Abstract

Lin, Yanxia. Advanced Gate Stacks for Strained Silicon Devices (under the direction of Dr. Veena Misra and Dr. Mehmet C. Öztürk).

Due to the mobility enhancement provided by strained Si for both electrons and holes, as well as the scaling requirement and potential issues of polysilicon gate electrodes, alternative gate stacks are being pursued for strained Si devices, which warrant investigation for better understanding on the integration of high-κ dielectrics and metal gate electrodes. Mobility enhancement of strained Si devices has been reported even with ultra-thin SiO₂. However, additional scattering mechanisms related with high-κ dielectrics and strained Si may result in mobility degradation, which requires a fundamental study. Furthermore, impacts of integration of metal gate electrodes with strained Si channels are not fully understood.

In this work, an investigation of the degradation of electrical properties of several candidate metal gate electrodes on high κ dielectrics on strained Si was performed and compared with that of bulk Si samples. This work consists of three parts. Strained Si layers were grown on relaxed SiGe virtual substrates by ultrahigh vacuum rapid thermal chemical vapor deposition (UHV/RTCVD). High-κ dielectrics and metal gates were formed by physical vapor deposition (PVD) methods. The first part of the study focused on the optimization of experimental conditions and the investigation of results from material analysis. The second part of this study compared electrical data from MOS capacitors fabricated with metal gate electrodes on strained Si with SiO₂ as the gate dielectric with that of HfO₂. Different strained Si
thickness and different Ge concentration in the virtual substrate were employed to study the effects of strain and Ge out-diffusion on electrical properties. Results from strained Si MOSFETs on SiO₂ or HfO₂ with TaN gate electrodes achieved by standard and advanced electrical characterization, including mobility measurement, two and three level charge pumping methods, were analyzed in the last part. It was found that electrical properties degraded as the strained silicon thickness decreased, which was attributed to the presence of Ge in the strained Si layer, and more degradation was observed with SiO₂ which may be due to Si consumption during oxidation. This trend of increasing degradation with decreasing strained silicon thickness did not change after rapid thermal annealing. Metal gate electrodes were found to exhibit as good performance on strained Si as on bulk Si. Strain does not lead to any degradation of the high-κ/strained Si interface. Ge diffusion is the dominant cause of the Dₓ increase, which explains that samples with thinner strained Si films show less device performance enhancement. Less degradation with HfO₂ samples was observed due to the low temperature formation process of high-κ dielectrics. The mechanisms responsible for mobility degradation in strained Si devices with advanced gate stacks were discussed.
ADVANCED GATE STACKS FOR STRAINED SILICON DEVICES

by

Yanxia Lin

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ELECTRICAL AND COMPUTER ENGINEERING

Raleigh

2005

APPROVED BY:

________________________________________  __________________________________________
Dr. Veena Misra                                           Dr. Mehmet C. Öztürk
Chair of the advisory committee          Co-Chair of the advisory committee

________________________________________  __________________________________________
Dr. C.M. Osburn                                           Dr. G.N. Parsons
To the memory of my dear father

Zhiying Lin
(1932 ~ 2002)

My mother

Wenyieng Weng

My sister

Yanli Lin

My niece

Yujing Fang

and

My husband

Tao Ouyang
Biography

Yanxia Lin received her Bachelor degree in Materials Science and Engineering from Tsinghua University in July 1997, and her M.S. degree in Materials Science from the Institute of Semiconductors of Chinese Academy of Sciences in July 2000, both in Beijing, China. Upon completion of her studies in China, she was admitted to the PhD program in the Department of Electrical and Computer Engineering at North Carolina State University. She started her doctoral work in the summer of 2001 as a research assistant under the direction of Dr. Mehmet C. Öztürk. One year later she started working on the strained Silicon project under the guidance of Dr. Veena Misra, with Dr. Öztürk as her co-advisor.

Following graduation Yanxia Lin will join Spansion LLC in Austin, Texas.
Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Veena Misra, for providing me the research opportunities, as well as her guidance, support and encouragement throughout the past four years. Without her supervision, I would not be in the stage where I am now. I am also very thankful to Dr. Mehmet C. Öztürk, for all the knowledge that he has taught me on research and life.

I am also honored to have Dr. C. M. Osburn and Dr. G. R. Parsons in my committee. I would like to thank them for their interest and encouragement throughout this research.

Many thanks to the AEMP cleanroom staff including Joan O'Sullivan, Harold Morton, Dr. Ginger Yu, Henry Taylor, Myrick Peacock for all of the work they have done in keeping the laboratory running and their kindly help in processing.

I would also like to acknowledge many past and present graduate students including Jing Liu, Hongxiang Mo, Jennifer Yopp, Heather Lazar, Nemanja Pesovic, Bei Chen, Jaehoon Lee, Youseok Suh, Rashmi Jha, Qiliang Li, Guru Mathur, Saurabh Chopra, Sriv Godwa and Yan Du for their help and many joyful discussions.

Finally, I also wish to acknowledge my parents and my sister for their endless love, encouragement and support. I would like to thank my husband, Tao Ouyang, for his patience, understanding and sacrifice which made this degree possible.
Table of Contents

LIST OF FIGURES....................................................................................................................................................... IX
LIST OF TABLES......................................................................................................................................................... XV

CHAPTER 1 INTRODUCTION

1.1 CMOS TECHNOLOGY SCALING.................................................................................................................................1
1.2 STRAINED SILICON TECHNOLOGY ..........................................................................................................................4
  1.2.1 Why Is Strained Silicon Required for Future CMOS Devices .................................................................4
  1.2.2 Biaxial Strained Silicon and Device Applications .........................................................................................8
  1.2.3 Uniaxial Strained Silicon and Device Applications ......................................................................................12
1.3 ALTERNATIVE HIGH DIELECTRIC CONSTANT GATE INSULATOR MATERIALS..................................................16
  1.3.1 Why Are High $\kappa$ Dielectrics Required ........................................................................................................16
  1.3.2 Hafnium Based Dielectrics ............................................................................................................................18
  1.3.3 Additional Problems with High $\kappa$ Dielectrics ............................................................................................19
1.4 METAL GATE ELECTRODES ................................................................................................................................20
  1.4.1 Why Do We Need Metal Gates .......................................................................................................................20
  1.4.2 Current Candidates: Advantages and Problems ..........................................................................................21
1.5 ADVANCED GATE STACKS ON STRAINED SILICON AND CURRENT CHALLENGES ........................................22
  1.5.1 Strained Silicon with Novel Gate Stacks .....................................................................................................22
  1.5.2 Current issues: Process and Device Design .................................................................................................23
1.6 OUTLINE OF THE DISSERTATION ........................................................................................................................24
1.7 REFERENCES............................................................................................................................................................26

CHAPTER 2 FABRICATION OF STRAINED SILICON DEVICES INCORPORATING ALTERNATIVE GATE STACKS

2.1 EPITAXY OF $\text{Si}_{1-x}\text{Ge}_x$ AND STRAINED SI LAYER ........................................................................................................33
  2.1.1 UHV-RTCVD System ........................................................................................................................................33
  2.1.2 Surface Preparation Prior to $\text{Si}_{1-x}\text{Ge}_x$ and Strained Si Deposition ..........................................................35
CHAPTER 3 MATERIAL ANALYSIS AND ELECTRICAL CHARACTERIZATION

3.1 MATERIAL ANALYSIS ............................................................................................................................... 46

3.2 ELECTRICAL ANALYSIS .............................................................................................................................. 49

3.2.1 Capacitance-Voltage Measurement ....................................................................................................... 50

3.2.2 Interface Trap Density Measurement .................................................................................................... 52

3.2.3 Current-Voltage Measurement ........................................................................................................... 57

3.2.4 Mobility Measurement ....................................................................................................................... 61

3.2.4.1 Split C-V Method ........................................................................................................................ 62

3.2.4.2 Corrections of Mobility Extraction .................................................................................................. 65

3.2.5 Charge Pumping .................................................................................................................................. 68

3.2.5.1 Two Level Charge Pumping ......................................................................................................... 68

3.2.5.2 Three Level Charge Pumping ........................................................................................................ 71

3.3 REFERENCES ........................................................................................................................................... 74

CHAPTER 4 MATERIALS ANALYSIS AND ELECTRICAL CHARACTERIZATION OF
STRAINED SI FILMS AND HIGH-K DIELECTRICS

4.1 PROPERTIES OF Si_{1-x}Ge_{x} AND STRAINED SI FILMS USED IN THIS WORK .................................... 76

4.1.1 Selectivity of Deposition ..................................................................................................................... 76

4.2 PROPERTIES OF HIGH-K DIELECTRICS: HfO_{2} .................................................................................. 86

4.2.1 Electrical Characteristics of HfO_{2} Metal-Oxide-Semiconductor (MOS) Capacitors ......................... 87

4.2.2 X-ray Photoelectron Spectroscopy (XPS) and Transmission Electron Microscopy (TEM) .................. 90
4.3 Effects of Facets on Electrical Properties ................................................................. 96
  4.3.1 Overlap and Non-overlap Capacitors ...................................................................... 97
  4.3.2 I-V Measurement of Overlap and Non-overlap Capacitors .................................. 98

4.4 Strain Analysis by Raman Spectroscopy ................................................................. 100

4.5 References .................................................................................................................. 103

CHAPTER 5 MATERIALS ANALYSIS AND ELECTRICAL CHARACTERIZATION OF
STRAINED SI METAL-OXIDE-SEMICONDUCTOR (MOS) CAPACITORS

5.1 Electrical Characterization of Strained Si MOS Capacitors .................................. 106
  5.1.1 Electrical properties of Samples after Forming Gas Anneal (FGA) ...................... 107
  5.1.2 Electrical properties of Samples after Rapid Thermal Anneal (RTA) .................... 110

5.2 Possible Mechanisms of Electrical Property Degradation .................................. 112

5.3 Summary .................................................................................................................. 119

5.4 References ................................................................................................................ 121

5.5 Impact of Ge on Integration of HfO₂ and Metal Gate Electrodes on Strained Si Channels ... 122

CHAPTER 6 ELECTRICAL CHARACTERIZATION OF STRAINED SI MOSFETS

6.1 Strained Si MOSFETs with SiO₂ Gate Dielectric and Polysilicon or Tan Gate Electrodes..... 138
  6.1.1 Basic Device Characteristics: C-V and I-V ............................................................ 139
  6.1.2 Mobility Extraction .............................................................................................. 141

6.2 Strained Si MOSFETs with Tan Gate: SiO₂ or HfO₂ .................................................. 142
  6.2.1 Basic Device Characteristics: C-V and I-V ............................................................ 143
  6.2.2 Mobility Extraction .............................................................................................. 146
  6.2.3 Interface and Bulk Traps ...................................................................................... 148

6.3 Mobility Degradation Mechanisms in Strained Si MOSFETs with Tan Gate Electrodes..... 154
  6.3.1 Mobility Correction for Interface Traps ............................................................... 154
  6.3.2 Mobility Degradation Related to High-κ Dielectrics ............................................ 159
  6.3.3 Mobility of Strained Si MOSFETs at Higher Temperatures .................................. 161
CHAPTER 7 SUMMARY AND FUTURE WORK

7.1 CONCLUSIONS ...........................................................................................................175
7.2 FUTURE WORK .........................................................................................................177
List of Figures

Figure 1-1  Schematic representation of the strain induced conduction band splitting in silicon. .........................................................5

Figure 1-2  Energy alignment of the Si conduction band with and without the tensile strain in (a) bulk and (b) inversion layer, respectively. .................................................6

Figure 1-3  Simplified hole valence band structure for longitudinal in plane direction (a) unstrained and (b) strained silicon. ..............................................................7

Figure 1-4  Schematic illustrations of (a) equilibrium lattices and (b) pseudomorphic strained Si on relaxed SiGe (c) band alignments between strained Si and the relaxed SiGe virtual substrate. ..............................................................................9

Figure 1-5  Measured (symbols) effective mobility enhancement ratios, $r$, compared to calculations for the phonon limited MOS mobility (solid line) for strained Si $n$-MOSFETs. .................................................................................................11

Figure 1-6  Comparison of hole mobility enhancement ratios in strained Si $p$-MOSFETs as a function of vertical effective field, $E_{eff}$. The numbers beside the data are the substrate Ge percent. .................................................................................................................12

Figure 1-7  Schematic showing different types of strain induced in the silicon channel. ...................13

Figure 1-8  Electron saturated drive current improvement verses nitride thickness. ..........................15

Figure 1-9  Band offset calculations for a number of potential high-$\kappa$ gate dielectric materials. ....18

Figure 2-1  A schematic illustration of UHV-RTCVD system used in this work. ............................34

Figure 2-2  The main process chamber (MPC) of the UHV-RTCVD system consists of a quartz bell jar, a top lamp bank, and a side lamp bank .........................................................35

Figure 2-3  A typical deposition sequence used in Si$_{1-x}$Ge$_x$ and Si epitaxy. ..................37

Figure 2-4  Schematic of the UHV RF sputtering system: (a) top view; (b) side view. ........39

Figure 3-1  XRD spectra of 100nm Si$_{1-x}$Ge$_x$ film deposited on a crystalline bulk silicon substrate where x is about 50%. .................................................................47

Figure 3-2  (a) The C-V characteristic of a MOS capacitor on P-type Si substrate. (b) The electrical equivalent circuit of a MOS capacitor. .........................................................50
Figure 3-3 High and low-frequency C-V curves show the offset $\Delta C/C_{ox}$ due to interface traps. 53

Figure 3-4 Equivalent circuits for conductance measurements..................................................55

Figure 3-5 $G_p/\omega$ versus $\omega$ for a single level, a continuum and experimental data. For all curves: $D_i=1.9\times10^{9}\text{cm}^{-2}\text{eV}^{-1}$, $\tau_i=7\times10^{-5}\text{s}$ .................................................................57

Figure 3-6 Energy band diagram description of (a) Fowler-Nordheim injection through a triangular barrier and (b) direct tunneling through a trapezoidal barrier.................60

Figure 3-7 Plot of overall mobility versus effective field limited by various scattering mechanisms. ..............................................................................................................................................62

Figure 3-8 Schematics of experimental setup to measure (a) gate to channel capacitance for Split C-V, (b) gate to substrate capacitance, and (c) total gate capacitance; and (d) gate-to-channel and gate-to-substrate capacitance as a function of gate voltage.....64

Figure 3-9 Schematic of an nMOSFET showing the current components in the $I_d-V_g$ measurement..................................................................................................................................................67

Figure 3-10 Schematic of (a) the charge pumping measurement configuration; (b) the square wave pulse used in two level charge pumping.................................................................69

Figure 3-11 Schematic of the waveforms and resulting $I_{cp}$ of (a) the Base Sweep and (b) the Amplitude Sweep of two level charge pumping measurements.................................................70

Figure 3-12 Schematic of the waveform used in the three level charge pumping to profile interface traps with energies between $E_c$ and $E_i$ ........................................................................72

Figure 4-1 Incubation time plotted as a function of insulator materials.......................................78

Figure 4-2 XRD scans for the deposited films with same gas flow but different thickness. Deposition parameters are listed in Table 4-1.........................................................................................81

Figure 4-3 AFM surface scans of deposited films: (a) epi Si: RMS ~0.17 nm (Z scale 30 nm); (b) $\text{Si}_{1-x}\text{Ge}_x$: RMS ~0.59 nm (Z scale 30 nm); (c) strained Si (9 nm): RMS ~0.75 nm (Z scale 30 nm); (d) strained Si (15 nm): RMS ~0.60 nm (Z scale 30 nm); (e) strained Si (20 nm): RMS ~0.62 nm (Z scale 20 nm). .........................................................................................85

Figure 4-4 RMS roughness of deposited films plotted as a function of strained Si thickness. ....86
Figure 4-5  C-V and I-V characteristics of four HfO$_2$ samples whose process conditions are listed in Table 4-3: (a) C-V data; (b) $J_g$ plotted as a function of voltage across the oxide ($V_{ox}$); (c) $J_g$ of SiO$_2$ from reference; (d) deviation of $J_g$ observed from the I-V data of sample Hf3.  

Figure 4-6  The XPS spectra of (a) Hf 4$f$ and (b) Si 2$p$ core-levels of two HfO$_2$/Si samples.  

Figure 4-7  The XPS spectra of (a) Hf 4$f$ and (b) Si 2$p$ core-levels of a RTA HfO$_2$/strained Si sample compared to the HfO$_2$/bulk Si system.  

Figure 4-8  A TEM image of a strained Si sample with high $\kappa$ and metal gate electrode stacks. All layers were labeled in (a) and (b) while details of HfO$_2$ and the interfacial layer were shown in (c).  

Figure 4-9  A SEM image of selectively deposited SiGe layer confined in the active area, achieved by S. Chopra working in the same group as the author in NCSU.  

Figure 4-10  Schematic illustration of cross-sections of overlap and non-overlap capacitors.  

Figure 4-11  Leakage current densities of overlap and non-overlap capacitors with (a) SiO$_2$ and (b) HfO$_2$ as gate dielectrics.  

Figure 4-12  Raman spectra of strained Si samples with different thicknesses: (a) a typical Raman spectrum of a strained Si sample; (b) different strain level was achieved by varying strained Si thickness; (c) Raman peak shifts plotted as a function of strained Si thickness.  

Figure 5-1  C-V curves from MOS capacitors with SiO$_2$ and HfO$_2$ gate dielectrics and TaN gate electrodes. The measured area is 50$\mu$m by 50$\mu$m.  

Figure 5-2  (a) Equivalent oxide thickness and (b) leakage current density of SiO$_2$ and HfO$_2$ samples are plotted vs. strained Si thickness. The gate electrodes are TaN.  

Figure 5-3  Hysteresis of SiO$_2$ and HfO$_2$ samples are plotted vs. strained Si thickness, with TaN as the gate electrodes.  

Figure 5-4  The density of interface traps ($D_{it}$) plotted as a function of (a) the trap energy; (b) strained Si thickness.
Figure 5-5  Electrical parameters of TaN gate MOS capacitors after RTA plotted vs. strained Si thickness: (a) leakage current density at 1V beyond $V_{FB}$; (b) EOT; (c) hysteresis...111

Figure 5-6  Interface trap density ($D_{it}$) after RTA is plotted as a function of strained Si thickness. The gate electrode is TaN. ..........................................................112

Figure 5-7  Ge profiles from SIMS show the effect of high temperature process on Ge out-diffusion. Samples have 20nm strained Si layer. ...........................................114

Figure 5-8  XPS spectrum of Ge 3d core level of a 9nm strained Si/ 3.5nm SiO$_2$ sample with the oxide removed. ........................................................................................................115

Figure 5-9  The density of interface traps ($D_{it}$) plotted as a function of (a) the trap energy; (b) Ge concentration with varying strained Si thickness. TaN is used as the gate electrode. .........................................................116

Figure 5-10  Raman peak shifts plotted vs. strained Si thickness with varying Ge content in the virtual substrate. ..............................................................................................................117

Figure 5-11  SIMS profiles of Si and Ge of two strained Si samples with different Ge content in the SiGe buffer layer. It can be seen that more Ge out-diffusion into the strained Si channel would be expected in sample B than in sample A ........................................118

Figure 5-12  $D_{it}$ plotted as a function of Raman peak shift with varying Ge content in the SiGe virtual substrate. Sample A and B refer to the conditions listed in Figure 5-11 ......119

Figure 6-1  Basic electrical characteristics of bulk Si and strained Si (16nm) MOSFETs with POLY and TaN gate electrodes: (a) C-V curves; (b) drain currents..................140

Figure 6-2  Mobilities of polysilicon and TaN metal gates on SiO$_2$ dielectrics extracted by Split C-V analysis. .................................................................................................................................142

Figure 6-3  (a) C-V curves of TaN nMOSFETs on bulk Si or 20nm strained Si with SiO$_2$ or HfO$_2$; (b) EOT values plotted as a function of strained Si thickness.........................144

Figure 6-4  Comparison of gate leakage current for strained Si and bulk Si nFETs with SiO$_2$ and HfO$_2$. ......................................................................................................................145

Figure 6-5  Drain currents of TaN gate MOSFETs on SiO$_2$ and HfO$_2$..............................146
Figure 6-6 Effective mobility of SiO$_2$ nFETs plotted vs. effective field. Three different strained Si thicknesses were employed. Bulk Si nFET was used as a control. .........................147

Figure 6-7 Effective mobility of HfO$_2$ nFETs plotted vs. effective field. Three different strained Si thicknesses were employed. ....................................................................................148

Figure 6-8 Charge pumping currents plotted vs. base sweep biases for TaN/HfO$_2$ devices on (a) bulk Si; (b) 9nm strained Si. The measurements were carried out at 100kHz. ............150

Figure 6-9 Charge pumping currents plotted vs. amplitude sweep biases for TaN/HfO$_2$ devices on (a) bulk Si; (b) 20nm strained Si. The measurements were carried out at 100kHz. ..............................................................151

Figure 6-10 $D_a$ measured from two level charge pumping at 100kHz for nFETs on SiO$_2$ and HfO$_2$ with different strained Si thickness. ...........................................................................151

Figure 6-11 $D_a$ extracted from three level charge pumping as a function of (a) bandgap; (b) strained Si thickness for strained Si MOSFETs with TaN gates on SiO$_2$ and HfO$_2$ ..............................................................153

Figure 6-12 Effective electron mobility after $D_a$ corrections for TaN strained Si MOSFETs on SiO$_2$ and HfO$_2$. .................................................................................................................................157

Figure 6-13 Effective electron mobility extracted with $D_a$ corrections for nMOSFETs on (a) SiO$_2$ and (b) HfO$_2$. .................................................................................................................................158

Figure 6-14 Peak mobility enhancement factor $\gamma$ plotted vs. strained Si thickness. ...............159

Figure 6-15 Possible sources of scattering in high-$\kappa$ gate stacks. ........................................160

Figure 6-16 HfO$_2$ limited mobility component for bulk Si and strained Si devices plotted as a function of effective field. ..............................................................161

Figure 6-17 Peak mobility plotted as a function of temperature for TaN gate MOSFETs on SiO$_2$ and HfO$_2$ dielectrics with bulk Si and/or strained Si channels. .........................162

Figure 6-18 $D_{tr}$-corrected effective mobility and calculated phonon limited mobility using Matthiessen’s rule for bulk Si and 20nm strained Si nMOSFETs on HfO$_2$. ...............165
Figure 6-19 “Phonon-limited” effective mobility components of n-channel devices plotted as a function of effective field on bulk Si and strained Si. Theoretical values calculated in Ref. A and ref. B are also included...........................165

Figure 6-20 The phonon, surface roughness and interface scattering mobility components of a strained Si sample simulated by using Hauser’s MOB2D model compared to those of a bulk Si sample. .................................................................................................167

Figure 6-21 Modeled peak mobility plotted vs. temperature with different interface scattering density (I.S.) and surface roughness (S.R.) parameter values extracted from the Hauser NCSU MOB2D model and Lazar’s Thesis...................................................169
List of Tables

Table 1-1  High performance logic technology requirements -- Near Term .............................................2
Table 1-2  Potential solutions to improve device performance .................................................................3
Table 1-3  Effects of stress on the MOSFET performance in different directions for a <110> oriented channel .............................................................................................................................14
Table 2-1  Typical process conditions for Si and Si1-xGe_x deposition used in this work ..................37
Table 2-2  Metal gate electrode sputtering conditions and rates ..............................................................42
Table 2-3  Summary of strained Silicon MOSFET process flow ...........................................................43
Table 4-1  Summary of XRD extracted Ge concentration in deposited films as a function of film thickness. The deposition temperature was 500°C and the deposition pressure was 700 mTorr maintained by adjusting the throttle valve .................................................................81
Table 4-2  Experimental conditions considered in this study to optimize deposition parameters without using the throttle valve ..........................................................................................83
Table 4-3  Process parameters used in HfO2 formation experiments ......................................................87
Table 4-4  Electrical properties extracted from C-V data and leakage current densities before and after RTA of HfO2 samples .........................................................................................88
Table 6-1  Extracted device parameters for SiO2 MOSFETs with polysilicon and TaN gate electrodes. Strained Si (SSI) thickness is the starting thickness before gate oxidation. .........................................................140
Table 6-2  Extracted device parameters of TaN gate MOSFETs on SiO2 and HfO2 ..........................146
Table 6-3  Constants used to extract average D_it values from two level charge pumping measurement ..................................................................................................................................149
Table 6-4  Parameters used in three level charge pumping to determine the D_it of TaN gate devices on SiO2 and HfO2 ..............................................................................................................152
Chapter 1 Introduction

1.1 CMOS Technology Scaling

Integrated circuit (IC) technology has been improving for over 40 years, following Moore’s Law by consistently scaling the design rules, increasing the chip and wafer size, and cleverly improving the designs of devices and circuits [1, 2]. For both memory and logic chips, the speed and density increased exponentially while the power dissipation and cost per function decreased. For example, the speed of microprocessors has been doubling approximately every three years, increasing from 2 MHz for the Intel® 8080 in the mid-1970s to over 1 GHz for the current chips [3]. Continuous MOSFET scaling resulted in advancing from the ~ 8 \( \mu \text{m} \) technology in 1972 to the current 90 nm technology, which corresponds to a reduction of ~ 0.87 per year [4]. If the scaling continues at this rate, the IC industry will face increasing difficulties due to the fundamental limits of certain devices and materials.

The Semiconductor Industry Association (SIA) has been publishing roadmaps since 1992. The most recent 2004 International Technology Roadmap for Semiconductors (ITRS) predicts the next 14 years [2]. In comparison to the 1999 ITRS [5], one major change is acceleration in the scaling of the physical gate length (\( L_g \)) driven by the industry’s need to maximize the chip speed [4]. Table 1-1 presents all the parameters required by High performance logic technology in the next several years as projected by the 2004 ITRS.
For digital circuits, the gate delay is defined as $\tau = CV / I$ where $C$ is the gate capacitance, $V$ is the voltage swing and $I$ is the drive current. While the drive current is increasing linearly with scaling the transistor off current is increasing exponentially. Both a large drive current and a small standby power are needed. Device performance can improve by 1) inducing a larger charge density for a given gate voltage drive; 2) enhancing the carrier transport; 3) ensuring device scalability to
achieve a shorter channel length; and 4) reducing parasitic capacitances and parasitic resistances [6]. Table 1-2 summarizes these opportunities/challenges and corresponding proposed technology options. These options can be classified into two categories: new materials and new device structures, which are usually related.

### Table 1-2 Potential solutions to improve device performance [6].

<table>
<thead>
<tr>
<th>Source of improvement</th>
<th>Parameters affected</th>
<th>Method</th>
</tr>
</thead>
</table>
| Charge density        | • S (inverse subthreshold slope)  
                       | • Q\textsubscript{inv} at a fixed \textsubscript{I}\textsubscript{off} | 1. Double-gate FET.  
                       | 2. Lowered operating temperature. |
| Carrier transport     | • Mobility (\(\mu\textsubscript{eff}\))  
                       | • Carrier velocity  
                       | • Ballistic transport | 1. Strained silicon.  
                       | 2. High-mobility and saturation-velocity materials (e.g., Ge, InGaAs, InP).  
                       | 3. Reduce mobility degradation factors (e.g., reduced transverse electric field, reduced Coulomb scattering due to dopants, reduced phonon scattering).  
                       | 4. Shorter channel length.  
                       | 5. Lowered operating temperature. |
| Parasitic resistance  | • \(R\textsubscript{ext}\) | 1. Extended/raised source/drain.  
                       | 2. Low-barrier Schottky contact. |
| Parasitic capacitance | • \(C\textsubscript{jn}\)  
                       | • \(C\textsubscript{GD}, C\textsubscript{GS}, C\textsubscript{GB}\) | 1. SOI.  
                       | 2. Double-gate FET. |
| Ensuring device scalability to a shorter channel length | • Generalized scale length (\(\lambda\))  
                       | • Channel length (\(L\textsubscript{g}\)) | 1. Maintaining good electrostatic control of channel potential (e.g., double-gate FET, ground-plane FET, and ultrathin-body SOI) by controlling the device physical geometry and providing means to terminate drain electric fields.  
                       | 2. Sharp doping profiles, halo/pocket implants.  
                       | 3. High gate capacitance (thin gate dielectrics, metal gate electrode) to provide strong gate control of channel potential. |

As summarized in the table above, strained silicon could be one of the candidates to enhance the carrier transport. A high dielectric constant (\(\kappa\)) material is required to achieve low equivalent oxide thickness (EOT). In addition, metal gate electrodes are employed to integrate with high-\(\kappa\) dielectrics as well as silicon dioxide.
to ensure the device scalability and eliminate problems brought about by polysilicon gates. All the advantages and current issues of strained silicon, high-κ dielectrics, and metal gates will be discussed in detail in the following sections.

1.2 Strained Silicon Technology

1.2.1 Why Is Strained Silicon Required for Future CMOS Devices

Alternative channel materials that offer higher carrier mobilities are needed to improve the carrier transport properties of future MOSFETs. Unfortunately, replacing Si with a new material does not necessarily provide the performance enhancement expected from the higher mobility due to other shortcomings or performance challenges of the new material [7]. This is one of the key reasons behind the recent excitement over strained Si. Essentially, it is now well established that without changing the channel material, significant enhancements in device performance are possible using strained silicon [8-23]. The theory of mobility enhancement in strained Si is still evolving [10]. The most commonly accepted explanation is that under the biaxial tensile strain, the six-fold degenerate valleys in Si are split into two groups. The group with the lower energy is two fold degenerate (labeled as $\Delta_2$ in Figure 1-1), which is the primary contributor to carrier transport at low fields. The in-plane effective mass of the electrons occupying these bands is approximately equal to the Si transverse effective mass ($m^{*}_{t} = 0.19m_0$). On the other hand, the effective mass perpendicular to the transport plane is equal to the longitudinal effective mass ($m^{*}_{l} = 0.92m_0$). The schematic representation of the energy ellipses is shown in Figure 1-1[8]. The energy of the conduction-band minima of the four valleys on the in-plane <100> axes rises with respect to the energy of the two valleys on the <100> axes.
perpendicular to the plane [24, 25], as shown in Figure 1-2 (a) [26]. The energy between the two-fold degenerate and the four-fold degenerate valleys, $\Delta E_{\text{strain}}$, is given by $\Delta E_{\text{strain}} = 0.67x \text{ eV}$, where $x$ is the Ge content of the relaxed Si$_{1-x}$Ge$_x$ substrate [9]. It should be noted that even in an unstrained Si MOS inversion layer there is band splitting between the sub-band energies in the two and the four-fold valleys due to quantization in the inversion layer. In a strained Si MOS inversion layer, the band splitting of the conduction band $\Delta E_{\text{strain}}$ is superimposed on this quantization, as schematically shown in Figure 1-2 (b) [26]. The electrons populate the lower $\Delta_2$ valleys with lighter effective mass, which results in the reduction of the average conductivity effective mass.

Figure 1-1 Schematic representation of the strain induced conduction band splitting in silicon [8].
Figure 1-2  Energy alignment of the Si conduction band with and without the tensile strain in (a) bulk and (b) inversion layer, respectively.

The other mechanism of mobility enhancement proposed by Takagi et al. [26] is the suppression of intervalley phonon scattering due to the energy splitting between the two fold and the four fold valleys. Intervalley phonon scattering is an inelastic process associated with the absorption or the emission of the relevant phonons with a large wave vectors. Thus, if the amount of the band splitting between the two and the four fold valleys becomes larger than the energy of the relevant phonons, the scattering probability will be significantly reduced. Therefore, better enhancement of the electron low field mobility will be achieved with higher Ge content in the relaxed Si$_{1-x}$Ge$_x$ substrate.

In unstrained material, the valence band maximum is composed of three bands: the degenerate heavy-hole (HH) and light-hole (LH) bands at $k=0$, and the split-off (SO) band which is slightly lower in energy, as shown in Figure 1-3 [12]. The biaxial stress can be resolved into a hydrostatic and a uniaxial stress component.
The hydrostatic stress equally shifts all three valence bands, while the uniaxial stress lifts the degeneracy between LH and HH bands by lifting the LH band higher than HH. The SO band is also lowered with respect to the other two bands. This leads to the population of holes in the energetically favorable LH like band. Application of stress also changes the shape of the bands as shown in Figure 1-3 (b). Therefore, due to the band deformation, the in-plane transport mass becomes smaller and the interband scattering is also suppressed. Thus the hole mobility is improved.

![Figure 1-3](image)

**Figure 1-3** Simplified hole valence band structure for longitudinal in plane direction (a) unstrained and (b) strained silicon [27].

The main difference between the effects on electron and hole mobilities is that the mobility of holes can be enhanced only at lower electric fields while the enhancement can be achieved at higher vertical electric fields for electrons. Fischetti et al. showed that the loss in hole mobility enhancement at higher fields was due to reduction in the separation between the light hole and heavy hole bands ($\Delta_{LH-HH}$) [28]. Based on the experimental data, it was speculated that this was due to the confining surface potential operating against the applied biaxial stress and trying to reduce the separation between the LH and HH bands [28]. The utilization of uniaxial strained...
silicon devices can improve hole mobility at both low and high electric fields, which will be presented in section 1.2.3.

1.2.2 Biaxial Strained Silicon and Device Applications

Strain in Si can be introduced in various ways. The most commonly used method is to deposit a thin Si epitaxial layer on top of a thick relaxed SiGe buffer layer [11, 14, 29]. Silicon and germanium, both crystallizing in the diamond lattice, can form a continuous series of Si_{1-x}Ge_x solid solutions with x ranging from 0 to 1. The lattice constants are 0.5431 nm for Si and 0.5657 nm for Ge. The lattice mismatch between Si and Ge, herein referred to as f, is about 4.2 %, which is sufficiently small. Therefore the deposited first several atomic layers will be strained to match the substrate and a coherent interface will be formed. The structure stores a high amount of elastic strain energy because interatomic bond lengths in the epilayer are stretched or compressed compared to their equilibrium values. At some epilayer thickness, generally called the critical thickness h_c, it becomes energetically favorable to relieve the elastic strain energy by introducing misfit dislocations and allowing the epilayer to relax towards its bulk lattice parameter. The critical thickness has been first calculated by several groups based on different models [30-35]. When the thickness of Si_{1-x}Ge_x alloy exceeds the critical thickness, the alloy will be called a “relaxed” Si_{1-x}Ge_x film. The lattice parameter could be calculated according to the Vegard’s law, assuming the film is fully relaxed:

\[ a_{SiGe}(x) = a_{Si} \cdot (1-x) + a_{Ge} \cdot x \]  

(1.1)

The primary function of this relaxed SiGe layer is to serve as a “virtual substrate” creating tensile strain in the top Si epilayer, as shown in Figure 1-4.
However it has no impact on the improvement of device performance. The thickness of strained Si must be relatively thin so that the strain will not be relaxed through misfit dislocations. Figure 1-4 (c) shows the band offset for a strained-Si film grown on relaxed (001) SiGe substrate. In this case, a large band offset is obtained in both the conduction and valence bands relative to the relaxed Si$_{1-x}$Ge$_x$ layer and is called a type II band offset [25, 36]. This allows both electron and hole confinements in the strained Si layer, making it useful for both n- and p-type devices for strained-Si/SiGe based CMOS technology. The critical thickness of Si layers grown on relaxed uniform SiGe layers has also been calculated [37].

Figure 1-4 Schematic illustrations of (a) equilibrium lattices and (b) pseudomorphic strained Si on relaxed SiGe [8] (c) band alignments between strained Si and the relaxed SiGe virtual substrate [38].
Recent work has provided encouraging experimental data showing the mobility enhancement with different devices structures, including both n-channel and p-channel devices. Research on strained Si MOSFETs can be grouped into two categories: “bulk” strained Si on relaxed SiGe FETs (SS FETs) and strained Si on relaxed SiGe on insulator FETs (SGOI FETs). Recent work indicates that a conventional CMOS process flow can be adopted while still achieving the mobility and current drive enhancement [20]. Fabrication processes such as source/drain extensions and halos, channel ion implantation, and associated high temperature activation anneals are shown to have no adverse impact on device characteristics.

The first strained Si n-MOSFETs were fabricated on relaxed Si$_{0.7}$Ge$_{0.3}$ substrates and provided about 70% electron mobility enhancement with a vertical effective electric field ($E_{\text{eff}}$) up to 0.6 MV/cm [21]. At a lower $V_{\text{GS}}$, the current drive enhancement over the unstrained silicon control is as large as 50%, while at $V_{\text{GS}} = 0.8$ V, the current drive of a strained Si device is ~ 35 % higher [6]. Figure 1-5 shows both the experimental data and the theoretical values of the phonon limited electron mobility enhancement versus the substrate Ge content [22]. With the Ge content above 20 %, the mobility enhancement factor, $r$, saturates near 1.8, in agreement with calculations of the impact of strain on the mobility. Experiments also indicate that for electrons, the strain induced mobility enhancement factor is relatively constant with $E_{\text{eff}}$. For the same channel length, the current drive enhancements can be significantly changed due to the variation of doping profiles, strain induced band changes, and oxide thickness even for nanoscale MOSFETs.
The combination of high mobility strained Si devices with SOI structures will provide additional advantages such as reduced parasitic capacitances, improved isolation, and reduced short channel effects. Strained Si-on-insulator (SSOI) MOSFETs can be fabricated from strained silicon grown on relaxed Si$_{1-x}$Ge$_x$-on-insulator (SGOI) virtual substrates [12, 15-19, 23, 37, 38]. SGOI can be achieved via several approaches such as “etch-back” and “smart-cut” processes [12, 15-17, 23], SIMOX technology [14, 18, 37], and Ge condensation techniques [12, 19, 38].

The Si valence band degeneracy is also split by biaxial tensile strain induced by growth in relaxed Si$_{1-x}$Ge$_x$ (~ 40 meV/10 % Ge). Therefore, hole mobility enhancement can also be obtained with strained Si p-MOSFETs. Generally, larger strain is required to get valence band splitting, which means higher Ge content is needed in the relaxed SiGe virtual substrate. Recent research data is shown in

Figure 1-5  Measured (symbols) effective mobility enhancement ratios, $r$, compared to calculations for the phonon limited MOS mobility (solid line) for strained Si n-MOSFETs [22].
Figure 1-6 [22] where the hole mobility enhancement is primarily achieved in the low $E_{\text{eff}}$ range ($<1$ MV/cm) and the mobility enhancement ratio approaches 1 with the Ge content below 30 % in the underlying relaxed SiGe substrate. Unlike electron mobility, hole mobility enhancement is reduced for higher $E_{\text{eff}}$. To improve the hole mobility further, the substrate Ge content should be increased. This could be the solution for the single strained Si channel structures, which is feasible by using SIMOX and Ge condensation techniques.

![Figure 1-6 Comparison of hole mobility enhancement ratios in strained Si p-MOSFETs as a function of vertical effective field, $E_{\text{eff}}$. The numbers beside the data are the substrate Ge percent.](image)

1.2.3 Uniaxial Strained Silicon and Device Applications

There are various methods to induce uniaxial stress, which differ from one company to another. The thermal mismatch of silicon, the isolation materials, novel
junction structures, and gate electrodes have been reported recently to cause sufficient local strain, which can alter device characteristics [28, 39-47].

Intel demonstrated strained silicon MOSFETs with uniaxial strain based on the structure shown in Figure 1-7[39, 40]. Piezoresistance coefficients in silicon were used to model the behavior of uniaxial stress in MOSFETs. This model is only valid for small stress values where the mobility enhancement mainly results from the change in the conductivity effective mass. This is a good assumption since Intel used low stress levels for its MOSFETs and fewer defects are created requiring less alteration to the existing technology. Detailed discussions of the effect of mechanical stress on the mobility can be found in literature [39-41]. A summary of the effects of various stress components on the channel mobility is listed in Table 1-3. It can be seen that in order to achieve mobility enhancement for a <110> channel, longitudinal compressive stress for pMOSFETs and longitudinal tensile and out of plane compressive stress for nMOSFETs are most effective. Therefore SiGe source/drain junctions on the PMOS and a tensile capping layer on the NMOS were employed by Intel to induce strain.

![Figure 1-7 Schematic showing different types of strain induced in the silicon channel [40].](image-url)
Table 1-3  Effects of stress on the MOSFET performance in different directions for a 
<110> oriented channel [39, 41, 42].

<table>
<thead>
<tr>
<th>Directions</th>
<th>NMOS</th>
<th>PMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Along channel (x)</td>
<td>+ + +</td>
<td>+ + + +</td>
</tr>
<tr>
<td>Across channel (z)</td>
<td>+ +</td>
<td>+ +</td>
</tr>
<tr>
<td>Vertical (y)</td>
<td>+ + + +</td>
<td>+</td>
</tr>
</tbody>
</table>

In addition to achieving a higher hole mobility enhancement at low vertical electric fields deduced from the piezoresistance coefficients [43], uniaxial strain also maintains this enhancement at higher electric fields, which has been demonstrated experimentally. Fischetti et al. has theoretically explained the loss of hole mobility enhancement at higher fields by using reduced separation between the LH and HH like bands [28]. However, for uniaxial stress, the confining surface potential does not reduce the strain induced band separation as it does in the biaxial case. This can be attributed to the band warping caused by the uniaxial stress to create an advantageous out of plane effective mass for the top energy band. Therefore, the hole population in the energetically favorable LH like band would be enhanced and hence the mobility enhancement at higher fields would be maintained.

For the corresponding NMOS device fabricated by Intel, a nitride capping layer which created longitudinal tensile and out of plane compressive stress in the silicon channel was used to obtain the electron mobility enhancement [39]. The enhancement in performance was shown to be strongly dependent on the thickness of the capping layer. As shown in Figure 1-8, a capping layer thickness of ~75 nm showed 10 % improvement in drive current.
Texas Instruments (TI) also presented PMOS transistors with a channel length of 37 nm built on successful integration of a recessed SiGe epitaxial layer at the drain extension location. Compressive stress induced by the SiGe layer resulted in 35 % improvement in current drive [44]. Yang et al. reported a high performance CMOS flow in which nitride contact liners were used as dual stress liners (DSL) to induce both tensile and compressive stress achieving mobility enhancement for both electrons and holes. This DSL approach results in 15 % and 32 % effective drive current enhancement for nFET and pFET, respectively and a saturated drive current enhancement of 11 % for nFET and 20 % for pFET [45]. Strain Enhancing Laminated Si$_3$N$_4$ (SELS) was employed by Fujitsu [46] using a new process flow in which SELS was formed selectively only on the nMOS gate. Multiple layers of Si$_3$N$_4$ created higher strain at the corner sidewalls enhancing the channel strain. Drive
currents of $1120\mu A/\mu m$ and $690\mu A/\mu m$ at $V_{dd}=1V$/$I_{off}=100nA/\mu m$ were demonstrated for a 37 nm gate nMOS and a 45 nm gate pMOS, respectively. Uniaxial strain was also induced by selective epitaxy of silicon carbide (SiC) in the source and drain (S/D) regions of sub 100nm gate nMOS transistors [47]. The carbon mole fraction was 1.3 % such that the lattice mismatch between SiC and Si was 0.65%, resulting in tensile strain along the Si channel and compressive strain normal to the channel. Both the tensile and compressive stress contributed to substantial electron mobility enhancement and ~50 % enhancement in drive current was obtained for a gate length of 50 nm.

1.3 Alternative High Dielectric Constant Gate Insulator Materials

1.3.1 Why Are High $\kappa$ Dielectrics Required

One of the biggest concerns in modern CMOS technology is the gate dielectric. Silicon dioxide (SiO$_2$) has been the ideal gate dielectric because of its amorphous structure, a large band gap of ~ 9 eV, insolubility in water, ability to serve as a separator between metal conducting layers and silicon substrates, and its compatibility with Si having low interface state densities [48]. As the gate oxide begins to scale down to the ultra thin oxide regime, the direct tunneling current increases [49]. In addition, quantum mechanical effects and the polysilicon depletion effect can be amplified as the SiO$_2$ thickness is reduced, which results in an increase in the effective dielectric thickness [50, 51]. Due to these limitations of SiO$_2$, alternative gate dielectrics with higher dielectric constant are required such that physically thicker dielectrics can be used to reduce the probability of electrons and holes tunneling.
Besides the high permittivity, additional requirements for alternate high-κ dielectrics include a large band gap for appropriate barrier height, thermodynamic stability on Si, noncrystalline film morphology, ability to form a high quality interface with Si (with low fixed charge and low interface charge), and compatibility with gate electrodes and existing CMOS processing. Current research is focused on dielectrics with moderate κ since the permittivity must be balanced with the band offset, which is the barrier height for tunneling processes. Figure 1-9 shows the key properties (dielectric constants and band offsets) of potential high-κ dielectric candidates [52, 53]. Therefore, to achieve low gate leakage, those materials with band offsets less than 1 eV will not be desirable candidates as gate insulators. Most of the high-κ metal oxide systems investigated so far have unstable interfaces with Si. Therefore an interfacial layer is formed which plays a dominant role in the device electrical properties. It is necessary to study the thermodynamics of these systems thereby attempting to control the interface between high-κ and Si [54]. Currently, most of the high-κ materials show D↑ of $10^{11}$–$10^{12}$ cm$^{-2}$-eV$^{-1}$ and fixed charge density >$10^{12}$ cm$^{-2}$ at the interface, which is higher than the typical midgap interface density of Si ($2 \times 10^{10}$ cm$^{-2}$) [55-58]. It is desirable to select a material which remains amorphous throughout the entire CMOS process since amorphous films exhibit isotropic electrical properties and will not suffer from grain boundaries. The last concern is the compatibility, i.e. the deposition process for high-κ dielectrics must be compatible with conventional CMOS processing based on the consideration of cost and throughput.
1.3.2 Hafnium Based Dielectrics

For the past several years, HfO\(_2\) and its silicates have received significant attention as alternative gate dielectrics due to their thermodynamic stability on Si and their large barrier heights [56, 59-61]. Lee et al. achieved an ultra thin HfO\(_2\) gate dielectric with 0.9 nm EOT and a Pt gate using an optimized reactive dc magnetron sputtering process [56]. It is found that the dielectric constant of HfO\(_2\) was ~28 and that the EOT was stable up to 700 °C. Lee et al. also reported a 1.04 nm EOT HfO\(_2\) with polysilicon gate without any barrier layer [59]. They reported that their HfO\(_2\) film remained high quality after high temperature dopant activation (950 °C for 30 seconds) and had very low leakage current (0.23 mA/cm\(^2\) at V\(_g\) = 1 V). Wilk et al. reported an EOT less than 1.8 nm for a 5 nm Hf\(_6\)Si\(_{25}\)O\(_{65}\) film, which yields a dielectric constant ~11 [60]. It is not conclusive, however, that HfO\(_2\) or Hf silicates will be the
best candidate as the gate insulator for 65 nm gate length CMOS. Researchers are still continuously searching for a material which can fulfill all the gate dielectric requirements.

1.3.3 Additional Problems with High $\kappa$ Dielectrics

Challenges arise from the processing and integration of high-$\kappa$ dielectrics, especially during those steps performed at high temperature. The intrinsic limitation of high-$\kappa$ dielectrics is that a SiO$_2$ rich interfacial layer will be formed during the deposition of high-$\kappa$ on Si, which makes it very difficult to scale the EOT. Recently, reports of nitrogen based materials and nitrogen annealing have shown the decrease or elimination of the interfacial layer [62-65]. Thermal stability is another key requirement for high-$\kappa$ dielectrics since junction activation is always performed at a high temperature in conventional CMOS processing. The possible formation of silicates between the dielectric and Si substrate as well as other probable reactions between high-$\kappa$ dielectric with metal gates will inevitably increase the EOT and have other adverse impacts on the channel. High temperature processes can also change the film morphology of high-$\kappa$ dielectrics, resulting in crystalline or polycrystalline materials and thus increased leakage.

A lot of attention has been placed on the mechanisms of mobility degradation in high-$\kappa$ devices. NMOS device mobility degradation is a more significant problem than that of PMOS degradation. Fischetti et al. reported that the high permittivity of materials like HfO$_2$ and ZrO$_2$ result in the presence of a large part of soft optical phonons providing a long-range scattering of electrons in the Si inversion layer [66]. Thus the electron mobility could be reduced by as much as a factor of 3 due to this
unavoidable scattering mechanism. Remote charge scattering may be another component limiting the carrier mobility in high-κ devices [67]. An in depth study to understand these additional scattering mechanisms in high-κ gate stacks is necessary to minimize the mobility degradation.

1.4 Metal Gate Electrodes

1.4.1 Why Do We Need Metal Gates

There are several limitations of the present polysilicon gate for submicron CMOS technology [68]. Modern CMOS processing uses n+ polysilicon gates for NMOS and p+ polysilicon gates for PMOS, which is normally achieved by ion implantation and subsequent annealing. During the doping process, dopant penetration into the thin gate dielectric may occur, especially for boron, and can shift the device threshold voltage. For deep submicron devices, ultra shallow junctions are required. Therefore, both the implantation energy and the dopant activation temperature are reduced, resulting in a low active dopant concentration in the externally doped polysilicon. The doping level of polysilicon can also be reduced due to dopant segregation during silicidation and dopant evaporation during activation annealing [69]. A polysilicon depletion layer is formed at the polysilicon/gate oxide interface and can decrease both the drive current and transconductance of the transistor [69]. The poly-Si depletion becomes more severe as the effective dielectric thickness is reduced. Also, the sheet resistance of the polysilicon gate can be high because of the scaling of polysilicon thickness which then limits the MOSFET circuit speed. Replacing gate electrodes with metal gates will be able to eliminate these issues.
Furthermore, Hobbs et al. have reported that Fermi level pinning problems occur when polysilicon gate electrodes are deposited on high-$\kappa$ dielectrics [70]. For Hf based dielectrics, the interfacial Si-Hf bonds create dipoles which pin the Fermi level right below the polysilicon conduction band, hence increasing the threshold voltage. Therefore, metal gate electrodes are required for alternative gate dielectrics.

### 1.4.2 Current Candidates: Advantages and Problems

Candidates for new metal electrodes are required to have good thermal and chemical stability, as well as process compatibility with current CMOS technology and future high-$\kappa$ dielectrics. A desirable metal gate should have an appropriate work function for NMOS or PMOS devices. Therefore, the work function needs to be within 0.2 eV of the conduction and valence band edges of Si [71]. In addition, a low diffusivity to oxygen and other dopants of the metal gate is necessary. Dual metal gates or midgap metal gate electrodes can be used in CMOS processing. There has been research on fabricating CMOS with a single midgap metal gate to simplify the process. However, due to the tradeoff of low channel doping requested by low threshold voltage and control of short channel effects, the mid-gap metal gate is not suitable for submicron devices [72].

Work functions of different metals have been measured by evaluating the flat band voltages on SiO$_2$, ZrO$_2$ and ZrSiO$_4$ [71]. It was found that the work functions of Al, Ta, Mo, Ti, Hf, and Zr were near the conduction band of Si, hence they could be potential candidates for NMOS. On the other hand, the work functions of Pt, Ru, Rh, Co, Pb, and RuO$_2$ were near the valance band, so these metals appear to be suitable for PMOS as gate electrodes. There are also conducting metal nitrides such
as WNₙ, TiNₓ, TaNₓ, TaSiₓNᵧ, which may be able to act as good metal electrodes. Elemental metals with lower work functions were found to have problems with stability due to high free energy of formation [62], while high work function metals provide better stability, but may have adhesion issues.

1.5 Advanced Gate Stacks on Strained Silicon and Current Challenges

1.5.1 Strained Silicon with Novel Gate Stacks

K. Rim et al. first reported the experimental data on the integration of biaxial strained Si and high-κ gate dielectrics [20]. The strained Si nMOSFETs with HfO₂ exhibited 60 % higher mobility than the unstrained Si device with HfO₂ and 30 % higher mobility than the conventional unstrained Si with SiO₂ when compared with the universal MOSFET mobility model [26]. At a given Eeff, the gate leakage current of the HfO₂/strained Si device is significantly reduced while the mobility is enhanced compared to the conventional SiO₂/unstrained Si device.

Datta et al. demonstrated the integration of strained Si NMOS with HfO₂ and TiN metal gate showing enhanced electron mobility at 1 MV/cm with an ultra-thin EOT of 1.4 nm without any intentional SiON interfacial buffer layer [73]. Both the mobility degradation due to additional phonon scattering at the HfO₂/Si interface and the screening of the remote phonon-electron interaction by metal gate electrodes were investigated experimentally.

Strained Si NMOS devices with a NiSi FUSI gate were presented by Xiang et al. for the first time [74]. NiSi gate electrode is Ni rich although it is a stoichiometric monosilicide close to the gate dielectric. Strained Si was grown on relaxed SiGe and the NiSi metal gate was achieved by full silicidation of the polysilicon gate. Further
enhancement was shown without any degradation in gate oxide integrity. Krivokapic et al. demonstrated the strained Si devices by using ultra thin fully depleted SOI with NiSi metal gates and mesa isolation [75]. From the electrical results, it was concluded that the channel was under uniaxial compressive stress for wider devices and tensile stress for narrower devices. About 30% performance improvement was accomplished in a 75 nm device due to the process induced strain.

1.5.2 Current issues: Process and Device Design

Several issues arise during the fabrication of biaxial strained Si devices. Thermal stability is the fundamental requirement for strained Si devices since only the preservation of strain can make any performance improvement possible. Ge diffusion can be enhanced by strain [76]. Boron diffusion is retarded in strained Si while arsenic diffusion is enhanced significantly [77]. Other issues related with integration include the potential strain relaxation during sidewall spacer formation due to the free boundary formed by reactive ion etching (RIE) [78], as well as the threshold voltage ($V_{th}$) adjustment because of the reduction of $V_{th}$ in strained Si on SiGe. For uniaxial strained Si devices, the main consideration is the strain variations resulting from the processing. Stress in the channel can be affected by many factors such as the height of polysilicon gate, the sidewall width, and the gate length [79]. It is also found that the mobility enhancement factor of uniaxial strained Si devices is dependent on the substrate dopant concentration ($N_{sub}$): the higher $N_{sub}$, the less mobility enhancement [80].

There are many important issues that need to be understood before the new gate stack materials can be used in MOSFETs with strained silicon channels. These
include the interfacial layer formation at the strained Si/high-κ dielectric interface and the effect of metal gate electrodes on the channel strain. Recently reported strained Si MOSFETs with advanced gate stacks have shown good performance [20, 73, 74]. In addition, these studies have shown that the mobility degradation commonly observed with high-κ dielectrics can be partially compensated by employing a strained Si channel. However, there are additional scattering mechanisms limiting the mobility enhancement in this system including increased phonon scattering attributed to HfO2 [73]. It has been shown that additional Coulomb scattering due to high Dit reduces the hole mobility in p-channel MOSFETs and enhanced phonon scattering can decrease the electron mobility in n-channel MOSFETs, both due to the presence of Ge atoms in the channel [81]. Therefore, it is necessary to fully understand the impact of Ge on the properties of MOSFETs with strained Si channels, when there is a relaxed Si1-xGex layer in close proximity, which is the focus of this work.

1.6 Outline of the Dissertation

This research focuses on the fabrication and characterization of strained Si devices with alternative gate stacks. Chapter 2 provides a description of the experiments used to fabricate the devices investigated in this work. Chapter 3 presents a detailed introduction to materials analysis and electrical characterization of advanced strained Si. In chapter 4, both material and electrical characteristics of epitaxial SiGe and strained Si films as well as the high-κ dielectric (HfO2) are evaluated. Chapter 5 discusses the electrical properties of strained Si MOS capacitors while analysis of transistor characteristics, including strained Si
MOSFETs fabricated with SiO$_2$, HfO$_2$ and/or metal gate electrodes, is given in Chapter 6. Chapter 7 concludes the research and provides directions of future work in this area.
1.7 References


[28] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band k center dot p calculation of the hole mobility in silicon inversion layers:


Chapter 2 Fabrication of Strained Silicon Devices Incorporating Alternative Gate Stacks

In this chapter, the process flow used in fabrication of strained silicon MOSFETs will be presented. Process parameters of critical process steps including epitaxy of strained silicon, high-κ dielectric formation, and deposition of metal gate electrodes will be provided, as well as the description of deposition systems used in this work.

2.1 Epitaxy of Si$_{1-x}$Ge$_x$ and Strained Si Layer

In-situ boron doped Si and Si$_{1-x}$Ge$_x$ films were selectively deposited in an ultra-high vacuum rapid thermal chemical vapor deposition (UHV-RTCVD) system designed and built at NCSU. The construction of this system was carried out by Nemanja Pesovic and Inkuk Kang, who were both former graduate students in Dr. Ozturk's research group.

2.1.1 UHV-RTCVD System

This UHV-RTCVD system is capable of processing 4", 6", and 8" wafers. As shown in Figure 2-1, the system consists of three chambers: a sample entrance chamber (SEC), an intermediate chamber (IC) and a main process chamber (MPC). The SEC is pumped by a dry molecular drag pump which maintains a base pressure of 10$^{-5}$ Torr. Both the IC and the MPC are pumped by cryogenic pumps with the base pressure as low as 10$^{-9}$ Torr, and they can be baked up to 100°C to desorb the water from the chamber walls. The IC is designed to minimize the contamination of MPC during sample transport. During processing, the MPC is pumped by a turbo-
molecular/molecular-drag combination pump, which can support process pressures up to 3 Torr. This pump is backed up by a dry mechanical pump. All pumps used on the system are oil free pumps and hence hydrocarbon contamination can be minimized. The MPC is a stainless steel chamber with a double-wall structure. A quartz bell jar is placed on top of the MPC sealed by two O-rings which are differentially pumped by a small molecular-drag pump. A pressure of $10^{-6}$ Torr is maintained between these two O-rings.

Figure 2-1 A schematic illustration of UHV-RTCVD system used in this work.

During deposition, a single wafer is placed on a quartz wafer holder and then raised to the desired height. Two tungsten-halogen lamp banks are used to heat the wafer. One of these lamp banks is placed on top of the bell jar consisting of eighteen 2 kW lamps arranged in two layers while the other lamp bank consists of sixteen 1 kW lamps placed around the quartz jar. Two optical pyrometers ($\lambda = 4.9 \, \mu m$) are focused to the edge and center of the backside of the wafer to monitor the
temperature. The pyrometer focused at the wafer center is used in a closed loop feedback control system to control the wafer heating. The gases used in this work for selective epitaxy of undoped or boron doped Si or Si$_{1-x}$Ge$_x$ films included Si$_2$H$_6$, GeH$_4$ (10% diluted in H$_2$), B$_2$H$_6$ (3% diluted in H$_2$) and H$_2$. All gases were UHP grade as defined by the gas supplier Voltaix Inc. A schematic illustration of the MPC is shown in Figure 2-2.

![Figure 2-2](image)

Figure 2-2 The main process chamber (MPC) of the UHV-RTCVD system consists of a quartz bell jar, a top lamp bank, and a side lamp bank.

2.1.2 Surface Preparation Prior to Si$_{1-x}$Ge$_x$ and Strained Si Deposition

Before deposition, the wafer clean process consists of two steps: an ex-situ clean followed by an in-situ clean in the growth chamber. The ex-situ clean includes a standard RCA clean with SC1 (NH$_4$OH+H$_2$O$_2$+5H$_2$O) and SC2 (HCl+H$_2$O$_2$+5H$_2$O) followed by a 30 second dip in 1% HF solution. The HF solution removes the
chemical oxide grown during the RCA clean and passivates the dangling bonds on the Si surface with hydrogen. If the wafer is rinsed in deionized water for a very short time (e.g. 15 seconds), the hydrogen passivation is partially lost and oxide islands can be formed on the wafer surface. However, the oxide can be easily desorbed during a low-thermal budget in-situ bake since its coverage is less than a monolayer. After the HF dip, the wafer is dried with N₂ and loaded into the SEC immediately. It typically takes about 10 minutes for the SEC to reach a pressure of 10⁻⁵ Torr. Then, the wafer cassette, which can hold four wafers, is transferred to the IC. After reaching low 10⁻⁸ Torr in this chamber, one of the wafers from the cassette is transferred into the MPC, which is pumped by a cryopump. After reaching the base pressure of 10⁻⁹ Torr, the chamber automatically switches to the turbo-molecular/molecular-drag combination pump.

The in-situ clean is carried out by heating the wafer to 800ºC for 10 - 15 seconds in vacuum and annealing the wafer at this temperature for 15 seconds to desorb the residual oxide from the ex-situ clean. Detailed discussion about this process can be found in publications of Sanganeria et al. and Celik et al. [1, 2] who formerly worked in this laboratory. Their studies showed that it is sufficient to reduce the oxygen to the concentration below the SIMS detection level and the carbon level below 10¹⁸ cm⁻³ by annealing the wafer at 800ºC for 10 seconds in vacuum. Epitaxial growth is initiated by gas switching (i.e. by switching on the gasses as soon as the process temperature is reached) and terminated by switching off the lamps as well as the gas flows simultaneously. A typical deposition cycle is illustrated in Figure 2-3.
2.1.3 Selective Deposition of Boron Doped Si and Si$_{1-x}$Ge$_x$

Immediately after the in-situ clean, the wafer is ready for the deposition of Si or Si$_{1-x}$Ge$_x$ films. Typical process conditions for undoped and boron doped Si and Si$_{1-x}$Ge$_x$ epitaxy used in this work are listed in Table 2-1. The deposition time was limited to one minute at temperatures higher than 750 °C to avoid excessive quartz heating which would eventually result in deposition on the bell jar. Discussion on the selective deposition regarding different deposition conditions and the quality of the epitaxial films will be included in Chapter 4.

Table 2-1 Typical process conditions for Si and Si$_{1-x}$Ge$_x$ deposition used in this work.

<table>
<thead>
<tr>
<th>Deposition Conditions</th>
<th>Undoped</th>
<th>Boron-Doped</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Si$_2$H$_6$</strong></td>
<td>2.5 sccm</td>
<td>0.03 sccm</td>
</tr>
<tr>
<td><strong>GeH$_4$</strong></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>B$_2$H$_6$</strong></td>
<td>--</td>
<td>0.03 sccm</td>
</tr>
<tr>
<td><strong>H$_2$</strong></td>
<td>675 sccm</td>
<td>--</td>
</tr>
<tr>
<td><strong>Temperature (°C)</strong></td>
<td>800</td>
<td>800</td>
</tr>
<tr>
<td><strong>Pressure (mTorr)</strong></td>
<td>~35</td>
<td>~35</td>
</tr>
</tbody>
</table>

Figure 2-3 A typical deposition sequence used in Si$_{1-x}$Ge$_x$ and Si epitaxy.
2.2 High κ Dielectric and Metal Gate Electrode Deposition

All the metal gate electrodes used in this work were deposited by RF sputtering. In addition, the formation of high-κ dielectrics was carried out in the sputtering tool followed by post deposition annealing (PDA).

2.2.1 UHV-Sputtering System

Sputtering is a physical deposition process. During sputtering, the momentum of the incident ions, which have high kinetic energy, is transferred to the surface atoms of the sputtering target, removing atoms from the surface of the target and landing on the wafer sitting underneath. As more atoms accumulated from the target, a thin film forms [3]. The most common method to provide incident ions is to introduce an inert gas flow into the chamber to a pressure of 1 - 100 mTorr and set a glow discharge to launch plasma.

The UHV RF sputtering system was designed and constructed in this laboratory, which can process both 4” and 6” wafers. Both a top view and a side view of the system are shown in Figure 2-4. The system consists of two stainless-steel chambers: a load lock and a main process (sputtering) chamber. A base pressure of low 10^{-8} Torr is maintained in the process chamber such that any potential contamination resulted from outgassing from the substrate or the chamber walls could be avoided. Three RF magnetron sputtering guns are equipped in the chamber so that three different targets can be employed simultaneously to deposit alloys or multiple film stacks.

A 10~15 minute pre-sputtering should be performed before the deposition on the wafers to remove any contamination on the target surface. Immediately prior to
loading, the wafers are dipped in a diluted 1% HF for 30 seconds, rinsed in deionized water, and finally blow-dried with N₂. After loading the wafer, the load lock is pumped down to mid 10⁻⁷ Torr. Then the wafer is transferred into the sputtering chamber. The turbo pump will pump this chamber until the pressure reaches low 10⁻⁸ Torr, which is the ideal base pressure to start sputtering. During sputtering, an argon flow of 40 sccm is used to introduce the plasma and the pressure is set at 5.5 mTorr. The wafer rotates smoothly to ensure a better film uniformity. The RF power used for sputtering ranges from 25 to 100 Watts depending on the target materials, the alloy composition, and the desired deposition rate. Experimental data in our study shows that lower sputtering power will result in less sputtering damage and will be discussed in following chapters.

Figure 2-4 Schematic of the UHV RF sputtering system: (a) top view; (b) side view.
2.2.2 High-κ Dielectric Formation

As discussed in Chapter 1, hafnium oxide (HfO$_2$) is one of the most promising alternative dielectrics due to its unique characteristics. It has a relatively high dielectric constant (κ~25) and shows a good stability on silicon [4]. It is resistive to impurity diffusion and interfacial intermixing and it can form a barrier high enough to prevent tunneling. The HfO$_2$ films used in some strained Si MOS capacitors were deposited at the University of Texas at Austin, TX. It was formed by Hf reactive sputtering in a DC magnetron sputtering system and followed by a post-deposition anneal in N$_2$ at 500°C for 5 minutes in a furnace [5]. The high-κ gate dielectrics of other strained Si MOS capacitors and MOSFETs were deposited in our laboratory. A thin layer of Hf was sputtered in the UHV RF sputtering system and then annealed in the Tylan furnace at 500°C for 5 minutes. Before the temperature was ramped, the furnace was purged with N$_2$ at 7200 sccm for 10 minutes. The target equivalent oxide thickness was about 2 nm.

Due to the nature of the process condition of HfO$_2$ formation adopted in this work, an interfacial layer will develop intentionally or unintentionally underneath the high-κ dielectric (HfO$_2$). It may be a silicate which can result from the reactive sputtering process [6-8]. It has been reported that surface nitridation can be used to minimize this interfacial layer in order to achieve low EOT as well as to improve the device performance [9-13]. Additional interfacial growth may result from the transportation of some samples (MOS capacitors) from TX to NC or the ex-situ furnace annealing after sputtering. More discussions of the quality of HfO$_2$ film and the corresponding impact on electrical performance will be presented in Chapter 4.
2.2.3 Metal Gate Electrodes used in this work

Several metal gate candidates are being pursued to satisfy the workfunction requirement for NMOS and PMOS devices. Based on the previous studies in this group, both tantalum nitride (TaN) and ruthenium tantalum (RuTa) alloys are chosen as the metal gate electrodes for the capacitors and transistors. Typically, a tungsten (W) capping layer is sputtered directly after metal gate deposition in order to decrease the probing resistance. It also serves as an additional barrier layer to block the implantation damage on the underlying dielectric in transistor fabrication. The common thickness for the gate electrode is 40~50 nm while the W capping layer is usually 50 nm. The sputtering conditions and sputtering rate for the metal gates used in this work can be found in Table 2-2.

Ru has a workfunction near the valence band which makes it a good candidate for p-channel devices. Ta has a workfunction near the conduction band of silicon so it is appropriate for n-channel devices. It has been reported that the RuTa alloy has a workfunction near the conduction band, which is appropriate for n-channel devices [14]. RuTa alloys can be deposited by co-sputtering both Ru and Ta targets simultaneously [14, 15]. The composition of RuTa alloys can be controlled by varying the sputtering power of each gun which will result in changes of alloy composition and hence a change in workfunction. As the ratio of the sputtering power of Ta to the power of Ru increases, the concentration of Ta increases but the value of workfunction decreases. The workfunctions of pure Ru and Ta are 5.1 eV and 4.2 eV, respectively. Both Ru$_{50}$Ta$_{60}$ and Ru$_{90}$Ta$_{10}$ were used as metal gate electrodes which have been found to have workfunctions of 5.1eV and 4.3eV,
respectively on SiO$_2$. This shows that the RuTa alloys can serve as metal gates for both n-channel and p-channel devices only by changing the composition. Similar workfunction differences were shown when Ru$_{50}$Ta$_{50}$ and Ru$_{90}$Ta$_{10}$ were deposited on HfO$_2$.

Tantalum nitride (TaN) is a promising material to be used as a gate electrode of n-channel devices. The interfacial layer formation can be effectively deterred by incorporating nitrogen into the Ta film to provide an excellent barrier to oxygen diffusion. However, the incorporation of nitrogen also results in the increase the workfunction [16]. Heuss et al. reported a workfunction of 4.5 eV achieved by depositing TaN on SiO$_2$ using 5% N$_2$ flow rate. These values were extracted by both C-V and barrier height analysis [16].

<table>
<thead>
<tr>
<th>Metal Gate</th>
<th>Metal Gate Conditions</th>
<th>Sputtering Rate (nm/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaN</td>
<td>Ta 100 W, N$_2$ 2 sccm</td>
<td>2</td>
</tr>
<tr>
<td>Ru$<em>{50}$Ta$</em>{50}$</td>
<td>Ru 50W, Ta 50W</td>
<td>1.67</td>
</tr>
<tr>
<td>Ru$<em>{90}$Ta$</em>{10}$</td>
<td>Ru 90W, Ta 10W</td>
<td>1.67</td>
</tr>
<tr>
<td>W</td>
<td>100W</td>
<td>1.67</td>
</tr>
</tbody>
</table>

2.3 Process Flow of Strained Si MOSFETs

The standard MOSFET process flow modified for strained silicon devices is provided in Table 2-3. The GEM mask set used in the fabrication of these MOSFETs was designed by Heather Lazar and Qian Zhao in Dr. Misra’s group. A manual of this GEM mask can be found in H. Lazar’s thesis [17]. The highest temperature processing step used in the fabrication for all these devices was the dopant activation after ion implantation, which was carried out at 950°C for 30 seconds. The metal gate electrodes were patterned by wet etching. A standard forming gas anneal
(FGA) at 400°C for 30 minutes was performed prior to the last step, i.e. the contact metal evaporation.

Table 2-3 Summary of strained Silicon MOSFET process flow.

<table>
<thead>
<tr>
<th>Step</th>
<th>Process Description</th>
<th>Important Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wafer scribing</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>RCA clean</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Field oxidation</td>
<td>~300 nm</td>
</tr>
<tr>
<td>4</td>
<td>JTB clean</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Pre-coat bake</td>
<td>115°C, 5 min</td>
</tr>
<tr>
<td>6</td>
<td>Photolithography</td>
<td>GEM Active Mask</td>
</tr>
<tr>
<td>7</td>
<td>Descum</td>
<td>3 min</td>
</tr>
<tr>
<td>8</td>
<td>Field oxide etch</td>
<td>BOE, 10sec overetch</td>
</tr>
<tr>
<td>9</td>
<td>Strip Resist</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>RCA clean</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Epitaxy of Si₁₋ₓGeₓ and Si film</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>JTB clean*</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Gate dielectric formation</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Metal gate or polysilicon gate deposition</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Pre-coat bake</td>
<td>115°C, 5 min</td>
</tr>
<tr>
<td>16</td>
<td>Photolithography</td>
<td>GEM Poly Mask</td>
</tr>
<tr>
<td>17</td>
<td>Metal gate or polysilicon etch</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Strip Resist</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Low temperature oxide deposition</td>
<td><del>10</del>15 nm</td>
</tr>
<tr>
<td>20</td>
<td>Source/Drain implantation</td>
<td>As, 5×10¹⁵ cm⁻², 40keV</td>
</tr>
<tr>
<td>21</td>
<td>Rapid thermal anneal for S/D activation</td>
<td>950°C, 30 sec</td>
</tr>
<tr>
<td>22</td>
<td>Low temperature oxide deposition</td>
<td>~300 nm</td>
</tr>
<tr>
<td>23</td>
<td>Pre-coat bake</td>
<td>115°C, 5 min</td>
</tr>
<tr>
<td>24</td>
<td>Photolithography</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Descum</td>
<td>3 min</td>
</tr>
<tr>
<td>26</td>
<td>Oxide contact hole etch</td>
<td>BOE</td>
</tr>
<tr>
<td>27</td>
<td>Strip Resist</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Forming gas anneal</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Pre-coat bake</td>
<td>115°C, 5 min</td>
</tr>
<tr>
<td>30</td>
<td>Photolithography</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Descum</td>
<td>3 min</td>
</tr>
<tr>
<td>32</td>
<td>Contact metal evaporation</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Contact metal lift off</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Backside oxide etch</td>
<td>BOE swab</td>
</tr>
<tr>
<td>35</td>
<td>Backside metal evaporation</td>
<td>Recommended</td>
</tr>
</tbody>
</table>

Considering the potential strained Si consumption, a JTB clean is used instead of the standard RCA clean. For bulk Si MOSFETs, it is recommended to use RCA clean and grow sacrificial oxide before gate oxidation. Please refer to [17] for a standard MOSFET process flow.
2.4 References


Chapter 3 Material Analysis and Electrical Characterization

3.1 Material Analysis

Various characterization methods, including X-ray diffraction (XRD), Raman spectroscopy, X-ray Photoelectron Spectroscopy (XPS), atomic force microscopy (AFM), scanning electron microscopy (SEM) and transmission electron microscopy (TEM), were employed to study different aspects of the film quality, such as alloy composition, lattice strain and the surface morphology.

The crystalline quality and the composition of the samples were examined by XRD θ-2θ scan [1, 2]. This measurement provides about the crystallinity of the layers as well as the composition of Si_{1-x}Ge_{x} alloys which can be calculated from the position of the SiGe peak, as shown in Figure 3-1. However, since the position of the SiGe peak moves with the amount of strain in the film, the Ge concentration of a strained layer cannot be determined by XRD. In addition, if the expected film thickness is far less than the measurement penetration depth, e.g. 10nm, a very weak signal is detected which leads to a low signal-to-noise ratio. Therefore, XRD serves as a fundamental technique for studying the crystallographic orientation of epitaxial silicon and composition of relaxed Si_{1-x}Ge_{x} film in this work.
Raman spectroscopy is a non-destructive technique based on the study of inelastic light scattering [3]. When photons strike a crystal, they can be scattered either elastically, as in Raleigh scattering where the energy of the outgoing beam remains unchanged, or inelastically, which results in a shift in the energy of the outgoing photons (Raman signal) from the original incoming photons. These shifts in energy are quantized as phonons and can provide information about the material being studied. The collected data results in a plot of intensity vs. energy, where peaks correspond directly to discrete phonon modes in the crystal. This technique is non-contacting and non-destructive and enables analysis through transparent layers. Therefore, Raman spectroscopy can be used to characterize solid, liquid or gaseous samples. In the solid state, it is applicable to samples in bulk form, thin films, and microscopic structures (e.g. device structures or particulates). Typical applications
include identification of chemical species and bonding interactions, distinguishing crystalline, polycrystalline and amorphous phases, detection of the crystallographic orientation, lattice strain measurement, micro-contamination analysis, etc. In this work, Raman spectroscopy was used to study the strain level in the epitaxial silicon film on top of the relaxed Si$_{1-x}$Ge$_x$ virtual substrate.

The surface morphology can be studied by AFM and plan view of SEM. As-grown samples can be used directly with both methods. For most of the cases in this study, AFM was employed to analyze the deposition rate by step-height measurements and surface roughness comparison of the film and surrounding oxide to confirm selectivity. A facet can be observed at the interface of the epitaxial film and the oxide from both AFM and SEM images.

When the sample feature size decreases to less than 100nm, it is very difficult to capture the structural details by SEM. Therefore, TEM will be necessary [4, 5]. Both cross-sectional and plan-view samples were prepared by mechanical polishing, followed by Ar-ion milling. The crystallographic orientation can be studied in detail by selected area diffraction (SAD), bright-field (BF) and central dark-field imaging (CDF). Cross-sectional TEM (XTEM) and High-resolution transmission electron microscopy (HRTEM) imaging give a projected shape of devices/stacks and faceting, which makes it possible to precisely determine the lateral size and height of nanostructures. A plan-view TEM study provides information on the uniformity and the nanostructure lateral distribution, if applicable.

Secondary ion mass spectrometry (SIMS) was used to analyze the distribution of elements during SiGe and strained silicon deposition with in-situ boron
doping and following oxidation and rapid thermal annealing (RTA). From the depth profiles of Si, Ge, and B, the thickness of the strained silicon layer was determined and diffusion of B and Ge was studied.

X-ray photoelectron spectroscopy (XPS) was also employed for chemical analysis (ESCA) [6, 7]. XPS uses soft x-rays to illuminate the sample in an ultrahigh vacuum such that electrons will be knocked out of inner-shell orbitals. The kinetic energy of these photoelectrons is determined by the difference between the energy of the x-ray radiation and the electron binding energy. This gives a spectrum with a series of photoelectron peaks. The binding energy of each peak is different for each element. The composition of the material can be determined by evaluating the area under each peak provided appropriate sensitivity factors are known. Since the shape of each peak and the binding energy can be slightly altered by the chemical state of the emitting atom, XPS can be employed to detect the chemical bonding states. XPS is not sensitive to hydrogen or helium, but can detect all other elements. In our study, XPS was employed to investigate the composition and bonding in HfO$_2$ films and potential Ge diffusion in the strained Si layer.

3.2 Electrical Analysis

A variety of electrical characterizations were performed in order to obtain the electrical properties of MOS capacitors and MOSFETs studied in this work. These characterization techniques include capacitance-voltage, current-voltage and charge pumping measurements. Electrical parameters such as flatband voltage ($V_{FB}$), equivalent oxide thickness (EOT), hysteresis, mobility, threshold voltage ($V_T$), subthreshold slope, interface trap density, as well as others can be obtained from
these measurements. The principles and advantages are discussed briefly in the following sections.

3.2.1 Capacitance-Voltage Measurement

The capacitance-voltage (C-V) measurement is a fundamental but very important technique to assess the electrical properties of a dielectric film or a device. Capacitance can be defined as the change in the charge amount, $dQ$, due to a change in the voltage, $dV$:

$$ C = \frac{dQ}{dV} $$

(3.1)

This method is known as differential capacitance. This change in voltage is typically a small-signal ac voltage imposed on a dc sweep bias that is applied to the MOS device. C-V measurements can be performed at high and low frequencies where low-frequency measurements are often accomplished by using a quasi-static technique. Both techniques have similar accumulation and depletion attributes, but differ in the inversion regime, which is shown in Figure 3-2(a).

![Figure 3-2](image)

Figure 3-2  (a) The C-V characteristic of a MOS capacitor on P-type Si substrate. (b) The electrical equivalent circuit of a MOS capacitor.
The electrical equivalent circuit of a MOS capacitor or MOSFET is the series combination of two capacitances: the oxide capacitance $C_o$, which is independent on gate voltage, and the Si capacitance $C_s$, which is dependent on gate voltage and defined according to Eq. (3.1), as shown in Figure 3-2(b).

In the accumulation region, holes are accumulated at the surface such that the MOS capacitor behaves like a parallel-plate capacitor and the fixed oxide capacitance $C_o$ is dominant. At $V_{FB}$ there is no charge in the device. But if a small voltage is applied, there will be charge that appears at a Debye length away from the oxide. So the capacitance at $V_{FB}$ is a combination of $C_o$ and $C_s$. In reality, $C_s$ is also present during accumulation. However, when there are lots of carriers, Debye length decreases, so $C_s$ will increase to infinity, which would cause the accumulation capacitance to be equal to $C_o$ only. As the voltage becomes less negative, ionized acceptor atoms are accumulated underneath the oxide such that the semiconductor surface is depleted. Thus the total capacitance per unit area $C'$ is:

$$C' = \left( \frac{1}{C'_o} + \frac{1}{C'_s} \right)^{-1} = \left( \frac{x_o}{\varepsilon_o} + \frac{W}{\varepsilon_s} \right)^{-1}$$

(3.2)

where $x_o$ is the oxide thickness, $W$ is the width of the depletion layer, $\varepsilon_o$ and $\varepsilon_s$ are permittivities of oxide and semiconductor, respectively. After inversion is reached, $C_s$ depends on the measurement frequency. For a slow D.C. sweep with low frequency A.C. signal ($f < 1$ Hz), strong inversion layer can be achieved due to the generation and recombination of large amount of minority carriers. For a slow D.C. sweep with high frequency A.C. signal ($f > 1$ KHz), depletion region is modulated because the generation-recombination processes is too slow to keep up with the change in the D.C. gate bias. The capacitor goes into deep depletion and equation (3.2) continues...
to be valid. Eventually the generation process supplies carriers to the inversion layer and the capacitance returns to the normal high frequency value.

The difference in the metal-semiconductor work function can be determined from the shift in the C-V curve from the ideal curve. Negative $\Phi_{ms}$ shifts the C-V curve to the negative direction. Various types of charges will also affect the C-V characteristics. The fixed interface charge, $Q_F$, is caused by uncompensated silicon-silicon bonds and on the order of $10^{10}$ cm$^{-2}$. The oxide trapped-charge, $Q_{OT}$, is due to defects in the SiO$_2$ network and usually negligible in modern MOS devices. Both of them can be calculated from the shift of the C-V curves by comparing them with the ideal curves. The mobile charge, $Q_M$, is due to alkaline ions (e.g., Na$^+$) and usually with a very low concentration ($\sim 10^9$ cm$^{-2}$) in current technology. By bias-temperature stress, $Q_M$ can be moved from the SiO$_2$-Si interface to the metal-SiO$_2$ interface. The shift in $V_{FB}$ indicates the existence of mobile charge and the magnitude of the shift can be used to estimate $Q_M$. The interface trapped-charge, $Q_{IT}$, is attributed to unterminated Si bonds and these interface traps have energy levels distributed in the band gap. The charge state of the traps depends on the Fermi level. The methods of measuring interface trap density will be discussed extensively in the following section.

### 3.2.2 Interface Trap Density Measurement

A number of techniques have been developed over the years to measure the interface trap density ($D_{it}$) [8]. Only three methods will be discussed in this chapter, all of which were employed in this work. In this section, we will describe the $D_{it}$
measurements which can be performed on MOS capacitors. Charge pumping
method, which requires a MOSFET as the test structure, will be discussed later.

Interface traps cause a non-uniform shift because their charge state \( Q_{IT} \) changes with bias and position of the surface Fermi level. \( Q_{IT} \) can be estimated by comparing the difference between the high-frequency \((hf)\) C-V (when the traps cannot keep up with the high frequency A.C. signal) and the low-frequency \((lf)\) C-V curves. \( D_{it} \) is defined in terms of the measured \( lf \) and \( hf \) curves as:

\[
D_{it} = \frac{C_{ox}}{q} \left( \frac{C_{if} / C_{ox}}{1 - C_{if} / C_{ox}} - \frac{C_{hf} / C_{ox}}{1 - C_{hf} / C_{ox}} \right)
\]

(3.3)

where \( C_{ox} \) is the oxide capacitance and both \( C_{if} \) and \( C_{hf} \) are measured as functions of the gate voltage. This method can only give \( D_{it} \) over a limited range of the band gap, typically from the onset of inversion to about 0.2 eV from the majority carrier band edge, where the A.C. measurement frequency equals the inverse of the interface trap emission time constant. The range will be closer to the band edge if the frequency increases. Typical \( hf \) and \( lf \) C-V curves are shown in Figure 3-3.

![Figure 3-3](image)

Figure 3-3 High and low-frequency C-V curves show the offset \( \Delta C/C_{ox} \) due to interface traps.
The conductance method was first proposed by Nicollian and Goetzberger in 1967, which is now considered to be the most sensitive and complete method to determine $D_{it}$ [9]. The equivalent parallel conductance, $G_p$, is measured as a function of bias voltage and frequency by using a standard LCR meter. Since the change of $G_p$ is related to interface trap capture and emission of carriers, $D_{it}$ can be calculated even in the presence of leakage current [10]. The only disadvantage of this method is that it is very time-consuming. The equivalent circuits for conductance measurements are shown in Figure 3-4, as well as a short review of these circuits [11].
Figure 3-4    Equivalent circuits for conductance measurements [10-12]: (a) General circuit for a MOS capacitor in depletion or accumulation for a single interface state level: $C_x$ = series combination of oxide and gate capacitance, $G_t$ = tunnel conductance including the gate to conduction band, semiconductor generation-recombination, gate to interface state and gate to valence band conductances, $C_m$ = interface state capacitance, $C_d$ = depletion layer capacitance, $C_{acc}$ = accumulation layer capacitance, $R_{pt}$ = majority carrier interface trap resistance, $R_s$ = series resistance associated with bulk, contacts and tabs. (b) Circuit (a) transformed for a MOS capacitor in depletion or accumulation for a continuum of interface states. (c) Circuit (b) transformed to show capacitance corrected for series resistance ($C_c$) and the a.c. conductance ($G_{ac}$): (d) Circuit (c) transformed to show $C_c$ and conductance corrected for series resistance ($G_c$): (e) Circuit (d) transformed to show measured capacitance ($C_m$) and measured conductance ($G_m$).

The following relationships can be determined by comparing these circuits in Figure 3-4 (d) and (e),

$$C_m = \frac{C_c}{(G_t R_s + 1)^2 + \omega^2 C_c^2 R_s^2}$$  \hspace{1cm} (3.4)
\[ G_m = \frac{G_c (G_c R_s + 1) + \omega^2 C_c^2 R_s}{(G_c R_s + 1)^2 + \omega^2 C_c^2 R_s^2} \] (3.5)

\[ C_c = \frac{C_m}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2} \] (3.6)

\[ G_c = \frac{\omega^2 C_m C_c R_s - G_m}{G_m R_s - 1} \] (3.7)

where, \( C_m \) is the measured capacitance, \( G_m \) is the measured conductance, \( C_c \) is the capacitance corrected for series resistance \( R_s \), \( G_c \) is the conductance corrected for series resistance \( R_s \), and \( \omega = 2\pi f \) where \( f \) is the measurement frequency. The units of conductance are S/cm\(^2\) in these equations. Typical experimental \( G_p/\omega \) versus \( \omega \) curves are shown in Figure 3-5 [13]. An approximate expression of interface trap density in terms of the maximum of measured \( G_p/\omega \) is:

\[ D_u \approx \frac{2.5}{q} \left( \frac{G_p}{\omega} \right)_{\text{max}} \] (3.8)

and

\[ \frac{G_p}{\omega} = \frac{\omega G_c C_{ox}^2}{G_c^2 + \omega^2 (C_{ox} - C_c)^2} \] (3.9)

where \( G_c \) and \( C_c \) are calculated from Eq.(3.6) and (3.7). The series resistance \( (R_s) \) could be easily calculated based on material parameters [10] and the insulator capacitance \( (C_{ox}) \) was assumed to be equal to the capacitance measured in accumulation. In our study, \( C_{ox} \) was calculated based on the equivalent oxide thickness obtained from Hauser CVC simulation which includes the effect of quantum-mechanical confinement.

For MOS capacitors with ultra-thin and alternative gate dielectrics, the gate leakage current could be high. The conductance technique can still provide detailed
interface state properties of such MOS capacitors. However, there are additional limitations and potential errors under this circumstance. The effect of error of $C_{ox}$ is negligible on the extracted interface state density. The effect of errors in $R_s$ increases with increasing bias towards accumulation. The higher $R_s$, the lower sensitivity to changes in $D_{it}$, especially for those located near the majority band edge. Increasing the tunneling current reduces the sensitivity to changes in $D_{it}$ especially for interface states located nearer midgap. For extremely large gate currents, the interface states are no longer in equilibrium with the substrate, and the theory commonly used to extract interface state density from conductance might not be valid any more.

![Graph](image)

**Figure 3-5** $G_p/\omega$ versus $\omega$ for a single level, a continuum and experimental data. For all curves: $D_{it}=1.9\times10^9$ cm$^{-2}$ eV$^{-1}$, $\tau_{it}=7\times10^{-5}$ s.

### 3.2.3 Current-Voltage Measurement

Current-voltage (I-V) measurements are usually performed in order to assess off-state leakage current and device performance. In this work, the currents from the gate, source, drain and substrate can be recorded simultaneously by using an
HP4155b parametric analyzer. Typical measurements include gate current versus gate voltage ($I_g-V_g$), drain current versus gate voltage ($I_d-V_g$), and drain current versus drain bias ($I_d-V_d$).

In today’s low-power applications, one of the primary challenges is to reduce the gate leakage current ($I_g$) while improving device performance. This is typically monitored by measuring the leakage current density ($J_g$) of MOS structures in accumulation at 1V beyond the flatband voltage ($V_{FB}$). Quantum mechanical approaches were used in current transport models to demonstrate leakage mechanisms related to the phenomenon that as gate dielectric thickness is reduced to increase device drive, the gate leakage increases.

There are two dielectric current transport models widely used today: the direct tunneling (DT) model and the Fowler-Nordheim (FN) tunneling model. The primary difference between the two models is the tunnel barrier shape: a triangular barrier for FN tunneling and a trapezoidal barrier for DT, as shown in Figure 3-6 (a) and (b), respectively. Current density under the FN tunneling can be mathematically explained by [14, 15]:

$$I_{FN} = A_G A \xi_{ox}^2 \exp\left(\frac{-B}{\xi_{ox}}\right)$$  \hspace{1cm} (3.10)

where $A_G$ is the gate area, $\xi_{ox}$ is the electric field in dielectric, and A and B are constants dependent upon the barrier height at the Silicon-dielectric interface ($\Phi_B$), and the effective electron mass in the oxide ($m_{ox}$) as given below:

$$A = \frac{q^3 (m/m_{ox})}{8 \pi h \Phi_B} = 1.54 \times 10^{-6} \frac{(m/m_{ox})}{\Phi_B} \frac{A}{\sqrt{V}}$$  \hspace{1cm} (3.11)
where \( q \) is the electron charge, \( m \) is the free electron mass and \( h \) is Planck’s constant.

The direct tunneling current is given by the expression \([16]\):

\[
I_{DT} = A_G A \xi_{ox}^2 \exp\left(-B [1 - (1 - q V_{ox} / \Phi_B)^{1.5}] \right)
\]  

(3.13)

A simplified expression often used for direct tunneling is \([17]\):

\[
I_{DT} = \frac{A_G A}{t_{ox}^2} \exp\left(-4 \pi m_{ox} \frac{q^2}{h} \frac{\sqrt{2m_{ox} q}}{\Phi_B} \sqrt{\Phi_B - V_{ox}} \right)
\]  

(3.14)

where \( A \) is a constant, \( t_{ox} \) is the physical thickness of the SiO₂, \( m_{ox} \) is the effective mass of an electron in SiO₂, \( h \) is Planck’s constant, \( \Phi_B \) is the barrier height of SiO₂ to Si substrate, and \( V_{ox} \) is the voltage drop across the SiO₂ portion (as shown in Figure 3-6 (b)). The total gate leakage current is the sum of \( I_{FN} \) and \( I_{DT} \) with the higher current dominating. Experimental data shows that for \( t_{ox} \) thinner than 3.5nm, direct tunneling is dominant for low gate voltages \([13]\).
Figure 3-6  Energy band diagram description of (a) Fowler-Nordheim injection through a triangular barrier and (b) direct tunneling through a trapezoidal barrier.

In MOSFET devices, $I_d-V_d$ and $I_d-V_g$ measurements provide significant information about the device performance. $I_d-V_d$ measurement is done by measuring the drain current as the drain voltage is swept typically from 0 V to an inversion operating bias ($V_{DD}$). The saturation current ($I_{dsat}$) can be obtained which is used to investigate the operational device performance. The $I_d-V_g$ characteristic can be used to explain two regimes of MOSFET operation: the linear regime and saturation regime. In the linear region, the MOSFET channel region functions like a resistor and the drain current is proportional to the drain voltage. This channel resistance is controlled by the applied gate bias. Once the drain bias is greater than the difference of gate bias and the threshold voltage ($V_g-V_t$), the drain current saturates. Two mathematical expressions are used to explain each region of operation:

\[
I_d = \frac{W}{L} \mu_{eff} C_{ox} (V_g - V_t) V_D \quad \text{for the linear regime} \tag{3.15}
\]

and

\[
I_d = \frac{W}{L} \mu_{eff} C_{ox} (V_g - V_t)^2 \quad \text{for the saturation regime} \tag{3.16}
\]
The threshold voltage, $V_t$, is defined as:

$$V_t = V_{FB} + \Phi_b + \frac{q2\varepsilon_S N_A \Phi_b}{C_{ox}}$$

(3.17)

and

$$\Phi_b \cong 2 \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right)$$

(3.18)

where $N_A$ is the acceptor doping density and $n_i$ is the intrinsic carrier concentration.

As shown in Eq. (3.17), $V_t$ can be affected by changes in the flatband voltage caused by charges.

### 3.2.4 Mobility Measurement

The carrier mobility has significant impact on the device performance since both the carrier velocity and the device current depend on the mobility. Higher mobility materials will be able to have a higher frequency response as well as a higher current. There are several types of mobilities being studied, including the conductivity mobility, Hall mobility and magnetoresistance mobility, which are all bulk mobilities. In this case, the carrier mobility is mainly determined by lattice or phonon scattering and ionized impurity scattering. At very low temperature, neutral impurity scattering becomes more important. For some semiconductors, there is additional piezoelectric scattering. In our study, MOSFET mobility influenced by additional scattering mechanisms is investigated. In a MOSFET, all current carriers are confined within a narrow inversion layer. Thus, in addition to the phonon scattering and impurity scattering, Coulomb scattering and surface roughness scattering also reduce the momentum of the electrons and hence, reduce its mobility. At lower transverse fields, the Coulomb scattering introduced by the impurities in the channel, the oxide charges and interface states is dominant and is regarded as one of the
main reasons for the low mobility of high-κ dielectrics [18]. The surface roughness scattering mechanism dominates at high transverse fields. All these scattering mechanisms contribute to decreasing the mobility. According to Mathiessen’s rule, the net mobility $\mu$ is defined as [19]

$$\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \cdots$$

(3.19)

and the lowest mobility dominates, as shown in Figure 3-7.

![Figure 3-7](image)

**Figure 3-7** Plot of overall mobility versus effective field limited by various scattering mechanisms.

### 3.2.4.1 Split C-V Method

The effective mobility $\mu_{\text{eff}}$ can be obtained by measuring the drain current in the linear region [20]:

$$\mu_{\text{eff}} = \frac{L}{W} \cdot \frac{I_d(V_g)}{V_d Q_{\text{inv}}(V_g)} = \frac{g_d L}{W Q_{\text{inv}}}$$

(3.20)

where $W$ and $L$ are the precise channel width and length of the device, respectively, and $Q_{\text{inv}}$ is the inversion charge density, which is a function of the gate bias. The drain conductance $g_d$ is defined as
In the past, for a MOSFET with a relatively thick gate dielectric many people used the following relationship to obtain $Q_{\text{inv}}$ in strong inversion: 

$$Q_{\text{inv}} = C_{\text{ox}}(V_g - V_t),$$

where $C_{\text{ox}}$ is the oxide capacitance and $V_t$ is the threshold voltage of the device. However, for MOSFETs with thin gate oxides this is a poor approximation, thus the split capacitance–voltage (C–V) technique was introduced in the early 1980s [21] to extract the inversion charge density more accurately by measuring the gate-channel capacitance as a function of gate voltage $V_{gs}$:

$$Q_{\text{inv}} = \int_{-\infty}^{V_{gs}} C_{gc} dV_{gs}$$

(3.22)

where $C_{gc}$ is the gate to channel capacitance. $C_{gc}$ is measured by connecting the capacitance meter between the gate and the source-drain with the substrate grounded, as seen in the schematic of Figure 3-8 (a). When $V_{gs} < V_t$, the measured capacitance is equal to the total overlap capacitance of the gate to the source and drain, $2C_{ov}$. The capacitance increases as the channel starts to invert. As the voltage is increased further, the capacitance saturates to some inversion value which is a combination of the overlap capacitance and the channel capacitance, $2C_{ov} + C_{ch}$. $C_{gc}$ is achieved by subtracting $2C_{ov}$ from the measured capacitance and $Q_{\text{inv}}$ is given as a function of $V_{gs}$ by integrating this $C_{gc}$-$V_{gs}$ curve.

The gate to substrate capacitance, $C_{gb}$, can be measured by connecting the capacitance meter between the gate and the substrate with the source and drain tied together and grounded. The equipment setup for measuring $C_{gb}$ and total gate capacitance can be seen in Figure 3-8 (b) and (c), respectively. The depletion
charge, \( Q_d \), can be acquired by integrating \( C_{gb} \). Typical \( C_{gc} \) and \( C_{gb} \) curves are shown in Figure 3-8 (d). The drain conductance, \( g_d \), can be determined from Eq. (3.21) by measuring the \( I_d-V_g \) current characteristics using a single low drain bias, usually 10–50 mV. The drain bias has to be as low as possible to make the error in \( \mu_{\text{eff}} \) negligible, since \( C_{gc} \) is measured with no drain bias and the inversion charge decreases as drain bias increases.

![Diagram](image)

**Figure 3-8** Schematics of experimental setup to measure (a) gate to channel capacitance for Split C-V, (b) gate to substrate capacitance, and (c) total gate capacitance; and (d) gate-to-channel and gate-to-substrate capacitance as a function of gate voltage.
3.2.4.2 Corrections of Mobility Extraction

Besides the split C-V method, effective mobility can also be extracted using the NCSU MOB2D model [22]. However, corrections are necessary considering the presence of resistances and leakage currents.

Series resistance ($R_s$) degrades the current-voltage behavior, as well as serious errors in capacitance-voltage measurements [8, 13]. $R_s$ can be calculated in terms of the measured conductance ($G_{ma}$) and capacitance ($C_{ma}$) in accumulation:

$$R_s = \frac{G_{ma}}{G_{ma} + \omega^2 C_{ma}^2},$$

where $\omega = 2\pi f$ and $f$ is the frequency. When parallel model is used in C-V measurement if $R_s$ is involved, both the capacitance and the conductance have to be corrected according to $R_s$

$$C_c = \frac{C_m}{(1 - G_m R_s)^2 + \omega^2 C_m^2 R_s^2},$$

and

$$G_c = \frac{G_m C_m R_s - G_m}{G_m R_s - 1}$$

(3.23)  

(3.24)

Source-drain resistance ($R_{SD}$) also affects the mobility. Since $R_{SD}$ affects $I_d$ and $\mu_{eff}$ depends on the drain conductance $g_d$, it is obvious that $\mu_{eff}$ also depends on $R_{SD}$ if using $g_d$ to determine the mobility. In the presence of $R_{SD}$, $g_d$ becomes

$$g_d(R_{SD}) = \frac{g_d 0}{1 + g_d 0 R_{SD}},$$

(3.25)

where $g_d 0$ is the drain conductance for $R_{SD}=0$. It is clearly shown that if $R_{SD}$ is not negligible, $g_d$ will be reduced hence the mobility will be degraded.

As the gate dielectric physical thickness reduces, the gate leakage component becomes more significant which will affect the effective mobility extraction significantly. A physical based approach on correcting the $I_d - V_g$ curve in
the presence of a large gate leakage current was presented in [23]. The model assumes that the substrate current is negligible since the substrate current is usually small in an \( I_d - V_g \) measurement at typical operating regimes. In addition, the gate leakage current density is assumed to be constant over the channel region. As the gate leakage current increases, electrons that are supposed to participate in the drain current across the channel are now being lost to the gate leakage current, which results in a non-constant channel current. It is possible to employ a simple model shown in Figure 3-9 to correct the drain current. Only the gate leakage comes from the source and drain is illustrated because of the assumption of negligible substrate current. Due to the constant current assumption and a symmetric device structure, the gate leakage current from the source (\( I_{s,g} \)) equals that from the drain (\( I_{d,g} \)). In addition, the source and drain each provide one-half of the gate leakage. Thus \( I_{s,g} = I_{d,g} = I_g/2 \). The measured channel drain current (\( I_{dm} \)) can be expressed as: \( I_{dm} = (I_{d,ch} - I_g/2) \). Therefore, the leakage corrected channel drain current (\( I_{d,ch} \)) is:

\[
I_{d,ch} = I_{dm} + I_g/2
\]  

(3.26)

This corrected channel drain current will be used to extract mobility in Split C-V method or using the NCSU MOB2D model. Since the extraction algorithm uses non-linear least squares fitting to process experimental data, the inaccuracy in \( I_{dm} \) will result in a poor model fit which leads to errors in the determination of the effective mobility. Therefore, a corrected \( I_{d,ch} \) is required for mobility extraction especially for devices with high gate leakage current. Another way to correct for gate leakage is to use the true equation for \( g_d \) where two drain currents are measured at two different drain biases, subtracted, and divided by the difference of the drain biases [18, 22]:

\[ g_d = \frac{I_{d2} - I_{d1}}{V_{g2} - V_{g1}} \]
Figure 3-9 Schematic of an nMOSFET showing the current components in the $I_d-V_g$ measurement.

Other corrections may be necessary for devices with high interface trap densities and high leakage current to determine the inversion charge and the reasons for any degradation. The interface traps can respond to the AC modulation signal in the C-V capacitance measurement, which introduces an additional parallel capacitance [24] that can result in an overestimation of $Q_{inv}$. This error can be minimized by using a higher frequency [24], or corrected by subtracting out the effect of the interface traps if $D_{it}$ can be measured accurately [25]. On the other hand, the interface traps can follow the DC voltage sweep, so that a change in gate voltage results in changes not only in the inversion charge, but also in the charges trapped in interface [26]. Therefore a stretchout effect in C-V characteristics is observed, hence inversion charge will be overestimated. A simple method has been proposed to correct this error without having to measure the interface-trap density [18], which is similar to the NCSU MOB2D model. The correction for interfacial trapping can be done further if either the average $D_{it}$ or $D_{rt}$ as a function of surface band bending could be achieved. Detailed discussion on these corrections will be included in the
following chapters to investigate experimental data of advanced gate stack MOSFETs and the possible mechanisms responsible for the mobility degradation.

3.2.5 Charge Pumping

As discussed above, the quality of the interface can be quantified through some parameters such as $D_{it}$. The charge pumping method, developed by Brugler and Jesper in 1969 [27] is the most popular method to investigate the interface trapped charge in MOSFETs. There are various forms of charge pumping measurements, including two level, three level and low frequency, which are all based on similar principles. A periodic voltage pulse is applied to the MOSFET gate with sufficient amplitude for the device to be biased into inversion and accumulation. The substrate current, or charge pumping current ($I_{cp}$), is measured from which $D_{it}$ can be obtained.

3.2.5.1 Two Level Charge Pumping

In the conventional type of charge-pumping measurement, $I_{cp}$ is measured while the gate is biased under a string of voltage pulses of fixed amplitude, rise time, fall time, frequency, and duty cycle. The $I_{cp}$ is actually a recombination current as the device is driven back and forth between accumulation and inversion. As the transistor goes into inversion, interface traps are filled with electrons from the source and drain. As the device goes back into accumulation, untrapped electrons are released to the source and drain and those trapped electrons recombine with majority holes to create $I_{cp}$. When the gate voltage returns to the positive bias, the electrons flow into the interface again to be captured. Maximum exchange occurs when the device swings through $V_{fb}$ and $V_t$. 
As shown in Figure 3-10 (a), the MOSFET source and drain are tied to ground or with a small reverse bias ($V_{R}$) applied. In two level charge pumping, the gate bias pulse is a periodic a trapezoidal waveform with a finite rise and fall time. Over one entire period of the applied waveform, the net influx of majority carriers is measured as $I_{cp}$, which is proportional to the average density of interface traps at the semiconductor – dielectric interface. $I_{cp}$ can be calculated as:

$$I_{cp} = 2qD_i f A_G kT \left[ \ln \left( n_i \right) + \frac{V_{fb} - V_g}{\Delta V_g \sqrt{t_r t_f}} \right]^{\sigma_n \sigma_p}$$

(3.28)

where $q$ is the electron charge, $f$ is the frequency, $A_G$ is the gate area, $v_{th}$ is the thermal velocity of the carriers, $n_i$ is the intrinsic carrier density, $\sigma_n$ and $\sigma_p$ are the capture cross section coefficients for electrons and holes, $\Delta V_g$ is the peak to peak amplitude of the pulse, and $t_r$ and $t_f$ are the rise and fall times seen in Figure 3-10 (b).

![Figure 3-10](image)

**Figure 3-10** Schematic of (a) the charge pumping measurement configuration; (b) the square wave pulse used in two level charge pumping.

The two level charge pumping can be done by varying the base sweep from accumulation to inversion with constant amplitude (Base Sweep), or with a fixed base voltage in accumulation but variable amplitude sweep (Amplitude Sweep). In
the Base Sweep, $I_{cp}$ is monitored at each base voltage and can be plotted against base voltage ($I_{cp}$ vs. $V_{base}$). In the Amplitude Sweep, $I_{cp}$ is plotted vs. the magnitude of top voltage ($V_{top}$), which may indicate the frequency dependence. From the maximum $I_{cp}$ of these plots, average $D_{it}$ can be determined. Demonstration of Base Sweep and Amplitude Sweep can be seen in Figure 3-11. At a lower frequency, electrons have more time to inject further into the bulk trapping sites as well as release from those sites to contribute to $I_{cp}$, hence the ability to access trap sites deeper in the dielectrics increases such that a higher trap density will be detected. Therefore, the Amplitude Sweep technique can be used to quantify the bulk trapped charge ($N_t$) as:

$$N_t = I_{cp}/(qA_G f) \quad (3.29)$$

because trapped charge further into the gate stack can be electrically sensed [28].

![Schematic of the waveforms and resulting $I_{cp}$ of (a) the Base Sweep and (b) the Amplitude Sweep of two level charge pumping measurements.](image-url)
Another recombination process may also affect the measured $I_{cp}$. If $t_f$ is too short, the inversion layer charge will not have enough time to drift back to the source and drain when moving into accumulation. Some minority carriers will recombine with the incoming majority carriers resulting in an increased substrate current hence the average $D_{it}$ will be overestimated. This additional geometrical current is dependent on the device geometry and can be minimized by the use of small channel length and appropriately long rise and fall times.

### 3.2.5.2 Three Level Charge Pumping

The three level charge pumping was first introduced in 1987 [29] and has been studied continuously by a number of groups [30-32]. It can be used to determine the distribution of $D_{it}$ as a function of the bandgap ($D_{it}(E)$) [29]. The three level charge pumping is performed in the same configuration as two level charge pumping, but the waveform of the voltage pulse applied to the gate has three biasing levels. The waveform shown in Figure 3-12 is used to obtain the $D_{it}$ with energies between the intrinsic energy level of silicon ($E_i$) and the conduction band edge ($E_c$) for a n-channel device. For a p-channel device the waveform must be inverted. $V_{gbh}$ is larger than $V_{th}$ such that the device can be biased into strong inversion. Hence the interface traps are filled with minority carriers. $V_{gbl}$ is smaller than $V_{fb}$ which drives the transistor into strong accumulation. $V_{step}$ is a variable bias which drives the device between the midgap voltage ($V_{mg}$) and $V_{th}$. The midgap voltage is defined as the gate voltage which results in $q\psi_s=E_i$. When the device is pulsed to $V_{step}$ and $t_{step}$ is much longer than the emission time constant of the interface traps, traps above the Fermi level will emit the captured minority carriers and only those traps below the
Fermi level will be available to recombine with the incoming majority carriers when the device is pulsed back into accumulation. Therefore, I_{cp} indicates the number of traps filled with minority carriers at a certain V_{step}. Assuming the relation between V_{step} and ψ_{step} is known, D_{it} as a function of surface potential can be determined in terms of I_{cp} as seen in Eq. (3.30):

\[
D_{it}(q\psi_{step}) = \frac{1}{qfA_{eff}} \frac{dI_{cp}}{d(q\psi_{step})}
\]  

where A_{eff} is the effective gate area, f is the frequency of the pulse, and ψ_{step} is the surface band bending when the gate is at V_{step}. The frequency equals the inverse of the total period, T. One advantage of this method is that D_{it} can be determined without the knowledge of capture cross section coefficients. However, geometric effects still have to be minimized in three level charge pumping since the effects increase as V_{step} reduces. All charge pumping measurements performed in this work used devices and parameters (e.g. W/L=50µm/5µm) to avoid geometrical effects as much as possible.

Figure 3-12  Schematic of the waveform used in the three level charge pumping to profile interface traps with energies between E_c and E_i.
All the methods discussed in this chapter were utilized to analyze the properties of strained silicon, dielectrics and metal gates integrated on advanced MOSFET devices. Next, both the material properties and the electrical performance of the MOS capacitors fabricated with strained silicon will be discussed.
3.3 References


Chapter 4 Materials Analysis and Electrical Characterization of Strained Si Films and High-κ Dielectrics

In this chapter, results from both the material analysis and the electrical characterization on the in-situ boron doped Si\textsubscript{1-x}Ge\textsubscript{x} alloys and epitaxial Si films by selective deposition are reviewed. Effects of the process parameters on the deposition selectivity and surface morphology are investigated, including temperatures, gas flows, and pressures. Current-voltage characterization is performed to study the quality of epitaxial films and the effects of faceting. The integrity of high-κ dielectrics is studied by electrical characterization of MOS capacitors. Raman spectra are presented to examine the strain level of samples as well as the potential strain relaxation.

4.1 Properties of Si\textsubscript{1-x}Ge\textsubscript{x} and Strained Si Films Used in this work

4.1.1 Selectivity of Deposition

It has been more than two decades that many research groups have investigated selective silicon epitaxy [1-6]. The selectivity is expected to be different in regard to the various bonding structures of insulating material, e.g. SiO\textsubscript{2}, Si\textsubscript{3}N\textsubscript{4}, etc [4]. The role of the deposition parameters and different insulators will be briefly discussed in the following section based on previous work on selective deposition of both silicon and Si\textsubscript{1-x}Ge\textsubscript{x} alloys. In addition to undoped films, in-situ doping has been widely used for both BJT and MOSFET applications [1, 4-8]. The role of the dopants on the selectivity will also be examined.
Loss of selectivity is determined by nuclei formation and coalescence on the insulator surface which ensures the minimization of free surface energy. The critical nucleus size is defined as [9]:

\[ r_{\text{critical}} = \frac{2\gamma}{nkT \ln \frac{P}{P_e}} \] (4.1)

where \( \gamma \) is the surface free energy determined by the insulator material, \( n \) is the number of atoms per unit volume of the surface, \( k \) is Boltzman's constant, \( T \) is the vapor temperature, \( P \) is the partial pressure of the deposition species, and \( P_e \) is the equilibrium vapor pressure. Adsorbed atoms on the surface have to reach \( r_{\text{critical}} \) to form stable nuclei. The selectivity is improved as the \( r_{\text{critical}} \) increases since it is more difficult for the ad-atoms to reach this threshold for stability. Therefore, lower deposition temperatures and pressures are preferred to achieve selectivity. In addition, \( P_e \) is controllable by choosing different deposition precursors, e.g. silane (SiH₄) and disilane (Si₂H₆) with Cl₂ [8, 10], or dichlorosilane (SiH₂Cl₂) with HCl [2, 3]. Generally, lower deposition pressures are used for non-SiH₂Cl₂ based systems. It was shown by Vescan et al. that the partial pressure of the deposition species was more important than the absolute deposition pressure [11]. Partial pressure of H₂O and O₂ are reduced in low-pressure deposition processes such that the quality of deposited epitaxial films can be improved [12, 13].

The impact of temperature on the selectivity of Si epitaxy with respect to SiO₂ is different for SiH₄ and SiH₂Cl₂ based systems [4]. As the temperature was increased, higher nuclei density was observed with SiH₄, but an inverse trend was shown with SiH₂Cl₂. The same trend was also observed on Si₃N₄ while the nuclei
density was higher [14]. This was attributed to the lower number of available dangling bonds on the SiO₂ surface [14, 15]. Insulators with smaller density of dangling bonds are better to achieve selectivity [16, 17]. The dependence of 'incubation time' as a function of insulator materials is illustrated in Figure 4-1.

![Figure 4-1](image_url) Incubation time plotted as a function of insulator materials [4].

Hydrogen is generally used as the carrier gas in epitaxial systems. The role of hydrogen on selectivity was studied by Fitch et al. and they proposed that the selectivity could be improved by adding H₂ since the Si-OH bonds were replaced by Si-H bonds such that the number of dangling bonds on the oxide surface decreases [3]. However, the role of hydrogen is strongly dependent on other deposition parameters.

Selective Ge deposition was first reported by Ozturk et al. [18]. Selective Si₁₋ₓGeₓ deposition was reported by the same group and applied to fabrication of
elevated source/drain MOSFETs [19, 20]. \( \text{Si}_2\text{H}_6 \) and \( \text{GeH}_4 \) are the most commonly used precursors by many research groups [1, 6, 7, 21, 22]. It was found that the addition of \( \text{GeH}_4 \) increased the ‘incubation time’ thus selectivity could be enhanced. It is proposed that Ge adsorbed atoms could form GeO and decrease their density on the \( \text{SiO}_2 \) surface [6, 10]. Additionally, the partial pressure of the Si precursor is reduced due to the presence of \( \text{GeH}_4 \), thus the nucleation on insulator surface is also reduced [23, 24].

The optimal deposition temperature is determined by a set of requirements. For a given Ge concentration, there is a maximum deposition temperature to avoid 3-D growth. Integration with advance high-\( \kappa \) dielectrics may also pose a limit to the processing temperature. It should be noticed that the selectivity is not the only concern for Si and \( \text{Si}_{1-x}\text{Ge}_x \) epitaxy in this work. The deposition conditions must provide acceptable growth rate, desirable film composition, good film quality, as well the lowest density of nuclei on the insulator. Therefore, the deposition condition used in this work, as introduced in Chapter 3, was optimized based on all the considerations listed above.

In-situ boron doped \( \text{Si}_{1-x}\text{Ge}_x \) films deposited by Gannavaram were naturally selective up to 35-50 nm, depending on the \( \text{Si}_2\text{H}_6/\text{GeH}_4 \) ratio in the gas phase [19]. Effects of process conditions on selective deposition of heavily boron doped \( \text{Si}_{1-x}\text{Ge}_x \) was investigated by Pesovic using \( \text{Si}_2\text{H}_6 \) and \( \text{GeH}_4 \) as the precursors [25]. In this work, initial experiments were designed based on former group members’ work [25, 26]. The Ge content in the films was estimated using XRD analysis. Assuming the
film was fully relaxed, the Ge percentage in each film was extracted using Vegard’s law [27]:

\[ a_{\text{SiGe}} = x \cdot a_{\text{Si}} + (1-x) \cdot a_{\text{Ge}} \]  \hfill (4.2)

where \( a_{\text{SiGe}} \) is the lattice spacing of the mixed \( \text{Si}_{1-x}\text{Ge}_x \) film, \( a_{\text{Si}} = 0.543\text{nm} \) is the silicon lattice constant, \( a_{\text{Ge}} = 0.565\text{nm} \) is the Ge lattice constant and \( x \) is the average fraction of Ge concentration in the film. The value of \( x \) can be solved from this equation if \( a_{\text{SiGe}} \) is determined from the moving peak of SiGe alloy in XRD spectrum.

The films grown for XRD analysis must be sufficiently thick to achieve full relaxation and to eliminate the error of the Ge content extracted from XRD analysis resulting from the contribution of strain. The critical thickness value has been calculated based on different models [28-31]. However, the critical thickness could be different with respect to different deposition technologies. An experiment was designed to determine the critical thickness for a target Ge concentration. Deposition parameters and film thickness as well as the Ge content extracted directly from the SiGe peak position from XRD are listed in Table 4-1. Corresponding XRD scans are shown in Figure 4-2. It is clearly seen that as the thickness for \( \text{Si}_{1-x}\text{Ge}_x \) film increases, the alloy peak position moves toward the bulk Si peak, indicating a decrease in lattice spacing along the direction normal to the surface of the \( \text{Si}_{1-x}\text{Ge}_x \) alloy. Since the unit crystal volume is constant, the lattice spacing normal to crystal growth direction is increasing, suggesting that the alloy is approaching full relaxation. For samples with a thickness of 80 nm or more, the films are fully relaxed, hence the extracted Ge fraction gives the real film composition.
Table 4-1  Summary of XRD extracted Ge concentration in deposited films as a function of film thickness. The deposition temperature was 500°C and the deposition pressure was 700 mTorr maintained by adjusting the throttle valve.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Si$_2$H$_6$ flow (sccm)</th>
<th>GeH$_4$ flow (sccm)</th>
<th>H$_2$ flow (sccm)</th>
<th>Time (s)</th>
<th>Thickness (nm)</th>
<th>Ge content</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>5</td>
<td>25</td>
<td>345</td>
<td>30</td>
<td>~25</td>
<td>0.649</td>
</tr>
<tr>
<td>X1</td>
<td>5</td>
<td>25</td>
<td>345</td>
<td>60</td>
<td>~50</td>
<td>0.549</td>
</tr>
<tr>
<td>X4</td>
<td>5</td>
<td>25</td>
<td>345</td>
<td>75</td>
<td>~65</td>
<td>0.530</td>
</tr>
<tr>
<td>X5</td>
<td>5</td>
<td>25</td>
<td>345</td>
<td>90</td>
<td>~80</td>
<td>0.499</td>
</tr>
<tr>
<td>X3</td>
<td>5</td>
<td>25</td>
<td>345</td>
<td>120</td>
<td>~100</td>
<td>0.493</td>
</tr>
</tbody>
</table>

Figure 4-2  XRD scans for the deposited films with same gas flow but different thickness. Deposition parameters are listed in Table 4-1.

As proved experimentally, the Ge content in Si$_{1-x}$Ge$_x$ alloy is proportional to the ratio of partial pressure of Si$_2$H$_6$ and GeH$_4$. The deposition rate is also determined by these partial pressures. A higher partial pressure will result in a higher growth rate. Additional hydrogen can be added to reach the desired deposition pressure. The
growth rate can be enhanced by increasing the total deposition pressure, keeping the partial pressure ratio of the precursors constant. The impact of hydrogen on selectivity and film quality was investigated by conducting an experiment in which Si$_{1-x}$Ge$_x$ films was grown at different H$_2$ partial pressures, while the input partial pressures of Si$_2$H$_6$ and GeH$_4$ were kept constant [25]. It was found that the growth rate decreased monotonically as H$_2$ flow increases. This was attributed to the insufficient hydrogen desorption from the growth surface due to the low deposition temperature (500°C). Although H$_2$ can desorb from Si at temperatures as low as 400°C, higher temperatures may be necessary to achieve the same desorption rate considering the hydrogen pressure in the background [25].

Since the pressure in the process chamber generally varies linearly with the flow rate and varies inversely with the conductance of the throttle valve, opening or closing the exhaust throttle valve can change the overall conductance of the vacuum pump and the exhaust line so that the desired process pressure can be achieved and maintained independent of the gas flow. In our work, throttling by a gate-valve was not used in order keep the pumping speed as high as possible thus reducing the carbon and oxygen contamination during growth. Experiments were designed to optimize the deposition conditions. AFM was employed to measure the surface roughness of the deposited layers. The areas examined were 5 x 5 $\mu$m$^2$, and the images were post-processed using a flattening algorithm. Deposition parameters, growth rate and film roughness are listed in Table 4-2.
Table 4-2  Experimental conditions considered in this study to optimize deposition parameters without using the throttle valve.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Si$_2$H$_6$ flow (sccm)</th>
<th>GeH$_4$ flow (sccm)</th>
<th>H$_2$ flow (sccm)</th>
<th>B$_2$H$_6$ flow (sccm)</th>
<th>T (°C)</th>
<th>P (mTorr)</th>
<th>Growth Rate (nm/min)</th>
<th>RMS roughness (nm) (film/oxide)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Si</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>800</td>
<td>32</td>
<td>37.23</td>
<td>0.20/0.41</td>
</tr>
<tr>
<td>B: Si</td>
<td>2.5</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>800</td>
<td>39</td>
<td>14.19</td>
<td>0.17/0.32</td>
</tr>
<tr>
<td>C: Ge</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>0.003</td>
<td>500</td>
<td>~260</td>
<td>30.01</td>
<td>5.73/0.01</td>
</tr>
<tr>
<td>D: SiGe</td>
<td>5</td>
<td>75</td>
<td>0</td>
<td>0.003</td>
<td>500</td>
<td>~270</td>
<td>21.17</td>
<td>2.33/0.01</td>
</tr>
<tr>
<td>E: SiGe</td>
<td>10</td>
<td>75</td>
<td>0</td>
<td>0.003</td>
<td>500</td>
<td>~275</td>
<td>43.10</td>
<td>0.60/0.79</td>
</tr>
<tr>
<td>F: SiGe</td>
<td>15</td>
<td>45</td>
<td>0</td>
<td>0.003</td>
<td>550</td>
<td>~150</td>
<td>73.24</td>
<td>1.01/2.62</td>
</tr>
<tr>
<td>G: SiGe</td>
<td>15</td>
<td>70</td>
<td>0</td>
<td>0.003</td>
<td>550</td>
<td>~255</td>
<td>37.87</td>
<td>1.40/5.24</td>
</tr>
</tbody>
</table>

From the results, single crystalline Si films can be achieved at 800°C with very good selectivity with respect to SiO$_2$. Considering that the target thickness of the strained Si layer can be quite thin (~10 nm), condition B is more desirable. Pure Ge deposition shows very good selectivity, however, the film is very rough which is due to the large lattice-constant mismatch. Selectivity becomes worse as the Si$_2$H$_6$ flow rate increases, which is attributed to the increase in the ad-atom density [6, 10, 23, 24]. On the other hand, the surface roughness improves as the Si$_2$H$_6$ partial pressure increases due to the reduction in the lattice mismatch. The surface roughness of samples deposited at condition F or G was higher and degradation of selectivity was observed, which suggests that the deposition temperature of 550 °C may be too high for the Ge content of these samples.

The Si$_{1-x}$Ge$_x$ growth rate was investigated as a function of the B$_2$H$_6$ partial pressure for different GeH$_4$ flows [25]. It was found that the growth rate increased significantly with GeH$_4$ flow, while B$_2$H$_6$ did not appear to have an impact on the growth rate. This can be explained by the catalytic effect of GeH$_4$ in Si$_{1-x}$Ge$_x$ growth.
because the hydrogen desorption energy was lowered [5, 32-34]. Growth rate also strongly depends on the deposition temperature.

The Ge content of undoped films increased with GeH₄ flow, but saturated at around 80% [25]. Strain compensation was observed with very high in-situ boron doping, which was demonstrated by comparing the SiGe peak in XRD scans of doped and undoped films with same thicknesses [20, 25, 26]. However, the doping level employed in this work was much lower (~10¹⁷ cm⁻³), therefore the boron compensation effect is negligible.

Considering all of the criteria, Condition E was chosen as the most commonly used deposition condition for selective Si₁₋ₓGeₓ epitaxy in our study, while condition B was used to deposit epitaxial Si films. It is important to note that these experiments were carried out to determine the optimum deposition conditions for the device study instead of completing an in-depth study of the deposition processes. A detailed study of selective epitaxy of Si and Si₁₋ₓGeₓ films can be found in other sources, e.g. N. Pesovic’s thesis [25].

Typical AFM surface scans of in-situ boron doped Si films, relaxed Si₁₋ₓGeₓ films and strained Si films grown at condition B and E are shown in Figure 4-3. The extracted unprocessed bulk Si substrate Root Mean Square (RMS) surface roughness after 1% HF dip was ~0.2 nm. The RMS roughness of epi Si and relaxed Si₁₋ₓGeₓ film are ~0.2 nm and ~0.6 nm, while strained Si samples with thickness of 9, 15, and 20 nm show an RMS roughness of 0.75, 0.60, and 0.62 nm, respectively. Therefore, the epitaxial Si films exhibit roughness values comparable to that of bulk Si films. The roughness of the strained Si samples is mainly determined by the
underlying relaxed Si$_{1-x}$Ge$_x$ virtual substrate. Also note that no correlation between
the surface roughness and strained Si thickness was observed, as shown in Figure 4-4.

Figure 4-3  AFM surface scans of deposited films: (a) epi Si: RMS ~0.17 nm (Z scale 30 nm);
(b) Si$_{1-x}$Ge$_x$: RMS ~0.59 nm (Z scale 30 nm); (c) strained Si (9 nm): RMS ~0.75 nm (Z scale 30
nm); (d) strained Si (15 nm): RMS ~0.60 nm (Z scale 30 nm); (e) strained Si (20 nm): RMS ~0.62 nm (Z scale 20 nm).

![RMS roughness of deposited films plotted as a function of strained Si thickness.](image)

**Figure 4-4** RMS roughness of deposited films plotted as a function of strained Si thickness.

### 4.2 Properties of High-κ Dielectrics: HfO₂

The high-κ dielectrics (HfO₂) used in this work were deposited by the physical vapor deposition (PVD) method of sputtering. In the beginning, HfO₂ was formed by Hf reactive-sputtering followed by a post-deposition anneal in N₂ at 500°C for 5 minutes in furnace [35], which was done at the University of Texas at Austin and used in the work published in *Applied Physics Letters*. This can be seen in Chapter 5 section 5.5. Due to the limitation of the sputtering tool in Austin, all the HfO₂ films used in strained Si MOS capacitors and MOSFETs were formed in our own facilities. Experiments were designed based on UT’s condition [35] to approach a feasible HfO₂ formation process. Both electrical characterization and materials analysis were carried out to examine the integrity of HfO₂.
4.2.1 Electrical Characteristics of HfO₂ Metal-Oxide-Semiconductor (MOS) Capacitors

Experimental conditions used for the HfO₂ formation process are listed in Table 4-3. The sputtering rate was calibrated by measuring the stepheight of four Hf samples. The sputtering time of each sample was 3, 5, 10 and 20 minutes, respectively. The average sputtering rate of Hf was determined to be ~2 nm per minute.

Table 4-3 Process parameters used in HfO₂ formation experiments.

<table>
<thead>
<tr>
<th></th>
<th>Hf sputtering power (W)</th>
<th>Hf sputtering time (s)</th>
<th>Ar gas flow (sccm)</th>
<th>Annealing condition</th>
<th>EOT (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hf₁</td>
<td>50</td>
<td>60</td>
<td>40</td>
<td>RTA* 60s + FA**</td>
<td>1.26</td>
</tr>
<tr>
<td>Hf₂</td>
<td>50</td>
<td>60</td>
<td>40</td>
<td>FA** only</td>
<td>1.47</td>
</tr>
<tr>
<td>Hf₃</td>
<td>50</td>
<td>70</td>
<td>40</td>
<td>RTA* 75s + FA**</td>
<td>1.46</td>
</tr>
<tr>
<td>Hf₄</td>
<td>50</td>
<td>70</td>
<td>40</td>
<td>FA** only</td>
<td>1.60</td>
</tr>
</tbody>
</table>

* RTA = rapid thermal anneal, which was carried out in PEAK RTA tool at 600°C in N₂;
** FA = furnace anneal, which was carried out in the Tylon furnace at 500°C for 5 minutes in N₂.

MOS capacitors were fabricated with tungsten (W) as the gate electrode. The equivalent oxide thickness (EOT) of each sample was extracted from Hauser’s CVC program which is a least squares fit to the C-V data and is shown in Figure 4-5 (a) and (d) [36]. As seen, all samples achieved an EOT of less than 2 nm. The leakage current density (J₉) is plotted vs. V₀x in Figure 4-5 (b). Compared to the J₉ of SiO₂ samples given in literature (as shown in Figure 4-5 (c)) [37], where V₀x was defined as V₉-1.1 volt, all samples exhibited similar J₉ as that of a SiO₂ sample with 2.5 nm EOT, indicating significant decrease in gate leakage. It should be noted that there was deviation in EOT and J₉ for measurements taken from all samples, typically as shown in Figure 4-5 (d), which is probably due to a thickness non-uniformity of the HfO₂ films resulting from the very short time of sputtering needed for that target
thickness. In addition, comparing the EOT and $J_g$ data of sample Hf1/2 and Hf3/4, respectively, one can question that whether the RTA step is necessary to form the HfO₂. Since the design of the PEAK RTA tool does not allow the process chamber and the sample-holding plate to be cleaned right before loading the as-deposited Hf samples, there might be additional contamination to the high-$\kappa$ surface. The PEAK RTA tool is a pumped system, such that there might be less oxygen in the ambient to achieve full oxidation of Hf. Based on these concerns, the RTA step was excluded from the final HfO₂ process. Hf was sputtered at 50 Watts for 60 or 70 seconds followed by a furnace anneal at 500 °C for 5 minutes. Before ramping up the temperature, $N_2$ was flowing at 7200 sccm to purge the furnace for another 5 minutes. Including the ramp-up and ramp-down steps, the total furnace anneal process time is about 20 minutes. An RTA at 950 °C for 30 seconds was performed on some devices to investigate the thermal stability of the dielectrics. C-V and I-V data are summarized in Table 4-4.

Table 4-4 Electrical properties extracted from C-V data and leakage current densities before and after RTA of HfO₂ samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$V_{fb}$ (V)</th>
<th>EOT (nm)</th>
<th>$J_g$ (A/cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hf 60s (RTA600°C,60s +FA)</td>
<td>-0.04~0.01</td>
<td>1.834~2.43</td>
<td>5E-6~2E-2</td>
</tr>
<tr>
<td>Hf 70s (FA only)</td>
<td>-0.16~0.2</td>
<td>1.66~1.76</td>
<td>2E-6~2E-4</td>
</tr>
<tr>
<td>After RTA 950°C,30s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hf 60s (RTA600°C,60s +FA)</td>
<td>-0.13~0.38</td>
<td>2.001~2.504</td>
<td>2e-5~6e-3</td>
</tr>
<tr>
<td>Hf 70s (FA only)</td>
<td>-0.367~0.38</td>
<td>1.918~2.21</td>
<td>1e-4~0.01</td>
</tr>
</tbody>
</table>

The final Hf sputtering condition of the HfO₂ process was set at 70 seconds followed by a furnace anneal at 500 °C for 5 minutes.
Figure 4-5 (a) ~ (c)
4.2.2 X-ray Photoelectron Spectroscopy (XPS) and Transmission Electron Microscopy (TEM)

In general, due to the nature of the process condition of the HfO₂ formation, there is always an interfacial layer developed intentionally or unintentionally between high-κ and the bulk Si (or strained Si in this case) regardless of the deposition method. Compared to other deposition methods such as ALD and MOCVD, high-κ films formed by the PVD method may have slightly thicker interfacial layers. TEM can be employed to study the quality of interfacial layer and HfO₂ film, including the thickness of each layer and the interface between HfO₂, the interfacial layer and bulk Si or strained Si substrate. XPS was employed to give us a better understanding on the composition of HfO₂ films by investigating the peak positions and intensities corresponding to different bonds.
The XPS spectra of Hf 4f and Si 2p from two HfO₂/bulk Si samples are shown in Figure 4-6. The HfO₂ film in sample Hf433 was formed by sputtering Hf for 60 seconds followed by annealing in the furnace at 500°C for 5 minutes. Sample Hf433 was sputtered for 70 seconds followed by the same anneal. As seen in the figure, there is no significant difference between these two samples. From the Hf 4f spectra plotted in Figure 4-6 (a), the 4f 7/2 peak at 16.7 eV corresponds to the HfO₂. Therefore the film is mainly composed of HfO₂. The peak at 14.2 eV is the Hf-Si bonding at the interface. After sputtering for 20 minutes (~1 nm), no significant change in the peak intensity was observed indicating that the interfacial layer was not very thick. In the Si 2p spectrum shown in Figure 4-6 (b), the peak at 99.3 eV is from the Si substrate while the peak at 103 eV resulting from the Si-O-Hf bonding implying a presence of a silicate-like (HfSiOₓ) interfacial layer. After 20 minutes sputtering, the Si peak was getting stronger as dielectrics became thinner. However, the interfacial silicate peak did not change appreciably. Thus, the dielectric is mostly composed of HfO₂. The interfacial silicate layer exists but is not dominant. TEM will help in providing detailed information on the structure of the gate dielectrics.
The XPS spectra of Hf 4f and Si 2p core-levels of two HfO₂/Si samples are shown in Figure 4-7 and compared to the HfO₂/bulk Si system. The strained Si sample was annealed at 950 °C for 30 seconds in N₂ after HfO₂ formation. It was then submitted to a standard forming gas anneal (FGA) which was performed to provide information about the final gate-dielectric composition of transistors and the same thermal budget as the high-κ MOSFETs. The sputtering time of Hf was 60 seconds. XPS spectra of sample Hf434 were also plotted as a comparison. Similar Hf-O peaks were observed from both samples. However, in the Si 2p spectrum, the ratio of the intensity of the Si-O-Hf and Si-Si peaks in the RTA sample was found to be higher than the ratio in sample Hf433, which may suggest the growth of an
interfacial silicate layer during RTA. Again, HRTEM will be necessary to achieve an explicit image of the structure and composition of this gate dielectric stack.

Figure 4-7 The XPS spectra of (a) Hf 4f and (b) Si 2p core-levels of a RTA HfO₂/strained Si sample compared to the HfO₂/bulk Si system.

A TEM image of the physical structure of these gate stacks is shown in Figure 4-8. This TEM cross section image was obtained from a MOSFET sample, and there are nine different layers in the gate stack which were labeled in Figure 4-8 (a) and (b). It can be seen that after RTA the W capping layer was partially oxidized. The interface between HfO₂ and TaN is not as smooth as the interface between HfO₂ and the interfacial layer or between the interfacial layer and the strained Si layer underneath, which may be attributed to partial crystallization of HfO₂ after high temperature annealing. The interfacial layer is about 2.15 nm for the TaN gated sample, with the HfO₂ thickness as about 9 nm, as shown in Figure 4-8 (c). Based on the EOTs obtained from samples after RTA which was only about 2nm, this
indicates that the interfacial layer is a medium $\kappa$ dielectric, for example, HfSi$_x$O$_y$. Since the interfacial layer is fairly thick, the dominant mobility degradation mechanism would be attributed to the high $\kappa$ dielectric [38].

Figure 4-8 (a)
Figure 4-8 (b)

Figure 4-8 (c)

Figure 4-8 A TEM image of a strained Si sample with high $\kappa$ and metal gate electrode stacks. All layers were labeled in (a) and (b) while details of HfO$_2$ and the interfacial layer were shown in (c).
4.3 Effects of Facets on Electrical Properties

Faceting of a crystal surface can be caused by strongly anisotropic surface tension and driven by a surface diffusion mechanism. In a typical chemical vapor deposition process, the flux of material on a growing crystal surface is from the diffusion boundary layer whose shape follows the shape of the surface. Therefore, the surface tension depends on the local curvature [39]. The impact of the resulting pyramidal structures can influence the electrical properties and the thermal stability of the epitaxial films. In selective epitaxial growth, faceting is typically observed, which is determined by the orientation of sidewalls [40].

An SEM image of selectively deposited Si$_{1-x}$Ge$_x$ layer with amorphous Si (α-Si) deposition on top is shown in Figure 4-9, which was investigated by S. Chopra, who is also working in Dr. Ozturk’s group. The deposition condition is same as Condition E introduced previously. A [111] SiGe facet is observed. The α-Si layer is conformal with the topology of the structure due to the nature of non-selective deposition of α-Si on SiO$_2$. If a crystalline Si layer is deposited on the Si$_{1-x}$Ge$_x$ film, the same facet is expected.
4.3.1 Overlap and Non-overlap Capacitors

There are two typical capacitors designed in the GEM task. The first type is the overlap capacitor and it is called such because there is a portion of the gate area overlapping with the field oxide region, i.e., the active area is smaller than the gate pads. The other type is the non-overlap capacitor, which has a smaller gate pad than the active area such that there is no overlapped region between the gate and the field oxide. The structures of these two capacitors are illustrated in Figure 4-10. It is clear that with overlap capacitors, additional perimeter leakage effects exist, which is not a concern for non-overlap capacitors.
4.3.2 I-V Measurement of Overlap and Non-overlap Capacitors

Leakage current densities of overlap and non-overlap capacitors are shown in Figure 4-11, with (a) SiO$_2$ and (b) HfO$_2$ as gate dielectrics. Solid lines refer to the $J_g$ of non-overlap capacitors while curves with symbols refer to the $J_g$ of overlap capacitors. A significant difference in $J_g$ between overlap and non-overlap capacitors can be observed in Figure 4-11 (a) in that overlap capacitors show much higher $J_g$. For all the strained Si samples with SiO$_2$ gate dielectric, $J_g$ was similar for both overlap and non-overlap capacitors, and it was higher than that of the bulk Si control. The same trend was observed regardless of the gate electrode employed. A possible cause of this phenomenon is the perimeter leakage in overlap capacitors mentioned previously. For epitaxial Si samples, the overlap capacitors showed much higher $J_g$, which suggests that perimeter leakage is still dominant. In strained Si samples, other leakage mechanisms might also be present. I-V data of strained Si with HfO$_2$ samples are shown in Figure 4-11 (b). Compared to the SiO$_2$ samples,
epitaxial Si samples with HfO$_2$ showed a smaller difference in $J_g$ with overlap and non-overlap capacitors. This may be attributed to the fact that $J_g$ is fairly high in both cases due to its thin HfO$_2$ EOT of ~1.2 nm. However, higher $J_g$ was observed on overlap capacitors than non-overlap capacitors for strained Si samples. Facets of selectively deposited SiGe and strained Si films may result in a non-uniform HfO$_2$ film and possibly an additional leakage path. For non-overlap capacitors, epitaxial Si films are as good as the bulk Si films considering the leakage. Based on the conclusions of these experiments, it was decided to use non-overlap capacitors to extract all relevant electrical parameters.

![Graphs](image)

**Figure 4-11** Leakage current densities of overlap and non-overlap capacitors with (a) SiO$_2$ and (b) HfO$_2$ as gate dielectrics.
4.4 Strain Analysis by Raman Spectroscopy

Raman spectroscopy was employed in this work to examine the strain in the epitaxial Si film on top of the relaxed Si$_{1-x}$Ge$_x$ virtual substrate. A typical Raman spectrum of strained Si is shown in Figure 4-12(a). There are four distinct peaks corresponding to the Si-Si bonds in the bulk Si layer, Si-Si bonds, Si-Ge bonds and Ge-Ge bonds in the Si-Ge layer, respectively. A small shoulder of the bulk Si-Si peak at 520 cm$^{-1}$ was observed, which is associated with Si-Si bonds in the strained Si layer as reported in literature [41]. The Ge content is calculated to be ~48% using the equations given in [41]. A separate peak of strained Si can be obtained by Gaussian multiple-peak fitting with a background correction, whose position can give the information of the relative amount of strain. Raman analysis proved that different strain levels were achieved by depositing strained Si films with different thicknesses, as shown in Figure 4-12 (b). In addition, the intensity of the Si-Si peak from the strained Si layer decreased as the strained Si thickness decreased, which is due to the lower amount of signal collected. The Raman peak shift is defined as:

$$\Delta \omega = \omega_{\text{Si-Si, bulk}} - \omega_{\text{Si-Si, strained Si}}$$  \hspace{1cm} (4.3)

where $\omega_{\text{Si-Si, bulk}}$ is always 520 cm$^{-1}$. Since the amount of strain is linear proportional to $\Delta \omega$, the change in $\Delta \omega$ can give us an explicit understanding on the strain thermal stability. Raman peak shifts are plotted as a function of strained Si thickness in Figure 4-12 (c). For each strained Si thickness, there are three samples examined by Raman: as-deposited, immediately after dielectric formation, and after dielectric formation followed by RTA at 950 °C for 30 seconds in N$_2$. 


Figure 4-12  Raman spectra of strained Si samples with different thicknesses: (a) a typical Raman spectrum of a strained Si sample; (b) different strain level was achieved by varying strained Si thickness; (c) Raman peak shifts plotted as a function of strained Si thickness.
Negligible differences in Raman peak shifts were observed among the strained Si with HfO$_2$ samples after HfO$_2$ formation, the strained Si with HfO$_2$ samples after RTA and the as-deposited strained Si samples. However, for strained Si/SiO$_2$ samples, the Si-Si peaks shifted to the left, which is probably due to Si consumption. The oxidation rate of strained Si is expected to be very similar to that of bulk Si and the Si consumption is $\sim 0.44 \cdot t_{ox}$, where $t_{ox}$ is the thermal oxide thickness [42]. $\Delta \omega$ was found to be smaller after RTA, which suggests that partial relaxation after RTA occurred and/or Ge was incorporated into the strained Si layer. More discussion will be conducted in Chapter 5.
4.5 References


Chapter 5 Materials Analysis and Electrical Characterization of Strained Si Metal-Oxide-Semiconductor (MOS) Capacitors

Many important issues need to be addressed before advanced gate stacks including high-κ dielectrics and metal gate electrodes are applied to strained silicon devices. These include the interfacial layer formation at the strained Si/high-κ dielectric interface and the effect of metal gate electrodes on the channel strain. It is also necessary to fully understand the impact of Ge on the properties of devices with strained Si channels, when a relaxed Si$_{1-x}$Ge$_x$ layer exists in close proximity. A fundamental study was first carried out on MOS capacitors formed on strained Si layers. The impact of the strained silicon thickness on dielectric properties was investigated for both HfO$_2$ and SiO$_2$. Mechanisms responsible for degradation of the electrical properties of the MOS gate stacks during high temperature oxidation and/or rapid thermal annealing (RTA) were investigated.

5.1 Electrical Characterization of Strained Si MOS Capacitors

Strained Si layers were obtained by selective epitaxy of a thin Si layer on top of a relaxed Si$_{1-x}$Ge$_x$ buffer in windows defined in a 100 nm thick isolation oxide. In general, the Ge content was ~50% unless specified. In fabrication of MOS capacitors, either SiO$_2$ or HfO$_2$ layers were used as the gate dielectric. SiO$_2$ was grown by dry oxidation in a furnace at 800°C for 8 minutes. HfO$_2$ was formed using the conditions introduced in Chapter 4. TaN and Ru-Ta alloy metal gate electrodes were formed by sputtering. Electrical measurements were performed before and after RTA in N$_2$ at 950°C for 30 sec. The last process step for all samples was
annealing in forming gas (10% H₂ in N₂) at 400°C for 30 min in a conventional tube furnace (FGA). Mostly the area of measured capacitor is 50x50 µm².

5.1.1 Electrical properties of Samples after Forming Gas Anneal (FGA)

C-V curves obtained from MOS capacitors with SiO₂ and HfO₂ gate dielectrics and TaN gate electrodes were shown in Figure 5-1. Equivalent oxide thickness (EOT) and flatband voltage (VFB) were extracted using Hauser’s program [1]. It can be seen that while the control sample of MOS capacitors fabricated on the bulk Si exhibits normal C-V behavior throughout the entire voltage range, samples with strained Si display a change in slope near the onset of inversion, indicating a trend of Dit increasing as strained Si thickness decreases. EOTs of SiO₂ and HfO₂ samples are plotted in Figure 5-2 (a) as a function of strained Si thickness. All SiO₂ samples show similar EOTs of ~3.5nm while the EOTs of HfO₂ samples are ~2.3nm. No correlation between EOT and strained Si was observed.

Figure 5-1 C-V curves from MOS capacitors with SiO₂ and HfO₂ gate dielectrics and TaN gate electrodes. The measured area is 50µm by 50µm.
Figure 5-2  (a) Equivalent oxide thickness and (b) leakage current density of SiO$_2$ and HfO$_2$ samples are plotted vs. strained Si thickness. The gate electrodes are TaN.

The gate leakage current densities ($J_g$) of the fabricated MOS capacitors were measured using a HP 4155B Semiconductor Parameter Analyzer. The $J_g$ at 1 volt beyond the $V_{FB}$ was plotted as a function of strained Si thickness in Figure 5-2 (b). Capacitors on strained Si exhibited higher $J_g$ than the bulk Si control samples regardless of the gate dielectrics even though the EOTs were similar. Thinner strained Si samples were found to have higher leakage.

Figure 5-3  Hysteresis of SiO$_2$ and HfO$_2$ samples are plotted vs. strained Si thickness, with TaN as the gate electrodes.
Hysteresis was observed with both SiO$_2$ and HfO$_2$ samples formed on strained Si layers and was found to decrease with increasing strained Si thickness as shown in Figure 5-3. Our measurements with other metal gates revealed that none of the electrodes considered in this study had any contribution on hysteresis.

The density of interface traps (D$_{it}$) were extracted for different strained Si layer thicknesses via the conductance method [2] and plotted as a function of the trap energy level with respect to the valence band edge in Figure 5-4 (a). The average D$_{it}$ was also plotted as a function of the final strained Si thickness in Figure 5-4 (b). During gate oxidation, the amount of Si consumed from the substrate is approximately equal to 44% of the SiO$_2$ thickness, which is valid for both bulk and strained Si [3]. A very thin interfacial SiO$_2$ layer (less than 1nm) is also expected to form during HfO$_2$ formation [4] which consumes some of the substrate. However, this amount is negligibly small compared to that consumed during SiO$_2$ formation. Therefore, for the SiO$_2$ samples the final strained Si thickness was found by subtracting the consumed Si thickness from the as-grown strained Si thickness, while for the HfO$_2$ samples the original as-grown strained Si thickness was used. It can be observed that the extracted D$_{it}$ increases with strained Si thickness decreasing for both SiO$_2$ and HfO$_2$, and it is consistently higher for SiO$_2$ for the same as-grown Si thickness. Discussion on the degradation mechanisms is included in following section (5.2).
5.1.2 Electrical properties of Samples after Rapid Thermal Anneal (RTA)

With RTA at 950°C for 30 sec in N₂, HfO₂ capacitors exhibited higher leakage levels in the low voltage range (within the range of ~1V beyond the V_{FB}). This trend was similar for all capacitors with HfO₂ gate dielectrics including those fabricated on bulk Si substrates, as shown in Figure 5-5 (a). The change in EOT during RTA was found to be 1~2Å, as shown in Figure 5-5 (b), which cannot explain the large increase in J_g. This may be attributed to metal gate/high-κ interactions. Previous studies indicated that metal/high-κ interactions and metal diffusion through the crystalline high-κ grain boundaries may introduce bulk traps, and/or reaction layers that can increase the gate leakage [5]. Negligible change in J_g of SiO₂ samples was observed with RTA. Hysteresis after RTA and a final FGA was also measured and the results are shown in Figure 5-5 (c). It can be seen that after RTA, all capacitors showed negligible hysteresis and it was no longer a function of the strained Si
thickness. The small hysteresis observed after RTA indicates that the high temperature anneal improves the dielectric quality by oxidation and/or by reducing the bulk trap density [6]. The $D_{it}$ measurements were repeated after RTA and a final FGA, which indicated slightly lower $D_{it}$ for both gate dielectrics, but still following the same trend with the Si thickness, as shown in Figure 5-6.

![Figure 5-5](image-url)  
(a): up-left;  
(b): up-right;  
(c): bottom-left.

**Figure 5-5** Electrical parameters of TaN gate MOS capacitors after RTA plotted vs. strained Si thickness: (a) leakage current density at 1V beyond $V_{FB}$; (b) EOT; (c) hysteresis.
5.2 Possible Mechanisms of Electrical Property Degradation

Comparing the electrical properties measured from samples with different strained Si thickness as well as those before and after RTA, some interesting observations should be underscored. Firstly, thinner strained Si samples show degradation in electrical properties, including higher $J_g$, higher $D_{it}$ and higher hysteresis. Secondly, strained Si with SiO$_2$ samples show higher $D_{it}$ compared with HfO$_2$ with same starting strained Si thickness. After RTA, hysteresis became negligible, and $D_{it}$ decreases slightly. However, same correlation between $J_g / D_{it}$ and strained Si thickness is observed.

A potential explanation for this phenomenon is the degradation of the oxide quality due to Ge diffusion into the strained Si layer [7], which may explain the higher $D_{it}$ levels observed for thinner strained Si layers. The fact that HfO$_2$ resulted in lower $D_{it}$ than SiO$_2$ may also be attributed to thinning of the strained Si layer during SiO$_2$ formation and the resulting increase in the Ge concentration near the Si/SiO$_2$ interface.
interface [8]. The fact that SiO$_2$ results in a higher $D_{it}$ than HfO$_2$ for the same strained Si thickness may be attributed to the higher Ge concentration near the Si/SiO$_2$ interface due to Si consumption as well as the enhanced Ge diffusion during oxidation. It is well known that the $D_{it}$ of the SiO$_2$-SiGe system is much higher than that of the SiO$_2$-Si system [9] due to larger density of intermediate oxidation states at the SiO$_2$-SiGe interface [10]. Therefore, the reason that lower $D_{it}$ was observed with HfO$_2$ samples than SiO$_2$ ones with same starting strained Si thickness is due to the less Si consumption as well as the low process temperature.

To study the presence of Ge underneath the gate oxide, secondary ion mass spectroscopy (SIMS) was performed on three different samples: i) as-grown without a gate dielectric or a high temperature process step, ii) after growing a 6 nm SiO$_2$ layer by dry oxidation at 850°C for 30 min, and iii) after HfO$_2$ formation followed by RTA. The thicknesses of these three samples are the same (~20nm) and the Ge content in the virtual substrate is ~50%. From the Ge profiles given in Figure 5-7, a Ge tail was observed in the strained Si layer even without SiO$_2$ formation or RTA. This behavior was previously attributed to the lower surface energy resulting from having an ad-layer of Ge atoms at the growth surface during Si$_{1-x}$Ge$_x$ epitaxy resulting in a Ge rich surface when the growth is terminated. It was proposed that during growth, the Ge atoms on the growth surface exchange sites with the underlying Si atoms resulting in Ge incorporation in the silicon film [11]. It was also shown that formation of this Ge layer could be suppressed by surfactant-mediated epitaxy, which involves using a different species, to lower the surface energy [5]. Ge concentration increasing in thinner strained Si layer was confirmed by SIMS analysis.
(as shown in Figure 5-7). It also indicated that during SiO$_2$ formation, Ge concentration in Si increased even further. It has been proposed that Ge diffusion in Si is dominated by the monovacancy mechanism [5], which could be enhanced due to the vacancy injection during oxidation. No significant diffusion observed after HfO$_2$ formation and RTA.

Figure 5-7 Ge profiles from SIMS show the effect of high temperature process on Ge out-diffusion. Samples have 20nm strained Si layer.

XPS analysis of Ge 3$d$ spectrum was detected on the surface of the thinnest strained Si sample (9nm) which experienced gate oxidation at 800°C for 8 minutes. The thin oxide layer was removed by 1% HF solution before the sample was loaded to the XPS vacuum chamber. As shown in Figure 5-8, the peak of Ge 3$d$ demonstrated the presence of Ge in the Si/oxide interface. After sputtering for 15 minutes (~1nm), the change of peak intensity is negligible, indicating that the Ge signals were only collected from the strained Si capping layer. However, due to the XPS resolution limit, no Ge was detected on the surface of thicker strained silicon layers.
Another possibility for increasing $D_{it}$ might be related to stress distribution at the Si/dielectric interface. In bulk Si, it has been shown that the midgap $D_{it}$ is proportional to the thickness-averaged stress in SiO$_2$ and Si [6]. The stress in the SiO$_2$ is compressive and the Si substrate is under tensile stress [6]. With intentionally grown tensile-strained Si on SiGe, the stress distribution is expected to be less and thus the $D_{it}$ is expected to be less especially for thinner strained Si samples in which higher amount of strain was proved by Raman spectroscopy. Therefore, the increase in $D_{it}$ for decreasing strained Si thickness is mainly attributed to the presence of Ge at the interface.

To further investigate the impact of Ge at the interface on the electrical properties, MOS capacitor with 3.5nm SiO$_2$ and TaN gate electrodes were fabricated with varying strained Si thickness and lower Ge content in the Si$_{1-x}$Ge$_x$ virtual substrate. Interface trap density was measured by conductance method and the results were plotted vs. the trap energy in Figure 5-9 (a). Same correlation between
$D_{it}$ and the strained Si thickness was observed as was seen with the 50% Ge samples: as the strained Si thickness decreases, $D_{it}$ increases. Lower $D_{it}$ was observed with samples having lower Ge content. The presence of Ge in the interface will have stronger effect on the electrical performance when the strained Si film is quite thin, as shown in Figure 5-9 (b) while thicker strained Si samples showed much less deviation with the varying Ge content.

Figure 5-9 The density of interface traps ($D_{it}$) plotted as a function of (a) the trap energy; (b) Ge concentration with varying strained Si thickness. TaN is used as the gate electrode.
Strain relaxation could also have an impact on the interface state density. Raman peak shifts of as-deposited and after gate oxidation strained Si samples with three different substrate Ge concentrations were shown in Figure 5-10. Smaller peak shift was observed with strained Si samples with less Ge in the virtual substrate, demonstrating less amount of strain. The Ge profiles of two samples obtained from SIMS are shown in Figure 5-11. The strained Si thickness is about 25nm with 3.5nm SiO$_2$ capped on top. As can be seen, Ge out-diffusion was observed with both samples. However, less amount of Ge was observed in the strained Si layer for the sample which had ~30% Ge in the SiGe substrate.

**Figure 5-10** Raman peak shifts plotted vs. strained Si thickness with varying Ge content in the virtual substrate.
Figure 5-11  SIMS profiles of Si and Ge of two strained Si samples with different Ge content in the SiGe buffer layer. It can be seen that more Ge out-diffusion into the strained Si channel would be expected in sample B than in sample A.

Figure 5-12 plots the conductance $D_{it}$ vs. Raman peak shift with varying Ge content in the SiGe buffer layer. Comparing sample A which has 18.5 nm strained Si film and ~50% Ge in SiGe, with sample B which as 7.5 nm strained Si and ~30% Ge content in SiGe, sample A shows higher amount of strain but lower $D_{it}$ than sample B. As can be estimated from Figure 5-11, more Ge diffusion would be expected in sample B. As shown in Figure 5-12, with increasing Raman peak shift, the $D_{it}$ values also increase indicating that $D_{it}$ also increases with strain amount. However, for a given Raman peak shift, i.e. a given amount of strain, the rate of increase in $D_{it}$ with decreasing strained Si thickness is significantly more than the increasing rate of $D_{it}$ with increasing strain (Ge content in SiGe). This suggests that Ge diffusion is the dominant cause of the $D_{it}$ increase.
Figure 5-12  
$D_{it}$ plotted as a function of Raman peak shift with varying Ge content in the SiGe virtual substrate. Sample A and B refer to the conditions listed in Figure 5-11.

5.3 Summary

To summarize, the interface trap density ($D_{it}$) was found to be increasing as the strained silicon thickness decreased, which was due to the presence of Ge in the strained Si layer. Strained Si capacitors with SiO$_2$ show higher $D_{it}$ which may be attributed to Si consumption during oxidation, leading to a higher density of Ge at the interface. Leakage current density ($J_g$) and hysteresis were also observed to increase with decreasing strained silicon thickness. This correlation between $D_{it} / J_g$ and strained Si thickness did not change after RTA. Both Ru-Ta and TaN gate electrodes were found to exhibit as good performance on strained Si as on bulk Si. For a given amount of strain, the rate of increase in $D_{it}$ with increasing strained Si thickness is significantly more than the increasing rate of $D_{it}$ as the Ge concentration
in the virtual substrate is increased, which suggests that Ge diffusion is the dominant cause of the $D_{ri}$ increase.
5.4 References


5.5 Impact of Ge on integration of HfO₂ and metal gate electrodes on strained Si channels

This section is adapted from the paper “Impact of Ge on integration of HfO₂ and metal gate electrodes on strained Si channels” which was submitted to Applied Physics Letters and accepted for publication in August 2005 (Appl. Phys. Lett. 87, 071903 (2005)).
Impact of Ge on integration of HfO$_2$ and metal gate electrodes on strained Si channels

Yanxia Lin, Mehmet C. Öztürk, and Bei Chen

Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695

Se Jong Rhee and Jack C. Lee

Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78758

Veena Misra$^a$

Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695

(Received 7 February 2005; accepted 6 July 2005)

Tensile-strained Si epitaxial layers (7.5nm-17nm) were grown on relaxed Si$_{0.5}$Ge$_{0.5}$ virtual substrates by ultrahigh-vacuum rapid thermal chemical vapor deposition. Metal-oxide-silicon capacitors were fabricated with SiO$_2$ or HfO$_2$ as gate dielectrics and Ru-Ta alloy or TaN as the metal gate electrodes. The results indicate that the interface trap density ($D_{it}$) increased as the strained silicon thickness decreased, which was attributed to the presence of Ge in the strained Si layer. Higher $D_{it}$ was observed with SiO$_2$ which may be due to Si consumption during oxidation, leading to a higher density of Ge at the interface. Leakage current density ($J_g$) was also observed to increase with increasing strained silicon thickness. This trend of increasing $D_{it}$ and $J_g$ with decreasing strained silicon thickness did not change after rapid thermal annealing. Both Ru-Ta and TaN gate electrodes were found to exhibit as a good performance on strained Si as on bulk Si.

$^a$Electronic mail: vmisra@ncsu.edu
Metal-oxide-silicon field-effect transistors (MOSFETs) with strained Si channels are presently attracting considerable attention due to their potential in providing significant improvements in transistor performance. Thin Si layers grown on relaxed Si<sub>1-x</sub>Ge<sub>x</sub> buffer can be under biaxial tensile strain which results in mobility enhancement for both electrons and holes<sup>1</sup>. Uniaxial strain is also a viable option, which has recently been implemented in a 90 nm complementary metal-oxide-semiconductor (CMOS) technology using recessed Si<sub>1-x</sub>Ge<sub>x</sub> source/drain junctions and nitride capping layers for p-channel and n-channel MOSFETs, respectively<sup>2</sup>.

In recent years, the silicon industry has invested considerably in finding a high-κ dielectric material as a replacement for silicon dioxide for continued MOSFET scaling. Metal gate electrodes are also being considered as replacements to polycrystalline silicon to eliminate problems stemming from gate depletion and Fermi level pinning, particularly for p<sup>+</sup> polysilicon<sup>3</sup>. It is anticipated that continued scaling of the MOSFET will require the integration of high-κ and metal gate electrodes with strained Si channels. However, there are many important issues that need to be understood before the new gate stack materials can be used in MOSFETs with strained silicon channels. These include the interfacial layer formation at the strained Si/high-κ dielectric interface and the effect of metal gate electrodes on the channel strain. Strained Si MOSFETs with HfO<sub>2</sub>/polysilicon<sup>4</sup>, SiO<sub>2</sub>/NiSi<sup>5</sup>, and HfO<sub>2</sub>/TiN<sup>6</sup> gate stacks have been reported with good performance. These studies have shown that the mobility degradation commonly observed with high-κ dielectrics can be partially compensated by employing a strained Si channel. However, there are additional scattering mechanisms limiting the mobility enhancement in this system including increased phonon scattering attributed to HfO<sub>2</sub><sup>6</sup>. It has been shown the presence of Ge atoms in the channel can
reduce the carrier mobility due to additional Coulomb scattering in $p$-channel MOSFETs, and enhanced phonon scattering in $n$-channel MOSFETs. Therefore, it is necessary to fully understand the impact of Ge on the properties of MOSFETs with strained Si channels, when there is a relaxed Si$_{1-x}$Ge$_x$ layer in close proximity.

This letter presents the results of a fundamental study on metal-oxide-semiconductor (MOS) gate stacks with HfO$_2$ gate dielectrics and TaN or Ru-Ta gate electrodes formed on strained Si layers. The impact of the strained silicon thickness on dielectric properties was investigated for both HfO$_2$ and SiO$_2$. Mechanisms responsible for degradation of the electrical properties of the MOS gate stacks during high-temperature oxidation and/or rapid thermal annealing (RTA) were investigated.

Strained Si layers were obtained by selective epitaxy of a thin Si layer on top of a relaxed Si$_{0.5}$Ge$_{0.5}$ buffer in windows defined in a 100 nm thick isolation oxide. Both Si and Si$_{0.5}$Ge$_{0.5}$ layers were grown by ultrahigh-vacuum rapid thermal chemical vapor deposition (UHV-RTCVD) using pure Si$_2$H$_6$ and 10% GeH$_4$ in H$_2$ as the gaseous precursors. Selective Si$_{0.5}$Ge$_{0.5}$ epitaxy was performed at 500°C while the Si layers were grown at a higher temperature of 800°C. The deposition pressure was 285 mTorr for Si$_{0.5}$Ge$_{0.5}$ and 35 mTorr for Si, respectively. Selective epitaxy confines both the Si$_{0.5}$Ge$_{0.5}$ and strained Si films to a small area effectively suppressing the formation of misfit dislocations. In fabrication of MOS capacitors, either SiO$_2$ or HfO$_2$ layers was used as the gate dielectric. SiO$_2$ was grown by dry oxidation in a furnace at 800°C for 30 min. HfO$_2$ was formed by Hf sputtering followed by a postdeposition anneal in N$_2$ at 500°C for 5 min in furnace. Ru-Ta alloy and TaN metal gate electrodes were formed by sputtering. Electrical measurements were obtained before and after RTA in argon at 800°C for 30 s. The last process step for all samples was
annealing in forming gas (10% H₂ in N₂) at 400°C for 30 min in a conventional tube furnace (forming gas anneal (FGA)). Raman spectroscopy was used to study the impact of the high-temperature processes on the residual strain, which was found to be stable during both gate oxidation and RTA. Secondary ion mass spectrometry (SIMS) was used to obtain the Si and Ge profiles.

Figure 1 shows the capacitance-voltage (C-V) curves obtained from MOS capacitors with SiO₂ gate dielectrics and TaN gate electrodes using an HP 4284A Precision LCR meter. The equivalent oxide thickness (EOT) and flat-band voltage (V_{FB}) were extracted using a program developed by Hauser at NCSU. The C-V curves obtained with HfO₂ gate dielectrics and/or Ru-Ta gate electrodes exhibited similar behavior to those shown in Fig. 1, with very similar EOT values (~6 nm for SiO₂ and ~2 nm for HfO₂ samples, respectively). The flat-band voltage levels obtained for different gate electrodes were in accord with the expected work function values. It can be seen that while the MOS capacitors fabricated on bulk and epitaxial Si control samples exhibit normal C-V behavior throughout the entire voltage range, samples with strained Si channels exhibit a slope change near the onset of inversion. A magnified version of this region has been included in Fig. 1, which shows that the slope is decreasing with the strained Si thickness. The observed change is indicative of changes in the interface trap density (D_{it}) and suggests that the interface traps can change their charge states fast enough in response to changes in the gate bias. We have extracted D_{it} for different strained Si layer thicknesses via the conductance method and plotted as a function of the strained Si thickness in Fig. 2. During gate oxidation, the amount of Si consumed from the substrate is approximately equal to 44% of the SiO₂ thickness, which is valid for both bulk and strained Si. A very thin interfacial SiO₂ layer (less than 1nm) is also
expected to form during HfO₂ formation also consuming some of the substrate, however, this amount is negligibly small compared to that consumed during SiO₂ formation. Therefore, for the SiO₂ samples, the final strained Si thickness was found by subtracting the consumed Si thickness from the as-grown strained Si thickness, while for the HfO₂ samples the original as-grown strained Si thickness was used. It can be seen that the extracted Dit is increasing with decreasing strained Si thickness for both SiO₂ and HfO₂, but it is consistently higher for SiO₂ for a given strained Si thickness. The Dit measurements were repeated after RTA and a final FGA, which indicated a slightly lower Dit for both gate dielectrics, but still following the same trend with the Si thickness. A potential explanation for the rise in Dit with decreasing Si thickness is the degradation of the oxide quality due to Ge diffusion into the strained Si layer. The fact that SiO₂ results in a higher Dit than HfO₂ for the same strained Si thickness may be attributed to the higher Ge concentration near the Si/SiO₂ interface due to Si consumption as well as the enhanced Ge diffusion during oxidation. It is well known that the Dit of the SiO₂-SiGe system is much higher than that of the SiO₂-Si system due to larger density of intermediate oxidation states at the SiO₂-SiGe interface. Therefore, the reason that lower Dit was observed with HfO₂ samples than SiO₂ ones with same starting strained Si thickness is due to the less Si consumption as well as the low process temperature.

It is also possible that if the interface between the oxide and strained Si is not perfectly flat, there will be an excess density of suboxide bonds which will increase the Dit. Since both dielectric formation and subsequent RTA temperatures were below 900°C, viscoelastic relaxation could not take place at the interface, preserving the high density of the suboxide bonds due to the rougher surface resulting in a higher Dit. However, the root-mean-square roughness obtained by an atomic force microscope is around 0.6-0.7 nm for
strained Si samples with three different thicknesses, which is comparable as has been reported, indicating no significant correlation between higher $D_{it}$ and decreasing strained Si thickness.

To study the presence of Ge underneath the gate oxide, SIMS was performed on three different samples: (i) as-grown without a gate dielectric or a high-temperature process step, (ii) after growing a 6 nm SiO$_2$ layer by dry oxidation at 850°C for 30 min, and (iii) after HfO$_2$ formation followed by RTA. A Ge tail was observed in the strained Si layer even without SiO$_2$ formation or RTA. This behavior was previously attributed to the lower surface energy resulting from having an adlayer of Ge atoms at the growth surface during Si$_{1-x}$Ge$_x$ epitaxy resulting in a Ge rich surface when the growth is terminated. It was proposed that during growth, the Ge atoms on the growth surface exchange sites with the underlying Si atoms resulting in Ge incorporation in the silicon film. It was also shown that formation of this Ge layer could be suppressed by surfactant-mediated epitaxy, which involves using a different species, to lower the surface energy. Increasing the Ge concentration in a thinner strained Si layer was confirmed by SIMS analysis. It also indicated that during SiO$_2$ formation, Ge concentration in Si increased even further. It has been proposed that Ge diffusion in Si is dominated by the monovacancy mechanism, which could be enhanced due to the vacancy injection during oxidation. X-ray Photoelectron Spectroscopy (XPS) analysis of Ge 3$d$ spectrum was detected in the thinnest strained Si sample with gate oxide removed by HF, as shown in the inset of Fig. 2, demonstrating the presence of Ge in the Si/oxide interface. However, the Ge content at the interface is so low that XPS was not able to detect Ge signals in other two samples.
Another possibility for increasing $D_{it}$ might be related to stress distribution at the Si/dielectric interface. On bulk Si, it has been shown that the midgap $D_{it}$ is proportional to the thickness-averaged stress in SiO$_2$ and Si.$^{27}$ The stress in the SiO$_2$ is compressive and the Si substrate is under tensile stress.$^{27}$ With intentionally grown tensile-strained Si on SiGe, the stress distribution is expected to be less and thus the $D_{it}$ is expected to be less especially for thinner strained Si samples in which higher amount of strain was proved by Raman spectroscopy. Therefore, the increase in $D_{it}$ for decreasing strained Si thickness is mainly attributed to the presence of Ge at the interface.

Strain relaxation could also have an impact on the interface state density, however, using Raman spectroscopy the thermal stability of the strain was confirmed for the annealing conditions used in this study.

The gate leakage current density ($J_g$) of the fabricated MOS capacitors was measured using an HP 4155B Semiconductor Parameter Analyzer. The current density at 1V beyond the flat band with FGA alone, and RTA plus FGA is plotted as a function of the strained Si thickness in Fig. 3. As shown, both HfO$_2$ and SiO$_2$ exhibited increasing leakage current with decreasing strained Si thickness. As discussed above, the presence of Ge in the strained Si channel and even in the dielectric can introduce more traps which may enhance the tunneling current. Similar to $D_{it}$, the leakage current trend with strained Si thickness did not change with RTA. However, an increase in the leakage current was observed for all HfO$_2$ samples including bulk Si, and this is attributed to metal gate/high-$\kappa$ interactions. Previous studies indicated that metal/high-$\kappa$ interactions and metal diffusion through the crystalline high-$\kappa$ grain boundaries may introduce bulk traps, and/or reaction layers that can increase the gate leakage.$^{28}$ A slight decrease in $J_g$ of SiO$_2$ samples was observed with RTA, which may be
due to annealing of defects introduced during metal sputtering. Finally, similar $J_g$ behavior was observed for all metal gates considered in this study.

The authors would like to thank Eric Harley in Dr. McNeil’s group at University of North Carolina at Chapel Hill for Raman analysis. This research was partially sponsored by the National Science Foundation (No. ECS 0301238).
References


FIG. 1. C-V plots of MOS capacitors fabricated on bulk Si and strained Si with SiO$_2$ dielectrics and TaN gate electrodes. The change in the slope near the onset of inversion indicates that $D_{it}$ increases as the strained Si thickness decreases and that the SiO$_2$ samples show higher $D_{it}$.

FIG. 2. Midgap interface trap density ($D_{it}$) vs. strained silicon thickness. The gate electrode is TaN. The inset shows the Ge $3d$ peak detected by XPS in the thinnest strained Si sample after gate oxidation and with the gate oxide removed.

FIG. 3. Leakage current density ($J_g$) vs. strained silicon thickness with and without RTA. Both sets were annealed in forming gas as the last step. TaN was used as the gate electrode.
Figure 1

C/Cox vs. Vg (V) for various Si layers and oxide thicknesses. The graph shows the impact of different Si thicknesses (SSi 7.5nm, SSi 10nm, SSi 12.5nm, SSi 17nm) and Si monitor on C/Cox. The inset highlights the effect of TaN gate at 100kHz and SiO$_2$ ~6nm on the C/Cox curve. The graph indicates that as Vg increases, the C/Cox decreases for all layers.

Key:
- Si monitor
- epi Si
- SSi 12.5nm
- SSi 17nm
- SSi 7.5nm
- SSi 10nm

TaN gate @100kHz
SiO$_2$ ~6nm
After 15min sputtering (~1nm)

Surface

Intensity (A.U.)

Ge3d 29.4eV

Strained Si thickness (nm)

SiO₂

HfO₂

-binding Energy (eV)

8

10

12

0

2

4

6

8

10

12

14

16

18

20

Dₚ(10¹² eV/cm²)

Yanxia Lin et al. Figure 2
Figure 3

[Graph showing the relationship between strained Si thickness (nm) and $J_g$ (A/cm$^2$) @ ($V_{FB}$-1)V for different oxide and HfO$_2$ conditions with and without RTA.]

Yanxia Lin et al.
Chapter 6 Electrical Characterization of Strained Si MOSFETs

As discussed in Chapter 1, strained Si grown on relaxed SiGe is under biaxial tensile, which results in mobility enhancement for both electrons and holes [1]. The incorporation of high-κ dielectrics with strained Si devices provides additional benefits of low gate leakage current and enables further scaling in order to meet the requirements of the International Technology Roadmap for Semiconductors [2]. Concurrent with the dielectric, metal gate electrodes are also being investigated to eliminate gate depletion and Fermi level pinning problems associated with polysilicon electrodes [3, 4]. In this chapter, the performance of strained Si n-MOSFETs integrated with metal gate and polysilicon electrodes and HfO₂ and SiO₂ was evaluated. The roles of various scattering mechanisms on the strained silicon mobility were investigated to gain insight into the scalability of the strained silicon layer.

6.1 Strained Si MOSFETs with SiO₂ Gate Dielectric and Polysilicon or TaN Gate Electrodes

The process flow of strained Si MOS transistor fabrication is summarized in Chapter 2. SiO₂ was thermally grown at 800°C for 10 minutes. Polysilicon (POLY) gate electrodes were deposited by LPCVD and patterned using GEM POLY mask (see appendix B of H. Lazar’s thesis [5]). TaN, as the metal gate electrode, was deposited via UHV reactive sputtering of Ta in 5% N₂ in Ar plasma [6]. C-V and I-V characteristics were obtained using an HP4284a LCR Meter and an HP4155b Semiconductor Parameter Analyzer, respectively. Two level and three level charge pumping measurements were performed using a Keithley 4200 and an HP 8112A.
pulse generator. The Hauser NCSU CVC program [7] was used to extract parameters such as flatband voltage ($V_{FB}$) and effective oxide thickness (EOT). Mobility values were extracted using split C-V method [8]. Hauser MOB2D [9] program was employed for mobility simulation and extraction of electrical parameters of transistors.

### 6.1.1 Basic Device Characteristics: C-V and I-V

For C-V, I-V and split C-V measurements, areas of 50 µm x 50 µm were selected to avoid the effects of overlap and series resistance. Frequencies of 100 kHz were chosen to minimize the interface trap response to the AC signals and avoid the limit of LCR meter resolution. Typical C-V curves of bulk Si and strained Si (16nm) samples with both POLY and TaN gate electrodes are shown in Figure 6-1 (a). As can be seen, negligible difference between the maximum capacitance in inversion region, $C_{inv}$, and accumulation region, $C_{acc}$, is observed for each device, indicating that gate depletion does not affect the measurement at this EOT range for both electrodes. Extracted values for EOT, $V_{FB}$, and $V_t$ are shown in Table 6-1. The devices show similar EOT for both POLY and TaN gates. The difference in $V_{FB}$ and $V_t$ are attributed to the difference in fixed charges, interface charges, the substrate doping and the gate work function between the samples. Using strained Si will also reduce the threshold voltage, as was previously reported [10]. The drain currents of these devices, shown in Figure 6-1 (b), exhibit good device performance, with strained Si devices showing enhanced drain current values.
Figure 6-1  Basic electrical characteristics of bulk Si and strained Si (16nm) MOSFETs with POLY and TaN gate electrodes: (a) C-V curves; (b) drain currents.

Table 6-1  Extracted device parameters for SiO₂ MOSFETs with polysilicon and TaN gate electrodes. Strained Si (SSI) thickness is the starting thickness before gate oxidation.

<table>
<thead>
<tr>
<th></th>
<th>POLY</th>
<th>TaN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bulk Si 16nm</td>
<td>Bulk Si 16nm</td>
</tr>
<tr>
<td>EOT (nm)</td>
<td>5.59</td>
<td>5.56</td>
</tr>
<tr>
<td>VFB (V)</td>
<td>-0.97</td>
<td>-0.76</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td>Nsub (x10¹⁷cm⁻³)</td>
<td>0.32</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>SSI 12nm</td>
<td>SSI 12nm</td>
</tr>
<tr>
<td>EOT (nm)</td>
<td>5.55</td>
<td>5.55</td>
</tr>
<tr>
<td>VFB (V)</td>
<td>-1.07</td>
<td>-0.75</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>-0.19</td>
<td>0.14</td>
</tr>
<tr>
<td>Nsub (x10¹⁷cm⁻³)</td>
<td>3.2</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>SSI 8nm</td>
<td>SSI 8nm</td>
</tr>
<tr>
<td>EOT (nm)</td>
<td>5.57</td>
<td>5.56</td>
</tr>
<tr>
<td>VFB (V)</td>
<td>-1.09</td>
<td>-0.77</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>-0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>Nsub (x10¹⁷cm⁻³)</td>
<td>3.1</td>
<td>2.7</td>
</tr>
</tbody>
</table>
6.1.2 Mobility Extraction

In order to investigate the device performance of strained Si MOSFETs and further understand any degradation in mobility due to the integration of TaN gate electrodes, mobility values were extracted using Split C-V method described in Chapter 3, with drain current correction. Figure 6-2 shows the extracted mobilities for POLY gate and TaN gate on SiO₂. A slight degradation of mobility compared to the universal mobility curve was observed with TaN gate devices and is attributed to the sputtering damage introduced during the metal gate deposition. However, this problem can be tuned by process control since other metal gate processes have provided similar mobility values on SiO₂ as polysilicon [11]. Enhanced electron mobility was achieved with strained Si devices with SiO₂ gate dielectrics, regardless of the gate electrodes. However, a lower degree of mobility enhancement was obtained with thinner strained Si channels and will be discussed later. Nevertheless, these results suggest that strained Si devices on SiO₂ integrated with metal gates are viable for n-channel MOSFETs. Metal gates can also be able to influence the net strain in the channel due to differences in lattice constants and thermal expansion coefficients. The lattice constant of sputtered TaN thin film has been reported to be ~0.4nm depending on the N content [12]. Therefore, it is a valid assumption that TaN gate electrodes will not introduce uniaxial tensile strain to the channel.
Figure 6-2  Mobilities of polysilicon and TaN metal gates on SiO$_2$ dielectrics extracted by Split C-V analysis.

6.2 Strained Si MOSFETs with TaN Gate: SiO$_2$ or HfO$_2$

MOSFETs studied in this section were fabricated with TaN gate electrodes only. Thermal gate oxide was obtained by dry oxidation at 800°C for 8 minutes, while
HfO$_2$ was formed by sputtering Hf at 50 watts for 70 seconds followed by furnace-annealing at 500°C for 5 minutes, as described in Chapter 4. Basic C-V and I-V characteristics were performed and split C-V method was employed to extract the effective mobility. Charge pumping measurements were also carried out to investigate the property of the interface between dielectrics and channels.

### 6.2.1 Basic Device Characteristics: C-V and I-V

C-V curves measured at 100 kHz are shown in Figure 6-3 (a). The inversion equivalent oxide thickness (EOT), as extracted by using Hauser CVC program [7], was found to be 3.5nm and 2.3nm for SiO$_2$ and HfO$_2$, respectively. The EOT values were plotted as a function of strained Si thickness in Figure 6-3 (b). No correlation between the EOT and strained Si thickness was observed. Gate leakage current was plotted in Figure 6-4 as a function of effective dielectric field for bulk Si and one strained Si channel thickness (15nm). A significant reduction of leakage current was observed with HfO$_2$ devices as compared to the direct tunneling model of SiO$_2$ with same EOT, confirming that HfO$_2$ is physically thicker with a higher dielectric constant.
Figure 6-3 (a) C-V curves of TaN nMOSFETs on bulk Si or 20nm strained Si with SiO$_2$ or HfO$_2$; (b) EOT values plotted as a function of strained Si thickness.

Figure 6-3 (a) C-V curves of TaN nMOSFETs on bulk Si or 20nm strained Si with SiO$_2$ or HfO$_2$; (b) EOT values plotted as a function of strained Si thickness.
Figure 6-4  Comparison of gate leakage current for strained Si and bulk Si nFETs with SiO$_2$ and HfO$_2$.

The drain current characteristics of these devices are shown in Figure 6-5. Higher output current was obtained with strained Si devices. However, HfO$_2$/strained Si devices still showed slightly lower $I_{dsat}$ than SiO$_2$/bulk Si, which may be attributed to the additional charge trapping effects as well as the lower mobility in typically observed with high-$\kappa$ dielectrics [23]. Electrical parameters extracted from C-V and I-V data are listed in Table 6-2. It should be noticed that the strained Si thickness used in this section is the final strained Si thickness after oxidation. The variation of $V_{FB}$ observed here is probably due to the fixed charge and interface charge in the high-$\kappa$ dielectric and/or strained Si sample. Due to differences in substrate doping, the actual reduction of $V_t$ is not as much as theoretical calculations [10].
Table 6-2 Extracted device parameters of TaN gate MOSFETs on SiO2 and HfO2.

<table>
<thead>
<tr>
<th></th>
<th>SiO2</th>
<th></th>
<th>HfO2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bulk Si</td>
<td>SSI 18.5nm</td>
<td>SSI 13.5nm</td>
<td>SSI 7.5nm</td>
</tr>
<tr>
<td>EOT (nm)</td>
<td>3.54</td>
<td>3.46</td>
<td>3.42</td>
<td>2.22</td>
</tr>
<tr>
<td>VFB (V)</td>
<td>-0.44</td>
<td>-0.59</td>
<td>-0.61</td>
<td>-0.42</td>
</tr>
<tr>
<td>Vt (V)</td>
<td>0.41</td>
<td>0.17</td>
<td>0.14</td>
<td>0.363</td>
</tr>
<tr>
<td>Nsub x10^17cm^-3</td>
<td>0.3</td>
<td>3.2</td>
<td>3.0</td>
<td>0.34</td>
</tr>
</tbody>
</table>

6.2.2 Mobility Extraction

Effective electron mobility was extracted using split C-V method on large area n-channel MOSFETs (W/L=50µm: 50µm) [13, 14]. Figure 6-6 and Figure 6-7 show the comparison between the effective mobility of bulk Si and strained Si devices with SiO2 and HfO2, respectively. Strained Si devices show expected mobility enhancement compared to bulk Si devices, regardless of the gate dielectrics. The mobility of bulk Si/HfO2 is degraded compared to the universal curve. However, with the incorporation of strained Si, the effective mobility recovered back to the universal value at Eeff >1MV/cm, or even higher. Transistors with thinner strained Si channel
showed less performance enhancement, which was not consistent with the reported electron mobility enhancement as a function of channel thickness [15, 16]. As will be discussed in section 6.3, this may be related to either higher interface trap densities in thinner strained Si channels or other issues with the channel which could be possibly resulted from Ge segregation and out diffusion during the processing.

Figure 6-6  Effective mobility of SiO$_2$ nFETs plotted vs. effective field. Three different strained Si thicknesses were employed. Bulk Si nFET was used as a control.
Figure 6-7 Effective mobility of HfO\textsubscript{2} nFETs plotted vs. effective field. Three different strained Si thicknesses were employed.

6.2.3 Interface and Bulk Traps

The presence of interface traps in strained Si MOS capacitors has been investigated via conductance method in Chapter 4. It was assumed that interfacial states resulted from the inherent nature of 800°C thermal oxidation and HfO\textsubscript{2} as well as the Ge out diffusion from the SiGe virtual substrate. Since the formation of HfO\textsubscript{2} was sputtering process followed by furnace annealing, an HfSi\textsubscript{x}O\textsubscript{y} - like interfacial layer was formed which exhibited higher interface traps than SiO\textsubscript{2}. For thermal oxide grown at 800°C, the viscoelastic relaxation could not take place at the interface, preserving the high density of the suboxide bonds due to the rougher surface resulting in a higher \( D_a \)\textsuperscript{[17]. Therefore, strained Si devices with both SiO\textsubscript{2} and HfO\textsubscript{2} may be affected by interfacial trapping. Both two-level and three-level charge
pumping methods described in Chapter 3 were employed on MOSFETs to determine the interface trap densities (D_{it}). Two-level charge pumping can only give the average D_{it}, while three level charge pumping can provide a distribution of D_{it} as a function of bandgap. MOSFETs with areas of 50 \mu m x 5\mu m were used in charge pumping measurements to avoid additional geometrical currents. Important constants used to extract the average D_{it} are listed in Table 6-3. The capture cross section coefficients \( \sqrt{\sigma_n\sigma_p} \) of similar HfO\textsubscript{2} devices were ascertained from [14] where charge pumping measurements as a function of frequency were used to obtain average capture cross sectional coefficients. The capture cross section values for HfO\textsubscript{2} were more than an order of magnitude higher than those of SiO\textsubscript{2}.

**Table 6-3** Constants used to extract average D_{it} values from two level charge pumping measurement.

<table>
<thead>
<tr>
<th></th>
<th>( \sqrt{\sigma_n\sigma_p} ) [cm(^2)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO\textsubscript{2}</td>
<td>( 5 \times 10^{-16} )</td>
</tr>
<tr>
<td>HfO\textsubscript{2}</td>
<td>( 9.4 \times 10^{-15} ) [18]</td>
</tr>
<tr>
<td>( V_a ) [V]</td>
<td>( 2.0 )</td>
</tr>
<tr>
<td>( V_{th} ) [cm/s]</td>
<td>( 1 \times 10^7 )</td>
</tr>
<tr>
<td>( n_i ) [cm(^3)]</td>
<td>( 1.5 \times 10^{10} )</td>
</tr>
<tr>
<td>( t_r, t_f ) [ns]</td>
<td>( 20 )</td>
</tr>
</tbody>
</table>

Typical base sweep charge pumping charges \((Q_{cp})\) were plotted versus base level bias \((V_{gb})\) for TaN on HfO\textsubscript{2} devices with bulk Si (Figure 6-8 (a)) and strained Si (Figure 6-8 (b)) channels. The extracted average interface charge densities \((D_{it})\) are listed in the plots. \( Q_{cp} \) and average \( D_{it} \) were also plotted versus amplitude biases, which can be seen in Figure 6-9 for both bulk Si and strained Si devices on HfO\textsubscript{2}. The \( D_{it} \) extracted from two level charge pumping data was plotted as a function
of strained Si thickness shown in Figure 6-10. It should be noticed that the average
$D_{it}$ values are similar for measurements performed at different frequencies.

Figure 6-8 Charge pumping currents plotted vs. base sweep biases for TaN/HfO$_2$ devices on (a) bulk Si; (b) 9nm strained Si. The measurements were carried out at 100kHz.
Figure 6-9 Charge pumping currents plotted vs. amplitude sweep biases for TaN/HfO\textsubscript{2} devices on (a) bulk Si; (b) 20nm strained Si. The measurements were carried out at 100kHz.

Figure 6-10 \(D_{it}\) measured from two level charge pumping at 100kHz for nFETs on SiO\textsubscript{2} and HfO\textsubscript{2} with different strained Si thickness.

Since average \(D_{it}\) may overestimate the real value of interface trap density due to the effect that most interface traps are near the threshold where the amount of trapped charge is a large fraction of the total charge, it is very important to locate
the $D_{it}$ distribution as a function of the bandgap [19]. Three level charge pumping measurements were performed on strained Si nMOSFETs with TaN gate electrodes on both SiO$_2$ and HfO$_2$ for the first time. The parameters used to extract $D_{it}$ for the three level charge pumping measurement are listed in Table 6-4 [20]. The measured device areas were 50 $\mu$m x 5 $\mu$m.

**Table 6-4** Parameters used in three level charge pumping to determine the $D_{it}$ of TaN gate devices on SiO$_2$ and HfO$_2$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{step}$ [(\mu s)]</td>
<td>740</td>
</tr>
<tr>
<td>$t_r$, $t_l$ [(\mu s)]</td>
<td>2</td>
</tr>
<tr>
<td>$t_{in}$, $t_{f}$ [(\mu s)]</td>
<td>0.1</td>
</tr>
<tr>
<td>$f$ [Hz]</td>
<td>1343</td>
</tr>
</tbody>
</table>

The plot of the distribution of interface traps with regard to the intrinsic energy level ($E_i$) is shown in Figure 6-11 (a). The extracted $D_{it}$ values were also plotted versus strained Si thickness in Figure 6-11 (b). It is observed that both two level and three level charge pumping techniques provide similar $D_{it}$ values. Furthermore, it is clearly observed that the interface trap densities increase as strained Si thickness decrease, which may imply that increasing both strain and Ge may be responsible for the $D_{it}$ increase. However, as discussed in Chapter 5, the increase of Ge in the channel with decreasing strained silicon thickness is the more dominant mechanism for $D_{it}$ increase. With the same starting strained Si thickness, SiO$_2$ samples show higher $D_{it}$ because of the strained Si consumption during gate oxidation as well as Ge diffusion. Compared to the $D_{it}$ results of strained Si capacitors achieved by the conductance method in Chapter 5, lower interface trap densities were always obtained by using charge pumping methods due to the nature of this technique: the
applied voltages are generated as pulse signals so that only fast traps can respond. Therefore, it will be very unlikely to overestimate the amount of interface traps at a specific energy level.

Figure 6-11  
(a) $D_{it}$ extracted from three level charge pumping as a function of (a) bandgap; (b) strained Si thickness for strained Si MOSFETs with TaN gates on SiO$_2$ and HfO$_2$. 

(b)
6.3 Mobility Degradation Mechanisms in Strained Si MOSFETs with TaN Gate Electrodes

As discussed in the previous sections, mobility enhancement was achieved with strained Si devices compared to bulk Si ones, or to the universal mobility, regardless of the gate dielectrics. Also, as expected, the mobility of bulk Si/HfO$_2$ was degraded as compared to the universal curve. In addition, transistors with thinner strained Si channel showed less performance enhancement. However the percent change in mobility with strained thickness was larger for SiO$_2$ dielectrics as compared to HfO$_2$ dielectrics. Since many factors are responsible for mobility degradation, this section discusses the application of correction techniques suitable for decoupling various parameters in an effort to obtain the true mobility. The advantages and disadvantages of integration of strained Si with high-κ dielectrics and metal gate electrodes can be investigated once these corrections are applied.

6.3.1 Mobility Correction for Interface Traps

All the mobility values shown so far were obtained with the high leakage current correction introduced in Chapter 3. For devices with advanced gate stacks including high-κ dielectrics and/or metal gate electrodes, other corrections may be necessary to calculate the inversion charge and avoid the error on extracting effective mobility. As already discussed, high level of interface traps was observed with strained Si devices with both SiO$_2$ and HfO$_2$. During split C-V measurements, the interface traps can respond to the AC signal so that the inversion charge can be
overestimated. Therefore it is critical to correct the mobilities with regard to interface trap densities.

Two methods can be used to correct mobilities for Dit. A technique developed at Yale uses a calculated theoretical inversion charge to extracted mobility which does not involve the effects of interface traps. The gate to channel capacitance is defined as [21]

$$C_{gc} = \left( \frac{1}{C_{ox}} + \frac{1}{C_{inv}} + \frac{C_d}{C_{ox}C_{inv}} \right)^{-1} \quad (6.1)$$

where $C_{ox}$ is the oxide capacitance and $C_d$ is the depletion capacitance. The inversion capacitance $C_{inv}$ is given as

$$C_{inv} = \frac{dQ_{inv}}{d\psi_s} \quad (6.2)$$

It is assumed the interface traps cannot respond to the high frequency AC signal so that the corresponding capacitance, $C_{it}$, can be negligible. Therefore, corrected inversion charge, $Q_{inv}$, can be obtained by the integration of $C_{gc}$ calculated from equation (6.1) and used to extract the mobility without knowing the values of Dit. This theory is similar to the Hauser MOB2D model developed at NCSU [9]. The other method to correct mobility for Dit is to include the interface trap capacitance $C_{it}$. If the value of Dit is known as a function of surface potential ($\psi_s$), then $C_{it}$ can be determined by [5]

$$C_{it} = q \frac{dD_{it}}{d\psi_s} \quad (6.3)$$

and the gate to channel capacitance is defined as
An accurate inversion charge can be acquired by integrating $C_{gc}$ from (6.4) which can be used to extract the mobility.

Both methods were used to correct the mobility for $D_{it}$ and similar mobility values were achieved with the Yale method and the $D_{it}$-C$_{it}$ method, as shown in Figure 6-12. Corrected effective mobility was plotted as a function of $E_{eff}$ in Figure 6-13 (a) and Figure 6-13 (b). The mobility enhancement factor $\gamma$ is defined as the ratio of the maximum mobility of strained Si devices to that of bulk Si device. As shown in Figure 6-14, corrected mobility of transistors with SiO$_2$ shows larger deviation from uncorrected values, indicating the effect of a higher $D_{it}$. However, even the corrected mobility values do not recover back to the expected level of 80% electron mobility enhancement, especially for the thinnest strained Si samples.
Figure 6-12  Effective electron mobility after $D_{it}$ corrections for TaN strained Si MOSFETs on SiO$_2$ and HfO$_2$. 

SiO$_2$: 3.5nm  
TaN/SiO$_2$/Strained Si 18.5nm  

HfO$_2$: 2.3nm  
TaN/HfO$_2$/Strained Si 20nm
Figure 6-13  Effective electron mobility extracted with $D_n$ corrections for nMOSFETs on (a) SiO$_2$ and (b) HfO$_2$. 
6.3.2 Mobility Degradation Related to High-κ Dielectrics

Now that the \(D_{\text{it}}\) corrected mobilities are obtained, the results can be further separated out into various components. In general, MOSFET mobility can be expressed as following, based on Matthiessen’s rule [13, 14]:

\[
\frac{1}{\mu} = \sum_i \frac{1}{\mu_i}
\]  \hspace{1cm} (6.5)

For typical Si/SiO\(_2\) devices, the mobility is limited by Coulomb scattering, phonon scattering and surface roughness scattering, with the lowest one dominating. For high-κ transistors, additional mobility degradation mechanisms may dominate. Possible sources of mobility degradation in high-κ gate stacks are illustrated in Figure 6-15 [22]. Various scattering mechanisms have to be considered for the mobility reduction, including remote phonon scattering and scattering resulted from...
fixed charges, surface roughness and phase separation, etc. The remote phonon scattering is reported to be unavoidable for high-κ system [23], while other types of scattering may be eliminated by improving the process.

By using the following model:

\[
\frac{1}{\mu_{\text{high }K\text{-limited}}} = \frac{1}{\mu_{\text{high }K/\text{bulk Si or SSi}}} - \frac{1}{\mu_{\text{SiO}_2/\text{bulk Si or SSi}}} \tag{6.6}
\]

the HfO\(_2\) limited mobility component was calculated and plotted as a function of effective field in Figure 6-16 for each strained silicon thickness. No obvious dependence of high-κ limited mobility on strained Si thickness was observed. It has been shown in Figure 6-3 (b) that there is no correlation between the EOT and strained Si thickness either. Therefore, no additional degradation mechanisms were introduced by the integration of high-κ with strained Si. However, the decrease of mobility with decreasing strained silicon thickness is still present and will be addressed next.

Figure 6-15 Possible sources of scattering in high-κ gate stacks [22].
6.3.3 Mobility of Strained Si MOSFETs at Higher Temperatures

In order to understand the degradation mechanisms under different temperature and different applied fields, Split C-V analysis was performed at higher temperatures. As is known, different scattering mechanisms show different temperature dependence [24-26]. Coulombic scattered mobility increases with increasing temperature while phonon scattering decreases as the temperature increases. The surface scattering limited mobility is independent of temperature. Therefore, at higher temperature, the effective mobility will decrease and be controlled by the phonon scattering limited mobility component at both low and intermediate effective fields. All split C-V measurements were performed on bulk Si or strained Si MOSFETs with SiO\(_2\) at room temperature (20°C), 100°C and 150°C.
The bulk Si control sample and two strained Si samples with HfO$_2$ were also measured to compare the temperature-dependent mobility degradation mechanisms of high-$\kappa$ and SiO$_2$.

![Figure 6-17](image_url)

**Figure 6-17** Peak mobility plotted as a function of temperature for TaN gate MOSFETs on SiO$_2$ and HfO$_2$ dielectrics with bulk Si and/or strained Si channels.

The peak values of extracted mobility are plotted as a function of temperature in Figure 6-17 for SiO$_2$ and one HfO$_2$ device. All samples show mobility degradation as temperature increases. However the slopes of those samples are noticeably different. It can be seen that for bulk Si samples, the slope on SiO$_2$ is larger than that on HfO$_2$, which has been reported with HfO$_2$ with Al gates [24]. In reference [21], this behavior is claimed to be attributed to the low-energy soft optical phonons in the HfO$_2$ gate dielectric so that the optical phonon scattering has a weak dependence on temperature. Hence the mobility component limited by soft phonon scattering shows less dependence on temperature. It can also be explained by other mechanisms like
interface scattering which may be dominant in this case. For all SiO\textsubscript{2} samples, strained Si devices demonstrate smaller slopes than that of bulk Si. This mobility enhancement with strained Si comes primarily from the decrease in intervalley scattering and the reduced transverse effective mass. The intervalley phonon scattering is more important at higher temperatures, therefore the mobility of bulk Si control decreases more rapidly than that of strained Si ones, indicating that alternate scattering mechanisms may be dominating. The slope of bulk Si with SiO\textsubscript{2} is larger than strained Si with SiO\textsubscript{2}, as seen in Figure 6-17. An even more reduced slope with temperature was observed with HfO\textsubscript{2} on strained Si samples, which can also be attributed to soft optical phonons. However, it can also be explained by other scattering mechanisms, for example, significant interface scattering. More discussions will be included in the following sections.

6.3.4 The Impact of Ge on Mobility Degradation

The Coulombic component of mobility is comprised of scattering from ionized impurities in the channel and charges located at the interface states. Theoretical calculation of the Coulombic scattering can be found in literature [27, 28]. It is hypothesized that Coulombic scattering in bulk Si is same as in strained Si [29], which was found to be consistent with experimental data [30]. The surface roughness component of mobility depends on the inverse square of the effective field [13, 14, 27][22, 26, 27]. Therefore, it will only affect the high field mobility and has little impact on the mobility in the effective field range studied in this work. Subtracting the Coulombic component and surface roughness component from the extracted mobility with D_{it-C_{it}} correction leaves the mobility due to phonon scattering.
for bulk Si and strained Si devices and can be seen in Figure 6-18. Compared to theoretical phonon-limited mobility values in strained Si [31, 32], thinner strained Si MOSFETs show lower phonon-limited components. Since ~0.8% biaxial strain can introduce sufficient band splitting in the conductance-band, all carriers will occupy the lower-energy two-fold degenerated subbands ($\Delta_2$) and the intravalley phonon scattering can be completely suppressed. Therefore the electron mobility enhancement will then saturate even with higher strain [31]. The lowest strain level in this study is ~0.7% for the strained Si 18.5nm with SiO$_2$ device. Therefore similar phonon-limited mobility should be expected for all strained Si devices, while the thickest strained Si devices should show slightly lower values due to less amount of strain. An opposite trend was obtained in Figure 6-19, indicating that there are other components limiting the improvement of carrier mobility. Also, as discussed before, even after $D_{lt}$ corrections, the mobility did not recover. Another responsible mechanism could be Ge diffusion into the channel, as discussed in Chapter 5, resulting in changes of bulk lattice scattering, as well as additional surface roughness scattering and defect scattering. Thus we can redefine this "phonon-limited" mobility component as the reciprocal of the sum of $1/\mu_{\text{phonon}}+1/\mu_{\text{Ge}}$. The impact of $\mu_{\text{Ge}}$ is more significant at $E_{\text{eff}} > 0.6\text{MV/cm}$, which indicates a possible similarity with $\mu_{\text{SR}}$ (the surface roughness scattering limited mobility component). Further investigation is needed to fully understand this degradation mechanism.
Figure 6-18 D_{it}-corrected effective mobility and calculated phonon limited mobility using Matthiessen's rule for bulk Si and 20nm strained Si nMOSFETs on HfO₂.

Figure 6-19 “Phonon-limited” effective mobility components of n-channel devices plotted as a function of effective field on bulk Si and strained Si. Theoretical values calculated in Ref. A [32] and ref. B [31] are also included.
6.3.5 Understanding Scattering Mechanisms in strained Si devices

As shown in this chapter, strained Si devices with thinner strained Si channels on both SiO$_2$ and HfO$_2$ show less mobility enhancement. However, electron mobility enhancement factor is expected to be 1.7~1.8 for strained Si channels thicker than 5nm, which confines most electrons in the strained Si channel [15, 16]. To investigate that which mobility components dominate the mobility degradation, a least squares curve fitting technique developed by Hauser at NCSU was employed to separate out individual mobility components [9]. Figure 6-20 shows the electron mobility extracted from experimental data of the transistor with SiO$_2$ dielectric and TaN gate electrode on 13.5 nm strained Si channel, as well as four mobility components simulated via Hauser’s MOB2D model: bulk impurity limited mobility component, $A$, interface scattering limited mobility component, $B$, surface phonon limited component, $C$, and surface roughness limited component, $D$. All these components were obtained from the simulation using strained Si device parameters, such as $V_t$, EOT and surface doping density, and physical parameters of bulk Si, such as effective mass, Si spring constant and typical surface roughness of SiO$_2$ on bulk Si (24Å$^2$). By changing certain simulation modes, the MOB2D model can be used to optimize some physical parameters such that a better curve fitting could be achieved. All mobility components were separated out and labeled with subscripts of $s$, referring to strained Si. The bulk impurity limited mobility, i.e. the Coulomb scattering limited mobility, remained same for both devices. All other three mobility components were expected to improve for strained Si devices due to the reduction in in-plane electron effective mass resulted from tensile strain. As shown in Figure 6-20,
compared to those of bulk Si, both the phonon and surface roughness mobility terms improved, however, a significant decrease in the interface scattering component was observed with strained Si.

Figure 6-20  The phonon, surface roughness and interface scattering mobility components of a strained Si sample simulated by using Hauser’s MOB2D model[9] compared to those of a bulk Si sample.

Interface scattering components can be affected by \( D_{ii} \), fixed charge, compensated charge and scattering at the Si-SiGe interface. Due to the Ge segregation near the interface of gate dielectric and the strained Si channel, the dielectric integrity may also be affected, especially for SiO\(_2\) which was formed through a process with higher thermal budget and more Si consumption than that of HfO\(_2\). It has been reported that Ge will not incorporate within the thermal oxide but...
only act as a catalyst during thermal oxidation if the Ge concentration is high enough [33]. Therefore, unreacted excess Ge is expected to be remained piled-up near the dielectric interface, which explains the higher level of interface traps observed with strained Si samples. As shown in Figure 6-14, although the $D_{it}$ correction increased the mobility enhancement value for a given strained Si thickness, the correlation between the mobility and strained Si thickness still persisted. Therefore the decrease in mobility with decreasing strained Si thickness cannot be solely attributed to $D_{it}$. Scattering from the Si-SiGe interface was not considered to be a major effect either since the strained Si layers were thicker than the inversion layer thickness. To determine the value of fixed charge, a fixed thickness of SiO$_2$ (~11nm) was grown and etched back to various thicknesses (8nm, 6nm, etc.). By using this technique, the Si-SiO$_2$ interface can be ensured to be same for these samples with varying dielectric thickness, therefore, fixed charge will be calculated using the slope of the plot of $V_{FB}$ vs. EOT. Fixed charge densities of $3 \times 10^9$ cm$^{-2}$ and $1 \times 10^{13}$ cm$^{-2}$ were measured for bulk Si and 20nm strained Si sample, respectively. Thinner strained Si samples are expected to have higher fixed charges because Ge out diffusion will have stronger impact. The value of fixed charge of strained Si sample is high enough to dominate the interface scattering component and lower the mobility of thinner strained Si samples. However, high $D_{it}$s were observed in other strained Si samples and larger errors were introduced when extracting $V_{FB}$ and EOT by using Hauser CVC model [7]. Thus accurate fixed charge values of thinner strained Si samples were unavailable. As shown in Figure 6-17, a reduced slope with temperature was observed with SiO$_2$ on strained Si as well HfO$_2$ on bulk Si. An even
more reduced slope with temperature was observed with HfO\textsubscript{2} on strained Si samples. The reduced slope with temperature can be attributed to soft optical phonons and the intervalley phonon scattering. It is also possible to explain the reduced slope by high values of interface scattering [5]. Based on the simulated data shown in Figure 6-21, a significant reduction in the dependence of mobility on temperature is observed with an interface scattering parameter of $1.5 \times 10^{11}$, which indicates that the mobility degradation in HfO\textsubscript{2} on both bulk Si and strained Si samples could be due to a major impact of interface scattering.

![Figure 6-21 Modeled peak mobility plotted vs. temperature with different interface scattering density (I.S.) and surface roughness (S.R.) parameter values extracted from the Hauser NCSU MOB2D model [9] and Lazar’s Thesis [5].](image)

Ge diffusion into the strained Si channel will result in additional limitations on mobility enhancement due to potential increase in alloy scattering and degradation of the heterostructures [34-36]. Besides the higher $D_r$ observed with strained Si
samples which is related to Ge segregation near the dielectric interface, Ge diffusion can also cause intermixing between the virtual substrate and the channel, which makes alloy scattering possible. Both electron and hole mobilities of SiGe are lower than those of Si and Ge, which is attributed to the alloy scattering in bulk unstrained SiGe samples [37]. For thinner strained Si layers, more Ge may diffuse into the channel, resulting in higher possibility of mobility degradation caused by alloy scattering. In addition to the Ge impact, mobility enhancement may also be limited by the potential increases in surface roughness or material relaxation occurring through the misfit dislocations, which can act as scattering centers [38] and leakage paths [39]. Scattering resulted from strain variation should also be concerned.

Finally, whether metal screening is occurring can only be determined by comparing the mobility of a metal gated MOSFET with that of a polysilicon gate on ultra thin dielectrics since no screening on any optical phonons should be observed with polysilicon electrode due to its large screening length. TaN gates did not show any impact on the mobility on either SiO₂ or HfO₂ in this work. No additional correlation between the carrier mobility of strained Si and the metal gate electrodes was observed either. However, due to the EOTs of the gate dielectrics studied here, it is unclear if phonon screening is occurring. More investigation on devices with aggressively scaled dielectrics is necessary.

6.4 Summary

In summary, strained Si nMOSFETs with TaN gate electrodes and either SiO₂ or HfO₂ gate dielectric have been investigated. Similar mobility enhancement was observed with strained Si devices with TaN gate electrodes as that of devices with
polysilicon gates. Therefore, TaN gate electrodes do not appear to have an impact on the channel strain or the effective electron mobility. Based on the similar electrical properties of HfO₂ dielectrics on bulk Si and strained Si, strain does not introduce any degradation of the high-κ/strained Si interface. Higher strain level was confirmed by Raman for thinner strained Si samples. However, less mobility enhancement was achieved. The presence of Ge at the interface and even in the gate dielectric is the major cause of this behavior, which results in degradations of electrical properties including higher D_it, higher fixed charge, higher interface scattering and less device performance enhancement. These degradations are less severe with HfO₂ samples because the high-κ dielectric is formed at a lower temperature.
6.5 References


Chapter 7 Summary and Future Work

7.1 Conclusions

A thin Si layer grown on relaxed Si_{1-x}Ge_{x} is under biaxial tensile strain and it can provide mobility enhancement for both electrons and holes. Incorporation of high-κ dielectrics and metal gate electrodes on strained silicon is necessary to fulfill the continuous scaling requirements with the additional benefit of low gate leakage current, as well as the elimination of poly depletion and Fermi level pinning issues. In this work, strained-Si MOS capacitors and MOSFETs were fabricated with SiO_{2} and HfO_{2} as gate dielectrics and Ru-Ta alloy and TaN as metal gate electrodes. Strained Si layers were grown on relaxed SiGe virtual substrates by ultrahigh vacuum rapid thermal chemical vapor deposition (UHV/RTCVD). Deposition conditions were optimized to achieve best selectivity and desirable growth rate. HfO_{2} was deposited by physical vapor deposition (PVD) of thin Hf layers followed by annealing at 500 °C in N_{2} for 5 minutes. The physical thickness of HfO_{2} and the interfacial layer were determined by TEM as 2.1nm and 9nm, respectively. The composition of the interfacial layer is Hf silicate (HfSi_{x}O_{y}), which was confirmed by XPS. Different amounts of strain were achieved by varying the strained Si thickness. For a given starting strained Si thickness, SiO_{2} samples show higher strain level than HfO_{2} samples, which is due to Si consumption during oxidation. After RTA, no change in strain level was observed for HfO_{2} samples, while for SiO_{2} samples slight decrease in strain was observed, which suggests that partial relaxation after RTA may be occurring and/or that Ge is being incorporated into the strained Si layer.
The results of strained Si MOS capacitors indicate that the interface trap density \( (D_{it}) \) increased as the strained silicon thickness decreased, which was attributed to the presence of Ge in the strained Si layer. Higher \( D_{it} \) was observed with SiO\(_2\) which may be due to Si consumption during oxidation, leading to a higher density of Ge at the interface. Leakage current density \( (J_g) \) and hysteresis were also observed to increase with decreasing strained silicon thickness. This trend of increasing \( D_{it} \) and \( J_g \) with decreasing strained silicon thickness did not change after rapid thermal annealing. Both Ru-Ta and TaN gate electrodes were found to exhibit as good performance on strained Si as on bulk Si. Ge segregation and out-diffusion into the strained Si channel was proved by both SIMS and XPS. For a given amount of strain, the rate of increase in \( D_{it} \) with increasing strained Si thickness is significantly more than the increasing rate of \( D_{it} \) as the Ge concentration in the virtual substrate increases, which suggests that Ge diffusion is the dominant cause of the \( D_{it} \) increase.

Several conclusions can be drawn from the investigation of strained Si nMOSFETs with TaN gate electrodes and either SiO\(_2\) or HfO\(_2\) gate dielectric. The electrical properties of HfO\(_2\) dielectrics on strained Si are similar to those on bulk Si. Strain does not lead to any degradation of the high-\( \kappa \)/strained Si interface. The lattice constant of TaN is smaller than that of Si, thus it can not introduce additional carrier enhancement. MOSFETs with TaN gate electrodes show similar mobilities as those with polysilicon gates. Therefore, TaN gate electrodes do not appear to have an impact on the channel strain or the effective electron mobility. Although high amount of strain was confirmed by Raman, thinner strained Si samples show less
mobility enhancement. The major contributor of this behavior is the presence of Ge at the interface and even in the gate dielectric, which results in degradations of electrical properties including higher $D_{it}$, higher fixed charge, higher interface scattering and less device performance enhancement. These degradations are less severe with HfO$_2$ samples because the high-$\kappa$ dielectric formation is a low thermal budget process.

7.2 Future Work

The integration of high $\kappa$ dielectric and TaN gate electrode on strained Si channel for n-type transistor is implemented. However, more work has to be done to thoroughly understand the mobility degradation mechanism and the impact of metal gate electrodes on strained Si devices.

First, more investigation of the epitaxial strained Si films is necessary to monitor the quality of strained Si channel as well as the accurate thickness. Plan-view and cross-sectional TEM would be very helpful. Commercial strained Si wafers can also be used as controls.

Second, fabricating strained Si devices without having SiGe in proximity will be able to provide an effective way to separate the potential impacts of strain and Ge out-diffusion on mobility degradation. Both uniaxial strained Si devices where strain is introduced by process induced stress instead of the presence of Si$_{1-x}$Ge$_x$ alloy and strained Si on insulator (SSOI) devices can be best candidates for this proposal if feasible. In addition, it will be necessary to investigate the potential impact of Ge diffusion in the short-channel strained Si devices, even nanoscale devices as FINFETS, where strain is introduced from the SiGe junctions.
Third, since TaN is known to have smaller lattice constant than Si, it is impossible to study the possibility of carrier enhancement caused by metal gate electrodes. Therefore, potential metal gates such as elemental metal or metal alloys should be chosen to integrate with strained Si to investigate the integration of metal gates with strained Si devices.

Finally, more advanced characterizations of mobilities can be employed to provide a better understanding on mobility degradation mechanisms. For example, mobility measurements at lower temperatures can rule out the contribution of bulk phonon mobility. Using pulsed measurements can also locate the impact of bulk traps on mobility. Fabrication of p-channel strained Si transistors has yet to be included in order to achieve a full view of implementation of strained Si devices integrated with advanced gate stacks.