STRAINED-SI TECHNOLOGY FOR RF POWER LDMOSFET

by

Niamh Waldron
B.Sc., Applied Physics
University of Limerick, 1996

Submitted to the Department of Electrical Engineering and Computer Science In Partial Fulfillment of The Requirements for the Degree of Master of Science in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

September 2002

© 2002 Massachusetts Institute of Technology
All rights reserved

Signature of Author....................................... 
Department of Electrical Engineering and Computer Science 
August 30, 2002

Certified by.......................................................... Jesús A. del Alamo 
Professor of Electrical Engineering 
Thesis Supervisor

Accepted by............. Arthur C. Smith 
Chairman, Department Committee on Graduate Students
Strained-Si Technology for RF Power LDMOSFET

by

Niamh Waldron

Submitted to the Department of Electrical and Computer Engineering
on August 30, 2002 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Electrical Engineering and Computer Science

ABSTRACT

This thesis studied the application of strained-Si technology to RF power LDMOSFETs. Key issues for its implementation were determined to be thermal budget restrictions, gate oxide formation and impact ionization effects.

2D simulations were carried out to explore the design space of the strained-Si LDMOSFET. In order to address the thermal budget restrictions, use of a high-tilt implant for the body doping was investigated. For a dose of $1.5 \times 10^{13} \text{cm}^{-2}$, the conditions for the body implant that resulted in the best output characteristics, as determined by $g_m$, DIBL and $r_o$, were 50 keV energy with a tilt of 60°. The major trade-off of the n-drift region was that of breakdown vs. on-resistance.

Loss of strained-Si in CMOS during the gate oxide formation was found to be a potential issue for System-on-Chip (SOC) applications. Two options for the implementation of a 10 nm gate oxide were assessed. Option one was a 750 °C dry/wet/dry thermal oxidation on a thick strained-Si layer. Option two was a composite oxide consisting of a thin dry oxidation followed by an LTO deposition. Capacitor structures were fabricated and tested. Both options exhibited good characteristics as determined by C-V, leakage and Dit measurements.

TLM structures were fabricated to investigate impact ionization effects in the strained-Si/SiGe heterostructure. Preliminary analysis of the structures show that there is a significant difference in II generation between the control bulk Si and strained-Si samples. For the same source current levels, the strained-Si samples had body current that was an order of magnitude higher than bulk Si. Lower saturation current levels were observed in the strained-Si structures compared to bulk Si. Self-heating had an effect in the strained-Si samples but was not thought to be solely responsible for the lower current levels.

Thesis Supervisor: Jesús A. del Alamo

Title: Professor of Electrical Engineering
Acknowledgements

Firstly I would like to thank my thesis advisor Jesús del Alamo for his support and guidance over the course of this work. His enthusiasm, patience, advice and time have been much appreciated. Hopefully I won’t wear his red pen out too much in the next few years!

I would also like to thank the staff of the MTL for all their help in the fab and in particular Bernard Alamariu, Bob Bicchieri and Paul Tierney.

Thanks to the del Alamo group – Jim, Anita, Joyce, Samuel, Joerg, Joachim and Tetsuya – for making MIT an interesting and fun place to work. This goes for all on the 6th floor. I’d like to especially thank Jim Fiorenza for all those lively LDMOSFET chats and for helping me get started when I first got here.

Thanks to all in the SiGe group and particularly Judy Hoyt for all her advice and helpful discussions on SiGe issues. The heterostructures used in this work were grown by Minjoo L. Lee and Arthur Pitera – thanks guys.

Finally, I’d like to thank my parents, Joan and Derrie, who have always supported and encouraged me in everything I’ve chosen to do.

This research was funded by the Defence Advanced Reseatch Projects Agency (DARPA).
Contents

1. Introduction .......................................................................................................... 15
  1.1 Strained-Si Technology ................................................................................. 16
  1.2 Direction of Thesis ........................................................................................ 19

2. Design of an RF LDMOSFET for Strained-Si Technology .................. 21
  2.1 The Silicon LDMOSFET ............................................................................. 21
  2.2 Process Considerations for the Strained-Si/SiGe System ......................... 23
  2.3 Proposed Structure of the Strained-Si LDMOSFET ............................... 27
  2.4 Heterostructure Design ............................................................................. 29
  2.5 Doping Design ............................................................................................ 30
    2.5.1 Body Doping ......................................................................................... 31
    2.5.1.1 Implant Splits for strained-Si Body Doping ................................. 34
    2.5.1.2 Under Source Body Link .................................................................... 42
    2.5.2 Drift Region Design ............................................................................. 43
    2.5.2.1 Doping Splits ................................................................................. 48
  2.6 Conclusions .................................................................................................... 52

3. Strained-Si LDMOSFET Gate Oxide ......................................................... 55
  3.1 Gate Oxide Options ...................................................................................... 55
    3.1.1 Conventional Dual Gate Oxide Process ............................................. 55
  3.2 Gate oxide options ...................................................................................... 57
    3.2.1 Test Structure Process Flow ................................................................. 58
  3.3 Analysis of Results ...................................................................................... 59
    3.3.1 C-V Measurements ............................................................................. 60
    3.3.2 I-V Measurements .............................................................................. 60
    3.3.3 D_T Measurements ............................................................................ 61
  3.4 Conclusions .................................................................................................... 63

4. Impact Ionization Effects in Strained-Si/SiGe ...................................... 65
  4.1 Design of Test Structure ............................................................................. 65
    4.1.1 Substrate Preparation ......................................................................... 67
    4.1.2 Process Flow ......................................................................................... 68
  4.2 Device Layout and Parasitics Extraction .................................................. 70
    4.2.1 R_{link} Extraction ............................................................................. 72
    4.2.2 R_{N+} Extraction ............................................................................. 73
    4.2.3 R_c Extraction .................................................................................. 73
    4.2.4 Normalization of the Width .................................................................. 74
4.3 Initial Analysis of Results.................................................................76
  4.3.1 Parasitic Extraction from Initial Measurements.........................76
  4.3.2 Junction Quality........................................................................78
  4.3.3 TLM Test Structure Measurements...........................................78
  4.3.4 Impact Ionization in the TLM structures...................................84
  4.4 Conclusions..................................................................................86

5. Conclusions.....................................................................................87

Appendix A.........................................................................................91
Appendix B.........................................................................................97
Bibliography.......................................................................................99
List of Figures

Figure 1-1: Output power of selected semiconductor technologies as a function of frequency. Light shaded region indicates expected Si RF-SOC capabilities of a SiGe-BiCMOS technology. Dark shaded region indicates goals for a new all-FET SiGe-SOC technology............................................................................................. 16

Figure 1-2: Intrinsic transconductance and carrier velocity as a function of gate length for strained and unstrained Si MOSFETs [4]...............................................................17

Figure 2-1: Schematic cross section of the Motorola Rf LDMOSFET [10] ................ 22

Figure 2-2: Outline of the strained-Si/SiGe heterostructure for (a) a bulk design and (b) SGOI. ..........................................................................................................................24

Figure 2-3: Critical thickness of a pseudomorphic SiGe layer grown on Si as a function of Ge fraction [12]...................................................................................................... 25

Figure 2-4: Schematic cross section of the proposed strained-Si RF-power LDMOSFET Inset shows the epitaxial layer structure. ............................................................... 28

Figure 2-5: Mobility enhancement in strained-Si as a function of Ge content in the relaxed SiGe buffer layer [21]. ......................................................................................... 30

Figure 2-6: Example of a MEDICI grid used in the process simulations. The mesh and regions are shown in (a) and the different doping regions in (b).................................31

Figure 2-7: Simulated (a) doping (b) field and (c) electron concentration profiles for a graded and uniform LDMOSFET channel biased at $V_{gs} = 3.6V$, $V_{ds} = 7V$. .......... 33

Figure 2-8: Comparison of the transconductance of a graded vs. uniformly doped channel. The uniform channel is doped to match the threshold voltage.................35

Figure 2-9: Arsenic 55 keV and Boron 60 keV as implanted and post 1000 °C RTA spike anneal doping profiles.................................................................36

Figure 2-10: Threshold voltage for each of the body doping splits...............................37

Figure 2-11: Boron doping profiles along channel for 50 keV implants with different tilt angles. The edge of the gate is at 0 um. The 0° tilt implants did not penetrate under the gate and so had a uniform channel doping $1x10^{16}$ cm$^{-3}$ .................38
Figure 2-12: Boron doping profiles along channel for 60° tilt implants for different energies. Note the lower doping concentration level for the 60 keV implant. .......38

Figure 2-13: Peak doping concentration along the channel interface for different body doping implant conditions. The 0 tilt implants are not shown as their peak doping concentration is that of the background doping 1x10^16 cm^-3 ........................................................................39

Figure 2-14: Peak transconductance of each of the body doping splits at V_ds = 3.6 V. ....39

Figure 2-15: \( L_{eff} \) for each of the body doping splits. The 0° tilt \( L_{eff} \) is that of the drawn device ~0.6um and is not shown .................................................................................40

Figure 2-16: DIBL performance for each of the body doping splits .........................41

Figure 2-17: Output resistance of each of the body doping splits. ..............................41

Figure 2-18: Boron doping profiles underneath the N+ source region away from the gate edge for different implant energies for 60° tilt. The x-axis represents the depth into the substrate. ..................................................................................................................43

Figure 2-19: Electric field contours for \( V_{GS} = 0 \) and \( V_{DS} = 8\) V are shown in (a) with the inset scale varying from 2x10^5 V/cm to 7.6x10^5 V/cm. The corresponding impact ionization generation rate is shown in (b) with the inset scale varying from 12 to 20 on a log scale. Source and drain region are outlined. ...........................................44

Figure 2-20: Effect of the drift region length (a) and dose (b) on the lateral electric field at the gate edge. ..................................................................................................................46

Figure 2-21: Output characteristics demonstrating the effect of adding a drift region to the device. Once the drift region resistance becomes dominant, the current drive becomes limited and the transconductance drops off rapidly .....................................47

Figure 2-22: Offstate breakdown voltage for different drift region designs. ...............49

Figure 2-23: Resistance associated with each of the drift regions......................................50

Figure 2-24: \( I_{DSAT} \) achieved for different drift region designs. ..........................................................50

Figure 2-25: \( f_t \) for different drift region designs. The scale is normalized to the 0.25um length, 1x10^13 cm^-2 dose value ................................................................................................................51

Figure 3-1: Flow for a typical dual gate oxide process .............................................56

Figure 3-2: Diagram of the capacitor test structure used to evaluate the different gate oxide options ..........................................................59
Figure 3-3: High frequency capacitance measurements for both gate oxide options. The fully thermally grown option is shown on the left and the composite oxide scheme on the right. The $C / C_{ox}$ minimum for is different for both options as the substrate doping was $\sim 2 \times 10^{17}$ cm$^{-3}$ for the composite oxide scheme and $\sim 1 \times 10^{15}$ cm$^{-3}$ for the fully thermal oxide.

Figure 3-4: Gate oxide leakage measurements for both gate oxide schemes.

Figure 3-5: Examples of the Gp/w curves obtained for the (a) fully thermal oxide and (b) composite oxide schemes. The data for (a) has been taken from the Si control sample.

Figure 4-1: Test structure design to examine II effects in strained-Si. Electrons are accelerated from the drain and create electron-hole pairs as they collide with the lattice. The hole current collected by the body is a measure of the impact ionization.

Figure 4-2: Simulation of the II test structure showing the increasing hole current with increasing field (voltage).

Figure 4-3: Process flow for the II test structure.

Figure 4-4: Layout of (a) full device showing metal pads and (b) a more detailed look at the actual resistor structure.

Figure 4-5: Cross section of actual device showing the parasitic resistances. These must be extracted from the data to get the true voltage across the n-region.

Figure 4-6 Kelvin structure used to measure $R_{\text{link}}$.

Figure 4-7 VDP structure for $R_{\text{sh}}$ extraction.

Figure 4-8 Cross-bridge Kelvin structure for $R_{\text{c}}$ extraction.

Figure 4-9: Plot of resistance of Kelvin structures vs. width as drawn and as corrected for offset. Reading the width from the resistance curve directly from the measured data would result in significant errors at narrow widths.

Figure 4-10: I-V characteristics of a 1um long TLM structure for as drawn widths of 1um and 10um. There is a larger correction for wider device as $R_{\text{parasitic}}$ does not scale with width.

Figure 4-11: N+/substrate diode breakdown for the bulk and strained Si samples. The breakdown is repeatable across the different splits.
Figure 4-12: Comparison of the different length TLM structures for n- doping level of $1 \times 10^{13}$ cm$^2$. Filed is calculated from $V_{\text{intrinsic}}/L$.................................................................80

Figure 4-13: Comparison of the different length TLM structures for n- doping level of $5 \times 10^{12}$ cm$^2$. Filed is calculated from $V_{\text{intrinsic}}/L$.................................................................81

Figure 4-14: Comparison of the different length TLM structures for n- doping level of $5 \times 10^{12}$ cm$^2$. Filed is calculated from $V_{\text{intrinsic}}/L$.................................................................82

Figure 4-15: Normalized Is vs. Field for different widths of the 1um TLM structure for $1 \times 10^{13}$ cm$^{-2}$ dose. The strained Si samples exhibit self heating effects. Bulk Si does not show this effect. ..................................................................................84

Figure 4-16: (a) I-V characteristics of a 1um TLM structure. The body current is due to $I_B$ across the n- region. The relative magnitude of the $I_B$ is determined by $M$ which is plotted in (b)..............................................................................85
List of Tables

Table 4-1: Extracted resistances for each of the parasitic elements of the TLM structure. $R_{\text{parasitic}}$ is the combined total of all the parasitic elements. Details for all drawn widths are given in Appendix B. .................................................................76

Table 4-2: Resistance of the n- region as measured from a 100um x 10um TLM structure ...............................................................................................................................................................................82

Table B-1: Extracted parasitic resistances for all TLM widths...........................................99
CHAPTER 1

Introduction

The rapid growth in the wireless communication market in recent years has driven the development of RF technology. The electronics market has seen a steep rise in the sale of products for applications such as cellular phones, wireless networking, internet communications and satellite transmitters. The implementation of the power amplifier function of these systems in the intermediate frequency range of 5-20 GHz has predominantly been fulfilled by GaAs technology. Advancements in SiGe HBT may allow performance of up to 10 GHz depending on the power level of the application, but as of yet there is no current prospect for Si-based technology to operate in the >10GHz range.

This project proposes to examine the emerging strained Si/SiGe MOSFET technology for the development of high performance RF power devices that can operate in the 10-20 GHz range at power levels of 100s mWs (Fig. 1.1). Such a device technology will rival the performance of GaAs technology but at a lower cost and with the potential for system-on-chip integration. The lower cost aspect of the technology will be a critical factor for the penetration of 3rd generation and future wireless technology products into the very competitive high-end consumer marketplace.
Figure 1-1: Output power of selected semiconductor technologies as a function of frequency. Light shaded region indicates expected Si RF-SOC capabilities of a SiGe-BiCMOS technology. Dark shaded region indicates goals for a new all-FET SiGe-SOC technology.

1.1 Strained-Si Technology

The strained Si/SiGe MOSFET is a newly emerging technology that exploits the enhanced fundamental transport properties of strained Si resulting in superior performing devices. Improvements of over 80% in current drive and tranconductance over standard silicon devices have been reported [1], [2].

The strained-Si/SiGe is a heterostructure system that consists of pseudomorphic Si layer grown on a relaxed Si$_{1-x}$Ge$_x$ buffer layer. There is a mismatch between the lattice constant of the underlying relaxed Si$_{1-x}$Ge$_x$ and Si. If the Si layer is grown thin enough, its lattice constant will conform to that of the buffer layer hence forming a pseudomorphic strained layer. The strain induced in the thin Si layer results in a splitting of the
degeneracy of the conduction band. This is the physical basis for its improved transport properties. The band splitting leads to decreased phonon scattering and a decrease in the effective transport mass which result in an enhanced carrier mobility.

However, if just the low field mobility were to increase, benefits would not be very significant in deep submicron devices where velocity saturation effects dominate. Experimentally, however, this is what is observed (Fig. 1.2). Increased saturation velocity and transient velocity overshoot [3] have been postulated to explain the higher performance of strained-Si over conventional Si in these scaled devices.

![Figure 1-2: Intrinsic transconductance and carrier velocity as a function of gate length for strained and unstrained Si MOSFETs [4].](image)

Strained-Si technology has been researched by commercial semiconductor manufacturers for its application to devices in the sub-100nm CMOS nodes. Both Toshiba and IBM have reported encouraging results using both Strained-Si-on-Insulator (SGOI) and bulk approaches. Mizuno et al at Toshiba have reported a structure based on a Si$_{0.9}$Ge$_{0.1}$ buffer which shows an 84% and 38% increase in the drive current of the p-
and n-MOSFET respectively [5]. Rim et al. at IBM [6] have also demonstrated a current drive increase of 35% in an n-MOSFET with an $L_{\text{eff}} < 70\text{nm}$. Excellent turn-off characteristics of the strained-Si device in comparison to the Si control were observed with subthreshold slopes of 82 and 85mV/dec and DIBL of 70 and 80mV (with a gate overdrive of 0.8V) respectively.

The strained-Si system obviously has potential for digital logic applications. The question remains then, does it have potential for analog RF applications? If so, then the attractive possibility of RF System on Chip (SOC) exists. The most problematic function of an RF system is that of the power amplifier (PA). The receiving chain of the RF system - the LNA, mixer, VCO and filters have been successfully integrated on the same chip but the PA still predominantly tends to be separate. Silicon-based PAs have struggled to be able to produce the required frequency response at the power levels required for a radio system (100 mW to 2 W). The lack of a semi-insulating substrate has hampered the efficiency of the devices compared to III-V based technologies. Si also has inferior transport properties compared to these systems. However, the SiGe HBT [7] and the Si based LDMOSFET [8] have recently shown to be promising for fulfilling the PA function at the low end of the power spectrum (Fig. 1.1). Therefore it is a natural desire to want to harness the improved speed performance of the strained-Si system for RF-SOC applications to further extend the use of Si into the higher frequency ranges.

The motivation of this thesis is then to examine the issues involved with integrating the PA function with strained-Si technology using an LDMOSFET device design. Both bulk strained-Si and the SGOI system are considered. Use of the SGOI
system has the additional benefits normally associated with SOI, namely lower parasitic capacitance and power dissipation [9].

1.2 Direction of Thesis

The technology that will be investigated in this thesis for the application of the strained-Si system for RF power applications will be the Laterally Double Diffused Metal-Oxide-Semiconductor Field Effect Transistor (LDMOSFET). The LDMOSFET has already shown to be very promising for fulfilling the power amplifier function on standard Si technology [8]. The LDMOSFET has the advantage that it is highly integratable into a conventional digital CMOS flow unlike the BiCMOS option which adds considerable complexity and expense to the fabrication process. The lower complexity and fewer additional steps also has the benefit of reducing overall cost.

This thesis is organized as follows: Chapter 2 will examine the issues associated with the design of an RF LDMOSFET using strained Si technology. The unique features of the device will be explained in detail and the impact of the process modifications due to the use of strained-Si will be explored. From this study two key technology and device physics issues highlighted were the implantation of the gate oxide and the possibility of higher impact ionization rates. Test structures were fabricated to investigate both of these concerns and the results and analysis of these experiments are presented in Chapter 3 and Chapter 4 respectively. Chapter 5 summarizes the study of strained-Si for RFLDMOSFET and suggests directions for future work in this area.
CHAPTER 2

Design of an RFLDMOSFET for Strained-Si Technology

The design of a RF power device is primarily concerned with its breakdown performance, output conductance, transconductance, on resistance and speed as determined by $f_T$ and $f_{max}$. This chapter considers the main features of an RF power LDMOSFET design and how they impact on the aforementioned parameters. Emphasis is placed on identifying how such a device may be fabricated on strained-Si/SiGe technology, what the main processing issues are and how they may be addressed.

2.1 The Silicon LDMOSFET

The LDMOSFET has been successfully employed as an RF power amplifier on silicon technologies, both bulk [10] and SOI [8]. A typical device as used by Motorola is shown in Figure 2.1. NMOS is the preferred device type over PMOS for its superior speed performance.

It can be seen that the design has a number of features that are rather distinct from those of a standard digital CMOS process such as the graded channel and n-drift region. They are employed to increase the breakdown of the device while maintaining the speed required for RF applications.
The n-drift region is lightly doped and acts as an extended drain region. The drain-substrate depletion region is thus spread across a longer region, reducing the maximum electric field and improving the off-state breakdown of the device. However, because it effectively increases the distance that an electron must travel from source to drain, increasing the transit time, this reduces the $f_T$ of the device. Furthermore, the resistance of this region may also limit the maximum drive current achievable. Therefore there is an inherent design trade-off between breakdown, speed and on resistance.

The P-ch implant and drive-in result in a graded doping profile across the channel region of the device. The doping varies from high concentration at the source end to a low concentration at the drain end. As will be explained in a later section, this results in a higher transconductance than expected for a device of the same drawn gate length but with uniform channel doping. The lower p epitaxial doping level at the drain end of the device is also beneficial for breakdown. In addition to this, the p substrate reduces the drain-substrate capacitance which is important for efficiency in an RF device.
A low resistance connection between the body of the device and the P+ substrate back contact is formed by the P+ sinker. It is important to have a low resistance back or body contact to avoid premature breakdown in the on-state. When high voltages are applied to the drain, electrons crossing the channel undergo impact ionization events resulting in holes being generated. If these holes are not extracted out of the device by a body contact they will cause the body potential to rise thus reducing $V_{sb}$. This in turn leads to a decrease in the threshold voltage and subsequent increase in the drain current giving rise to what is termed a 'kink' in the output characteristics. Additionally, the extra current generation and resultant higher body potential lead to a positive feedback loop causing premature breakdown.

The design of an LDMOSFET on strained-Si technology will essentially take the same form as the Si device but there are a number of additional points that need to be considered and these are discussed in the next section.

2.2 Process Considerations for the Strained-Si/SiGe System

A simple sketch of the strained-Si/SiGe heterostructure, both bulk and SGOI, is shown in Fig. 2.2. In the bulk version a graded SiGe buffer layer is first grown on a Si substrate wafer. The Ge content is graded up to the desired final composition $\text{Si}_{1-x}\text{Ge}_x$. The growth continues with this composition to its final thickness and finally the pseudomorphic strained-Si layer is grown. In the SGOI heterostructure, the strained-Si/$\text{Si}_{1-x}\text{Ge}_x$ sits on an oxide layer which in turn sits on a Si handle wafer. This structure may be prepared by means of the Smart Cut process [11].

While the strained Si/SiGe heterostructure (both bulk and SGOI) system is essentially compatible with mainstream CMOS manufacturing process techniques, there
are a number of differences between this system and standard bulk Si that must be taken into account during device design. This applies to both standard MOS and LDMOS with the LDMOS requiring more consideration due to its higher complexity.

Compared to a standard CMOS flow, the main constraint for strained-Si technology is its thermal budget. The thermal budget is limited because of potential strain relaxation and alloy scattering. There is a critical thickness (t_c) associated with pseudomorphically grown layers as shown in Figure 2.3. Layers grown beneath this thickness are stable at all temperatures but typically the strained-Si will be grown thicker than this (~180 A) and will be metastable. Therefore at too high temperatures the layer will begin to relax losing its strain through the propagation of dislocations. However, even more importantly before the relaxation of the strained layer there could be up diffusion of germanium from the buffer layer causing alloy scattering and effectively eliminating any transport enhancements that arise from the increased mobility.
Figure 2-3: Critical thickness of a pseudomorphic SiGe layer grown on Si as a function of Ge fraction. It is expected that Si grown on SiGe will have similar properties.

The limitations of the thermal budget initially resulted in pessimistic forecasts of the commercial viability of strained-Si, citing the projected increased parasitic resistances due to lower dopant activation as a show-stopper [13]. Since then, many sources have shown that strained-Si can withstand furnace cycles of 750 °C and 800 °C resulting in reliable gate oxides [6] and RTP spike anneals of 1000 °C or a 900 °C thermal cycle for dopant activation [14]. However, the use of the standard LOCOS isolation is prohibited and requires that the isolation oxide be deposited or that a shallow trench isolation (STI) scheme is employed. In specific relation to the LDMOSFET, the reduced thermal budget
means that the body doping implant may not be driven in using a high temperature cycle which is typical. Instead a high angled implant must be employed.

As can be seen from Figure 2.3, the strained-Si layer thickness is very thin. It is on the order of 100 Å to 200 Å. During the fabrication process it becomes even thinner due to Si loss in cleaning and etch steps. This issue of Si loss is not unique to strained-Si technology as very thin body SOI will also have the same order of Si thickness [15]. Cleaning processes have been developed in order to minimize silicon loss. However, these Si devices will typically only have gate oxides of about 2 nm. The LDMOSFET is designed as a power device and will therefore have to withstand higher voltages necessitating the use of a thicker gate oxide. How this gate oxide may be implemented and any issues related to it will be discussed in more detail in Chapter 3.

Dopant diffusivities in SiGe are different to that in silicon. Boron diffusion is retarded by up to a factor of 10 [16] and arsenic and phosphorus diffusion is enhanced by up to a factor of 7 [17]. It is obviously important then to use the correct parameters when modeling the device for optimum doping profiles. For the SGOI option for LDMOS this may be of concern when designing the body contact as the boron needs to diffuse ahead of the N+ source doping in order to form a good link to the body of the device.

To date the literature has only reported on the design of strained-Si MOS structures for digital applications and therefore has not been too concerned with its performance at the higher voltages a power amplifier may expect to see during operation. The bandgap of SiGe is narrower than that of standard Si and therefore is expected to have a lower breakdown and higher impact ionization (II) coefficients. This has already been demonstrated in photodiodes [18]. Part of the transport enhancements in strained-Si
are due to reduced scattering. This means electrons will have a longer mean free path before a collision occurs with the lattice; also, the collision will involve a higher energy. This in turn may lead to higher $I$ coefficients. The effect of these two issues on the breakdown of the device are of great concern and Chapter 4 is dedicated to the study of $I$ effects in strained-Si technology.

### 2.3 Proposed Structure of the Strained-Si LDMOSFET

A cross section of a proposed structure for an SGOI version of a strained-Si LDMOSFET is shown in Figure 2.4. The device maintains the same key features as the standard Si version previously described (n-drift region, graded body doping profile). The constituent layers of the epitaxial stack are also shown. The epitaxial structure consists of a SiGe relaxed buffer layer on which the thin strained-Si layer is deposited. This thin Si layer forms the active region of the device where the channel inversion layer is created. The gate oxide is grown on the strained-Si layer. In the SOI option this entire stack then sits on a buried oxide forming what is termed an SGOI (Strained Si/SiGe on Insulator) structure.

There are a number of advantages with using an SGOI structure over a bulk design. Clearly the same benefits associated with SOI come into play. These include lower capacitance and power dissipation and the fabrication of higher Q-passives compared to a bulk process [9].
SGOI also allows for the use of STI technology for isolation. Thermal budget considerations prohibit the use of the LOCOS process and the use of a deposited oxide to provide isolation is unproven in mainstream commercial processes. Another advantage of using SGOI is that the thickness of the SiGe buffer layer may be reduced from that required for a bulk device. The drain-substrate depletion region must be kept clear of the dislocation rich bulk Si-SiGe interface to avoid leakage. As the LDMOS is subject to higher voltages this requires that the buffer layer be thick.

Depending on the voltage range, the buffer may have to be about 3um thick to withstand voltages of 20 V and above. The thermal conductivity of SiGe is about 10 times less than that of Si [19] (0.15 W/K.cm vs. 1.5 W/K.cm) and having such a thick layer can lead to self-heating which is a major concern for power devices. Rim has
observed self heating effects in buffer layers of the order of 1.5um [20]. Use of SGOI technology allows for the Si/SiGe interface to be etched away prior to the flip and bond to the handle oxide wafer and so thinner buffer layers for the LDMOSFET may be realized.

2.4 Heterostructure Design

The heterostructure design is important in terms of strain engineering to get the desired improved transport properties. Figure 2.5 shows the relationship between the Ge concentration level in the underlying SiGe “virtual substrate” layer and carrier mobility in the pseudomorphic strained-Si layer. For electrons it can be seen that the improvement in mobility saturates at a germanium fraction of between 20% and 30%. The choice of a Si$_{0.8}$Ge$_{0.2}$ buffer layer is therefore a good one in order to achieve improved performance while keeping the germanium content low in order to limit any potential up diffusion of germanium into the strained layer.

For this composition, the thickness of the starting strained-Si can be grown over its critical thickness value to about 200 A [1]. Much of thickness will be lost in the fabrication process and will be consumed by the gate oxide which is nominally set at 100A for this design. Realistically this is about as thick as the gate oxide can be grown while leaving enough strained-Si to support the inversion layer. The final thickness of the strained-Si layer is targeted at 100 A.

The dielectric field strength of SiO$_2$ is of the order of 10 MV/cm and thinner films (<10 nm) have shown larger breakdown fields in the region of 15 MV/cm [13]. A 100A gate oxide could then theoretically support up to 15V across it, a comfortable value for PA applications.
2.5 Doping Design

The doping profile of the device, dominated by the graded body and the n-drift region, will determine it’s electrical parameters – threshold voltage ($V_T$), transconductance ($g_m$), output resistance ($r_o$), drain-induced-barrier lowering (DIBL), cut-off frequency ($f_T$) and saturation drain current ($I_{dss}$). This section will investigate the effect of the body doping and n-drift region design on each of these parameters by means of simulation.

Doping profiles were generated in the 2D process simulator SUPREM IV. Dopant diffusivity coefficients specific to SiGe were taken from Eguchi [17] and Kuo [16]. These profiles were then fed into the 2D device simulator MEDICI. The universal mobility
model was used for the transport properties of strained-Si. Parameters for the model were taken from Rim [20]. The polysilicon gate length was set at 0.6um and the background doping at $1\times 10^{16}$ cm$^{-3}$. A bulk design was implemented with contact being made via the backside of the wafer. It was considered a valid approach not to include the buried oxide for a full SOI simulation as the SOI is not expected to impact on the intrinsic electrical performance of the device [22]. An example of a device grid is shown in Figure 2.6.

These simulations are not intended to provide an absolute solution for the device design but rather to gain insight into the effects associated with varying the doping profiles.

![Figure 2-6: Example of a MEDICI grid used in the process simulations. The mesh and regions are shown in (a) and the different doping regions in (b).](image)

### 2.5.1 Body Doping

The graded body doping profile along the channel dominates the $V_T$, $g_m$, $r_o$ and DIBL of the device. Use of a graded channel usually results in enhancements in these parameters as compared to a device of the same physical polysilicon gate length with uniform channel doping.
In such a standard device the effective length of the channel ($L_{\text{eff}}$) is determined by the drawn gate length or the distance between the metallurgical source/channel and drain/channel junctions. However, in a graded channel device of the same polysilicon gate length it is more difficult to exactly define $L_{\text{eff}}$. This is especially true when the channel doping is graded linearly across the gate length. However, the devices that will be discussed in this study have a very abrupt grading. $L_{\text{eff}}$ will be taken as the point at which the channel doping drops sharply back down to the background doping level. From this definition, for the same polysilicon length, the $L_{\text{eff}}$ graded channel devices will invariably be shorter than the uniformly doped device.

Figure 2.7 shows a comparison of a graded to uniformly doped channel device in strong inversion. The doping level of the uniformly doped channel was chosen to match the $V_T$ of both devices. The net doping profile in the body is shown in Figure 2.7(a) and the lateral field and electron concentration along the channel are plotted in Figures 2.7(b) and 2.7(c) respectively.

It can be seen that the field near the source in the graded channel device is higher. This can be explained by considering that the device can be viewed as an enhancement-mode NMOS in series with a depletion-mode NMOS at the drain end. For a given gate to source voltage, the inversion layer charge (electron concentration) in the depletion mode device is higher.
Figure 2-7: Simulated (a) doping (b) field and (c) electron concentration profiles for a graded and uniform LDMOSFET channel biased at $V_{gs} = 3.6V$, $V_{ds} = 7V$. 
than that of the enhancement-mode part (Figure 2.7(c)). Therefore the resistance of the channel near the source is higher than towards the drain and the voltage preferentially drops along this end. The higher field results in a higher drain current as the electrons are being pulled at a higher velocity from the source.

From a first order analysis the transconductance of a MOSFET is given by

\[
g_m = W_g C_{ox} v_n(s)
\]

where \(W_g\) is the width of the gate, \(C_{ox}\) is the oxide capacitance per unit width and \(v_n(s)\) is velocity at the source. The graded body design therefore allows for higher \(g_m\) values by allowing \(v_n(s)\) to get closer to the saturation velocity \(v_{sat}\). Saturation velocity effects, of course, set the limit on the improvements that can be achieved by using the graded channel approach.

Figure 2.8 shows the \(g_m\) of the graded channel device vs. the uniformly doped channel device. There is an improvement of 50% in \(g_m\) and it can be seen by the flatter plateau of the graded channel \(g_m\) that it is indeed more affected by velocity saturation effects.

The fact that the graded channel design can exhibit the higher performance characteristics of a shorter device but with a longer drawn gate length (and subsequently higher breakdown) is one of the key enabling factors of RFLDMOS.

2.5.1.1 Implant Splits for strained-Si Body Doping

As discussed earlier the restricted thermal budget of strained-Si technology is of concern for implementing the graded doping profile. In a probable process flow the only thermal cycle that could act as a drive-in for the body implant is the RTA spike anneal at 1000°C.
Considering that the diffusion of B is retarded in SiGe with respect to Si then it can be expected that the implanted profile is not going to diffuse much laterally or in depth from its original state. Figure 2.9 shows a SUPREM simulation of an as-implanted and diffused arsenic and boron profile, which shows that indeed the boron does not move from its implanted state. Therefore the body doping will need to be directly placed under the gate by means of a high-tilt implant.

A series of simulations were carried out to determine the effect of the implant tilt angle and energy on the device characteristics. The drift region length is 0.5um with a dose and energy of 3x10^{12} cm^{-2} and 55 keV respectively. The body dose was set at 1.5x10^{13} cm^{-2}. Tilt angles of 0°, 30°, 45° and 60° are used each with energies of 20, 30, 40, 50 and 60 keV. The implant profiles were generated using the Monte Carlo option in SUPREM IV in order to accurately simulate the high-tilt angles.
Figure 2-9: Arsenic 55 keV and Boron 60 keV as implanted and post 1000 °C RTA spike anneal doping profiles.

Figure 2.10 shows the threshold voltage achieved for each option. As would be expected the 0° tilt implants have the lowest $V_T$ which corresponds to that of the background doping level of $1 \times 10^{16} \text{ cm}^{-3}$ as no dopants penetrated under the gate. The $V_T$ then increases with increasing tilt angle as the larger angled implants are more effective at getting more Boron under the gate (Fig 2.11). For the higher tilt angles of 45° and 60° the threshold voltage decreases with the highest energy of 60 keV. Fig. 2.12 illustrates the effect of implant energy on the doping concentration along the gate oxide/strained Si interface for the 60° tilt implants. For the 60 keV energy, the peak of the implant goes further into the substrate leaving a lower doping concentration at the surface compared to the lower energy implants. This reduces $V_T$. The effect of the implant conditions on the peak surface doping concentration for all splits is summarized in Fig 2.13.
The peak transconductances obtained for each of the high angled splits at a drain bias of 3.6 V is shown in Figure 2.14. For all angled implants, $g_m$ is within about 15% of each other due to velocity saturation effects. Changing the dopant profiles around the source end of the channel, in order to engineer the field to higher values, is only going to have a limited effect if the electrons are already approaching velocity saturation. This is especially true for strained-Si, as the higher mobility will result in velocity saturation occurring at lower lateral fields. However it can be seen from Fig 2.15 that the $L_{eff}$ is longer for progressively higher energies. This results in a slight drop off in transconductance, indicating that there still some room to move along the velocity curve.

DIBL is plotted in Fig. 2.16. Each of the splits exhibits good performance, varying from 3 mV/V to 17 mV/V. It might be expected that the uniform-graded channel as represented by the 0° implants would have significantly worse DIBL than the high-tilt
Figure 2-11: Boron doping profiles along channel for 50 keV implants with different tilt angles. The edge of the gate is at 0 um. The 0° tilt implants did not penetrate under the gate and so had a uniform channel doping $1 \times 10^{16}$ cm$^{-3}$.

Figure 2-12: Boron doping profiles along channel for 60° tilt implants for different energies. Note the lower doping concentration level for the 60 keV implant.
options because of the lower channel doping. However, the $L_{\text{eff}}$ of these devices is 0.6μm as opposed to the ~0.2μm $L_{\text{eff}}$ of the other splits (Fig 2.15). The doping at the source end of the channel sets the threshold voltage. The higher local doping around this area
prevents the spread of the drain depletion region under this section of the gate and thus limits the amount of barrier lowering and $V_T$ roll-off. In fact the doping profile under this end of the gate resembles that of the retrograde-profiles used in short channel devices for the very purpose of improving the DIBL performance. The boron concentration is lower at the surface than deeper into the substrate. The variations in DIBL from split to split are small and most likely due to small differences in how the dopants are distributed immediately under the point at which the threshold voltage is determined.

Output resistance is plotted in Fig. 2.17. $r_o$ was determined at $V_T+1.5V$ so that it would be a valid comparison at similar current levels. It can be seen that higher tilt angles and higher energies result in a higher $r_o$. Ideally $r_o$ should be infinite. However the saturation current is a function of the drain voltage due to two effects – DIBL and Channel Length Modulation (CLM).
Through DIBL, $V_T$ drops as $V_{ds}$ increases thus increasing the current and the CLM causes a decrease in the channel length as the pinch off point encroaches further into the channel. Decreasing the channel length results in a higher field at the source as
\( V_{\text{dsat}} \) is dropped across a shorter distance. This also results in more current. The effect is limited if the device is already in the velocity saturation regime.

From Fig. 2.16 the DIBL of each of the devices is low and comparable across all splits. Therefore the variations seen in \( r_0 \) (Figure 2.17) must be due to CLM. Higher tilt angles and energies push the dopants further into channel region under the gate. As a result, \( L_{\text{eff}} \) increases and these devices will be less affected by the CLM effect. The doping level in the channel also affects CLM. Lower doping levels allow the pinch off point to encroach further into the channel region resulting in higher CLM. This effect can be seen from the 60 keV energy splits. They have the longest \( L_{\text{eff}} \) (Fig 2.15) for the 45° and 60° tilts but have a lower doping concentration (Fig. 2.13) resulting in higher CLM and lower \( r_0 \).

### 2.5.1.2 Under Source Body Link

The LDMOSFET investigated in this thesis is to be ultimately fabricated using SGOI technology. In this case, the back of the wafer then may no longer be used as a body contact. In order to reduce the body resistance and prevent the kink effect, a separate contact will have to be made to the body. This requires that there be a substantial p-type doping level beneath the source to form this link.

The higher energy implants are more suited to this purpose as they result in higher doping concentrations beneath the source (Fig. 2.18). Based on this, the thickness of the SiGe buffer layer may be thinned to about 2000 Å.
2.5.2 Drift Region Design

The drift region of the device is what gives the LDMOS its superior breakdown performance. Off-state breakdown in the MOSFET occurs at the gate edge where the channel/drain junction is formed. In a MOSFET in saturation, this is where the peak of the lateral electric field appears. As the field grows higher and higher with increasing $V_{ds}$, carriers in this region are accelerated to higher and higher velocities. An electron can then gain enough energy to generate an electron/hole pair when it collides with a silicon atom in the lattice. The additional carriers generated by this impact ionization can themselves
Figure 2-19: Electric field contours for $V_g = 0$ and $V_{ds} = 8\text{V}$ are shown in (a) with the inset scale varying from $2\times10^5 \text{V/cm}$ to $7.6\times10^5 \text{V/cm}$. The corresponding impact ionization generation rate is shown in (b) with the inset scale varying from 12 to 20 on a log scale. Source and drain region are outlined.
create further electron/hole pairs leading to avalanche breakdown. This effect is illustrated in Fig. 2.19.

The key to increasing the breakdown voltage of a MOSFET is to decrease the electric field at the gate edge. This can be achieved by inserting a lightly doped region between the gate edge and the heavily doped drain contact region thus effectively increasing the distance over which the drain/substrate depletion region is sustained. Therefore the length of the drift region is one the critical design parameters for breakdown. This is shown in Figure 2.20(a).

The doping level of this region is also very important. The abruptness of the channel/drain junction has a significant impact on the breakdown. It is a well known fact that abrupt junctions have a lower breakdown than that of graded junctions due to differences in the electric field profile [23]. Lower doping levels enable a smoother grading from the channel to the drain region and result in lower peak electric fields. This is shown in Fig 2.20(a).

Unfortunately, there is a price to be paid for the improved breakdown performance. The presence of the drift region effectively increases the length of the device between the source and drain contacts. This results in an increase in the transit time between the source and drain that slows down the device and yields a lower frequency response.

The resistance of the drift region also limits the maximum drive current. This is of great concern for power devices which have to deliver anywhere between 200mW to 4W for wireless applications [24]. If the current drive is reduced then the device must be made wider to meet the power specification. For the LDMOS design, increasing the
Figure 2-20: Effect of the drift region length (a) and dose (b) on the lateral electric field at the gate edge.
Ids vs. Vds for Vgs = 2V, 2.3V, 2.8V, 3.2V, 3.6V, 4V, 4.4V, 4.8V

Figure 2-21: Output characteristics demonstrating the effect of adding a drift region to the device. Once the drift region resistance becomes dominant, the current drive becomes limited and the transconductance drops off rapidly.

width of the device essentially means adding more fingers to the cell. This increases parasitics and also has a detrimental effect on the frequency response.
The effect of adding a drift region on the device output characteristics is shown in Fig. 2.21. At higher gate voltages the increased channel charge reduces the channel resistance to a level where the on-resistance of the device is dominated by the drift region. As a resistor this region becomes limited by velocity saturation as to how much current it can carry. Once it becomes saturated, even if the inversion layer charge is increased by increasing $V_{gs}$, the current in the device cannot increase because the drift region cannot sustain it. Once this point is reached the transconductance falls off sharply which also limits the operating window of the device.

2.5.2.1 Doping Splits

The effect of the length and dose of the drift region on the electrical characteristics of the device was investigated by simulation. The energy was set at 55 keV and the dose was varied between $1 \times 10^{12}$ cm$^{-2}$ and $1 \times 10^{13}$ cm$^{-2}$ arsenic. This resulted in doping levels between $2 \times 10^{17}$ cm$^{-3}$ and $2 \times 10^{18}$ cm$^{-3}$. The drift length was also varied between 0.25um and 1um. All devices had a gate length of 0.6um and a body doping dose of $1.5 \times 10^{13}$ cm$^{-2}$ implanted at an energy of 50 keV and tilt 60°. Figures 2.22 to 2.25 show the impact of each drift region design on the main device characteristics.

As to be expected, the breakdown of the device improves with increasing length and decreasing doping concentration. The effect of increasing the length diminishes as the doping concentration increases. The breakdown values were simulated using the impact ionization rates for silicon, the validity of this will be discussed further in Chapter 4.
Decreasing the doping to improve the breakdown must be weighed against the increase in the on resistance. Figures 2.23 and 2.24 show the resistance of the drift region and the corresponding $I_{dsat}$ achieved for each option. As expected, the drift regions with high resistance resulted in a lower $I_{dsat}$.

The impact of the drift design on the $f_T$ of the device was also investigated. MEDICI does not support the use of the universal mobility model for AC simulations. The $f_T$ of a MOSFET is given by

$$f_T = \frac{g_m}{2\Pi C_{gate}}$$

Therefore the $g_m$ of each device was taken from the DC characteristics obtained from the universal mobility model using the values for strained Si and the gate capacitance was obtained from AC simulations run using the standard Lombardi scattering mobility model.
Resistance of Different n-drift Regions

Figure 2-23: Resistance associated with each of the drift regions.

\[ I_{dsat} @ V_{ds} = 3.6V, \ V_{gs} = 5V \]

Figure 2-24: \( I_{dsat} \) achieved for different drift region designs.
Figure 2-25: $f_T$ for different drift region designs. The scale is normalized to the 0.25um length, $1 \times 10^{13}$ cm$^{-2}$ dose value.

$f_T$ was determined at a bias point of $V_{gs} = 2.9$ V and $V_{ds} = 3.6$ V. The values calculated for $f_T$ were then normalized with respect to result obtained for the 0.25um length and $1 \times 10^{13}$ cm$^{-2}$ dose split.

The $f_T$ of the device for the lower resistance splits did not change very significantly with the variations in the drift region design. The higher resistance drift regions show a much lower $f_T$ due to the transconductance dropping off sharply at this bias point as was demonstrated in Fig. 2.25.

An absolute value for $f_T$ was not quoted due to the need to normalize the results because of MEDICI simulation issues. Based on how the $f_T$ was estimated for these comparisons an approximate value would be 12-13 GHz. This value is based on the standard drift-diffusion models and does not take into account any momentum relaxation effects that can lead to velocity overshoot [3]. The $f_T$ may increase once these effects are
taken into account. A more practical way to improve the $f_T$ would be to decrease the gate length from the 0.6um used in this study.

In order to achieve a high breakdown voltage the doping of the drift region must be kept low or the length increased but this has been demonstrated to result in a direct trade-off with $R_{on}$, $I_{dsat}$ and $f_T$. $R_{on}$ and $I_{dsat}$ can be recovered by increasing the width of the device to achieve the desired performance. However, this in turn reduces the frequency response of the device by increasing the parasitics. The conflicting relationship between these parameters determines the design space for the n-drift region.

### 2.6 Conclusions

This chapter has discussed the issues involved with implementing an LDMOSFET using strained-Si technology and an initial design space for the intrinsic device was established.

The heterostructure of choice for the strained-Si LDMOSFET design was determined to be that of SGOI. The use of SGOI allows for all the benefits associated with standard SOI. In particular for strained-Si technology, it should improve self-heating effects and STI technology may be used.

The restricted thermal budget requires the use of a high tilt angle to form the graded body region. Higher tilt angles in general resulted in better performing devices but there is a limit to the improvements that higher implant energies will achieve. If the energy of the implant is too high, then the peak of the doping profile is situated too deep into the substrate and the surface doping concentration is reduced. This was found to result in a poorer output resistance performance. Also higher energies resulted in a longer
Leff slightly reducing the transconductance. Excellent DIBL performance was achieved for all options. Based on the doping splits carried out, the optimum implant conditions for the body were determined to be a 50 keV implant with a tilt of 60°.

The main trade-off in the design of the drift region was found to be between the breakdown of the device and the on-resistance. The on-resistance needs to kept low so that it does not limit the drive current and thus the power output of the device. \( R_{on} \) is kept low by decreasing the length of the drift region and increasing its doping level. Both of these actions conflict with the breakdown requirement of increasing the length and decreasing the doping. A good choice for the drift region length and dose based on the simulations would be 0.5\( \mu m \) and 5\( \times 10^{12} \) cm\(^2\). The device may always be made wider by adding more fingers to the unit cell in order to meet the power output requirements. However, this is always an issue with RF devices as the power efficiency is impacted significantly by parasitic elements.
CHAPTER 3

Strained-Si LDMOSFET Gate Oxide

A ground rule in the design of strained-Si LSMOSFET is that it be compatible as possible with a strained-Si digital CMOS process in order to enable SOC technology. A key issue then is that the CMOS devices are expected to have a gate oxide thickness of the order of 2nm or less while the LDMOSFET will require a gate oxide on the order of 10nm. A dual gate oxide approach will have to be devised whereby both flavours of devices can have a good quality oxide but that is compatible with the constraints of strained-Si processing. This chapter describes two options that have been evaluated for this purpose. Both of these gate oxide options are easily integrated into a dual gate oxide process flow.

3.1 Gate Oxide Options

3.1.1 Conventional Dual Gate Oxide Process

A conventional dual gate oxide process flow is first presented to outline the issues involved with the gate oxide processing for a SOC strained-Si system. A typical process flow is shown in Figure 3.1.
The dual gate oxide process proceeds as follows. First a thick oxide is grown across the entire wafer. The oxide is then coated with photoresist and patterned so that the area where the thin oxide will ultimately be grown is exposed. The oxide in this area is then removed by means of a buffered oxide etch (BOE). The photoresist is then stripped and the remaining oxide surface is then cleaned via a controlled dilute HF dip-off. This clean is required so that the surface of the thick gate oxide is cleared of residual organic contaminants from the photoresist which would represent a reliability concern. The process is completed with a second oxidation grown to required thickness of the thin
oxide. The thick oxide regions will also grow during this step, so the initial oxide growth must be set so that the combination of the two cycles results in the required thickness for the thick oxide.

For a standard bulk Si process flow the amount of Si loss due to the oxide growth is not a concern. A 10nm thermally grown oxide will consume 4.5 nm of Si. Given that the thickness of the strained-Si layer is on the order of 10 nm to 15 nm, this Si loss is considerable. Even more so when it is considered that the cleaning steps that occur in the process prior to the gate oxide formation will result in even further Si loss. Such a flow then, would constitute a problem for the CMOS devices as they will be fabricated in the ‘thin oxide’ regions that suffer the most severe oxide loss.

### 3.2 Gate oxide options

Given the concerns of Si loss due to a conventional dual oxide process, two options have been considered in this thesis for the gate stack processing in order to circumvent this issue.

The first option consists of growing the strained-Si layer further beyond the critical thickness to about 22 nm and then thermally growing the full 10nm of the thick gate oxide via a dry/wet/dry process. It is necessary to use a wet oxidation step to keep the time of the 750 °C thermal cycle down. Growing the strained layer to this thickness has the associated risk of strain dislocation as the layer becomes less stable at higher thicknesses. However this should be alleviated by the actual growing of the gate oxide itself. The oxide consumes part of the strained layer thus reducing its thickness back down to a more typical value prior to the highest thermal step of the RTP spike anneal.
The second option consists of a composite oxide scheme. The strained-Si layer is grown to about 15 nm. The thick gate oxide is then grown by a two-step process. The first step being a thin thermal oxidation and the second step a low temperature oxide (LTO) deposition to bring the stack to the required thickness. The reasoning behind this is that it is the Si/SiO₂ interface that is critical to the gate oxide quality. LTO is known to form a poor interface but by moving it away from the direct Si/SiO₂ interface by first growing a thermal oxide, it is hoped that this will improve its performance.

Both of these options may be implemented using the same generic flow as the conventional dual gate oxide process.

3.2.1 Test Structure Process Flow

Both of these gate oxide options were evaluated via means of capacitor test structures. Two different heterostructures were used for the experiment. Option 1 (the thermal oxide) used a lightly doped (~ 1x10¹⁵ cm⁻³ p-type) stack which had 220 Å of strained-Si. The buffer layer in this structure did not have a CMP step. The heterostructure used for Option 2 (the composite oxide scheme) was doped to 1x10¹⁷ cm⁻³ p-type throughout the stack. The strained-Si layer thickness was 120 Å and in this case the wafers did receive the CMP etch-back step. Bulk Si controls were run with both splits.

The process flow for the capacitors is shown in Figure 3.2. It is a simple one mask process. First the gate oxide is grown. For option 1, which is the fully thermal oxide, this is a 750C dry/wet/dry process. The initial cycle is a 30 min. dry growth at 750 °C which results in 2.5 nm of oxide. The 10 min. wet cycle brings this up to the desired 10 nm. The final dry cycle does not result in any further significant growth (~ 1-2 Å) but passivates
the top of the oxide. For option 2, the composite oxide scheme, a 30 min. dry oxidation cycle at 750 °C for 30 mins that grows 2.5 nm is followed by a 400 °C LTO deposition step which brings the thickness up to 10 nm.

After oxide growth, a 150 nm polysilicon layer is deposited and then receives a $3 \times 10^{15} \text{cm}^{-2} 15$ keV implant which is activated with a 1000 °C RTP spike anneal. Ideally the polysilicon would be deposited in-situ doped in order to avoid poly depletion effects but that was not an option for this experiment. A 1000 A/1 um Ti/Al metal layer completes the stack which is then dry etched back to the oxide. Al is finally e-beamed on the backside of the wafer to form a good body contact and the metal is alloyed at 400 °C for 30 minutes in forming gas.

![Figure 3-2: Diagram of the capacitor test structure used to evaluate the different gate oxide options.](image)

### 3.3 Analysis of Results

The capacitors fabricated with the gate oxide options described earlier were characterized to determine the quality of the oxide. The results of this characterization are described in the next sections.
3.3.1 C-V Measurements

High frequency C-V measurements were carried out using a HP4192 LCR measurement unit. Both oxide growth options showed good characteristics as evidenced in Figure 3.3.

The $C / C_0$ minimum is different between the two samples because of the difference in the substrate doping. The slight hump in the curve near the midgap region is due to the band offset between the underlying $Si_{1-x}Ge_x$ buffer layer and the strained-Si layer [25].

3.3.2 I-V Measurements

I-V measurements were carried out to determine the leakage current of the gate oxide as an indication of its integrity. Typical I-V characteristics for both options are
shown in Figure 3.4. Both show low levels of leakage which would be acceptable for a manufacturing process. However it should be noted that there were a number of sites on the fully thermally grown oxide sample that exhibited premature breakdown when stressed. This wafer had a noticeably bad cross-hatch pattern as it did not receive the CMP step during the buffer layer growth process. A bad cross hatch pattern corresponds to a rough, uneven surface that may transfer into an oxide grown on that surface. It is speculated that this is responsible for the poorer oxide yield on this wafer.

3.3.3 $D_{it}$ Measurements

The density of interface traps ($D_{it}$) gives a measure of the quality of the Si/SiO$_2$ interface. Interface traps are essentially defects located at the Si/SiO$_2$ interface caused by the termination of the periodic silicon lattice along this plane. In general, the lower the $D_{it}$ the better the quality of the oxide.
Figure 3-5: Examples of the Gp/w curves obtained for the (a) fully thermal oxide and (b) composite oxide schemes. The data for (a) has been taken from the Si control sample.

The Dᵣ was obtained by means of the parallel conductance method described in Nicollian and Brews [26]. The extracted parallel conductance curves are shown in Figure 3.5. Unfortunately, the series resistance due to the substrate of the fully thermally grown
oxide strained-Si sample dominated the results and the parallel conductance could not be extracted from the data. Instead the data was taken from the Si control sample that was run with the split. This was justified by the fact that the composite oxide Si control sample gave similar $D_{it}$ results to the composite oxide strained-Si sample.

The $D_{it}$ of the composite oxide was $4 \times 10^{11} \text{cm}^2\text{eV}^{-1}$ as measured near flatband and $1.5 \times 10^{11} \text{cm}^2\text{eV}^{-1}$ near midgap. For the fully thermal oxide the results were $1.7 \times 10^{11} \text{cm}^2\text{eV}^{-1}$ and $1 \times 10^{11} \text{cm}^2\text{eV}^{-1}$ near flatband and midgap respectively. Typical values of $D_{it}$ for modern CMOS devices are of the order of $10^{10} \text{cm}^2\text{eV}^{-1}$ [27]. The values obtained for both options are reasonable then considering that cleaning processes and deposition / growth cycles have not yet been optimized for this flow.

### 3.4 Conclusions

Two gate oxide options were investigated for the strained-Si LDMOSFET. One consisted of a fully thermally grown oxide by a dry/wet/dry cycle on a thick strained-Si layer grown to 220 Å. The second option was a composite oxide composed of a thin dry thermal oxide and an LTO deposition. Both options could be used in a typical dual oxide process flow.

Characterization of the gate oxide options was carried out via means of high frequency capacitance, leakage and $D_{it}$ measurements. Both options exhibited good characteristics and would be considered acceptable for an LDMOS process.
CHAPTER 4

Impact Ionization Effects in Strained-Si/SiGe

The n-drift layer region design is critical to the breakdown performance of a strained-Si LDMOSFET. Previous work has looked at breakdown effects in SiGe layers [18]. This work indicates that the impact ionization (II) coefficients for both electrons and holes increase relative to that of Si, but nothing has been published relating to such effects in strained-Si. This chapter describes and discusses the results of an experiment designed to examine II effects in strained-Si.

4.1 Design of Test Structure

The basic structure designed for the investigation of impact ionization and breakdown effects is shown in Figure 4.1. An n-drift region is contacted by N+ regions. As the voltage is increased from drain to source, the field accelerates electrons that will eventually gain enough energy to cause impact ionization. The generated hole current is then collected by the body contact. This structure mimics the Transmission Line Model (TLM) device often used in III-Vs. A 2D MEDICI simulation (Figure 4.2) demonstrates the principle of the device. As the applied voltage is increased and therefore also the field, impact ionization takes place and the body current grows until the device ultimately breaks down.

The body current is a critical measurement parameter in the experiment and
therefore it is imperative that the junctions be of high quality so that leakage current does not obscure any results. In order to achieve this criterion there has to be a trade off between the thickness of the underlying buffer layer and the background doping level.

Due to the high voltage levels applied to the drain, the N+/p- depletion region will spread into the substrate. The graded buffer / Si handle wafer interface is rich in dislocations so it is important that the depletion region is kept away from this area. Due to limitations in the epi growth process, this had to be kept at a reasonable value of a few microns. How far the depletion region spreads for a given voltage will be dependent on the p-type substrate doping level. The lighter the doping the wider the depletion region. Therefore the doping can be increased to restrict the widening of the depletion region. However, increasing the p- doping level will reduce the breakdown of the N+/p- diode. This breakdown must be high enough so that impact ionization effects due to the TLM operation can be observed. A choice of a 4um thick buffer layer with a doping of ~2-3x10^{16} cm^{-3} was chosen as a good compromise to balance the two effects.
4.1.1 Substrate Preparation

The starting substrate consists of a Si$_{0.8}$Ge$_{0.2}$ buffer layer grown on a P+ handle silicon wafer. The buffer layer was initially grown and then ground back by means of Chemical Mechanical Polishing (CMP) to a thickness of 0.5um. The layer was then regrown to a thickness of 4um. The grindback and regrowth process is employed to reduce the effect of the “cross hatch” pattern observed in SiGe buffer layers that leads to high surface roughness effects [28]. The pseudomorphic strained-Si layer was then grown on the buffer with a targeted thickness of 200 Å.

Bulk Si controls were run with the lot. They were P/P+ epi wafers. The epi layer thickness was 4.5um and its doping level was set at 2-3x10$^{16}$ cm$^{-3}$ in order to match the SiGe heterostructure as closely as possible.
4.1.2 Process Flow

The fabrication process flow is outlined in Figure 4.3. Active area definition is performed via means means of a 5000 A LTO deposition and etch. The etch consists of an anisotropic plasma etch followed by a BOE dip off. The dip off is used to avoid causing plasma damage to the strained-Si layer. This also has the benefit of producing a sloped sidewall profile which mimics that usually seen in a LOCOS process. The LTO provides isolation for the devices. Once the strained-Si has been exposed, a thin 40 A dry oxide is grown at 800 °C for 30 minutes. Originally this thickness was targeted for 100 A to match that of the expected oxide thickness of the LDMOSFET but there was no available wet oxidation process at the time of processing. The purpose of this oxide is to passivate the surface of the wafer and to provide a protective layer for the future implants.

A blanket low dose As implant (splits at this point were $1\times10^{12}$ cm$^{-2}$, $5\times10^{12}$ cm$^{-2}$ and $1\times10^{13}$ cm$^{-2}$ ) at 20 keV is then performed. This sets the doping and thickness of the region where $\Pi$ will take place. The wafers are then masked for the N+ As $3\times10^{15}$ cm$^{-2}$ 20 keV contact implant.

Post the implants another 1500 A LTO layer is deposited. The dopants in the drift region and contact regions are activated by means of an RTP 1000 °C spike anneal. A backetch is carried out prior to this step to remove the LTO as the RTP process is emissivity dependent.

Contact cuts are made and a 1000 A Ti / 7500 A Al metal stack is deposited and etched. The Ti is used as a barrier layer to prevent junction spiking. Backside aluminium metallization is carried out via e-beam deposition. The metal is then alloyed in forming gas for 30 minutes at 400 °C.
Figure 4-3: Process flow for the II test structure.
4.2 Device Layout and Parasitics Extraction

The layout of the TLM device is shown in Figure 4.4. Body contact is made via the wafer backside. A kelvin structure is used for the metal pads so that the contact resistance of the measurement probes is automatically factored out.

![Diagram of TLM device layout](image)

**Figure 4-4:** Layout of (a) full device showing metal pads and (b) a more detailed look at the actual resistor structure.

The intrinsic n- portion of the device is contacted by two short N+ regions termed "links" that are of the same width of the resistor portion of the device and are 2um long on either side of it. Use of these links makes the definition of the n- drift region length more accurate. The photo step for the N+ implant can be misaligned by up to 2um in either X
direction without having an effect on the electrical characteristics. The link regions then reach a 10umx10um N+ region which is contacted by four 2umx2um contacts.

It is important that the actual voltage across the intrinsic portion of the device be ascertained so that the field can be accurately determined. Figure 4.5 shows a cross section of the device and the parasitic resistances associated with the layout. Aside from the intrinsic resistance of the n- region there are resistive contributions from each of the links (R_{link}), the N+ source and drain contact regions (R_{N+}) and the actual silicon/metal contact resistance itself (R_c). R_{N+} consists of that portion of the contact regions up to the first row of contacts which is a 2umx10um section or 0.2 squares. Although there are four contacts for both the source and drain it is the first row of two that will conduct the majority of the current to the metal pad.

![Figure 4-5: Cross section of actual device showing the parasitic resistances. These must be extracted from the data to get the true voltage across the n- region.](image)

\[
V_{\text{intrinsic}} = V - I \times 2 \times (R_{\text{link}} + R_{\text{N+}} + R_c/2)
\]

Therefore the contribution of the contacts to the overall resistance will be half that of a single contact on either side. The actual voltage (V_{\text{intrinsic}}) across the n- region is thus given by

\[
V_{\text{intrinsic}} = V - I \times 2 \times (R_{\text{link}} + R_{\text{N+}} + R_c/2)
\]
Where I is the current through the device at a given applied voltage (V). Depending on the device width, length and doping the voltage correction was on the order of 1% to 25%.

Test structures were laid out on the chip in order to measure each of these parasitic components and are described in the next sections.

### 4.2.1 $R_{\text{link}}$ Extraction

The link resistance is extracted by means of a simple kelvin resistor structure (Figure 4.6). A current, I, is forced through a 150um long resistor with different widths mirroring those of the TLMs. The potential is then tapped at two points in the line. The resistance per um of length is then given by the voltage difference (V) of the taps divided by the distance between them L (in this case 48um). Therefore $R_{\text{link}}$ is given by $2*(I/V)/48$. Typical values of $R_{\text{link}}$ were 176 Ω for the narrowest devices and 28 Ω for the widest devices.

![Figure 4-6 Kelvin structure used to measure $R_{\text{link}}$.](image)
4.2.2 \(R_{N^+}\) Extraction

The sheet resistance of the \(N^+\) Source/Drain regions is determined by the measurement of a Van Der Pauw structure (VDP) as shown in Figure 4.7. Current (I) is forced through two diagonally opposite pads and the voltage (V) is sensed across the other two. \(R_{N^+}\) is then \(\frac{R\Omega}{5}\). A typical value for \(R_{N^+}\) was 31 \(\Omega\).

\[
R_{\text{c}} = \frac{\pi}{\ln 2} \left( \frac{V_3 - V_2}{I_1 - I_4} \right)
\]

Figure 4-7 VDP structure for \(R_{\text{sh}}\) extraction.

4.2.3 \(R_c\) Extraction

The contact resistance is measured using a Kelvin cross structure (Figure 4.8). Again current (I) is forced through two diagonally opposite pads and the voltage (V) is sensed across the other two. A typical value for \(R_c\) was 19 \(\Omega\).
4.2.4 Normalization of the Width

In order to make valid comparisons between each of the splits in the experiment it is necessary to normalize the current with respect to the width of the device. The TLMs are drawn with widths defined by the active area lithography of 1um though 10um. The LTO etch following this litho step consists of a dry and wet portion, with the latter being isotropic. Therefore there is undercutting of oxide beneath the mask and the active area width is larger than drawn. There was a large LTO (>1000 A) thickness variation within wafer and wafer to wafer in this run. A significant overetch in the BOE step was required to ensure that the oxide was cleared. Based on etch rates the offset in the width was estimated to be between 0.5um and 0.8um for a given wafer. In order to accurately
determine this offset for each wafer, the Kelvin resistance structures outlined in 4.3.1 were used.

Each of the resistors are the same length, therefore two widths $W_x$ and $W_y$ are related to the measured resistances $R_x$ and $R_y$ by

$$\frac{W_y}{W_x} = \frac{R_x}{R_y}$$

or

$$W_xR_x = W_yR_y = \text{constant}.$$ 

Assuming a constant offset $W_{\text{offset}}$ this can be further generalized to

$$(W + W_{\text{offset}})R = \text{const.}$$

where $W$ is the drawn width. A least squares fit may then be applied to the Kelvin resistance data to obtain the offset. A comparison of the resistance vs. drawn width and corrected width curves is shown in Fig. 4.9. Typical offsets were between 0.65um and 0.9um.

![Plot of Resistance vs Width for Actual Data and Drawn Width + Extracted Offset (W_{offset})](image)

**Figure 4.9:** Plot of resistance of Kelvin structures vs. width as drawn and as corrected for offset. Reading the width from the resistance curve directly from the measured data would result in significant errors at narrow widths.

75
4.3 Initial Analysis of Results

4.3.1 Parasitic Extraction from Initial Measurements

The TLM structures were laid out with different lengths and widths varying between 1um to 100um for the length and 1um to 10um for the width. IV characteristics of these structures were obtained using a HP 4155B measurement unit.

Before any analysis can be carried out, the parasitic resistances as described in section 4.3 must be de-embedded from the IV characteristics. The parasitic resistances and their individual components for each of the splits are given in Table 4.1. R\text{N+} is fairly consistent across the splits at 31 Ω. The R\text{N+} of the 5x10^{12} cm^{-2} doped strained-Si sample is slightly lower at 27.5 Ω but this may be explained by variations in the implant process.

R\text{c} was also consistent across the splits. Differences in R\text{link} between the bulk and strained-Si samples for the same doping levels can be accounted for by differences in the width offset W\text{offset} and again process variation.

<table>
<thead>
<tr>
<th>Dose (cm^{-2})</th>
<th>Bulk Si 1x10^{12}</th>
<th>Strained Si 1x10^{12}</th>
<th>Bulk Si 5x10^{12}</th>
<th>Strained Si 5x10^{12}</th>
<th>Bulk Si 1x10^{13}</th>
<th>Strained Si 1x10^{13}</th>
</tr>
</thead>
<tbody>
<tr>
<td>W\text{offset} (um)</td>
<td>0.56</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.67</td>
<td>0.65</td>
</tr>
<tr>
<td>R\text{N+} (Ω)</td>
<td>31</td>
<td>31</td>
<td>31.2</td>
<td>27.5</td>
<td>31</td>
<td>30.4</td>
</tr>
<tr>
<td>R\text{C} (Ω)</td>
<td>19</td>
<td>20</td>
<td>19</td>
<td>17.5</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Drawn Width = 1um</td>
<td>R\text{link} (Ω)</td>
<td>190</td>
<td>161</td>
<td>180</td>
<td>149</td>
<td>186</td>
</tr>
<tr>
<td>R\text{parasitic} (Ω)</td>
<td>460</td>
<td>400</td>
<td>441</td>
<td>370</td>
<td>454</td>
<td>403</td>
</tr>
<tr>
<td>Drawn Width = 10um</td>
<td>R\text{link} (Ω)</td>
<td>29</td>
<td>28</td>
<td>29</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td>R\text{parasitic} (Ω)</td>
<td>139</td>
<td>136</td>
<td>140</td>
<td>124</td>
<td>140</td>
<td>135</td>
</tr>
</tbody>
</table>

Table 4-1: Extracted resistances for each of the parasitic elements of the TLM structure. R\text{parasitic} is the combined total of all the parasitic elements. Details for all drawn widths are given in Appendix B.
From the data it can be seen that the wider devices have a lower overall parasitic resistance ($R_{\text{parasitic}}$) as would be expected. However, $R_{\text{parasitic}}$ does not scale with width because $R_e$ and $R_{N^+}$ are constant for each device. These resistances will then have a relatively larger impact on the wider devices than on the narrower devices where $R_{\text{parasitic}}$ is dominated by $R_{\text{link}}$.

**Comparison of as measured vs. de-embedded data for $W=1\mu m$**

**Comparison of as measured and de-embedded data for $W=10\mu m$**

**Figure 4-10**: I-V characteristics of a 1um long TLM structure for as drawn widths of 1um and 10um. There is a larger correction for wider device as $R_{\text{parasitic}}$ does not scale with width.

Figure 4.10 compares the IV characteristics for the raw and de-embedded measurements for 1um and 10um widths at a length of 1um for the $1\times10^{13}$ cm$^{-2}$ samples.
This represents the worst case scenario as the intrinsic resistance will be at its lowest and the current levels at its highest for these widths. For the 10um sample there is up to a 25% difference between the applied and intrinsic voltages. This falls to 15% for the 1um sample.

4.3.2 Junction Quality

The I-V characteristics of the N+/substrate diode was determined in order to qualify the quality of the junction. The leakage current of the source/substrate and drain/substrate junctions adds to the body current and therefore must be small so that it does not obscure the hole current caused by impact ionization.

The silicon control samples break down at about 15.4 V and the strained silicon ones at 13.6 V (Figure 4.11). The breakdown of this junction sets the maximum measurement range of the TLM structures. The strained-Si samples also have higher reverse leakage current but this would be expected as the SiGe substrate has a smaller bandgap than Si. This leakage current sets a lower limit on the resolution of the body current ($I_b$) of $10^{-10}$ A for bulk Si and $10^{-8}$ A for strained-Si. However, this leakage current is consistent and repeatable and can therefore be subtracted out of the $I_b$ measurement to get a true value of the II generated hole current.

4.3.3 TLM Test Structure Measurements

A comparison between the bulk and strained Si de-embedded TLM structures for different n- doping levels is shown in Figures 4.12 to 4.14. Devices of length 100um, 10um, 5um and 1um are graphed to show the comparison across a range of fields. They
Figure 4-11: N+/substrate diode breakdown for the bulk and strained Si samples. The breakdown is repeatable across the different splits.

are all nominally 10um wide with the source current normalized with respect to width (after correcting for $W_{\text{offset}}$). The field is taken to be the extracted intrinsic voltage, $V_{\text{intrinsic}}$, divided by the length of the structure in question. This is based on the assumption that the field is uniform across the length of the n-region (this may not necessarily be the true and in this case the field can be viewed as being an average value across the structure).

In order to compare the enhancements due to the strained Si, the resistance of the 100um x 10um TLM was measured at low voltage levels (0 V to 0.2 V). The 100um length was chosen to give the lowest field and to minimize any errors due to the parasitic
resistances and variations in the length due to the lithography process. The 10um width was chosen so that the currents levels would be well above the background leakage current. This was especially important for the $1 \times 10^{12}$ cm$^{-2}$ doped sample. The results are summarized in Table 4.2.

The enhancement is highest for the lowest doping level and then falls significantly as the dose is increased. It is not clear as to why this occurs. In strained-Si MOSFETs the transconductance enhancement levels fall off at high doping levels [29], possibly due to coulombic scattering though this has not been absolutely proved. Upon further
examination of the characteristics it also clear that this enhancement decreases at higher fields and the strained Si samples saturate at significantly lower current levels than the bulk samples. For the $1 \times 10^{13}$ cm$^{-2}$ doped split, the saturation current of the bulk Si is more than twice that of its strained-Si counterpart. If this was due to large error in the dose then this difference in current levels should also be observed at lower field levels but it is not.
Figure 4-14: Comparison of the different length TLM structures for n- doping level of 5x10^{12} cm^{-2}. Filed is calculated from $V_{\text{intrinsic}}/L$.

Table 4-2: Resistance of the n- region as measured from a 100um x 10um TLM structure

<table>
<thead>
<tr>
<th>Doping Level (cm^{-2})</th>
<th>Strained-Si R (Ω)</th>
<th>Bulk Si R (Ω)</th>
<th>Enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x10^{12}</td>
<td>9.38x10^6</td>
<td>1.63x10^6</td>
<td>74%</td>
</tr>
<tr>
<td>5x10^{12}</td>
<td>5.28x10^4</td>
<td>6.87x10^4</td>
<td>30%</td>
</tr>
<tr>
<td>1x10^{13}</td>
<td>3.35x10^3</td>
<td>4.01x10^3</td>
<td>20%</td>
</tr>
</tbody>
</table>
The strained Si samples exhibit a negative slope in their current saturation region that is not evident in the bulk Si controls. Figure 4.15 shows this effect for the 1um TLM at different widths. As the width increases the normalized current decreases for the strained Si samples and shows a negative slope. This is a typical characteristic of self-heating. Considering the buffer layer thickness is of the order of 4.5um thick, this is not a surprising result. What is interesting is that the current does not extrapolate back to the same point for each of the widths as would be expected in a classic self-heating scenario. When the current is extrapolated back from this slope it gives the saturation current that would be achieved if there were no self heating present. This would result in about a 10% increase in the strained Si saturation current but this still does not come anywhere near that achieved by the bulk Si. So while self heating does have an effect, it in itself does not seem to account for the difference in the current levels.

Another interesting observation from the data is that TLMs do not saturate at the same current levels. As the length of the TLM decreases, the saturation current at the same nominal field increases. This implies that the saturation of the devices is not due to velocity saturation. Some other effect is coming in to play and is dominating the characteristics. As of yet it has not been determined what this effect is. One possibility that could be investigated is that devices are being pinched off by the body, with the body effectively acting as the gate of a JFET. If this is case, then the field across the TLM will not be uniform, making analysis of the impact ionization much more complex.
4.3.4 Impact Ionization in the TLM structures

Figure 4.16 shows the drain, source and body currents measured from a 1μm TLM structure and the associated II multiplication factor for strained and bulk Si for a dose of $1 \times 10^{13}$ cm$^{-2}$. The multiplication factor $M$ is defined as the ratio between total current and the current that initialized the multiplication process ($I_d$ and $I_s$ respectively in this case). The strained Si samples exhibit higher II (about an order of magnitude) even though the initial current is lower. This results in the strained Si device breaking down earlier (at a voltage of 7.73 V compared to 9.93 V for Si). Because of the uncertainty of
Figure 4-16: (a) I-V characteristics of a 1μm TLM structure. The body current is due to II across the n-region. The relative magnitude of the II is determined by M which is plotted in (b).
whether the TLM structures are exhibiting JFET characteristics, it is not clear what type of fields these voltages are producing across the n- region. Until this issue can be resolved the impact ionization coefficients cannot be extracted from the data as a precise knowledge of the field is required to carry out this analysis. Also if the fields are non uniform, this further complicates the analysis considerably.

4.4 Conclusions

This chapter presented the results from test structures fabricated to investigate impact ionization effects in strained-Si. The TLM structures function to a first order as resistors. The device junctions are clean and non leaky, allowing clear impact ionization components to be identified. Differences between the body current for the strained Si and bulk Si samples were observed indicating differences in impact ionization behaviour between the two systems.

There are effects occurring in the TLMs that have not yet been explained. These include why the saturation current of the strained Si samples at high fields is lower than that of bulk Si and what is the main physical effect causing the devices to saturate in the first place. Once these questions have been satisfactorily answered then an attempt can be made to extract the II coefficients from the data.

These results are to be viewed as a preliminary analysis of the data and are in need of confirmation and more detailed study.
CHAPTER 5

Conclusions

The main thrust of this thesis has been to determine how an RF power LDMOSFET may be implemented on strained-Si technology. Key issues for the realization of this technology were identified as the restricted thermal budget, gate oxide formation and potentially worse impact ionization effects than bulk Si. These issues were explored in detail by means of simulation and the fabrication of appropriate test structures.

The heterostructure design proposed for the LDMOSFET was the SGOI option with a Si$_{0.8}$Ge$_{0.2}$ buffer layer. SGOI was chosen as the heterostructure of choice for a number of reasons. Use of a buried oxide layer gives all the advantages normally associated with SOI, namely lower parasitic capacitances and power dissipation. It also allows for the use of STI isolation technology. In particular, for strained-Si technology, it allows the underlying buffer layer to be thinned considerably which is advantageous for self-heating effects.

The graded channel design of the LDMOSFET cannot be formed by a long high temperature drive-in cycle in strained-Si technology because of the limitations of the thermal budget. Instead a high angled implant must be used to set the body doping
profile characteristics. 2D simulations were carried out to investigate the conditions of this implant and of the drift region design.

It was found that the high tilt implants for the body formation (45° and 60°) showed the best characteristics. Using a high energy implant at these tilt angles was effective up to a point. If the peak of the implant is too deep in the substrate, then the surface doping concentration drops and $r_0$ is reduced because of CLM effects. A high $r_0$ is essential for power devices in order to achieve high gain.

The major trade-off in the n-drift region design was found to be between the breakdown voltage and the on-resistance. The doping of this region should be kept low and the length of the region long to improve the breakdown performance. However, this increases the on-resistance of the device and limits $I_{dsat}$. To recover the current drive, and hence the power output, the device may be made wider. This is undesirable for a power device as it increases the parasitic elements that hurt power efficiency. A reasonable dose and length for the n-drift region was found to be $5 \times 10^{12} \text{ cm}^{-2}$ and 0.5um.

In order to implement SOC, CMOS digital devices will have to fabricated on the same wafer. A dual gate oxide scheme would then have to be implemented as the gate oxide thickness of the LDMOSFET is nominally set at 10nm and that of the CMOS devices at 2nm. Two options were investigated for the LDMOSFET that would be integratable into a dual oxide scheme for SOC. Option one was a 750°C dry/wet/dry thermal oxidation on a thick strained-Si layer. Option two was a composite oxide consisting of a thin dry oxidation followed by an LTO deposition. Both options showed good characteristics based on C-V, leakage and $D_{it}$ measurements of capacitor test structures.
The lower bandgap of strained-Si and reduced scattering in this layer were determined as potentially increasing impact ionization and reducing the breakdown voltage. TLM test structures were fabricated to investigate this effect. The devices showed current enhancements over bulk Si at low lateral fields but these enhancements disappeared at higher fields. In fact the saturation current levels of the strained-Si samples were only about half that of the bulk-Si samples. Self-heating was observed in the strained-Si structures but it is not thought that this effect alone can account for the differences in the current levels.

Higher impact ionization was observed in the strained-Si devices compared to bulk-Si. For the same source current level, the measured hole current was an order of magnitude higher in the strained-Si samples resulting in a lower breakdown voltage for these structures. A more detailed analysis of the data needs to be carried out before the impact ionization coefficients can be extracted from the measurements.
Example of Source Code for 2D MEDICI Simulations.

This is a file that will generate a mesh for a standard or strained Si LDMOSFET. The profile is read in from SUPREMIV. The regions are defined to have a strained Si layer, relaxed SiGe buffer, and a buried oxide. The oxide region may just be referred to as silicon in the code to generate a bulk mesh. The material parameters may be turned on or off as required to have a genuine heterostructure or all Si substrate.

---

$********************************** INPUT PARAMETERS**********************************$

$ Lg $Lg$ Gate Length
$ Ls $Ls$ Source Length
$ Ld $Ld$ Drain Length
$ Ldrift $Ldrift$ Drift Region Length
$ Lc $Lc$ Contact Length
$ GtoCon $GtoCon$ Gate to Contact Distance
$ tox $tox$ Gate oxide thickness
$ tstrain $tstrain$ Thickness of strained Si layer if heterostructure is to be used. Otherwise, will double as thickness of inversion layer for the mesh generation.
$ tbuffer $tbuffer$ Buffer layer thickness
$ tpol$tpoly$ Thickness of gate polysilicon
$ tsub $tsub$ Wafer thickness
$ tburox $tburox$ Buried oxide thickness

$**********************************

loop steps=1

$ ASSIGN INPUT NAME OF PROFILE
assign name=input cl="ldmos"

$ ASSIGN OUTPUT NAME OF MESH
assign name=output c2="ldmos"

$ ASSIGN DEVICE PARAMETERS
assign name=Lg n.val=0.6
assign name=Ls n.val=3
assign name=Ld n.val=3
assign name=Lc n.val=1.5
assign name=Ldrift n.val=0.5
assign name=GtoCon n.val=1
assign name=tox n.val=0.0100
assign name=tstrain n.val=0.0100
assign name=tbuffer n.val=0.1000
assign name=tpoly n.val=0.1500
assign name=tburox n.val=0.1000
assign name=tsub n.val=3

$ GENERATE MESH

mesh RECTANGU smooth.k=1
assign name=left n.val=-1*@Ls
assign name=right n.val=@Lg+@Ldrift+@Ld
assign name=gateleft n.val=0
assign name=gater n.val=@Lg
assign name=driftend n.val=@Lg+@Ldrift
assign name=top n.val=-1-@tox-@tpoly
assign name=polytop n.val=-@tox-@tpoly
assign name=goxtop n.val=-@tox
assign name=strain n.val=@tstrain
assign name=boxtop n.val=@tstrain+@tbuffer
assign name=boxbot n.val=@tstrain+@tbuffer+@tburox
assign name=bottom n.val=@tsub

x.mesh n=1 l=@left
x.mesh n=\@Ls*3 \quad l=\@gateleft
x.mesh n=\@Ls*3+40 \quad l=\@gater
x.mesh n=\@Ls*3+40+20 \quad l=\@driftend
x.mesh n=\@Ls*3+40+20+\@Ld*3 \quad l=\@right

y.mesh n=1 \quad l=\@top
y.mesh n=5 \quad l=\@polytop
y.mesh n=8 \quad l=\@goxtop
y.mesh n=15 \quad l=0
y.mesh n=15+40 \quad l=\@strain
y.mesh n=15+40+20 \quad l=\@boxtop
y.mesh n=15+40+20+4 \quad l=\@boxbot
y.mesh n=15+40+20+4+10 \quad l=\@bottom/3
y.mesh n=15+40+20+4+10+5 \quad l=\@bottom

$\text{ELIMINATE ROWS AND COLUMNS}$

$\text{LTO ABOVE GATE}$
eliminate rows x.min=\@gateleft x.max=\@gater y.min=\@top y.max=\@polytop
eliminate rows x.min=\@gateleft x.max=\@gater y.min=\@top y.max=\@polytop
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@top y.max=\@polytop
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@top y.max=\@polytop
eliminate columns x.min=\@gater x.max=\@driftend y.min=\@boxtop y.max=\@bottom
eliminate columns x.min=\@gater x.max=\@driftend y.min=\@boxtop y.max=\@bottom
eliminate columns x.min=\@gater x.max=\@driftend y.min=\@bottom y.max=\@goxtop
eliminate columns x.min=\@gater x.max=\@driftend y.min=\@bottom y.max=\@goxtop
$\text{BUFFER BENEATH GATE}$
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@boxtop y.max=\@bottom
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@boxtop y.max=\@bottom
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@boxtop y.max=\@bottom
eliminate columns x.min=\@gateleft x.max=\@gater y.min=\@boxtop y.max=\@bottom

$\text{STRAINED SI LAYER IN SOURCE AND DRAIN}$
eliminate rows x.min=\@left x.max=\@gateleft y.min=0.001 y.max=\@strain
$\textbf{DEFINE REGION NAMES}$

region name=STRAIN silicon
+x.min=@left x.max=@right y.min=0 y.max=@strain

region name=BUFFER silicon
+x.min=@left x.max=@right y.min=@strain y.max=@boxtop

region name=GATEOX oxide
+x.min=@left x.max=@right y.min=@goxtop y.max=0

region name=SOI oxide
+x.min=@left x.max=@right y.min=@boxtop y.max=@boxbot

region name=SUBSTRATE silicon
+x.min=@left x.max=@right y.min=@boxbot y.max=@bottom

region name=LTO oxide
+x.min=@left x.max=@right y.min=@top y.max=@goxtop

$\textbf{DEFINE CONTACTS}$

elec name=gate
+x.min=@gateleft x.max=@gater y.min=@polytop y.max=@goxtop
+void

elec name=source
+x.min=-@GtoCon-@Lc x.max=-@GtoCon y.min=@top y.max=0
+void

elec name=drain
+x.min=@gater+@GtoCon+@Lc x.max=@gater+@GtoCon+@Lc y.min=@top y.max=0
+void

elec name=bulk bottom

$\textbf{READ IN DOPING PROFILE}$

profile uniform p-type n.peak=lel5 y.min=@boxbot y.max=@bottom
profile tsuprem4 in.file="PROFILE/"@input".prof" + x.min=@left x.max=@right y.min=0 y.max=@boxbot
$ Define gate contact as N+ Poly
contact num=gate n.poly

$ DEFINE MATERIAL PARAMETERS IF REQUIRED (OTHERWISE COMMENT OUT)

material region=STRAIN
$+eg.model=0
$+eg300=0.99
$+affinity=4.24
$+permittivity=11.9

material region=BUFFER
$+x.mole=0.3
$+eg.model=0
$+eg300=0.97
$+affinity=4.05
$+permittivity=13.1

models temp=300 unimob fldmob consrh auger bgn fermi impact.i

$ SPECIFY MOBILITY MODEL

$ These are for bulk Si device

mobility silicon fldmob=1 vsatn=7e6 MUN.UNI=635.2 ECN.UNI=6.857e5 EXN.UNI=0.9064

$ SAVE OUTPUT MESH FILE

save mesh w.models out.file="MESH/"@output".mesh"

$ INITIAL SOLUTION

symbolic gummel carriers=0
method iccg damped itlimit=40
solve

symbolic gummel carriers=2
method iccg damped itlimit=40
solve

symbolic newton carriers=2
method autonr n.damp px.toler=50e-5 cx.toler=50e-5
solve V(GATE)=0 V(DRAIN)=0

save solution
+ out.file="SOLS/"@output".init" w.models

1.end
## APPENDIX B

<table>
<thead>
<tr>
<th>Dose (cm$^2$)</th>
<th>Bulk Si 1x10$^{12}$</th>
<th>Strained Si 1x10$^{12}$</th>
<th>Bulk Si 5x10$^{12}$</th>
<th>Strained Si 5x10$^{12}$</th>
<th>Bulk Si 1x10$^{13}$</th>
<th>Strained Si 1x10$^{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{N+}$ (Ω)</td>
<td>31</td>
<td>31</td>
<td>31.2</td>
<td>27.5</td>
<td>31</td>
<td>30.4</td>
</tr>
<tr>
<td>$R_s$ (Ω)</td>
<td>19</td>
<td>20</td>
<td>19</td>
<td>17.5</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>$W_{offset}$ (um)</td>
<td>0.56</td>
<td>0.91</td>
<td>0.67</td>
<td>0.87</td>
<td>0.65</td>
<td>0.84</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>190</td>
<td>161</td>
<td>180</td>
<td>149</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>460</td>
<td>400</td>
<td>441</td>
<td>370</td>
<td>454</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>111</td>
<td>103</td>
<td>109</td>
<td>95</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>303</td>
<td>285</td>
<td>299</td>
<td>264</td>
<td>310</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>82</td>
<td>78</td>
<td>80</td>
<td>71</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>245</td>
<td>234</td>
<td>242</td>
<td>215</td>
<td>248</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>64</td>
<td>62</td>
<td>64</td>
<td>56</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>210</td>
<td>202</td>
<td>208</td>
<td>185</td>
<td>211</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>54</td>
<td>52</td>
<td>53</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>188</td>
<td>183</td>
<td>188</td>
<td>168</td>
<td>190</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>46</td>
<td>45</td>
<td>46</td>
<td>41</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>174</td>
<td>169</td>
<td>174</td>
<td>155</td>
<td>175</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>40</td>
<td>39</td>
<td>40</td>
<td>36</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>161</td>
<td>157</td>
<td>161</td>
<td>144</td>
<td>162</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>152</td>
<td>148</td>
<td>152</td>
<td>136</td>
<td>153</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>32</td>
<td>141</td>
<td>32</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>145</td>
<td>31</td>
<td>145</td>
<td>130</td>
<td>145</td>
</tr>
<tr>
<td>Width = 1um:</td>
<td>$R_{\text{link}}$ (Ω)</td>
<td>29</td>
<td>28</td>
<td>29</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{parasitic}}$ (Ω)</td>
<td>139</td>
<td>136</td>
<td>140</td>
<td>124</td>
<td>140</td>
</tr>
</tbody>
</table>

Table B-1: Extracted parasitic resistances for all TLM widths.
Bibliography


10. Ma, G.; Burger, W.; Shields, M.,“High efficiency 0.4 /spl mu/m gate LDMOS power FET for low voltage wireless communications”, Microwave Symposium Digest, 1999 IEEE MTT-S International, Volume: 3, 1999 Page(s): 1195 -1198 vol.3


