Gate Stack And Channel Engineering: Study Of Metal Gates And Ge Channel Devices

2007

Ravi Todi

University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Electrical and Electronics Commons

STARS Citation

https://stars.library.ucf.edu/etd/3387

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
GATE STACK AND CHANNEL ENGINEERING:
STUDY OF METAL GATES AND Ge CHANNEL DEVICES

by

RAVI M. TODI
B.E.E.E. Mumbai University, 2002
M.S.E.E., University of Central Florida, 2004
M.S.M.E., University of Central Florida, 2005

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the School of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2007

Major Professors: Kevin R. Coffey
Kalpathy B. Sundaram
ABSTRACT

The continued scaling of device dimensions in complementary metal oxide semiconductor (CMOS) technology within the sub-100 nm region requires an alternative high dielectric constant (high-κ) oxide layer to counter high tunneling leakage currents, a metallic gate electrode to address polysilicon depletion, boron penetration and high polysilicon sheet resistance, and high mobility channel materials to boost the CMOS performance. Metal gates can also offer improved thermal and chemical stability, but their use requires that we improve our understanding of how the metal alloy phase, crystallographic orientation, and composition affect the electronic properties of the metal alloy-oxide interface. To replace n++ and p++ polysilicon gate electrodes and maintain scaled device performance requires metal gate electrodes with work functions within 0.2 eV of the silicon conduction and valence band edges, i.e., 5.0-5.2 and 4.1-4.3 eV, for PMOS and NMOS devices, respectively. In addition to work function and thermal/chemical stability, metal gates must be integrated into the CMOS process flow. It is the aim of this work to significantly expand our knowledge base in alloys for dual metal gates by carrying out detailed electrical and materials studies of the binary alloy systems of Ru with p-type metal Pt. Three n-type metals systems, Ru-Ta, Ru-Hf and Ru-Nb have also been partially investigated. This work also focuses on high mobility Ge p-MOSFETs for improved CMOS performance.

DC magnetron sputtering has been used to deposit binary alloy films on thermally grown SiO₂. The composition of the alloy films have been determined by Rutherford backscattering spectrometry and the identification of phases present have been made using x-ray and electron diffraction of samples. The microstructure of the phases of interest has been examined in the
transmission electron microscope and film texture was characterized via x-ray diffraction. The electrical characterization includes basic resistivity measurements, and work function extraction. The work function has been determined from MOS capacitor and Schottky diodes.

The need for electron and hole mobility enhancement and the progress in the development of high-\(\kappa\) gate stacks, has lead to renewed interest in Ge MOSFETs. The p-MOS mobility data for Ge channel devices have been reported. The results indicate greater than 2 x improvements in device mobility as compared to standard Si device. A low frequency noise assessment of silicon passivated Ge p-MOSFETs with a TiN/TaN/HfO\(_2\) gate stack has been made. For the first time we also report results on low frequency noise characterisation for a Ge P\(^+\)-n junctions with and without Ni germanidation.
dedicated to my grandparents and parents

Shri Ramlal Todi & Late Smt. Mohini Devi Todi

Shri Mahendra Kumar Todi and Smt. Kusum M. Todi

who have always been inspirational sources in helping me realize my goals
ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge and thank my advisors Dr. Kalpathy B. Sundaram and Kevin R. Coffey for their guidance and support which was crucial for the completion of this work. Thanks to Prof. John Shen and Prof. J. S. Yuan for serving on my dissertation committee and for their valuable advice and comments that helped me improve the quality of my dissertation. I would like to acknowledge Prof. Katayun Barmak from Department of Materials Science and Engineering at Carnegie Mellon University for valuable discussions from time to time. Special acknowledgement is due to my promoter at K. U. Leuven Prof. Cor Claeyss, for facilitating my summer internship at IMEC. I also thank Ge III-V program manager Dr. Marc Meuris for his support and guidance during my stay at IMEC.

I am grateful to Edward Dein and Kevin Casey for giving me my first lessons in the cleanroom and for helpful suggestions at all times. I also wish to acknowledge advanced microfabrication facility (AMF) and the materials characterization facility (MCF) at the advanced materials processing and analysis center (AMPAC) for laboratory support at the University of Central Florida (UCF). I gratefully acknowledge the faculty and staff at AMPAC, the School of Electrical Engineering and Computer Science and the College of Engineering and Computer Science at UCF for academic support for nearly five years.

I would like to thank my current and former colleagues at UCF; Mr. Arun Vijayakumar, Mr. Sumant Sood, Mr. Parag Gadkari, Mr. Andrew Warren, Dr. Rumyana Petrova, Mr. Vu Lam, Ms. Chaitali China, Mr. Bo Yao and Vinit Todi, who helped me tremendously throughout all the experiments. I would also like to acknowledge my colleagues at IMEC; Dr. Gareth Nicholas, Mr. Jan Van Steenbergen, Mr. Brice De Jaeger, Dr. Paul Zimmerman and Mr. Sushant Sonde for
their help with experiments while at IMEC. Special thanks are due to my daily mentor Dr. Eddy Simon for being patient with me and for his tremendous scientific and technical help.

At this time I would also like to acknowledge all my peers for their valuable suggestions at all times. I would like to acknowledge the student government association and the division of graduate studies at UCF for numerous travel awards that helped me present research at various international conferences. I would also acknowledge professional organizations like IEEE, AVS and ECS for providing me with the opportunity for professional development. I

Completing a Ph.D. dissertation is a humongous task that I would have not been able to complete without the love and moral support of my parents, family and friends. Finally, with deep gratitude, I would like to thank my spiritual Guru for his divine guidance at all times.
# TABLE OF CONTENTS

LIST OF FIGURES ....................................................................................................................... xi

LIST OF TABLES .......................................................................................................................... xvi

LIST OF ACRONYMS/ABBREVIATIONS .................................................................................. xvii

CHAPTER ONE: INTRODUCTION ................................................................................................ 1

CHAPTER TWO: LITERATURE REVIEW ....................................................................................... 4
  2.1 CMOS and Its Scaling ............................................................................................................ 4
    2.1.1 Conventional CMOS Scaling ......................................................................................... 6
    2.1.2 Non Conventional CMOS Scaling ................................................................................ 10
  2.2 High –κ Dielectrics ................................................................................................................ 18
    2.2.1 Gate Stack Compatibility ............................................................................................... 23
  2.3 Metal Gates ............................................................................................................................ 26
    2.3.1 Work Function ............................................................................................................... 28
    2.3.2 Thermal and Chemical Stability .................................................................................... 29
    2.3.3 High conductivity and high active electrical concentration ........................................ 30
    2.3.4 Compatibility with CMOS integration .......................................................................... 31
    2.3.5 Deposition technique .................................................................................................... 32
  2.4 Germanium Devices .............................................................................................................. 34
    2.4.1 Mobility Enhancement in Silicon .................................................................................... 36
    2.4.2 Germanium: Challenges and Opportunities .................................................................. 39

CHAPTER THREE: METHODOLOGY ......................................................................................... 43
  3.1 Experimental Study: Metal Gates ......................................................................................... 43
3.1.1 Selection of Metal Alloys ............................................................................................ 45
3.1.2 Film deposition ............................................................................................................ 53
3.1.3 Materials characterization ............................................................................................ 56
3.1.4 Electrical characterization ............................................................................................ 56
3.2 Experimental Study: Ge Channel Devices ........................................................................ 60
  3.2.1 Device Fabrication ....................................................................................................... 61
  3.2.2 Device Characterization ............................................................................................... 65
    3.2.2.1 Mobility Extraction ............................................................................................... 65
    3.2.2.2 Low Frequency Noise Measurements .................................................................. 67
CHAPTER FOUR: RESULTS AND DISCUSSIONS: METAL GATES ........................................ 69
  4.1 Materials Characterization ................................................................................................ 69
    4.1.1 Composition and thickness ........................................................................................ 69
    4.1.2 Thermal Treatment: Agglomeration ......................................................................... 72
    4.1.3 Microstructure and phase identification ..................................................................... 75
  4.2 Electrical Characterization ................................................................................................ 80
    4.2.1 Resistivity .................................................................................................................... 80
    4.2.2 Work function: MOS C-V .......................................................................................... 81
    4.2.3 Work function: Schottky I-V ....................................................................................... 84
    4.2.4 Work function: Schottky C-V ..................................................................................... 88
    4.2.5 Barrier Height: Schottky temperature – I-V ................................................................. 92
CHAPTER FIVE: RESULTS AND DISCUSSIONS: Ge CHANNEL DEVICES ......................... 98
  5.1 Ge PMOSFET Mobility Data ............................................................................................. 98
  5.2 LF Noise Assessments ..................................................................................................... 103
5.2.1 LF Noise results for Ge p-MOSFETs ................................................................. 104
5.2.2 LF Noise Results for Ge p⁺-n Junctions ......................................................... 108

CHAPTER SIX: CONCLUSION .................................................................................. 114
6.1 Dissertation Remarks ....................................................................................... 114
6.2 Outlook ............................................................................................................... 116

LIST OF REFERENCES ............................................................................................. 118
LIST OF FIGURES

Figure 2.1: Complementary MOS (CMOS) inverter. (a) Circuit diagram. (b) Circuit layout. (c) Cross section along dotted A-A’ line of (b). .......................................................... 5

Figure 2.2: Energy diagram of threshold voltages for NMOS and PMOS devices using (a) midgap metal gates and (b) dual metal gates. .................................................. 25

Figure 2.3: Selection criteria for metal as alternate gate electrode for CMOS devices......... 28

Figure 2.4: The first point contact transistor invented in 1947 at Bell laboratories. .......... 35

Figure 2.5: First integrated circuit invented in late 1950s. ...................................... 35

Figure 2.6: (a) HOT implemented on SOI (b) HOT implemented on bulk silicon. .......... 38

Figure 3.1: Possible metal choices and their classification as NMOS, PMOS or Midgap based on the elemental workfunction ................................................................. 47

Figure 3.2: Enthalpies per mole of oxygen for the most stable oxides for gate metal materials. Range (± 10%) for the enthalpies of Al₂O₃ and HfO₂ are indicated for comparison. ..... 48

Figure 3.3: Melting point (°C) for the gate metal materials. ....................................... 49

Figure 3.4: Pt-Ru binary alloy phase diagram. ......................................................... 50

Figure 3.5: Ru-Ta binary alloy phase diagram. ....................................................... 52

Figure 3.6: Ru-V binary alloy phase diagram ......................................................... 52

Figure 3.7: Nb-Ru binary alloy phase diagram ....................................................... 53

Figure 3.8: Target voltage versus deposition power plot .............................................. 55

Figure 3.9: Energy Band diagram of a MOS system, (a) V_G = 0, (b) Flatband .............. 58

Figure 3.10: Schematic of the final multi-layer gate dielectric stack with epitaxial Si interfacial layer ........................................................................................................... 62
Figure 4.1: RBS spectrum for Pt-Ru film indicating a film thickness of 270 Å with at.% of 50-50 for Pt and Ru.

Figure 4.2: Composition (by RBS) of Ru-Pt alloys as a function of Pt percent power.

Figure 4.3: Composition (by RBS) of Ru-Ta alloys as a function of Ta percent power.

Figure 4.4: Composition (by RBS) of Ru-Nb alloys as a function of Nb percent power.

Figure 4.5: Composition (by RBS) of Ru-V alloys as a function of V percent power.

Figure 4.6: SEM images of 20nm of Ru thin film annealed in Ar+3%H2 environment: (a) as-deposited before annealing; (b) 600 °C; (c) 700 °C; (d) 900 °C.

Figure 4.7: SEM images of 20 nm of Pt thin film annealed in Ar+3%H2 environment: (a) as-deposited before annealing; (b) 450 °C; (c) 750 °C; (d) 900 °C.

Figure 4.8: Figs. (a) and (c) show bright field TEM images of the 50 nm thick Pt and Ru film samples before 400 °C annealing and figs. (b) and (d) shows TEM results after 400 °C annealing.

Figure 4.9: X-ray diffraction of the as deposited Pt-Ru thin films.

Figure 4.10: X-ray diffraction of the annealed Pt-Ru thin films.

Figure 4.11: Peak position (2θ) for the as deposited and annealed Pt-Ru films are plotted as a function of Pt composition.

Figure 4.12: Resistivity as a function of at. % Pt concentration for as-deposited and annealed Ru-Pt alloys.

Figure 4.13: The C-V data for Ru-Pt alloy films as gate electrode with a 10.5 nm SiO2 as gate dielectric.

Figure 4.14: Flatband voltage vs. oxide thickness for different Ru-Pt alloy compositions.

Figure 4.15: Ru-Pt alloy work function plotted as a function of at. % Pt.
Figure 4.16: Semilog plot of I-V for Pt-Ru alloy Schottky diodes

Figure 4.17: Barrier height extracted from the I-V data of the Pt-Ru Schottky diode, plotted as a function of at. % Pt.

Figure 4.18: Work function values obtained by Schottky I-V plotted as a function of at. % Pt.

Figure 4.19: C-V data for Pt Schottky diodes with four different areas.

Figure 4.20: $1/C^2$ vs. V plot for Pt-Ru alloy Schottky diode measured at room temperature.

Figure 4.21: Work function values obtained by Schottky C-V plotted as a function of at. % Pt.

Figure 4.22: Work function values obtained by MOS C-V, Schottky I-V and Schottky C-V plotted as a function of at. % Pt.

Figure 4.23: I-V characteristics for 30%Pt and 70%Ru alloy in a 300 µm$^2$ Schottky device at different temperatures.

Figure 4.24: Extracted barrier heights for (a) 100% Pt and (b) 100% Ru at increasing temperatures for 300 µm$^2$, 400 µm$^2$, 500 µm$^2$ and 600 µm$^2$ area Schottky devices.

Figure 4.25: Extracted barrier heights for Pt-Ru alloy compositions with temperature for different area (300 µm$^2$, 400 µm$^2$, 500 µm$^2$ and 600 µm$^2$) devices, (a) 50%Ru, (b) 70%Ru and (c) 80%Ru.

Figure 4.26: Extracted barrier heights for different temperatures (25°C, 50°C, 75°C, 100°C and 125°C) and areas (300 µm$^2$ to 600 µm$^2$) plotted as a function of at. % Ru concentration.

Figure 5.1: C-V data for a 10 µm X 10 µm PMOS obtained at 100 kHz, showing excellent characteristics and an EOT of 1.2 nm.

Figure 5.2: Linear $I_{ds}$-$V_g$ and $G_{m}$-$V_g$ obtained at $V_d = -50$mV, for the 10 µm X 10 µm p-MOS of Fig. 5.1.
Figure 5.3: Hole mobility vs. effective field for a 10 µm X 10 µm p-MOS as in Fig. 5.1 and 5.2, with and without a PMA.

Figure 5.4: C-V characteristics without PMA, showing large difference in C-V shape from device to device.

Figure 5.5: Mobility characteristics for un-annealed devices showing large variation, indicating strong influence of the $N_{it}$.

Figure 5.6: C-V characteristics after 450°C, 5 min PMA in N$_2$ ambient, showing an order of magnitude improvement in terms of variations in C-V shape from device to device.

Figure 5.7: Mobility characteristics for post PMA devices showing reduced variation in device to device mobility, indicating large decrease in $N_{it}$.

Figure 5.8: $I_d$–$V_g$ for Ge PMOS devices with $L_g$ from 0.25 µm to 0.125 µm. Subthreshold slope and GIDL are highlighted for the 0.25 µm device.

Figure 5.9: Low-frequency noise spectra for a low- (155 cm$^2$/Vs) and a high-mobility (204 cm$^2$/Vs) Ge PMOSFET with $L=1$ µm and corresponding with $V_{ds}=-50$ mV and a $V_{gs}=-0.57$ V.

Figure 5.10: Normalized noise spectral density $S_{Id}/I_d^2$ versus gate voltage overdrive $V_{gs}-V_{th}$ at $f=25$ Hz for an $L=1$ µm Ge PMOSFET.

Figure 5.11: Input-referred noise spectral density $S_{V_g}$ versus gate voltage overdrive for a $L=0.25$ µm PMOSFET at $f=25$ Hz.

Figure 5.12: Correlation between $S_{V_g}$ at 25 Hz and the mobility $\mu$ for the $L=1$ µm PMOSFETs studied.

Figure 5.13: Inverse mobility versus $qD_{it}$ derived from CP and $qD_{ot}$ extracted from 1/f noise for the $L=1$ µm PMOSFETs studied.
Figure 5.14: Low-frequency noise spectra for a germanided and a non-germanided $p^+\text{-}n$ junction at a forward bias of 0.5 V. The area is 15000 $\mu$m$^2$ and the perimeter $P=10300$ $\mu$m........ 109

Figure 5.15: a) $I_F$ versus $V_F$ and b) $S_I$ at 1 Hz versus $I_F$ for a large area square diode ($A=30000$ $\mu$m$^2$) with or without NiGe. ........................................................................................................................................ 110

Figure 5.16: a) $I_F$ versus $V_F$ and b) $S_I$ at 1 Hz versus $I_F$ for a large perimeter diode ($P=10300$ $\mu$m) with or without NiGe. ............................................................................................................................................. 111

Figure 5.17: $S_I$ at 1 Hz and 100 $\mu$A for the germanided and non-germanided $p^+\text{-}n$ junctions versus area................................................................................................................................................. 113

Figure 6.1: Schematic cross section of a proposed ideal CMOS structure for the maximum enhancement of both electron and hole mobility................................................................. 117
LIST OF TABLES

Table 2.1: Scaling rules for constant field scaling................................................................. 9
Table 2.2: Scaling rules for different types of scaling.............................................................. 10
Table 2.3(a): Single-gate Non-classical CMOS Technologies............................................ 12
Table 2.3(b): Multiple-gate Non-classical CMOS Technologies........................................ 13
Table 2.4: Comparison of relevant properties of high κ candidates.................................... 23
Table 2.5: Comparison of some basic properties of Si, Ge and GaAs at 300 K...................... 40
Table 3.1: Deposition details for Pt-Ru alloy thin films...................................................... 55
Table 3.2: List of Gate dielectric process steps. ................................................................. 62
Table 3.3: Si – compatible process flow for Ge devices......................................................... 64
# LIST OF ACRONYMS/ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD</td>
<td>Atomic Layer Deposition</td>
</tr>
<tr>
<td>BOX</td>
<td>Buried Oxide</td>
</tr>
<tr>
<td>CET</td>
<td>Capacitance Equivalent Thickness</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>C-V</td>
<td>Capacitance-Voltage</td>
</tr>
<tr>
<td>CVD</td>
<td>Chemical Vapor Deposition</td>
</tr>
<tr>
<td>DI</td>
<td>De-Ionized</td>
</tr>
<tr>
<td>DIBL</td>
<td>Drain Induced Barrier Lowering</td>
</tr>
<tr>
<td>EOT</td>
<td>Effective Oxide Thickness</td>
</tr>
<tr>
<td>FCC</td>
<td>Face Centered Cubic</td>
</tr>
<tr>
<td>FET</td>
<td>Field Effect Transistor</td>
</tr>
<tr>
<td>FD</td>
<td>Fully Depleted</td>
</tr>
<tr>
<td>GIDL</td>
<td>Gate Induced Drain Leakage</td>
</tr>
<tr>
<td>GOI</td>
<td>Germanium On Insulator</td>
</tr>
<tr>
<td>HCP</td>
<td>Hexagonal Closed Pack</td>
</tr>
<tr>
<td>HDD</td>
<td>Highly Doped Drain</td>
</tr>
<tr>
<td>High-κ</td>
<td>High dielectric constant</td>
</tr>
<tr>
<td>HOT</td>
<td>Hybrid Orientation Technology</td>
</tr>
<tr>
<td>ITRS</td>
<td>International Technology Roadmap for Semiconductors</td>
</tr>
<tr>
<td>I-V</td>
<td>Current-Voltage</td>
</tr>
</tbody>
</table>

xvii
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF</td>
<td>Low Frequency</td>
</tr>
<tr>
<td>MOSFET</td>
<td>Metal Oxide Semiconductor Field Effect Transistor</td>
</tr>
<tr>
<td>n-MOS</td>
<td>n-channel Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>PD</td>
<td>Partially Depleted</td>
</tr>
<tr>
<td>PMA</td>
<td>Post Metallization Anneal</td>
</tr>
<tr>
<td>p-MOS</td>
<td>p-channel Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>PVD</td>
<td>Physical Vapor Deposition</td>
</tr>
<tr>
<td>RBS</td>
<td>Rutherford Backscattering Spectroscopy</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning Electron Microscope</td>
</tr>
<tr>
<td>STI</td>
<td>Shallow Trench Isolation</td>
</tr>
<tr>
<td>SOI</td>
<td>Silicon on Insulator</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission Electron Microscope</td>
</tr>
<tr>
<td>UHV</td>
<td>Ultra High Vacuum</td>
</tr>
<tr>
<td>UHP</td>
<td>Ultra High Pure</td>
</tr>
<tr>
<td>UTB</td>
<td>Ultra Thin Body</td>
</tr>
<tr>
<td>XRD</td>
<td>X-ray Diffraction</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

For over four decades, silicon has been the key ingredient used in the manufacturing of semiconductors that power today's advanced electronic devices. With the push for increasingly smaller devices, semiconductor manufacturers are faced with two major challenges: reducing power consumption and boosting device performance, well documented in the International Technology Roadmap for Semiconductors (ITRS) [1]. The high gate leakage of nitrided SiO$_2$ (SiON) and the depletion of polysilicon gate electrodes limit their usefulness in advanced complementary metal oxide semiconductor (CMOS) devices. Advanced high performance devices require high dielectric constant (high-$
\kappa$) gate dielectrics and metal gate electrodes to address these challenges. Although there has been a multitude of high-$\kappa$ dielectric materials investigated to date, the community is focusing on Hf-based materials such as HfO$_2$, HfO$_2$/SiN and HfSiON [2]. While the industry is converging on the high-$\kappa$ gate solution, it is not yet united behind a single approach for dual function metal gates.

Now nearly for about half a century the silicon industry has been delivering a continuously improving performance at an ever reducing cost, chasing what has been called the “Moore’s Law”. This was achieved primarily by physical scaling of the silicon device. As the industry enters the nanometer regime, where the transistor gate length drops down to few nm and the gate oxide thickness below 1 nm, physical limitations such as off-state leakage current and power density pose a potential threat to the performance enhancement that can be obtained simply by means of geometrical scaling. While high-$\kappa$ dielectric and metal gates are areas of active research, silicon on insulator (SOI) and strain engineering has quickly emerged as a new scaling vector for performance enhancement to extend the life of traditional silicon technology.
and are extensively been used in today’s 90 and 65 nm device technologies. It is envisaged that higher mobility materials will be needed to continue running on the Moore’s track. Germanium; the semiconductor once considered secondary to silicon due to its unfavorable surface properties, is now making a comeback as an enabling material for the next generation of device technology, primarily due to its high carrier mobility. Ge along with other high mobility materials such as III/V compounds are projected to be the new range of semiconductor materials, that will provide the platform for continued performance improvement at the 22nm technology node and beyond.

Investigation of various metal gates and the study of their materials and electrical properties is a major focus of this dissertation and results will be presented in later chapter. This dissertation also attempts to study high mobility channel devices. Results will be presented for Ge p-MOS devices for high hole mobility. Some low frequency 1/f noise measurement results on silicon passivated Ge p-MOSFETs and p⁺-n junctions with and without Ni germanidation are also presented.

Chapter 2 is intended to provide the reader with a comprehensive background on the current state of the art in the literature of the CMOS scaling efforts. It presents a historical perspective of classical scaling and a sneak preview of the future trends in non-classical scaling. It also discusses the opportunities and challenges for high κ, metal gates and Ge channel devices based on current literature review.

Chapter 3 describes the actual experimental process, equipment, and techniques used to study the metal gates and Ge channel devices. This section focuses on the materials processing and characterization, as well as describes various electrical characterization techniques used in this research work.
Chapter 4 and 5 are the core of this dissertation and include results and discussions. Chapter 4 focuses on the metal gates studies and presents results for binary metal alloys. This chapter presents materials studies such as XRD and TEM for the binary metal alloys. This chapter also discusses workfunction results obtained by different techniques of MOS C-V, and Schottky I-V, C-V and temperature dependent I-V.

Chapter 5 is dedicated to the Ge channel device studies. This chapter presents I-V and low frequency noise characterization for Ge p+ n diode and also presents some I-V and mobility data for Ge pMOS devices fabricated on a 2 µm thick epitaxial Ge layers on Si wafer.s The mobility results are compared to the Si universal mobility curve. LF noise assessment of silicon passivated Ge p-MOSFETs with TiN/TaN/HfO₂ gate stack has also been presented.

Chapter 6 summarizes the present work and provides a future outlook for continued research in this direction.
CHAPTER TWO: LITERATURE REVIEW

This chapter provides the necessary background information on CMOS technology and its scaling. It introduces the need for high-κ, metal gates and Ge channel devices. This chapter also discussed the opportunities and challenges for each of these based on current literature review.

2.1 CMOS and Its Scaling

The metal oxide semiconductor field effect transistor (MOSFET) is composed of an MOS capacitor and two p-n junctions placed immediately adjacent to the MOS diode. Since its first demonstration in 1960s, the MOSFET has developed quickly and has become the most important device for advanced integrated circuits such as microprocessors and semiconductor memories. This is because the MOSFET consumes very low power and has a high yield of working devices. Of particular importance is the fact that MOSFET can be readily scaled down and will take up less space than a bipolar transistor using the same design rule.

Complementary MOS (CMOS) refers to a complementary p-channel and n-channel MOSFET pair. CMOS logic is the most popular technology utilized in the present day integrated circuit design. The main reasons for the success of CMOS are low power consumption and good noise immunity.

Figure 2.1(a) shows CMOS inverter. The gate of the upper p-MOS device is connected to the gate of lower n-MOS device. Both devices are enhancement-mode MOSFETs with the threshold voltage $V_{Th}$ less than zero for the p-MOS device and $V_{Th}$ greater than zero for the n-
MOS device (threshold voltages are typically about 1/4th $V_{DD}$). When the input voltage, $V_i$, is at ground or at small positive value, the p-MOS device is turned on and the n-MOS device is off.

Figure 2.1: Complementary MOS (CMOS) inverter. (a) Circuit diagram. (b) Circuit layout. (c) Cross section along dotted A-A’ line of (b).

Hence, the output voltage, $V_o$, is very close to $V_{DD}$ (logic 1). When the input is at $V_{DD}$, the p-MOS is turned off, and the n-MOS is turned on. Therefore, the output voltage $V_o$ equals zero (logic 0). The CMOS inverter has unique feature: in either logic state, one device in the series path from $V_{DD}$ to ground is nonconductive. The current that flows in either steady state is a small leakage current, and only when both devices are on during switching does a significant current flow through the CMOS inverter. Thus, the average power dissipation is small, on the
order of nanowatts. Fig. 2.1(b) shows a layout of the CMOS inverter and Fig. 2.1(c) shows the device cross section along the A-A' line.

2.1.1 Conventional CMOS Scaling

Historically, the scaling of minimum feature sizes has been the major means for improving circuit speed, reducing power dissipation and increasing packing density of MOSFET devices. To meet the demands for increased integration density, device dimensions were shrunk drastically over the past years and the desired performance enhancement was primarily achieved by geometric scaling. The motivation for scaling of logic device dimensions can be seen by considering a simple model for the drive current of a MOSFET. The drive current can be written as

\[ I_D = \frac{W}{L} \mu C_{ox} (V_G - V_T - \frac{V_D}{2}) V_D \]  

(2.1)

where \( W \) is the width of the transistor channel, \( L \) is the channel length, \( \mu \) is the channel carrier mobility (assumed constant), \( C_{ox} \) is the capacitance density associated with the gate dielectric when the underlying channel is in the inverted state, \( V_G \) and \( V_D \) are the voltages applied to the transistor gate and drain, respectively, and the threshold voltage is given by \( V_T \). It can be seen that the drain current is proportional to the average charge across the channel (with the potential \( V_D/2 \)) and the average electric field \( (V_D/L) \) along the channel direction. Initially, \( I_D \) increases linearly with \( V_D \) and eventually saturates to a maximum when \( V_{D,sat} = V_G - V_T \) to yield

\[ I_{D,sat} = \frac{W}{L} \mu C_{ox} \frac{(V_G - V_T)^2}{2} \]  

(2.2)
The term \((V_G-V_T)\) is limited due to reliability and room temperature operation constraints, since too large a \(V_G\) would create an undesirable, high electric field across the oxide. Furthermore, \(V_T\) cannot easily be reduced below about 200 mV due to sub-threshold current and temperature fluctuation constraints. From a CMOS circuit performance point of view, a performance metric considers the dynamic response (i.e., charging and discharging) of the transistors and the supply voltage provided to the element at a representative (clock) frequency. The switching time of a CMOS inverter is limited by both the fall time required to discharge the load capacitance by the n-FET drive current and the rise time required to charge the load capacitance by the p-FET drive current. The switching response time \(t\),

\[
    t = C_{LOAD} \frac{V_{DD}}{I_D}.
\]

The load capacitance in the case of a single CMOS inverter is simply the gate capacitance, if one ignores parasitic contributions such as junction and interconnects capacitance. The above equation can be further simplified to:

\[
    t = CV = \frac{C_{AN}WLV_{gs}}{L \mu C_{OX}(V_{gs}-V_T)} \approx \frac{L^2}{\mu}.
\]

Obviously, decreasing the channel length is the most effective way to improve the device speed.

As the device is aggressively scaled down, short channel effects, such as (1) threshold voltage lowering, (2) channel length modulation, (3) carrier velocity saturation, (4) drain induced barrier lowering (DIBL), (5) punch-through (6) breakdown and (7) hot carrier generation, have to be controlled [3]. Threshold voltage lowering is of serious concern for switching properties and off-state power consumption. With decreasing channel length, the source and drain depletion regions become a more significant part of the channel. Thus, due to this shared charge from the
source and drain depletion regions, less charge (to be induced by the gate) is needed to invert the transistor. This results in lowering of the transistor threshold voltage. Channel length modulation is the effect of reduction in the length of the channel, as the pinch off region becomes a significant part of the channel. Thus, the channel length varies from the “physical” channel length with the applied voltages. This results in dependence of $I_{DS}$ on $V_{DS}$ even in saturation region. Carrier velocity saturation refers to a situation where, beyond a critical electric field, carrier velocity becomes constant and no longer depends on the applied electric field. This essentially leads to lower mobility, and hence, reduced current. DIBL refers to Drain Induced Barrier Lowering, where the barrier for the carriers to come to the channel from the source is reduced due to field lines penetrating from the drain to the source. This effect is enhanced at higher drain bias. As the channel length reduces, the depletion regions of source and drain approach each other. If the substrate doping is not high enough, eventually these two depletion regions could touch each other, a situation called punch-through. Due to the high field in the region near the drain, the carriers gain large velocities and may cause impact ionization. These newly generated carriers may accelerate in the high fields and generate more carriers. This can eventually result in an avalanche. This effect is called hot carrier generation and can lead to a rapid increase in the current and eventually, to a transistor break down.

Various scaling rules have been proposed including constant field scaling, constant voltage scaling, quasi-constant voltage scaling and generalized scaling. In commonly used constant field scaling, the maximum magnitude and shape of the internal electric field remain constant, and thus the shorter device has dimensions that are define a linearly scaled version of the larger device. The scaling rules for the constant field scaling are listed in Table 2.1 [4]. The limitation with this type of scaling is that the fraction of voltage required to turn the device off
from on is very high, since weak inversion region width does not scale. Also the external chip interface may require the voltage level to be maintained constant. Hence, sometimes voltages are not scaled and are maintained constant. This type of scaling is called constant voltage scaling. However in constant voltage scaling, the reduced device dimensions can increase the electric fields.

Table 2.1: Scaling rules for constant field scaling.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Scaling Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device Dimensions</td>
<td>1/k</td>
</tr>
<tr>
<td>Area</td>
<td>1/k²</td>
</tr>
<tr>
<td>Packing Density</td>
<td>k²</td>
</tr>
<tr>
<td>Doping Concentration</td>
<td>k</td>
</tr>
<tr>
<td>Bias Voltage and $V_T$</td>
<td>1/k</td>
</tr>
<tr>
<td>Bias Currents</td>
<td>1/k</td>
</tr>
<tr>
<td>Power Dissipation/circuit</td>
<td>1/k²</td>
</tr>
<tr>
<td>Power Dissipation/area</td>
<td>1</td>
</tr>
<tr>
<td>Capacitance</td>
<td>1/k</td>
</tr>
<tr>
<td>Capacitance/area</td>
<td>k</td>
</tr>
<tr>
<td>Charges</td>
<td>1/k²</td>
</tr>
<tr>
<td>Charges/Area</td>
<td>1</td>
</tr>
<tr>
<td>Electric field intensity</td>
<td>1</td>
</tr>
<tr>
<td>Body factor coefficient</td>
<td>1/k</td>
</tr>
<tr>
<td>Speed</td>
<td>k</td>
</tr>
<tr>
<td>Power-delay product</td>
<td>1/k³</td>
</tr>
</tbody>
</table>

To reduce field, quasi-constant voltage scaling is often used wherein device dimensions are scaled according to constant field scaling rules, but voltages are scaled less dramatically. However, in that case the depletion region widths do not scale in the same ratio as the device dimensions. This problem is avoided by using a different scaling factor for the substrate doping. This is called generalized scaling. The different scaling rules are summarized in Table 2.2.
Table 2.2: Scaling rules for different types of scaling.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Constant Field Scaling</th>
<th>Constant Voltage Scaling</th>
<th>Quasi-constant voltage scaling</th>
<th>Generalized Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W,L$</td>
<td>$1/k$</td>
<td>$1/k$</td>
<td>$1/k$</td>
<td>$1/k$</td>
</tr>
<tr>
<td>$T_{ox}$</td>
<td>$1/k$</td>
<td>$1/k'$</td>
<td>$1/k$</td>
<td>$1/k$</td>
</tr>
<tr>
<td>$N_A$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k$</td>
<td>$k^2/k'$</td>
</tr>
<tr>
<td>$V,V_T$</td>
<td>$1/k$</td>
<td>$1$</td>
<td>$1/k'$</td>
<td>$1/k'$</td>
</tr>
</tbody>
</table>

For the past four decades the silicon industry has been delivering a continuously improving performance at an ever reducing cost by the physical scaling of silicon devices. As the transistors shrinks to the nanometer regime performance improvement simply by means of geometric scaling of the device is becoming more and more challenging. This has forced engineers to find new and innovative solutions to extend CMOS down to, and perhaps beyond, the 22 nm node by introducing non-conventional scaling. The following sub-section summarizes various non-classical CMOS scaling trends, currently in active research.

2.1.2 Non Conventional CMOS Scaling

Non-conventional or non-classical CMOS includes those advanced MOSFET structures shown in Table 2.3 (a) and (b), which, combined with material enhancements, such as new gate stack materials, provides a path to scaling CMOS to the end of the roadmap. For digital applications, scaling challenges include controlling leakage currents and short-channel effects, increasing drain saturation current while reducing the power supply voltage, and maintaining control of device parameters (e.g., threshold voltage and leakage current) across the chip and from chip to chip. For analog/mixed-signal/RF applications, the challenges additionally include
sustaining linearity, low noise figure, high power-added efficiency, and good transistor matching.

Researchers are pursuing two avenues to meeting these challenges—new materials and new transistor structures. New materials include those used in the gate stack (high-κ dielectrics and electrode materials), those used in the conducting channel that have improved carrier transport properties, as well as new materials used in the source/drain regions with reduced resistance and improved carrier injection properties. New transistor structures seek to improve the electrostatics of the MOSFET, provide a platform for introduction of new materials, and accommodate the integration needs of new materials. The following provides a brief introduction and overview to each of these non-classical CMOS structures given in Table 2.3 (a) and (b).
Table 2.3(a): Single-gate Non-classical CMOS Technologies.

<table>
<thead>
<tr>
<th>Device</th>
<th>Transport-Enhanced MOSFETs</th>
<th>UTB SOI MOSFETs</th>
<th>Source-Driven Engineered MOSFETs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>Strained Si, Ga, SiGe, etc.</td>
<td>Fully depleted SOI with body thinner than 10 nm</td>
<td>Schottky source/drain</td>
</tr>
<tr>
<td>Application/Driver</td>
<td>HP CMOS (12)</td>
<td>HP, LDP, and LSTP CMOS (12)</td>
<td>HP CMOS (12)</td>
</tr>
<tr>
<td>Advantages</td>
<td>• High mobility</td>
<td>• Improved subthreshold slope</td>
<td>• Low source/drain resistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No floating body</td>
<td>• Reduced S/D and DIBL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Potentially lower Eoff</td>
<td>• Reduced parasitic gate capacitance</td>
</tr>
<tr>
<td>Particular Strength</td>
<td>• High mobility</td>
<td>• Low diode leakage</td>
<td>• No need for abrupt S/D doping or activation</td>
</tr>
<tr>
<td></td>
<td>without change in device architecture</td>
<td>• Low junction capacitance</td>
<td>• Very low gate capacitance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No significant change in design with respect to bulk</td>
<td></td>
</tr>
<tr>
<td>Potential Weakness</td>
<td>Material defects and device leakage (only for bulk)</td>
<td>Very thin silicon required with low defect density</td>
<td>Ultra thin SOI required for channel and S/D</td>
</tr>
<tr>
<td></td>
<td>Process complexity and thermal budget</td>
<td>Vth adjustment difficult</td>
<td>High current drain resistance</td>
</tr>
<tr>
<td></td>
<td>Operating temperature</td>
<td>Selective doping required for inverted S/D</td>
<td>Reliability</td>
</tr>
<tr>
<td>Scaling Issues</td>
<td>Rashba spacing usually smaller than Si</td>
<td>Controllability of Si film thickness</td>
<td>Process becomes more severe with larger source/drain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Process becomes more severe with larger source/drain</td>
<td>No particular scaling issue</td>
</tr>
<tr>
<td>Design Challenges</td>
<td>None</td>
<td>None</td>
<td>Compact model needed</td>
</tr>
<tr>
<td>Gain/Leak in Layout</td>
<td>No difference</td>
<td>No difference</td>
<td>Compact model needed</td>
</tr>
<tr>
<td>compared to Bulk</td>
<td></td>
<td></td>
<td>No difference</td>
</tr>
<tr>
<td>Impact on ss/Dy compared to Bulk</td>
<td>Improved by 25–30% (from MASTAR, supposing $E_{min}$)</td>
<td>Improved by 15–20% (from MASTAR, supposing $E_{min}$)</td>
<td>Improved by 10–15% (from MASTAR, supposing $P_{min}$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improved by 15–20% (from MASTAR, supposing $E_{min}$)</td>
<td>Improved by 10–15% (from MASTAR, supposing $P_{min}$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improved by 15–20% (from MASTAR, supposing $E_{min}$)</td>
<td>Improved by 10–15% (from MASTAR, supposing $P_{min}$)</td>
</tr>
<tr>
<td>Aging Suitability</td>
<td>Not clear</td>
<td>Potential for slight improvement</td>
<td>Not clear</td>
</tr>
<tr>
<td>$E_{min}$ advantage compared to Bulk</td>
<td></td>
<td>Potential for slight improvement</td>
<td>Not clear</td>
</tr>
</tbody>
</table>
Transport-enhanced MOSFETs [5-7] are those structures for which increased transistor drive current for improved circuit performance can be achieved by enhancing the average velocity of carriers in the channel. Approaches to enhancing transport include mechanically straining the channel layer to enhance carrier mobility and velocity, and employing alternative
channel materials such as silicon-germanium, germanium, or III-V compound semiconductors with electron and hole mobilities and velocities higher than those in silicon. A judicious choice of crystal orientation and current transport direction may also provide transport enhancement [8]. However, an important issue is how to fabricate transport enhanced channel layers (such as a strained Si layer) in several of the non-classical CMOS transistor structures [e.g., the multiple gate structures discussed in Table 2.3 (b)]. Researchers have recently demonstrated that a strained Si-on-insulator (SOI) substrate technology can be used to combine the advantages of the ultra-thin body (UTB) structure and enhanced carrier transport [9-11].

The UTB SOI MOSFET [12-14] consists of a very thin \( t_{\text{Si}} \leq 10 \text{ nm} \), fully depleted (FD) transistor body to ensure good electrostatic control of the channel by the gate in the off state. Typically, the ratio of the channel length to the channel thickness will be \( \geq 3 \). Therefore, an extremely thin \( t_{\text{Si}} < 4 \text{ nm} \) Si channel is required to scale CMOS to the 22-nm node. The use of a lightly doped or undoped body provides immunity to \( V_T \) variations due to statistical dopant fluctuations, as well as enhanced carrier mobilities for higher transistor drive current. The localized and ultra-thin buried oxide (BOX) FET [15-18] is an UTB SOI like FET in which a thin Si channel is locally isolated from the bulk-Si substrate by a thin (10–30 nm) buried dielectric layer. This structure combines the best features of the classical MOSFET (e.g., deep source/drain contact regions for low parasitic resistance) with the best features of SOI technology (improved electrostatics). The increased capacitive coupling between the source, drain, and channel with the conducting substrate through the ultra-thin BOX has the potential of reducing the speed of the device, and improving its electrostatic integrity. The former may be traded against the latter by reducing the channel doping, which eventually leads to moderately improved speed for a constant \( I_{\text{off}} \). Engineering the source/drain is becoming critically important.
to maintaining the source and drain resistance to be a reasonable fraction (10%) of the channel resistance. Consequently, a new category of source/drain engineered MOSFETs [19-21] is introduced to address this issue. Two sub-category structures are described for providing engineered source/drain structures. First is the Schottky source/drain structure [22]. In this case, the use of metallic source and drain electrodes minimizes parasitic series resistance and eliminates the need for ultra-shallow p–n junctions. Metals or silicides that form low (near zero) Schottky barrier heights in contact with silicon (i.e., a low-work-function metal for NMOS, and a high-work-function metal for PMOS) are required to minimize contact resistance and maximize transistor drive current in the on state. A UTB is needed to provide low leakage in the off state. Second is the reduced fringing/overlap gate FET. As MOSFET scaling continues, the parasitic capacitance between the gate and source/drain detrimentally affects circuit performance, and its impact becomes more significant as the gate length is scaled down. For gate lengths below ~20 nm, transistor optimization for peak circuit performance within leakage current constraints will likely dictate a structure wherein the gate electrode does not overlap the source or drain to minimize the effect of parasitic fringing/overlap capacitance. Due to lengthening of its electrical channel, the non-overlapped gate structure does not require ultra-shallow source/drain junctions in order to provide good control of short-channel effects. Also, the increase of source/drain resistance usually expected for the non-overlap transistor is reduced with decreasing gate length, thus providing a new optimization paradigm for extremely short devices.

As illustrated in Table 2.3 (b) and described in the following, a variety of multiple-gate non-classical CMOS structures [23] have been proposed and demonstrated to help manage electrostatic integrity in ultra scaled CMOS structures. In the first of these structures, the $N$-gate ($N > 2$) MOSFET [24], current flows horizontally (parallel to the plane of the substrate) between
the source and drain along vertical channel surfaces, as well as one or more horizontal channel surfaces. The large number of gates provides for improved electrostatic control of the channel, so that the Si body thickness and width can be larger than for the UTB SOI and double-gate FET structures, respectively. The gate electrodes are formed from a single deposited gate layer and are defined lithographically. They are tied together electrically and are self-aligned with each other as well as the source/drain regions. The principal advantage of the structure resides in the relaxation of the needs on the thinness of the Si body or the vertical fin. The challenge is in slightly poorer electrostatic integrity than with double-gate structures, particularly in the corner regions of the channel.

Several double-gate MOSFET structures [25, 26] have been proposed to further improve engineering of the channel electrostatics and, in some cases, to provide independent control of two isolated gates for low-power and, perhaps, mixed-signal applications. Four typical double-gate structures are described in the following. First is the tied double-gate, sidewall conduction structure [27]. This is a double-gate transistor structure in which current flows horizontally (parallel to the plane of the substrate) between the source and drain along opposite vertical channel surfaces. The width of the vertical silicon fin is narrow (smaller than the channel length) to provide adequate control of short-channel effects. A lithographically defined gate straddles the fin, forming self-aligned, electrically connected gate electrodes along the sidewalls of the fin. The principal advantage with this structure is the planar bulk-like layout and process. In fact, this structure can be implemented on bulk Si substrates [28]. The major challenge is with fabrication of thin fins that need to be a fraction (one third to one half) of the gate length, thus requiring sub-nm lithographic techniques.
The second structure is the tied double-gate planar FET [29]. In this structure, current flows horizontally (parallel to the plane of the substrate) between the source and drain along opposite horizontal channel surfaces. The top and bottom gate electrodes are deposited in the same step and are defined lithographically. They may or may not be self-aligned to each other, and are electrically connected to one another. The source/drain regions are typically self-aligned to the top gate electrode. The principal advantages of this structure reside in the potential simplicity of the process (closest to bulk planar process) and in the compactness of the layout (same as for bulk planar) as well as in its compatibility with bulk layout (no need for redesigning libraries). It is also important that the channel thickness is determined by epitaxy, rather than etching, and, thus, is very well controlled. The challenge resides in the doping of the poly in the bottom gate (shadowed by the channel), but this problem disappears automatically when switching to a metal-like gate electrode. Another major challenge is in the fabrication process, particularly for those structures requiring alignment of the top and bottom gate electrodes.

The third structure is the independently switched double-gate (ground-plane) FET [30]. This structure is similar to the tied double-gate planar FET, except that the top and bottom gate electrodes are electrically isolated to provide for independent biasing of the two gates. The top gate is typically used to switch the transistor on and off, while the bottom gate is used for dynamic (or static) $V_T$ adjustment. The principal advantage is in the very low $I_{off}$ this structure offers. The disadvantage is in rather poor sub-threshold behavior and in the relaxed layout. An independently switched double-gate transistor can also be implemented in a vertical structure by disconnecting the gates of the double-gate, sidewall conduction structure by chemical mechanical polishing [31].
The fourth structure is the “vertical conduction” transistor [32]. In this case, current flows between the source and drain in the vertical direction (orthogonal to the plane of the substrate) along two or more vertical channel surfaces. The gate length, hence the channel length, is defined by the thickness of the single deposited gate layer, rather than by a lithographic step. The gate electrodes are electrically connected, and are vertically self-aligned with each other and the diffused source/drain extension regions. The principal advantage with this structure is that the channel length is defined by epitaxy rather than by lithography (possibility of very short and well controlled channels). The disadvantage is this structure requires a challenging process and the layout is different from that for bulk transistors.

Scaling CMOS to and beyond the 22-nm technology node (requiring a physical gate length of 9-nm or less) will probably require the introduction of several new material and structural changes to the MOSFET to sustain performance increases of 17% per year. Material changes will include strained silicon n- and p-channels and a new gate stack including a high-κ dielectric and a metal gate electrode. Materials with high mobilities such as Ge and other III-V compound semiconductors will need to be explored extensively. The following sub-sections discuss the current state of the art, opportunities and challenges for high-κ, metal gate and Ge channel devices.

2.2 High –κ Dielectrics

The gate stack is the most important and critical part of the sub-micron MOSFETs. The gate stack is composed of (a) dielectric-silicon interface, (b) bulk dielectric and (c) gate electrode. In order to improve the device performance, the gate oxide has been scaled
aggressively to increase the gate capacitance thus to improve the device performance. The aggressively scaled gate oxide is necessary to invert the surface to a sufficient sheet charge density to obtain the desired current for the given supply voltage and to avoid short channel behavior. Most of the short channel effects can be made less severe by having an electrically thinner gate oxide, thereby increasing the gate oxide capacitance and increasing the gate control of the channel. The threshold voltage lowering is reduced with lower gate oxide thickness, where the lowering in the threshold voltage is given by $Q_B/C_{OX}$, where the shared charge from the depletion regions of source and drain is $Q_B$. Since it applies at higher drain voltages, DIBL also reduces with decreasing gate oxide thickness. However, there are several major challenges facing gate oxide thickness scaling below 1.5nm. These include; (1) the gate leakage current through the gate oxide increases exponentially with decreasing gate thickness [33] (2) inadequate reliability for SiO$_2$ below 1.5nm, (3) boron penetration in PMOSFETs surface-channel with P$^+$ gate increases significantly [34] and (4) increased effect of polysilicon depletion. New materials have to be introduced in the basic CMOS structure to replace the existing ones to enable continued scaling of devices.

For the gate capacitance discussion, consider a parallel plate capacitor (ignoring quantum mechanical and depletion effects from a Si substrate and gate),

$$ C = \frac{\kappa \varepsilon_0 A}{t} $$

(2.5)

where $\kappa$ is the dielectric constant (also referred to as the relative permittivity) of the material, $\varepsilon_0$ is the permittivity of free space ($= 8.85 \times 10^{-12} \text{ F/m}$), $A$ is the area of the capacitor, and $t$ is the thickness of the dielectric. This expression for $C$ can be rewritten in terms of equivalent oxide thickness, $t_{eq}$ and $\kappa_{ox}$ ($=3.9$, dielectric constant of SiO$_2$) of the capacitor. For example, if the
The dielectric is SiO$_2$, $t_{eq} = 3.9 \varepsilon_0 (A/C)$, and a capacitance density of $C/A = 34.5$ fF/µm$^2$ corresponds to $t_{eq} = 10$ Å.

One efficient way to reduce leakage current is to use a high dielectric constant $\kappa$, or high-κ dielectric film. High-κ films allow the use of a physically thicker film while electrically acting as a thin dielectric. A material such as Ta$_2$O$_5$ with dielectric constant around 25 can be made about 6 times thicker than SiO$_2$ for the same SiO$_2$ equivalent oxide thickness, which is easy to control and manufacture. A physically thicker film would increase the tunneling barrier thickness and decrease the gate leakage. Such high-κ materials would offer significant reductions in tunneling currents while still producing low $t_{eq}$.

\[
\frac{t_{eq}}{K_{ox}} = \frac{t_{high-K}}{K_{high-K}}
\]

\[
t_{high-K} = \frac{K_{high-K}}{K_{ox}} t_{eq} = \frac{K_{high-K}}{3.9} t_{eq}
\]

Therefore, the MOSFET drive current, which is dependent on the electrical thickness of a dielectric, can be kept essentially constant while the tunneling current may be effectively reduced. This type of approach can potentially be quite powerful since gate current density has an inverse exponential dependence on dielectric thickness. However, in the case of high-κ dielectrics some possible trade-offs exist which may actually increase the gate tunneling current. One issue is the band gap, or more importantly the barrier height, tends to decrease with increasing dielectric constant [35]. The other tradeoff may lie in the effective mass within the dielectric where the effective mass may follow the band gap of the film and therefore decrease accordingly [36]. Subsequently, the decrease in barrier height along with the decrease of the effective mass would cause an increase in tunneling current at a particular bias, and these factors
could then offset the reduction in current caused by the increased physical thickness of the
dielectric [37]. Therefore, as the dielectric constant of the dielectric increases, the physical
thickness increase is compensated by a decrease in the barrier height and three issues must be
considered. In addition some interfacial oxide between the high-\(\kappa\) dielectric and the silicon
substrate can be formed during gate stack deposition. With same equivalent oxide thickness, the
tunneling currents through the stacks can be much higher than that through the pure high-\(\kappa\)
dielectric because of significantly reduced barrier height associated with the interface region
[38].

Unfortunately, many additional problems are associated with these dielectrics including
thermal stability on silicon substrate, compatibility with gate electrode materials, small band gap,
undesired band misalignment, leakage current, low channel mobility, and lower than expected
dielectric constants of the as-deposited films [39, 40]. For example, in addition to appropriate
barrier height and dielectric constant, any alternative dielectric must form a stable interface with
silicon as well as the gate contact material at any subsequent processing temperature. The quality
of the dielectric-silicon interface is critical for achieving high channel mobility. It is not clear the
required low interface state densities, low fixed charge and smooth surface can be achieved with
any interface material combination other than silicon–SiO\(_2\). If an interfacial SiO\(_2\) oxide layer is
required, it may be extremely difficult to obtain equivalent oxide thickness below 1nm.
Furthermore, it seems that intrinsic interface fixed charges exist between these high-\(\kappa\) materials
and silicon and also in internal interfaces [41]. The sign of fixed charge depends on the
properties of the dielectrics. For example, the fixed charge within ZrO\(_2\) dielectric were found to
be negative and to induce a \(\sim\) 0.6 nm interface SiO\(_2\) oxide layer [42].
Many materials systems are currently under consideration as potential replacements for SiO$_2$ as the gate dielectric material for sub-micron CMOS technology. A systematic consideration of the required properties of gate dielectrics indicates that the key guidelines for selecting an alternative gate dielectric are:

- permittivity, band gap, and band alignment to silicon
- thermodynamic stability
- film morphology
- interface quality
- compatibility with the current/expected materials to be used in CMOS device processing.
- process compatibility
- reliability

Many dielectrics appear favorable in some of these areas, but very few materials are promising with respect to all of these guidelines [43]. As an alternative to oxide/nitride systems, much work has been done on high-\(\kappa\) metal oxides as a means to provide a substantially thicker (physical thickness) dielectric for reduced leakage and improved gate capacitance. In the search to find suitable high-\(\kappa\) gate dielectrics for use beyond silicon oxynitride systems, several approaches have been used in fabricating potential material candidates.
Table 2.4: Comparison of relevant properties of high κ candidates.

<table>
<thead>
<tr>
<th>Material</th>
<th>Dielectric constant(k)</th>
<th>Band Gap $E_G$(eV)</th>
<th>$\Delta E_c(eV)$ to Si</th>
<th>Crystal structure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>3.9</td>
<td>8.9</td>
<td>3.2</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Si$_3$N$_4$</td>
<td>7</td>
<td>5.1</td>
<td>2</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>9</td>
<td>8.7</td>
<td>2.8$^a$</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Y$_2$O$_3$</td>
<td>15</td>
<td>5.6</td>
<td>2.3$^a$</td>
<td>Cubic</td>
</tr>
<tr>
<td>La$_2$O$_5$</td>
<td>30</td>
<td>4.3</td>
<td>2.3$^a$</td>
<td>Hexagonal, cubic</td>
</tr>
<tr>
<td>Ta$_2$O$_5$</td>
<td>26</td>
<td>4.5</td>
<td>1-1.5</td>
<td>Orthorhombic</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>80</td>
<td>3.5</td>
<td>1.2</td>
<td>Tetrac.$^c$ (rutile, anatase)</td>
</tr>
<tr>
<td>HfO$_2$</td>
<td>25</td>
<td>5.7</td>
<td>1.5$^a$</td>
<td>Mono.$^b$,tetrag.c, cubic</td>
</tr>
<tr>
<td>ZrO$_2$</td>
<td>25</td>
<td>7.8</td>
<td>1.4$^a$</td>
<td>Mono.$^b$,tetrag.c, cubic</td>
</tr>
</tbody>
</table>

$^a$ Calculated by Robertson
$^b$ Mono. = monoclinic.
$^c$ Tetrac. = tetragonal.

Considering the potential problems and limitations in using a ~5 Å SiO$_2$ layer in a dielectric stack, it is highly desirable to employ an advanced gate dielectric which is stable on Si, and exhibits an interface quality to Si which is comparable to that of SiO$_2$. This would avoid the need for an interfacial layer and at the same time, the high permittivity of the material could be fully realized. Table 2.4 is a compilation of several potential high-κ dielectric candidates, with the columns indicating the most relevant properties [44, 45].

2.2.1 Gate Stack Compatibility

A significant issue for integrating any advanced gate dielectric into standard CMOS is that the dielectric should be compatible with Si-based gates, rather than require a metal gate. Si-based gates are desirable because dopant implant conditions can be tuned to create the desired threshold voltage $V_T$ for both n-MOS and p-MOS, and the process integration schemes are well established in industry. Nearly all the potential advanced gate dielectrics investigated to this
point, however, require metal gates. This is expected because the instability with Si will exist at both the channel and the poly-Si gate interfaces.

Specifically, metal gates such as TiN and Pt have been used with most of the high-κ gate dielectrics mentioned above to prevent reactions at the gate interface. Attempts have been made to use doped poly-Si gates with Ta$_2$O$_5$, by depositing a CVD SiO$_2$ reaction barrier, which is of lower electrical quality than thermal oxide, on top of the Ta$_2$O$_5$ layer [46]. The presence of SiO$_2$ at both the channel and gate interface predictably limited the device performance, and the lowest obtainable oxide equivalent was $t_{eq} = 23\text{Å}$. Even initial attempts to use poly-Si gates with ZrO$_2$ [47, 48] have been unsuccessful as reaction layers have been observed at the interface. Al$_2$O$_3$ has been shown to be stable with respect to reaction with the poly-Si gates throughout typical CMOS processing, however, both boron and phosphorous dopant diffusion have been observed with Al$_2$O$_3$ gate dielectrics, which cause significant, undesired shifts of $V_{FB}$ and $V_T$ values [49, 50].

Metal gates are very desirable for eliminating dopant depletion effects and sheet resistance constraints. In addition, use of metal gates in a replacement gate process can lower the required thermal budget by eliminating the need for dopant activation anneals in the poly-Si electrode [51, 52]. There are two basic approaches toward achieving successful insertion of metal electrodes: a single midgap metal or two separate metals. The energy diagrams associated with these two approaches are shown in Fig. 2.2. The first approach is to use a metal (such as TiN) that has a work function that places its Fermi level at the midgap of the Si substrate, as shown in Fig. 2.2(a). These are generally referred to as “midgap metals.” The main advantage of employing a midgap metal arises from a symmetrical $V_T$ value for both NMOS and PMOS, because by definition the same energy difference exists between the metal Fermi level and the
conduction and valence bands of Si. This affords a simpler CMOS processing scheme, since only one mask and one metal would be required for the gate electrode.

![Figure 2.2: Energy diagram of threshold voltages for NMOS and PMOS devices using (a) midgap metal gates and (b) dual metal gates.](image)

For the case of sub-0.13 µm bulk CMOS devices, however, a major drawback is that the band gap of Si is fixed at 1.1 eV, thus the threshold voltage for any midgap metal on Si will be ~0.5 V for both NMOS and PMOS. Since voltage supplies are expected to be <1.0 V for sub-0.13 µm CMOS technology, $V_T$ ~0.5V is much too large, as it would be difficult to turn on the device. Typical threshold voltages for these devices are expected to be 0.2–0.3 V. Compensation implants can be made in the channel to lower the $V_T$, but other concerns then arise regarding increased impurity ion scattering, which would degrade the channel carrier mobility. Furthermore, midgap work function metal gate systems have been predicted not to provide a performance improvement worthy of the added process complexity to replace Si-based gates [53].

The second main approach toward metal electrodes involves two separate metals, one for PMOS and one for NMOS devices. As shown in Fig. 2.2(b), two metals could be chosen by their
work functions, $\Phi_m$, such that their Fermi levels line up favorably with the conduction and valence bands of Si, respectively. In the ideal case depicted in Fig. 2.2(b), the $\Phi_m$ value of Al could achieve $V_T \sim 0.2V$ for NMOS, while the higher $\Phi_m$ value of Pt could achieve $V_T \sim 0.2V$ for PMOS. In practice, Al is not a feasible electrode metal because it will reduce nearly any oxide gate dielectric to form an Al$_2$O$_3$-containing interface layer. Other metals with relatively low work functions, such as Ta, V, Nb, etc., however, are feasible gate metals for NMOS. Similarly for PMOS, Pt, Ru, etc. are possible choices for the gate metal. As an alternative to elemental metals, conducting metal oxides such as IrO$_2$ and RuO$_2$ can provide high $\Phi_m$ values in addition to affording the use of standard etching and processing techniques. Alloys of these metals and similar conducting oxides can also potentially be fabricated to achieve a desired work function. More work must be done to better understand alternative metal electrodes for dual metal approaches, as a means to alleviate potentially limiting properties of doped poly-Si.

For CMOS scaling in the longer term, however, current roadmap predictions indicate that poly-Si gate technology will likely be phased out beyond the 45 nm node, after which a metal gate substitute appears to be required. It is therefore also desirable to focus efforts on dielectric materials systems which are compatible with potential metal gate materials. The following section discusses some opportunities and challenges in implementing for metal gates.

### 2.3 Metal Gates

Along with the replacement of SiO$_2$ with alternative gate dielectrics, replacement of dual-doped polysilicon is desired due to limited active doping density which will cause the loss of drive voltage in polysilicon and increase $t_{eq}$ values [54, 55]. As scaling of MOS devices
continues further, this polysilicon depletion effect will be a major problem and will most likely be unacceptable for 45 nm and beyond devices. As devices scale down, polysilicon depletion becomes more significant. With the anticipated scaling by constant field scaling, the oxide field remains essentially constant. This means that the charge per unit area in the polysilicon gate remains constant with scaling. This may result in the voltage drop in the polysilicon as long as gate-doping density remains constant. Thus, since voltage levels decrease, the polysilicon voltage drop becomes a larger fraction of available device voltage and current drive is subsequently reduced. From literature reports, it is anticipated that it will be difficult to get electronically active doping densities much above $10^{20}$/cm$^3$ for n-type poly and above the mid $10^{19}$/cm$^3$ for p-type poly. This represents a significant degradation in current drive capability. In addition, boron out-diffusion from the P$^+$ doped silicon gates can travel through the gate dielectric and accumulate in the n-Si substrate where it can change the threshold voltage and reduce dielectric reliability, thereby again degrading device characteristics. Finally, it is expected that polysilicon will not be stable on most high-$\kappa$ dielectric materials above since it can react to form silicides. Therefore, the use of stable metallic gate electrodes may solve the above issue.

The international semiconductor roadmap for semiconductors has provided guidelines for gate electrode carrier concentration and sheet resistance that are needed to successfully achieve sub-45 nm CMOS operation. The most important selection criteria for an alternative metal gate electrode are shown in Fig. 2.3 and are briefly discussed in the following sub-sections.
2.3.1 Work Function

Appropriate work function $\Phi_m$ for NMOS and PMOS must be achieved in order to get low threshold voltages. For next generation CMOS technology, dual metals are required as replacement to achieve low threshold voltages, i.e., high work function (Fermi level near the silicon valence band edge) for PMOS and low work function (Fermi level near the silicon conduction band edge) for NMOS to achieve appropriate threshold voltage without compromising off-state leakage current. A device simulation taking into account quantum effects on device drive current suggests that the $\Phi_m$ for NMOS and PMOS gate electrodes must be around 4 eV and 5 eV, respectively [56]. Midgap gates are not viable because the resulting threshold voltage is too high and the threshold voltage can not be reduced by lowering the
substrate doping. There are several NMOS candidate metals with work functions near 4 eV such as Al, Ta, Mo, Zr, Hf, V, Ti and several metals with work functions near 5 eV such as Co, Pd, Ni, Re, Ir, Ru, Pt for p-MOS. There are some general trends in anticipated properties that can be drawn from the studies of metal-silicon (Schottky) contacts. Generally, work functions with respect to vacuum increase as the electronegativity of the metal increases. For examples, metal silicides and metal nitrides have higher work functions than their elemental metal counterparts. Metal nitrides with a work function of 5 eV are attractive since they offer better diffusion barrier properties and better thermal stability compared to the metal. However, at present, no metal nitrides with large work function values are available. Moreover problems like nitrogen diffusion into the gate dielectric or the substrate from the gate nitride may occur. Work function modulation may also occur by adding oxygen into the metals while still maintaining high conductivity.

Usually, the work function of a conducting metal oxide is greater than that of the corresponding metal because of the Fermi level change. As the metal is oxidized, it loses its electrons resulting in a decrease of Fermi level. Conducting metal oxides such as RuO$_2$ and IrO$_2$ have been shown to have large work functions. However, it is difficult to find a conducting metal oxide whose work function is near 4 eV, a value required for NMOS electrode.

### 2.3.2 Thermal and Chemical Stability

Thermal and chemical stability is one of the most important issues for metal gate electrodes. Metal gates should be able to survive device processing and must be stable with respect to the underlying gate dielectric. However, most metal gate electrodes suffer from some
drawbacks such as high temperature instability resulting in inter-diffusion of silicon-oxygen. Many elemental metals form silicides on dielectrics after high temperature annealing if the dielectrics contain SiO₂. Also, many metals are not stable on emerging high-κ dielectric materials such as Ta₂O₅, SiO₂-ZrO₂, and Al₂O₃. Several thermal and chemical instabilities have been observed when these metals are placed on SiO₂ and on alternate dielectrics. Any reaction or intermixing at electrode/dielectric interface will make the issue more complicated due to the unpredictable modification of work function, resistivity and EOT. Thermal and chemical stabilities of conducting oxides have also been widely studied in literatures. In general, no reactions are found to take place between conducting oxides and other dielectric materials at temperatures as high as 900 °C [57]. This is attributed to the excellent diffusion barrier property of conducting oxides that can prevent inter-diffusion at interfaces. It is expected that these excellent diffusion barrier properties will greatly assist in achieving chemical and thermal stability of the interface between advanced gates and high-κ dielectrics, which currently suffer from many instabilities.

2.3.3 High conductivity and high active electrical concentration

In order to eliminate gate depletion effects, the carrier concentration of the gate electrode must be high. The use of metal gates eliminates the gate depletion effects, resulting from the high carrier concentration and low resistivity. The resistivity of metal films is in the range of 3 μΩ.cm to 100 μΩ.cm. Conducting oxide such as RuO₂ and IrO₂ have resistivity as low as 30 μΩ.cm. It has been reported that metal films have carrier concentration above 10²² cm⁻³, which is significantly higher than the heaviest doped poly-Si films [58].
2.3.4 Compatibility with CMOS integration

For CMOS technology, a preferred solution to the integration of two gate electrode metals may involve the use of a metal and compound of the same metal, such as metal alloy (e.g., W and WN$_x$) so that a simple reaction might locally convert the metal into its compound and thus significantly simplify device integration. However, this approach relies on finding a metal and its nitride or alloy that has a 1 eV difference in work function. Si based gates are desirable because dopant implant conditions can be tuned to create the desired threshold voltage $V_T$ for both NMOS and PMOS, and the process integration schemes are well established in industry. But many of the new dielectric materials are unstable in direct contact with polysilicon. In addition, metal gates are very desirable for eliminating poly depletion effects and sheet resistance constraints. Thus it is likely that the entire gate stack will have to be replaced with metal gates together with high-κ dielectrics. A common midgap work function metal gate shifts the threshold voltage higher by 0.5 V compared to highly doped polysilicon. The threshold voltage can be brought down to 0.2-0.3 V by a 10 nm counter doping layer with a net counter doping concentration in the range of $3 \times 10^{18}$ cm$^{-3}$. But this will result in 10% subthreshold slope degradation. If the channel doping is lowered to provide a suitable lower threshold voltage, the channel doping is too low to control short-channel effects. For this reason, two different gate metals are required with work functions near $E_C$ and $E_V$. The incorporation of two different metal gates in a CMOS flow enormously complicates the fabrication process. In dual-metal-gate CMOS process, two dielectric depositions, two electrode depositions and two non-critical photo lithography steps can be expected.
2.3.5 Deposition technique

Various methods such as reactive sputtering, solution chemical techniques, and chemical vapor deposition (CVD) have been used for the deposition of thin films. Among these, CVD has advantages for film fabrication in IC processing due to its good step coverage, high deposition rate and compatibility to large-scale processing. Since any impurity contained in a CVD metal can degrade the dielectric during a subsequent high temperature annealing process, sputtering deposition of the metal gate electrode is more favorable than CVD for the conventional IC processing. On the other hand, the degradation of gate dielectric induced by the sputtering deposition of the gate electrode is also reported [59]. Reactive sputtering is the most commonly used technique in depositing thin film metal gates.

Metal gates such as TiN and Pt have been used with most of the high-κ gate dielectrics mentioned above to prevent reaction at the gate interface. Attempts have been made to use doped poly-Si gates with Ta₂O₅, by depositing a CVD SiO₂ reaction barrier, which is of lower electrical quality than thermal oxide, on top of the Ta₂O₅ layer [60]. The presence of SiO₂ at both the channel and gate interface predictably limited the device performance, and the lowest obtainable oxide equivalent was $t_{eq} = 23$ Å. Even initial attempts to use poly-Si gates with ZrO₂ [61, 62] have been unsuccessful as reaction layers have been observed at the interface. Al₂O₃ has been shown to be stable with respect to reaction with the poly-Si gates throughout typical CMOS processing, however, both boron and phosphorous dopant diffusion have been observed with Al₂O₃ gate dielectrics, which cause significant, undesired shifts of flat band voltage, $V_{FB}$, and threshold voltage, $V_T$, values [63, 64].
As mentioned earlier, replacing the polysilicon in the gate typically requires a dual metal approach; a metal with a PMOS work function and a second with NMOS work function. The work function ($\Phi_m$) of the metal must be near the conduction band edge and valance band edge of silicon for NMOS and PMOS [65], respectively, in order to satisfy the low power and low supply voltage demand. Ideally, metal gate electrodes will have work functions that are within 0.2 eV of the silicon conduction and valance band edges, namely work functions of 5.0-5.2 and 4.1-4.3 eV, for PMOS and NMOS, respectively [66, 67]. However, metals with midgap work functions (4.6-4.8 eV), such as CoSi$_2$ and NiSi, are also being considered since they can be more easily integrated into the process flow [68-71]. The work function of reactive sputtered Ti$_{1-x}$Al$_x$N$_y$ showed a range from 4.36 to 5.13 eV with a nitrogen partial flow [72]. Recently, TaSiN [73] and RuO$_2$ [74] films were proposed for n-type and p-type gate electrodes respectively. While the TaSiN gate showed good thermal stability on SiO$_2$ up to 1000 °C, the p-type candidate brought out some issues due to its volatile and unstable features at high temperatures [75, 76]. There are publications describing TaN, TiN, W, W/TiN, Mo, Pt, Ta/Pt as gate electrodes [77-82]. Electrical properties of Ru based alloy gate electrodes for dual metal gate CMOS have been recently reported [83].

Investigation of various metal gates and the study of their materials and electrical properties is a major focus of this dissertation and results will be presented in the later chapter. The following sub-section introduces the need for high mobility channel materials and describes the current stat of the art for Si devices in terms of strain engineering. It then briefly discusses the opportunities and challenges with Ge channel devices.
2.4 Germanium Devices

Germanium, the semiconductor material used by Bardeen and Brattain to develop the first point-contact transistor [84] in 1947, was the material of choice for many years due to it having the best crystalline quality achievable at that time. For more than a decade, till the late 1950s – early 1960s, Ge had been the workhorse for the development of bipolar transistors. To provide a historical perspective, Fig. 2.4 shows the image of the first transistor invented at Bell Laboratories and Fig. 2.5 shows the image of first integrated circuit. The two point contacts at the bottom of the triangular quartz crystal were made from two strips of gold foil separated by about 50 µm and pressed onto the Ge surface. With one gold contact forward biased and the other reversed biased, the *transistor* action was observed and this was the beginning of what is know as the *electronic age*.

With the development of field effect transistors in late 1960s, Ge was considered secondary to silicon and was largely abandoned by the semiconductor industry, mainly due to its unfavorable surface properties and to the fact that, in contrast with Si, it has no stable natural oxide required for MOS operation. At present Ge is now making a comeback as an enabling material for the next generation of device technology, primarily due to its high carrier mobility. Ge along with other high mobility materials such as III/V compounds are projected to be the new range of semiconductor materials, that will provide the platform for continued performance improvement at the 22 nm technology node and beyond. In the following sub section, an attempt has been made to present the current status of silicon mobility enhancement and future perspective for Ge and III/V materials for continued CMOS scaling.
Figure 2.4: The first point contact transistor invented in 1947 at Bell laboratories.

Figure 2.5: First integrated circuit invented in late 1950s.
2.4.1 Mobility Enhancement in Silicon

It is often said that, engineers perform better under stress. Is it true for the silicon too? Well, in the last 5 years there has been an exponential increase in the research on silicon mobility enhancement through strain engineering and results have been encouraging [85]. For the past four decades the industry has been chasing Moore’s Law primarily by miniaturizing (scaling) the silicon device. We are now entering the nanometer regime at the 35 nm node. The gate oxide thickness is dipping below 1 nm and physical limitations such as off-state leakage current and power density pose a real threat to performance enhancements derived simply by means of geometrical scaling [86]. While high-κ dielectric and metal gates are areas of active research, strain engineering has quickly emerged as a new scaling vector for performance enhancement to extend the life of traditional silicon technology.

Straining silicon involves ways of manipulating the silicon channel by altering the crystal lattice so as to improve carrier mobility. Stretching or compressing the silicon crystal lattice alters its electronic band structure [87]. Due to the crystal symmetry of Si, the mobility of electrons and holes varies with the direction of the applied electric field and, therefore, the performance of MOSFETs fabricated in Si depends on the crystal orientation of the substrate. Strained Si enhances the performance of CMOS devices solely by increasing the carrier mobility without having to make them smaller.

The change in carrier speed depends on the direction of strain as well as the type of channel. The two major types of induced strain are biaxial and uniaxial. Fundamental understanding of the mobility enhancement and the behavior of electrons and holes under the influences of applied strain, biaxial or uniaxial in various directions, is still an active research
area [88]. For NMOS devices, longitudinal tensile strain (strain along the channel, making it longer) allows electrons to move more quickly and smoothly, increasing the drive current and thus improving transistor-switching speed. In the case of PMOS devices compressive strain along the channel (making it shorter) is the best way to enhance mobility. Strain influences each type of charge carrier (holes and electrons) differently [89]. We can broadly classify the current straining technology into three major categories, namely pre-process strain, process induced strain and post-process strain.

Pre-process or substrate strain introduces global strain into the silicon prior to any device fabrication. The most established global strain method involves the use of a SiGe (silicon-germanium) alloy. In Biaxial Tensile Strain, the interatomic distances in the silicon crystal are stretched, generally increasing the mobility of electrons making n-type transistors faster. p-type transistors, in which holes are the main charge carriers, are not significantly enhanced with tensile stress and in some cases their mobility may even be reduced. On the other hand the use of compressive stress, where the interatomic distances are shortened, can dramatically increase the mobility of holes.

Most process induced strain is uniaxial. The main advantage of uniaxial compressive stress to the biaxial counterpart is that the hole mobility enhancement does not degrade at high vertical fields where commercial MOSFETs operate. Several approaches have been attempted by IC manufacturers to incorporate uniaxial strain into their current fabrication processes. In addition to using the substrate to provide channel strain, strain effects induced during the fabrication process are also used to enhance the mobility. The most successful process induced strain technique is the use of embedded SiGe in the source/drain regions of the transistor. Other
techniques include shallow-trench isolation (STI), silicidation processes, and addition of a silicon nitride (Si$_3$N$_4$) cap layer.

A strain effect induced after the fabrication process can also be used to enhance channel mobility, and involves straining the silicon lattice after the IC chips have been completely processed. The process starts with thinning the processed wafer to ultra-thin dimensions of less than 10 µm, followed by transferring the ultra-thin silicon layer to a polymer film and then mechanically straining the silicon membrane. After straining, the membrane is glued to a final substrate using an adhesive where it is held in tension [90].

![Diagram of hybrid-orientation technology (HOT)](image)

Figure 2.6: (a) HOT implemented on SOI (b) HOT implemented on bulk silicon.

In addition to the aforementioned techniques, orientation of the transistors on the silicon substrate can also be used to improve the overall performance of strained Si transistors. The current microelectronics industry standard is to fabricate CMOS chips on (100) orientated silicon. Detailed research work has indicated that other crystallographic orientations can be used to enhance the transistor output characteristics. Recently IBM has introduced the so-called *hybrid-orientation technology* (HOT), that introduces mixed-crystal orientations on a single
wafer to take advantage of the fact that PMOS transistors operate best when fabricated on silicon with a (110) orientation, while NMOS transistors operate best with a (100) orientation [91]. Both, silicon on insulator (SOI) and bulk versions of HOT have been recently demonstrated [92] as shown in Fig. 2.6.

### 2.4.2 Germanium: Challenges and Opportunities

Even with the implementation of strain engineering, high-k dielectrics and metal gates, the International Technology Roadmap for Semiconductors (ITRS) predicts that semiconductors with even higher mobility will be needed to continue scaling and performance improvements beyond the 22 nm technology node. In this regards, Ge is receiving a lot of attention, as it is relatively compatible with Si-technology, and has much higher carrier mobility (~2x for electrons and 4x for holes). Many of the III/V compound semiconductors such as InSb, InAs, InGaAs, GaAs, etc. have been known to have even higher electron mobility (of the order of 10,000 cm²/V-s) and have been studied for optical applications or high-frequency devices for quite some time now. While there are many compound semiconductor materials available that have a high electron mobility, materials with good hole mobility, necessary to make p-MOS devices, are more difficult to find, making Ge the prime material choice for this application. Table 2.5 provides comparison of some of the important properties of Si, Ge and GaAs.

While it is recognized that Ge has a lot of potential as a high mobility channel material for future MOSFETs, it also became clear that Ge, with its low bandgap (0.67 eV) and low melting point (937 °C) not only poses challenges for device design, but also for process integration. The fact that Ge wafers have poor mechanical strength and are much more expensive
than Si wafers is an important question, which might be answered by using Ge on a Si platform, such as Ge on insulator (GOI). Before Ge can be considered for mainstream CMOS integration, several critical issues have to be resolved. The following discussion presents some of these key challenges and possible solutions for Ge channel devices.

Table 2.5: Comparison of some basic properties of Si, Ge and GaAs at 300 K.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Si</th>
<th>Ge</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoms/cm³</td>
<td>$5.02 \times 10^{22}$</td>
<td>$4.42 \times 10^{22}$</td>
<td>$4.42 \times 10^{22}$</td>
</tr>
<tr>
<td>Atomic weight</td>
<td>28.09</td>
<td>72.6</td>
<td>144.63</td>
</tr>
<tr>
<td>Breakdown field (V/cm)</td>
<td>$\sim 3 \times 10^5$</td>
<td>$\sim 1 \times 10^5$</td>
<td>$\sim 4 \times 10^5$</td>
</tr>
<tr>
<td>Crystal structure</td>
<td>Diamond</td>
<td>Diamond</td>
<td>Zincblende</td>
</tr>
<tr>
<td>Density (g/cm³)</td>
<td>2.329</td>
<td>5.326</td>
<td>5.317</td>
</tr>
<tr>
<td>Dielectric constant</td>
<td>11.9</td>
<td>16.0</td>
<td>13.1</td>
</tr>
<tr>
<td>Effective density of states in conduction band, $N_c$ (cm⁻³)</td>
<td>$2.86 \times 10^{19}$</td>
<td>$1.04 \times 10^{19}$</td>
<td>$4.7 \times 10^{17}$</td>
</tr>
<tr>
<td>Effective density of states in valance band, $N_v$ (cm⁻³)</td>
<td>$1.04 \times 10^{19}$</td>
<td>$6.0 \times 10^{18}$</td>
<td>$7.0 \times 10^{18}$</td>
</tr>
<tr>
<td>Optical phonon energy (eV)</td>
<td>0.063</td>
<td>0.037</td>
<td>0.035</td>
</tr>
<tr>
<td>Effective mass (conductivity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrons ($m_e/m_o$)</td>
<td>0.26</td>
<td>0.082</td>
<td>0.067</td>
</tr>
<tr>
<td>Holes ($m_p/m_o$)</td>
<td>0.69</td>
<td>0.28</td>
<td>0.57</td>
</tr>
<tr>
<td>Electron affinity, $\chi$ (V)</td>
<td>4.05</td>
<td>4.0</td>
<td>4.07</td>
</tr>
<tr>
<td>Energy Gap (eV)</td>
<td>1.12</td>
<td>0.67</td>
<td>1.42</td>
</tr>
<tr>
<td>Intrinsic carrier concentration (cm⁻³)</td>
<td>$1.45 \times 10^{10}$</td>
<td>$2.4 \times 10^{13}$</td>
<td>$1.8 \times 10^6$</td>
</tr>
<tr>
<td>Intrinsic resistivity (Ω·cm)</td>
<td>$2.3 \times 10^8$</td>
<td>47</td>
<td>$10^8$</td>
</tr>
<tr>
<td>Lattice constant (Å)</td>
<td>5.431</td>
<td>5.646</td>
<td>5.653</td>
</tr>
<tr>
<td>Melting point (°C)</td>
<td>1415</td>
<td>937</td>
<td>1240</td>
</tr>
<tr>
<td>Minority carrier lifetime (s)</td>
<td>$2.5 \times 10^{-3}$</td>
<td>$10^{-5}$</td>
<td>$\sim 10^{-8}$</td>
</tr>
<tr>
<td>Mobility (cm²/V·s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrons ($\mu_n$)</td>
<td>1500</td>
<td>3900</td>
<td>8500</td>
</tr>
<tr>
<td>Holes ($\mu_p$)</td>
<td>450</td>
<td>1900</td>
<td>450</td>
</tr>
<tr>
<td>Thermal Diffusivity (cm²/sec)</td>
<td>0.9</td>
<td>0.36</td>
<td>0.24</td>
</tr>
<tr>
<td>Thermal conductivity (W/cm⁻²°C)</td>
<td>1.5</td>
<td>0.6</td>
<td>0.46</td>
</tr>
</tbody>
</table>
One of the reasons for abandoning Ge in the early sixties was the lack of a stable native oxide. GeO$_2$ thermally desorbs above 420 °C and is water soluble. GeO also desorbs below 450 °C. A possible solution for this is the use of Ge oxynitride (GeON) where the introduction of nitrogen in the layer results in improved thermal and chemical stability. During the past ten years or so, much work has been done to introduce a deposited high-κ dielectric as a replacement for the thermally grown SiO$_2$ to reduce gate leakage in conventional Si CMOS technology. The introduction of a deposited high-κ dielectric provides an exciting opportunity for Ge channel devices, where in the past this was not an option [93].

An important challenging task for Ge high-κ MOS devices is the Ge surface preparation and the control of the interface between the deposited high-κ layer and the Ge channel. The conventional hydrogen terminated surface passivation technique, commonly used for Si, appears to be less effective for Ge. Several different techniques such as, desorption of the Ge native oxide in ultra high vacuum system at high temperature followed by in situ deposition of high-κ, surface nitration by direct deposition using either atomic N exposure or a high temperature NH$_3$ gas treatment or by using very thin layer of epitaxially grown strained Si to form Si-SiO$_2$ interface between high-κ and Ge channel, are actively being investigated [94].

Owing to low melting temperature of Ge, use of metal gates is essential as conventional poly-Si gate electrodes require high temperature (>900 °C) dopant activation anneals. In the junction formation process the diffusion of p-type dopants such as boron is suppressed while the diffusion of n-type dopants such as P is enhanced in Ge compared to bulk Si. This favors the formation of ultra-shallow junctions in p-channel Ge MOSFETs, while presenting a challenge for shallow-junction formation in n-channel Ge MOSFETs.
With the currently available optimized gate stack, surface passivation and junction formation processes, very good results have been demonstrated recently for PMOS devices, holding the promise that these devices can outperform their Si based counterparts [95]. However, the fabrication of good NMOS devices has been less successful so far. There are studies suggesting that the main reason for this seems to be closely related to the presence of specific surface defects that directly degrade the channel mobility and limits the drive current.

Therefore it is an attractive proposition to use high electron mobility III/V compounds such as (In)GaAs for NMOS devices and Ge for PMOS devices in order to come to a CMOS technology. Most III-V materials show very high bulk electron mobility, compared to the silicon electron mobility. GaAs is an interesting candidate since it has almost the same lattice parameter as Ge and therefore can be defect-free epitaxially grown directly on Ge. Researchers have already demonstrated the feasibility of selective growth of GaAs based materials with low defect density on Si wafers through the use of a Ge interlayer. Although many encouraging results have already been obtained, extensive research efforts are still needed in order to achieve good surface passivation and high-κ dielectric deposition on these III/V substrates, which is essential to make devices with low gate leakage.
CHAPTER THREE: METHODOLOGY

In this chapter the experimental procedures and techniques used to fabricate and characterize the MOS devices for metal gates and Ge channel are reported. This chapter also outlines various materials and electrical characterization techniques used to study these devices. The chapter is broadly divided into two main sub-sections; one on metal gates investigations and the other on Ge channel device studies, with both having different experimental procedures for device fabrication and characterization.

3.1 Experimental Study: Metal Gates

There are several criteria that must be considered in the selection of metal gate electrodes. Three of the most important ones are:

(i) the metal work function,

(ii) the thermal/chemical stability of the metal in contact with the dielectric at temperatures relevant to CMOS processing, and

(iii) integration challenges.

The work function and thermal and chemical stability are discussed in section 2.3.1 and 2.3.2. The integration challenges for metal gates are discussed below.

The replacement of silicon dioxide or nitrided oxide with a high-\( \kappa \) dielectric does not change the conventional CMOS process flow, but the incorporation of two different work function metals (for PMOS and NMOS devices on the same wafer) does increase the integration complexity [96]. This is because dual metal gates will require two metal deposition steps, as opposed to the single step required in dual polysilicon gate technology. In addition, depending on
the exact integration scheme, non-critical lithography steps are required to remove the first metal or the second metal by means of selective etching (dry or wet) or chemical mechanical polishing [97]. It may be necessary to develop a scheme that will allow etching of both the n-type and p-type metals in the gate stack. For example, in the scheme reported by Zhang et al. [98], after the deposition of HfO$_2$ by atomic layer deposition (ALD), the NMOS metal TaSi$_N$$_x$ was deposited by physical vapor deposition (PVD). Then, using a hard mask, the p-MOS regions were opened up and the TaSiNx was removed by a wet etch that was highly selective to HfO$_2$. After removal of the remainder of the hard mask, Ru was deposited as the PMOS metal and a plasma etch process was used to etch the dual metal gate stack simultaneously, yielding Ru layers for PMOS and TaSi$_N$$_x$/Ru layers for the NMOS. In another integration scheme, Ru/Ta and RuTa/Ru bilayer stacks were deposited and the top metal layer was removed in selected regions. Annealing induced mixing of the two layers allowed the work function to become n-type (Ru/Ta) or p-type (RuTa/Ru) as needed, while the regions with single metal layers (Ru, or RuTa) provided the opposite type of work function [99].

Other integration criteria include, but are not limited to: (i) the ease of hydrogen diffusion through the material for dielectric interface passivation [100], (ii) the diffusivity of oxygen through the gate electrode and the resultant (unwanted) increase in EOT, and (iii) the deposition method of the metal to avoid extensive damage to the underlying dielectric. Hence, plasma-free deposition processes such as atomic layer deposition (ALD), chemical vapor deposition (CVD) or thermal evaporation of gate electrodes are preferred in order to minimize the damage to the gate dielectric. In this respect, Ru is an attractive choice since it can be deposited by ALD [101], and Ru and Ru-Ta alloys allow adequate hydrogen diffusion for dielectric passivation upon annealing in forming gas (H$_2$ + N$_2$ gas mixtures) [102].

44
Ultimately, it will likely be the integration challenges, far more complex and extensive than presented here, that will determine which metal gate electrodes are implemented in future CMOS technology. However, the selection of materials for metal gates depends heavily on the availability of detailed materials and electrical data. Such data include alloy work function, crystal structure (i.e., phase identity), grain structure, texture, thermal and chemical stability, etc. While the relationship of alloy composition to work function, the chemical stability of alloys in contact with silicon dioxide, and the fabrication and electrical characterization of dual metal gate FETs have been studied for some alloys, our scientific understanding of the gate metal is far from complete. Thus, it is the aim of the proposed work to significantly expand our knowledge base in alloys for dual metal gates by carrying out detailed electrical and materials studies in the binary alloy systems of Ru with p-FET metal, namely Ru-Pt and n-FET metals, namely Ru-Ta, Ru-Hf and Ru-Nb. These alloy systems are chosen, in part, for their technological importance, but also for their scientific relevance. Our rationale for choosing these three alloy systems is discussed in the next section.

3.1.1 Selection of Metal Alloys

The element Ru has an attractive combination of characteristics that make it a strong candidate for PMOS devices and hence a worthwhile choice for our proposed studies. These characteristics are that ruthenium

(i) can be deposited by ALD,

(ii) can be patterned by dry etching,

(iii) is thermally stable for anneals up to 1000 °C,
(iv) has high diffusivity for hydrogen and low diffusivity for oxygen,
(v) has low resistivity \((7.3 \, \mu \Omega \text{cm, bulk Ru})\), and
(vi) forms strongly \(<0001>\) fiber textured films (i.e., the basal planes (0002) in the grains are parallel to the film plane - noting that Ru has the A3, i.e., hexagonal close packed, crystal structure).

Our understanding of the band structure of metals leads immediately to an understanding that a metal’s work function is dependent upon crystal orientation, and this is clearly understood in ultrahigh vacuum measurements of metallic work functions, but it is rarely considered in CMOS technology. The dependence can be large, as seen, for example, for Ta, which has a reported overall work function of 4.25 eV, but for which the reported work function is 4.00 eV for the (111) plane and 4.80 eV for the (110) plane [103]. This is a difference of 0.8 eV and nearly spans the n- to p-type work function regime of interest for CMOS. Two approaches to eliminating or minimizing the local grain to grain orientation and the associated work function variation are to use either an amorphous metal gate electrode or a highly textured polycrystalline one. Since most deposited metals form as polycrystalline films, a promising solution is to identify metals (or metal alloys) that have the tendency to form highly textured films. Ru is one such metal, making it well suited for polycrystalline metal gates and for the studies proposed here.

Considering the periodic table of elements to select the possible metal candidates for dual workfunction metal gate electrodes, we have the following options:
From Fig. 3.1 it is obvious that there are too many choices to select from. Thus to eliminate few options, we also consider the thermodynamic stability of these metals on the two most promising high-κ materials, Al₂O₃ and HfO₂.

The thermodynamic stability of various gate metals can be determined by considering enthalpies of formation for their oxides. Thus by simple means we can eliminate Ti, Zr, Hf and Al from possible choices.
Figure 3.2: Enthalpies per mole of oxygen for the most stable oxides for gate metal materials. Range (± 10%) for the enthalpies of $\text{Al}_2\text{O}_3$ and $\text{HfO}_2$ are indicated for comparison.

The melting points of the various gate metals are shown in Fig. 3.3, and these represent another selection criteria for the gate electrode metal. For a conventional CMOS process these materials would have to withstand 1000°C, several seconds anneals on the dielectrics during the S/D dopant activation. At such high processing temperatures thin films of metallic materials can agglomerate. Agglomeration occurs when grain boundary energies overcome surface or interface energies, leading to roughening of the thin films. It can be decreased by keeping the grains small (lower processing temperatures), increasing the film thickness, decreasing the grain boundary energies (adding impurities), or increasing surface and/or interface energies with respect to grain boundary energies, thus reducing diffusivities (adding a capping layer). Agglomeration of the
materials can be expected if the homogeneous temperature (anneal temperature divided by the
melting point) gets close to 2/3. Therefore, those materials with melting points less than 1500 °C
can be expected to be thermally unstable during 1000°C activation anneals.

![Melting Points Graph](image)

**Figure 3.3:** Melting point (°C) for the gate metal materials.

Thus by further eliminating metals like Mn, Co, Ni, we narrow down our choices to the
following few options.

NMOS: V, Nb, and Ta

PMOS: Ru, Rh, Re, Ir and Pt

Since most of the PMOS metals are noble metals, We have select Ru and Pt for our studies.

Given our choice of Ru as the p-type metal of interest for PMOS devices, we plan to
examine binary alloys of Ru with n-type metals to form n-type alloys. The integration
advantages of binary metal alloys that combine an n-type metal and a p-type metal for NMOS
devices were discussed above. It was also noted that by judicious choice of the alloying elements and alloy composition, thermally and chemically stable metals with n-type work functions can be identified. In the present work, Ru alloys with Pt, Ta, V and Nb have been investigated. Fig. 3.4 shows the binary phase diagram of Pt-Ru.

![Pt-Ru Phase Diagram](image)

Figure 3.4: Pt-Ru binary alloy phase diagram.

When the composition of the alloy has a Ru content less than 62% at.% a FCC Pt phase is expected and when greater than 80% at.% a HCP Ru phase is expected, based on the equilibrium phase diagram as shown in Fig.3.4. Both FCC and HCP phases of Pt-Ru are expected to be present when the Ru content is between 62% and 80% at.% A detailed study of the resulting phases is important, as they can be expected to impact the important properties of the gate electrode material, such as its work function and thermal stability. Therefore, for alloy compositions with at.% Ru less than 62%, Pt like properties are expected such as large work function and excellent thermal stability. Similarly, films with at.% Ru greater than 80% are
expected to exhibit Ru like properties. However, when the at.% Ru concentration is between 62% and 80% a new mixed phase region is obtained whose electrical properties are extensively studied in this work.

In the Ru-Ta system, the region of the phase diagram where the RuTa compound, with a reported work function of 4.3 eV is obtained corresponds to the approximate middle of the composition range (Fig. 3.5). At the 50:50 composition, the RuTa phase with a B2 ordered crystal structure is thermodynamically stable only above approximately 1100 °C. At lower temperatures, the RuTa’ phase with the L10 ordered crystal structure is the stable phase, and, below 800 °C, it is a third polymorph, namely the orthorhombic RuTa’’ phase, that is the thermodynamically stable phase. Further, with the ability of phases to exist metastably, the Ru-Ta binary system is even more complex than is apparent from the equilibrium phase diagram in Fig. 3.5. Since a change of phase can alter the work function and shift the transistor threshold voltage, the question of what phases are actually present and how stable they are during thermal treatment deserves attention. In contrast to the Ru-Ta system, the Ru-V system is significantly simpler. Intermediate in complexity between these two systems is the Ru-Nb system that has two polymorphs, one with the B2 structure and one with the L10 structure at/near the 50:50 compositions. These two polymorphs have a wide compositional stability range (~17 at% for B2). By investigating these binary systems, we will be able to assess the importance of the compositional homogeneity range, the number of polymorphs and the associated polymorphic transformations on alloy work function and resistivity.
Figure 3.5: Ru-Ta binary alloy phase diagram.

Figure 3.6: Ru-V binary alloy phase diagram.
3.1.2 Film deposition

Silicon substrates (n-type) were used for the experiments. In order to remove the native oxide layer on top of the Si wafer, chemical etching was done for some of the Si (100) wafers by immersing the wafer in a 9:1 buffered oxide etch solution for 30 seconds. Contaminants present at the surface of the silicon wafers at the start of processing, or accumulated during processing have to be removed at specific processing steps in order to obtain high performance and high reliability semiconductor devices and to prevent contamination of semiconductor equipment. SC1 and SC2 cleaning (modified RCA cleaning) was used to remove organic, ionic and heavy metal contaminants from the silicon wafers.
The SC1, SC2 procedure has two major steps executed sequentially:

2. Removal of ionic and heavy metal atomic components using a solution of 6:1:1 H₂O/H₂O₂/HCl solution.

The SC1 solution was prepared by heating the solution of NH₄OH and H₂O in the ratio of 6:1 till a temperature of 75 °C. 1 part of H₂O₂ was then added to the solution just prior to immersing the Si wafers. The wafers were cleaned for 15 minutes. This was followed by a DI water rinse for 5 minutes.

For SC2, the SC2 solution was prepared by heating the HCl and H₂O in the ratio 6:1 to 75°C. H₂O₂ was added just prior to immersing the SC1 cleaned wafers in the solution. A 15 minutes SC2 clean was followed by 5 minutes DI rinse, after which the wafers were blown dry using a N₂ gun. Prior to any device processing all silicon substrates were cleaned using the same procedure.

Co-sputtering of Pt-Ru, Ta-Ru, Nb-Ru and V-Ru thin films were done on a thermally grown 100 nm thick SiO₂ layer on a Si substrate using ultra high vacuum dc magnetron sputtering. The system base pressure was approximately 1 X 10⁻⁸ Torr and the purity of the Ar process gas was maintained by a hot reactive metal getter. Process gas contamination in the deposition chamber at the typical deposition pressure of 4 mTorr was confirmed to be less than 10 ppm (instrumental sensitivity limit) by closed ion source quadrupole mass spectrometry. The deposition time for each run was kept fixed for 3 minutes while the film composition was varied by changing the dc target power form 0 W to 200 W in steps of 20 W. The target voltage was recorded for each deposition.
Table 3.1 summarizes the deposition details for Pt-Ru alloy series. Similar tables were generated for Ta-Ru, Nb-Ru and V-Ru. Target voltage versus power plot for each metal target is shown in Fig. 3.8. The voltage increases linearly with the power, indicating that the deposition rate is linearly proportional to the deposition power. The target voltage – power relationship is somewhat nonlinear for lower power range.

![Target voltage versus deposition power plot](image)

Figure 3.8: Target voltage versus deposition power plot.
3.1.3 Materials characterization

Alloy films in the thickness range of 20 to 50 nm were successfully deposited by the cosputtering procedure described above. Rutherford backscattering spectroscopy (RBS) was used to determine the deposition rate, thickness and composition of the alloy films. Transmission electron microscopy (TEM) was used to examine the grain size of the as-deposited films and later to study the effect of annealing on the grain growth. X-ray diffraction (XRD) was used to investigate the crystal structure and preferred orientation of the as-deposited and annealed films. For grain growth, XRD and resistivity measurements, the films were annealed in Ar+3% H₂ at 400 °C for four hours. All other electrical measurements, including MOS C-V and Schottky I-V and C-V were performed on un-annealed devices.

3.1.4 Electrical characterization

To extract the work function of the binary alloys, MOS capacitor structures were fabricated. The n-type silicon wafers were cleaned and etched by a buffered oxide etch diluted with de-ionized water for 3 minutes to remove any native oxide. The gate oxide was thermally grown in dry oxygen ambient to a thickness in the range of 20 Å to 150 Å, followed by a passivation annealing in hydrogen ambient. Aluminum was thermally evaporated, in a cryofilament evaporator, on the back side of the silicon wafer and annealed in Ar ambient at 500°C for 30 minutes to establish an ohmic contact. The metal alloys were deposited after back side Al ohmic contacts were established. HP 4192 impedance analyzer was used to measure high frequency C-V data at 1 MHz. The flat band voltage, \( V_{FB} \), was extracted from the C-V characteristics based on the North Carolina State University (NCSU) C-V program [104] and
plotted against the effective oxide thickness \((EOT)\) to determine the work function for the binary alloy system.

The definition of work function is the energy required to move an electron from the bulk of the material (at the Fermi level \(E_F\)) to a point just outside the surface (at the vacuum level \(E_o\)). By definition, this property of the material is independent of the surface and is essentially a bulk property. This is an idealistic definition, however, an accurate values can be difficult to obtain. For example, because of reconstruction at the surface or surface impurities, there will be differences in the surface energy, which make the work function dependent on surface orientation and surface impurities.

The work function difference between the metal and semiconductor in the MOS system is defined as,

\[
\Phi_{ms} = -(\Phi_s - \Phi_m) = \Phi_m - \Phi_s \quad (3.1)
\]

which neglects the metal/insulator and insulator/semiconductor interfaces and the attendant charge redistribution that may occur at the interfaces. In the energy band diagram in Fig.3.9, \(E_c\) and \(E_v\) denote the conduction and valence bandedges, respectively, \(E_i\) is the intrinsic Fermi level and \(E_{FS}\) and \(E_{FM}\) are the Fermi levels in silicon and metal. At flatband,

\[
q\Phi_s = E_{so} + E_{VL} - E_g/2 - q\Phi_B \quad (3.2)
\]

\[
q\Phi_m = E_{mo} + E_{VL} \quad (3.3)
\]

where \(E_{VL}\) is the vacuum energy level measured with respect to the conduction band edge of the oxide. \(E_{so}\) and \(E_{mo}\) are zero field barriers from the silicon valence band and metal Fermi level to the SiO\(_2\) conduction band, respectively and \(E_g\) is the energy gap. The bulk potential \(\Phi_B\) is defined as,

\[
\Phi_B = kT/q \ln(N_D/\eta_i) \quad ............ \text{ n type} \quad (3.4)
\]
\[ \Phi_B = \frac{kT}{q} \ln \left( \eta_i / N_A \right) \]  

Here \( \eta_i \) is the intrinsic carrier concentration and \( N_A \) and \( N_D \) are p- and n-type doping concentrations.

Figure 3.9: Energy Band diagram of a MOS system, (a) \( V_G = 0 \), (b) Flatband.

From the above expressions,

\[ q\Phi_{ms} = E_{mo} - E_{so} + \frac{E_g}{2} + q\Phi_B \]

\[ = q\Phi_{ms}^* - \frac{E_g}{2} + q\Phi_B \]  

The new quantity \( q\Phi_{ms}^* \) is independent of doping but related to \( \Phi_{ms} \) through

\[ q\Phi_{ms}^* = E_{mo} - E_{so} + E_g = q\Phi_m - q\chi_s \]  

where \( \chi_s \) is the electron affinity of the semiconductor. This new quantity is useful when comparing the data obtained from wafers of varying doping levels.
Φ<sub>ms</sub> leads to a charge redistribution in the MOS system. This implies a voltage drop \( V_{ox} \) across the oxide and a voltage drop \( \Psi_s \) across the semiconductor as shown in Fig.3.9. A counterbalancing gate bias is necessary to produce the flatband condition \( (V_{ox} = \Psi_s = 0) \) where \( V_{ox} \) and \( \Psi_s \) are potentials across the SiO\(_2\) and silicon, respectively. This gate bias is known as flatband voltage \( V_{FB} \).

The flatband voltage in the MOS structure is given by,

\[
V_{FB} - \Phi_{ms} = - \frac{Q_{surf}}{C_{ox}} \tag{3.8}
\]

\[
V_{FB} = \Phi_{ms} - \frac{Q_{surf}}{\varepsilon_{ox}} \left( \frac{t_{ox}}{\varepsilon_{ox}} \right) \tag{3.9}
\]

where \( Q_{surf} \) is surface charge, \( C_{ox} \) is oxide capacitance, \( t_{ox} \) is oxide thickness and \( \varepsilon_{ox} \) is the permittivity of the oxide. A plot of \( V_{FB} \) vs \( t_{ox} \) yields an intercept of \( \Phi_{ms} \).

If the work function difference is solely a bulk property, it should be independent of processing conditions or crystal orientation; it should depend only on \( \Phi_m \), \( \chi_s \) and \( N_D \) or \( N_A \).

The accurate determination of the metal function, as shown in Fig. 3.9, requires a careful consideration of interface states and bulk trapped charges in the dielectric layer that may also influence the capacitance transition voltage of a MOS device. The characterization of a series of such MOS capacitors fabricated with different gate oxide thicknesses is required to accurately quantify the gate metal work function in the presence of these and other effects, and this characterization also provides other properties such as gate dielectric leakage current, EOT, and trapped charge densities.

The workfunction is also determined from the Schottky barrier height, extracted from the \( I-V \) and \( C-V \) characteristics of the Schottky barrier diode. It is well known that the Schottky barrier height depends upon the quality of the semiconductor material, the method of surface preparation prior to metallization and the metallization itself. The n-type silicon wafers were
cleaned and the back side aluminum ohmic contact was first established as described above. The silicon surface was hydrogen terminated by HF just before introducing the wafer in the vacuum chamber for sputter deposition of Pt-Ru alloy metals to form the Schottky diode. *I-V* characteristics of the Pt-Ru Schottky contacts were obtained under dark conditions, at room temperature (295 K), using Keithley Instruments picoammeter/voltage source model 6487. A Keithley Instruments *C-V* analyzer model 590 was used to measure the capacitance–voltage characteristics of the Pt-Ru Schottky contacts, and the barrier height was extracted to calculate the metal work function.

Under the Schottky-Mott theory it is assumed that ideal conditions are obtained when forming the intimate contact with the metal and semiconductor and both dipole surface values remain unchanged. We also assume that no chemical reactions take place, and no physical strains are created during the contact procedure of the metal and semiconductor. Knowing the electron affinity of the semiconductor material and solving for the Schottky barrier height will lead us to determination of the work function of Pt-Ru alloy compositions, deposited on silicon substrate.

### 3.2 Experimental Study: Ge Channel Devices

It is believed that below 45-nm node the conventional bulk CMOS can be scaled, however, without appreciable performance gains. To continue the scaling of Si CMOS in the sub-45 nm regime, innovative device structures and new materials have to be created in order to continue the historical progress in device performance. Examples of novel device structures currently being investigated have been discussed in chapter-2. The examples of novel materials are; i) high mobility channel materials like strained Si and Ge, ii) high κ gate dielectrics and iii)
metal gate electrodes. Heterogeneous integration of these materials on Si with novel device structures will be needed to drive the technology node beyond 22 nm. This will require new fabrication technology solutions that are generally compatible with current and forecasted Si manufacturing. In this sub-section we describe a silicon compatible Ge channel device fabrication process flow and the characterization of mobility and low frequency 1/f noise.

3.2.1 Device Fabrication

The need for electron and hole mobility enhancement and the progress in the development of high-κ gate stacks, has lead to renewed interest in Ge MOSFETs. A key challenge is to develop a proper Ge surface passivation technique to obtain low interface state density and high carrier mobility. Recent publications have proposed several options such as NH$_3$ anneal [105, 106], SiH$_4$ anneal [107] or plasma treatment in PH$_3$ [108] with mixed success. In the current work, we have used a thin epitaxial Si layer between Ge and high κ layer to develop an effective Ge surface passivation technique.

The gate dielectric stack process consists of several steps, as shown in Table 3.2. To allow epitaxial Si growth, the Ge native oxide and other residues are first removed from the Ge surface by a HF 2 % dip followed by an anneal in a H$_2$ ambient in the epi-reactor. Immediately after the H$_2$ anneal, the Si layer is grown using SiH$_4$ under a N$_2$ ambient at 40 Torr. By varying the temperature and time of the anneal, the thickness of the Si layer can be precisely controlled within ± 2 e$^{14}$ deposited Si atoms/cm$^2$ (one monolayer of Si atoms is defined in this work as 6.5 e$^{14}$ atoms/cm$^2$, corresponding to a Si layer thickness of 1.2 Å). A good thickness control will prove crucial to achieve optimal C-V characteristics, as will be shown further.
Table 3.2: List of Gate dielectric process steps.

<table>
<thead>
<tr>
<th>Step Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 s HF 2 % dip</td>
</tr>
<tr>
<td>60 s, 650 °C H₂ anneal</td>
</tr>
<tr>
<td>500 °C/575 °C Si growth with SiH₄ in N₂ ambient</td>
</tr>
<tr>
<td>Room temperature N₂O plasma oxidation</td>
</tr>
<tr>
<td>300 °C ALD HfO₂ dielectric deposition</td>
</tr>
</tbody>
</table>

Once the Si layer is grown on the Ge substrate, the techniques developed for the high-κ dielectric stack on Si can be re-used. In this work, the Si layer is partially oxidized in a N₂O plasma at room temperature. Finally, a 300 °C ALD HfO₂ gate dielectric is deposited. A schematic of the final multi-layer gate dielectric stack with epitaxial Si interfacial layer is shown in Fig. 3.10.

Figure 3.10: Schematic of the final multi-layer gate dielectric stack with epitaxial Si interfacial layer.
PMOS devices are fabricated using a Si – compatible process flow using 200 mm (100) Ge-on-Si wafers as starting material. The process flow is described in Table 3.3. The Ge top layer of these wafers is grown epitaxially directly on Si and, to minimize the threading dislocation density at the surface the Ge layer, it is grown 2 µm thick. The process starts with p-well and channel implants, resulting in a channel surface doping of \( \sim 5 \times 10^{16} \text{ cm}^{-3} \). Field isolation is defined by a box isolation scheme. The gate deposition sequence starts with the Ge surface passivation, consisting of an ultra thin epitaxial Si layer which is partially oxidized after growth, as described above. To avoid further oxidation, this Si passivation layer is immediately capped by a 4 nm ALD HfO\(_2\) gate dielectric from an ASM pulsar 2000 reactor, followed by 10 nm TaN and 80 nm TiN PVD depositions. The resulting EOT derived from a high-frequency C-V measurement was 1.3 nm, corrected for quantization effects. Activation of the highly doped drain (HDD) and extension implantations (B) was achieved by rapid thermal anneal at 500 °C for 5 min. Finally NiGe source/drain regions are formed followed by a backend processing with TiN contact pads.
In addition to p-MOS, some shallow p⁺-n junction were also fabricated in 2 μm Ge layers epitaxially deposited on 200 mm diameter silicon wafers to investigate in forward operation the low frequency (LF) noise behavior. Both square and perimeter type diodes have been studied in order to have an idea about the geometrical origin of the noise. Junctions with and without a nickel-germanide contacts were fabricated and are compared. It is shown that devices with NiGe
exhibit a lower current noise spectral density \( S_1 \) at a frequency \( f=1 \) Hz compared with their non-
germanided counterparts. This is ascribed to the lower series resistance, leaving defect-assisted
carrier fluctuations as the main origin of the 1/f noise.

Junctions were formed in relaxed germanium epitaxial layers on silicon containing a
density of threading dislocations in the range of \( \sim 10^7 \) cm\(^{-2}\). A 570 keV P implantation to a dose of \( 5\cdot10^{11} \) cm\(^{-2}\) was employed to form the n-well region. This was followed by a 35 keV Ge pre-
amorphisation followed by 7.5 keV B to a dose of \( 4\cdot10^{15} \) cm\(^{-2}\). Junction activation was performed
by a 500 °C 5 minute anneal in N\(_2\) ambient. A layer of 30 nm Ni was sputtered, followed by a
rapid thermal anneal step at 350 °C to form a germanide. Non-germanided samples had a TiN-
based back-end metallization.

3.2.2 Device Characterization

Extensive device characterizations were performed to extract mobility data from the
standard I-V measurements and to understand the correlation between the low field mobility and
the oxide trap density derived from the 1/f noise magnitude.

3.2.2.1 Mobility Extraction

The mobility of carriers in the channel of a MOSFET is lower than in bulk
semiconductors because there are additional scattering mechanisms. Since carriers in the channel
are very close to the semiconductor-oxide-interface, they are scattered by surface roughness and
by coulombic interaction with fixed charges in the gate oxide. When the carriers travel in the
inversion layer from the source to the drain, they encounter microscopic roughness on an
atomistic scale at the oxide-semiconductor interface and undergo scattering because any deviation from a perfectly periodic crystal potential results in scattering. The mobility degradation increases with gate bias because a higher gate bias draws the carriers closer to the oxide-semiconductor interface, where they are more influenced by the interfacial roughness.

It is very interesting to note that if we plot the effective carrier mobility in the MOSFET as a function of the average transverse electric field in the middle of the inversion layer, we get what is know as a “universal” mobility degradation curve for any MOSFET, which is independent of the technology or device structural parameters such as oxide thickness and channel doping.

Effective carrier mobility, $\mu_{\text{eff}}$, is traditionally defined in terms of the measured drain current, $I_d$, of a MOSFET at low drain voltage, $V_d$, in the linear region as

$$I_d = \left( \frac{W}{L} \right) \mu_{\text{eff}} Q_{\text{inv}} V_d$$

(3.10)

where $W/L$ is the width/length ratio of the device. Effective mobility can thus be expressed as

$$\mu_{\text{eff}} = \frac{g_d}{\left( \frac{W}{L} \right) Q_{\text{inv}}}$$

(3.11)

where $g_d = I_d/V_d$ is the channel conductance evaluated for small drain voltages. For accurate extraction of effective mobility values, it is obvious that accurate values of $W/L$ and $Q_{\text{inv}}$ must be used in this expression. As lithography dimensions decrease, the uncertainty in $W$ and $L$ increases and large device geometries are typically used to increase the accuracy of this ratio.

The inversion layer charge must be accurately known in Eq. (3.11) to obtain an accurate evaluation of mobility. At sufficiently large voltages above the channel threshold voltage the inversion layer charge approaches,
\[ Q_{\text{inv}} \rightarrow C_{\text{ox}} (V_{gs} - V_T) \]  

(3.12)

where \( V_T \) is the threshold voltage. This approximation is frequently used in evaluating the mobility. There are basically two approaches to improve the accuracy of Eq. (3.12) for use in determining the effective mobility using Eq. (3.11). One is to perform a split capacitor measurement and obtain the inversion layer charge from the voltage integral of a measured split capacitance as originally described in Koomen [109]. This approach has been used by several investigators in combination with Eq. (3.11) to extract effective mobility values. A second approach is to theoretically calculate the inversion layer charge as a function of gate voltage using established semiconductor device equations for depletion layer charge and inversion layer charge. Basically if one knows oxide thickness (or oxide capacitance), bulk doping density and threshold voltage, \( Q_{\text{inv}} \) can be numerically evaluated at any desired effective surface field or gate voltage and combined with experimental \( I_d-V_{gs} \) data in Eq. (3.11) to extract the effective mobility.

HP 4146 Semiconductor parameter analyzers have been extensively used to measure \( I_d-V_{gs} \), and \( I_d-V_{ds} \) characteristics for Ge channel devices with various channel lengths and widths. We have also used split capacitance measurements and the extensive calculations to accurately determine the inversion charge, in order to extract the device channel effective mobility.

3.2.2.2 Low Frequency Noise Measurements

One of the tools to study the quality of the gate oxide that is gaining wider acceptance is the low-frequency (LF) noise. It is based on the so-called McWhorter theory, which was originally developed on germanium devices [110]. LF noise has been extensively used for the
characterization of high-$\kappa$ gate dielectrics. Improvements in the processing have led to progressively better performing p-MOSFETs, so that it is worthwhile to evaluate the LF noise of our state-of-the-art devices, with Si passivation, metal gate and HfO$_2$, corresponding with an EOT of 1.2 nm. It is shown that the LF noise in linear operation and for most of the gate voltage ($V_{gs}$) range studied is of the $1/f^\gamma$ type with $\gamma \sim 1$. In contrast to the case of metal-gate/HfO$_2$ p-MOSFETs on silicon, the dominant noise mechanism appears to be number fluctuations ($\Delta n$ model). At the same time, a good correlation is observed between the oxide trap density derived from $1/f$ noise ($D_{ot}$) and the interface state density ($D_{it}$), following from Charge-Pumping (CP) measurements on the same devices. The large device-to-device scatter found across the wafer indicates that there is still room for improvement by a better control of the high-$\kappa$ and interface defects. This will also reduce the scatter in the low-field mobility, which is shown to closely follow the scatter in $D_{it}$ and $D_{ot}$.

On-wafer LF noise measurements have been performed using the BTA9812 pre-amplifier and noise analyzer, combined with a HP35665A spectrum analyzer. W=10 $\mu$m wide p-MOSFETs were characterized in linear operation, i.e., corresponding with a drain voltage of -50 mV, whereby the gate voltage $V_{gs}$ was varied from weak to strong inversion. Around 10 different devices with length $L=1$ and 0.25 $\mu$m have been evaluated. The low-field mobility $\mu$ of the same devices has been extracted from a linear $I_d$-$V_{ds}$ characteristic. Also CP has been performed on the same transistors at a pulse frequency of 1, 2 and 4 MHz while varying the gate bias from -1 to +1.5 V with 1 V amplitude and fall and rise times of 50 ns.

Noise measurements were also performed on the p$^+$-n diodes in forward bias between the top junction and the bottom substrate in the bias range from 0 to 0.8 V in steps of 50 mV, using a probe station, BTA hardware and the NoisePro software from Cadence.
CHAPTER FOUR: RESULTS AND DISCUSSIONS: METAL GATES

This chapter presents results from the metal gates study. The chapter is mainly divided into two sub-sections, one on materials characterization and the other on electrical characterization.

4.1 Materials Characterization

This section presents results and discussion from the materials characterization studies on metal gates. The presented result includes composition and thickness analysis by RBS, agglomeration studies by SEM, grain growth and crystal structure by TEM and XRD.

4.1.1 Composition and thickness

RBS was used to determine the deposition rate, film thickness and composition. RUMP simulations software was used for curve fitting the experimental data. Fig. 4.1 shows a typical RBS spectrum illustrating the clear separation between Ru, Pt and substrate portions of the spectrum. Eleven samples with different alloy compositions were similarly analysed using RBS to determine the exact alloy composition and are plotted in Fig. 4.2 for the case of Pt-Ru. As shown in Fig. 4.2, a composition of 50% Ru and 50% Pt is obtained near about 30% Pt power, indicating that Pt has a higher sputtering rate as compared to Ru. The deposition rate at a typical deposition pressure of 4 mTorr, 20 sccm of Ar flow and 200 W of DC power, for Pt was ~ 2.5 Å/sec and that for Ru was ~ 1.2 Å/sec. Based on the equilibrium phase diagram (Fig. 3.4), when the composition of the alloy has a Ru content less than 62% at.% a FCC Pt phase is expected and
when greater than 80% at.% a HCP Ru phase is expected. Both FCC and HCP phases of Pt-Ru are expected to be present when the Ru content is between 62% and 80% at.%. A detailed study of the resulting phases is important, as they can be expected to impact the important properties of the gate electrode material, such as its work function and thermal stability.

Figure 4.1: RBS spectrum for Pt-Ru film indicating a film thickness of 270 Å with at.% of 50-50 for Pt and Ru.

Figure 4.2: Composition (by RBS) of Ru-Pt alloys as a function of Pt percent power.
A similar, linear correlation of composition to the sputtering power was observed for Ru-Ta, Ru-Nb and Ru-V alloy thin films as shown in Figs. 4.3, 4.4 and 4.5 respectively.

Figure 4.3: Composition (by RBS) of Ru-Ta alloys as a function of Ta percent power.

Figure 4.4: Composition (by RBS) of Ru-Nb alloys as a function of Nb percent power.
By using RBS analysis in similar way the film thickness for all alloy compositions were determined. The deposition rates for a typical 200 W of power are as summarized below.

- Pt ~2.5Å/sec
- Nb ~1.9Å/sec
- Ta ~1.7Å/sec
- V ~1.6Å/sec
- Ru ~1.2Å/sec

4.1.2 Thermal Treatment: Agglomeration

Thin metallic films are routinely produced and employed in various industries as interconnects, gate electrodes etc, a number of questions arise on their long term stability. One interesting general problem relates to stability against shape changes in the thin films, particularly when shape changes tend to break up the continuous film into group of islands. This
phenomenon is termed as agglomeration. Agglomeration is a thermally activated, two step processes of void nucleation and void growth which eventually leads to film breakup forming islands on the film.

Ru and Pt films with thickness of 50 nm was deposited on a thermally grown 100 nm thick SiO$_2$ layer on a Si substrate using ultra high vacuum dc magnetron sputtering. The thermal stability of the films was studied by heating the films and characterizing the resultant film microstructure by scanning and transmission electron microscopies. The annealing was performed in an ambient of one atmosphere of an Ar + 3%H$_2$ reducing gas mixture in a quartz tube furnace. Temperatures in the range of 400 °C to 900 °C were explored and a common annealing time of 30 min was used.

The basic film morphology of the annealed and as-deposited samples was examined in a JOEL 6400F scanning electron microscope (SEM). Additionally, samples of the as-deposited and 300° C annealed films with a thickness of 50 nm were examined by transmission electron microscopy (TEM). The normal, capillary agglomeration process was observed for Ru thin film sample and an example of this is illustrated by the SEM images for the case of a 20 nm Ru thin film. Fig. 4.6 (a) shows the featureless as-deposited thin film. Fig. 4.6 (b) shows the image of a film annealed at 600 °C, where some rough texture can be observed along with initial voids, the latter presumably nucleated at grain boundaries. Further annealing to 700 °C results in the initial voids growing to form obvious holes in the film that expose parts of the substrate, as shown in Fig. 4.6 (c). At 900 °C this process is further developed, and these holes have grown to impinge upon each other forming group of islands connected together by bridges, as shown in Fig. 4.6 (d). Further annealing leads to film breakup.
Voids in the Pt thin films were typically observed to grow in branched manner, unlike Ru, described as “fractals”. The main feature of which is that the voided regions appear to be initially elongated, linear, structures and not the regions of minimum perimeter associated with capillary forces. An illustrative example is shown in Fig. 4.7, which consists of SEM images of a 20 nm thick Pt film. The as-deposited Pt film is shown in Fig. 4.7 (a). After annealing at 450 °C, as shown in Fig. 4.7 (b), voids were found to nucleate in Pt thin films creating a hole in the film and exposing the substrate. Fig. 4.7 (b) also shows the distinctive fractal appearance of the voids. Further annealing would result in growth and impingement of these voids, Fig. 4.7 (c), and eventually island formation at 900 °C.
Figure 4.7: SEM images of 20 nm of Pt thin film annealed in Ar+3%H₂ environment: (a) as-deposited before annealing; (b) 450 °C; (c) 750 °C; (d) 900 °C.

4.1.3 Microstructure and phase identification

The basic film morphology of the annealed and as-deposited samples was examined in a JOEL 6400F scanning electron microscope (SEM) as shown above in Figs. 4.7 and 4.8. Additionally, samples of the as-deposited and 400 °C annealed films with a thickness of 50 nm were examined by transmission electron microscopy (TEM). One of two techniques was used to prepare plan view TEM specimens: focused ion beam in situ lift out and chemical etching. For the chemical etching, hydrofluoric acid was used to remove the SiO₂ under layer and float off the
metal film onto a TEM grid. Analysis was carried out in an FEI Tecnai F30 TEM operating at 300 keV and the images were recorded on a CCD camera.

Figure 4.8: Figs. (a) and (c) show bright field TEM images of the 50 nm thick Pt and Ru film samples before 400 °C annealing and figs. (b) and (d) shows TEM results after 400 °C annealing.

Fig. 4.8 show bright field TEM images of the 50 nm thick Pt and Ru film samples before and after annealing at 400 °C. Only modest grain growth was observed for the Ru film, while a significantly larger grain growth was observed for the Pt. Grains in Pt films were observed to
grow by about 100%, from a 17 nm average grain size in as-deposited film to a 34 nm average in annealed film. In the Ru thin film, the grain growth was a modest 18%, from 12.1 nm to 14.6 nm. This reduced extent of grain growth in Ru is consistent with Ru having the highest melting temperature of the two metals.

The crystal structure of the films was examined by XRD using Cu Kα radiation and a 2θ step size of 0.05°. A fiber texture was present in the films, with a strong Pt (111) FCC peak for the Pt film that gradually shifts towards the position of the Ru (0002) HCP peak with the increase in Ru content in the film as shown in Fig. 4.9. After annealing in Ar+3% H₂ at 400 °C the films showed an increase in intensity for both peaks, indicating an increase in the fiber texture due to grain growth as shown in Fig. 4.10.

![Figure 4.9: X-ray diffraction of the as deposited Pt-Ru thin films.](image-url)
Due to the strong texture present in our as-deposited and the annealed films, the X-ray diffraction data does not conclusively prove that the pure Pt films are FCC phase or that the pure Ru films have the HCP structure. Our film texture is such that the close-packed planes, for both crystal structures, are parallel to the substrate plane. The commonly used Bragg-Brentano X-ray diffraction geometry that we used in our experiments can identify the spacing between these planes but can not distinguish their stacking order. The stacking order would evidence itself in the presence or absence of X-ray diffraction peaks that could be examined with non-specular X-ray diffraction geometry, for example using parallel beam optics, but not with the Bragg-Brentano (powder diffraction) geometry used here. However, it is well known that pure Ru thin films always adopt the HCP phase, and that pure Pt thin films always adopt the FCC phase. From
the plot of the spacing of the close packed planes as a function of composition (Fig. 4.11) we can see evidence for the solubility of Ru in FCC Pt, and vice versa, and then a sudden change in the lattice parameter. The two regions of slowly changing spacing with composition we can reasonably presume to be single phase FCC and HCP regions. The region of slowly changing spacing with composition is more pronounced in the annealed films, as can be seen in Fig. 4.11. The compositional range of rapidly changing spacing may be a two-phase region, as would be suggested from the equilibrium phase diagram, or it may be a material that is highly defective, containing a high density of stacking faults, such that it has a character intermediate to FCC and HCP crystal structures. More extensive structural characterization of films within this intermediate composition range would be required to resolve this question.

![Figure 4.11: Peak position (2θ) for the as deposited and annealed Pt-Ru films are plotted as a function of Pt composition.](image)

Figure 4.11: Peak position (2θ) for the as deposited and annealed Pt-Ru films are plotted as a function of Pt composition.
4.2 Electrical Characterization

This section focuses on the electrical characterization for the Pt-Ru alloy films. The blanket films were characterized for the electrical resistivity measurements, for as deposited and then after thermal treatment. This section also presents a systematic comparison study of work function for the Pt-Ru binary alloy metals, extracted from capacitance–voltage characteristics of metal oxide semiconductor capacitors and the current–voltage and capacitance-voltage characteristics of Schottky barrier diodes. This sub-section also presents results from the temperature dependent I-V measurements made on the Schottky diode structures.

4.2.1 Resistivity

The resistivity of the as-deposited Pt-Ru alloys with various compositions and of the films annealed at 400 °C for four hours in Ar + 3% H₂ was measured by a four-point resistivity probe and is plotted in Fig. 4.12. The resistivity of the as-deposited films initially increased with increasing Pt content, and then remained stable at mid compositions and finally started decreasing as the Pt content of the films increased. For the annealed films a similar trend is observed, but with lower values of resistivity. The decrease in resistivity is attributed to the grain growth observed by TEM. Generally, the resistivity of bulk metal is significantly lower, than its thin film counterpart. In this work, the pure Ru and Pt thin films showed higher resistivities than bulk metal due to their randomly oriented polycrystalline microstructure. With the addition of Pt into Ru thin film, the resistivity increased in correlation to the amount of Pt contained. In the Ru-Pt alloy system, the Pt atoms can be regarded as impurities which diminish the degree of crystalline perfection of pure metal, and hinder the movement of the electrons thereby increasing
the scattering effect. There is a larger decrease in the Ru resistivity values and this could be due to the fact that the Ru films had higher resistivity to start with. This is also evident from the initial grain sizes as found by TEM. The un-annealed Pt films had an average grain size of about 17 nm as compared to the average Ru grain size of about 12.1 nm.

![Graph showing resistivity as a function of at. % Pt concentration for as-deposited and annealed Ru-Pt alloys.](image)

Figure 4.12: Resistivity as a function of at. % Pt concentration for as-deposited and annealed Ru-Pt alloys.

### 4.2.2 Work function: MOS C-V

To extract the work function of the binary alloys of Pt-Ru, MOS capacitor structures were fabricated. The n-type silicon wafers were cleaned and etched by a buffered oxide etch diluted with de-ionized water for 3 minutes to remove any native oxide. The gate oxide was thermally grown in dry oxygen ambient to a thickness in the range of 20 Å to 150 Å, followed by a passivation annealing in hydrogen ambient. Aluminum was thermally evaporated, in a cryo filament evaporator, on the back side of the silicon wafer and annealed in Ar ambient at 500°C.
for 30 minutes to establish an ohmic contact. The metal alloys were deposited after back side Al ohmic contacts were established. HP 4192 impedance analyzer was used to measure high frequency $C-V$ data at 1 MHz. The flat band voltage, $V_{FB}$, was extracted from the $C-V$ characteristics based on the North Carolina State University (NCSU) $C-V$ program and plotted against the effective oxide thickness ($EOT$) to determine the work function for the Pt-Ru binary alloy system.

Fig. 4.13 shows the high frequency $C-V$ data for Ru-Pt binary alloy system as metal gate electrode with a 10.5 nm of SiO$_2$ as gate dielectric. The inset in Fig. 3.13 shows the systematic shift of the $V_{FB}$ as the alloy composition is changed from 100 % Pt to 100 % Ru. The actual data is indicated by the data point; where as the abnormal shape in the transition region is due to a linear curve fit.

![Figure 4.13: The C-V data for Ru-Pt alloy films as gate electrode with a 10.5 nm SiO$_2$ as gate dielectric.](image-url)
The flat band voltage, $V_{FB}$, was extracted from the $C-V$ characteristics based on the NCSU $C-V$ program and plotted against the $EOT$ to determine the work function for the Pt-Ru binary alloy system. Fig. 4.14 shows the $V_{FB}$ vs. EOT for different Ru-Pt alloy compositions. Data in Fig. 4.14 is extrapolated to zero oxide thickness to extract the workfunction and is plotted as a function of at. % Pt in Fig. 4.15. From the extracted values of the workfunction it is evident that the workfunction values for Ru-Pt binary alloy can be varied over a range of 4.8 eV to 5.2 eV for the desired tunability of the workfunction. Also from Fig. 4.15 it is clear that the workfunction is not a linear function of the alloy composition but instead follows the equilibrium phase diagram, having regions of weak compositional dependence in the single phase region of the phase diagram and a steep compositional dependence across the intermediate compositional range.

Figure 4.14: Flatband voltage vs. oxide thickness for different Ru-Pt alloy compositions.
Figure 4.15: Ru-Pt alloy work function plotted as a function of at. % Pt.

4.2.3 Work function: Schottky I-V

The simple Schottky-Mott model for Schottky barrier formation is given by:

$$\Phi_b = \Phi_m - \chi_s$$

(4.1)

where $\chi_s$ is the electron affinity of the semiconductor, defined as the potential difference between the bottom of the conduction band and the vacuum level at the semiconductor surface. If this model is accepted, then a linear dependence should exist between the Schottky barrier height and metal work function [111]. The work function of a metal is defined as the energy required to remove an electron from the Fermi level to a state of rest outside the surface. The metal work function consists of two components, a volume term and a surface dipole term. Similarly, the electron affinity of a semiconductor also contains a surface dipole term. These surface dipole
terms are governed by the way the electronic charge is distributed at the surface of the solids and are therefore clearly influenced by the status of the surface. It is, however, impossible to measure the volume and the surface contributions of the work function separately. When a metal and semiconductor are brought onto intimate contact, the atomic positions and the charge distribution of the surface in contact will change in an unknown way and affect the above mentioned surface contributions. The Schottky-Mott theory, therefore, does not account well for barrier heights at such interfaces, as it assumes $\Phi_m$ and $\chi_s$ or the difference between them, remains unchanged when contact is made [112]. In an attempt to understand the limits of the models, we compare the Pt-Ru alloy metal work function values obtained from Schottky theory to those extracted from a MOS capacitor theory.

It is well known that the Schottky barrier height depends upon the quality of the semiconductor material, the method of surface preparation prior to metallization and the metallization itself. The n-type silicon wafers were cleaned and the back side aluminum ohmic contact was first established as described above. The silicon surface was hydrogen terminated by HF just before introducing the wafer in the vacuum chamber for sputter deposition of Pt-Ru alloy metals to form the Schottky diode. $I-V$ characteristics of the Pt-Ru Schottky contacts were obtained under dark conditions, at room temperature (295 K), using Keithley Instruments picoammeter/voltage source model 6487.

Under the Schottky-Mott theory it is assumed that ideal conditions are obtained when forming the intimate contact with the metal and semiconductor and both dipole surface values remain unchanged. It is also assumed that no chemical reactions take place, and no physical strains are created during the contact procedure of the metal and semiconductor. Knowing the electron affinity of our semiconductor material and solving for the Schottky barrier
height will lead us to determine the work function for Pt-Ru alloy compositions, deposited on silicon substrate.

In a Schottky barrier diode, the ideal barrier height of $\Phi_{BO}$ is approached only when the diode is strongly forward biased. The actual barrier height $\Phi_B$ is less than $\Phi_{BO}$ due to image force barrier lowering, voltage drop across any interfacial layers between the metal and semiconductor, and other possible effects. The $I-V$ data is analyzed under the assumption that the dominating current transport mechanism is thermionic emission. The thermionic current-voltage relationship of a Schottky barrier diode, neglecting series and shunt resistance, is given by:

$$I = AA^*T^2 e^{-q\Phi_B/kT} (e^{qV/nkT} - 1) = I_s e^{-q\Phi_B/kT} (e^{qV/nkT} - 1) = I_s (e^{qV/nkT} - 1)$$

where $I_s$ is the saturation current, $A$ the diode area, $A^*$ the Richardson’s constant, $\Phi_B$ the effective barrier height, and $n$ the ideality factor. The barrier height is most commonly calculated from the current $I_s$, determined by extrapolating the semilog $I$ versus $V$ to $V = 0$. The barrier height $\Phi_B$ can be calculated from $I_s$, using Eq. (4.2) as:

$$\Phi_B = \frac{kT}{q} \ln \left( \frac{AA^*T^2}{I_s} \right)$$

Semilog $I$ versus $V$ plots for Pt-Ru alloy Schottky diodes with an area of 650 micron square are shown in Fig. 3.16. The barrier height determined by the above procedure is $\Phi_B$ for zero bias, and is plotted as a function of at. % Pt in Fig. 4.17. At least sixteen contacts, obtained from two runs of the same alloy composition were measured in each case to obtain the mean Schottky barrier height values plotted in Fig. 4.17. The largest statistical variation applicable to the various barrier height value was less than 2%. Using Eq. (4.1) the Pt-Ru alloy work function values were calculated and are also plotted in Fig. 4.18 as a function of at. % Pt.
Figure 4.16: Semilog plot of I-V for Pt-Ru alloy Schottky diodes

Figure 4.17: Barrier height extracted from the I-V data of the Pt-Ru Schottky diode, plotted as a function of at. % Pt.
4.2.4 Work function: Schottky C-V

The workfunction is also determined from the Schottky barrier height, extracted from the $C-V$ characteristics of the Schottky barrier diode. A Keithley Instruments $C-V$ analyzer model 590 was used to measure the capacitance–voltage characteristics of the Pt-Ru Schottky contacts, and the barrier height was extracted to calculate the metal work function.

The flatband barrier height was obtained from $C-V$ measurements on the Pt-Ru Schottky diode structures. A representative $C-V$ data for pure Pt Schottky diodes with four different areas (600µm X 600µm, 500µm X 500µm, 400µm X 400µm and 300µm X 300µm) is as plotted in Fig. 4.19. The barrier height is given by:

$$\Phi_b = V_{bi} + \frac{kT}{q} + \xi$$  \hspace{1cm} (4.4)
with,

\[ \xi = \frac{kT}{q} \ln \left( \frac{N_c}{N_d} \right) \]  

(4.5)

The built-in potential, \( V_{bi} \), is obtained from the voltage intercept of the curve, \( 1/C^2 \) versus bias voltage, \( V \). \( N_c \) and \( N_d \) represents the effective density of states in conduction band and the free-carrier density of semiconductor material, respectively. \( \Phi_B \) is approximately the flat-band barrier height because it is determined from the \( 1/C^2-V \) curve for \( 1/C^2 \to 0 \) or \( C \to \infty \), indicating sufficient forward bias to cause flatband conditions in the semiconductor.

A \( (C/A)^2 \) versus \( V \) plot for Pt-Ru Schottky diodes are shown in Fig. 4.20, and are linearly extrapolated to determine the voltage intercept. The barrier height and the work function values are obtained using Eq. (4.4) and Eq. (4.1) respectively. The obtained work function values are also plotted as a function of at. % Pt in Fig. 4.21.

Figure 4.19: C-V data for Pt Schottky diodes with four different areas.
Figure 4.20: $1/C^2$ vs. V plot for Pt-Ru alloy Schottky diode measured at room temperature.

Figure 4.21: Work function values obtained by Schottky C-V plotted as a function of at. % Pt.
The work function values obtained by the three techniques of MOS C-V, Schottky I-V and Schottky C-V are plotted as a function of at. % Pt in Fig. 4.22. The barrier height values obtained by Schottky $I-V$ measurements are lower compared to those obtained from the $C-V$ measurements. This observation is consistent with the published literature, where number of studies have been undertaken to compare barrier heights determined by various different techniques. Any damage at the interface affects the $I-V$ behavior as defects may act as recombination centers or as intermediate states for trap-assisted tunnel currents. $C-V$ measurements are less prone to such defects, however defects can alter the space-charge region width and hence the intercept voltage. Oxide charges, such as fixed oxide charge, mobile oxide charge, oxide trapped charge and interface trapped charge, all of these are known to have a significant impact on the EOT and work function extraction for a MOS capacitor structures. It is also essential the $V_{FB}$ vs. EOT plots be linear and independent of the frequency.

![Graph showing work function values obtained by MOS C-V, Schottky I-V and Schottky C-V plotted as a function of at. % Pt.](image)

Figure 4.22: Work function values obtained by MOS C-V, Schottky I-V and Schottky C-V plotted as a function of at. % Pt.
It is clear from the Fig. 4.22, that there are obvious discrepancies amongst the three techniques for extracting work function values, suggesting that a comprehensive investigation of these defect states is needed. However, it is clear that the work function values, obtained from all three techniques, i.e., MOS C-V and Schottky I-V and C-V, show a similar trend in following the equilibrium phase diagram, showing regions of weak compositional dependence in the single phase regions of the phase diagram and a steep compositional dependence across the intermediate compositional range.

4.2.5 Barrier Height: Schottky temperature – I-V

Compositions varying from 100% Pt to 100% Ru were considered, and specific alloy compositions were selected to span the Pt-Ru equilibrium phase diagram. A single phase Pt is expected if the Ru content in the film is less than 62 at. % Ru and a single phase Ru is expected for films with Ru content more than 80 at. % Ru. A two phase region of Pt-Ru is expected between 62 and 80 at. % Ru based on the equilibrium phase diagram. Fig. 4.23 shows typical I-V curves for 30 %Pt electrode on a 300 µm² device area. A general trend of increasing forward bias current with temperature can be seen as expected. All five different samples each with a different ratio of Pt to Ru showed a similar increasing trend for the forward bias current with an increase in temperature. Based on theoretical considerations of the Mott-Schottky theory it is expected that a linear dependence should exist between the Schottky barrier height and metal work function.
Figure 4.23: I-V characteristics for 30%Pt and 70%Ru alloy in a 300 µm² Schottky device at different temperatures.

Fig. 4.24 shows the extracted barrier height values for 100% Pt and 100% Ru devices at different temperatures. It can be clearly observed that the barrier height for Pt increases considerably from 0.95eV to 1.2 eV over the measured temperature range. It is interesting to note that for temperatures greater than 75°C the extracted barrier height for 100% Pt is larger than the silicon band gap of 1.12 eV. Further, in general semiconductor theory it is expected that the silicon band gap will reduce as a function of increasing temperature. The fact that, the barrier height is larger than the band gap indicates the formation of an inversion layer at the interface. The presence of a thin inversion layer with high charge density in this type of diodes causes the maximum electric field in the space-charge region to become very high, even with moderate doping densities. Detailed physical analysis and several reasoning for this phenomenon is discussed elsewhere [113, 114]. However for the Ru devices the increase is comparatively much smaller (≈ 6%). This behavior of the barrier height can be explained based on Eq. (4.3), where it can be noted that $\phi_B$ is directly proportional to the change in temperature and also proportional to
$\ln(AA^*T^2/I_s)$. The effect due to increasing temperature clearly overshadows the change due to increase in saturation current.

Figure 4.24: Extracted barrier heights for (a) 100% Pt and (b) 100% Ru at increasing temperatures for 300 µm$^2$, 400 µm$^2$, 500 µm$^2$ and 600 µm$^2$ area Schottky devices.
The extracted barrier height values for different Pt-Ru alloy compositions at different temperatures are shown in Fig. 4.25 (a) – (c). It can be observed that there is still an increase in the barrier height with temperature for the alloy phases but this increase is not as prominent as in the case of 100 %Pt based devices. The calculated increase in barrier height is less than 10 % and lowers with increasing Ru content in the alloy. This change in barrier height is considerably lower and nearly close to the observed increase in 100 % Ru based devices as the amount of Ru increases in the alloy composition. This improved barrier height stability with temperature for the Ru-Pt alloy compositions can be attributed to the presence of increasing Ru in the material. This indicates that there is obvious difference between interface states for pure Pt and Ru, with Ru having better interface that is stable with increasing temperature. The issue of interface states needs more detailed investigation for better understanding of the physical phenomenon.
Figure 4.25: Extracted barrier heights for Pt-Ru alloy compositions with temperature for different area (300 µm², 400 µm², 500 µm² and 600 µm²) devices, (a) 50%Ru, (b) 70%Ru and (c) 80%Ru.
Another significant observation from the extracted values of barrier height in Figs. 4.24 and 4.25 is that different area devices produced almost identical values at different temperatures. Hence it can be concluded that device area has almost negligible affect on the barrier height measurement for these devices. Fig. 4.26 shows the extracted barrier heights for different temperatures (25°C, 50°C, 75°C, 100°C and 125°C) and areas (300 µm², 400 µm², 500 µm² and 600 µm²) plotted as a function of at. % Ru. It can be clearly seen that the barrier height does not follow a linear compositional dependence as the Pt-Ru alloy composition is changed from 100 % Pt to 100 % Ru. As expected from the equilibrium phase diagram of Pt-Ru, the barrier height is a weaker function of alloy composition in the single phase Pt and Ru regions and a stronger compositional dependence can be observed for the intermediate compositional range.

Figure 4.26: Extracted barrier heights for different temperatures (25°C, 50°C, 75°C, 100°C and 125°C) and areas (300 µm² to 600 µm²) plotted as a function of at. % Ru concentration.
CHAPTER FIVE: RESULTS AND DISCUSSIONS: Ge CHANNEL DEVICES

This chapter presents results from the Ge channel device studies. This chapter is mainly divided into two sub-sections, one on mobility extraction and comparison with the state of the art Si devices and low frequency noise assessments.

5.1 Ge PMOSFET Mobility Data

10 µm X 10 µm Ge devices were fabricated using the Si like process flow as described in section 3.2.1. Sample C-V data is shown in Fig. 5.1, yielding CETs and EOTs of 1.6 nm and 1.2 nm respectively. Figure 5.2 shows the corresponding $I_{ds}$-$V_g$ and $G_m$ data.

![Figure 5.1: C-V data for a 10 µm X 10 µm PMOS obtained at 100 kHz, showing excellent characteristics and an EOT of 1.2 nm.](image)

Figure 5.1: C-V data for a 10 µm X 10 µm PMOS obtained at 100 kHz, showing excellent characteristics and an EOT of 1.2 nm.
Figure 5.2: Linear $I_{ds}$-$V_g$ and $G_m$-$V_g$ obtained at $V_d = -50$ mV, for the 10 µm X 10 µm p-MOS of Fig. 5.1.

Figure 5.3: Hole mobility vs. effective field for a 10 µm X 10 µm p-MOS as in Fig. 5.1 and 5.2, with and without a PMA.
Fig. 5.3 shows ~2.7 X mobility enhancement over Si at both low fields and high fields, with peak mobility ~ 300 cm$^2$/Vs for Ge devices exposed to a 450°C post metallization anneal (PMA). It is evident that the mobility enhancement is maintained even for large fields.

The PMA is critical to obtaining homogeneous PMOS performance. Without the PMA, large differences in the C-V shape are present from device to device, Fig. 5.4. These differences correspond to a more than 10-fold spread (from 0.5 to 8 x 10$^{12}$ cm$^{-2}$) in N$_{it}$. The PMOS performance (mobility) in turn directly correlates to the N$_{it}$, Fig. 5.5. With PMA, the difference in C-V shape and device performance are reduced by an order of magnitude, as shown in Fig. 5.6 and Fig. 5.7 respectively.

Figure 5.4: C-V characteristics without PMA, showing large difference in C-V shape from device to device.
Figure 5.5: Mobility characteristics for un-annealed devices showing large variation, indicating strong influence of the $N_{it}$.

Figure 5.6: C-V characteristics after 450°C, 5 min PMA in N$_2$ ambient, showing an order of magnitude improvement in terms of variations in C-V shape from device to device.
Figure 5.7: Mobility characteristics for post PMA devices showing reduced variation in device to device mobility, indicating large decrease in $N_{it}$.

Ge PMOS devices with gate length ($L_g$) down to 0.25 $\mu$m show well-behaved $I_{ds}$-$V_g$ characteristics with more than 4 decades between $I_{on}$/$I_{off}$ and subthreshold slope of $\sim$ 100 mV/dec at $V_{ds} = -50$ mV, Fig. 5.8. Gate induced drain leakage (GIDL currents arises in the high electric field under the gate/drain overlap region. GIDL occurs at a low $V_g$ and high $V_d$ bias and generates carriers into the substrate and drain from surface traps or badn-to-band tunneling) limits $I_{off}$ due to un-optimized $V_t$, which can be adjusted metal gate work function engineering. Device with $L_g < 0.25$ $\mu$m suffers from gradually increasing short channel effect since no halo implants are used, as illustrated for subthreshold slope and $I_{off}$ in Fig. 5.8.
Figure 5.8: $I_d - V_g$ for Ge PMOS devices with $L_g$ from 0.25 µm to 0.125 µm. Subthreshold slope and GIDL are highlighted for the 0.25 µm device.

5.2 LF Noise Assessments

The following two sub-sections report results from the LF noise assessments on the silicon passivated Ge p-MOSFETS and on the geometry dependence of 1/f noise in the shallow $p^+ - n$ junctions with and without Ni germanidation. A correlation between the low-field mobility and the oxide trap density derived from the 1/f noise magnitude and the interface trap density obtained from charge pumping is reported and explained by considering remote Coulomb scattering.
5.2.1 LF Noise results for Ge p-MOSFETs

As can be derived from Fig. 5.9, the noise spectra are typically of the flicker noise type, with some traces of a Generation-Recombination (GR) component at higher frequencies. It is interesting to note that the current noise spectral density $S_{id}$ is higher for the lower-mobility 1 $\mu$m PMOSFET. Before exploring this trend in more detail, let us first find out the origin of the 1/f-like noise. According to the behavior of the normalized noise at $f=25$ Hz, i.e., $S_{id}/I_d^2$ versus the gate voltage overdrive $V_{gt}=V_{gs}-V_{th}$, represented in Fig. 5.10, it is concluded that up to $V_{gt}=-0.5$ V the noise behaves according to the number fluctuations or $\Delta n$ model [115]. This follows from the proportionality with $V_{gt}^{-2}$. The increase of the normalized noise at higher $|V_{gt}|$ may be related to the parasitic series resistance of the p-channel transistors [116].

Figure 5.9: Low-frequency noise spectra for a low- ($155 \text{ cm}^2/\text{Vs}$) and a high-mobility ($204 \text{ cm}^2/\text{Vs}$) Ge PMOSFET with $L=1$ $\mu$m and corresponding with $V_{ds}=-50$ mV and a $V_{gs}=-0.57$ V.
Figure 5.10: Normalized noise spectral density $S_{id}/I_d^2$ versus gate voltage overdrive $V_{gs-Vth}$ at $f=25$ Hz for an L= 1µm Ge PMOSFET.

It should be noted that for the 1/f noise of metal gate/high-κ p-channel transistors, one usually finds a behavior which is more in line with the mobility fluctuations ($\Delta \mu$) theory [117, 118], although it has been shown that on Si$_{0.8}$Ge$_{0.2}$ surface channel p-MOSFETs with ALD Al$_2$O$_3$/TiN stack correlated number fluctuations appear to be dominant [119].

As the 1/f noise is due to trapping-detrapping events, one can derive a density of oxide traps from the input-referred noise spectral density $S_{vg}=S_{id}/g_m^2$ in flat-band operation ($g_m$ being the device transconductance), as indicated in Fig. 5.11 and according to:

$$S_{VFB} = \frac{q^2 kT N_{ot}}{WLC_{ox}^2 \alpha_f f}$$  \hspace{2cm} (5.1)
with q the elementary charge, k Boltzmann’s constant, T the absolute temperature, WL the device area, \( C_{ox} \) the effective oxide capacitance density, derived from a C-V plot in inversion and \( \alpha_t \) the tunnelling parameter. The oxide trap density \( N_{ot} \) (in \( \text{cm}^{-3} \text{eV}^{-1} \)) can be converted in a surface density (\( D_{ot} \)) by assuming an energy interval of \( 4kT \) [120], while for \( \alpha_t \) a value of \( 0.72 \cdot 10^8 \text{cm}^{-1} \) has been assumed [121], yielding a tunnelling depth at 25 Hz of ~2.5 nm.

It has been noted in Fig. 5.9 that the noise magnitude is higher for the low-mobility transistors. This trend has been investigated in more detail on a larger number of devices across the wafer. As can be seen for the \( L=1 \mu m \) PMOSFETs in Fig. 5.12, a higher \( S_{Vg} \) is indeed found for a lower mobility. A similar trend is obtained for the 0.25 \( \mu m \) transistors. Such behaviour has been reported before for different types of n-channel transistors [122-124] and points to the impact of oxide traps on both the 1/f noise and the mobility through remote Coulomb scattering.

![Graph](image)

Figure 5.11: Input-referred noise spectral density \( S_{Vg} \) versus gate voltage overdrive for a \( L=0.25 \mu m \) PMOSFET at \( f=25 \text{Hz} \).

106
Figure 5.12: Correlation between $S_{V_g}$ at 25 Hz and the mobility $\mu$ for the L=1 $\mu$m PMOSFETs studied.

This idea is further supported by the results of Fig. 5.13, representing the inverse mobility $(1/\mu)$ versus the oxide trap charge $q_{D_{ot}}$ (C/cm$^2$) derived from 1/f noise and the surface-state charge density $q_{D_{it}}$ following from CP. It is evident from Fig. 5.13 that while the trap densities are of the same order of magnitude, they are typically a factor two higher for the oxide traps. Also the spread is significantly higher compared with the interface traps. The scattering coefficient derived from the initial slope of the $1/\mu$ versus $q_{D_{ot}}$ trend yields an acceptable value in the range of $10^4$ Vs/C [125-127].
Figure 5.13: Inverse mobility versus qD_{it} derived from CP and qD_{ot} extracted from 1/f noise for the L=1 µm PMOSFETs studied.

5.2.2: LF Noise Results for Ge p⁺-n Junctions

As shown in Fig. 5.14, the noise spectra are of the 1/f^γ type, with γ close to 1, while the frequency-independent shot noise occurs outside the accessible frequency range (>10^5 Hz). Occasionally, Lorentzian humps due to generation-recombination noise have been found. When comparing the germanided with the non-germanided junctions in Fig. 5.15 or 5.16, it is first of all clear that the NiGe helps to reduce the series resistance. This follows from the higher forward current observed in Fig. 5.15a or 15.16a at a V_F in the range 0.1 to 0.4 V. The improvement appears to be more significant for the square large-area diode of Fig. 5.15 compared with the large-perimeter diode of Fig. 5.16.
Figure 5.14: Low-frequency noise spectra for a germanided and a non-germanided p⁺-n junction at a forward bias of 0.5 V. The area is 15000 µm² and the perimeter P=10300 µm.

The current noise spectral density $S_I$ at 1 Hz versus $I_F$ follows approximately an $I_F^2$ dependence over most of the forward current range studied. It has been demonstrated before that for silicon junctions, $S_I$ is proportional to $I_F^\beta$, [128-130] with $\beta$ in the range of 1.2 to 2. Moreover, the noise spectral density for the germanided samples is smaller than for the non-germanided counterparts, whereby the difference increases for increasing $I_F$. This is in contrast to the case of CoSi$_2$ n⁺-p junctions, showing an increase in the 1/f noise due to GR fluctuations of holes in the shallow n⁺-junction region [131]. At the same time, the 1/f noise improvement is more pronounced for the large-area junctions compared with the large-perimeter ones.
Figure 5.15: a) $I_F$ versus $V_F$ and b) $S_I$ at 1 Hz versus $I_F$ for a large area square diode ($A=30000 \ \mu m^2$) with or without NiGe.
Figure 5.16: a) $I_F$ versus $V_F$ and b) $S_I$ at 1 Hz versus $I_F$ for a large perimeter diode ($P=10300 \, \mu m$) with or without NiGe.
It has been shown in the past for silicon junctions that the origin of the excess 1/f noise is related to generation-recombination (GR) events by defects either in the depletion region [132] or at the surrounding surface regions [133, 134]. For pure GR events in the bulk depletion region, $\beta=2$ is expected. Lower values can be explained by the fact that the forward recombination current is only a fraction of the total forward current, so that we anticipate an $I_F^{2/m}$ dependence with $m$ the diode ideality factor ($\geq 1$). However, as derived from the forward characteristics in Figs 5.15a and 5.16a, the series resistance of the junctions can also play an important role and may generate 1/f noise fluctuations [135].

In order to further identify the origin of the 1/f noise in the Ge $p^+\text{-}n$ junctions, the noise spectral density versus the junction area $A$ at a fixed current $I_F=100 \mu A$ has been studied (Fig. 5.17). The germanided diodes exhibit a clear increase with $A$, which follows a power law, i.e., $I_F$ is proportional with $A^{0.39}$. This further supports a defect-assisted GR origin of the 1/f noise. Candidates for this are the threading dislocations in the relaxed epitaxial germanium layers. We may not rule out, however, a possible contribution from residual, unannealed n-well implantation damage.

On the other hand, there is no straightforward area-dependence for the non-germanided junctions in Fig. 5.17. This strongly suggests that the 1/f noise is dominated by fluctuations in the series resistance. In summary, it can be stated that the application of nickelgermanide ohmic contacts not only lowers the series resistance of a Ge $p^+\text{-}n$ junction but improves the 1/f noise. In germanided samples, the 1/f noise is shown to be governed by defect-assisted GR events in the junction depletion region.
Figure 5.17: $S_I$ at 1 Hz and 100 $\mu$A for the germanided and non-germanided p$^+$-n junctions versus area.
CHAPTER SIX: CONCLUSION

The key findings of this dissertation have been summarized in this chapter under dissertation remarks. Based on the understanding of this work and following the literature, few recommendations have been made for further studies in the outlook section of this chapter.

6.1 Dissertation Remarks

DC magnetron sputtering has been used to deposit binary alloy films on thermally grown SiO$_2$. The composition of the alloy films have been determined by RBS and the identification of phases present were made by x-ray and electron diffraction of samples annealed to different stages of reaction. The microstructure of the phases of interest was examined in the TEM and film texture was characterized by XRD. The electrical characterization includes basic resistivity measurements and work function extraction. The work function has been determined from MOS capacitor and Schottky diodes.

The results indicate that the change of film properties (i.e. resistivity, work function and crystal structure) with composition is consistent with the equilibrium phase diagram and the work function in the FCC and HCP single phase regions is only weakly dependent on alloy composition while a strong dependence is observed in the intermediate compositional range of mixture (62% to 80% Ru). It is also observed that work function values obtained from the I-V analysis are lower than those extracted from the C-V data. Both I-V and C-V methods are extremely sensitive to defects, such as spatial inhomogeneities, insulating layers between metal and semiconductor, oxide charges, doping inhomogeneities, surface damage and tunneling, and must be considered in more details. However, it is clear that the work function values, obtained
from all three techniques show a similar trend in following the equilibrium phase diagram. The alloy work function follows the equilibrium phase diagram by showing regions of weak compositional dependence in the single phase regions and a steep compositional dependence across the intermediate compositional range of the phase diagram.

The need for electron and hole mobility enhancement and the progress in the development of high-κ gate stacks, has lead to renewed interest in Ge MOSFETs. The p-MOS mobility data for Ge channel devices have been reported. The results indicate greater than 2 x improvements in device mobility as compared to standard Si device. Mobilities of 315 cm²/Vs were obtained for Ge devices processed in a conventional Si-compatible flow without enhancement techniques such as strain. Using Si surface passivation and HfO₂ dielectric very promising results have been demonstrated, even for short channel devices. Such performance is obtained without halo or other optimization, suggesting significant remaining headroom for future improvements.

Low frequency noise assessment of silicon passivated Ge PMOSFETs with TiN/TaN/HfO₂ gate stack has been investigated. For the first time we also report results on low frequency noise characterisation for a Ge P⁺-n junctions with and without Ni germanidation. It has been shown that the LF noise of TiN/TaN/HfO₂/SiO₂ gate stack PMOSFETs on epitaxial Ge-on-Si substrates is predominantly 1/fγ noise, behaving according to the number fluctuations theory. It has also been observed that both the interface trap density derived from CP and the oxide trap density derived from 1/f noise are well correlated with the low-field mobility, pointing to the impact of Coulomb scattering on the latter parameter. It also suggests that a better control of the uniformity (quality) of the gate dielectric and the interfacial layer should lead to a better, more uniform mobility.
From the study of p$^+$-n junctions with and without Ni germanide, it can be stated that the application of Ni germanide ohmic contacts not only lowers the series resistance of a Ge p$^+$-n junction but improves the 1/f noise. In germanided samples, the 1/f noise is shown to be governed by defect-assisted GR events in the junction depletion region.

**6.2 Outlook**

To be able to realize high performance metal gates and Ge-III/V technology, it is essential to develop a method to introduce these materials on a Si substrate. Therefore, the technology for the growth of Ge and (In)GaAs active areas on a Si substrate must be developed. Various techniques are under investigation. One study is concentrating on the selective growth of GaAs on Ge, which on itself can be selectively grown on Si. Another possible method is based on the Ge condensation technique whereby a SiGe layer is deposited on top of a SOI wafer and the Ge concentration is increased by specially tuned oxidation steps followed by subsequent etching. Finally, the use of epitaxial oxides to make compliant substrates should also be considered as a possible option. In all cases defect densities in the material are a prime concern. The possible integration of more complex III-V materials or materials with even higher electron mobility on this type of substrate to further increase the NMOS performance would be an interesting area of research and needs to be explored. Also the introduction of strain must be considered since strained Ge shows dramatic effective hole mobility enhancements of the order of ~20X that of Si in PMOS and can be directly implemented on a Si platform. As a summary Fig. 6.1 shows a schematic of what could be the future of CMOS!
GaAs or other III/V nMOS  Strained Ge pMOS  Si CMOS

GaAs or other III/V selectively grown on Ge

Strained Ge selectively grown on Si or GeOI substrate

Si wafer

Figure 6.1: Schematic cross section of a proposed ideal CMOS structure for the maximum enhancement of both electron and hole mobility.
LIST OF REFERENCES


[105] W.P. Bai et al., Ge MOS characteristics with CVD HfO$_2$ gate dielectrics and TaN gate electrode, in: proc. of the symposium on VLSI technology, 2003, pp. 121-122.


