WORK FUNCTION AND PROCESS INTEGRATION ISSUES OF METAL GATE MATERIALS IN CMOS TECHNOLOGY

REN CHI

NATIONAL UNIVERSITY OF SINGAPORE

2006

WORK FUNCTION AND PROCESS INTEGRATION ISSUES OF METAL GATE MATERIALS IN CMOS TECHNOLOGY

REN CHI

B. Sci. (Peking University, P. R. China) 2002

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

OCTOBER 2006

ACKNOWLEGEMENTS

First of all, I would like to express my sincere thanks to my advisors, Prof. Chan Siu Hung and Prof. Kwong Dim-Lee, who provided me with invaluable guidance, encouragement, knowledge, freedom and all kinds of support during my graduate study at NUS. I am extremely grateful to Prof. Chan not only for his patience and painstaking efforts in helping me in my research but also for his kindness and understanding personally, which has accompanied me over the past four years. He is not only an experienced advisor for me but also an elder who makes me feel peaceful and blessed. I also greatly appreciate Prof. Kwong from the bottom of my heart for his knowledge, expertise and foresight in the field of semiconductor technology, which has helped me to avoid many detours in my research work. I do believe that I will be immeasurably benefited from his wisdom and professional advice throughout my career and my life. I would also like to thank Prof. Kwong for all the opportunities provided in developing my potential and personality, especially the opportunity to join the Institute of Microelectronics, Singapore to work with and learn from so many experts in a much wider stage. My best wishes will be with Prof. Chan and Prof. Kwong always.

I would also like to greatly acknowledge Prof. Li Ming-Fu, Dr. Yeo Yee-Chia from NUS and Prof. Kang Jinfeng from Peking University for their valuable suggestions and inspirational discussions which had been indispensable for my research work. My special acknowledgement goes to Dr. Yu Hongyu, who is my senior in the lab previously and currently with IMEC at Belgium, for his self-giving help in most of the technical problems that I had encountered in the first two years of my research.

I also owe the opportunity to collaborate with so many talented graduate students in Silicon Nano Device Lab at NUS. Many thanks to Dr. Hou Yongtian, Dr. Chen Jinghao, Dr. Yu Xiongfei, Mr. Whang Sung Jin, Mr. Wang Xinpeng, Mr. Shen Chen, Mr. Hwang Wan Sik, Mr. Liow Tsung-Yang, Mr. Lim Eu-Jin, Mr. Faizhal Bin Bakar, and Mr. Peng Jianwei for their useful discussions and kind assistances during the course of my research, as well as the friendships that will be cherished always. I would also like to extend my best appreciation to all other SNDL teaching staff, graduate students, and technical staff for the good academic environment created.

A significant part of my research was performed in Institute of Microelectronics (IME), Singapore. Many of my thanks also go to the managers and technical staff in the Semiconductor Process Technologies (SPT) lab of IME. I would like to appreciate Dr. Balasubramanian Narayanan, Dr. Lo Guo-Qiang, Dr. Rakesh Kumar, and Dr. Feng Han-Hua for all the supports during my stay at IME. I also must acknowledge Dr. Alastair David Trigg for the help in AES analysis, Dr. Tung Chih-Hang for the help in TEM characterization, and Dr. Loh Wei-Yip, Dr. Agarwal Ajay, Dr. Lakshimi Kanta Bera, Dr. Yu Ming-Bin, and Dr. Subramaniam Balakumar for their knowledge and experiences which had helped me so much. My gratitude also goes to the excellent team of the technical staff in the IME cleanroom for their skillful and responsible work. Without these, I would not have gained so much during the course of my doctoral research.

I also need to thank Dr. Pan Jishen in Institute of Materials Research and Engineering (IMRE), Singapore, for the help in XPS analysis, and Dr. Thomas Osipowicz in the Department of Physics, NUS, for the help in RBS analysis.

Last but not least, to my family, especially the love of my life, Zhang Li, for their love and enduring supports.

TABLE OF CONTENTS

Acknowledgements i

Table of Contents iii

Summary viii

List of Tables x

List of Figures xi

List of Symbols xviii

List of Abbreviations xx

Chapter 1. Introduction 1

1.1. Overview 1

1.2. MOSFET Scaling: Challenges and Opportunities 2

1.2.1. Leakages in Deep-Submicrometer MOSFET 4

1.2.2. Vertical Scaling of MOSFET Gate Stack 6

1.2.3. Innovations in Device Structures 10

1.2.4. Mobility Enhancement for Performance Gain 11

1.3. Summary 13

References 15

Chapter 2. Developments in Metal Gate Materials for CMOS Technology 22

2.1. Limitations of Poly-Si Electrode 22

2.1.1. Poly-Si Depletion Effect 23

2.1.2. Dopant Penetration Effect...24

2.1.3. Gate Electrode Resistivity...25

2.1.4. Compatibility with High-κ Dielectrics...27

2.2. Post Polysilicon Era: Metal Gate Technology..30

2.2.1. Historical Perspective of Metal Gate Electrodes 30

2.2.2. Considerations for Metal Gate Candidates 31

2.2.2.1. Work Function Requirement..31

2.2.2.2. Thermal Stability Considerations 33

2.2.2.3. Process Integration Issues...34

2.2.2.4. Co-optimization of Metal Gate/High-κ Gate Stack 36

2.2.3. Research Status of Metal Gate Technology..37

2.2.3.1. Direct Metal Gates 38

2.2.3.2. Binary Metal Alloys...40

2.2.3.3. Fully-Silicided (FUSI) Metal Gates...40

2.3. Challenges in Metal Gate Technology..43

2.3.1. Understanding of the Metal-Dielectric Interface 43

2.3.2. Developing Appropriate Metal Gate Materials..44

2.3.3. Dual Metal Gate Integration Issues..44

2.4. Research Scope and Major Achievements in this thesis.................................45

References...48

Chapter 3. The Metal-Dielectric Interface and Its Impact on the Effective Work Function of Metal Gates...56

3.1. Introduction...56

3.2. Theoretical Backgrounds 58

3.2.1. Work Function of Metal Materials 58

3.2.2. Definition of Effective Work Function 58

3.2.3. Factors Affecting the Work Function of Metals 60

3.2.4. Fermi-Level Pinning: Schottky Model and Bardeen Model 61

3.2.5. Metal Induced Gap States (MIGS) Theory and Its Limitations 62

3.2.6. Work Function Measurement Techniques 65

3.3. Experimental 67

3.4. Results and Discussions 68

3.4.1. Work Function Thermal Instability of TaN 68

3.4.2. General Trends in the Process Dependentce of $\Phi_{m,eff}$ on SiO_2 and high-κ Dielectrics 74

3.4.3. Model: Fermi Level Pinning Induced by Extrinsic States 77

3.4.4. Investigation of Hf-Si Bond Induced Extrinsic States 79

3.5. Conclusion 85

References 87

Chapter 4. Lanthanide-Incorporated Metal Nitrides Electrodes for NMOS Applications 91

4.1. Introduction 91

4.2. Experimental 93

4.3. Material Characteristics of Lanthanide-MN_x 96

4.3.1. Composition Analysis 96

4.3.2. Auger Electron Spectroscopy (AES) Study 98

4.3.3. X-ray Photoelectron Spectroscopy (XPS) Study 100

4.3.4. X-ray Diffraction (XRD) Study 103

4.3.5. Resistivity 105

4.4. Work Function Tunability 108

4.5. Thermal Stability Study 113

4.6. MOSFET Characteristics 121

4.7. Conclusion 128

References 129

Chapter 5. Process Integration for Dual Metal Gate Electrodes 132

5.1. Introduction 132

5.2. A Gate-First Dual Metal Gate Integration Scheme by High-Temperature Metal Intermixing Technique 140

5.2.1. Motivation 140

5.2.2. Process Integration Flow and Device Fabrication 141

5.2.3. Feasibility Study of the High-Temperature Intermixing Method 143

5.2.4. Compatibility with High-*κ* Dielectrics 147

5.2.5. Dual Work Function Metal Gate Integration using InM 151

5.2.6. Summary 155

5.3. A Gate-Last Dual Metal Gate Integration Process Employing a Novel HfN Replacement Gate 156

5.3.1. Motivation 156

5.3.2. Proposed Integration Flow and Device Fabrication 157

5.3.3. Results and Discussions 158

5.3.4. Summary 164

5.4. Conclusion 164

References 166

Chapter 6. Conclusion 169

5.1. Summary 169

5.1.1. Understanding the Metal-Dielectric Interface 169

5.1.2. Lanthanide-Incorporated Metal Nitride Gate Electrodes 170

5.1.3. Process Integration of Dual Metal Gates 171

5.2. Suggestions for Future Work 173

References 176

Appendix

List of Publications 177

SUMMARY

Rapid advances in CMOS technology have led to aggressive scaling of the MOSFET gate stack. Conventional poly-Si/SiO_2 gate stack is approaching some practical limits, and novel metal gate materials and high-κ dielectrics may need to be introduced into IC industry as will novel process integration technologies. Immense challenges arise in material engineering and process integration of novel metal gate electrodes. This thesis attempts to address some of these challenges.

The metal-dielectric interface is important since it directly affects the effective work function of metal gates. The influence of the metal-dielectric interface on the effective work function has been investigated systematically in this thesis. It is found that the creation of *extrinsic states* at the metal-dielectric interface, which appears to be thermodynamically driven, could be the major cause for the instability of metal gate effective work function during the high-temperature annealing process. The chemical bond configurations at the metal-dielectric interface could be correlated with the creation of *extrinsic states*. In general, the Hf-Si bond tends to create *extrinsic states* upon annealing while Hf-Hf or Si-Si bonds' effect is less pronounced. A model considering the impact of *extrinsic states* has also been proposed to qualitatively explain the dependence of metal effective work function on the annealing process.

One of the most urgent issues for metal gate technology is to find a way to tune the work function of metal gates for CMOS applications. We demonstrate, for the first time, that lanthanide elements can be very useful in modulating the work function of refractory metal-nitride gate electrodes, which provides a new way for metal gate work function engineering. In this work, lanthanide elements with very low work function are incorporated into metal-nitride materials to get the best trade-off between thermal stability and low work function. By varying the lanthanide

concentration in lanthanide-incorporated metal-nitrides, a work function value of 4.2~4.3 eV can be obtained even after a 1000 ^{o}C RTA treatment. This is promising for NMOS devices using a gate-first bulk-Si CMOS process. The good thermal stability has been attributed to the high nitrogen concentration in these lanthanide-incorporated metal-nitrides, therefore the N concentration needs to be carefully engineered in process. Good transistor characteristics have also been demonstrated using these novel metal gate materials.

Dual metal gate integration issues for advanced CMOS devices are also discussed in this thesis. A novel dual metal gate integration process using a high-temperature metal intermixing technique is first demonstrated for gate-first CMOS flow. In this process, a TaN buffer layer is used to protect the gate dielectric during the selective metal etching process. The work function of the TaN buffer layer can be modulated for CMOS by a subsequent metal intermixing process at high-temperature, which is compatible with the conventional gate-first process flow. By using this integration scheme, dual work function of 4.15 and 4.72 eV has been achieved in TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks, respectively.

Another dual metal gate integration process proposed in this thesis is a gate-last replacement gate process employing HfN as a novel dummy gate electrode. In this process, a high-quality HfN/HfO_2 gate stack with HfO_2 EOT less than 1 nm is first fabricated using a gate-first process. The dummy HfN gate can then be selectively removed from HfO_2 so that other metal gate candidates with suitable work functions for bulk-Si CMOS can be integrated. In a prototype demonstration, large work function difference for about 0.8 eV has been achieved by using Ta and Ni to replace HfN for NMOS and PMOS devices, respectively, with no degradation in the EOT, gate leakage, and TDDB characteristics of the ultra-thin HfO_2 gate dielectric.

LIST OF TABLES

Table 1.1	Technology roadmap for the scaling of dielectrics thickness in next ten years.	7
Table 2.1	Specifications for the scaling of gate electrode, derived from *ITRS*-2005.	26
Table 3.1	Experimental splits of different laminated stacks consisting Hf (N) and Si layers on slanted SiO_2 and HfO_2 dielectrics.	80
Table 3.2	V_{FB} shifts for TaN/HfN/Si and TaN/Si/HfN stacks on SiO_2 and HfO_2 after RTA at 1000°C for 5 sec, observed from the lower parts of the *C-V* curves shown in Fig. 3.10 & Fig. 3.11.	84
Table 4.1	Work functions of some lanthanide metals.	92
Table 4.2	Experiment splits and the compositions for the Lanthanide-MN_x films.	94
Table 4.3	Experiment splits and the compositions for $Ta_{0.9}Tb_{0.1}N_y$ with different N_2 flow rates during reactive sputtering deposition.	94
Table 4.4	Work function and barrier height of lanthanide-incorporated TaN on SiO_2 as a function of rapid thermal anneal (RTA) temperatures.	112

LIST OF FIGURES

Fig. 1.1 CPU transistor counts from 1970s to present, showing the device scaling according to Moore's Law; © Intel corp. 2

Fig. 1.2 Historical scaling trends of supply voltage (V_{DD}), threshold voltage (V_{th}) and gate-oxide thickness (t_{ox}) vs. channel length (L_g) for CMOS logic technologies, showing the different scaling factors for supply voltage and device dimension. 3

Fig. 1.3 Schematic cross section of MOSFET showing the major leakage current paths. I_1 for direct tunneling through gate oxide; I_2 for subthreshold leakage; I_3 for BTBT; I_4 for GIDL; and I_5 for punchthrough. 4

Fig. 1.4 Gate leakage current density of some high-κ dielectrics as a function of EOT, compared with the gate leakage specifications at 100°C for high-performance (HP), low-operating-power (LOP), and low-standby-power (LSTP) applications according to *ITRS* 2004. 9

Fig. 2.1 (a) The energy band diagram of an NMOS device showing the poly-Si gate depletion effect; (b) Equivalent circuit for the gate stack of MOSFET. *C* denotes the total gate capacitance which determines the inversion charge density Q_i in the channel, C_{poly}, C_{ox}, C_{Si} represent the capacitance from the poly depletion, gate oxide, and substrate, respectively. C_{Si} is further broken up into a depletion charge capacitance C_d and inversion-layer capacitance C_i. *CET* represents the capacitance equivalent thickness of the MOS gate stack, and ψ_s is the surface potential. 23

Fig. 2.2 Work function of some metal elements collected from experiments. 32

Fig. 2.3 Work function modulation by various mechanisms in some NiSi-based-silicide metal gates; data from [75], [77], [80]-[84]. 42

Fig. 3.1 Band diagram of a MOS structure in flat-band condition (a) without interface dipoles and (b) with interface dipoles at metal-dielectric interface. 59

Fig. 3.2 Schematic energy band diagram (left) and the characteristics of the gap states (right) for metal gate on dielectrics. The character of MIGS becomes more acceptor- (donor-) like toward the *E*c (*E*v), as indicated by the solid (dashed) line. 63

Fig. 3.3 Plots of V_{FB} versus EOT of (a) TaN/SiO_2 and (b) TaN/HfO_2 69

devices before and after 1000°C RTA treatment, from which the effective work function of TaN can be extracted.

Fig. 3.4 The comparison of (a) EOT (with three different gate-oxide thickness) and (b) effective work function of TaN as a function of RTA temperature with and without HfN capping layer on top of the TaN/SiO_2 stack 71

Fig. 3.5 (a) Gate leakage measurement of HfN/TaN/SiO_2 devices and (b) barrier height extraction by F-N tunneling analysis before and after 1000°C RTA treatment. 72

Fig. 3.6 The variation of metal gate work function Φ_m with the annealing temperature on SiO_2 dielectric. 74

Fig. 3.7 Work function of metal gates on HfO_2 before and after high-temperature annealing. HfN_x-1 and HfN_x-2 denotes HfN_x with different N concentration. 76

Fig. 3.8 Schematic energy band diagram for a metal gate on a dielectric, showing the mechanism of Fermi-level pinning by *extrinsic states*. (a) When $E_{F,m}$ is above the pinning level, (b) When the $E_{F,m}$ is below the extrinsic pinning level. The conduction band edge and the valence band edge of the dielectric are denoted by $E_{c,d}$ and $E_{v,d}$, respectively. 78

Fig. 3.9 Work function of the TaN/Hf/Si laminated stack on SiO_2 and HfO_2 after annealing at different conditions: as-deposited, 420°C FGA and RTA at 1000°C for 5 sec followed by FGA. 81

Fig. 3.10 *C-V* measurements of the TaN/HfN/Si laminated stack on (a) SiO_2 and (b) HfO_2/SiO_2 dielectrics after different annealing. 83

Fig. 3.11 *C-V* measurements of the TaN/Si/HfN laminated stack on (a) SiO_2 and (b) HfO_2/SiO_2 dielectrics after different annealing. 83

Fig. 3.12 XTEM image of a TaN/HfN/Si stack on SiO_2 after 1000°C RTA for 5 sec. 85

Fig. 4.1 Illustration of the idea to modulate the work function of metal nitrides by incorporating lanthanide elements for n-MOSFET applications. 93

Fig. 4.2 Relationship between Tb concentration in $Ta_{1-x}Tb_xN_y$ and the sputtering power applied on the Tb target during co-sputtering deposition. 96

Fig. 4.3 RBS spectrum of $Ta_{0.92}Tb_{0.08}N_{1.0}$ film, where the concentration of each species are determined from the simulation by XRUMP [9]. 97

Fig. 4.4 AES depth profiling for $TaN/Ta_{0.94}Tb_{0.06}N_y/SiO_2$ (a-b) and $TaN/Ta_{0.95}Er_{0.05}N_y/SiO_2$ (c-d) gate stacks before and after 1000 ^oC RTA in N_2 ambient. 99

Fig. 4.5 The XPS spectra of the (a) N 1*s*, (b) Ta 4*f*, and (c) Tb 4*d* region for the as-deposited $Ta_{1-x}Tb_xN_y$ films with different Tb concentrations: (1) TaN; (2) $Ta_{0.97}Tb_{0.03}N_y$; (3) $Ta_{0.94}Tb_{0.06}N_y$; (4) $Ta_{0.9}Tb_{0.1}N_y$; (5) $Ta_{0.87}Tb_{0.13}N_y$. 101

Fig. 4.6 The XPS core level spectra in the N 1*s* region for the as-deposited (a) $Ta_{1-x}Er_xN_y$, (b) $Ta_{1-x}Yb_xN_y$, and (c) $Hf_{1-x}Tb_xN_y$ films with different lanthanide concentrations. 102 & 103

Fig. 4.7 XRD spectrums of $Ta_{1-x}Tb_xN_y$ materials with different Tb concentrations before and after 1000 ^oC RTA anneal for 20 sec, compared with that of TaN. 104

Fig. 4.8 XRD spectrums of $Ta_{0.9}Tb_{0.1}N_y$ films with different N concentrations measured in the following conditions: as-deposited, after 900 ^oC RTA and after 1000 ^oC RTA. 105

Fig. 4.9 Resistivity of $Ta_{0.9}Tb_{0.1}N_y$ films as a function of N_2/Ar flow rate ratio during the sputtering, with and without RTA performed. 106

Fig. 4.10 Resistivity of lanthanide-incorporated TaN materials as a function of lanthanide type and concentration before and after RTA treatments in N_2 ambient. 107

Fig. 4.11 High-frequency *C-V* characteristics (100 kHz) of $Ta_{1-x}Tb_xN_y$ gated MOS capacitors with different Tb concentration in $Ta_{1-x}Tb_xN_y$ on SiO_2. The measurements are taken after a 420^oC forming gas anneal. 108

Fig. 4.12 High-frequency C-V characteristics (100 kHz) of MOS-capacitors with $Ta_{0.94}Tb_{0.06}N_y$ gate electrode after 420^oC FGA and after 1000^oC RTA. The lines show the simulated *C-V* curves which takes quantum mechanical effect into account. 109

Fig. 4.13 V_{FB} vs. EOT plots of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$, $Ta_{0.95}Er_{0.05}N_y/SiO_2$, and $Hf_{0.8}Tb_{0.2}N_y/SiO_2$ gate stacks (a) after 420^oC FGA and (b) after 1000^oC RTA treatment, compared with that of TaN and HfN. 110

Fig. 4.14 Work function values of some MN_x and lanthanide-MN_x gate electrodes as a function of lanthanide type and concentrations under different annealing conditions, showing the tunability of MN_x work functions by incorporating lanthanide. 111

Fig. 4.15 *C-V* characteristics of TaN and $Ta_{0.9}Tb_{0.1}N_y$ metal gates on ALD HfAlO dielectrics after FGA at 420^oC for 30 min. and 113

RTA at 1000^oC for 5 sec.

Fig. 4.16 XTEM images of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack on (100) Si substrate after 420^oC FGA for 30 min, 900^oC RTA for 30 sec and 1000^oC RTA for 30 sec. 114

Fig. 4.17 EOT variation of the $Ta_{1-x}Tb_xN_y/SiO_2$ gate stacks before and after 1000^oC RTA for 20 sec, as a function of Tb concentrations in $Ta_{1-x}Tb_xN_y$. 115

Fig. 4.18 Gate leakage characteristics of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack with PMA performed at different temperatures. 116

Fig. 4.19 Gate leakage characteristics of $Ta_{0.95}Er_{0.05}N_y/SiO_2$ gate stack with PMA performed at different temperatures. 117

Fig. 4.20 TDDB characteristics of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack (SiO_2~3.2 nm) after PMA at different temperatures, measured under negative constant voltage stress (CVS) at room temperature. 117

Fig. 4.21 EOT variation as a function of annealing temperature for $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different N_2 flow rates during the deposition of $Ta_{0.9}Tb_{0.1}N_y$. Two groups of oxides with initial thickness of ~3.3 nm and ~ 5.8 nm were investigated. 118

Fig. 4.22 Typical *I-V* characteristics of $Ta_{0.9}Tb_{0.1}N_y$ gated MOS capacitors with different N_2 flows during metal gate deposition, measured after FGA at 420^oC for 30 min. 119

Fig. 4.23 Work function of $Ta_{0.9}Tb_{0.1}N_y$ gate electrodes with different N concentrations as a function of annealing temperature. 120

Fig. 4.24 XTEM image of $TaN/Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stack with $Ta_{0.9}Tb_{0.1}N_y$ thickness of ~ 40 Å on SiO_2 dielectric. 122

Fig. 4.25 *C-V* characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness. The measurement was taken on the samples with 1000^oC RTA for 10 sec performed. 123

Fig. 4.26 (a) V_{FB} versus EOT plots of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness after 900^oC RTA for 20 sec. (b) V_{FB} versus EOT plots of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness after 1000^oC RTA for 10 sec. 123 & 124

Fig. 4.27 Process flow of the damascene process used to pattern the TaN/ $Ta_{0.9}Tb_{0.1}N_y$ metal gate stack in MOSFET fabrication. 125

Fig. 4.28 Typical high-frequency *C-V* measurement of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET. 126

Fig. 4.29 $I_{DS} \sim V_{DS}$ characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET, with the substrate doping concentration of $N_A = 5 \times 10^{15}$ cm^{-3}. 126

Fig. 4.30 $I_{DS} \sim V_{GS}$ characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET, with the substrate doping concentration of $N_A = 5 \times 10^{15}$ cm^{-3}. 127

Fig. 4.31 Effective electron mobility in $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET. 127

Fig. 5.1 Process flow of dual metal gate integration by direct etching method. PR denotes photoresist, HM denotes hard-mask, and HK denotes high-κ dielectric. 133

Fig. 5.2 Process flow of the dual metal gate/dual high-κ integration scheme. (a) Metal-A/HK-A deposition; (b) Metal-A/HK-A selective etching from one side of CMOS; (c) Metal-B/HK-B deposition; (d) hard-mask-B deposition and patterning; (e) Metal-B/HK-B selective removal; (f) hard-mask removal, thick poly-Si top-up, and gate patterning. 135

Fig. 5.3 Process flow of the dual metal gate integration via metal inter-diffusion. 136

Fig. 5.4 Process flow of the FUSI process. (a) CMOS fabrication conventionally; (b) oxide re-flow and planarization by CMP; (c) hard-mask stripping followed by ion-implantation or poly-Si etch-back; (d1) deposition of a same metal, e.g. Ni, for both NMOS and PMOS; (d2) deposition of different metals for NMOS and PMOS, respectively (in parallel with step (d1)); (e) silicidation and unreacted metal stripping. 137

Fig. 5.5 Process flow of the replacement gate process. (a) CMOS fabrication with poly-Si as dummy gate; (b) oxide re-flow and planarization; (c) dummy poly-Si & SiO_2 removal; (d) filling the groove with new high-κ and metal gate; (e) metal CMP to pattern the metal gate; (f) dual metal gate CMOS formation by repeat steps (c)-(e). 138

Fig. 5.6 Dual metal gate integration process flow by high-temperature metal intermixing technique: (1) TaN buffer layer deposition; (2) P-type metal gate stack (e.g TaN/Ti/HfN) formation followed by selective etching; (3) N-type metal gate stack formation (e.g TaN/Tb/TaN) and capping layer deposition; (4) gate etching, S/D implantation, and dopant activation annealing (also for intermixing). 141

Fig. 5.7 XTEM images of the TaN/Tb/TaN stack on SiO_2 (a) as-deposited and (b) after 1000^oC RTA in N_2 ambient for 5 sec. 143

Fig. 5.8 (a), (b) STEM and (c), (d) EDX depth profiles of the TaN/Tb/TaN stack on SiO_2 as-deposited and after 1000°C RTA in N_2 ambient for 5 sec. In (a)& (b), the dark layer in the sandwich structure denotes Tb element. 144

Fig. 5.9 Work function versus annealing temperature for different TaN/Metal stacks. Thickness of TaN or TaN_x (less N%) bottom layers are about 2.0~2.5 nm, and that of Tb or Ir are about 2.5 nm. The N_2 gas flow rate during deposition of thin TaN layer is 5 sccm, while that for TaN_x (less N%) is 4 sccm. All samples are capped with thick TaN film of ~100 nm. 146

Fig. 5.10 Gate leakage characteristics of TaN/Tb stack after different RTA treatments. The corresponding WF of the sample is denoted by open circle in Fig. 5.9. 146

Fig. 5.11 *C-V* characteristics of n-MOSFETs with TaN, TaN/Tb/TaN and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ metal gates on HfTaON after RTA at 1000°C for 1 sec. 148

Fig. 5.12 I_{DS}-V_{GS} characteristics of n-MOSFETs with TaN, TaN/Tb/TaN (InM), and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ gates on HfTaON high-κ dielectric. 149

Fig. 5.13 V_{th} distribution of n-MOSFETs with TaN, TaN/Tb/TaN (InM) and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ gates on HfTaON/HfO_2 dielectric. (W/L=320 μm / 5 μm) 150

Fig. 5.14 Effective electron mobility in n-MOSFETs with TaN, TaN/Tb/TaN (InM) and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ metal gates on HfTaON dielectric, measured by split *C-V* measurement. (W/L = 200 μm / 20 μm) 150

Fig. 5.15 *C-V* characteristics of TaN/Ti/HfN metal stack with and without Tb on top, where the HfN thickness is (a) ~15 Å and (b) ~100 Å. 152

Fig. 5.16 AFM images of TaN (~ 2 nm) deposited on the bare-Si wafer before and after wet etching in DHF (1:200) for 30 sec. 152

Fig. 5.17 XTEM images of as-deposited dual metal gate stacks on a single wafer: TaN/Tb/TaN (left) for NMOS and TaN/Ti/HfN (right) for PMOS on SiO_2. 153

Fig. 5.18 (a) *C-V* and (b) *I-V* characteristics of TaN/Tb/TaN (N-type) and TaN/Ti/HfN (P-type) metal gate stacks on SiO_2 in as-deposited condition. 153

Fig. 5.19 *C-V* characteristics of TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) stacks on SiO_2 before and after 1000°C RTA for 1 sec. 154

Fig. 5.20 Work Function extraction for TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks on SiO_2 after metal intermixing process. 155

Fig. 5.21 Proposed replacement gate process using HfN as dummy gate: (a) CMOS fabrication using TaN/HfN/HfO_2 as the gate stacks; (b) high selective etching of TaN and HfN by wet chemicals; (c) new metal gate deposition and CMP planarization; (d) dual metal gate integration by repeating steps (b)-(c). 157

Fig. 5.22 Etching properties of the HfN/HfO_2 gate stack (open triangle symbol) and the HfO_2 (solid symbol) film after 1000°C RTA process in diluted HF solution (1:100). The etch rate of HfN is determined by surface profiler, and the remaining HfO_2 thickness is measured by ellipsometer. 159

Fig. 5.23 AFM images of HfO_2 with different process history: as-deposited HfO_2 film, HfO_2 after 1000 °C RTA anneal, and HfO_2 in HfN/HfO_2 stack with HfN removed by DHF solution after 1000 °C RTA anneal. 160

Fig. 5.24 *C-V* and *I-V* (inset) characteristics of the "control" HfN/HfO_2 devices and "re-deposited" HfN/HfO_2 devices with HfO_2 EOT~0.83 nm. The *C-V* curves were measured at 100 kHz and 1 MHz on devices with an area of 50×50 μm^2. 161

Fig. 5.25 High frequency *C-V* curves of the HfN/HfO_2 "control" devices and "re-deposited" Ta/HfO_2, Ni/HfO_2 devices. The inset compares the *C-V* curves measured from the "re-deposited" Ni/HfO_2 devices with ultra-thin HfO_2 (EOT~0.9 nm) and that of a "control" HfN/HfO_2 device. All the *C-V* curves were measured at 100 kHz on devices with an area of 50×50 μm^2. 162

Fig. 5.26 Comparison of TDDB and gate leakage (inset) characteristics between the "control" HfN/HfO_2 devices and "re-deposited" HfN/HfO_2, Ta/HfO_2, Ni/HfO_2 devices. For the TDDB study, CCS with a current density of ~8 A/cm^2 was performed on devices with an area of 100×100 μm^2 at room temperature. 163

LIST OF SYMBOLS

C_d	depletion-layer capacitance
C_i	inversion-layer capacitance
C_{ox}	gate-oxide capacitance
C_{poly}	poly-depletion capacitance
C_{si}	Si substrate capacitance
E_c	conduction band
$E_{F,m}$	Fermi level of metal
E_{ox}	electric field in gate oxide
E_v	valence band
$\hbar$	reduced Planck constant
J_{FN}	F-N tunnelling current density
K	Boltzmann constant
L_g	gate length
m	body-effect coefficient
m^*	effective electron mass
N_b	substrate doping concentration
q	elementary charge
Q_d	depletion charge density
Q_i	inversion charge density
Q_{ox}	equivalent oxide charge density at oxide/Si interface
S	Schottky pinning parameter
T_{gate}	gate electrode thickness
t_{ox}	gate-oxide thickness

Symbol	Description
V_{DD}	supply voltage
V_{FB}	flatband voltage
V_{th}	threshold voltage
γ	electronic specific heat coefficient
ε	permittivity
ε_0	permittivity in vacuum
μ_{eff}	effective carrier mobility
ρ	resistivity
	effective density of states at Fermi-level
κ	permittivity (or dielectric constant)
ϕ_{b}	Schottky barrier height
ϕ_{ox}	barrier height between gate electrode and gate oxide
$\Phi_{CNL,d}$	charge neutrality level of dielectric
Φ_m	metal work function
Φ_{MS}	work function difference between gate electrode and substrate
$\Phi_{m,eff}$	effective work function
$\Phi_{m,vac}$	metal work function in vacuum
Φ_s	semiconductor work function
$\Phi_{\mathrm{TaN/HfO2}}$	effective work function of TaN in TaN/HfO_2 stack
$\Phi_{\mathrm{TaN/SiO2}}$	effective work function of TaN in TaN/SiO_2 stack
χ_{s}	electron affinity in semiconductor

LIST OF ABBREVIATIONS

AES	Auger electron spectroscopy
AFM	atomic force microscopy
ALCVD	atomic-layer chemical vapor deposition
ALD	atomic-layer deposition
BEoL	back-end of line
BTBT	band-to-band tunnelling
BTI	bias-temperature-instability
CCS	constant current stress
CES	constant-field scaling
CET	capacitance equivalent thickness
CMOS	complementary metal-oxide-semiconductor
C-V	capacitance-voltage
CVD	chemical vapor deposition
CVS	constant-voltage scaling
	constant-voltage stress
DG	double-gate
DHF	diluted hydrofluoric (acid)
DIBL	drain-induced barrier lowering
DOF	depth-of-focus
E_{CNL}	charge-neutrality level
EDX	energy dispersive X-ray
EELS	Electron energy loss spectroscopy
EOT	equivalent oxide thickness

EWF	effective work function
FLP	Fermi-level pinning
FUSI	fully-silicided (metal gate)
F-N	Fowler-Nordheim (tunnelling)
FGA	forming-gas annealing
GIDL	gate-induced-drain leakage
GOI	gate oxide integrity
HFCV	high-frequency *C*-*V*
HK	high-κ (dielectric)
HM	hard-mask
HOT	hybride-orientation-technology
HP	high-performance
HRTEM	high-resolution transmission electron microscopy
IC	integrated circuits
I/I	ion implantation
InM	(metal) intermixing
IPE	internal photoemission
ITRS	International Technology Roadmap for Semiconductors
I-*V*	current-voltage
Lanthanide-MN_x	lanthanide-incorporated metal nitride
LOP	low-operation-power
LSTP	low-standby-power
MG	metal gate
MIGS	metal-induced gap state
MN_x	(refractory) metal nitride

MOCVD	metal-organic chemical vapor deposition
MOSFET	metal-oxide-semiconductor field-effect transistor
PC	phase-controlled (silicide)
PDA	post-deposition-annealing
PMA	post-metal-annealing
PMD	post-metal-dielectric
PR	photoresist
PVD	physical vapor deposition
RBS	Rutherford backscattering spectrometry
RF	radio-frequency
RSF	relative sensitivity factor
RTA	rapid thermal annealing
SBH	Schottky barrier height
SC-1	standard cleaning-1 (NH_4OH+H_2O_2+H_2O) solution
S/D	source/drain
SIIS	silicidation induced impurity segregation
SS	subthreshold swing
SSDOI	strained-Si directly on insulator
SSOI	strained-Si on insulator
STI	shallow trench isolation
UPS	ultraviolet photoemission spectroscopy
UTBSOI	ultra-thin-body silicon-on-insulator
UV	ultraviolet
Vo	oxygen vacancy
WF	work function

XPS	X-ray photoelectron spectroscopy
XRD	X-ray diffraction
XTEM	cross-sectional transmission electron microscope

Chapter 1

Introduction

1.1 Overview

Since the invention of the first integrated circuit (IC) in 1958, the semiconductor industry has undergone unprecedented growth through the latter half of the 20th century. Today, silicon-based IC products have infiltrated every corner of our daily life. Driven by the demand for IC chips with higher speed, greater functionality, and lower cost, the physical dimensions of metal-oxide-semiconductor field-effect transistor (MOSFET), the basic element in IC chips, have been scaled down continuously over the past 40 years so that more transistors can be integrated on a single chip. In 1965, Gordon Moore of *Intel* predicted the trend of MOSFET scaling, which is popularly known as Moore's Law: the number of transistors on a chip doubles about every two years, as shown in Fig. 1.1 [1]-[2]. This trend has been made possible by the advancements in semiconductor process technology from ~ 8 μm in 1972 to the current 65 nm technology. As a result, cost per function has decreased at an average rate of ~ 25-30% per year per function [3]. According to the prediction of the latest 2005 International Technology Roadmap for Semiconductors (*ITRS*), in the year of 2015, the physical gate length (L_g) for high-performance logic applications will shrink down to 10 nm [3], which is about 10,000 times smaller than the diameter of a hair!

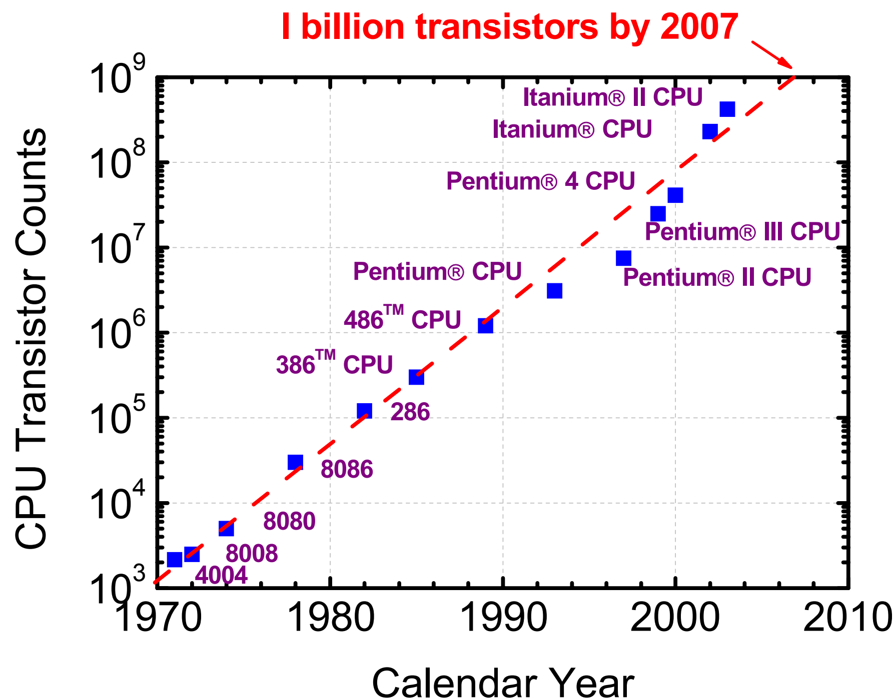


Fig. 1.1 CPU transistor counts from 1970s to present, showing the device scaling according to Moore's Law; © Intel corp. [2].

1.2 MOSFET Scaling: Challenges and Opportunities

Several scaling rules were proposed to guide the scaling of MOSFETs, such as constant-field scaling (CES), constant-voltage scaling (CVS), and generalized scaling [4]-[6]. In practice, the generalized scaling rule was followed in the modern complementary metal-oxide-semiconductor (CMOS) technology. The principle of the generalized scaling is to scale the electric field and the physical dimensions (both lateral and vertical) of MOSFET by different factors, α and κ, respectively [6]. Under this protocol, the supply voltage (V_{DD}) typically scales slower than the channel length, which leads to the increase of electric field by a factor of α, as well as the increase of power density by a factor of α^2 to α^3 [6]. Fig. 1.2 illustrates the scaling of V_{DD},

threshold voltage (V_{th}), and gate-oxide thickness (t_{ox}) as a function of channel length [7], showing the different scaling factors of V_{DD} compared with that of L_g. As a result, the power dissipation of chips became higher and higher.

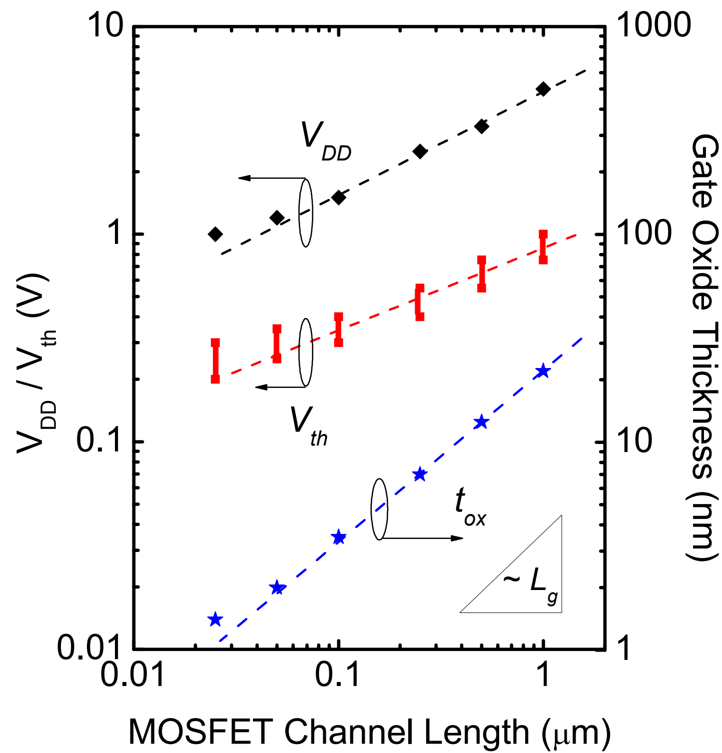


Fig. 1.2 Historical scaling trends of supply voltage (V_{DD}), threshold voltage (V_{th}) and gate-oxide thickness (t_{ox}) vs. channel length (L_g) for CMOS logic technologies, showing the different scaling factors for supply voltage and device dimension [7].

The power-performance trade-off has become the major road-block for the continuous scaling of CMOS into deep-submicron regimes. In order to squeeze the maximum performance gain from the continuous scaling while maintaining the power consumption at an acceptable level, innovative device structures and new materials have been explored extensively in recent years [8].

1.2.1 Leakages in Deep-Submicrometer MOSFET

High leakage current is becoming a most serious issue for aggressively scaled MOSFETs in the sub-50-nm regime, and is very likely to be the show-stopper for the MOSFET scaling eventually. The major leakage components in short-channel devices are 1) tunneling current through the thin gate oxide; 2) subthreshold leakage between source and drain; 3) band-to-band tunneling (BTBT) current through drain-well junction; 4) gate-induced-drain leakage (GIDL), and 5) punchthrough leakage [9]. Fig. 1.3 is the schematic cross section of a MOSFET, illustrating various leakage components.

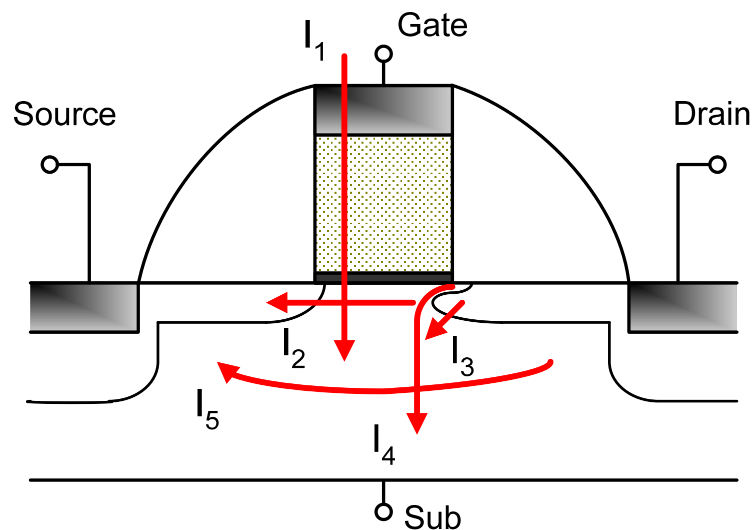


Fig. 1.3 Schematic cross section of MOSFET showing the major leakage current paths. I_1 for direct tunneling through gate oxide; I_2 for subthreshold leakage; I_3 for BTBT; I_4 for GIDL; and I_5 for punchthrough.

The leakage problems in MOSFET rise with the decreasing distance between the four terminals in the vertical and horizontal directions. As the source and drain terminals approach each other and the distance becomes comparable with the MOS depletion width in the channel region, the conventional 1-D field pattern for long-channel devices, where the electric field in the channel region is controlled by the gate

electrode only, will no longer be valid. Instead, the source and drain fields penetrate deeply into the middle of the channel, which lowers the potential barrier between the source and drain and causes a substantial increase of the subthreshold current. This is referred to as drain-induced barrier lowering (DIBL) effect [10]. The DIBL effect will be further amplified when a high drain voltage is applied, leading to the decrease of V_{th} and dramatic increase of the subthreshold current by diffusion.

The subthreshold leakage current (I_{sub}) is considered as one of the major contributors to the "off-state" power dissipation (or passive power), as described by the following equation [11]:

$$P_{off} = W_{tot} V_{DD} I_{off} = W_{tot} V_{DD} I_0 \exp\left(-\frac{qV_{th}}{mKT}\right) \qquad (1\text{-}1)$$

where W_{tot} is the total turn-off device width, I_{off} is the average off-current per device width, I_0 is the extrapolated current per width at threshold voltage, and m is body-effect factor typically ranging from 1.1 to 1.4 [11]. For the modern CMOS circuits, the passive power can even exceed the active switching power eventually [12].

In order to regulate the I_{sub} to a tolerable level so that the lateral scaling of MOSFET can be implemented, the influence of drain electric field to the channel region must be minimized so that the "gate control" can overwhelm that of the drain. There are several approaches to achieve this goal in device design: increase the gate capacitance (C_{ox}), reduce the source/drain (S/D) junction depth, or engineer the doping profiles in the channel region. Increasing channel doping level is a typical way to control I_{sub}. In order to minimize the side-effects associated with the high channel doping, including the high electric-field, high BTBT leakage, large subthreshold swing (SS) and low mobility, non-uniform channel engineering techniques such as retrograde channel doping [13] and halo implantation [14] have been introduced into the MOSFET fabrication. Moreover, increasing C_{ox} and/or

reducing S/D junction depths can also help to control the short-channel effects, but the trade-off is the increased gate leakage and S/D series resistance. Today, transistor with 10 nm L_g has been demonstrated with manageable subthreshold current [15].

1.2.2 Vertical Scaling of MOSFET Gate Stack

In addition to the lateral scaling, the gate stack of MOSFET also needs to be scaled down to provide a better gate control to the channel and improve the drive capability of a MOSFET. The saturation drive current of a MOSFET can be depicted by the following equation:

$$I_{ds,sat} = \mu_{eff} C_{ox} \frac{W}{L} \frac{(V_g - V_{th})^2}{2m} = \mu_{eff} \frac{\varepsilon_0 \varepsilon_{SiO_2}}{CET} \frac{W}{L} \frac{(V_g - V_{th})^2}{2m} \tag{1-2}$$

where C_{ox} is the gate capacitance, μ_{eff} is the effective carrier mobility, ε_{SiO2} is the permittivity of gate oxide, *CET* is the capacitance equivalent thickness of the gate stack which includes the contributions from the poly-depletion and quantum mechanical effects, and m is the body-effect coefficient [11]. From this equation, it is clear that the drive current can be improved by reducing the gate oxide thickness. Table 1.1 summarizes the requirements for today's and tomorrow's gate dielectric in different applications according to *ITRS* 2004, showing the aggressive scaling of gate oxide thickness.

The outstanding properties of SiO_2 have enabled the vertical scaling of Si-based MOSFET for several decades. SiO_2 possess many ideal dielectric properties, such as amorphous structure, thermodynamical and electrical stability, wide band gap of ~ 9 eV, smooth interface with Si, and so on. However, when the physical thickness of SiO_2 shrinks to less than about 3 nm, the direct tunneling current [16]-[17] through the thin SiO_2 will become significant and rise exponentially as the thickness of SiO_2

decreases. This has become one of the major issues for MOSFET scaling. Although incorporating nitrogen into SiO_2 to form SiON can slightly reduce the gate leakage, the situation is still getting worse and worse as the gate oxide thickness shrinks towards the sub-1-nm regime [18]-[19]. On the other hand, from the material point of view, the minimum thickness needed for SiO_2 to maintain its bulk properties (i.e. band-gap) is about 7 Å [20], which is too thick for the requirement of high-end applications after year 2010 according to Table 1.1. Therefore, an alternative way out of the quandary needs to be carved out.

Table 1.1 Technology roadmap for the scaling of dielectrics thickness in next ten years [3].

Year of Production	***2004***	***2007***	***2010***	***2013***	***2016***
Technology node	hp90	hp65	hp45	hp32	hp22
Physical gate length for MPU (nm)	37	25	18	13	9
Physical gate length for low-operating-power (nm)	53	32	22	16	11
Physical gate length for low-standby-power (nm)	65	37	25	18	13
EOT for MPU (nm)	1.2	0.9	0.7	0.6	0.5
EOT for low-operating-power (nm)	1.5	1.2	0.9	0.8	0.7
EOT for low-standby-power (nm)	2.1	1.6	1.3	1.1	1
Gate leakage at 100^oC for high performance (A/cm^2)	460	920	1,833	7,692	18,556
Gate leakage at 100^oC for low-operating-power (A/cm^2)	1.9	5.2	10.6	20.8	90.9
Gate leakage at 100^oC for low-standby-power (mA/cm^2)	4.6	21.6	80	150	254

Using high-permittivity (κ) dielectrics to replace SiO_2 or SiON would be a potential solution to enable the further scaling of the gate stack in MOSFET [21]-[22]. The advantage of high-κ gate dielectrics rather than SiO_2 is to provide a physically thicker film for leakage current reduction while improving the gate capacitance by higher permittivity, as described in equation 1-3:

$$EOT = \frac{\varepsilon_{SiO_2}}{\varepsilon_{high-\kappa}} T_{high-\kappa,Phy} \tag{1-3}$$

where EOT is the equivalent oxide thickness of the high-κ dielectric, ε_{SiO2} and $\varepsilon_{high\text{-}\kappa}$ are the permittivity of SiO_2 and the high-κ dielectric, respectively, and $T_{\text{high-}\kappa\text{,Phy}}$ is the physical thickness of the high-κ film.

The candidate high-κ materials should have suitable permittivity ($\kappa \approx$ 15-25), large barrier height for both electron and hole, high crystallization temperature, good thermal stability and good interface quality with Si substrate and gate electrodes, and high carrier mobility for both electrons and holes [21]. Among various high-κ materials investigated, Hf-based high-κ dielectrics have drawn considerable attention due to their appropriate κ values and relatively high barrier heights for both electrons and holes. High-κ materials such as HfO_2 [23]-[24], HfAlO [25], HfSiO [26]-[27], HfON [28], and HfSiON [29]-[30] have been extensively studied. Fig. 1.4 shows the scalability of some higk-κ materials compared with the *ITRS* requirements. It clearly shows that the gate leakage reduction by 2 to 4 orders compared to SiO_2 can be achieved using high-κ dielectrics.

However, there are still some challenges for high-κ dielectric to replace SiO_2. The first one is mobility degradation. Coulomb scattering due to the pre-existing and trapped charges in high-κ [31]-[32] and remote phonon scattering associated with the ionic properties of the "soft" metal-oxygen bonds in high-κ films [33]-[34] have been

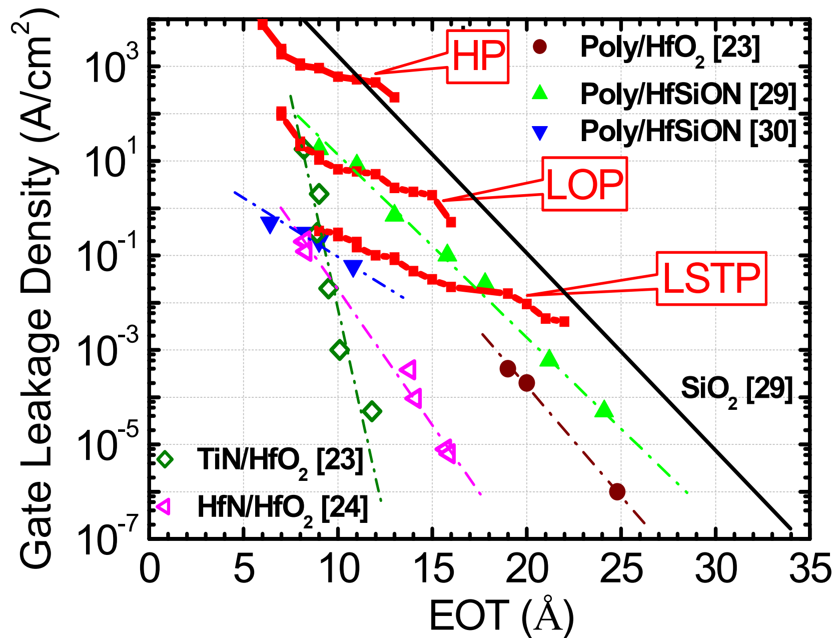


Fig. 1.4 Gate leakage current density of some high-*κ* dielectrics as a function of EOT, compared with the gate leakage specifications at 100°C for high-performance (HP), low-operating-power (LOP), and low-standby-power (LSTP) applications according to *ITRS* 2004 [3].

proposed to account for the electron mobility degradation in high-*κ* stacks. Among various high-*κ* dielectric candidates, $HfSiO_x$ shows most promising mobility characteristics [35], but the permittivity is relatively low. Some methods have been demonstrated to improve the mobility characteristics, such as improving the microstructure of high-*κ* films [36] or inserting a SiO_x(N) layer under the high-*κ* [37]-[38]. The second issue for high-*κ* dielectric is the Fermi-level pinning (FLP) problem at the high-*κ*/poly-Si interface [39]-[40], which causes a high V_{th} for MOSFETs, especially for p-MOSFET. This problem may exclude the use of poly-Si/high-*κ* gate stack in high-performance applications where low V_{th} of around +/-0.25 V are required to maintain enough gate overdrive. Hf-Si bond-induced interface dipole [39] and/or oxygen vacancy induced charge transfer across the poly-Si/high-*κ* interface

[41]-[42] have been proposed to explain the FLP phenomenon. Some techniques such as F^- implantation could be useful in mitigating the FLP problem [43]. Finally, some reliability issues, such as charge trapping and bias-temperature-instability (BTI) in high-κ dielectric [44]-[45], are still not well understood. Therefore, scientific understanding and technology development of high-κ materials are still expected before its implementation in CMOS technology.

Poly-Si depletion effect is another factor affecting the device scaling in the vertical direction [46]. The poly-depletion layer accounts for an additional thickness of about 4 Å to the capacitance equivalent thickness (CET) of the gate stack and results in a significant loss of gate control. This problem is particularly serious when the gate oxide scales to the sub-1-nm regime. Metal gate technology can be used to eliminate the poly-Si depletion effect and hence improve the device performance. Meanwhile, metal gate electrodes can also address some other concerns associated with the poly-Si gate electrodes. These make the metal gate technology one of the hottest research areas in recent years. The detailed backgrounds about metal gate technology will be discussed in Chapter 2.

1.2.3 Innovations in Device Structures

In order to manage the short-channel effects in the aggressively scaled devices, many novel device structures have been proposed and investigated, including ultra-thin-body silicon-on-insulator (UTBSOI), double-gate (DG), FinFET, triple-gate, Ω-gate FET, nanowire FET etc. [47]-[53]. In these device structures, the gate to channel potential coupling can be greatly improved by their special device geometry compared with planar bulk-Si CMOS, so that the short-channel characteristics can be controlled. Consequently, the intrinsic silicon channel can be adopted, which enables

lower channel electric field, lower BTBT leakage, sharper subthreshold slope and better carrier mobility to be achievable. These advantages make them very attractive as potential technology options for the future high-performance applications.

However, there are still many challenges for these novel device structures. One of these challenges is threshold voltage adjustment. Due to the small amount of depletion charges and the intrinsic Si channel used, the desirable gate work function for these devices should be close to the mid-gap of Si [3]. Conventional poly-Si gate will not work properly in this situation and novel metal gate electrodes with mid-gap work functions are required [54]. Secondly, the high source/drain series resistance caused by the thin silicon body used in these 3-D structures is another concern which may affect the overall performance of these novel FETs. Thirdly, the carrier transport characteristics in the ultra-thin Si channels will be very sensitive to the Si-body thickness T_{Si} [55]-[57], rendering a requirement for very strict process control of T_{Si}. The manufacturing tolerance for T_{Si} would be added up with the tolerance in defining the gate length, resulting in even smaller process windows and higher manufacturing cost in fabricating these novel structures compared with the conventional planar devices. Therefore these novel structures may only be used for some kernel parts in MPU or ASIC chips.

1.2.4 Mobility Enhancement for Performance Gain

Another way to improve the performance of MOSFET is to enhance the carrier mobility, which may provide a key to escape from the power/speed box in device scaling [58]. Basically there are three avenues to achieving the enhanced mobility for MOSFET: inducing strain to the channel region, utilizing the high mobility surface orientation, or employing new channel materials with high mobility

and high saturation velocity. Sometimes these techniques are combined to get the utmost gain in performance.

There are two groups of technologies to introduce strain into the channel of MOSFET. One is the local strain or called process induced strain technologies, including the strain from shallow trench isolation (STI), Si_3N_4 stress liners, silicide induced strain, and embedded SiGe or SiC stressors in the source/drain region, etc [62]-[67]. These techniques are based on the conventional bulk-Si CMOS process and thus have the advantages like low-cost and easy integration. Some of these techniques have already been adopted in the latest 65-nm CMOS technology for the mass production. The other group of technologies is global-strain technologies, in which the strain is induced from substrate, such as strained-Si on relaxed-SiGe, strained-Si on insulator (SSOI), strained-Si directly on insulator (SSDOI), and so on [68]-[71]. One of the concerns for the global-strain techniques is the difficulty to optimize the n-MOSFET and p-MOSFET individually. Another concern is the cost issue because strained-Si substrates with very low defect level are required.

The surface orientation and channel direction (current flow direction) can also affect the carrier mobility in MOSFET. In conventional bulk-Si CMOS technology, CZ-Si with (100) surface orientation are commonly used. M. Yang *et al.* developed a novel hybride-orientation-technology (HOT) to integrate (100) and (110) surface orientation on a same wafer to get ideal mobility for n-MOSFET and p-MOSFET, respectively [72]. Recent progress demonstrates that the HOT technology can also be integrated with other uniaxial local strain technologies for more performance enhancement [73].

High-mobility and high-saturation-velocity semiconductors, such as SiGe, Ge, InP, and GaAs, have also attracted considerable attentions as the possible candidates

for channel materials [74]-[76]. Compared with Si, the physical properties of these materials are still not very well understood yet. Many process issues also need to be addressed, including the dielectric/channel interface engineering and the S/D junction formation [77]. Moreover, integration of these materials into the conventional Si-based CMOS process flow would be another challenge. More comprehensive study on these materials will be appreciated in the future.

1.3 Summary

In summary, immense challenges arise as the Si-based CMOS technology enters the sub-50-nm node, accompanied by the great opportunities for novel materials and integration technologies. CMOS scaling will no longer be driven by photolithography alone; rather it will be driven by the innovations in developing advanced materials/structures and the ability to integrate these novel materials and structures into CMOS fabrication processes. Reliable and cost-effective solutions are expected to unlash the power-performance deadlock. Among the various technology choices discussed above, metal gate and high-κ technologies are among the most urgent technologies required by the semiconductor industry to keep step with Moore's law in the future ten years [3].

The work in this thesis will be focused on understanding and developing the advanced metal gate technology, which is particular important for the scaling of MOSFET in the vertical direction. According to *ITRS* 2005, metal gate technology has been considered as one of the performance booster for CMOS in 32-nm technology node and beyond. The overall objective of this thesis is to gain insights into the major issues and to address some of the critical challenges in implementing the metal gate electrodes in the future CMOS platform. The background information

and recent developments of the metal gate technology will be introduced in Chapter 2, where the major issues and challenges will be highlighted. According to the different challenges and the efforts to address these problems, the work in this thesis will be divided into three aspects. In Chapter 3, the role of metal-dielectric interface in affecting the effective work function of metal gates will be investigated systematically. After that, a novel approach to modify the work function of metal nitride materials will be presented in Chapter 4, which would be valuable for the applications in NMOS devices using the conventional gate-first process flow. The dual metal gate integration issues will be discussed in Chapter 5, where two novel integration schemes will be proposed and demonstrated in order to address some of the challenges faced in the process integration of dual metal gates. Finally, the major results achieved in this thesis will be summarized in Chapter 6, as well as some suggestions on future research work.

References

1. G. E. Moore, "Progress in digital integrated electronics," in *IEDM Tech. Dig.*, pp. 11-13, 1975.

2. From *Intel corp.*; http://www.intel.com/technology/silicon/mooreslaw/index.htm

3. *International Technology Roadmap of Semiconductors (ITRS)*, Semiconductor Industry Association (SIA), SanJose, CA 95110. Available: http://public.itrs.net/

4. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SC-9, p. 256, 1974.

5. G. Baccarani, M.R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to ¼ micrometer MOSFET design," *IEEE Tran. Electron. Dev.*, vol. 31, pp. 452, 1984.

6. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 4, pp. 167-170.

7. Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. & Dev.*, vol. 46, pp. 213-222, 2002.

8. H.-S. P. Wong, "Beyond the conventional transistors," *IBM J. Res. & Dev.*, vol. 46, pp. 133-168, 2002.

9. K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proc. IEEE*, vol. 91, pp. 305-327, 2003.

10. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 3, pp. 143-144.

11. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 3, pp. 126-130.

12. E. J. Nowak, "Maintaining the benefits of CMOS scaling when scaling bogs down," *IBM J. Res. & Dev.*, vol. 46, pp. 169-180, 2002.

13. Indranil De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE Tran. Electron. Dev.*, vol. 46, pp. 1711-1717, 1999.

14. S. Thompson, P. Packan, and M. Bohr, "MOS scaling: transistor challenges for the 21st century," *Intel Tech. Journal*, vol. 2, no.3, 1998.

15. H. Wakabayashi, T. Ezaki, M. Hane, T. Ikezawa, T. Sakamoto, H. Kawaura, S. Yamagami, N. Ikarashi, K. Takeuchi, T. Yamamoto, and T. Mogami, "Transport

properties of sub-10-nm planar-bulk-CMOS devices," in *IEDM Tech. Dig.*, pp. 429-432, 2004.

16. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 2, pp. 96-97.

17. N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, "Modeling study of ultrathin gate oxides using direct tunneling current and capacitance-voltage measurements in MOS devices," *IEEE Tran. Electron. Dev.*, vol. 46**,** pp. 1464-1471, 1999.

18. X. Guo, and T. P. Ma, "Tunneling leakage current in oxynitride: dependence on oxygen/nitrogen content," *IEEE Electron Device Lett.*, vol. 19, no. 6, pp. 207-209, 1998.

19. S. Inaba, T. Shimizu, S. Mori, K. Sekine, K. Saki, H. Suto, H. Fukui, M. Nagamine, M. Fujiwara, T. Yamamoto, M. Takayanagi, I. Mizushima, K. Okano, S. Matsuda, H. Oyamatsu, Y. Tsunashima, S. Yamada, Y. Toyoshima, and H. Ishiuchi, "Device performance of sub-50 nm CMOS with ultra-thin plasma nitrided gate dielectrics," in *IEDM Tech. Dig.*, pp. 651-654, 2002.

20. D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, "The electronic structure at the atomic scale of ultrathin gate oxides," *Nature* (London) vol. 399, pp.758-761, 1999.

21. G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-k gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.,* vol. 89, pp. 5243-5275, 2001.

22. Y.-C. Yeo, T.-J. King, and C. Hu, "MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations," *IEEE Tran. Electron. Dev.*, vol. 50**,** pp. 1027-1035, 2003.

23. W. Tsai, L. Ragnarsson, L. Pantisano, P. J. Chen, B. Onsia, T. Schram, E. Cartier, A. Kerber, E. Young, M. Caymax, S. De Gendt, and M. Heyns, "Performance comparison of sub-1-nm sputtered TiN/HfO_2 nMOS and pMOSFETs," in *IEDM Tech. Dig.* pp. 311-314, 2003.

24. H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du, and D. Kwong, "Thermally robust high quality HfN/HfO_2 gate stack for advanced CMOS devices," in *IEDM Tech. Dig.*, pp. 99-102, 2003.

25. S. G. Park, Y. K. Lee, S. B. Kang, H. S. Jung, S. J. Doh, J.-H. Lee, J. H. Choi, G. H. Kim, G. H. Choi, U. I. Chung, and J. T. Moon, "Performance improvement of MOSFET with HfO_2-Al_2O_3 laminate gate dielectric and CVD-TaN metal gate deposited by TAIMATA," in *IEDM Tech. Dig.*, pp. 327-330, 2003.

26. G. D. Wilk and R. M. Wallace, "Electrical properties of hafnium silicate gate dielectrics deposited directly on silicon," *Appl. Phys. Lett.*, vol. 74, pp. 2854-2856, 1999.

27. M. A. Quevedo-Lopez, M. El-Bouanani, B. E. Gnade, R. M. Wallace, M. R. Visokay, M. Douglas, M. J. Bevan, and L. Colombo, "Interdiffusion studies for $HfSi_xO_y$ and $ZrSi_xO_y$ on Si," *J. Appl. Phys.,* vol. 92, pp. 3540-3550, 2002.

28. C. S. Kang, H.-J. Cho, K. Onishi, R. Choi, R. Nieh, S. Goplan, S. Krishnan, and J. C. Lee, "Improved thermal stability and device performance of ultra-thin (EOT<10Å) gate dielectric MOSFETs by using hafnium oxynitride (HfO_xN_y)," in *VLSI Tech. Dig.*, pp. 146-147, 2002.

29. A.L.P. Rotondaro, M.R. Visokay, J.J. Chambers, A. Shanware, R. Khamankar, H. Bu, R.T. Laaksonen, L.Tsung, M. Douglas, R. Kuan, M.J. Bevan, T. Grider, J. McPherson, and L. Colombo, "Advanced CMOS transistors with a novel HfSiON gate dielectric," in *VLSI Tech. Dig.*, pp. 148-149, 2002.

30. M. Koike, T. Ino, Y. Kamimuta, M. Koyama, Y. Kamata, M. Suzuki, Y. Mitani, A. Nishiyama, and Y. Tsunashima, "Effect of Hf-N bond on properties of thermally stable amorphous HfSiON and applicability of this material to sub-50 nm technology node LSIs, " in *IEDM Tech. Dig.*, pp. 107-110, 2003.

31. M. Hiratani, S. Saito, Y. Shimamoto, K. Torii, "Effective electron mobility reduced by remote charge scattering in high-kappa gate stacks" *Jpn. J. Appl. Phys.* vol. 41, pp.4521-4522, 2002.

32. F. Lime, K. Oshima, M. Casse, G. Ghibaudo, S. Cristoloveanu, B. Guillaumot, H. Iwai, "Carrier mobility in advanced CMOS devices with metal gate and HfO2 gate dielectric," *Solid-State Electronics*, vol. 47, pp. 1617-1621, 2003.

33. M. Fischetti, D. Neumayer, and E. Cartier, "Effective electron mobility in Si inversion layers in MOS systems with a high-k insulator: The role of remote phonon scattering," *J. Appl. Phys.,* vol. 90, pp. 4587-4608, 2001.

34. W. J. Zhu and T. P. Ma, "Temperature dependence of channel mobility in HfO_2-gated NMOSFETs," *IEEE Electron Device Lett.*, vol. 25, pp. 89-91, 2004.

35. Y. Kim, H. J. Lim, H. Jung, J. Lee, J. Park, S. K. Han, J. H. Lee, S. Doh, J. P. Kim, N. I. Lee, Y. Chung, H. Y. Kim, N. K. Lee, S. Ramanathan, T. Seidel, M. Boleslawski, G. Irvine, B. Kim, H. Lee, and H. Kang, "Characteristics of ALD $HfSiO_x$ using new Si precursors for gate dielectric applications," in *IEDM Tech. Dig.*, pp. 511-514, 2004.

36. X. Yu, C. Zhu, X. P. Wang, M. F. Li, A. Chin, A. Y. Du, W. D. Wang, and D.-L. Kwong, "High mobility and excellent electrical stability of MOSFETs using a novel HfTaO gate dielectric," in *Symp. VLSI Tech. Dig.*, pp. 110-111, 2004.

37. A. Morioka, H. Watanabe, M. Miyamura, T. Tatsumi, M. Saitoh, T. Ogura, T. Iwamoto, T. Ikarashi, Y. Saito, Y. Okada, H. Watanabe, Y. Mochiduki, and T.

Mogami, "High mobility MISFET with low trapped charge in HfSiO films," in *Symp. VLSI Tech. Dig.*, pp. 165-166, 2003.

38. H. Ota, N. Yasuda, T. Yasuda, Y. Morita, N. Miyata, K. Tominaga, M. Kadoshima, S. Migita, T. Nabatame, and A. Toriumi, "Study on oxynitride buffer layers in HfO_2 metal-insulator-semiconductor structures for improving metal-insulator-semiconductor field-effect transistor performance," *Jpn. J. Appl. Phys.*, vol. 44, pp.1698-1703, 2005.

39. C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, G. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, and P. Tobin, "Fermi level pinning at the polySi/metal oxide interface", in *Symp. VLSI Tech. Dig.*, pp. 9-10, 2003.

40. C. W. Yang, Y. K. Fang, C. H. Chen, S. F. Chen, C. Y. Lin, and C. S. Lin, M. F. Wang, Y. M. Lin, T. H. Hou, C. H. Chen, L. G. Yao, S. C. Chen, and M. S. Liang, "Effect of polycrystalline-silicon gate types on the opposite flatband voltage shift in n-type and p-type metal–oxide–semiconductor field-effect transistors for high-k-HfO_2 dielectric," *Appl. Phys. Lett.* vol. 83, pp. 308-310, 2003.

41. K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Kohno, T. Chikyo, H. Kitajima, and T. Arikado, "Physics in Fermi level pinning at the polySi/Hf-based high-k oxide interface," in *Symp. VLSI Tech. Dig.*, pp. 108-109, 2004.

42. H. Takeuchi, H. Y. Wong, D. Ha, and T. King, "Impact of oxygen vacancies on high-k gate stack engineering," in *IEDM Tech. Dig.*, pp. 829 - 832, 2004.

43. M. Inoue, S. Tsujikawa, M. Mizutani, K. Nomura, T. Hayashi, K. Shiga, J. Yugami, J. Tsuchimoto, Y. Ohno, and M. Yoneda, "Fluorine incorporation into HfSiON dielectric for V_{th} Control and its impact on reliability for Poly-Si gate pFET," in *IEDM Tech. Dig.*, pp. 413-416, 2005.

44. D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.,* vol. 94, pp. 1-18, 2003.

45. B. H. Lee, C. D. Young, R. Choi, J. H. Sim, G. Bersuker, C. Y. Kang, R. Harris, G. A. Brown, K. Matthews, S. C. Song, N. Moumen, J. Barnett, P. Lysaght, K. S. Choi, H. C. Wen, C. Huffman, H. Alshareef, P. Majhi, S. Gopalan, J. Peterson, P. Kirsh, H. Li, J. Gutt, M. Gardner, H. R. Huff, P. Zeitzoff, R. W. Murto, L. Larson, and C. Ramiller, "Intrinsic characteristics of high-k devices and implications of fast transient charging effects (FTCE)," in *IEDM Tech. Dig.*, pp. 859-862, 2004.

46. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 2, pp. 74-78.

47. H.-S. Wong, D. Frank, Y. Taur, and J. Stork, "Design and performance considerations for sub-0.1 μm double-gate SOI MOSFETs," in *IEDM Tech. Dig.*, pp. 747, 1994.

48. B. Doris, M. Ieong, T. Kanarsky, Y. Zhang, R. A. Roy, O. Dokumaci, Z. Ren, F. Jamin, L. Shi, W. Natzle, H. Huang, J. Mezzapelle, A. Mocuta, S. Womack, M. Gribelyuk, E. C. Jones, R. J. Miller, H. P. Wong, and W. Haensch, "Extreme scaling with ultra-thin Si channel MOSFETs," in *IEDM Tech. Dig.*, pp. 267-270, 2002.

49. D. Hisamoto, W.-C. Lee; J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, "A folded-channel MOSFET for deep-sub-tenth micron era," in *IEDM Tech. Dig.*, pp. 1032-1034, 1998.

50. Y. Choi, N. Lindert, P. Xuan, S. Tang, D. Ha, E. Anderson, T. King, J. Bokor, and C. Hu, "Sub-20nm CMOS FinFET technologies," in *IEDM Tech. Dig.*, pp. 421-424, 2001.

51. R. Chau et al., "Advanced depleted-substrate transistors: single-gate, double-gate and tri-gate," in *Proc. Int. Conf. on Solid State Dev. and Mat.(SSDM)*, pp. 68-69, 2002.

52. J. Kavalieros, B. Doyle, S. Datta, G. Dewey, M. Doczy, B. Jin, D. Lionberger, M. Metz, W. Rachmady, M. Radosavljevic, U. Shah, N. Zelick and R. Chau, "Tri-gate transistor architecture with high-k gate dielectrics, metal gates and strain engineering," in *Symp. VLSI Tech. Dig.*, pp. 62-63, 2006.

53. F. Yang, D. Lee, H. Chen, C. Chang, S. Liu, C. Huang, T. Chung, H. Chen, C. Huang, Y. Liu, C. Wu, C. Chen, S. Chen, Y. Chen, Y. Chen, C. Chen, B. Chan, P. Hsu, J. Shieh, H. Tao, Y. Yeo, Y. Li, J. Lee, P. Chen, M. Liang, and C. Hu, "5nm-gate nanowire FinFET," in *Symp. VLSI Tech. Dig.*, pp. 196-197, 2004.

54. J. Kedzierski, E. Nowak, T. Kanarsky, Y. Zhang, D. Boyd, R. Carruthers, C. Cabral, R. Amos, C. Lavoie, R. Roy, J. Newbury, E. Sullivan, J. Benedict, P. Saunders, K. Wong, D. Canaperi, M. Krishnan, K. Lee, B. A. Rainey, D. Fried, P. Cottrell, H. P. Wong, M. Ieong, and W. Haensch, "Metal-gate FinFET and fully-depleted SOI devices using total gate silicidation," in *IEDM Tech. Dig.*, pp. 247 - 250, 2002.

55. K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental study on carrier transport mechanism in ultrathin-body SOI n-and p-MOSFETs with SOI thickness less than 5 nm," in *IEDM Tech. Dig.*, pp. 47-50, 2002.

56. S. Takagi, J. Koga, and A. Toriumi, "Subband structure engineering for performance enhancement of Si MOSFETs," in *IEDM Tech. Dig.*, pp. 219-222, 1997.

57. T. Low, M. F. Li, W. J. Fan, S. T. Ng, Y. Yeo, C. Zhu, A. Chin, L. Chan, and D. L. Kwong, "Impact of surface roughness on silicon and germanium ultra-thin-body MOSFETs," in *IEDM Tech. Dig.*, pp. 151-154, 2004.

58. Chenming Hu, "Device Challenges and opportunities," in *Symp. VLSI Tech. Dig.*, pp. 4-5, 2004.

59. S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, "Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal–oxide–semiconductor field-effect transistors," *J. Appl. Phys.,* vol. 80, pp. 1567-1577, 1996.

60. K. Rim, J. L. Hoyt, and J. F. Gibbons, "Fabrication and analysis of deep submicron strained-Si NMOSFETs," *IEEE Tran. Electron. Dev.*, vol. 47, pp. 1406-1415, 2000.

61. S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C.-H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Youssef El-Mansy, "A 90-nm logic technology featuring strained-silicon," *IEEE Tran. Electron. Dev.*, vol. 51, pp. 1790-1797, 2004.

62. S. Tiwari, M. Fischetti, P. Mooney, and J. Welser, "Hole mobility improvement in silicon-on-insulator and bulk silicon transistors using local strain," in *IEDM Tech. Dig.*, pp. 939-941, 1997.

63. A. Shimizu, K. Hachimine, N. Ohki, H. Ohta, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, "Local mechanical-stress control (LMC): A new technique for CMOS-performance enhancement," in *IEDM Tech. Dig.*, pp. 433-436, 2001.

64. H. S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, *et al.*, "Dual stress liner for high performance sub-45nm gate length SOI CMOS manufacturing," in *IEDM Tech. Dig.*, pp. 1075-1078, 2004.

65. A. Steegen, M. Stucchi, A. Lauwers, and K. Maex, "Silicide induced pattern density and orientation dependent transconductance in MOS transistors," in *IEDM Tech. Dig.*, pp. 497-500, 1999.

66. T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr, "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors," in *IEDM Tech. Dig.*, pp. 978-980, 2003.

67. K. W. Ang, K. J. Chui, V. Bliznetsov, A. Du, N. Balasubramanian, M. F. Li, G. Samudra, and Y. Yeo, "Enhanced performance in 50 nm N-MOSFETs with silicon-carbon source/drain regions," in *IEDM Tech. Dig.*, pp. 1069-1071, 2004.

68. K. Rim, J. Chu, H. Chen, K.A. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H.-S. Wong, "Characteristics and device design of sub-100 nm strained Si N- and PMOSFETs," in *Symp. VLSI Tech. Dig.*, pp. 98-99, 2002.

69. T. Mizuno, S. Takagi, N. Sugiyama, J. Koga, T. Tezuka, K. Usuda, T. Hatakeyama, A. Kurobe, and A. Toriumi, "High performance strained-Si p-MOSFETs on SiGe-on-insulator substrates fabricated by SIMOX technology," in *IEDM Tech. Dig.*, pp. 934-936, 1999.

70. T. Tezuka, N. Sugiyama, T. Mizuno and S. Takagi, "High-performance strained-Si-on-insulator MOSFETs by novel fabrication processes utilizing Ge-condensation technique," in *Symp. VLSI Tech. Dig.*, pp. 96-97, 2002.

71. K. Rim, K. Chan, L. Shi, D. Boyd, J. Ott, N. Klymko, F. Cardone, L. Tai, S. Koester, M. Cobb, D. Canaperi, B. To, E. Duch, I. Babich, R. Carruthers, P. Saunders, G. Walker, Y. Zhang, M. Steen, and M. Ieong, "Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs," in *IEDM Tech. Dig.*, pp. 47-52, 2003.

72. M. Yang, M. Ieong, L. Shi, K. Chan, V. Chan, A. Chou, E. Gusev, K. Jenkins, D. Boyd, Y. Ninomiya, D. Pendleton, Y. Surpris, D. Heenan, J. Ott, K. Guarini, C. D'Emic, M. Cobb, P. Mooney, B. To, N. Rovedo, J. Benedict, R. Mo, and H. Ng, "High performance CMOS fabricated on hybrid substrate with different crystal orientations," in *IEDM Tech. Dig.*, pp. 453-456, 2003.

73. C. D. Sheraw, M. Yang, D. M. Fried, G. Costrini, T. Kanarsky, W-H. Lee, V. Chan, M. V. Fischetti, J. Holt, L. Black, M. Naeem, S. Panda, L. Economikos, J. Groschopf, A. Kapur, Y. Li, R. T. Mo, A. Bonnoit, D. Degraw, S. Luning, D. Chidambarrao, X. Wang, A. Bryant, D. Brown, C-Y. Sung, P. Agnello, M. Ieong, S-F. Huang, X. Chen, and M. Khare, "Dual stress liner enhancement in Hybrid Orientation Technology," in *Symp. VLSI Tech. Dig.*, pp. 12-13, 2005.

74. H. Shang, H. Okorn-Schmidt, K. K. Chan, M. Copel, J. A. Ott, P. M. Kozlowski, S. E. Steen, S. A. Cordes, H. P. Wong, E. C. Jones, and W. E. Haensch, "High mobility p-channel germanium MOSFETs with a thin Ge oxynitride gate dielectric," in *IEDM Tech. Dig.*, pp. 441-444, 2002.

75. H. Shang, J. O. Chu, S. Bedell, E. P. Gusev, P. Jamison, Y. Zhang, J. A. Ott, M. Copel, D. Sadana, K. W. Guarini, and M. Leong, "Selectively formed high mobility strained Ge PMOSFETs for high performance CMOS," in *IEDM Tech. Dig.*, pp. 157-160, 2004.

76. C. H. Huang, M. Y. Yang, Albert Chin, W. J. Chen, C. X. Zhu, B. J. Cho, M.-F. Li, and D. L. Kwong, "Very low defects and high performance Ge-on-insulator p-MOSFETs with Al_2O_3 gate dielectrics," in *Symp. VLSI Tech. Dig.*, pp. 119-120, 2003.

77. N. Wu, Q. Zhang, C. Zhu, D. S. H. Chan, M.-F. Li, N. Balasubramanian, A. Chin, and D. L. Kwong, "Alternative surface passivation on germanium for metal-oxide-semiconductor applications with high-k gate dielectric," *Appl. Phys. Lett.* vol. 85, pp. 4127-4129, 2004.

Chapter 2

Developments in Metal Gate Materials for CMOS Technology

2.1 Limitations of Poly-Si Gate Electrode

Before discussing the metal gate technology, we first briefly review the advantages and limitations of the conventional polycrystalline-Si (poly-Si) gate electrode in today's CMOS technology.

Poly-Si has been used as the gate electrode material in MOSFETs since 1970s. One advantage of the poly-Si gate is its excellent thermal stability with SiO_2, which enables the self-aligned gate-first process to be implemented in IC manufacturing. Another inherent advantage of poly-Si is the easy adjustment of its work function by dopant implantation, which enables the dual work function poly-Si gate process to be realized. These merits make the poly-Si a superior gate electrode material for the gate-first CMOS process.

However, as the transistors scales into the high-performance (HP)-45-nm technology node and beyond, some fundamental limits of poly-Si become more and more serious and tends to retard the advancing of CMOS performance, as will be discussed below.

2.1.1 Poly-Si Depletion Effect

The poly-Si depletion effect happens when a MOSFET is working in the inversion region [1]. Fig. 2.1 illustrates the poly-Si depletion effect in a NMOS device with n^+ poly-Si gate. When a positive bias is applied on the n^+ poly-Si gate, a depletion layer with a finite thickness will be formed in the poly-Si gate side near the poly-Si/oxide interface, leading to a non-negligible band bending and voltage drop in the poly-Si gate. As a result, the actual voltage applied on the gate capacitor becomes smaller, such that less inversion charges will be formed in the channel side. This causes a loss of gate control and a reduction of MOSFET drive current. The result is equivalent to introducing a capacitor C_{poly} in series with C_{ox}.

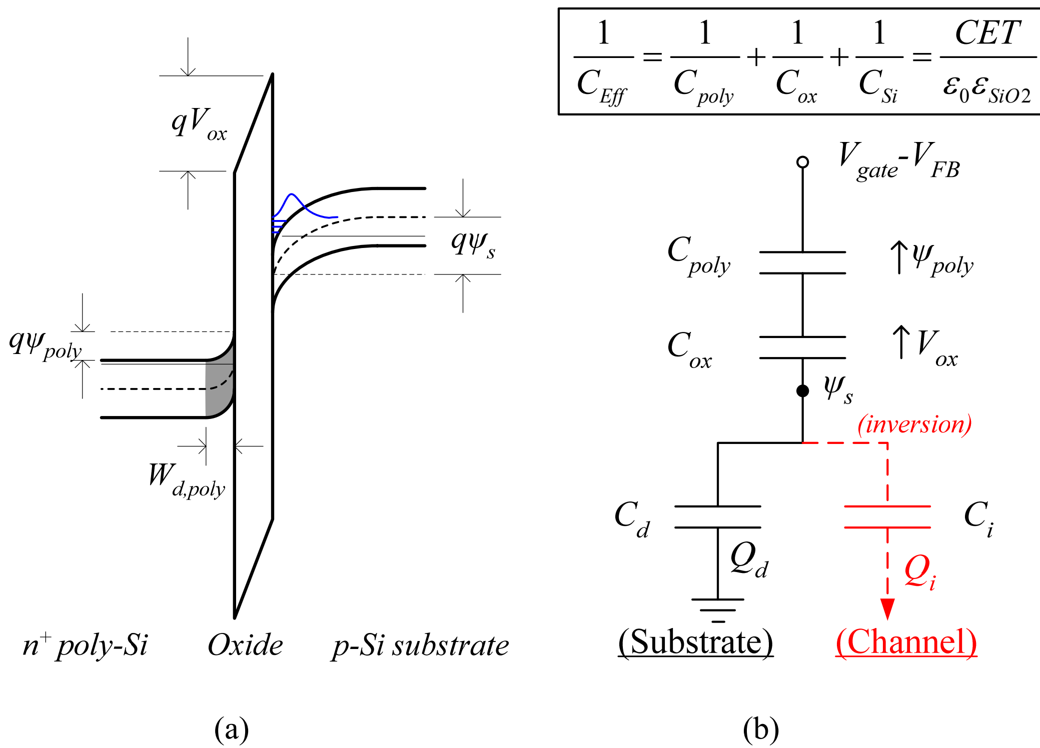


Fig. 2.1 (a) The energy band diagram of an NMOS device showing the poly-Si gate depletion effect; (b) Equivalent circuit for the gate stack of MOSFET. C_{Eff} denotes the total gate capacitance which determines the inversion charge density Q_i in the channel, C_{poly}, C_{ox}, C_{Si} represent the capacitance from the poly-depletion, gate oxide, and substrate, respectively. C_{Si} is further broken up into a depletion-layer capacitance C_d and inversion-layer capacitance C_i. *CET* represents the capacitance equivalent thickness of the MOS gate stack, and ψ_s is the surface potential.

The equivalent circuit diagram is shown in Fig. 2.1 (b). The existence of the poly-Si depletion capacitance C_{Poly} is equivalent to using a capacitor with thicker gate oxide. Assuming a heavily doped poly-Si with active doping concentration of 1.5 × 10^{20} cm^{-3} (actually this could be difficult for p^+ poly-Si with boron), channel doping concentration of 5 × 10^{18} cm^{-3} [2], and inversion charge of 1 × 10^{13} cm^{-2}, the poly-Si depletion layer thickness will be around 1.2 nm, which is equivalent to 0.4-nm SiO_2 in terms of the capacitance. Obviously, this none-scalable additive component makes all other efforts in the vertical scaling of MOSFET less pronounced, since the gate oxide thickness has been reduced to about 1.2 nm in today's technology.

The equivalent thickness due to poly-depletion can be reduced by increasing the doping concentration in poly-Si, but this may aggravate the dopant penetration problem, which will be discussed in the following section. Moreover, the active dopant concentration will finally be restricted by the solid solubility of dopants in poly-Si. In order to eliminate the poly-depletion effect ultimately, the poly-Si gate needs to be replaced by metal gate electrodes in which the electron density is high enough to make the gate depletion layer negligible.

2.1.2 Dopant Penetration Effect

As discussed in Chapter 1, the vertical scaling of MOSFET gate stack requires continuous reduction of gate oxide thickness below 1 nm. On the other hand, the doping in poly-Si is becoming very heavy (more than 10^{20} cm^{-3}) in order to minimize the poly-depletion effect and reduce the sheet resistance of poly-Si. As a result, dopant penetration from the heavily doped poly-Si gate into the channel would be a severe issue, especially for boron in p^+ poly-Si gate because boron can easily segregate into SiO_2. The dopant penetration effect will affect the threshold voltage of

transistors and raise some reliability concerns. It was suggested that using high-κ dielectrics with physical thickness larger than SiO_2 may help to address this problem. However, it has been reported that dopant penetration also happens for some high-κ materials like HfO_2 [3], implying that this issue still need to be addressed for high-κ dielectrics.

One way to suppress the dopant penetration problem is to incorporate nitrogen into gate oxide. It has been shown that nitrogen in SiON [4] and high-κ HfSiON [5] dielectrics can effectively block the boron from penetrating into the channel region, which was suggested to be due to the particular Si-O-N bonding lattice formed in silicon nitride and oxynitride [6]. However, it has been found that a large amount of nitrogen near the dielectric/Si interface would degrade the channel mobility and thus the drive current [7]. Therefore, a sandwich nitrogen profile would be advantagous, and precise nitrogen profile control will be the key to optimize the SiON dielectric further [8]. But this would be more and more difficult as the gate oxide thickness scales into the sub-1-nm regime.

Therefore, using a dopant-free metal gate electrode to replace heavily doped poly-Si gate would be a potential solution to avoid the dopant penetration problem ultimately.

2.1.3 Gate Electrode Resistivity

Another important parameter for a gate electrode is the resistivity. The significance of the gate electrode resistivity is that it directly determines the gate *RC* delay of digital circuits, particularly in radio-frequency (RF) applications [9]. Typically the sheet resistance of a gate electrode needs to be kept below 5 Ω/□ according to *ITRS* roadmap. However, as the channel length L_g becomes smaller and

smaller, the gate electrode thickness (T_{gate}) should be scaled down with L_g in order to maintain roughly a constant aspect ratio ($T_{gate}/L_g \approx 2$) [2]. On the other hand, the contact silicide thickness also has to be scaled down because the maximum silicon consumption in the salicide process should not exceed half of the S/D junction depth X_j [2]. As a result, it is necessary to reduce the resistivity of the poly-Si layer steadily to meet the sheet resistance specification for the gate electrodes. These relationships are summarized in Table 2.1.

Table 2.1 Specifications for the scaling of gate electrode, derived from *ITRS*-2005 [2].

Year	***2004***	***2006***	***2008***	***2010***	***2012***
Technology Node / *DRAM* half pitch (nm)	90	70	57	45	36
Physical gate length for *MPU/ASIC* (nm)	37	28	22	18	14
Average gate electrode sheet resistance (Ω/□)	< 5	< 5	< 5	< 5	< 5
Contact silicide thickness (nm)	20	19	15	12	9
Gate electrode thickness (nm) (poly-Si or metal gate)	74	56	46	36	28
Contact silicide sheet resistance (for NiSi) (Ω/□)	7.9	8.6	10.5	13.5	17.3
Poly-Si sheet resistance (Ω/□)	13.6	11.9	9.5	7.9	7.0
Poly-Si resistivity (μΩ-cm)	100.8	66.9	43.9	28.6	19.7

The way to reduce the resistivity of poly-Si is to increase the doping concentration. However, the active doping concentration in poly-Si would be limited by the solid solubility of the dopants in Si. For the commonly used dopants such as B,

P, and As, the solid solubility at 1100°C are ~ 5×10^{20} cm^{-3}, ~ 1.2×10^{21} cm^{-3}, and ~ 2×10^{21} cm^{-3}, respectively [10]. Therefore, the minimum resistivity values for poly-Si with B, P, and As dopants would be about 268 μΩ-cm, 74 μΩ-cm, and 58 μΩ-cm, respectively. Obviously, these values are not adequate to meet the specifications summarized in Table 2.1. This suggests that the choice of gate materials need to be reconsidered and metal gates with lower resistivity than poly-Si could be useful to solve this problem.

2.1.4 Compatibility with high-*κ* dielectrics

As discussed in Chapter 1, a high-*κ* dielectric would probably be finally required to breakthrough the scaling limits of SiO_2 and SiON dielectrics in terms of gate leakage. Beside the considerations from the high-*κ* dielectrics themselves, however, the interface quality between poly-Si and some high-*κ* materials is also a big challenge for the implementation of high-*κ* with conventional n^+/p^+ poly-Si electrodes.

First and foremost, the Fermi-level pinning (FLP) phenomenon will lead to an undesirable shift of the flat-band voltage (V_{FB}) when n^+ and p^+ poly-Si gate electrodes are in contact with many Hf-based high-*κ* dielectrics. This is particularly severe for p-MOSFET ($\Delta V_{FB} \approx 0.5$-$0.6$ V) [11]-[13], and causes asymmetric high threshold voltages for n- and p-MOSFETs, making them difficult to be used for circuit design [14]. The origin of the FLP problem has been attributed to diverse effects. Dopant penetration from poly-Si gate into the dielectric [11] or the formation of HfB_2 at the interface between p-type poly-Si and HfO_2 [15] were proposed to explain the high V_{th} in p-MOSFET. However, some later studies showed that the high V_{th} in p-MOSFET had been established during the poly-Si deposition, irrespective of the dopant-type and the activation process [16]. Electron energy loss spectroscopy (EELS)

measurement also showed no evidence of HfB_2 formation at the poly-Si/HfO_2 interface [17]. C. W. Yang *et al.* suggested that the FLP effect can be attributed to the formation of acceptor- and donor-like interface states, but the origin of the interface states was not identified [14]. C. Hobbs *et al.* proposed a Hf-Si bond induced dipole theory to explain the observed FLP phenomenon at the poly-Si/HfO_2 interface [18]-[19]. According to this theory, electrons can transfer from Hf to Si and create an interface dipole layer, which "pins" the effective work function (EWF) of p^+ poly-Si at a position near the conduction band (E_c) of Si. The formation of the dipole layer at the p^+-poly-Si/HfO_2 interface was then evidenced by a potential imaging method using atomic force microscopy (AFM), and the "pinning" position was found to be about 0.4 eV below E_c [20]. Recently, K. Shiraishi *et al.* proposed another model in which the V_{FB} shift for p^+ poly-Si on Hf-based dielectric was attributed to the oxygen vacancy (Vo) promoted charge transfer across the interface [17]. In this model, the interactions between poly-Si and HfO_2 will cause oxygen transport from the ionic high-κ material into poly-Si, leaving an oxygen vacancy Vo near the poly-Si/high-κ interface and two weakly bonded electrons in the high-κ film. These electrons can easily transfer into the conduction band of Si and leave positively charged Vo^+ in the dielectric side, resulting in the formation of dipole layer across the poly-Si/high-κ interface and consequent shift of V_{FB}. However, it is difficult to explain the opposite shift of V_{FB} for n^+ and p^+ poly-Si gates using this model.

Apart from the disagreements in understanding the origin of the FLP effect, it is also a practical challenge to minimize this effect in the device fabrication process and to obtain reasonably low V_{th} for p-MOSFET. M. Koyama *et al.* demonstrated a method to reduce the FLP effect by carefully engineering the gradient of Hf in HfSiON film, where a Hf-ratio below 10% near the poly-Si/HfSiON interface is

required to achieve 0.64 V V_{FB} difference between n^+ and p^+ poly-Si [21]. But the disadvantage of this method is the over-all dielectric constant of the high-κ dielectric will be sacrificed. Another way to reduce the high V_{th} for p-FET is to cap the Hf-based high-κ by a thin AlO_x layer in the p-FET region [22]-[23]. However, this will lead to a different equivalent oxide thickness (EOT) for n-FET and p-FET, as well as immense challenges in process integration. Finally, adding Ge into the poly-Si gate or using n^+ Ge as gate electrode was proposed as a potential solution to minimize FLP due to the smaller possibility to form Vo at the Ge/high-κ interface [24], but the process integration will be a potential issue.

The thermodynamic instability between poly-Si and many high-κ materials will be another concern for the applications of poly-Si with high-κ dielectrics. Poly-Si was found to be reactive with some of the high-κ materials, e.g. ZrO_2 and HfO_2, which leads to excess leakage current and poor charge trapping properties in the high-κ films. Inserting an Al_2O_3, Si_3N_4 or amorphous-Si layer between poly-Si and HfO_2 had been demonstrated to improve the thermodynamic stability and lower the gate leakage density of poly-Si/HfO_2 stack by several orders [13], [25]-[26]. Incorporating nitrogen into the gate dielectric to form HfON or HfSiON is another approach to improve the thermal/electrical stability of poly-Si/high-κ gate stacks [27]-[29], [5].

In summary, the conventional poly-Si gate has been an ideal gate electrode material in the SiO_2 era, whereas many issues arise when the gate stack of MOSFET scales into the sub-1-nm regime. Metal gate technology is believed to be a promising technology solution because it not only eliminates the poly-Si depletion and dopant penetration effects, but also greatly reduces the gate electrode resistivity. Therefore the metal gate technology has been extensively studied in recent years.

2.2 Post Polysilicon Era: Metal Gate Technology

2.2.1 Historical Perspective of Metal Gate Electrodes

The technology evolution always advances in a cyclical way. In the early era of MOSFET fabrication, aluminium (Al) metal was selected as the gate electrode material of MOS devices until 1968. At that time, a low-thermal-budget gate-last process, in which the source/drain was formed prior to the gate deposition and patterning, was used for MOSFET fabrication with an Al gate electrode. However, this scheme makes it quite difficult to align the gate with the channel accurately, which limits the development of more complex logic circuits. A self-aligned gate-first process was then proposed by R. W. Bower in 1966. However, no one could make the idea work since the melting point of Al is only about 660°C. This is not adequate for the dopant activation in Si processing, which is typically higher than 900°C. Fortunately, F. Faggin in Fairchild® invented a silicon gate process and adopted this technology into their Fairchild 3708 chip firstly in 1968, which shows improved performance over its Al gate counterpart. From then on, the poly-Si gate technology as well as the self-aligned process has been developed into a standard process for MOSFET fabrication.

Today, poly-Si electrodes have encountered many challenges and are expected to be replaced by metal gates in the sub-50-nm regime. However, the biggest challenge is the difficulty to identify the appropriate metal(s). Many issues in material selection and process integration need to be thought out before we reenter the metal gate era.

2.2.2 Considerations for Metal Gate Candidates

2.2.2.1 Work Function Requirements

One of the most important parameters for metal gate candidates is the work function (WF) because it directly affects the threshold voltage of a MOSFET. The threshold voltage of a MOSFET is typically given by the following expression [30]:

$$V_{th} = V_{FB} + 2\phi_B + \frac{Q_d}{C_{ox}} = V_{FB} + 2\phi_B + \frac{\sqrt{4\varepsilon_{Si} q N_b \phi_B}}{C_{ox}} \tag{2-1}$$

where φ_B is the difference between the Fermi-level and the intrinsic level in Si, ε_{Si} is the permittivity of Si, Q_d is the total depletion charge in the channel region, N_b is the doping concentration of the Si substrate (for uniform channel doping), C_{ox} is the gate oxide capacitance per unit area, and V_{FB} is the flat band voltage across the MOS stack. V_{FB} can be further given by the following equation:

$$V_{FB} = \Phi_{MS} - \frac{Q_{ox}}{C_{ox}} = (\Phi_M - \Phi_S) - \frac{Q_{ox}}{C_{ox}} \tag{2-2}$$

where Φ_{MS} denotes the work function difference between the metal gate and silicon substrate, Q_{ox} represents the equivalent oxide charge density at the oxide/Si interface. It is therefore clear that V_{th} is jointly controlled by N_b and Φ_M for a given technology (for a given C_{ox} and Q_{ox}). For sub-50-nm bulk-Si devices, the optimal work function values required for NMOS and PMOS should be about 4.05~4.25 eV and 4.97~5.17 eV, respectively [31]. In other words, the Fermi level of metal gates should be within 0.2 eV from the band-edges of Si. For the fully-depleted and multi-gate devices, e.g. FDSOI or FinFET, the V_{th} will be determined by the gate work function only since the channel region is almost intrinsic; accordingly metal gates with work functions of ± 0.15 eV from the midgap position of Si will be best served for high-performance applications (± 0.1 eV for low-standby-power (LSTP) applications) [2], [32]. Recent

study shows that when the body thickness of UTBSOI devices shrinks to less than 5 nm, band-edge work function will again be required due to the carrier quantization effect [33].

Fig. 2.2 summarizes the vacuum work function values of various metal materials in nature. It shows that metals such as Ta, Hf, Zr, In, Al, and Nb could be useful for NMOS, while Ni, Ir, Re, Rh, and Pt could be used in PMOS, and W, Cr, and Mo could be possible metal gate candidates for fully-depleted or multi-gated devices. Note that these inferences are only from the vacuum work function point of view, and may not be valid or practical in a real MOS system. Actually, the effective work function of metal gates in MOSFET will be affected by many factors, such as crystallinity of metals, metal-dielectric interface, metal deposition techniques, and even the thermal process used in the device fabrication. These effects will be discussed in Chapter 3.

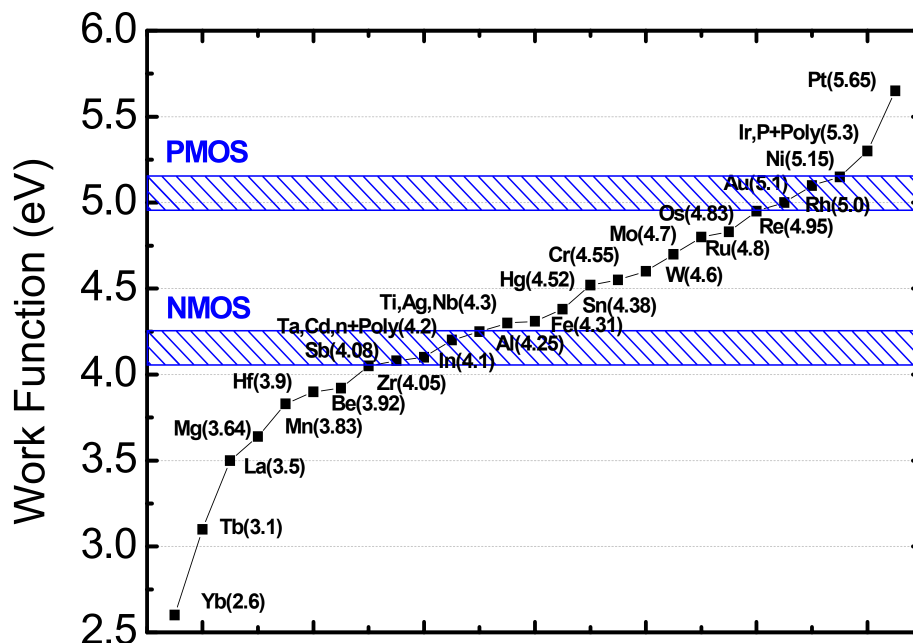


Fig. 2.2 Work function of some metal elements collected from experiments [34].

2.2.2.2 Thermal Stability Considerations

Besides the consideration for proper work function, another concern for metal gate candidates is the thermal stability in process. The thermal stability requirements for the metal candidates will vary with the different process sequences used in fabricating the MOSFET. In a gate-first process, the metal gates are formed prior to the source/drain implantation and dopant activation. This implies that the metal gate material and the metal-dielectric interface should be robust enough to sustain a high thermal budget used for the dopant activation, which is typically above 1000°C.

The most crucial thermal stability concern for the metal gates is the reaction or inter-diffusion between metal gate and the underlying gate dielectric. The interface reaction is thermodynamically driven, and likely to happen at the interface where the atoms have large differences in electronegativity and radius [35]. Once the reaction happens, a shift in the device properties with annealing conditions will be expected, in terms of gate leakage, CET, threshold voltage, gate oxide charge density and gate oxide reliability. It has been found that many metals with low work function, like Ta, Hf, Ti, etc, tend to react with the gate dielectric at high temperature, leading to a reduction of EOT and an increase in gate leakage [36]-[37]. On the other hand, some high work function metals such as Pt, Ir, and Ni tend to diffuse or penetrate through the gate oxide during the high temperature annealing process. The penetration of metals may introduce serious degradations to the gate oxide integrity (GOI), therefore it needs to be carefully managed.

In addition to the metal reaction/diffusion, some other thermal stability issues such as microstructure change, stress generation, and oxygen penetration at high-temperature should also be carefully avoided. Phase change or grain growth may affect the work function of metal gates and roughen the gate-dielectric interface,

leading to a change of V_{th} and channel mobility upon annealing [38]-[39]. From this point of view, an amorphous metal gate material would be desirable. The process related stress generation is another concern for some metal gate candidates. Due to the extreme high temperature and the ultra-fast temperature ramp up/down rate (200-250°C/s in spike annealing case) used in a state-of-the-art CMOS manufacturing process, the thermal induced stress would be a serious problem for metal gates, especially for those which have very different expansion properties with Si and the underlying dielectric. The stress in the MOS stack will result in some adherence problems, and even cause the metal film to crack or peel after annealing. In the light of this, it is necessary to choose a gate material with a thermal expansion coefficient close to that of bulk Si. Moreover, some metals also show poor barrier properties with respect to the diffusion of oxygen at high temperature [40]. This can lead to growth of an interfacial layer under the high-κ dielectric due to the penetration of oxygen residues or moisture from the gas ambient during annealing. This effect makes it challenging to scale the EOT down to the sub-1-nm regime.

2.2.2.3 Process Integration Issues

Process integration is another tough challenge for the implementation of metal gate electrode in CMOS fabrication. The considerations for the process integration are closely tied up with the selection of metal gate candidates, which forms a negative feedback loop until the best solution arrives.

The process integration issues include the deposition techniques, etching and post-etching cleaning issues, and the dual metal gate integration process. For the dual metal gate integration, a simple, reliable, and cost-effective scheme to integrate dual metal gates on the same wafer would be desirable. Basically there are three groups of

integration schemes: gate-first, gate-last, and fully-silicided (FUSI) gate processes. The gate-first process is a relatively cost-effective and reliable process, but it requires excellent thermal stability for both n^+ and p^+ metal gate candidates, as well as refined metal etching and post-etching cleaning techniques [41]-[42]. The gate-last process has the lowest thermal budget requirement and hence imposes fewer demands on the selection of metal gate materials, but the process is relatively complex [43]-[44]. The FUSI gate process is most compatible with the conventional poly-Si CMOS technology [45]-[46]. More details on the dual metal gate integration schemes will be discussed in Chapter 5.

The metal gate deposition techniques can affect the properties of metal gate electrode in many aspects, such as film morphology, resistivity, work function, thermal stability, and even the gate stack reliability [47]-[50]. The most commonly used method to deposit metal gate is the physical vapor deposition (PVD) technique, including sputtering, evaporation, etc [51]. One advantage of the sputtering method is that it allows easy control of the film composition in metal alloys, metal nitrides, metal carbides, etc. However, the charged ions used in the sputtering process may induce plasma damage problems [52]-[53]. The evaporation method can avoid the plasma damage problem, but it is not as flexible as sputtering in modifying the composition of metal gate candidates. For all the PVD deposition techniques, a fundamental limit is the step coverage issue in high aspect ratio structures [51], which may limit the applications of PVD techniques in 3-D device structures (FinFET, Ω-gate, nanowire, etc). However, this problem can be addressed by using chemical vapor deposition (CVD) methods. The CVD methods not only have the advantages such as good step coverage and low damage to the dielectric, but they provide a number of variables, viz. temperature, pressure and gas flow which could be useful to

control the film microstructure [51]. Among many kinds of CVD techniques, metal-organic chemical vapor deposition (MOCVD) and atomic-layer chemical vapor deposition (ALCVD) are two most important forms. MOCVD have the merits of high deposition rate, good uniformity control, etc. However, it will introduce some impurities like carbon into the films due to the use of organic precursors. Recently, ALCVD (or ALD) has drawn considerable attention and has been utilized to deposit many high-κ and metal gate films. The major advantage of the ALD technique is that it provides the ability to control the atoms layer by layer such that the film concentration and even the interface chemistry can be well engineered. This advantage makes ALD a powerful tool in scientific research. However, one concern for the ALD technique is its relative long lead time and hence low throughput due to the need to purge the ambient in every deposition cycle.

Etching and post-etching cleaning of metal gate electrodes is another practical issue for the integration of metal gate in CMOS process. To achieve high selectivity, vertical profile, and small feature size in the metal gate etching process, the selection of masks (photoresist or hard mask), etchant and the post-etching cleaning process need to be optimized systematically according to the properties of the specific metal gate materials. These practical issues can even affect the selection of metal gate candidates.

2.2.2.4 Co-optimization of Metal Gate/High-κ Gate Stack

According to *ITRS* 2005, it is most likely that metal gate technology will be introduced into production together with high-κ dielectric. Therefore the co-optimization of metal gate/high-κ gate stack would be of significance.

The first issue that needs to be considered is the effective work function of metal gate on high-*κ* dielectric. It has been found that the effective work functions of some metal gates measured on high-*κ* dielectrics deviate from the work function values measured in vacuum and exhibits large dependence on the underlying dielectric, which makes the work function tuning and threshold voltage adjustment on high-*κ* dielectric more complicated. The dependence of the metal work function on the permittivity of the gate dielectric has been explained by Yeo *et al* using a metal-induced gap state (MIGS) theory [54]. Recently, our study indicates that the extrinsic states relating to the interface chemistry and the thermal process will also affect the effective work function of metal gates [55]-[56]. This topic will be further discussed in detail in Chapter 3.

The impact of the metal gate to the carrier mobility in MOSFET has also drawn considerable attention recently. It has been observed that the electron mobility in TiN/high-*κ* stacks is higher than that of poly-Si/high-*κ* gate stacks, and this was attributed to the screening of remote phonon scattering by metal gate [57]. Some recent research by Akasaka *et al.* reveals that the chemistry of the metal gate electrode can also affect the electron mobility [58]. Since mobility degradation is one of the most crucial issues for high-*κ* gate dielectrics, co-optimization of metal/high-*κ* stack to achieve mobility close to that of the poly-Si/SiO_2 stacks will become an indispensable criterion in evaluating potential metal gate electrodes.

2.2.3 Research Status of Metal Gate Technology

According to the different stratagems used to achieve dual work function for CMOS applications (referring to Chapter 5 for more details), metal gate candidates can be grouped into several categories, such as direct metal gates, binary metal alloys,

and FUSI metal gates. Here the direct metal gates refer to those metal candidates whose work function is directly determined by the metal material itself after deposition, without a following process step to modify its work function intentionally. These metal gate candidates are typically used in a gate-first process. For the binary (or ternary) metal alloys, the work function are not directly determined by the as-deposited layered metal stacks; instead, a subsequential annealing process will be carried out to form new alloys and achieve the final work function. Depending on the thermal budgets of the alloying process, these metal gates can either be used in a gate-first integration process or gate-last process. As for FUSI metal gates, the work function is determined after the silicidation of poly-Si gate. For all these metal gate solutions, remarkable progress has been achieved in recent research.

2.2.3.1 Direct Metal Gates

The direct metal gates include some pure metals, metal nitrides, metal carbides and conductive metal oxides. The development of these direct metal gates is mainly targeted at applications using a gate-first integration scheme, therefore the thermal stability is a pre-requisite. From this point of view, metal nitrides and metal carbides, such as TaN, TiN, HfN, WN, and TaC, would be of priority due to their excellent thermal stability compared with many pure metal materials [40],[50]. In addition, these metal gate candidates also tend to form a stable interface with high-κ dielectric, rendering excellent EOT scalability to be achievable. To date, most of the reported metal gate/high-κ stacks fabricated using the gate-first approach with sub-1-nm EOT are based on the metal nitride gate electrodes [40],[50].

However, the work functions of most metal nitride materials are close to the mid-gap position of Si. In order to modulate the work function of metal nitride

materials towards the band-edge of Si, several approaches have been explored. Adding Si [59]-[60] and Al [61] into metal nitrides to form ternary compound have been demonstrated to be useful for NMOS and PMOS, respectively, but the work function tuning range is still not ideal. In our work, we investigated a novel approach to incorporate lanthanide into metal nitrides to reduce the work function of TaN or HfN, which shows promising characteristics for NMOS (to be discussed in Chapter 4). Besides these methods, another way to tune the work function of metal nitride materials is to modify the nitrogen concentration, which can be achieved either by tuning the N_2 flow rate during deposition [62] or by a post-deposition N implantation process [63]-[64]. The later approach is quite attractive from the process integration point of view. However, the work function can only be tuned in the range of 4.5 ~ 4.95 eV, which is not adequate for bulk-Si CMOS [65].

In terms of the process integration of direct metal gates for dual metal gate application, the most commonly studied approach is to selectively remove the first-deposited metal layer in one side (NMOS or PMOS), followed by the deposition of the second metal material. Many metal gate combinations have been demonstrated using this approach, such as Ti/Mo, TiN/TaSiN, TaSiN/Ru, etc. [66]-[67], [41]. Note that the etching of the first-deposited metal layer from the underlying gate dielectric may introduce some reliability concern if the process is not optimized. In our work, we proposed a new integration scheme which can avoid the etching of metal film from the dielectric and hence avoid the exposure of the gate dielectric during process. This novel process will be discussed in Chapter. 5.

2.2.3.2 Binary Metal Alloys

Binary metal alloys are another group of metal gate candidates which have been studied from the very beginning of metal gate research. Polishchuk *et al*. first proposed a Ti-Ni inter-diffusion process to achieve dual work function values for n-MOSFET and p-MOSFET [68]. The advantage of this approach is that there is no need to expose the dielectric to any etching process. Instead, the work function tuning is realized by a metal inter-diffusion process. This approach had been further studied by many researchers using other material combinations, such as Ru-Ta alloy [69], Ta-Pt alloy [70], Ti-Pt alloy [71], and Hf-Mo alloy [72] etc. Wide range work function modulation (more than 1 eV) and precise work function control could be achieved using this approach. However, an inherent concern of this method is the thermal stability, which is typically limited by the low work function metals used in the alloys. As a result, the alloying temperature is typically around 400°C to 600°C. This greatly limits the applications of these metal candidates in the gate-first CMOS process. However, they are still promising metal gate solutions for a gate-last integration process where only the thermal budget for back-end of line (BEoL) processes is required.

2.2.3.3 Fully-Silicided (FUSI) Metal Gates

All the metal gate solutions mentioned above need notable modification to the existing poly-Si based CMOS process, which implies high cost and high risk for the technology migration. For the semiconductor industry, a simple approach with little modification to the current CMOS process flow would be more attractive in the near-term. Therefore, the fully-silicided metal gate technology has drawn considerable attention and steady progresses have been achieved in recent year. Tavel *et al.* first

demonstrated a FUSI $CoSi_2$ process with low sheet resistance (~2Ω/□) [45]. However, the work function of $CoSi_2$ is about 4.6 eV, rendering the V_{th} of bulk-Si MOSFET too high to be acceptable. Min *et al.* reported that the work function of FUSI NiSi can be modulated from about 4.6 eV to 5.0 eV by the dopants in poly-Si [46], which makes the FUSI gate process more advantageous than any other dual metal gate solutions. The silicidation induced impurity segregation (SIIS) and pile-up at the silicide-dielectric interface was believed to be the main mechanism for the work function modulation in NiSi [73]-[75]. Various impurities have been investigated in tuning the work function of NiSi, as summarized in Fig. 2.3 [76]. Applications of mid-gap NiSi in FinFET were also demonstrated successfully with good electrical characteristics [77].

However, the work function tuning of NiSi through the dopant segregation approach seems no longer efficient on high-κ dielectric [78]-[79]. This was again attributed to the Fermi-level pinning effect induced by Hf-Si bonds at the NiSi/high-κ interface, similar with that in poly-Si/high-κ interface [79]. In order to solve this problem, phase-controlled (PC) silicide technology was developed [78]. The Ni_3Si phase seems to be useful for p-MOSFET with the high work function of 4.8 eV on HfSiON dielectric, while $NiSi_2$ phase would be used for n-MOSFET with the low work function of 4.4 eV [78]. However, the work function tuning range (< 0.4 eV) is still not large enough for planar bulk-Si devices. Therefore, the PC-silicide process will be potentially useful only for those novel 3-D devices featuring intrinsic channel and high-κ dielectric. In addition, controlling the phases of silicide in process is a challenging task, particularly for MOSFETs with a very small gate length [80].

In order to achieve a larger work function window in FUSI metal gates and address the dopant fluctuation problem associated with the SIIS approach, some

ternary silicides were proposed. The basic idea is to introduce the third element into the $NiSi_x$ system to (i) reduce the number of Hf-Si bonds at $NiSi_x$/high-κ interface and (ii) set the work function. Many material systems have been studied, such as $Ni_{1-x}Yb_xSi_y$, $Ni_{1-x}Ta_xSi_y$, or $Ni_{1-x}Al_xSi_y$ for NMOS and $Ni_{1-x}Pt_xSi_y$ for PMOS [81]-[83], as summarized in Fig. 2.3.

In some extreme cases such as substituted Al (SA) gate or partial silicided Pt_xSi (x >10) processes, the silicon is almost replaced by metals. The work function will thus be determined by that of the metals at the metal-dielectric interface [84]-[85].

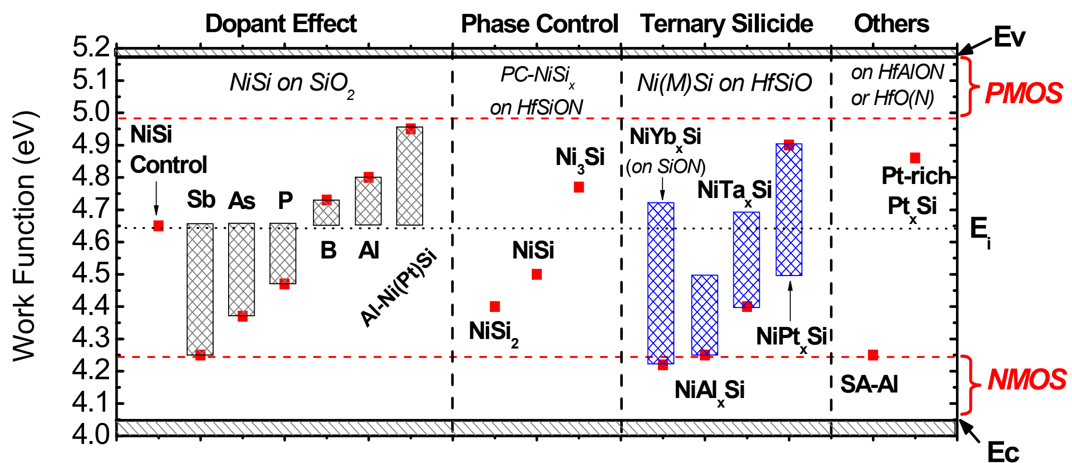


Fig. 2.3 Work function modulation by various mechanisms in some NiSi-based-silicide metal gates; data from [76], [78], [81]-[85].

In summary, though many successes have been achieved in developing potential metal gate materials for next generation transistors, the right materials and the appropriate integration flows have not been identified yet. Fundamental understanding on the work function tuning mechanisms is still needed, as well as the solutions which can reach every aspects of the matter.

2.3 Challenges in Metal Gate Technology

2.3.1 Understanding of the Metal-Dielectric Interface

The interface between the gate electrode and gate dielectric is of importance since it determines the effective work function of the metal gate electrode as well as many other characteristics of the MOSFET. However, the knowledge on this crucial interface is still insufficient to guide us in developing the metal-gated MOSFET. The impacts of many factors to the properties of metal-dielectric interface still need to be identified, including the metal crystallinity, interface chemistry, dangling bonds and defects, oxygen vacancy in dielectrics, interface reactions, metal and dopant penetrations, and so on. Simple models depicting the general behavior of this interface will be appreciated.

Among these issues, one of the most critical challenges is to understand the role of the metal-dielectric interface in determining the effective work function of metals. Although Yeo *et al.* proposed the interface dipole theory to explain the dependence of the effective work function of metal gates on the underlying dielectric, it is still not clear whether it can be used to explain the process dependence of metal work function during high-temperature annealing, which finally determines the V_{th} of the MOSFET. Therefore, the factors contributing to the thermal instability of metal work functions need to be identified and the mechanisms need to be understood. Moreover, engineering of the material system at the metal-dielectric interface is also of technical importance for work function tuning of metal gates.

2.3.2 Developing Appropriate Metal Gate Materials

As we discussed above, there are many criteria in the material selection and process integration of novel metal gate candidates. The materials fulfilling all the requirements have not been identified yet, and novel metal gate candidates with promising properties are still expected.

For the direct metal gate candidates, one of the biggest difficulties is how to get the best trade-off between the thermal stability concern and the low work function requirement for NMOS applications. Developing a thermally stable N-type metal gate candidate and finding a simple approach to integrate it into a CMOS process will be of significance from this point of view. For P-type metal candidates, the thermal stability issue will be less pronounced. However, co-optimization of P-type metal gates with high-κ dielectrics will be a challenging issue because it has been observed that the work function of many P-type materials will decrease significantly on high-κ dielectrics, probably due to the oxygen vacancies induced interface dipoles [86]. A proper method to minimize the oxygen vacancies and obtain a high work function above 5.0 eV would be desired.

2.3.3 Dual Metal Gate Integration Issues

Besides the difficulties in developing the individual metal gate materials for both n- and p-MOSFETs, the integration of dual metal gate in a CMOS process flow is always a challenge for metal gate technology. Many factors need to be considered when integrating dual metal gates, such as process complexity, cost, GOI degradation, process controllability, EOT scalability, and so on. These considerations will in turn affect the selection of metal materials.

For the gate-first CMOS integration flow, one of the major concerns for the conventional direct-etching integration scheme is the exposure of gate dielectrics to the etching environment during the metal selective etching process, as we discussed above in Section 2.2.3.1. A proper integration flow to avoid this concern is needed. For the gate-last process, there have been some solutions for the dual metal gate integration process, thanks to the low thermal budget used. However, the co-optimization of the metal-gate/high-κ stacks should be emphasized in this case, regarding the EOT scalability, carrier mobility, and device reliability. For the FUSI metal gate process, integration of different silicide materials and the precise control of the dopants, phases, and stress in the silicidation process will still be the key issues.

2.4 Research Scope and Major Achievements in This Thesis

The overall objective of this work is to: (i) study the role of the metal-dielectric interface in determining the effective work function of metal gate electrodes, and (ii) develop appropriate metal materials as well as proper integration schemes for the implementation of metal gates in next generation transistors.

For the metal-dielectric interface study, this work will be focused on understanding the impact of dielectric types and different thermal treatments to the effective work functions of metal gates. The behavior of metal effective work function as a function of annealing temperature was investigated in several combinations of metals and dielectrics. It was observed that the work function thermal instability problem is more serious in metal/SiO_2 systems than in metal/HfO_2 stacks. A metal-dielectric interface model that takes the role of *extrinsic states* into account was proposed to qualitatively explain the work function thermal instability. The generation of extrinsic states seems to be thermodynamically driven, and

becomes more pronounced as the annealing temperature increases. The interface chemistry plays an important role in determining the generation of extrinsic states. These results may provide some useful understanding of the properties of metal-dielectric interface and be of practical significance in engineering the metal-dielectric interface for future study. These contents will be discussed in Chapter 3.

Based on the above understanding, the key to modulating the metal gate work function is to engineer the metal-dielectric interface with proper materials. In Chapter 4, a novel approach to modifying the work function of metal nitride gate electrodes by incorporating lanthanide elements are investigated systematically for the first time. Thanks to the very low work function of lanthanide elements (Tb, Er, Yb, etc.), the work function of metal nitride gate electrodes (TaN, HfN, etc.) can be tuned gradually from mid-gap position down to ~ 4.2 eV by incorporating lanthanide. Good thermal stability can also be achieved simultaneously due to the very low lanthanide concentration and the enhanced nitrogen levels in these materials. These results suggest that lanthanide-incorporated metal nitrides could be a promising metal gate candidate for n-MOSFETs.

Moreover, in order to address some of the challenges associated with the existing dual metal gate integration processes, several potential dual metal gate integration flows will be proposed and discussed in Chapter 5. The first one is a gate-first integration process using a novel high-temperature metal intermixing technique to achieve dual work function. Unlike the conventional dual metal gate integration scheme where selective etching of the metal material from the gate dielectric is required, the new integration scheme avoids the exposure of the gate dielectric during the metal etching process by using an ultra-thin TaN buffer layer on top of the gate dielectric. The work function of the TaN buffer layer can then be modulated by the

intermixing of TaN with other metals in a high-temperature annealing process, which is compatible with the source/drain dopant activation process in the gate-first CMOS integration flow. The second integration scheme discussed in Chapter 5 is a gate-last process utilizing HfN as the dummy gate material. The aim of this approach is to achieve large work function difference and optimal high-κ properties simultaneously for the integration of dual metal gates with high-κ materials. The use of the HfN dummy gate enables high quality high-κ dielectric with EOT of less than 1 nm to be achieved, as well as the work function difference of ~ 0.8 eV by using Ta and Ni as the electrodes for NMOS and PMOS, respectively. These attempts could be of practical values for the community in developing the dual metal gate solutions for future CMOS technology. Note that the intention in this part of study is not to solve all the potential issues from the material selection to the process integration in nanometer level, but to propose the possible technology approaches and discuss their advantages and limitations.

References

1. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 2, pp. 74-78.

2. *International Technology Roadmap of Semiconductors (ITRS)*, 2005 edition, Semiconductor Industry Association (SIA), SanJose, CA 95110. Available: http://public.itrs.net/

3. K. Onishi, L. Kang, R. Choi, E. Dharmarajan, S. Gopalan, Y. Jeon, C. S. Kang, B. H. Lee, R. Nieh, and J. C. Lee, "Dopant penetration effects on polysilicon gate HfO_2 MOSFET's," in *Symp. VLSI Tech. Dig.*, pp. 131-132, 2001.

4. T. Morimoto, H. S. Momose, Y. Ozawa, K. Yamabe, and H. Iwai, "Effect of boron penetration and resultant limitations in ultra thin pure-oxide and nitrided-oxide gate flims," in *IEDM Tech. Dig.*, pp. 429-432, 1990.

5. M. Koyama, A. Kaneko, T. Ino, M. Koike, Y. Kamata, R. Iijima, Y. Kamimuta, A. Takashima, M. Suzuki, C. Hongo, S. Inumiya, M. Takayanagi, and A. Nishiyama, "Effects of nitrogen in HfSiON gate dielectric on the electrical and thermal characteristics," in *IEDM Tech. Dig.*, pp. 849-852, 2002.

6. K. A. Ellis and R. A. Buhrman, "Time-dependent diffusivity of boron in silicon oxide and oxynitride," *Appl. Phys. Lett.* vol. 74, pp. 967-969, 1999.

7. H. Yang and G. Lucovsky**, "**Integration of ultrathin (1.6~2.0 nm) RPECVD oxynitride gate dielectrics into dual poly-Si gate submicron CMOSFETs," in *IEDM Tech. Dig.*, pp. 245-248, 1999.

8. D. Ishikawa, S. Sakai, K. Katsuyama, and A. Hiraiwa, "Nitride-sandwiched-oxide gate insulator for low power CMOS," in *IEDM Tech. Dig.*, pp. 869-872, 2002.

9. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 5, pp. 247-250.

10. S. M. Sze, *Physics of Semiconductor Devices*, 2nd Edition, John Wiley & Son, 1981, ch. 2, p. 69.

11. L. Kang, K. Onishi, Y. Jeon, B. H. Lee, C. Kang, W. Qi, R. Nieh, S. Gopalan, R. Choi, and J. C. Lee, "MOSFET Devices with polysilicon on single-layer HfO_2 high-K dielectrics," in *IEDM Tech. Dig.*, pp. 35-38, 2000.

12. S.Pidin, Y.Morisaki, Y.Sugita, T.Aoyama, K.Irino, T.Nakamura, and T.Sugii, "Low standby power CMOS with HfO_2 gate oxide for 100-nm generation," in *Symp. VLSI Tech. Dig.*, pp. 28-29, 2002.

13. V. S. Kaushik, E. Rohr, S. De Gendt, A. Delabie, S. Van. Elshocht, M. Claes, X. Shi, Y. Shimamoto, L.-A. Ragnarsson, T. Witters, Y. Manabe, M. Heyns,

"Effects of interactions between HfO_2 and poly-Si on MOSCAP and MOSFET electrical behavior," *Proc. Int. Workshop Gate Insulator (IWGI)*, pp. 62-63, 2003.

14. C. W. Yang, Y. K. Fang, C. H. Chen, S. F. Chen, C. Y. Lin, and C. S. Lin, M. F. Wang, Y. M. Lin, T. H. Hou, C. H. Chen, L. G. Yao, S. C. Chen, and M. S. Liang, "Effect of polycrystalline-silicon gate types on the opposite flatband voltage shift in n-type and p-type metal–oxide–semiconductor field-effect transistors for high-k HfO2 dielectric," *Appl. Phys. Lett.* vol. 83, pp. 308-310, 2003.

15. T. Aoyama *et al.*, *Int. Workshop on Gate Insulator (IWGI)*, pp. 175-176, 2003.

16. E. Cartier, V. Narayanan, E. P. Gusev, P. Jamison, B. Linder, M. Steen, K. K. Chan, M. Frank, N. Bojarczuk, M. Copel, S. A. Cohen, S. Zafar, A. Callegari, M. Gribelyuk*, M. P. Chudzik, C. Cabral Jr., R. Carruthers , C. D'Emic, J. Newbury, D. Lacey, S. Guha and R. Jammy, "Systematic study of pFET Vt with Hf-based gate stacks with poly-Si and FUSI gates," in *Symp. VLSI Tech. Dig.*, pp. 44-45, 2004.

17. K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Kohno, T. Chikyo, H. Kitajima, and T. Arikado, "Physics in Fermi level pinning at the polySi/Hf-based high-k oxide interface," in *Symp. VLSI Tech. Dig.*, pp. 108-109, 2004.

18. C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, and P. Tobin, "Fermi level pinning at the polySi/metal oxide interface," in *Symp. VLSI Tech. Dig.*, pp. 9-10, 2003.

19. C. Hobbs, L. Fonseca, A. Knizhnik, V. Dhandapani, S. Samavedam, W. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, M. Lovejoy, R. Rai, E. Hebert, H. Tseng, S. Anderson, B. White, and P. Tobin, "Fermi-level pinning at the polysilicon/metal oxide interface," *IEEE Tran. Electron. Dev.*, vol. 51, pp. 971-984, 2004.

20. R. Ludeke, V. Narayanan, E. P. Gusev, E. Cartier, and S. J. Chey, "Potential imaging of Si/HfO_2/polycrystalline silicon gate stacks: Evidence for an oxide dipole," *Appl. Phys. Lett.*, vol. 86, 122901, 2005.

21. M. Koyama, Y. Kamimuta, T. Ino, A. Kaneko, S. Inumiya, K. Eguchi, M. Takayanagi, and A. Nishiyama, "Careful examination on the asymmetric V_{fb} shift problem for poly-Si/HfSiON gate stack and its solution by the Hf concentration control in the dielectric near the poly-Si interface with small EOT expense," in *IEDM Tech. Dig.*, pp. 499-502, 2004.

22. H.-S. Jung, J.-H. Lee, S. K. Han, Y.-S. Kim, H. J. Lim, M. J. Kim, S. J. Doh, M. Y. Yu, N.-I. Lee, H.-L. Lee1, T.-S. Jeon, H.-J. Cho, S. B. Kang, S. Y. Kim, I. S. Park, D. Kim, H. S. Baik, Y. S. Chung, "A highly manufacturable MIPS (metal inserted poly-Si stack) technology with novel threshold voltage control," in *Symp. VLSI Tech. Dig.*, pp. 232-233, 2005.

23. H.-J. Li and M. Gardner, "Dual high-k gate dielectric with poly gate electrode: HfSiON on nMOS and Al_2O_3 capping layer on pMOS," *IEEE Electron Device Lett.*, vol. 26, pp. 441-443, 2005.

24. H. Takeuchi, H. Y. Wong, D. Ha, and T. King, "Impact of oxygen vacancies on high-k gate stack engineering," in *IEDM Tech. Dig.*, pp. 829-832, 2004.

25. D. C. Gilmer, R. Hegde, R. Cotton, R. Garcia, V. Dhandapani, D. Triyoso, D. Roan, A. Franke, R. Rai, L. Prabhu, C. Hobbs, J. M. Grant, L. La, S. Samavedam, B. Taylor, H. Tseng, and P. Tobin, "Compatibility of polycrystalline silicon gate deposition with HfO_2 and Al_2O_3/HfO_2 gate dielectrics," *Appl. Phys. Lett.*, vol. 81, pp. 1288-1290, 2002.

26. Y. Morisaki, T. Aoyama, Y. Sugita, K. Irino, T. Sugii, and T. Nakamura, "Ultra-thin (Teffinv=1.7 nm) poly-Si-gated $SiN/HfO_2/SiON$ high-k stack dielectrics with high thermal stability (1050°C)," in *IEDM. Tech. Dig.*, pp. 861-864, 2002.

27. R. M. Wallace, R. A. Stoltz, G. D. Wilk, "Zirconium and/or hafnium oxynitride gate dielectric," *US Patent* 6,013,553, Jan, 2000.

28. R. M. Wallace, R. A. Stoltz, G. D. Wilk, "Zirconium and/or hafnium silicon-oxynitride gate dielectric," *US Patent* 6,020,243, Feb, 2000.

29. C. H. Choi, S. J. Rhee, T. S. Jeon, N. Lu, J. H. Sim, R. Clark, M. Niwa, and D. L. Kwong, "Thermally stable CVD HfO_xN_y advanced gate dielectrics with poly-Si gate electrode," in *IEDM Tech. Dig.*, pp. 857-860, 2002.

30. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, U.K., Cambridge Univ. Press, 1998, ch. 3, pp. 118-119.

31. I. De, D. Johri, A. Srivastava, and C. M. Osburn, "Impact of gate workfunction on device performance at the 50 nm technology node," *Solid State Electronics*, vol. 44, no. 6, pp. 1077-1080, Jun. 2000.

32. B. Cheng, B. Maiti, S. Samavedam, J. Grant, B. Taylor, P. Tobin, and J. Mogab, "Metal gates for advanced sub-80-nm SOI CMOS technology," in *IEEE Intl. SOI Symp. Proc.*, pp. 91-92, 2001.

33. T. Low, M. F. Li, W. J. Fan, S. T. Ng, Y. Yeo, C. Zhu, A. Chin, L. Chan, and D. L. Kwong, "Impact of surface roughness on silicon and germanium ultra-thin-body MOSFETs," in *IEDM Tech. Dig.*, pp. 151-154, 2004.

34. H. B. Michaelson, "The work function of the elements and its periodicity," *J. Appl. Phys.,* vol. 48, pp. 4729-4733, 1977.

35. J. K. Schaeffer, S. B. Samavedam, D. C. Gilmer, V. Dhandapani, P. J. Tobin, J. Mogab, B.-Y. Nguyen, B. E. White, Jr., S. Dakshina-Murthy, R. S. Rai, Z.-X. Jiang, R. Martin, M. V. Raymond, M. Zavala, L. B. La, J. A. Smith, R. Garcia, D. Roan, M. Kottke, and R. B. Gregory, "Physical and electrical properties of metal

gate electrodes on HfO2 gate dielectrics," *J. Vac. Sci. Technol. B*, vol. 21, no. 1, pp. 11-17, Jan./Feb. 2003.

36. V. Misra, G. P. Heuss, and H. Zhong, "Use of metal-oxide-semiconductor capacitors to detect interactions of Hf and Zr gate electrodes with SiO_2 and ZrO2," *Appl. Phys. Lett.*, vol. 78, no. 26, pp. 4166-4168, 2001.

37. Takeo Ushiki, Kunihiro Kawai, Ichiro Ohshima, and Tadahiro Ohmi, "Chemical reaction concerns of gate metal with gate dielectric in Ta gate MOS devices: an effect of self-sealing barrier configuration interposed between Ta and SiO_2," *IEEE Trans. Electron Devices*, vol. 47, no.11, pp. 2201-2207, Nov. 2000.

38. S. H. Bae, W. P. Bai, H. C. Wen, S. Mathew, L. K. Bera, N. Balasubramanian, N. Yamada, M. F. Li, and D. L. Kwong, "Laminated metal gate electrode with tunable work function for advanced CMOS," in *Symp. VLSI Tech. Dig.*, pp. 188-189, 2004.

39. M.-F. Wang, T.-Y. Huang, Y.-C. Kao, H.-C. Lin, and C.-Y. Chang, "Impact of thermal stability on the characteristics of complementary metal oxide semiconductor transistors with TiN metal gate," *Jpn. J. Appl. Phys.*, vol. 41, pp. 546-551, 2002.

40. H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M.-F. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du, and D.-L. Kwong, "Thermally robust high quality HfN/HfO2 gate stack for advanced CMOS devices," in *IEDM Tech. Dig.*, pp. 99-103, 2003.

41. S. B. Samavedam, L. B. La, J. Smith, S. Dakshina-Murthy, E. Luckowski, J. Schaeffer, M. Zavala, R. Martin, V. Dhandapani, D. Triyoso, H. H. Tseng, P. J. Tobin, D. C. Gilmer, C. Hobbs, W. J. Taylor, J. M. Grant, R. I. Hegde, J. Mogab, C. Thomas, P. Abramowitz, M. Moosa, J. Conner, J. Jiang, V. Arunachalam, M. Sadd, B.-Y. Nguyen, and B. White, "Dual-metal gate CMOS with HfO_2 gate dielectric," in *IEDM Tech. Dig.*, pp. 433-436,2002.

42. Z. B. Zhang, S. C. Song, C. Huffman, J. Barnett, N. Moumen, H. Alshareef, P. Majhi, M. Hussain, M.S. Akbar, J.H. Sim, S.H. Bae, B. Sassman, and B.H. Lee, "Integration of dual metal gate CMOS with TaSiN (NMOS) and Ru (PMOS) gate electrodes on HfO_2 gate dielectric," in *VLSI Tech. Dig.*, pp. 50-51. 2005.

43. I. Polishchuk, P. Ranade, T.-J. King, and C. Hu, "Dual work function metal gate CMOS technology using metal interdiffusion," *IEEE Electron Device Lett.*, vol. 22, pp. 444-446, Sep. 2001.

44. C. Ren, H. Y. Yu, J. F. Kang, X. P. Wang, H. H. H. Ma, Yee-Chia Yeo, D. S. H. Chan, M.-F. Li, and D.-L. Kwong, "A dual-metal gate integration process for CMOS with sub-1 nm EOT HfO2 by using HfN replacement gate," *IEEE Electron Device Lett.*, vol. 25, pp. 580-582, Aug. 2004.

45. B. Tavel, T. Skotnicki, G. Pares, N. Carrière, M. Rivoire, F. Leverd, C. Julien, J. Torres, and R. Pantel, "Totally silicided ($CoSi_2$) polysilicon: A novel approach to

very low-resistive gate (~2Ω/□) without metal CMP nor etching," in *IEDM Tech. Dig.*, pp. 825-828, 2001.

46. M. Qin, V. M. C. Poon, and S. C. Ho, "Investigation of polycrystalline nickel silicide films as a gate material," *J. Electrochem. Soc.*, vol. 148, no. 5, pp. G271-G274, 2001.

47. Y. H. Kim, C. H. Lee, T. S. Jeon, W. P. Bai, C. H. Choi, S. J. Lee, L. Xinjian, R. Clarks, D. Roberts, and D. L. Kwong, "High quality CVD TaN gate electrode for sub-100nm MOS devices," in *IEDM Tech. Dig.*, pp. 667-670, 2001.

48. J. C. Hooker, R.J.P Lander, F.N. Cubaynes, T. Schram, F. Roozeboom, J. van Zijl, M. Maas, F.C. van den Heuvel, E.P. Naburgh, J.G.M. van Berkum, Y. Tamming, T. Dao, K. Henson, M. Schaekers, A. van Ammel, Z. Tokei, M. Demand, C. Dachs, "Tantalum-based gate electrode metals for advanced CMOS applications," *Proc. 207th ECS meeting*, 632, Quebec, Canada, 2005.

49. D.-G. Park, K.-Y. Lim, H.-J. Cho, T.-H. Cha, J.-J. Kim, J.-K. Ko, I.-S. Yeo, and J. W. Park, "Novel damage-free direct metal gate process using atomic layer deposition," in *VLSI Tech. Dig.*, pp. 65-66. 2001.

50. D.-G. Park, Z.J. Luo, N. Edleman, W. Zhu, P. Nguyen, K. Wong, C. Cabral1, P. Jamison1, B.H. Lee, A. Chou, M. Chudzik, J. Bruley, O. Gluschenkov, P. Ronsheim, A. Chakravarti, R. Mitchell, V. Ku, H. Kim1, E. Duch1, P. Kozlowski1, C.D'Emic1, V. Narayanan1, A. Steegen, R. Wise, R. Jammy1, R. Rengarajan, H. Ng, A. Sekiguchi, and C.H. Wann, "Thermally robust dual-work function ALD-MN_x MOSFETs using conventional CMOS process flow," in *Symp. VLSI Tech. Dig.*, pp. 186-187, 2004.

51. S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era*, *Volumn 1*, 2nd Edition, California, Lattice Press, 2000, ch. 11& ch. 6.

52. T. Ushiki , M.-C. Yu, K. Kawai, T. Shinohara, K. Ino , M. Morita and T. Ohmi "Gate oxide reliability concerns in gate-metal sputtering deposition process: an effect of low-energy large-mass bombardment", *Microelectronics Reliability*, vol. 39, pp. 327-332, 1999.

53. T. Yamada, M. Moriwaki, Y. Harada, S. Fujii, and K. Eriguchi, "Effects of the sputtering deposition process of metal gate electrode on the gate dielectric characteristics," *Microelectronics Reliability*, vol. 41, pp. 697-704, 2001.

54. Y.-C. Yeo, T.-J. King, and C. Hu, "Metal-dielectric band alignment and its implications for metal gate complementary metal-oxide-semiconductor technology," *J. Appl. Phys.*, vol. 92, no. 12, pp. 7266-7271, Dec. 2002.

55. C. Ren, H. Y. Yu, J. F. Kang, Y. T. Hou, M.-F. Li, W. D. Wang, D. S. H. Chan, and D.-L. Kwong, "Fermi-level pinning induced thermal instability in the effective work function of TaN in TaN/SiO_2 gate stack," *IEEE Electron Device Lett.*, vol. 25, pp. 123-125, Mar. 2004.

56. H. Y. Yu, C. Ren, Y.-C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M.-F. Li, D. S. H. Chan, and D.-L. Kwong, "Fermi pinning induced thermal instability of metal gate work functions," *IEEE Electron Device Lett.*, vol. 25, pp. 337-339, May 2004.

57. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-k/metal–gate stack and its MOSFET characteristics," *IEEE Electron Device Lett.*, vol. 25, pp. 408-410, Jun. 2004.

58. Y. Akasaka, K. MiyagawaA, T. Sasaki,K. Shiraishi, S. Kamiyama, O. Ogawa, F. Ootsuka and Y. Nara, "Impact of electrode-side chemical structures on electron mobility in metal/HfO_2 MISFETs with sub-1nm EOT," in *VLSI Tech. Dig.*, pp. 228-229. 2005.

59. Y.-S. Suh, G. Heuss, H. Zhong, S.-N. Hong, and V. Misra, "Electrical characteristics of $TaSi_xN_y$ gate electrodes for dual gate Si-CMOS devices," in *VLSI Tech. Dig.*, pp. 47-48, 2001.

60. H. C. Wen, H. N. Alshareef, H. Luan, K. Choi, P. Lysaght, H. R. Harris, C. Huffman, G. A. Brown, G. Bersuker, P. Zeitzoff, H. Huff, P. Majhi, and B. H. Lee, "Systematic investigation of amorphous transition-metal-silicon-nitride electrodes for metal gate CMOS applications," in *Symp. VLSI Tech. Dig.*, pp. 46-47, 2005.

61. T.-H. Cha, D.-G. Park, T.-K. Kim, S.-A. Jang, I.-S. Yeo, J.-S. Roh, and J. W. Park, "Work function and thermal stability of $Ti_{1-x}Al_xN_y$ for dual metal gate electrodes," *Appl. Phys. Lett.*, vol. 81, pp. 4192-4194, 2002.

62. C. S. Kang, H.-J. Cho, Y. H. Kim, R. Choi, K. Onishi, A. Shahriar, and J. C. Lee, "Characterization of resistivity and work function of sputtered-TaN film for gate electrode applications," *J. Vac. Sci. Technol. B*, vol. 21, pp. 2026-2028, 2003.

63. H. Wakabayashi, Y. Saito, K. Takeuchi, T. Mogami, and T. Kunio, "A dual-metal gate CMOS technology using nitrogen-concentration-controlled TiN_x film," *IEEE Trans. Electron Devices*, vol. 48, pp. 2363-2369, 2001.

64. P. Ranade, Y.-C. Yeo, Q. Lu, H. Takeuchi, T.-J. King, C. Hu, "Molybdenum as a gate electrode for deep sub-micron CMOS technology", *Mat. Res. Soc. Symp.*, vol. 611, pp.C3.2.1-C3.2.6, 2000.

65. P. Ranade, Y. Choi, D. Ha, A. Agarwal, M. Ameen, and T. King, "Tunable work function molybdenum gate technology for FDSOI-CMOS," in *IEDM Tech. Dig.*, pp. 363-366, 2002.

66. Y.-C. Yeo, Q. Lu, P. Ranade, H. Takeuchi, K. J. Yang, I. Polishchuk, T.-J. King, C. Hu, S. C. Song, H. F. Luan, and D.-L. Kwong, "Dual-metal gate CMOS technology with ultra-thin silicon nitride gate dielectric," *IEEE Electron Device Lett.*, vol. 22, no. 5, pp. 227-229, May 2001.

67. Z. B. Zhang, S. C. Song, C. Huffman, J. Barnett, N. Moumen, H. Alshareef, P. Majhi, M. Hussain, M.S. Akbar, J.H. Sim, S.H. Bae, B. Sassman, and B.H. Lee, “Integration of dual metal gate CMOS with TaSiN (NMOS) and Ru (PMOS) gate electrodes on HfO_2 gate dielectric,” in *VLSI Tech. Dig.*, pp. 50-51, 2005.

68. I. Polishchuk, P. Ranade, T.-J. King, and C. Hu, “Dual work function metal gate CMOS technology using metal interdiffusion,” *IEEE Electron Device Lett.*, vol. 22, pp. 444-446, Sep. 2001.

69. J. Lee, H. Zhong, Y. Suh, G. Heuss, J. Gurganus, B. Chen, and V. Misra, “Tunable work function dual metal gate technology for bulk and non-bulk CMOS,” in *IEDM Tech. Dig.*, pp. 359-362, 2002.

70. B.-Y. Tsui, and C.-F. Huang, “Wide range work function modulation of binary alloys for MOSFET application,” *IEEE Electron Device Lett.*, vol. 24, pp. 153-155, Mar. 2003.

71. I. S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H. J. Kim, D. Cha, J. Huang, M. J. Kim, B. E. Gnade, J. Kim, and R. M. Wallace, “A novel methodology on tuning work function of metal gate using stacking bi-metal layers,” in *IEDM Tech. Dig.*, pp. 303-306, 2004.

72. T.-L. Li, C.-H. Hu, W.-L. Ho, H. C.-H. Wang, and C.-Y. Chang, “Continuous and precise work function adjustment for integratable dual metal gate CMOS technology using Hf–Mo binary alloys,” *IEEE Trans. Electron Devices*, vol. 52, pp. 1172-1179, 2005.

73. W. P. Maszara, Z. Krivokapic, P. King, J. Goo, and M. Lin, “Transistors with dual work function metal gates by single full silicidation (FUSI) of polysilicon gates,” in *IEDM Tech. Dig.*, pp. 367-370, 2002.

74. J. H. Sim, H. C. Wen, J. P. Lu, and D. L. Kwong, “Dual work function metal gates using full nickel silicidation of doped poly-Si,” *IEEE Electron Device Lett.*, vol. 24, pp. 631-633, Oct. 2003.

75. J. Kedzierski, D. Boyd, C. Cabral, Jr., P. Ronsheim, S. Zafar, P. M. Kozlowski, J. A. Ott, and M. Ieong, “Threshold voltage control in NiSi-gated MOSFETs through SIIS,” *IEEE Trans. Electron Devices*, vol. 52, pp. 39-46, 2005.

76. C. Cabral, Jr., J. Kedzierski, B. Linder, S. Zafar, V. Narayanan, S. Fang, A. Steegen, P. Kozlowski, R. Carruthers , and R. Jammy, “Dual workfunction fully silicided metal gates,” in *VLSI Tech. Dig.*, pp. 184-185, 2004.

77. J. Kedzierski, E. Nowak, T. Kanarsky, Y. Zhang, D. Boyd, R. Carruthers, C. Cabral, R. Amos, C. Lavoie, R. Roy, J. Newbury, E. Sullivan, J. Benedict, P. Saunders, K. Wong, D. Canaperi, M. Krishnan, K. Lee, B. A. Rainey, D. Fried, P. Cottrell, H. P. Wong, M. Ieong, and W. Haensch, “Metal-gate FinFET and fully-depleted SOI devices using total gate silicidation,” in *IEDM Tech. Dig.*, pp. 247-250, 2002.

78. K. Takahashi, K. Manabe, T. Ikarashi, N. Ikarashi, T. Hase, T. Yoshihara, H. Watanabe, T. Tatsumi, and Y. Mochizuki, "Dual workfunction Ni-silicide/HfSiON gate stacks by phase-controlled full-silicidation (PC-FUSI) technique for 45nm-node LSTP and LOP devices," in *IEDM Tech. Dig.*, pp. 91-94, 2004.

79. M. Terai, K. Takahashi, K. Manabe, T. Hase, T. Ogura, M. Saitoh, T. Iwamoto, T. Tatsumi, and H. Watanabe, "Highly reliable HfSiON CMOSFET with phase controlled NiSi (NFET) and Ni_3Si (PFET) FUSI gate electrode," in *VLSI Tech. Dig.*, pp. 68-69, 2005.

80. J. A. Kittl, A. Veloso, A. Lauwers, K. G. Anil, C. Demeurisse, S. Kubicek, M. Niwa, M. J. H. van Dal, O. Richard, M. A. Pawlak, M. Jurczak, C. Vrancken, T. Chiarella, S. Brus, K. Maex and S. Biesemans, "Scalability of Ni FUSI gate process: phase and Vt control to 30 nm gate lengths," in *VLSI Tech. Dig.*, pp. 72-73, 2005.

81. H. Y. Yu, J. D. Chen, M. F. Li, S. J. Lee, D. L. Kwong, M. van Dal, J.A. Kittl, A. Lauwers, E. Augendre, S. Kubicek, C. Zhao, H. Bender, B. Brijs, L. Geenen, A. Benedetti, P. Absil, M. Jurczak, and S. Biesemans, "Modulation of the Ni FUSI workfunction by Yb doping: from midgap to n-type band-edge," in *IEDM Tech. Dig.*, pp. 630-633, 2005.

82. N. Biswas, S. Novak, B. Chen, D. Lichtenwalner, M. Ozturk and V.Misra, "$Ni_xTa_{1-x}Si$ and $Ni_xPt_{1-x}Si$ ternary alloys for work function tuning on SiO_2, $HfSiO_x$ and HfO_2 dielectrics," in *IEDM Tech. Dig.*, pp. 650-653, 2005.

83. Y. H. Kim, C. Cabral. Jr, E. P. Gusev, R. Carruthers, L. Gignac, M. Gribelyuk, E. Cartier, S. Zafar, M. Copel. V. Narayanan, J. Newbury, B. Price, J. Acevedo, P. Jamison, B. Linder, W. Natzle, J. Cai, R. Jammy, and M. Ieong, "Systematic study of workfunction engineering and scavenging effect using NiSi alloy FUSI metal gates with advanced gate stacks," in *IEDM Tech. Dig.*, pp. 642-645, 2005.

84. C. S. Park, B. J. Cho, L. J. Tang, and D. Kwong, "Substituted aluminum metal gate on high-K dielectric for low work-function and Fermi-level pinning free," in *IEDM Tech. Dig.*, pp. 299-302, 2004.

85. T. Nabatame, M. Kadoshima, K. Iwamoto, N. Mise, S. Migita, M. Ohno, H. Ota, N. Yasuda, A. Ogawa, K. Tominaga, H. Satake, and A. Toriumi, "Partial silicides technology for tunable work function electrodes on high-k gate dielectrics-Fermi level pinning controlled $PtSi_x$ for HfO_x(N) pMOSFET," in *IEDM Tech. Dig.*, pp. 83-86, 2004.

86. E. Cartier, F. R. McFeely, V. Narayanan, P. Jamison, B. P. Linder, M. Copel, V. K. Paruchuri, V.S. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, G. Shahidi, "Role of oxygen vacancies in V_{FB}/V_t stability of pFET metals on HfO_2," in *VLSI Tech. Dig.*, pp. 230-231, 2005.

Chapter 3

The Metal-Dielectric Interface and Its Impact on the Effective Work Function of Metal Gates

3.1 Introduction

As discussed in Chapter 2, metal gate is more advantageous than the conventional poly-Si gate due to the elimination of the gate-depletion and dopant penetration problems, therefore it is seen as a performance booster for future CMOS transistors. However, one of the major challenges for the metal gate technology is how to get desirable work function (Φ_m) values in MOSFETs. Although the work function of a metal material in vacuum can be precisely measured as a physical parameter of the metal material itself, the effective work function ($\Phi_{m,eff}$, EWF) of the metal gate in a real MOS structure will be affected by many other factors such as the crystallinity, deposition techniques, impurities, and most importantly, the underlying gate dielectrics [1]-[2]. Moreover, integrating metal gates in a gate-first CMOS fabrication process requires the gate-stack to undergo a high-temperature source/drain (S/D) dopant activation annealing, which may lead to a change in the microstructures of the metal gates or some interactions at the metal-dielectric interface, and hence affects the effective work function. All these effects make the engineering of the

metal gate effective work function in MOSFET a complex and challenging task. From a scientific perspective, identifying the major factors which affects the effective work function of metal gates and understanding the mechanisms determining how these factors contribute to the effective work functions would be of significance.

The dependence of the metal gate effective work function on the gate dielectric materials was explained by Yeo *et al.* [1]-[2]. An interface dipole theory was proposed where the *intrinsic states* or *metal-induced gap states* (MIGS) are believed to be the major factor in determining the effective work function of metals on high-κ materials. However, it is not clear whether this model can be applied to explain the instability of the metal gate work function during the thermal annealing processes used in the CMOS fabrication, which finally determines the transistor threshold voltage.

In this chapter, we examined the dependence of the metal gate effective work function on the underlying gate dielectric materials as well as the process temperature. It was found that the creation of *extrinsic states*, which is typically related to the interface defects arising from the interface reactions, could be the major reason responsible for the thermal instability of the metal gate effective work functions. Different from the *intrinsic states*, the creation of these *extrinsic states* are thermodynamically driven and can result in a dipole layer at the metal-dielectric interface upon annealing, which modifies the $\Phi_{m,eff}$ of metal gates. A model considering the contribution of these *extrinsic states* is used to qualitatively explain the phenomenon of the process-induced work function instability. This insight could be useful in the understanding and engineering of this critical metal-dielectric interface.

3.2 Theoretical Backgrounds

3.2.1 Work Function of Metal Materials

The work function of a solid (either a metal or a semiconductor) in vacuum ($\Phi_{m,vac}$) is defined as the energy difference between the Fermi level in the solid and the vacuum level. For a semiconductor, the work function is a statistical concept and stands for the weighted average of the amount of energy required to raise an electron from the valence band and the conduction band to the vacuum level, respectively [3].

There are two parts for the work function calculation using quantum mechanical effect [3]: (1) A volume contribution which stems from the energy of an electron due to the periodic potential of the crystal and interaction of the electron with other electrons; (2) A surface contribution which is due to a possible surface dipole. As the electron charge distribution around the atoms at the solid surface is not symmetrically disposed around the nucleus, the center of the positive and the negative charge will not coincide, leading to a surface dipole. Therefore, any change in the surface electron charge distribution of a solid will result in the modification of the surface dipole layer, and hence the work function value of the solid.

3.2.2 Definition of Effective Work Function

The effective work function ($\Phi_{m,eff}$) of a metal gate in MOSFET represents the effective value of metal work function from the device operation point of view. $\Phi_{m,eff}$ directly determines the flat band (V_{FB}) and threshold voltage (V_{th}) of a MOSFET, and hence is of importance for device design. Although $\Phi_{m,vac}$ is a physical parameter of the metal material itself, the effectiveness of the metal gate electrode in determining V_{FB} and V_{th} of MOSFET may be affected by many other factors, e.g. the dipole layer

at the metal-dielectric interface. Fig. 3.1 illustrates the impact of interface dipoles in modulating the effective work function of metal gates. As shown in Fig. 3.1 (a), the V_{FB} is defined as the gate voltage required to flatten the energy bands in metal, dielectric and silicon, and is equal to the Fermi level difference between the metal and silicon if we assume that the oxide charge is negligible. However, the interface dipole layer will introduce an additional electric field at the metal-dielectric interface, and hence modify the alignment of the metal Fermi level to gate dielectric, as shown in Fig. 3.1 (b). As a result, the V_{FB} will be modified, as well as the V_{th} of transistors.

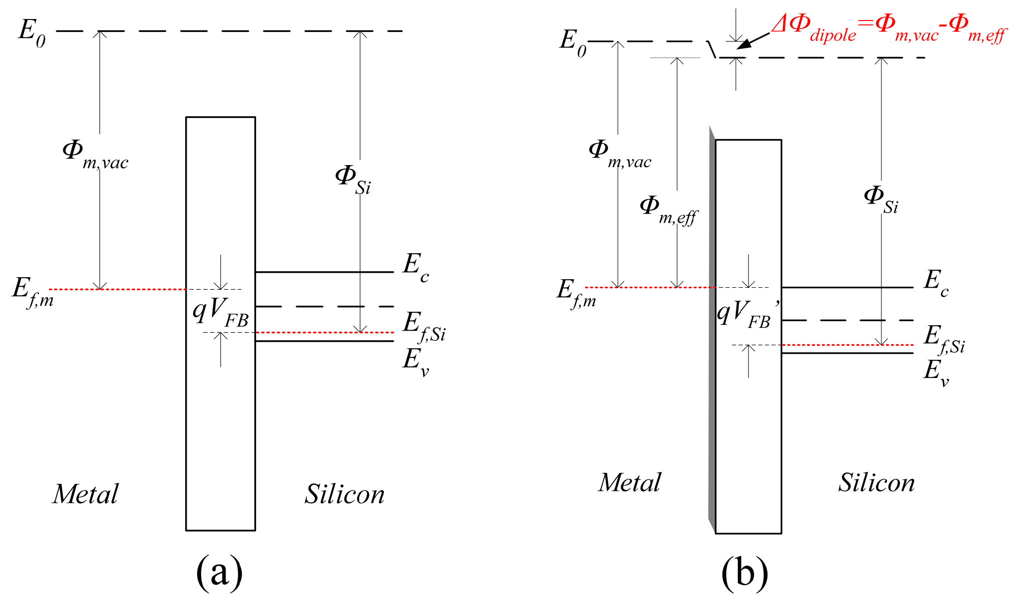


Fig. 3.1 Band diagram of a MOS structure in flat-band condition (a) without interface dipoles and (b) with interface dipoles at metal-dielectric interface.

From the device operation point of view, this result is equivalent to using a new metal material with a modified work function $\Phi_{m,eff}$ as the gate electrode, as illustrated in Fig. 3.1 (b). It is clear that the effective work function consists of both the contributions from the metal itself and that from the metal-dielectric interface, or other factors if any. Form the device engineering point of view, the effective work function is more important than the work function value measured in vacuum.

To avoid the possible confusion between these two concepts, the metal work function measured in vacuum will be deliberately denoted by $\Phi_{m,vac}$, while the effective work function from a device perspective will be expressed as $\Phi_{m,eff}$ in this chapter.

3.2.3 Factors Affecting the Work Function of Metals

In order to study the behavior of the $\Phi_{m,eff}$ in a MOS structure, we need to first understand the factors which may affect the effective work function of metals. These factors can be grouped into two categories: bulk contribution and interface contribution.

Let's start with the factors which affect the bulk properties on the metal work function. First, the different crystal orientations of a metal material will result in different work functions [4]-[5]. For Al, the work function difference between the different crystal orientations could be about 0.2 eV [4]; while that for Mo was reported to be about 0.5 eV [5]. The reason could be attributed to the different surface atom densities on difference facets [4]. Second, the microstructure of the metal film, e.g. grain size and film texture, may affect its work function. Generally, the film with an amorphous structure tends to have a low work function [6], while the poly-crystalline structures with lots of grains may have a relatively high work function [7]. Third, the impurities in the metal film can also have impact, and this effect could be more pronounced if the impurities pile up near the metal-dielectric interface. Pan *et al.* reported that impurities like P, C, and Si can modify the work function of TaN or TiN metal gates [8]. In addition, impurities like O and N are also believed to be capable of affecting the metal work function.

Besides these factors, the metal-dielectric interface contribution is also important for the $\Phi_{m,eff}$ of metal gates in a MOS stack. The contribution of the interface is realized through the interface dipoles, which is typically associated with the energy states or defects near the metal-dielectric interface. The creation of the interface states/defects can even "pin" the Fermi-level of metals at a certain position, which is commonly cited as the "Fermi-level pinning" (FLP) effect. However, the origin of these interface states/defects, as well as their material and process dependence, are still not well understood. This will be the main objective of this chapter.

3.2.4 Fermi Level Pinning: Schottky Model and Bardeen Model

The Fermi-level pinning effect is initially studied in the metal-semiconductor contact. A Schottky Model is first proposed to describe the ideal metal-semiconductor contact [9]. In this model, there is no charge transfer across the metal-semiconductor interface, and the Schottky Barrier Height (SBH, ϕ_b) for electron is determined only by $\Phi_{m,vac}$ and the electron affinity of the semiconductor χ_S. Therefore there is,

$$\phi_b = \Phi_{m,Vac} - \chi_s \qquad (3\text{-}1)$$

However, it was experimentally observed that the above equation is not generally obeyed. Bardeen then proposed a Surface State model to explain this observation [10]. Bardeen's model assumes that there exists high density of surface states at the semiconductor surface (1 per surface atom), as well as a gap δ of atomic-scale dimension between the metal and semiconductor. The filling of these surface states results in a charged dipole layer across the small gap δ, which "pins" the metal Fermi level to a certain energy level. However, later it was further pointed out by

Heine [11] that the high density of surface states did not exist in the fundamental gap for most metal-semiconductor interfaces. Instead, the *metal-induced gap states* (MIGS) or simply *intrinsic states* could be responsible for the FLP phenomenon observed.

3.2.5 Metal-Induced Gap States (MIGS) Theory and Its Limitations

When the conduction band of the metal overlaps the band gap of semiconductor, the wave functions of the electrons in the metal will decay into the band gap of the semiconductor, resulting in states in the forbidden gap of the semiconductor. These states are known as *intrinsic states* or called *metal-induced gap* states (MIGS). These MIGS can exist not only in the metal-semiconductor interface, but also in a metal-dielectric interface [1]-[2], [12].

Since the *intrinsic states* are split off from the valence and the conduction band, their character varies across the gap from mostly donor-type close to the top of the valence band E_v to mostly acceptor-type close to the bottom of the conduction band E_c [13]. Charge transfer generally happens between the metal and these *intrinsic states* in the dielectric. Filling the acceptor-type state gives an excess negative charge, while leaving the donor-type state empty gives an excess positive charge. These excess charges create a dipole at the interface, and this dipole layer tends to drive the band lineup toward a position that would give zero dipole charge, which is called charge neutrality level E_{CNL} [14]. Fig. 3.2 illustrates the characteristics of the *intrinsic states* at a metal-dielectric interface and the mechanism to create dipoles across the interface. As shown in Fig. 3.2, the energy level at which the dominant character of the interface states changes from acceptor-like to donor-like is defined as the charge-neutrality level (E_{CNL}) [14]. In other words, E_{CNL} is the balanced point of the weights

of the conduction-band and valence-band density of states. High-density states near the valence (conduction) band edge tend to drive the E_{CNL} towards the conduction (valence) band edge.

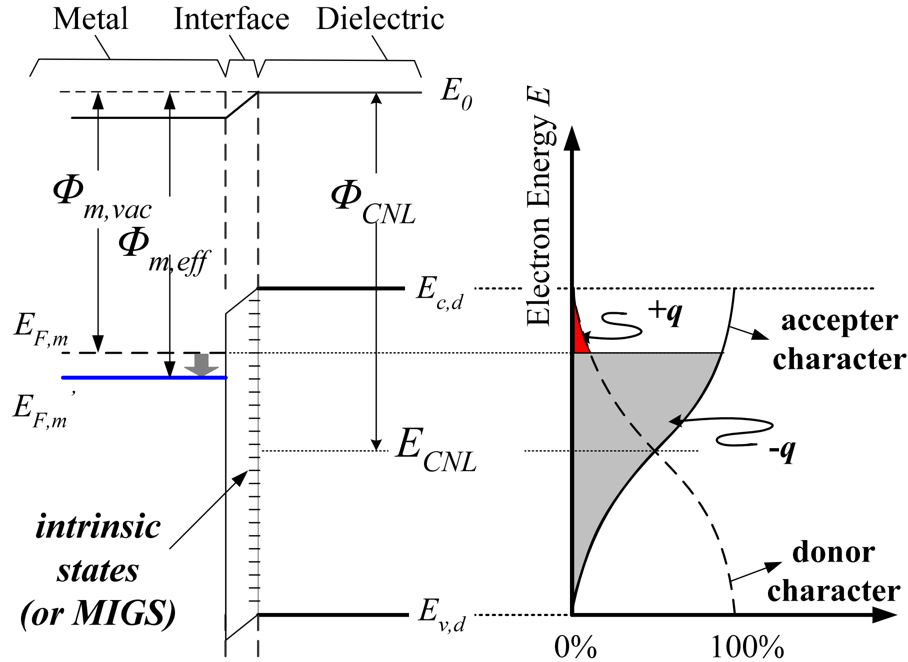


Fig. 3.2 Schematic energy band diagram (left) and the characteristics of the gap states (right) for metal gate on dielectrics. The character of MIGS becomes more acceptor- (donor-) like toward the Ec (Ev), as indicated by the solid (dashed) line [2].

Fig. 3.2 depicts the case where the metal Fermi level ($E_{F,m}$) is above the dielectrics E_{CNL}. Electrons from the metal tend to transfer across the interface to fill the MIGS where the energy is below $E_{F,m}$. Consequently, negative charges are created in the dielectrics side, and a dipole layer is then formed at the interface. $E_{F,m}$ would be driven towards the E_{CNL} by this interface dipole, resulting in the $\Phi_{m,eff}$ shift. According to the MIGS theory, the relationship between $\Phi_{m,eff}$ and $\Phi_{m,vac}$ is given by the following equation [2]:

$$\Phi_{m,eff} = \Phi_{CNL,d} + S(\Phi_{m,vac} - \Phi_{CNL,d}) \quad (3\text{-}2)$$

where S is the Schottky pinning parameter [12], describing the dielectric screening. In other words, S determines the strength of the FLP effect. W. Mönch found that the parameter of S empirically obeys the following equation [13]:

$$S = \frac{1}{1 + 0.1(\varepsilon_\infty - 1)^2} \tag{3-3}$$

where ε_∞ stands for the electronic part of the dielectric constant. The smaller the S parameter for a material, the more effective this material is to pin the metal Fermi level. When S is equal to zero, $E_{F,m}$ would be fully pinned to the E_{CNL}, and this is the “Bardeen limit”. When S is equal to 1, there is no pinning of $E_{F,m}$, and this is called the “Schottky limit”.

According to this MIGS theory, the Fermi-level pinning effect on high-κ dielectric will be dominated by the S parameter as well as the E_{CNL} of the gate dielectric materials [2]. On a SiO_2 dielectric, the FLP effect would be negligible due to the large S value of 0.95. On the other hand, significant FLP would be expected on HfO_2 or ZrO_2 because the S values are as small as 0.52 [2]. Such a model has been successful in fitting the experimental results in many metal-dielectric systems [2].

However, it is assumed in the *intrinsic states* or MIGS model that the metal-semiconductor or metal-dielectric interface is perfect without any physical/chemical defects. This may not be the case in real Schottky contact or MOS structures. W. E. Spicer *et al.* proposed an Interface Defect Model to explain the FLP phenomenon of III-V materials [15], where the *extrinsic defects* caused by the deficit of some atoms (e.g. As in GaAs) are believed responsible for the FLP. Nevertheless, there has been no research on the impact of the *extrinsic states* on the $\Phi_{m,eff}$ of metal gates in MOS structure, as well as the material and process dependence. This will be investigated and discussed in this chapter.

3.2.6 Work Function Measurement Techniques

Before starting the discussions on the metal gate work function, we will first briefly review the techniques to measure the $\Phi_{m,vac}$ and $\Phi_{m,eff}$ of metal gates. To measure the vacuum work function of metals, the commonly used technique is the photoemission method. Since the work function of a metal is determined by the valence band electrons instead of the core level electrons, ultraviolet (UV) light with hv of 5~100 eV is typically used to excite the photoelectrons from valence band, instead of the high-energy x-ray which can lead to too high a kinetic energy for the excited photoelectrons from the valence band. Therefore this technique is also called ultraviolet photoemission spectroscopy (UPS). From an analysis of the kinetic energy and angular distribution of the photoelectrons, information on the electronic structure of the material under investigation, including the $\Phi_{m,vac}$, can be extracted. Note that the UPS measurement is surface sensitive so that the absorbates on the surface need to be cleaned properly.

To get the $\Phi_{m,eff}$ of a metal material, an internal photoemission (IPE) technique has been developed and applied by Afanas'ev *et al* [16]. In this technique, an MOS structure with very thin transparent metal layer is required to enable the photon to reach the metal-dielectric interface. A gate bias will be applied on the metal electrode so that the leakage current through the dielectric can be monitored. The electrons in metal will be excited by the photons and give rise to a significant current when the photon energy hv is large enough for electrons to overcome the potential barrier at the metal-dielectric interface. Therefore, the barrier height can be determined from the threshold of the IPE current, so that the $\Phi_{m,eff}$ can be deduced. The disadvantage of this technique is that it requires special samples with transparent metal layers.

A straightforward approach to determine the $\Phi_{m,eff}$ of metal gate electrode in MOS structure is to measure the capacitance-voltage (C-V) characteristics. The following equation describes the V_{FB} of metal gate in a MOS structure:

$$V_{FB} = (\Phi_m - \Phi_s) - \frac{Q_{ox}}{C_{ox}} = (\Phi_m - \Phi_s) - \frac{Q_{ox}}{\varepsilon_o \varepsilon_{SiO2}} EOT \qquad (3\text{-}4)$$

where Q_{ox} is the equivalent oxide charge per unit area, C_{ox} is the oxide capacitance, and Φ_S is the work function of Si substrate. By plotting the V_{FB} versus the EOT determined from the C-V measurements, the work function difference between the metal and the silicon substrate can be obtained from the intercept on the y-axis. $\Phi_{m,eff}$ can therefore be calculated. This method provides an easy and convenient way to extract the effective work function of metal gates from the real MOS devices. Note that one of the assumptions for Eq. 3-3 is that the oxide charges should be located at the dielectric-Si interface and the amount of Q_{ox} does not change with the dielectric thickness. This assumption is reasonable for the conventional SiO_2/Si stack where the charge in bulk-SiO_2 is negligible. However, this assumption may not be valid in the high-κ dielectrics due to the relatively high Q_{ox} level in the bulk high-κ film as well as the existence of an interfacial layer under high-κ, which makes the distribution of oxide charge in high-κ stack more complex. Therefore some corrections may be needed to improve the accuracy of this approach [17].

Another method to determine the $\Phi_{m,eff}$ of metal gate is to extract the barrier height between metal and dielectric by analyzing the Fowler-Nordheim (F-N) tunneling current through the dielectric, as depicted in Eq. 3-4:

$$\frac{J_{FN}}{E_{ox}^{\ 2}} = \frac{q^3}{16\pi^2 \hbar \phi_{ox}} \exp\left(-\frac{4\sqrt{2m^*}\phi_{ox}^{\ 3/2}}{3\hbar q}\frac{1}{E_{ox}}\right) \qquad (3\text{-}5)$$

where J_{FN} is the F-N tunneling current density, E_{ox} is the electric field in the oxide, ϕ_{ox} is the barrier height between metal and oxide, and m^* is the effective electron

mass in oxide. By plotting ln (J_{FN}/E_{ox}^2) versus $1/E_{ox}$, a straight line should be obtained so that the barrier height ϕ_{ox} can be extracted from the slope of this plot. However, this method requires a prior knowledge of the effective mass in the dielectric, which may limit its applications on some high-κ dielectrics where the effective mass has not been very well determined.

3.3 Experimental

The metal-oxide-semiconductor capacitors were fabricated using p-Si (100) substrates (B, $6\text{x}10^{15}/\text{cm}^{-3}$). After the definition of the active area with 4000Å field oxide, and a standard DHF-last RCA pre-gate clean process, either thermally grown SiO_2 or MOCVD HfO_2 with different thicknesses was grown. Capacitors with TaN, TiN, HfN, TaTi, TaTiN and WN metal gates were fabricated. The various metal gate electrodes were deposited by dc sputtering. HfN was used as the capping layer on TaN, TaTi, and TaTiN gate electrodes to minimize oxygen diffusion through the gate stack during high-temperature post process, and hence minimize the EOT variation induced by oxygen diffusion. To study the thermal stability of Φ_m, the capacitors were annealed by rapid thermal annealing (RTA) in N_2 gas at 800-1000 ºC for ~20 s.

The *C-V* characteristics were measured on large area ($100\text{x}100\ \mu\text{m}^2$) MOS capacitors with an HP 4285A LCR meter at a high frequency (100 kHz). EOT and the flat-band voltage V_{FB} were determined through simulation using a model which takes the quantum mechanical effect into account [18]. The current-voltage measurements were performed using an HP 4156A semiconductor parameter analyzer. Values of metal gate work function $\Phi_{m,eff}$ were extracted from plots of V_{FB} versus the gate dielectric EOT.

3.4 Results and Discussions

3.4.1 Work Function Thermal Instability of TaN

In order to identify the major factors which influence the $\Phi_{m,eff}$ during process, we first compare the effective work function of TaN in TaN/SiO_2 and TaN/HfO_2 gate stacks ($\Phi_{TaN/SiO2}$ and $\Phi_{TaN/HfO2}$) before and after RTA treatments. As shown in Fig. 3.3, the V_{FB} vs. EOT plots exhibit linear relationship with almost identical slopes before and after RTA treatments. The fixed oxide charge density extracted from Fig. 3.3 is about $+1.6 \times 10^{11}$ cm^{-2} for TaN/SiO_2 and $-1.38 \times 10^{12} cm^{-2}$ for TaN/HfO_2 devices, and both are independent of RTA temperature. These results indicate that the V_{FB} shift after RTA is caused by the effective work function change instead of the increase in the fixed oxide charge. This allows accurate extraction of $\Phi_{TaN/SiO2}$ and $\Phi_{TaN/HfO2}$ from V_{FB} versus EOT plots. The values of $\Phi_{TaN/SiO2}$ and $\Phi_{TaN/HfO2}$ before and after RTA are noted in Fig. 3.3.

Comparing Fig. 3.3 (a) and Fig. 3.3 (b), it is interesting to note that for low temperature (420°C forming gas) annealed devices, $\Phi_{TaN/SiO2}$ (4.4eV) is quite close to $\Phi_{TaN/HfO2}$ (4.34eV) (also consistent with published data [19]-[20]), indicating that the interface structure and composition of TaN/SiO_2 and TaN/HfO_2 are very similar and the impact of MIGS is not very significant. However, the amount of work function change before and after 1000 °C RTA anneal is considerably different for the TaN/SiO_2 and TaN/HfO_2 systems. The increase of $\Phi_{TaN/SiO2}$ (from 4.4eV to 4.7eV) is much larger than that of $\Phi_{TaN/HfO2}$ (from 4.34eV to 4.41eV), indicating that the effective work function during RTA is strongly influenced by the interaction between metal and the underlying dielectric. As we discussed in Section 3.2.3, the structural change, such as the crystallization effect, may lead to a change of the metal gate work

function during the thermal annealing process. However, this may not be able to explain the large difference observed between $\Delta\Phi_{TaN/SiO2}$ and $\Delta\Phi_{TaN/HfO2}$ because both

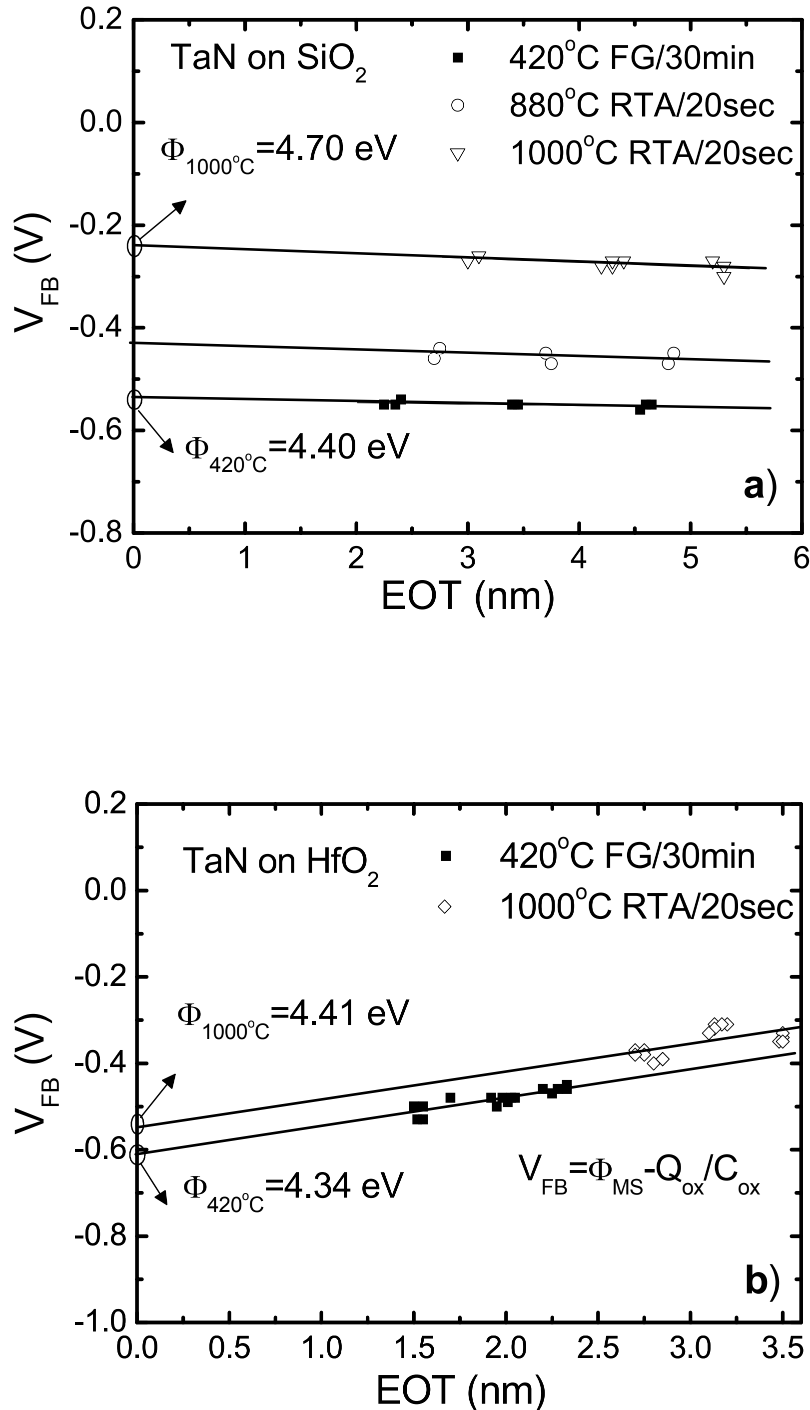


Fig. 3.3 Plots of V_{FB} versus EOT of (a) TaN/SiO_2 and (b) TaN/HfO_2 devices before and after 1000 °C RTA treatment, from which the effective work function of TaN can be extracted.

TaN/SiO_2 and TaN/HfO_2 devices received the same RTA and hence the same crystallization effects. Interfacial reaction between the metal electrode and the gate dielectric and the formation of an interfacial layer (metal silicide or high-κ) could be another reason accounting for the phenomenon observed. The interface reaction between the metal gate and gate dielectric has been studied in metal gates such as Hf, Ta, and $TaSi_xN_y$ [21]-[23]. One way to monitor the interfacial reaction is to check the EOT variation before and after high-temperature annealing. However, it is important to recognize that oxygen in the RTA ambient can diffuse through the gate stack and oxidize the Si substrate at the dielectric/Si interface, causing EOT to increase during RTA.

In order to eliminate the oxygen diffusion effects, we capped TaN with an HfN layer (50nm) to block the O_2 diffusion during RTA because HfN has been demonstrated to be an excellent O_2 diffusion barrier [24]. Fig. 3.4 shows both the EOT and work function of TaN devices as a function of RTA temperature with and without HfN capping layer. With the HfN capping layer, the EOT stability of TaN/SiO_2 stack is improved significantly even after 1000 °C RTA anneal. This clearly indicates that the EOT increase observed in TaN/SiO_2 stack after 1000 °C RTA is mainly caused by the O_2 diffusion and not due to the interfacial layer formation between TaN and SiO_2 during RTA, which would have affected EOT. In other words, the reaction between TaN and SiO_2 could be very slight and results in no detectable EOT change. However, even with the HfN capping layer, the effective work function still changes significantly after 1000 °C RTA, as shown in Fig. 3.4 (b).

The gate leakage measurements of HfN/TaN/SiO_2 gate stack under various RTA conditions are depicted in Fig. 3.5 (a). The leakage current decreases with the increasing RTA temperature, which is attributed to the increase of the effective work

function of TaN. The barrier height between TaN and SiO_2 is extracted before and after RTA treatment using the F-N tunneling ayalysis, as shown in Fig. 3.5 (b). The increase of the barrier height after RTA can be well correlated with the work function increase obtained from the *C-V* measurement.

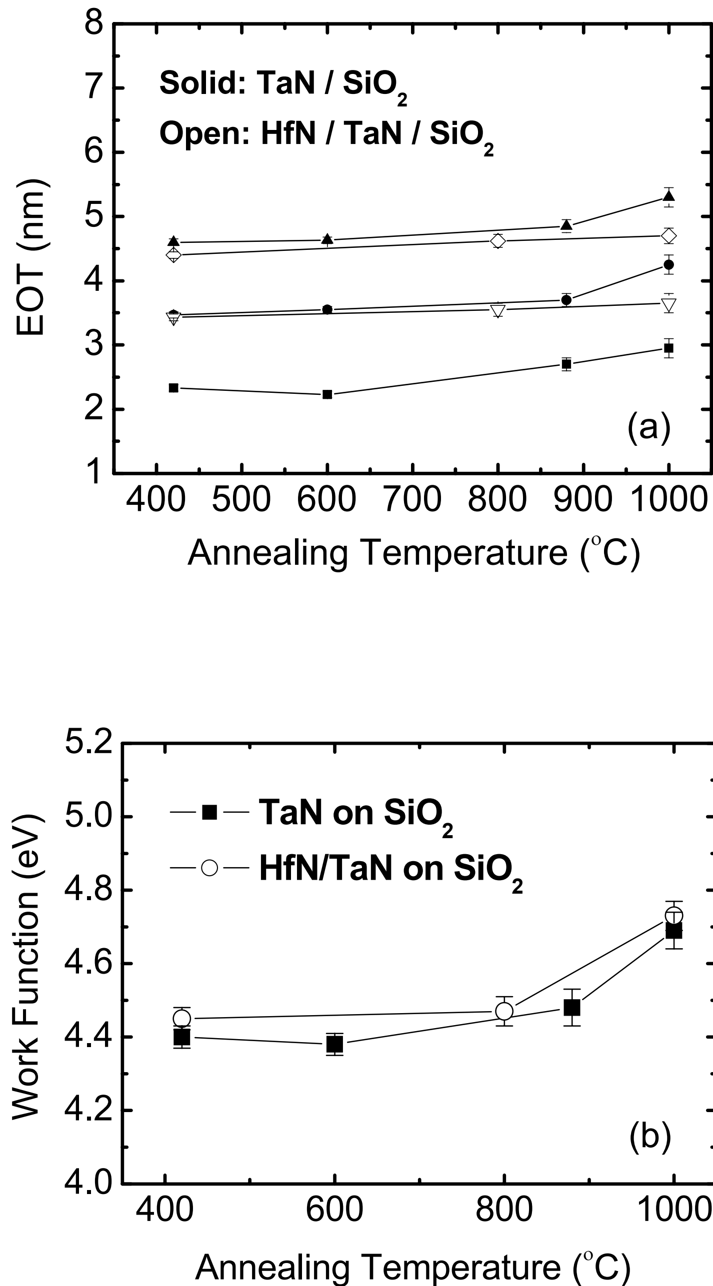
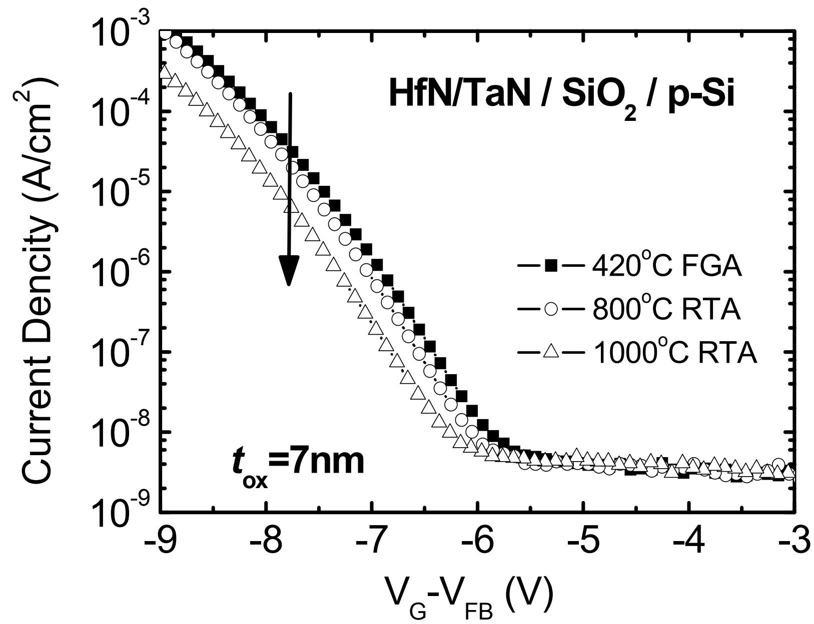
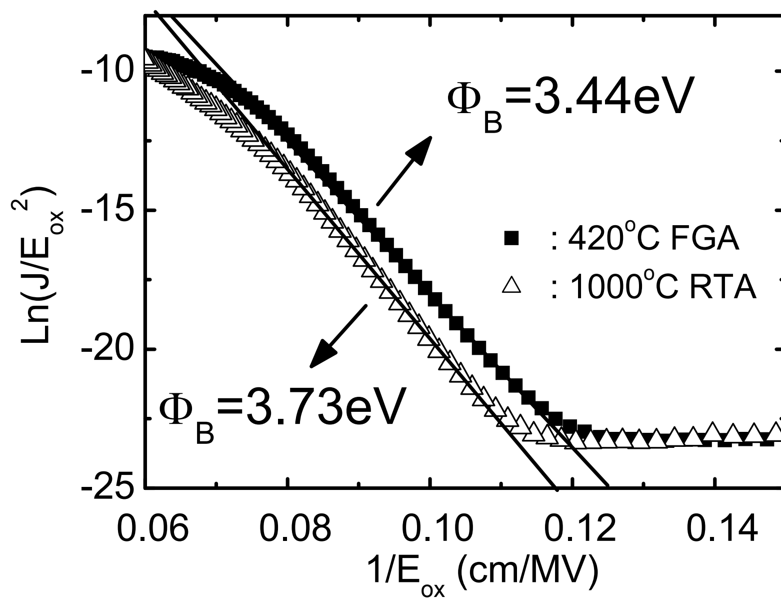


Fig. 3.4 The comparison of (a) EOT (with three different gate-oxide thickness) and (b) effective work function of TaN as a function of RTA temperature with and without HfN capping layer on top of the TaN/SiO_2 stack.



(a)



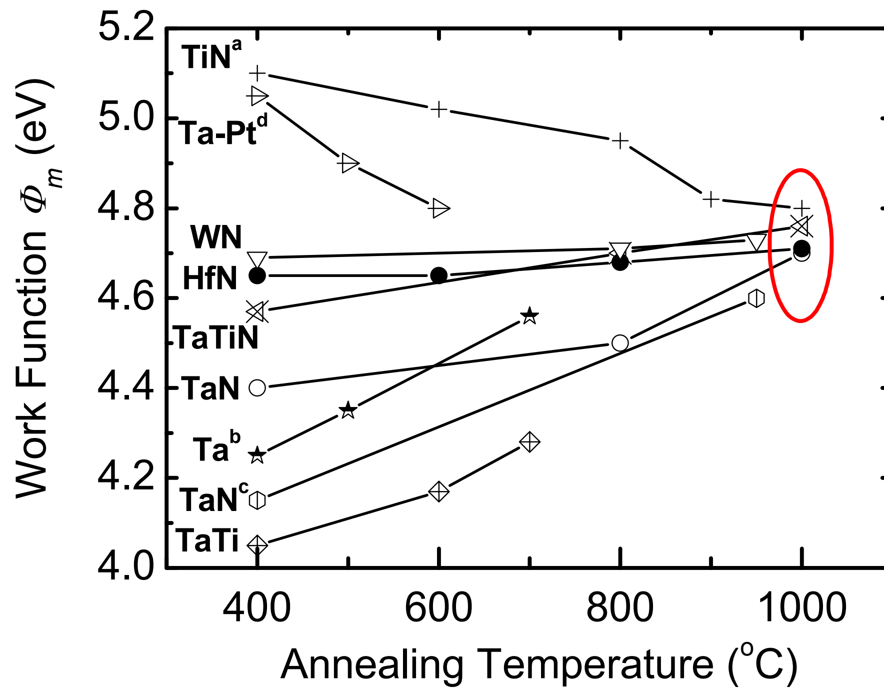
(b)

Fig. 3.5 (a) Gate leakage measurement of HfN/TaN/SiO_2 devices and (b) barrier height extraction by F-N tunneling analysis before and after 1000 °C RTA treatment.

Another mechanism which would affect the effective work function of TaN during the high-temperature annealing is the reaction between TaN and SiO_2 that only occur within an "interface" distance and thus have no impact on EOT. Comparing the changes in $\Phi_{TaN/SiO2}$ and $\Phi_{TaN/HfO2}$ caused by RTA, it suggests that the Fermi level of TaN at the TaN/SiO_2 interface is pinned from 4.4eV at 420°C to 4.7eV at 1000°C. However, this effect is not seen at the TaN/HfO_2 interface, as shown by Fig. 3.3 (b) where $\Phi_{TaN/HfO2}$ remains almost a constant, independent of the RTA temperature. We believe that the Ta(N)-Si bonds at the TaN/SiO_2 interface are responsible for the $\Phi_{TaN/SiO2}$ increase during the high-temperature RTA, similar to the pinning effect induced by Hf-Si bonds in poly-Si/HfO_2 devices [25]. At low temperatures (420°C), there are few Ta(N)-Si bonds due to the lack of activation, so the pinning effect is weak and $\Phi_{TaN/SiO2}$ is very close in value to $\Phi_{TaN/HfO2}$. As annealing temperature increases, the formation of Ta(N)-Si bonds are more likely to occur than that of Ta-Hf bonds due to the larger electronegativity difference between Ta and Si atoms compared to that between Ta and Hf atoms (Pauling's electronegativity: Ta~1.5, Si~1.9, and Hf~1.3). This may result in more interface defects or *extrinsic states* at the TaN/SiO_2 interface. Therefore, more charge transfer across the TaN/SiO_2 interface and stronger extrinsic pinning effect could be expected in TaN/SiO_2 stack. The lack of $\Phi_{m,eff}$ variation during RTA process in TaN/HfO_2 stake could be attributed to the chemical similarities between metal Ta(N) and metal oxide HfO_2, and hence the small number of *extrinsic states* created at the TaN/HfO_2 interface.

3.4.2 General Trends in the Process Dependence of $\Phi_{m,eff}$ on SiO_2 and High-*κ* Dielectrics

Besides TaN, the thermal instability issue regarding the $\Phi_{m,eff}$ value of many other metal gates are also studied in this work. To reveal the general trends of the $\Phi_{m,eff}$'s behaviors during process, the $\Phi_{m,eff}$ of many metals on SiO_2 are plotted in Fig.3.6 as a function of the RTA temperature. The experimental data from the literature [26]-[29] are also included in this figure.



[a] taken from Ref. [26]; [b] taken from Ref. [27];

[c] taken from Ref. [28]; [d] taken from Ref. [29].

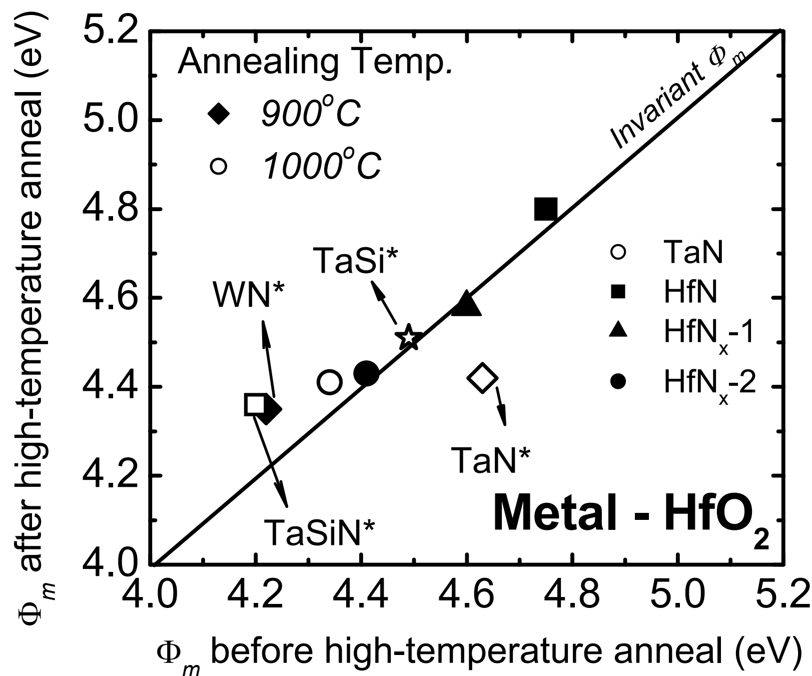
Fig. 3.6 The variation of metal gate work function Φ_m with the annealing temperature on SiO_2 dielectric.

From the above figure, a noteworthy trend is that the work function of some metals like TiN, Ta, TaTi, TaN, TaPt, and TaTiN converge towards the mid-gap position of Si as the annealing temperature increases. This suggests that the metal

gate Fermi levels are pinned at about 4.7-4.8 eV after high-temperature treatment. According to the theory of *intrinsic states* (MIGS), the FLP effect will be negligible for SiO_2 dielectric due to the high *S* parameter of SiO_2 [2]. Therefore, the observed change in $\Phi_{m,eff}$ with increasing temperature is likely to be due to the creation of a high density of *extrinsic states* upon annealing. The creation of these *extrinsic states* seems to be correlated with the interface reactions between metal gates and SiO_2 since the annealing temperature directly determines the magnitude of the $\Phi_{m,eff}$ variation in process. Like the case of TaN that has been discussed earlier in this chapter, the *extrinsic states* could be associated with some bonding defects between metals and Si. Some other possible source of the *extrinsic states* could be oxygen-related vacancies [30], nitrogen-related vacancies, or the penetrations of metal atoms into dielectrics. Given the chemical similarity of Ti, Hf, and Ta, it is plausible that *extrinsic states* with similar characteristics are formed between SiO_2 and these metals. This may explain why the $\Phi_{m,eff}$ tends to converge towards a similar energy position around the mid-gap of Si.

However, a different trend has been obtained on high-κ dielectric. In Fig. 3.7, the $\Phi_{m,eff}$ values of various metals on HfO_2 dielectric before high-temperature anneal are plotted on the horizontal axis, while that after high-temperature anneal are plotted on the vertical axis. If the work function of a metal gate does not change appreciably upon annealing at high temperature, it contributes a data point on the solid invariant line, otherwise the data point will deviate from the line. It is observed in Fig. 3.7 that the $\Phi_{m,eff}$ of many metals on HfO_2 dielectric does not change significantly upon annealing. Although the work function of some metals such as TaN or TaSiN on HfO_2 departs slightly from the invariant line, the deviation is quite small compared with that on SiO_2. This suggests that the RTA process temperature has less

significant impact on the metal-HfO_2 interface than that on the metal-SiO_2 interface. This phenomenon is possibly due to two reasons. Firstly, the *intrinsic states* at the metal-HfO_2 interface has already "pinned" the $\Phi_{m,eff}$ strongly, and the role of *extrinsic states* is comparatively diminished. Secondly, it could be possible that the creation of *extrinsic states* or interfacial bonding defects upon thermal annealing is less significant for the metal-HfO_2 interface compared with the metal-SiO_2 interface. The likelihood of *extrinsic states* generation could be related to the chemical constituents in the gate dielectric and the gate electrode. For example, given the similarity in atomic radii and electronegativity of metal atoms in the metal gate investigated in this work and the metal atoms (the hafnium) in the HfO_2 gate dielectric, it is plausible that chemical reactions are less likely to occur at the metal-HfO_2 interface [31].



* taken from Ref. [31].

Other data points are measured from this work.

Fig. 3.7 Work function of metal gates on HfO_2 before and after high-temperature annealing. HfN_x-1 and HfN_x-2 denotes HfN_x with different N concentration.

These results on SiO_2 and HfO_2 suggest that the creation of *extrinsic states* in process and the associated Fermi-level pinning problem could be an important consideration in tuning the effective work function of metal gates in MOSFET. The creation of *extrinsic states* is determined by the thermal/chemical stability of the metal-dielectric interface. Elevation of the annealing temperature probably increases the density of *extrinsic states* and hence the effectiveness in pinning the Fermi level of the metal gates.

3.4.3 Model: Fermi Level Pinning Induced by Extrinsic States

A metal-dielectric interface model considering the contribution of the *extrinsic states* to the $\Phi_{m,eff}$ is hence proposed to qualitatively explain the thermal instability of metal gate work function on SiO_2, as shown in Fig. 3.8. The creation of the *extrinsic states* is thermodynamically driven, and the states will have a certain energy level in the band gap of the gate dielectric, which is determined by the properties of both the gate electrode and the gate dielectric materials. Fig. 3.8 (a) illustrates the case where the metal Fermi level $E_{f,m}$ is above the energy level of the *extrinsic states*, and thus the *extrinsic states* at the pinning location (at an energy $\Phi_{pin,ex}$ below the vacuum level) are filled with electrons from the metal. This creates an interface dipole that is charged negatively in the dielectric side, driving $E_{f,m}$ towards the pinning position. Conversely, for the case where the metal Fermi level $E_{f,m}$ is below the energy level of the *extrinsic states*, as shown in Fig. 3.8 (b), the existing electrons at the pinning level tend to redistribute toward the metal side, resulting in an interface dipole that is charged positively in the dielectric side. As a result, the $E_{f,m}$ will move up towards the pinning position. This explains why the $\Phi_{m,eff}$ of various metals with different $\Phi_{m,vac}$ values tends to converge towards a similar position, as depicted in Fig. 3.6. Note that

the $\Phi_{m,eff}$ variation will be less pronounced if $E_{f,m}$ is close to the pinning level of the *extrinsic states*, like the situation of HfN and WN on SiO_2 as illustrated in Fig. 3.6.

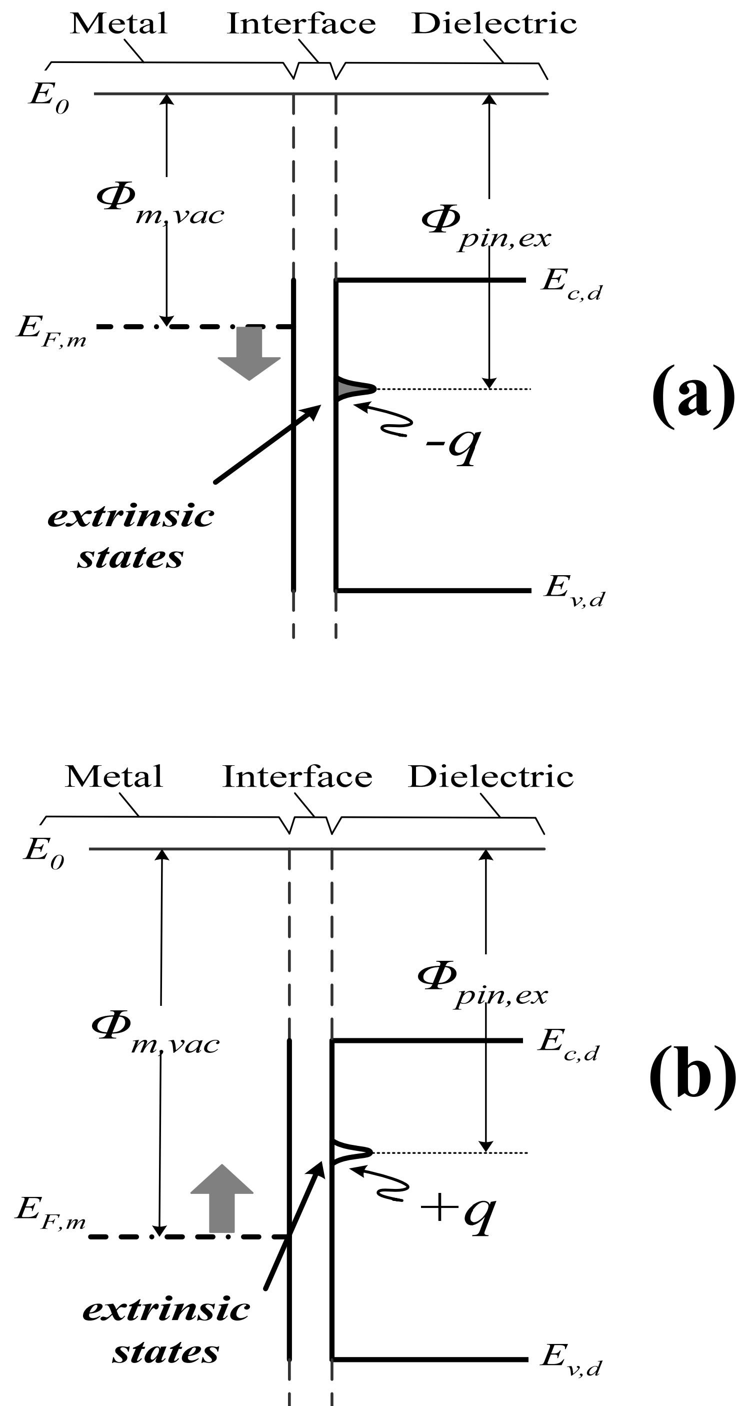


Fig. 3.8 Schematic energy band diagram for a metal gate on a dielectric, showing the mechanism of Fermi-level pinning by *extrinsic states*. (a) When $E_{F,m}$ is above the pinning level, (b) When the $E_{F,m}$ is below the extrinsic pinning level. The conduction band edge and the valence band edge of the dielectric are denoted by $E_{c,d}$ and $E_{v,d}$, respectively.

The source of the *extrinsic states* could be related to the interfacial bonding defects between the gate electrode and the gate dielectrics, or some vacancies (e.g. oxygen vacancy) near the interface [30]. In order to avoid or minimize the creation of *extrinsic states* in device design and processing, the source of the *extrinsic states* need to be identified. However, it is difficult for the conventional physical analytical tool such as XPS, AES, or SIMS to investigate this problem due to their limited resolution. Therefore we will discuss some preliminary study using electrical characterization.

3.4.4 Investigation of Hf-Si Bond Induced Extrinsic States

As discussed in the above sections, the creation of the *extrinsic states* could be responsible for the thermal instability of metal gate work function on SiO_2. The existence of *extrinsic states* can also be found in the poly-Si/HfO_2 interface, as described in [32]-[33]. However, there has been no consensus on the origin of the *extrinsic states* at the gate-dielectric interface. The oxygen vacancy induced interface dipoles was proposed by K. Shiraishi *et al*. to explain the flatband voltage (V_{FB}) shift of p^+ poly-Si gate on HfO_2 [34], and this mechanism has been found to be as important for many p-type metal candidates [30], [36]. However, this theory has difficulty in explaining the symmetric shift of $\Phi_{m,eff}$ for n^+ and p^+ poly-Si electrodes [33]. On the other hand, some researchers believe that the Hf-Si bonds at the poly-Si/HfO_2 or FUSI/high-κ interfaces are responsible for the FLP effects observed in these gate stacks [33], [37]. However, it is still not clear how the Hf-Si bond affects the $\Phi_{m,eff}$ and whether the direction of the Hf-Si bond at the interface is a factor or not.

To investigate the origin of the *extrinsic states* at the gate-dielectric interface and to study their properties, we investigated the behavior of $\Phi_{m,eff}$ in novel laminated metal gate structures consisting of ultra-thin Si and Hf (N) layers, where the bond

configurations at the gate-dielectric interface can be deliberately varied by using different laminated layers on SiO_2 and HfO_2 dielectrics.

Capacitors with different laminated structures on SiO_2 and HfO_2 dielectrics were fabricated, as summarized in Table 3.1. For the splits with high-κ dielectric, HfO_2 with a thickness of 10 Å was deposited by MOCVD on top of the slanted SiO_2 to minimize the impact of oxide fixed charge associated with thick high-κ films. Gate electrodes comprising different laminated stacks were then deposited. From these splits, different interface bond configurations including Si-Si, Hf-Si, Si-Hf, and Hf-Hf bonds can be achieved. After gate patterning, RTA was then performed at 1000°C for 5 s in some splits, followed by a final forming gas anneal (FGA) at 420 ^{o}C for 30 min.

Table 3.1 Experimental splits of different laminated stacks consisting Hf (N) and Si layers on slanted SiO_2 and HfO_2 dielectrics.

No.	Gate Dielectric	Laminated Stacks	Thickness (Å)
1A	SiO_2	TaN /Hf /Si	1000 /10 /10
2A	SiO_2	TaN /HfN /Si	1000 /40 /10
3A	SiO_2	TaN /Si /HfN	1000 /10 /10
1B	SiO_2 + HfO_2 (10 Å)	TaN /Hf /Si	1000 /10 /10
2B	SiO_2 + HfO_2 (10 Å)	TaN /HfN /Si	1000 /40 /10
3B	SiO_2 + HfO_2 (10 Å)	TaN /Si /HfN	1000 /10 /10

The behavior of $\Phi_{m,eff}$ in TaN/Hf/Si stack was first investigated on SiO_2 and SiO_2/HfO_2 dielectrics before and after 1000 ^{o}C RTA. As shown in Fig. 3.9, $\Phi_{m,eff}$ of TaN/Hf/Si stack on SiO_2 and HfO_2 are about 4.15 and 4.2 eV, respectively, in the as-deposited condition, and the values become ~ 4.6 eV and ~ 4.62 eV after RTA treatment at 1000 ^{o}C. The variation of $\Phi_{m,eff}$ upon annealing could be attributed to two factors: "bulk" effect and/or "interface" effect. The "bulk" effect is typically related to the microstructure or composition change of the metal gate, while the

"interface" effect represents the contribution of a dipole layer possibly formed at the metal-dielectric interface. The dipole formation at the metal-dielectric interface could be arising from both *intrinsic states* and *extrinsic states*, and the later one is typically associated with the thermal process used in device fabrication. Since the $\Phi_{m,eff}$ of TaN/Hf/Si stack in the as-deposited condition shows only small difference on SiO_2 and HfO_2 dielectrics, it can thus be inferred that the contribution from the *intrinsic states* to the $\Phi_{m,eff}$ is not a major factor in our case [2]. On the other hand, the contribution of *extrinsic states* associated with the interface defects could become more pronounced after RTA annealing, leading to the $\Phi_{m,eff}$ instability problem after RTA treatments [38]. However, it should be noted that a silicidation process between the laminated ultra-thin Si and Hf layers is also likely to happen during RTA, resulting in the formation of $HfSi_x$ material and a change of the bulk work function. Therefore, it is difficult to distinguish the "interface" effect with the "bulk" effect in this case.

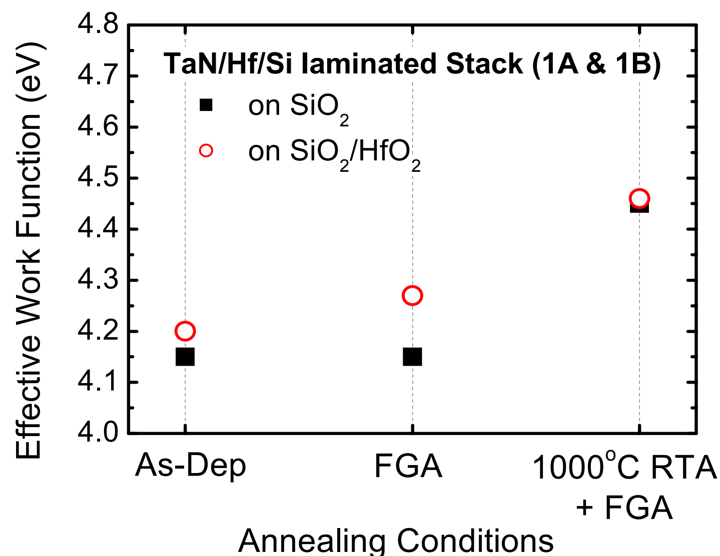


Fig. 3.9 Work function of the TaN/Hf/Si laminated stack on SiO_2 and HfO_2 after annealing at different conditions: as-deposited, 420 ^{o}C FGA and RTA at 1000 ^{o}C for 5 sec followed by FGA.

In order to separate the contributions of "interface" and "bulk-material" to the variation of $\Phi_{m,eff}$ during the RTA process, the Hf layer is replaced by HfN in the laminated structures so that the silicidation process can be minimized during the RTA process. As shown in Fig. 3.10 and Fig. 3.11, the variation in the $\Phi_{m,eff}$ of these metal stacks is observed to be dependent to the different dielectrics, suggesting that the "interface contribution" become dominant. Interestingly, a "kink" can be observed in some of the *C-V* measurements after RTA annealing, which seems to be a kind of "non-uniformity" problem since both the smooth *C-V* curves and the "kinked" curves can be obtained from the same sample, as shown in Fig. 3.10 (b) & Fig. 3.11 (b). Similar "kinked" *C-V* curves have also been reported in Ref. [35] and are attributed to the different localized work function values in the gate electrode comprising different metal materials. In our case, a localized crystallization effect among the ultra-thin laminated layers could have happened during the high-temperature process, so that the grain growth can break the continuity of the well-ordered bottom layer. The crystallized grains may have different compositions compared with the well-ordered stacks, leading to the different localized $\Phi_{m,eff}$ values as well as the "kinks" in the *C-V* measurements. The morphology of the gate-dielectric interface has been examined by XTEM, as illustrated in Fig. 3.12. The result confirms that the localized crystallization effects are visible after RTA and some of the grains can breakthrough the bottom layer and come into physical contact with the dielectric, which is consistent with the explanation above. Based on this understanding, therefore, it can be inferred that the upper part of the "kinked" *C-V* curve possibly represents the $\Phi_{m,eff}$ of the localized newly-grown grains, while the lower part represents that of the well-ordered laminated regions.

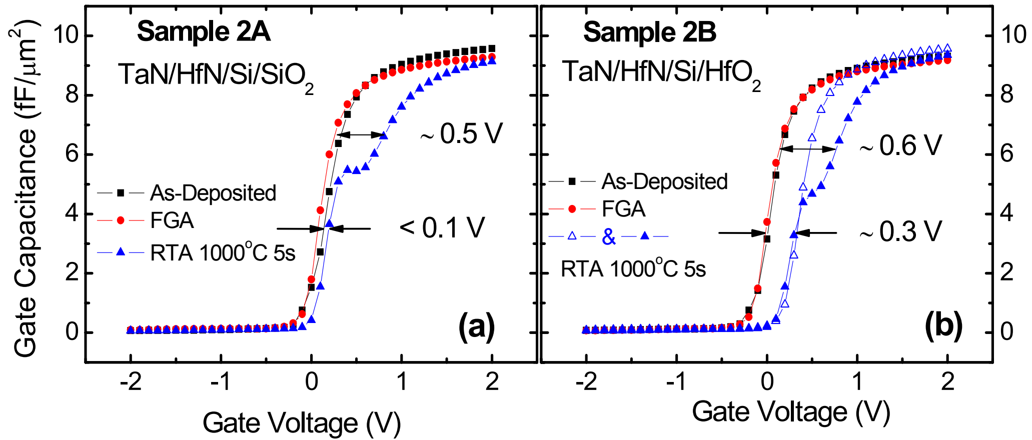


Fig. 3.10 *C-V* measurements of the TaN/HfN/Si laminated stack on (a) SiO_2 and (b) HfO_2/SiO_2 dielectrics after different annealing.

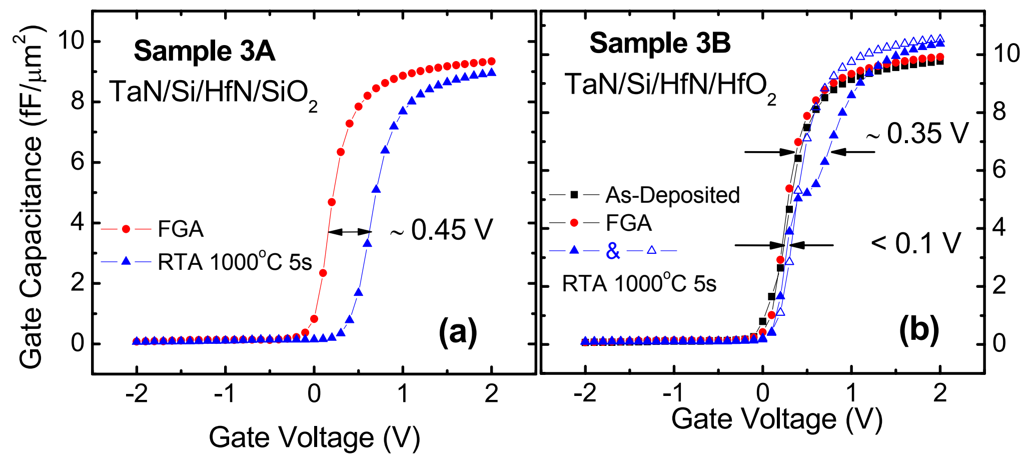


Fig. 3.11 *C-V* measurements of the TaN/Si/HfN laminated stack on (a) SiO_2 and (b) HfO_2/SiO_2 dielectrics after different annealing.

Despite the “kinks” observed in the *C-V* measurements, the V_{FB} shift observed from the lower part of the *C-V* curves in Fig. 3.10 and Fig. 3.11 shows a clear dependence on different gate dielectrics and the different bond configurations at the interface, as summarized in Table 3.2. First, the magnitude of the V_{FB} shift after RTA is found to be considerably different for the devices with SiO_2 and HfO_2 dielectrics,

suggesting that the bond configuration at the gate-dielectric interface contributes to the V_{FB} shift. Secondly, as depicted in Table 3.2, the Hf-Si bonds at the gate-dielectric interface always lead to significant V_{FB} shift after RTA, while the Si-Si or Hf-Hf bond configuration results in negligible V_{FB} shift. This implies that the creation of the *extrinsic states*, which lead to the dipole formation and V_{FB} shifts, is determined by the chemical properties of elements at the gate-dielectric interface. A metal-metal or Si-Si atoms terminated interface is not likely to induce interface defects during the high-temperature treatment so that the creation of extrinsic states is negligible, enabling better $\Phi_{m,eff}$ stability to be achievable compared with the Hf-Si terminated interface. Thirdly, irrespective of the direction of Hf-Si (or Si-Hf) bonds across the interface, the V_{FB} and $\Phi_{m,eff}$ always shift in a same direction (comparing devices 2B and 3A), implying that the dipole layer formation is mainly due to the creation of some energy states rather than the charge transfer from one kind of atom to another [33]. Finally, these results are difficult to be explained by the oxygen vacancy theory, where a negative V_{FB} shift should be expected for device with HfO_2 dielectric. Therefore it is believed that the Si-metal bond related interface defects should be the major source for the *extrinsic states* at the gate-dielectric interface.

Table 3.2 V_{FB} shifts for TaN/HfN/Si and TaN/Si/HfN stacks on SiO_2 and HfO_2 after RTA at 1000 ^{o}C for 5 sec, observed from the lower parts of the *C-V* curves shown in Fig. 3.10 & Fig. 3.11.

	SiO_2			**HfO_2**		
	No.	ΔV_{FB} (V)	bonds @ interface	No.	ΔV_{FB} (V)	bonds @ interface
TaN /HfN /Si	2A	**< 0.1**	Si-Si	2B	**~ 0.3**	Si-Hf
TaN /Si /HfN	3A	**~ 0.45**	Hf-Si	3B	**< 0.1**	Hf-Hf

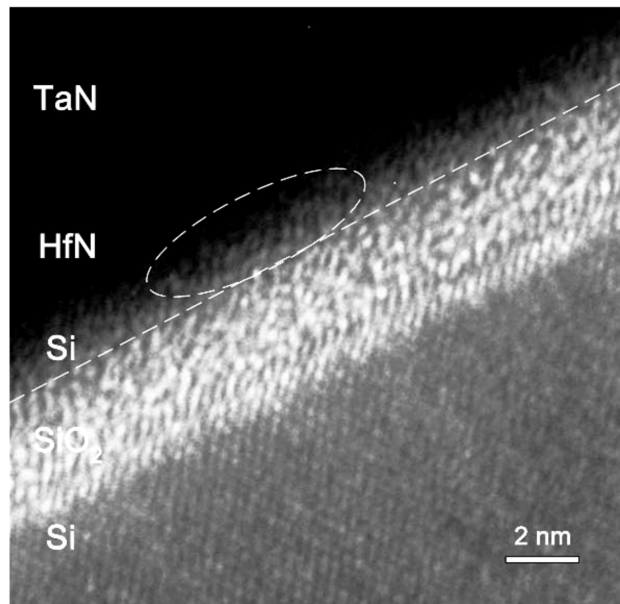


Fig. 3.12 XTEM image of a TaN/HfN/Si stack on SiO_2 after 1000 ^oC RTA for 5 sec.

In short summary, by investigating the behavior of the WF in laminated metal gate structures consisting of Si and Hf (N) layers, it was found that the presence of a Hf-Si bond is correlated with the creation of *extrinsic states* which can lead to the EWF instability during high-temperature annealing process, while Hf-Hf or Si-Si bonds do not. The creation of the bond-related interface states as well as the dipole layer formation is irrespective of the direction of Hf-Si bond across the interface. This understanding could be useful in engineering the gate-dielectric interface in MOSFETs.

3.5 Conclusion

The dependence of metal gate work function on annealing temperature was investigated in devices with several combinations of metal gate and gate dielectric

materials. It is identified that the *extrinsic states* created at the metal-dielectric interface is the major factor responsible for the instability of metal gate effective work function during high-temperature process. The creation of *extrinsic states* and the resulting Fermi level pinning problem appears to be thermodynamically driven, and becomes more pronounced when the annealing temperature is higher. A metal-dielectric interface model that takes the role of *extrinsic states* into account is proposed to qualitatively explain the work function instability phenomenon. In general, the generation of *extrinsic states* upon annealing can be correlated with the chemical bond configurations at the metal-dielectric interface. The presence of a metal-Si bond at the gate-dielectric interface can lead to the $\Phi_{m,eff}$ instability problem during high-temperature annealing process, while metal-metal or Si-Si bonds do not. Moreover, the creation of the bond-related interface states as well as the dipole layer formation is irrespective of the direction of the bonds across the interface. These insights regarding the metal-dielectric interface could be useful for work function tuning and interface engineering of the metal gate electrodes in future MOS devices.

References

1. Y.-C. Yeo, P. Ranade, T.-J. King, and C. Hu, “Effects of high-k gate dielectric materials on metal and silicon gate workfunctions”, *IEEE Electron Device Lett.*, vol. 23, pp. 342-344, 2002.

2. Y.-C. Yeo, T.-J. King, and C. Hu, “Metal-dielectric band alignment and its implications for metal gate complementary metal-oxide-semiconductor technology”, *J. Appl. Phys.*, vol. 92, pp. 7266-7271, 2002.

3. E. H. Rhoderick, and R. H. Williams, “*Metal-Semiconductor Contacts*,” 2nd Edition, Clarendon Press, Oxford, 1988.

4. R M Eastmenti and C H B Mee, “Work function measurements on (100), (110) and (111) surfaces of aluminium,” *J. Phys. F: Metal Phys.*, vol. 3, pp. 1728-1745, 1973.

5. S. Berg, P. O. Gartland and B. J. Slagsvold, “Photoelectric work function of a molybdenum single crystal for the (100), (110), (111), (112), (114), and (332) faces,” *Surface Science*, vol. 43, pp275-292, 1974.

6. P. Ranade, Y. Choi, D. Ha, A. Agarwal, M. Ameen, and T. King, “Tunable work function molybdenum gate technology for FDSOI-CMOS,” in *IEDM Tech. Dig.*, pp. 363-366, 2002.

7. S. H. Bae, W. P. Bai, H. C. Wen, S. Mathew, L. K. Bera, N. Balasubramanian, N. Yamada, M. F. Li, and D. L. Kwong, “Laminated metal gate electrode with tunable work function for advanced CMOS,” in *Symp. VLSI Tech. Dig.*, pp. 188-189, 2004.

8. J. Pan, C. Woo, M.-V. Ngo, P. Besser, J. Pellerin, Q. Xiang, and M.-R. Lin, “A low-temperature metal-doping technique for engineering the gate electrode of replacement metal gate CMOS transistors,” *IEEE Electron Device Lett.*, vol. 24, pp. 547-549, Sep. 2003.

9. W. Schottky, Phys. Z. vol. 41, pp. 570, 1940.

10. J. Bardeen, “Surface states and rectification at a metal semiconductor contact,” *Phys. Rev.*, vol. 71, pp. 717-727, 1947.

11. V. Heine, “Theory of Surface States,” *Phys. Rev*, vol. 138, pp. A1689-A1696, 1965.

12. John Robertson, “Band offsets of wide-band-gap oxides and implications for future electronic devices,” *J. Vac. Sci. Technol. B*, vol. 18, pp.1785-1791, 2000.

13. W. Mönch, “Role of virtual gap states and defects in metal-semiconductor contacts,” *Phys. Rev. Lett.*, vol. 58, pp. 1260-1263, 1987.

14. J. Tersoff, “Theory of semiconductor heterojuction: The role of quantum dipoles,” *Physical Review B*, vol. 30, No. 8, pp. 4874, 1984.

15. W. E. Spicer, P. W. Chye, P. R. Skeath, C. Y. Su and I. Lindau, “New and unified model for Schottky barrier and III-V insulator interface states formation,” *J. Vac. Sci. Technol.*, vol. 16, pp.1422-1433, 1979.

16. V K Adamchuk and V. V. Afanas’ev, “Internal photoemission spectroscopy of semiconductor-insulator interface,” *Pro. In Surf. Sci.*, vol. 41, pp. 111-211, 1992.

17. R. Jha, J. Gurganos, Y. H. Kim, R. Choi, J. Lee, and V. Misra, “A capacitance-based methodology for work function extraction of metals on high-k,” *IEEE Elecron Device Lett.*, vol. 25, pp. 420-423, 2004.

18. Keven Yang, Ya-Chin King, and Chenming Hu, “Quantum effect in oxide thickness determination from capacitance measurement,” in *Symp. VLSI Tech. Dig.*, pp. 77-78, 1999.

19. D.-G. Park, T.-H. Cha, K.-Y. Lim, H.-J. Cho, T.-K. Kim, S.-A. Jang, Y.-S. Suh, V. Misra, I.-S. Yeo, J.-S. Roh, J. W. Park, and H.-K. Yoon, “Robust Ternary Metal Gate Electrodes for Dual Gate CMOS Devices,” in *IEDM Tech. Dig.*, pp. 671-674, 2001.

20. J. C. Hooker, R. J. P. Lander, Z. M. Rittersma, T. Schram, G. S. Lujan, J. van Zijl, E. van den Heuvel and F. Roozeboom, “Materials and Electrical Characterization of Metal Gate Electrodes on High-k Dielectrics for Advanced CMOS Technologies,” in *Proc. Int. Conf. Solid State Devices and Materials*, pp. 174-175, 2002.

21. V. Misra, G. P. Heuss, and Huicai Zhong, “Use of metal-oxide-semiconductor capacitors to detect interactions of Hf and Zr gate electrodes with SiO_2 and ZrO_2,” *Appl. Phys. Lett.*, vol. 78, no. 26, pp. 4166-4168, 2001.

22. T. Ushiki, K. Kawai, I. Ohshima, and T. Ohmi, “Chemical Reaction Concerns of Gate Metal with Gate Dielectric in Ta Gate MOS Devices: An Effect of Self-Sealing Barrier Configuration Interposed Between Ta and SiO_2,” *IEEE Trans. Electron Devices*, vol. 47, no.11, pp. 2201-2207, Nov. 2000.

23. Y.-S. Suh, G. P. Heuss, and V. Misra, “Electrical characteristics of $TaSi_xN_y/SiO_2/Si$ structures by Fowler-Nordheim current analysis,” *Appl. Phys. Lett.*, vol. 80, no. 8, pp. 1403-1405, 2002.

24. H. Y. Yu, H. F. Lim, J. H. Chen, M. F. Li, C. X. Zhu, D.-L. Kwong, C. H. Tung, K. L. Bera, and C. J. Leo, “Robust HfN Metal Gate Electrode for Advanced MOS Devices Application,” in *Symp. VLSI Tech. Dig.*, pp. 151-152, 2003.

25. C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L.

Hebert, H. Tseng, B. White and P. Tobin, "Fermi Level Pinning at the PolySi/Metal Oxide Interface," in *Symp. VLSI Tech. Dig.*, pp. 9-10, 2003.

26. J. Westlinder, T. Schram, L. Pantisano, E. Cartier, A. Kerber, G. S. Lujan, J. Olsson, and G. Groeseneken, "On the thermal stability of atomic layer deposited TiN as gate electrode in MOS devices", *IEEE Elecron Device Lett.*, vol. 24, pp. 550-552, 2003.

27. J. Lee, H. Zhong, Y. Suh, G. Heuss, J. Gurganus, B. Chen, and V. Misra, "Tunable work function dual metal gate technology for bulk and non-bulk CMOS", in *Tech. Dig. IEDM*, 2002, pp. 359-362.

28. C. S. Kang, H. J. Cho, Y. H. Kim, R. Choi, K. Onishi, A. Shahriar, and J. C. Lee, "Characterization of resistivity and work function of sputtered-TaN film for gate electrode applications", *J. Vac. Sci. Technol. B*, vol. 21, pp. 2026-2028, 2003.

29. B. Y. Tsui and C. F. Huang, "Wide range work function modulation of binary alloys for MOSFET application", *IEEE Elecron Device Lett.*, Vol. 24, pp. 153-155, 2003.

30. E. Cartier, F. R. McFeely, V. Narayanan, P. Jamison, B. P. Linder, M. Copel, V. K. Paruchuri, V.S. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, G. Shahidi, "Role of oxygen vacancies in V_{FB}/V_t stability of pFET metals on HfO_2," in *VLSI Tech. Dig.*, pp. 230-231, 2005.

31. J. K. Schaeffer, S. B. Samavedam, D. C. Gilmer, V. Dhandapani, P. J. Tobin, J. Mogab, B. Y. Nguyen, B. E. White, Jr. S. Dakshina-Murthy, R. S. Rai, Z. X. Jiang, R. Martin, M. V. Raymond, M. Zavala, L. B. La, J. A. Smith, R. Garcia, D. Roan, M. Koeeke, and R. B. Gregory, "Physical and electrical properties of metal gate electrodes on HfO_2 gate dielectrics", *J. Vac. Sci. Technol. B*, vol. 21, pp. 11-17, 2003.

32. C. W. Yang, Y. K. Fang, C. H. Chen, S. F. Chen, C. Y. Lin, and C. S. Lin, M. F. Wang, Y. M. Lin, T. H. Hou, C. H. Chen, L. G. Yao, S. C. Chen, and M. S. Liang, "Effect of polycrystalline-silicon gate types on the opposite flatband voltage shift in n-type and p-type metal–oxide–semiconductor field-effect transistors for high-k HfO2 dielectric," *Appl. Phys. Lett.* vol. 83, pp. 308-310, 2003.

33. C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, and P. Tobin, "Fermi level pinning at the polySi/metal oxide interface," in *Symp. VLSI Tech. Dig.*, pp. 9-10, 2003.

34. K. Shiraishi, K. Yamada, K. Torii, Y. Akasaka, K. Nakajima, M. Kohno, T. Chikyo, H. Kitajima, and T. Arikado, "Physics in Fermi level pinning at the polySi/Hf-based high-k oxide interface," in *Symp. VLSI Tech. Dig.*, pp. 108-109, 2004.

35. I. S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H. J. Kim, D. Cha, J. Huang, M. J. Kim, B. E. Gnade, J. Kim, and R. M. Wallace, "A novel methodology on tuning work function of metal gate using stacking bi-metal layers," in *IEDM Tech. Dig.*, pp. 303-306, 2004.

36. J. K. Schaeffer, C. Capasso, L. R. C. Fonseca, S. Samavedam, D. C. Gilmer, Y. Liang, S. Kalpat, B. Adetutu, H. Tseng, Y. Shiho, A. Demkov, R. Hegde, W. J. Taylor, R. Gregory, J. Jiang, E. Luckowski, M. V. Raymond, K. Moore, D. Triyoso, D. Roan, B. E. White Jr., and P. J. Tobin, "Challenges for the integration of metal gate electrodes," in *IEDM Tech. Dig.*, pp. 287 - 290, 2004.

37. M. Terai, K. Takahashi, K. Manabe, T. Hase, T. Ogura, M. Saitoh, T. Iwamoto, T. Tatsumi, and H. Watanabe, "Highly reliable HfSiON CMOSFET with phase controlled NiSi (NFET) and Ni_3Si (PFET) FUSI gate electrode," in *VLSI Tech. Dig.*, pp. 68-69, 2005.

38. H. Y. Yu, C. Ren, Y.-C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M.-F. Li, D. S. H. Chan, and D.-L. Kwong, "Fermi pinning induced thermal instability of metal gate work functions," *IEEE Electron Device Lett.*, vol. 25, pp. 337-339, May 2004.

Chapter 4

Lanthanide-Incorporated Metal Nitrides for NMOS Applications

4.1 Introduction

Based on the understanding in Chapter 3, the metal-dielectric interface is a critical factor in determining the effective work function (EWF) of metal gates (MG). Since the nature of the extrinsic interface states tends to drive the effective work function of metals towards the mid-gap position of Si, adopting materials with work function (WF) values far below the conduction band E_c or far above the valence band E_v of Si as the gate electrodes would be a potential solution to achieve the required band-edge work function values after the rigorous thermal process.

As we know, most lanthanide series elements, including lanthanum (La), gadolinium (Gd), terbium (Tb), dysprosium (Dy), erbium (Er), ytterbium (Yb) etc., exhibit work function values of less than 4.0 eV, as summarized in Table 4.1. However, these metal materials could be quite reactive in the presence of oxygen or moisture so that they are difficult to handle and integrate into the gate-first CMOS process flow. On the other hand, refractory metal nitrides (MN_x), such as tantalum nitride (TaN), hafnium nitride (HfN) and titanium nitride (TiN), have been widely studied due to their relatively good thermal stability compared with other metal gate

candidates [1]-[3]. In addition, the metal nitride gate electrodes also show superior compatibility with high-κ gate dielectrics, enabling excellent EOT scalability of below 1 nm to be achievable [2]-[4]. These advantages make the metal nitrides a most important category among all the metal gates being developed. However, the limitation of MN_x materials is their mid-gap work functions, which leads to the V_{th} of the device too high to be acceptable for plenary bulk-Si devices. Therefore, it is intuitional to think of the possibilities of combining the advantages of lanthanide and MN_x metal gates together.

Table 4.1 Work functions of some lanthanide metals.

Metals	La	Gd	Tb	Dy	Er	Yb
Work Function (eV)	3.5[a]	3.17[a]	3.1[a]	3.25[b]	3.25[b]	2.6[c]

[a] taken from Ref. [5];
[b] taken from Ref. [6];
[c] taken from Ref. [7].

In this work, we systematically investigated a novel approach to tune the work function of MN_x gate electrodes (TaN, HfN, etc) by incorporating lanthanide series elements (Tb, Er, and Yb) into MN_x films, as illustrated in Fig. 4.1. The results show that the WF of lanthanide-incorporated metal nitrides (lanthanide-MN_x) can be successfully modulated continuously from mid-gap values down to 4.2 ~4.3 eV with good thermal stability achieved even after rapid thermal annealing (RTA) treatments at 1000 ^{o}C. The deposition methods, material and electrical properties, and the MOSFET characteristics of the novel lanthanide-MN_x metal gates will be presented and discussed in this chapter.

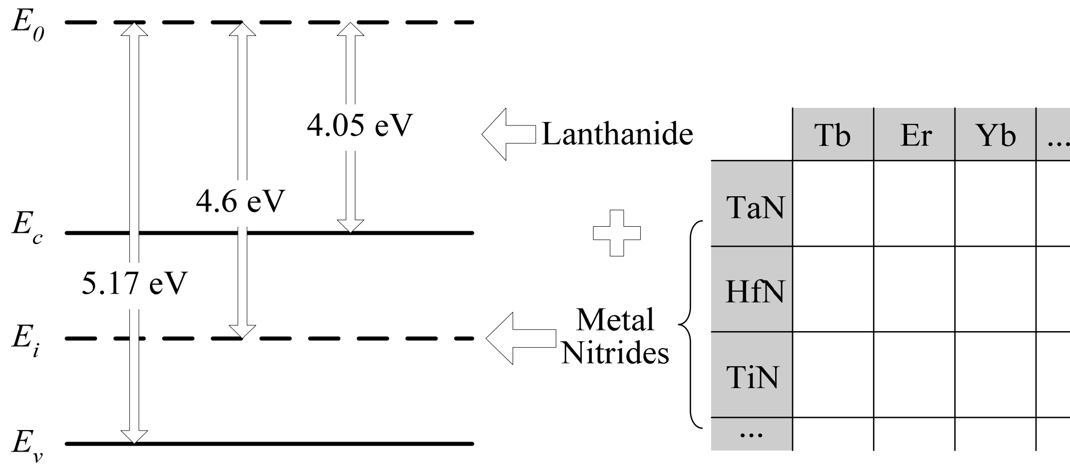


Fig. 4.1 Illustration of the idea to modulate the work function of metal nitrides by incorporating lanthanide series elements for n-MOSFET applications.

4.2 Experimental

Lanthanide-MN_x films were deposited by reactive co-sputtering of refractory metals (Ta, Hf, etc) with lanthanide metals (Tb, Er, Yb, etc) in an Ar + N_2 ambient using a PVD facility. The background pressure in the process chamber is about 3×10^{-7} Torr or less. For the deposition of refractory metal nitride TaN and HfN, a dc power of 450 W was applied on a 3-inch plenary Ta or Hf target with purity of 99.99 %, and the N_2/Ar flow rate was kept at 5/25 sccm. For the deposition of lanthanide-MN_x, a relatively long burn-in step (a dummy sputtering of lanthanides with a dc power of 150 W in pure Ar ambient for 10~15 min) was performed on the lanthanide targets prior to the co-sputtering process in order to remove the possible surface moisture or oxides on the lanthanide targets. During the co-sputtering, the power applied on the Ta (or Hf) target was kept at a constant value of 450 W, whereas that on the lanthanide Tb, Er, or Yb targets was varied from 70 W to 200W to modulate the lanthanide concentration in the films (Table 4.2). The N_2/Ar flow in these experiments was kept at 5/25 sccm unless otherwise stated. In some experimental splits, the N_2/Ar flow rate ratio was varied intentionally from around

3/25 to 8/25 in sccm to evaluate the impact of N concentration on the properties of lanthanide-MN_x. The deposition conditions and the compositions of the films investigated are summarized in Table 4.2 and Table 4.3.

Table 4.2 Experimental splits and the compositions for the Lanthanide-MN_x films.

No.	**MN_x**	**Lan.**	**Power on Ta or Hf (watt)**	**Power on Lantha. (watt)**	**N_2/Ar (sccm)**	**Composition (by XPS)**	**Composition (by RBS)**
1	TaN	N.A.	450	0	5/25	TaN	$Ta_{1.05}N_{0.95}$
2		Tb	450	70	5/25	$Ta_{0.97}Tb_{0.03}N_y$	N.A.
3			450	100	5/25	$Ta_{0.94}Tb_{0.06}N_y$	N.A.
4			450	150	5/25	$Ta_{0.9}Tb_{0.1}N_y$	$Ta_{0.92}Tb_{0.08}N_{1.0}$
5			450	200	5/25	$Ta_{0.87}Tb_{0.13}N_y$	N.A.
6		Er	450	100	5/25	$Ta_{0.97}Er_{0.03}N_y$	N.A.
7			450	150	5/25	$Ta_{0.95}Er_{0.05}N_y$	N.A.
8		Yb	450	60	5/25	$Ta_{0.97}Yb_{0.03}N_y$	N.A.
9	HfN	N.A.	450	0	5/25	HfN	N.A.
10		Tb	450	100	5/25	$Hf_{0.89}Tb_{0.11}N_y$	N.A.
11			450	150	5/25	$Hf_{0.8}Tb_{0.2}N_y$	N.A.

Table 4.3 Experimental splits and the compositions for $Ta_{0.9}Tb_{0.1}N_y$ with different N_2 flow rates during reactive sputtering deposition.

No.	**Lanthanide-MN_x**	**Power on Ta or Hf (watt)**	**Power on Lantha. (watt)**	**N_2/Ar (sccm)**	**N/(Ta+Tb) (by RBS)**
1	$Ta_{0.9}Tb_{0.1}N_y$	450	150	3/25	0.96
2	$Ta_{0.9}Tb_{0.1}N_y$	450	150	5/25	1.0
3	$Ta_{0.9}Tb_{0.1}N_y$	450	150	8/25	1.06

MOS capacitors and n-MOSFETs with lanthanide-MN_x gates and SiO_2 or HfAlO gate dielectrics were fabricated on p-Si (100) substrates (6-9 Ω·cm). For MOS capacitor fabrication, $Ta_{1-x}Tb_xN_y$, $Ta_{1-x}Er_xN_y$, $Ta_{1-x}Yb_xN_y$ and $Hf_{1-x}Tb_xN_y$ metal gates with thicknesses of about 500 Å were deposited using the method described above, followed by the deposition of an *in-situ* TaN capping layer with a thickness of 1000 Å on top of all the lanthanide-MN_x films in order to reduce the sheet resistance of the whole gate electrode stacks and prevent the oxidation of lanthanide-MN_x surface. After gate patterning, some of the capacitors were subjected to RTA treatment from 800°C to 1000°C in N_2 ambient to evaluate the thermal stability. For n-MOSFETs fabrication, a damascene gate process was used to pattern the $Ta_{0.9}Tb_{0.1}N_y$/TaN stack, which will be discussed in Section 4.6. The thickness of $Ta_{0.9}Tb_{0.1}N_y$ was reduced to 100 Å to minimize the difficulties in gate patterning. The As implantation with a dose of 4×10^{15} was performed followed by a RTA at 1000 ^{o}C for 5 sec to form the source/drain (S/D) regions of the MOSFET. NiSi salicide was adopted to reduce the S/D parasitic resistance. Finally, all the samples were subjected to a backside Al metallization and a forming gas annealing (FGA) at 420 ^{o}C for 30 min.

Capacitance-voltage (*C-V*) and current-voltage (*I-V*) characteristics were measured using a HP4284A LCR meter and HP4156A semiconductor parameter analyzer, respectively, on MOS capacitors with area of 100×100 μm^2. EOT and flat-band voltage (V_{FB}) were obtained by fitting the *C-V* measurements with the theoretical *C-V* curves, which take the quantum mechanical effect into account. Auger electron spectroscopy (AES), X-ray photoelectron spectroscopy (XPS), Rutherford backscattering spectrometry (RBS), X-ray diffraction (XRD) and high-resolution transmission electron microscopy (HRTEM) were also performed on blanket films deposited on thin (2~3 nm) SiO_2 dielectric for material characterizations.

4.3 Material Characteristics of Lanthanide-MN_x

4.3.1 Composition Analysis

The compositions of the films investigated were analyzed by various techniques including XPS, AES, and RBS. For $Ta_{1-x}Tb_xN_y$, the Tb/ (Ta+Tb) ratios are determined by XPS analysis according to the Ta 4*f* and Tb 4*d* peaks, as summarized in Table 4.2. It is found that the Tb concentration in $Ta_{1-x}Tb_xN_y$ exhibits a linear dependence on the sputtering power applied on the Tb target, as depicted in Fig. 4.2. This suggests good controllability of lanthanide concentration by adjusting the sputtering power. The Tb/ (Ta+Tb) ratios in some $Ta_{1-x}Tb_xN_y$ films were also measured by the RBS technique, as indicated in Table 4.2. For the most frequently used deposition condition, the Tb/(Ta+Tb) fraction determined by RBS is about 8 %, which is quite consistent with the value of 10% measured by XPS after taking the possible experiment errors induced by different characterization techniques into account.

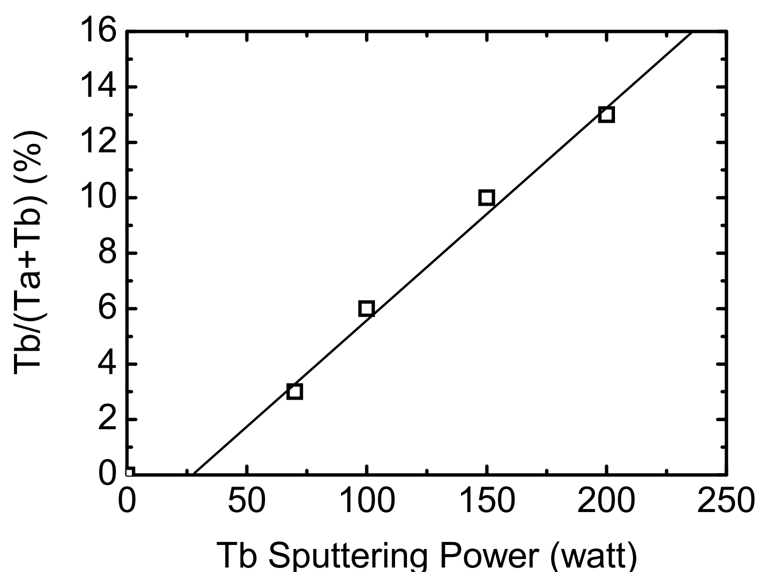


Fig. 4.2 Relationship between Tb concentration in $Ta_{1-x}Tb_xN_y$ and the sputtering power applied on the Tb target during co-sputtering deposition.

However, the nitrogen concentration is relatively difficult to determine very accurately. For XPS analysis, the N 1*s* peak partially overlaps with Ta 4*p*3 peak, which implies that curve fitting of the measured spectrum is necessary to differentiate the contributions from the two elements. However, it should be noted that the Ta/N ratio determined from Ta 4*p*3 and N 1*s* peak should be same with that determined from the Ta 4*f* and N 1*s* peak theoretically. In practice, unfortunately, the best curve fitting may not fulfill the later constraint very well, resulting in some uncertainty in determining the N concentration. For the RBS measurement, the channel of N spectrum is masked by the background signal from the Si substrate because the atomic number of N is smaller than that of Si [8]. Moreover, the RBS yield of the N element is relatively low because of the small scattering cross section of N atoms, leading to the low signal detected from N, as shown in Fig. 4.3. As a result, accurate determination of N concentration from RBS analysis could also be difficult.

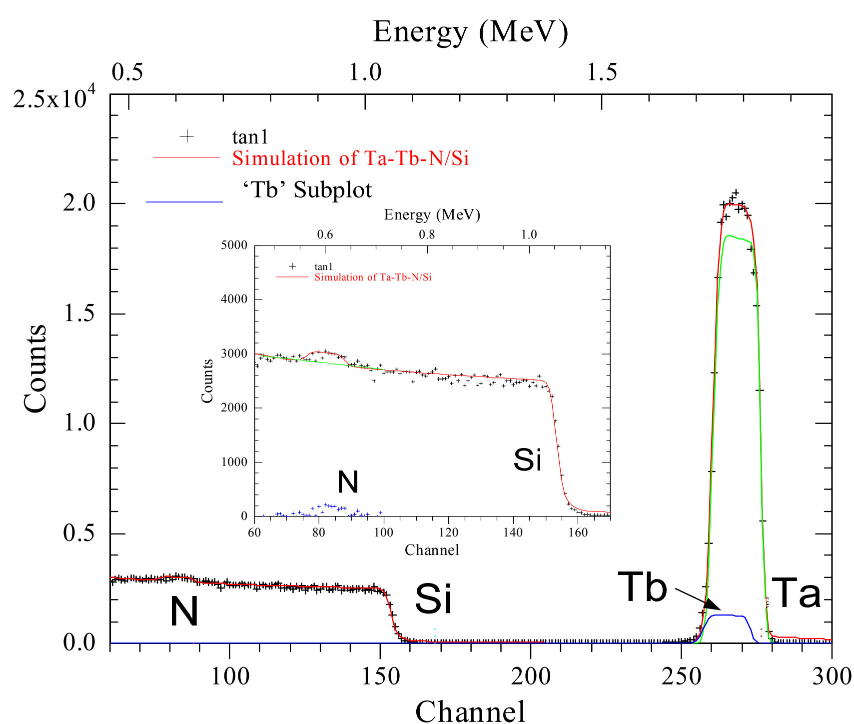


Fig. 4.3 RBS spectrum of $Ta_{0.92}Tb_{0.08}N_{1.0}$ film, where the concentration of each species are determined from the simulation by XRUMP [9].

Although it is difficult to predict the N concentration very accurately, the general trend of N concentration in the different experimental splits can still be observed. For the TaN control sample, the N % determined from XPS, RBS, and AES analysis is about 38%, 47%, and 50%, respectively. The RBS and AES measurements give quite close values for N content. On the other hand, for the $Ta_{0.9}Tb_{0.1}N_y$ film, the measured N % values are about 43%, 50%, and 55%, respectively, by XPS, RBS, and AES. Therefore, it can be concluded that the N concentration in $Ta_{0.9}Tb_{0.1}N_y$ is higher than that in TaN. Indeed, an increase of N concentration can be observed while increasing the lanthanide concentration in the lanthanide-MN_x films.

In the following discussion, the compositions determined by XPS will be used as a standard metrology. The actual N fraction in lanthanide-MN_x will not be explicitly labeled to avoid any possible confusion. Instead, where needed, the N_2/Ar flow rates during the sputtering deposition will be used as an indication of the relative level of N content. The default N_2/Ar flow rates will be 5/25 sccm unless otherwise stated.

4.3.2 Auger Electron Spectroscopy (AES) Study

The depth profiles of the TaN/lanthanide-MN_x gate stacks were studied by AES analysis. Fig. 4.4 presents the AES profiles of the TaN/$Ta_{0.94}Tb_{0.06}N_y$/SiO_2 and TaN/$Ta_{0.95}Er_{0.05}N_y$/SiO_2 gate stacks before and after 1000 °C RTA. The spectra and the relative sensitivity factor (RSF) for Ta, N, O, and Si elements are obtained from the standard reference book [10], while that of Tb, Er, and Yb elements are measured from the pure metal samples deposited in this experiment. The RSF for these lanthanide elements are relatively low, rendering the noise level of Tb or Er to be

relatively higher compared with O. From the depth profiles shown in Fig. 4.4, it can be observed firstly that the Ta, Tb (Er), and N profiles appear to be very uniform across the TaN and $Ta_{0.94}Tb_{0.06}N_y$ ($Ta_{0.95}Er_{0.05}N_y$) films, without notable changes before and after the 1000 °C RTA. Secondly, at the interface between the undoped TaN and the $Ta_{0.94}Tb_{0.06}N_y$ ($Ta_{0.95}Er_{0.05}N_y$), there is a clear change of the N percentage. The nitrogen concentration in $Ta_{0.94}Tb_{0.06}N_y$ (or $Ta_{0.95}Er_{0.05}N_y$) is observed to be remarkably higher than that in the TaN film, and keeps stable even after 1000 °C RTA treatment. Note that the N concentrations in other lanthanide-MN_x films also tend to increase with the lanthanide concentration, as judged from both AES and XPS studies.

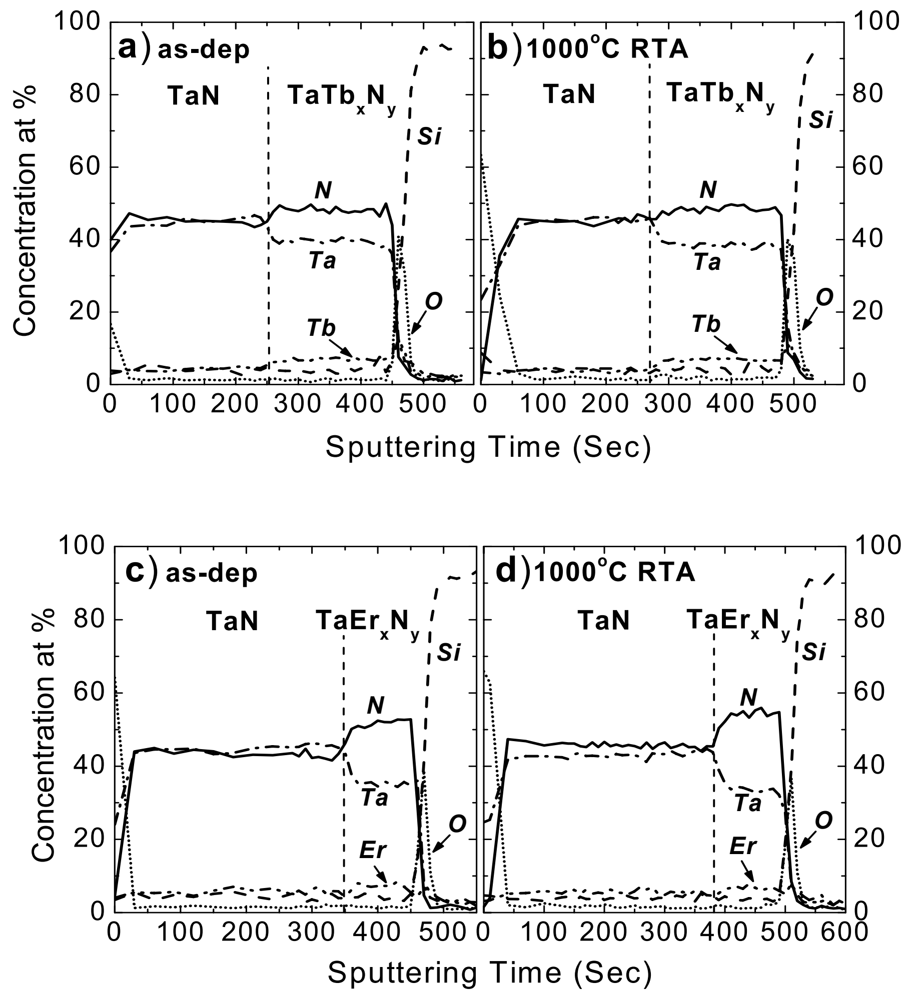


Fig. 4.4 AES depth profiling for $TaN/Ta_{0.94}Tb_{0.06}N_y/SiO_2$ (a-b) and $TaN/Ta_{0.95}Er_{0.05}N_y/SiO_2$ (c-d) gate stacks before and after 1000 °C RTA in N_2 ambient.

Since the N_2/Ar flow rates, the process pressure and the sputtering power applied on the Ta target were all kept constant during the deposition of $Ta_{0.94}Tb_{0.06}N_y$ (or $Ta_{0.95}Er_{0.05}N_y$) and TaN films, this abrupt change of nitrogen concentration implies that the presence of Tb (Er) makes it possible to incorporate more N into the $Ta_{1-x}Tb_xN_y$ (or $Ta_{1-x}Er_xN_y$) films. The low work functions of lanthanide series elements and the relatively large electronegativity differences to nitrogen could be responsible for the affinity of lanthanide elements to nitrogen. The enhanced N concentration could be of importance to the properties of lanthanide-MN_x, especially its thermal stability.

4.3.3 X-ray Photoelectron Spectroscopy (XPS) Study

XPS analysis was also performed to study the binding energy and composition of the lanthanide-MN_x films investigated in this experiment. Fig. 4.5 shows the core level spectra of N 1*s*, Ta 4*f*, and Tb 4*d* regions for the as-deposited $Ta_{1-x}Tb_xN_y$ films with different Tb/(Ta+Tb) ratios. The ratio of Tb/(Ta+Tb) was obtained by XPS analysis from the Ta 4*f* and Tb 4*d* spectra, and summarized in Table 4.2. As shown in Fig. 4.5 (a), the binding energy of the N 1*s* peak shifts gradually from 498.7 eV to about 487.5 eV as the Tb concentration increases, while that of the Ta 4*p*3/2 peak remains almost unchanged. This implies that N may form new bonds with the incorporated Tb, either in the form of Tb-N or Ta-N-Tb compound, as the Tb concentration in $Ta_{1-x}Tb_xN_y$ increases. In Fig. 4.5 (b), the Ta 4*f*7/2 peak in TaN shows a binding energy of 23.0 eV, which is higher than the value of 21.9 eV for pure Ta, suggesting the existence of Ta-N bonds [11]. As the Tb concentration increases, the Ta 4*f* peak shows only a 0.3-eV-shift which could be attributed to the formation of

Ta-N-Tb compound. The signals from the Tb 4*d* peaks in Fig. 4.5 (c) appear noisy due to the low Tb content.

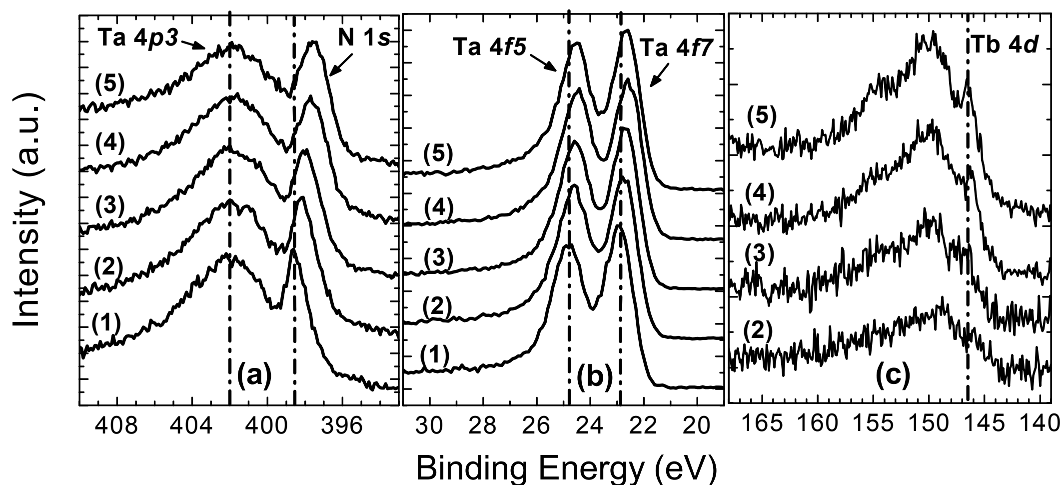
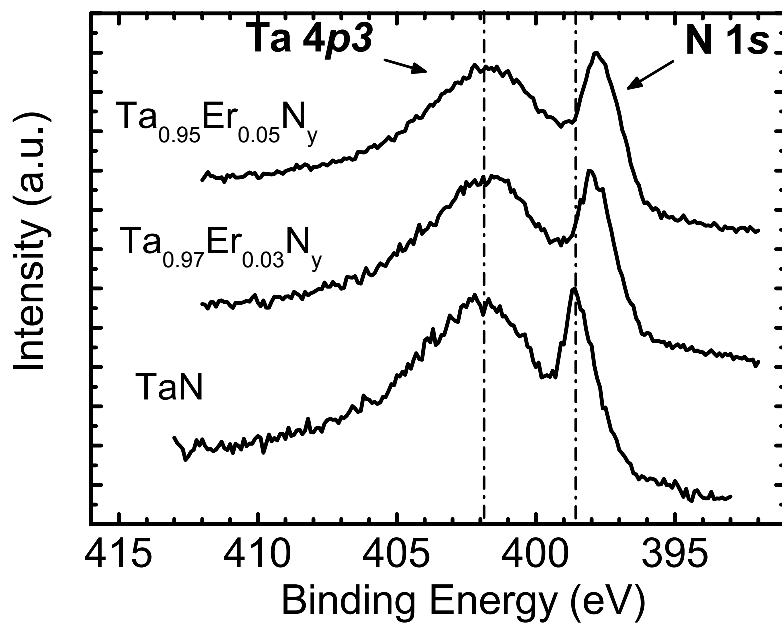
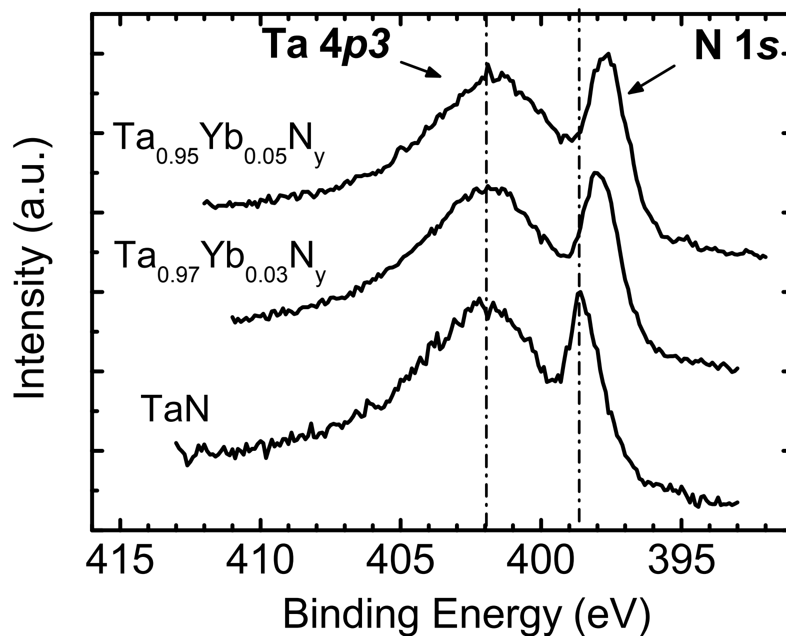


Fig. 4.5 The XPS spectra of the (a) N 1*s*, (b) Ta 4*f*, and (c) Tb 4*d* region for the as-deposited $Ta_{1-x}Tb_xN_y$ films with different Tb concentrations: (1) TaN; (2) $Ta_{0.97}Tb_{0.03}N_y$; (3) $Ta_{0.94}Tb_{0.06}N_y$; (4) $Ta_{0.9}Tb_{0.1}N_y$; (5) $Ta_{0.87}Tb_{0.13}N_y$.

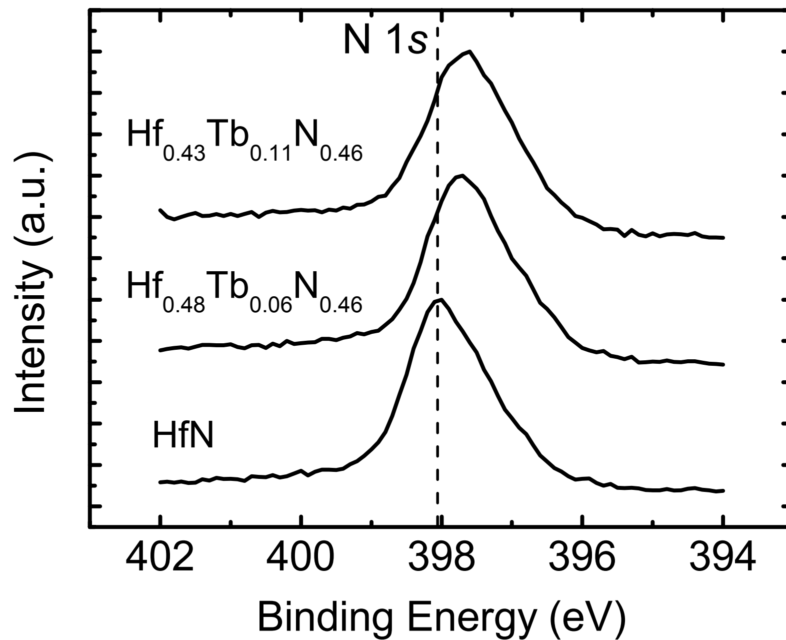
The similar shift of N 1*s* peak as a function of nitrogen concentration can also be observed in $Ta_{1-x}Er_xN_y$, $Ta_{1-x}Yb_xN_y$ and $Hf_{1-x}Tb_xN_y$ materials, as depicted in Fig. 4.6. These results suggest that the nitrogen may form some complex compounds with lanthanide, either in the form of lanthanide-N or Ta (Hf)-N-lanthanide, as the lanthanide elements are introduced into the binary Ta (Hf)-N system. We believe that the formation of these compounds could be helpful for improving the stability of the lanthanide-MN_x since the lanthanide metals themselves are quite reactive. The role of nitrogen in lanthanide-MN_x will be further discussed in Section 4.5.



(a)



(b)



(c)

Fig. 4.6 The XPS core level spectra in the N 1*s* region for the as-deposited (a) $Ta_{1-x}Er_xN_y$, (b) $Ta_{1-x}Yb_xN_y$, and (c) $Hf_{1-x}Tb_xN_y$ films with different lanthanide concentrations.

4.3.4 X-ray Diffraction (XRD) Study

The film morphology of lanthanide-MN_x was studied by XRD analysis. Fig. 4.7 shows the XRD spectra of $Ta_{1-x}Tb_xN_y$ films with different Tb contents, compared with that of TaN. We can observe that TaN exhibits crystalline structure even in the as-deposited condition. Moreover, a peak shift can be clearly observed in the spectra of TaN after 1000 ^{o}C RTA treatment, suggesting the phase of TaN films is not stable and some re-crystallization or condensation effect happens during the high-temperature annealing process. On the other hand, a broad diffraction peak can be found in the XRD spectrum of $Ta_{1-x}Tb_xN_y$ when the Tb/(Ta+Tb) ratio is higher than 10 %. This indicates that the crystallization of TaN can be effectively suppressed by the Tb additive, and very fine grains or amorphous morphology can be achieved.

More importantly, the amorphous structure can be maintained up to 1000°C, which is adequate for the S/D dopant activation in gate-first process.

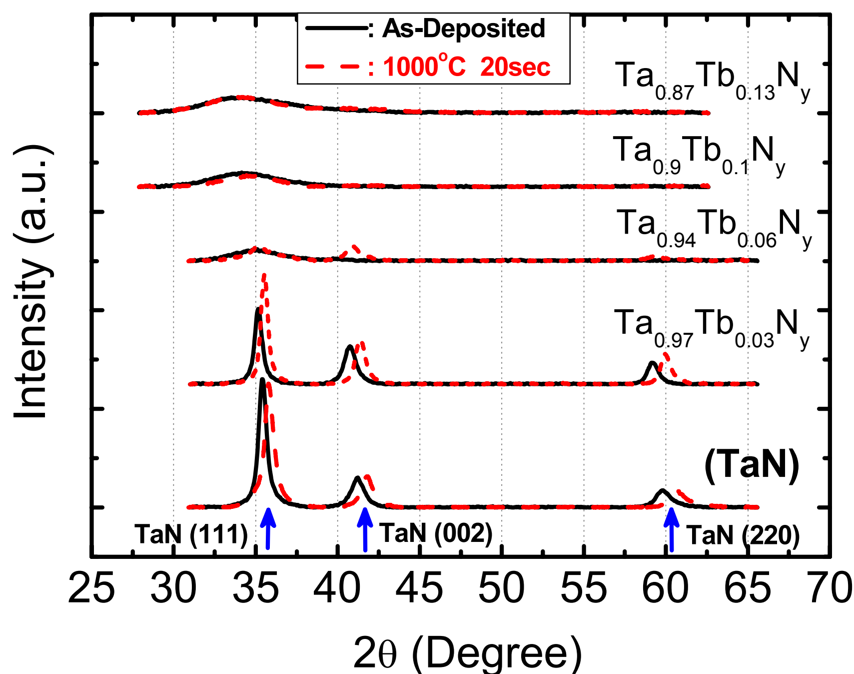


Fig. 4.7 XRD spectrums of $Ta_{1-x}Tb_xN_y$ materials with different Tb concentrations before and after 1000 ^{o}C RTA anneal for 20 sec, compared with that of TaN.

For a metal gate electrode, amorphous morphology as well as good phase stability against the high thermal budget used in the CMOS process will be desirable since it will result in less stress and uniformity problems for the large-scale IC fabrication [12]. The phase stability of $Ta_{1-x}Tb_xN_y$ was also investigated as a function of N concentration, as shown in Fig. 4.8. $Ta_{0.9}Tb_{0.1}N_y$ gate electrodes with different N content were prepared by varying the N_2/Ar flow rates (N_2/Ar = 3/25, 5/25, and 8/25 sccm) during sputtering, as summarized in Table 4.3. According to the XRD spectra, the crystallization of TaN can be significantly suppressed by adding Tb into TaN. However, it is observed in the sample with the highest-N-split (N_2/Ar = 8/25) that a peak at TaN (200) position tends to increase after the high temperature RTA

process, implying that some re-crystallization may happen at high temperature. Therefore the N concentration needs to be well engineered to get a stable amorphous morphology in $Ta_{1-x}Tb_xN_y$.

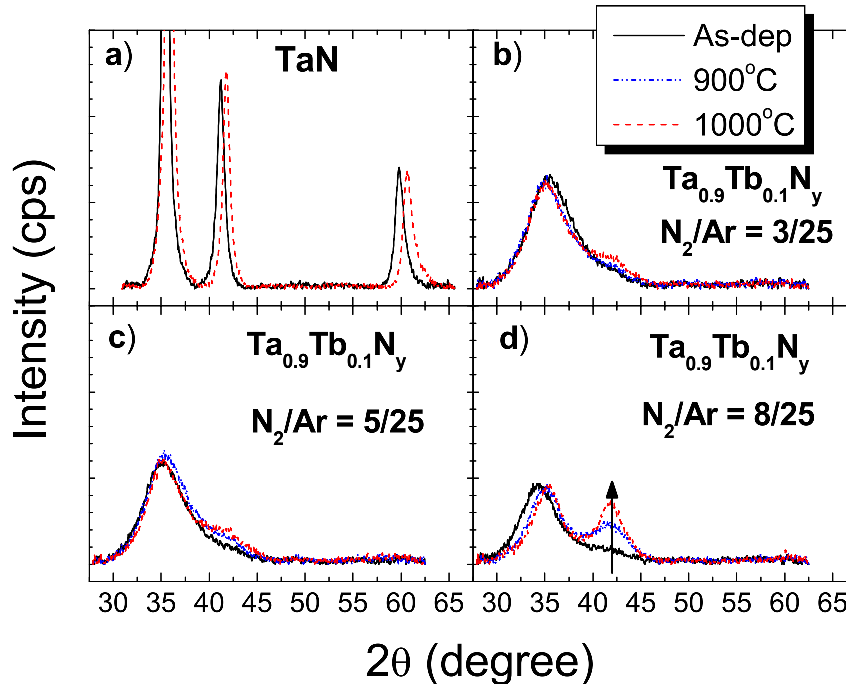


Fig. 4.8 XRD spectrums of $Ta_{0.9}Tb_{0.1}N_y$ films with different N concentrations measured in the following conditions: as-deposited, after 900 °C RTA and after 1000 °C RTA.

4.3.5 Resistivity

The resistivity of lanthanide-MN_x was investigated as a function of N and lanthanide contents, respectively. We first discuss the dependence of the resistivity on the N concentration in lanthanide-MN_x. Fig. 4.9 compares the resistivity of the $Ta_{0.9}Tb_{0.1}N_y$ films with different nitrogen flow rates during deposition (see Table 4.3). When the nitrogen flow increases from 3 sccm to 8 sccm, the resistivity of $Ta_{0.9}Tb_{0.1}N_y$ increases from 416 μΩ·cm to 965 μΩ·cm for the as-deposited films. An increase of resistivity with N concentration can also be observed in the transition MN_x

materials such as TaN, TiN and HfN [13]. Microstructure change and the formation of some poorly conducting phases in the MN_x film could be one of the reasons [14]. Besides this, the formation of a Tb-N compound, which probably behaves as a narrow gap insulator [15], may also account for the increased resistivity as a function of N content in the case of $Ta_{1-x}Tb_xN_y$. The higher resistivity values after the high temperature (900 °C, 1000 °C) RTA process could be correlated with the oxidation of the metal surface during the RTA process. This is due to the existence of oxygen residues in the process chamber, and can be suppressed by capping the lanthanide-MN_x by TaN or other electrodes which are difficult to be oxidized.

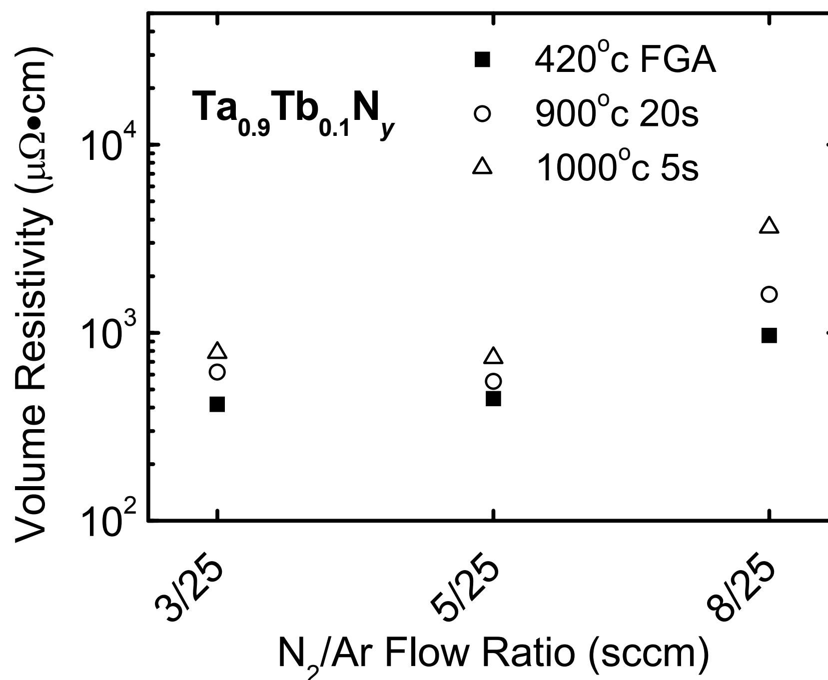


Fig. 4.9 Resistivity of $Ta_{0.9}Tb_{0.1}N_y$ films as a function of N_2/Ar flow rate ratio during the sputtering, with and without RTA performed.

The dependence of the film resistivity on the lanthanide concentration in lanthanide-MN_x materials was also studied. Fig. 4.10 summarizes dependence of the

resistivity of lanthanide-TaN films on the lanthanide type and concentration. As shown in Fig. 4.10, the resistivity of lanthanide-TaN is higher than that of TaN and tends to increase with the lanthanide concentration. This could be correlated with the N concentration in the lanthanide-TaN films which is higher than that in TaN, as observed from the AES analysis in Section 4.3.2. Again, an increase of the resistivity with annealing temperature can be observed in Fig. 4.10, which could be correlated with the oxygen traces in the N_2 ambient during the RTA process. A capping layer with very low resistivity and high immunity to oxidation would be helpful in achieving the low sheet resistance required by *ITRS*.

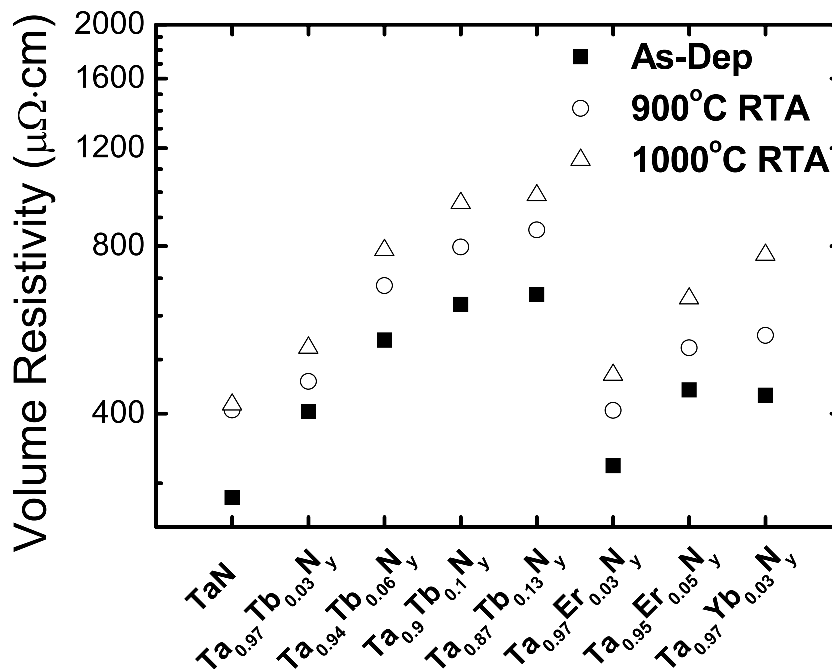


Fig. 4.10 Resistivity of lanthanide-incorporated TaN materials as a function of lanthanide type and concentration before and after RTA treatments in N_2 ambient.

4.4 Work Function Tunability

The work function tunability of lanthanide-MN_x was studied by varying the lanthanide content in lanthanide-MN_x. Fig. 4.11 shows the high-frequency *C-V* (HFCV) measurement of the $Ta_{1-x}Tb_xN_y$ gated MOS capacitors with varied Tb concentrations in $Ta_{1-x}Tb_xN_y$ (corresponding to Table 4.2). Compared with that of TaN, a parallel shift in the *C-V* curves of $Ta_{1-x}Tb_xN_y/SiO_2$ MOS capacitors is observed, indicating continuous work function tuning by varying the Tb concentrations. Fig. 4.12 compares the *C-V* measurement of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ MOS capacitors after 420 ^oC FGA for 30 min and 1000 ^oC RTA for 20 sec. The *C-V* measurements fit well with the simulated *C-V*, implying good interface quality. A shift in V_{FB} for about + 0.15 V is observed after 1000 ^oC RTA treatment, possibly due to the Fermi-level pinning effect similar with that in TaN/SiO_2 stack as discussed in Chapter 3 [16].

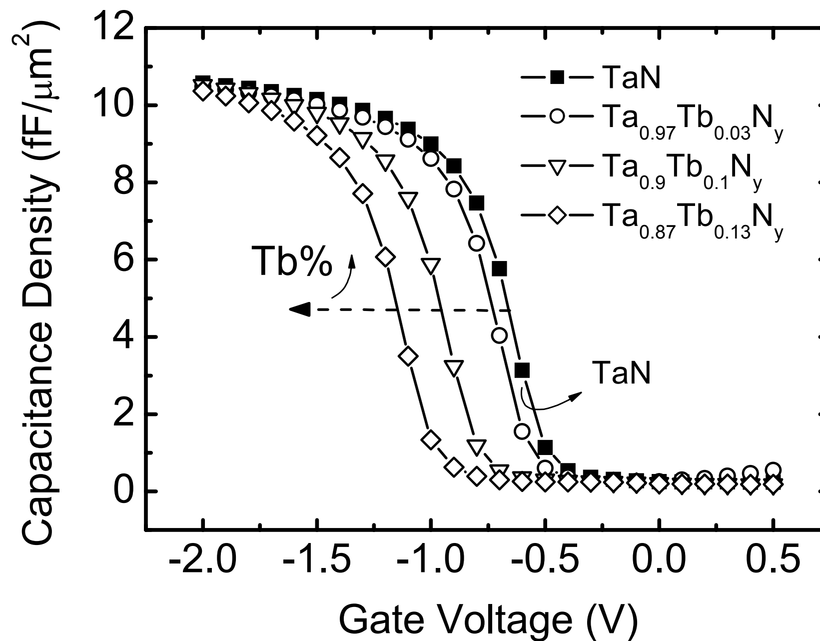


Fig. 4.11 High-frequency *C-V* characteristics (100 kHz) of $Ta_{1-x}Tb_xN_y$ gated MOS capacitors with different Tb concentration in $Ta_{1-x}Tb_xN_y$ on SiO_2. The measurements are taken after a 420°C forming gas anneal.

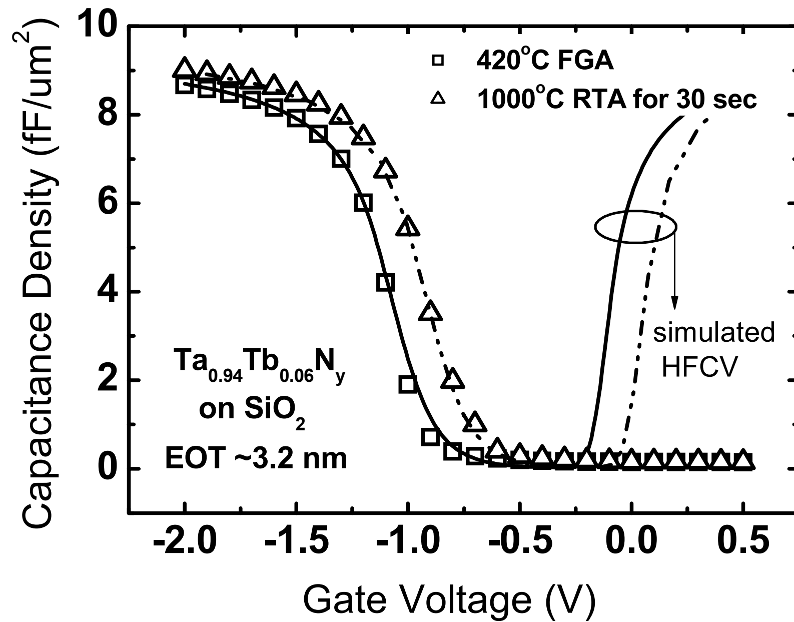


Fig. 4.12 High-frequency C-V characteristics (100 kHz) of MOS-capacitors with $Ta_{0.94}Tb_{0.06}N_y$ gate electrode after 420 °C FGA and after 1000 °C RTA. The lines show the simulated *C-V* curves which takes quantum mechanical effect into account.

The work function of all the experimental splits investigated in this work were extracted from the V_{FB} versus EOT plots which exclude the contribution of oxide fixed charge to the V_{FB} of MOS capacitor such that accurate WF can be obtained. Fig. 4.13 compares the V_{FB} versus EOT plots of $Ta_{0.94}Tb_{0.06}N_y$, $Ta_{0.95}Er_{0.05}N_y$, and $Hf_{0.8}Tb_{0.2}N_y$ metal gates with that of TaN and HfN on SiO_2 dielectric after 420 °C FGA and 1000 °C RTA processes. The V_{FB} vs. EOT plots for these gate stacks show similar slopes after different post-metal-annealing (PMA) treatments, indicating the gate oxide quality is not affected much by the PMA processes. It is observed clearly that the WF of MN_x is lowered by adding lanthanide into MN_x metal gates. The WF values of $Ta_{0.94}Tb_{0.06}N_y$, $Ta_{0.95}Er_{0.05}N_y$, and $Hf_{0.8}Tb_{0.2}N_y$ metal gates are determined to be 4.08 eV, 4.17 eV, and 4.23 eV after 420 °C FGA, and they increase to 4.23 eV, 4.3 eV, and 4.31 eV, respectively, after 1000 °C RTA treatment. The WF increase after

the high temperature anneal process could be due the Fermi pinning effect resulting from some extrinsic defect states generated during the high temperature anneal process at the metal gate/SiO_2 interface [17].

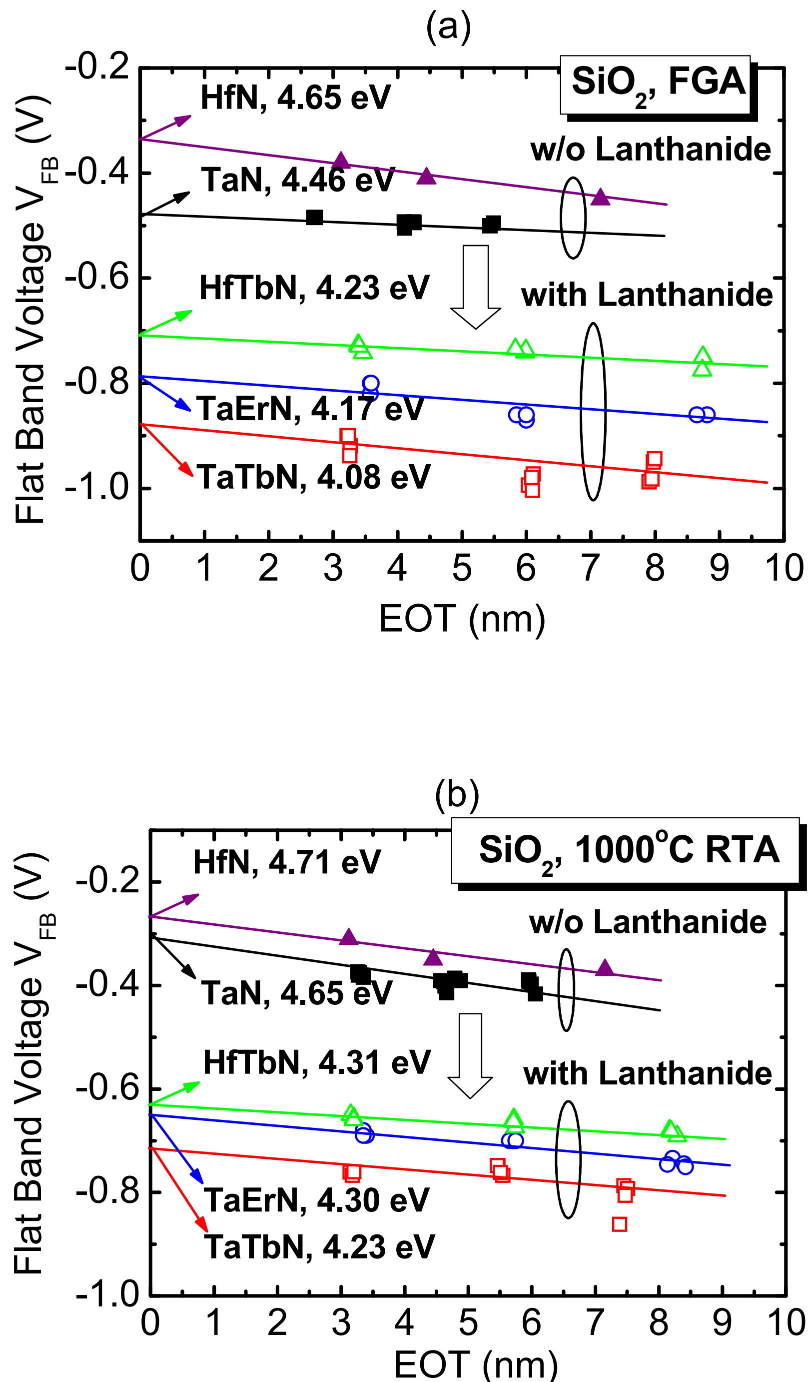


Fig. 4.13 V_{FB} vs. EOT plots of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$, $Ta_{0.95}Er_{0.05}N_y/SiO_2$, and $Hf_{0.8}Tb_{0.2}N_y/SiO_2$ gate stacks (a) after 420 °C FGA and (b) after 1000 °C RTA treatment, compared with that of TaN and HfN.

Fig. 4.14 summarizes the WF values of some lanthanide-MN_x as functions of lanthanide type and concentration after different PMA treatments. It shows clearly that WF of MN_x, such as TaN and HfN, can be tuned continuously from mid-gap values down to 4.2~4.3 eV by incorporating a small amount of lanthanide elements into these MN_x. It is believed that the presence of lanthanide at the metal gate-dielectric interface is responsible for drawing the effective work function of MN_x down to a lower value, since lanthanide elements generally possess very low WF values (see Table 4.1). Note that the metal gate work function of 4.2~4.3 eV is achievable even after 1000 °C RTA treatment, which is promising as N-type metal gate candidates in the gate-first bulk-Si CMOS process. Further tuning of the work function is still possible by increasing the concentration of lanthanide elements.

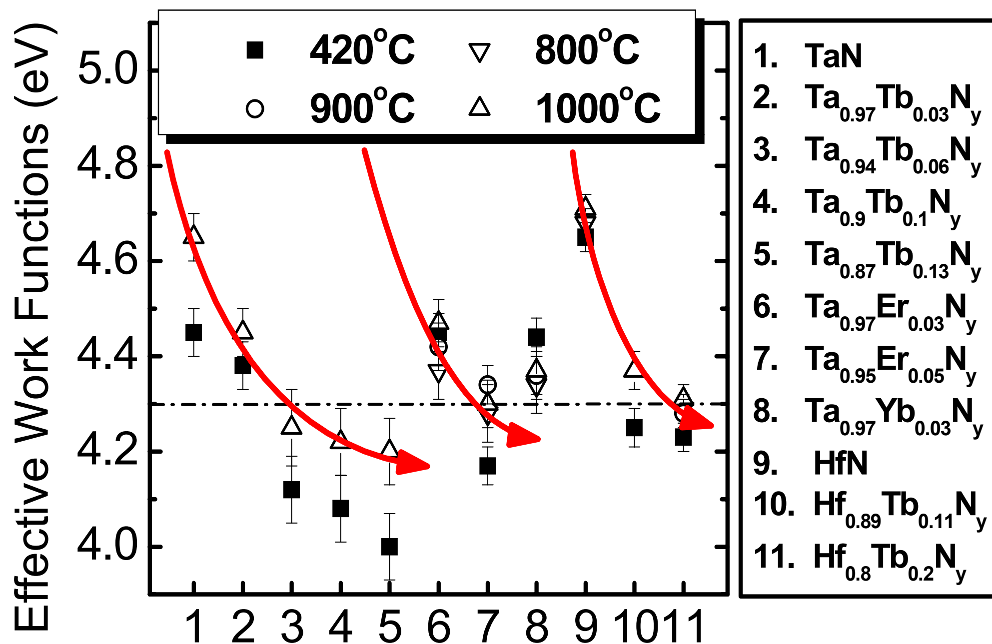


Fig. 4.14 Work function values of some MN_x and lanthanide-MN_x gate electrodes as a function of lanthanide type and concentrations under different annealing conditions, showing the tunability of MN_x work functions by incorporating lanthanide.

Table 4.4 summarizes the barrier heights measured from several lanthanide-MN_x/SiO_2 stacks as well as the corresponding work function values. The work-function values are obtained from V_{FB} versus EOT plots. The barrier height values are extracted by Fowler-Nordheim current analysis on capacitors with ~6 nm SiO_2 dielectrics. The effective mass of 0.4 (m_{ox}/m_0) was chosen for the tunneling electrons [18]. Most barrier height values correlate well with the extracted work function values, suggesting the work function of MN_x has indeed been modulated by the lanthanide.

Table 4.4 Work function and barrier height of lanthanide-incorporated TaN on SiO_2 as a function of RTA temperatures.

Gate electrode	Work function (eV)				Barrier height (eV)			
	420 °C	800 °C	900 °C	1000 °C	420°C	800 °C	900 °C	1000 °C
TaN	4.4		4.5	4.7	3.44		3.5	3.73
$Ta_{0.94}Tb_{0.06}N_y$	4.08		4.15	4.23	3.12		3.25	3.36
$Ta_{0.95}Er_{0.05}N_y$	4.17	4.28	4.34	4.30	3.14	3.23	3.32	3.29
$Ta_{0.97}Yb_{0.03}N_y$	4.44	4.34	4.36	4.37	3.40	3.25	3.41	3.05

The work function tunability of lanthanide-MN_x on high-κ dielectric was also investigated. Fig. 4.15 shows the *C-V* characteristics of $Ta_{0.9}Tb_{0.1}N_y$ gated capacitors on HfAlO dielectric before and after 1000 °C RTA process. The *C-V* measurements fit well with the *C-V* curves produced by simulation, suggesting good interface quality between HfAlO and Si substrate. Compared with the TaN/HfAlO stack, the $Ta_{0.9}Tb_{0.1}N_y$/HfAlO gate stack shows a negative shift in V_{FB} by 0.2~0.3 V irrespective of the PMA temperature, which could be attributed to the lower WF of $Ta_{0.9}Tb_{0.1}N_y$ compared with that of TaN. Considering the EWF value of TaN on HfAlO (~ 4.6 eV

on HfAlO in our previous work; ~ 4.5 ± 0.1 eV on ALD HfAlO in Ref. [19]), the EWF of $Ta_{0.9}Tb_{0.1}N_y$ on HfAlO could be around ~ 4.3 eV. The Q_f difference between the two gate stacks, if any, is not likely to take the major role on the V_{FB} difference observed since the high-κ films used in the two gate stacks are identical. Further study will be needed to study the work function tunability of lanthanide-MN_x on high-κ dielectrics in more detail.

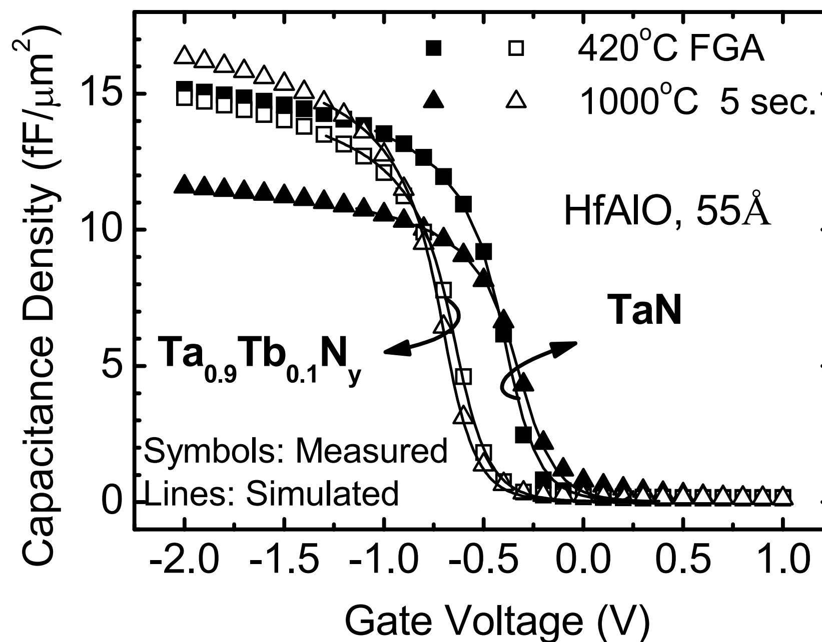


Fig. 4.15 *C-V* characteristics of TaN and $Ta_{0.9}Tb_{0.1}N_y$ metal gates on ALD HfAlO dielectrics after FGA at 420 °C for 30 min. and RTA at 1000 °C for 5 sec.

4.5 Thermal Stability Study

Besides appropriate work function values, another important consideration in choosing a suitable metal gate candidate is the thermal stability of the metal material against the high-thermal budget used in the gate-first CMOS process. Therefore, another effort in this chapter is to study the thermal stability of the lanthanide-MN_x metal gates and its dependence.

Fig. 4.16 shows the XTEM pictures of a $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack after 420 ^oC FGA, 900 ^oC RTA and 1000 ^oC RTA. No significant change has been observed in the thickness of SiO_2, indicating the good stability of $Ta_{0.94}Tb_{0.06}N_y$ on SiO_2. This is also consistent with the *C-V* measurements shown in Fig. 4.12 where the EOT of the $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ stack does not show much change with and without 1000 ^oC RTA process.

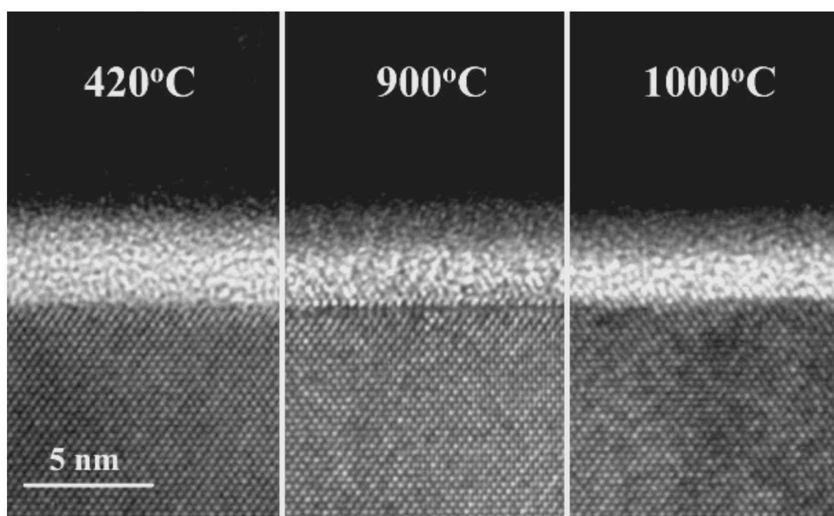


Fig. 4.16 XTEM images of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack on (100) Si substrate after 420 ^oC FGA for 30 min, 900 ^oC RTA for 30 sec and 1000 ^oC RTA for 30 sec.

Fig. 4.17 compares the EOT stability of the $Ta_{1-x}Tb_xN_y/SiO_2$ gate stacks with various Tb content after a 1000 ^oC RTA treatment. As shown in Fig. 4.17, TaN/SiO_2 shows a significant increase in EOT for about 0.6 nm after 1000 ^oC RTA. This increase is believed due to the poor oxygen diffusion barrier properties of TaN, allowing the oxygen residues to penetrate through the film. However, the EOT variation observed in the $Ta_{1-x}Tb_xN_y/SiO_2$ stacks are much smaller than that in the TaN/SiO_2 stack, and it tends to decrease with the concentration of Tb in $Ta_{1-x}Tb_xN_y$. Similar results can also be observed in other lanthanide-MN_x-gated MOS devices.

The improved EOT stability in lanthanide-MN_x-gated MOS capacitors implies that lanthanide-MN_x is a good diffusion barrier to O_2, which could be related to the amorphous structure of the $Ta_{1-x}Tb_xN_y$ film as discussed in the section 4.3.4. An EOT reduction by 1~2 Å was also observed in Fig. 4.17 when the Tb concentration becomes high. This could be due to some reactions of the excess Tb with SiO_2 at the $Ta_{1-x}Tb_xN_y/SiO_2$ interface, which can be improved by increasing the N content in the film. The role of N to the thermal/chemical stability of lanthanide-MN_x will be discussed later in this section.

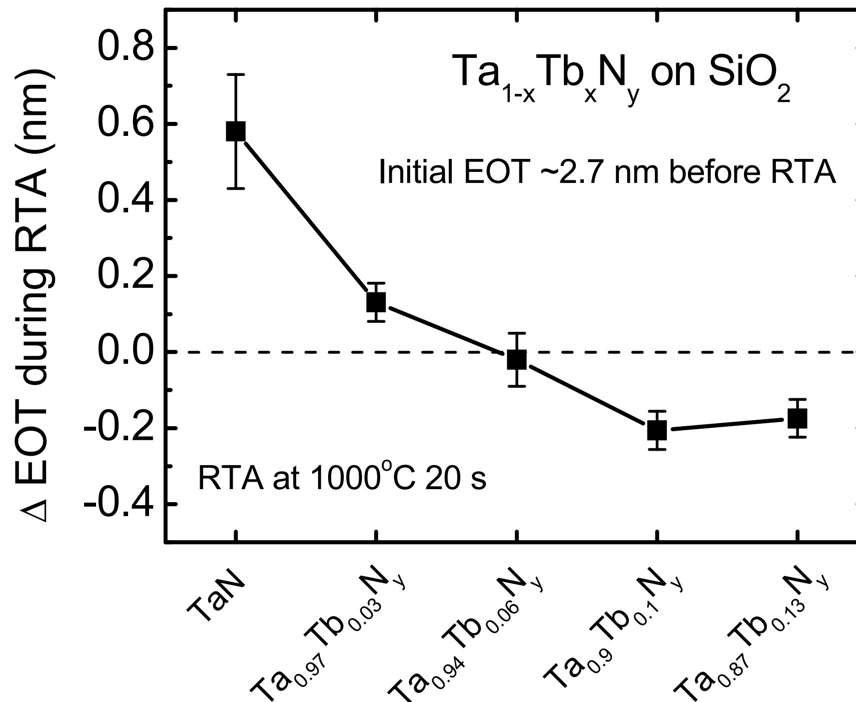


Fig. 4.17 EOT variation of the $Ta_{1-x}Tb_xN_y/SiO_2$ gate stacks before and after 1000 ^oC RTA for 20 sec, as a function of Tb concentrations in $Ta_{1-x}Tb_xN_y$.

Fig. 4.18 and Fig. 4.19 present the gate leakage characteristics of the $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ and $Ta_{0.95}Er_{0.05}N_y/SiO_2$ stacks after thermal treatments at different temperatures. The *I-V* characteristic does not show significant change after annealing

at different temperatures, suggesting good thermal and chemical stability of the $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ (or $Ta_{0.95}Er_{0.05}N_y/SiO_2$) interface. The TDDB reliability characteristics of lanthanide-MN_x metal gates after different RTA treatments were also investigated. The $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ capacitors with different PMA performed were stressed under a constant negative bias to examine the potential impact of high temperature annealing on the reliability of the $Ta_{0.94}Tb_{0.06}N_y$ /SiO_2 gate stack, as shown in Fig. 4.20. No significant reliability degradation is observed even after RTA at 1000 ^oC for 30 secs, demonstrating the excellent thermal stability of $Ta_{0.94}Tb_{0.06}N_y$. The appropriate N content in the lanthanide-MN_x metal gates, which can suppress the reaction of lanthanide with the underlying SiO_2, is believed to be responsible for the thermal stability of lanthanide-MN_x.

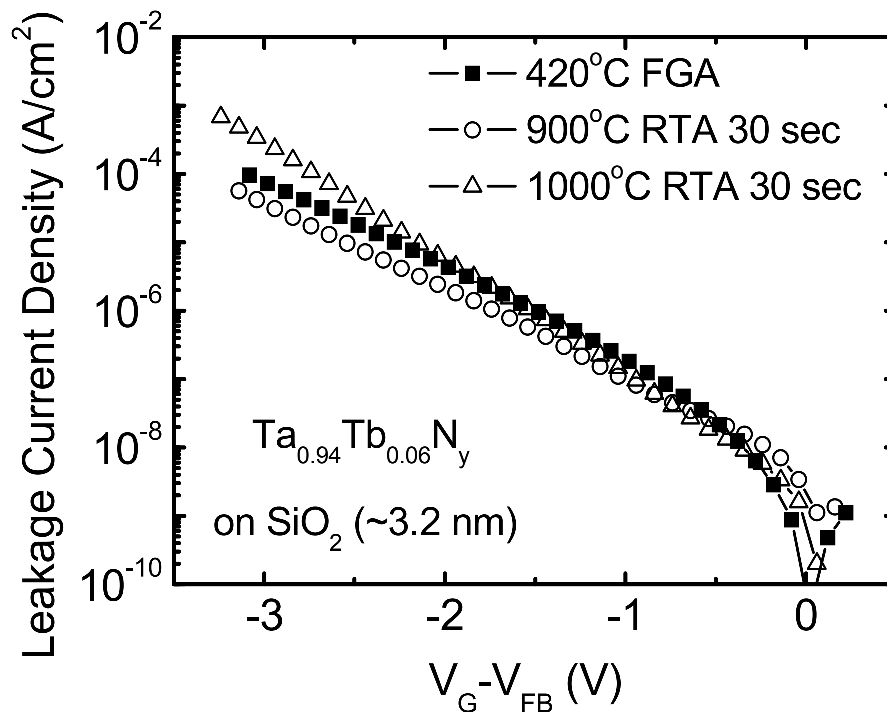


Fig. 4.18 Gate leakage characteristics of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack with PMA performed at different temperatures.

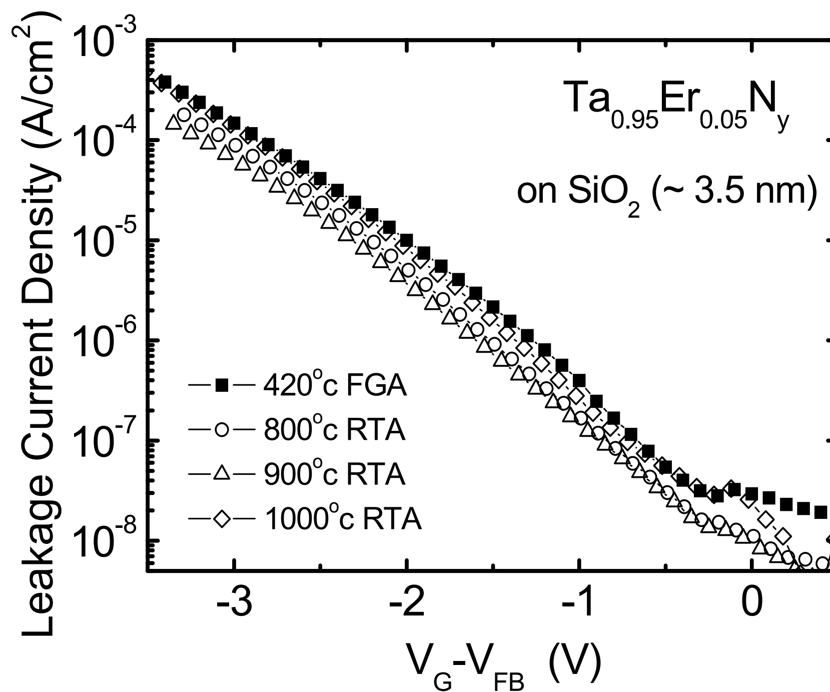


Fig. 4.19 Gate leakage characteristics of $Ta_{0.95}Er_{0.05}N_y/SiO_2$ gate stack with PMA performed at different temperatures.

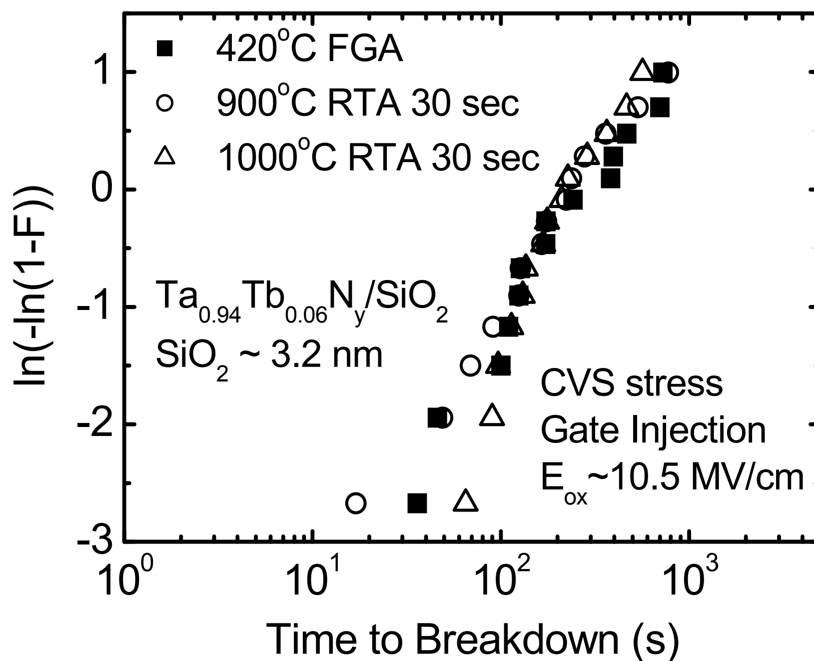


Fig. 4.20 TDDB characteristics of $Ta_{0.94}Tb_{0.06}N_y/SiO_2$ gate stack (SiO_2~3.2 nm) after PMA at different temperatures, measured under negative constant voltage stress (CVS) at room temperature.

In order to study the role of N in the thermal stability of lanthanide-MN_x materials, we varied the N_2/Ar flow rates during the deposition of $Ta_{0.9}Tb_{0.1}N_y$ (N_2/Ar = 3/25, 5/25, and 8/25 sccm) on purpose. Fig. 4.21 compares the EOT stability among $Ta_{0.9}Tb_{0.1}N_y$ gates with different N_2 flow rates during deposition. It is observed that $Ta_{0.9}Tb_{0.1}N_y$ with low N concentration (N_2/Ar = 3/25) shows poor EOT stability against the high thermal budget used in the process. This is attributed to some reactions between Tb and/or Ta metals with the underlying SiO_2 dielectrics. Fortunately, this kind of interface reaction can be suppressed by increasing the nitrogen concentration in lanthanide-MN_x. As shown in Fig. 4.21, the EOT of $Ta_{0.9}Tb_{0.1}N_y$ (N_2/Ar = 8/25)/SiO_2 gate stack keeps almost constant even after 1000 ^{o}C RTA treatment, implying a stable interface between $Ta_{0.9}Tb_{0.1}N_y$ (N_2/Ar = 8/25) and SiO_2.

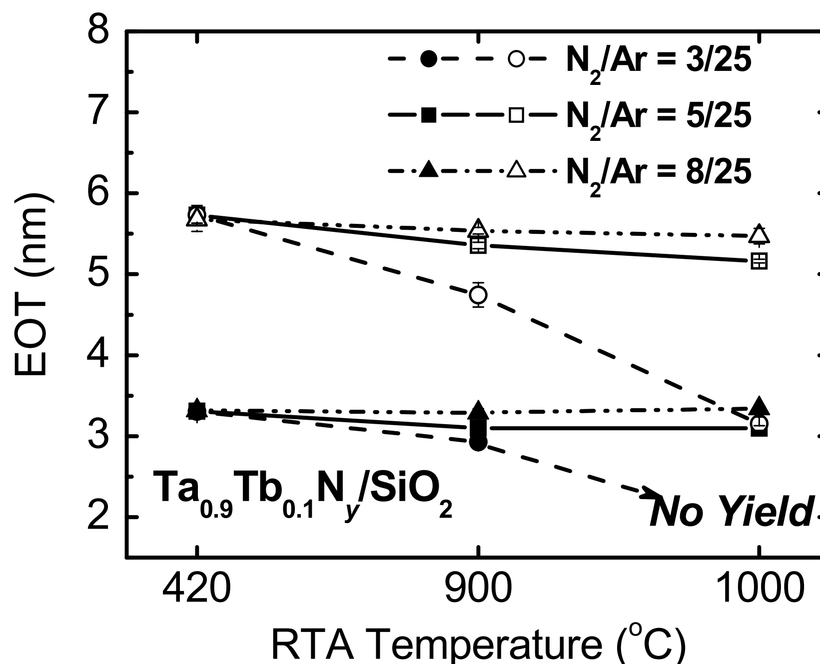


Fig. 4.21 EOT variation as a function of annealing temperature for $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different N_2 flow rates during the deposition of $Ta_{0.9}Tb_{0.1}N_y$. Two groups of oxides with initial thickness of ~3.3 nm and ~ 5.8 nm were investigated.

The higher N concentration can also contribute to the leakage current characteristics in $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ stack, as depicted in Fig. 4.22. $Ta_{0.9}Tb_{0.1}N_y$ with lower N_2 flow rate ($N_2/Ar = 3/25$) and consequently lower N concentration exhibits higher gate leakage than the other two splits. These results suggest that the N concentration in lanthanide-MN_x could be responsible for the chemical-thermal stability of the lanthanide-MN_x/SiO_2 interface.

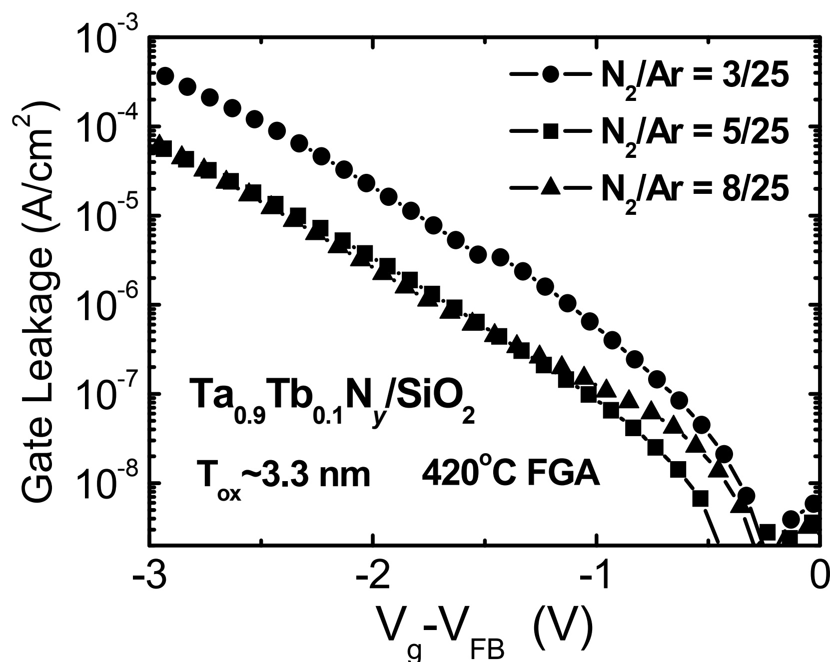


Fig. 4.22 Typical *I-V* characteristics of $Ta_{0.9}Tb_{0.1}N_y$ gated MOS capacitors with different N_2 flows during metal gate deposition, measured after FGA at 420 ^{o}C for 30 min.

If the N concentration is too high, however, it will lead to high work function values for lanthanide-MN_x. Fig. 4.23 compares the WF of $Ta_{0.9}Tb_{0.1}N_y$ metal gates with different N concentrations as a function of annealing temperature. The result shows that $Ta_{0.9}Tb_{0.1}N_y$ with $N_2/Ar = 5/25$ exhibits a low WF of ~ 4.2 eV after 1000 ^{o}C RTA process. However, the split with highest nitrogen flow ($N_2/Ar = 8/25$) shows

higher WF (4.4 ± 0.05 eV) than the others. The reason for the dependence of the metal work function on the N concentration is still not very clear, but a possible reason could be the reduction of Tb atoms near the metal-dielectric interface due to the increased N concentrations. A more plausible explanation could be that the number of metallic Tb species, which contributes to the low WF, decreases when the N concentration in $Ta_{0.9}Tb_{0.1}N_y$ increases due to the formation of Tb-N compounds, which may possess insulator properties [15].

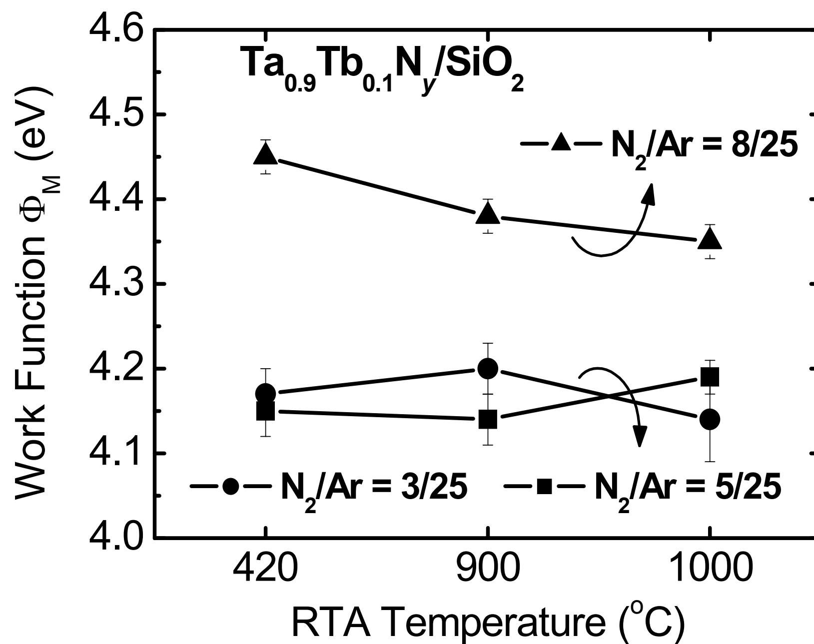


Fig. 4.23 Work function of $Ta_{0.9}Tb_{0.1}N_y$ gate electrodes with different N concentrations as a function of annealing temperature.

In summary, the thermal stability of lanthanide-MN_x is found to be related with the N content. Therefore carefully engineering of the N content according to the lanthanide additives in the lanthanide-MN_x metal gate could be important in the process to achieve a best trade-off between the good thermal stability and a proper work function value desired.

4.6 MOSFET Characteristics

A further effort of this chapter is to discuss some process integration issues of the lanthanide-MN_x metal gate in CMOS process. As mentioned before, lanthanide-MN_x has higher resistivity than the MN_x metal gates. This leads to high sheet resistance of the gate electrode stack and is not desirable. One way to address this problem is to minimize the thickness of lanthanide-MN_x in the metal gate stack. An additional advantage to use a thin lanthanide-MN_x layer is that this minimizes the difficulties in patterning these ternary metal nitride films since the lanthanide and the refractory metal species (e.g. Tb and Ta) may possess different etching properties (etching rate, selectivity, byproduct properties, etc.).

However, the thickness of a lanthanide-MN_x layer must be large enough to determine the WF, otherwise the capping layer material may affect the WF of the gate stack [20]. To evaluate the dependence of work function on $Ta_{0.9}Tb_{0.1}N_y$ thickness, $Ta_{0.9}Tb_{0.1}N_y$-gated MOS capacitors with different $Ta_{0.9}Tb_{0.1}N_y$ thickness (~ 40 Å, 100 Å, 150 Å and 200 Å) were fabricated. Fig. 4.24 shows the cross-section image of the gate stack consisting of TaN capping layer, 40 Å $Ta_{0.9}Tb_{0.1}N_y$ and SiO_2 dielectric. The *C-V* characteristics of these devices are measured after 1000°C PMA and plotted in Fig. 4.25. It is observed that the device with ~ 40 Å $Ta_{0.9}Tb_{0.1}N_y$ shows different V_{FB} compared with the other samples.

The WF values of $Ta_{0.9}Tb_{0.1}N_y$ with different thickness are extracted in Fig. 4.26. It is found that $Ta_{0.9}Tb_{0.1}N_y$ with a thickness of ~ 40 Å exhibits higher WF value than the other splits with thicker $Ta_{0.9}Tb_{0.1}N_y$ of 100 ~ 200 Å. It is also interesting to note that the WF difference tends to increase with the annealing temperature. As shown in Fig. 4.26, the WF difference for $Ta_{0.9}Tb_{0.1}N_y$/TaN stack with varying $Ta_{0.9}Tb_{0.1}N_y$ thicknesses is only about 0.08 eV if the RTA is performed

at 900 ^oC, but it becomes 0.18 eV when the RTA is performed at 1000 ^oC for 10 sec. This implies that the different WF for $Ta_{0.9}Tb_{0.1}N_y$/TaN stacks with varied $Ta_{0.9}Tb_{0.1}N_y$ thicknesses could be attributed to some interaction between the $Ta_{0.9}Tb_{0.1}N_y$ and the capping TaN layer during the high-temperature annealing process, causing the effective work function of $Ta_{0.9}Tb_{0.1}N_y$/TaN metal gate stack to be shifted. The interaction becomes more severe when the annealing temperature becomes higher. The behaviors of EWF in the stacked thin metal layers are quite interesting and have been reported in many studies [20]-[22]. Carrier redistribution in thin continuous metal stacks [20] and/or inter-diffusion of the metal species in bi-layer structures [22] have been proposed to explain the dependence of EWF on the metal layer thicknesses in the stacked metal structures. The mechanisms in our case, however, are still not very clear and need to be further investigated from a scientific perspective. To guarantee the low EWF value, $Ta_{0.9}Tb_{0.1}N_y$ with thickness of 100 Å is chosen for the MOSFET fabrication in this work.

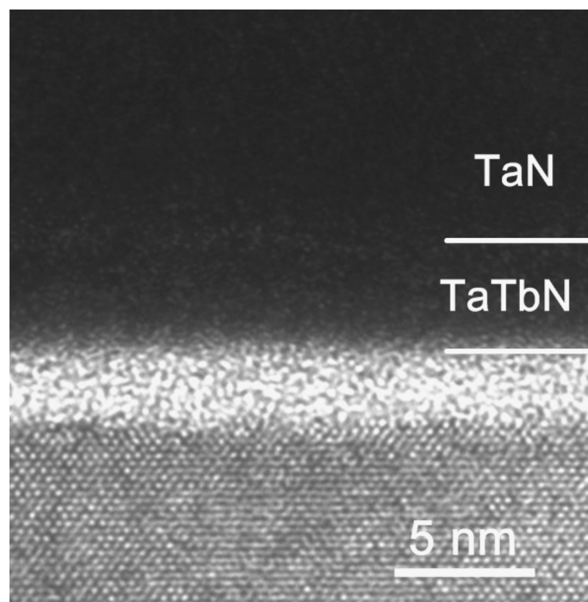


Fig. 4.24 XTEM image of TaN/$Ta_{0.9}Tb_{0.1}N_y$/SiO_2 gate stack with $Ta_{0.9}Tb_{0.1}N_y$ thickness of ~ 40 Å on SiO_2 dielectric.

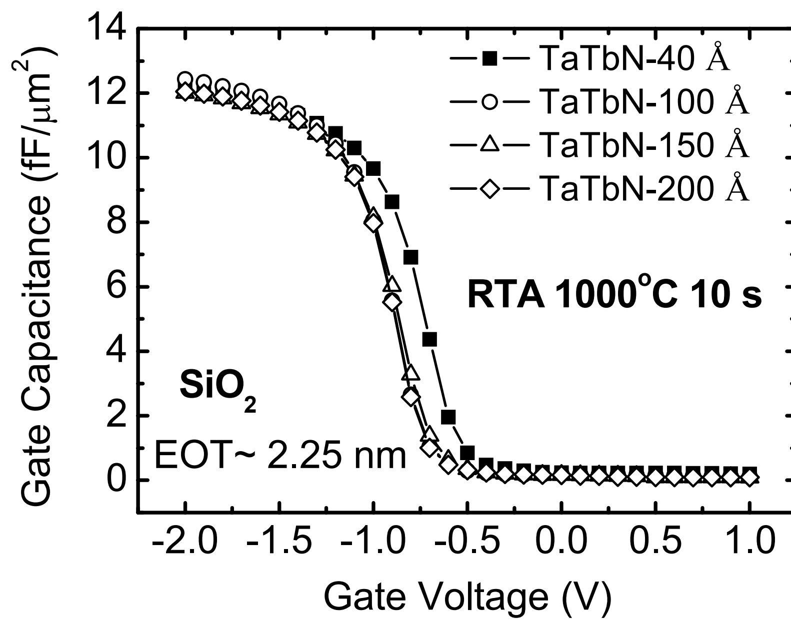


Fig. 4.25 *C-V* characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness. The measurement was taken on the samples with 1000 ^oC RTA for 10 sec performed.

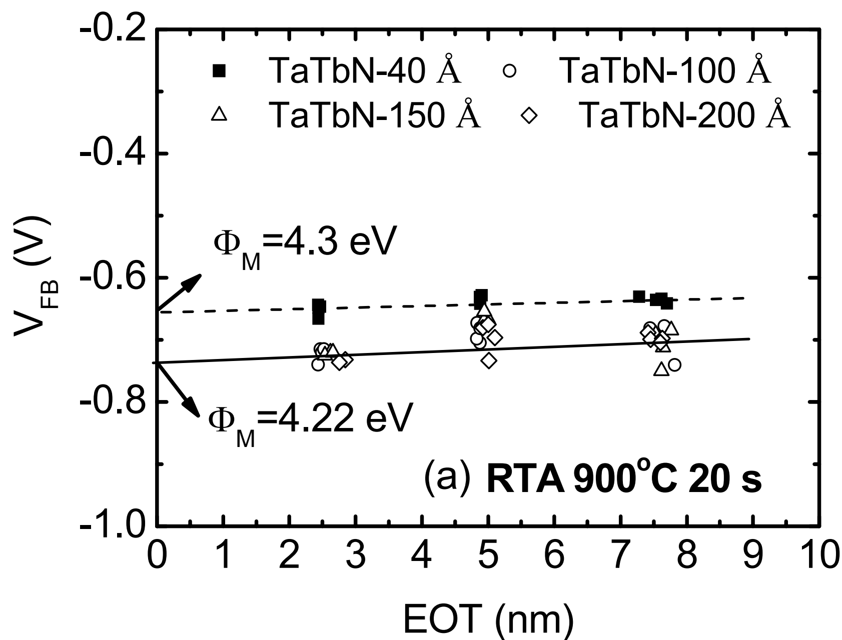


Fig. 4.26 (a) V_{FB} versus EOT plots of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness after 900 ^oC RTA for 20 sec.

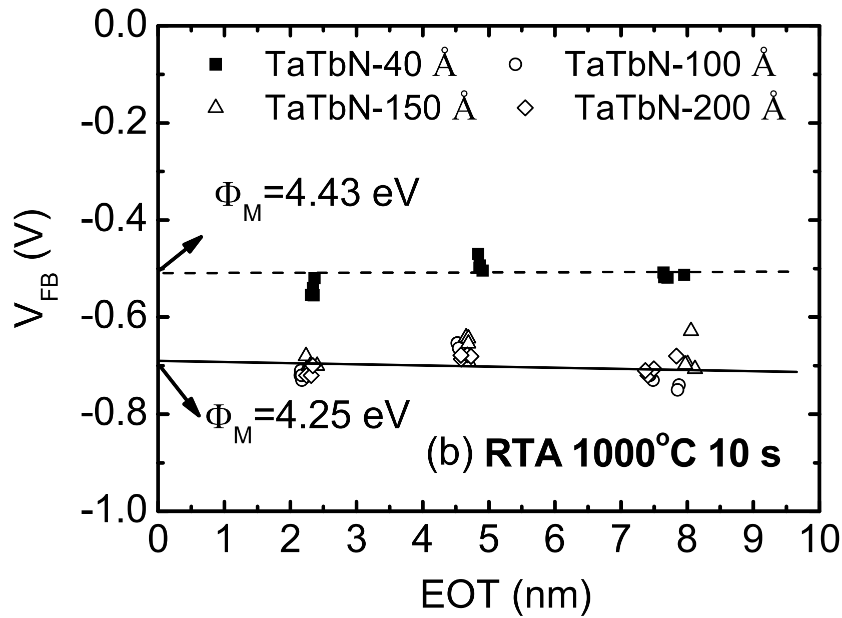


Fig. 4.26 (b) V_{FB} versus EOT plots of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stacks with different $Ta_{0.9}Tb_{0.1}N_y$ thickness after 1000 °C RTA for 10 sec.

In the MOSFET fabrication, a damascene-gate process was used to pattern the gate electrode in order to avoid the S/D recess problem associated with the non-optimized plasma etching process, as illustrated in Fig. 4.27. First, a thick passivation oxide layer of ~1500 Å was grown on p-Si substrate, followed by the opening of the gate area on the passivation oxide. Second, the thermal oxide with thickness of about 30 Å was grown as the gate dielectric, followed by the deposition of 100-Å-thick $Ta_{0.9}Tb_{0.1}N_y$ and an *in-situ* deposited TaN capping layer of about 2000 Å. Third, a CMP process was used to polish away the metals on top of the passivation oxide. Finally, the passivation oxide was removed by diluted hydrofluoric (DHF) acid solution, followed by a standard process to form spacer and S/D region. Note that such a process is only a temporary solution when the plasma etching process for these novel metal gate materials has not been well developed. Further development of a dry etching solution will be required in future study.

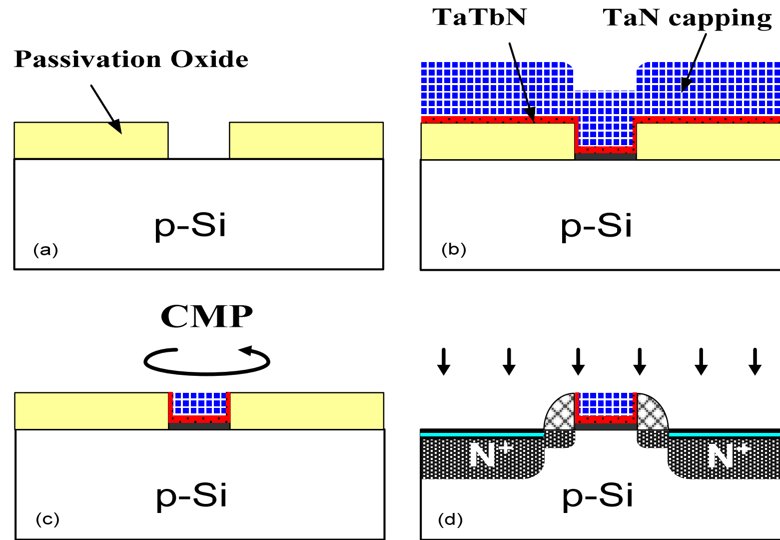


Fig. 4.27 Process flow of the damascene process used to pattern the TaN/ $Ta_{0.9}Tb_{0.1}N_y$ metal gate stack in MOSFET fabrication.

Fig. 4.28 to Fig. 4.31 present the characteristics of n-MOSFET with $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gate stack. The source/drain activation anneal was performed at 1000°C for 5 sec in N_2 ambient. A typical *C-V* curve measured from the n-MOSFET is depicted in Fig. 4.28, with no gate-depletion observed. Fig. 4.29 and Fig. 30 present the $I_{DS} \sim V_{DS}$ and $I_{DS} \sim V_{GS}$ characteristics of n-MOSFET with $Ta_{0.9}Tb_{0.1}N_y$ metal gate. The threshold voltage is only about 0.07 V thanks to the low WF of $Ta_{0.9}Tb_{0.1}N_y$ (4.25 ± 0.05 eV). It was reported that N can diffuse out from TaN_x during the high-temperature process when the N concentration is high [23], resulting in the degradation of the SiO_2/Si interface. However, an excellent subthreshold swing (SS) of 68 mV/dec has been obtained from $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET, suggesting that the interface quality may not be affected in the $Ta_{0.9}Tb_{0.1}N_y$-gated devices. This could be attributed to the presence of Tb which is more capable to combine with N, as discussed in section 4.3.2 and 4.3.3. The effective electron mobility in the $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ stack has also been investigated, as shown in Fig. 4.31. No mobility degradation has been observed, suggesting good interface quality.

These results demonstrate that $Ta_{0.9}Tb_{0.1}N_y$ could be potential N-type metal gate candidates for next-generation transistors. More studies on the characteristics of lanthanide-MN_x with high-κ dielectrics will be needed in the future work.

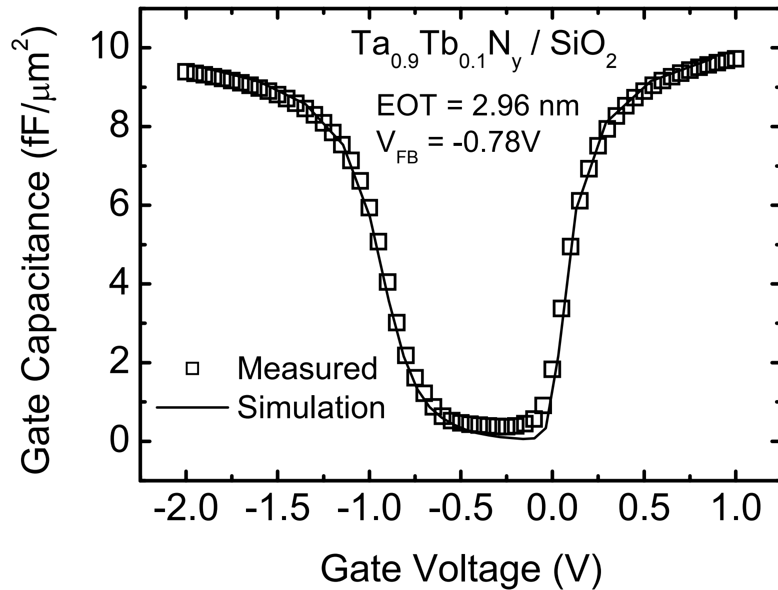


Fig. 4.28 Typical high-frequency *C-V* measurement of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET.

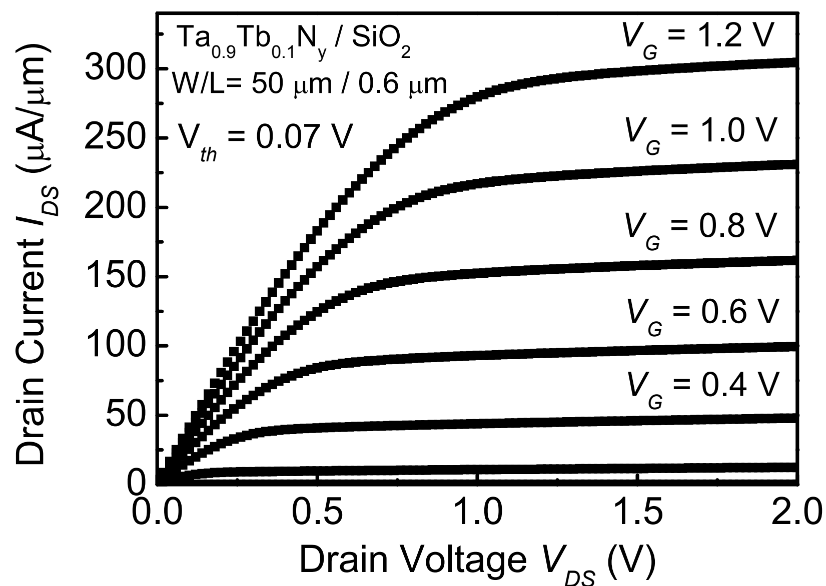


Fig. 4.29 $I_{DS} \sim V_{DS}$ characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET, with the substrate doping concentration of $N_A = 5 \times 10^{15}$ cm^{-3}.

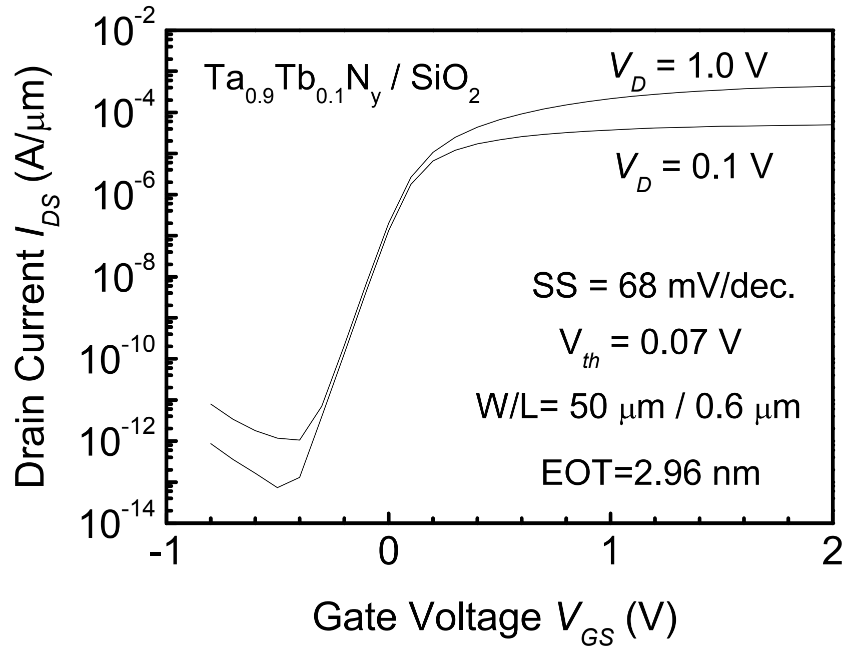


Fig. 4.30 $I_{DS} \sim V_{GS}$ characteristics of $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET, with the substrate doping concentration of $N_A = 5 \times 10^{15}$ cm^{-3}.

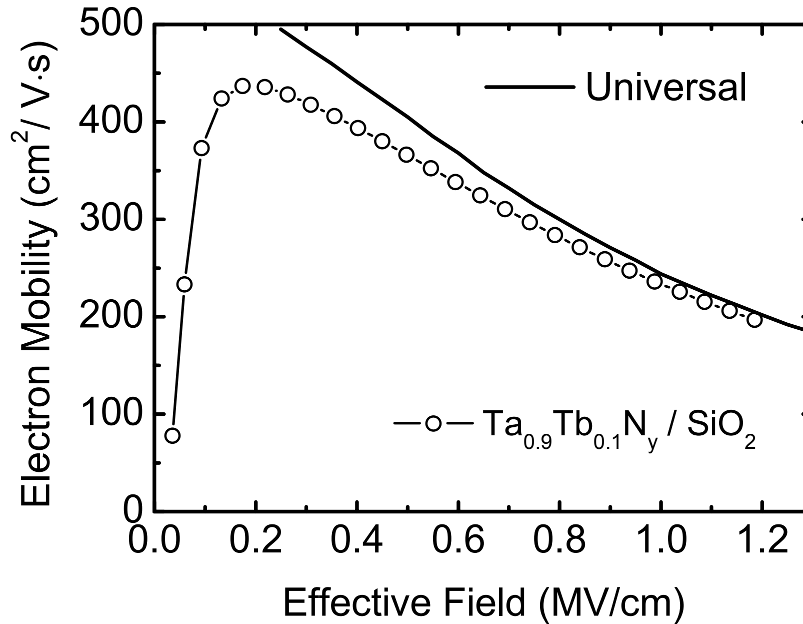


Fig. 4.31 Effective electron mobility in $Ta_{0.9}Tb_{0.1}N_y/SiO_2$ gated n-MOSFET.

4.7 Conclusion

In summary, for the first time, we successfully demonstrated a novel method to tune the work function of MN_x gate electrodes by incorporating lanthanide-series elements. The material and electrical properties of lanthanide-incorporated metal nitrides have been systematically studied. The results show that the work function of the lanthanide-MN_x metal gates can be continuously tuned by varying the lanthanide concentration, and a work function value of 4.2~4.3 eV can be obtained even after a 1000 ^{o}C RTA treatment, which is attractive for the applications in n-MOSFET using a gate-first process. Good thermal stability with respect to EOT, leakage current and TDDB characteristics have also been achieved up to 1000°C in the TaTbN/SiO_2 stacks. The nitrogen concentration is identified as an important parameter to the properties of lanthanide-MN_x, especially for the thermal stability; therefore the N concentration needs to be carefully engineered in lanthanide-MN_x. MOSFETs with $Ta_{0.9}Tb_{0.1}N_y$ as the metal gate electrode have also been successfully demonstrated, with excellent transistor characteristics reported. All these results indicate that lanthanide-MN_x can be a promising N-type metal gate candidate for the next generation CMOS technology.

References

1. Y. H. Kim, C. H. Lee, T. S. Jeon, W. P. Bai, C. H. Choi, S. J. Lee, L. Xinjian, R. Clarks, D. Roberts, and D. L. Kwong, "High quality CVD TaN gate electrode for sub-100 nm MOS devices," in *IEDM Tech. Dig.*, pp. 667-670, 2001.

2. D.-G. Park, Z.J. Luo, N. Edleman, W. Zhu, P. Nguyen, K. Wong, C. Cabral, P. Jamison, B.H. Lee, A. Chou, M. Chudzik, J. Bruley, O. Gluschenkov, P. Ronsheim, A. Chakravarti, R. Mitchell, V. Ku, H. Kim, E. Duch, P. Kozlowski, C. D'Emic, V. Narayanan, A. Steegen, R. Wise, R. Jammy, R. Rengarajan, H. Ng, A. Sekiguchi, and C.H. Wann, "Thermally robust dual-work function ALD-MN_x MOSFETs using conventional CMOS process flow," in *VLSI Tech. Dig.*, pp. 186-187, 2004.

3. H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M.-F. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du, and D.-L. Kwong, "Thermally robust high quality HfN/HfO_2 gate stack for advanced CMOS devices," in *IEDM Tech. Dig.*, pp. 99-103, 2003.

4. W. Tsai, L.-Å Ragnarsson, L. Pantisano, P. J. Chen, B. Onsia, T. Schram, E. Cartier, A. Kerber, E. Young, M. Caymax, S. De Gendt, and M. Heyns, "Performance comparison of sub 1 nm sputtered TiN/HfO_2 nMOS and pMOSFETs," in *IEDM Tech. Dig.*, pp. 311-314, 2003.

5. H. B. Michaelson, "The work function of the elements and its periodicity," *J. Appl. Phys.,* vol. 48, pp. 4729-4733, 1977.

6. T. Durakiewicz, S. Halas, A. Arko, J. J. Joyce, and D. P. Moore, "Electronic work-function calculations of polycrystalline metal surfaces revisited," *Phys. Rev. B*, vol. 64, 045101, Jun. 2001.

7. M. K. Fung, S. W. Tong, S. L. Lai, S. N. Bao, and C. S. Lee, W. W. Wu, M. Inbasekaran, and J. J. O'Brien, S. Y. Liu, and S. T. Lee, "Role of ytterbium and ytterbium/cesium fluoride on the chemistry of poly (9,9-dioctylfluorene-co-benzothiadiazole) as investigated by photoemission spectroscopy," *J. Appl. Phys.*, vol. 94, pp. 2686-2694, Aug. 2003.

8. Dieter K. Schroder, *Semiconductor material and device characterization*, 2nd Edition, John Wiley & Sons, Inc, New York, 1998, ch. 10, pp. 689-697.

9. Refer to www.genplot.com for XRUMP simulator.

10. T. Sekine, Y. Nagasawa, M. Kudoh, Y. Sakai, A. S. Parkes, J. D. Geller, A. Mogami, and K. Hirata, "Handbook of Auger Electron Spectroscopy," *JEOL Inc*., Tokyo,1982.

11. K. Sasaki, A. Noya, and T. Umezawa, "Stoichiometry of Ta-N film and its application for diffusion barrier in the Al_3Ta/Ta-N/Si contact system," *Jpn. J. Appl. Phys.*, vol. 29, pp. 1043-1047, 1990.

12. M.-F. Wang, T.-Y. Huang, Y.-C. Kao, H.-C. Lin, and C.-Y. Chang, "Impact of thermal stability on the characteristics of complementary metal oxide semiconductor transistors with TiN metal gate," *Jpn. J. Appl. Phys.*, vol. 41, pp. 546-551, Feb. 2002.

13. C. S. Kang, H.-J. Cho, Y. H. Kim, R. Choi, K. Onishi, A. Shahriar, and J. C. Lee, "Characterization of resistivity and work function of sputtered-TaN film for gate electrode applications," *J. Vac. Sci. Technol.* B, vol. 21, pp. 2026-2028, Sep./Oct. 2003.

14. X. Sun, E. Kolawa, J.-S. Chen, J. S. Reid, and M.-A. Nicolet, "Properties of reactively sputter-deposited Ta-N thin films," *Thin Solid Film*, vol. 236, pp. 347-351, 1993.

15. C. M. Aerts, P. Strange, M. Horne, W. M. Temmerman, Z. Szotek, and A. Svane, "Half-metallic to insulating behavior of rare-earth nitrides," *Phys. Rev. B*, vol. 69, 045115, 2004.

16. C. Ren, H. Y. Yu, J. F. Kang, Y. T. Hou, M.-F. Li, W. D. Wang, D. S. H. Chan, and D.-L. Kwong, "Fermi-level pinning induced thermal instability in the effective work function of TaN in TaN/SiO_2 gate stack," *IEEE Electron Device Lett.*, vol. 25, pp. 123-125, Mar. 2004.

17. H. Y. Yu, C. Ren, Y.-C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M.-F. Li, D. S. H. Chan, and D.-L. Kwong, "Fermi pinning induced thermal instability of metal gate work functions," *IEEE Electron Device Lett.*, vol. 25, pp. 337-339, May 2004.

18. M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2," *J. Appl. Phys.*, vol. 40, pp. 278-283, 1969.

19. S. G. Park, Y. K. Lee, S. B. Kang, H. S. Jung, S. J. Doh, J.-H. Lee, J. H. Choi, G. H. Kim, G. H. Choi, U. I. Chung, and J. T. Moon, "Performance improvement of MOSFET with HfO_2-Al_2O_3 laminate gate dielectric and CVD-TaN metal gate deposited by TAIMATA," in *IEDM Tech. Dig.*, pp. 327-330, 2003.

20. I. S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H. J. Kim, D. Cha, J. Huang, M. J. Kim, B. E. Gnade, J. Kim, and R. M. Wallace, "A novel methodology on tuning work function of metal gate using stacking bi-metal layers," in *IEDM Tech. Dig.*, pp. 303-306, 2004.

21. H.N. Alshareef, K. Choi, H.C. Wen, H. R. Harris, H. Luan, P. Lysaght, P. Majhi, and B.-H. Lee, "Gate work function modification using ultra-thin metal interlayers," *Proc. ECS 207th meeting*, 630, Quebec, Canada, 2005.

22. C. H. Lu, G. M. T. Wong, M. D. Deal, W. Tsai, P. Majhi, C. O. Chui, M. R. Visokay, J. J. Chambers, L. Colombo, B. M. Clemens, and Y. Nishi, "Characteristics and mechanism of tunable work function gate electrodes using a

bilayer metal structure on SiO_2 and HfO_2," *IEEE Electron Device Lett.*, vol. 26, pp. 445-447, July 2005.

23. Y.-S. Suh, G. P. Heuss, J.-H. Lee, and Veena Misra, "Effect of the composition on the electrical properties of $TaSi_xN_y$ metal gate electrodes," *IEEE Electron Device Lett.*, vol. 24, pp. 439-441, July 2003.

Chapter 5

Process Integration of Dual Metal Gate Electrodes

5.1 Introduction

Integration of dual metal gate electrodes into the CMOS process is another major challenge for the development of metal gate technology. The criteria for an ideal dual metal gate integration process include: (i) achieving dual work function (WF) values for NMOS and PMOS respectively; (ii) with no process-induced damages or reliability concerns to the gate oxide; and (iii) less process complexity. Moreover, good scalability is also required for the ideal dual metal gate integration process for the implementation of metal gates in nanometer CMOS fabrication. Depending on the different integration schemes used, the issues and considerations regarding the metal gate integration would be different.

Generally, the schemes to integrate the dual work function metal gates onto a single wafer can be reduced to three categories: gate-first approach, gate-last approach, and FUSI gate integration process. Among the three categories of integration schemes, the gate-first approach is most challenging and many issues like metal selection, metal etching and cleaning need to be considered systematically. The gate-last integration schemes show less difficulty in material selection, but the problem is the process complexity and cost. The FUSI process requires least change to the

conventional CMOS process, but the work function tuning issues on high-κ dielectric need to be addressed.

In this chapter, we will first review the existing dual metal gate integration processes ever reported and discuss the problems and concerns related to these processes. This will facilitate the following discussions on the motivation and innovations of two proposed integration schemes in this work.

The first dual metal gate CMOS integration process was reported by Lu *et al.* by a direct-etching method, where Ti and Mo were used as the gate electrodes for NMOS and PMOS, respectively [1]. The basic steps of this integration scheme are illustrated in Fig. 5.1. One metal gate material (Metal-A) is deposited on the whole wafer firstly, followed by selective etching from one side (NMOS or PMOS). Then the other metal gate (Metal-B) is deposited, and finally it is the gate patterning and source/drain (S/D) formation. Sometimes a poly-Si capping layer is deposited on top of Metal-B to reduce the thicknesses of Metal-A and Metal-B layers; or a planarization process can be inserted after step (c) to reduce the height difference between NMOS and PMOS for the depth-of-focus (DOF) consideration in gate lithography. This approach is a gate-first integration scheme and is generally used in integrating those direct metal gate candidates with good thermal stability, like TiN-TaSiN, TaSiN-Ru, etc [2]-[3].

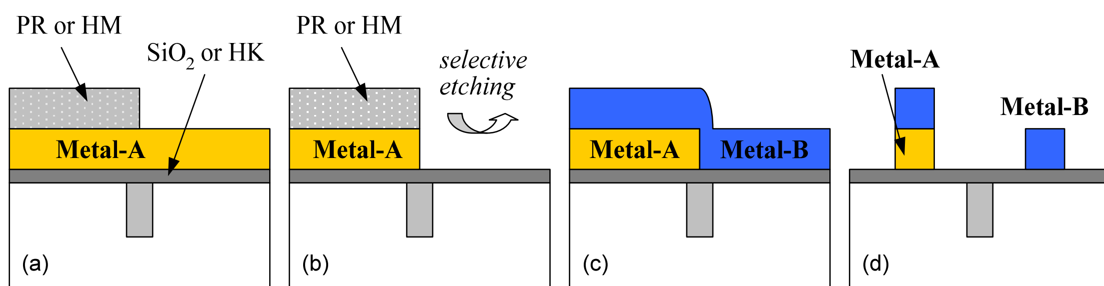


Fig. 5.1 Process flow of dual metal gate integration by direct etching method. PR denotes photoresist, HM denotes hard-mask, and HK denotes high-κ dielectric.

The most challenging step in this direct-etching integration scheme is the selective etching of Metal-A from the gate dielectric and the subsequential process to strip the photoresist (PR) or hard-mask (HM) used to pattern the Metal-A (Fig. 5.1 (b)). Dry etching of Metal-A from gate dielectric will be easy to cause a loss of gate dielectric, leading to different EOT for NMOS and PMOS, as observed in [1]. A wet etching process will be more gentle and safe for the gate dielectric, but still there may be some reliability concerns because the gate dielectric is exposed to a series of wet chemicals during processing. Moreover, appropriate chemical solution need to be identified for the specific metal material so that high selectivity can be achieved among Metal-A, dielectric and hard-mask (HM) [4]. In addition, a proper plasma etching process also needs to be developed to enable good selectivity to be achieved on NMOS and PMOS regions simultaneously, given that the different metal materials with different thicknesses are formed in NMOS and PMOS regions (referring to step (d)).

In order to avoid possible etching damage to the gate dielectric, a dual high-κ (HK) dual metal gate integration process was proposed [5]. The basic idea of this process is to remove the high-κ dielectric together with Metal-A during the metal etching process, and then deposit a new high-κ dielectric layer and Metal-B material as the new gate stack, as illustrated in Fig. 5.2. Note that the Metal-B and HK-B layers must be selectively removed from the top of Metal-A to make the Metal-A stack conductive, as depicted in step (d)-(e) of Fig. 5.2. By using this process, the gate dielectric in either NMOS or PMOS region will not be exposed during the metal etching process, thus less reliability problem would be expected. An additional advantage of this integration scheme is that the metal gate/high-κ stacks for n-MOSFET and p-MOSFET can be engineered independently to obtain optimized

transistor characteristics. However, one more lithography step and CMP step is required in this scheme and thus the process complexity increases.

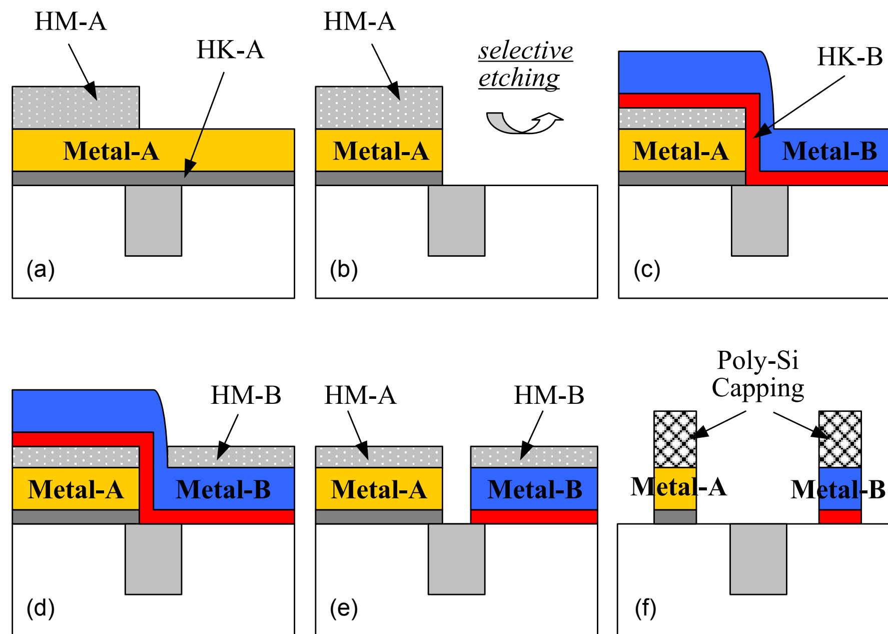


Fig. 5.2 Process flow of the dual metal gate/dual high-κ integration scheme. (a) Metal-A/HK-A deposition; (b) Metal-A/HK-A selective etching from one side of CMOS; (c) Metal-B/HK-B deposition; (d) hard-mask-B deposition and patterning; (e) Metal-B/HK-B selective removal; (f) hard-mask removal, thick poly-Si top-up, and gate patterning.

In the above process flows, the first-deposited metal gate (Metal-A) are all etched away in order to put another metal gate to get dual work function. But actually this may not be necessary. Another stratagem to achieve dual work function for CMOS is to "modulate" the work function of the first-deposited metal, instead of "replace" it by others. The metal inter-diffusion method is such an integration scheme with no step to expose the gate dielectric during the whole process, and hence less process complexity [6]. As shown in Fig. 5.3, a metal gate (Metal-A) is first deposited on top of the gate dielectric on the whole wafer. Then the work function of Metal-A can be modulated by putting the 2nd metal (Metal-B) on top of Metal-A and

performing a subsequential annealing process to form a new metal alloy (Alloy-AB). In some extreme cases, Metal-B can even segregate at the metal-dielectric interface and push the Metal-A atoms away from the interface.

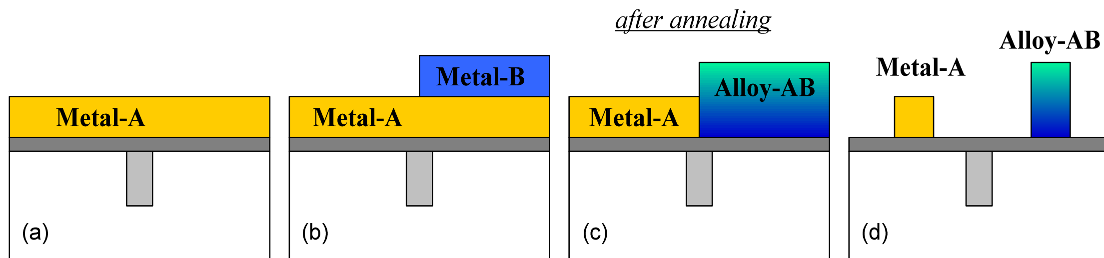


Fig. 5.3 Process flow of the dual metal gate integration via metal inter-diffusion.

The metal inter-diffusion approach has been applied to many metal combinations, such as Ru-Ta alloy [7], Ta-Pt alloy [8], Hf-Mo alloy [9], and even Mo-Si pair [10] where Mo and $MoSi_x$ is formed for PMOS and NMOS, respectively. However, the concern of this integration scheme is the thermal stability of the N-type metal gate candidate. Though a P-type metal gate candidate with good thermal stability can be utilized as the bottom metal (Metal-A), the stability of the Alloy-AB is still a concern for the applications of this approach in a gate-first CMOS process. Therefore it may be difficult to apply such an integration scheme into a gate-first CMOS process. However, it is at least quite useful in a gate-last process.

Another example to “modulate” the work function with no need to etch away the first-deposited layer is the FUSI gate process, where the conventional poly-Si gate is used as the bottom protective layer instead of the Metal-A in the inter-diffusion process discussed above. The use of poly-Si as the bottom layer makes the FUSI process most compatible with the conventional CMOS process flow. As illustrated in Fig. 5.4, a poly-Si gated CMOSFET is first fabricated using a conventional gate-first

CMOS process. After S/D silicidation, the devices are wrapped by the post-metal-dielectric (PMD). A planarization step is then performed using oxide CMP until the HM on top of poly-Si is exposed. After HM stripping and pre-silicidation cleaning, metal film(s) is deposited on top of poly-Si followed proper rapid thermal annealing (RTA) steps to form the desired silicide phases. Some optional process steps, such as an ion implantation (I/I) process (to introduce impurities) or a poly-Si etch-back step (to adjust the poly-Si thickness for desired silicide phases) [11], can be introduced before the metal deposition, as shown in Fig. 5.4 (c). From the integration point of view, the FUSI process is less challenging thanks to its compatibility with the conventional CMOS process. However, the work function tunability on high-κ dielectric still needs to be improved, as well as the precise process control of the impurities and phases in MOSFETs with sub-20-nm L_g.

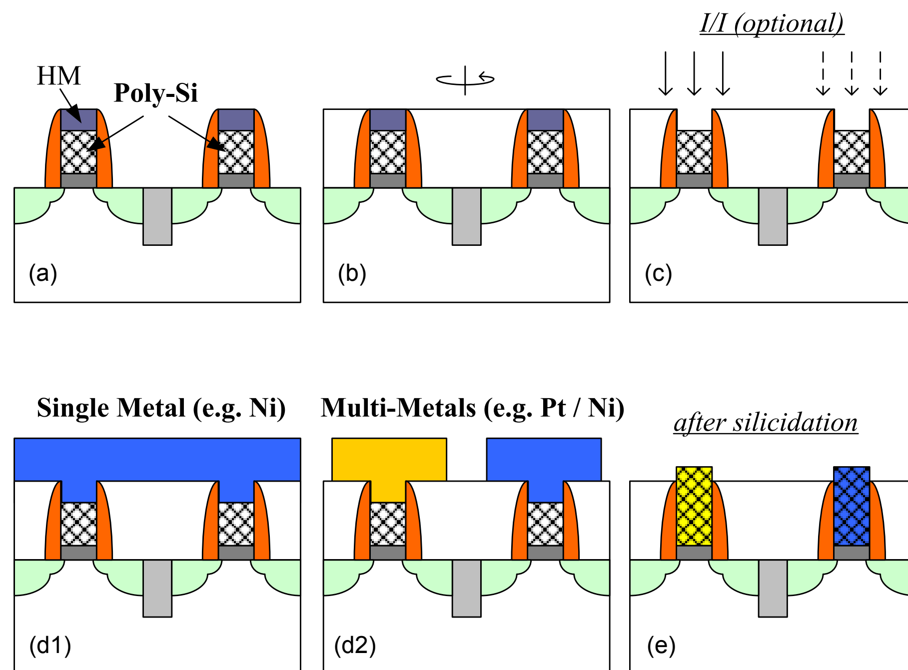


Fig. 5.4 Process flow of the FUSI process. (a) CMOS fabrication conventionally; (b) oxide re-flow and planarization by CMP; (c) hard-mask stripping followed by ion-implantation or poly-Si etch-back; (d1) deposition of the same metal, e.g. Ni, for both NMOS and PMOS; (d2) deposition of different metals for NMOS and PMOS, respectively (in parallel with step (d1)); (e) silicidation and unreacted metal stripping.

Finally, the gate-last replacement gate process has also been widely studied for long time. The process flow of the replacement gate process is described in Fig. 5.5. First, CMOS transistors with dummy SiO_2 gate dielectric and poly-Si gate electrodes are fabricated using a conventional process, followed by the PMD deposition and planarization. After that, the dummy poly-Si and SiO_2 dielectric are selectively removed to create a groove. New high-κ and metal gate materials are then filled into the groove to replace the original SiO_2 and poly-Si gate, followed by CMP to pattern the gate electrode. Dual metal gates can be integrated one by one by utilizing the above "replacement gate" process for two times. Alternatively, one can open the grooves of NMOS and PMOS together and adopt a metal inter-diffusion process to obtain dual work function metal gates.

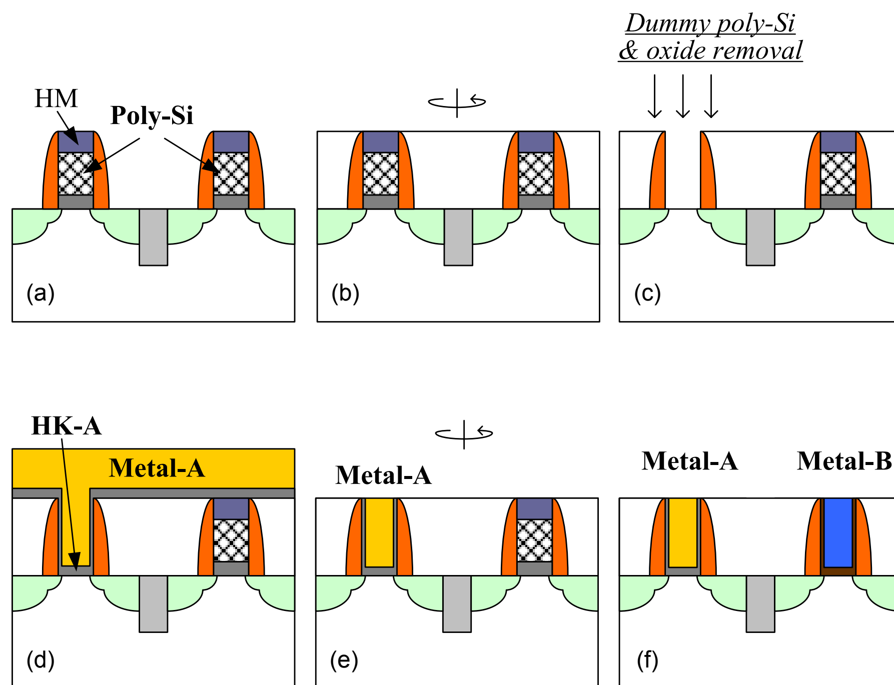


Fig. 5.5 Process flow of the replacement gate process. (a) CMOS fabrication with poly-Si as dummy gate; (b) oxide re-flow and planarization; (c) dummy poly-Si & SiO_2 removal; (d) filling the groove with new high-κ and metal gate; (e) metal CMP to pattern the metal gate; (f) dual metal gate CMOS formation by repeat steps (c)-(e).

From the process integration perspective, the replacement gate process can bypass the metal selective etching step required in the direct-etching approach, and allow the metal/high-κ gate stacks for NMOS and PMOS to be optimized individually. From the material selection point of view, many metal gate candidates, including both the direct metal gates and the binary inter-diffusion metal alloys, can all be used in a replacement gate process, thanks to the low thermal budget used in this process. Therefore, the replacement gate process could be a valuable solution for the metal gate integration. Unfortunately, the concerns of process complexity and cost may prevent this process being utilized for mass-production.

In summary, integration of dual metal gate in CMOS process is a very challenging task and many material and process issues need to be considered systematically. Innovations in both the novel materials and process integration schemes will be highly valued. In order to address or improve some of the issues and limitations in the existing integration schemes mentioned above, we propose two novel dual metal gate integration methods in this chapter. The first one is a gate-first dual metal gate CMOS integration process with the aim to solve the etching damage issue associated with the direct-etching method. In this process, an ultra-thin TaN buffer layer is utilized to protect the gate dielectrics during the metal etching process and a subsequential high-temperature intermixing technique is used to tune the work function of TaN for dual-WF CMOS. The other one is a novel replacement gate process utilizing HfN as the dummy gate to enable the large work function tuning range and excellent EOT scalability to be achieved simultaneously. It is believed that these attempts will be of practical values for the development of the dual metal gate CMOS technology.

5.2 A Gate-First Dual Metal Gate Integration Scheme by High-Temperature Metal Intermixing Technique

5.2.1 Motivation

Several approaches have been studied to integrate dual work function metal gate (MG) electrodes into the CMOS platform using a gate-first process flow, as discussed in the Section 5.1. The conventional direct-etching method is to selectively remove the first-deposited metal from the gate dielectric, followed by the deposition of the second metal material [1]-[3]. However, this may introduce a potential reliability concern as the gate dielectric is exposed to plasma or several chemical solutions during the metal etching process. To avoid the exposure of the gate dielectric during processing, the metal inter-diffusion process was proposed, as depicted in Fig. 5.3 [6]-[9]. But the concern is that the thermal stability of these metal alloys is always limited by the properties of the low WF metal counterpart and not adequate for a gate-first CMOS process. Park *et al.* has proposed a novel method in which an ultra-thin AlN buffer layer is used to protect the gate dielectric during metal etching, and is later consumed by Ta and Hf in an alloying process [12]. However, obtaining a thermal stability which is adequate for the source/drain (S/D) dopant activation step in a gate-first CMOS process is still challenging.

In order to develop a dual metal gate integration process which can not only avoid the gate dielectric being exposed during the process but also fulfill the thermal stability requirement for the gate-first CMOS process, the thermal/chemical stabilities of the buffer metal layer which protects the gate dielectric need to be carefully optimized. Using the well-known thermally stable materials, such as TaN or TiN, as the buffer layer could be helpful to address this problem.

In this work, we report a dual metal gate integration scheme using ultra-thin TaN as the buffer layer and a subsequently high-temperature metal intermixing (InM) process to modulate the WF of TaN. TaN/Tb/TaN and TaN/Ti/HfN [13] stacks are employed for NMOS and PMOS, respectively. Thanks to the good diffusion barrier properties of TaN, the temperature for the metal InM process can be elevated to 1000°C, comfortable for the gate-first CMOS process. Several factors affecting the metal intermixing process will also be discussed. In principle, this technique can be further generalized to other MN_x/metal stacks for performance optimization.

5.2.2 Proposed Integration Flow and Device Fabrication

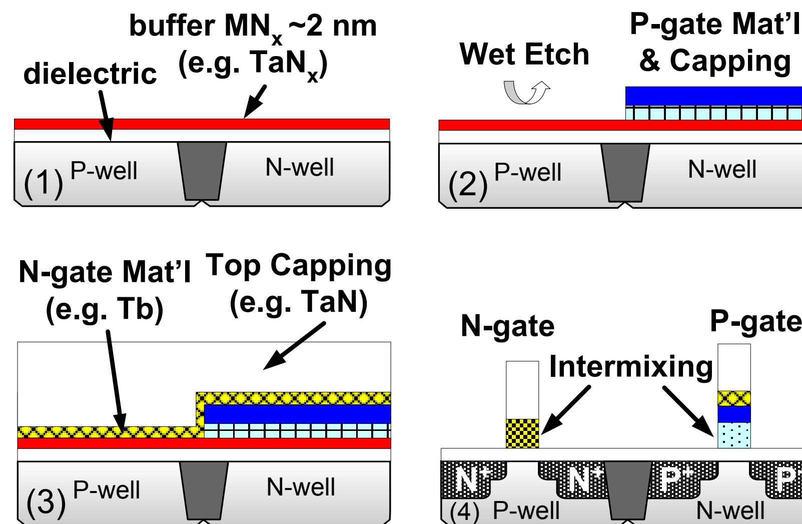


Fig. 5.6 Dual metal gate integration process flow by high-temperature metal intermixing technique: (1) TaN buffer layer deposition; (2) P-type metal gate stack (e.g TaN/Ti/HfN) formation followed by selective etching; (3) N-type metal gate stack formation (e.g TaN/Tb/TaN) and capping layer deposition; (4) gate etching, S/D implantation, and dopant activation annealing (also for intermixing).

Fig. 5.6 shows the proposed dual MG integration process by high-temperature metal intermixing technique. In this prototype demonstration, we used TaN/Tb/TaN and TaN/Ti/HfN structures for NMOS and PMOS, respectively. Firstly, an ultra-thin

TaN buffer layer with thickness of ~2 nm was deposited on the gate dielectric using a precisely controlled PVD facility, Anelva C-7100 GT PVD system, followed by the deposition of Ti and HfN to form laminate gate stack [13]. Secondly, HfN and Ti were selectively removed by diluted HF (DHF) solution (1:200) from NMOS region. Terbium with thickness of ~2.5 nm was then deposited on TaN buffer layer to form TaN/Tb stack in NMOS region, followed by the deposition of an *in-situ* TaN capping layer to finalize the metal gate stack. The TaN capping can also be replaced by other metals or a TaN/Poly-Si stack. After gate patterning, metal intermixing was carried out during the S/D activation annealing process at 950°C or 1000°C for 1~5 sec.

To study the feasibility of this high temperature metal intermixing process in tuning the WF of TaN, MOS capacitors with TaN/Tb/TaN and TaN/Ir/TaN gate electrode stacks on thermally grown SiO_2 were fabricated. TaN with thickness of 2~2.5 nm was first deposited on SiO_2 by dc sputtering of Ta target in Ar/N_2 ambient, where the nitrogen N_2 flow rate was varied during sputtering. Tb or Ir of ~2.5 nm was then *in-situ* deposited on top of TaN, followed by a thick TaN capping layer. After gate patterning, the capacitors were subjected to RTA up to 1000°C for metal intermixing study. On the other hand, in order to investigate the feasibility of this InM process in adjusting V_{th} of the MOSFET with high-κ dielectric, HfTaON/HfO_2 (30Å/15Å) dielectric stack (HfTaON on top) was deposited directly on p-Si by PVD process for MOSFET fabrication [14], followed by post-deposition-annealing (PDA) in N_2/O_2 ambient (5% O_2) to improve the film quality. TaN, TaN/Tb/TaN, and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ were then deposited as the gate electrodes of n-MOSFETs for comparison. Intermixing of TaN/Tb/TaN stack was carried out at 1000°C for 1s together with the S/D activation annealing. NiSi was performed in S/D regions to reduce the sheet resistance. All the samples received a forming gas annealing (FGA)

at 420 ^{o}C for 30 min prior to fab-out. Cross-sectional transmission electron microscope (XTEM) and energy dispersive X-ray (EDX) were utilized to study the intermixing process. Atomic force microscopy (AFM) was used to characterize the film roughness. WF was extracted from EOT versus flat band voltage (V_{FB}) plots.

5.2.3 Feasibility Study of the High-Temperature Intermixing Method

Fig. 5.7 presents the XTEM images of TaN/Tb stack capped by thick TaN layer before and after high-temperature intermixing annealing. As shown in the XTEM images, smooth and clear interface between TaN and SiO_2 can be found even after 1000 ^{o}C RTA, indicating good barrier properties of TaN. Unlike HfN/Ti/TaN structure studied in [13], the XTEM image of the TaN/Tb/TaN stack does not show significant crystallization effect after 1000 ^{o}C RTA treatment. This could be due to the large atom size difference between Tb and Ta [15].

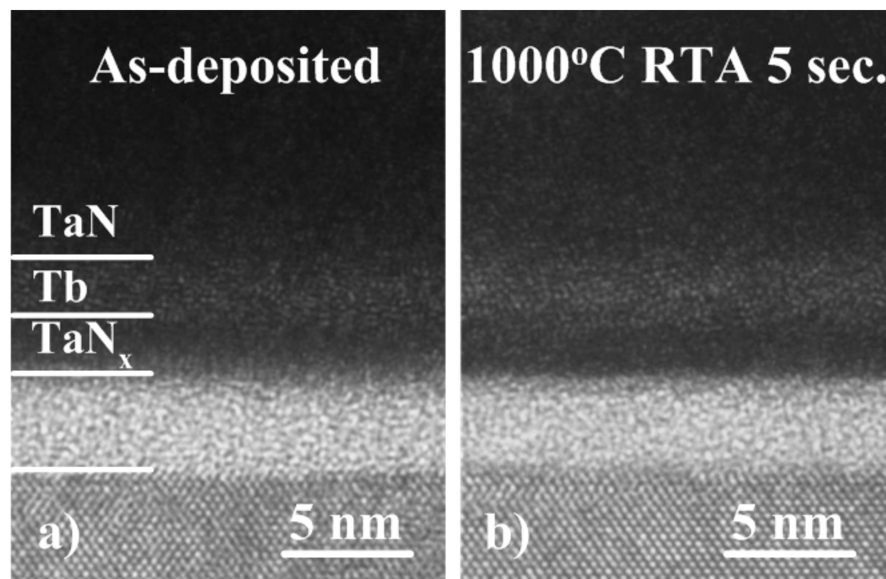


Fig. 5.7 XTEM images of the TaN/Tb/TaN stack on SiO_2 (a) as-deposited and (b) after 1000 ^{o}C RTA in N_2 ambient for 5 sec.

Intermixing of Tb with TaN buffer layer after 1000 °C RTA can be evidenced by STEM/EDX depth profiles, as shown in Fig. 5.8. The Si-Ta profiles represent counts from Si (K) line and/or Ta (M) line due to the overlapping of the two signals at ~1.7 keV in EDX spectrum. The EDX count of oxygen is low because of the lower X-ray generation rate of oxygen compared with that of Ta or Tb. It is observed that the centers of Tb and Ta profiles separates for ~2.5 nm for the as-deposited condition (Fig. 5.8 (c)), corresponding with the XTEM result in Fig. 5.7 (a). After annealing, the profile of Tb becomes broader and tends to mix with that of Ta (N), as shown in Fig. 5.8 (d). Note that a small amount of Tb at the metal-dielectric interface would be sufficient to modify the WF of TaN towards the conduction band of Si as demonstrated in Chapter 4.

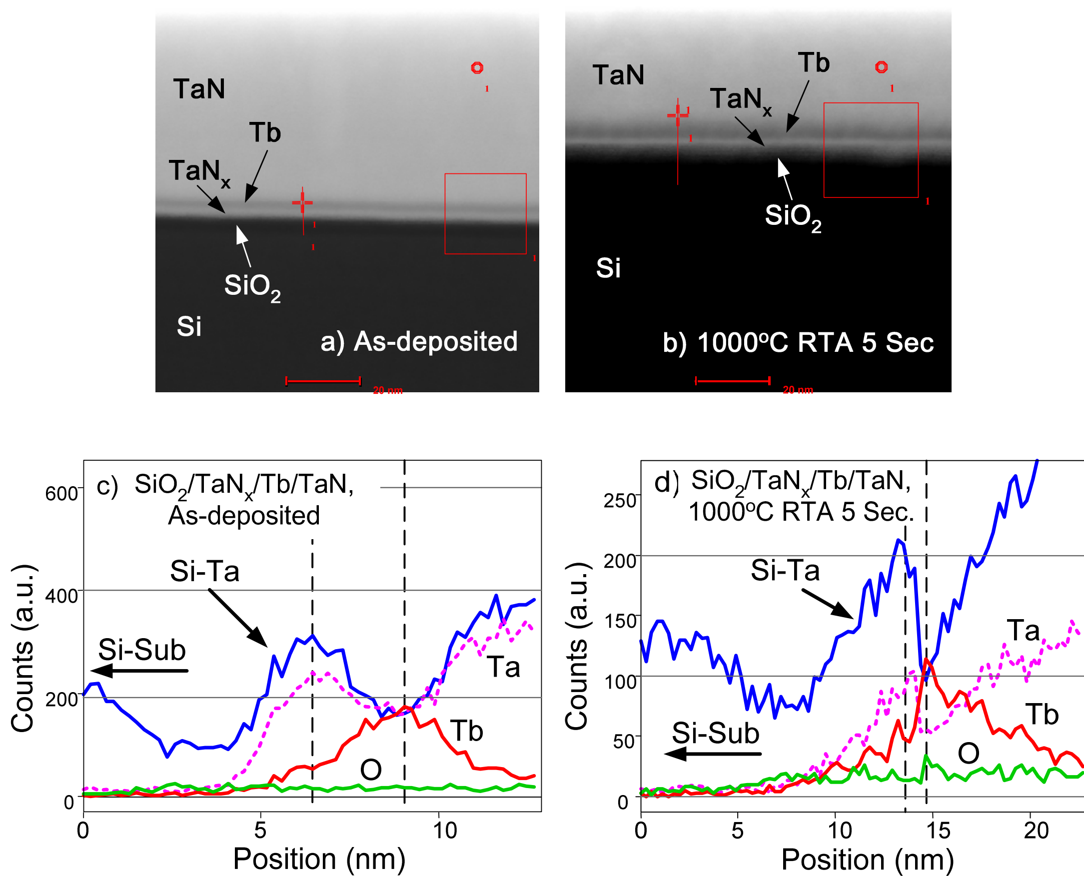


Fig. 5.8 (a), (b) STEM and (c), (d) EDX depth profiles of the TaN/Tb/TaN stack on SiO_2 as-deposited and after 1000 °C RTA in N_2 ambient for 5 sec. In (a) & (b), the dark layer in the sandwich structure denotes Tb element.

Fig. 5.9 summarizes the WF tunability of the metal intermixing approach. WF of pure Tb and Ir elements are ~3.1 eV and ~5.27 eV, respectively. WF of ~4.25 eV and ~4.75 eV were achieved in the TaN/Tb and TaN/Ir stacks, respectively, indicating that the WF of TaN can be adjusted by the proposed high-temperature InM technique. There are several factors which may affect the intermixing process as well as the final WF values of the metal stacks. The first one is the thickness ratio between TaN and the pure metal layers. If TaN is too thick or Tb (Ir) is not thick enough, metal on top of TaN will be prevented from reaching the metal/dielectric interface due to the diffusion barrier properties of TaN. For example, no significant V_{FB} shift was observed in a TaN/Tb stack with 40 Å TaN under 25 Å Tb even after 1000 ^{o}C RTA for 5 sec. Secondly, the properties of the TaN bottom layer can also affect the intermixing process. By reducing the N_2 flow rate from 5 sccm to 4 sccm during sputtering deposition of the TaN_x buffer layer, it was observed that WF of TaN_x with less nitrogen (f_{N2} = 4 sccm) can be modified more easily with a lower thermal budgets (Fig. 5.9), suggesting that N concentration determines the immunity of TaN_x to intermixing. Thirdly, the choice of the pure metal used to tune the WF is also of importance. Although Ir can be used to tune the WF of TaN, the magnitude of WF tuning is smaller than that of Tb; this will be discussed later. Moreover, TaN/Ir devices show worse uniformity and larger EOT increase compared with TaN/Tb capacitors. Breakdown was also observed in TaN/Ir gated capacitors after 1000°C RTA for 10 sec, presumably due to the high diffusivity of the Ir element [16]. Therefore the P-type metal candidate may need to be further developed. On the other hand, TaN/Tb stack with optimized metal thickness shows stable gate leakage even after 1000°C 10 sec anneal as depicted in Fig. 5.10, implying good thermal stability suitable for gate-first CMOS process.

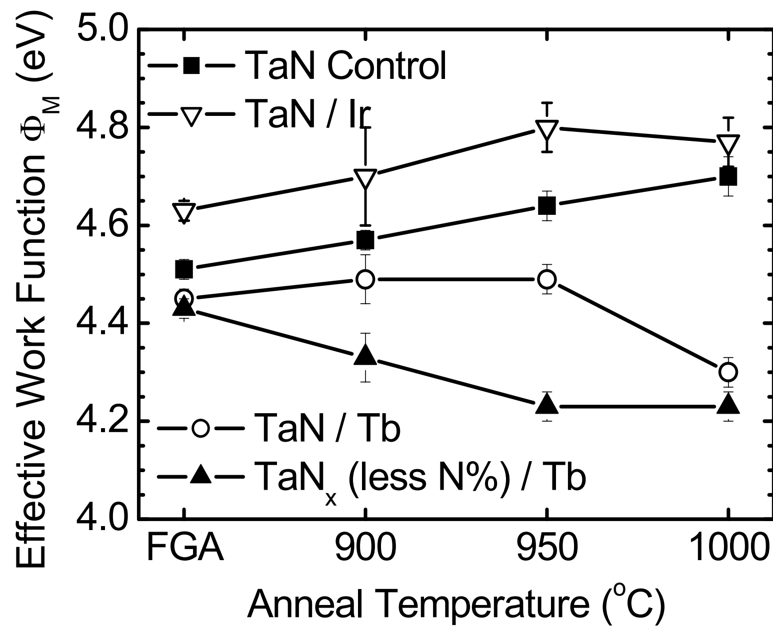


Fig. 5.9 Work function versus annealing temperature for different TaN/Metal stacks. Thickness of TaN or TaN_x (less N%) bottom layers are about 2.0~2.5 nm, and that of Tb or Ir are about 2.5 nm. The N_2 gas flow rate during deposition of thin TaN layer is 5 sccm, while that for TaN_x (less N%) is 4 sccm. All samples are capped with thick TaN film of ~100 nm.

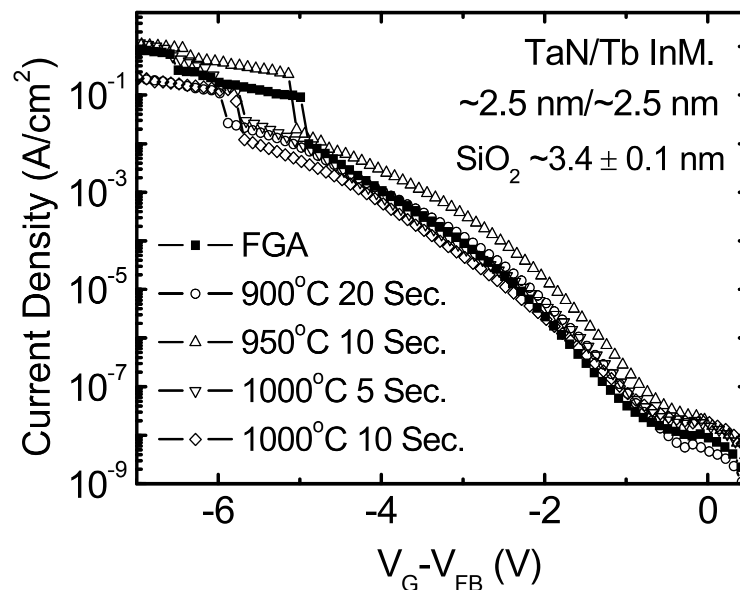


Fig. 5.10 Gate leakage characteristics of TaN/Tb stack after different RTA treatments. The corresponding WF of the sample is denoted by open circle in Fig. 5.9.

A self-consistent charge transfer calculation [17] suggests that the WF of a binary metal AB alloy is determined by not only the WF of each species, but also the effective density of states at the Fermi level, as approximately expressed below [8][9]:

$$\Phi_m(x) = x\Phi_{m,A} + (1-x)\Phi_{m,B} + x(1-x)\frac{(\Phi_{m,A} - \Phi_{m,B})(\rho_A - \rho_B)}{x \cdot \rho_A + (1-x)\rho_B} \quad (5\text{-}1)$$

where x is the fraction of metal A in AB alloy, ρ_A and ρ_B are the effective densities of states at the Fermi level for metal A and B, which are proportional to their respective electronic specific heat coefficients γ. In the dilute-A limit, where a small concentration x of metal A is incorporated into metal B to tune the WF, Eq. (5-1) can be simplified as follow:

$$\Phi_m(x) \cong \Phi_{m,B} + x(\Phi_{m,A} - \Phi_{m,B})\rho_A / \rho_B \quad (5\text{-}2)$$

It is thus preferable to have $\rho_A/\rho_B > 1$ (hence $\gamma_A > \gamma_B$) for the introduction of a small amount of A to modify the WF of metal B significantly. In other words, the element with the larger effective density of states at the Fermi Level would have a higher weighting in determining the final WF of AB alloy, and vice versa. It is known that γ of Ta, Ir and Tb are 5.9, 3.1 and 9.05 [18], respectively. Therefore, the different behaviors of Tb and Ir in modifying WF of Ta(N) can be plausibly explained by the theory above, although the metal system in our study is not strictly a binary alloy and the situation could be more complicated than described. Moreover, larger WF tunability might be achievable by replacing Ir with Pt (γ_{Pt}=6.8) or using TiN (γ_{Ti}=3.35) to replace TaN, as demonstrated in [19].

5.2.4 Compatibility with High-*κ* Dielectrics

Although the InM approach is effective in modulating work function on SiO_2,

it is still not clear whether this method is compatible with high-κ dielectric or not. Therefore we further investigated the effectiveness of this InM method in adjusting the V_{th} of MOSFET with high-κ HfTaON dielectric. HfTaON has been reported as a promising high-κ candidate with high mobility, good thermal and electrical stability [14]. In this work, HfTaON/HfO_2 (30Å/15Å) stack was used as the gate dielectric to reduce the N concentration at the Si surface. TaN, TaN/Tb/TaN, and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ were used as the gate electrodes for comparison. The InM annealing and S/D activation were performed at 1000°C for 1s at the same time. Fig. 5.11 compares the *C-V* measurements of the n-MOSFETs with different metal gates. It is found that the V_{FB} of TaN/Tb/TaN-gated MOSFET shifts negatively compared with that of the TaN-gated devices, showing the effectiveness of the InM method in tuning WF on high-κ dielectric. However, the EOT of the $Ta_{0.9}Tb_{0.1}N_y$ and TaN/Tb/TaN gated nFET is smaller than that of TaN gated devices, possibly due to the scavenging effect of Tb element [20].

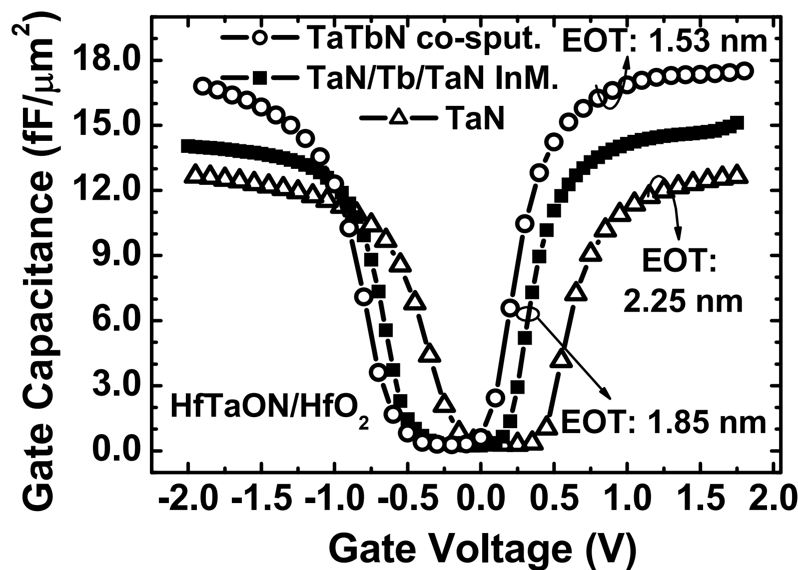


Fig. 5.11 *C-V* characteristics of n-MOSFETs with TaN, TaN/Tb/TaN and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ metal gates on HfTaON after RTA at 1000 ^{o}C for 1 sec.

Fig. 5.12 compares the I_{DS}-V_{GS} characteristics of n-FET with three kinds of metal gates: TaN, TaN/Tb/TaN by InM, and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ with uniform Tb composition. Good subthreshold swing of ~75 mV/dec has been achieved. It is noted that a significant shift of V_{th} can be obtained by introducing Tb into TaN gate, either by the InM approach or by the co-sputtering process. Statistic V_{th} distributions of the three kinds of devices are compared in Fig. 5.13. The results show that the V_{th} can be adjusted successfully with a magnitude of ~300mV by the InM method, while larger V_{th} difference of ~400mV can be achieved using the pre-doped $Ta_{0.9}Tb_{0.1}N_y$ metal gate. This implies that only a small amount of Tb diffuses through TaN and reaches the MG/high-κ interface during the InM annealing. Further adjustment of V_{th} could be realized by optimizing the thickness ratio of TaN and Tb layers.

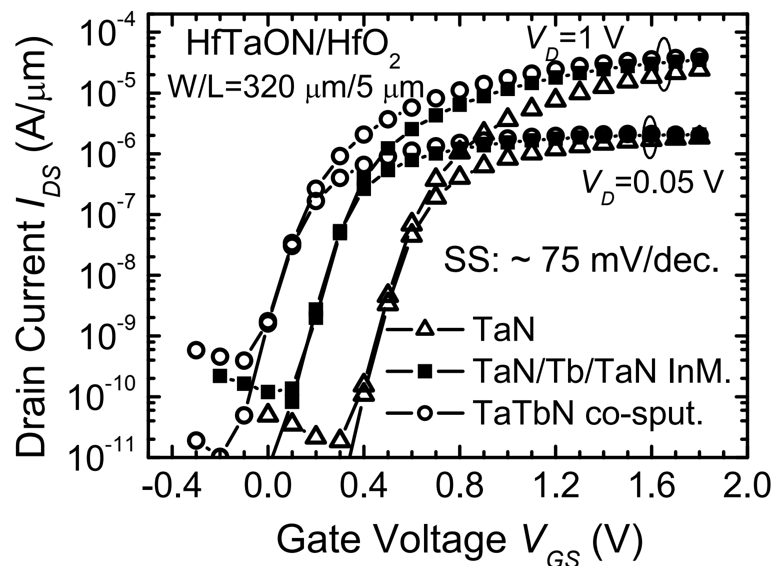


Fig. 5.12 I_{DS}-V_{GS} characteristics of n-MOSFETs with TaN, TaN/Tb/TaN (InM), and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ gates on HfTaON high-κ dielectric.

The electron mobility in the three kinds of devices has also been evaluated and shown in Fig. 5.14. It was observed that the electron mobility in TaN/Tb/TaN gated n-MOSFET is a little bit lower than that in TaN and $Ta_{0.9}Tb_{0.1}N_y$ gated n-MOSFETs.

This could be correlated with the higher D_{it} level in TaN/Tb/TaN gated devices, measured by charge pumping method. The reason is still not very clear, but a possible reason could be the stress effect associated with the InM process. Therefore the annealing condition needs to be further optimized to minimize this problem.

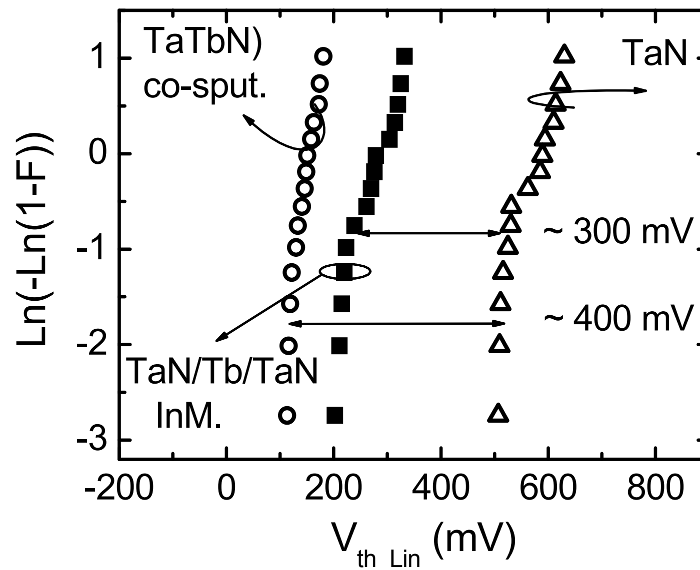


Fig. 5.13 V_{th} distribution of n-MOSFETs with TaN, TaN/Tb/TaN (InM) and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ gates on $HfTaON/HfO_2$ dielectric. (W/L=320 μm / 5 μm)

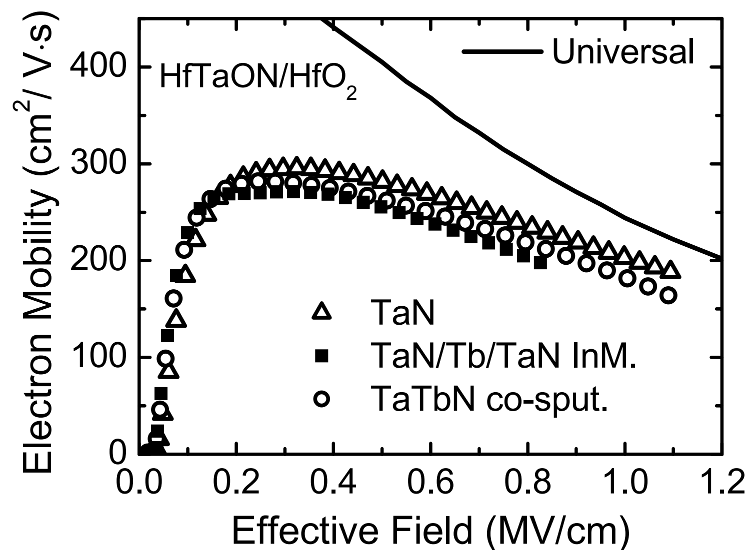


Fig. 5.14 Effective electron mobility in n-MOSFETs with TaN, TaN/Tb/TaN (InM) and co-sputtered $Ta_{0.9}Tb_{0.1}N_y$ metal gates on HfTaON dielectric, measured by split *C-V* measurement. (W/L = 200 μm / 20 μm)

5.2.5 Dual Work Function Metal Gate Integration using InM

Dual metal gate integration using the proposed InM method is also successfully demonstrated. TaN/Tb/TaN and TaN/Ti/HfN [13] laminate structures were integrated on a single wafer for NMOS and PMOS, respectively, as a prototype demonstration of the integration scheme depicted in Fig. 5.6. In this process, it is important to note that the capping layer on top of the P-type metal gate should be thick enough to prevent the N-type material in affecting the WF of the P-type metal gate stack (Fig. 5.6). In the other words, the HfN layer in the TaN/Ti/HfN stack should be thick enough to prevent Tb from diffusing down. It is observed in Fig. 5.15 (a) that HfN of 15 Å is not thick enough to block Tb from reaching the metal-dielectric interface, as evidenced by the negatively V_{FB} shift after 1000 ^{o}C RTA. To eliminate this problem, we increases the HfN thickness to about 100 Å, and found that the Tb layer no longer affect the EWF of the TaN/Ti/HfN stack in PMOS region, as shown in Fig. 5.15 (b). Therefore HfN with thickness of 100 Å was used in our process, though this may not be an optimal value.

Now we start to discuss the integration process. After gate dielectric growth, TaN/Ti/HfN (~2nm/1nm/10nm) stack was first deposited on SiO_2. HfN and Ti were then selectively removed by DHF (1:200) solution in NMOS region, where the ultra-thin TaN of ~2nm was used to protect the underlying gate dielectric in metal etching process. In order to check whether the DHF process can cause any change to the TaN buffer layer, root mean square (rms) roughness of the ultra-thin TaN film has been examined by AFM measurement. As illustrated in Fig. 5.16, the rms roughness of the TaN film shows no significant difference after the DHF process, indicating TaN is robust enough to the DHF solution used in our process.

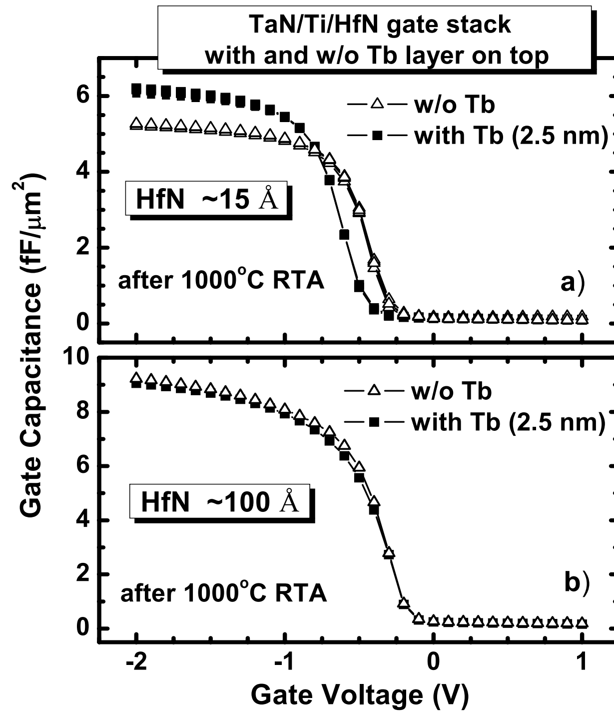


Fig. 5.15 *C-V* characteristics of TaN/Ti/HfN metal stack with and without Tb on top, where the HfN thickness is (a) ~15 Å and (b) ~100 Å.

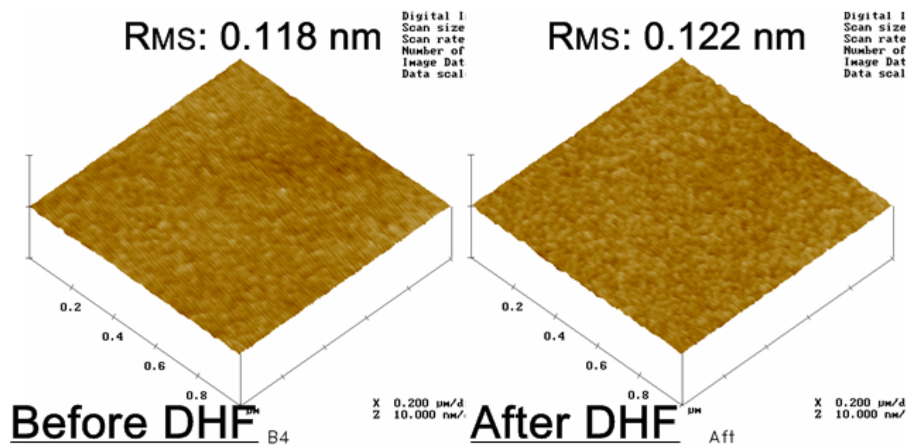


Fig. 5.16 AFM images of TaN (~ 2 nm) deposited on the bare-Si wafer before and after wet etching in DHF (1:200) for 30 sec.

After removal of HfN and Ti layer, ultra-thin Tb and the TaN capping was then deposited to form TaN/Tb/TaN in NMOS region, followed by gate patterning. Fig. 5.17 illustrates the final structures of the N-type and P-type metal gate stacks

integrated on a same wafer. *C-V* and *I-V* characteristics were measured in the as-deposited condition and no obvious difference observed between the capacitors with TaN/Tb/TaN (N-type) and TaN/Ti/HfN (P-type) metal gate stacks, as shown in Fig. 5.18. This indicates clearly that the wet etching process by DHF does not cause any change to the thin TaN buffer layer as well as the underlying dielectric.

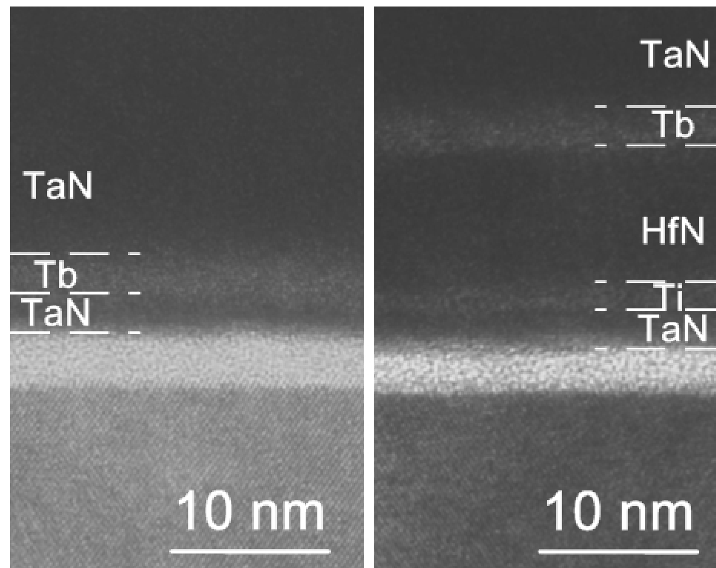


Fig. 5.17 XTEM images of as-deposited dual metal gate stacks on a single wafer: TaN/Tb/TaN (left) for NMOS and TaN/Ti/HfN (right) for PMOS on SiO_2.

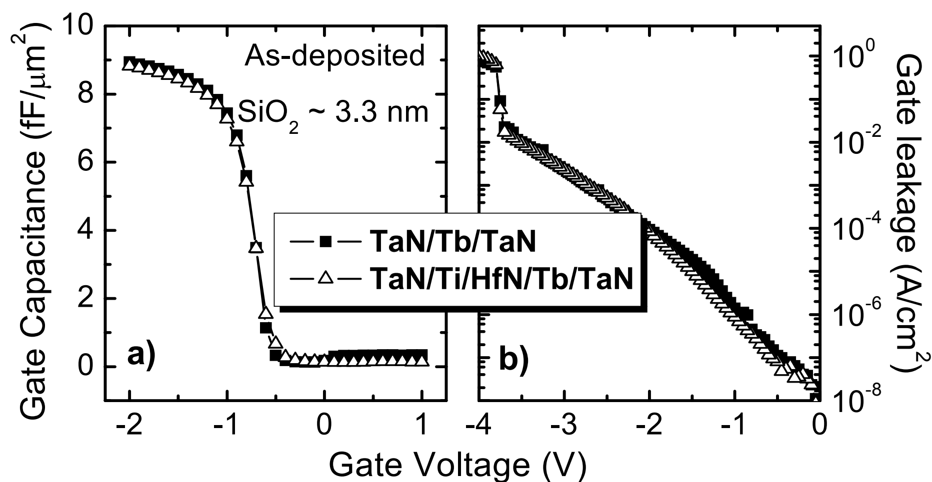


Fig. 5.18 (a) *C-V* and (b) *I-V* characteristics of TaN/Tb/TaN (N-type) and TaN/Ti/HfN (P-type) metal gate stacks on SiO_2 in as-deposited condition.

Finally the devices were subjected to a RTA annealing at 1000 ^{o}C for 1 sec for metal intermixing. This annealing also simulates the S/D dopant activation process used in a gate-first CMOS fabrication. Fig. 5.19 shows the *C-V* measurements of the TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) stacks before and after InM annealing. A V_{FB} difference for above 0.5 V can be achieved through InM, suggesting that the work function of TaN buffer layer is modulated by intermixing with the metals on top. The work function values of the two metal gate stacks are determined to be 4.15 eV and 4.72 eV, respectively, from the V_{FB}-EOT plots shown in Fig. 5.20. Note that the work function of 4.15 eV is ideal for NMOS. But the WF value of 4.72 eV is still not optimal for PMOS and the investigation on other laminate combinations [13] could be one of the possible solutions.

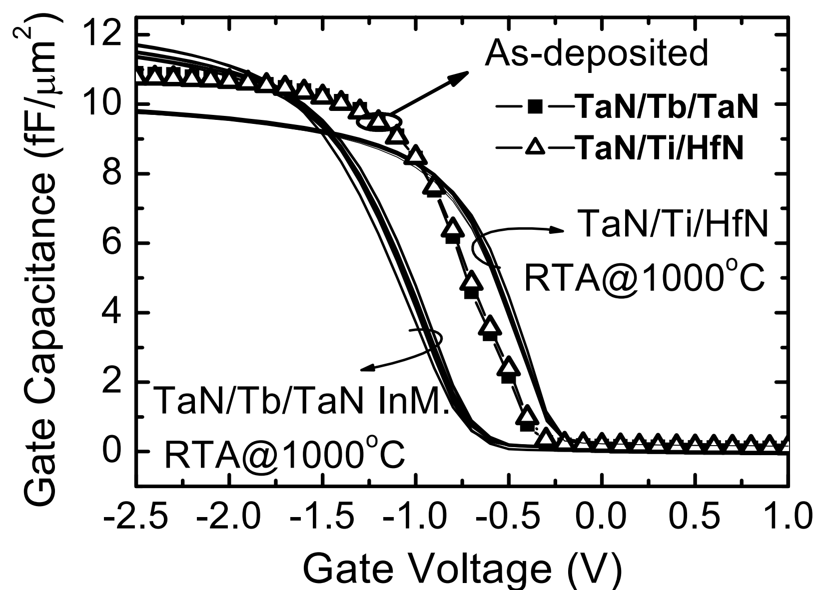


Fig. 5.19 *C-V* characteristics of TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) stacks on SiO_2 before and after 1000 ^{o}C RTA for 1 sec.

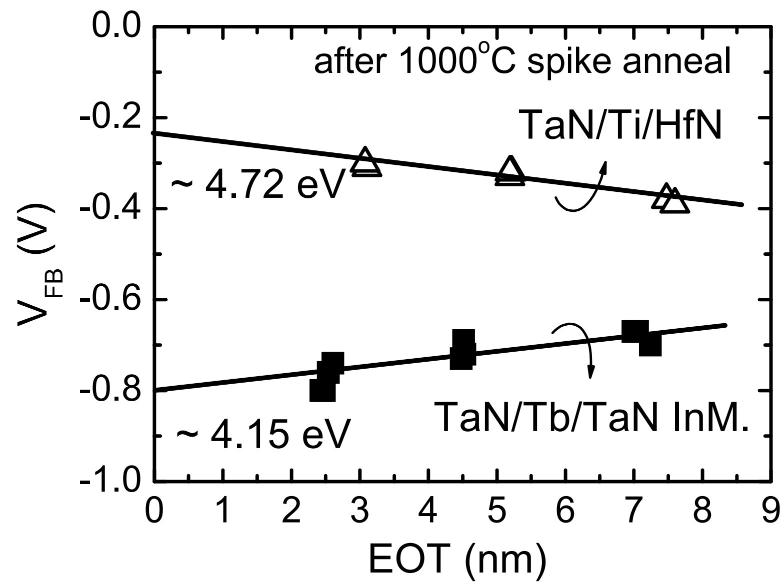


Fig. 5.20 Work Function extraction for TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks on SiO_2 after metal intermixing process.

5.2.6 Summary

In summary, we demonstrate a CMOS compatible dual work function metal gate integration process by using a high-temperature metal intermixing technique. In this process, an ultra-thin TaN buffer layer is used to protect the gate dielectric during metal etching process, and a subsequential high-temperature metal intermixing process is performed to modify the work function of TaN for the requirement of CMOS. Work function of 4.15 and 4.72 eV can be achieved by intermixing of TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks, respectively, after a high-temperature annealing process which can be performed together with the S/D activation process used in conventional CMOS fabrication. Successful V_{th} adjustment for ~300 mV compared with TaN has also been demonstrated on HfTaON dielectric using this metal intermixing technique.

5.3 A Gate-Last Dual Metal Gate Integration Process Employing a Novel HfN Replacement Gate

5.3.1 Motivation

As we discussed in Chapter 2, dual metal gate electrodes with band-edge work functions within ± 0.2 eV from the E_c or E_v of Si are required for the threshold voltage considerations in plenary bulk-Si CMOS technology. Immense challenges have been faced in identifying the potential metal gate candidates with band-edge work functions and integrating them into CMOS process. To date, it is still quite difficult to integrate dual metal gates with work function difference (ΔWF) larger than 0.8 eV on high-κ dielectrics using a gate-first process due to the thermal stability and Fermi-pinning considerations. On the other hand, the gate-last integration process shows great advantages in achieving wide-range work function modulation, thanks to its low thermal budget process flow. From this point of view, the gate-last integration process still stands the chance to be a potential integration scheme for the dual metal gate CMOS and is hence worthy to be studied [21]-[24].

The process flow for a conventional gate-last integration process is illustrated in Fig. 5.5. In this process, the metal gate electrodes and high-κ dielectric are used to replace the poly-Si dummy gate and underlying sacrificial SiO_2, respectively, after the formation of the S/D regions. Nevertheless, a high-temperature annealing of the high-κ dielectric is desirable for better carrier mobility, less fixed charge and less *C-V* hysteresis [24]-[26]. A high-temperature post-deposition-annealing (PDA) of high-κ dielectric without metal gate capping, however, will cause an unwanted EOT increase which is a serious concern for aggressive CMOS scaling [24]. As of the starting of this work, there had been no solution to achieve high-quality high-κ dielectric with sub-1-nm EOT and wide-range WF difference above 0.8 eV simultaneously.

In this work, a novel replacement gate process employing HfN as the dummy gate electrode is demonstrated to integrate dual metal gate electrodes on HfO_2 dielectric with sub-1-nm EOT. The HfN/HfO_2 gate stack shows excellent thermal stability during the S/D dopant activation annealing, allowing sub-1 nm EOT and good dielectric characteristics to be achieved [27]. In addition, the high etching selectivity of HfN with respect to HfO_2 enables the high-quality HfO_2 dielectric to be left after the removal of dummy HfN electrode, with no degradation to the leakage or TDDB reliability characteristics. Replacement of the dummy HfN gate by Ta and Ni has also been demonstrated, with ΔWF of ~0.8 eV achieved. These results make the HfN dummy gate process more attractive than the conventional poly-Si dummy gate process.

5.3.2 Proposed Integration Flow and Device Fabrication

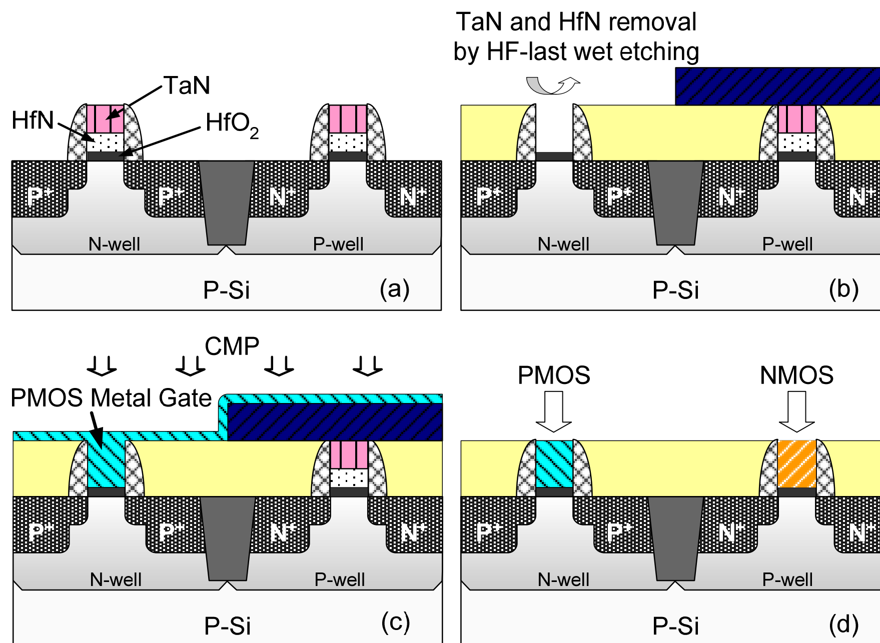


Fig. 5.21 Proposed replacement gate process using HfN as dummy gate: (a) CMOS fabrication using $TaN/HfN/HfO_2$ as the gate stacks; (b) high selective etching of TaN and HfN by wet chemicals; (c) new metal gate deposition and CMP planarization; (d) dual metal gate integration by repeating steps (b)-(c).

Fig. 5.21 illustrates a simplified process flow of the proposed HfN replacement gate process. The feasibility of this process was demonstrated using MOS capacitors. First, the HfN/HfO_2 gate stack was fabricated [27]. Metal-organic chemical vapor deposition (MOCVD) HfO_2 with thickness of ~ 25 Å was deposited on p-Si (100) substrate (N_A=6×10^{15} cm^{-3}) followed by PDA at 700 °C in N_2 ambient. A TaN/HfN stack (1000 Å/500 Å) was then deposited on HfO_2 by dc magnetron sputtering. After gate patterning, all the devices were subjected to RTA in N_2 at 1000°C for 20 sec, which is adequate for S/D dopant activation. The TaN/HfN dummy stack was then removed by standard cleaning-1 (SC-1) (NH_4OH+H_2O_2+H_2O) solution at 80°C and DHF (1:100) solution at room temperature, respectively. It should be noted that the SC-1 solution does not attack the HfN metal. Finally, Ta and Ni were deposited on HfO_2 gate dielectric as the new gate electrodes. For some devices, HfN was re-deposited onto HfO_2 to compare with those fresh devices to check the impact of the DHF etching process to the dielectric properties. Backside Al metallization and FGA were also performed to all the devices.

5.3.3 Results and Discussions

HfN has been demonstrated as a thermally robust metal gate electrode on HfO_2 [27], making it feasible to scale down the EOT of the HfN/HfO_2 gate stack to less than 1 nm after 1000 °C RTA annealing. Good transistor electrical characteristics have also been achieved [28]. The work function of HfN, however, is close to the mid-gap level of silicon. By replacing the HfN with Ta or Ni using the proposed replacement gate process, dual metal gates with large work function difference can be realized while maintaining the good properties of the HfO_2 gate dielectric.

To study the feasibility of the proposed HfN dummy gate process, we first investigate the selectivity of HfN with respect to HfO_2 in DHF solution. Fig. 5.22 compares the etch rates of HfN and HfO_2 in DHF (1:100) solution. The etch rate of HfN is around 12 nm/min while that of HfO_2 is almost negligible if 1000 °C PDA treatment has been performed. This demonstrates the very high etching selectivity of the HfN over HfO_2. Fig. 5.23 examines the surface morphology of the HfO_2 films under three different conditions: as-deposited HfO_2 film, HfO_2 after 1000 °C RTA anneal, and HfO_2 in HfN/HfO_2 stack with HfN removed by DHF solution after 1000 °C RTA anneal. The rms variation induced by the DHF etching process is only about 0.05 nm, indicating that the HfN removal process has negligible physical impact to the HfO_2 film.

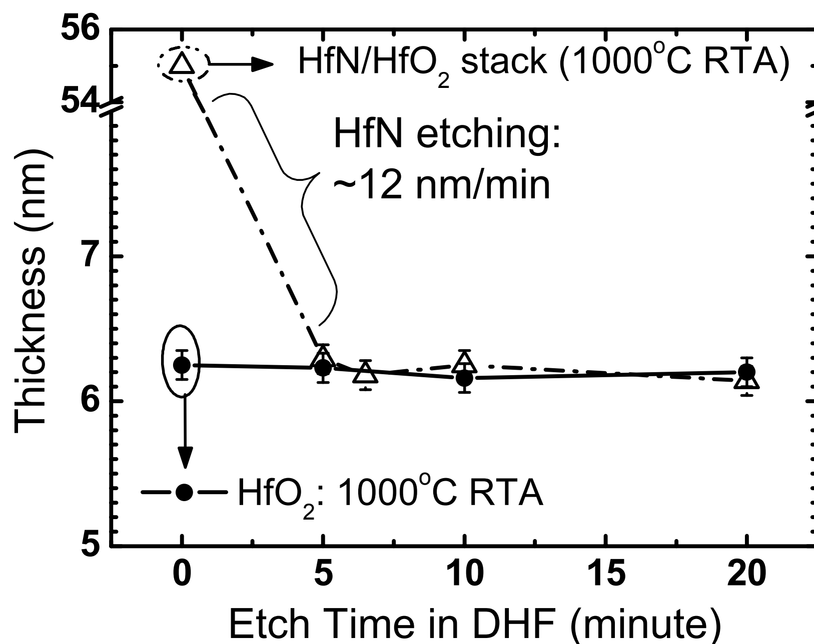


Fig. 5.22 Etching properties of the HfN/HfO_2 gate stack (open triangle symbol) and the HfO_2 (solid symbol) film after 1000 °C RTA process in diluted HF solution (1:100). The etch rate of HfN is determined by surface profiler, and the remaining HfO_2 thickness is measured by ellipsometer.

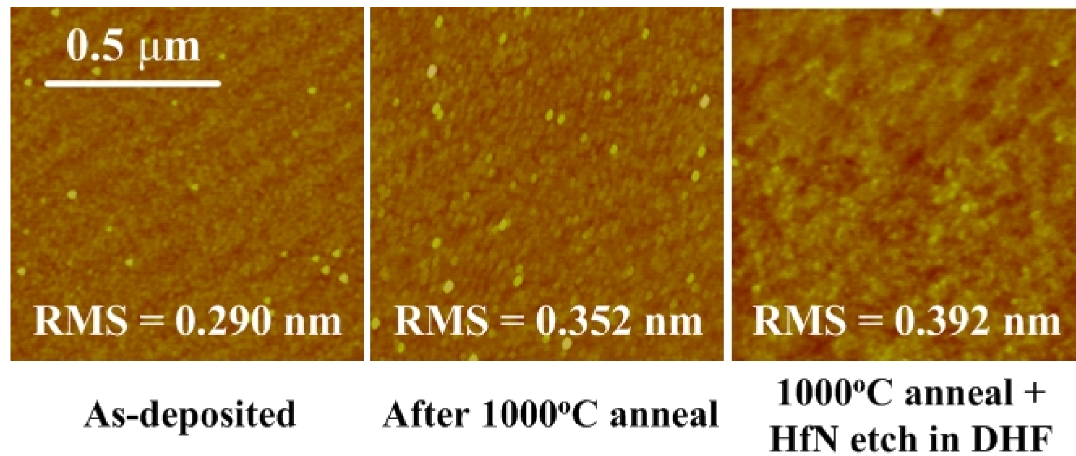


Fig. 5.23 AFM images of HfO_2 with different process history: as-deposited HfO_2 film, HfO_2 after 1000 °C RTA anneal, and HfO_2 in HfN/HfO_2 stack with HfN removed by DHF solution after 1000 °C RTA anneal.

To further examine the potential influence of the wet etching process to the electrical properties of HfO_2, HfN metal gate is re-deposited onto HfO_2 dielectric after removing the HfN dummy gate. Fig. 5.24 compares the *C-V* and *I-V* characteristics of the "control" HfN/HfO_2 devices (1000 ^oC RTA, fresh device without HfN removal process) and the "re-deposited" HfN/HfO_2 devices (re-deposit HfN after the removal of HfN dummy gate by DHF solution). The *C-V* characteristics were measured at different frequencies (100 KHz and 1 MHz). No significant frequency dispersion observed from both of the samples, indicating the series resistance from the gate and substrate is not an issue to the *C-V* measurement. The EOT of the two gate stacks can thus be extracted and a low EOT value of ~0.83 nm is obtained by fitting the *C-V* measurement with simulated curves which takes quantum mechanical effect into account. It is found that the EOT and the gate leakage of the two devices are almost identical, as shown in Fig. 5.24. These results suggest that the ultra-thin HfO_2 film is not etched or damaged. The small V_{FB} difference between the "control" and "re-deposited" HfN/HfO_2 devices could be due to the different thermal history employed

to the HfN metal gate in these two devices, or a possible plasma damage effect during the re-deposition of HfN.

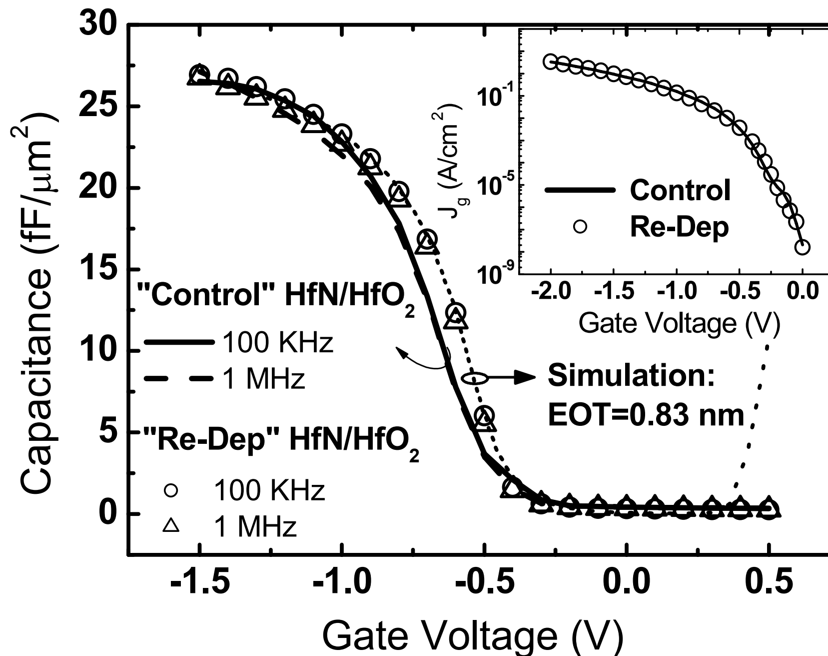


Fig. 5.24 *C-V* and *I-V* (inset) characteristics of the "control" HfN/HfO_2 devices and "re-deposited" HfN/HfO_2 devices with HfO_2 EOT~0.83 nm. The *C-V* curves were measured at 100 kHz and 1 MHz on devices with an area of 50×50 μm^2.

The above observations allow the HfN to be used as a dummy gate material on HfO_2 using the proposed replacement gate process. Replacement of HfN dummy gate by Ta and Ni are then demonstrated for the applications of NMOS and PMOS, respectively. Fig. 5.25 shows the high frequency *C-V* measurements for the Ta/HfO_2 and Ni/HfO_2 devices using the HfN replacement gate process. The *C-V* measurements of all the devices fit well with the simulation curves, indicating negligible changes of the HfO_2/Si interface quality during the HfN removal process. The *C-V* hysteresis for all the devices is less than 20 mV thanks to the good dielectric quality of HfO_2. The work function shifts attributed to Ta and Ni with respect to HfN

are -0.35 eV and +0.45 eV, respectively. The work function difference of 0.8 eV between the two gate electrodes could be adequate for good device performance for bulk-Si CMOS transistors. More importantly, the working devices with "re-deposited" Ni on ultra-thin HfO_2 (EOT ~0.9nm) are successfully demonstrated using this novel replacement gate process (inset of Fig. 5.25), indicating good scalability of this process for next generation MOSFETs. However, the "re-deposited" Ta-gated device did not survive on such a thin HfO_2, which is presumably associated with the higher tunneling current through HfO_2 due to the relatively lower WF of Ta than that of Ni, and/or some possible plasma damage effects. Process optimization will be needed in future study.

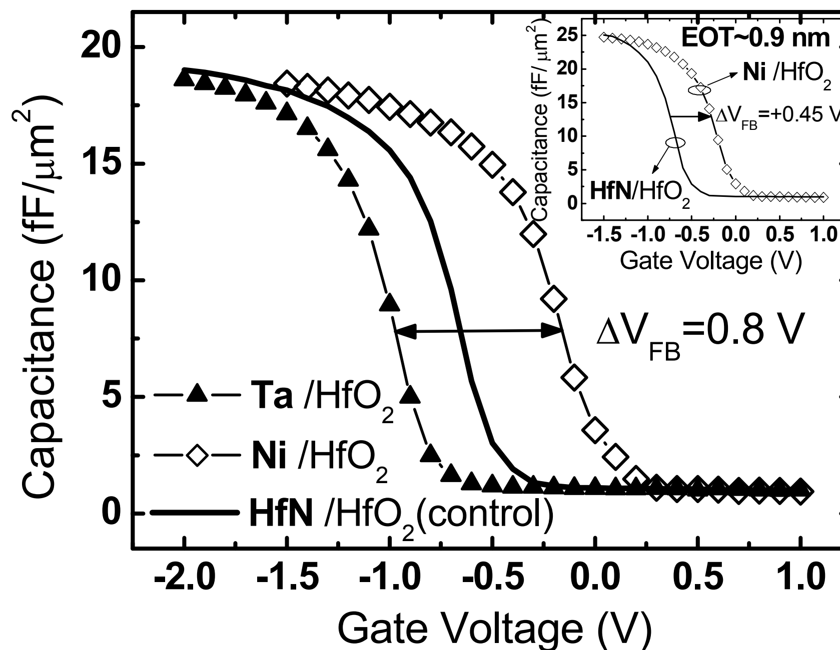


Fig. 5.25 High frequency *C-V* curves of the HfN/HfO_2 "control" devices and "re-deposited" Ta/HfO_2, Ni/HfO_2 devices. The inset compares the *C-V* curves measured from the "re-deposited" Ni/HfO_2 devices with ultra-thin HfO_2 (EOT~0.9 nm) and that of a "control" HfN/HfO_2 device. All the *C-V* curves were measured at 100 kHz on devices with an area of 50×50 μm^2.

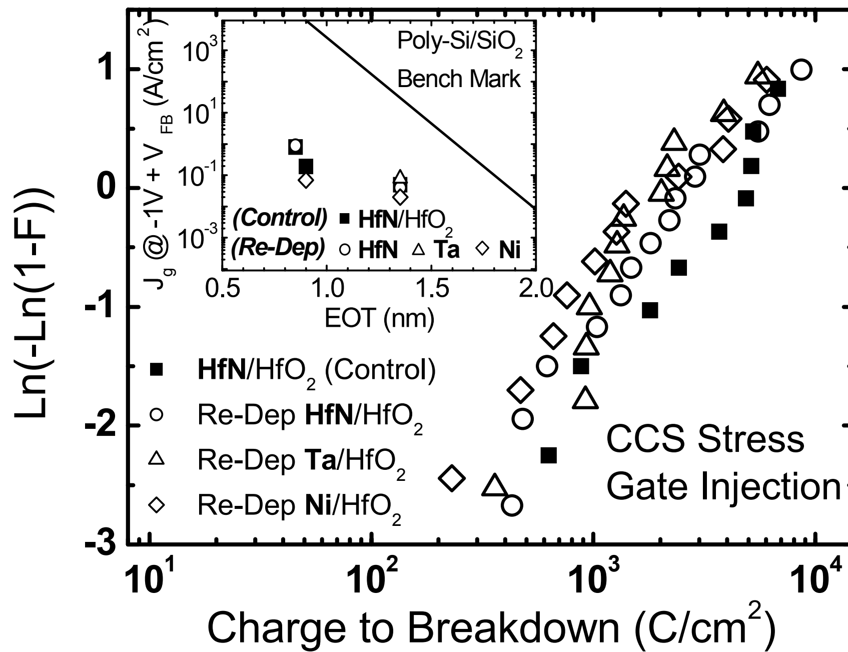


Fig. 5.26 Comparison of TDDB and gate leakage (inset) characteristics between the "control" HfN/HfO_2 devices and "re-deposited" HfN/HfO_2, Ta/HfO_2, Ni/HfO_2 devices. For the TDDB study, CCS with a current density of ~8 A/cm^2 was performed on devices with an area of 100×100 μm^2 at room temperature.

Finally, we compare the gate leakage and TDDB characteristics of the "re-deposited" HfN/HfO_2, Ta/HfO_2 and Ni/HfO_2 devices with that of the "control" HfN/HfO_2 stack, as depicted in Fig. 5.26. It is observed that the gate leakage of the Ta/HfO_2 (or Ni/HfO_2) devices is slightly higher (or lower) than that of the HfN/HfO_2 "control" devices (inset of Fig. 5.26). This can be attributed to the lower (or higher) work function of Ta (or Ni) with respect to HfN. The TDDB characteristics are also investigated by applying a constant current stress (CCS) to the samples with different gate electrodes. The EOT of these devices is about 1.35 nm. No significant degradation in TDDB characteristics was observed by using this HfN replacement gate process, as shown in Fig. 5.26. More reliability study will be required to qualify this process systematically.

It should be noted that other metal gate candidates, like Ta-Ru or Hf-Mo alloys, can also be integrated using this HfN replacement gate process as potential solutions for the dual metal gate CMOS technology towards sub-1-nm EOT regime.

5.3.4 Summary

In summary, we proposed a novel replacement gate process employing HfN as the dummy gate material for the integration of dual metal gates on HfO_2 gate dielectric with sub-1-nm EOT. The excellent thermal stability of the HfN/HfO_2 gate stack and the high etch selectivity between HfN and HfO_2 allows an ultra-thin, high-quality, and damage-free HfO_2 gate dielectric to be achievable. Replacing the HfN dummy gate by Ta and Ni metal gates using the proposed process has been successfully demonstrated, with large work function difference for about 0.8 eV achieved.

5.4 Conclusion

In this Chapter, we demonstrated two integration schemes for dual metal gate CMOS integration. The first one is a novel gate-first integration process using a high-temperature metal intermixing technique. In this process, a TaN buffer layer is used to avoid the gate dielectric being exposed during the metal etching process. This addresses the etching damage concerns associated with the conventional direct-etching integration scheme. The work function of the TaN buffer layer can be modulated by a following high-temperature metal intermixing process, which is compatible with the conventional gate-first CMOS process flow. By using this integration scheme, dual work function of 4.15 and 4.72 eV have been achieved in

TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks, respectively. Successful V_{th} adjustment on HfTaON dielectric has also been demonstrated.

The second integration scheme proposed in this thesis is a gate-last replacement gate process employing HfN as a novel dummy gate electrode, which enables the high-quality HfO_2 gate dielectric with sub-1 nm EOT and a wide work function tuning range to be achieved simultaneously for bulk-Si CMOS applications. High quality HfN/HfO_2 stack with HfO_2 EOT less than 1 nm can be achieved, due to the good thermal stability of the HfN/HfO_2 stack, and large work function difference for about 0.8 eV can be realized by using Ta and Ni to replace the HfN dummy gate electrode. The EOT, gate leakage, and TDDB characteristics of the ultra-thin HfO_2 dielectric are observed not to be affected by the HfN dummy gate removal process.

These proposed novel integration schemes may provide some useful discussions to address some of the major issues associated with the conventional integration schemes, and are believed to make a contribution to the development of the dual metal gate integration processes for the future CMOS technology.

References

1. Q. Lu, Y.-C. Yeo, P. Ranade, H. Takeuchi, T.-J. King, C. Hu, S. C. Song, H. F. Luan, and D.-L. Kwong, "Dual-metal gate technology for deep-sub-micron CMOS transistors," in *VLSI Tech. Dig.*, pp. 72-73, 2000.

2. S. B. Samavedam, L. B. La, J. Smith, S. Dakshina-Murthy, E. Luckowski, J. Schaeffer, M. Zavala, R. Martin, V. Dhandapani, D. Triyoso, H. H. Tseng, P. J. Tobin, D. C. Gilmer, C. Hobbs, W. J. Taylor, J. M. Grant, R. I. Hegde, J. Mogab, C. Thomas, P. Abramowitz, M. Moosa, J. Conner, J. Jiang, V. Arunachalam, M. Sadd, B.-Y. Nguyen, and B. White, "Dual-metal gate CMOS with HfO_2 gate dielectric," in *IEDM Tech. Dig.*, pp. 433-436, 2002.

3. Z. B. Zhang, S. C. Song, C. Huffman, J. Barnett, N. Moumen, H. Alshareef, P. Majhi, M. Hussain, M.S. Akbar, J.H. Sim, S.H. Bae, B. Sassman, and B.H. Lee, "Integration of dual metal gate CMOS with TaSiN (NMOS) and Ru (PMOS) gate electrodes on HfO_2 gate dielectric," in *VLSI Tech. Dig.*, pp. 50-51, 2005.

4. M. M. Hussain, N. Moumen, J. Barnett, J. Saulters, D. Baker, M. Akbar and Z. Zhang, "Metal wet etch process development for dual metal gate CMOS using standard industry tools," *Proc. ECS 207th meeting*, 629, Quebec, Canada, 2005.

5. S. C. Song, Z. B. Zhang, M. M. Hussain, C. Huffman, J. Barnett, S. H. Bae, H. J. Li, P. Majhi, C. S. Park, B. S. Ju, H. K. Park, C. Y. Kang, R. Choi, P. Zeitzoff, H. H. Tseng, B. H. Lee, and R. Jammy, "Highly manufacturable 45nm LSTP CMOSFETs using novel dual high-k and dual metal gate CMOS integration," in *VLSI Tech. Dig.*, pp. 16-17, 2006.

6. I. Polishchuk, P. Ranade, T.-J. King, and C. Hu, "Dual work function metal gate CMOS technology using metal interdiffusion," *IEEE Electron Device Lett.*, vol. 22, pp. 444-446, Sep. 2001.

7. J. Lee, H. Zhong, Y. Suh, G. Heuss, J. Gurganus, B. Chen, and V. Misra, "Tunable work function dual metal gate technology for bulk and non-bulk CMOS," in *IEDM Tech. Dig.*, pp. 359-362, 2002.

8. B.-Y. Tsui, and C.-F. Huang, "Wide range work function modulation of binary alloys for MOSFET application," *IEEE Electron Device Lett.*, vol. 24, pp. 153-155, Mar. 2003.

9. T.-L. Li, C.-H. Hu, W.-L. Ho, H. C.-H. Wang, and C.-Y. Chang, "Continuous and precise work function adjustment for integratable dual metal gate CMOS technology using Hf–Mo binary alloys," *IEEE Trans. Electron Devices*, vol. 52, pp. 1172-1179, 2005.

10. T.-L. Li, W.-L. Ho, H.-B. Chen, H. C.-H. Wang, C.-Y. Chang, and C. Hu, "Novel dual-metal gate technology using Mo-$MoSi_x$ combination," *IEEE Trans. Electron Devices*, vol. 53, pp. 1420-1426, 2006.

11. A. Veloso, T. Hoffmann, A. Lauwers, S. Brus, J.-F. de Marneffe, S. Locorotondo, C. Vrancken, T. Kauerauf, A. Shickova, B. Sijmus, H. Tigelaar, M. A. Pawlak, H.Y. Yu, C. Demeurisse, S. Kubicek, C. Kerner, T. Chiarella, O. Richard, H. Bender, M. Niwa, P. Absil, M. Jurczak, S. Biesemans and J. A. Kittl, "Dual work function phase controlled Ni-FUSI CMOS (NiSi NMOS, Ni_2Si or $Ni_{31}Si_{12}$ PMOS): manufacturability, reliability & process window improvement by sacrificial SiGe cap," in *VLSI Tech. Dig.*, pp. 96-97, 2006.

12. C. S. Park, B. J. Cho, A. Y. Du, N. Balasubramanian, and D.-L. Kwong, "A novel approach for integration of dual metal gate process using ultra thin aluminum nitride buffer layer," in *VLSI Tech. Dig.*, pp. 149-150, 2003.

13. W. P. Bai, S. H. Bae, H. C. Wen, S. Mathew, L. K. Bera, N. Balasubramanian, N. Yamada, M. F. Li, and D.-L. Kwong, "Three-layer laminated metal gate electrodes with tunable work functions for CMOS applications," *IEEE Electron Device Lett.*, vol. 26, pp. 231-233, Apr. 2005.

14. X. Yu, C. Zhu, M. Yu, M. F. Li, A. Chin, C. H. Tung, D. Gui, and D.-L. Kwong, "Advanced MOSFETs using HfTaON/SiO_2 gate dielectric and TaN metal gate with excellent performances for low standby power application," in *IEDM Tech. Dig.*, 2005, pp. 31-34.

15. H. J. Lee, T. Cagin, W. L. Johnson, W. A. Goddard III, "Criteria for formation of metallic glasses: the role of atomic size ratio," *J. Chem. Phys.*, vol 119, pp. 9858-9870, Nov. 2003.

16. J. K. Schaeffer, S.B. Samavedam, D.C. Gilmer, V. Dhandapani, P.J. Tobin, J. Mogab, B.-Y. Nguyen, B.E. White, Jr., S. Dakshina-Murthy, R.S. Rai, Z.-X. Jiang, R. Martin, M.V. Raymond, M. Zavala, L.B. La, J.A. Smith, R. Garcia, D. Roan, M. Kottke, and R.B. Gregory, "Physical and electrical properties of metal gate electrodes on HfO_2 gate dielectrics," *J. Vac. Sci. Technol. B*, vol. 21, pp. 11-17, Jan./Feb. 2003.

17. C. D. Gelatt, Jr. and H. Ehrenreich, "Charge transfer alloys: AgAu," *Phys. Rev. B*, vol. 10, pp. 398-415, Jul. 1974.

18. O. V. Lounasmaa and P. R. Roach, "Specific heat of terbium metal between 0.37 and 4.2 ^{o}K," *Phys. Rev.*, vol. 128, pp. 622-626, Oct. 1962.

19. I. S. Jeon, J. Lee, P. Zhao, P. Sivasubramani, T. Oh, H. J. Kim, D. Cha, J. Huang, M. J. Kim, B. E. Gnade, J. Kim, and R. M. Wallace, "A novel methodology on tuning work function of metal gate using stacking bi-metal layers," in *IEDM Tech. Dig.*, pp. 303-306, 2004.

20. H. Kim, P. C. Mclntyre, C. O. Chui, K. C. Saraswat, and S, Stemmer, "Engineering chemically abrupt high-k metal oxide/silicon interfaces using an oxygen-gettering metal overlayer," *J. Appl. Phys.*, vol. 96, pp. 3467-3472, 2004.

21. A. Chatterjee, R.A. Chapman, K. Joyner, M. Otobe, S. Hattangady, M. Bevan, G.A. Brown, H. Yang, Q. He, D. Rogers, S.J. Fang, R. Kraft, A.L.P. Rotondaro,

M. Terry, K. Brennan, S.-W. Aur, J.C. Hu, H.-L. Tsai, P. Jones, G. Wilk, M. Aoki, M. Rodder, and I.-C. Chen, “CMOS metal replacement gate transistors using tantalum pentoxide gate insulator,” in *IEDM Tech. Dig.*, pp. 777-780, 1998.

22. A. Yagishita, T. Saito, K. Nakajima, S. Inumiya, Y. Akasaka, Y. Ozawa, K. Hieda, Y. Tsunashima, K. Suguro, T. Arikado, and K. Okumura, “High performance damascene metal gate MOSFET’s for 0.1 μm regime,” *IEEE Trans. Electron Devices*, vol. 47, pp. 1028-1034, May 2000.

23. J. Pan, C. Woo, C.-Y. Yang, U. Bhandary, S. Guggilla, N. Krishna, H. Chung, A. Hui, B. Yu, Q. Xiang, and M.-R. Lin, “Replacement metal-gate NMOSFETs with ALD TaN/EP-Cu, PVD Ta, and PVD TaN electrode,” *IEEE Electron Device Lett.*, vol. 24, pp. 304-305, May 2003.

24. B. Guillaumot, X. Garros, F. Lime, K. Oshima, B. Tavel, J. Chroboczek, P. Masson, R. Truche, A. M. Papon, F. Martin, J. F. Damlencourt, S. Maitrejean, M. Rivoire, C. Leroux, S. Cristoloveanu, G. Ghibaudo, J. L. Autran, T. Skotnicki, and S. Deleonibus, “75nm damascene metal gate and high-k integration for advanced CMOS devices,” in *IEDM Tech. Dig.*, pp. 355-358, 2002.

25. Y. Kim, G. Gebara, M. Freiler, J. Barnett, D. Riley, J. Chen, K. Torres, J. Lim, B. Foran, F. Shaapur, A. Agarwal, P. Lysaght, G. A. Brown, C. Young, S. Borthakur, H.-J. Li, B. Nguyen, P. Zeitzoff, G. Bersuker, D. Derro, R. Bergmann, R. W. Murto, A. Hou, H. R. Huff, E. Shero, C. Pomarede, M. Givens, M. Mazanec, and C. Werkhoven, “Conventional n-channel MOSFET devices using single layer HfO_2 and ZrO_2 as high-k gate dielectrics with polysilicon gate electrode,” in *IEDM Tech. Dig.*, pp. 455-458, 2001.

26. G. D. Wilk, M. L. Green, M.-Y. Ho, B. W. Busch, T. W. Sorsch, F. P. Klemens, B. Brijs, R. B. van Dover, A. Kornblit, T. Gustafsson, E. Garfunkel, S. Hillenius, D. Monroe, P. Kalavade, and J. M. Hergenrother, “Improved film growth and flatband voltage control of ALD HfO_2 and Hf-Al-O with n^+ poly-Si gates using chemical oxides and optimized post-annealing,” in *VLSI Tech. Dig.*, pp. 88-89, 2002.

27. H. Y. Yu, J. F. Kang, J. D. Chen, C. Ren, Y. T. Hou, S. J. Whang, M.-F. Li, D. S. H. Chan, K. L. Bera, C. H. Tung, A. Du, and D.-L. Kwong, “Thermally robust high quality HfN/HfO_2 gate stack for advanced CMOS devices,” in *IEDM Tech. Dig.*, pp. 99-103, 2003.

28. J. F. Kang, H. Y. Yu, C. Ren, M.-F. Li, D. S. H. Chan, X. Y. Liu, and D.-L. Kwong, “Ultrathin HfO_2 (EOT<0.75 nm) gate stack with TaN/HfN electrodes fabricated using a high-temperature process,” *Electrochemical and Solid-State Lett.*, vol. 8, G311-G313, 2005.

Chapter 6

Conclusion

6.1 Summary

This work has sought to address some of the most pressing issues in advanced metal gate technology. As the time of writing this dissertation, the 65 nm CMOS technology is being introduced into production, and the 45 nm CMOS technology is under development at most leading semiconductor manufacturers. Concurrently, the whole semiconductor community is initiating the research and development efforts for the 32 nm technology node and beyond, where it is likely that some of the materials and process integration challenges discussed herein must be addressed. The task for finding a replacement material for poly-silicon gate electrode is by no means trivial. Numerous problems related to the material selection and the process integration of metal gates (deposition, lithography, etching, etc.) need to be addressed before a suitable metal gate solution can be transferred to manufacturing.

6.1.1 Understanding the Metal-Dielectric Interface

The metal-dielectric interface is important since it directly affects the effective work function of metal gates. Although the *metal-induced gate states* and the related interface dipole theory has been pretty successful in describing the dependence of

metal effective work function on underlying dielectric, the variation of metal gate work function during the thermal annealing process used in CMOS fabrication still need to be understood and addressed with scientific and technological importance.

In Chapter 3, the dependence of metal gate work function on the annealing temperature was investigated on different dielectrics. The *extrinsic states* created at the metal-dielectric interface was identified to be the major factor responsible for the instability of metal gate effective work function during the high-temperature annealing process. The creation of *extrinsic states* appears to be thermodynamically driven and becomes more pronounced when the annealing temperature is higher, making the Fermi-level pinning effect to be more serious after RTA at higher temperature. The chemical bond configurations at the metal-dielectric interface could be correlated with the creation of *extrinsic states*. In general, the Hf-Si bond tends to create *extrinsic states* upon annealing while Hf-Hf or Si-Si bonds would be less pronounced. A model considering the impact of *extrinsic states* has also been proposed to qualitatively explain the dependence of metal effective work function on annealing process. This understanding on the metal-dielectric interface could be useful for work function tuning and interface engineering of the metal gate electrodes in future MOSFETs.

6.1.2 Lanthanide-Incorporated Metal Nitride Gate Electrodes

Identifying the proper metal gate candidates for CMOS application is a key challenge for the implementation of metal gates in 32 nm technology node and beyond. Many issues need to be considered systematically in developing the appropriate metal gate materials. One of the most challenging issues for the development of a N-type metal gate candidate is the difficulty to achieve good

thermal stability required for gate-first process and a low work function suitable for NMOS applications simultaneously.

Chapter 4 discussed a possible solution of the N-type metal gate materials for the applications in a gate-first CMOS process. For the first time, we demonstrated that lanthanide elements can be very useful in modulating the work function of metal gates, and this provides a new way for metal gate work function engineering. In this work, lanthanide elements with very low work function are combined with metal nitride materials to get a best trade-off between thermal stability and low work function. The physical and electrical properties of lanthanide-incorporated metal nitrides are systematically studied. By varying the lanthanide concentration in lanthanide-MN_x, the work function of lanthanide-MN_x can be tuned continuously and a work function value of 4.2~4.3 eV can be obtained even after a 1000 ^oC RTA treatment, promising for NMOS devices using a gate-first CMOS process flow. Our study also indicates that the nitrogen concentration is an important parameter in influencing the properties of lanthanide-MN_x, especially its thermal stability. Hence the nitrogen content needs to be carefully engineered in process. Moreover, we also demonstrated the compatibility of this method with high-κ dielectrics. It is believed that these results will be of considerable practical value for the development of metal gate technology.

6.1.3 Process Integration of Dual Metal Gates

Another effort of this thesis is a discussion on the dual metal gate integration issues for CMOS applications. Introducing novel materials into the well-established CMOS platform is always accompanied by innovations in process integration technology. However, to date, there has been no integration scheme which addresses

all the concerns in dual metal gate integration. The goal of this thesis has been to identify the issues and limitations in the existing dual metal gate integration schemes and propose novel integration schemes to address these potential problems.

Two dual metal gate integration schemes have been studied in Chapter 5. The first one is aimed at addressing the etching damage issues associated with the conventional direct-etching method used for dual metal gate integration. A novel gate-first dual metal gate CMOS integration process by using a high-temperature metal intermixing process was demonstrated for the first time. In this process, a TaN buffer layer is used to protect the gate dielectric and avoid the exposure of gate dielectric during the selective metal etching process. The work function of the TaN buffer layer can be modulated by a subsequent high-temperature metal intermixing technique, which is compatible with the conventional gate-first CMOS process flow. By using this integration scheme, dual work function of 4.15 and 4.72 eV has been achieved in TaN/Tb/TaN (NMOS) and TaN/Ti/HfN (PMOS) metal stacks, respectively. Successful V_{th} adjustment on high-κ HfTaON dielectric has also been demonstrated using this method.

In the second part of Chapter 5, the feasibility of a gate-last replacement gate process employing HfN as a novel dummy gate electrode has also been investigated. The advantage of this process is that the high-quality HfO_2 gate dielectric with EOT scalability down to sub-1 nm regime can be integrated with dual metal gates with a wide range work function difference for above 0.8 eV, which is very attractive for future bulk-Si CMOS devices. Due to the good thermal stability of HfN/HfO_2 stack, high-quality HfO_2 dielectric with EOT less than 1 nm can be achieved using a gate-first process. The dummy HfN gate can then be selectively removed from HfO_2 so that other metal gate candidates with suitable WF for bulk-Si CMOS can be integrated

on the high-quality HfO_2 dielectric. As a prototype demonstration, Ta and Ni are used to replace HfN and large work function difference for about 0.8 eV has been demonstrated. The EOT, gate leakage, and TDDB characteristics of the ultra-thin HfO_2 dielectric are also not affected by the HfN dummy gate removal process, suggesting the potential of this process for bulk-Si CMOS applications.

3.2 Suggestions for Future Work

The work in this thesis has been very exploratory. More detailed investigation and more rigorous characterization will be necessary to further optimize the processs described in this thesis. Suggestions for future work will be directly or indirectly related to the concerns described earlier in this thesis.

In Chapter 3, a metal-dielectric interface model taking the role of *extrinsic states* into account has been proposed to explain the work function thermal instability. However, in order to precisely engineer the effective work function of metal gates in process, the source of the *extrinsic defects* need to be further identified and the properties of these *extrinsic states*, such as their energy levels, need to be studied by physical analysis or atomistic simulation. On the other hand, approaches to minimizing the impact of the *extrinsic defects* and improving the thermal stability of metal gate work functions during process would be an important question for discussion in future work.

For the development of lanthanide-MN_x metal gates, the etching and cleaning issues need to be well optimized in order to examine the feasibility of these materials in short-channel transistor fabrication. Moreover, the co-optimization of these metal

gates with state-of-the-art high-κ candidates, like HfSiON or HfSiO, also needs to be emphasized in future work. The mobility and charge-trapping characteristics in MOSFETs with lanthanide-MN_x metal gates should be investigated. In addition, the possible interface interaction between lanthanide elements and the underlying dielectrics needs to be further characterized and its impact on device reliability should be evaluated.

In addition, developing a suitable P-type metal gate material with work function higher than 5.0 eV would be an urgent task for the dual metal gate CMOS technology. It has been reported that the effective work function of many P-type metal gate candidates, such as Pt, Ru, and Re, on high-κ HfO_2 will be affected by the oxygen vacancy induced dipoles, rendering the resulting threshold voltage too high for p-MOSFET [1]. This problem could possibly be addressed by engineering the high-κ dielectric to reduce the oxygen vacancies near the metal-dielectric interface [2]. Therefore, the co-optimization of the metal/high-κ stack should be emphasized in future research to get optimal device performance.

Moreover, for the proposed dual metal gate integration scheme using high-temperature metal intermixing technique, some interesting phenomenon are still worthy to be studied in future work. The dependence of the work function modulation range on a number of variables, including the buffer layer thickness, the buffer layer composition, the selection of metal materials, and the annealing conditions, should be systematically investigated in more detail. A model to depict the work function tunability of this intermixing process should be developed to guide

the selection of metal materials. The performance of the P-type metal stacks also needs to be further optimized.

Finally, the applications of metal gate electrodes in multi-gate or other non-planar CMOS devices should also be studied. FUSI gate process could be a potential solution for these kinds of applications. Process integration of FUSI metal gates in these 3-D structures could be challenging due to the difficulties in controlling the phase/dopant precisely in 3-D structures, and hence it needs to be carefully developed, particularly for devices in 32-nm technology node and beyond where the gate length will be less than 15 nm.

References

1. E. Cartier, F. R. McFeely, V. Narayanan, P. Jamison, B. P. Linder, M. Copel, V. K. Paruchuri, V.S. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, G. Shahidi, "Role of oxygen vacancies in V_{FB}/V_t stability of pFET metals on HfO_2," in *VLSI Tech. Dig.*, pp. 230-231, 2005.

2. X. P. Wang, C. Shen, M.-F. Li, H. Y.Yu, Y. Sun, Y. P. Feng, A. Lim, W.S. Hwang, A. Chin, Y. C. Yeo, G. Q. Lo and D. L. Kwong, "Dual metal gates with band-edge work functions on novel HfLaO High-k gate dielectric," in *VLSI Tech. Dig.*, pp. 12-13, 2006.

APPENDIX

Publications Related to This Thesis

Referred Journal Publications

1. **C. Ren**, H. Y. Yu, J. F. Kang, Y. T. Hou, M.-F. Li, W. D. Wang, D. S. H. Chan, and D.-L. Kwong, "Fermi-level pinning induced thermal instability in the effective work function of TaN in TaN/SiO_2 gate stack," ***IEEE Electron Device Lett.***, vol. 25, pp. 123-125, Mar. 2004.

2. **C. Ren**, H. Y. Yu, J. F. Kang, X. P. Wang, H. H. H. Ma, Yee-Chia Yeo, D. S. H. Chan, M.-F. Li, and D.-L. Kwong, "A dual-metal gate integration process for CMOS with sub-1 nm EOT HfO_2 by using HfN replacement gate," ***IEEE Electron Device Lett.***, vol. 25, pp. 580-582, Aug. 2004.

3. **C. Ren**, H. Y. Yu, X. P. Wang, H. H. H. Ma, D. S. H. Chan, M.-F. Li, Yee-Chia Yeo, C. H. Tung, N. Balasubramanian, A. C. H. Huan, J. S. Pan, and D.-L. Kwong, "Thermally robust $TaTb_xN$ metal gate electrode for n-MOSFETs applications," ***IEEE Electron Device Lett.***, vol. 26, pp. 75-77, Feb. 2005.

4. **C. Ren**, D. S. H. Chan, X. P. Wang, Faizhal B. B., M.-F. Li, and Y.-C. Yeo, A. D. Trigg, A. Agarwal, N. Balasubramanian, J. S. Pan, P. C. Lim, A. C. H. Huan, and D.-L. Kwong, "Physical and electrical properties of lanthanide-incorporated tantalum nitride for n-channel metal-oxide-semiconductor field-effect-transistors," ***Appl. Phys. Lett.***, vol. 87, 073506, Aug. 2005.

5. **C. Ren**, Faizhal B. B., D. S. H. Chan, M.-F. Li, Y.-C. Yeo, A. D. Trigg, N. Balasubramanian, and D.-L. Kwong, "Work function tuning of metal nitride electrodes for advanced CMOS devices," ***Thin Solid Films***, vol. 504, pp. 174-177, May 2006.

6. **C. Ren**, D. S. H. Chan, M.-F. Li, W. Y. Loh, S. Balakumar, C. H. Tung, N. Balasubramanian, and D.-L. Kwong, "Work function tuning and material characteristics of lanthanide -incorporated metal nitride gate electrodes for NMOS device applications," ***IEEE Trans. Electron Devices***, vol. 53, pp. 1877-1884, Aug. 2006.

7. **C. Ren**, D. S. H. Chan, W. Y. Loh, S. Balakumar, A. Y. Du, C. H. Tung, G. Q. Lo, R. Kumar, N. Balasubramanian, and D.-L. Kwong, "Work function tuning of TaN by high-temperature metal intermixing technique for gate-first CMOS process," to be published in ***IEEE Electron Device Lett.***

8. H. Y. Yu, **C. Ren**, Y.-C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M.-F. Li, D. S. H. Chan, and D.-L. Kwong, "Fermi pinning induced thermal instability of metal gate work functions," ***IEEE Electron Device Lett.***, vol. 25, pp. 337-339, May 2004.

Referred Conference Publications

1. **C. Ren**, J. F. Kang, H. Yang, D. D. Han, X. Y. Liu, R. Q. Han, D. S. H. Chan, and D.-L. Kwong, "Carrier transport and anomalous C-V characteristics in PVD AL_2O_3 gate dielectrics," in ***Int. Conf. on Materials for Advanced Technologies (ICMAT-2003)***, Dec. 2003, Singapore.

2. **C. Ren**, D. S. H. Chan, Fazhal B. B., H. M.-F. Li , Y.-C. Yeo, A. D. Trigg, A. Agarwal, N. Balasubramanian, J. S. Pan, P. C. Lim, and D.-L. Kwong, "Lanthanide-incorporated metal nitrides with tunable work function and good thermal stability for NMOS devices," in ***Symposium on VLSI Technology (VLSI-2005)*,** pp. 42-43, Jun. 2005, Kyoto, Japan.

3. **C. Ren**, Faizhal B. B., D. S. H. Chan, M.-F. Li, Y.-C. Yeo, A. D. Trigg, N. Balasubramanian, and D.-L. Kwong, "Work function tuning of metal nitride electrodes for advanced CMOS devices," in ***Int. Conf. on Materials for Advanced Technologies (ICMAT-2005)***, July. 2005, Singapore.

4. **C. Ren**, D. S. H. Chan, W. Y. Loh, J. W. Peng, S. Balakumar, Y. Jiang, C. H. Tung, A. Y. Du, G. Q. Lo, R. Kumar, N. Balasubramanian, and D.-L. Kwong, "CMOS compatible dual metal gate integration with successful V_{th} adjustment on high-k HfTaON by high-temperature metal intermixing," accepted by *36th* ***European Solid-State Device Research Conference (ESSDERC-2006)***, Sep. 2006, Montreux, Switzerland.

5. H.Y. Yu, **C. Ren**, J. F. Kang, Y.-C. Yeo, D. S. H. Chan, M.-F. Li, and D.-L. Kwong, "Thermal stability of metal gate work functions," in ***Int. Conference on Solid State Devices & Materials (SSDM-2004)***, pp. 712-713, Sep. 2004, Tokyo, Japan.

6. J. F. Kang, **C. Ren**, H. Y. Yu, X. P. Wang, M-F. Li, D. S. H. Chan, Yee-Chia Yeo, Y. Y. Wang, and D.-L. Kwong, "A novel dual-metal gate integration process for sub-1 nm EOT HfO_2 CMOS devices," in ***Int. Conference on Solid State Devices & Materials (SSDM-2004)***, pp. 198-199, Sep. 2004, Tokyo, Japan.