the device channel as a function of the spacer width [Payet 2005]

2.5.5 Altered energy band properties

Research has shown that uniaxial tensile stress applied to the channel region brings about nMOS performance enhancements and compressive stress brings about pMOS enhancement [Ito 2000] [Payet 2005] [Thompson 2004b]. Both the x and z components of strain in the channel contribute to mobility enhancement.

For a pMOS device, in order to split the valence bands by around 60meV, which is the required amount for scattering rates to be lowered, high stresses must be applied to the channel region (>1GPa). This is the case whether the strain induced is biaxial or uniaxial. The difference with uniaxial strain is that the effective mass of the holes is also greatly reduced at all levels of applied stress. This reduction in mass does not occur when biaxial stress is used. For low uniaxial strain the effective mass can be reduced by as much as ~40%.

High effective electric fields also cause problems for biaxially strained Si as when the carriers are confined at the S-Si/oxide interface the mobility enhancements are lost. The surface confinement of carriers in uniaxial strained silicon actually increases the splitting of the valence bands further increasing the mobility enhancements [Thompson 2004b].

Figure 2-29 shows the effects of uniaxial strain on an already biaxially strained channel. If a tensile nitride layer is deposited then the conduction band separation is increased. The two components of the uniaxial strain have an additive effect on the band separation. As the biaxial strain has already taken care of the interband scattering further mobility enhancement comes from the lowered effective mass. If a compressive nitride layer is utilised the separation of the energy bands is reduced allowing interband scattering to once again taken place reducing the mobility enhancements.

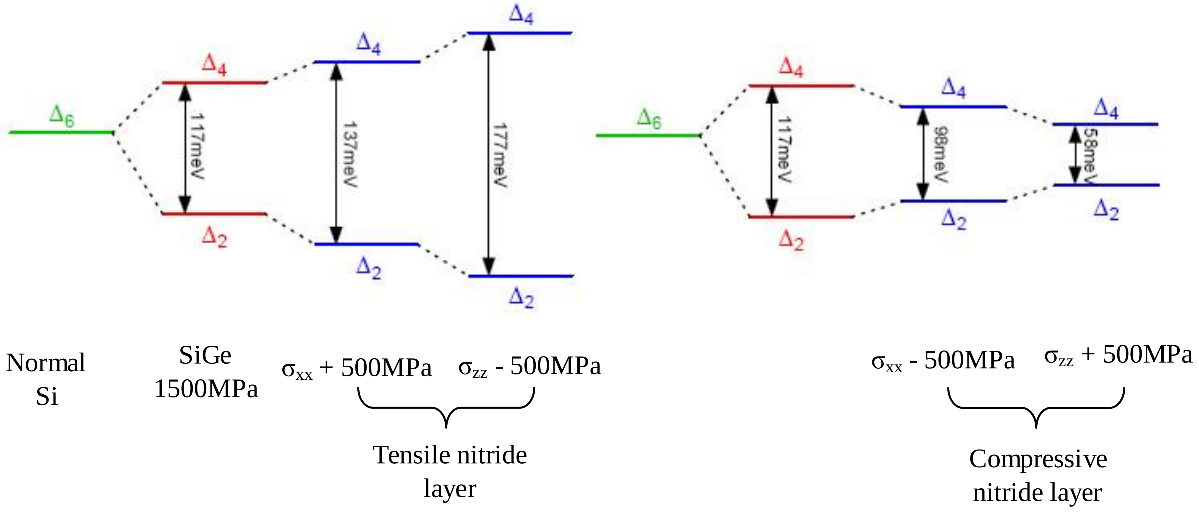

Figure 2-29: Conduction band splitting due to induced strain in the channel region for an nMOS device [Payet 2005]

2.6 Strain effects on threshold voltage

The strain effects on nMOS have been heavily researched and expressions for the shift in threshold voltage have been derived [Lim 2004] [Thompson 2004b] for uniaxial:

$$e\Delta V_T(\sigma) = (m-1)\left[\Delta E_g(\sigma) + k_B T \ln \frac{N_v(0)}{N_v(\sigma)}\right] \tag{2.15}$$

Field

and for biaxial

$$e\Delta V_T(\sigma) = \Delta E_c(\sigma) + (m-1)\left[\Delta E_g(\sigma) + k_B T \ln \frac{N_v(0)}{N_v(\sigma)}\right] \tag{2.16}$$

Field

There are two reasons for the larger shift in threshold voltage for biaxial strain than uniaxial. Firstly the band gap narrowing is larger for biaxial strain as the valence band edge is shifted more. The second is because for uniaxial strain there is no electron affinity term. This is omitted as the nitride layer puts both the poly-Si gate and channel under strain and so both have equal electron affinity changes.

Lim et al show, using recommended deformation potentials [Fischetti 1996] and expressions derived for the effect of strain on the band edges, that the electron affinity term for biaxial stress is the largest factor influencing threshold voltage shift although the bandgap narrowing term is significant. The bandgap narrowing for uniaxial strain is much lower than for biaxial strain as are the changes in the density of states in the valence band. The total shift from the derived equation is shown in *Figure 2-30.*

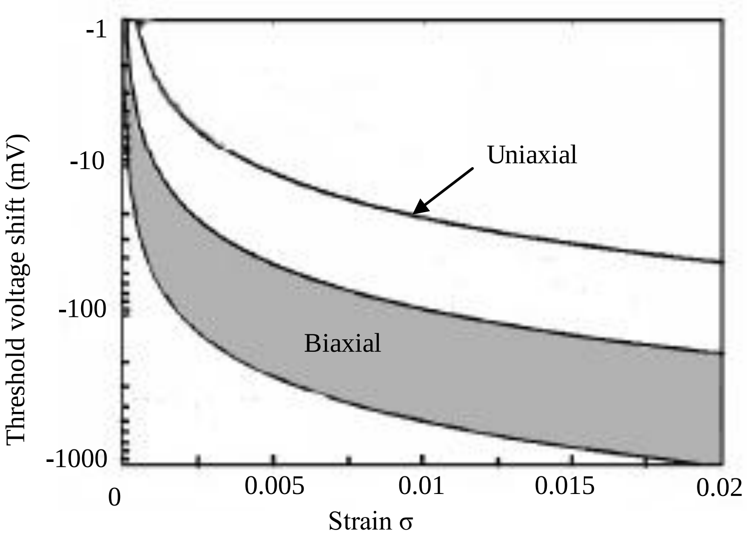

Figure 2-30: Strain induced threshold voltage shift

For biaxial strain two different sets of deformation potentials are used to bracket the uncertainty in the electron affinity which brings about a spread of possible shifts in threshold voltage for a given strain.

The biaxial strain can be related to the percentage of Ge in the VS [Goo 2003] giving us the relationship for nMOS as shown in *Figure 2-31*. For a Ge content of 10% it is therefore expected that the shift from the biaxial strain would be ~100mV.

While there has been a fair amount of research with regards to nMOS, pMOS devices and the effect stress has on them, has not been so widely published. It has been found that biaxial strain has a similar but smaller effect on the threshold voltage (*Figure 2-31*). However the valence band acts as a barrier for the holes by confining some of the inversion charge at the S-Si/SiGe interface as a parasitic channel. As a result the shift is believed to be lower at around 40mV for 10% Ge as opposed to 67mV [Goo 2003].

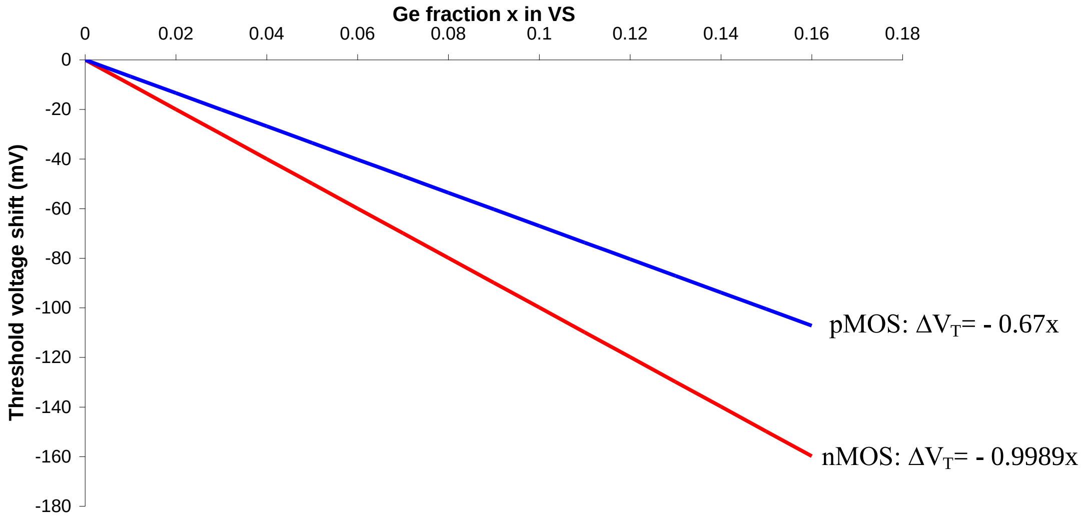

Figure 2-31: Threshold voltage shift as a function of the Ge fraction in the VS

Literature searches indicate that nothing has been published on the effect of uniaxial strain on pMOS devices.

2.7 MOSFET scaling issues & Short Channel Effects

The force driving the industry is the need for higher device speeds and more densely packed circuits. These targets can be achieved by shrinking the physical size of the devices. Unfortunately this is not that simple and it is possible for short devices to cause far more problems than long ones.

The desire is therefore to reduce all the device dimensions and the voltages applied by a scaling factor χ. In doing this the electric fields will be proportional to those present in long channel devices. This constant field scaling factor was proposed in 1974 [Dennard 1974]. Due to advanced lithography techniques and dopant schemes it is possible to scale the physical dimensions of the MOSFET, however, it is not possible to scale down the basic physical parameters of the materials being used such as the work functions, the defect densities, the oxide charges, and carrier sizes for example. These factors will limit the extent of the scaling.

If we scale the device dimensions by χ then:

- Current saturation scales by $1/\chi$
- Gate oxide capacitance scales by $1/\chi$
- Power dissipation scales by $1/\chi2$
- Max frequency of operation scales by χ
- Current density scales by χ

These parameters all scale beneficially except for the manner in which current density is scaled. A high current density can lead to Ohmic heating and electromigration problems in the conductive regions of an integrated circuit that connect the individual transistors. These ‘interconnects’ are made with metal, silicides or highly doped polysilicon. The charge carriers are going extremely fast and so have high kinetic energies (and hence momentums) and upon collision with the lattice atoms of the interconnects, this momentum can be transferred. The lattice atoms are physically moved which might result in a break in the track or if two tracks are made to touch (they are very close together) it could lead to a short circuit. This is the electromigration effect.

In reality the constant field scaling proposed by *Dennard et al* has not been followed as his methodology does not take into account other performance and reliability issues such as devices isolation, and so the electric fields in MOSFETs have been allowed to increase as device dimensions have reduced. Instead two differing scenarios have been forged. The first satisfies high performance demands through the use of a high supply voltage. The second satisfies the need for low power applications with the use of a reduced supply voltage (*Figure 2-32*).

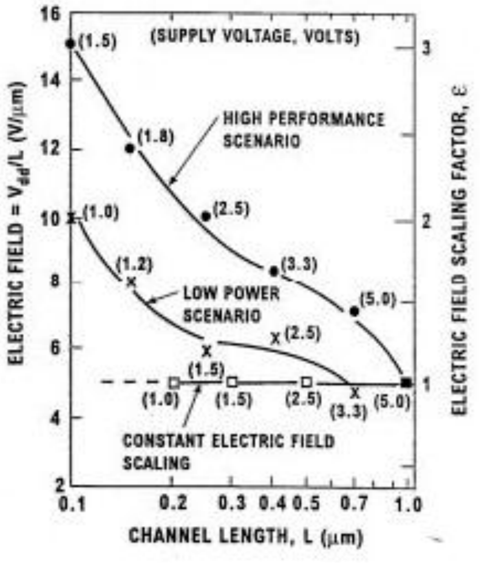

Figure 2-32: Scaling scenarios as device dimensions have been reduced [Davari 1995]

Other issues associated with the scaling of the MOSFET are:

- Punchthrough
- Drain induced barrier lowering
- Hot electrons
- Velocity saturation
- Gate leakage

and collectively these are known as Short Channel Effects (SCEs). In most cases they can be alleviated through increased doping.

It should also be noted that in *Chapter 2.2* the 'Gradual Channel Approximation' was used where we assumed current flow to be purely one dimensional parallel with the gate length. In very short devices this simplification is no longer valid as the current also flows from the sides of the source and the drain becoming highly two dimensional. This means that far more computational power is required to properly simulate such devices.

2.7.1 Punchthrough

In long channel devices when looking at the size of the depletion region in a device we can assume the region to be rectangular and completely controlled by the gate. In short channel devices the source and drain are responsible for creating some of the depletion region in the channel. This leads to an increased depletion region under the inversion layer for the same gate voltage, a larger surface potential and so the channel is more attractive to carriers. The threshold voltage starts to dramatically lower for a given gate bias as L becomes small. This is known as roll-off (*Figure 2-33*).

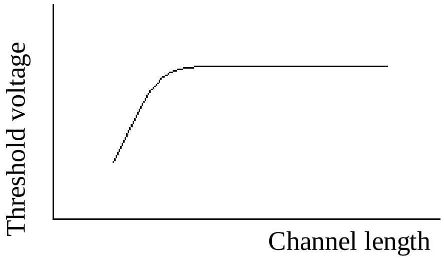

Figure 2-33: Threshold voltage variation as a function of channel length

For a given channel length if the drain bias is increased the depletion regions around the source and drain are increased in size and the threshold voltage will also exhibit roll-off characteristics (*Figure 2-34*).

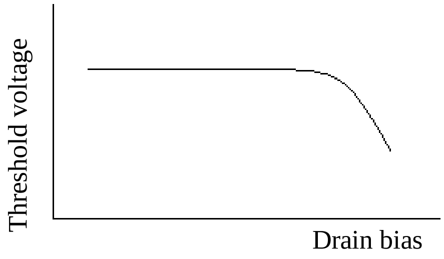

Figure 2-34: Threshold voltage variation as a function of drain bias

When the dimensions are reduced to a critical value, it is possible for the source and drain depletion regions to merge and this is known as ‘punchthrough’. The gate is no longer in control of the channel and is a major limitation to device scaling.

2.7.2 Drain Induced Barrier Lowering

In essence Drain Induced Barrier Lowering (DIBL) is a reduction in the gates control over the current flowing in the channel. When the drain source bias is increased the depletion region at the drain end of the channel becomes larger. This reduces the height of the energy barrier that carriers need to overcome in order to flow along the channel. For long devices (*Figure 2-35 (a)*) the increased bias has little effect on the barrier height at the source end, however, for short devices (*Figure 2-35 (b)*) it can dramatically reduce the height of the energy barrier ($\text{DIBL}_1 << \text{DIBL}_2$). This reduces the threshold voltage and can result in current flow when the device is in the ‘off’ state.

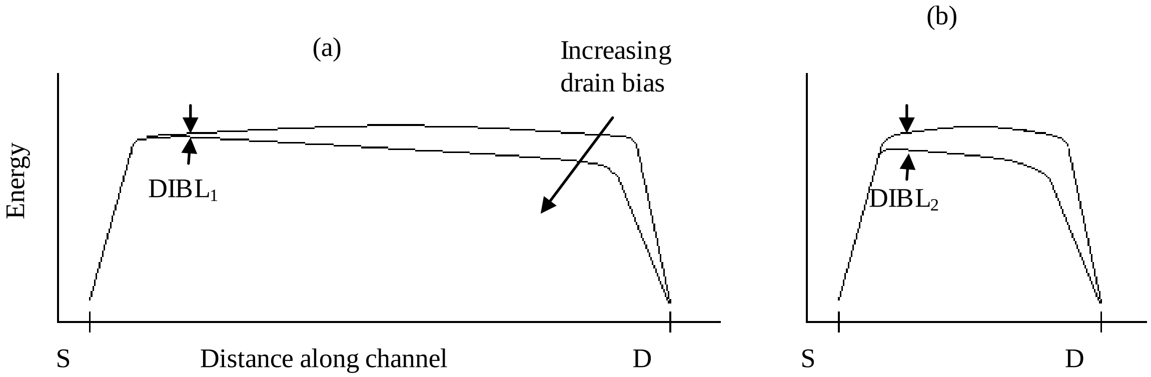

Figure 2-35: Energy versus length along the devices channel in (a) a long channel device and (b) a short channel device

2.7.3 ‘Hot Electrons’

As mentioned earlier, as the device dimensions are reduced the electric fields will increase unless the supply voltages are also lowered. As this has not happened, the high electric fields present is short devices accelerates the carriers to high energy states. The large values of kinetic energy now associated with the carriers allow them to bury themselves in the gate oxide creating excess charge in this region. The flatband voltage is affected and the I-V characteristics of the device will change over time.

These ‘hot electrons’ also deteriorate the device by breaking bonds in the semiconductor/oxide interface. Overall the device looses its intended functionality.

Devices are designed in an attempt to maintain similar electric fields in all areas to prevent localized cases of ‘hot electron’ damage. One way of achieving this is through the use of the Lightly Doped Drin structure (LDD).This involves the use of a drain extension extending from the main drain region. The concentration of the doping in the extension region is an order or two lower in magnitude. A larger depletion region can therefore exist at the drain and the electric field here is reduced. For symmetry this extension must also be included at the source end however it can lead to high values for the series resistance.

2.7.4 Velocity saturation

The mobility for charge carriers is given by *Equation 2.11* and can be related to the carrier drift velocity by the equation

$$v = \mu E_{lat} \tag{2.17}$$

Field

As the devices get smaller and the electric fields increase in magnitude the carrier drift velocity will also increase. From *Equation 2.17* we could assume that the charge carriers can be accelerated to any speed providing a large enough electric field could be applied but this is not the case. The drift velocity actually saturates (*Figure 2-36*). This begins to occur around the carrier thermal velocity, v_{th}, given by

$$v_{th} = \sqrt{\frac{3k_B T}{m^*}} \tag{2.18}$$

Field

Any extra energy from the electric field after this point goes into heating the semiconductor lattice and scattering rates rise due to an increased density of states. So saturation occurs because the charge carriers loose energy as fast as they can gain it from the field.

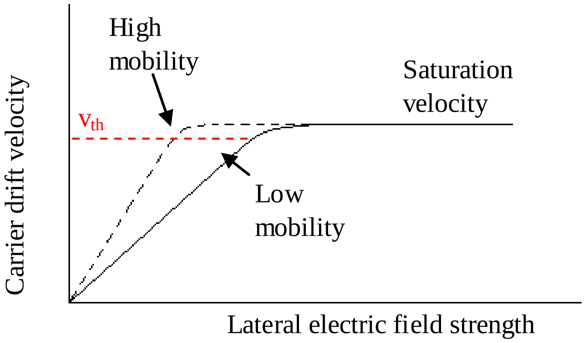

Figure2-36: Carrier drift velocity versus the lateral electric field strength and the condition for saturation velocity

Problems arise is MOSFETs as there are generally regions of high and low electric field. The device will therefore be limited by mobility in areas of low electric field (at the source end of the channel) and by the saturation drift velocity in areas of high electric field (at the drain end).

By reducing the channel length sufficiently into the realm of short channel devices the carrier transport which is usually in thermal equilibrium with the channel, starts to break down. This is because there is a small finite relaxation time associated with the carriers and when the channel length is very short the carriers do not have time to come to equilibrium with the lattice. This phenomenon is known as ‘ballistic overshoot’ (*Figure 2-37*).

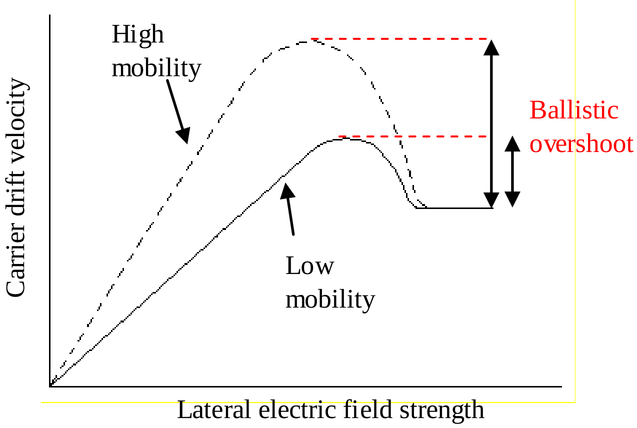

Figure2-37: Carrier drift velocity versus the lateral electric field strength and how a short channel length can bring lead to ballistic overshoot

This velocity overshoot will occur at the drain end of the channel where the electric fields are the highest and so shorter channel lengths help to combat limitations at high electric fields.

Strain in the channel region helps to reduce the limitation at all electric fields. While it has been shown that strain introduced into the channel region doesn't affect the saturation velocity of carriers [Roldan 1997] the increased mobility will increase the carrier velocity at the source and the drain end of the device.

2.7.5 Gate leakage

Oxide thickness' (T_{ox}) for the current technology nodes are 12-16Å (1.2-1.6nm), which is approximately 4-5 atomic layers of SiO_2. The thin oxide is desirable to obtain a substantial current drive at lower voltages and to help alleviate SCEs such as DIBL.

Gate leakage (I_G) occurs as there is a finite possibility of an electron tunnelling through the SiO_2 dielectric layer. The probability of tunnelling and hence I_G is a strong exponential function of T_{ox} and the voltage potential across the gate. A 2Å difference in oxide thickness can lead to an order of magnitude change in I_G. The problem is typically an order of magnitude less for pMOS devices compared to nMOS as holes in the latter devices require a higher energy to tunnel and thus there are fewer holes present for the same T_{ox} and V_g.

'High-k' dielectrics have been proposed as a solution to the problem of gate leakage. The same capacitance as a thin layer of SiO_2 can be achieved with a thicker layer of the high-k material, potentially reducing the leakage. Fabrication issues predominantly to do with the poor quality interfaces that are formed still need to be solved before these materials are introduced into CMOS production lines.

2.8 Chemical Mechanical Polishing

Due to the nature of growing a relaxed layer of SiGe on a Si substrate dislocations are generated. These dislocations can interact [Nash 2005] and propagate to the surface creating a certain degree of surface roughness (*Figure 2-38(a)*). The rms height (Δ) is of the order of tens of nanometers and in some cases visible with the use of an optical microscope. This surface roughness causes carriers to scatter and a method known as Chemical Mechanical Polishing (CMP) was first introduced in 1998 to try and smooth the surface [Currie 1998]. Once the graded section of a VS has been deposited a CMP stage is normally performed before the uniform layer of SiGe is grown. This removes the 'cross hatch' pattern (*Figure 2-38(b)*). As the uniform layer of SiGe is then grown on top, the threading dislocations can glide again, so no new ones need to nucleate. This reduces the threading dislocation density that reaches the top surface and the roughness can be halved. It is important to use a low thermal budget for the rest of the device fabrication as elevated temperatures can reintroduce the crosshatching.

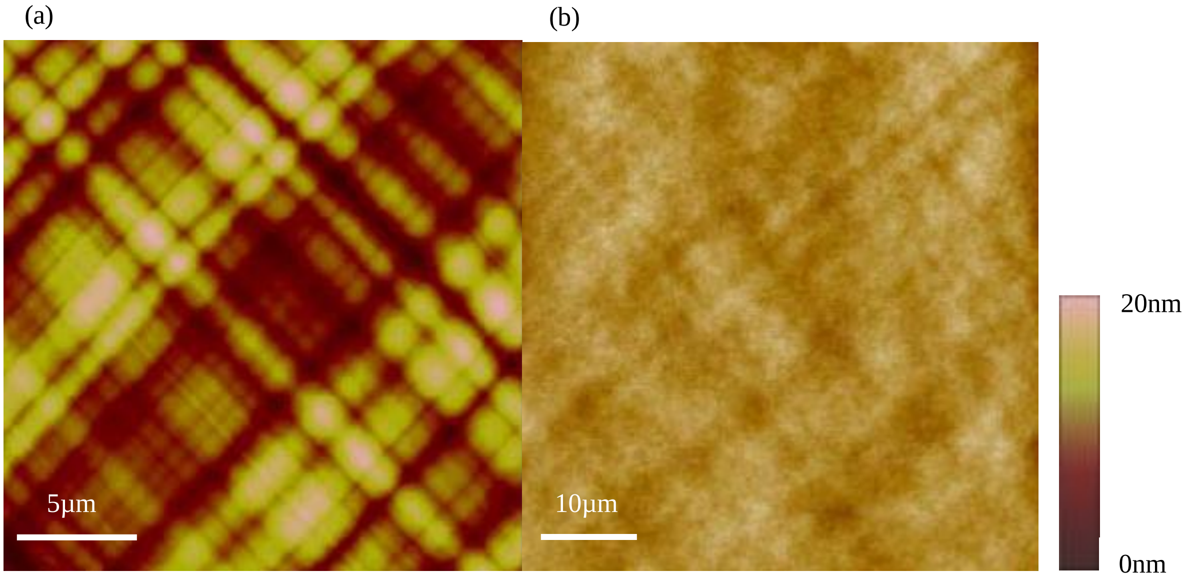

Figure 2-38: (a) Cross hatching clearly visible prior to CMP stage, (b) after the CMP stage the surface roughness has been reduced

3 Experimental Method

3.1 Electrical characterisation

Two batches of devices manufactured by ST Microelectronics (France) were characterised. Room temperature I-V and C-V measurements were performed predominantly on an *Agilent 4156C precision semiconductor parameter analyzer*. A *Keithley 590 CV analyser*, *Keithley 230 programmable voltage source* (only required if testing exceeds a ± 20V range) combined with a *Keithley 5951 remote input coupler* setup was used for high frequency capacitance measurements. The high frequency equipment was controlled by Capital Equipment Corporation's Testpoint™ control software. This was first developed by Dr Martin Prest and later modified over the years by other member of the Warwick University Nano-Silicon research group.

The Device Under Test (DUT) was housed in a Karl Suss probing station that is an earthed Faraday cage designed to eliminate electromagnetic interference. The probing station also maintains a completely black environment preventing the generation of electron-hole pairs which can lead to increased leakage current in the 'off' state. Four contact points were made with each of the devices. Surface contacts were made with 0.02" Tungsten needles (American Probe & Technologies Inc, 0.14-0.16" taper, 1μm radius, 1.25" length) if the Substrate contact was on the bottom of the wafer this was connected via the metal chuck onto which the DUT was placed. A Karl Suss membrane vacuum pump secured the wafer in all setups. The contacts were connected to the testing equipment with biaxial cables. A microscope was used for positioning.

In an attempt to prevent damage to the devices that had very thin oxides, care was taken when handling any sample. This involved using plastic tweezers to hold samples to prevent static discharge from metal ones. Another precaution was to only use a very low light setting when connecting the DUT and this was turned off before testing began. To prevent a debilitating discharge damaging the thin oxide, contacts were always made in the order:

Substrate → Source → Drain → Gate

After the tests were run the contacts were removed in the reverse order.

To ensure accurate results calibration of the Keithley equipment was performed using a precision standard capacitor and coaxial probes in order to minimize the parasitic capacitance.

3.1.1 I-V measurement

Two types of I-V measurement were carried out:

I_D-V_G: The source and substrate were grounded and a constant bias was applied across the source and drain. The drain current was plotted as a function of a variable gate bias. Using a low source-drain bias (V_D=50mV) the linear region of operation was observed where there was a constant electric field along channel and therefore a constant carrier population. By applying a high bias the device entered the saturation region of operation (V_D=1.5V) replicating CMOS performance in a real circuit where the source and drain are biased at power supply levels.

I_D-V_D: The gate was held at a constant potential with respect to the grounded source and substrate. The drain current was observed as a function of the drain source voltage. By repeating over a range of gate biases a family of I_D-V_D curves could be built up for a single device.

By performing both types of I-V measurement it was possible to extract a large number of device parameters such as the threshold voltage, subthreshold slope, and to determine if any self heating had occurred in the channel. Combined with C-V data, mobility could also be extracted.

3.1.2 Split C-V measurement

By examining the induced current in the source/drain contacts and the substrate contact separately it was possible to calculate separately the mobile channel charge density and the bulk charge density. If the source/drain capacitance is being measured the substrate contact is earthed. If the substrate capacitance is being measured the source/drain contact is earthed.

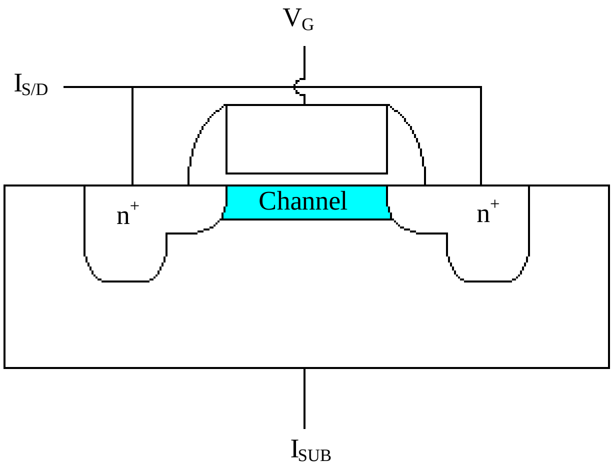

Figure 3-1: Schematic for contacts required to obtain split C-V n-type MOSFET data

Two curves were observed dependant on where the current was measured (*Figure 3-2*). When the channel was populated by the minority carriers then increasing the bias more positively pulled more electrons in from the source and drain which have a plentiful supply of electrons. This resulted in a current flow in the $I_{S/D}$ contact while the I_{Sub} contact would have no current flowing through it due to the depletion region screening. When the channel was not populated by minority carriers and was therefore in the accumulation or depletion state, changing bias caused majority carriers to move from the substrate region, leading to a current on I_{Sub} and no current flowing on $I_{S/D}$.

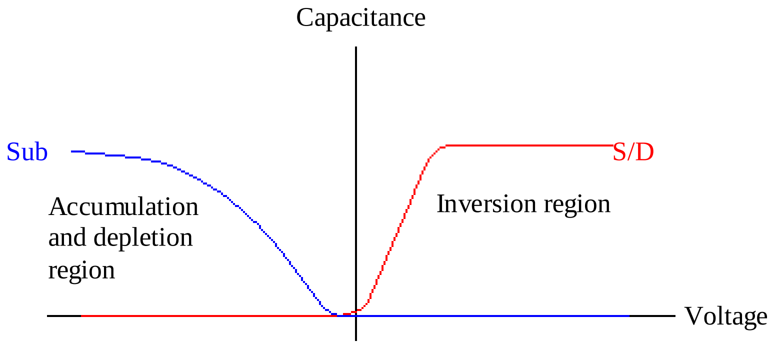

Figure 3-2: Split C-V characteristics for a n-type MOSFET

3.2 Nomarski Imaging

Optical microscopes, limited by the Raleigh criterion, are not capable of resolving the tiny features present on a typical mask structure. Differential interference contrast or Nomarski imaging (a modification of phase contrast microscopy) allows for such small features to be viewed.

A Reichart-Jung Nomarski interference microscope coupled with a Nikon DN100 digital camera was used. The light illuminating the object was plane polarised then split into two beams which passed over the sample. After reflection the beams were then recombined prior to being directed to the eye piece or digital camera. Any height differences caused the beams to be out of phase with each other upon recombination. The phase changes were converted to amplitude changes providing the image contrast. Features in the order of nanometres can be resolved; however no quantitative information relating to the height of features was given.

3.3 Device characterisation

The main criteria used to evaluate modern day MOSFETs are:

- Effective mobility & Oxide thickness
- Threshold voltage
- Onset of SCEs
- I_D-V_G characteristics
- Subthreshold slope
- I_D-V_D characteristics
- On current
- Off current
- Source/drain resistance
- Effective channel length
- Transconductance

These parameters can all be obtained from I_D-V_G, I_D-V_D and C-V data. The methods for extracting the parameters are described in the following sections of this chapter.

3.3.1 Threshold voltage

The threshold voltage of a device is very important and needs to be calculated as accurately as possible. It specifies the gate voltage that needs to be applied in order to set up an inversion region in the channel and turn on the device. The physical meaning of threshold voltage is given in *Chapter 2.2* by $e\phi_s = 2e\phi_f - V_{Sub}$ and with the possibility of a strain element the threshold voltage becomes

$$V_T \propto 2e\phi_F - V_{sub} - g(\sigma) \qquad (3.1)$$

Field

There are a number of methods that have been developed to calculate $V_{T,}$ the simplest being the Constant Current method (CC). A current is chosen based on the device dimensions and the device is said to be ‘on’ when that current has been reached. The current is empirically defined as:

$$I_D = \frac{W}{L}.10^{-7} \qquad (3.2)$$

Field

This method is only really suitable for long channel devices, as short channel devices tend to have a long tail on their I_D-V_G curve. Other methods include [Tsuno 1999] [El-Kareh 1990]:

- Linear extrapolation method (LE) & Square-root-of-current extrapolation method (SRE)
- Exponetial-extrapolation of subthreshold current method or match point method (MP)
- Second derivative of the logarithm of drain current method (SDL)
- Transconductance change method (TC)

None of these methods satisfied the requirement that V_T could be found for all MOSFETs down to sub 0.1μm dimensions; gave V_T as defined correctly for inversion and/or were relatively easy to use. However the transconductance g_m-linear extrapolation method (GMLE) allowed the above criteria to be met and was the method used to extract V_T for this work. Using the transconductance and differential of the transconductance, the threshold voltage is defined by the GMLE method as:

$$V_T \equiv V_{G(\max)} - \frac{g_{m(\max)}}{\mathrm{Diff}_{(\max)}} \tag{3.3}$$

Field

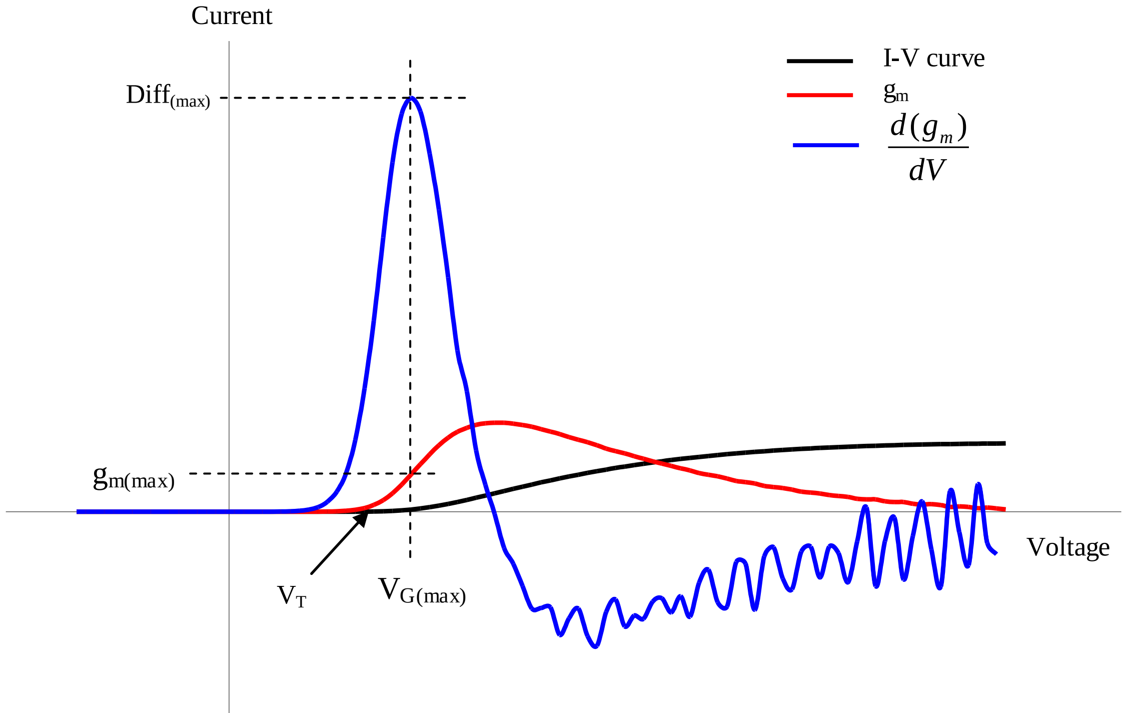

Figure 3-3: Graphical representation of the GMLE method for calculating the MOSFET threshold voltage

Noise associated with the second derivative leads to errors in the determination of V_T. Even weak noise contamination (i.e. small perturbations) of the raw data points can result in a highly unstable and oscillatory shape of the differentiated data curve. A regularization technique has been proposed [Choi 2004] in order to eliminate the noise effects brought about through the differentiation process. However these methods have not been used in this work, as noise contamination was not present in the critical voltage range used to derive V_T.

Once the threshold voltage was extracted a 4 parameter sigmoid equation of the form

$$y = y_0 + \frac{a}{1 + e^{-(x - x_0)/b}} \tag{3.4}$$

was used to create a line of best fit for the data. This allowed the threshold voltage, prior to the onset of SCEs, to be calculated accurately.

To determine the channel length at which SCEs are causing the threshold voltage to 'roll-off' *Equation 3.4* was rearranged giving

$$x = x_0 + b \ln\left[\frac{y - y_0}{a + y_0 - y}\right] \tag{3.5}$$

The condition chosen for the definite onset of the SCEs was when the threshold voltage changed by 10mV from its value at long channel lengths.

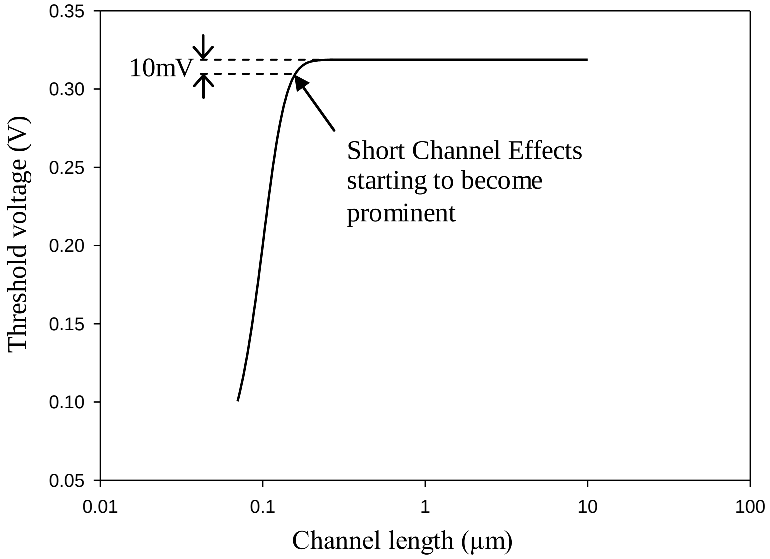

Figure 3-4: An example of threshold voltage data for similar devices over a range of channel lengths and the point at which the onset of SCEs has been defined

3.3.2 On/Off current

The ‘on current’ (I_{on}) was derived from graphs of log(I_D) as a function of V_G. The value, $I_{on,}$ is defined as the highest value for

$$\frac{d(\log(I_D))}{d(V_G)} = 0 \tag{3.6}$$

Field

The devices tested did not reach saturation in the gate voltage range used and so the on current has been taken at a gate overdrive (V_G-V_T) of 1.5V.

The ‘off current’, $I_{off,}$ is defined as the value of I_D, when V_G=O. The strained devices showed large threshold voltage shifts and so the off current has been taken instead at zero overdrive voltage.

3.3.3 Subthreshold slope

The subthreshold slope, *S*, (units of mV/dec) looks at the behaviour in the linear region between I_{on} and I_{off} for an I_D-V_G curve, and is defined as:

$$S = \left[\frac{d(\log(I_D))}{d(V_G)}\right]^{-1} \tag{3.7}$$

Field

The subthreshold slope allows the variation in gate voltage necessary to produce a one decade change in the drain current to be calculated. A surface channel device, like those discussed in this thesis, has a theoretical minimum value of 60mV/dec at 300K and is given by

$$S = \ln(10).\frac{nk_BT}{e} \tag{3.8}$$

Field

With $n = 1 + (C_{dep} + C_{it})/C_{gc}$

3.3.4 Source/drain resistance & Effective channel length

The total resistance of a MOSFET is the sum of the resistance in the channel and the resistance of the source and drain regions

$$R_m = R_{ch} + R_{S/D} \tag{3.9}$$

Field

$R_{S/D}$ is a combined series resistance. It is a major component of the total resistance in a MOSFET when heavily doped semiconductor regions are used under the metal contacts. This is due to the difficulty carriers have when moving from the doped region to the inversion layer in the channel. Any additional LDD structures cause a massive increase in this resistance. The metal contacts contribute a small amount to the resistance as does the Schottky barrier. The drain current will be adversely affected by the source drain resistance.

The resistance of the source and drain regions should remain constant during operation, however the channel resistance will change as it is a function of the applied gate bias

$$R_m = \frac{L - \Delta L}{W \mu C_{ox} (V_G - V_T)} + R_{S/D} \tag{3.10}$$

Field

The channel resistance introduces the term for the effective length of the channel

$$L_{eff} = L - \Delta L \tag{3.11}$$

Field

The effective channel length will be different from the written channel length, L, by an amount ΔL. This accounts for any process bias such as:

i. Print bias
ii. Etch bias
iii. Lateral diffusion of source/drain dopants

For large devices the value of ΔL is not as important as it is only a fraction of the effective channel length. However the characteristics of small-scale devices are more sensitive to channel length variation.

The Terada-Muta technique involves using a double regression method (*Figure 3-5*) to extract both ΔL and source /drain resistances for a set of devices of varying written channel lengths for a given channel width. I_D-V_G data for devices of the same type (i.e nMOS, '*High speed*', Si control), operating in the linear regime with different gate lengths, were used. The total channel resistance for each device at a range of gate biases was plotted. Linear regression was applied to the data points with the same gate bias. From *Equation 3.10*, it can be seen that when V_G is small, then $R_m = R_{ch} + R_{S/D}$. When V_G is large, then $R_m \approx R_{S/D}$. The points at which the regression lines intercept gives $R_{S/D}$ and ΔL (*Figure 3-5 (a)*). Rather than attempting to read these values off the graph a second regression can be done using as input values these initial slopes and intercepts. The equation of this new best fit line provides the required two values, $R_{S/D}$ and ΔL (*Figure 3-5 (b)*).

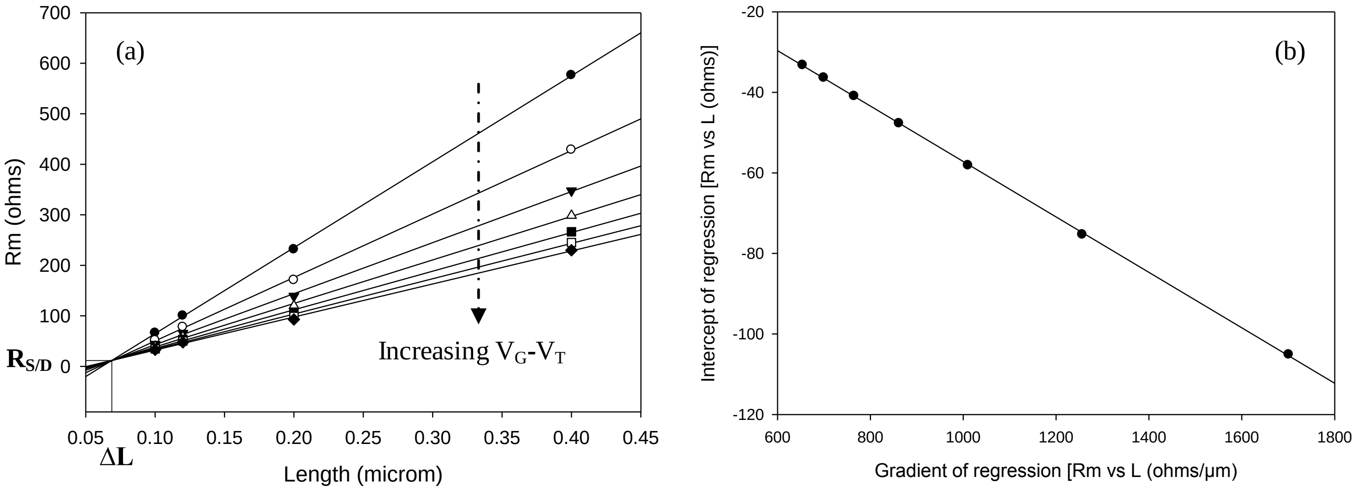

Figure 3-5: Representative plots of (a) resistance versus length and (b) the second regression used to find the source drain resistance and effective length of a MOSFET

It was essential to use sharp clean probe needles in collecting the raw data. This was because the probe to metal pad resistance would be included in the $R_{S/D}$ values, thus needed to be kept to a minimum. During the performed analysis it was possible to achieve an overall correlation coefficient of greater than 0.999 for the regression data.

3.3.5 Effective mobility

The mobility is obtained as a function of $V_G/N_s/E_{eff}$ from a combination of I_D-V_G and split C-V measurements. The drain current in relation to the mobile charge current density is

$$I_D = \frac{W\mu_{eff}Q_nV_D}{L} - W\mu_{eff}\frac{dQ_n}{dx} \qquad (3.12)$$

Field

with μ_{eff} (effective mobility as this is the average mobility of carriers along the channel length) measured at low drain voltage (V_D = 50mV) as at such values it is possible to assume the channel charge to be uniform from the source to the drain and hence the derivative is zero.. Then rearranging the equation gives the following expression for effective mobility

$$\mu_{eff} = \frac{L}{WQ_n}\frac{I_D}{V_D} \qquad (3.13)$$

Field

The mobile channel charge density can be approximated by $Q_n = C_{ox}(V_G - V_T)$ however this relies on the threshold voltage being accurately measured. It was therefore better to obtain a direct measure from the $I_{S/D}$ split C-V data and perform an integration to find the area under the curve where

$$Q_n = \int_{-\infty}^{V_{GS}} C_{GC}dV_G \qquad (3.14)$$

Field

The mobile channel charge density was calculated from the split C-V data using the trapezium rule (*Figure 3-6*) as an approximation for the integration with

$$Q_n = \frac{1}{2}h\sum_{1}^{j}(f_j + f_{j+1}) \qquad (3.15)$$

Field

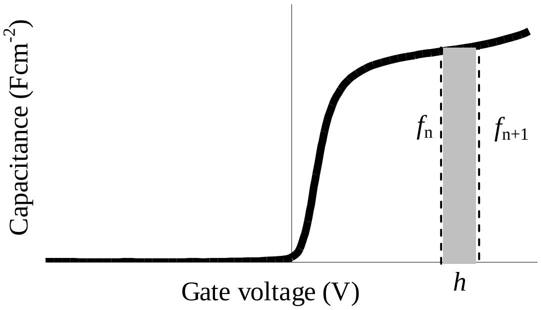

Figure 3-6: Source/drain capacitance versus gate voltage for an nMOS device

This allowed the effective mobility to be plotted as a function of the gate voltage and the mobile channel charge density.

The effective electric field was extracted using the relationship

$$E_{eff} = \frac{Q_b + \eta Q_n}{K_s \varepsilon_0} \tag{3.16}$$

Field

For electrons η=1/2 and holes η=1/3, and accounts for the averaging of the electric field over the electron distribution in the inversion layer. The bulk charge density is calculated in a similar way as for Q_n from the I_{sub} split C-V data (*Figure 3-7*) using

$$Q_b = \frac{1}{2} h \sum_{1}^{j} (f_j + f_{j+1}) \tag{3.17}$$

Field

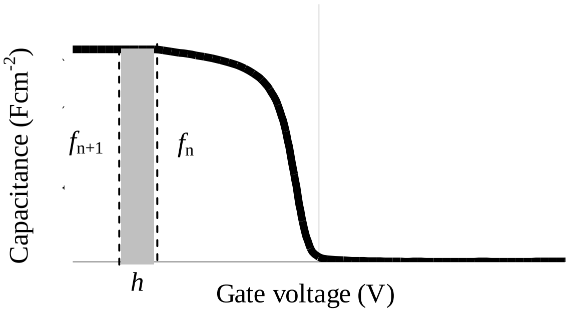

Figure 3-7: Substrate capacitance versus gate voltage for an nMOS device

3.3.6 Oxide thickness

The oxide thickness can be calculated from C-V data along with the dimensions of the device.

$$T_{ox} = \frac{\varepsilon_0 \varepsilon_d}{C_{ox}}.10^7 \qquad (3.18)$$

Field

Where ε_d = 3.9 for SiO_2/nitride oxide, and the 10^7 component of the equation gives the thickness in nm.

4 Wafer fabrication & Specifications

4.1 STMicroelectronics ‘2005’ batch

The bulk of this research has been performed on a wafer batch made at STMicroelectronics in France under the supervision of Fabrice Payet. The wafers allowed for a large number of comparisons to be made for both nMOS and pMOS devices at short channel lengths (L=1μm→70nm). In normal CMOS devices the pMOS devices are laid down with a wider channel so that drive currents can be matched. All devices tested in this batch and the ‘2002’ batch, have an identical width of 10μm. Three wafer forms were used:

1. Standard Si control
2. $Si_{0.8}Ge_{0.2}$ Virtual substrate induced S-Si
3. $Si_{0.8}Ge_{0.2}$ Virtual substrate combined with a nitride cap (CESL) induced S-Si

For each form of Si three slightly different doping schemes were used in an attempt to optimize device operation (*Figure 5-1*).

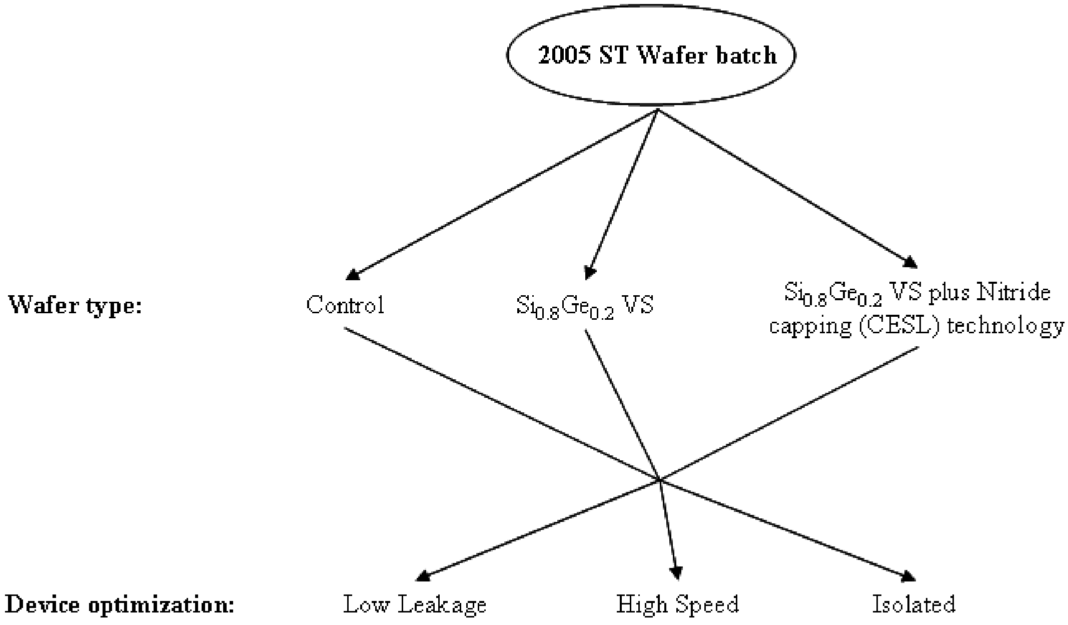

Figure 4-1: Wafers under test in ‘2005’ batch

The ‘*Isolated*’ devices are doped as described in *Table 4-1* and each device has independent contacts. The ‘*Low Leakage*’ devices have received additional doping compared to the ‘*Isolated*’ devices to increase the threshold voltage so that at zero gate bias there is minimal drain current (for low power applications), while the ‘*High Speed*’ devices have been additionally doped to lower the threshold voltage (for high performance applications). For a given gate voltage the ‘*High Speed*’ devices will then give the highest drive current. These shifts in the threshold voltage are illustrated in *Figure 4-2*. The specifics of these additional V_T implants are kept in confidence by STMicroelectronics.

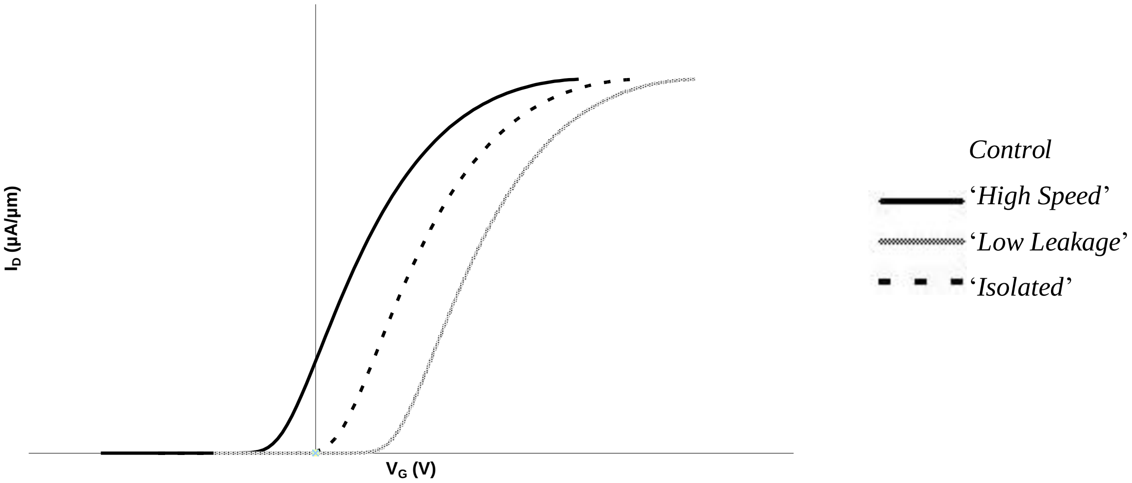

Figure 4-2: The effect of a device optimization shift in the threshold voltage for high performance and low power applications

The ‘*Low Leakage*’ and ‘*High Speed*’ devices have shared contacts. Therefore the ‘*Isolated*’ devices are expected to exhibit lower source/drain resistances because of the shorter interconnects between the actual MOSFET under test and the contact pads

The schematics of the devices on the three wafers can be seen in *Figure 4-3*.

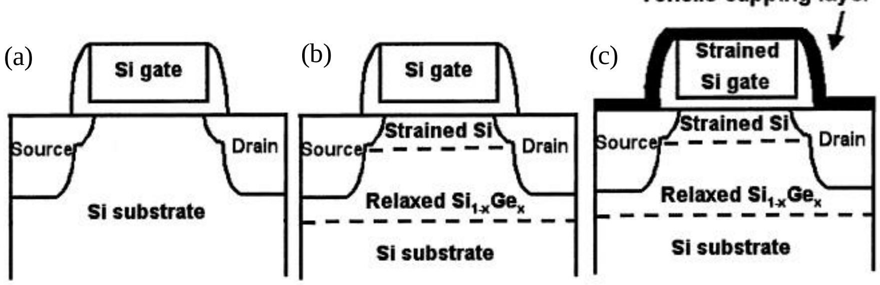

Figure 4-3: Vertical profiles of (a) control, (b) $Si_{1-x}Ge_x$ induced strain, (c) combined $Si_{1-x}Ge_x$ and process induced by CESL strain devices [Lim 2004]

4.1.1 Device processing

All devices were fabricated on a wafer orientated in the (100) direction such that the gate lengths were in the <110> direction. To create the strained devices the following main growth and processing steps were undertaken.

Wafers 2 & 3:

1. VS formed with final Ge content of 20%
2. Si epitaxial stage depositing 15nm (3nm consumed in later processing steps)
3. nitride gate oxide deposited by Plasma Nitration (PN) at a thickness of 1.2nm at 900-950^{0}C

Wafer 3 only:

4. PECVD deposition of 80nm CESL with a tensile strain of 960MPa
5. Ge ion implantation over pMOS devices

An oxide nitride has been used rather than the more standard SiO_2 in an attempt to reduce leakage currents. The difference in the value of the dielectric constant between these two materials is considered negligible. A NiSi salicidation process has been used along with TiN contacts in an attempt to reduce source/drain resistances.

In order to avoid problems due to dislocations and cross hatching forming as well as unwanted Ge diffusion, a reduced thermal budget scheme had been implemented (*Figure 4-4*). Solid blue lines indicate no change from normal thermal budgets. Stages that have been reduced in terms of the thermal budget are indicated by a blue dotted line (old scheme) and a solid red line (new scheme). The second stage has been omitted altogether for these wafers.

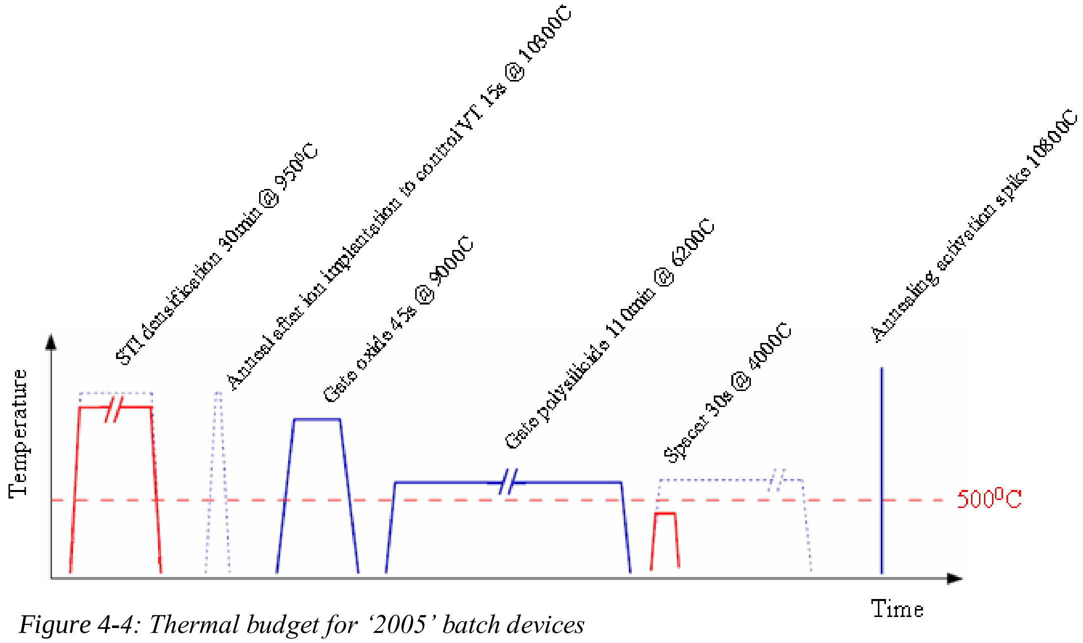

Figure 4-4: Thermal budget for '2005' batch devices

The doping stage to create the source and drain regions was performed towards the end of the processing. The regions that were doped are illustrated in *Figure 4-5* with the numerical details in the *Table 4-1*. These values correspond to the '*Isolated*' devices, the '*High Speed*' and '*Low Leakage*' devices having undergone additional V_T optimisation steps the details of which are confidential. The dopant concentrations and implant energies used were different for the Control (in black) and the strained wafers (in red).

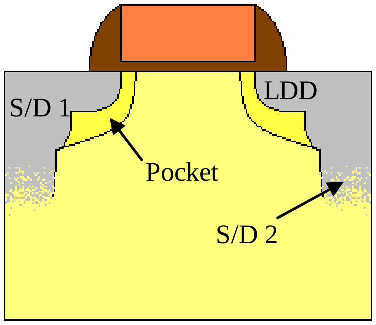

Figure 4-5: Dopant regions in the MOSFET

	pMOS		nMOS	
	Dopant	Specification	Dopant	Specification
V_T optimization	As	5x10^{12} cm^{-2} @ 60KeV 2x10^{12} cm^{-2} @ 60KeV	B	7.5x10^{12} cm^{-2} @ 8KeV 1x10^{13} cm^{-2} @ 8KeV
LDD implantation	B	4x10^{14} cm^{-2} @ 1KeV 8x10^{14} cm^{-2} @ 1KeV	As	6.5x10^{14} cm^{-2} @ 1KeV 4x10^{14} cm^{-2} @ 1KeV
Pocket implantation	As	3x10^{13} cm^{-2} @ 50KeV 3x10^{13} cm^{-2} @ 40KeV	BF_2	4x10^{13} cm^{-2} @ 30KeV 3x10^{13} cm^{-2} @ 30KeV
S/D 1 implantation	B	2x10^{15} cm^{-2} @ 1KeV 2x10^{15} cm^{-2} @ 2KeV	As	2x10^{15} cm^{-2} @ 10KeV 2x10^{15} cm^{-2} @ 5KeV

Table 4-1: Dopant scheme for 'Isolated' devices

Different implantations were used for a number of reasons. As described in *Chapter 2.4.4*, the dopants diffuse at varied rates in the VS. In the case of the pMOS devices the LDD dose has been increased as the normal dose does not diffuse enough. The V_T implant has been altered in an attempt to compensate for threshold voltage shift due to the Si/SiGe band offset. Alterations have also been made because of the reduced thermal budget used. Doping of the S/D 2 region is normally done in order to suppress short channel effects but has been omitted as the diffusion issues result in further degradation of the threshold voltage.

4.1.2 Device specification

A detailed vertical profile for the SiGe VS can be seen in *Figure 4-6*.

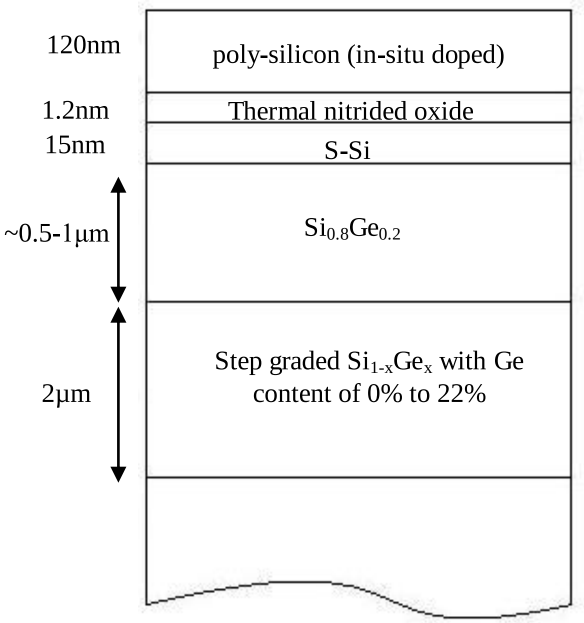
120nm
poly-silicon (in-situ doped)
1.2nm
Thermal nitrided oxide
15nm
S-Si
~0.5-1µm
$Si_{0.8}Ge_{0.2}$
2µm
Step graded $Si_{1-x}Ge_x$ with Ge content of 0% to 22%

Figure4-6: Vertical profile for the SiGe VS and gate structure

All values stated here are nominal. TEM analysis was performed [Payet 2005] prior to device fabrication on the VS and the actual thickness of the $Si_{0.8}Ge_{0.2}$ layer was found to be 500nm (*Figure 4-7*).

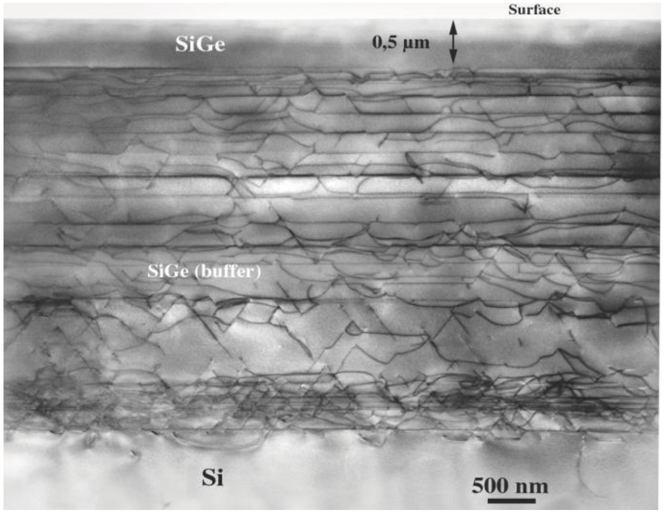

Figure 4-7: TEM image of the VS showing an extensive dislocation network [Payet 2005]

Further TEM analysis indicated that the S-Si layer had been reduced to 12nm due to the processing stages required to create the devices [Payet 2005]. Global TEM allowed the entire device to be seen (*Figure 4-8*). The amount of S-Si that can theoretically be grown on top of a 20% VS is minimal (*Figure 2-11*). In order to reduce S-Si consumption of these devices the STI (shallow trench isolation) was fabricated directly into the SiGe relaxed layer prior to the S-Si epitaxy stage.

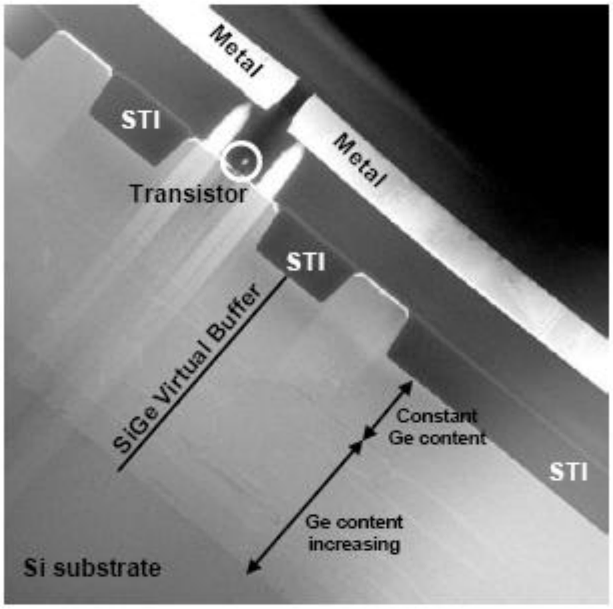

Figure 4-8: Global TEM of the MOSFETs under test [Payet 2005]

4.2 STMicroelectrics ‘2002’ batch

This batch also from STMicroelectronics formed a major part of a colleague’s PhD thesis [Nicholas 2002] in which it was fully characterised. In an attempt to see how far virtual substrates have developed over the last 3 years I have fully characterised these devices again at room temperature to compare them to the ‘2005’ batch. I have retested the devices rather than obtaining the data already collected on them in order to make an unbiased comparison and to exclude errors introduced by the use of different equipment or extraction techniques. The three wafers that comprise this batch (*Figure 4-9*) have nMOS and pMOS devices on them with short channel lengths (L=1µm→70nm). The two strained wafers are under biaxial strain from a SiGe VS.

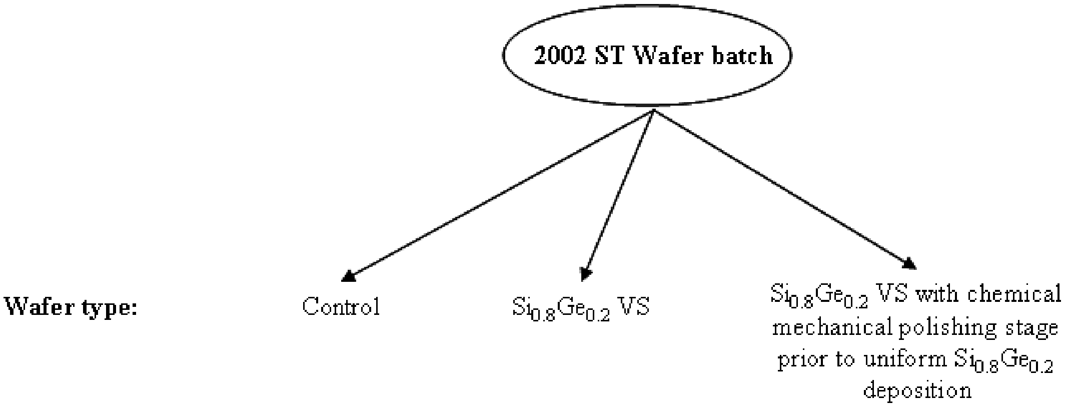
2002 ST Wafer batch
Wafer type:
Control
$Si_{0.8}Ge_{0.2}$ VS
$Si_{0.8}Ge_{0.2}$ VS with chemical mechanical polishing stage prior to uniform $Si_{0.8}Ge_{0.2}$ deposition

Figure 4-9: Wafers under test in the '2002' batch

The third wafer has undergone a CMP stage after the graded layer of SiGe has been deposited in an attempt to prevent dislocations moving into the top uniform section. The schematics of the devices on the three wafers can be seen in *Figure 4-10*.

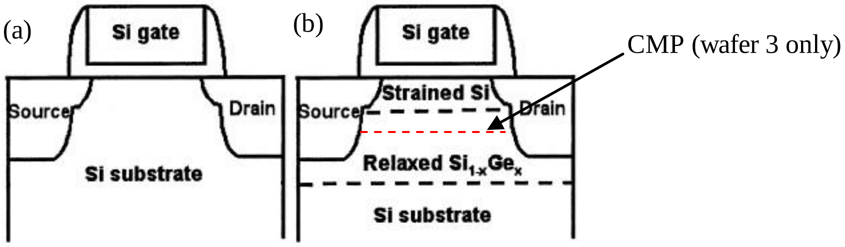

Figure 4-10: Vertical profiles of (a) control, (b) $Si_{1-x}Ge_x$ induced strains wafers including the additional CMP stage for wafer 3 [Lim 2004]

4.2.1 Device processing

All devices were fabricated on a wafer orientated in the (100) direction such that the gate length are in the <110> direction. To create the strained devices the following main growth and processing steps were undertaken.

Wafers 2 & 3:

1. VS graded buffer with final Ge content of 20%

Wafer 3 only:

2. CMP

Wafer 2 & 3:

3. 40nm uniform VS with a Ge content of 20%
4. Si epitaxial stage depositing 15nm
5. Well implantation
6. Channel implant
7. Gate deposition
8. MDD implant
9. S/D implant
10. RTA anneal (Si 30s @ 900^{0}C, S-Si 40s @ 850^{0}C)

The contact pads are made from aluminium.

As for the '2005' batch we see in *Table 4-2* that a slightly different dopant concentration and implant energy was used for the strained wafers (in red).

	pMOS		nMOS	
	Dopant	Specification	Dopant	Specification
Well implantation	P	8x10^{12} cm^{-2} @ 480keV 4x10^{12} cm^{-2} @ 120keV	B	8x10^{12} cm^{-2} @ 220keV 4x10^{12} cm^{-2} @ 50keV
Channel implant	P	10^{12} cm^{-2} @ 70keV	B	10^{12} cm^{-2} @ 30keV
Gate deposition	B	> 5x10^{19} cm^{-3}	As	> 5x10^{19} cm^{-3}
MDD implant	BF$_2$	10^{14} cm^{-2} @ 12keV	P	10keV 10^{14} cm^{-2} @ 10keV
S/D implant	B	2x10^{15} cm^{-2} @ 5keV	As	2x10^{15} cm^{-2} @ 40keV

Table 4-2: Dopant scheme

4.2.2 Device Specification

A detailed look at the vertical profile for the SiGe VS can be seen in *Figure 4-11*.

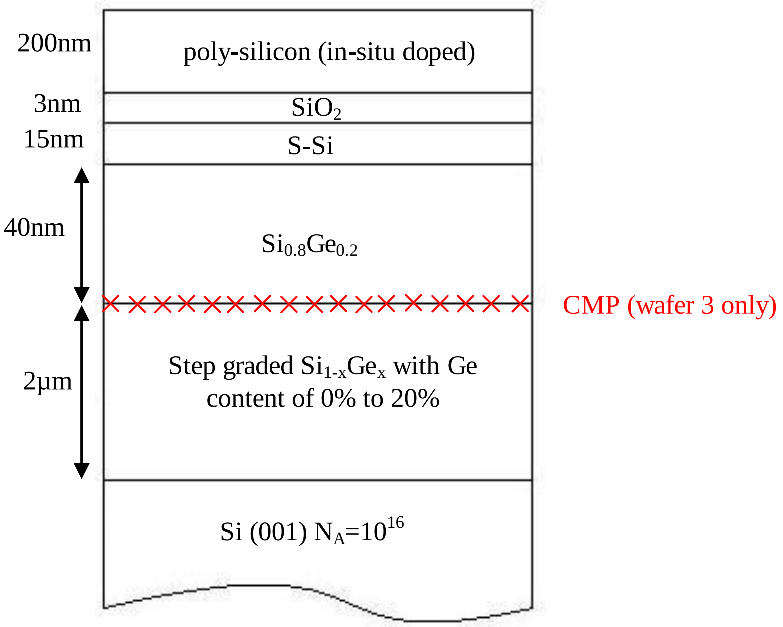

Figure 4-11: Vertical profile for the SiGe VS and gate structure

TEM analysis [Nicholas 2002] has shown that the S-Si layer thickness has been reduced to 13nm due to the device fabrication stages. The oxide thickness has been shown from C-V analysis to be ~1.6nm.

5 Device analysis

5.1 Introduction

The overall performance of a modern day MOSFET is determined by evaluating a number of device parameters and its operational behaviour. As specified in *Chapter 3.3* these are:

- Effective mobility & Oxide thickness
- Threshold voltage
- Onset of SCEs
- I_D-V_G characteristics
- Subthreshold slope
- I_D-V_D characteristics
- On current
- Off current
- Source/drain resistance
- Effective channel length
- Transconductance

These criteria have been assessed for the ‘2005’ and ‘2002’ wafers, as part of this MSc, using appropriate methods described in *Chapter 3* and the results are presented in the subsequent sections.

The mask used to create the individual devices has been repeated over the entire 8 inch wafers under test. In an attempt to obtain a fair test a minimum of 5 identical MOSFETs were tested for each device type, ie the ‘*High Speed*’, *SGN, L=200nm* device was tested at least 5 times in the centre of the wafer. It was chosen that only the centre of the wafers would be tested so that the data would not include variations that occur across a typical wafer. The outlying devices were then excluded and the remaining data averaged to provide the results presented in this chapter. The following abbreviations have been used throughout this chapter when discussing the data collected. All graphs have then been colour coordinated as listed unless labelled otherwise.

Control – devices build on a regular Si wafer. Graph data presented in **black**.

SG – devices built on a SiGe VS platform. Graph data presented in blue.

SGN – devices built on a SiGe VS platform with an additional strain introduced by a nitride layer. Graph data presented in red.

SGCMP – devices built on a SiGe Vs platform with an additional CMP stage. Graph data presented in light blue.

5.2 Nomarski Imaging

The initial step undertaken upon receipt of the samples was Nomarski imaging to provide high resolution images of the devices to be tested under this research.

Field

Field

5.2.1 '2005' batch

The '*High speed*' and '*Low leakage*' devices have the same structural form (*Figure 5-1* and magnified in *Figure 5-2*). Each transistor has a shared gate and source contact and an individual drain contact. The drain contacts numbered D_1-D_9 allow for transistors with different gate lengths to be tested.

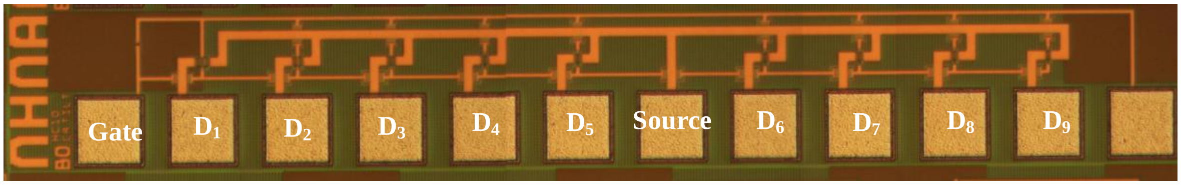

Figure 5-1: 'High Speed/Low Leakage' MOS Plan view photograph of the device layout used in this investigation

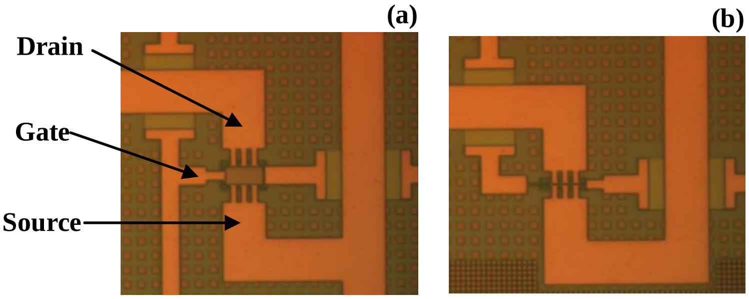

Figure 5-2: Magnified (a) 10µm x 5µm and (b) 10µm x 70nm devices from Figure 5-1

The '*Isolated*' devices have independent gate, source and drain and contacts (*Figure 5-3*).

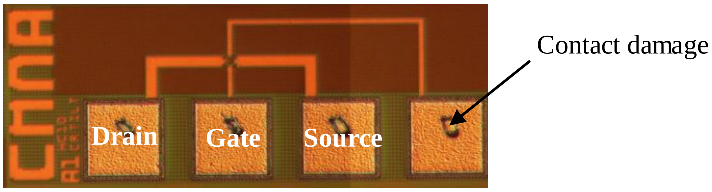

Figure 5-3: 'Isolated' MOS Plan view photograph of the device layout used in this investigation

All three devices types use the back of the wafer as a shared substrate contact. Damage caused when the tungsten probes make contact with the pads in order to test the devices can be seen in *Figure 5-3*.

5.2.2 '2002' batch

The devices were printed into groups of longer gate length devices (*Figure 5-4*) with lengths from 10µm→300nm (drain contacts D_1-D_7 and shared source, drain and substrate contacts), and groups of shorter gate length devices (*Figure 5-5*) with lengths from 300nm→125nm. No uniformity exists with respect to the contact pads for the shorter devices. Some devices having shared contacts and others have independent contacts. The substrate contact however for all the devices is made on the top surface of the wafer.

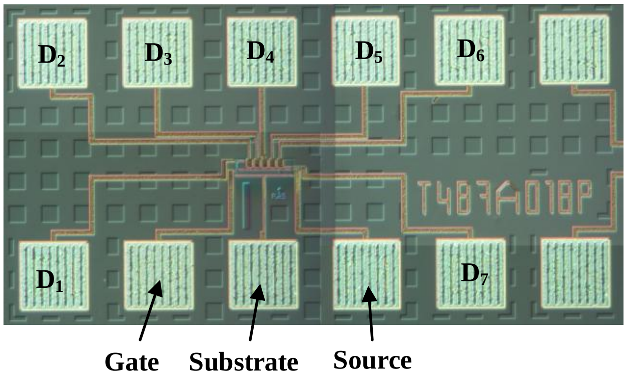

Figure 5-4: MOS Plan view photograph of the device layout used in this investigation for channel lengths greater than 300nm, contact pads not labeled are drain contacts for different devices

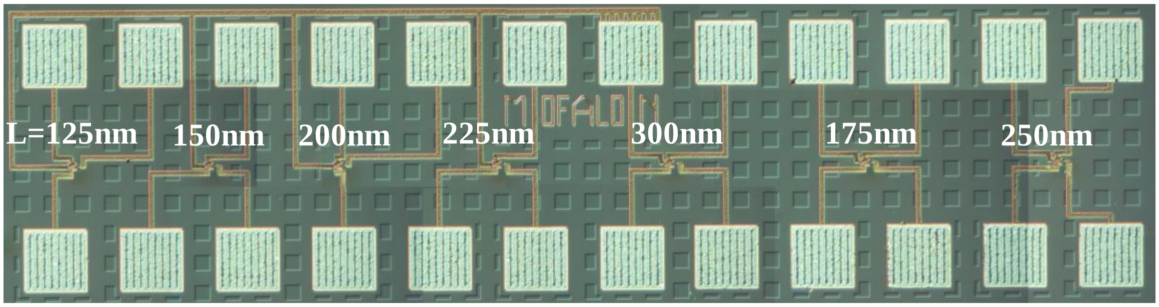

Figure 5-5: Same as for 5-4 except for channel lengths less than 300nm

5.3 Effective mobility & Oxide thickness

The mobility of the charge carriers is one of the most important parameters extracted from a MOSFET as it has a direct relationship to the current drive attainable.

5.3.1 '2005' batch

Prior to sending the '2005' batch of wafers to Warwick University, STMicroelectronics had already partially characterised them and had extracted an oxide capacitance value of 1.63x10^{-6}Fcm^{-2} [Contaret 2006]. In house attempts to obtain split C-V data from this batch failed due to high gate leakage currents that could not be compensated for with the equipment available. Therefore in order to extract the effective mobility, the oxide capacitance value obtained by STMicroelectronics along with the approximation for the mobile charge sheet density, $Q_n = C_{ox}(V_G - V_T)$, was used.

Figure 5-6 shows the effective carrier mobility for the nMOS devices. As expected the *SGN* devices for each optimization group outperformed the *SG* and the *Control*. The biaxial strain from the VS brings about a mobility enhancement of between 1.27-1.52 times at its peak over the *Control*, and the addition of an extra uniaxial strain component results in a mobility enhancement from the combined technologies of between 1.32-1.57 over the *Control* devices.

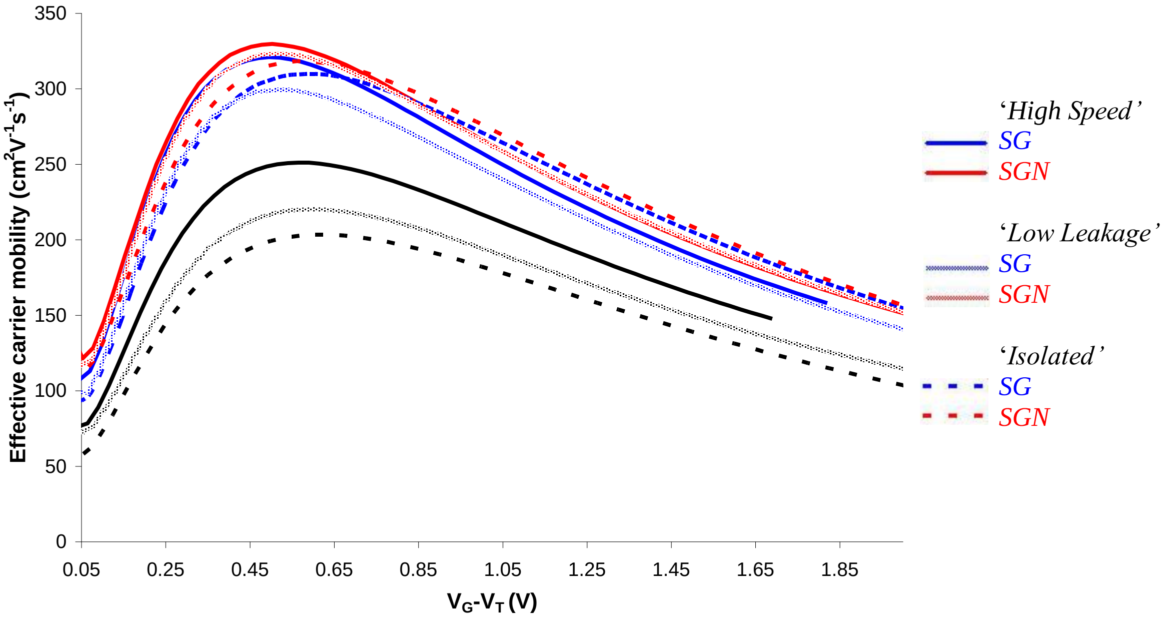

Figure 5-6: nMOS effective carrier mobility versus gate overdrive

It can be seen for pMOS (*Figure 5-7*) that for all device optimization groups the *SG* devices had a marginally lower or almost identical mobility than the *Control* samples. This is suspected to be due to the surface confinement of carriers degrading any mobility enhancement brought about by the biaxial strain. The mobility of the *SGN* devices is surprising as it was expected that they would behave like an *SG* device. This is because a Ge ion bombardment stage was used in the manufacture of the pMOS devices (*Chapter 4.1.1*) and its effect is described in *Chapter 2.5.1* (*Figure 2-17*). The pMOS *SGN* devices therefore should have a nitride layer that is no longer applying stress to the channel region and since the rest of the vertical structure is the same as that of the *SG* devices the mobility should be almost identical. This is clearly not the case. It is possible that the tensile strained nitride layer has become compressively strained in the areas over the pMOS devices due to the Ge bombardment. The increased band splitting due to the now compressive uniaxial strain could then explain the rise in mobility of the *SGN* device by 1.5 times. This effect where the Ge ions reverse the strain of the nitride layer has been seen before [Shimizu 2004].

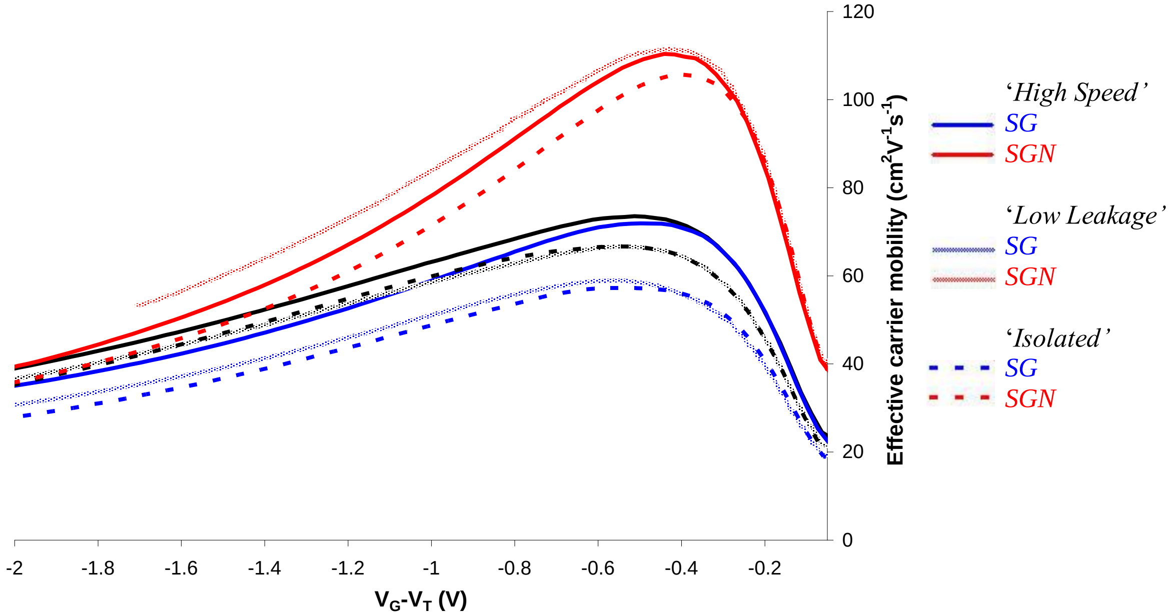

Figure 5-7: pMOS effective carrier mobility versus gate overdrive

In an attempt to reduce gate leakage these devices were fabricated with a nitride oxide. Using the oxide capacitance (in heavy accumulation) value of 1.63x10^{-6}Fcm^{-2}, and the method for extraction as described in *Chapter 3.3.6*, it was found that the equivalent SiO_2 oxide thickness is 2.1nm. The actual thickness of the nitride oxide is however as stated in *Chapter 4.1.2*.

5.3.2 '2002' batch

Split C-V data was extracted for large gate area devices of 100x100 μm^2 for the nMOS devices (*Figure 5-8*) and the pMOS devices (*Figure 5-9*).

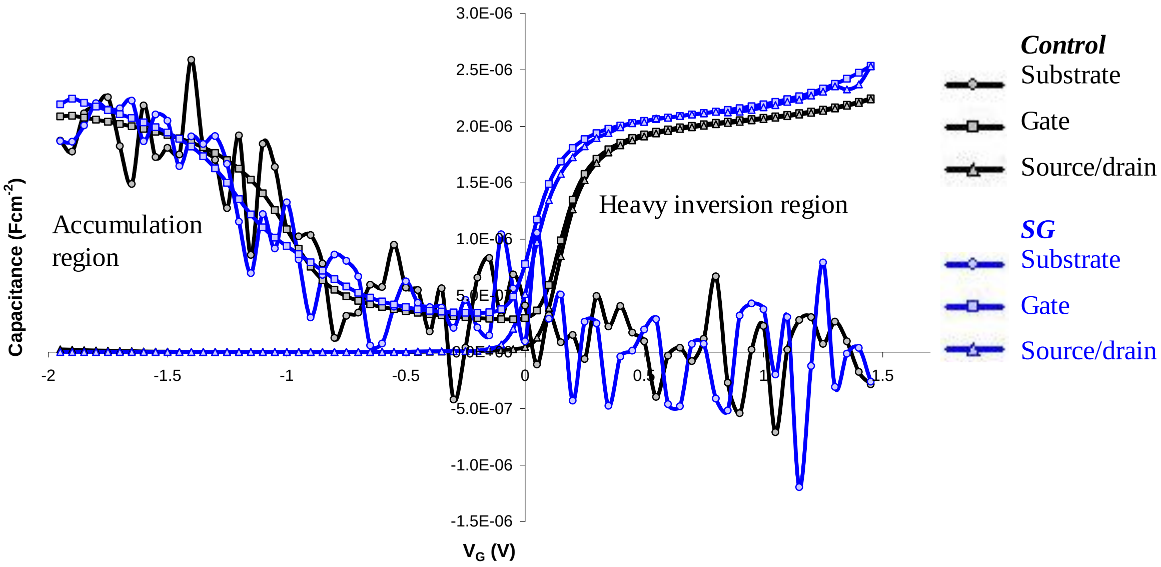

Figure 5-8: Split C-V data for 100x100 μm^2 nMOS devices

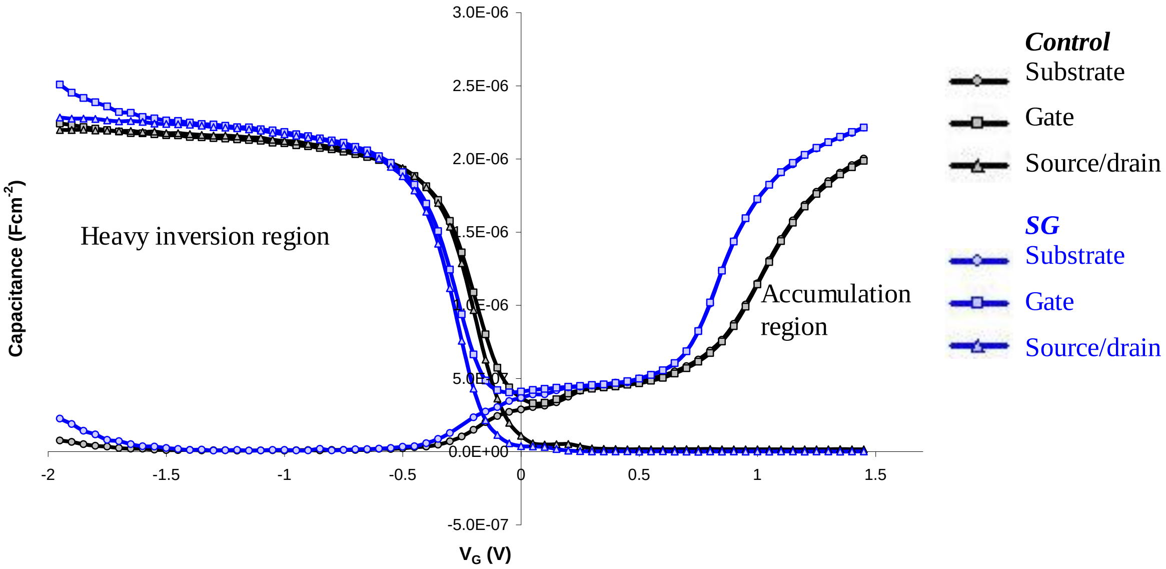

Figure 5-9: Split C-V data for 100x100 μm^2 pMOS devices

The split C-V data when combined with the I_D-V_G data gives the mobility as a function of voltage (*Figures 5-10 & 12* for nMOS and pMOS respectively), and as a function of the effective electric field (*Figures 5-11 & 13* for nMOS and pMOS respectively).

At its peak the mobility enhancement in the nMOS *SG* devices is a factor of 2 over the *Control*. However as dimensions are reduced and the effective electric field is increased the enhancement is diminished but still present.

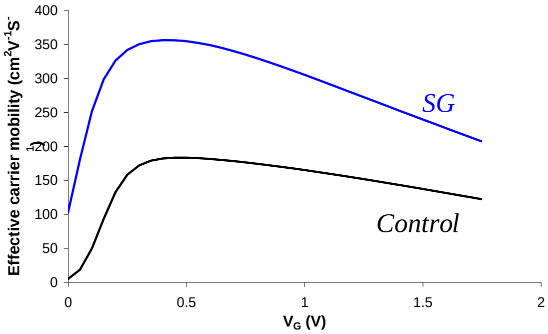

Figure 5-10: nMOS effective carrier mobility versus gate overdrive

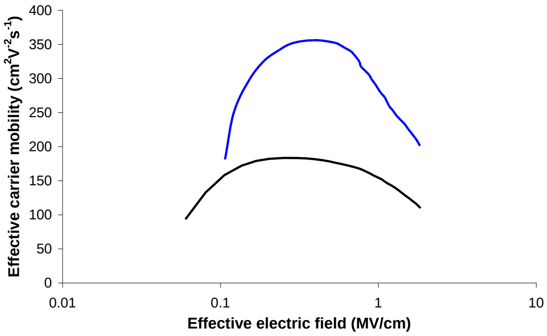

Figure 5-11: nMOS effective carrier mobility versus effective electric field

For the pMOS devices the strain only brings a maximum enhancement of 1.15 times in hole mobility. For short channel lengths where the effective electric field starts to increase the mobility of the holes in the strained silicon actually drops below that of normal silicon. This is again due to the surface confinement of the carriers.

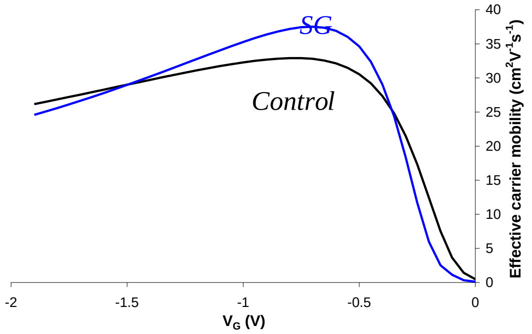

Figure 5-12: pMOS effective carrier mobility versus gate overdrive

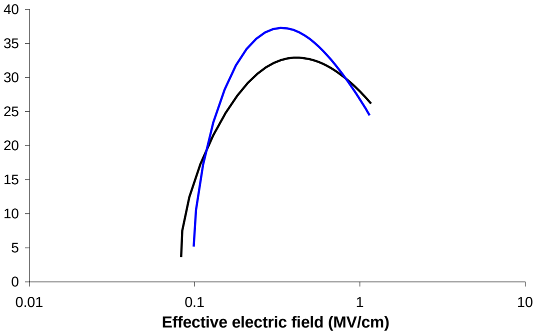

Figure 5-13: pMOS effective carrier mobility versus effective electric field

Figures 5-8 & 9 give the oxide capacitance in heavy accumulation to be 2.08x10^{-6}Fcm^{-2} and 2.17x10^{-6}Fcm^{-2} for the *Control* and *SG* wafers respectively. This equates to a *Control* wafer oxide thickness of 1.66nm and a *SG/SGCMP* wafer oxide thickness of 1.59nm. These are roughly half the nominal value stated in the device specifications. While a 50% error between the desired deposition thickness and the actual thickness is quite considerable, the values obtained from the C-V data are correct and correspond with those previously extracted for these devices [Nicholas 2004].

5.3.3 Batch comparisons

Both wafer batches saw improvements from the suppression of scattering for the nMOS devices when the channel was placed under biaxial strain from the VS. This value was smaller for the '2005' batch and is likely to be due to the larger electric fields in the newer devices. The additional tensile strain from the nitride layer had a small but positive effect on the mobility due to the increased band splitting reducing scattering but also due to the lowered effective mass of the carriers.

The mobility of the '2005' pMOS *SG* devices was lower than that for the C*ontrol* devices and this was to be expected due to the reasons mentioned earlier. The '2002' *SG* devices have been affected similarly but to a lesser extent as for the dimensions the electric fields were lowered. The nitride layer in the '2005' batch *SG* pMOS devices, is believed to be applying compressive stress to the channel and this has had a huge positive effect on the mobility increasing it 1.5 times that of the *Control* devices.

5.4 Threshold voltage

Not all devices will enter the inversion state at the same applied gate bias (V_G) and so should only be compared to each other at like gate overdrive values (V_G-V_T) to account for this. The introduction of strain into the channel region alters the gate bias that is required to set up the inversion condition (*Chapter 2.2*). This shift in the threshold voltage is directly related to the way the band energies are affected. The semiconductor surface band bending is equal to the gate bias applied plus any additional component causing a shift of the bands, in this case the application of strain, such that

$$e\phi_s \propto V_G + g(\sigma) \qquad (5.1)$$

Rearranging and using the condition for inversion gives a new threshold voltage of

$$V_T \propto 2e\phi_F - g(\sigma) \tag{5.2}$$

Field

Thus strain applied to the system should cause the threshold voltage to drop.

5.4.1 ‘2005’ batch

As described in *Chapter 4.1* the devices have been engineered such that $|V_T|$ is greatest for the ‘*Low Leakage*’ devices and smallest for the ‘*High Speed*’ with the ‘*Isolated*’ ones somewhere in-between. In the case of the n-type MOSFETs (*Figure 5-14*) this has been achieved with an approximate two percent drop in the threshold voltage for the ‘*High Speed*’ devices over the ‘*Isolated*’ ones, and an approximate 20% increase over the ‘*Isolated*’ devices for the ‘*Low Leakage*’ ones. The pMOS (*Figure 5-15*) ‘*Low Leakage*’ devices showed the correct behavior relative to the ‘*Isolated*’ ones. The magnitude of the threshold voltage increase was approximately 35% for the *Control* and *SG* devices and 20% for the *SGN* devices. Something however has gone slightly wrong with the ‘*High Speed*’ devices. The doping worked correctly for the *Control* batch showing a two percent drop in threshold voltage compared to the ‘*Isolated*’ set, but then this reverses to around a 15% increase for the strained devices. This is likely to be due to the large difference in the diffusion ratios of Arsenic (the dopant used) in SiGe and S-Si of 9 times and 1.4 times respectively [Payet 2005] making it very hard to control where the dopants finally reside after implantation andhence the affect they have on the threshold voltage.

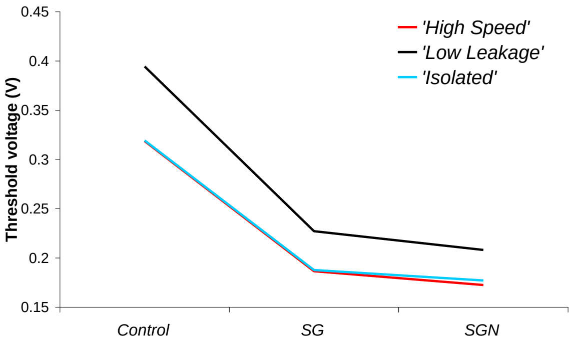

Figure 5-14: Threshold voltages for the different nMOS devices and optimization groups

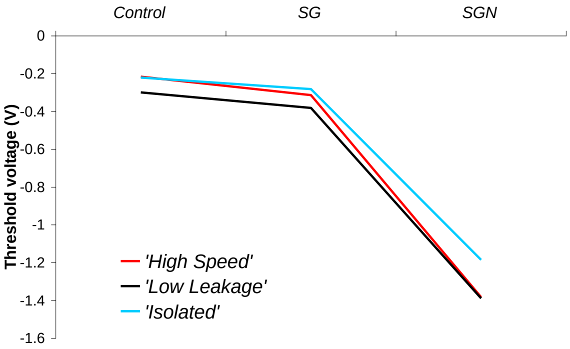

Figure 5-15: Same as 5-14 except for pMOS

The biaxial and uniaxial stress applied to the channel region should also reduce the threshold voltage for nMOS and pMOS devices if theory is correct (*Chapter 2.6*). *Figures 5-16 & 17* for nMOS and pMOS devices respectively clearly illustrate that the addition of strain in the channel region has caused the threshold voltage to lower. While these two graphs only represent the '*Low Leakage*' devices the other two optimization types show almost identical shapes and shifts, only the magnitudes are different. The shift in threshold voltage makes it harder to switch an nMOS device off and a pMOS device on.

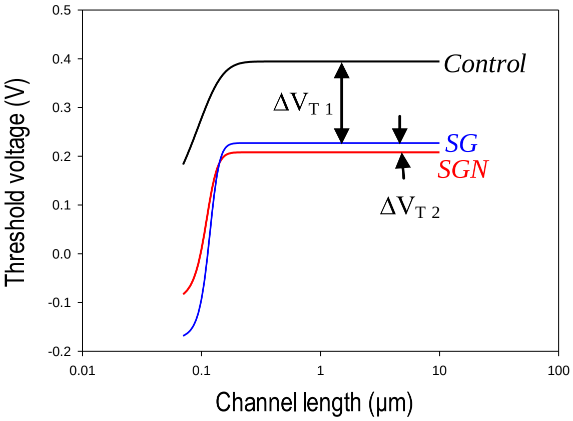

Figure 5-16: Threshold voltage versus channel length for nMOS 'Low Leakage' devices

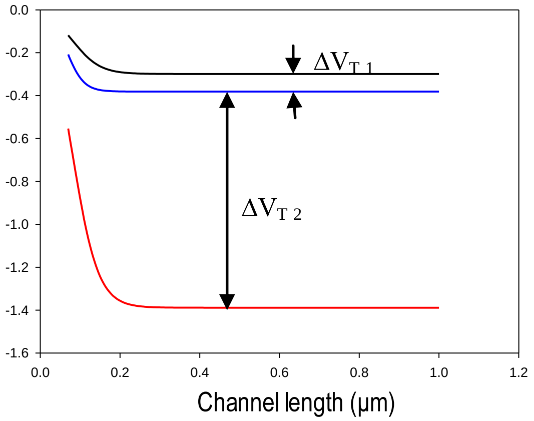

Figure 5-17: Threshold voltage versus channel length for pMOS 'Low Leakage' devices

The actual recorded data for the nMOS '*Low Leakage*' *SG* device as an example is presented in *Figure 5-18*. The spread in the threshold voltages was typically ± 10-20 mV as can be seen however for the pMOS *SGN* devices the spread increased dramatically to ± 100-200mV for the '*Low Leakage*' and 'High Speed' device sets.

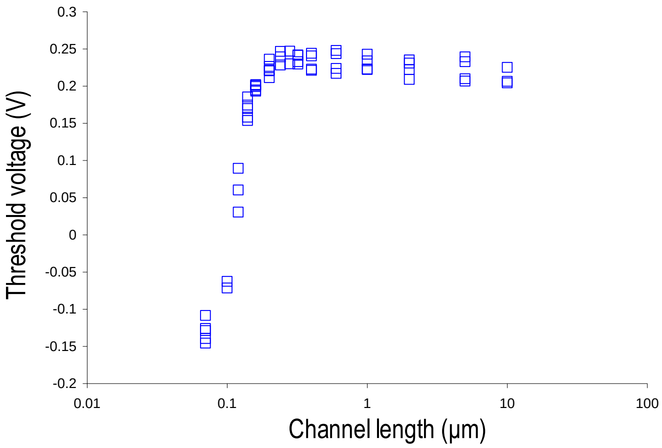

Figure 5-18: 'Low Leakage' SG devices to illustrate the scattering of the data points for each batch set

	SG-Control=$\Delta V_{T\,1}$ (mV)		*SGN*-SG=$\Delta V_{T\,2}$ (mV)	
	nMOS	pMOS	nMOS	pMOS
'High Speed'	-132 ± 12	-97 ± 15	-14 ± 9	-1069 ± 121
'Low Leakage'	-167 ± 14	-82 ± 22	-19 ± 12	-1008 ± 201
'Isolated'	-132 ± 14	-61 ± 16	-11 ± 10	-903 ± 22

Table 5-1: Threshold voltage shift due to technology boosters prior to the onset of SCEs

For an nMOS device (*Table 5-1*) the biaxial strain on the channel causes a large shift in the threshold voltage of approximately 130mV to 170mV, with the addition of further uniaxial strain only increasing this shift marginally by up to approximately 20mV. Averaging out the data it was demonstrated that the additional shift from the uniaxial strain is approximately an order of magnitude less than that introduced by biaxial strain.

Using the relation derived by Goo et al (*Chapter 2.7*) the shift for the *SG* devices is 30-70mV less than predicted for a SiGe VS with a 20% Ge content, however this can be accounted for, by the uncertainty in the electron affinity resulting in a range of possible threshold voltage shifts for the same strain value (*Figure 2-30*). Experimental data published on biaxially[Xiang 2003] [Sugii 2002] and uniaxially[Zhao 2003] strained n-type devices has shown similar results however it should be pointed out that in these cases the shift for uniaxial strain was observed independently from the biaxial strain.

For pMOS the magnitude of the relevant shifts was reversed with the biaxial strain causing a shift approximately 12 times less than that caused by the additional uniaxial strain. It is predicted by *Goo et al* (*Chapter 2.6*) that for a 20% Ge content in the VS, the pMOS devices should produce a threshold shift of approximately -134mV before taking any parasitic channels into consideration, and -80mV after. My data supports this and the spread in the values can again be explained by the uncertainty in the electron affinity.

Literature searches indicate that nothing has been published on the effect uniaxial strain has of the threshold voltage for pMOS devices. Due to the Ge ion bombardment it is believed that the devices have been additionally compressively strained along the channel length. However some ions will have not only entered the nitride layer causing relaxation, but some will have penetrated through into the gate structure. Therefore it is not only the strain causing a threshold voltage shift but also the damage done to the gate stack, and any equations such as those derived by Lim et al (*Chapter 2.6*) which only take energy levels and the density of states of the channel into account, will no longer hold valid. Uniaxial strain is most likely to shift the threshold voltage by the same order of magnitude as for nMOS devices so it is believed that the deposition of Ge ions into the pMOS *SGN* gate stack is the cause for the extreme shift of approximately 1V.

5.4.2 '2002' batch

As predicted in *Chapter 2.6* the strained nMOS (*Figure 5-19*) and pMOS (*Figure 5-20*) devices show a drop in the threshold voltage with the numerical values found in *Table 5-2*.

Using the relation derived by *Goo et al* (*Chapter 2.6*) the shift for the nMOS *SG* devices is approximately 70mV less than predicted for a SiGe VS with a 20% Ge content. This variation is accounted for by the uncertainty in the electron affinity [Lim 2004] [Thompson 2004b] as shown in *Chapter 2.6*, and corroborated by experimental data published [Xiang 2003] [Sugii 2002].

For the pMOS *SG* device the threshold shift difference between my experimental value of -63.1mV and the -80mV predicted by *Goo et al* (*Chapter 2.7*) can also be put down to the uncertainty in the electron affinity.

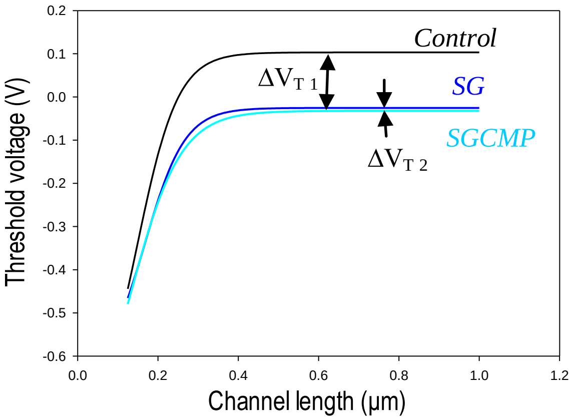

Figure 5-19: Threshold voltage versus channel length for nMOS devices

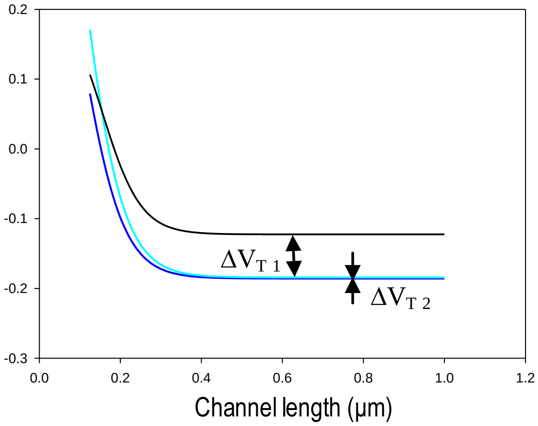

Figure 5-20: Threshold voltage versus channel length for pMOS devices

The CMP stage has reduced the total threshold voltage shift by 7mV for the nMOS *SGCMP* devices and increased it by 2mV in the pMOS devices. When the errors in the values presented in *Table 5-2* are also included it can be seen that the devices are

SG-Control=$\Delta V_{T\,1}$ (mV)		*SGCMP*-SG=$\Delta V_{T\,2}$ (mV)	
nMOS	pMOS	nMOS	pMOS
-128.5 ± 10	-63.1 ± 12	-6.9 ± 13	1.5 ± 6

Table 5-2: Threshold voltage shift due to technology boosters prior to the onset of SCEs

fluctuating around a value of ΔVT2=0V suggesting that CMP has no impact on this device parameter.

5.4.3 Batch comparisons

The nMOS and pMOS devices from both batches show similar shifts in the threshold voltage when the channel is biaxially strained in the *SG* devices and the magnitudes are supported by theory. The use of a CMP stage does not appear to have an adverse affect on the threshold voltage, and the additional shift from the nitride layer on an nMOS device is minimal. The shift when a tensile turned compressively strained nitride layer is used to apply stress to the channel of a pMOS device is however problematic and will require a lot of optimization to control which will degrade the performance enhancements. An alternative to the method used for these devices, is selective deposition of a intrinsically compressively strained nitride layer over the pMOS devices which should result in performance increases with only a modest change in V_T.

5.5 Onset of SCEs

The indication that SCEs are starting to degrade performance can be seen in the data for the threshold voltage at the point where the voltage starts to drop. This is known as the threshold voltage roll off and is caused primarily due to punchthough and DIBL effects. While ballistic overshoot is a positive effect of having a short channel length the other negative effects have a larger impact on device performance.

5.5.1 ‘2005’ batch

It is desirable that the channel length (L_{SCE}) at which SCEs start to cause problems be as short as possible. This allows for high current drives without the need for additional doping in the processing stages to correct the threshold voltage. The data (*Table 5-3*) suggests that strain has no consistent effect on the point at which SCEs start to occur in these short channel devices. However it is obvious that the additional doping intended to shift the threshold voltage for ‘*High Speed*’ and ‘*Low Leakage*’ devices has caused for an earlier onset of the threshold voltage roll-off (i.e. at a longer channel length) as well as introduced more fluctuation in the data. For nMOS altering the doping to change the threshold voltage has increased the value of L_{SCE} by approximately 40-45% over that of the ‘*Isolated*’ devices, and for pMOS 70-90%.

		Control L_{SCE} (nm)	*SG* L_{SCE} (nm)	*SGN* L_{SCE} (nm)
‘*High Speed*’	nMOS	157 ± 44	171 ± 26	164 ± 15
	pMOS	170 ± 10	163 ± 29	212 ± 65
‘*Low Leakage*’	nMOS	164 ± 15	159 ± 8	152 ± 16
	pMOS	196 ± 24	143 ± 25	238 ± 100
‘*Isolated*’	nMOS	112 ± 5	116 ± 4	121 ± 6
	pMOS	102 ± 1	101 ± 1	119 ± 8

Table 5-3: Gate length at which SCEs start to massively degrade the threshold voltage of the device

5.5.2 ‘2002 batch

The channel length at which threshold voltage roll-off starts to occur seems unaffected when the channel region is placed under biaxial strain (*Table 5-4*). Attempts to minimise dislocations by performing CMP on the VS has increased L_{SCE} by 5-10%. Fewer dislocations in the channel and VS will alter the diffusion rates of the dopants and hence the size and shape of the depletion zones created by the source and drain.

	Control L_{SCE} (nm)	*SG* L_{SCE} (nm)	*SGCMP* L_{SCE} (nm)
nMOS	373 ± 35	372 ± 115	408 ± 44
pMOS	323 ± 34	316 ± 47	332 ± 36

Table 5-4: Gate length at which SCEs start to massively degrade the threshold voltage of the device

5.5.3 Batch comparisons

Strain it would seem has little effect on the channel length at which SCEs start to cause problems. This has been shown within each batch set. However for the '2002' batch the L_{SCE} is approximately double that of the '2005' batch. As both batches have been made by STMicroelectronics only 3 years apart this difference must be due to refinements in the manufacturing process. The alteration in dopant energies and concentrations can be seen in *Tables 4-1 & 2*. For example the S/D implant concentration is the same for both batches but the energy used has been reduced by a factor of four. Such changes in combination with a reduced thermal budget, has had a positive effect on the device operation in the latest batch.

5.6 I_D-V_G characteristics

Using *Equation 2.6* it was to be expected that, when comparing like devices, as the physical channel length was reduced the drive current would increase for the same gate bias. This is because $I_D \ \alpha \ 1/L$. This phenomenon occurred in the linear and saturation regimes of operation for the nMOS (*Figures 5-21 & 22* respectively) and pMOS (*Figures 5-23 & 24* respectively) devices for both wafer batches.

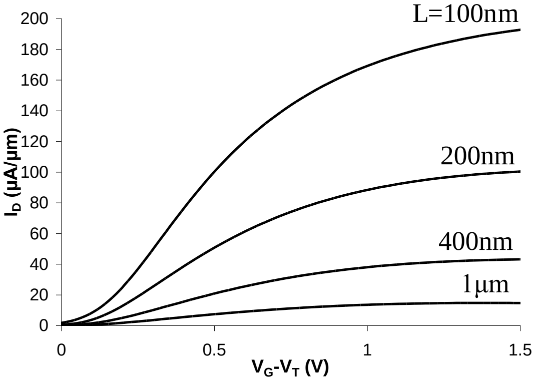

Figure 5-21: Family of curves showing the increase in drain current for a given gate overdrive as the gate length is varied for a nMOS 'High speed' Control device under linear conditions

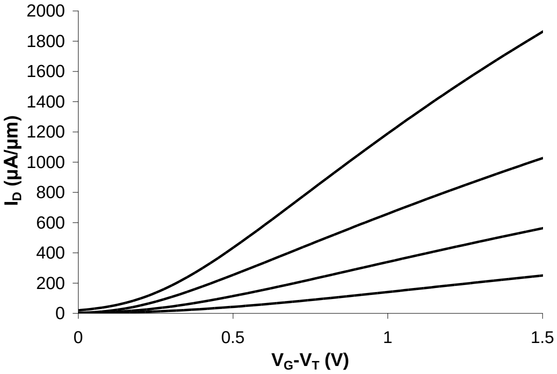

Figure 5-22: As for 5-21 but for saturation conditions

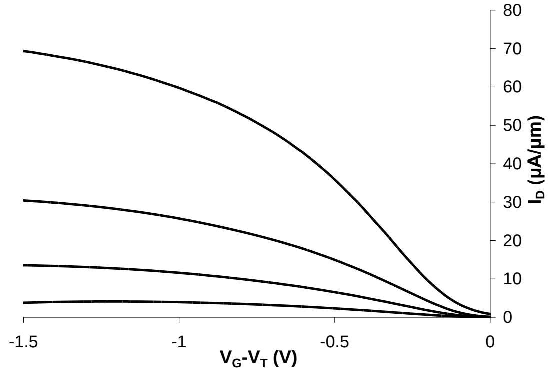

Figure 5-23: As for 5-21 but for pMOS

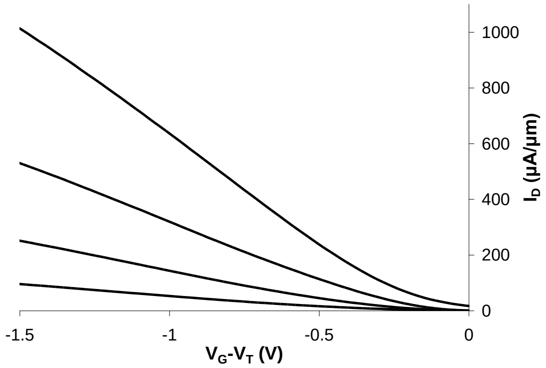

Figure 5-24: As for 5-21 but for pMOS under saturation conditions

All device types show the same patterns of behaviour as illustrated in *Figures 5-21* to *5-24*, however the shape of the curves and the numerical values differ for the different wafer types and optimizations.

5.6.1 '2005' batch

The linear region of operation of the 3 types (*Control, SG,* and *SGN*) of nMOS '*High Speed*' devices at a gate length of 200nm can be seen in *Figure 5-25*. This gate length is prior to the onset of SCEs. The strained devices show improved current drive for all gate biases over the control. The *SGN* devices perform marginally better than the SG ones. The data from the '*Low Leakag*e' and '*Isolated*' devices is equivalent to this.

Figure 5-26 shows the saturation region of operation of the 3 types of nMOS '*High Speed*' devices at a gate length of 200nm. The *SGN* device performs worse that the *SG* device but still with increases in drive current of 5% and 10% respectively over the *Control* device up to a gate bias of around 1.4V. After this the *Control* device outperforms the other two. The data from the '*Low Leakage*' and '*Isolated*' devices is equivalent to this.

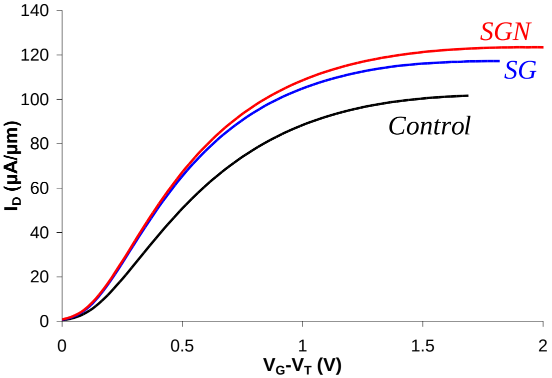

Figure 5-25: Drain current versus gate overdrive for nMOS 'High Speed' device (L=200nm) under linear conditions

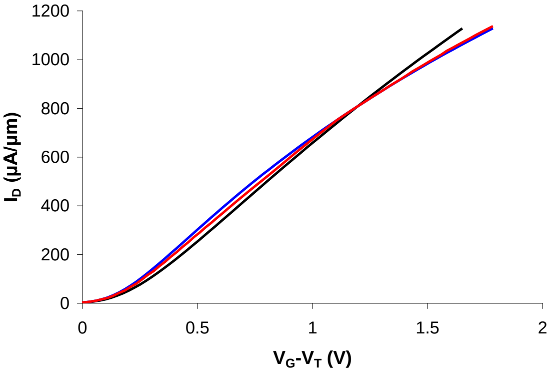

Figure 5-26: Drain current versus gate overdrive for nMOS 'High Speed' device (L=200nm) under saturation conditions

The results for the pMOS '*High Speed*' devices are shown in the linear (*Figure 5-27*) and saturation (*Figure 5-28*). The *SG* devices exhibit a degraded current drive over the *Control* at all overdrive voltages with the largest drop of 11% in drive current at -2V in linear operation and a 20% drop in drive current at -2V in saturation. From theory (*Chapter 2.4.2*), any mobility enhancement gains from the biaxial strain should be lost due to surface confinement of carriers at high effective fields. This reduction in mobility enhancement was seen in *Chapter 5.3.1*, and this is reflected in the lowered drain currents. The *SGN* device has increased current drive showing a maximum gain of approximately 40% at around -0.7V in the linear region and approximately 36% at a similar gate overdrive in saturation. This is as suggested earlier due to the nitride layer becoming intrinsically compressively strained due to Ge ion bombardment resulting in positive performance changes. The data from the '*Low Leakage*' and '*Isolated*' devices is similar to this.

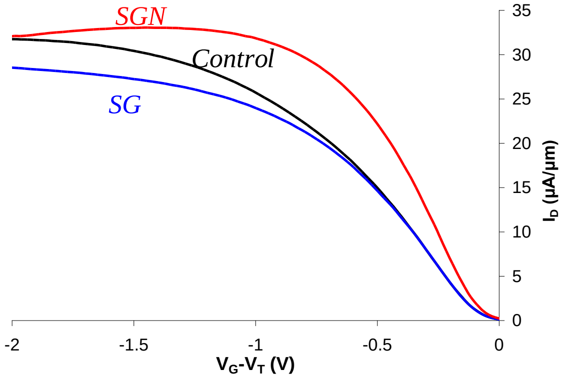

Figure 5-27: Drain current versus gate overdrive for pMOS 'High Speed' device (L=200nm) under linear conditions

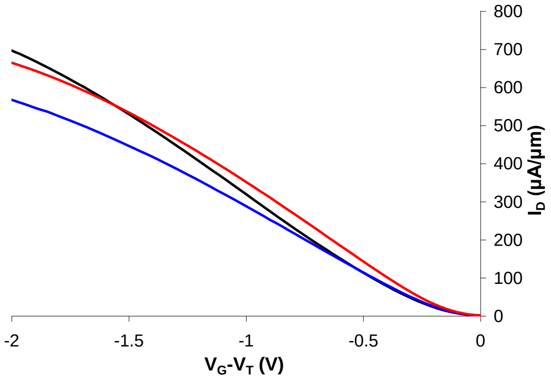

Figure 5-28: Drain current versus gate overdrive for pMOS 'High Speed' device (L=200nm) under saturation conditions

5.6.2 '2002' batch

In the linear (*Figure 5-29*) region the nMOS *SG* and *SGCMP* devices show a faster turn on giving a higher drain current than the *Control* device for low gate overdrive biases (<0.55V). At higher gate overdrive biases the strained devices exhibited lower drain currents. This same pattern can also be seen in the saturation region of operation (*Figure 5-30*).

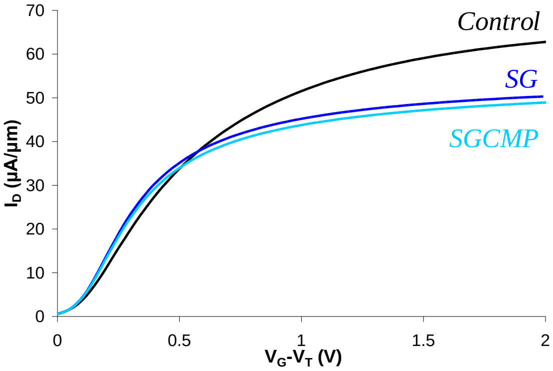

Figure 5-29: Drain current versus gate overdrive for nMOS device (L=200nm) under linear conditions

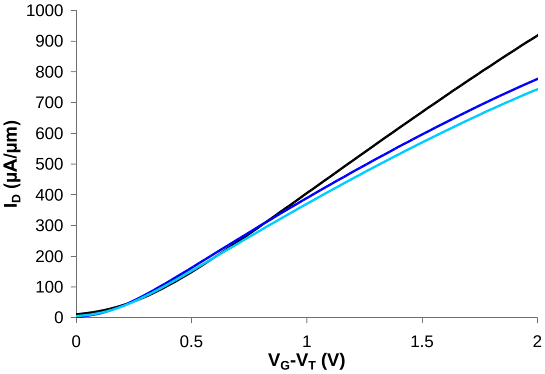

Figure 5-30: Drain current versus gate overdrive for nMOS device (L=200nm) under saturation conditions

The pMOS strained devices show improved drain currents in the linear region (*Figure 5-31*) up to gate overdrives of around -1.3V whereas in the saturation region (*Figure 5-32*) all three device types showed similar drive current values up until -0.7V at which point the *Control* device began outperforming the other two.

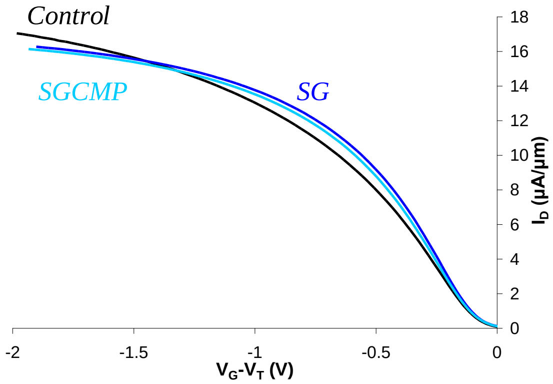

Figure 5-31: Drain current versus gate overdrive for pMOS device (L=200nm) under linear conditions

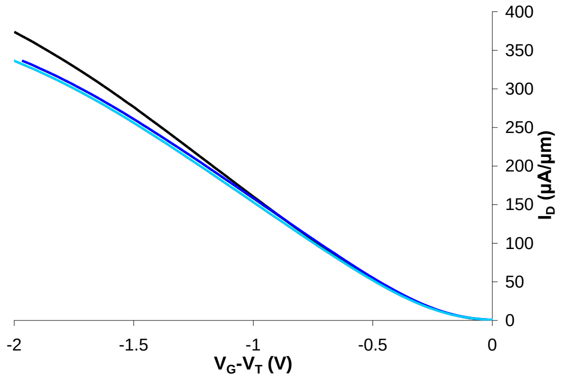

Figure 5-32: Drain current versus gate overdrive for pMOS device (L=200nm) under saturation conditions

The additional CMP processing step theoretically should improve drain currents for both nMOS and pMOS devices by reducing scattering at the interface. These results do not support this as for all regimes and gate lengths the, *SGCMP* devices showed worse performance than the *SG* ones.

5.6.3 Batch comparisons

The exact behaviour of the devices as the gate overdrive is increased is quite different between the two batches. For example the '2005' *SG* devices in *Figure 5-25* outperform the *Control* devices for all gate overdrive values whereas for the equivalent '2002' devices in *Figure 5-29* the drive current of the *Control* devices outperforms the *SG* devices at gate overdrives over 0.55V.

The most important change is however in the magnitude of the drive currents. The ITRS set a roadmap value for devices (in the saturation region with $V_G=V_D=1.1V$) fabricated in 2002 at 900μA/μm for nMOS and 405μA/μm for pMOS (pMOS defined as 0.4-0.5 times that of the nMOS drive current) for a channel length of 65nm [ITRS 2001]. The roadmap values for devices fabricated in 2005 were set as 1020μA/μm for nMOS and 460μA/μm for pMOS with a channel length of 32nm [ITRS 2005]. *Table 5-5* shows the range of drive currents that have been produced by the devices tested for this thesis. The conditions for calculating the data in *Table 5-5* and that presented in the ITRS publications is not quite the same. The voltage I have used is slightly larger however as the channel lengths are also larger it roughly evens things out giving a loose indication that the devices under test for this thesis were 'state of the art'.

	'2002' I_D (μA/μm)	'2005' I_D (μA/μm)
nMOS	570-670	990-1020
pMOS	260-280	450-540

Table 5-5: Saturation region drive current range covered by the different device types with $V_G=V_D=1.5V$

5.7 Subthreshold slope

Values for subthreshold slope in devices are extremely important as they show how much control the gate has over the channel region. For long channel lengths the subthreshold slope is usually low at around 70mV/dec and gets higher for shorter channel devices especially when SCEs occur. *Figure 5-33* is an illustrative I_D-V_G curve which is used to extract the subthreshold slope parameter in the linear region (highlighted in yellow and then enlarged). It has been predicted that the additional strain in the channel for a VS Ge content of 20% such as that under investigation for this thesis should lead to a nMOS increase in the subthreshold slope of ~8% and no increase for pMOS [Goo 2003].

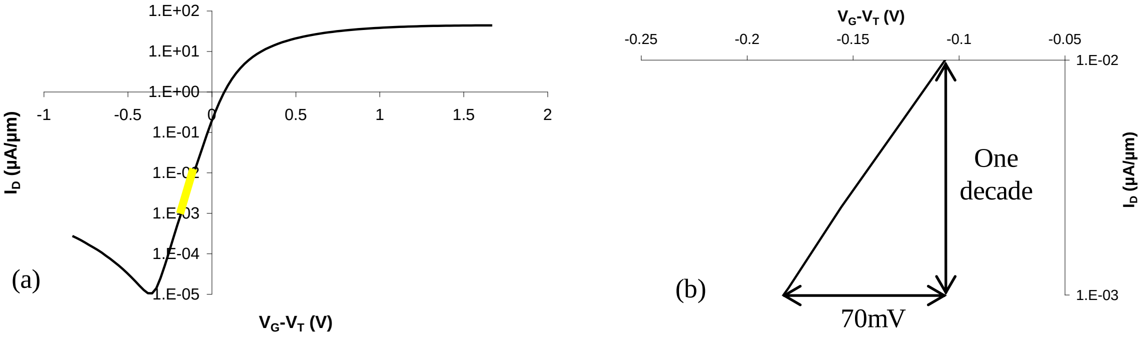

Figure 5-33: Current drive (on a log scale) versus gate overdrive for (a) all the data and (b) an enlargement of the highlighted section in (a) at which the subthreshold slope value is taken

For all the graphs in this section the x-axis only shows the channel length up to 500nm. Devices were tested up to 1µm however the data at these longer lengths is identical in magnitude to that presented from 400-500nm and so has been omitted so that more detail can be observed for the shorter lengths where fluctuations are taking place.

5.7.1 '2005' batch

As predicted [Goo 2003] the subthreshold slope for nMOS channel lengths of 200nm up to 1µm were approximately 10% larger for the *SG* and *SGN* devices than the *Control* devices at 80mV/dec (*Figure 5-34*). As SCEs start to affect performance at gate lengths given in *Table 5-3* the subthreshold slope value starts to dramatically increase to values in excess of 200mV/dec and up to as high as 395mV/dec for the '*High Speed*' *SG* devices. The increase for the *Control* devices is not as dramatic reaching a high of only 150mV/dec. Devices with large subthreshold slope values of 200mV/dec or greater are considered useless, as the change in gate bias required to manipulate the drive current is too great. There is little variation between the values for the *SG* and *SGN* devices. The '*Isolated*' devices performed slightly better in the channel length region of 100-200nm due to their later onset of SCEs.

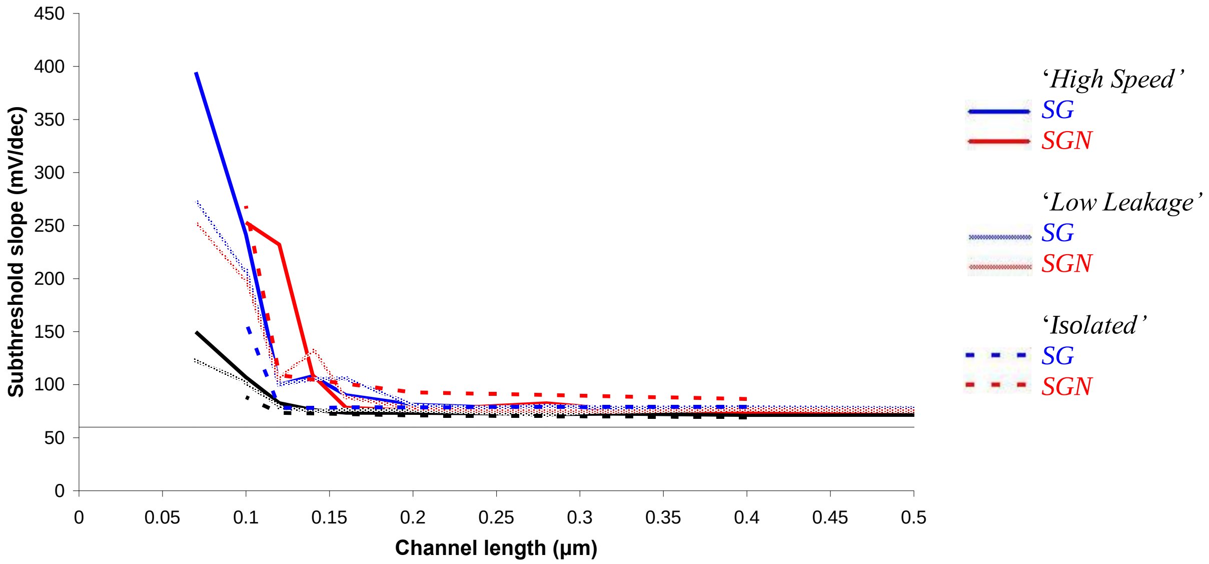

Figure 5-34: Subthreshold slope versus channel length for nMOS devices

The pMOS devices (*Figure 5-35*) show a reduced spread in values with the *Control*, *SG* and *SGN* devices all giving subthreshold values of 68-69mV/decfor channel lengths above 250nm as has previously been found [Goo 2003]. Values start to increase first for the *SGN* devices as a result of the higher channel length at which SCEs start to degrade performance (*Table 5-3*), and then the curve is followed by the other device types.

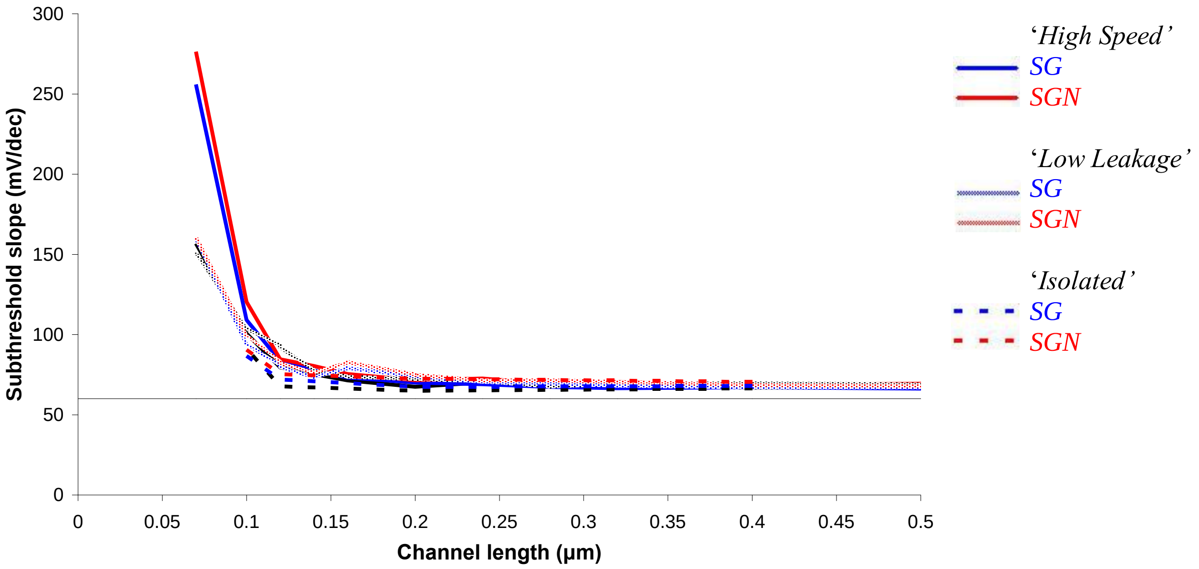

Figure 5-35: Subthreshold slope versus channel length for pMOS devices

Overall it would appear that for channel lengths prior to the onset of SCEs, the nMOS devices are worse than the pMOS devices with a 5% increase in values of subthreshold slope.

5.7.2 '2002' batch

A slightly higher than predicted [Goo 2003] increase in the subthreshold slope of the *SG* devices compared to the *Control* devices was observed of 14% (*Figure 5-36*) at channel lengths prior to the onset of SCEs (*Table 5-4*). The additional CMP stage performed on the *SGCMP* devices has further increased this by another 14%. As SCEs start to occur in the devices at a channel length of approximately 400nm we see the subthreshold slope values start to rise. For the *Control* devices this is a gentle slope which climbs dramatically at 200nm. For the strained devices an immediate jump to values of 220-230mV/dec is observed before this too begins to climb in an exponential manner. For channel lengths less than 175nm, all the device types would be considered useless as they have surpassed the upper limit for a 'good' device In terms of the subthreshold voltage value

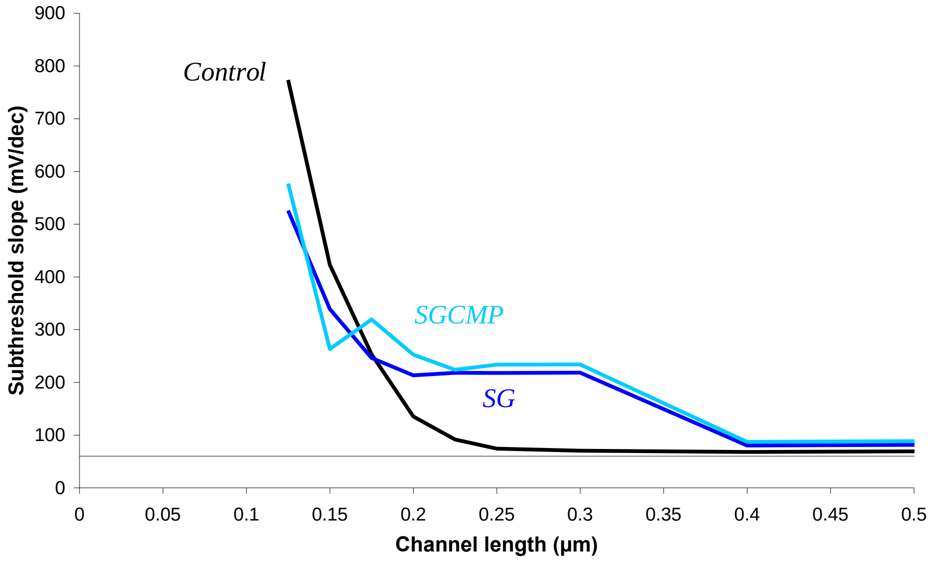

Figure 5-36: Subthreshold slope versus channel length for nMOS devices

The addition of strain in the pMOS devices also shows an increase in the subthreshold slope (*Figure 5-37*) of around 5% compared to the *Control* devices. The *SGCMP* devices show an increase in subthreshold slope of approximately 6% over the *SG* devices prior to the onset of SCEs The values for all three device types start to rise for channel lengths just above 300nm, and this ties in with the values for the onset of SCEs given in *Table 5-4*.

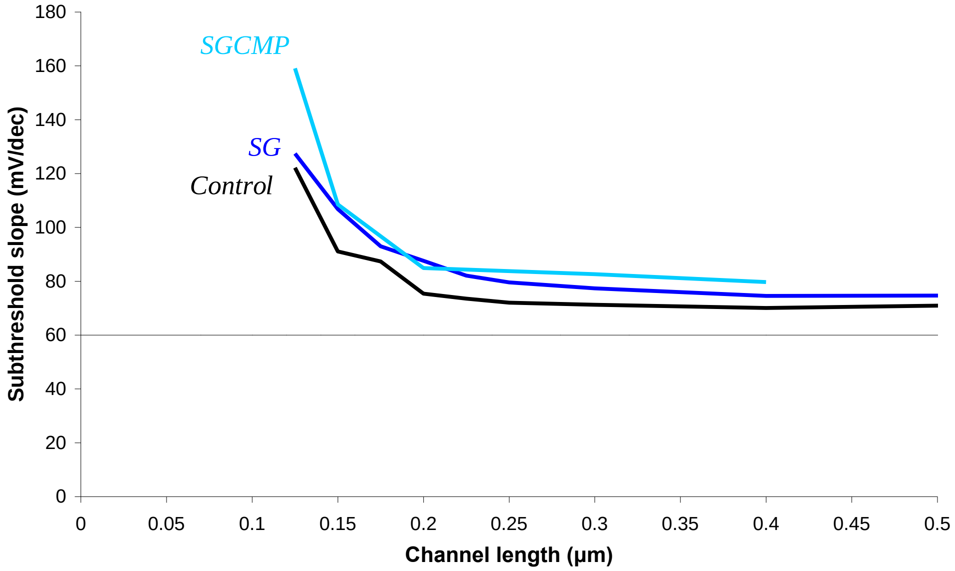

Figure 5-37: Subthreshold slope versus channel length for pMOS devices

5.7.3 Batch comparisons

Both batch sets had similar values for the nMOS and pMOS devices with the pMOS devices performing marginally better at longer channel lengths and considerably better after SCEs have started to degrade performance. The type of strain or the addition of a CMP stage seems to have little impact of the subthreshold slope values. The point at which the subthreshold slopes start to dramatically increase is different between the two batches. This difference reflects the earlier onset of SCEs as described in *Chapter 5.5*.

5.8 I_D-V_D characteristics

The I_D-V_D characteristics of a device can indicate whether any 'self-heating' is occurring in the channel. This is shown by a drop in drive current as the drain bias is increased for a given gate bias.

5.8.1 '2005' batch

It was expected from theory (*Chapter 2.4.5*) that the strained devices would exhibit the effects of self heating due to the SiGe VS thermally isolating the channel. A negative drain conductance for the *SG* and *SGN* nMOS devices (*Figure 5-38*) implies that there is considerable 'self-heating' taking place. Even though the conductance at low voltages is not negative the drain currents will still have been reduced for a given drain bias and so if 'self heating' were not occurring, the gradient of the curves would be larger.

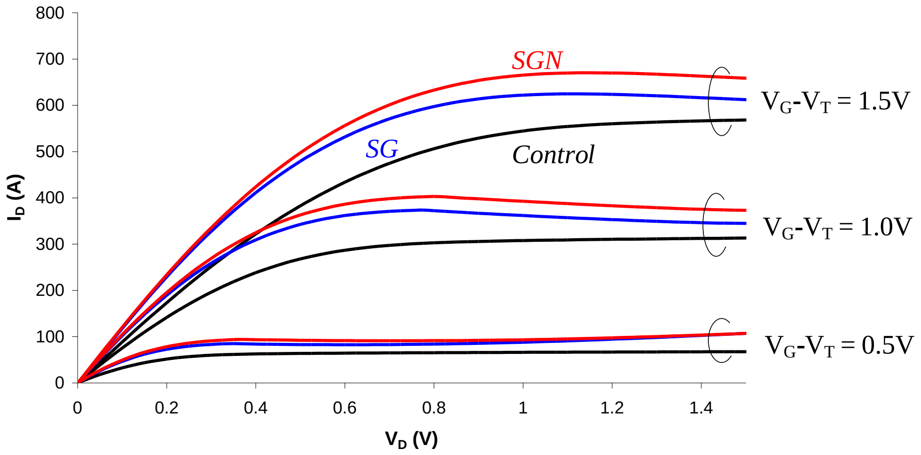

Figure 5-38: Drain current versus drain bias for nMOS 'High Speed' device (L=400nm) at 3 different gate overdrive values

The drain conductance remains positive for the pMOS devices (*Figure 5-39*) suggesting that 'self-heating' is not occurring. A pMOS device will produce roughly half the power output of an nMOS device [Jenkins 2002] as a result of its lower drive current ($P = I^2R$). Therefore while 'self heating' is still taking place in the channel, higher drain biases must be applied before its effect becomes apparent. If the currents were matched for the nMOS and pMOS devices, the pMOS devices would actually dissipate more heat due the lower mobility. It was not possible to collect any I_D-V_D data for the *SGN* devices as once I_D-V_G measurements had been made (in order to determine the threshold voltage) and the devices were re-tested oxide breakdown occurred. This phenomenon then allows current to flow from the source to the gate and the device is destroyed.

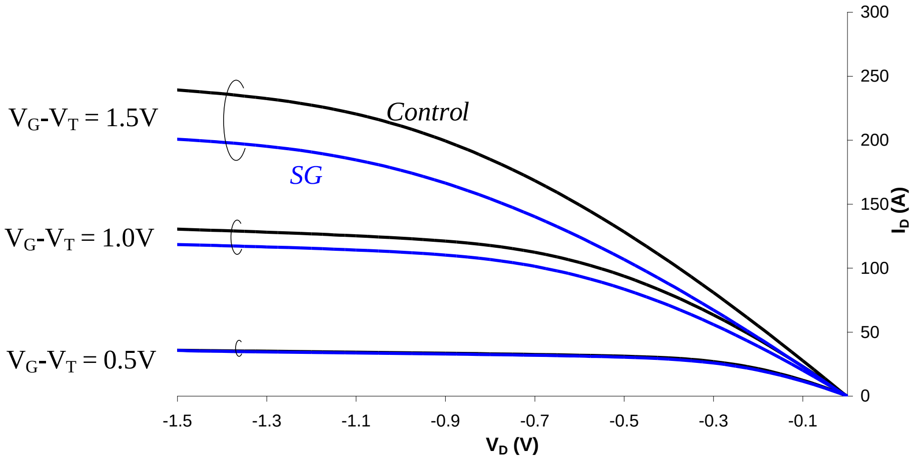

Figure 5-39: Drain current versus drain bias for pMOS 'High Speed' device (L=200nm) at 3 different gate overdrive values

5.8.2 ‘2002’ batch

The negative drain conductance for the strained nMOS devices (*Figure 5-40*) implies ‘self-heating’ due to channel isolation has become a problem. While the *SGCMP* devices show a slight drain current improvement for a given drain bias over the *SG* devices the drain conductance is identical. It would appear that smoothing the surface of the SiGe VS prior to the deposition of the uniform Ge composition layer has little effect on its poor thermal conductivity.

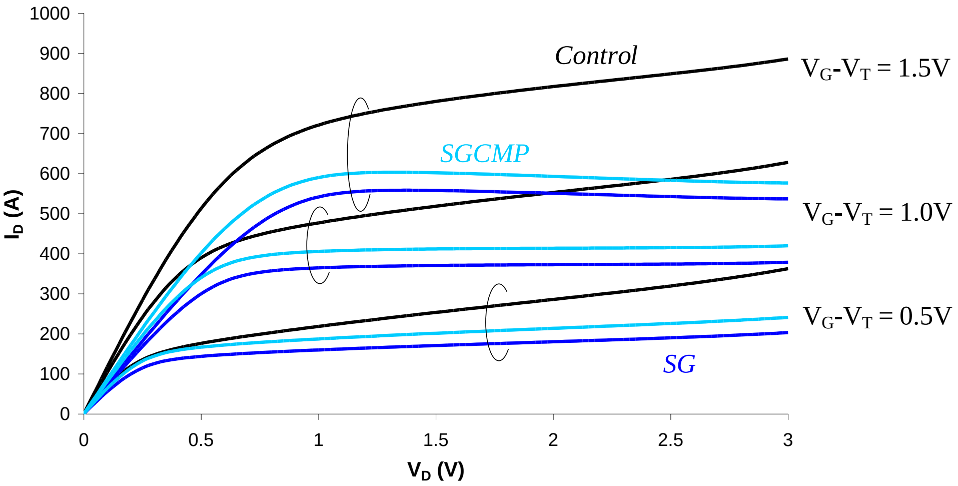

Figure 5-40: Drain current versus drain bias for nMOS device (L=200nm) at 3 different gate overdrive values

5.8.3 Batch comparisons

Both batches of nMOS devices showed that when a SiGe VS is used to induce strain in the channel it causes degradation in the drain current due to its low thermal conductivity allowing the channel to heat up. The pMOS devices did not indicate ‘self-heating’ to be a huge problem at these drain biases but if more power is dissipated the effect will become apparent.

Pulsed measurements can be performed where I_D-V_D data is obtained as a function of wafer temperature which allows for the extraction of any temperature rise due to ‘self-heating'. Such measurements were previously performed on the '2002' batch [Nicholas 2004] and showed that 'self-heating' resulted in a 10% drop in drive currents for nMOS devices. Other literature has shown that the decrease in drive current can be as much as 15% [Jenkins 2002] or 20% [Jenkins 1995] with the channel temperature rising by up to 100^0C.

5.9 On currents

The drive current produced by a device is a big indicator of performance with larger drive currents desired. The on current is defined as the current at a suitably large gate bias such that the channel is most definitely populated. The *SG*, *SGN*, and *SGCMP* devices have been plotted to show the percentage increase in the on currents over that of the *Control* device under linear conditions (V_D=50mV).

5.9.1 '2005' batch

The difference device optimizations performed on the '*Low Leakage*', '*High Speed*', and '*Isolated*' devices are briefly discussed in *Chapter 4.1*. The additional doping is done to shift V_T, and as the on current values have been taken at 1.5V gate overdrive (ie. V_G-V_T), the V_T element has been removed from the data and all the device groups should perform similarly.

The nMOS devices are shown in *Figure 5-41*. The *SG* devices are responsible for a big jump in the on current giving an enhancement of around 60% at a channel length of 1µm. This reduces until around 100-150nm at which point the *Control* devices outperform the others. This is the channel length at which the onset of SCEs has been seen to occur (*Table 5-3*). At long channel lengths it was expected that the difference in performance between the *SGN* and *SG* devices would be negligible as the nitride layer is unable to act on such a long channel as illustrated in *Figure 2-22*. As the channel length is reduced the nitride layer makes more of an impact until around the 200nm length where it is predicted that the entire channel is additionally strained and the performance increases rise. This effect is seen in the results. Comparing the *SGN* to the *SG* devices it can be seen that the additional uniaxial strain only improves on currents by around 0-1% at a channel length of between 600nm and 1µm. The *SGN* over *SG* improvement rises gently to approximately 3% at 280nm and then starts to climb in a more pronounced manner reaching approximately a 10% increase in drive current at 240nm and maintaining this up until 140nm and is shown in all three device groups (‘*High Speed’, ‘Low Leakage’* and *‘Isolated’*).

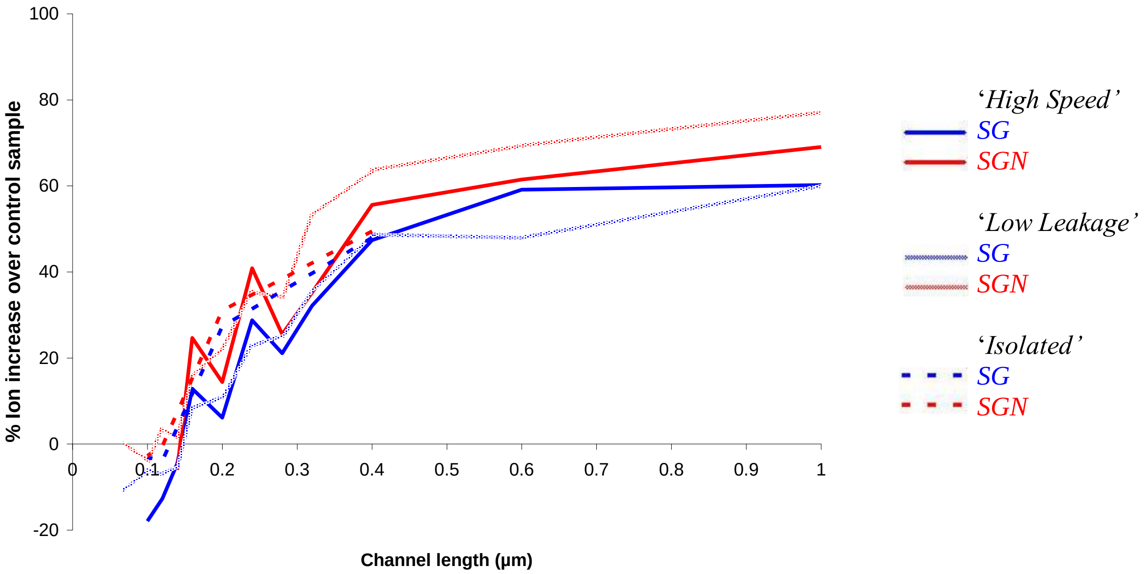

Figure 5-41: Percentage increase of nMOS Ion current for strained devices relative to the Control versus gate length

The pMOS (*Figure 5-42*) *SG* devices exhibit a reduction in performance when compared to the *Control* devices of approximately 40% at a gate length of 1µm which slowly improves to a reduction of approximately 15% at 100nm. Little improvement over the *Control* is expected due to the surface confinement of holes at high effective electric fields. The light to heavy band splitting is cancelled [Thompson 2004b] and the mobility enhancements are lost [Mikkelsen 1982] as shown in *Figure 5-7*. However this level of degradation appears a little too severe especially as for reasons discussed earlier the *SG* devices should match the performance of *SGN* devices at long channel lengths. It is possible that the additional degradation could be due to this wafer having a larger surface roughness and more impurities within the channel.

The *SGN* devices start at an on current level 5% larger than the *Control* devices and this is likely to be due to the VS only as it has been shown that at long gate lengths the nitride layer is not able to place the channel region under the desired strain. As the channel length is reduced past 200nm which is predicted [Payet 2005] as the maximum length that can be completely strained by the nitride layer, the percentage increase over the *Control* devices starts to climb rapidly. The addition of the nitride layer in the *SGN* devices, along with the VS component, has increased the on currents such that they provide improvements over the *Control* devices by as much as 55% for gate length of 100nm. The increase with each successively shorter device is reduced after 150nm which is around the point at which SCEs are becoming an issue.

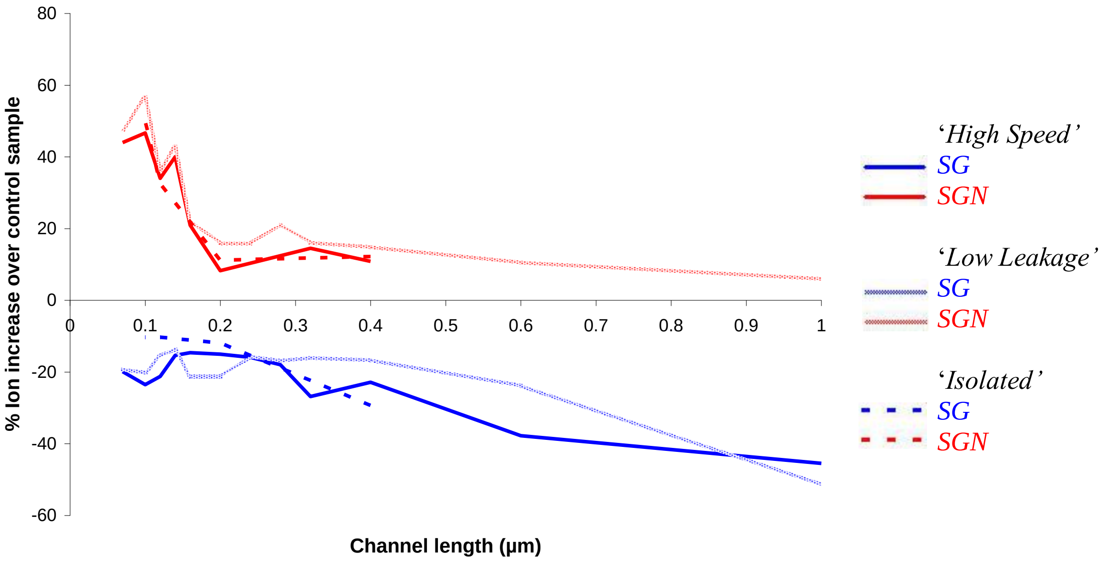

Figure 5-42: Percentage increase of pMOS Ion current for strained devices relative to the Control versus gate length

5.9.2 '2002' batch

The *SG* and *SGCMP* nMOS devices (*Figure 5-43*) both show on current increases over the control for channel lengths longer than 350nm. This is the channel length at which we have observed the onset of SCES (*Table 5-4*). It would appear that a reduction in the benefits from using a biaxially strained channel occurs as the gate length is shortened at sub micron dimensions. The *SGCMP* devices performed marginally worse that the *SG* devices by approximately 3%.

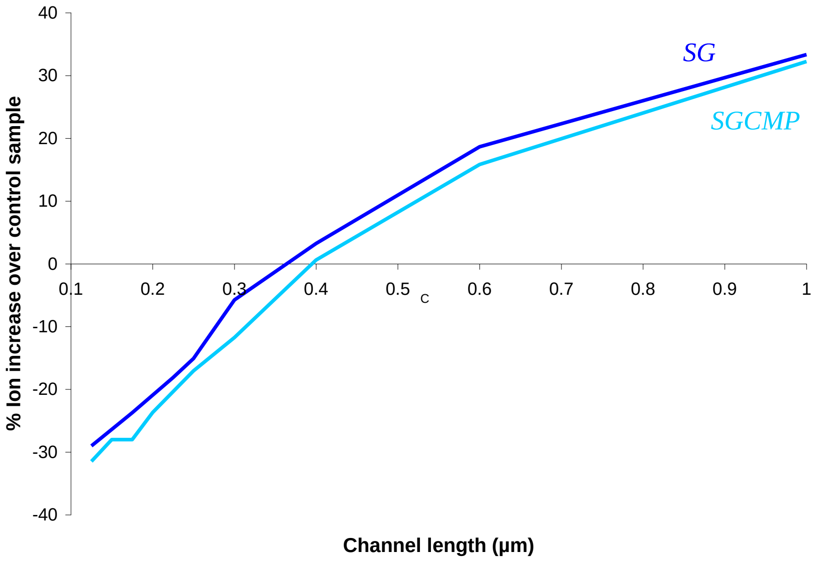

Figure 5-43: Percentage increase of nMOS Ion current for strained devices relative to the Control versus channel length

The *SG* pMOS devices (*Figure 5-44*) show performance enhancements of around 2% right down to the shortest channel lengths tested when the fluctuations are smoothed. The additional CMP stage seems irrelevant for pMOS.

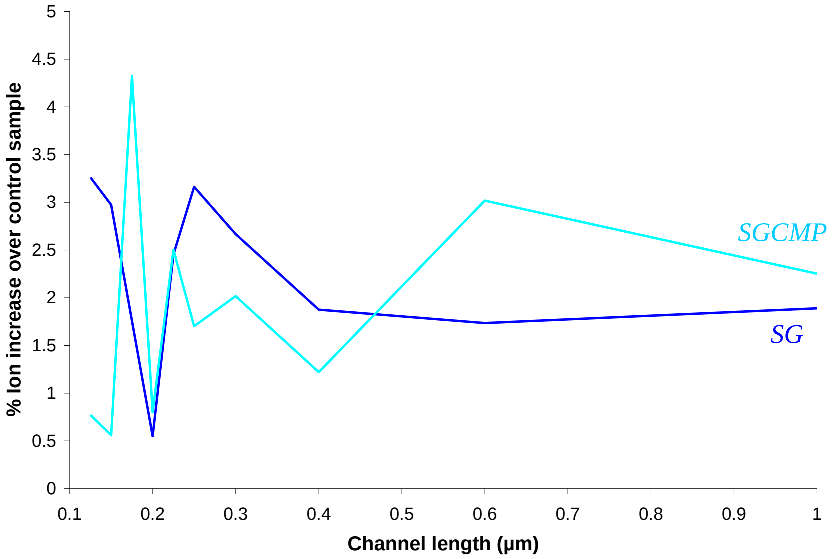

Figure 5-44: Percentage increase of pMOS Ion current for strained devices relative to the Control versus channel length

5.9.3 Batch comparisons

In both batches the nMOS *SG* devices perform similarly showing enhancement which then degrades until it becomes negative at the point where it has been determined that SCEs really start to cause problems. The addition of a nitride layer has little benefit until channel lengths are reduced to approximately 240nm and it would appear that smoothing the VS with CMP can actually degrade performance.

For pMOS the ‘2002’ batch behaves as expected with minimal improvement of around 2%. The ‘2005’ *SGN* devices, which at long gate lengths are essentially *SG* devices as the nitride layer is unable to act on the channel, also show a similarly small improvement of 5%. The ‘2005’ *SG* data is surprising but has been obtained by a colleague working independently on the same wafers. It is likely that VS on this wafer is unusually rough or has a greatly increased number of impurities in the channel and VS.

5.10 I_{off} & I_{on}/I_{off} ratios

Off currents that are large are usually considered bad as they indicate that the MOSFET cannot be switched off by removing the bias from the gate. At V_G=0V the channel is still in the inversion state with charge carriers able to flow from the source to the drain. The I_{on}/I_{off} ratio gives a clear indication of the difference between the I_{on} and I_{off} drain currents. A large ratio is indicative of a good device but such ratios can be misleading.

5.10.1 ‘2005’ batch

The ‘*Low Leakage*’ devices have been fabricated so that the I_D-V_D curve is shifted as in *Figure 4-2* to have a low leakage current at V_G=0V, and the ‘*High Speed*’ devices fabricated to have a higher drive current at a given gate bias, which in turn will make the off currents much larger at V_G=0V. Apart from this V_T optimization the devices are identical. The off current data presented in this section have been taken at V_G-V_T=0V and so like devices in different optimization groups should show almost identical behaviour. This is indeed seen for the nMOS devices (*Figure 5-45*). The *Control* devices all seem to coincide with each other with off currents of ~4x10^{-5} µA/µm at a channel length of 1µm and this value gradually increases up to a maximum value of ~1x10^{-1} µA/µm at a channel length of 1µm. No clear pattern can be identified with regards to the *SG* and *SGN* devices suggesting the addition of the nitride layer has no effect on the off currents. The off currents are considerably higher than for the *Control* devices, in the L=200-300nm region with off currents 1000 times higher. High off currents can normally be attributed to high threading dislocation densities. This data suggests that the VSs are plagued by high numbers of dislocations.

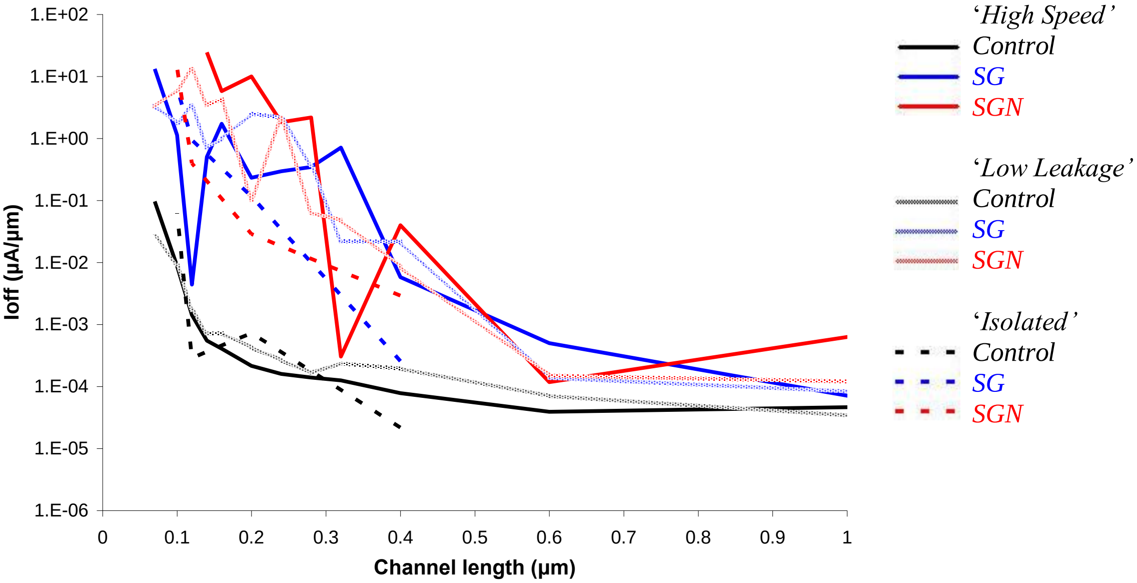

Figure 5-45: Off currents versus cannel length for nMOS devices

The pMOS (*Figure 5-46*) *Control* devices had off currents almost identical in magnitude and rate of change with channel length to the nMOS ones. The *SG* and *SGN* devices had on average an off current only ten times higher than the *Control* devices and the rate at which I_{off} increased was also comparable. Again there appears to be negligible difference between the two types of strained devices indicating the addition of a nitride layer makes little impact. In comparison the *SG* and *SG* off currents match those of the nMOS devices at the longest channel length but remain lower by as much as a 100 times until the onset of SCEs at around L=100nm at which they rise to meet the values of the nMOS devices. As each wafer has both nMOS and pMOS devices on them these results suggest that the pMOS devices are not as affected by the possible high threading dislocation density.

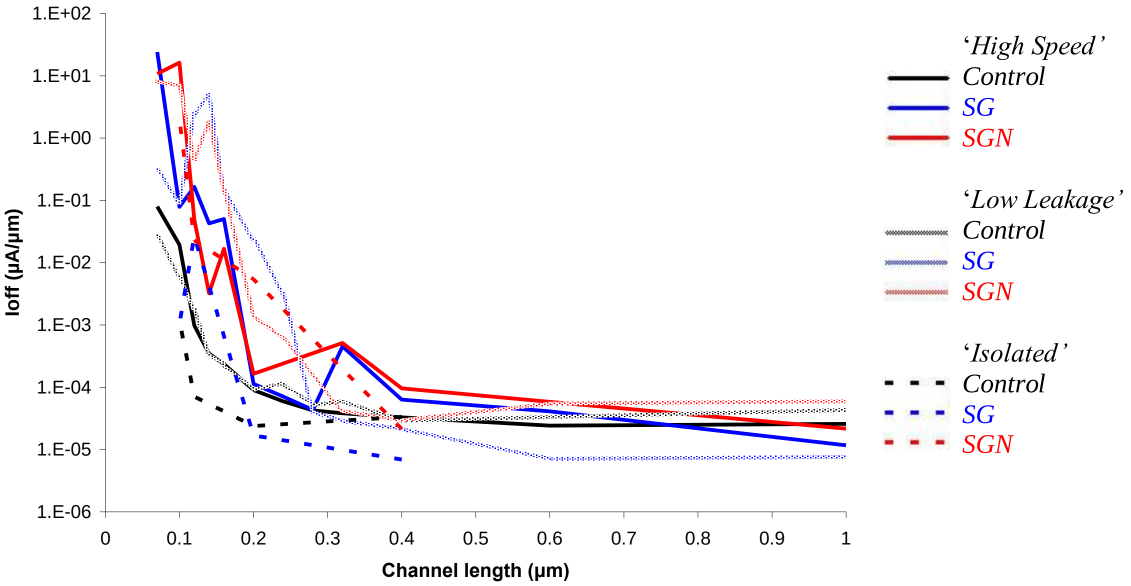

Figure 5-46: Off currents versus cannel length for pMOS devices

Combining the off current and on current data (*Table 5-6*) the *Control* devices continuously outperform the strained devices. This is because although the on currents for the strained devices are slightly higher in general than the *Control* devices, the off currents are a great deal higher by around 10 times for pMOS and up to 1000 times for nMOS. This gives a very poor I_{on}/I_{off} ratio.

		nMOS I_{on}/I_{off}		pMOS I_{on}/I_{off}	
		120nm	1µm	120nm	1µm
'High speed'	*Control*	$1x10^5$	$3x10^5$	$6x10^4$	$1x10^5$
	SG	$3x10^4$	$3x10^5$	$3x10^2$	$1x10^5$
	SGN	-	$4x10^4$	$2x10^3$	$6x10^4$
'Low Leakage'	*Control*	$9x10^4$	$4x10^5$	$3x10^4$	$7x10^4$
	SG	$4x10^1$	$3x10^5$	$2x10^1$	$2x10^5$
	SGN	$1x10^1$	$2x10^5$	$2x10^2$	$5x10^4$
'Isolated'	*Control*	$3x10^5$	-	$4x10^5$	-
	SG	$1x10^2$	-	$9x10^2$	-
	SGN	$3x10^2$	-	$2x10^3$	-

Table 5-6: I_{on}/I_{off} ratios for devices at channel lengths of 1µm and 120nm

5.10.2‘2002’ batch

The off currents for nMOS devices (*Figure 5-47*) remain at a constant level for all three different devices types up until channel lengths of around 300nm, when they rise dramatically. Although a delayed response, as this rise is at a length 70nm shorter than the length identified as being the onset of SCEs (*Table 5-4*), SCEs are still the cause for the rise in off currents. The strained devices have off currents 100 times greater than the *Control* devices, indicating that strain has an adverse affect as once again the dependence on the threshold voltage has been removed. The *SGCMP* devices have off currents double that of the *SG* devices.

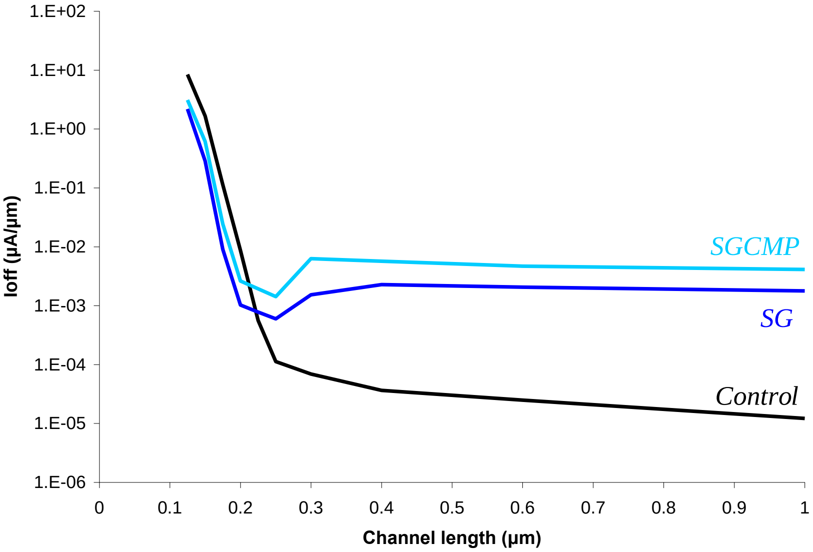

Figure 5-47: nMOS Ion/Ioff ratio versus gate length

The pMOS devices maintain relatively low off currents at all channel lengths and the strain added to the channel does not increase the values much over the *Control* devices. The *SGCMP* again consistently show higher off currents.

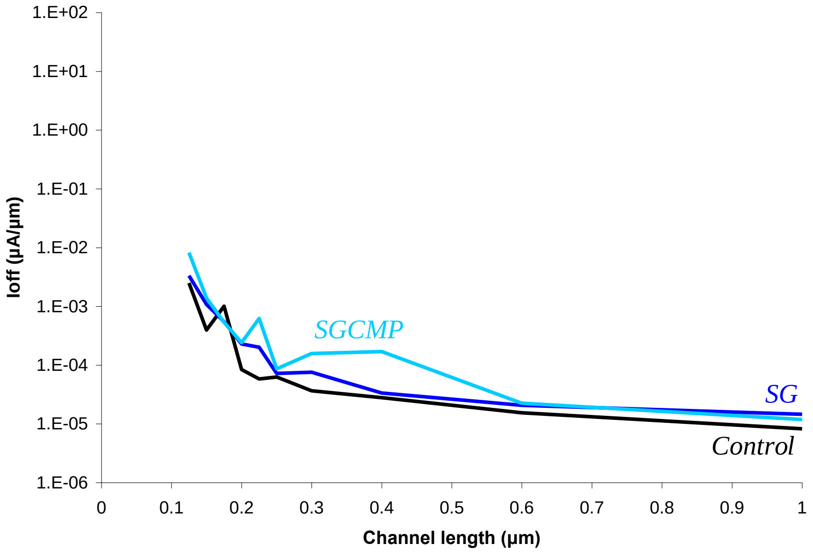

Figure 5-48: nMOS Ion/Ioff ratio versus gate length

The I_{on}/I_{off} ratios (*Table 5-7*) of the nMOS devices suffer as a result of the high leakage current giving only around one decade of difference. The low leakage of the pMOS devices allowed for 3 decades of difference even at the shortest channel lengths.

	nMOS I_{on}/I_{off}		pMOS I_{on}/I_{off}	
	125nm	1µm	125nm	1µm
Control	9	$1x10^6$	$7x10^3$	$4x10^5$
SG	$2x10^1$	$1x10^4$	$6x10^3$	$3x10^5$
SGCMP	$1x10^1$	$5x10^3$	$2x10^3$	$3x10^5$

Table 5-7: I_{on}/I_{off} ratios for devices at channel lengths of 1µm and 125nm

5.10.3Batch comparisons

Both wafer batches highlight the bad leakage currents that plague devices at the very short channel lengths. Strain seems to have a large negative effect on the nMOS devices, but much less of an effect on the pMOS devices. There was negligible difference between the uniaxial and biaxial strained devices in the ‘2005’ batch indicating that while strain impacts the leakage current the type of strain does not. From the ‘2002’ batch it can be seen that leakage is additionally worsened by utilising a CMP stage in the processing of the devices.

5.11 Source/drain resistance

The method used to calculate the source/drain resistance (*Chapter 3.3.4*) assumes that the values for the resistance will be equal for all channel lengths of a particular device type. This is not the case as very short channel devices often show a higher resistance than short channel devices. This makes the determination of $R_{S/D}$ somewhat subjective and dependant on the researcher deciding which device sets to use. These short comings are known and while other methods of determining $R_{S/D}$ are available through the use of capacitance data [Sheu 1984], the Terada-Muta technique is still the most commonly used in the field and so has been used for the resistance data presented in this thesis.

5.11.1 '2005' batch

The values for the source/drain resistance (*Table 5-8*) are very low are will therefore not be a limiting factor with regards to device performance. These low resistances are likely to be due to the use of the metal contacts for these devices. The addition of strain in the channel for nMOS seems to have a large impact increasing the resistance by 40-80% over the *Control* devices. Further strain from the nitride layer has little or no impact.

In the case of the pMOS devices it could be said that the addition of strain also increases the resistance by the same amount as for nMOS if we don't include the errors in the values. The source/drain resistance varies more over the range of channel lengths tested for pMOS and this is reflected in the errors present, making any claims tenuous.

		nMOS $R_{S/D}$ (Ωμm)	pMOS $R_{S/D}$ (Ωμm)
'High speed'	*Control*	100 ± 10	110 ± 30
	SG	180 ± 10	150 ± 150
	SGN	190 ± 30	210 ± 100
'Low Leakage'	*Control*	110 ± 10	90 ± 80
	SG	150 ± 10	110 ± 150
	SGN	150 ±10	120 ± 50
'Isolated'	*Control*	80 ± 10	80 ± 10
	SG	120 ± 10	100 ± 80
	SGN	120 ± 10	110 ± 10

Table 5-8: Source/drain resistance

5.11.2‘2002’ batch

At high doping concentrations, as at the source and drain, resistivity of n-type material is about half that of p-type, mostly because of the higher mobility. This is reflected in the difference between the nMOS and pMOS strained devices (*Table 5-9*). The nMOS *Control* devices however had a contact resistance one third that of the pMOS. This is most likely due to better incorporation of n-type dopants in the Si.

The other trend that can be seen from these devices is that the resistance has increased for the nMOS devices when strain is added, and under the same conditions of strain the resistance is reduced for pMOS devices.

	nMOS $R_{S/D}$ (Ωμm)	pMOS $R_{S/D}$ (Ωμm)
Control	490 ± 10	1600 ± 30
SG	800 ± 10	1380 ± 60
SGCMP	660 ± 10	1450 ± 30

Table 5-9: Source/drain resistance

5.11.3Batch comparisons

As for all other device parameters the ITRS set targets for the source/drain resistance that needed to be met to maintain the current rate of progress. For devices fabricated in 2002 the target was set at 180Ωμm [ITRS 2001] and 2005 also at 180Ωμm [ITRS 2005]. The ‘2002’ batch tested exhibited values up to 4.5 times higher for nMOS and 8.5 times for pMOS devices. However new developments for the ‘2005 batch in contact technology, the use of a NiSi salicidation process, TiN contacts and a reduced S/D implant energy has massively reduced source/drain resistances. The new fabrication methods have also allowed the resistances of the nMOS and pMOS devices to be matched.

5.12 Effective channel length

The effective channel length is given by *Equation 3.11*, and allows the actual channel length that charge carriers use, to be calculated if ΔL is known.

5.12.1‘2005’ batch

Table 5-10 gives the difference (ΔL) between the written channel length and the effective channel length. These values are all positive which means that the effective channel length is ΔL shorter than the written length. Looking at the nMOS devices the strained channel devices indicate that the error in the channel length is smaller when compared to the *Control* devices. The dopant used in the nMOS devices is Arsenic. This diffuses at a greater rate in S-Si and SiGe than Si and so to account for this a smaller doping concentration of Arsenic has been used for the LDD implantation (*Table 4-1*) in the strained devices. This concentration is therefore obviously too low if the device dimensions are to be matched between the wafers.

For the pMOS devices the opposite effect can be seen with ΔL greater for the strained devices. Boron, the dopant used for the LDD implantation for pMOS diffuses at a slower rate in SiGe than Si and at a slightly greater rate in S-Si. As the source and drain structures go into the SiGe VS, the dosage was doubled (*Table 4-1*) for the strained devices. With this increased concentration the geometry of the *Control* and *SG* devices in terms of channel length is very close at between 1-10% depending on which optimization group you are looking at. There has however been an increase in ΔL for the *SGN* devices over the *SG* ones that is not seen for the nMOS devices. The reason behind this is unclear.

		nMOS ΔL (nm)	pMOS ΔL (nm)
‘High speed’	*Control*	60 ± 1	52 ± 1
	SG	45 ± 1	58 ± 8
	SGN	52 ± 1	99 ± 5
‘Low Leakage’	*Control*	67 ± 1	56 ± 4
	SG	31 ± 1	59 ± 5
	SGN	33 ± 1	99 ± 4
‘Isolated’	*Control*	77 ± 1	71 ± 2
	SG	67 ± 1	72 ± 1
	SGN	66 ± 1	78 ± 1

Table 5-10: The difference between the written channel length and the effective channel length for nMOS and pMOS devices

5.12.2‘2002’ batch

Table 5-11 shows that the strained nMOS and pMOS devices have effective channel lengths closer to the written channel length when compared to the *Control* devices. For nMOS, the same doping concentrations were used for the *Control*, *SG* and *SGCMP* devices, and likewise for pMOS (*Table 4-2*).

	nMOS ΔL (nm)	pMOS ΔL (nm)
Control	61 ± 2	60 ± 2
SG	23 ± 3	43 ± 4
SGCMP	38 ± 3	46 ± 2

Table 5-11: The difference between the written channel length and the effective channel length for nMOS and pMOS devices

5.12.3Batch comparisons

This device parameter is dependant on the dopants used and the materials that have been used to create the device. Using a strain technology causes a change the material structure and hence a change in the rate of diffusion in the areas affected. If the magnitude of the strain is known then diffusion rates can be quite accurately calculated. Any desired ΔL can therefore be realistically achieved. This can be seen for the pMOS devices. In the ‘2005’ batch the *SG* devices have a ΔL greater than the *Control* and for the ‘2002’ batch they have a ΔL less than the *Control*.

5.13 Transconductance

The transconductance values of a device in the saturation region (V_D=1.5V) are extremely important as they give an indication of how a research device would perform in a more realistic CMOS environment. Transconductance is given in *Equation 5.3*.

$$\begin{aligned} g_m &= \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D=\text{constant}} \\ &= \frac{C_G}{t_r} \end{aligned} \qquad (5.3)$$

and since

$$f = \frac{1}{t_r} \qquad (5.4)$$

we find that

$$f \propto g_m \qquad (5.5)$$

By increasing the transconductance of a device it is possible to increase its switching speed. Such an improved production chip can now perform more calculations in a given period of time.

The average maximum transconductance value for each device type has been plotted as a function of the channel length in this chapter.

5.13.1 ‘2005’ batch

The nMOS (*Figure 5-49*) strained devices all produce higher g_m values than the *Control* devices until a channel length of 320nm. For shorter channel lengths we see the ‘*High Speed*’ *SG* and ‘*Low Leakage*’ *SGN* devices outperformed while the others still show improvements over the *Control* devices. Since all three optimization groups should be identical taking an average of all of them leads to the result that the strained devices do perform better by around 25% at the longer channel lengths investigated and this slowly increases to a maximum enhancement of ~65% at L=100nm. Shorter channel lengths start to exhibit a reduced transconductance and this is likely to be due to SCEs crippling the devices.

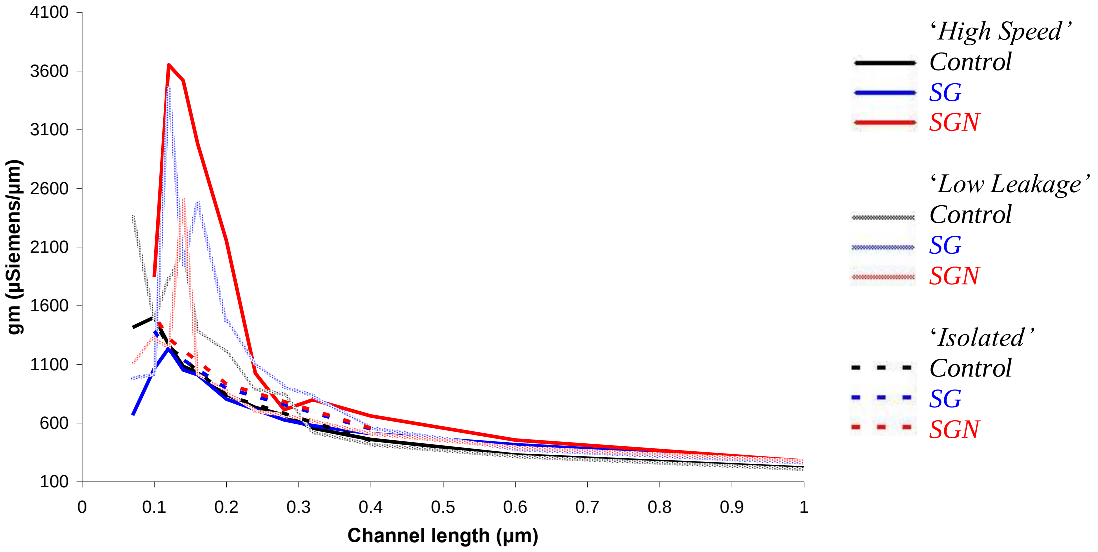

Figure 5-49: Saturation transconductance versus channel length for nMOS devices

From the relevant I_D-V_G graphs (*Figure 5-28* represents a channel length of 200nm) it was expected that the *SG* devices would have a lower transconductance than the *Control* and this is observed (*Figure 5-50*). On average the biaxially strained devices performed 20% worse. This is put down to the surface confinement issues that have already been mentioned.

The *SGN* devices show an almost identical level of degradation at channel lengths less than approximately 300nm after which the tranconductance rises outperforming the *Control* devices by around 20%. This coincides with the longest channel length the nitride layer can place completely under uniaxial strain. Once again at channel lengths around 160nm as defined in *Table 5-3*, the transconductance starts to drop for all devices as the SCEs start to reduce the gates ability to control it.

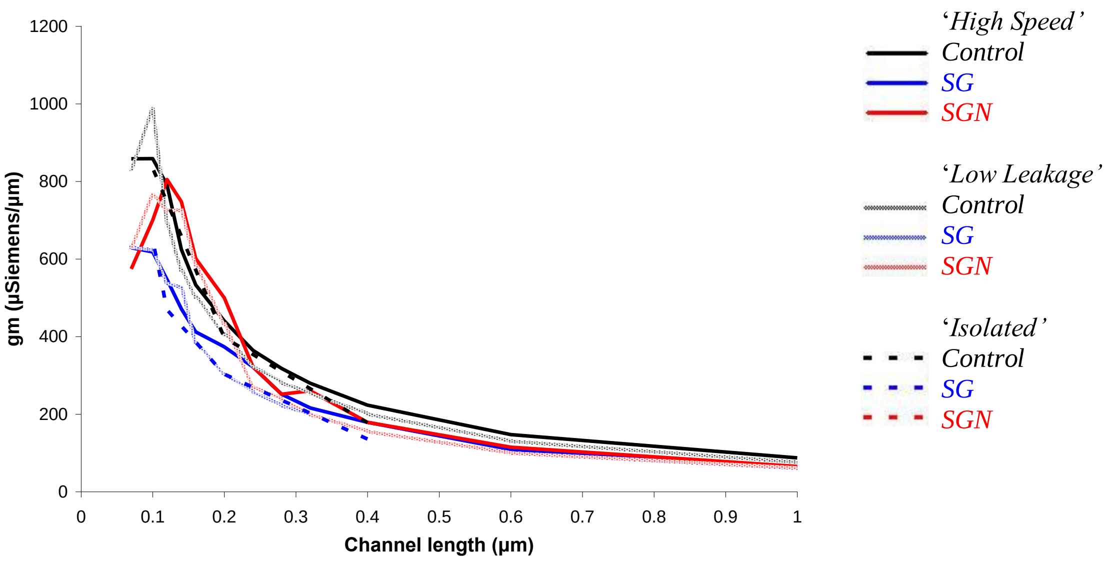

Figure 5-50: Saturation transconductance versus channel length for pMOS devices

5.13.2 '2002' batch

The *SG* and *SGCMP* (*Figure 5-51*) give a higher transconductance up until a channel length of 500nm. After this the *Control* devices perform up to 20% better than those with strained channels. The data also suggests that addition of a CMP stage makes a slight difference of ~1% to the maximum transconductance of the devices if the jumps in values at lengths below 300nm are ignored.

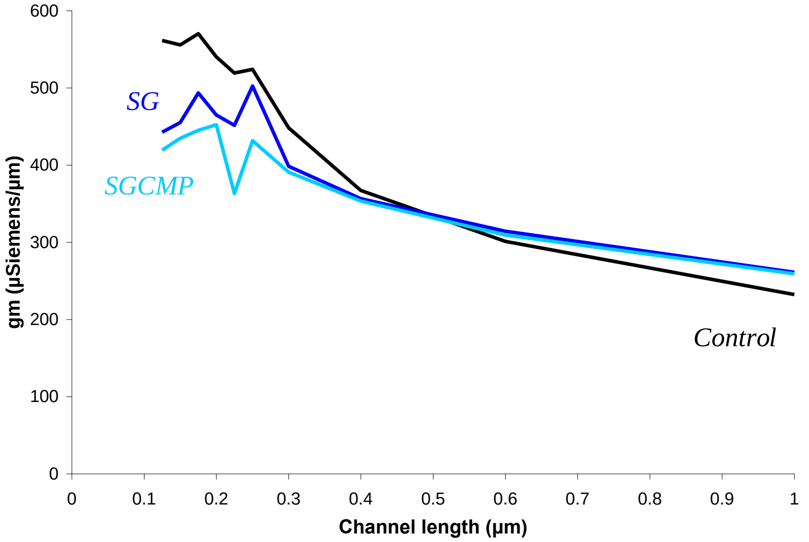

Figure 5-51: Saturation transconductance versus channel length for nMOS devices

pMOS devices (*Figure 5-52*) show that under more realistic CMOS conditions the strained devices perform on average 8% worse than the *Control* devices. In this case the CMP has no impact on the transconductance.

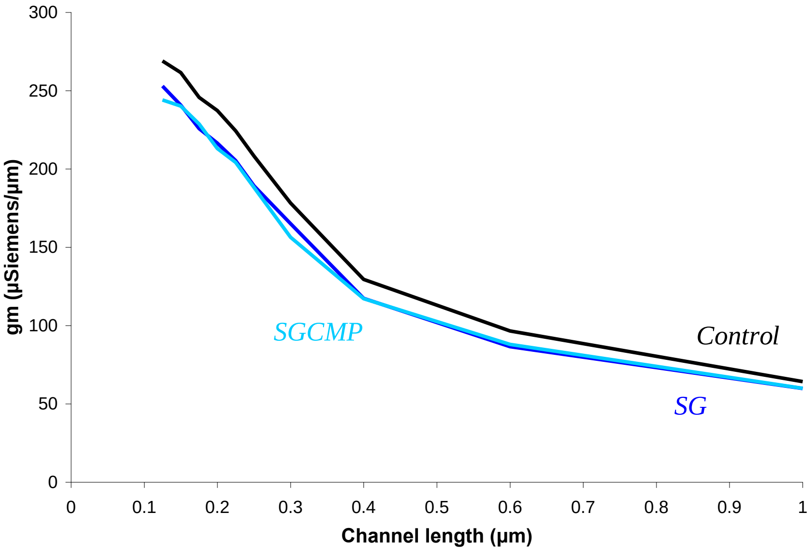

Figure 5-52: Saturation transconductance versus channel length for pMOS devices

5.13.3 Batch comparisons

Both batches show transconductance improvements for the nMOS *SG* devices at longer gate lengths (greater than approximately 400nm) but these are lost as the channel is shortened. The addition of a nitride layer boosts the transconductance when the stress is able to influence the entire length of the channel. CMP seems only to degrade the already rather low values.

Both batches also show similar pMOS behaviour, with the *SG* devices being outperformed at all channel lengths however the ‘2002’ batch did show less degradation in comparison. Once again as channel lengths are reduced to around 200nm the nitride layers influence on the channel brings about enhancements. CMP has no effect in the case of pMOS.

Overall the transconductance values are lower than expected and this is believed to be due to the thermal isolation of the channel from ‘self heating’ which has had a massive negative effect on device performance.

5.14 Device exclusion

For all device sets and wafers, results obtained for devices with channel lengths longer than 1μm have been excluded as they are faulty producing gate leakage currents that can no longer be considered negligible. Prior to sending the '2005' devices to Warwick, STMircroelectrics published a paper illustrating these large gate leakage currents and negative drive currents [Contaret 2006].

6 Conclusion

6.1 Summary

Two batches of strained silicon MOSFETSs have been thoroughly characterised at room temperature. The channel regions were strained by either a relatively high Germanium composition virtual substrate, or by a combination of such a virtual substrate and a deposited nitride layer.

It has been found that at the device geometries tested, strained silicon from a SiGe VS can offer improvements of between 1.2 and 2 times for nMOS and 1 to 1.15 for pMOS, to the carrier mobility (and therefore drive current) over conventional silicon at low drain biases and channel lengths longer than approximately 200nm. The nMOS devices coincide with data previously published with an enhancement of 1.75 times [Rim 2000], as do the modest improvements for pMOS [Thompson 2004b] [Mikkelsen 1982] which are attributed to the surface confinement of carriers at high effective electric fields. As the channel length is reduced the enhancements are lost and this has been attributed to issues relating to the source and drain regions known as Short Channel Effect. When high drain biases were applied in an attempt to simulate more realistic CMOS conditions any enhancements were seen to be further reduced due to excessive ‘self heating’ of the channel.

If a nitride layer is then additionally deposited on top of the gate stack for a MOSFET with a channel length of approximately 200nm or shorter, improvements in performance over those already brought about from the biaxial strain can be obtained at all drain biases. For nMOS the drive current improvement was up to 1.1 times that of the biaxially strained devices and 1.55 times for pMOS. Due to all the factors that determine the exact strain placed on the channel from a nitride layer as described in *Chapter 2.5.4*, it is hard to make direct comparisons to other published work but improvements of these magnitudes have been seen [Ootsuka 2000]. At channel lengths longer than 200nm the nitride layer is not able to fully place the channel under any additional uniaxial strain and so its effects are not seen and this has been corroborated [Payet 2005].

Placing the channel under strain by either of the novel techniques investigated does bring about a number of other negative effects on MOSFET behaviour. Leakage currents are increased, most likely due to the high levels of threading dislocations in the VS and the very thin oxides, as are subthreshold slope values and source and drain resistances. The threshold voltage is also negatively impacted by strain and is shifted by up to 170mV when the channel is biaxially strained and shifted a further 10-20mV when the nitride layer is also employed. These match values published [Lim 2004] [Thompson 2004b] [Goo 2003]. Surprisingly the pMOS devices produced a shift of 1V. This is thought to be due to gate stack damage from the Ge ions and ΔV_{T2} would have been minimal had an intrinsically compressively strained nitride been deposited.

Chemical Mechanical Polishing of the VS during the fabrication stage was also investigated. It has been found that CMP either has no effect on the characteristics of a MOSFET built upon a VS, or actually degrades these characteristics.

6.2 Suggestions for further work

The benefits and the problems associated with biaxial strained silicon devices have been well documented. Further investigation is required with regards to the use of a nitride capping layer as a channel stressor. Currently there are large unknown areas relating to this field especially with regards to pMOS devices. Recreating the devices tested in this thesis with only the nitride layer as opposed to the combined strain method would allow the enhancements and degradation effects attributed to the nitride layer alone to be unequivocally determined. This needs to be done for both compressive and tensile strained nitrides because the type of strain only positively affects either nMOS or pMOS devices [Payet 2005] [Ootsuka 2000] [Ito 200].

Pulsed measurements of the devices tested for this thesis would be invaluable to understanding strained silicon. It has already been seen that drive currents can be reduced by 10% [Nicholas 2004] to 20% [Jenkins 1995] because of the underlying SiGe but newer growth techniques for the ‘2005’ batch may have improved on this. It is also possible that using a nitride layer further thermally isolates the channel region when used in conjunction with a VS.

6.3 Conclusion

Biaxial tensile strained silicon is able to offer benefits to the industry which are currently considered to out way the drawbacks, the largest being cost, hence why it has already been adopted as a technology booster by Intel®. The idea of using a nitride layer is a newer technology, but has already shown promise of eliminating all of the problems associated with the use of a VS to obtain biaxial strain, while hinting at huge enhancements that could be harnessed. Uniaxial strain is easy to implement at low cost, and does not require any special substrate which would isolate the channel. A perceived issue of this method is that it is layout dependant and so benefits are only seen for certain device geometries. These benefits are only realised for device geometries with short channels (less than 200nm) so make it ideal as target gate lengths set by the ITRS are well below this value. The nitride can also be grown with intrinsic strain tailored to offer either nMOS enhancement with a tensile nitride layer, or pMOS enhancement with a compressive nitride layer.

Biaxial strain technology is already allowing the industry to continue moving forward and uniaxial strain appears to be the next logical direction to take. With these two strain technologies in combination with others such as high-k dielectrics and metal contacts the future of the MOSFET seems bright.

7 References

[Arghavani 2004] R.Arghavani, Z.Yuan, N.Ingle, K-B.Jung, M.Seamons, S.Venkataraman, V.Banthia, K.Lilja, P.Leon, G.Karunasiri, S.Yoon, A.Mascarenhas, "*Stress Management in Sub-90-nm Transistor Architecture*", IEEE Trans. Elec. Dev. Vol. 51 No. 10, pg 1740-1743, Oct 2004

[Braga 1994] N.Braga, A.Buczkowski, H.R.Kirk, G.A.Rozgonyi, "*Formation of cylindrical n/p junction diodes by arsenic enhanced diffusion along interfacial misfit dislocations in p-type epitaxial Si/Si(Ge)*" Appl. Phys. Lett. Vol. 65, pg 1410, 1994

[Braunstein 1958] R.Braunstein, A.R.Moore, F.Herman, "*Intrinsic Optical Absorption in Germanium-Silicon Alloys*", Phys. Rev. Vol. 109, No. 3, pg.695-710, Feb 1958

[Choi 2004] W.Y.Choi, H.KIM, B.Lee, J.D.Lee, B.Park, "*Stable Threshold Voltage Extraction Using Tikhonov's Regularization Theory*", IEEE Trans Elec. Dev. Vol.51 No.11, pg1833-1839, Nov 2004

[Contaret 2006] T.Contaret, K.Romanjek, T.Boutchacha, G.Ghibaudo, F.Bœuf, "*Low frequency noise characterisation and modelling in ultrathin oxide MOSFETs*", Solid-State Elec. 50, pg. 63-68, 2006

[Currie 1998] M.T.Currie, S.B.Samavedam, T.A.Langdo, C.W.Leitz, E.A.Fitzgeral, "*Controlling threading dislocation densities in Ge on Si using graded SiGe layers and chemical-mechanical polishing*", Appl. Phys. Lett., Vol 72 No. 14, April 1998

[Davari 1995] B.Davari, R.H.Dennard, G.GShahidi, "*CMOS scaling for high performance and low power-the next tem years*", IEEE Proceedings, Vol 83, Issue 4, pg. 595-606, April 1995

[Dennard 1974] R.H.Dennard, F.H.Gaensslen, V.L.Rideout, E.Bassous, A.R.LeBlanc, "*Design of ion-implanted MOSFET's with very small physical dimensions*", IEEE Journal of Solid State Circuits, Vol 9, Issue 5, pg.256-268, Oct 1974

[Dismukes 1993] J.P.Dismukes, L.Ekstrom, R.J.Paff , "Lattice Paramter and Density in Germainum-Silicon Alloys", J. Phys. Chem. Vol. 68 No. 10, pg. 3021-3027, 1993

[El-Kareh 1990] B.El-Kareh, W.R.Tonti, S.L.Titcomb, "*A Submicron MOSFET parameter extraction technique*" IBM J. Res. Develop. Vol.34, No.2/3, pg.243-249, March/May 1990

[Fiorenza 2004] J.G.Fiorenza et al, "*Film thickness constraints for manufacturable strained silicon CMOS*", Semicond. Sci. Technol., Vol. 19, pg L4-L8, 2004

[Fischetti 1996] M.V.Fischetti, S.E.Laux, "*Band structure, deformation potentials, and carrier mobility in strained-Si, Ge and SiGe alloys*", J. Apply. Phys. Vol.80, pg. 2234-2253, 1996

[Fong 1976] C.Y.Fong, W.Weber, J.C.Phillips, "*Violation of Vegard's law in covalent semiconductor alloys*", Phys. Rev. B, Vol. 14, pg.5387-5391, 1976

[Fossum 2003] J.G.Fossum, W.Zhuang, "*Performance projections of scaled CMOS devices and circuits with strained-Si-on-SiGe channels*", IEEE Trans. Elec. Dev. Vol. 50, pg. 1042-1048, April 2003

[Gehrsitz 1999] S.Gehrsitz, H.Sigg, N.Herres, K.Bachem, K.Köhler, F.K.Reinhart, "*Compositional dependence of the elastic constants and the lattice parameter of $Al_xGa_{1-x}As$*", Phy. Rev. B, Vol. 60 No.16, pg11601-11610, 1999

[Gomeniuk 1999] Y.V. Gomeniuk, V.S. Lysenko, I/N. Osiyuk, I.P. Tyagulski, M.Ya. Valakh, V.A. Yukhimchuk, M. Willander, C.J. Patel, "*Properties of SiGe/Si heterostructures fabricated by ion implantation technique*", Semiconductor Physics, Quantum Electronics & Optoelectronics, V.2 N.3, pg 74-80, 1999

[Harrison 2002] P.Harrison, "Quantum Wells, Wires and Dots: Theoretical and computational Physics", A Wiley Publication, 2002

[Huang 2005] L. Huang, M. B. H. Breese, E.J. Teo, "*Characterisation of 60^o misfit dislocations in SiGe alloy using nuclear microscopy*", Nuclear Instruments and Methods in Physics Research B 231, pg 452-456, 2005

[Ito 200] S.Ito et al, "*Mechanical Stress Effect of Etch-Stop Nitride and its Impact on Deep Submicron Transistor Design*", IEEE IEDM, pg 247-250, 2000

[ITRS 2001] "*International Technology Roadmap for Semiconductors*" S.I.Association, Edition 2001

[ITRS 2005] "*International Technology Roadmap for Semiconductors*" S.I.Association, Edition 2005

[Jenkins 1995] K.A.Jenkins, J.Y.C.Sun, "*Measurement of I-V curves of Silicon-on-Insulator (SOI) MOSFETs Without Self-Heating*", IEEE Elec. Dev. Lett., Vol.16 No.4, pg. 145-147, April 1995

[Jenkins 2002] K.A.Jenkins, K.Rim, "*Measurement of the Effect of Self-Heating in Strained-Silicon MOSFETs*", IEEE Elec. Dev. Lett. Vol. 23, No. 6, June 2002

[Jenkins 1995] T.E. Jenkins, "*Semiconductor Science: Growth and Characterization Techniques*", Prentice Hall Publication, 1995

[Kotchetkov 2001] D.Kotchetkov, J.Zou, A.A.Balandin, D.I.Florescu, F.H.Pollak, "*Effect of dislocations on thermal conductivity of GaN layers*", App. Phys. Lett. Vol. 79, No. 26, Dec. 2001

[Leitz 2002] C.W.Leitz, M.T.Currie, M.L.Lee, Z.Y.Cheng, D.A.Antonaidis, E.A.Fitzgerald, "*Hole mobility enhancements and alloy scattering-limited mobility in tensile strained Si/SiGe surface channel metal-oxide-semiconductor field-effect transistors*", J. Appl. Phys., Vol.92 No.7, pg. 3745-3751, Oct 2002

[Lim 2004] J.Lim, S.E.Thompson, J.G.Fossum, "*Comparison of Threshold-Voltage Shifts for Uniaxial and Biaxial Tensile-Stressed n-MOSFETs*", IEEE Elec. Dev. Lett. Vol. 25, No. 11, pg 731-733, Nov 2004

[Lyutovich 2004] K. Lyutovich, M. Oehme, F. Ernst, "*Growth of ultra-thin and highly relaxed SiGe layers under in-situ introduction of point defects*", Eur. Phys. J. Appl. Phys. 27, pg 341-344, 2004

[Mackenzie 2004] K.D.Mackenzie, B.Reelfs, M.W.DeVre, R.Westerman, D.J.Johnson, "*Characterization & Optimization of Low Stress PECVD Silicon Nitride For Production GaAs Manufacturing*", Compound Semiconductor Manufacturing Technology, 2004 Digest

[Matthews 1974] J.W.Matthews, A.E.Blakeslee, "*Defects in epitaxial multilayers: I. Misfit dislocations*", J. Cryst. Growth, Vol. 27, pg 118, 1974

[Mikkelsen 1982] J.C.Mikkelsen Jr, J.B.Boyce, "*Atomic-Scale Structure of Random Solid Solutions: Extended X-ray Absorption Fin-Structure Study of* $Ga_{1-x}In_xAs$", Phys. Rev. Lett. Vol. 49, pg 1412-1415, 1982

[Nabarro 1967] R.N.Nabarro, "Theory of Crystal Dislocation", Oxford University Press Publication, pg464, 1967

[Nash 2005] L.J.Nash, "*Characterisation of Terrace Graded Virtual Substrates with* $Si_{1-x}Ge_x$ $0.15 \leq x \leq 1$", PdD, September 2005

[Nicholas 2004] G.Nicholas, "*Investigation into the Electrical Properties of Tensile Strained Silicon MOSFETs*", PhD, June 2004

[Nikulin 1996] Y.Nikulin, A.W.Stevenson, H.Hashizume, "*Model-independent determination of the strain distribution for a Si0.9Ge0.1/Si* *superlattice using x-ray diffractometry data*”, Phys. Rev. Vol. 53, pg8277-8282, 1996

[Oberhuber 1998] R.Oberhuber, G.Zandler, P.Vogl, “*Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET’s*”, Physical Review B, Vol. 58, No.15, October 1998

[Ootsuka 2000] F.Ootsuka, S.Wakahara, K.Ichinose, A.Honzawa, S.Wada, H.Sato, T.Ando, H.Ohta, K.Watanabe, T.Onai, “*A Highly Dense, High-Performance 130nm node CMOS Technology for Large Scale System-on-a-chip Applications*”, IEDM Tech. Dig., pg 57-578, 2000

[Parker 2004] G.Parker, “*Introductory Semiconductor Device Physics*” Institute of Physics Publication, Pg206-208, 2004

[Payet 2005] F.Payet, “ *Modelisation et Integration de Transistors a Canal de Silicium Constraint pour les Nœuds Technologiques CMOS 45nm et en deça*”, Université de Provence – Aix Marseille I, 2005

[People 1985] R.People, J.C.Bean, “*Calculation of critical layer thickness versus lattice mismatch for GexSi1-x/Si strained-layer heterostructures*”, Appl.Phys.Lett., Vol.47, No.3, pg.322-324, May 1985

[Rim 2000] K.Rim, J.L.Hoyt, J.F.Gibbons, “*Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFETs*”, IEEE Trans. Elec. Dev. Vol.47, No. 7, July 2000

[Roldan 1997] J.B.Roldan, F.Gamiz, J.A.Lopez-Villaneuva, J.E.Carceller, “*Understanding the improved performance of strained Si/Si1-xGex channel MOSFETs*”, Semicond. Sci. Technol. 12, pg. 1603-1608, 1997

[Schaffler 1997] F.Schaffler, “*High-mobility Si and Ge structures*”, Semiconductor Science and Technology, Issue 12, pg.1515-1549, 1997

[Sheu 1984] B.J.Sheu, P.K.Ko, “*A capacitance method to determine channel lengths for conventional and LDD MOSFETs*”, IEEE Elec Dev. Lett. Vol. 5, No. 11, pg.491-492, Nov 1984

[Shimizu 2004] A.Shimizu, K.Hachimine, N.Ohki, H.Ohta, M.Kogucki, Y.Nonaka, H.Sato, F.Ootsuka, “*Local Mechanical-Stress Control (LMC): A New Technique for CMOS-Performance Enhancement*”, IEEE 2004

[Singh 1994] J.Singh, “*Semiconductor Devices:An Introduction*”, A McGraw-Hill Publication, pg 405-436, 1994

[Sugii 2002] N.Sugii et al, “*Performance enhancement of strained-Si MOSFETs fabricated on a chemical-mechanical-polished SiGe substrate*”, IEEE Trans Elec. Dev, Vol. 49, Pg. 2237-2243, Dec 2002

[Thompson 2004a] S.E. Thompson et al, “*A Logic Nanotechnology Featuring Strained-Silicon*”, IEEE Elec. Dev. Let. Vol 25 No.4, April 2004

[Thompson 2004b] S.E.Thompson, G.Sun, K.Wu, J.Lim, T.Nishida, “*Key Differences For Process – induced Uniaxial vs. Substrate – induced Biaxial Stressed Si and Ge Channel MOSFETs*”, IEEE, 2004

[Thompson 2004c] S.E.Thompson et al, “A 90-nm Logic Technology Featuring Strained-Silicon”, IEEE Trans. Elec. Dev. Vol.51, No.11 Nov 2004

[Toda 2001] A.Toda, N.Ikarashi, H.Ono, S.Ito, T.Toda, K.Imai, “*Local lattice strain distribution around a transistor channel in metal-oxide – semiconductor devices*”, Appl.Phys.Lett. Vol.79, No.25, pg 4243-4245, Dec 2001

[Toh 2005] S.L.Toh, K.P.Loh, C.B.Boothroyd, K.Li, C.H.Ang, L.Chan, “*Strain analysis in silicon substrates under uniaxial and biaxial stress by convergent beam electron diffraction*”, J.Vac.Sci.Technol.B 23(3), May/June 2005

[Tsuno 1999] M.Tsuno, M.Suga, M.Tanaka, K.Shibahara, M.Miura-Mattausch, M.Hirose, “*Physically-Based Threshold Voltage Determination for MOSFET’s of All Gate Lengths*” IEEE Trans. Elec. Dev. Vol.46, No.7, Pg.1429-1434, July 1999

[Vdovin 2002] V. I. Vdovin, M. Mühlberger, M. M. Rzaev, F. Schäffler, T. G. Yugova, “*Interface structure formation in SiGe/Si(001) heterostructures with low-temperature buffer layers*”, J. Physc. Condens. Matter 14, pg 13313-13318, 2002

[Wolf 1995] S.Wolf, ‘*Silicon Processing for the VLSI era Volume 3: The Submicron MOSFET’*, Lattice Press, pg 213-222, 1995

[Xiang 2003] Q.Xiang, J.Goo, J.Pan,B.Yu, S.Ahmed, M.R.Lin, “*Strained silicon nMOS with nickel-silicide gate*”, Symp. VLSI Tech. Dig pg101-102, 2003

[Zhao 2003] W.Zhao, J.He, R.E.Belford, L.E.Wernersson, A.Seabaugh, “*Performance depleted SOI MOSFETs under uniaxial tensile strain*”, IEE Trans. Elec. Dev. Vol. 51, Pg. 317-323, Feb 2003