ADVANCED GATE STACK MATERIALS AND PROCESSES FOR SUB-100 nm CMOS APPLICATIONS

by

Qiang Lu

Memorandum No. UCB/ERL M03/4

2 December 2002

ADVANCED GATE STACK
MATERIALS AND PROCESSES
FOR SUB-100 nm CMOS APPLICATIONS

by

Qiang Lu

Memorandum No. UCB/ERL M03/4

2 December 2002

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Advanced Gate Stack Materials and Processes for Sub-100 nm CMOS Applications

by

Qiang Lu

B.S. (Peking University, Beijing) 1996
M.S. (University of California, Berkeley) 2001

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering
and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Chenming Hu, Chair
Professor Tsu-Jae King
Professor Timothy Sands

Fall 2002

Advanced Gate Stack Materials and Processes for Sub-100 nm CMOS Applications

by

Qiang Lu

Abstract

Advanced Gate Stack Materials and Processes for Sub-100 nm CMOS Applications

by

Qiang Lu

Doctor of Philosophy in Engineering –
Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Chenming Hu, Chair

Continued CMOS technology scaling beyond the sub-100 nm node encounters serious challenges posed by the intrinsic limits of the materials used in silicon CMOS. This dissertation investigates the solutions to the critical issues of the gate stack by introducing new gate stack materials, i.e., using high-permittivity (high-*k*) gate dielectrics to replace SiO_2, and metal or poly-SiGe gates to replace the poly-Si gate.

A number of high-*k* gate dielectrics are studied. Silicon nitride, formed by jet-vapor deposition (JVD) or rapid-thermal CVD (RTCVD), and sputter-deposited HfO_2 are so far the promising candidates. Silicon nitride is thermally more stable, while HfO_2 offers lower gate leakage for a given equivalent oxide thickness (EOT) and consequently better scalability. It is also shown that the HfO_2 gate dielectric with a nitrided interface has sufficient hot electron reliability.

A gate material with a reduced or negligible depletion effect is needed to improve device performance. The poly-SiGe gate is for the first time demonstrated to be stable with a HfO_2 gate dielectric. Reduced gate depletion is achieved by better dopant activation in the poly-SiGe gate. In addition, the poly-SiGe gate results in thinner EOT than a poly-Si gate by suppressing the interfacial layer growth during high temperature annealing. These benefits combined with the CMOS compatibility make the poly-SiGe gate a very attractive near term alternative to the poly-Si gate.

A metal gate eliminates the gate depletion and boron penetration problems of the poly-Si gate. A key challenge is to identify CMOS compatible metals with appropriate work-functions. A Mo gate is studied with several alternative gate dielectrics, and is demonstrated to be a promising gate material for p-MOSFETs. In addition, the work-function of Mo can be adjusted to meet the requirements for n-MOSFETs by nitrogen implantation, making it possible to achieve dual gate work-functions using a single metal gate. CMOS processes based on dual-metal gates (Mo and Ti) or single metal gate (employing a Mo gate with nitrogen implantation to the gate of the n-FETs) are demonstrated. The use of metal gates requires substantial changes to the existing CMOS technology, so it is more likely a long-term solution to the gate electrode problem.

The dissertation abstract of Qiang Lu is approved:

Professor Chenming Hu Date
Committee Chair

To my family

Table of Contents

Acknowledgments

First and foremost I would like to thank my research advisors, Professor Chenming Hu and Professor Tsu-Jae King. Professor Hu gave me the much-needed advices when I was looking for research opportunities as a first-year graduate student, and generously offered to have me join the Device Group. Professor Hu and Professor King are both the ideal kind of advisors a graduate student can possibly find, and I am just lucky enough to have the privilege to have been a student of both of them. I have benefited tremendously from their continuous encouragement and support. In addition to their insightful and timely technical guidance throughout my graduate research, their diligence and enthusiasm for the scientific pursuit also have profoundly influenced me.

I would like to thank Professor Avideh Zakhor, for serving on my qualifying exam committee. I also enjoyed taking her excellent courses on digital signal processing and multimedia.

I also would like to thank Professor Robert Littlejohn of the Physics Department for serving on my qualifying exam committee. I was very fortunate to have taken the Quantum Mechanics courses he taught, from which I continue to benefit in my research.

I am very grateful to Professor Timothy Sands of the Materials Science and Engineering Department for serving on my dissertation committee.

I must thank Professor Charles K. P. Cheung of Rutgers University, who was my advisor during my summer internship at Bell Labs in 1998. That internship experience was productive and exciting, and I benefited enormously from working with Dr. Cheung. Since then, he has been closely watching my progress and providing me with most valuable advices on various aspects of my research and career.

I am very thankful to Hideki Takeuchi-san, who has generously helped me with equipments and techniques in the cleanroom. His warmheartedness as well as his broad knowledge makes him a most valuable asset in our lab, and a most responsive resort in case of emergencies. I also would like to thank Dr. Xiaofan Meng and other staff members of the Microfabrication Laboratory, whose help and support was indispensable to the success of my research projects.

I would like to thank Judy Fong, the grants administrator of our group. Her excellent support has made a key contribution to the success of our research.

I am indebted to the former students of the Device Group, Dr. Kai Chen, Dr. Bin Yu, Dr. Donggun Park, Dr. Sundar K. Iyer, Dr. Ya-Chin King, Dr. Wen-Chin Lee, for helping me efficiently get through the initial learning stage as a new graduate student. Their encouragement and advices are greatly appreciated.

I would like to acknowledge my talented fellow students of the FEP Research Group for their various contributions to this dissertation. My close collaborations with Yee-Chia Yeo and Ron Lin, from the efficient teamwork to the interesting conversations while waiting on a process step, were very enjoyable and memorable. I am also very grateful to Kevin J. Yang, Pushkar Ranade, Igor Polishchuk and Leland Chang, for instructive technical discussions as well as their amiable company, which has made the everyday life in the cubicle more pleasant and productive.

I am also thankful to the members or former members of the Device Group, Peiqi Xuan, Min She, Qing Ji, Charles Kuo, Yu Cao, Yang-Kyu Choi, Dae-Won Ha, Xuejue Huang, Xiaodong Jin, Kanyu Cao, Hui Wan, Pin Su, Gang Liu, Shiying Xiong, Dr.

Weidong Liu, Dr. Jakub Kedzierski, Dr. Michael Orshansky, Dr. Stephan Tang, Dr. Nick Lindert, Dr. Xuemei Xi, for their friendship and support.

On several research projects involving novel gate dielectrics, I benefited greatly from the collaboration with Professor Tso-Ping Ma (Yale University), Professor Jack C. Lee (University of Texas, Austin), Professor Dim-Lee Kwong (University of Texas, Austin) and their groups. Their supports were an essential part of those research projects.

Ultimately, my deepest gratitude goes to my family. Their love, support and encouragement have been with me through every single step in my life, not excepting this dissertation.

Chapter 1

Introduction

1.1 The IC industry and CMOS scaling

Since the introduction of the first commercial MOSFETs in the early 60's, the microelectronics industry has enjoyed decades of phenomenal growth. The feature size of an individual MOSFET shrank from tens of microns in the 60's to roughly a tenth of a micron today, and the number of transistors on a single silicon chip increased by several orders of magnitude. Consequently, very complicated functions can be integrated and manufactured at a lower cost per function. The advancement of the silicon integrated circuit (IC) technology is the key to the ubiquity of communication and computing devices, which laid the foundation of the Internet and reshaped many aspects of human life and society. CMOS technology is dominant in microelectronics today, and tremendous efforts are being made to maintain its momentum of growth.

MOSFET scaling has been the driving force of the microelectronic industry for more than 30 years. The scaling trend is described by the well-known Moore's law, which states that the reduction of IC device dimensions and the increase of number of transistors on a single silicon chip both follow exponential relations with the technology generation [1.1]. Although at the time of its publication, the Moore's law was only based on the observation of the semiconductor industry data of a single decade, it turned out to be amazingly accurate in the more than 30 years to follow. The International Technology Roadmap for Semiconductors (ITRS) was later established to project future technology specifications and to coordinate the tool development efforts across different sectors of the IC industry. The ITRS projects that in the coming years the exponential growth will continue at the historical pace. Consequently, this poses considerable technical challenges to the semiconductor technology and manufacturing [1.2].

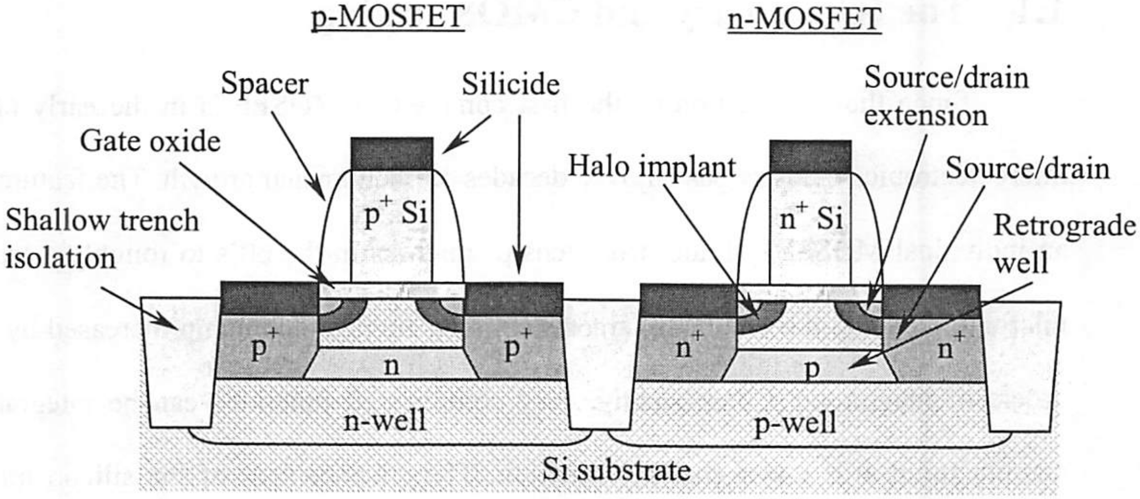


Figure 1.1 A schematic cross-sectional view of state-of-the-art bulk Si CMOS transistors with major front-end components shown. After reference [1.3].

Figure 1.1.shows the schematic cross-sectional view of start-of-the-art n- and p-channel MOSFETs, with only front-end components shown [1.3]. The main features of these transistors are shallow trench isolation, ultra-thin gate oxide, shallow implant source/drain extension, pocket/halo implant, retrograde well and Co silicide contact. The physical gate length is below 100 nm. The major challenges of continued scaling for front-end processes are:

1). Lithography

Improved lithography capability makes it possible to print smaller lateral features, and plays a key role in device scaling. The manufacturing of ICs consisting of millions of transistors with sub-50 nm gate lengths requires patterning tools that can provide high resolution and high yield beyond the limit of today's optical lithography. Prototype tools are already available, but a manufacturable solution is yet to be developed.

2). Gate stack (ultra-thin gate oxide and gate electrode)

An ultra-thin gate oxide is needed to maintain good device performance and to keep off-state leakage current low. In most advanced CMOS processes today, the gate oxide (SiO_2) thickness is already below 20 Å. To electrically realize the full benefits of the thin gate oxide, increasingly high active dopant concentration in the poly-Si gate is needed so as to minimize the reduction of gate-to-channel coupling due to the gate depletion effect. High active dopant concentration is difficult to achieve due to the limitations of both the solid solubility of the dopants in silicon and the activation thermal budget. High temperature annealing may cause the boron dopants in the p^+ gate to diffuse through the ultra-thin gate oxide into the channel, affecting normal device characteristics. Therefore, this constraint makes high gate doping level even more difficult to achieve.

3). Ultra-shallow source/drain extension junction

As the gate length of CMOS transistors is reduced, it becomes more difficult to shut off the transistors. Increased off-state leakage results from a sub-surface conduction path formed by the two-dimensional distribution of the electrical field near the junction depletion regions. A shallow junction is critical to controlling the short-channel effects. In addition, steep abruptness of the junctions is required for reducing the short channel effects [1.2].

4). Channel/substrate dopant profile optimization

For bulk MOSFETs, the dopant profiles in the channel and the substrate need to be engineered to suppress the leakage path associated with the short channel effect. Retrograde well, pocket, halo or super halo implants are examples of this approach [1.3]. As the device dimensions get smaller, these techniques may bring more undesirable side-effects, and the optimization becomes more challenging. While functioning bulk CMOS transistors with gate length down to 15 nm have been demonstrated [1.4], novel device structures are being intensively explored as an alternative to bulk CMOS for scaling toward the 10 nm regime [1.5].

5). Low source/drain/contact resistance

The channel resistance of a MOSFET gets smaller with shorter gate length, so the effect of parasitic resistances, such as the contact resistance, becomes more pronounced. Techniques of reducing parasitic resistance include silicide contact (CoSi, NiSi) [1.6], elevated source/drain [1.7], etc.

Although the above is not an exhaustive list of the research needs, the difficulty of CMOS scaling is obvious. Among these challenges, the gate stack scaling is an outstanding one, as the gate oxide (SiO_2) thickness is already the dimension of a few molecules, approaching the fundamental physical limit imposed by the nature.

1.2 Gate stack scaling issues

The dual n^+/p^+ poly-Si gate and SiO_2 gate dielectric is the standard gate stack used in current CMOS technologies. While the basic material system remained the same throughout many technology generations, the film thicknesses have been significantly reduced and many improvements have been made to the process modules. The gate SiO_2 thickness decreased from ~1000 Å in the 60's to ~15 Å in the state-of-the-art technologies. Thin gate oxide thickness is critical for the suppression of the short-channel effects as well as the reduction of power supply voltage, therefore, even thinner gate

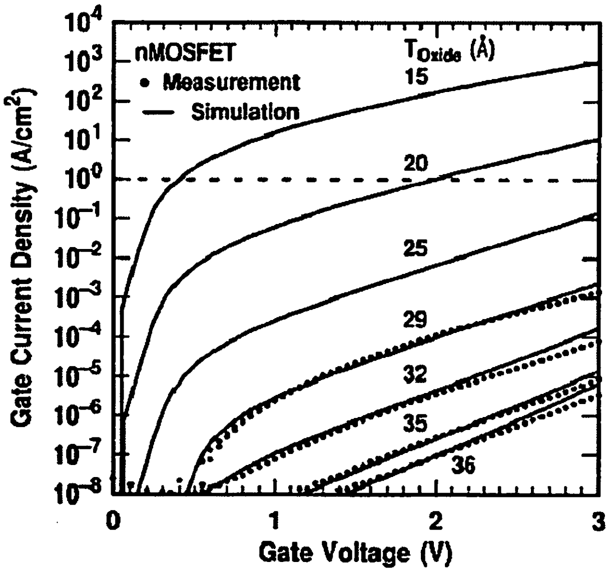


Figure 1.2 Measured and simulated direct tunneling gate current of n-MOSFETs with ultra-thin SiO_2 gate dielectric. From reference [1.9].

oxide will be needed for the future CMOS technology. However, when scaled to below ~15 Å, conventional thermal SiO_2 results in unacceptably high gate tunneling current, which increases power consumption and distorts normal transistor characteristics [1.8]. Figure 1.2 shows the measured and simulated I_G-V_G characteristics of n-MOSFETs with gate SiO_2 thickness down to 15 Å [1.9]. The leakage level of 1 A/cm^2 is highlighted. Assuming such a gate current limit and 2 V power supply, the thinnest usable SiO_2 thickness is about 20 Å. More recently, the gate leakage requirements have been relaxed and power supply voltage reduced, but ultra-thin gate oxide remains to be the most serious problem.

In the past, higher gate capacitance can be achieved simply by using physically thinner gate oxide, which is the dominant factor in the gate capacitance. With gate oxide

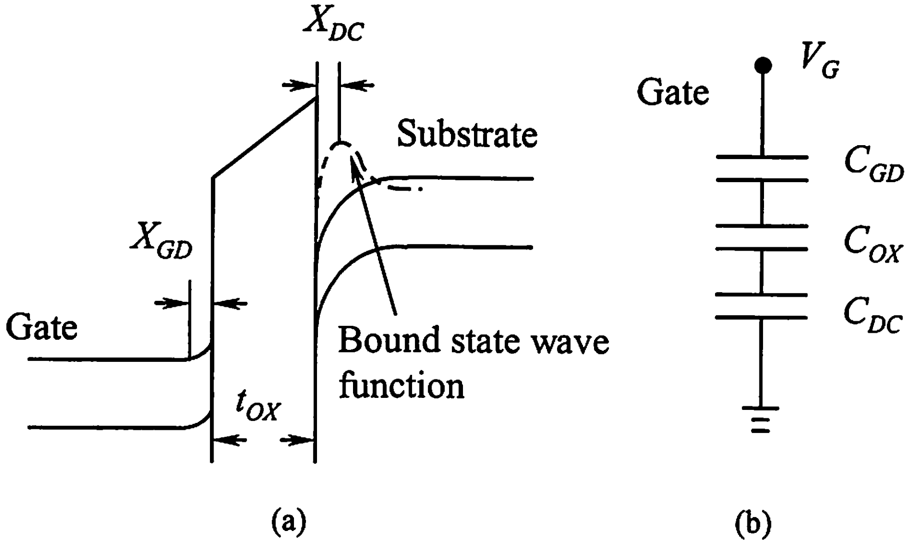


Figure 1.3 **(a)** The band diagram of an n-channel MOSFET with n$^+$ poly-Si gate in inversion regime. The inversion electrons populate the bound states in the potential well near the surface of the substrate. The charge centroid is located at X_{DC} from the surface of the substrate. The poly-Si gate has a depletion width of X_{GD}. **(b)** The equivalent circuit of the gate capacitance stack. C_{GD}, C_{OX} and C_{DC} are contributed by gate depletion region, gate oxide and inversion charge centroid offset, respectively.

thickness scaled to below 20Å, however, two other factors affecting the gate capacitance can no longer be ignored. One is the capacitance related to the gate depletion region under inversion gate bias, which electrically increases the equivalent oxide thickness by a few angstroms [1.11]. The other is the quantum confinement effect of the carriers in the inverted channel. As shown in Figure 1.3 (a), in the inversion regime, the band bending near the surface of the silicon substrate and the barrier of the gate oxide forms a potential well, in which the inverted channel exists. When the potential well is deep enough, there can exist bound states of the inversion carriers due to the confinement along the direction perpendicular to the substrate surface. Therefore, carrier energy levels related to this degree of freedom becomes discrete, depending on the eigenstate (sub-band) that the carriers occupy, and the carriers in the channel become two dimensional instead of three-dimensional free particles, or a two-dimensional electron gas. The consequence of this quantum confinement effect is that the maximum of the eigen-wave-function of the carriers on the *i*-th sub-band occurs at a distance ($X_{DC,\,i}$) from the substrate surface. The distance depends on the gate bias and the sub-band of the bound state. The electrical effect of this spatial charge distribution can be represented by a charge centroid located at X_{DC} from the substrate surface. Figure 1.3 (b) shows the equivalent circuit of the components contributing to the gate capacitance in the inversion regime. The effective gate capacitance is related to these components by

$$\begin{aligned}\frac{1}{C_G} &= \frac{1}{C_{GD}} + \frac{1}{C_{OX}} + \frac{1}{C_{DC}} \\ &= \frac{X_{GD}}{\varepsilon_0 \varepsilon_{Si}} + \frac{t_{OX}}{\varepsilon_0 \varepsilon_{OX}} + \frac{X_{DC}}{\varepsilon_0 \varepsilon_{Si}}\end{aligned} \tag{1.1}$$

where the definitions of the variables follow those given in Figure 1.3 (a) and (b). ε_0=8.85×10^{-12}F/m is the dielectric constant of vacuum, and ε_{Si}=11.9 is the relative dielectric constant of silicon. In the above equation, an arbitrary gate dielectric with physical thickness t_{OX} and relative dielectric constant ε_{OX} is assumed. The equivalent oxide thickness (EOT) of the gate dielectric is defined by

$$EOT = t_{OX} \cdot \frac{\varepsilon_{SiO2}}{\varepsilon_{OX}} \qquad (1.2)$$

where ε_{SiO2} =3.9 is the relative dielectric constant of SiO_2. So in terms of capacitance, a gate dielectric of thickness t_{OX} is equivalent to SiO_2 with a thickness given by the EOT value. The quantity used to describe the overall effective gate capacitance is defined by

$$CET = \frac{\varepsilon_0 \varepsilon_{SiO2}}{C_G} \approx EOT + \frac{(X_{GD} + X_{DC})}{3} \qquad (1.3)$$

where CET stands for capacitance-equivalent thickness. Both X_{GD} and X_{DC} are bias-dependent, so the CET also depends on the gate bias V_G. Typically under normal device operation, gate depletion and quantum effect each contribute a few angstroms to the CET, so in the case of an ultra-thin gate dielectric, the CET can be significantly larger (percentage wise) than EOT, which is determined by the gate dielectric. Thinner CET and higher gate voltage translate to higher inversion charge density and better device performance, therefore, with reduced power supply voltage for each technology generation, the gate stack scaling not only involves the thinning of gate dielectric (EOT), but also requires the reduction of the gate depletion effect (X_{GD}).

In addition, the boron penetration problem also becomes more serious with ultra-thin SiO_2. In recent years, oxynitride gate dielectric, SiO_2 with nitrogen incorporation, has been used to reduce boron penetration [1.10]. But this problem will exist as long as p^+ poly-Si gate is used. These problems require a solution in the very near future.

Table 1.1 lists some of the key gate-stack related requirements projected by the ITRS for future technology generations. The near term ITRS projections are specified for every year, and the long term ones are specified only for each technology generation (three year period). Requirements within each technology generation, which also progress on a yearly basis, are not listed in the table below.

Table 1.1 Gate stack requirements projections for the future technology generations by the ITRS [1.2]. The status of the solutions to each requirement is indicated by the corresponding background color.

Year of production	2004	2007	2010	2013	2016
Technology node (nm)	90	65	45	32	22
Physical gate length MPU (nm)	37	25	18	13	9
MPU EOT (Å)	9-14	6-11	5-8	4-6	4-5
MPU gate leakage at 100°C (nA/μm)	100	1000	3000	7000	10000
Low operating power (LOP) EOT (Å)	14-18	10-14	8-12	7-11	6-10
LOP gate leakage at 100°C (pA/μm)	300	700	1000	3000	10000
Low standby power (LSTP) EOT (Å)	18-22	12-16	9-13	8-12	7-11
LSTP gate leakage at 100°C (pA/μm)	1.0	1.0	3.0	7.0	10.0
Active poly doping (cm^{-3})	1.5×10^{20}	1.87×10^{20}	1.8×10^{20}	2.5×10^{20}	2.99×10^{20}

Solution exists

Solution known

Solution unknown

Each major technology generation spans three years, with the technology node named by the half pitch for that generation. The physical gate lengths of the high performance transistors are significantly smaller than the half pitch, while both roughly follow a 0.7× reduction over the previous generation. As apparent in the table, thinner EOT is required for shorter gate length, however, the EOT scaling is sub-exponential due to the technical difficulties. In the ITRS, different specifications were set for high-performance (MPU), low operating power (LOP) and low standby power (LSTP) devices, so that some of the stringent requirements can be relaxed where possible, and better optimization of the process technology for different applications can be facilitated. For high performance applications, gate leakage limit is higher so as to allow the use of very thin EOT to achieve smaller channel length and better drive current. Increased gate tunneling current is becoming a serious source of power consumption, therefore the gate leakage specifications are more stringent for low operating power and low standby power devices, as reflected in the more conservative EOT scaling for these two types of devices. Low operating power and low standby power devices also have larger gate length than the MPU devices of the same generation to trade off some performance for reduced power consumption. The active poly gate doping is based on the assumption that the gate depletion contributes 25% of the EOT, so the required gate doping increases with thinner EOT. For the majority of these specifications, manufacturable solutions are not known yet, so research efforts in these areas are urgently needed. As these gate stack problems are all due to the intrinsic material properties, the search for new gate and gate dielectric materials has been the main focus of the research activities in this field.

There are two major aspects of the challenges for continued gate stack scaling. On the materials side, appropriate new materials are to be identified to replace the conventional SiO_2 gate dielectric and poly-Si gate electrode. The corresponding process modules, such as thin film deposition, etching, metrology, also need to be developed and qualified for Si CMOS processing. On the process integration side, CMOS processes need to be developed so as to incorporate the new materials into a CMOS fabrication process. Material properties and thermal stability issues may require considerable modification to existing CMOS processes. In this dissertation, both aspects will be discussed.

1.3 Organization

This dissertation addresses the aforementioned issues of gate stack scaling for future generations of CMOS technology, and proposes solutions based on new materials and CMOS processes. Chapter 2 discusses alternative gate dielectric materials. After introduction to the general requirements of alternative gate dielectrics, silicon nitride and hafnium oxide will be presented in more detail. These two materials represent the dielectrics with medium and high dielectric constant, respectively, and they also may well be the most promising materials for short and long term replacements for SiO_2. Common problems with alternative gate dielectrics in CMOS technology and key progress will be highlighted. The hot carrier reliability of n-FETs using hafnium oxide with a nitrided interface is also investigated.

Chapter 2 raises the two major issues of alternative gate dielectrics with poly-silicon gate, i.e., the increase in EOT upon high-temperature annealing and the gate depletion effect, which diminishes the benefits of using alternative gate dielectrics. Chapter 3 discusses the use of poly-SiGe as the gate electrode to alleviate the above problems. Poly-SiGe gate for SiO_2 gate dielectric had been studied and demonstrated to improve the gate dopant activation and reduce the boron penetration for p^+ gate devices. In this work, poly-SiGe is investigated with hafnium oxide gate dielectric, and compared to the poly-Si gate control devices fabricated in the same CMOS process. The effects of nitrided HfO_2/Si-substrate interface are also studied with both gate materials. It was found that poly-SiGe gate not only improves gate dopant activation compared to poly-Si gate, but also reduces the gate dielectric EOT after high temperature process. A possible mechanism for this phenomenon is proposed based on additional experimental results. The nitrided interface is shown to provide several benefits.

Although the poly-SiGe gate shows improvements over the poly-silicon gate, it does not completely eliminate the gate depletion and boron penetration problems, which are both due to intrinsic materials limitations. Chapter 4 discusses the use of metal gate electrode in CMOS devices, an approach that in principle can completely solve those problems. The chapter begins with the discussion of the general requirements of metal gate electrode, followed by the results of using Mo as a gate metal for p-FETs. In addition to the choice of metal gate materials, CMOS process integration is also a considerable challenge. The gate-first metal gate integration is similar to the existing CMOS technology, and will be emphasized in this dissertation. Two different integration schemes are proposed based on this approach. In the direct-deposited dual metal gate process, two different metals are used for the p-FETs and n-FETs, respectively. It demonstrated the benefits of using metal gate electrode, and also revealed some problems of using two dissimilar metal electrodes. A simpler process using a single metal gate with tunable gate work-function for both n- and p-FETs are then presented.

The dissertation is concluded with a summary of the major results and possible future research directions.

1.4 References

[1.1] G. Moore, "Cramming more components onto integrated circuits", Electronics, Vol. 38, No. 8, April 1965.

[1.2] The International Technology Roadmap for Semiconductors (2001 version). Available at http://public.itrs.net/

[1.3] S. Thompson, M. Alavi, M. Hussein, P. Jacob, P. Kenyon, P. Moon, M. Prince, S. Sivakumar, S. Tyagi, M. Bohr, "130 nm logic technology featuring 60 nm transistors, low-*k* dielectrics, and Cu interconnects", *Intel Technology Journal*, Vol. 6, No. 2, pp. 5-13, 2002.

[1.4] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, M.-R. Lin, "15nm gate length planar CMOS transistor", *International Electron Devices Meeting*, pp. 937-939, Dec. 2001.

[1.5] H.-S. P. Wong, "Beyond the conventional transistor", *IBM Journal of Research and Development*, Vol. 46, No. 2/3, pp. 133-168, March/May 2002.

[1.6] Q. Xiang, C. Woo, E. Paton, J. Foster, B. Yu and M.-R. Lin, "Deep sub-100nm CMOS with ultra low gate sheet resistance by NiSi", *Symposium on VLSI Technology*, pp. 76-77, June 2000.

[1.7] H. Shibata, Y. Suizu, S. Samata, T. Matsuno, K. Hashimoto, K, "High performance half-micron PMOSFETs with 0.1 μm shallow p^+n junction utilizing selective silicon growth and rapid thermal annealing", *International Electron Devices Meeting*, pp. 590-593, Dec 1987.

[1.8] H. S., Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, H. Iwai, "Tunneling gate oxide approach to ultra-high current drive in small geometry MOSFETs", *International Electron Devices Meeting*, pp. 593-596, Dec 1994.

[1.9] S.-H. Lo, D. A. Buchanan, Y. Taur and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFETs", IEEE Electron Device letters, Vol. 18, NO. 5, pp. 209-211, May 1997.

[1.10] M. J. Ma, J. C. Chen, Z. H. Liu, J. T. Krick, Y. C. Cheng, C. Hu, P. K. Ko, "Suppression of boron penetration in p^+ polysilicon gate p-MOSFETs using low-temperature gate-oxide N_2O anneal", *IEEE Electron Device Letters*, Vol. 15, No. 3, pp. 109-111, March 1994.

[1.11] C. Hu, "Gate oxide scaling limits and projection", *International Electron Devices Meeting*, pp. 219-322, Dec 1996.

Chapter 2

Alternative gate dielectrics for CMOS

2.1 Introduction

SiO_2 has been used as the standard gate dielectric for MOSFETs. Thermally grown SiO_2 has high film quality and an excellent interface with the Si substrate, and very precise control of the SiO_2 thickness as well as uniformity can be achieved. With CMOS scaling, the gate oxide thickness also needs to be reduced in order to control the short-channel effects. In short-channel devices, the coupling between the drain and the channel becomes stronger, resulting in undesirable effects such as threshold voltage roll-off and increased off-state leakage current. Thinner gate oxide, i.e., larger gate capacitance, can enhance the coupling between the gate and channel, so that off-state leakage current can be reduced. Several problems, however, arise when the gate SiO_2 thickness is reduced to below ~30 Å. With such a thin physical barrier, quantum-mechanical tunneling of electrons and holes between the gate and the substrate becomes significant even at relatively low gate voltage (the direct tunneling regime). High gate leakage current increases power consumption and may also disturb normal transistor characteristics. The gate tunneling current increases roughly exponentially with reduced gate oxide thickness, therefore the above problem is seriously worsened for oxide thickness below 20 Å [2.1]. While the reliability of ultra-thin SiO_2 is still an open question, it is generally agreed that the high gate leakage current will be the ultimate limiting factor for gate SiO_2 scaling.

Another problem related to thin gate oxide is boron penetration. In p-channel MOSFETs, the p^+ doped gate contains a high concentration of boron dopants, which in a high temperature process can diffuse relatively easily through thin gate SiO_2 into the channel [2.1]. These boron dopants cause threshold voltage shifts, and also degrade device reliability. Some techniques for improving the SiO_2 gate dielectric, such as introducing nitrogen to SiO_2 to reduce boron penetration, can alleviate the problems to a certain extent but do not provide a long-term solution. In addition, with thinner gate oxide, the uniformity and process control becomes more difficult, and device characteristics are more sensitive to process variations. Therefore, pure SiO_2 is not suitable for further scaling into the sub-20 Å regime, and a likely solution is to use a new gate dielectric. If an insulator with higher dielectric constant (*k*) can be used to replace SiO_2, then for a given EOT specification, physically thicker gate dielectric can be used so that the gate tunneling current may be reduced.

2.1.1 General Requirements for Alternative Gate Dielectrics

Besides having a higher dielectric constant than SiO_2, the new gate dielectric material needs to meet some basic requirements as necessitated by CMOS technology. To ensure low gate leakage current, reasonably large bandgap and favorable band alignment with Si is needed so that there exist significant tunneling barriers for both electrons and holes. From a process integration point of view, the new gate dielectric needs to be compatible with existing CMOS process technology, i.e., can be deposited and patterned using methods that are practical for manufacturing. Thermal and chemical stability is another important issue. In a CMOS process, the gate dielectric will be in contact with the Si substrate and the gate electrode, and will undergo annealing at above 1000°C. It is essential that the gate dielectric remain intact after the full CMOS process.

The oxides and nitrides of metals are a major category of inorganic dielectrics that have high dielectric constants, therefore, they naturally became the candidates of research for alternative gate dielectrics. Table 1 summarizes the properties of some commonly studied alternative gate dielectrics. It should be noted that the film deposition methods, composition and thermal process history can all affect the material properties, so some experimental results may show significant differences from the values listed below.

Table 2.1 Material properties of some high-*k* dielectrics considered for gate insulator application.

Dielectric	*k*	Bandgap (eV)	Electron barrier (eV)	References
Si_3N_4	7-8	5.3	2.4	[2.3],[2.4]
Al_2O_3	11	8.8	2.8	[2.4],[2.5]
Ta_2O_5	25	4.4	0.3	[2.4],[2.6]
TiO_2	20-30	3.1	0	[2.4],[2.7],[2.8]
La_2O_3	27	6.0	2.3	[2.4],[2.9]
Y_2O_3	11	6.0	2.3	[2.4],[2.10]
ZrO_2	22	5.8	1.4	[2.4],[2.10],[2.11]
Zr-silicate	12	6.0	1.5	[2.4],[2.11]
HfO_2	22	6.0	1.5	[2.4],[2.10],[2.12]
Hf-silicate	11	6.0	1.5	[2.4],[2.13]

Many materials listed above have a bandgap larger than 5 eV. Since the bandgap of silicon is 1.1 eV, a reasonably symmetrical band alignment will ensure barrier heights for both electrons and holes to be larger than 1 eV, as required for low Schottky emission current [2.4]. The electron barrier heights in Table 1.1 are theoretical predictions, and may be different from experimental results. In addition, a thin interfacial layer with different composition than the high-*k* film may exist between the Si substrate and the high-*k* dielectric, modifying the actual tunneling barrier heights. The actually elemental composition of some materials, e.g., silicon nitride, Hf-silicate, Zr-silicate, can vary in a fairly large range. And the corresponding material properties can also be significantly different.

In the early stage of high-*k* gate dielectric studies, the majority of the research efforts were focused on the high-*k* dielectrics that had been studied for possible applications in dynamic random access memory (DRAM), such as Ta_2O_5, TiO_2, $BaSrTiO_3$, etc. They all have significantly higher dielectric constant than SiO_2. But besides the small bandgaps, they all were later shown to be thermodynamically unstable in direct contact with Si at high temperature. The reaction between these dielectrics and Si can form dielectrics with lower *k* value or conductors (silicides). For examples, Ta_2O_5 can react with Si and form SiO_2 and silicide:

$$\frac{13}{2}Si + Ta_2O_5 \rightarrow 2TaSi + \frac{5}{2}SiO_2$$

The formation of SiO_2 in series with the high-*k* dielectrics increases the overall electrical thickness of the gate dielectric stack, therefore diminishes the advantages of using the high-*k* gate dielectrics. And the formation of the silicide can either prevent the vertical electrical field from controlling the channel by screening or cause a short between the gate and channel, resulting in device failure. The thermal instability problem can be tackled either by selecting gate dielectrics that are thermodynamically stable with Si or by using a stable thin buffer layer between the dielectric and the Si substrate and using gate electrodes that are thermally stable with the gate dielectric.

Along the first approach, an exhaustive theoretical study by D. Schlom et al investigated the thermodynamic stability of all binary oxides and nitrides of metals on the periodic table, and narrowed down a much shorter list of candidate materials [2.10]. The critical reactions involved are:

$$Si + MO_x \rightarrow M + SiO_2$$

$$Si + MO_x \rightarrow MSi_z + SiO_2$$

$$M + SiO_2 \rightarrow MO_x + MSi_y$$

where M stands for the metal element. Using positive change of Gibbs free energy ΔG in all the above reactions as the criterion for thermal stability, different metal oxides and nitrides were evaluated. The majority of the elements on the periodic table were ruled out, and only about 20 metal oxides and nitrides are left.

While equations (1.2) and (1.3) seem to suggest that higher k is always desirable, other effects actually impose constraints on the feasible choice of k. As the gate length of MOSFETs with high-k gate dielectrics will be below 50 nm, the aspect ratio of the gate dielectric T_{OX}/L_G can approach 1, so that there exists significant fringing field from both sides of the gate dielectric. Figure 2.1 illustrates the difference in the fringing field between SiO_2 and a high-k gate dielectric. Assuming that the transistor is of the 65 nm node with 9 Å EOT, the gate dielectric aspect ratio for k=40 is 0.37, but only 0.036 for SiO_2. Physically, the fringing field lines from the sides of the high-k gate dielectric provide additional couplings between the channel and the source/drain regions, therefore exacerbating the short-channel effects. Simulations using a two-dimensional device simulator indicated that for a given gate length and EOT, higher k results in worse threshold voltage roll-off and sub-threshold swing degradation [2.14]. For a given T_{OX}/L_G ratio, the degradation in short channel performance is comparable for different gate length, so this aspect ratio can be used as a good indicator for the fringing field effect. Theoretical study of device scaling using high-k gate dielectrics and low-k spacers suggested a fairly narrow design space, which must be considered for aggressively scaled MOSFETs [2.15].

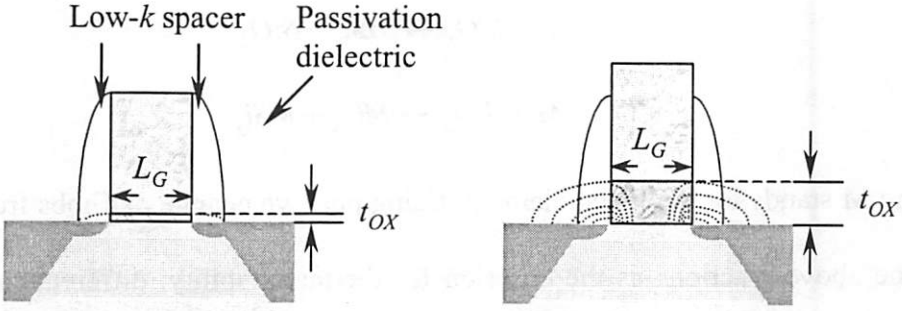


Figure 2.1 Illustration of the enhanced fringing field effect of high-k gate dielectrics. Assuming L_G=25 nm and EOT=9 Å., the high aspect ratio of k=40 gate dielectric (right) causes significant fringing field to the source/drain areas, while such an effect is negligible for SiO_2 (left).

In the subsequent sections, experimental results of using Ta_2O_5 gate dielectrics will be briefly discussed to demonstrate that a dielectric that is not stable with Si can still be used with a buffer layer underneath and a non-Si gate electrode. But due to the complexity of this approach, high-k gate dielectric that is thermally stable with Si is preferred for CMOS integration, and will be the focus of this chapter.

Silicon nitride is among the first alternative gate dielectrics to be considered. It has been routinely used in Si CMOS, and is thermally stable with Si. New techniques have been developed to improve the quality of ultra-thin silicon nitride to meet the requirement for gate insulators. But a major limitation for silicon nitride is the relatively low dielectric constant (~7-8), which limits its scalability. So metal oxides, many of which have much higher dielectric constant than silicon nitride, attracted a lot of research efforts since the late 90's. Ta_2O_5 and TiO_2 were among the first candidates, but were gradually disqualified due to the thermal stability problems. Al_2O_3 is another material of interest. Similar to Si_3N_4, it has a relatively low dielectric constant (~10), so may not be feasible for more than two future generations. More recent studies are focused on ZrO_2 and HfO_2, which are relatively stable with Si thus more compatible with CMOS process. The silicates of these metals, e.g., $ZrSi_xO_y$ and $HfSi_xO_y$, have also been studied and demonstrated to have good dielectric properties. But the drawback is the relatively low dielectric constant (~10-12), which constrains their scalability. In this chapter, CMOS results of silicon nitride and HfO_2 gate dielectrics, representing the two categories (medium-*k* and high-*k*), will be discussed to illustrate the key issues in alternative gate dielectrics, and the need for sufficiently high-*k* to ensure scalability.

2.2 Ta_2O_5 gate dielectric

Ta_2O_5 has a high *k* value (~25), and has been studied for DRAM capacitor applications. Due to its known instability with Si, in the experiment of MOSFETS using a Ta_2O_5 gate dielectric, a buffer layer had to be introduced at the bottom surface of the Ta_2O_5 layer to separate it from the Si substrate, and a TiN gate electrode was used. The gate stack structure is shown is Figure 2.2. Silicon oxynitride (SiO_xN_y) at the bottom interface provides a good interface with the Si substrate and blocks the reaction between

Si and Ta_2O_5. TiN is stable with Ta_2O_5, so a TiN gate was used to replace the standard poly-Si gate. The *in situ* n^+ doped poly-Si cap provides mechanical support, and prevents the TiN from oxidation during subsequent high temperature processes.

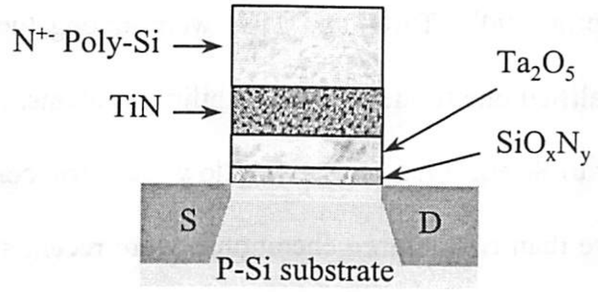


Figure 2.2 Schematic cross-sectional view of the transistor structure consisting of Ta_2O_5/SiO_xN_y dielectrics, n^+ poly-Si/TiN gate electrode and source (S) and drain (D) on p-type Si substrate

N-channel MOSFETs were fabricated using a lightly doped p-type Si substrate. After active region definition and the LOCOS process, the Si-substrate received a rapid thermal nitridation (RTN) in an NH_3 ambient at 800°C for 30 s to form the oxynitride layer. 60 Å or 90 Å Ta_2O_5 was deposited by chemical vapor deposition (CVD) using $Ta(OC_2H_5)_5$ and O_2 at 420°C and 400 mTorr. A post-dielectric-deposition anneal was performed using rapid thermal processing (RTP) at 800°C for 30 s in an O_2 or N_2O ambient. A 600 Å TiN gate was deposited on some wafers by reactive sputtering, then all wafers were capped by LPCVD *in situ* n^+ doped poly-Si. Another RTP (750°C 20 s in N_2) was performed after poly-Si deposition. Gate patterning was done using I-line lithography and Cl_2 reactive ion etching (RIE). Source and drain regions were formed by arsenic ion implantation ($5\times10^{15}/cm^2$, 60 keV) with an activation anneal of 800°C 30 min in N_2. The devices were finished with metallization and forming gas anneal.

The EOT values of different devices were determined by fitting the *C-V* measurements using a quantum-mechanical *C-V* simulator [2.17]. The samples with TiN gate showed thinner EOT than the control samples with an n^+ poly-Si gate. Figure 2.3 shows the measured gate tunneling currents of Ta_2O_5/SiO_xN_y stacks compared to SiO_2 with similar EOTs. The wafer that received post-dielectric deposition in N_2O showed an EOT of 24 Å, while that annealed in O_2 showed 18 Å EOT. Significant reduction of gate leakage was achieved in both cases using Ta_2O_5.

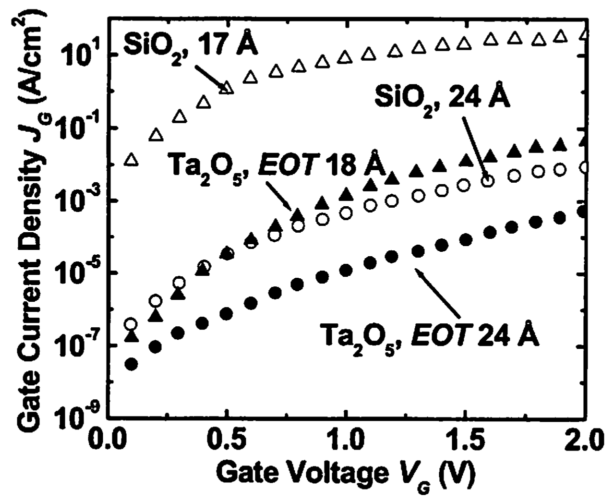


Figure 2.3 Measured gate leakage current density of the Ta_2O_5/SiO_xN_y stacks and SiO_2 with comparable electrical thickness. For sub-20Å EOT, Ta_2O_5 shows ~10^3x reduction of gate leakage than SiO_2.

To investigate the interface quality of the gate stacks, electron mobilities were extracted from transistor I_{DS}-V_{GS} and split *C-V* measurements on different wafers (Figure 2.4). Due to the reaction between Ta_2O_5 and poly-Si, the wafer with an n^+ poly-Si gate (no TiN) showed much thicker EOT (38Å) after S/D activation anneal. The samples with 24 Å and 18 Å EOT correspond to TiN gate electrode and post-dielectric-deposition anneal in N_2O and O_2, respectively. Because of the different EOT values, these devices operate under very different effective vertical field, while their electron mobilities are all comparable to the universal mobility model [2.16], which is based on high quality SiO_2/Si interface. This demonstrates the effectiveness of using a SiO_xN_y layer on the Si substrate to achieve good interface quality.

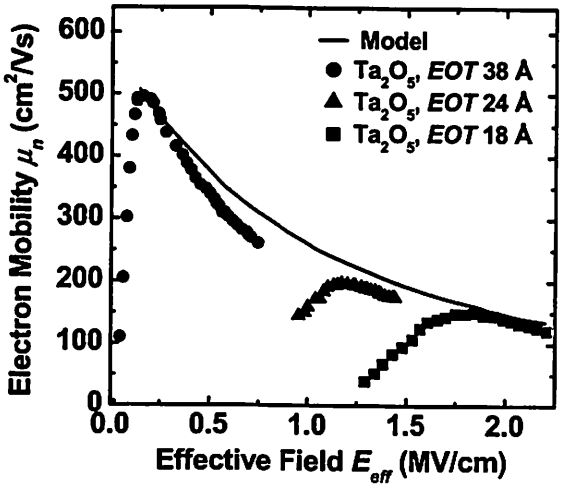


Figure 2.4 The electron mobilities of Ta_2O_5/SiO_xO_y gate stacks with various EOTs (due to different process conditions) are comparable to the universal mobility model [2.16], indicating that good interface quality was achieved using the oxynitride interfacial layer.

Although these encouraging results showed that a high-*k* gate dielectric that is unstable with Si can still be used by using non-Si gate electrode and a buffer layer at the bottom interface, the drawbacks are obvious. The gate electrode that is stable with the high-*k* dielectric may not have the desirable work-function. In the case of Ta_2O_5, the mid-gap work-function of TiN (~4.6-4.7eV) can result in high threshold voltages for both p- and n-MOSFETs, unless the channel is very lightly doped. As high channel doping is required for the control of short channel effects, a TiN gate electrode is not appropriate for bulk-Si CMOS. Moreover, the bottom buffer layer needs to be sufficiently thick in order to prevent the reaction between the high-*k* dielectric and Si substrate. As this buffer layer typically has lower *k*, it limits the scaling of the gate stack EOT. Therefore, alternative gate dielectrics that are thermally stable with Si are preferable.

2.3 Silicon nitride gate dielectric – medium-*k*

2.3.1 General properties and deposition techniques, JVD vs. RTCVD

Silicon nitride has a dielectric constant of ~7-8, roughly twice as that of SiO_2. It can be easily deposited by low-temperature chemical vapor deposition (LPCVD) and etched by reactive ion etching. Silicon nitride is used extensively in Si CMOS processes, therefore it does not have the compatibility issues of other alternative gate dielectrics. However, the quality of silicon nitride deposited by conventional LPCVD is not good enough for the gate dielectric application. It is known to have a high trap density, which severely affects MOSFET characteristics. So new deposition techniques are needed to produce high-quality ultra-thin silicon nitride. Two approaches were demonstrated and will be discussed in this work. One is the jet vapor deposition (JVD) [2.3] and the other is rapid thermal chemical vapor deposition (RTCVD) [2.18].

The basic concept of the JVD technique is to use a high-speed carrier gas to transport the reacting species onto the substrate, so that the diffusion limitations of the mass transport that occur in other deposition techniques can be overcome [2.3]. This mechanism results in highly directional and localized deposition. The carrier gas is formed using He pumped by high-pressure mechanical pump and injected through a nozzle. With proper design, the carrier gas jet can be supersonic. The precursor gases

(SiH_4 and N_2) were carried to the surface by the He jet, and the high kinetic energy provides the energy needed for film deposition, therefore high-temperature is not required.

In the RTCVD process, a rapid thermal oxidation in NO gas is first used to form a thin layer of passivation oxide on the silicon substrate. This passivation oxide improves the interface quality of the gate dielectric, thereby improving MOSFET characteristics. Then on top of this oxide layer, silicon nitride is deposited using SiH_4 and NH_3 precursor by rapid thermal processing. The entire gate dielectrics stack consists of silicon nitride on top of a thin oxide layer (6-7Å). Post-deposition anneals in NH_3 and N_2O are needed to improve the gate dielectric stack properties [2.18]. These rapid thermal processing steps enable precise control of the temperature profiles and greatly improve the gate dielectric's quality over the LPCVD process.

It should be noted that the silicon nitride films formed using the above techniques are not exactly stoichiometric. Auger depth profile of the JVD nitride indicates that in addition to Si and N, there is also fair amount of oxygen in the film [2.3]. The distributions of Si, N and O are uniform throughout the thickness of the film, with relative atomic concentrations of 41%, 47% and 12%, respectively. The as-deposited RTCVD silicon nitride is Si rich, containing lots of traps, and can be easily oxidized. The post-deposition anneal in NH_3 can increase the nitrogen incorporation and resistance to oxidation, although the exact composition has not been calibrated. Unlike the JVD nitride, the RTCVD nitride gate dielectric is a two-layer stack, with an intentionally introduced oxide interface at the Si substrate surface.

2.3.2 CMOS process using Si_3N_4 gate dielectric and poly-Si gate

Prior to this work, MOS capacitors and long-channel MOSFETs using silicon nitride gate dielectrics formed by JVD or RTCVD had been reported. In this experiment, silicon nitride gate dielectrics by the two deposition techniques are studied in the same deep sub-micron CMOS process, to compare their characteristics and to demonstrate sub-100 nm transistors using silicon nitride gate dielectrics.

Four-inch p-type <100> silicon wafers with the same specifications were used to fabricate the CMOS transistors. After n-well formation and the LOCOS process to form isolation, the wafers were split into two groups and received JVD and RTCVD silicon nitride film deposition at Yale University (Prof. T. P. Ma's group) and U.T. Austin (Prof. D. L. Kwong's group), respectively. In the JVD process, silicon nitride film deposition was followed by an 800°C 5 min post-deposition anneal in an N_2 ambient. In the RTCVD process, the passivation oxide layer was grown by RTP in an NO ambient at 800°C for 20 s, followed by silicon nitride film deposition using SiH_4 and NH_3 at 800 °C and rapid thermal anneals in NH_3 (950° 30 s) and then in N_2O (850° 30 s). The wafers were then returned to UC Berkeley for further processing. A 1700 Å undoped poly-Si gate was deposited by LPCVD. I-line lithography and photoresist ashing in an O_2 plasma were used to obtain sub-100 nm gate lengths. Gate etching was done in a HBr:He:O_2 plasma, which provided good selectivity (>50:1) against silicon nitride. Individual devices were inspected by Scanning Electron Microscopy (SEM) after the poly-gate etch to verify the gate length. Figure 2.5 show an SEM picture of an 80 nm gate. Uniform fine gate patterns can be achieved using this technique.

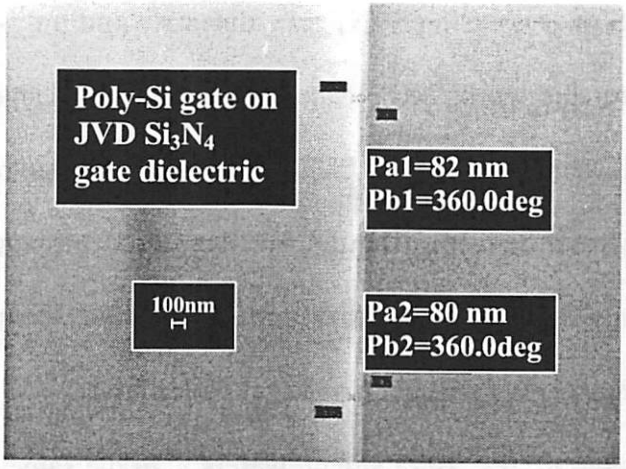


Figure 2.5 SEM image of an 80-nm gate pattern.

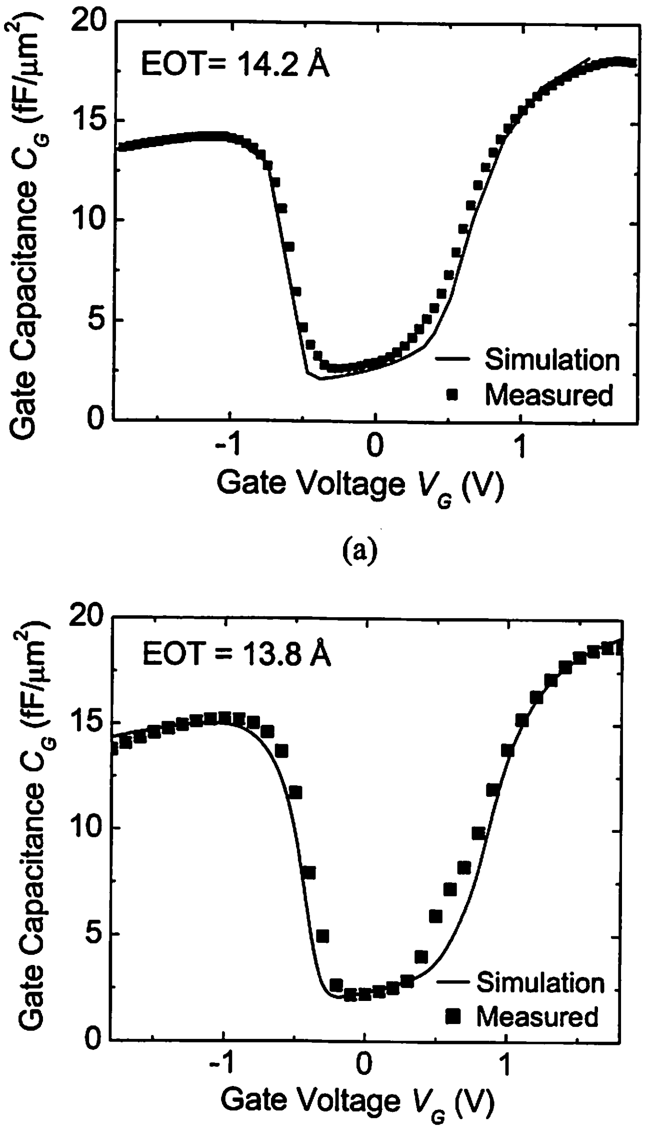
In this CMOS process, the "reverse-LDD" process was used [2.19], in which the source/drain extension regions are formed after the source/drain regions, so that they undergo less diffusion than in the "normal LDD" process. This is advantageous to achieving good characteristics of the short channel devices. After gate etching, a silicon nitride spacer was formed on the sides of the gate, and the source and drain regions were implanted with phosphorus for n-MOSFETs and boron for p-MOSFETs. Then the spacer was removed using a wet etch, and the implantations of the S/D extension regions were performed. Rapid thermal anneals at up to 1050 °C in N_2 were used for dopant activation. After passivation, metallization and forming gas anneal, a water vapor anneal (WVA) at 400°C was performed [2.20] and found to slightly improve drive current but not gate leakage current. A detailed process flow is included at the end of the chapter (see appendix A2.1).

2.3.3 p-FET characteristics

The p-FETs using both silicon nitride gate dielectrics showed good yield, while n-FETs characteristics were not normal. Electrical measurements suggest that the cause is likely due to incomplete removal of the nitride spacers, which blocked the source/drain extension implants. The n-FETs were more seriously affected due to the slower diffusion of the n-type dopants. Therefore only the p-FETs results are presented in the following discussions.

Figure 2.6 shows the gate capacitance characteristics *C-V* of the p-FETs with JVD and RTCVD silicon nitride gate dielectrics. Measurements were done on 10 μm by 10 μm transistors. A *C-V* simulator that takes into account the quantum-mechanical effect and the poly-Si gate depletion effect was used to fit the experimental data and to extract the EOT and active gate dopant concentration [2.17]. Overall good agreement between the experimental data and the simulation can be seen, although some distortion of the curves due to interface traps is visible. The EOTs of both gate dielectrics are 14 Å. The effect of the gate depletion is also obvious in the figures. In the inversion regime, due to the depletion layer in the gate, the inversion gate capacitance is significantly lower than in the accumulation regime. Consequently the total inversion charge is reduced, resulting in lower drive current.

The gate leakage currents of the p-FETs are shown in Figure 2.7. The leakage in the inversion regime is of particular interest, as it is relevant in normal operations. Also shown in the figure for reference is the simulated leakage current of 14 Å SiO_2 using an empirical direct tunneling model, the accuracy of which has been validated by extensive experimental data and numerical simulation [2.21]. The gate leakage of the two nitride



(a)

(b)

Figure 2.6 *C-V* characteristics of p-FETs with (a) JVD and (b) RTCVD silicon nitride gate dielectrics. Quantum *C-V* simulation was used to extracted the EOTs [2.17].

gate dielectrics are roughly two orders of magnitude lower than that of SiO_2 with the same equivalent thickness, which shows the significant advantage of using silicon nitride compared to ultra-thin SiO_2. The direct tunneling model can also be applied to a uniform layer of silicon nitride gate dielectric, with appropriate fitting parameters [2.22]. Using

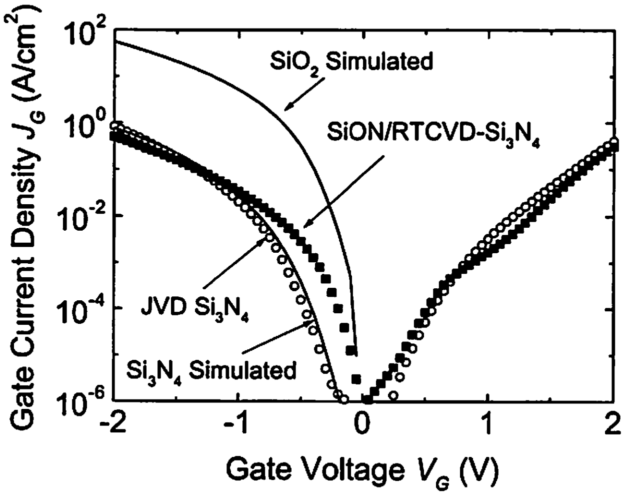


Figure 2.7 Measured gate leakage currents of the 14Å EOT JVD and RTCVD silicon nitride gate dielectrics. Both nitrides show roughly 100X lower gate leakage compared to 14 Å SiO$_2$ (by simulation [2.21]).

this model, the direct tunneling leakage current of even thinner silicon nitride can be projected. Figure 2.8 shows the gate leakage current at a fixed gate voltage as a function of the gate dielectric EOT for n-FETs with SiO$_2$ or silicon nitride gate dielectrics. For the same device parameters (substrate doping, threshold voltage, etc.) and gate voltage, n-FETs set the leakage limit of SiO$_2$ while p-FETs set the limit of silicon nitride [2.21][2.22]. As the gate leakage specifications in the ITRS do not differentiate between n-FETs and p-FETs, these two cases are compared in Figure 2.8. While the leakage current is lower for silicon nitride, the advantage diminishes as the EOT is reduced. In view of the ITRS gate leakage specifications (Table 1.1), silicon nitride may be applicable for high performance transistors down to the 45 nm node or even further, but is apparently inadequate for low operating/standby power transistors. Gate dielectrics to enable further reduction of gate leakage current are critical for those lower applications.

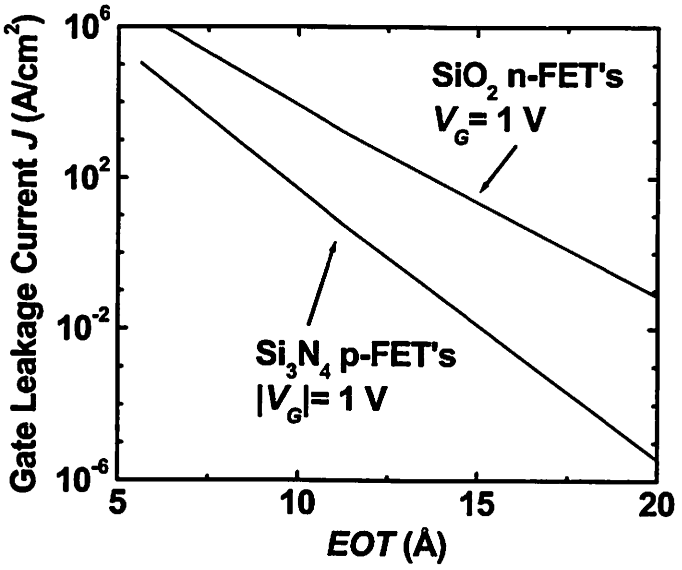


Figure 2.8 Gate leakage current at $|V_G|$=1.0 V as a function of gate dielectric EOT for MOSFETs with SiO_2 or silicon nitride gate dielectrics. The limiting cases, i.e., p-FETs with silicon nitride and n-FETs with SiO_2, are compared. The benefit of using silicon nitride diminishes for thinner EOTs.

Transistor characteristics of the p-FETs are shown in Figures 2.9 (I_{DS} – V_{GS}) and Figure 2.10 (I_{DS} – V_{DS}). The p-FETs using JVD or RTCVD silicon nitride showed very similar characteristics, and normal transistor behaviors were obtained for both types of dielectrics. The similar drive currents and the same EOTs of the two types of p-FETs suggest that the channel carrier mobilities, and therefore the interface qualities are comparable for the two gate dielectrics. A possible reason could be that after the full CMOS processing, particularly the high temperature S/D activation annealing, the film compositions near the interface are similar for the two types of nitrides. It is worth mentioning that the drive currents of the p-FETs are low in view of the gate length and EOT, and further analysis indicated that the low channel carrier mobility was the main reason.

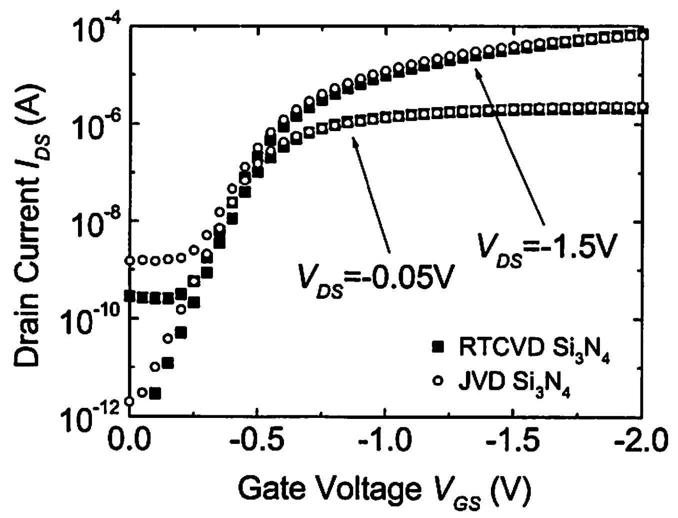


Figure 2.9 p-MOSFETs with JVD and RTCVD silicon nitrides have similar I_{DS}-V_{GS} characteristics.

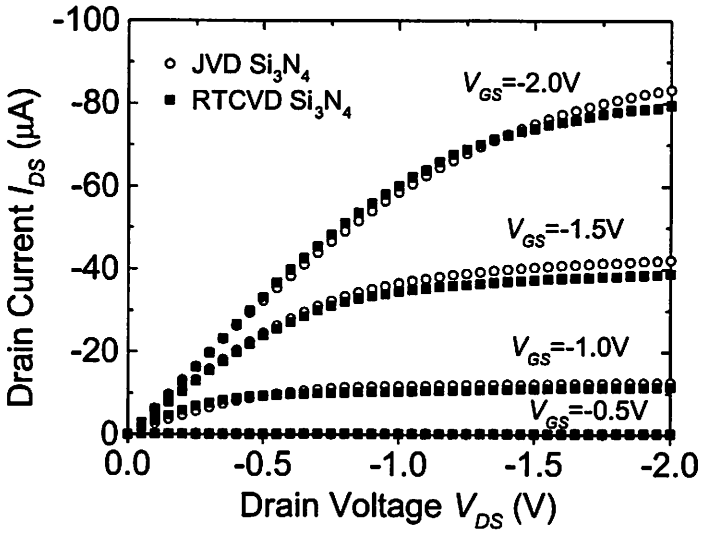


Figure 2.10 p-MOSFETs with JVD and RTCVD silicon nitrides have comparable drive currents, possibly due to the similar gate dielectric interface quality.

This experiment demonstrated the low leakage benefit of using a silicon nitride gate dielectric. However, due to its modest dielectric constant, silicon nitride is not likely to be a material for use in sub-10 Å EOT technology nodes. Therefore, dielectric materials with higher permittivity need to be investigated.

2.4 HfO_2 gate dielectrics – high-*k*

2.4.1 Device fabrication

Hf is one of the a few metals that theoretical study indicated to have both oxide and nitride that are stable with Si. Bulk HfO_2 has a *k* value of ~22 and a fairly large bandgap, with significant barrier height for both electrons and holes [2.4]. HfO_2 can be deposited by several methods, such as reactive sputtering, LPCVD and atomic layer deposition (ALD). After some initial successful reports of MOS capacitor and long channel MOSFET using a HfO_2 gate dielectric, we investigated the CMOS integration of HfO_2.

CMOS transistors with HfO2 gate dielectric were fabricated using a process similar to the one mentioned in previous section, with a few modifications. HfO_2 was deposited at UT Austin by Prof. Jack Lee's group using reactive sputtering from a Hf target with modulated O_2 [2.23], and the wafers were transported back to Berkeley for the rest of the CMOS processing. In the gate dielectric formation step, after standard HF/RCA cleaning, the wafers first received a NH_3 nitridation, followed by reactive sputtering deposition of HfO_2. Previous unsuccessful attempts revealed that HfO_2 is very sensitive to moisture, and can quickly degrade after excessive exposure to air. So the wafers were sealed in vacuum immediately after HfO_2 deposition, and the delay before gate deposition was less than 24 hours. Using reactive sputtering, the HfO_2 is amorphous as-deposited, and crystallizes upon high temperature (800°C or above) anneal. When a poly-Si gate is used, annealing at very high temperature results in thicker EOT and increased leakage current [2.23]. In this experiment, a conservative gate dopant activation anneal (950°C 25 s) was used to order to reduce the above undesirable effects. Empirically it was found that after the high temperature anneal, HfO_2 cannot be etched by dilute HF, so the contact hole wet etch was performed before the gate dopant activation RTA. After metallization, all devices received a forming gas anneal at 400°C for 30 min.

The photo resist trimming technique was also used in gate lithography to obtain sub-100 nm gate length. Figure 2.11 shows an SEM image of a 70 nm gate pattern after gate etching. Due to the variation of the photo-resist ashing process, the gate lengths of a number of sub-100 nm devices were individually measured using SEM.

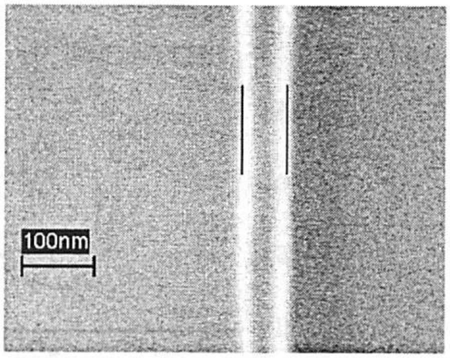
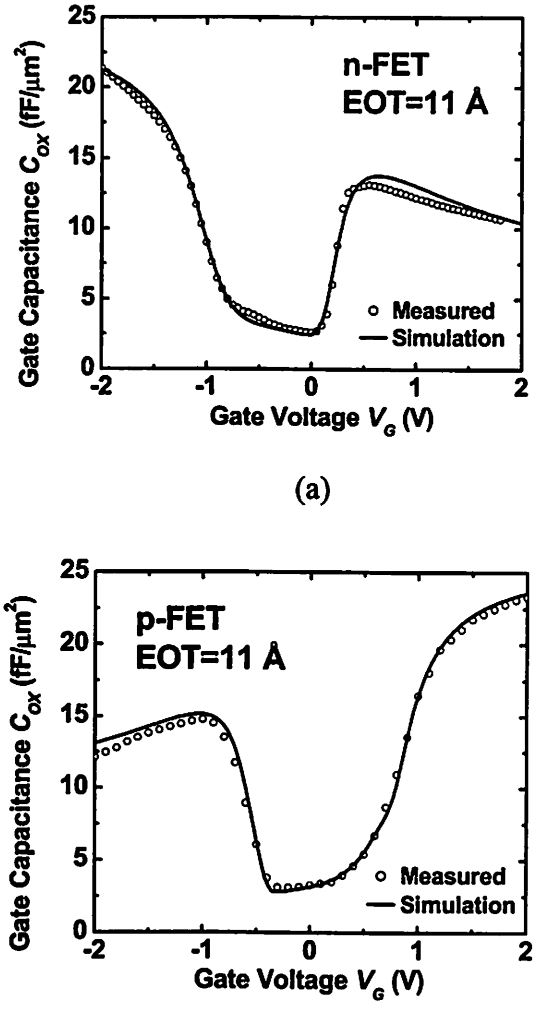


Figure 2.11 Minimum gate length of 70 nm was confirmed by SEM after gate etch.

2.4.2 Device characterization

The *C-V* characteristics of n-FETs and p-FETs are shown in Figure 2.12. An EOT of 11 Å for both p-FETs and n-FETs was extracted by quantum mechanical *C-V* simulation [2.17]. The good fitting of the simulated curve suggests good HfO_2 film quality. However, serious poly-Si gate depletion effects resulted from the conservative dopant activation condition. The depletion layer in the poly-Si gate typically increases the gate stack EOT by a few angstroms determined by the poly-Si doping and electric field,



(a)

(b)

Figure 2.12 *C-V* characteristics of (a) n-FETs and (b) p-FETs with HfO_2 gate dielectric. EOT of 11Å was extracted using quantum *C-V* simulation [2.17].

regardless of the gate dielectric thickness [2.24]. So the inversion gate capacitance for thinner gate dielectric is reduced by higher percentage due the gate depletion. In this case, only roughly 60% of the accumulation gate capacitance is obtained in inversion regime. This directly impacts the transistor drive current.

The physical structure of the gate stack was analyzed using transmission electron microscopy (TEM). It is evident that a low-electron-density layer is present at the bottom interface of the HfO_2 (Figure 2.13), but the thickness of this layer cannot be determined to a high degree of precision without detailed image simulation. This interfacial layer could be a nitride, an oxide or a silicate (containing Hf, Si, N and O) layer formed during the high temperature anneal. Assuming that k=22 (bulk value) for the HfO_2, the k of the interfacial layer is roughly estimated to be ~5-7 using the physical thicknesses estimated from the image and the 11 Å gate stack EOT. This k value suggests a silicon nitride layer.

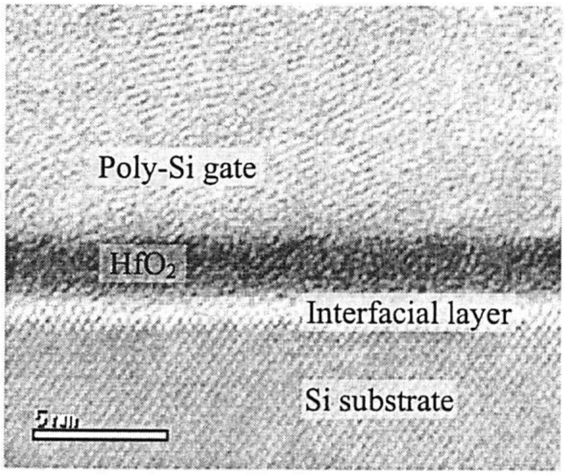


Figure 2.13 High-resolution TEM cross-sectional image of the gate stack, showing the poly-Si gate, the HfO_2 gate dielectric with a distinctive bottom interfacial layer, and the Si substrate.

The threshold voltages and subthreshold swings (S) of the transistors with different gate length are shown in Figure 2.14 and 2.15, respectively. It can be seen that the short-channel effects have been well controlled down to sub-100 nm gate length for n-FETs. The threshold voltages of the p-FETs are relatively high, and can be improved by optimization of the V_T implants. In addition, further improvements of the short channel behavior are needed for the p-FETs.

The transistor characteristics I_{DS}-V_{DS} and I_{DS}-V_{GS} of 70nm gate length p- and n-FETs are shown in Figures 2.16 and 2.17 respectively. Both n- and p-FETs showed very well behaved transistor characteristics, but the drive currents are considerably lower than expected for this EOT value and gate length. One main reason is the gate depletion problem, which results in lower gate over drive than the actual applied gate voltage.

In addition, the channel carrier mobility values are found to be both low. Figure 2.18 shows the electron and hole mobilities as a function of the effective vertical field. The universal mobility model, which represents the mobility achieved with high quality SiO_2 gate dielectric, is also shown in the figure for comparison [2.25]. It can be seen that

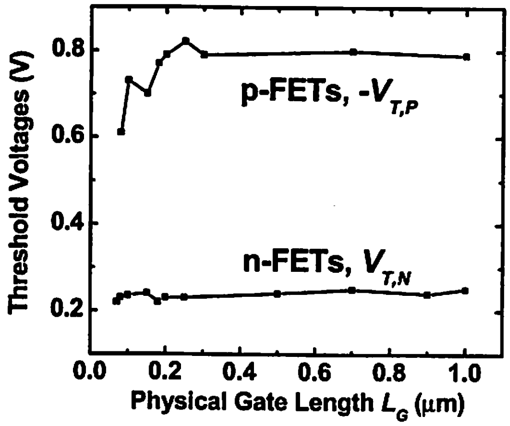


Figure 2.14 The threshold voltages of n-FETs ($V_{T,N}$) and p-FETs ($V_{T,p}$) with different gate lengths.

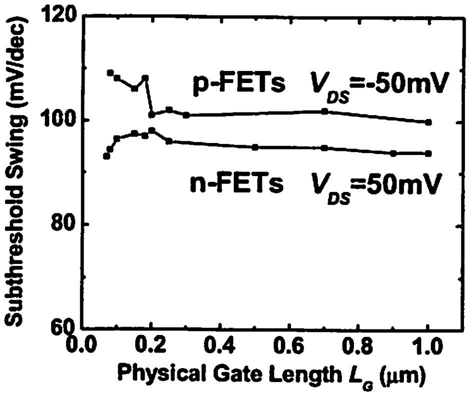


Figure 2.15 The linear subthreshold swings of n-FETs and p-FETs with different gate lengths.

in the high field range showed less degradation from the universal model, and a crossover of the two curves occurred at about 2.5 MV/cm. This crossover phenomenon had been reported for n-MOSFETs with a nitrided-SiO_2 gate dielectric [2.28]. This similarity suggests that the nitridation pretreatment of the Si substrate may have played a more important role in determining the carrier mobility than the upper HfO_2 film. The high field mobility degradation indicates that the interface roughness for the HfO_2 gate dielectric with a nitrided substrate is also a problem.

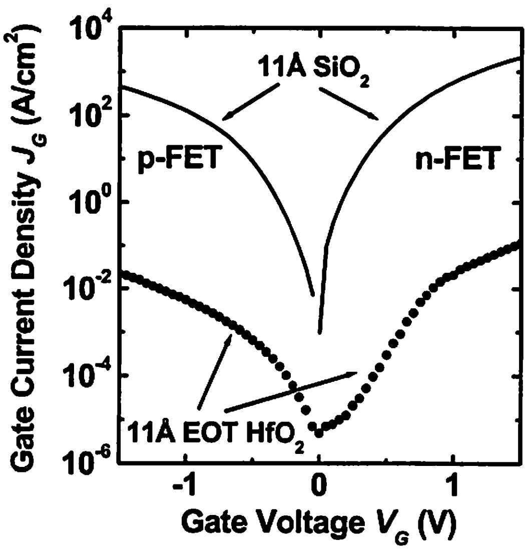


Figure 2.19 Gate leakage current of n-FETs and p-FETs with 11 Å EOT under inversion gate bias. Simulated gate leakage current of 11 Å SiO_2 is shown for comparison.

The gate leakage currents of the HfO_2 dielectric with an 11 Å EOT are shown in Figure 2.19, and compared with the simulated curve for 11 Å SiO_2. It can be seen that in the inversion regime, both n-FETs and p-FETs with HfO_2 gate dielectrics shows ~10^4× reduction in gate leakage compared to SiO_2. This significant advantage allows the EOT of the HfO_2 gate dielectric to be further reduced, ensuring its good scalability.

2.5 Hot carrier reliability of HfO_2 n-FETs

2.5.1 Hot carrier reliability measurement

Hot carrier reliability is a very important aspect of IC devices and circuits. In short-channel devices, the electric field near the drain region is high enough to create carriers with high energy, or hot carriers. The vertical electrical field in the channel accelerates the hot carriers toward the interface of the gate dielectric, where they release the energy and create damage to the gate dielectric interface, such as increased interface traps and charge trapping in the gate oxide. Hot carrier stress results in degraded device characteristics, e.g., shifted threshold voltage, lower transconductance g_m, lower drain current, etc, and these performance degradations directly affect the circuits' functionality. The hot carrier reliability is sensitive to the gate dielectric material and process, therefore, studies of the hot carrier reliability are very important for proving the feasibility of alternative gate dielectrics.

In order to observe significant hot carrier effects, short channel transistors are necessary. In this experiment, 0.15 μm gate length n-FETs with HfO_2 gate dielectric and nitrided HfO_2/Si interface are used. The devices were fabricated in the process mentioned in the previous section. The n-FETs have an 11.2 Å EOT, with a ~10 Å interfacial layer which is nitrogen rich. Before hot carrier stress was applied, the fresh device characteristics were measured, then hot carrier stress was applied for a certain amount of time, followed by measurement of the post-stress device characteristics a few minutes after the stress was stopped, which allows for the transient charge de-trapping to settle. All stress and measurements were performed at room temperature. There are two

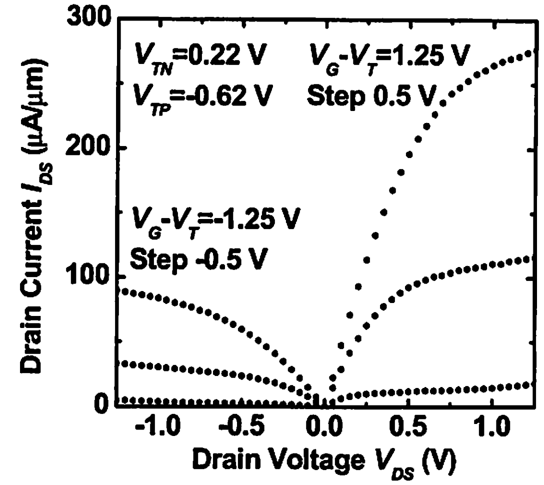


Figure 2.16 I_{DS}-V_{DS} of n-FETs and p-FETs with 70 nm gate length.

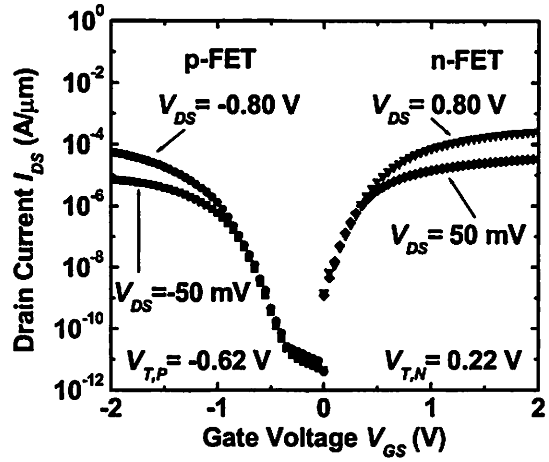
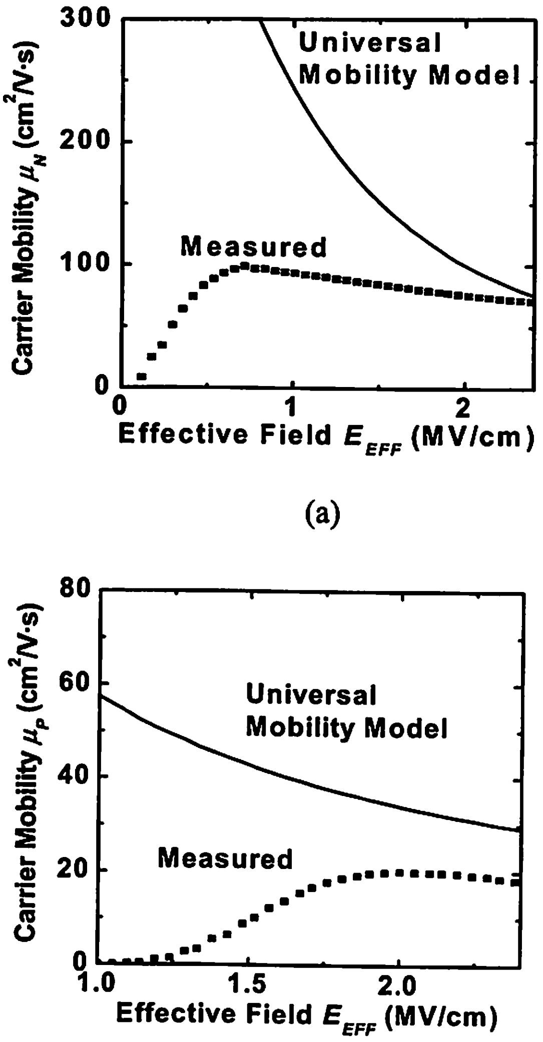


Figure 2.17 I_{DS}-V_{GS} of n-FETs and p-FETs with 70 nm gate length.

the channel electron mobility for the HfO_2 gate dielectric is much lower than the universal model at low field, and the difference gets smaller in the high field range. The hole mobility for the HfO_2 gate dielectric shows degradation from the universal model by roughly a constant factor in a wide field range. In the low field range, the dominant scattering mechanism is Coulombic scattering, so the low field mobility degradation for HfO_2 suggests enhanced Coulombic scattering, possibly by fixed charge or interface



(a)

(b)

Figure 2.18 (a) Electron and (b) hole mobilities as a function of effective vertical field of HfO_2 gate dielectric. Both showed significant degradation from the universal mobility model [2.27].

trapped charge, which are commonly observed in high-*k* gate dielectrics [2.26]. In the high field range, the dominant scattering mechanism is surface roughness scattering. This mechanism is affected by the inversion carriers' locations relative to the gate dielectric and substrate interface. The different behaviors of the electrons and holes are possibly due to the fact that electrons and holes have different effective masses, therefore the quantum confinement and carrier spatial distributions are different. The electron mobility commonly used hot carrier stress conditions, i.e., the peak substrate current stress and the maximum drain voltage stress. The substrate current is created in the impact ionization process caused by the hot electrons in the channel, and is therefore a good indicator of the hot electron damage. The coincidence of maximum degradation in device characteristics and the peak substrate current has been observed [2.29]. In recent years, however, there have also been reports that the worst hot electron stress condition can switch from peak substrate current to maximum gate voltage when the effective channel length is scaled toward 0.1 μm [2.30]. The worst-case stress condition depends on the specific process technology and the gate length. In this experiment with the HfO_2 gate dielectric, the maximum gate voltage condition is not practical. Given the thin EOT of the HfO_2 gate dielectric, the breakdown voltages are fairly low, and the maximum gate voltage stress will cause very significant Fowler-Nordheim stress to the gate dielectric, therefore the device degradation can be a combined effect of the Fowler-Nordheim stress and the hot

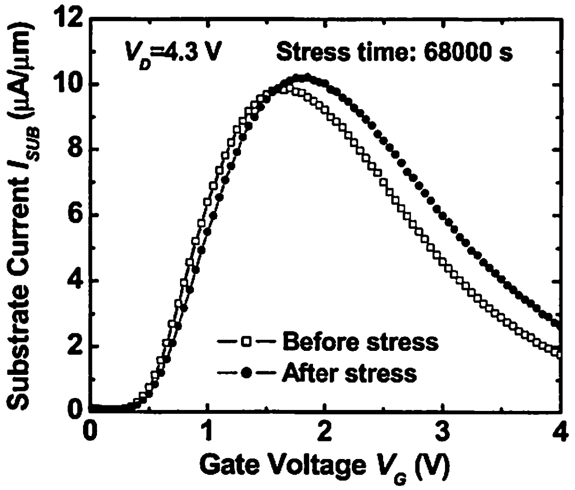


Figure 2.20 A typical substrate current I_{SUB} vs. gate voltage characteristic of fresh and stressed device. Drain current degraded after the stress, which shifted the peak of the substrate current to slightly higher gate voltage. In a series of stress this small change was not taken into account.

carrier stress, making the results difficult to interpret. So in this experiment, peak substrate current was used as the stress condition.

A typical substrate current vs. gate voltage characteristic is shown in Figure 2.20. For a given drain voltage, the substrate current initially rises with the gate voltage due to the increase of the drain current, and falls at higher gate voltage due to the decrease of maximum electric field in the channel. The peak substrate current corresponds to a gate voltage of 1/3 – 1/2 of the drain voltage. It can also be seen that the location and the height of the peak slightly changed after the stress. This is due to the degradation of the drain current and possibly an increase in the maximum field near the drain region. In the multiple stress-measurement cycles, this small change was not adjusted in the stress setup, i.e., the stress condition was kept at the initial peak substrate current throughout a series of stress.

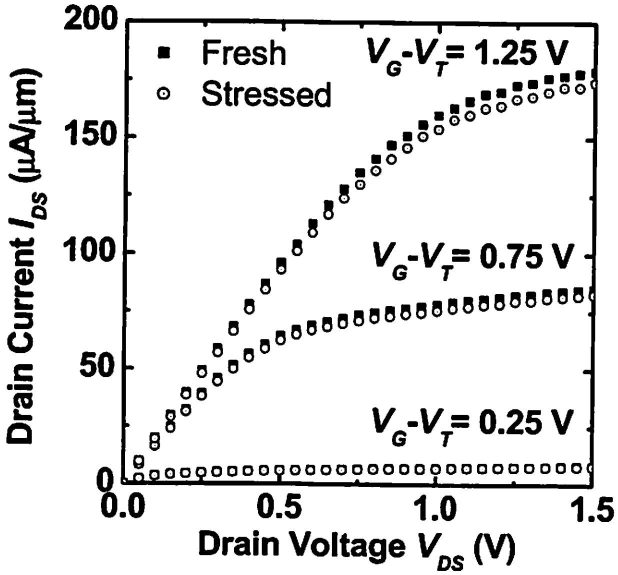


Figure 2.21 Transistor characteristics of a fresh device and the same device after 2000 s stress. V_T shift is negligible, and linear as well as saturation drain current at different gate voltage are degraded by a similar percentage.

Figure 2.21 shows the transistor I_{DS} - V_{DS} curve of a fresh n-FET and the same device after 2000 s stress at the peak substrate current of 0.19μA/μm. It can be seen that the drain current at different gate bias voltages are degraded roughly by the same percentage. The change in threshold voltage of the device after the stress was very small, and has insignificant effect on the drive current.

Four stress conditions were used in this experiment. The corresponding peak substrate current and drain voltage are shown in Figure 2.22. For each stress condition, the key device parameters were monitored after each period of stress until the device lifetime was reached. The lifetime is defined as the stress time when the saturation drain current I_{DM} is degraded by 5%, as I_{DM} is a very important parameter for device operation. The above range of stress conditions roughly span two orders of magnitude of observed device lifetime.

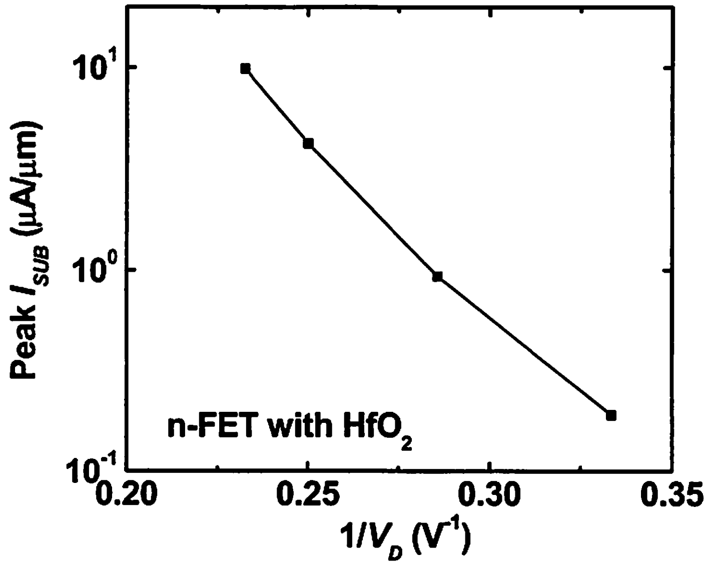


Figure 2.22 The peak substrate current as a function of the inverse of the drain voltage for n-FETs with HfO_2 gate dielectric.

Changes in key device parameters are plotted against accumulated stress time in Figure 2.23. Linear drain current I_{DL}, saturation drain current I_{DM}, and threshold voltage V_T are shown in the figure. The linear and saturation drain current change can be fitted very well by a power-law dependence on the stress time t,

$$\Delta I_{DL,DM} = C \cdot t^n \tag{2.1}$$

where C is a constant and is different for I_{DL} and I_{DM}. The range of n is about ~0.42-0.48. This power-law relation has been widely observed for n-MOSFETs with SiO_2 gate dielectrics stressed at peak substrate current condition. The n value is also similar to those reported for SiO_2 [2.29] [2.31].

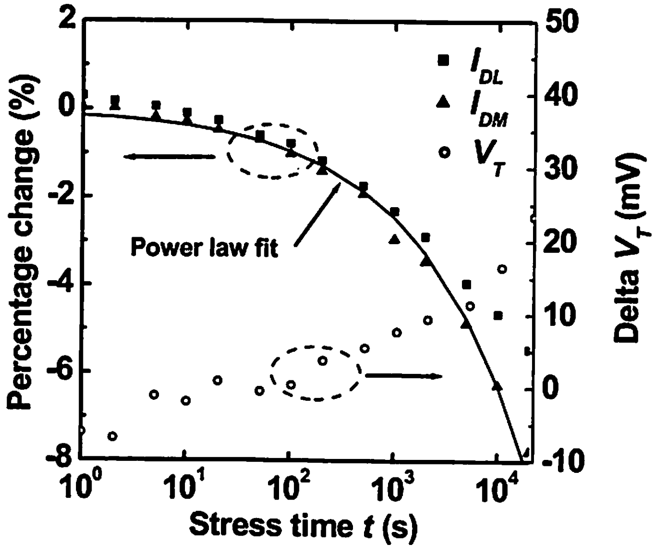


Figure 2.23 Typical trend of the relative change in linear (I_{DL}) and saturation (I_{DM}) drain current and V_T shift as a function of accumulated stress time. The percentage change in drain currents can be well fitted by a power law dependence on the stress time. The V_T shift is small, and does not show a clearly preferred functional form of time dependence.

After a significant amount of stress, the increase in threshold voltage was relatively small. The time dependence may be better fitted by a linear rather than a power-law form, although there is not a clearly best fit. It is worth mentioning that during the initial part of the stress, the V_T shift apparently increased faster than a power-law form.

It is interesting to compare the above results with the degradation of SiO_2 n-MOSFETs. It is known that under the peak I_{SUB} stress condition, the major degradation mechanism for SiO_2 n-MOSFETs is interface trap generation, and the $n \approx 0.5$ power-law applies to the degradation of drain currents, transconductance, etc. [2.31]. In the case of HfO_2 with a nitrided interface, the above power-law degradation of drain currents also suggests that interface trap generation is the dominant damage mechanism. However, the fact that the V_T shift is faster than a power-law behavior during the initial stage of the stress suggests that other types of damage mechanisms may exist. The positive change in V_T is possibly due to electron trapping. After a considerable amount of stress, the interface trap generation process can also affect the V_T shift. When the devices reached their lifetime, i.e., at 5% degradation of saturation drain current, the V_T increase was still fairly small (~50 mV), which has very little effect on the saturation drain current. So the main reason for the degradation of drive current was reduced channel carrier mobility caused by interface trap generation, which enhances the scattering of the channel carriers.

2.5.2 Lifetime projection and comparison with SiO_2

In order to evaluate the reliability of the HfO_2 gate dielectric, n-MOSFETs with a SiO_2 gate dielectric were used as control devices. The SiO_2 n-MOSFETs have a physical gate oxide thickness of 16 Å and gate length of 0.18 μm. These SiO_2 devices were fabricated in a different CMOS process where the source/drain extension region design was optimized for performance, and may negatively impact hot carrier reliability. So a direct comparison with the HfO_2 devices based on stress voltages is difficult. However, as can be seen in the following section, a lifetime comparison based on substrate current can still be a valid way of evaluating the transistors from different fabrication processes.

The same stress-measurement cycles were performed on the control devices, and the lifetime followed the same definition as used for the HfO_2 devices. The peak substrate currents used in the stress are shown in Figure 2.24.

Based on the measured lifetime, the long-term hot carrier reliability of these devices can be projected. Figure 2.25 shows the linear extrapolation of the drain voltages for ten-year lifetime based on $1/V_D$. It can be seen that the HfO_2 devices has a significantly higher ten-year operating voltage than the control devices. As there is a fundamental tradeoff between hot carrier reliability and device performance, the different source/drain design of the HfO_2 and SiO_2 devices make this comparison difficult to interpret.

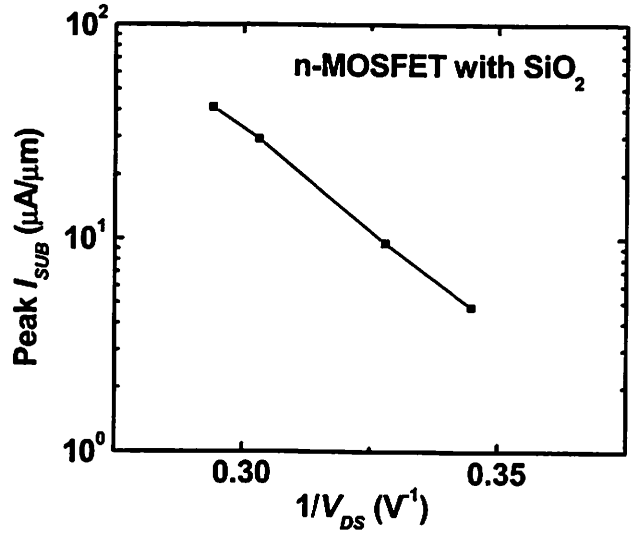


Figure 2.24 The peak substrate currents as a function of the inverse of the drain voltage for n-MOSFETs with SiO_2 gate dielectric.

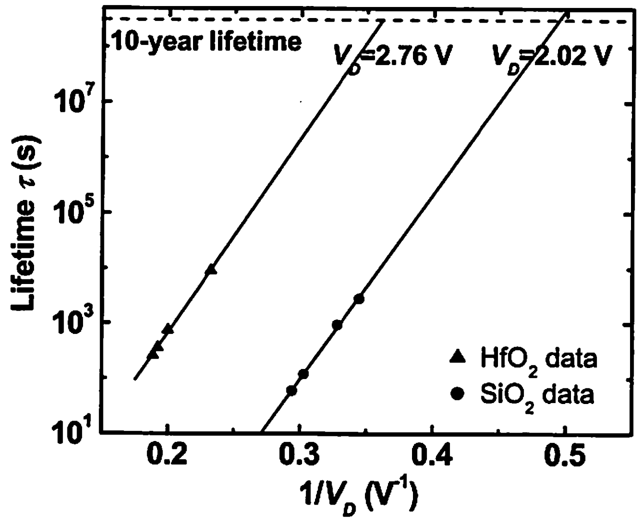


Figure 2.25 Hot carrier stress lifetime extrapolation of n-FETs with HfO_2 and SiO_2 gate dielectrics based on drain voltage. The different source/drain designs result in significantly different drain voltages for ten-year lifetime operation.

A meaningful comparison can be made based on the peak substrate current used for the stress, which is a quantitative indicator of the physical driving force that is responsible for hot carrier damage. The lifetime data are therefore plotted against the peak substrate current in Figure 2.26. To further examine the validity of this comparison, data from other published reports using SiO_2 gate dielectric with similar gate length but different source/drain designs are also shown in the figure [2.32][2.33]. It can be seen that these SiO_2 data from the literature and the data measured on the control devices are in a similar range when plotted against the peak substrate current, and the effects of the different source/drain designs among those devices can be largely reduced using this method. It should be noted that the data from the literature are based on different gate oxide thicknesses and processes, and the lifetime definitions are slightly different, so a universal trend line is not expected. Still, the lifetime of the HfO_2 n-FETs is apparently

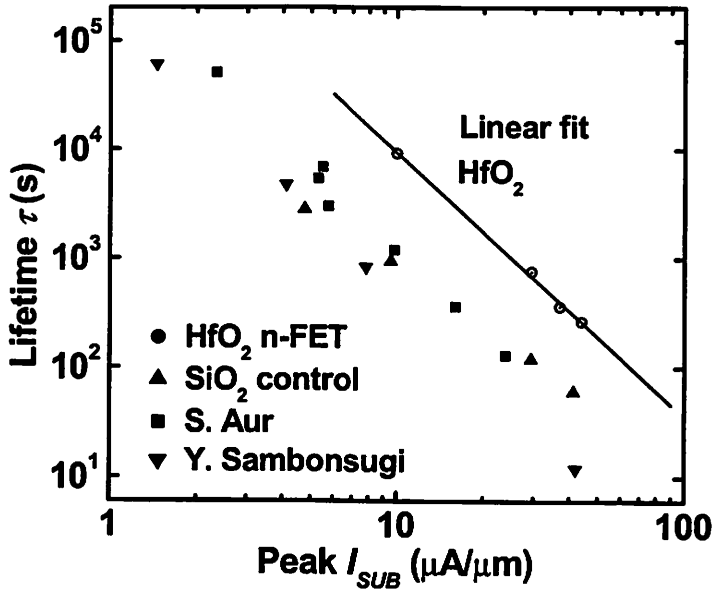


Figure 2.26 When plotted against the peak substrate current, the hot carrier stress lifetime of HfO_2 devices are higher than the SiO_2 data, which are either measured in this experiment or collected from the literature.

better than that of the SiO_2 devices. This suggests that the HfO_2 gate dielectric with nitrided interface allow sufficient design space to optimize the reliability-performance tradeoff.

The longer hot-carrier lifetime of the devices with the HfO_2 gate dielectric is not too surprising. As discussed in earlier sections, there is a thin interfacial layer formed by NH_3 nitridation underneath the HfO_2 gate dielectric. Given the relatively thin physical thickness of the two dielectric layers, the effect of bulk charge trapping should be relatively small, and the nitrided surface is the key factor affecting the hot carrier reliability. A report on the hot carrier reliability of SOI MOSFETs with JVD nitride gate dielectrics also showed better hot carrier lifetime than SiO_2 control devices [2.34]. As the nitrogen concentration in the JVD nitride is much lower than in a stoichiometric nitride, the interface composition may be similar to that obtained by NH_3 nitridation, as used in this experiment. Therefore the improved hot carrier reliability observed here is essentially the effect of the nitrided gate dielectric interface. In a study of the interface properties of NH_3 nitrided-SiO_2/Si interface [2.35], it was found that nitridation of the SiO_2/Si in general causes the degradation of the interface properties, such as increased surface roughness, higher interface trap density and larger hole capture cross-section. On the other hand, the hot carrier induced interface trap generation is much lower for nitrided SiO_2 due to the structural modification of the interface introduced by nitrogen. The hot carrier hardness of the interface was shown to depend strongly on nitridation condition, and lower temperature and longer nitridation time is preferred for MOSFET gate dielectric purpose. These findings suggest a possible way to further optimize the HfO_2 gate dielectric with a nitrided interface. Other benefits and effects of such a nitrided interface will be discussed in Chapter 3.

While this experiment did not reveal the hot carrier reliability of a "true" HfO_2/Si interface (which is very likely a thin layer of hafnium silicate after high temperature processes [2.23]), it showed that with NH_3 nitridation pretreatment of the substrate surface, the hot carrier hardness of the gate dielectric can be sufficient for the use in short-channel devices.

2.6 References

[2.1] M. S. Krishnan, L. Chang, T.-J. King, J. Bokor, and C. Hu, "MOSFETs with 9 to 13 A Thick Gate Oxides", *International Electron Devices Meeting*, pp. 241-244, Dec. 1999.

[2.2] M. Cao, P. V. Voorde, M. Cox, W. Greene, "Boron diffusion and penetration in ultrathin oxide with poly-Si gate", *IEEE Electron Device Letters*, Vol.19, No.8, pp. 291-293, Aug. 1998.

[2.3] T. P. Ma, "Making silicon nitride film a viable gate dielectric", *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 680-690, March, 1998.

[2.4] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices", *Journal of Vacuum Science and Technology*, B 18(3), pp. 1785-1791, May/Jun, 2000.

[2.5] D. Buchanan, E. P. Gusev, E. Cartier, H. Okorn-Schmidt, K. Rim, M. A. Gribelyuk, A. Mocuta, A. Ajmera, M. Copel, S. Guba, N. Bojarczuk, A. Callegari, C. D'Emic, P. Kozlowski, K. Chan, R. J. Fleming, P. C. Jamison, J. Brown, R. Arndt, "80nm poly-silicon gated n-FETs with ultra-thin Al_2O_3 gate dielectric for ULSI applications", *International Electron Devices Meeting*, pp.223-226, Dec. 2000.

[2.6] Y. Momiyama, H. Minakata, T. Sugii, "Ultra-thin Ta_2O_5/SiO_2 gate insulator with TiN gate technology for 0.1μm MOSFETs", *Symposium on VLSI Technology*, pp. 135-136, June 1997.

[2.7] S. A. Campbell, D. C. Gilmer, X. Wang, M. Hsieh, H.-S. Kim, W. L. Gladfelter and J. Yan, "MOSFET transistors fabricated with high permittivity TiO_2 dielectrics", *IEEE Transactions on Electron Devices*, vol. 44,no. 1, pp.104-109, Jan. 1997.

[2.8] X. Guo, T. P. Ma, T. Tamagawa and B. L. Halpern, "High quality ultra-thin TiO_2/Si_3N_4 gate dielectric for giga scale MOS technology", *International Electron Devices Meeting Tech. Dig.*, pp.377-380, 1998.

[2.9] A. Chin, Y. H. Wu, S. B. Chen, C. C. Liao, W. J. Chen, "High quality La_2O_3 and Al_2O_3 gate dielectrics with equivalent oxide thickness 5-10Å", *Symposium on VLSI Technology*, pp. 16-17, June, 2000.

[2.10] D. G. Schlom and J. H. Haeni, "A thermodynamic approach to selecting alternative gate dielectrics", *MRS Bulletin*, pp.198-204, March 2002.

[2.11] W.-J. Qi, R. Nieh, B. H. Lee, K. Onishi, L. Kang, Y. Jeon, J. C. Lee, V. Kaushik, B.-Y. Neuyen, L. Prabhu, K. Eisenbeiser, J. Finder, "Performance of MOSFETs with ultra-thin ZrO_2 and Zr-silicate gate dielectrics", *Symposium on VLSI Technology*, pp. 40-41, June, 2000.

[2.12] L. Kang, Y. Jeon, K. Onishi, B.-H. Lee, W.-J. Qi,, R. Nieh, S. Gopalan and J. C. Lee, "Single-layer thin HfO_2 gate dielectric with n^+-polysilicon gate", *Symposium on VLSI Technology*, pp. 44-45, June, 2000.

[2.13] G. D. Wilk and R. M. Wallace, "Electrical properties of hafnium silicate gate dielectrics deposited directly on silicon", *Applied Physics Letters*, Vol. 74, No. 19, pp. 2854–2856, 10 May 1999.

[2.14] B. Cheng, M. Cao, R. Rao, A. Inani, P. V. Voorde, W. M. Greene, J. M. C. Stork, Z. Yu, P. M. Zeitzoff and J. C. S. Woo, "The impact of high-*k* gate dielectrics and metal gate electrodes on sub-100 nm MOSFETs", *IEEE Transactions on Electron Devices*, Vol. 46, No. 7, pp. 1537-1542, july 1999.

[2.15] D. J. Frank and H. S. P. Wong, "Analysis of the design space available for high-*k* gate dielectrics in nanoscale MOSFETs", *Superlattices and Microstructures*, Vol. 28, No. 5/6, pp. 485-491, 2000.

[2.16] M. Yoshida, "MOSFET carrier mobility model based on gate oxide thickness, threshold and gate voltages", *Solid State Electronics*, Vol. 39, No. 10, pp.1515-1518, Oct. 1996.

[2.17] K. Yang, Y.-C. King, and C. Hu, "Quantum effects in oxide thickness determination from capacitance measurement", *Symposium On VLSI Technology*, pp. 77-78, June 1999.

[2.18] B. Y. Kim, H. F. Luan, and D. L. Kwong, "Ultra thin (<3 nm) high quality nitride/oxide stack gate dielectrics fabricated by in-situ rapid thermal processing", *Internaitional Electron Devices Meeting*, pp. 463-466, Dec. 1997.

[2.19] L. C. Parrillo, J. R. Pfiester, J.-H. Lin, E.O. Travis, and R. D. Sivan, "An advanced 0.5 mm CMOS disposable LDD spacer technology", *Symposium on VLSI Technology*, pp. 31-32, June 1989.

[2.20] X. Wang, M. Khare, T. P. Ma, "Effects of water vapor anneal on MIS devices made of nitrided gate dielectrics", *Symposium On VLSI Technology*, pp. 226-227, June 1996.

[2.21] W.-C. Lee and C. Hu, "Modeling gate and substrate currents due to conduction and valence band electron and hole tunneling", *Symposium On VLSI Technology*, pp. 198-199, June 2000.

[2.22] Y.-C. Yeo, Q. Lu, W. Lee, T. King, C. Hu, X. Wang, X. Guo, T.P. Ma, "Direct tunneling gate leakage current in transistors with ultra-thin silicon nitride gate dielectric", *IEEE Electron Device Letters*, Vol. 21, No. 11, pp. 540-542, November 2000.

[2.23] L. Kang, K. Onishi, Y. Jeon, B. H. Lee, C. Kang, W.-J. Qi, R. Nieh, S. Gopalan, R. Choi and J. Lee, "MOSFET devices with polysilicon on single-layer HfO_2 high-*k* dielectrics", *International Electron Devices Meeting*, pp.35-38, Dec 2000.

[2.24] C. Hu, "Gate oxide scaling limits and projection", *International Electron Devices Meeting*, pp. 219-322, Dec 1996.

[2.25] L. Kang, Y. Jeon, K. Onishi, B.-H. Lee, W.-J. Qi, R. Nieh, S. Gopalan and J. C. Lee, "Single-layer thin HfO_2 gate dielectric with n^+-polysilicon gate", *Symposium on VLSI Technology*, pp. - , June 2000.

[2.26] E. P. Gusev, D. A. Buchanan, E. Cartier, A. Kumar, D. DiMaria, S. Guba, A. Callegari, S. Zafar, P. C. Jamison, D. A. Neumayer, M. Copel, M. A. Gribelyuk, H. Okorn-Schmidt, C. D'Emic, P. Kozolowski, K. Chan, N. Bojarczuk, L.-A. Ragnarsson, P. Ronsheim, K. Rim, R. J. Fleming, A. Mocuta, and A. Ajmera, "Ultrathin high-k gate stacks for advanced CMOS devices", *International Electron Devices Meeting*, pp. 451-454, Dec. 2001.

[2.27] K. Chen, C. Hu, P. Fang, M. R. Lin and D. L. Wollensen, "Predicting CMOS speed with gate oxide voltage scaling and interconnect loading effects", *IEEE Tran. Electron Devices*, Vol. 44, pp. 1951-1957, Nov. 1997.

[2.28] Z. J. Ma, Z. H. Liu, Y. C. Cheng, P. K. Ko, and C. Hu, "New insight into high-field mobility enhancement of nitride-oxide n-MOSFETs based on noise measurement", *IEEE Transactions on Electron Devices*, Vol. 41, No. 11, pp. 2205-2209, Nov. 1994.

[2.29] C. Hu, S. C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation – model, monitor, and improvement", *IEEE Transactions on Electron Devices*, Vol. 32, No. 2, pp. 375-385, Feb. 1985.

[2.30] E. Li, E. Rosenbaum, J. Tao, G. C.-F. Yeap, M.-R. Lin, and P. Fang, "Hot carrier effects in nMOSFETs in 0.1 μm CMOS technology", *International Reliability Physics Symposium*, pp. - , April 1999.

[2.31] B. S. Doyle, K. R. Mistry, and J. Faricelli, "Examination of the time power law dependencies in hot carrier stressing of n-MOS transistors", *IEEE Transactions on Electron Devices*, Vol. 18, No. 2, pp. 51-53, Feb. 1997.

[2.32] Y. Sambonsugi and T. Sugii, "Hot-carrier degradation mechanism and promising device design of nMOSFETs with nitride sidewall spacer", *IEEE International Reliability Physics Symposium*, pp. 184-188, March 1998.

[2.33] S. Aur, T. Grider, V. McNeil, T. Holloway, and R. Eklund, "Remote plasma nitridation, deuterium anneal and pocket implant effects on NMOS hot carrier reliability", *Microelectronics Reliability*, Vol. 9, No. 5, pp. 673-679, May 1999.

[2.34] S. Mahapatra, V. R. Rao, J. Vasi, B. Cheng, and J. C. S. Woo, "Reliability studies on sub 100 nm SOI-MNSFETs", *IRW Final Report*, pp. 29-31, 2000.

[2.35] Z. Liu and Y. C. Cheng, "Properties of very thin thermally nitrided-SiO_2/Si interface based on conductance and hot-electron injection techniques", *IEEE Transactions on Electron Devices*, Vol. 36, No. 9, pp. 1629-1633, Sep. 1989.

Appendix 2.1

Sub-100 nm gate length n-well CMOS process flow

Original process flow designed by Wen-Chin Lee.
Modified for studying novel gate stack materials.
CMOS mask set originally designed by Ya-Chin King, with modifications to gate, contact and metal masks.

STEP	PROCESS	PROCESS CONDITIONS	EQUIPMENT	COMMENT
1.0	Wafers	4" (100 mm) bulk p-type Si		
2.0	Nitride/oxide			
2.1	Pre-clean	Piranha clean + 25:1 HF 20sec	Sink6	
2.2	Initial oxidation	SGATEOX, 1000°C, 21min. O_2, 15min. N_2	Tylan5/6	T_{ox}=200Å
2.3	Nitride	SNITC, 800°C, 35min.	Tylan9	T_{SiN}=1600Å
2.4	Label and Rinse	Label on backside, DI water rinse	Sink6	
3.0	Well Formation			
3.1	Prebake	Bake and HMDS coat	Primeoven	
3.2	PR coating	Standard I-line, program#1/1 on coater/oven	Svgcoat1/2	T_{PR}=1. 1μm
3.3	Exposure	Use posted exposure time and focus	Gcaws	Mask NMS
3.4	PEB	120°C, 60sec., program#1 for oven	Svgdev	
3.5	Development	Program#1 for developer station	Svgdev	
3.6	Descum	O_2 ashing for 1min., 50W	Technics-c	
3.7	Hard bake	120°C, 30 min	VWR oven	
3.8	Nitride etching	Program NITSTD1 with 90% auto endpoint	Lam1	
3.9	n-well implant	$^{31}P^+$ 150KeV 5.0×10^{12} cm^{-2}	Vendor	Rp=0.2μm
3.10	PR ashing	O_2 ashing for 5min., 240W	Technics-c	
3.11	Cleaning	Piranha clean	Sink8	
3.12	Cleaning	Piranha clean + 10:1 HF 1.5min	Sink6	Check dewet
3.13	Well Drive-in	WELLDR, 1100°C 2.5hrs O_2, 30 min.. N_2	Tylan2	X_j=1.5um, T_{OX}=2000Å
3.14	Oxide wet etching	10:1 HF 1 min	Sink6	Remove oxide on nitride
3.15	Nitride wet etching	H_3PO_4 acid, 150°C, 3hrs (fresh solution)	Sink7	Check oxide thickness
3.16	Oxide wet etching	5:1 BHF	Sink8	Check dewet
4.0	Nitride/oxide			
4.1	Clean	Piranha clean + 25:1 HF 20sec	Sink6	
4.2	Pad oxidation	SGATEOX, 1000°C, 21min. O_2, 15min. N_2	Tylan5	T_{ox}=200Å
4.3	Nitride	SNITC, 800°C, 35min	Tylan9	T_{SiN}=1600 Å
5.0	Active region definition			
5.1	Active Area Lithography	Cover active area	Svgcoat1/2, Gcaws, Svgdev	T_{PR}=1. 1μm Mask ND
5.2	Nitride etching	Program NITSTD1 with 90% auto endpoint	Lam1	
5.3	Hard bake	120°C, more than 2 hrs	VWR oven	
6.0	Field Implant (p-type)			
6.1	Field lithography	Use field mask to cover n-well	Svgcoat1/2, Gcaws, Svgdev	T_{PR}=1.1μm Mask NG
6.2	Field Implant	$^{11}B^+$ 80KeV 2.0×10^{13} cm^{-2}	Vendor	
6.3	PR ashing	O_2 ashing for 5min., 240 W	Technics-c	

6.4	Cleaning	Piranha clean	Sink8	
7.0	LOCOS			
7.1	Cleaning	Piranha clean + 25:1 HF 5min	Sink6	
7.2	Field oxidation	WETOXO2A, 1000°C, 2hrs, 20min. N_2	Tylan1/2	Tox=5800 Å
7.3	Oxide wet etching	10:1 HF 30 sec	Sink6	Remove oxide on nitride
7.4	Nitride wet etching	H_3PO_4 acid, 150°C, 3 hrs (fresh solution)	Sink7	Check oxide thickness
8.0	V_T Implant			
8.1	Cleaning	Piranha clean + 10:1 HF 1.5min	Sink6	Check dewet
8.2	Sacrificial oxidation	WETOXO2A, 900°C, 17min	Tylan1	T_{OX}=550 Å
8.3	Oxide wet etching	Piranha clean + 10:1 HF	Sink6	Check dewet
8.4	Screen oxidation	SGATEOX, 900°C, 35min. O_2, 15min. N_2	Tylan5	T_{OX}=100Å
8.5	Field lithography	Use field mask to cover n-well	Svgcoat1/2, Gcaws, Svgdev	T_{PR}=1.1µm Mask NG
8.6	nFET V_T implant	$^{49}BF_2^+$ 50KeV 1.2×10^{13} cm^{-2} Rp=0.04 µm	Vendor	Depending on T_{OX} and gate Φ_M
8.7	PR ashing	O_2 ashing for 5min, 240W	Technics-c	
8.8	Cleaning	Piranha clean	Sink8	
8.9	Well lithography	Use mask to open well area, cover n-FETs	Svgcoat1/2, Gcaws, Svgdev	T_{PR}=1.1µm Mask NMS
8.10	pFET V_T implant	$^{31}P^+$ 30KeV 2.0×10^{12} cm^{-2} Rp=0.04 µm	Vendor	Depending on T_{OX} and gate Φ_M
8.11	PR ashing	O_2 ashing for 5min, 240W	Technics-c	
8.12	Cleaning	Piranha clean	Sink8	
9.0	Gate dielectric			
9.1	Cleaning	Piranha clean + 10:1 HF 1min.	Sink6	Check dewet
9.2	Oxide wet etching	10:1 HF 1 min	Sink6	Remove sac ox, check dewet
9.3a	Oxide growth (SiO_2 control)	Recipe THIN_ANN, perform test run immediately before real run	Tylan5/6	TCA clean within 12 hours
9.3b	Oxide growth (high-*k*)	THIN_ANN, 800°C O_2 30 min, protection of Si during transportation	Tylan6	TCA clean, T_{ox}<100Å
9.4b	High-*k* dielectric deposition	Varied. Wafers with high-*k* must be vacuum sealed during transportation back		
9.5 b	Undoped poly-Si deposition	Recipe 11SUDPLYA	Tylan11	T_{Poly}=1500~1700 Å
10.0	Gate lithography	I-line lithography and PR ashing		
10.1	Prebake	Bake and HMDS coat	Primeoven	
10.2	PR coating	Standard I-line, program#1/1 on coater/oven	svgcoat1/2	T_{PR}=1.1µm
10.3	Exposure	To get best resolution and PR profile, do F/E test immediately before real exposure	Gcaws	Mask NP
10.4	PEB	120°C, 60sec, program#1 for oven	Svgdev	
10.5	Development	Program#1 for developer (NO hard bake)	Svgdev	
10.6	PR ashing	Lam5 offers better uniformity	Technics-c/Lam5	Reduce ~0.4 µm → ~0.1 µm
10.7	SEM inspect	Measure critical line width	Leo	
10.8	Hard bake	120°C, 10min	VWR oven	
11.0	Gate etching			
11.1	Dry etching	Recipe 5963, with CF_4 for BT, Cl_2, HBr for ME, HBr for OE	Lam5	
11.2	HF clean	100:1 HF 30 s to remove polymer	Sink7	
11.3	PR stripping	Fresh PRS3000 at 90 °C for 20min, program#2 (short rinse) on spindryer3	Sink5, spindryer3	Inspect PR residue
11.4	Cleaning	Pirahna clean	Sink8	
12.0	Disposable nitride spacer			
12.1	Cleaning	Pirahna clean	Sink6	

12.2	HTO deposition	9HOXN2OD, 800°C, 30min., SiH_2Cl_2 : N_2O=10sccm/50sccm; test run first	Tylan9	T_{ox}=75 Å
12.3	Nitride deposition	SNITC, 800°C, 24min	Tylan9	T_{SiN}=1000 Å
12.4	Nitride dry etching	Standard recipe#5100 on Lam5 or NITSTD1 on Lam1	Lam5/Lam1	Check remained field oxide thickness
13.0	N^+ implant	N^+ gate and S/D, n-well contact		
13.1	N^+-region lithography		Svgcoat1/2, Gcaws, Svgdev	Mask NI
13.2	N^+ implant	$^{31}P^+$, 12KeV, $3x10^{15}cm^{-2}$	Vendor	
13.3	PR stripping	O_2 plasma 250 W 5 min	Technics-c	Inspect PR residue
13.4	Cleaning	Pirahna clean	Sink8	
14.0	P^+ implant	P^+ gate and S/D, p-sub contact		
14.1	P-region lithography		Svgcoat1/2, Gcaws, Svgdev	Mask NB
14.2	P+ implant	$^{11}B^+$, 5KeV, $3x10^{15}cm^{-2}$	Vendor	
14.3	PR stripping	O_2 plasma 250 W 5 min	Technics-c	Inspect PR residue
14.4	Cleaning	Pirahna clean	Sink8 + sink6	
15.0	1st RTA			
15.1	Oxide wet etching	100:1 HF, 20sec.	Sink7	Remove oxide on nitride
15.2	Nitride wet etching	H_3PO_4 acid, 150°C, 3 hrs Include HTO/nitride test wafers. Make sure HTO layer is not etched through.	Sink7	H_3PO_4 acid attacks poly-Si gate
15.3	Cleaning	Pirahna clean	Sink8 + Sink6	
15.4	Gate and S/D annealing	900°C, 10sec + 1050°C, 5sec, N_2, 300sccm	Heatpulse3	Check ρ_S on test wafers
15.5	Oxide wet etching	25:1 HF, 7sec to remove liner SiO_2	Sink6	Reduce implant offset
16.0	LDD ion implant			Optimize for HCE
16.1	N-region lithography		Svgcoat1/2, Gcaws, Svgdev	Mask NI
16.2	N-LDD implant	$^{75}As^+$, 7KeV, $4x10^{14}cm^{-2}$	Vendor	
16.3	PR stripping	O_2 plasma 250 W 5 min	Technics-c	Inspect PR residue
16.4	Cleaning	Pirahna clean	Sink8	
16.5	P-region lithography		Svgcoat1/2, Gcaws, Svgdev	Mask NB
16.6	P-LDD implant	$^{49}BF_2^+$, 5KeV, $4x10^{14}cm^{-2}$	Vendor	
16.7	PR stripping	O_2 plasma 250 W 5 min	Technics-c	Inspect PR residue
16.8	Cleaning	Pirahna clean	Sink8	
17.0	Passivation			
17.1	Cleaning	Pirahna clean	Sink6	
17.2	LTO deposition	VDOLTOC, 450°C, 40min	Tylan12	T_{ox}=4000 Å
17.3	Densification + 2nd RTA	900°C, 10sec., N_2, 300sccm	Heatpulse3	Check ρ_S on test wafers
18.0	Contact definition			
18.1	Contact lithography		Svgcoat1/2, Gcaws, Svgdev	Mask NC
18.2	LTO dry etch	Recipe# 5003 BT only, leave ~500 Å to be removed by wet etch	Lam5	Check T_{OX} in contact hole region
18.3	PR stripping	O_2 plasma 250 W 5 min	Technics-c	
18.4	LTO wet etch	5:1 BHF ~ 10 s, Piranha clean	Sink8	

19.0	Metal deposition			
19.1	Cleaning	Pirahna + 25:1 HF 2min	Sink6	
19.2	Metal stack deposition	Ti: 1.1 kW, Ar 20 mTorr, 80cm/min, 1 pass TiN: 2.0 kW, Ar:N_2=10:10 mTorr, 36cm/min, 1 pass Al: 4.5 kW, 6mTorr, 36cm/min, 2 passes	CPA	Ti: 200Å TiN: 200Å Al-2%Si: 3000Å
20.0	Metal patterning			
20.1	Metal lithography	Hard bake and HMDS coat	Svgcoat1/2, Gcaws, Svgdev	Mask NM
20.2	Hard bake	120°C, 20min.	VWR oven	
20.3	Al wet etching	Fresh Al etchant 50°C ~20-30 s	Sink8	
20.4	TiN/Ti etching	Recipe# 5963 main etch step only, 30 s	Lam5	
21.5	PR Stripping	O_2 plasma 240 W, 5 min	Technics-c	
21.0	Sintering			
21.1	Cleaning	DI water rinse	Sink8	No pirahna or HF
21.2	Al sintering	Forming gas anneal 400°C, 30 min	Tylan13	Improve sheet resistance

Chapter 3

SiGe gate for HfO_2 gate dielectric

3.1 Introduction

The previous chapter demonstrated that very thin EOT can be achieved using high-*k* gate dielectrics, but the CET in the inversion region can still be considerably thicker than the EOT due to the poly-Si gate depletion and the quantum effects. While in the special case of the HfO_2 CMOS in chapter 2, the gate dopant activation could be significantly improved, this problem generally exists for all high-*k* gate dielectrics with a poly-Si gate. The reduction of the poly-Si gate depletion effect requires higher dopant activation anneal temperature and thermal budget, which enhance the reaction of the high-*k* gate dielectrics with the substrate or the gate, degrading many aspects of the gate stack properties. Possible approaches to solve this conflict are either to use a high-*k* gate dielectric with very stable properties during annealing at above 1000 °C, or to use a different gate electrode material which has less depletion effect than poly-Si. There is a known candidate material for the second approach, poly-Si_xGe_{1-x} (denoted as poly-SiGe for simplicity unless a specific Ge percentage is emphasized). Poly-SiGe has been proposed as an alternative gate material to the poly-Si gate for CMOS technology, with the benefit of better dopant activation and reduced boron penetration [3.1]. With a thin

SiO_2 gate dielectric, the poly-SiGe gate technology was optimized to obtain the best tradeoff between the gate depletion effect and short channel effects [3.2]. Although the benefits of poly-SiGe gate have been demonstrated with an SiO_2 gate dielectric, little is known about the behavior of the poly-SiGe gate in contact with high-*k* gate dielectrics. This chapter investigates the device and integration issues of poly-SiGe gate for high-*k* gate dielectrics.

3.1.1 Gate depletion effect

Poly-Si is a preferred gate material for CMOS technology due to its high temperature stability and good interface with SiO_2. It is important to obtain degenerate doping in the poly-Si gate to reduce the gate depletion effect. Different techniques can be used, such as *in situ* doping, ion implantation and solid source diffusion. For a single gate, *in situ* doping results in higher active dopant concentration [3.3], while in the case of a dual gate (n^+ and p^+), which is required by advanced CMOS technology, ion implantation must be used. High temperature for the dopant activation anneal is critical for obtaining high active dopant concentrations near the poly-Si gate/gate oxide interface, so the pre-doping technique [3.4], which decouples the S/D drain anneal from the gate activation anneal, can be used to achieve better device performance. But this advantage does not apply to high-*k* gate dielectrics, for which a more stringent thermal budget limit is imposed by stability concerns.

As mentioned in Chapter 1, the poly-Si gate depletion effect becomes pronounced when an ultra-thin gate dielectric is used, in which case the quantum confinement in the channel has a comparable effect on the CET. Therefore, a strict treatment of the poly-Si gate depletion effect requires self-consistently solving the Poisson equation across the entire gate stack and the Schrödinger equation in the substrate [3.5]. Computation intensive numerical methods are necessary for this approach. Despite this, a simple model can be derived to provide useful insights into the poly-Si gate depletion effect. Consider an n$^+$ poly-Si gated n-MOSFET biased in the inversion region (Figure 3.1), the flat-band voltage (V_{FB}), the voltage drop in the substrate (φ_S), the gate (φ_G), and across the gate oxide (V_{OX}) add up to the gate voltage V_G.

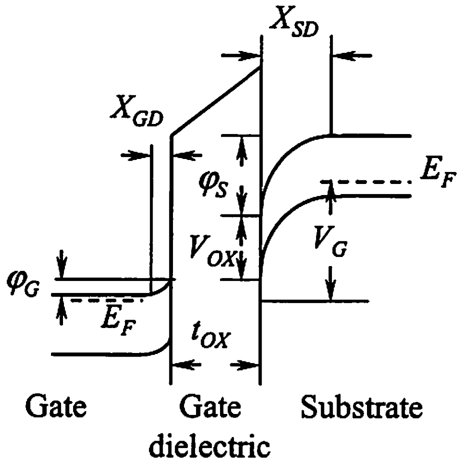


Figure 3.1 Voltage distribution of an n$^+$ poly-Si gated n-MOSFET biased in inversion region. The depletion region widths in the substrate and the gate are X_{SD} and X_{GD}, respectively.

$$V_G = V_{FB} + V_{OX} + \varphi_S + \varphi_G \qquad (3.1)$$

Firstly, in the case of low gate bias (depletion or weak inversion region), the very small amount of inversion charge in the channel can be neglected in the following equations. The boundary conditions at the two interfaces of the gate oxide relate V_{OX} to the electric field at the surface of (but within) the substrate (E_S) and gate (E_G).

$$\varepsilon_{OX} \cdot \frac{V_{OX}}{t_{OX}} = \varepsilon_{Si} \cdot E_S = \varepsilon_{Si} \cdot E_G \qquad (3.2)$$

Using the depletion approximation and assuming uniform doping concentrations in the gate (N_G) and the substrate (N_{SUB}), the electric field E_x can be solved from Gauss' Law. In the substrate,

$$\frac{dE_x}{dx}=\frac{qN_{SUB}}{\varepsilon_0\varepsilon_{Si}} \qquad (3.3)$$

therefore,

$$\varphi_S=\int_0^{X_{SD}}E_x\cdot dx=\frac{qN_{SUB}X_{SD}{}^2}{2\varepsilon_0\varepsilon_{Si}} \qquad (3.4)$$

And similarly,

$$\varphi_G=\frac{qN_GX_{GD}{}^2}{2\varepsilon_0\varepsilon_{Si}} \qquad (3.5)$$

From (3.2)-(3.5),

$$V_{OX}=\gamma_G\sqrt{\varphi_G}=\gamma_S\sqrt{\varphi_S} \qquad (3.6)$$

where $\gamma_{G,S}\equiv\sqrt{2\varepsilon_0\varepsilon_{Si}qN_{G,SUB}}\,/\,C_{OX}$.

Equation (3.6) indicates that $\varphi_G/\varphi_S=(\gamma_S/\gamma_G)^2=N_{SUB}/N_G$, which is typically ~$10^{-2}$. Therefore the gate depletion can be neglected unless strong inversion occurs. Beyond strong inversion, the inversion charge increases rapidly with the gate voltage, and φ_S saturates at $2\varphi_B=2(k_BT/q)ln(N_{SUB}/n_i)$. Therefore, (3.1) should be replaced by

$$V_G=V_{FB}+V_{OX}+2\varphi_B+\varphi_G \qquad (3.7)$$

Since $V_T=V_{FB}+2\varphi_B+\gamma_S\sqrt{2\varphi_B}$, (3.7) becomes

$$\varphi_G+V_{OX}-(V_G+\gamma_S\sqrt{2\varphi_B}-V_T)=0 \qquad (3.8)$$

The gate depletion region width X_{GD} can be solved from (3.5), (3.6) and (3.8).

$$X_{GD} = \frac{\varepsilon_{Si}\varepsilon_0}{C_{OX}}\left(\sqrt{1+\frac{2C_{OX}^2}{\varepsilon_{Si}\varepsilon_0 q N_G}(V_G - V_T + \gamma_S\sqrt{2\varphi_B})} - 1\right)$$

(for $V_G > V_T$) (3.9)

To account for the quantum effects, C_{OX} can be approximated by

$$C_{OX} = \frac{\varepsilon_0 \varepsilon_{OX}}{(T_{OX} + X_{DC}\varepsilon_{OX}/\varepsilon_{Si})} \qquad (3.10)$$

And the inversion charge centroid location X_{DC} can be either approximated as a constant or estimated using an analytical model [3.6]. Using typical device parameters in (3.9), N_G=1×10^{20} cm^{-3}, N_{SUB}=1×10^{18} cm^{-3}, C_{OX}=15×10^{-7} fF/μm^2, we can get V_T=0.26 V, and for V_G=1.26 V, X_{DG}=12 Å, which is electrically equivalent to 4 Å SiO$_2$. Figure 3.2 shows the EOT contribution from the gate depletion effect for different active gate doping concentrations and gate capacitance. Obviously a lower gate doping level results in a

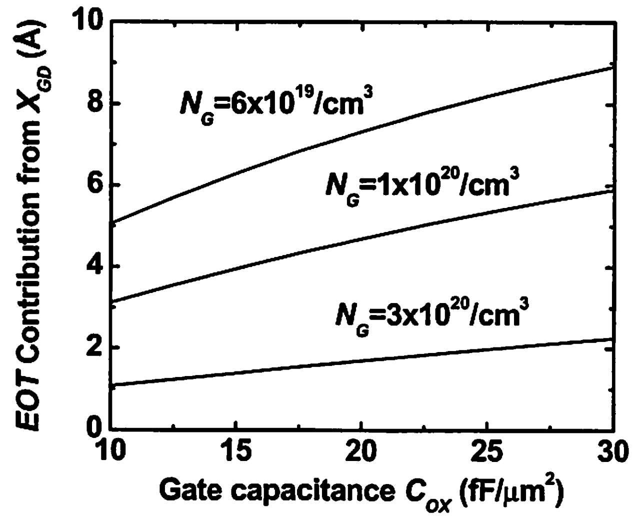


Figure 3.2 Contribution in EOT from the gate depletion effects as a function of the active gate doping and the gate capacitance. The same substrate doping of 1×10^{18}/cm^3 is assumed. Gate voltage is always set to V_T+1.0 V.

larger gate depletion region, i.e., a larger contribution to the EOT. When a thinner gate dielectric is used, γ_G becomes smaller, and (3.6) indicates that more voltage drop occurs in the gate depletion region. Therefore the gate depletion effect is more serious for thinner gate dielectrics. For an EOT below 10 Å, improving the active gate doping can have significant benefits in device characteristics. Using a metal gate electrode in principle can eliminate the gate depletion effect, but at this point, it is very difficult to implement in a manufacturable CMOS process. In this sense, the poly-SiGe gate CMOS technology is a good tradeoff between performance and integration complexity, therefore, it is a worthy candidate to be investigated for use on high-*k* gate dielectrics.

3.2 CMOS process of SiGe gate with HfO_2 gate dielectric

The Ge content in the poly-SiGe gate is an important design parameter in CMOS process integration. The dopant activation for boron doped p^+ gate is monotonically improved with higher Ge content, while the best dopant activation for a phosphorus doped n^+ gate occurs around 20% Ge [3.2]. In addition, Si and Ge have similar electron affinity, but significantly different bandgap (1.12 eV for Si and 0.66 eV for Ge). The bandgap of poly-SiGe varies with Ge content, therefore high Ge content results in a significant smaller bandgap than pure Si, and consequently a shifted valence band compared to the Si substrate. So the p^+ gate will have a lower work-function, and the threshold voltage of the p-MOSFETs will be relatively high. This can affect the choice of the channel doping, the short-channel performance, etc. In view of the above factors, a relatively low Ge content may be more appropriate for the experiment with high-*k* gate dielectrics. In this work, a poly-$Si_{0.75}Ge_{0.25}$ gate was used to fabricate CMOS transistors.

The CMOS process with a poly-SiGe gate is similar that used for a poly-Si gate and a HfO_2 gate dielectric, which was discussed in detail in Chapter 2. Control devices with a poly-Si gate were fabricated in the same process. In addition, the effects of the bottom nitride (BN) were also studied in this experiment. It was reported that the presence of a nitride layer at the bottom interface of the HfO_2 film, formed by NH_3 nitridation, can suppress the reaction between the HfO_2 and the Si substrate, therefore resulting in a thinner EOT after a high temperature anneal [3.7]. In the work by R. Choi et al [3.7], the gate electrode was TaN deposited by DC sputtering, and it would be interesting to see if such an effect exists when a poly-Si/SiGe gate is used. Figure 3.3 shows the four different gate stack structures used in this study. After identical processes for all the wafers up to the V_T-adjust implants, structures (a) and (b) received the same rapid thermal nitridation in NH_3 to form the bottom nitride layer. The HfO_2 layers are of

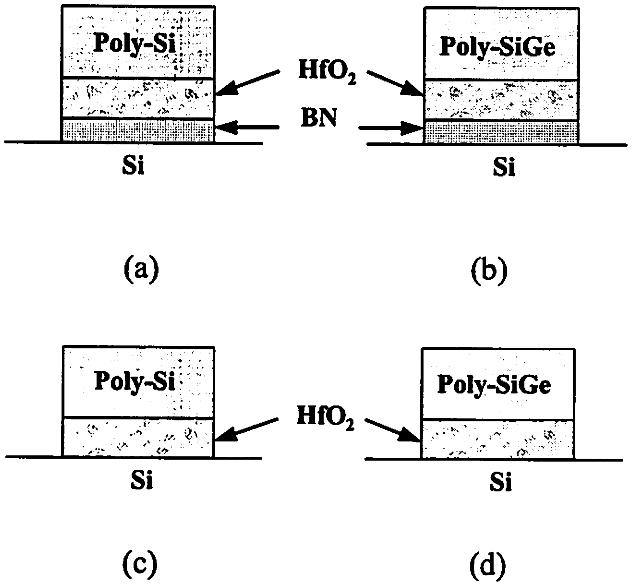


Figure 3.3 The gate stack structures used to study the effects of the poly-SiGe gate and bottom nitride (BN). The physical thickness of the HfO_2 layer is the same for (a) – (d), and the BN has the same physical thickness in (a) and (b).

the same thickness in all four splits, and were deposited by reactive sputtering in modulated oxygen. Following the gate dielectric formation, ~2000Å undoped Si or $Si_{0.75}Ge_{0.25}$ was deposited on the wafers by LPCVD. Prior to the poly-$Si_{0.75}Ge_{0.25}$ film deposition, a thin layer of Si (target thickness of ~50Å) was deposited using SiH_4 (550°C 1 min) to improve the nucleation of the poly-SiGe film. The poly-$Si_{0.75}Ge_{0.25}$ film was deposited at 550°C using SiH_4 and GeH_4. The Si gate was deposited using SiH_4 at 550°C. As deposited at this temperature, the $Si_{0.75}Ge_{0.25}$ film is polycrystalline, but the Si film is amorphous, and was crystallized by the subsequent high-temperature annealing. Keeping the same gate deposition temperature is important for making a fair comparison between the Si and the $Si_{0.75}Ge_{0.25}$ gate, as the reaction between the HfO_2 and the Si substrate is sensitive to temperature. Due to the much faster deposition rate of $Si_{0.75}Ge_{0.25}$, the deposition time of poly-$Si_{0.75}Ge_{0.25}$ was 20 min, in contrast to 75 min for Si. In order to evaluate the feasibility of these HfO_2 gate stacks in advanced CMOS processes as well as to reduce the gate depletion, the dopant activation anneal was done by RTA at 1000°C for 10 s for all wafers.

3.3 Device characterization

The effects of the poly-SiGe gate and the bottom nitride can be highlighted in side-by-side comparisons with the control devices. Figure 3.4 shows the *C-V* characteristics of the n-FETs with bottom nitride and poly-Si or poly-SiGe gate. In the accumulation/depletion region, the two curves essentially coincide. As in this region the gate capacitance is determined only by the gate dielectric and the channel doping, this suggests that the two devices have the same EOT and channel doping. Apparently the low Ge content in the gate didn't cause measurable degradation of the gate stack compared to the poly-Si gated control devices. Under positive gate bias, the poly-SiGe gated devices showed higher inversion capacitance, or thinner CET than the poly-Si gated control. From equation (1.3), and the same EOT and channel doping for the two devices, the thinner CET can be attributed to thinner gate depletion region, i.e., improved gate dopant activation. This confirms that the poly-SiGe gate is thermally stable with HfO2 during 1000°C anneal, and that the advantages in dopant activation can be retained.

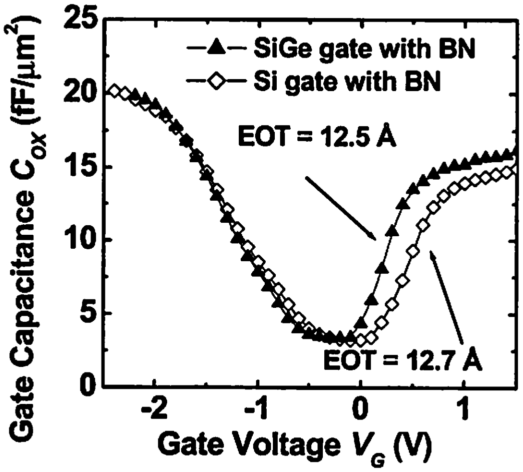


Figure 3.4 N-FETs' *C-V* characteristics with poly-Si or poly-SiGe gate and the bottom nitride (BN). The identical accumulation capacitances of both devices suggest that they have the same EOT, and the higher inversion capacitance of SiGe gated devices indicates improved gate dopant activation. EOT values were estimated for quantum *C-V* simulation [3.8].

When the bottom nitride is not present in the gate stack, the two gate electrodes resulted in very different device characteristics. As shown in Figure 3.5, in the accumulation region, the n-FETs with a poly-SiGe gate showed significantly higher gate capacitance, i.e, thinner EOT. The slightly lower gate capacitance in the depletion region (for V_G between –1 V and 0 V) of the poly-Si gated device is also due to its thicker EOT.

On the inversion side, the higher gate capacitance of the poly-SiGe gated devices is a combined result of the thinner EOT and the reduced gate depletion. It is worth mentioning that there are obvious V_T shifts between the two gate materials regardless of the bottom nitride. In all cases the threshold voltage is slightly shifted by the interface states that are related to the high-*k* gate dielectric, and for comparable interface states densities, the shift is larger for devices with thicker EOT. It is also possible that the two gate materials behaved differently in terms of dopant penetration.

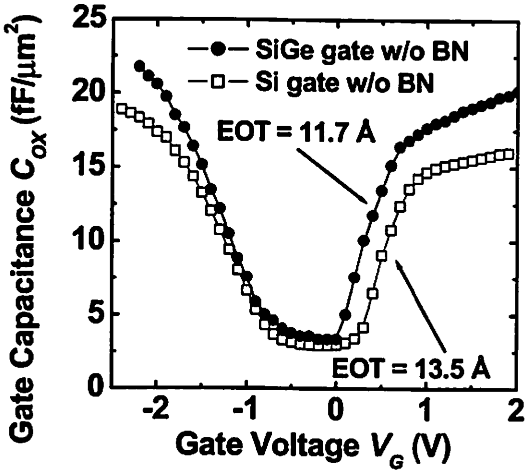


Figure 3.5 N-FETs' *C-V* characteristics with poly-Si or poly-SiGe and without the bottom nitride (BN). Higher accumulation capacitance the of SiGe gated devices indicates a thinner EOT, and the higher inversion capacitance of SiGe gated devices is due to improved gate dopant activation and thinner EOT.

Apparently, the bottom nitride had different effects on the devices when different gate materials are used. To examine its effects with the presence of poly-Si gate, the *C-V* curves of the Si gated n-FETs with or without bottom nitride are compared in Figure 3.6. It can be seen that the devices with the bottom nitride shows thinner EOT, in agreement with the previous report based on the TaN gate electrode [3.7]. Also visible in the figure

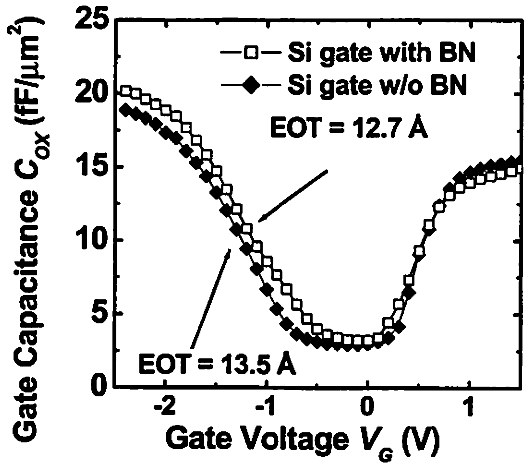


Figure 3.6 N-FETs' *C-V* characteristics with poly-Si gate and with or without the bottom nitride (BN). The device with the BN shows higher accumulation capacitance, i.e., thinner EOT.

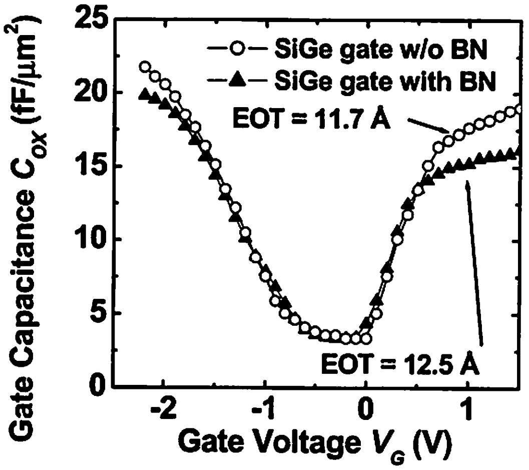


Figure 3.7 N-FETs' *C-V* characteristics with poly-SiGe gate and with or without the bottom nitride (BN). The device without the BN shows higher accumulation capacitance, i.e., thinner EOT.

is slightly more distortion of the bottom nitride curve, which is likely due to the fact that the bottom nitride interface typically has a higher interface trap density [3.9].

When a poly-SiGe gate is used, however, the effects of the bottom nitride seem to be reversed. As can be seen in Figure 3.7, the devices with the bottom nitride have thicker EOT than the ones without. And as in the case of the poly-Si gate, the devices with the bottom nitride show visible distortion in the *C-V* curve (V_G between −1 V to 0 V), which consistently shows that the bottom nitride may degrade the interface quality.

To elucidate these seemingly contradicting effects of the bottom nitride, high-resolution cross-sectional TEM analysis was performed on the different gate stack structures, and side-by-side comparisons were made between different gate materials. Figure 3.8 shows the two gate stacks with the bottom nitride layer after a full CMOS process, with a 950 °C 25 s RTA for dopant activation. A high-contrast interfacial layer between the HfO_2 and the Si substrate can be seen with both gate materials, while it appears to be slightly thinner for the poly-SiGe gate. The fairly low dielectric constant of the interfacial layer (estimated in Chapter 2) suggests that it is likely rich in oxygen, which promotes the formation of SiO_X. The poly-SiGe gate may reduce the amount of oxygen available at the substrate interface, resulting in a thinner interfacial layer. This interfacial layer reduces the overall EOT of the gate dielectric. The two devices shown in

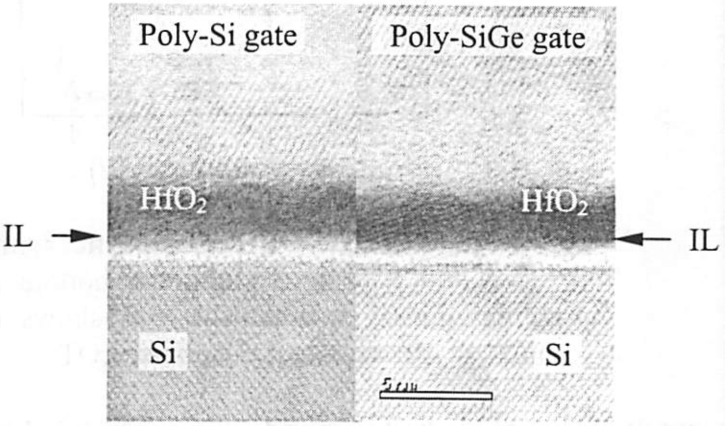


Figure 3.8 Cross-sectional TEM images of the gate stacks with poly-Si or poly-SiGe gate and with the bottom nitride. After a full CMOS process, including a 950°C 25 s dopant activation anneal, two gate stacks showed similar structures, with slightly thinner interfacial layer (IL) for the poly-SiGe gate stack.

Figure 3.8 have very similar EOT, so the larger difference in EOT seen in Figure 3.4 suggests that anneal at a higher temperature (1000°C 10 s) may lead to a larger difference in the interfacial layer thickness between the poly-Si and poly-SiGe gated devices.

The gate stacks without the bottom nitride are compared in Figure 3.9. The interfacial layer is still seen in the case of poly-Si gate, but it is essentially missing in the poly-SiGe gated device. The thinner EOT of the poly-SiGe gated devices can be explained by the absence of this interfacial layer, which has a lower dielectric constant therefore increases the overall EOT.

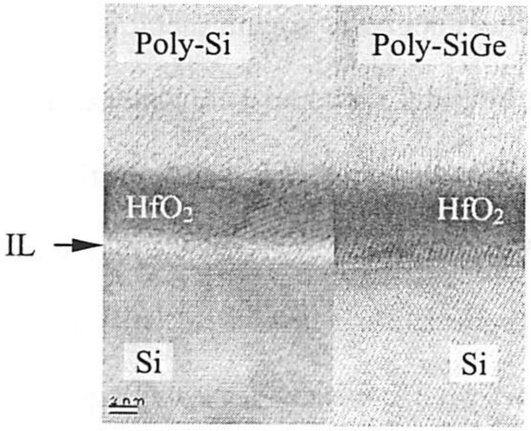


Figure 3.9 Cross-sectional TEM images of the gate stacks with poly-Si or poly-SiGe gate and without the bottom nitride. After a full CMOS process, including a 1000°C 10 s dopant activation anneal, a distinctive interfacial layer (IL) can be seen for the poly-Si gated stack, while it is missing for the poly-SiGe gate stack.

It is not clear yet why the interfacial layer was not formed when poly-SiGe gate is used. A possible reason may be that the poly-SiGe gate somehow reduces the amount of oxygen in the gate or in the gate dielectric, which can diffuse to the interface and form the interfacial layer during the subsequent high-temperature processes. With this assumption, a self-consistent explanation can be constructed to interpret the above electrical results. Both the bottom nitride and the poly-SiGe gate help reduce the final EOT by suppression of the interfacial layer forming during high temperature anneals. So when a poly-Si gate is used, the gate stack with the bottom nitride shows a thinner EOT after the 1000°C 10 s anneal (Figure 3.6), and similarly when there is no bottom nitride, the poly-SiGe gated devices showed thinner EOT (Figure 3.5). When the bottom nitride exists, the interfacial layer formation is already reduced, therefore the gate electrode does not make significant difference, and comparable EOTs resulted (Figure 3.5). When a poly-SiGe gate is used, it suppresses the interfacial layer formation, so whether the bottom nitride exists does not make a significant difference in this respect. However, the bottom nitride does physically introduce an additional dielectric layer (comparing (b) with (d) in Figure 3.3), therefore it increases the final EOT even if no additional interfacial reactions are involved. This explains the results in Figure 3.7. To prove or disprove the above argument, it is helpful to clarify the elemental compositions of the interfacial layer in each case. Specifically, the above explanation suggests that the oxygen concentration in the interfacial layer is higher when the bottom nitride does not exist, and that with the bottom nitride the interfacial layer composition is similar for poly-Si and poly-SiGe gated devices. In addition, the film composition should be relatively uniform across the whole gate dielectric layer when poly-SiGe gate is used without the bottom nitride. Accurate measurement of the thickness of the interfacial layer will also bring useful information, but it is difficult to define a clear boundary between the thin films. The mechanisms of the suppressed interfacial layer formation also need to be further investigated. It is proposed that the bottom nitride blocks the diffusion of oxygen, silicon and hafnium, therefore prevents excessive interfacial layer formation [3.7]. In the case of a poly-SiGe gate, it could be either due to some kind of oxygen gettering or diffusion barrier mechanism, or simply the faster deposition rate than poly-Si. For the same target gate thickness, the shorter deposition time may help reduce the amount of oxygen introduced during the LPCVD process, thus suppressing the interfacial layer growth in the subsequent steps. The effect of deposition time can be clarified by using a disilane poly-Si (Si_2H_6) gate, which has a comparable deposition rate to poly-SiGe. Better understanding of the interfacial layer formation process will provide guidance to balancing the tradeoff between EOT and interface quality, and optimizing the high-*k* dielectric stack.

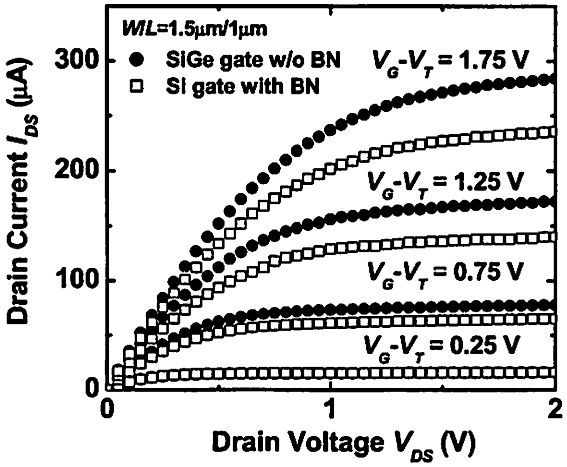


Figure 3.10 Transistor drive current I_{DS}-V_{DS} comparison between poly-SiGe gate without the BN and poly-Si gate with the BN, both of which had suppressed interface reaction during high temperature anneal. The higher drive current of the poly-SiGe gated devices is attributed to the thinner EOT and better interface quality.

The above results indicate that using either a poly-SiGe gate or a bottom nitride with HfO_2 can reduce the gate stack EOT after a high temperature anneal. To evaluate the transistor performance using these two techniques, I_{DS} – V_{DS} characteristics of n-FETs with these two types of gate stacks are compared in Figure 3.10. For the same gate overdrive, the SiGe gated non-bottom-nitride devices showed higher drive current than the devices with poly-Si gate and bottom nitride. This is due to the thinner inversion CET and higher electron mobility of the poly-SiGe gated devices. Bottom nitride can result in degraded interface quality, such as higher interface trap density and *C-V* hysteresis [3.9], therefore is a possible factor for degraded carrier mobility. On the other hand, as shown in Chapter 2, the carrier mobility without the bottom nitride is also significantly lower than the universal model, and this is still a major concern for alternative gate dielectrics.

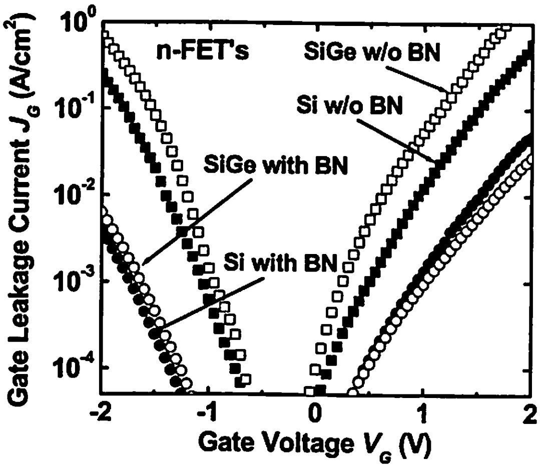


Figure 3.11 Gate leakage currents of n-FETs with poly-Si or SiGe gate, with or without the bottom nitride (BN). The bottom nitride significantly reduces the gate leakage current for both gate materials.

The n-FETs' gate leakage currents are shown in Figure 3.11. In both the inversion and the accumulation region, the two gate stacks with the bottom nitride showed more than an order of magnitude lower gate leakage than the gate stacks without the bottom nitride. The non-bottom-nitride devices with the poly-SiGe gate showed slightly higher leakage than the ones with a poly-Si gate mainly due to their thinner EOT. The strong correlation between the bottom nitride and lower gate leakage currents is also seen in the case of p-FETs (Figure 3.12). Both theoretical and experimental studies indicated that for

HfO_2 the tunneling barrier for holes is much larger than that for electrons [3.10][3.11], therefore, the p-FETs' gate leakage currents are lower than those of the n-FETs. For a given gate leakage current requirement, the EOT scaling of HfO_2 will therefore be limited by n-FETs.

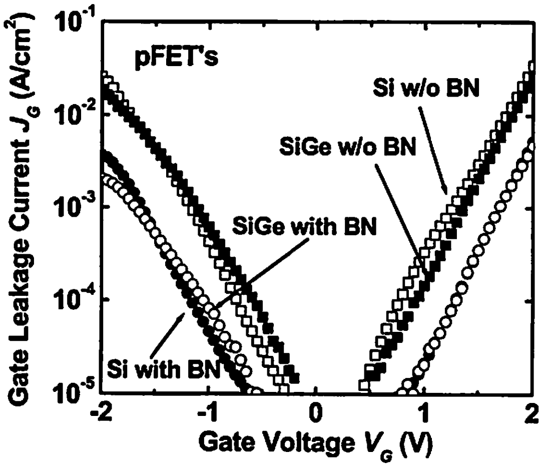


Figure 3.12 Gate leakage currents of p-FETs with poly-Si or SiGe gate, with or without the bottom nitride (BN). Significant reduction of the gate leakage current due to the bottom nitride is also observed.

Figure 3.13 plots the gate leakage current of n-FETs with HfO_2 gate dielectric at 1 V gate voltage for different EOT values. The HfO_2 data from the literature and this work form a trend line, which projects roughly two orders of 100× lower leakage than SiO_2 down to 5 Å EOT. A data point for a nitride/oxide stack is also shown for reference. In view of the low power application requirements in the International Technology Roadmap for Semiconductors, The HfO_2 gate dielectric may be feasible for a few technology generations.

The impact of the bottom nitride on gate dielectric breakdown voltage is shown in Figure 3.14. N-FETs of the same area were used for the test, in which the gate voltage was quickly ramped up from 0 V until hard breakdown occurred. The dielectric breakdown mechanism of HfO2 is not clear yet, and it remains to be seen if the dielectric breakdown of such ultra-thin films is driven by the electric field or gate voltage, an open question that similarly exists in the case of the very well studied ultra-thin SiO2 gate dielectric [3.13]. The fact that the gate dielectrics are bi-layer stacks (with the exception

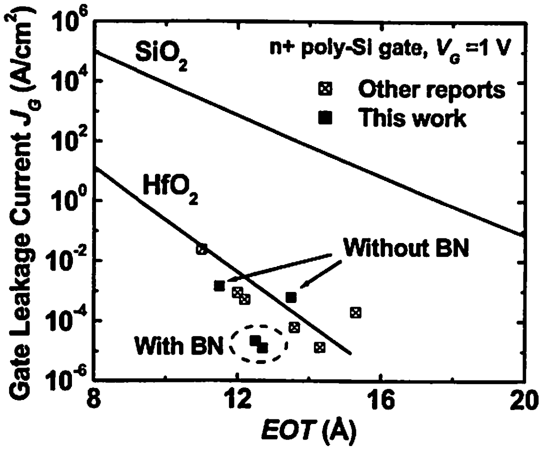


Figure 3.13 Gate leakage currents of HfO2 n-FETs at a fixed gate voltage for different EOT. The data (without bottom nitride) from the literature and this work form a consistent trend line. And the bottom nitride technique results in improvement over the trend line. Simulated SiO2 trend line is also shown for reference [3.12].

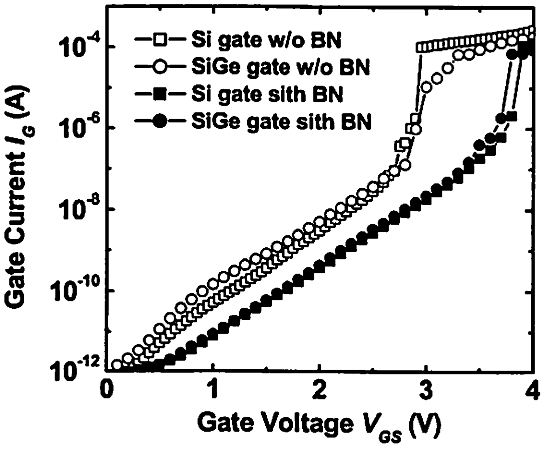


Figure 3.14 Gate current vs. gate voltage for the four gate stacks in a ramp breakdown test. The bottom nitride (BN) results in higher breakdown voltages for both gate materials.

of the poly-SiGe gated non-bottom-nitride devices) further complicates the breakdown issue. Without getting into the details of the breakdown mechanism, a comparison can still be made using the practical criterion of EOT. As the two devices with the bottom nitride have similar gate dielectric stacks and EOTs, it is relatively easy to interpret the results. The similar breakdown voltages suggest that the poly-SiGe gate does not cause significant change of gate dielectric reliability compared to the poly-Si gate. The poly-Si gated non-bottom-nitride device has a thicker EOT, therefore, it is clear that for poly-Si gated devices, the bottom nitride is advantageous in terms of both the breakdown voltages and the EOT. The bottom nitride resulted in higher breakdown voltage and thicker EOT for the poly-SiGe gated devices, therefore it is not clear if the bottom nitride is always beneficial in this case.

These encouraging results from the poly-SiGe gate may provide a near-term solution to the gate stack scaling issues, but are still insufficient to meet the long-term goals set by the ITRS. When the gate dielectric EOT is scaled below 10 Å, minimal gate depletion can be tolerated to achieve lower than 25% contribution to the EOT from the gate. Eventually, metallic gate materials will be needed if the CMOS scaling is to follow the ITRS.

3.4 References

[3.1] T.-J. King, J. R. Pfiester, J. D. Shott, J. P. McVittie, and K. C. Saraswat, "Polycrystalline-Si_xGe_{1-x}-gate CMOS technology", *International Electron Devices Meeting*, pp. 253-256, Dec. 1990.

[3.2] W.-C. Lee, T.-J. King, and C. Hu, "Optimized poly-Si_xGe_{1-x} gate technology for dual gate CMOS application", *Symposium on VLSI Technology*, pp. 190-191, June 1998.

[3.3] V. P. Lesnikova, A. S. Turtsevich, V. Y. Krasnitsky, V. A. Emelyanov, O .Y. Nalivaiko, S. V. Kravtsov, and T. V. Makarevich, "The structure, morphology and resistivity of i*n situ* phosphorus doped polysilicon films", *Thin Solid Films*, Vol. 247, pp. 156-161, 1994.

[3.4] Q. Xiang, J. Jeon, P. Sachdey, B. Yu, K. Saraswat, and M.-R. Lin, "Very high performance 40 nm CMOS with ultra-thin nitride/oxynitride stack gate dielectric and pre-deoped dual poly-Si gate electrodes", *International Electron Devices Meeting*, pp. 860-862, Dec. 2000.

[3.5] Y.-C. King, Ph.D. Dissertation, 1999.

[3.6] Y.-C. King, H. Fujioka, S. Kamobara, and C. Hu, "DC electrical oxide thickness model for quantization of the inversion layer in MOSFETs", Semiconductor Science and Technology, Vo. 13, No. 8, pp. 963-966, 1998.

[3.7] R. Choi, C. S. Kang, B. H. Lee, K. Onishi, R. Nieh, S. Gopalan, E. Dharmarajan, and J. C. Lee, "High-quality ultra-thin HfO_2 gate dielectric MOSFETs with TaN electrode and nitridation surface preparation", *Symposium on VLSI Technology*, pp. 15-16, June 2001.

[3.8] K. Yang, Y.-C. King, and C. Hu, "Quantum effects in oxide thickness determination from capacitance measurement", *Symposium On VLSI Technology*, pp. 77-78, June 1999.

[3.9] H.-J. Cho, C. S. Kang, K. Onishi, S. Gopalan, R. Nieh, R. Choi, E. Dharmarajan, and J. C. lee, "Novel nitrogen profile engineering for improved TaN/HfO_2/Si MOSFET performance", *International Electron Devices Meeting*, pp. 655-658, Dec. 2001.

[3.10] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices", *Journal of Vacuum Science and Technology*, B 18(3), pp. 1785-1791, May/Jun, 2000.

[3.11] W. Zhu, T. P. Ma, T. Tamagawa, Y. Di, J. Kim, R. Carruthers, M. Gibson, and T. Furukawa, "HfO_2 and HfAlO for CMOS: thermal stability and current transport", *International Electron Devices Meeting*, pp. 463-466, Dec. 2001.

[3.12] W.-C. Lee and C. Hu, "Modeling gate and substrate currents due to conduction and valence band electron and hole tunneling", *Symposium On VLSI Technology*, pp. 198-199, June 2000.

[3.13] P. E. Nicollian, W. R. Hunter, and J. Hu, "Experimental evidence for voltage driven breakdown models in ultra-thin gate oxides", *Proceedings of the International Reliability Physics Symposium*, pp. 7-15, Apr. 2000.

Chapter 4

Metal gate technology for CMOS

4.1 Introduction

It is clear that using high-*k* gate dielectrics alone will not solve the problem of CMOS gate stack scaling. For each technology generation, the power supply voltage is reduced, and the physical gate length is much shorter than the metal line half pitch. The negative impact of the poly-Si gate depletion on short-channel behavior and gate overdrive consequently becomes an increasingly serious problem. The poly-SiGe gate may alleviate the problem temporarily, but with continued device scaling, the stringent requirements of ITRS will necessitate the use of metal gate electrodes in the CMOS structure.

The metal gate is actually an old idea that predated the poly-Si gate technology. In fact, the first commercial MOS transistor circuits (by Fairchild and RCA separately in 1964) were fabricated using an aluminum gate electrode, which at that time was a simple and economical choice [4.1]. The poly-Si gate technology was developed in the 70's, and became the standard gate technology because of a number of advantages over the metal gate. To revive the idea of using metal gate, it is instructive to review why the poly-Si gate replaced the original metal gate technology.

Aluminum has a fairly low melting point, so it cannot undergo any high-temperature (>450 °C) processes. In the original aluminum gate MOSFET process (Figure 4.1), the aluminum gate was deposited and patterned after the formation of the source and drain regions. To allow for possible misalignments in this non-self-alignment process, sufficient overhang of the gate outside the contact hole is necessary. As the transistor dimensions shrank and the integration scale became larger, this overhang area caused a few problems. Firstly, it incurs considerable parasitic capacitance, and degrades the transistor switching speed. Secondly, it wastes a lot of area that can otherwise be availed to increase transistor layout density. In addition, using an Al gate electrode for p-MOSFETs results in very high threshold voltages, therefore it is not good for CMOS technology, where symmetrical low threshold voltages for n- and p-FETs are desirable.

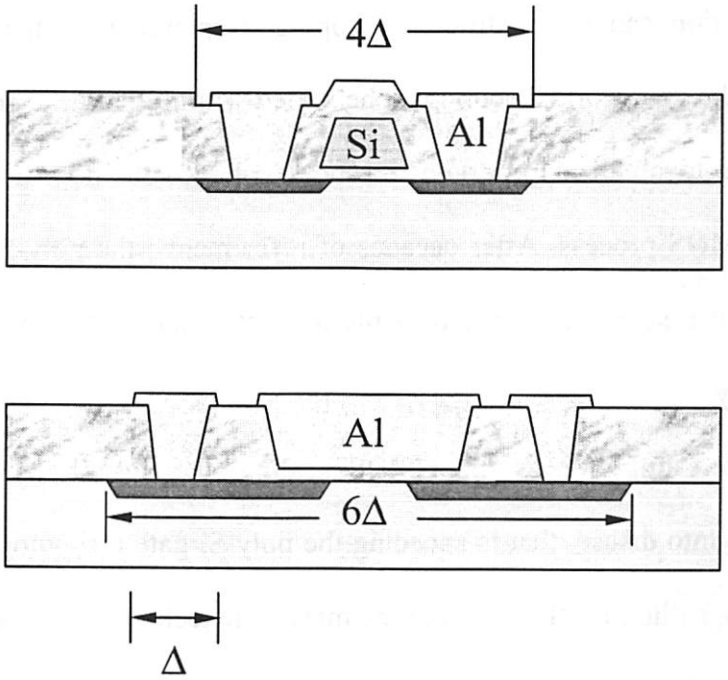


Figure 4.1 Cross-sectional view of (upper) a self-aligned poly-Si gate and (lower) a non-self-aligned metal gate MOSFET structure. The non-self-aligned process requires larger gate-to-source/drain overlap to allow for possible misalignment, therefore it results in larger layout area as well as higher parasitic capacitance. (After [4.2])

When poly-Si gate was introduced, all the above problems were solved. The exceptional thermal stability of poly-Si gate enables the self-aligned source/drain formation, so that the gate overhang area could be minimized, resulting in improved performance as well as smaller transistor footprint. Another unique feature of the poly-Si gate is that the gate work-function can be modified by doping. Depending on the type and dose of the dopants, the gate work-function can be varied within the full bandgap range. This makes it possible to obtain appropriate threshold voltages for both n- and p-MOSFETs in a simple CMOS process. After decades of refinements, the poly-Si gate technology is so mature that it seems almost impossible for metal gate electrodes to return to the CMOS technology.

Interestingly, it is the CMOS scaling, the very force that drove metal gate electrodes into disuse, that is speeding the poly-Si gate technology toward the end of its usefulness, calling for the return of the metal gate technology. In addition to the mounting problems of gate depletion and boron penetration, which have been discussed in preceding chapters, it is also found that poly-Si gate is not very stable with some high-*k* gate dielectric candidate materials. Theoretically, HfO_2 is thermally stable with Si [4.3]. But it was reported that for the same deposition conditions, a single-layer HfO_2 gate dielectric (without special substrate pre-treatment) with a Pt gate consistently resulted in ~4 Å thinner EOT than those with poly-Si gate, and this is possibly due to the poly-Si deposition and dopant activation annealing, which introduced a higher thermal budget [4.4]. So metal gate electrodes may offer the advantage of reducing the increase of the EOT during those high-temperature processes. Low resistance is another unique property of metal gate electrodes. A poly-Si gate with aggressively scaled thickness will have high gate resistance, which results in high gate electrode RC delay and affects RF performances of the transistors [4.5]. In addition, when the gate dielectric is physically thin enough, the remote charge scattering due to poly-Si gate may be an additional mechanism that can reduce the channel carrier mobility, while this does not happen with metal gate electrodes [4.6]. All these promising potentials of the metal gate technology suggest that this old idea be re-examined for the CMOS device in the nanometer regime.

4.2 Effects of metal gate on gate dielectric scaling

A metal gate can completely eliminate the voltage drop within the gate electrode, therefore it increases the actual gate overdrive for a given applied gate voltage. Although this leads to higher drain current, it also brings the side effect of higher gate leakage current. Qualitatively, for a given gate voltage, the voltage drop across the gate dielectric is higher for metal gate, therefore the gate tunneling current is also higher compared with the case of poly-Si gate. Consequently, for the same power supply voltage, metal gated devices may require a thicker gate dielectric to meet a given gate leakage specification, which essentially imposes an additional constraint on the scalability of the gate dielectric. In this section, we try to quantitatively estimate this effect to evaluate the net benefit of using metal gate.

Consider the gate tunneling current of a p-channel MOSFET biased in the inversion regime as shown in Figure 4.2. At low gate bias, the direct tunneling gate leakage is dominated by the valence band hole tunneling (J_{HVB}), and the conduction band electron tunneling (J_{ECB}) can be neglected. We will compare the gate leakage currents of this device when a p^{+}-poly-Si gate or a metal gate is used. The metal gate is assumed to have the same work-function as the p$^+$-poly-Si gate, so as to avoid the complication that the gate work-function difference also affects the comparison of gate leakage currents.

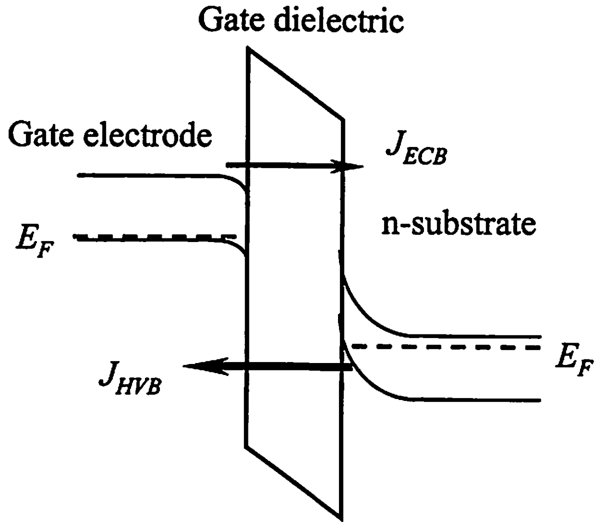
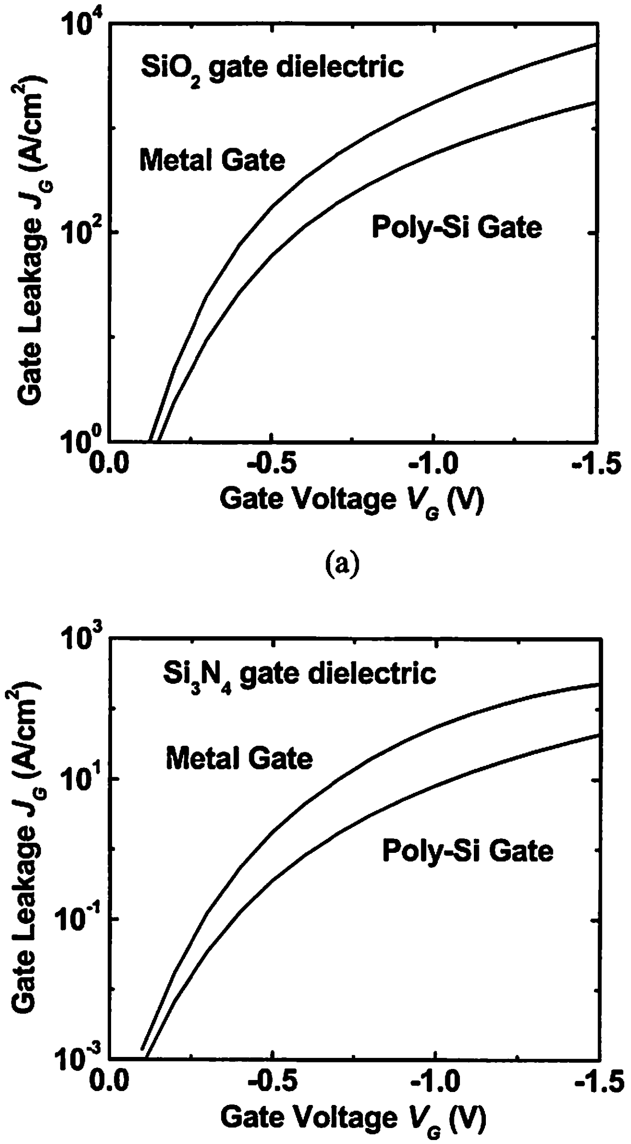


Figure 4.2 Energy band diagram of a p-MOSFET biased in the inversion regime. The dominant component of the gate tunneling current is the valence band hole tunneling (J_{HVB}), and the conduction band electron tunneling (J_{ECB}) is negligible. (After [4.8])

With this assumption, for both gate materials the gate Fermi level roughly coincides with the Si valence band (Figure 4.2). A semi-empirical model of direct tunneling gate current is used for the simulation [4.7]. This model applies to not only SiO_2, but also other gate dielectrics, e.g. silicon nitride, when appropriate model parameters are used [4.8]. In all the following $J_G - V_G$ simulations, the poly-Si gate doping was assumed to be $5\times10^{19}/cm^3$, and the substrate doping $3\times10^{17}/cm^3$.

Figure 4.3-(a) shows the simulated J_G - V_G curves for 10 Å SiO_2 with the two different gate electrodes. As expected, a metal gate resulted in higher (still <10×) gate leakage current than a poly-Si gate with the same gate work-function. As the poly-Si gate depletion effect is more serious for gate dielectrics with thinner EOT, the difference in J_G will become larger for thinner SiO_2.



(a)

(b)

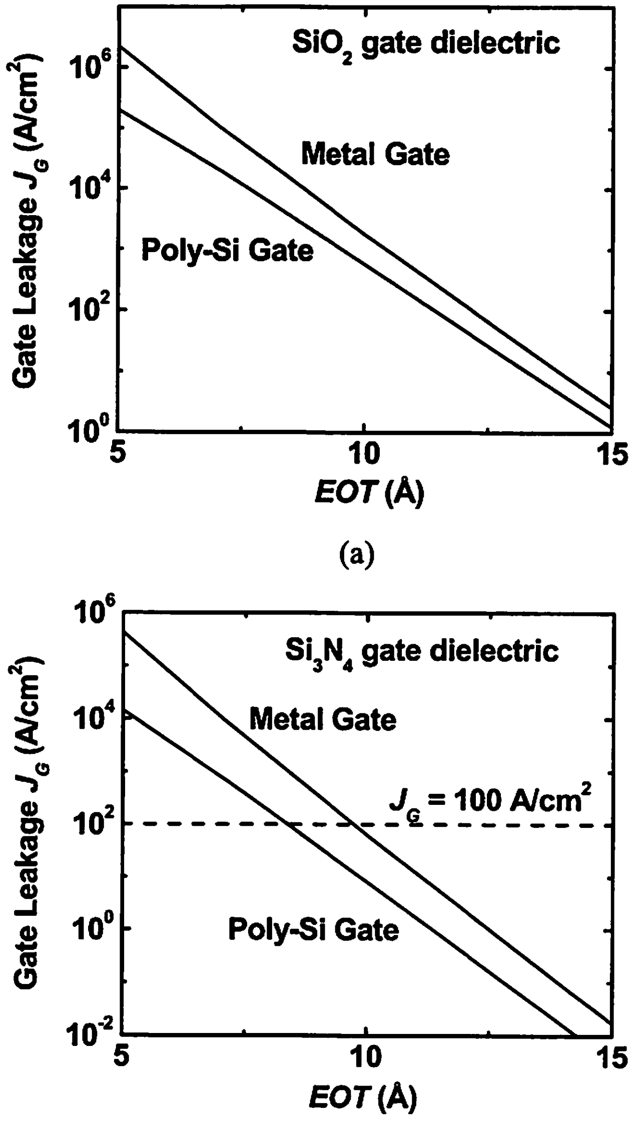
Figure 4.3 Simulated gate leakage current for (a) 10 Å SiO_2 and (b) 10 Å EOT silicon nitride gate dielectrics with poly-Si or metal gate electrodes. Increase in gate leakage due to metal gate is more pronounced for silicon nitride.

When the same comparison is made for a silicon nitride gate dielectric with a 10 Å EOT, the difference between the metal gate and the poly-Si gate becomes visibly larger (Figure 4.3-(b)). A simple physical explanation for this phenomenon can be found in the tunneling current model [4.7]. When the poly-Si gate depletion is eliminated, the major change is the increased voltage drop (V_{OX}) and electrical field (E_{OX}) in the gate dielectric region. Although in the leakage current equation there are multiplicative factors that depend on V_{OX} and E_{OX}, the major dependence comes from the exponential term, i.e., the tunneling probability T.

$$T = \exp\left\{ -\frac{4\sqrt{2m\phi_b^3}}{3\hbar E_{OX} q}\left[1-\left(1-\frac{V_{OX}q}{\phi_b}\right)^{3/2}\right]\right\} \tag{4.1}$$

Since V_{OX} and E_{OX} are linearly correlated, the change in E_{OX} in the denominator of the exponent is always partly cancelled by the corresponding change of the V_{OX} term, and this is the reason why despite the fairly large increase in E_{OX} due to the use of metal gate, the gate leakage current only increased moderately. In the above equation, the term $(V_{OX}\text{q})/\phi_b$ determines how sensitive T is to the change of V_{OX}. In the extreme case where $(V_{OX}\text{q}) << \phi_b$ (large barrier height), simple Taylor expansion shows that T has virtually no V_{OX} dependence, and on the other hand if $(V_{OX}\text{q}) \sim \phi_b$ (small barrier height), T becomes more sensitive to the change in V_{OX}. In the above simulations, V_{OX} is about 1 V. Note that the hole barrier height ϕ_b is 4.5 eV for SiO_2, but only 1.9 eV for silicon nitride [4.8][4.7], equation 4.1 explains why the two gate electrodes showed a larger difference on a silicon nitride gate dielectric.

The impact of a metal gate on gate dielectric scaling is illustrated in Figure 4.4. Using the same device parameters as in the previous simulations, the gate leakage vs. EOT trends are simulated for SiO_2 (Figure 4.4-(a)) and silicon nitride (Figure 4.4-(b)) for a fixed gate voltage of -1 V. It can be seen that for the same EOT, the metal gate causes a larger increase in gate leakage for silicon nitride than for SiO_2. In addition, the difference between the two electrodes becomes larger for thinner EOT. These results all agree with the previous physical analysis. The upward shift of the metal gate curve may be a significant limiting factor of the gate dielectric scaling. For example, to ensure $J_G \leq 10^2 \mathrm{A/cm}^2$ at $V_G = -1$ V, a metal gate roughly requires a 1.3 Å (EOT) thicker silicon nitride gate dielectric than a poly-Si gate (Figure 4.4 (b)). Using the method illustrated in Figure



(a)

(b)

Figure 4.4 Simulated gate leakage J_G vs. EOT trend with metal or poly-Si gate for (a) SiO_2 and (b) silicon nitride gate dielectrics. Due to the smaller hole barrier height, silicon nitride shows a larger increase in gate leakage due to the metal gate. For a fixed gate leakage requirement, larger EOT is needed for metal gate.

4.4 (b), the EOT required to meet a fixed gate leakage current specification (for example, J_G=100 A/cm^2 at V_G= -1 V) is extracted for poly-Si gates with different doping, and compared to a metal gate (Figure 4.5). The EOT for the poly-Si gate gets closer to that of the metal gate with increased gate doping, which results in a reduced gate depletion effect. It should be mentioned that different poly-Si gate doping intrinsically results in different threshold voltage. And the metal gate work-function was chosen to match the lowest threshold voltage among the poly-Si gate cases (which causes the gate leakage to increase slightly), so as not to favor the metal gate. Considering the above effect, the choice of poly-Si or metal gate electrodes involves the tradeoff between drive current and power-consumption. In the off-state (low gate bias), the poly-Si gate depletion effect is negligible, so the same gate dielectric EOT will be needed for the poly-Si or the metal gate to achieve the same off-state leakage current, assuming everything else is the same for the two gate technologies. In the on-state, however, the metal gated devices will have higher gate leakage current for the same EOT, along with the higher drive current than

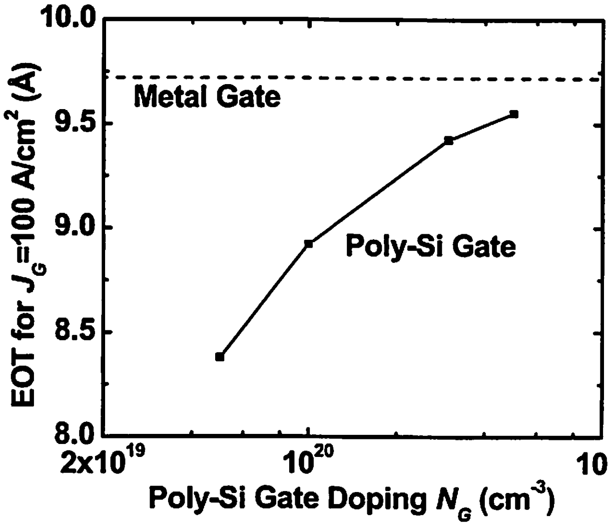


Figure 4.5 The EOT required to satisfy J_G = 100 A/cm^2 at V_G= -1 V for different poly-Si gate doping. The varied gate doping does not apply to metal gate, which is shown for comparison with the poly-Si gate.

that of the poly-Si gated devices. If the on-state gate leakage is set as the criterion, a slightly thicker gate dielectric EOT is needed for the metal gate than for the poly-Si gate. This might slightly degrade the off-state current compared to the poly-Si gate devices, and consequently affects the minimum scalable channel length of the transistor. For bulk MOSFETs, the minimum gate length is given by an empirical rule [4.9]:

$$L_{MIN} = A[t_j t_{OX}(w_S + w_D)^2]^{1/3} \qquad (4.2)$$

where t_j, w_S and w_D are determined by the doping profiles, and t_{OX} is generalized to be EOT. The 1/3 power weakens the effect of the thicker EOT resulted from using metal gate, e.g., only ~5% larger L_{MIN} with ~15% thicker EOT in the above silicon nitride example. In the case of fully depleted SOI, the scale length is [4.10]

$$\lambda = t_{Si}\sqrt{2\left(1+\frac{\varepsilon_{Si}t_{OX}}{4\varepsilon_{OX}t_{Si}}\right)} \qquad (4.3)$$

Similarly t_{OX} can be generalized to EOT. In aggressively scaled devices, the Si body thickness t_{Si} is roughly 10× larger than the EOT, therefore the thicker EOT due to the metal gate will have even less effect on the device scaling. Therefore, the use of metal gate will not likely be a serious concern in future device scaling.

The thicker EOT for a metal gate, on the other hand, does not necessarily result in worse drive current for the metal gated devices. Depending on the active dopant concentration of the poly-Si gate, a metal gate may still offer performance advantages. Following the results in Figure 4.5, the simulated *C-V* curves of the metal gated and poly-Si gated devices are compared on the basis of the same gate leakage current, i.e., 10^2 A/cm^2 at V_G = -1 V (Figure 4.6). The quantum *C-V* simulation shows that compared to the state-of-the-art poly-Si gate doping levels, the gate capacitance and inversion charge

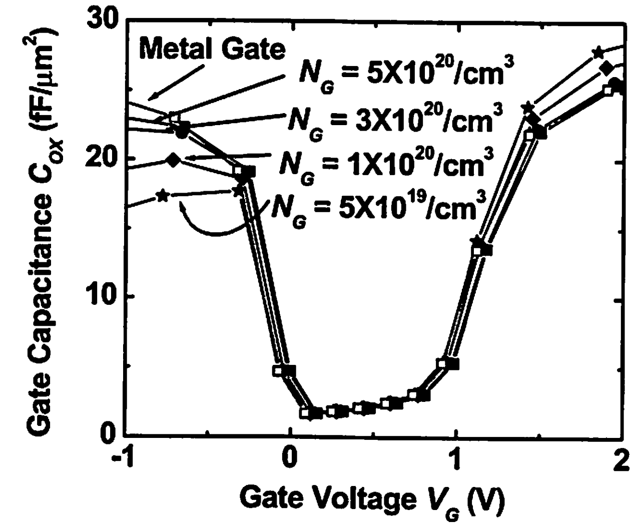


Figure 4.6 Simulated *C-V* characteristics of p-FETs with silicon nitride gate dielectric and metal gate (EOT= 9.7 Å), or poly-Si gate with various EOTs and active gate dopant concentrations. All these p-FETs have the same gate leakage current at V_G = -1 V. Although the metal gate requires thicker gate dielectric EOT to achieve the same gate leakage, it still offers higher inversion charge density unless the poly-Si gate has very high active dopant concentration.

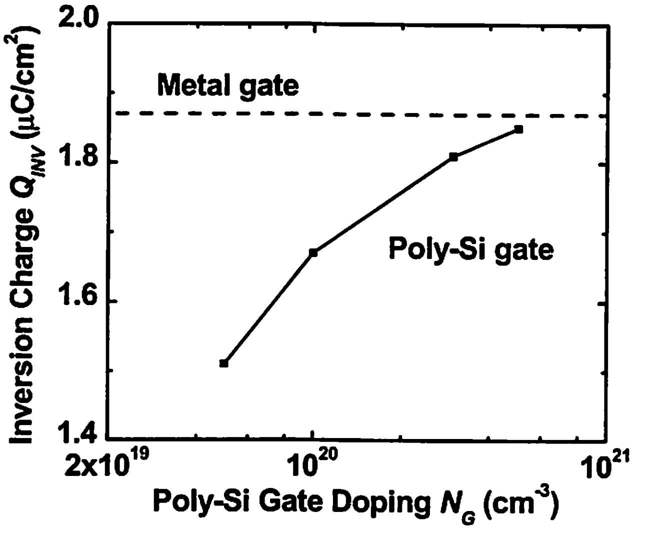


Figure 4.7 Simulated inversion charge density for p-FETs with silicon nitride gate dielectric and poly-Si gate (varied EOT) or metal gate (EOT= 9.7 Å). All these p-FETs have the same gate leakage current at V_G= -1 V. Despite the thicker EOT, metal gate offers higher inversion charge than poly-Si gate with commonly attainable doping level. Q_{INV} has been corrected for the small differences in threshold voltages.

density are still higher for the metal gated devices despite their thicker EOT. In this example, ~5×10^{20} cm^{-3} poly-Si gate doping is needed to achieve a similar inversion charge and on-state gate leakage as the metal gated devices. In general, the poly-Si gate depletion effect becomes pronounced only under high-gate bias, which may be leveraged to control the short-channel effects and the on-state leakage currents. However, the absolute drive current requirements still necessitate very high gate doping, as demonstrated in Figure 4.7. If we compared the ratio Q_{INV}/L_{MIN}, which is roughly an indicator of the MOSFET drive current, then in the above example, poly-Si gate doping of ~$5\text{-}6\times10^{20}$ cm^{-3} will be needed to match the metal gate for the same on-state gate leakage. In summary, for low-power applications, there might be a design window where a specific poly-Si gate doping results in an acceptable power-performance tradeoff. But for high-performance applications, where the gate leakage current requirement is relatively relaxed, the use of metal gate is definitely attractive and beneficial.

For other gate dielectrics, the above behaviors may quantitatively vary, depending on the gate dielectric properties, such as the barrier height, effective mass, metal gate work-function, etc. The effects on n-MOSFETs can be similarly understood. From these simulations, it can be seen that for high-*k* gate dielectrics, which have smaller band gap and typically lower barrier for both electrons and holes, the effects of a metal gate on scaling should be carefully taken into account in the design of gate stack. And the differentiation between high-performance and low-power applications is also necessary.

4.3 Requirements for metal gate materials

Obviously a practical metal gate technology will have to offer some features of the poly-Si gate in order to be integrated in existing CMOS technology. The necessary gate material properties also partly depend on the choice of CMOS process integration. The common requirements are suitable work-function and CMOS compatibility.

1). Appropriate work-function for CMOS applications

For bulk n-MOSFETs, the threshold voltage is determined by

$$V_T = V_{FB} + 2\varphi_B + \gamma_S\sqrt{2\varphi_B} \qquad (4.4)$$

with $\gamma_S \equiv \sqrt{2\varepsilon_0\varepsilon_{Si}qN_{SUB}}/C_{OX}$, $\varphi_B = (k_BT/q)\ln(N_{SUB}/n_i)$, and

$$\begin{aligned} V_{FB} &= \phi_M - \phi_S = (\phi_M - \phi_i) + (\phi_i - \phi_S) \\ &= \phi_M - (\phi_i + \varphi_B) \end{aligned} \qquad (4.5)$$

where ϕ_i is the midgap level of Si. For short-channel devices, the channel doping is typically in the range of 10^{17}-10^{18} cm^{-3}, so $2\varphi_B$ will be close to 1 V. With sub-20 Å EOT gate dielectric, γ_S is about ~0.3-0.4. To obtain a reasonably low V_T, V_{FB} $+2\varphi_B$ needs to be ~ 0 V. So according to (4.5), the gate work-function $\phi_M \approx \phi_i - \varphi_B$~4.0 eV, i.e., very close to the Si conduction band E_C. Similarly, the gate work-function of bulk p-MOSFETs need to be close to ~5.1 eV, close to the Si valence band E_V. N$^+$ and p$^+$ doped poly-Si gate have the appropriate work-functions to meet the above requirements. For metals electrodes, this criterion considerably narrows the range of choices. Table 4.1 lists some metals that have suitable work-functions for bulk Si CMOS [4.11]. It should be noted that unlike poly-Si, which shows consistent chemical properties that are not very sensitive to doping, there is a strong correlation between the chemical properties of metals (electronegativity scale) and their work-functions [4.14]. Lower work-function means that the electrons are relatively weakly bound, so the metals with lower work-function, such as Ti, Ta, are more reactive in high temperature processes. Consequently, their stability can be a problem for gate electrode applications. On the other hand, the metals with higher work-functions tend to be less reactive, so etching in some cases is difficult. This is an extra constraint for using metal gate materials.

Table 4.1 Some elemental metals with work-functions that may be suitable for bulk-Si CMOS gate electrode applications.

n-MOSFET		p-MOSFET	
Metal	ϕ_M (eV)	Metal	ϕ_M (eV)
Hf	3.9	Mo*	4.95 (110)
Zr	4.05	Co	5.0
Ta*	4.25	Pd*	5.12
Nb*	4.3	Ni*	5.15
Ti	4.33	Ir*	5.27
Zn*	4.33	Pt	5.65

*Varied work-functions reported, depending on orientation.

In recent years, some novel transistor structures, such as the FinFET and ultra-thin body (UTB) transistors, have been proposed to address the scaling issues in the nanometer regime [4.12][4.13]. In these extremely small devices, dopant fluctuation in the channel can cause significant variation in device characteristics. So in these device structures, the channel is very lightly doped, and the threshold voltage is mainly controlled by the gate work-function. It was found that gate work-functions of 0.2 eV above/below the Si midgap are desirable for n-/p-channel devices [4.15]. In this case, a number of silicides and some metals with midgap work-function are more suitable.

2). Compatibility with CMOS process

This requirement is related to a broad range of issues, such as thin film deposition, cleaning, patterning, and thermal stability, etc. To integrate metal gate electrodes into a CMOS manufacturing process, it is necessary to have the metal films deposited using practical methods, such as CVD, as opposed to some methods that are only possible in a research environment. A conventional patterning technique, e.g., reactive ion etching (RIE) of the metal gate is also required. Chemical-mechanical polishing (CMP) may be used depending on the integration approach. Sufficient thermal stability of a metal gate electrode makes it compatible with self-aligned source/drain process, so that the drawbacks of an Al gate can be overcome. These compatibility issues are closely related to the CMOS process integration approach to be used. More discussions on this point will be presented later in this chapter.

As n- and p-MOSFETs need different gate work-functions, metal gated CMOS will likely require two different metals. For p-MOSFETs, there are very limited choices of gate metals that can be easily patterned, while this problem is not as serious for n-MOSFETs. The next section will discuss a high work-function metal that showed a number of attractive properties as a gate electrode for p-MOSFETs.

4.4 Mo gate for alternative gate dielectrics

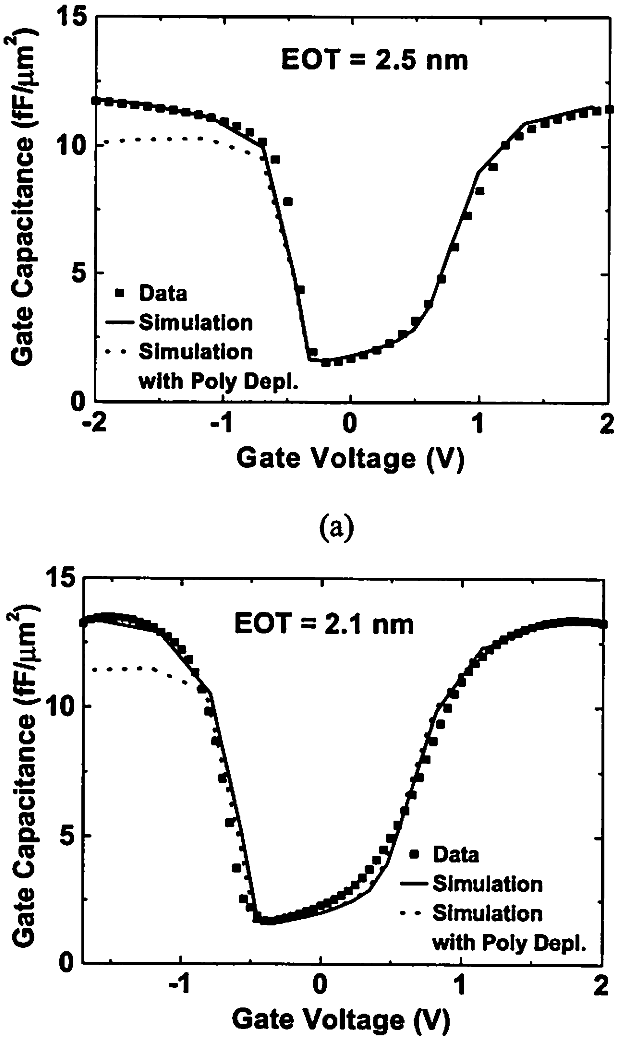
Mo is a refractory metal that exhibits a range of work-functions depending on the crystal orientation. For (110)-Mo, the reported work-function is 4.95 eV, close to the desirable value for a bulk-Si p-MOSFETs. The high melting point (2623 °C) of Mo makes it suitable for a gate-first, i.e., self-aligned CMOS process. Mo can be deposited using several methods that are commonly used in IC fabrication, such as sputtering deposition, CVD and evaporation. Unlike many metals that have high work-functions, Mo can be etched using several wet etch or RIE methods based on different chemistry. Mo and a few other refractory metals were considered as possible gate electrode materials in the very early stage of the silicon IC technology, and MOSFETs with Mo gates and thermally growth SiO_2 gate dielectrics were demonstrated in a self-aligned process [4.16]. However, the drastic changes in CMOS process technology over the more than thirty years necessitate a re-evaluation of Mo as a gate material. Given the process complexity of metal gate technology and that its benefit becomes significant only when a gate dielectric with very thin EOT is used, a metal gate is likely to be implemented with alternative gate dielectrics. Therefore, it is important to investigate the metal gate materials together with those promising high-*k* gate dielectric candidates. For this reason, before testing Mo in a full CMOS process, we investigated the feasibility of Mo as a gate electrode material for p-MOSFETs.

4.4.1 Mo gated p-FET process

A p-FET process with LOCOS isolation was used for this experiment. After LOCOS processing, V_T-adjust implant and formation of sacrificial oxide, the wafers were sent to two collaborators to receive gate dielectric deposition. JVD silicon nitride gate dielectric was deposited at Yale Univ. (Prof. T. P. Ma's group), and RT-CVD ZrO_2 and Zr-silicate gate dielectrics were deposited at UT Austin (Prof. D.-L. Kwong's group). The wafers were then returned to UC Berkeley for the rest of the p-FET process. 650 Å Mo was deposited by sputtering, followed by an anneal in Ar at 700°C for 10 min. A TiN barrier layer of 160 Å was deposited on top of the Mo by sputtering, and 1000 Å *in situ* n^+ doped LPCVD poly-Si was deposited to cap the gate stack. After I-line lithography, Cl_2-HBr based RIE was used to etch poly-Si and TiN, and CF_4 RIE was used to etch the Mo. Source and drain regions were formed by boron implantation with dose of 3×10^{15} cm^{-2}, and implantation energy of 7 keV. Dopant activation anneal conditions were chosen based on the known thermal stability of the gate dielectrics. A two-step RTA (900°C 10s + 1050°C 5s) was used for the wafer with JVD silicon nitride. A conservative furnace anneal of 800°C 30 min was used for the wafers with ZrO_2 and Zr-silicate gate dielectrics, which are know to degrade upon high-temperature (~950°C) anneals. The fabrication process was finished with LTO passivation, metallization and forming gas anneal (400°C 30 min).

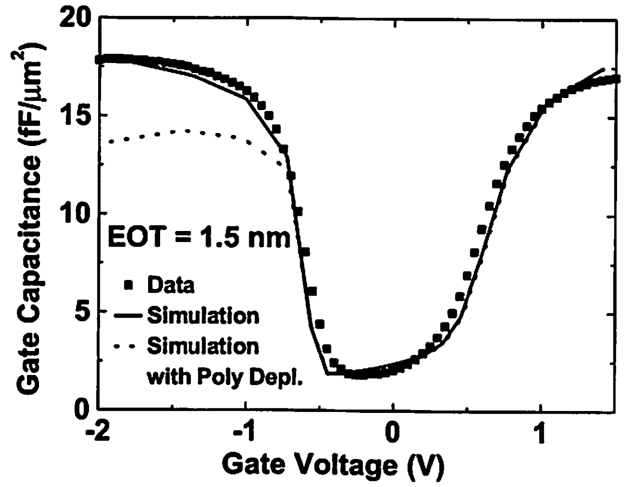
4.4.2 Mo gated p-FET characteristics

The p-FETs' *C-V* characteristics were measured using an HP-4282 LCR meter, and fitted with the quantum *C-V* simulator to extract the device parameters. Shown in Figure 4.8 (a)-(c) are the measured data (symbols), simulated metal gate *C-V* (solid lines) and simulated poly-Si gate *C-V* (dashed lines) for the three different gate dielectrics. The measured data are generally in good agreement with the simulated metal gate *C-V* curves, indicating good dielectric film quality with the Mo gate. In all three cases, the inversion capacitances are the same as the accumulation capacitances, which confirms that the gate depletion effect is completely eliminated by using the Mo gate. Compared to the simulated *C-V*s with a poly-Si gate, for which a reasonably high active gate dopant concentration (1×10^{20} cm^{-3}) is assumed, the Mo gate still resulted in more than a 20% increase in gate capacitance in the strong inversion region. The measured data also showed that for the same gate dopant concentration, the benefit of using a metal gate becomes more significant with thinner gate dielectric EOT, which is consistent with the theoretical analysis in Chapter 3.



(a)

(b)



(c)

Figure 4.8 P-FETs' *C-V* data of (a) ZrO_2, (b) Zr-silicate and (c) JVD silicon nitride gate dielectrics. Measured data (square symbols) shows good agreement with the quantum *C-V* simulation (solid lines). The benefit of higher inversion capacitance due to Mo gate is more pronounced for gate dielectrics with thinner EOT.

The effective work-function of the Mo gate can be extracted by fitting the measured *C-V* data with quantum *C-V* simulation [4.17]. As shown in Figure 4.9, the extracted work-function values depend on the underlying gate dielectric material. A

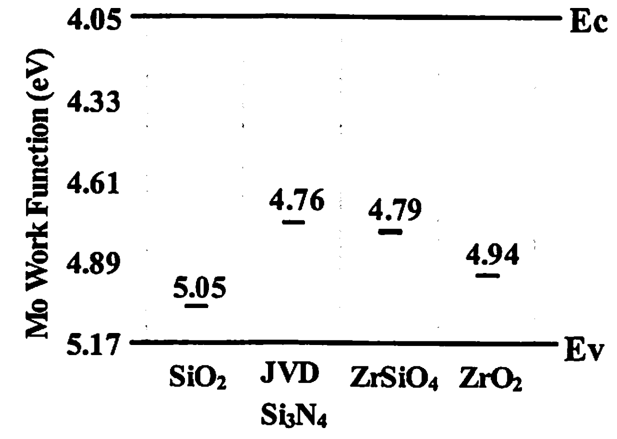


Figure 4.9 The effective work-functions of the Mo gate on different gate dielectrics extracted from p-FETs *C-V* measurements using quantum simulation. The work-function values of Mo are all within the lower half of the Si bandgap, and in some cases can potentially be used for bulk-Si p-FETs.

theoretical model explains this dependence in terms of the Fermi level pinning at the interface between the gate and the gate dielectric [4.18]. Due to the screening effects of the interfacial dipoles at the dielectric interface, the actual work-function that determines the device characteristics, i.e., φ_m in equation (4.5), can be considerably different from the metal work-function commonly measured by electron emission into vacuum. Therefore, in order to obtain optimal threshold voltages of the transistors, different high-*k* gate dielectric may have different requirements for metal gate materials as well as process integration. Similar to the thermal stability issue, the above effect also implies the results of metal gate electrode tend to be gate-dielectric specific, therefore they should not be inappropriately generalized.

The cross-sectional TEM analysis (Figure 4.10) reveals information about the film structures and interfaces of the meal gate stacks. The columnar grain structures of the Mo film can be seen in the lower resolution image (left), and the crystalline orientation was confirmed to be (110) by X-ray diffraction. This explains the high work-

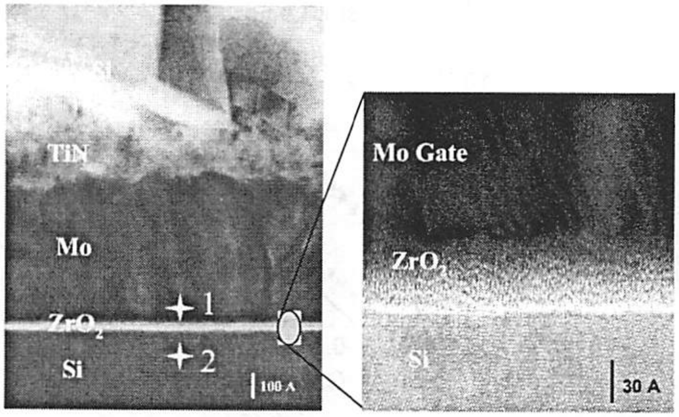


Figure 4.10 Cross-sectional TEM image (left) of a MOS gate stack showing ZrO_2 gate dielectric, Mo gate, TiN barrier layer and the poly-Si cap on top. The columnar morphology of the Mo gate is visible. The high-resolution view (right) near the ZrO_2laye r shows smooth and uniform interfaces of the gate dielectric. EDS analysis was performed in sites 1 and 2, close to the gate dielectric layer.

function values observed in the *C-V* measurements. The high-resolution TEM image (right) shows that the interfaces of the gate stack are uniform and smooth. Due to the low thermal budget for dopant activation anneal, the ZrO_2 gate dielectric remained amorphous after the full MOSFET processing, and the bottom interfacial layer is also very thin. Both phenomena are helpful for reducing the gate leakage and EOT. Local energy dispersive spectroscopy (EDS) analysis was also performed to analyze the elemental composition of the gate and the substrate (sites 1 and 2 in Figure 4.10) in search for possible unwanted interdiffusion or contamination. It was found that within the resolution limit of the analysis (~0.1%), there is no detectable Mo in the substrate, and insignificant amount of Si in the Mo gate. This is a proof of the good thermal stability of such a Mo-gated stack.

The gate leakage currents of the Mo gated p-FETs are plotted in Figure 4.11. These alternative gate dielectrics are known to have ~10^2x lower gate leakage than SiO_2 with comparable EOT, so the simulated SiO_2 gate leakages are also shown in the figure

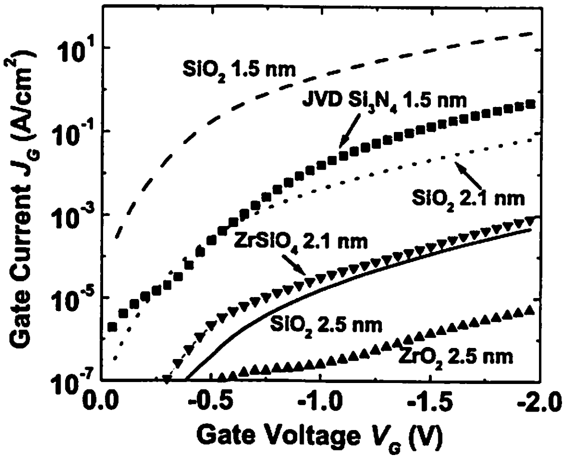


Figure 4.11 The gate leakage currents of Mo gated p-FETs with different gate dielectrics. Compared to simulation results for SiO_2 with the same EOTs, all these alternative gate dielectrics showed low gate leakage, indicating that the low gate leakage benefit can be retained in the Mo gate MOSFET process.

for reference. It can be seen that these non-SiO_2 gate dielectrics also showed ~10^2×reduction of the gate leakage current with Mo gate electrode, and this is indicates that the Mo gate process did not cause significant damage to the gate dielectrics.

The I_{DS}-V_{DS} and I_{DS}-V_{GS} characteristics of the long channel Mo gated p-FETs are shown in Figures 4.12-4.14. These well-behaved transistor characteristics indicate that

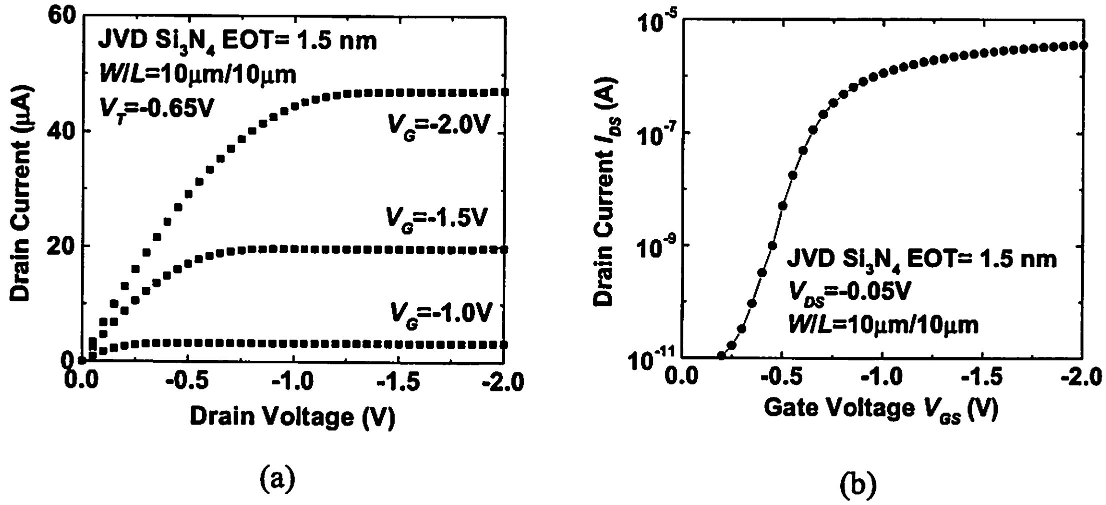


Figure 4.12 The transistor characteristics (a) I_{DS}-V_{DS} and (b) I_{DS}-V_{GS} of p-FETs with Mo gate and JVD silicon nitride gate dielectric.

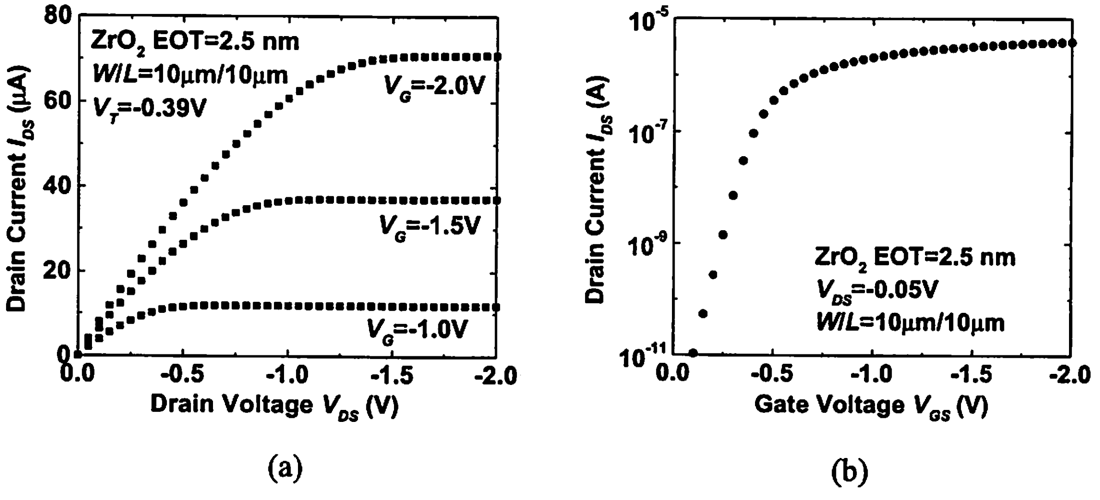


Figure 4.13 The transistor characteristics (a) I_{DS}-V_{DS} and (b) I_{DS}-V_{GS} of p-FETs with Mo gate and ZrO_2 gate dielectric.

the Mo gate can be used in a self-aligned MOSFET process under normal annealing conditions. The threshold voltages of these p-FETs are not well controlled in this lot, and further improvement can be made by better control of the gate dielectric EOT. The effect of the gate dielectric on the actual gate work-function also need to be calibrated and taken into account when choosing the channel implants and diffusion processes.

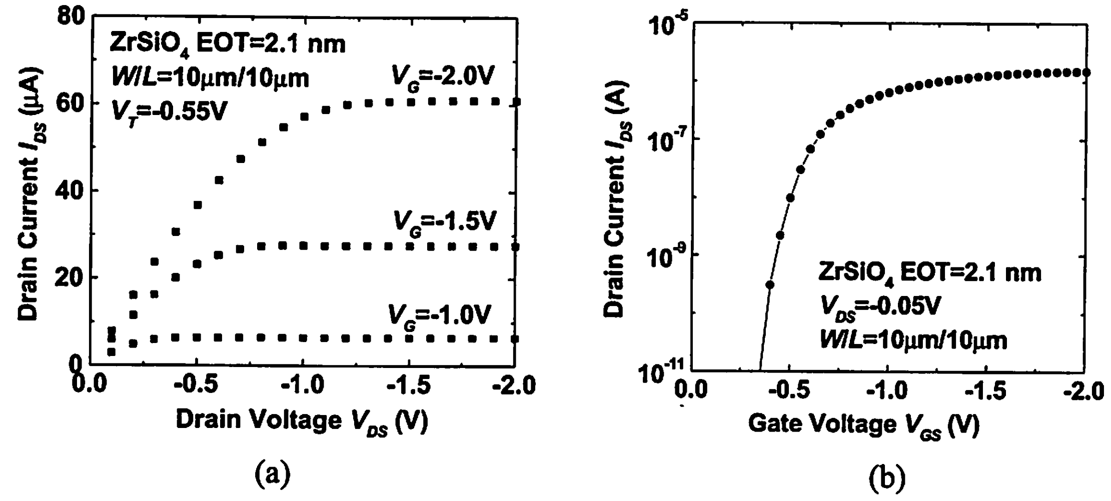


Figure 4.14 The transistor characteristics (a) I_{DS}-V_{DS} and (b) I_{DS}-V_{GS} of p-FETs with Mo gate and Zr-silicate gate dielectric.

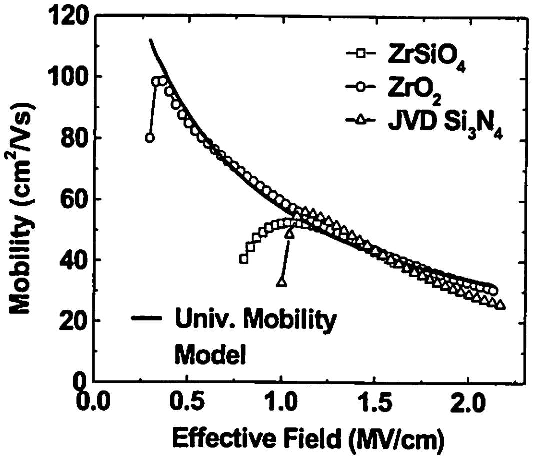


Figure 4.15 Field effective mobility of the p-FETs with Mo gate match the universal mobility model for SiO$_2$, which suggests good interface quality between the gate dielectrics and the silicon substrate.

The channel hole mobilities of these p-FETs are plotted against the effective electric field in Figure 4.15. Due to the different EOTs and threshold voltages, the p-FETs with different gate dielectrics operate under different effective field range. All three non-SiO_2 gate dielectrics showed hole mobilities that can match the universal model for SiO_2. This provides another evidence that the Mo gate remained stable during the high-temperature annealing.

Both material and electrical characterization of the Mo gated p-FETs indicate that the Mo gate can be used on different gate dielectrics in a self-aligned process, and the gate stack thermal stability is not limited by the Mo gate per se, i.e., with certain gate dielectrics (e.g., the JVD silicon nitride), anneals at above 1000°C can be tolerated. Although unlike the degenerately doped poly-Si gate, the effective work-function of Mo depends on the underlying gate dielectric and can be moved toward the Si midgap in some cases, there is still enough design window to obtain a feasible threshold voltage for short-channel bulk-Si p-FETs with carefully chosen gate dielectrics. The smaller gate work-function range required by novel CMOS structures with undoped channels further reduces the impact of the above effect. The unique features of easy deposition and patterning as demonstrated in the p-FET process suggest that a self-aligned process using Mo as the p-FET gate electrode is a practical approach to metal gated CMOS.

4.5 Dual metal gate CMOS process

4.5.1 Dual-metal gate CMOS process integration

The power supply voltage is reduced for every newer generation of CMOS technology, therefore, it is very important to achieve precisely controlled low threshold voltages for the most advanced technology to maintain sufficient performance. The discussions in section 4.3 shows that using a single midgap metal gate for both n- and p-channel MOSFETs will result in fairly high threshold voltages for both types of transistors, therefore it is not an ideal option for short channel bulk-Si MOSFETs, which require a high channel doping concentration to reduce the short channel effects. This theoretical analysis was confirmed by experimental results of single PVD TiN gate ($\varphi_m \approx$ 4.7 eV) for bulk-Si and fully depleted silicon-on-insulator (FD-SOI) CMOS technology [4.19]. Therefore, dual gate work-functions will be needed for high performance metal gate CMOS technology.

There are basically two approaches to integrating metal gate electrodes in a CMOS process with self-aligned source/drain formation. Based on the different orders of the gate stack and source drain formation, these two integration schemes are sometimes referred to as "gate-first" and "gate-last" CMOS process, respective. The gate-first process is similar to the standard dual poly-Si gate process, where the gate stack deposition and gate electrode patterning happen before the source/drain formation. In the gate-last process, a replacement gate stack is used to ensure the self-aligned source/drain formation, and CMP can be used to pattern the metal electrodes. The advantage of this process is that metals that are difficult to pattern by RIE may still be used and patterned by CMP. In addition, the gate dielectric is deposited after the source/drain dopant activation, therefore its exposure to high-temperature processes can be minimized, making it possible to use the gate dielectrics that are not thermally stable with the gate electrode at high temperature. A schematic process flow of the single metal gate replacement gate CMOS process is shown in Figure 4.16 [4.20]. After the shallow trench

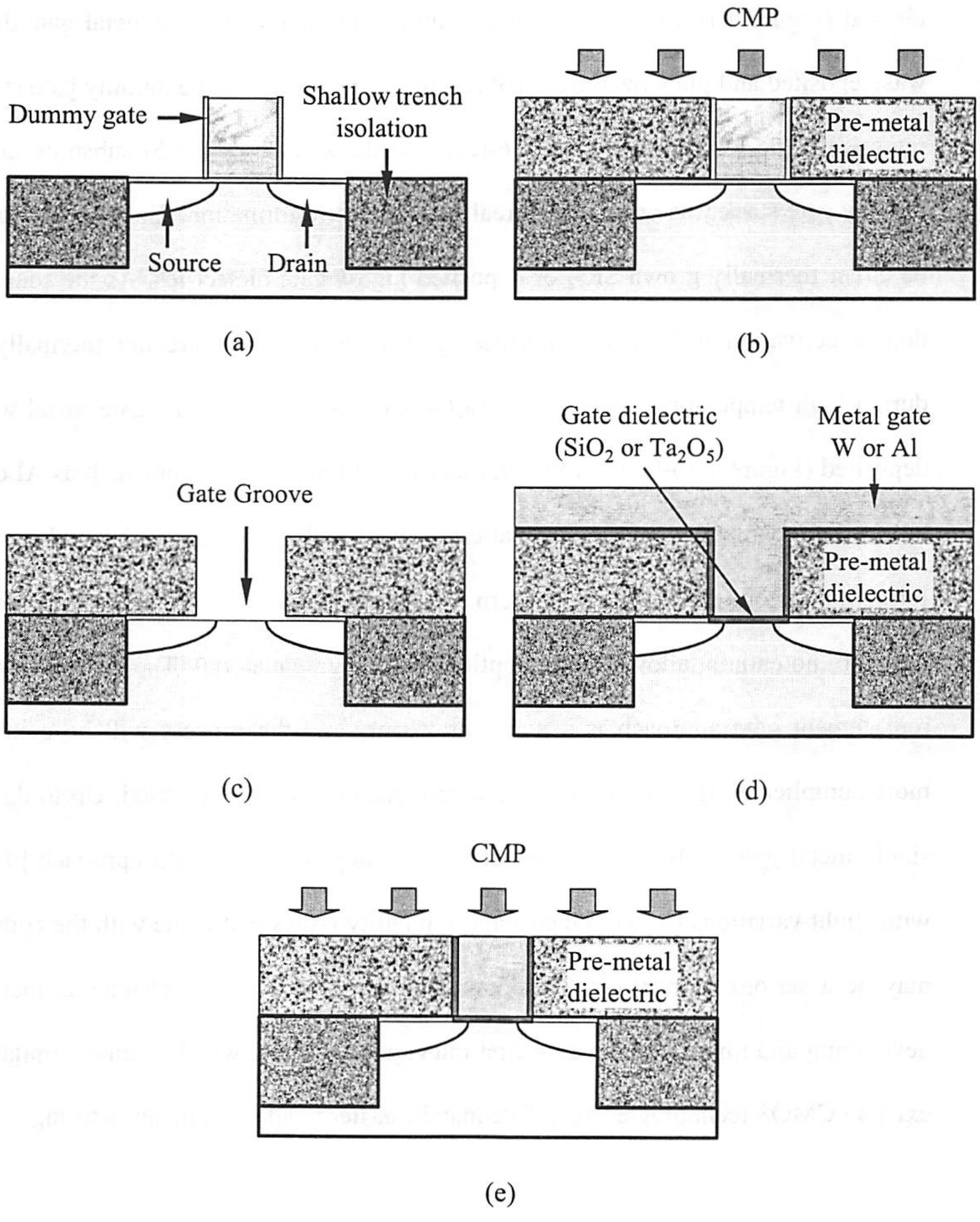


Figure 4.16 A schematic process flow of the single damascene metal gate CMOS fabrication process (after [4.20]).

isolation (STI) process, a dummy gate stack consisting of dummy gate oxide and dummy Si_3N_4/poly-Si gate was formed, so that the source and the drain implantations can be self-aligned (Figure 4.16-(a)). After source/drain dopant activation, pre-metal gate dielectric was deposited and planarized by CMP (Figure 4.16-(b)). Then the dummy gate stack was removed using wet etch (Figure 4.16-(c)), and the surface of the Si substrate under the dummy gate stack was prepared for real gate dielectric formation. The gate dielectric can be either thermally grown SiO_2 or deposited high-*k* gate dielectrics. As the source/drain dopant activation was already finished, gate dielectrics that are not thermally stable during high temperature anneal, e.g., Ta_2O_5, can also be used. The gate metal was then deposited (Figure 4.16-(d)). Similarly, metals with low melting point such as Al can also be used. The metal gate is then patterned by CMP to complete the real gate stack formation. The metal CMP can pattern some metals that are not easily etched by RIE, therefore potentially allowing more options for the gate material. The complexity of this replacement gate approach is a major drawback, and the process will be conceivably more complicated if two different metal gate electrodes are to be used. Up to date, only single metal gate CMOS has been published using this integration approach [4.20], or with slight variations [4.21].The cost and reliability issues that come with the complexity may be a serious barrier to the adoption of this approach. This chapter is focused on developing and improving the gate-first integration scheme, which is more similar to the existing CMOS technologies, therefore may be easier to adopt in manufacturing.

4.5.2 Directly-deposited dual metal gate CMOS process

The directly-deposited dual metal gate CMOS process is a gate-first process, in which the two different metal electrodes used for n- and p-FETs are directly deposited on the gate dielectric, as opposed to chemically/structurally transformed from another metal afterwards. The schematic process flow is shown in Figure 4.17. In the experiment, the gate dielectric was RTCVD silicon nitride, deposited by Prof. D. L. Kwong's group at UT Austin. Based on the positive results presented in the previous sections, Mo was used as the p-FET gate material. Ti, which has high melting point and can be easily deposited and patterned, was chosen to be the gate metal for n-FETs. The same n-well CMOS process as described in Chapter 2 was followed up to sacrificial oxidation. After RTCVD silicon nitride gate dielectric deposition, 230 Å Ti was deposited by sputtering on the whole wafer, then capped with 230Å sputter deposited TiN, which served as a barrier layer (Figure 4.17-(a)). A patterning process was performed to remove the two metal layers on the p-FETs' side so as to prepare for p-FET gate metal the deposition, while the n-FETs were protected by photoresist throughout this step (Figure 4.17-(b)). To improve the selectivity of this metal etch process and to reduce the possible damage to the p-FET gate dielectric, a wet etch (NH_4OH : H_2O_2 : H_2O = 1:1:5 by volume) was used. After photoresist removal by O_2 plasma and cleaning, 200 Å Mo was deposited by sputtering, and also capped with 230 Å sputter deposited TiN. The gate stack step formation was completed with 1000 Å LPCVD *in situ* n^+ doped poly-Si, which prevents the underlying metal layers from oxidation in the subsequent anneals (Figure 4.17-(c)). Then the gate lithography was performed, followed by gate etching using a three-step RIE (Figure 4.17-(d)). The gate etch started with Cl_2:O_2 RIE to etch the top n^+ poly-Si and the TiN, then switched to CF_4 to etch the Mo, and finally switched back to Cl_2:O_2 to etch the TiN/Ti at the bottom of the gate stack. The Cl_2:O_2 chemistry also slowly etches the RTCVD silicon nitride, therefore, it is important to use very thin TiN/Ti at the bottom so as to avoid breaking through the gate dielectric and attacking the Si-substrate. Beyond this step, the process is identical to a dual poly-Si CMOS process. After source/drain implantations for

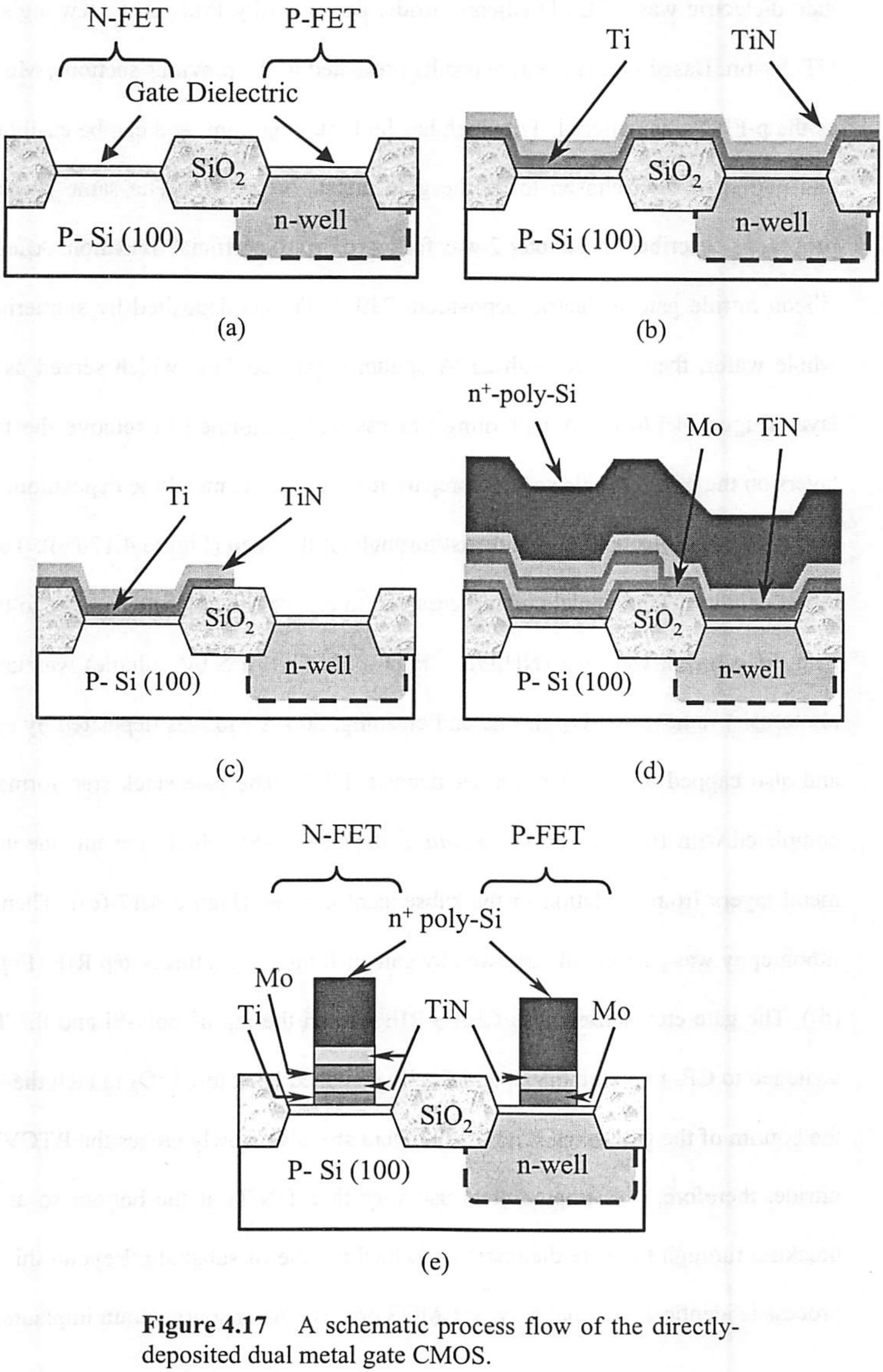


Figure 4.17 A schematic process flow of the directly-deposited dual metal gate CMOS.

n- and p-FETs, an RTA dopant activation anneal (950° 10s + 1025°C 5s) was performed. The process was completed with LTO passivation, metallization and forming gas anneal.

4.5.3 Dual metal gate CMOS characterization

As a basic test of the gate stack quality after the full CMOS process, the *C-V* characteristics of the Mo-gated p-FETs and Ti-gated n-FETs were measured and compared to quantum *C-V* simulation results (Figures 4.18 and 4.19). For both p- and n-FETs, the inversion capacitance is the same as the accumulation capacitance, indicating the elimination of the gate depletion effect. The Mo-gated p-FETs' *C-V* agrees well with the simulation results except for the distortion near the flat-band condition (0 V < V_G < 0.5 V). The distortion is due to interface traps, and has been seen in the poly-Si gated p-FETs with RTCVD silicon nitride gate dielectric of the same deposition conditions (Chapter 2, Figure 2.6-(b)), therefore it is more likely as result of the gate dielectric process itself rather than caused by the Mo gate. The Mo gate work-function extracted by

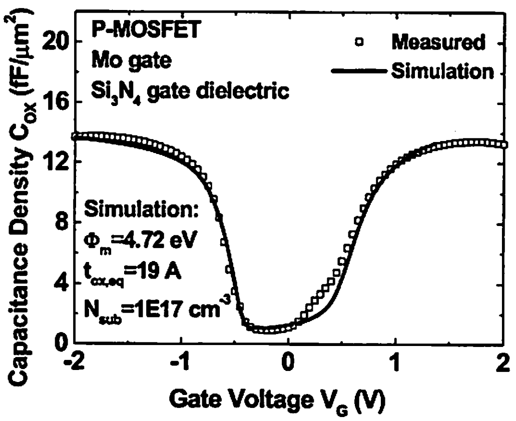


Figure 4.18 Measured *C-V* characteristics of the p-FETs with Mo gate and RTCVD gate dielectric. An effective gate work-function of 4.72 eV and an EOT of 19 Å were extracted by quantum *C-V* simulation.

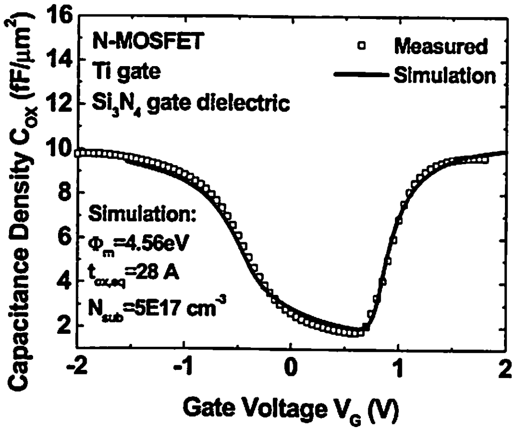


Figure 4.19 Measured *C-V* characteristics of the n-FETs with Ti gate and RTCVD gate dielectric. An effective gate work-function of 4.56 eV and an EOT of 28 Å were extracted by quantum *C-V* simulation.

quantum simulation is 4.72 eV, which is similar to the value (4.76 eV) obtained with the JVD silicon nitride. For the Ti-gated n-FETs, however, the extracted work-function is 4.56 eV, much higher than expected for a Ti gate. A possible reason is that Ti is not thermally stable with the silicon nitride gate dielectric during the aggressive anneal, and the reaction converted the n-FET's gate electrode to TiN. A study on similar material systems showed that the higher affinity of Ti to nitrogen is the driving force for such a nitrogen gettering effect during high-temperature anneal [4.22]. In the particular gate stack structure used in this experiment, the thin Ti gate between the silicon nitride gate dielectric and the TiN barrier may getter nitrogen from both neighboring layers, converting the Ti gate to TiN and also reducing the nitrogen content in the RTCVD silicon nitride gate dielectric. The percentage oxygen concentration in the gate dielectric then increased, which reduced the dielectric constant of the film and therefore contributed to the increase of the n-FETs' EOT. Another unexpected phenomenon is the large difference in EOT between the p- and the n-FETs. This is due to the wet etch of the Ti gate on the p-FETs' gate dielectric (Figure 4.17-(c)). Compared to RIE, the wet etch rate is less stable and more difficult to calibrate accurately, and the selectivity against silicon nitride is not sufficiently high. So the over-etch of the Ti gate electrode also attacked the p-FETs' gate dielectric, resulting a thinner EOT for the p-FETs. This difference in EOT can also be seen in the transistor I_{DS}-V_{DS} characteristics (Figure 4.120). Unlike the commonly observed 1:2~3 drive current ratio for p-FETs vs. n-FETs, the drive currents of the p- and n-FETs in this dual metal gate CMOS are comparable. This is caused by the thinner EOT of the p-FETs, which increases the inversion charge density and enhances the drive current. The undesirable thinning of the p-FETs' gate dielectric during the metal wet etch is potentially a reliability concern. The time-dependent dielectric breakdown (TDDB) design and qualification will become more complicated due to the different gate dielectric thickness. And the wet etch can cause more variation in the gate dielectric thickness and device performance.

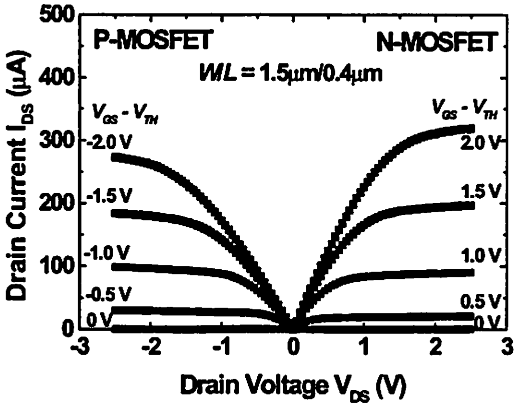


Figure 4.20 I_{DS}-V_{DS} characteristics of the dual metal gate CMOS transistors. Due to the much thinner EOT of the p-FETs, the difference between the n- and p-FETs' drive currents is much smaller than usually seen on CMOS transistors on the same wafer.

The transistor I_{DS}-V_{GS} characteristics of the dual metal gate CMOS are shown in Figures 4.21 and 4.22. Both n-FETs and p-FETs have very good turn-off behavior and low off-state leakage currents The subthreshold swing of the p-FETs is very close to the value expected for the EOT and substrate doping, while that of the n-FETs is relatively larger, partly due to the thicker EOT. Given the clean *C-V* curve of the n-FETs, this slightly larger subthreshold swing is more likely due to the junction design, which can be improved by modifying the source/drain formation process. It also can be seen that the p-FETs have a reasonably low threshold voltage, but the n-FETs' threshold voltage is high,

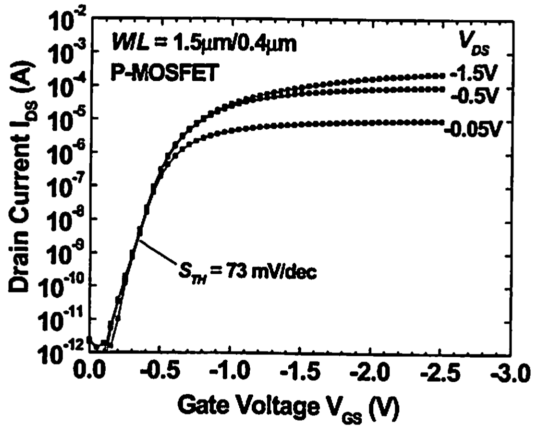


Figure 4.21 I_{DS}-V_{GS} characteristics of the Mo-gated p-FETs. Very good subthreshold swing is obtained.

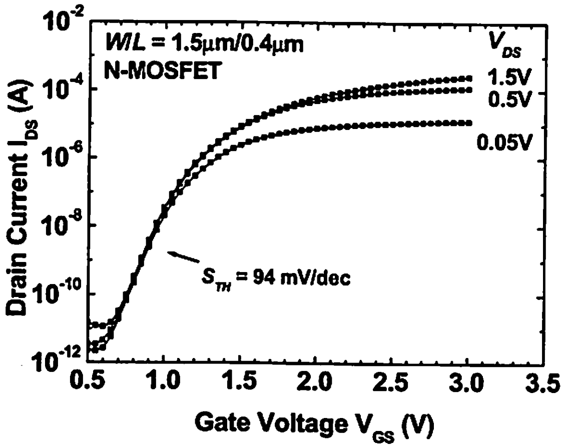


Figure 4.22 I_{DS}-V_{GS} characteristics of the Ti-gated n-FETs. The subthreshold swing is slight degraded.

due to the almost midgap n-FET effective gate work-function resulting from the high-temperature annealing. In order to achieve desirable threshold voltages, the appropriate gate work-function must be attained after the high-temperature processes.

The channel carrier mobilities of the dual metal gate p- and n-FETs are shown in Figures 4.23 and 4.24. Similar to the Mo-gated p-FETs discussed in section 4.4, the p-FETs with RTCVD silicon nitride gate dielectric and Mo gate in this CMOS process can also match the universal mobility model for holes despite the unwanted thinning during the metal wet etch, providing further evidence for the high quality of the gate dielectric as well as the good thermal stability of the gate stack during annealing at above 1000°C. In addition to the interface between the Si-substrate and the gate dielectrics, the interface between the gate and the gate dielectric also plays an important role in the channel carrier mobility. A rough upper interface of the gate dielectric can also cause the non-uniformity of the vertical field in the channel, thereby enhancing carrier scattering and degrading mobility under high field [4.23]. So the good mobility of the p-FETs is also indicative of a stable and smooth interface between the Mo gate and the RTCVD silicon nitride gate dielectric. Compared to the universal mobility model, the electron mobility of the n-FETs is degraded in low to medium field range and shows the crossover behavior at higher field, consistent with other results obtained with nitrided gate dielectric interfaces (Figure 2.18 and [4.24]). The Coulombic scattering of the channel electrons in the medium field range needs to be reduced to improve the mobility under normal device operation gate bias. In this experiment, the use of the metal gate electrodes did not seem to have caused significant mobility degradation, and improving the interface quality of the alternative gate dielectrics is still a major challenge regardless of the gate electrodes.

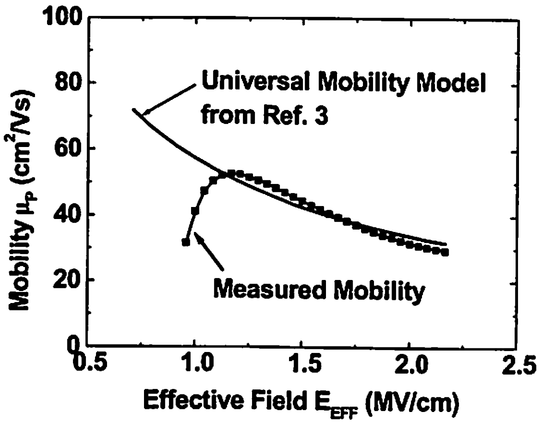


Figure 4.23 The channel hole mobility of the p-FETs with Mo/RTCVD nitride gate stack is comparable to the universal mobility model.

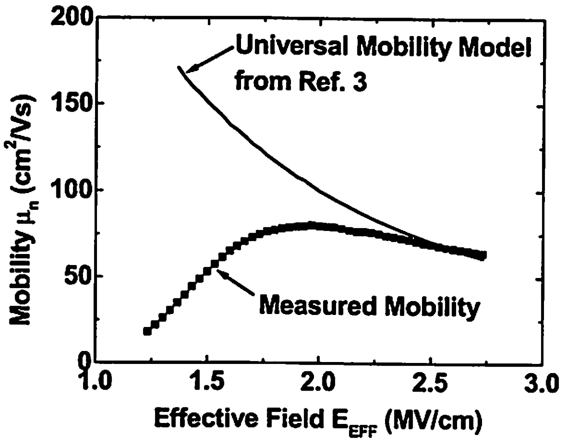


Figure 4.24 Compared to the universal mobility model, the channel electron mobility of the n-FETs with Ti/RTCVD nitride gate stack is degraded in medium field range, and crosses over the universal model at high field.

The above dual metal gate CMOS process successfully shows that a self-aligned source/drain process is possible with two different metal gate electrodes when the appropriate gate and gate dielectric materials are chosen. The device parameters used in the process, such as the channel doping, the gate stack thickness and the gate length, are very realistic in view of the current CMOS technology. The demonstrated benefits make it worth pursuing further. On the other hand, the above specific process integration approach has a couple of problems that affect its robustness and reliability. The unwanted etching of the p-FETs' gate dielectric during the wet etch of n-FETs' gate has been shown to be a reliability and process control concern. Although this in principle can be improved by using an etch process with higher selectivity against the gate dielectric, the process of depositing then removing a metal on the gate dielectric can always be a potential reliability issue. Another major difficulty is the gate etch step, where highly asymmetrical gate stacks of the n- and p-FETs are etched in same process (Figure 4.17-d and e). In such a simultaneous etch process of the two different gate stacks, it is not easy to ensure that the thicker n-FET gate stack is completely cleared while the etching of the thinner p-FET gate stack is precisely stopped before attacking the Si-substrate. The process window of this gate etch is not big enough to make it reliable in large scale integration. So at this point, a major challenge is to improve and to simplify the self-aligned metal CMOS process integration. A conceptually simple solution is to perform the gate lithography for n- and p-FETs separately, and to etch one type of gate stack with the other protected by photoresist. This, however, is not easy to implement due to the difficulty of the gate lithography as well as the additional cost. Therefore, symmetrical gate stacks, i.e., stacks consisting of the same thin film layers, for the n-and p-FETs are highly desirable. These problems, all of which originate from the need for using different metal gate electrodes for n- and p-FETs, contrast the unique advantages of the dual poly-Si gate technology, which allows dual gate work-functions to be achieved based on the same gate material. Similarly, without compromising the device performance by using a single midgap work-function metal, the means to achieve dual/tunable work-function using a single metal gate electrode can greatly simplify the metal gate CMOS process.

4.6 Single Mo gate with tunable work-function CMOS

4.6.1 Introduction

A technique to modulate the work-function of a metal gate electrode is critical to developing a simpler metal gate CMOS technology, in which the same metal gate material is deposited on both n- and p-FETs and then converted to the appropriate work-functions respectively. Such a conversion process may be chemical or structural change, as long as sufficient shift in the work-function can be obtained. As an analogy to the dual poly-Si process, doping the metal gate by implantation is a somewhat natural idea, although it is obviously not trivial to identify the suitable metal and dopant. An experimental study of the effects of doping on Mo's work-function demonstrated an interesting phenomenon that can be potentially used for such a purpose [4.25]. As shown in Figure 4.25, the *C-V* curves of MOS capacitors with Mo gates can be significantly shifted due to different implantation and annealing conditions. An anneal of 700°C 15 min without ion-implantation (No I/I) shifted the as-deposited *C-V* to more positive gate voltage, indicating a high gate work-function. When the gate is implanted with nitrogen and annealed at 700°C for 15 min, the *C-V* curves was shifted to more negative gate voltage side, indicating lowered gate work-function. Without deducting the effects of fixed charge, etc., the shift between implanted and unimplanted curves is larger than 1 V, making this effect attractive for application to a single metal CMOS process.

To fabricate CMOS transistors using a nitrogen implanted gate, it is necessary to get a better understanding of how the shift in gate work-function depends on the process parameters, such as annealing and implantation conditions. In the above MOS capacitor experiment, the Mo (~1500Å) and SiO_2 (~1000Å) films were too thick for a realistic MOSFET process, and the annealing thermal budget of 700°C 15 min is also impractically low. In order to decide on a suitable process condition for the above phenomenon to be used in a CMOS process, a set of MOS capacitor experiments were first conducted.

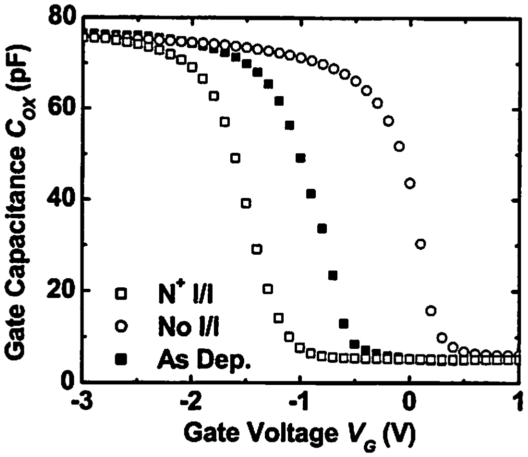


Figure 4.25 MOS capacitor *C-V* characteristics showing the effects of nitrogen implantation and annealing on the work-function of the Mo gate. The curves of no ion-implantation (No I/I) and with nitrogen implantation (N^+ I/I) both received 700°C 15 min anneal, and showed more than 1 V shift in flatband voltage. (From [4.25])

The MOS capacitors were fabricated on a lightly doped (~$1\times10^{15}/cm^3$) p-type Si substrate, with varied SiO_2 gate dielectric thickness (30-80 Å) and 650 Å sputter deposited Mo gate electrode. After the gate stack formation, lithography was performed to cover half of each wafer with photoresist. The wafers received nitrogen implants of $5\times10^{15}/cm^2$ dose and varying implant energy, and then were capped with sputter-deposited TiN and *in situ* n$^+$ doped LPCVD poly-Si gate, followed by RTA at 700 °C for 10 min. The shift in the gate work-function due to the nitrogen implantation was extracted from the *C-V* curves measured from the same wafer. The possible flatband shift due to interface states and fixed charge should be the same for devices on the same wafer, therefore the flatband shift between devices with or without the nitrogen implantation of the gate should be caused primarily by the implantation. Table 4.2 is a summary of the implantation energy and corresponding effective gate work-function shift. The simulated implantation range R_P and longitudinal straggle ΔR_P for each implantation energy using SRIM simulator are also shown in the table [4.26]. Initially, the shift in the gate work-function increases with higher implantation energy, until the implantation energy is so high as to cause too much damage to the gate dielectric resulting in device failure, evidenced by high gate leakage current and distorted *C-V* curves. The simulation assumes an amorphous Mo film, but the actually Mo film is polycrystalline with a columnar morphology, therefore the actual implantation range is likely larger than the simulated

Table 4.2 Experimentally measured shift in the effective work-function of the Mo gate and simulated implantation range for different nitrogen implantation

Implantation energy E_{imp} (keV)	Projected R_P (Å)	Straggle (Å)	$\Delta\Phi_m$ (eV)
15	205	112	0.26
22	279	148	0.40
29	353	180	0.56
35	415	209	Non-yielding
45	500	232	Non-yielding

value. Based on these results, a practical implantation energy condition for a CMOS process should be within 30 keV.

4.6.2 Single Mo dual gate work-function CMOS process

Using Mo as the gate material for both p- and n-FETs with nitrogen implantation to adjust the gate work-function, the gate-first CMOS process is very similar to the conventional dual-poly-Si gate CMOS. As illustrated in Figure 4.26, after standard well

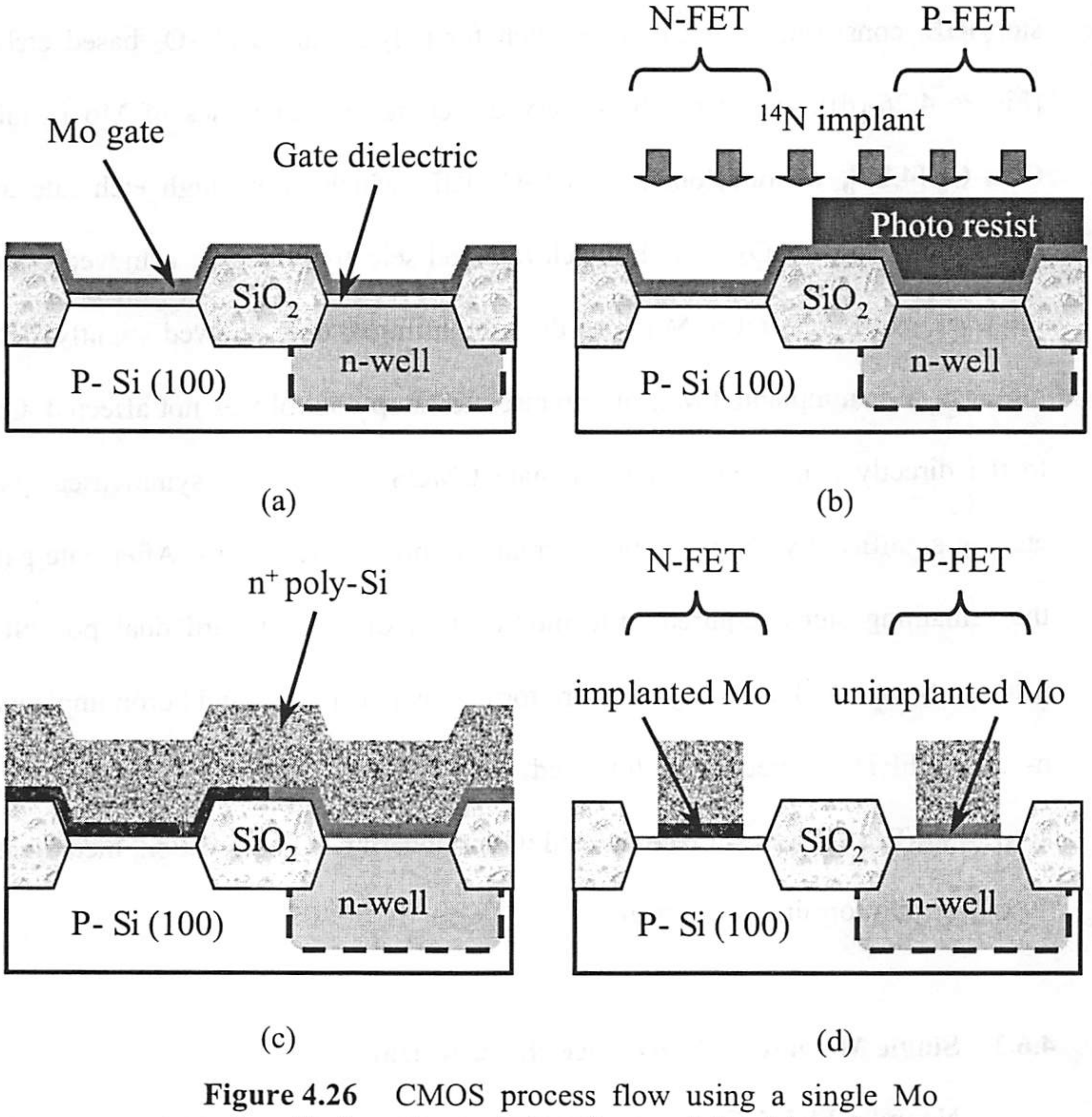
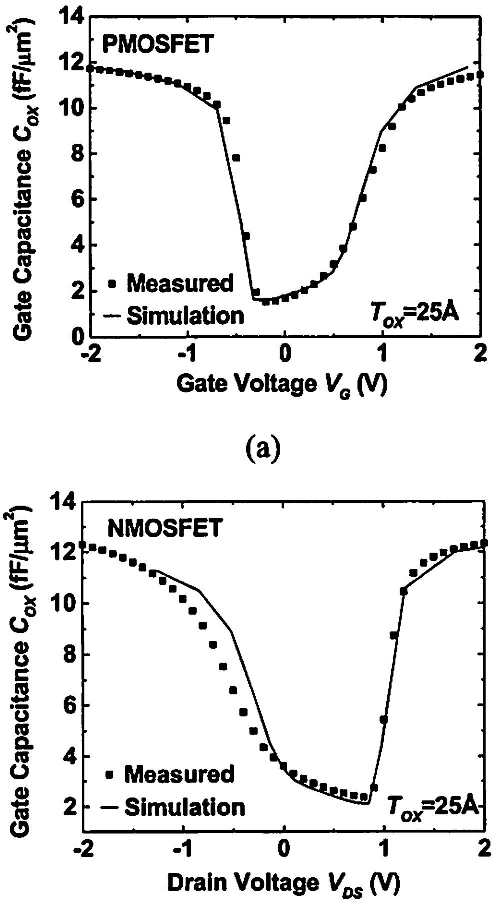


Figure 4.26 CMOS process flow using a single Mo gate with the gate work-function modulated by nitrogen implantation for the n-FETs.

formation, LOCOS processing, V_T implants, and gate dielectric formation, a 650 Å Mo gate layer was deposited on the whole wafer (Figure 4.26-(a)). A lithography step was performed to cover the p-FETs with photoresist while exposing the n-FETs to nitrogen implantation (Figure 4.26-(b)). The same dose ($5\times10^{15}/\text{cm}^{-2}$) but two different implant energies (15 and 29 keV) were used to study the process window for CMOS transistors. After stripping the photoresist, 1000 Å *in situ* n^+ doped poly-Si was deposited on top of the Mo film (Figure 4.26-(c)). After gate lithography, gate etching was done by a two-step RIE, consisting of a Cl_2 based etch for poly-Si and a Cl_2-O_2 based etch for Mo (Figure 4.26-(d)). Based on the observed etching characteristics of Mo in mixture of CCl_4-O_2 [4.27], we developed the Cl_2-O_2 RIE, which offers high etch rate and good selectivity against SiO_2. Variable etch rate and selectivity can be achieved by adjusting the Cl_2:O_2 flow rate ratio. Mo with the high nitrogen dose showed slightly slower etch rate than the unimplanted Mo, but precise etch stop control was not affected. Compared to the directly deposited dual metal gate CMOS process, the symmetrical gate stack etching significantly improves the tolerance to process variations. After gate patterning, the remaining steps required little modification of the standard dual poly-Si CMOS process. Source and drain regions were formed by phosphorus and boron implantation for n- and p-FETs, respectively, followed by 800°C 30 min furnace anneal for dopant activation. The devices were completed with standard LTO passivation, metallization and 400°C 30 min forming gas anneal.

4.6.3 Single Mo gate CMOS device characterization

N- and p-FETs' *C-V* characteristics with gate nitrogen implantation energies of 15 keV and 29 keV are shown in Figures 4.27 and 4.28, respectively. Due to the different process conditions for gate dielectric formation, the devices on the two wafers have different EOTs. On the same wafer, the n- and p-FETs show the same EOT, which is an improvement over the directly-deposited dual metal gate CMOS process. The *C-V* curves of the p-FETs on both wafers can be well fitted by quantum simulation, but the n-FETs' *C-V* curves for both implantation energies showed visible deviation from the simulated, or ideal, results, indicating possible damage to the gate dielectrics introduced by the high-dose nitrogen implantation. Comparing Figures 4.27-(b) and 4.28-(b), the distortion is apparently more serious for the n-FETs with thinner EOT, although the corresponding

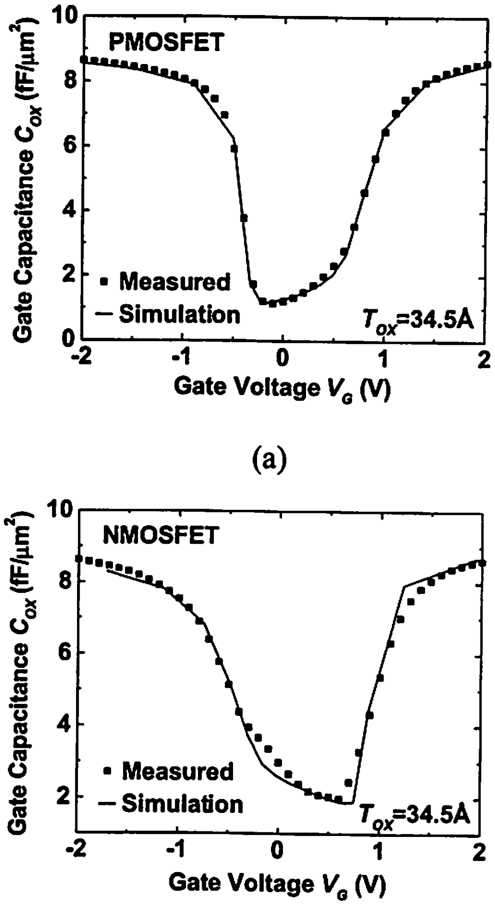


(a)

(b)

Figure 4.27 *C-V* characteristics of (a) p-FETs with unimplanted Mo gate and (b) n-FETs with 15 keV nitrogen implantaion into the Mo gate.

nitrogen implantation energy is actually lower. This suggests that the high-temperature annealing could also be a damaging mechanism in addition to the implantation. The physical cause of the change in work-function by nitrogen implantation is not fully clear yet, but current research findings indicate that it is a combined result of both chemical and structural changes. SIMS analysis shows that upon annealing, the nitrogen species pile up at the upper interface of the gate dielectric, and formation of MoN is also found [4.28]. The high concentration nitrogen may be incorporated into or diffuse through the gate dielectric, therefore degrading the MOS characteristics.



(a)

(b)

Figure 4.28 *C-V* characteristics of (a) p-FETs with unimplanted Mo gate and (b) n-FETs with 29 keV nitrogen implantaion into the Mo gate.

The effective gate work-functions of these devices were extracted using the quantum *C-V* simulation and summarized in table 4.3. The qualitative correlation between the implantation energy E_{imp} and the difference in the effective gate work-function $\Delta\Phi_m$ is similar to those listed in Table 4.2, but for the same E_{imp}, $\Delta\Phi_m$ is smaller. The major difference in the CMOS process is the higher dopant activation annealing temperature than the MOS capacitor process, so the low work-function of Mo induced by nitrogen implantation may be restored by high temperature processes. The similar trend is also consistently observed for varying nitrogen dose [4.28].

Table 4.3 Effective work-functions of the Mo gates extracted from the CMOS transistors.

Nitrogen implantation energy E_{imp} (keV)	15	29
p-FET gate Φ_m (eV)	4.94	4.95
n-FET gate Φ_m (eV)	4.70	4.53
$\Delta\Phi_m$ (eV)	0.24	0.42

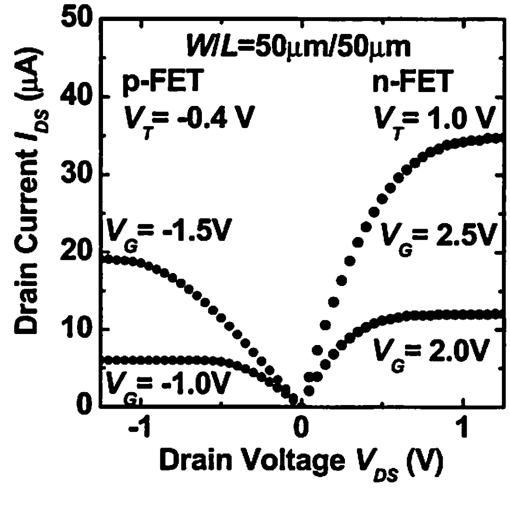


Figure 4.29 Long channel n- and p-FETs fabricated in the single Mo gate CMOS process (with nitrogen implantation into n-FET gates) show normal transistor characteristics.

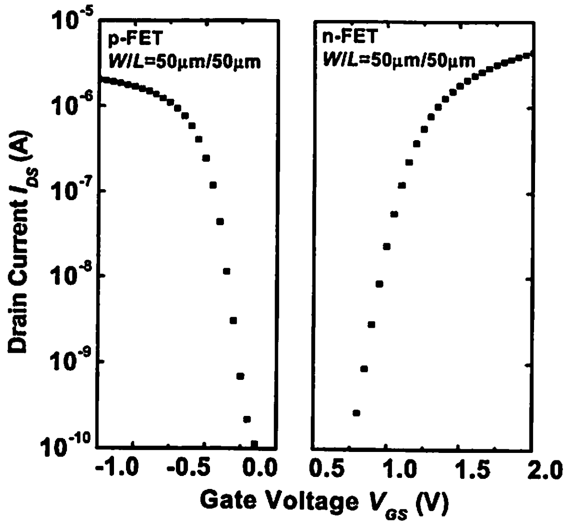


Figure 4.30 I_{DS}-V_{GS} characteristics of the long channel p-FETs with unimplanted Mo gate and n-FETs with nitrogen implanted Mo gate.

The long channel transistor I_{DS}-V_{DS} characteristics of p- and n-FETs with 29 keV nitrogen implantation are shown in Figure 4.29. Normal transistor behaviors are obtained, but the threshold voltages of the n- and p-FETs are not symmetrical. The n-FET gate work-function is still significantly higher than the ideal value for bulk n-FET gate, so the n-FET V_T is too high. The I_{DS}-V_{GS} curves of the long channel n- and p-FETs are shown in Figure 4.30. Both n- and p-FETs have low off-state leakage currents and normal subthreshold swings. Similar results are observed on the CMOS transistors with the thinner EOT and 15 keV nitrogen implantation energy. The V_T mismatch between n- and p-FETs is more serious due to the even higher n-FET gate work-function.

While these results demonstrated the concept of using a single metal gate with adjustable work-function for simplifying CMOS fabrication process, a couple of problems are outstanding. Unlike the dual poly-Si gate process, work-function reduction of Mo requires considerably higher implantation dose and implantation energy. So the damage to the gate dielectric and the substrate interface induced by the gate implantation process degrade device characteristics and reliability. Using thinner Mo gate and consequently lower implantation energy may reduce the implant straggle, thus alleviating the problem, but it is a costly and incomplete solution. A solid source diffusion approach was recently proposed as an alternative method to incorporate nitrogen into Mo gate [4.29]. Using over-stoichiometric TiN_{1+X} deposited on top of the thin Mo film as the solid source, nitrogen was diffused into the Mo during a high-temperature anneal process, which does not involve ion implantation damage. The work-function shift achieved using this method is very close to that obtained by nitrogen implantation. Another issue for this Mo-N system is that the range of the work-function adjustment is not large enough for optimal bulk-Si CMOS, as already seen from the high V_T of the n-FETs. MoN, which has a work-function of 4.4 eV, is likely the lower limit of this work-function adjustment mechanism. Considering the recovery of the effective work-function after high temperature processes, the Mo-N system may not be feasible for bulk Si CMOS even after optimization. However, the effective work-function of the n-FET gate achieved in the CMOS process is acceptable for a FinFET or fully depleted SOI transistor structures, and the Mo gate can also be used for CMOS transistors based on these structures.

Comparing the two metal gate CMOS integration schemes, the gate-first approach has the main advantage of the relative simplicity, and the gate-last approach offers more flexibility at the price of process complexity. The gate-first integration poses more challenges in terms of materials selection and processing, and is closely related to the research progress of high-*k* gate dielectrics. The two dual gate work-function metal gate CMOS processes discussed in this chapter suggest that the stability of the gate stack and the gate work-function is currently the major issue, and the reliability problems need to be addressed in the process modules, e.g., metal deposition and doping.

4.7 References

[4.1] C.-T. Sah, "Evolution of the MOS transistor – from conception to VLSI", *Proceedings of the IEEE*, Vol. 76, No. 10, pp. 1280-1326, Oct. 1988.

[4.2] D. G. Ong, *Modern MOS Technology: Processes, Devices, and Design*, p. 139, McGraw-Hill, 1984.

[4.3] D. G. Schlom and J. H. Haeni, "A thermodynamic approach to selecting alternative gate dielectrics", *MRS Bulletin*, pp.198-204, March 2002.

[4.4] L. Kang, Y. Jeon, K. Onishi, B.-H. Lee, W.-J. Qi, R. Nieh, S. Gopalan and J. C. Lee, "Single-layer thin HfO_2 gate dielectric with n^+-polysilicon gate", *Symposium on VLSI Technology*, pp. 44-45, June 2000.

[4.5] T. Sakurai, and T. Iizuka, "Gate electrode RC delay effects in VLSIs", *IEEE Transactions on Electron Devices*, Vol. 32, No. 2, pp. 370-374, Feb. 1985.

[4.6] M. Krishnan, Y.-C. Yeo, Q. Lu, "Remote charge scattering", *International Electron Devices Meeting*, Dec. 1998.

[4.7] W.-C. Lee and C. Hu, "Modeling gate and substrate currents due to conduction and valence band electron and hole tunneling", *Symposium On VLSI Technology*, pp. 198-199, June 2000.

[4.8] Y.-C. Yeo, Q. Lu, W.-C. Lee, T.-J. King, C. Hu, X. Wang, X. Guo, and T. P. Ma, "Direct tunneling gate leakage current in transistor with ultrathin silicon nitride gate dielectric", *IEEE Electron Device Letters*, Vol. 21, No. 11, pp. 540-542, Nov. 2000.

[4.9] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, "Generalized guide for MOSFET miniaturization", *IEEE Electron Device Letters*, Vol. EDL-1, No. 1, pp. 2-4, Jan. 1980.

[4.10] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs", *IEEE Transactions on Electron Devices*, Vol. 40, No. 12, pp. 2326-2329, Dec. 1993.

[4.11] *CRC handbook of chemistry and physics*, 66th Edition, R. C. Weast (Editor), CRC Press 1985.

[4.12] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang; J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi; K. Asano, V. Subramanian, T.-J. King, J. Bokor, C. Hu, "Sub 50-nm FinFET: PMOS". *International Electron Devices Meeting*, pp. 67-70, Dec. 1999.

[4.13] Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J. King, J. Bokor, C. Hu, "Ultra-thin body SOI MOSFET for deep-sub-tenth micron era", *International Electron Devices Meeting*, pp. 919-921, Dec. 1999.

[4.14] H. B. Michaelson, "Relation between an atomic electronegativity scale and the work function", IBM Journal of Research and Development, Vol. 22, No. 1, pp. 72-80, Jan. 1978.

[4.15] L. Chang, S. Tang, T.-J. King, J. Bokor, C. Hu, "Gate length scaling and threshold voltage control of double-gate MOSFETs", *International Electron Devices Meeting*, pp. 719-22, Dec. 2000.

[4.16] D. M. Brown, W. E. Engeler, M. Garfinkel, and P. V. Gray, "Refractory metal silicon device technology", *Solid-State Electronics*, Vol. 11, No. 12, pp. 1105-1112, Dec. 1968.

[4.17] K. Yang, Y.-C. King, and C. Hu, "Quantum effects in oxide thickness determination from capacitance measurement", *Symposium On VLSI Technology*, pp. 77-78, June 1999.

[4.18] Y.-C. Yeo, P. Ranade, Q. Lu, R. Lin, T.-J. King, and C. Hu, "Effects of high-k dielectrics on the workfunctions of metal and silicon gates", *Symposium on VLSI* Technology, pp. 49-50, June 2001.

[4.19] B. Maiti, P. J. Tobin, C. Hobbs, R. I. Hegde, F. Huang, D. L. O'Meara, D. Jovanovic, M. Mendicino, J. Chen, D. Connelly, O. Adetutu, J. Mogab, J. Candelaria, L. B. La, "PVD TiN metal Gate MOSFETs on bulk silicon and fully depleted silicon-on-insulator (FDSOI) substrates for deep sub-quarter micron MCOS technology", *International Electron Devices Meeting*, pp. 781-784, Dec. 1998.

[4.20] A. Yagishita, T. Saito, K. Nakajima, S. Inumiya, Y. Akasaka, Y. Ozawa, G. Minanmihaba, H. Yano, K. Hieda, K. Suguro, T. Arikado, and K. Okumura, "High performance metal gate MOSFETs fabricated by CMP for 0.1 μm regime", *International Electron Devices Meeting*, pp. 785-788, Dec. 1998.

[4.21] A. Chatterjee, R. A. Chapman, K. Joyner, M. Orobe, 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", *International Electron Devices Meeting*, pp. 777-780, Dec. 1998.

[4.22] I. Shalish and Y. Shapira, "Thermal stability of a Ti-Si-N diffusion barrier in contactwith a Ti adhesion layer for Au metallization", *Journal of Vacuum Science and Technology B*, Vol. 17, No. 1, pp. 166-173, Jan./Feb. 1999.

[4.23] J. Li and T.-P. Ma, "Scattering of silicon inversion layer electrons by metal-oxide interface roughness", *Journal of Applied Physics*, Vol. 62, No. 10, pp. 4212-4215, Nov. 1987.

[4.24] Z. J. Ma, Z. H. Liu, Y. C. Cheng, P. K. Ko, and C. Hu, "New insight into high-field mobility enhancement of nitride-oxide n-MOSFETs based on noise measurement", *IEEE Transactions on Electron Devices*, Vol. 41, No. 11, pp. 2205-2209, Nov. 1994.

[4.25] P. Ranade, Y.-C. Yeo, Q. Lu, H. Takeuchi, T-J. King, and C. Hu, "Molybdenum as a gate electrode for deep sub-micron CMOS technology," *MRS Symposium Proceedings*, v. 611, C3.2.1, Spring 2000, San Francisco.

[4.26] Available at http://www.srim.org/

[4.27] Y. Kurogi and K. Kamimura, "Molybdenum etching using CCl_4 O_2 mixture gas", *Japanese Journal of Applied Physics*, Vol.21, No.1, Part 1, pp. 168-172, Jan. 1982.

[4.28] P. Ranade, H. Takeuchi, T.-J. King, and C. Hu, "Work function engineering of molybdenum gate electrodes by nitrogen implantation", *Electrochemical and Solid-State Letters*, Vol. 4, No. 11, pp. 85-87, Nov. 2001.

[4.29] R. J. P. Lander, J. C. Hooker, J. P. van Zijl, F. Roozeboom, M. P. M. Maas, Y. Tamminga, and R .A. M. Wolters, "Control of a metal-electrode work function by solid-state diffusion of nitrogen", *Material Research Society Meeting Proceedings*, Vol. 716, pp. B.5.11.1-B.5.11.6, Spring 2002.

Chapter 5

Conclusion

5.1 Summary

The preceding chapters discussed a number of key problems related to gate stack scaling in future generations of CMOS technology, and proposed and demonstrated some possible solutions. These problems stem from the fundamental limitations of the gate stack materials currently used in the CMOS technology, therefore significant improvement of the CMOS gate stack scalability will definitely require the introduction of new gate stack materials in addition to innovations in device structure and process technology.

It becomes clear that high-*k* gate dielectrics will be needed for low-power applications, and very likely for the high-performance applications as well. We discussed the CMOS process integration of a number of alternative gate dielectrics, including silicon nitride, Ta_2O_5, ZrO_2, Zr silicate and HfO_2. In a conventional CMOS process flow, the gate stack is subjected to high temperature anneals for dopant activation, therefore the thermal stability of alternative gate dielectrics is a key concern. Stability of the gate dielectrics depends on both the gate dielectric properties as well as the gate material, which may react with the gate dielectric at elevated temperatures. In current research, poly-Si gate technology is still preferred because of its maturity and the minimal changes needed in the conventional CMOS process. In terms of compatibility with poly-Si gate, silicon nitride and HfO_2 were shown to be very promising candidates. Both materials showed reasonably good results after anneals at or above 1000°C, with silicon nitride exhibiting better thermal stability. Although not as robust in high-temperature processes, HfO_2 offers the advantage of higher dielectric constant, which is associated with lower gate leakage current and better scalability. The transistor results based on these gate dielectrics showed that gate-first CMOS processes are practical for integrating these new gate dielectrics. However, the serious performance penalty due to the poly-Si gate depletion effect started to surface when the gate dielectric EOT approaches 10 Å.

The conflicting requirements between better dopant activation for poly-Si gate and boron penetration continues to be a problem with the HfO_2 gate dielectric, and is further aggravated by the thermal stability issues of HfO_2 [5.1]. Without drastic changes to the existing CMOS fabrication processes, the use of low Ge content poly-SiGe gate can improve the gate dopant activation and reduce boron penetration. In addition, it was found that poly-SiGe gate suppressed the interfacial layer growth for HfO_2/Si interface during high-temperature annealing, resulting in thinner EOT.

Engineering the gate dielectric-substrate interface was shown to be an effective approach to improving the gate dielectric properties. When deposited on nitrided Si-substrate surface, HfO_2 gate dielectric showed significant improvements in gate leakage current, breakdown voltage and EOT after high-temperature processes. Hot electron reliability of short-channel n-FETs with HfO_2 gate dielectric and nitrided interface was studied and compared with that of SiO_2 gate dielectric. The hot carrier degradation characteristics of the HfO_2 n-FETs were similar to those observed for silicon nitride gate dielectric or nitrided oxide interface. Nitridation of the substrate-gate dielectric interface improves the resistance to hot carrier damage, therefore results in better device lifetime than SiO_2 gate dielectric for a given substrate current. This shows that some intentional interfacial layer can be utilized to improve the device characteristics relatively independent of the upper layer gate dielectric.

Currently, the channel carrier mobility degradation is a very serious concern for most of the candidate high-*k* gate dielectrics. The degradation mechanism is essentially enhanced Coulombic scattering due to fixed charge and high-density interface traps [5.3]. Nitrided gate dielectric-substrate interface increases the interface trap density and degrades carrier mobility. There also seems to be a tradeoff between the EOT and the channel carrier mobilities. Thicker EOT allows the room for a sufficiently thick SiO_2-like interfacial layer, which improves interface quality and helps recover the channel carrier mobilities. We demonstrated good hole mobility on a Mo/RTCVD silicon nitride gate stack with 15 Å EOT. The channel carrier mobilities with HfO_2 gate dielectric, however, are still much lower than the universal mobility model. Very recently, it was reported that a high temperature forming gas anneal could significantly improve the channel carrier mobility for HfO_2 gate dielectric, especially in the low effective field region [5.2]. This kind of process optimization will continue to be an important area for high-*k* gate dielectrics research in the near future.

It should be possible to improve the channel carrier mobilities to an acceptable level, if not completely matching the universal model. Then a source for further performance improvement will be the elimination of the gate depletion effect, as the gate overdrive is continually reduced with lower power supply voltage. So metallic gate materials will be needed in the long term. In this work, molybdenum was studied with a number of different gate dielectrics, and was shown to be thermally stable in a gate-first MOSFET process. With ZrO_2 and HfO_2, Mo film with (110) orientation exhibits a high effective work-function, which makes it an appropriate gate material for bulk-Si p-FETs. Based on these results, a dual metal gate CMOS process was demonstrated using Mo and Ti as the gate material for p- and n-FETs, respectively. With relatively thick silicon nitride gate dielectric, the hole mobility matched the universal model and the electron mobility showed slight degradation than the model. To simplify the fabrication process, a CMOS process based on single-metal gate for both n- and p-FETs was developed, with the work-function adjustable by ion implantation to achieve appropriate threshold voltages. The effective work-function of the n-FET Mo gate was reduced by 0.42 eV using high-dose nitrogen implantation. This is not sufficient for bulk-Si CMOS applications, but can still be acceptable for lower channel doping device structures or as a means of achieving multiple threshold voltages.

5.2 Contributions

This dissertation research made several contributions to the screening and process integration of novel gate dielectric and gate electrode materials for CMOS technology. A number of issues related to gate stack scaling were addressed, with successful experimental demonstration of possible solutions, quite a few of which were the firsts in the literature.

The study of Ta_2O_5, silicon nitride and HfO_2 in CMOS integration confirmed the advantage of using gate dielectrics that are thermally stable with Si, which enables simpler integration into existing CMOS process. Sub-100 nm gate length transistors with the thinnest reported EOTs using silicon nitride or HfO_2 gate dielectrics were demonstrated in this work. These device demonstrations investigated the behaviors of the gate dielectrics when subjected to an advanced CMOS process, where the process module requirements are more stringent than typically used in MOS capacitor or long channel transistor processes. Therefore they provided a real-world evaluation of those new gate dielectric materials. The study of the hot carrier reliability of n-FETs using a HfO_2 gate dielectric with nitrided interface was the first of such evaluation of HfO_2. This positive result together with the other benefits of the nitrided interface suggests that engineering the gate dielectric-substrate interface offers great potential to improve the high-*k* gate stack properties.

The serious impact on thin-EOT device performance due to the poly-Si gate depletion effect was emphasized in this work, and near term and long term solutions were proposed. For the first time, a poly-SiGe gate was investigated with high-*k* gate dielectrics. In addition to improved gate dopant activation compared to a poly-Si gate, the poly-SiGe gate was found to suppress the interfacial dielectric layer formation during high-temperature processes. Understanding of this phenomenon can provide useful information on controlling the interface between high-*k* gate dielectrics and the substrate.

Metal gate CMOS integration was focused on the gate-first approach. The dual metal gate CMOS was the first report integrating different metal gate electrodes for n- and p-FETs. A CMOS process using a single Mo gate with adjustable gate work-function was the first to demonstrate a significant effective work-function adjustment range that is practical for real CMOS application. A Mo gate was studied with a number of alternative gate dielectrics, and the first direct experimental evidence of the effects of gate dielectric on the observed effective gate work-function was reported. This leads to the understanding of another constraint for selecting the gate material for high-*k* gate dielectrics. Although Mo was studied for MOS applications more than thirty years ago, and was abandoned, the encouraging findings in a new context presented in this work make it compelling to revisit the Mo gate technology.

5.3 Recommendations for future work

Currently, process modules based on novel gate stack materials are still in a research stage, and a few critical issues stand out as the main barriers to their acceptance. In this section, some possible continuations of this thesis work that are closely related to those top priority problems are proposed. The proposed future works are in two general directions: device performance improvement and reliability.

5.3.1 Channel carrier mobility improvement for high-*k* gate dielectric

It is obvious that the mobility degradation due to high-*k* gate dielectrics seriously compromises the transistor performance. Currently there is evidence showing that enhanced Coulombic scattering due to fixed charge in some high-*k* gate dielectrics [5.3], [5.4] is a major cause of the low mobility. In addition, the poor interface quality in some high-*k* gate dielectric processes also significantly degrades the carrier mobility [5.5]. Potential solutions are:

1). Interface engineering for high-*k* gate dielectrics

The interface property of high-*k* gate dielectrics on the Si substrate can be effectively modified by a very thin layer of dielectric film with different chemical composition. An oxygen-rich interface is known to result in a SiO_2-like interface, and recovers the carrier mobilities. Such a technique has been demonstrated for silicon nitride gate dielectric to achieve a thin EOT and good carrier mobilities [5.6], but has not been applied successfully to high-*k* gate dielectrics such as HfO_2. In addition, the flatband voltage shift due to fixed charge for some high-*k* gate dielectrics can be reduced by a thin chemical oxide between the Si substrate and the high-*k* layer [5.7]. Thermodynamically the existence of such a SiO_X interfacial layer is not favored with a high-*k* layer on the top, therefore research is needed to improve the stability and process control of the interfacial layer. One example is that the effects of the poly-SiGe gate need to be clarified, and the findings will provide valuable insight into the interfacial layer growth.

2). Multi-component high-*k* gate dielectrics

Although HfO_2 is viewed as one of the most stable high-*k* gate dielectrics, pure HfO_2 still does not meet the CMOS process integration requirements, and further improvements are possible through the incorporation of other elements. It has been reported that doping Al_2O_3 films with Zr or Si can reduce electrical defects [5.8], and nitrogen doping in Al_2O_3 can reduce interface trap density and hysteresis [5.9]. Some other interesting reports of multi-component dielectrics include nitrogen-incorporated Zr or Hf oxides [5.10][5.11], Zr or Hf aluminates [5.7][5.12], and Hf-Si oxynitride [5.13]. These results showed the great potential of modifying the high-*k* dielectric film properties by introducing other elements to the dielectric film. At this point, however, the choice and quantity of the dopants are mostly based on empirical approaches. Systematic studies of these doping effects are needed so that some theoretical guidance may be established to get more advantages out of this technique.

3). Epitaxial high-*k* gate dielectrics on conventional or strained Si substrates

In principle, the most effective way to reduce scattering due to the dielectric-substrate interface is to use an epitaxial gate dielectric on a Si substrate. With a high quality single-crystalline gate dielectric, the carrier mobility will be significantly improved due to the much reduced surface roughness scattering. With very low defect density at the gate dielectric and substrate boundary, scattering caused by the fixed charge and interface states will also be minimized. At this point, there are very limited reports of such kinds of gate dielectrics applied to MOSFETs. An example is the $SrTiO_3/Ba_{0.75}Sr_{0.25}O$ hetero-junction gate dielectric on a Si substrate [5.14]. While the lattice matching with conventional Si substrate is an extra requirement that narrows the choices of dielectrics, this constraint is actually relaxed for strained Si substrates. Recently, mobility enhancement using a Si channel with biaxial tensile strain attracts serious consideration as a way to achieve better Si n-MOSFETs performance [5.15]. Many high-*k* metal oxides have slightly larger lattice constant than Si, so more candidates are available for epitaxial growth on strained Si [5.16].

Actual implementation of this concept is challenging. It requires deposition process controlled with atomic precision to achieve the perfect transition from covalent silicon lattice to the ionic oxide lattice. In view of the tremendous potential benefits of this approach, it is worth pursuing further.

5.3.2 Reliability of high-*k* dielectrics and metal gate stack

Time dependent dielectric breakdown (TDDB) reliability is an extremely important aspect of gate dielectrics. The highly reliable thermal SiO_2 is one of the major reasons for the predominance of Si as the substrate for MOS devices. The physics of TDDB has been a very challenging problem. As a matter of fact, while the SiO_2 gate dielectric is quickly approaching the end of its usefulness in CMOS, there are still ongoing debates on the true mechanisms of SiO_2 breakdown after over thirty years of investigation. For the gate stacks based on novel materials, two new problems arise.

1). TDDB reliability of high-*k* gate dielectrics

For TDDB of high-*k* dielectrics, there have been only limited experimental and little theoretical studies. Intrinsically high-*k* gate dielectrics have smaller bandgap, therefore lower breakdown field than SiO_2. Although this could be compensated by the larger physical thickness of the high-*k* gate dielectrics, reliability may still be a potential problem. In addition, unlike thermal SiO_2, the high-*k* gate dielectrics are likely to be a multi-layered structure, with a special bottom layer to improve the interface with the substrate. Therefore the breakdown process is more complicated, and existing models for SiO_2 must be extended to address the multi-layer dielectric breakdown. If a complete understanding of the high-*k* dielectric breakdown cannot be achieved, a thorough experimental confirmation of good TDDB reliability will be a precondition for high-*k* dielectrics' acceptance to CMOS manufacturing. This aspect needs to be included as a standard routine in the process development and integration of high-*k* gate dielectrics.

2). The impact of metal gate on gate dielectric reliability

A metal gate can affect the gate dielectric reliability in two aspects. Depending on the specific process, metal deposition may cause damage to the gate dielectric, such as sputtering damage, or defects created by diffusion/reaction during high temperature processes. A very recent work reported different charge trapping behaviors in ultra-thin HfO_2 using different metal gate electrodes, suggesting the effects of mechanical stress and damage to the gate dielectric introduced by the gate deposition process[5.17]. In addition, the effects of a metal gate on gate dielectric TDDB has not been studied much, and it can be a difficult problem given the complexity of SiO_2 breakdown theory. According to the anode-hole-injection model, the damage to the gate dielectric is caused by the holes created in the gate electrode by impact ionization and injection into the gate dielectric by the high electric field [5.18]. So the TDDB mechanism for a metal gate stack can be different from the poly-Si gate case. In the initial study, using metal gates on SiO_2 can help focus on the role of the metal gate, and eventually, metal gate/high-*k* stack structure will need to be studied.

5.4 References

[5.1] 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 MOSFETs”, *Symposium on VLSI Technology*, pp. 131-132, June 2001.

[5.2] K. Onishi, C. S. Kang, R. Choi, H.-J. Cho, S. Gopalan, R. Nieh, S. Krishnan, and J. C. Lee, "Effects of High-Temperature Forming Gas Anneal on HfO_2 MOSFET Performance", *Symposium on VLSI Technology*, pp. 22-23, June 2002.

[5.3] K. Torii, Y. Shimamoto, S. Saito, O. Tonomura, M. Hiratani, Y. Manabe, M. Caymax, J. W. Maes, "The mechanism of mobility degradation in MISFETs with Al_2O_3 gate dielectric", *Symposium on VLSI Technology*, pp. 188-189, June 2002.

[5.4] T. Yamaguchi, H. Satake, N. Fukushima, "Degradation of current drivability by the increase of Zr concentrations in Zr-silicate MISFET", *International Electron Devices Meeting*, pp. 663-666, Dec. 2001.

[5.5] K. Onishi, S. C. S. Kang, R. Choi, H.-J. Cho, S. Gopalan, R. Nieh, E. Dharmarajan, J. C. Lee, "Reliability characteristics, including NBTI, of polysilicon gate HfO_2 MOSFETs", *International Electron Devices Meeting*, pp. 659-662, Dec. 2001.

[5.6] S. Tsujikawa, T. Mine, Y. Shimamoto, O. Tonomura, R. Tsuchiya, K. Ohnishi, H. Hamamura, K. Torii, T. Onai, J. Yugami, "An ultra-thin silicon nitride gate dielectric with oxygen-enriched interface (OI-SiN) for CMOS with EOT of 0.9 nm and beyond", *Symposium on VLSI Technology*, pp. 202-203, June 2002.

[5.7] 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, 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", pp. 88-89, *Symposium on VLSI Technology*, pp. 202-203, June 2002.

[5.8] L. Manchanda, W. H. Lee, J. E. Bower, F. H. Baumann, W. L. Brown, C. J. Case, R. C. Keller, Y. O. Kim, E. J. Laskowski, M. D. Morris, R. L. Opila, P. J. Silverman, R. W. Sorsch, and G. R. Weber, "Gate quality doped high-k films for

CMOS beyond 100 nm: 3-10nm Al_2O_3 with low leakage and low interface states", *International Electron Devices Meeting*, pp. 605-608, Dec. 1998.

[5.9] Y. Tanida, Y. Tamura, S. Miyagaki, M. Yamaguchi, C. Yoshida, Y. Sugiyama, and Tanaka, "Effect of in-situ nitrogen doping into MOCVD-grown Al_2O_3 to improve electrical characteristics of MOSFETs with polysilicon gate", *Symposium on VLSI Technology*, pp. 190-191, June 2002.

[5.10] R. Nieh, S. Krishnan, H.-J. Cho, C. S. Kang, D. Gopalan, K. Onishi, R. Choi, and J. C. lee, "Comparison between ultra-thin ZrO_2 and ZrO_XN_Y gate dielectrics in TaN or poly-gated NMOSCAP and NMOSFET devices", *Symposium on VLSI Technology*, pp. 186-187, June 2002.

[5.11] C. S. Kang, H.-J. Cho, K. Onishi, R. Choi, R. Nieh, S. Gopalan, 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)", *Symposium on VLSI Technology*, pp. 146-147, June 2002.

[5.12] P. J. Chen, E. Cartier, R. J. Carter, T. Kauerauf, C. Zhao, J. Petry, V. Cosnier, Z. Xu, A. Kerber, W. Tsai, E. Young, S. Kubicek, M. Caymax, W Vandervorst, S. De Gendt, M. Heyns, M. Copel, W. F. A. Besling, P. Bajolet, J. Maes, "Thermal stability and scalability of Zr-aluminate-based high-k gate stacks", *Symposium on VLSI Technology*, pp. 192-193, June 2002.

[5.13] 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, L. Colombo, "Advanced CMOS transistors with a novel HfSiON gate dielectric", *Symposium on VLSI Technology*, pp. 148-149, June 2002.

[5.14] R. A. McKee, F. J. Walker, M. F. Chisholm, “Physical structure and inversion charge at a semiconductor interface with a crystalline oxide”, *Science*, Vol. 293, No. 5529, pp. 468-471, July 2001.

[5.15] N. Sugii, D. Hisamoto, K. Washio, N. Yokoyama, and S. Kimura, “Enhanced performance of strained-Si MOSFETs on CMP SiGe virtual substrate”, *International Electron Devices Meeting*, pp. 737-740, Dec. 2000.

[5.16] Private discussion with Prof. D. Schlom of Pennsylvania State University.

[5.17] W. J. Zhu, T. P. Ma, S. Zafar, and T. Tamagawa, “Charge trapping in ultrathin hafnium oxide”, *IEEE Electron Device Letters*, Vol. 23, No. 10, pp. 597-599, Oct. 2002.

[5.18] Y.-C. Yeo, Q. Lu, and C. Hu, "Gate oxide reliability: anode hole injection model and its applications", *Oxide Reliability A Summary of Silicon Oxide Wearout, Breakdown, and Reliability*, edited by D. J. Dumin, World Scientific, Singapore, ISBN 981-02-4842-3