STUDY ON APPLICATION OF HIGH-*K* DIELECTRIC MATERIALS FOR DISCRETE CHARGE STORAGE MEMORY

WANG YING QIAN

NATIONAL UNIVERSITY OF SINGAPORE

2006

STUDY ON APPLICATION OF HIGH-*K* DIELECTRIC MATERIALS FOR DISCRETE CHARGE STORAGE MEMORY

WANG YING QIAN

(M. Eng., Tsinghua University, China)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2006

Acknowledgement

I would like to express my deep and sincere appreciation to my thesis advisors Prof Yoo Won Jong and Dr Yeo Yee Chia, who provided constant support and invaluable guidance during this work. Their constructive suggestions and scientific excitement have been truly inspirational; their kindness and patience made all my time of research pleasant.

I am also extremely grateful to A/Prof Samudra Ganesh for his continual support and numerous valuable suggestions throughout my research work. I am very grateful to Prof Albert Chin in National Chiao Tung University, Taiwan for his inspiring discussion about the device engineering. My best regards are given to all other professors — Byung-Jin Cho, Mingfu Li, Chunxiang Zhu, Sungjoo Lee in my lab and Prof Dim-Lee Kwong in Institute of Microelectronics for their help and instruction. Many thanks to Dr An-Yan Du in Institute of Microelectronics for the TEM works.

My special thanks to my friends in SNDL, Chen Jing Hao, Hwang Wan Sik, Tan Kian Ming, Zerlinda, Yiang Kok Yong, Shen Chen, Zhang Qing Chun, Kim Sun Jung, Ren Chi, Li Rui, Wang Xinpeng, Wu Nan, Gao Fei, Yu Xiongfei, Chen Jingde, He Wei for their kind assistance in my research and many other aspects. Thanks for make my study life enjoyable and memorable. I am indebted to all the stuff of my lab; without their painstaking maintenance of the cleanroom this thesis will not be accomplished.

Finally, I would like the express my earnest appreciation to my parents. Thanks for your understanding and your loving support.

Abstract

The conventional flash devices use a continuous floating gate to store charges. This floating gate structure is very sensitive to the local defect of the tunnel oxide because all charges can be lost through a defect path，making the scaling of tunnel oxide the largest challenge for device scaling. Discrete charge storage memories including nanocrystal (NC) memory and SONOS type memory are the most promising candidates to substitute for floating gate memory. Thanks to their isolated charge storage nodes, the discrete charge storage memories are immune to local defect related leakages, therefore providing aggressive scaling capability. In this thesis, the following issues are addressed: formation of NCs, application of high-*k* dielectric materials for NCs memory and SONOS type memory device, and optimization of the SONOS cell structure.

Self-assembled Ge NCs are formed on HfO_2 and HfAlO by CVD with density of 10^{11} cm^{-2}. Additionally, Ge NCs with diameter about 5-10 nm embedded in HfAlO high-*k* dielectric are obtained by cosputtering method. The Ge NCs are thermally stable in HfAlO matrix. A nonvolatile memory device employing Ge NCs embedded in HfAlO dielectric exhibits excellent memory performance.

HfO_2 NCs are developed by annealing the HfSiO film at above 900^oC. $Hf_{0.5}Si_{0.5}O_2$ film containing HfO_2-$Hf_xSi_{1-x}O_2$ dual phase as a trapping layer is found to provide a faster programming speed at a lower programming voltage than Si_3N_4 film because of its higher dielectric constant and higher trap efficiency. Meanwhile, the HfO_2-$Hf_xSi_{1-x}O_2$ film also provides better retention property than HfO_2 because the presence of the amorphous phase $Hf_xSi_{1-x}O_2$ suppresses formation of grain boundary effectively thereby reducing lateral migration.

For the further optimization of the cell structure, besides the phase separated HfSiO trapping layer, the high-*k* tunneling and blocking oxide HfAlO and high work function gate electrode IrO_2 are integrated. Combining advantages of high-*k* HfAlO, good trapping capability of HfSiO, and high work function of the IrO_2 gate, the device with IrO_2/HfAlO/HfSiO/HfAlO gate stack achieves excellent retention with 10-year memory window decay ratio within 18%, high erasing speed with threshold voltage shift of 3V within 0.5ms at V_g = -12V, and additionally, lower operation voltage and lower reading voltage than other contending device structures.

Another optimizing SONOS type memory structure for NAND Flash application is explored by using the dual tunneling layer (Si_3N_4/SiO_2) along with a high-*k* HfO_2 charge storage layer. Combining advantages of the high trapping efficiency of high-*k* materials and the enhanced charge injection from the substrate through the dual tunneling layer, the device achieves fast program/erase speed and large memory window. The device demonstrates the excellent retention due to the physically thick dual tunneling layer and also the improved endurance without the increase of programming V_{th} throughout the cyclic test in comparison with SONOS Flash memory devices using a Si_3N_4 trapping layer.

Table of Contents

List of Figures

List of Tables

List of Symbols

A	area
α_g	capacitance coupling ratio
C	capacitance (F)
d	thickness
E	electrical field (V/cm)
h	Planck's constant (6.626×10^{-34} J s)
I	current (A)
I_d	drain current (A)
I_g	gate leakage current (A)
J	current density (A/cm^2)
L	channel length (μm)
m_{ox}	effective mass (kg)
Q	charge (C)
T	temperature
t	time
V	voltage (V)
V_d	drain voltage (V)
V_g	gate voltage (V)
V_{fb}	flatband voltage (V)
V_{th}	threshold voltage (V)
ε_0	permittivity of free space (8.854×10^{-14} F/cm)
ϕ_B	barrier height (eV)
κ	dielectric constant
ΔE_c	conduction band offset to Si
ΔE_v	valence band offset to Si
ΔG	Gibbs free energy change
ΔH	enthalpy changes
ΔS	entropy changes

List of Abbreviations

AFM	Atomic Force Microscopy
ALD	Atomic layer Deposition
CHE	Channel Hot Electron
CSL	Charge Storage layer
CVD	Chemical Vapor Depostion
DIBL	Drain-Induced-Barrier-Lowing
DPHSO	Dual Phase HfO_2-$Hf_xSi_{1-x}O_2$
DRAM	Dynamic Random Access Memory
DT	Direct Tunneling
DTL	Dual Tunneling Layer consisting of Si_3N_4/SiO_2
EDX	Energy Dispersive X-ray
EEPROM	Electrically Erasable and Programmable Read Only Memories
EOT	Equivalent Oxide Thickness
EPROM	Electrically Programmable Read Only Memories
F-N	Fowler-Nordheim
HRTEM	High Resolution Transmission Electron Microscopy
IPD	Interpoly Dielectric
ITRS	International Technology Roadmap for Semiconductors

LPCVD	Low Pressure Chemical Vapor Deposition
MONOS	Metal/ Oxide / Nitride / Oxide / Silicon
NC	Nanocrystal
PDA	Post Deposition Anneal
ROM	Read Only Memories
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectroscopy
SONOS	Silicon / Oxide / Nitride / Oxide / Silicon
SRAM	Static Random Access Memory
SS	Sub-threshold Swing
STEM	Scanning Transmission Electron Microscopy (STEM)
TEM	Transmission Electron Microscopy
XPS	X-ray Photoelectron Spectroscopy
XRD	X-ray Diffraction

Chapter 1

Introduction

1.1 Introduction to Semiconductor Memory Devices

Semiconductor memory is an essential component in the current electrical system, and its application covers cell phone, consumer, automotive, and computer systems. In the past few years, semiconductor memory occupied above 20% of the total semiconductor market and this percentage tends to increase continuously and is going to be 30% in near future, as shown in Fig. 1.1 [1].

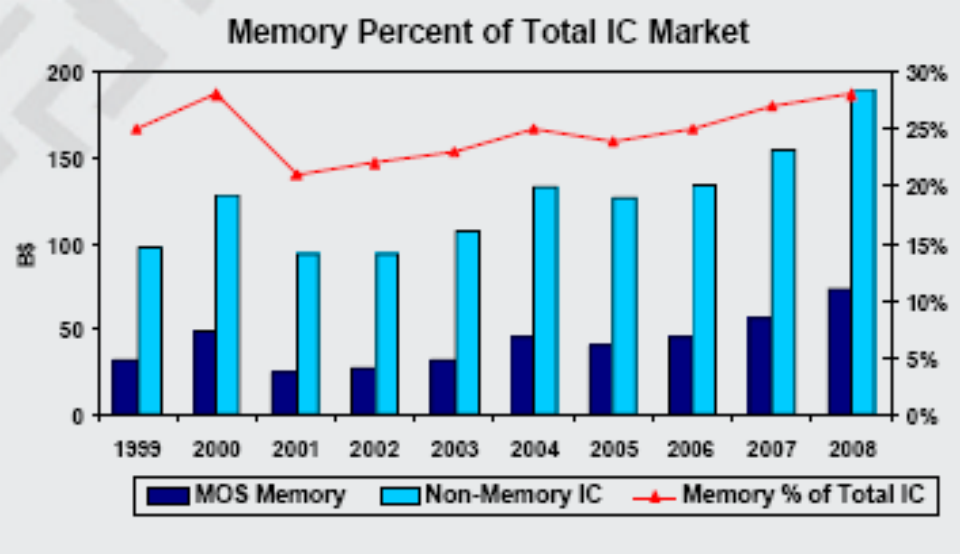


Figure 1.1. Revenues of semiconductor market versus year. The top line is the memory percentage of the total market.

There are varieties of semiconductor memories and their category is shown in Fig. 1.2. Basically, semiconductor memories are divided into two groups: volatile and nonvolatile memories. Volatile memories need to be refreshed constantly because they lose the stored information once the power supply is off. Nonvolatile memories can keep the data even without power supply [2].

Volatile memories include Static Random Access Memory (SRAM) and Dynamic Radom Access Memory (DRAM). SRAM retains its contents as long as electrical power is applied to the chip. In comparison, DRAM has an extremely short data lifetime typically about 100 ms. The data in DRAM are kept by refreshing it periodically. The advantage of SRAM is that it offers the fastest write/read speed (8 ns) among all memories. In contrast, DRAM is much slower (50 ns). However, DRAM is used more extensively than SRAM because of its attractive low cost-per-byte.

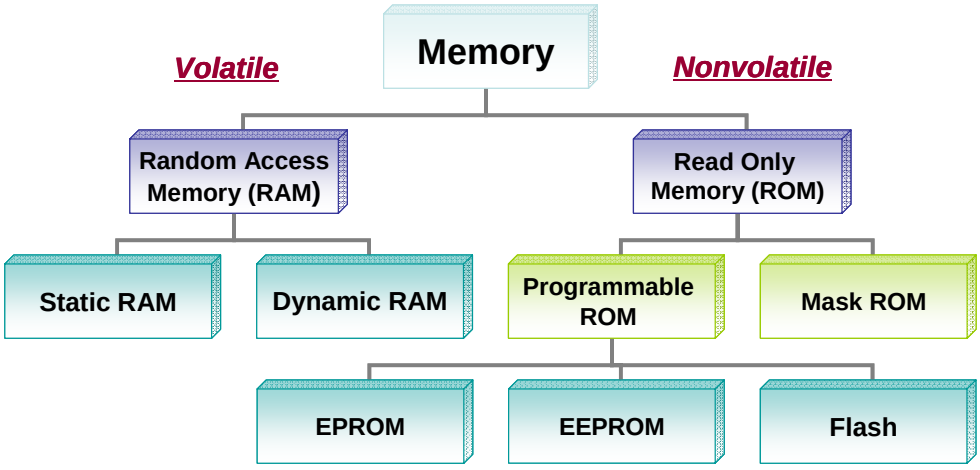


Figure 1.2. Branches of semiconductor memory family.

Nonvolatile memories include mask Read Only Memories (ROM) and reprogrammable memories such as Electrically Programmable Read Only Memories (EPROM), Electrically Erasable and Programmable Read Only Memories (EEPROM) and flash.

The mask ROM were devices that were programmed when they were manufactured at the factory with a special mask. The mask ROM was further developed to a progressive generation Programmable ROM (PROM) which consists of an array of fuses and can only be programmed once by a type of special equipment named programmer. EPROM is a reprogrammable memory device in that it can be erased by using an ultraviolet light source. EPROM is programmed using voltage rather than current as the PROM uses. The erasure ability of EPROM enables it to be reused and makes it an important part of the software development and testing process, although they are more expensive than PROM. EEPROM is a different device from EPROMs in that the erasure of EEPROM is accomplished electrically, but not by UV source. Each byte of the EEPROM can be written and erased separately, and the data in EEPROM can be remained as long as needed. EEPROM is actually a device combining features of both RAM and ROM because it can be read and written like RAM; besides, it can also maintain their contents when electrical power is off like ROM.

Flash is similar to an EEPROM except that flash are erased by blocks of different sizes (256 bytes to 16KB) while a regular EEPROM can be erased per single byte as mentioned earlier. Considering all the features of the memory devices, flash memory devices compromise the flexibility and cost best. As a result, the market of the flash

memory increased dramatically, as shown in Fig. 1.3 [1]. It was forecasted that the flash market will reach the size of the DRAM market in a few years.

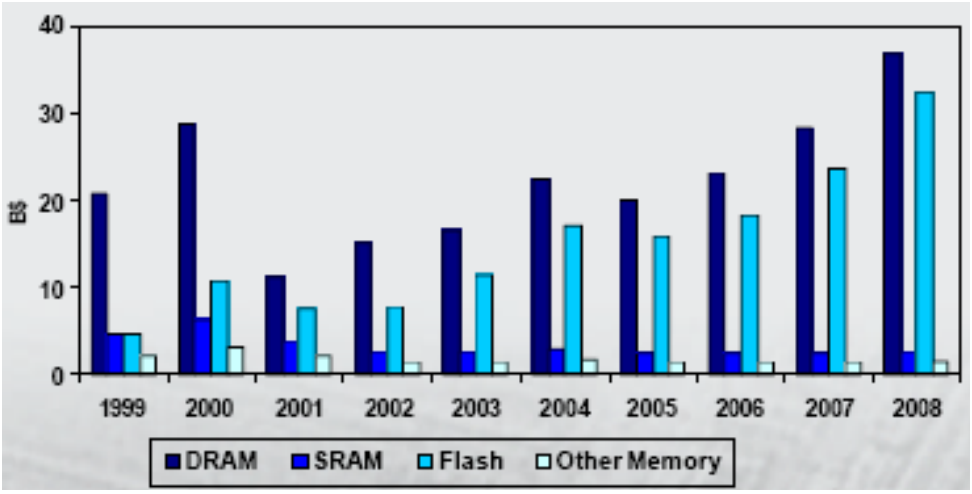


Figure 1.3. Revenues of memory market versus year.

1.2 Operation Mechanisms and Architectures of Flash

The conventional flash cell consists of one transistor. This transistor is simply a MOSEFT transistor except a floating-gate existing between two dielectric layers: tunneling oxide and interpoly dielectric (IPD). The schematic cross section of a floating gate device is shown in Fig. 1.4. When charges (electron in floating gate memory) are injected to the floating gate, the threshold voltage of the memory MOSFET will be modified and shift of the threshold voltage can be expressed by:

$$\Delta V_T = V_T - V_{T0} = -\overline{Q} / C_{FC}$$

where

$\overline{Q}$ is the amount of charges injected;

V_{T0} is the threshold voltage when $\overline{Q}$ = 0;

C_{FC} is the capacitance between floating gate and control gate [3].

Thus, if we apply a sense voltage which is between V_{T0} and V_T, the memory state can be determined by the measured current level, as shown in Fig. 1.5 [1]. The state of the lower threshold voltage corresponds to logic "1" state due to the high current sensed and the state of the higher threshold voltage corresponds to logic "0" state due to the low current sensed.

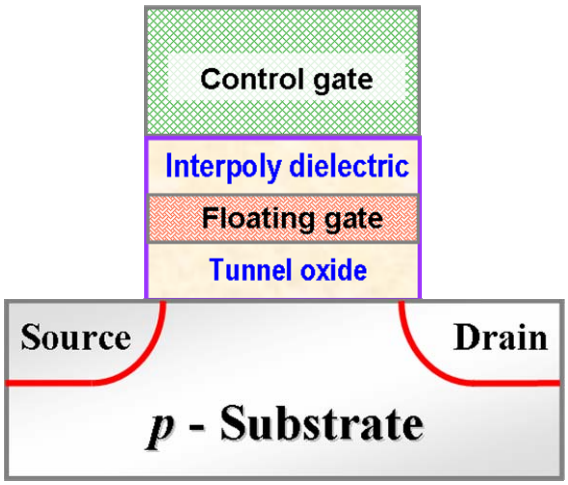


Figure 1.4. Schematic cross section of a flash cell transistor.

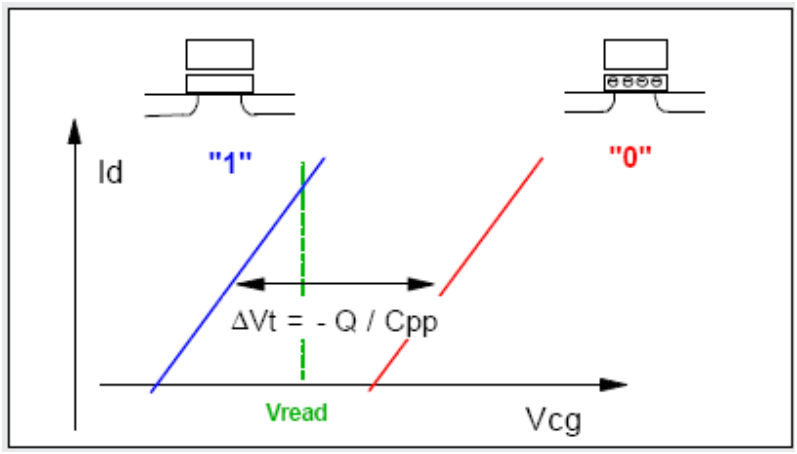


Figure 1.5. Reading scheme of the flash memory.

There are two main mechanisms used for injecting charges to the floating gate: channel hot electron (CHE) and Fowler-Nordheim (F-N). F-N tunneling is also used to remove the charges from the floating gate. The details of these two mechanisms are described blow.

1. Channel hot electron (CHE)

As shown in Fig. 1.6 [4], when electrons travel from source to drain under an electrical field larger than 100 kV/cm, they are “heated” and some of them can gain enough energy so that their kinetic energy is higher than the potential barrier between the oxide and silicon. If some of these electrons direct toward barrier and the electric field in the oxide is attracting them to the floating gate, they can pass the barrier to reach floating gate and can charge the floating gate to a more negative potential. A typical terminal condition of CHE injection is also indicated in Fig. 1.6, where the source and substrate are grounded and the gate is applied with a larger positive voltage than the drain.

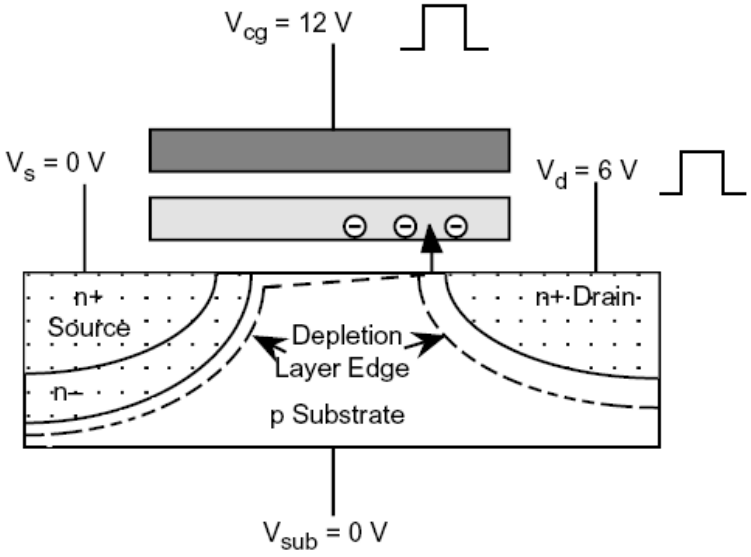


Figure 1.6. Schematic cross section of a floating gate cell when using channel hot electron injection for programming.

2. Fowler-Nordheim (F-N) tunneling

F-N tunneling is a quantum-mechanical tunnel which is induced by a high electric field. When a large electric field is applied across the SiO_2, as shown in Fig. 1.7, the electrons in the silicon conduction band see a triangular energy barrier with a width dependent on the electric field. When the width is small enough the electron can tunnel through the barrier from the silicon conduction band into the oxide conduction band without destroying SiO_2 dielectric properties [6]. F-N tunneling current is adequate enough for memory devices to inject electrons into the floating gate or push electrons out of the floating gate.

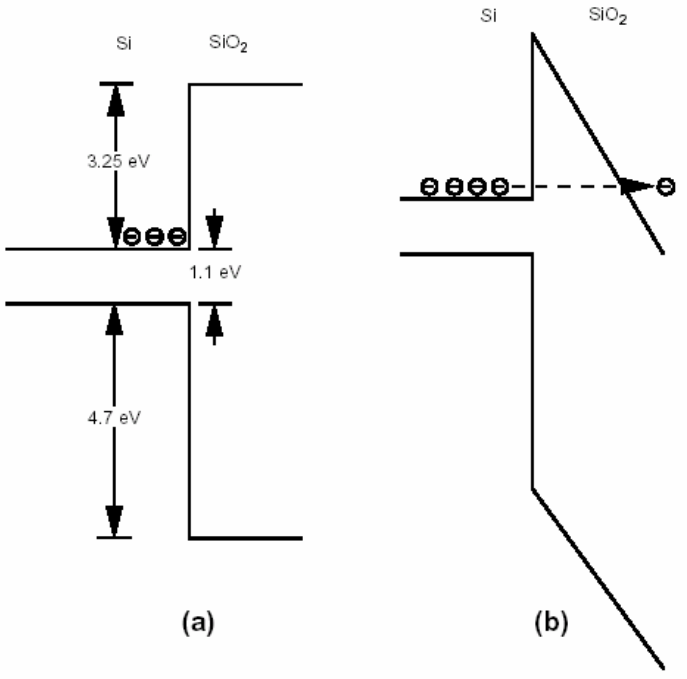
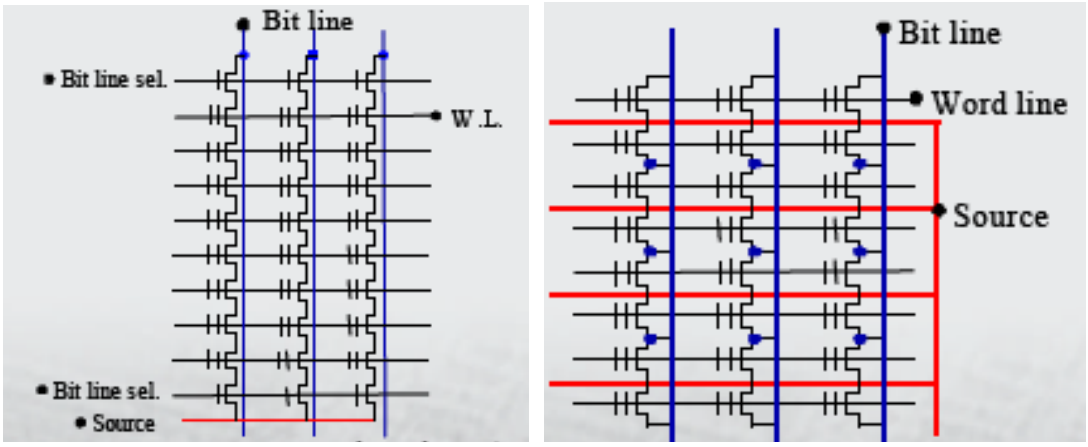


Figure 1.7. Band diagram of Si and SiO_2 interface (a) with no electric field and (b) with a strong applied electric field, whereby electrons tunnel through the triangular barrier.

The tunneling mechanisms used are actually related to the architectures of flash circuit. The market of flash memory is dominated by two types of architectures: NAND and NOR, as illustrated in Fig. 1.8 [1]. NAND uses F-N for both programming and erasing, and NOR uses CHE for programming and F-N for erasing.

In NAND flash, multiple cells (16 or 32 cells) are connected in series; therefore source and drain contacts are not needed for each cell. As a result, NAND has the smallest cell size among current semiconductor memories. NAND flash features high cell densities, high capacity, fast program and erase rates, and easiness of scaling down. In contrast, the advantages of NOR flash lies in the random access and byte write capability.

According to their features, NAND is used for high capacity data storage application, while NOR is suitable for small capacities application such as code storage and execution.



(a) **(b)**

Figure 1.8. Architecture of (a) NAND and (b) NOR flash.

1.3 Scaling Limitation of Floating Gate Flash Memories

Since the integrated circuit was invented, the semiconductor devices have been scaling down following the Moore's Law which states the transistor density of integrated circuit doubles every couple of years. “In the past, flash devices tended to lag behind the current CMOS technology generation, but that delay no longer exists”, stated in international technology roadmap for semiconductors (ITRS) 2005 [5]. There are still large challenges for the further scaling of flash devices which is related to the reduction of cell area and the thickness of tunneling oxide and interpoly dielectric (IPD).

One of the key challenges for scaling of flash memory is the scaling of tunneling oxide thickness because of the tradeoff between the program/erase performance and the

charge retention properties. The thin tunneling oxide brings high program/erase performance but causes degradation of charge retention properties, while the thick tunnel oxide guarantees retention properties but degrades write /erase performances.

Scaling of IPD thickness is another challenge because it needs to be thick to assure the retention and is required to be thin to keep the capacitance coupling ratio α_g which is defined by C_{FC}/C_{FS}, where C_{FC} is control gate to floating gate capacitance and C_{FS} is total floating gate to source, drain and substrate capacitance. α_g represents the ratio of the voltage drop through the tunnel oxide and IPD. α_g should be large enough so that the voltage drop through the tunneling oxide is large enough to achieve high charge exchange speed between the floating gate and channel. Meanwhile, α_g can not be too large considering the impact of charges in floating gate on the channel is reversely proportional to C_{FC}, i.e., $\Delta V_T = -\overline{Q}/C_{FC}$. The optimum range of α_g is kept to be 0.6-0.7 for the optimum program/erase performance. In current architecture, part of the α_g comes from the overlap area of floating gate and control gate along the sidewalls, as shown in Fig. 1.9 [5]. However, this part will become difficult when two adjacent floating gates are too close to allow the overlap between control gate and floating gate beyond 45-40nm technology generation. Therefore a strong reduction of IPD thickness will be required, which will bring a significant reduction of retention properties, making the scaling of IPD beyond 45-40nm technology generation more challenging.

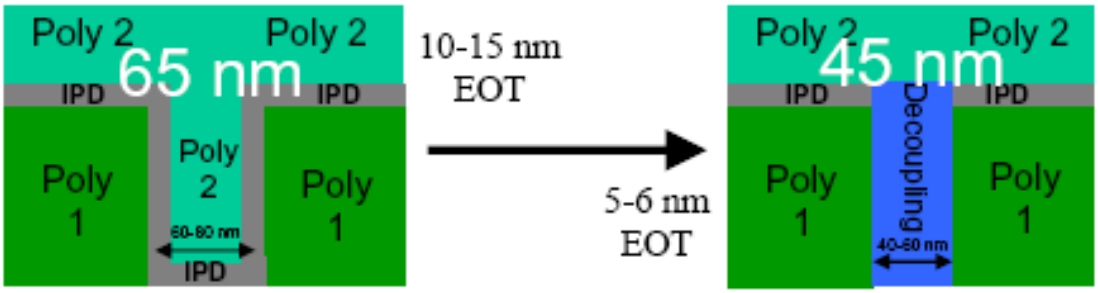


Figure 1.9. Comparison of coupling between the control gate (poly 2) and floating gate (poly 1) between 65 nm and 45nm technology node.

There are also some intrinsic limitations of flash memory scaling. First of all, for NOR flash which uses CHE programming, the drain voltage scaling is limited by the barrier between Si and SiO_2, given 3.2 eV [7]. Secondly, the large EOT (about 20 nm referring to the data in Table 1.1 in the year 2006) of the total gate dielectric in the floating gate cell and the high drain bias (typical 2 V in the reading mode and 4.5 V in the programming mode for NOR flash) limit scaling of the effective gate length L_{eff} because of the punchthrough and drain-induced-barrier-lowing (DIBL) effect when devices scale down. In addition, the transistors in peripheral circuit which provide the high voltage required by program/erase of the memory cell occupy a large part of the chip area, but the scaling of these transistors lags behind the memory cell because the program/erase voltage of memory cell has not decreased for last few technology generations [8]. Furthermore, the coupling of the unrelated floating gate and the poly word line also increases when the spacing between cells is decreased, these will cause the interference of cell threshold voltage and widen its distribution [9]. Targets of the flash memory nodes in the near-term years are summarized in Table 1.1 [5], according to ITRS 2005.

Table 1.1. Flash Nonvolatile memory technology requirement (ITRS 2005)

Year of Production	2006	2007	2008	2009	2010	2011	2012	2013
DRAM 1/2 Pitch (nm) contacted	70	65	57	50	45	40	36	32
MPU/ASIC metal (M1) 1/2 Pitch(nm)(contacted)	78	68	59	52	45	40	36	32
Flash technology generation NOR/NAND – F (nm)	70/65	65/55	57/50	50/45	45/40	40/35	35/32	32/28
Flash NOR tunnel oxide thickness (EOT-nm)	8-9	8-9	8-9	8-9	8	8	8	8
Flash NAND tunnel oxide thickness (EOT-nm)	7-8	6-7	6-7	6-7	6-7	6-7	6-7	6-7
Flash program/erase window min DVT SLC/MLC (V)	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4	1.5/2.4
Flash NOR interpoly dielectric thickness (EOT-nm)	13-15	13-15	13-15	13-15	◆6-13	◆6-13	◆6-13	4-6
Flash NAND interpoly dielectric thickness (EOT-nm)	13-15	10-13	10-13	10-13	◆5-12	◆5-12	◆5-12	4-6
Tunnel/ Interpoly max leakage current (A) at 2V for 10 years retention	1E-24	5E-25	5E-25	5E-25	2.5E-25	2.5E-25	2.5E-25	1.3E-25
Flash coupling ratio	0.6-0.7	0.6-0.7	0.6-0.7	0.6-0.7	0.6-0.7	0.6-0.7	0.6-0.7	0.6-0.7
Flash endurance (erase/write cycles)	1E+5	1E+5	1E+5	1E+5	1E+6	1E+6	1E+6	1E+6
Flash nonvolatile data retention	10-20	10-20	10-20	10-20	10-20	10-20	10-20	20
MPU physical gate length (nm)	28	25	22	20	18	16	14	13
Flash NOR typical L_g-stack (physical –μm)	0.135	0.13	0.12	0.12	0.11	0.11	0.1	0.1

Manufacturable solutions exist, and are being optimized
Manufacturable solutions are known
Interim solutions are known ◆
Manufacturable solutions are NOT known

1.4 Scope of Our Project

According to the aforementioned discussion, the aggressive scaling of flash memory device is challenging and extensive researches are required on exploration of new materials, new technology and new memory structure for flash memory application.

The alternative memory structures other than floating gate structures which are actively explored include SONOS (Silicon / Oxide / Nitride / Oxide / Silicon), FeRAM (Ferroelectric RAM), MRAM (Magnetic Random Access Memory) and PCRAM (Phase Change Random Access Memory). The FeRAM utilizes the positive or negative remnant polarization charge states of ferroelectric capacitor to record either data “1” or “0”. MRAM utilizes the electrical resistance that depends on spin state of a magnetic material to represent two states of memory device. PCRAM uses GST (GeSbTe) chalcogenide alloys as a memory element, and it senses the resistance difference of the amorphous and crystal state which can be changed by the heating current. However, both the MRAM and FeRAM need one transistor and another component in a cell, and PCRAM is not mature enough.

The discrete charge storage memory is a single transistor memory cell structure. The single transistor can be nanocrystal (NC) floating gate transistor or SONOS transistor. In these memories, the charge storage nodes are separated and independent each other, thus the charges stored are isolated and cannot easily redistribute amongst themselves. In conventional floating gate flash memory, if there is one defect chain across tunnel oxide, all of the charges in the floating gate can leak to the channel or source / drain through this leakage path. However, this leakage path can be effectively minimized by utilizing the discrete charge storage nodes, in which only charges stored directly above

the defect nodes will be drained. Hence the memory is more immune to defect related leakages and thinner tunnel oxide can be used to improve program speed or reduce the operation voltage [10, 11]. Another attractive advantage of the discrete charge storage memory is its high compatibility with the conventional semiconductor process. Therefore, the discrete charge storage memory is the most promising technology, especially in the 45 nm node.

In this dissertation, we will focus on the study of nanocrystal (NC) memory and SONOS type memory device. Especially, application of high-*k* dielectric materials and optimization of the cell structure will be investigated. The focus of the memory device will be on speed up, power reduction and reliable retention and endurance.

Formation of NCs with desired size and uniform distribution is the key technology for the fabrication of NC memory. However, studies on formation of NCs with high-*k* dielectrics are quite limited. Our efforts shall be made on the formation of NCs; in particular, Ge NCs are formed on the Hf- based high-*k* dielectrics by chemical vapor deposition (CVD) and in HfO_2 and HfAlO matrix by cosputtering. Additionally, dielectric HfO_2 NCs are formed by the self-driven phase separation of the silicate film annealed at high temperature.

We investigate Hf-based high-*k* dielectric materials as the tunnel or control oxide for NC memory as well as the SONOS type memory to optimize the structure of the device. The Hf-based high-*k* dielectric materials are the most promising materials as an alternative dielectric material to SiO_2 and they have been widely studied in recent years. In terms of memory application, high-*k* materials as tunneling oxide is expected to provide the following advantages: fast program speed since it offers lower barrier and

hence larger tunneling current; long retention because it is physically thicker than SiO_2 with the same EOT [12, 13].

In addition, the trap nature of the high-*k* dielectric materials is utilized for SONOS type memory. High-*k* materials are used as charge storage material instead of the nitride in SONOS type memory. The effect of the *k* value and band structure of the charge storage layer on the memory performance is discussed and an engineered tunneling oxide with dual layers Si_3N_4/SiO_2 is explored to optimize the SONOS type device with high-*k* trapping layer.

Besides charge storage layer, tunnel oxide and blocking oxide, we also study the effect of gate electrode on the performance of the memory device. High work function IrO_2 gate electrode is exploited and compared to TaN gate.

1.5 Organization of Thesis

This dissertation consists of eight chapters. The remainder of this thesis has been arranged as follows.

In Chapter 2, we will review the development of NCs and SONOS memory in terms of the theoretical understanding and structure evolution.

Chapter 3 will focus on the formation of Ge NCs by CVD on Hf-based dielectric films and covers the device integration of Ge NCs and high-*k* HfO_2 tunnel and blocking oxide. In Chapter 4, efforts are put to the other method co-sputtering to form Ge NCs with smaller sizes and higher density than CVD. The device fabrication and device performance are also described.

Chapter 5 proposes a phase-separated HfSiO as the charge storage layer. In particular, the HfO_2 NCs formation by annealing HfSiO film is demonstrated. The trapping properties of the phase-separated HfSiO film are compared to the Si_3N_4 and HfO_2 films. Chapter 6 covers the optimization of the SONOS type memory structure. Besides the HfSiO trapping layer, high-*k* HfAlO is investigated as tunneling and blocking oxide to reduce the programming/erasing voltage. In addition, a high work function gate electrode IrO_2 is explored as the control gate, and its influence on erasing and retention properties is analyzed.

Chapter 7 deals with the integration of high-*k* HfO_2 trapping layer with an engineered tunneling layer, i.e. the dual tunneling layer Si_3N_4/ SiO_2. The effect of the dual tunneling layer on the program/erase, retention and endurance properties is discussed.

An overall conclusion is given in Chapter 8 to summarize the major results. Besides, possible future work is proposed in the same chapter.

References

[1] P.Cappelletti, IEDM 2004 Short Course: Memory Devices.

[2] Roberto Bez, Emilio Camerlenghi, Alberto Modelli, Angelo Visconti, “Introduction to flash memory”, *Proceedings of the IEEE*, vol. 91(4), pp.489-502, 2003.

[3] Paolo Pavan, Roberto Bez, Piero Olivo, Enrico Zanoni, “Flash memory cells-An overview”, *Proceedings of the IEEE*, vol. 85(8), pp.1248-1271, 2003.

[4] IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays.

[5] International Technology Road Map for Semiconductors 2005: Semiconductor Industry Association.

[6] William D. Brown, Joe E. Brewer, “Nonvolatile semiconductor memory technology”, IEEE Press, 1998.

[7] Stefan Lai, “Flash memories: Where we were and where are going”, in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp. 971-973, 1998.

[8] M. She, Ph.D thesis, U.C. Berkeley, 2003.

[9] K. Kim , “Technology for sub-50nm DRAM and NAND flash manufacturing”, *IEDM Tech. Dig.* pp.333 - 336 (2005).

[10] A. Thean and J.-P. Leburton, “Flash memory: towards single- electronics”, *IEEE Potentials* 21, pp. 35-41, 2002.

[11] Min She, Tsu-Jae King, “Impact of crystal size and tunnel dielectric on semiconductor nanocrystal memory performance”, *IEEE Transactions on Electron Devices*, vol. 50, pp.1934-1940, 2003.

[12] D. W. Kim, T. Kim, and S. K. Banerjee, “Memory characterization of SiGe quantum dot flash memories with HfO_2 and SiO_2 tunneling dielectrics”, *IEEE Trans. Electron Devices,* vol. 50, pp.1823-1829, 2003.

[13] J.J. Lee, X. Wang, W. Bai, N. Lu, J. Liu, D. L. Kwong, “Theoretical and experimental investigation of Si nanocrystal memory device with HfO_2 high-k tunneling dielectric”, in *Proc. VLSI Technol. Symp.*, pp.33-34, 2003.

Chapter 2

Literature Review

2.1 Evolution of Nanocrystal Memory

Nanocrystal (NC) memory was first proposed by Tiwari *et al* [1] who made devices with Si NCs embedded in SiO_2. Figure 2.1 shows the cross-sectional schematic of the device in which the NCs are sandwiched between the tunnel oxide and control oxide. The key technique of the device fabrication is the self-assembly formed Si NCs which were achieved by chemical vapor deposition.

One of the main advantages of NC memory over the conventional floating gate memory is that the NC memory is immune to the oxide defects due to the distributed nature of the charge storage layer [1-4]. This allows the use of thinner tunnel oxide and thus lowering of operation voltages or increasing of program/erase speed. Another advantage of NC memory is its simple process in comparison to the floating gate memory. The NC memory reduced the amount of masks substantially, allowing low process complexity and low cost. Compared to the conventional logic process flow, the Si NC memory only need 4 more additional masks, while the floating gate memory needs 11 additional masks [2]. Moreover, due to the absence of drain to floating gate coupling, NC memories suffer less from drain induced barrier lowering, and result in further

scaling [4]. Additionally possible advantage is that the Coulomb blockade effect of the NCs makes the multi level per cell possible.

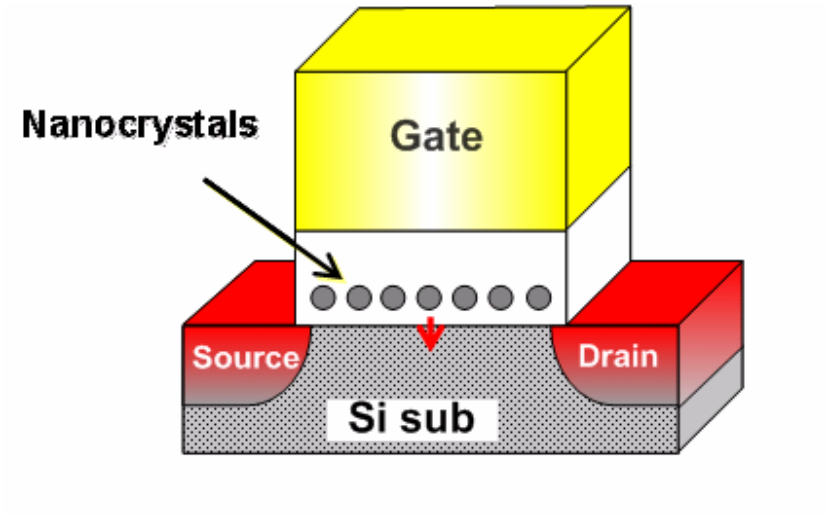


Figure 2.1. Schematic of the NC device.

The program and erase dynamics of NC memory are analyzed by Compagnoni *et al* as illustrated in Fig. 2.2 (a) and (b) [5, 6]. During program, electrons are injected from the substrate through tunnel oxide and emitted through control oxide, holes are emitted from NC to substrate through tunnel oxide. The charges are accumulated in the NCs when the injection and emission are different; ultimately a steady state will be reached when they are equal and the device is accompanied with a saturated threshold voltage V_{th}. At the steady state, the hole current is negligible and $I_{e,out} = I_{e,in}$. The charges stored at the steady state is: $Q = \varepsilon_1 A_1 F_1 - \varepsilon_2 A_2 F_2$, where ε is dielectric constants, A is the effective area of the dielectrics contributing to NC capacitance and F is the electric field across oxide. Subscripts 1 and 2 represent tunnel and control oxide respectively.

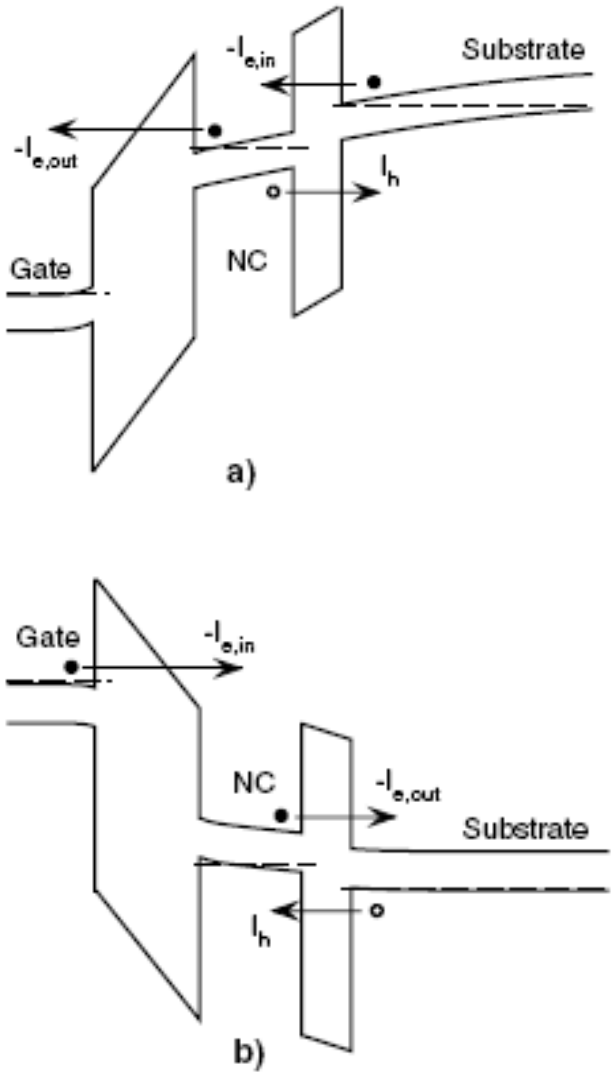


Figure 2.2. Band diagram for NC memory under a) program and b) erase modes. [4]

Figure 2.3 shows the calculated current for the case that tunnel oxide is 2.8 nm thick and control oxide is 6 nm thick and the device is applied with a 6 V and 16 V stresses [5]. Under the 6 V stress, the current across the tunnel oxide is resulted from direct tunneling and the current across the control oxide is resulted from FN tunneling. Thanks to the different tunneling mechanism, the electric field *F* across tunnel oxide and

control oxide are different at the steady state when $I_{e,out} = I_{e,in}$ and a large amount of charges Q can be stored. On the other side, when the memory is under 16 V stress, both currents across the tunnel oxide and control oxide are resulted from FN tunneling, leading to nearly same electric field F across tunnel and control oxide and small charges Q stored.

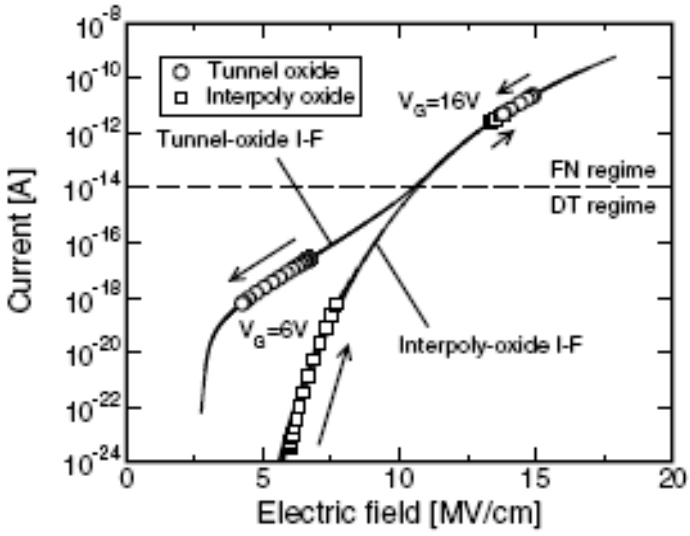


Figure 2.3. Calculated current-electric field (*I-F*) characteristics of tunnel and control oxides under gate bias at 6 V and 16 V.

In addition to CVD, other methods were also investigated to optimize the formation of NCs. The optimum NCs is with suitable size, high density and uniform distribution. The NCs should not be too small or too dispersed so that the memory device will not hold sufficient charge density. The NCs should not be too large or too dense neither so that the distributed charge storage nature is corrupted, although the large NCs bring higher program speed [7]. The optimized size of NC proposed by simulation is 5 nm [7]. The NCs obtained by CVD are strongly impacted by the surface properties of the

dielectric, and are limited since it forms crystals in plane, and there is a trade-off between nucleation and crystal growth. Aerosol deposition and ion implantation are explored to form NCs with high density [8, 9]. However, these methods suffer the problem of integration. The ion implantation need much higher thermal budget than conventional CMOS process and the control of the damage of ions to the dielectric is challenging. Aerosol deposition needs additional equipment to the conventional process line [4, 9].

Besides Si NCs, other NCs materials such as Ge, SiGe alloy and metal were also explored. The difference of Ge NCs from Si NCs is that it has narrower band gap than Si, also it has favorable phase separation nature from SiO_2 matrix and therefore other method of formation of Ge NCs are available. Oxidation of $Si_{1-x}Ge_x$ and cosputtering of $Ge+SiO_2$ to form Ge nanocrytals were reported and it was reported the Ge NCs demonstrates superior charge retention properties [10, 11]. However, formation of Ge NCs on SiO_2 surface by CVD which is the most promising and feasible technique has not been successful like Si NCs. The details of this problem will be discussed in Chapter 3.

Another breakthrough is the application of high-*k* dielectric material for the NC memory. As mentioned above, conventional NC memory uses SiO_2 as tunnel oxide. The improvement and scaling of this type of memory is physically limited by the trade off between the program efficiency and retention. Hence, use of a new material instead of SiO_2 is a possible way to overcome the limitation.

High-*k* dielectric such as HfO_2 is advantageous over SiO_2 for NC memory in that the HfO_2 is physically much thicker than SiO_2 with the same EOT, leading to several orders lower leakage current at the retention mode. When programming, the electron tunneling barrier through HfO_2 is only 1.5 eV, while is 3.5 eV though SiO_2. This leads to

a higher tunneling current and higher program efficiency when using HfO_2 as tunneling oxide. Figure 2.4 illustrates the band diagram of the Si NC memory embedded in high-*k* HfO_2. Considering 1.6 nm EOT tunnel oxide, the ratio between the leakage current across the tunneling oxide in program mode and in retention mode, $J_{g,program}$ /$J_{g,\ retention}$ of HfO_2 is about 4 orders larger than that of SiO_2, demonstrating the superiority of the HfO_2 for memory application [12]. In addition to Si NC, SiGe NCs were also successfully integrated with HfO_2 [13].

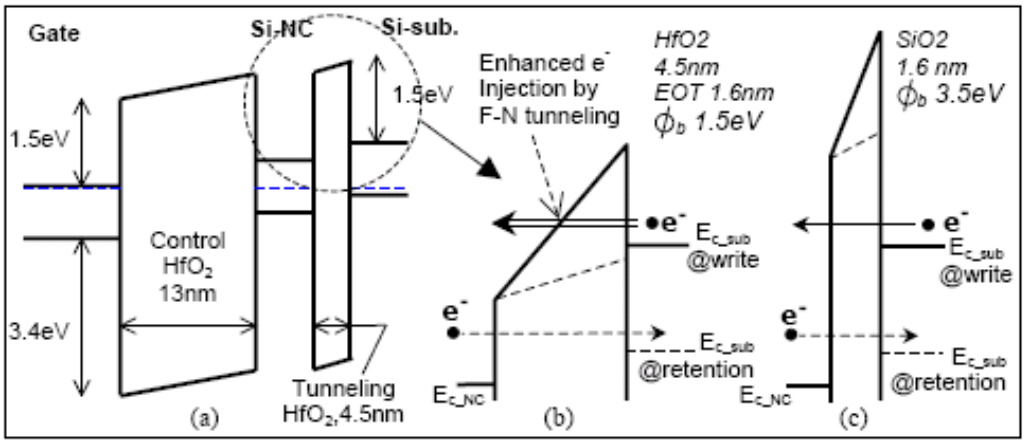


Figure 2.4. (a) Band diagram of memory device with Si NC memory embedded in HfO_2. (b) The band profile of tunneling HfO_2 in the program mode, in comparison with that of SiO_2 with the same EOT in (c). The dashed line in (b) and (c) indicate the band bending of the two dielectrics in the retention mode.

In summary, NC memory with high-*k* dielectrics features good retention, low operation voltage and resultingly excellent endurance properties.

Despite of the attractive scaling properties of NC memory, still a lot of challenges remain for its commercialization. Manufacturing problems include the control of the NC size and distribution and their reproducibility. In particular, when gate length scales, the

gate area reduces and only a few numbers of NCs will be accommodated in NC memory. This makes the production of NC devices more challenging. Physical issues include the small memory window of NC memory. In addition, the fundamental understanding of the scalability of NC memory is not sufficient, and the new technologies have to face the "nonprocessing related value-components" such as "the time-compression diseconomies" and the "intellectual property" [4]. All these issues need to be resolved before the mass-manufacture of NC memory.

2.2 Evolution of SONOS Type Memory

The SONOS type memory was developed in the 1960s. For many years, the development of SONOS memory was driven by the military and space applications because of their robustness of radiation. In recent years, the development of SONOS memory has been accelerated again by its attractiveness for high-density embedded nonvolatile memory applications, mainly because of their endurance of aggressive scaling, process compatibility with the CMOS technology, process simplicity and low integration costs.

The typical cross-section structure of SONOS memory is schematically shown in Fig. 2.5 [3]. A nitride film is sandwiched between a thin tunneling oxide and a thick block oxide layer. The band diagrams of device during programming are illustrated in Fig.2.6 [14]. Related charges transports are also indicated in the figure. During the programming, a positive bias is applied to the gate electrode, electrons are injected from the entire channel through the tunneling oxide by direct tunneling, some electrons will continue move and escape to the gate, and the others will be trapped by the nitride layer.

At the same time, minority of holes will transfer through the blocking oxide and some of them will be captured by the nitride traps and the others will escape from the nitride layer and tunnel to substrate channels. The erasing process corresponds to a reverse charge transport courses.

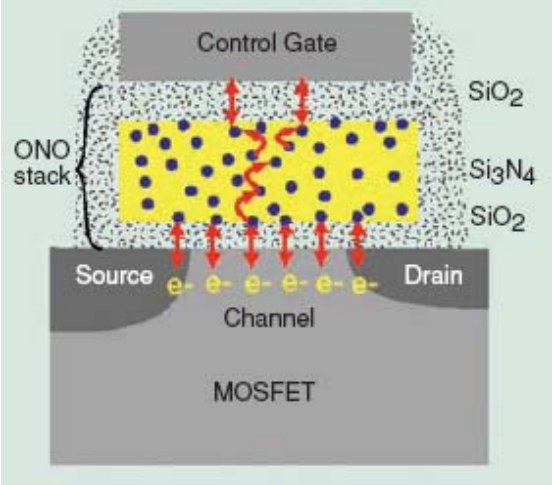


Figure 2.5. Schematic cross sectional structure of SONOS device.

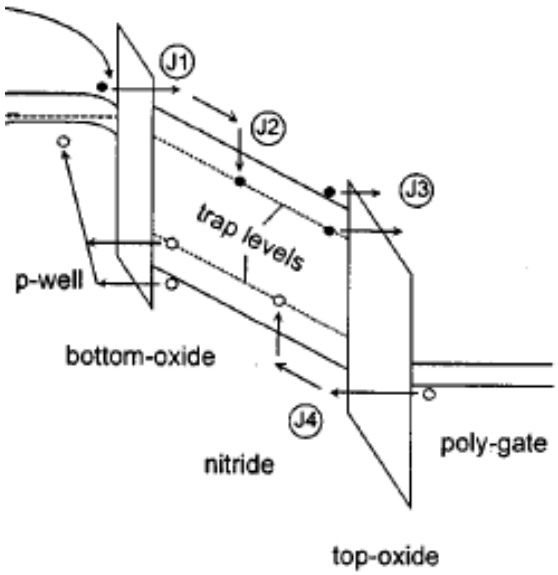


Figure 2.6. Band diagram of SONOS type memory in program mode.

To accelerate the commercialization and enhance its performance, a lot of efforts were made on the engineering of SONOS structure. The engineering involves gate, blocking oxide, trapping layer and tunneling oxide, furthermore the full structure.

1. Gate Engineering

The gate electrode material of SONOS memories is commonly n^+ poly Si. However, studies demonstrated that using metal or p^+ poly Si with higher work function than n^+ poly Si brings more benefits to SONOS memories [14, 15].

A typical phenomenon observed from SONOS is erase saturation, as shown in Fig. 2.7 [15]. The erase limitation is caused by the gate electron injection from the gate electrode, which is in competition with the holes injection from the channel. During erasing, negative charges stored in the nitride are reduced, causing the band structure to tune to enhance electron injection from gate and reduce holes injection from the substrate. A steady state will be reached when the two currents are balanced. The reduction of the gate electron injection can be achieved straightforwardly by increasing the barrier height, i.e., using an electrode with a high work function. In this case, the erasing speed will be increased and especially, the saturation level will be lower, resulting in larger memory window, as demonstrated by Reisinger *et al* and illustrated in Fig. 2.7 [15]. Alternatively, a thicker tunneling oxide can be used to achieve better retention without degradation of erase speed. Based on the similar idea, TaN and IrO_2 were explored to improve the erase of the device [16, 17].

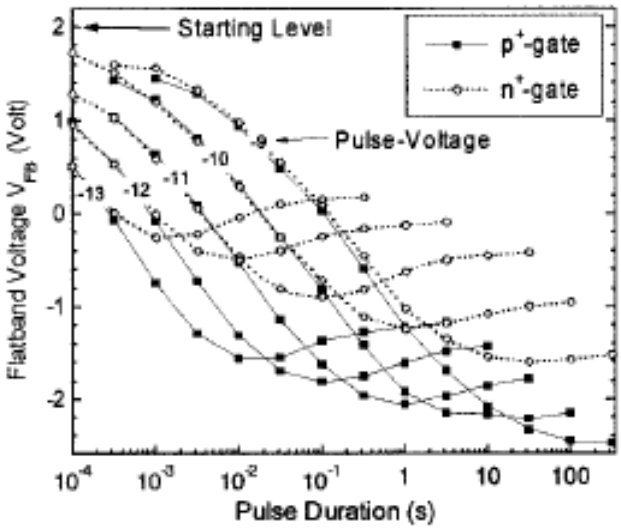


Figure 2.7. Erase characteristics of SONOS MOS capacitors with n^+ and p^+ gate. The tunnel oxide is 3 nm thick.

2. Blocking Oxide Engineering

The desired properties of blocking oxide are high dielectric constant and large conduction band offset to Si ΔE_c. When the dielectric constant of blocking oxide is increased, the voltage drop through it will be reduced while the voltage drop through tunneling oxide will be enhanced when programming and erasing because the voltage drop through a dielectric is reversely proportional to its capacitance. This will improve program/erase speed. On the other side, the larger conduction band offset ΔE_c provides a higher barrier for electrons to tunnel through. This will reduce electron injection from gate electrode when erasing and reduce the electrons emitting from trapping layer to gate electrode when programming, resulting in improved program/erase speed. In addition, the high dielectric constant and large ΔE_c also contributes a lower leakage current when the dielectric has the same EOT and is applied the same voltage through it; therefore the retention of the memory will be guaranteed.

The relevant properties of potential dielectric materials are listed in Table 2.1 and the band structures of some dielectric materials are illustrated in Fig. 2.8 [18, 19]. It is observed from Table 2.1 and Fig. 2.8 that the ΔE_c for most interested high-*k* gate dielectrics on Si are considerably smaller than SiO_2 except Al_2O_3. In particular, Al_2O_3 is a thermally stable material and can be readily integrated to SONOS device. Therefore, Lee *et al* made SONOS memory device with Al_2O_3 as blocking oxide, and proved Al_2O_3 offers excellent better blocking effect than SiO_2 because the erase saturation of device is successfully overcome by using Al_2O_3 instead of SiO_2 [20].

Table 2.1. Comparison of relevant properties for high-*k* candidates.

Material	Dielectric constant (κ)	Band gap E_G (eV)	ΔE_C (eV) to Si
SiO_2	3.9	8.9	3.2
Si_3N_4	7	5.1	2
Al_2O_3	9	8.7	2.8[a]
Y_2O_3	15	5.6	2.3[a]
La_2O_3	30	4.3	2.3[a]
Ta_2O_5	26	4.5	1–1.5
TiO_2	80	3.5	1.2
HfO_2	25	5.7	1.5[a]
ZrO_2	25	7.8	1.4[a]

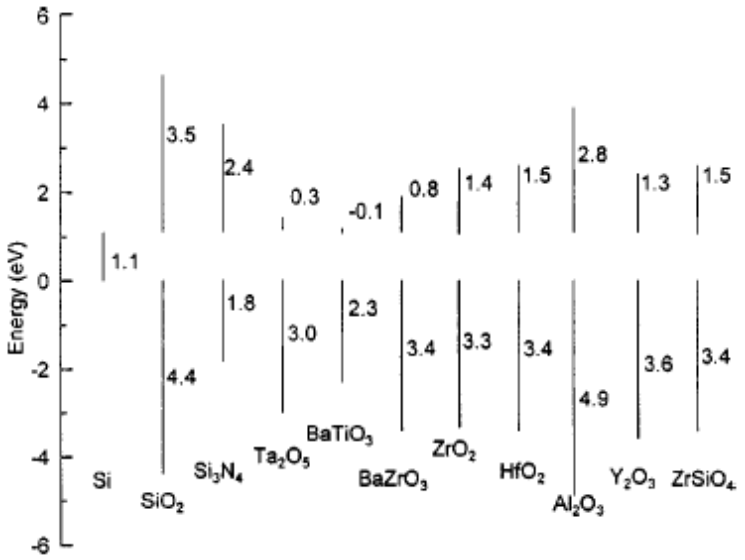


Figure 2.8. Calculated conduction and valence band offsets of the various gate dielectric materials. [19]

3. Trapping layer investigation

Another exploration of high-*k* dielectrics for SONOS type memory is to use high-*k* dielectrics as charge storage layer instead of nitride. The potential advantages of high-*k* charge storage layer lies in its higher dielectric constants than that of Si_3N_4. High-*k* material is physically much thicker than Si_3N_4 when having the same EOT. Therefore higher area trap densities can be expected. Additionally, the thicker high-*k* increases the scattering possibility, hence increasing the possibilities of charge capturing and charge trapping efficiency [21]. As the scaling of the semiconductor devices, the thickness of trapping layer is scaled as well. The physically thick high-*k* trapping layer thus will offer a more aggressive scalability. Alternatively, if using the same physical thickness, the device with high-*k* layer will have a lower program/erase voltage due to the smaller EOT of the high-*k* layer [22].

Moreover, when using high-*k* dielectric materials with smaller conductance band offset than nitride, the tunneling probability for electrons through blocking oxide is also reduced, increasing the trapping efficiency of device. Furthermore, the varieties of high-*k* dielectric candidate provide a wider choice of material, this make the adjustment of trap density and trap energy level possible.

The trapping properties of Al_2O_3, HfO_2 and HfAlO were examined by research works, and it was proved that all these materials provide better electron trapping capability, which demonstrate either by lower program/voltage or faster program/erase speed [21-23].

4. Tunneling oxide engineering

A concept of layered barrier was invented by Likharev and it was originally suggested for floating gate memory [24]. The layered barrier consists of multilayer dielectrics with different dielectric constants. The asymmetric barrier low-*k* /high-*k* and symmetric barrier low-*k* /high-*k* /low-*k* barriers were proposed by Govoreanu *et al* [25]. The benefit of the layered barrier for floating gate flash is that by proper design, the current across the barrier is more sensitive to electric field than that across single SiO_2 barrier. In other words, the engineered tunneling barrier offers higher current at higher electric field and lower current at lower electric field than single SiO_2, allowing higher program/erase speed and better retention for floating gate memory devices [25, 26]. Use of engineered tunnel barrier for SONOS type or NC memory is also beneficial in the same sense. Recently, engineered tunneling barrier was also explored for SONOS type memory [27]. In Chapter 7 of this thesis, we also study an engineered tunnel barrier for MONOS memory application.

2.3 Summary

The conventional floating gate flash memory faces significant challenge when scaling beyond 45 nm node. This has led to extensive research on the NC memory and SONOS type memory, which are the most promising potential candidates to take the place of the floating gate memory.

NC memory with Si embedded in SiO_2 has been studied extensively with respect to the formation of NCs, device integration, and the memory performance. The implementation of high-*k* dielectric instead of SiO_2 brings additional advantage to NC memory in that the use of high-*k* dielectric as tunnel oxide can improve the programming speed because of the reduction of the barrier height and simultaneously assure retention properties because of its large physical thickness. As the study of the NC memory using high-*k* dielectric materials is still limited, further study will be necessary.

Improvement of SONOS memory has been accomplished by engineering of gate electrode, blocking oxide, charge storage layer and tunnelling oxide. High work function gate electrode, high-*k* blocking oxide with high conduction band offset to Si, high-*k* trapping layer, and engineered tunnel oxide are proved to bring advantages to SONOS performance with respect to fast program/erase speed, long time retention as well as sufficient cycling endurance. Studies on the optimization of total gate stack may be needed for the further improvement.

References

[1] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan, "A silicon NCs based memory", *Appl. Phys. Lett.,* vol.68, pp. 1377-1379, 1996.

[2] R.Muralidhar, R.F. Steimle, M. Sadd, R. Rao, C.T. Swift, E.J. Prinz, J. Yater, L. Grieve, K. Harber, B.Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S.G.H. Anderson, M. Rossow, T. Merchant, M. Paransky, T. Huynh, D. Hadad, Ko-Min Chang, and B.E. White Jr. "A 6V embedded 90nm silicon NC nonvolatile memory", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.,* pp.26.2.1 -26.2.4, 2003.

[3] A. Thean and J.-P. Leburton, "Flash memory: towards single-electronics", *IEEE Potentials* 21, pp. 35-41, 2002.

[4] J. D. Blauwe, "NC nonvolatile memory devices", *IEEE Transaction on nanotechnology*, vol.1, No.1, pp. 72-77, 2002.

[5] C. Monzio Compagnoni, D. Ielmini, A.S. Spinelli, A. L. Lacaita, C. Gerardi, L. Perniola, B. De Salvo and S. Lombardo, "Program/erase dynamics and channel conduction in NC memories", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp.22.4.1 -22.4.4, 2003.

[6] C. Monzio Compagnoni, D. Ielmini, A.S. Spinelli, A. L. Lacaita, "Modeling of tunneling P/E for NC memories", *IEEE Transaction on Electron Devices,* vol. 52, No.4, pp. 569-576, 2005.

[7] M. She, T.-J. King, "Impact of crystal size and tunnel dielectric on semiconductor NC memory performance". *IEEE Transaction on Electron Devices*, vol. 50, No.9, pp. 1934-1940, 2003.

[8] J. D. Blauwe, M. Ostraat, M. L. Green, G. Weber, T. Sorsch, A. Kerber, F. Klemens, R. Cirelli, E. Ferry, J. L. Grazul, F. Baumann, Y. Kim, W. Mansfield, J. Bude, J. T. C. Lee, S. J. Hillenius, R. C. Flagan, and H. A. Atwater, "A novel, aerosol-NC floating-gate device for non-volatile memory applications", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp.683-686, 2000.

[9] C. J. Nicklaw, M.P. Pagey, S. T. Pantelides, D. M. Fleetwood, R. D. Schrimpf, K. F. Galloway, J.E. Witting, B. M. Howard, E. Taw, W.H. Mcneil, and J. F. Conley Jr., "Defects and NCs generated by Si implantation into a-SiO_2," *IEEE Trans. Nucl.Sci.,* vol. 47, pp.2269-2275, 2000.

[10] Y.-C. King, T.-J. King, and C. Hu, "MOS memory using germanium NCs formed by thermal oxidation of $Si_{1-x}Ge_x$," in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp. 115-118, 1998.

[11] W. K. Choi, W. K. Chim, C.L.Heng, L. W. Teo, Vincent Ho, V. Ng, D.A. Antoniadis and E.A. Fitzgerald, "Observation of memory effect in germanium NCs embedded in an amorphous silicon oxide matrix of a metal-insulator-semiconductor structure" *Appl. Phys. Lett.,* vol. 80, pp. 2014-2016 (2002).

[12] J. J. Lee, X. Wang, W. Bai, N. Lu, J. Liu, and D. L. Kwong, "Theoretical and experimental investigation of Si NC memory device with HfO_2 High-*k* tunneling dielectric," in *Proc. VLSI Technol. Symp.*, pp. 33–34, 2003,.

[13] D.-W. Kim, T. Kim and S. K. Banerjee, "Memory characterization of SiGe quantum dot flash memories with HfO_2 and SiO_2 tunneling dielectrics," *IEEE Trans. Electron Devices*, vol. 50, pp. 1823 - 1829, 2003.

[14] H. Bachhofer, H.Reisinger, E. Bertagnolli and H. von Philipsborn, "Transient conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures", *J. Appl. Phys.,* vol. 89, pp. 2791-2800, 2001.

[15] H. Reisinger, M. Franosch, B. Hasler, and T.Böhm, "A novel SONOS structure for nonvolatile memories with improved data retention", in *Proc. VLSI Technol. Symp.*, pp. 113–114, 1997.

[16] C.-H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp.26.5.1-26.5.4, 2003.

[17] C.H. Lai, A.C. Chin, K.C. Chiang, W.J. Yoo, C.F. Cheng, S.P. Mcalister, C.C. Chi and P. Wu, "Novel SiO_2/AlN/HfAlO/IrO_2 memory with fast erase, large ΔV_{th} and good retention", in *Proc. VLSI Technol. Symp.*, pp.210-211, 2005.

[18] G. D. Wilk, R. M. Wallace, J. M. Anthony, "High-*k* gate dielectrics: current status and materials properties considerations", *J. Appl. Phys.*, vol. 89, No. 10, pp.5243-5275, 2001.

[19] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices", *J. Vac. Sci. Technol.* B 18, pp.1785-1791, 2000.

[20] C.-H. Lee, S.-H. Hur, Y.-C. Shin, J.-H. Choi, D.-G. Park, and K. Kim, "Charge-trapping device structure of SiO_2/SiN/high-k dielectric Al_2O_3 for high-density flash memory", *Appl. Phys. Lett.* vol. 86, 152908 1-3, 2005.

[21] Y. N. Tan, W. K. Chim, B. J. Cho, W. K. Choi, "Over-erase phenomenon in SONOS-type flash memory and its minimization using a hafnium oxide charge storage layer," *IEEE Trans. Electron Devices*, vol. 51, pp.1143-1147, 2004.

[22] M. She, Ph.D thesis, U.C. Berkeley, 2003.

[23] T. Sugizaki, M. Kobayashi, M. Ishidao, H. Minakata, M. Yamaguchi, Y. Tamura, Y. Sugiyama, T. Nakanishi, and H. Tanaka, "Novel multi-bit SONOS type flash memory using a high-k charge trapping layer", in *Proc. VLSI Technol. Symp.,* pp.27-28, 2003.

[24] K. K. Likharev, "Layered tunnel barriers for nonvolatile memory devices," *Appl. Phy.Lett*., vol. 73, pp. 2137-2139, 1998.

[25] B. Govoreanu, P. Blomme, M. Rosmeulen, J. Van Houdt, and K. De Meyer, "VARIOT: A Novel Multilayer Tunnel Barrier Concept for Low-Voltage Nonvolatile Memory Devices", *IEEE Electron Device Lett.*, vol. 24, pp.99-101, 2003.

[26] S. H. Hong, J. H. Jang, T. J. Park, D. S. Jeong, M. Kim, C. S. Hwang and J. Y. Won, " Improvement of the current-voltage characteristics of a tunneling dielectric by adopting a $Si_3N_4/SiO_2/Si_3N_4$ multilayer for Flash memory application"*, Appl. Phys. Lett*., vol. 87, pp. 152106 1-3, 2005.

[27] H. T. Lue, S.-Y. Wang, E.-K. Lai, Y.-H. Shih, S.-C. Lai, L.-W. Yang, K.-C. Chen, J. Ku, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "BE-SONOS: a bandgap engineered SONOS with excellent performance and reliability", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.,* pp. 555 – 558, 2005.

Chapter 3

Ge Nanocrystals Formed by Chemical Vapor Deposition

3.1 Introduction

As discussed in Chapter 2, flash memory device employing nanocrystals (NCs) has received considerable attention as one of the most promising candidates for future nonvolatile, high density, and low power memory devices. Currently, commercial nonvolatile floating gate flash memory devices employ a continuous poly-Si floating gate and a thick SiO_2 tunnel oxide (7-9 nm) to achieve ten-year retention time, resulting in high programming voltage or slow programming speed. To meet future scaling requirements, the trade-off between data retention and program efficiency should be improved [1]. By using discrete NCs instead of continuous floating gate as charge storage nodes, local defect related leakage can be reduced [2]; therefore thinner tunnel oxide can be used to assure the ten-year nonvolatile properties of the memory device and the program efficiency of the memory device can also be improved simultaneously.

Self-assembled Si NCs can be successfully formed on SiO_2 films by chemical vapor deposition (CVD) which is a favorable chemical process in semiconductor manufacturing due to its advantages over other deposition processes such as flexibility, reproducibility, high purity and low cost; therefore the structure with Si NCs embedded

in SiO_2 is widely used for NC memory. Besides Si, Ge NCs is also promising for NC memory application. It has been reported that memory devices with Ge NCs demonstrated superior retention properties to that with Si NCs [3, 4]. However, Ge NCs have not been successfully self-assembled on SiO_2 by CVD like Si NCs. Indeed, it was found difficult to deposit significant amount of Ge on SiO_2 because the GeH_4 precursor does not react with a SiO_2 surface as SiH_4 does [5].

In this study, we are going to access the possibility of Ge NCs deposited on Hf-based high-*k* dielectric materials by CVD. The influence of the CVD process parameters such as CVD temperature, deposition time and flow rate of the gas precursor and the influence of the underlying layers on the characteristics of the Ge NCs are identified. Here, Ge is expected to provide an additional advantage over Si, i.e., excellent chemical stability with high-*k* material. In addition, the use of a high-*k* dielectric instead of conventional SiO_2 dielectric brings more advantages to memory device. The high-*k* dielectric has smaller conduction band offset with respect to Si than SiO_2 and is physically thicker than SiO_2 when having the same equivalent oxide thickness (EOT); hence both the programming efficiency and data retention of the memory device can be improved [4, 6].

In addition to the formation of NCs, we also fabricate memory capacitor by integrating Ge NCs with HfO_2 high-*k* dielectric material, and study their electric properties.

3.2 Experiment

P-type Si (100) substrates were used as started wafer in the experiment. HfO_2 films were formed on the substrate by the following process: 1) surface nitridation of the substrates at 700ºC for 60 sec. in an ammonia (NH_3) ambient, 2) deposition of hafnium oxide at 400ºC and 4×10^{-4} Torr using $Hf(OC(CH_3)_3)_4$ as a precursor, 3) post deposition anneal (PDA) at 700ºC for 60 sec. in N_2 ambient. HfAlO films were also formed following the same processes as HfO_2 except that the deposition of HfAlO was performed at 420ºC and 6.4×10^{-5} Torr using $HfAl(MMP)_2(OiPr)_5$ (where MMP is $OC(CH_3)_2CH_2OCH_3$, and OiPr is $Oi\text{-}C_3H_7$) as a precursor. The physical thickness of HfO_2 and HfAlO were about 60 Å after PDA measured by ellipsometer. CVD of Ge NCs was then performed on HfO_2 or HfAlO using germane (GeH_4) precursor at 0.4 Torr. The deposition temperature, time and GeH_4 gas flow rate were varied to investigate their impact on NC formation

Atomic force microscopy (AFM), X-ray diffraction (XRD), X-ray photoelectron spectroscopy (XPS), scanning electron microscopy (SEM) and ellipsometer were employed to characterize the Ge NCs.

3.3 Dependence of Ge NCs on Deposition Condition

1. Influence of deposition temperature on the formation of Ge NCs

Figures 3.1 (a)-(c) show the AFM images of Ge NCs deposited at 500ºC, 550ºC, and 600ºC, respectively, on HfO_2. The other deposition parameters (time = 30sec., flow rate = 80 sccm) are kept constant.

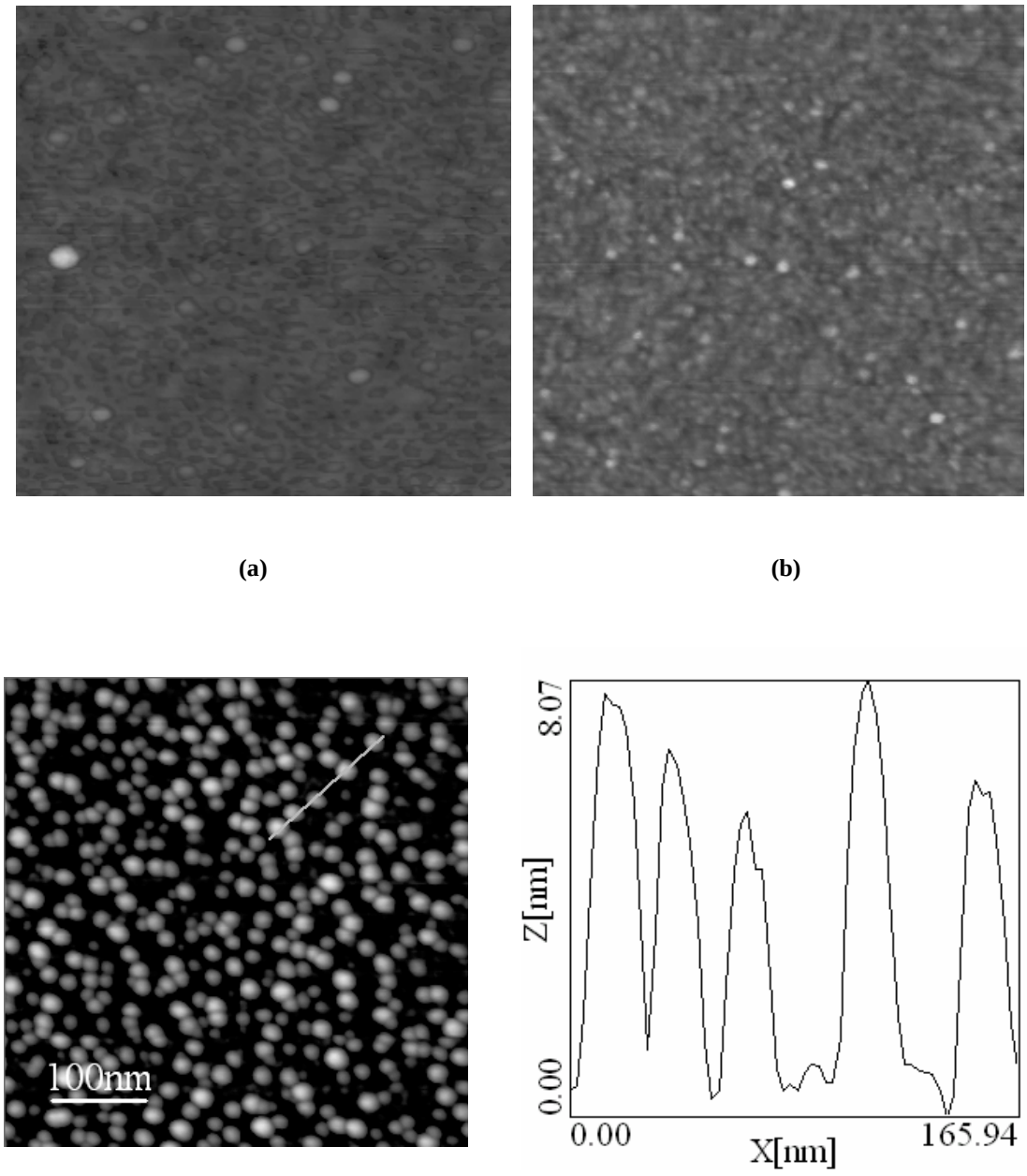


Figure 3.1. AFM images of Ge NCs deposited at (a) 500 ^{o}C, (b) 550 ^{o}C and (c) 600^{o}C on HfO_2 dielectric. (d) Surface profile along the line in (c). The Ge NC density is obtained to be about 10^{11} cm^{-2}, and the mean diameter of Ge NCs is 16 nm. The mean height of Ge NCs is 7 nm.

No observable Ge NCs were formed at a deposition temperature of 500ºC [Fig. 3.1 (a)]. Even at a deposition temperature of 550ºC, formation of Ge NCs does not substantially occur. However, when the deposition temperature is increased to 600ºC, Ge NC formation is evident [Fig. 3.1 (c)], with the density of NCs about 10^{11} cm^{-2}. The surface profile along the fine line in Fig. 3.1(c) is shown in Fig. 3.1 (d), indicating the mean diameter and height of the Ge NCs in Fig. 3.1(c) are 16 nm and 7 nm, respectively. A scanning electron microscopy (SEM) view of the sample of Fig. 3.1 (c) is shown in Fig. 3.2. Analysis of the SEM image obtained the same NC size and densities as from the AFM image analysis.

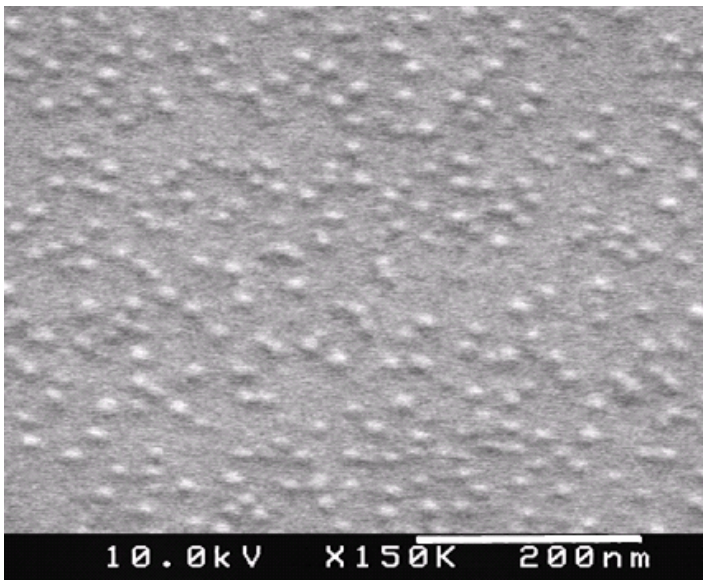


Figure 3.2. SEM image of Ge NCs on HfO_2 deposited at 600^{o}C.

To examine the chemical state of the deposited Ge NCs, an ex-situ XPS measurement was carried out on the sample of Fig. 3.1 (c). Both the Ge-Ge bond and the Ge-O bond are observed from the XPS spectrum (Fig. 3.3). The signal due to the Ge-Ge bond is predominant when the detector angle is 90^{o}, but the signal due to the Ge-Ge bond is comparable to that due to the Ge-O bond when the detector angle is 15^{o}. The difference of XPS profile at the different detector angle indicates that the core of NCs is Ge, and Ge-O bonds are on the surface of Ge NCs. GeO_x may be formed when the sample was upon exposure to air. Further confirmation of the presence of crystalline Ge was achieved by an XRD analysis, as shown in Fig. 3.4, where the as-deposited Ge film sample reveals peaks corresponding to the (111), (220) and (331) orientations of the diamond-like Ge crystal structure.

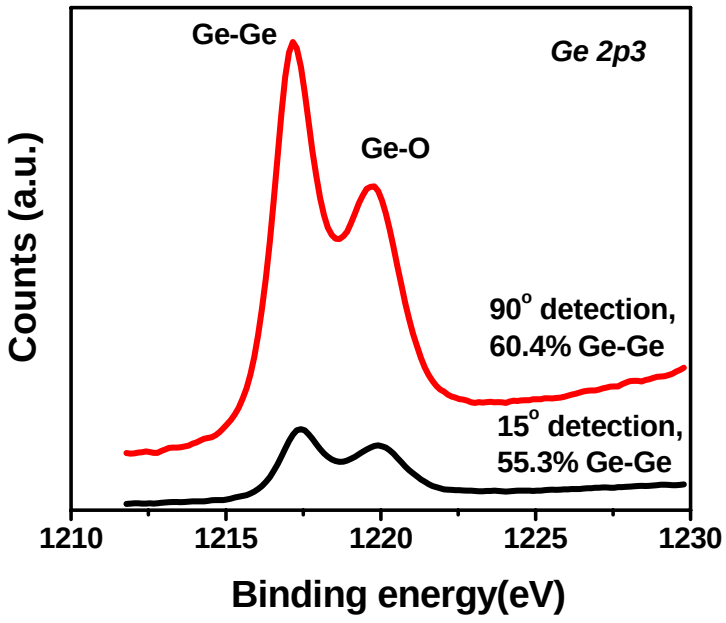


Figure 3.3. XPS spectrum of Ge NCs deposited at 600^{o}C, showing the existence of Ge-Ge and Ge-O bonds.

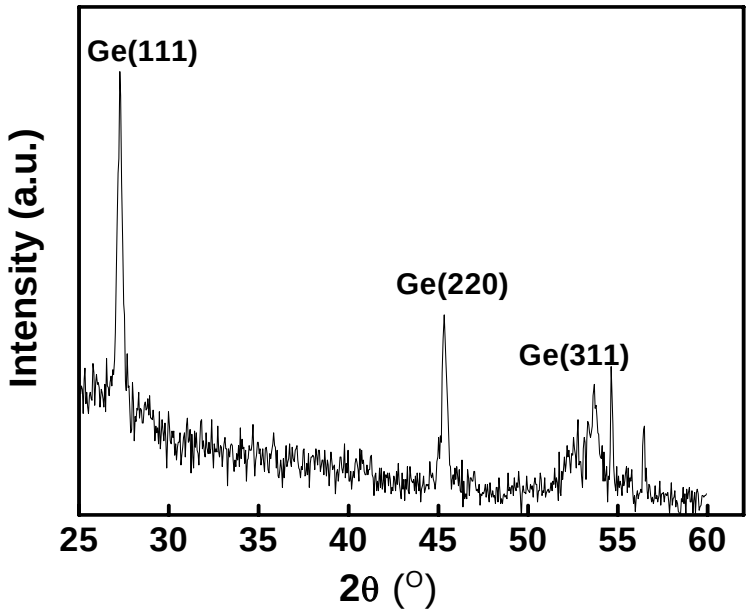


Figure 3.4. XRD profile of Ge deposited at 600°C, indicating the diamond-like crystals structure of Ge.

2. Influence of deposition time on the formation of Ge NCs

Figure 3.5 summarizes the mean diameter and density of the Ge NCs observed by AFM with the deposition temperature fixed at 600°C and the GeH_4 flow rate fixed at 80 sccm but varied deposition time. The Ge NCs density initially increases with time, reaching a peak at a deposition time of 30 sec., then decreases with time. The diameter of the Ge NCs follows a different trend: it does not change much for deposition time below 30 sec., but subsequently increases with time.

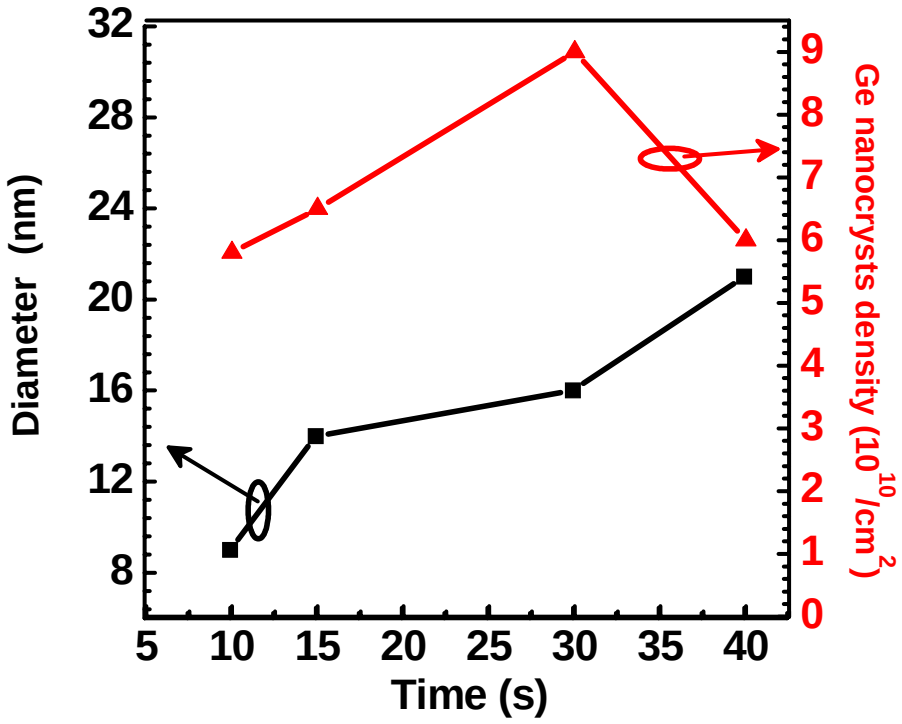


Figure 3.5. Mean diameter and surface density of Ge NCs on HfO_2 as a function of deposition time.

3. Influence of gas flow rate on the deposition of Ge NCs

Fig. 3.6 shows the dependence of the Ge NC diameter and density on the GeH_4 gas flow rate. The deposition temperature and deposition time are kept at 600^{o}C and 30 sec., respectively. There is an intrinsic trade-off between the amount of sources and the gas resident time when the gas flow rate is adjusted. This trade-off most likely results in a maximum density when the sources and the residence time are both optimized, as shown in Fig. 3.6.

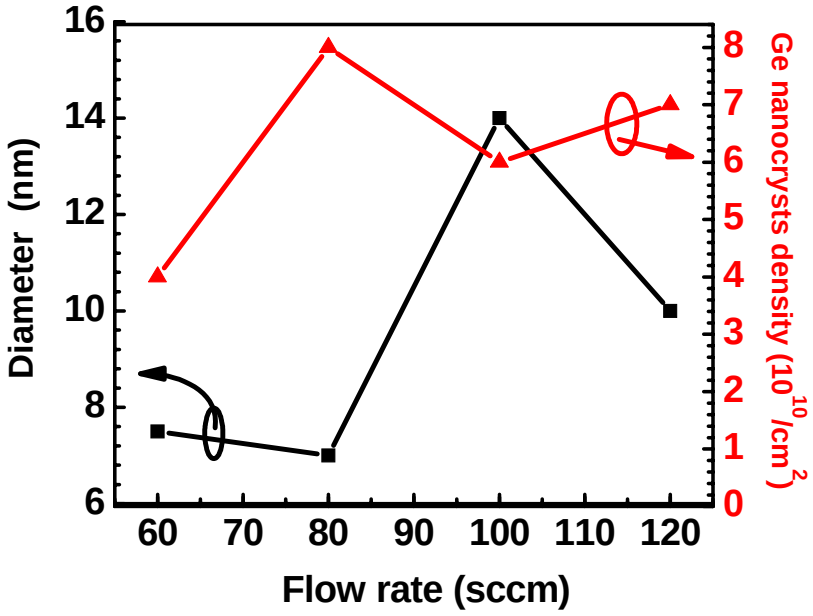


Figure 3.6. Mean diameter and surface density of Ge NCs on HfO_2 as a function of flow rate.

4. Formation of Ge NCs on HfAlO

The Ge NCs deposited on hafnium aluminum oxide HfAlO was also investigated and the AFM images of Ge NCs deposited on HfAlO at 580ºC, 590ºC and 600ºC are shown in Fig. 3.7. It was noted that the density of the Ge NCs on HfAlO increases as the temperature decreases, and the highest density of NCs was observed at 580^{o}C. This indicates that the optimum deposition temperature of Ge NCs on HfAlO is lower than that on HfO_2.

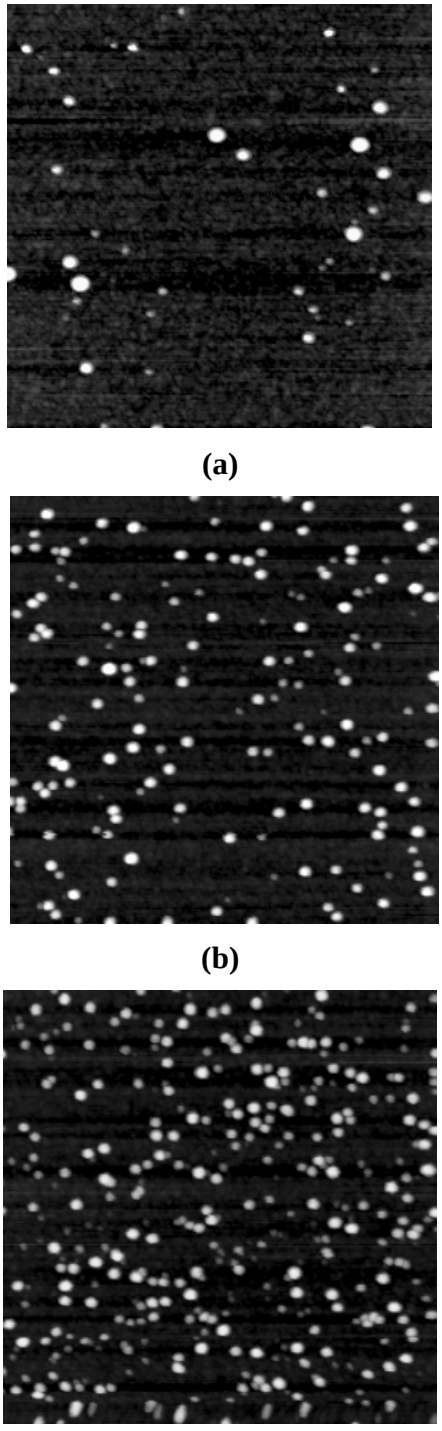


Figure 3.7. AFM images (1μm×1μm) of Ge NCs deposited on HfAlO at (a) 600 ^{o}C, (b) 590 ^{o}C and (c) 580^{o}C.

5. Influence of surface treatment on the formation of Ge NCs

In addition, the influence of NH_3 treatment on HfO_2 surface on Ge NCs formation was also investigated. This experiment followed the same process as HfO_2 film deposition, except that the PDA was replaced by annealing at 700ºC for 1min. in an ammonia (NH_3) ambient. Ge was deposited on the treated HfON surface with the deposition parameters the same as the sample for Fig. 3.2 which is deposited on HfO_2 surface. No much difference was found between the two underlying materials. Fig. 3.8 shows the SEM picture of Ge NCs on HfON.

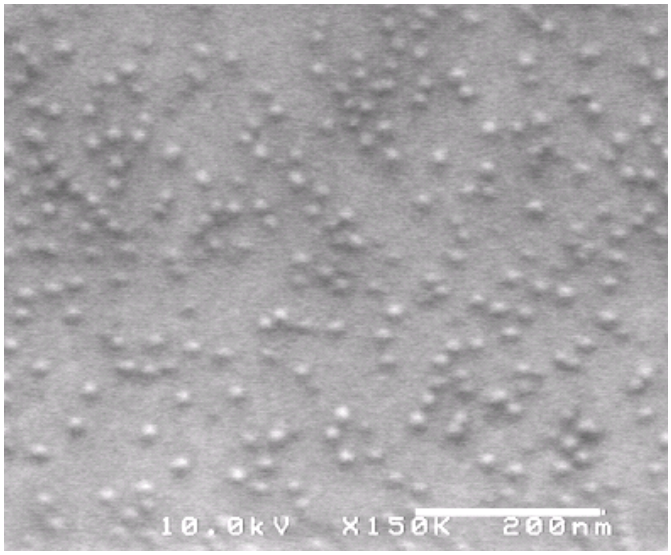


Figure 3.8. SEM image of Ge NCs on NH_3 treated HfO_2.

3.4 Discussion

Based on the above observation and the data from [7], we analyze the growth mechanism for Ge NCs on high-*k* gate dielectrics. The deposition of the NCs should follow the sequence of adsorption, reaction, nucleation, growth and agglomeration, as illustrated in Fig. 3.9. The reactions can be simplified and shown below:

$$GeH_4\,(gas) \rightarrow GeH_4\,(adsorbed) \rightarrow Ge + 2H_2\,(gas) \qquad (1)$$

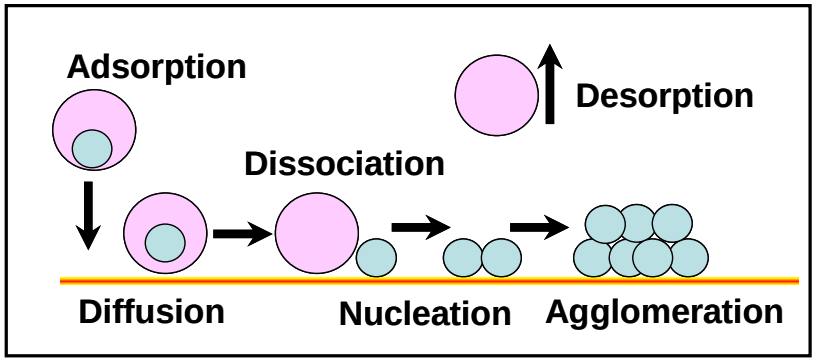


Figure 3.9. Schematic illustration of the process of CVD Ge NCs.

For a given dielectric surface on which Ge is deposited, increasing the deposition temperature reduces the rate of adsorption but accelerates the dissociation reaction. Between 550°C and 600ºC, it seems that the dissociation is the rate-limiting factor that influences the Ge NC deposition on HfO_2 rather than adsorption, resulting in the highest density at 600ºC when the deposition time and gas flow rate are fixed. After nuclei are formed, they grow through either crystal growth or agglomeration as the deposition time

is prolonged. The agglomeration will result in the decrease of the NCs density, as we observed in Fig.3.5.

For HfAlO, the highest density of Ge NCs was attained at a deposition temperature of 580^{o}C (Fig. 3.7), with the density decreasing with increasing deposition temperature from 580^{o}C to 600^{o}C. Considering the aforementioned process of NCs deposition, the adsorption step should be the rate-limiting step for Ge NC deposition on HfAlO.

It therefore appears that HfO_2 has a stronger ability to adsorb the precursors containing of Ge than HfAlO. Here, combining with the fact that it is difficult to form Ge NCs on SiO_2 by CVD, we are leaded to consider the influence of the high-*k* dielectric material on the Ge NCs formation using CVD. A dielectric material with a higher permittivity generally has more polarized bonds and this high polarization nature of high-*k* materials may make them more adhesive to precursors. The strong adsorption ability of a surface to the precursor is primary for CVD because the activation of the chemical reaction requires a certain high temperature and the precursor will desorbs if the adsorption ability if not strong enough at this temperature.

Our results were favorably compared by another works [8], where the reactions of precursor containing of Ge with SiO_2 or HfO_2 were studied in detail and the intermediate products during the reaction were detected. It was stated that GeO_x is produced and becomes an adhesion layer when Ge is deposited on HfO_2 surface; the desorption temperature of GeOx is 850 K (577 ^{o}C). On the other hand, GeO is formed when Ge is deposited on SiO_2, and it desorbs at a low temperature about 500 K (277^{o}C), resulting in the etching process of SiO_2 rather than deposition. Although our own model does not

consider the intermediate products during the reaction, the understanding of the impact of the adsorption ability of the surface on CVD is consistent with their work [8]. And their work also reveals that high-*k* HfO_2 has the better adsorption ability than SiO_2.

The absorption ability of SiO_2 is possibly too low to accumulate enough Ge to nulei. For HfAlO, the NCs are formed, but the highest density temperature is lower than HfO_2 possibly because the desorption temperature are lower on HfAlO than HfO_2.

The relation of the dielectric constant of substrate materials to CVD is also revealed in the case of Si NCs deposition. Baron *et al.* described the deposition of Si NCs from the activation energy for *nucleation and growth.* The rate of nucleation proportional to $\exp(-E_a/kT)$, where E_a is the nucleation activation energy, k is Bolzman constant, and T is process temperature. E_a for SiO_2, SiO_xN_y, Si_3N_4, and Al_2O_3 were reported to be 3.8 eV, 3.4 eV, 2.8 eV, and 1.8eV, respectively [8, 9]. Obviously, a dielectric with a higher-k provides a lower barrier for the nucleation of Si NCs. It also supports our understanding that the surface with higher dielectric constant benefits the formation of NCs by CVD possibility because of the strong adsorption ability of the high-*k* surface which increases the nucleation rate.

3.5 Capacitor Fabrication and Characterization

A capacitor with Ge NCs embedded in HfO_2 was fabricated. After a standard pre-gate cleaning and surface nitridation, 60 Å HfO_2 was deposited and annealed. Ge NCs were then deposited using the process conditions of Fig. 1 (c). 140 Å of HfO_2 was deposited as the control oxide. A post-deposition anneal was performed at 700ºC for 60

sec. in N_2 ambient. Finally, 1500 Å of TaN metal gate was reactively sputtered. A control capacitor with only 200 Å HfO_2 but no Ge NCs was also fabricated.

To understand the element distribution in the Ge NCs capacitor, secondary ion mass spectroscopy (SIMS) was carried out. The HfO_2/Ge/HfO_2 sandwich structure was clearly observed from the Ge and Hf profile, as shown in Fig. 3.10

.

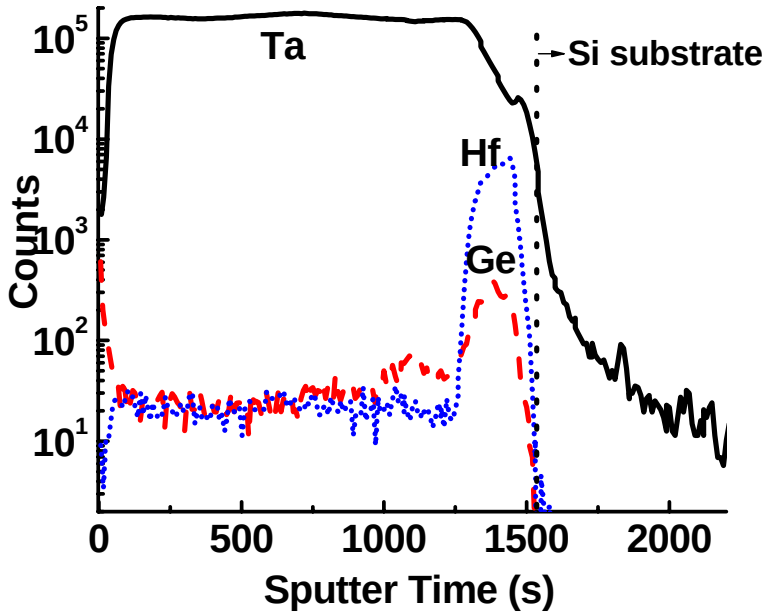


Figure 3.10. SIMS profiles of Ta, Hf and Ge in memory capacitor.

Capacitance-voltage characteristics of the fabricated capacitors were measured by scanning the gate voltage from -6 V to 6V and then 6V to -6V. As shown in Fig. 3.11, the control sample shows a 0.5 V hysteresis, while the Ge nanocrytals capacitor shows a 2V hysteresis. The hysteresis of control wafer is most likely due to traps in the HfO_2 dielectric. The Ge NCs significantly enhances the hysteresis, indicating the insertion of Ge NCs provides for a memory effect.

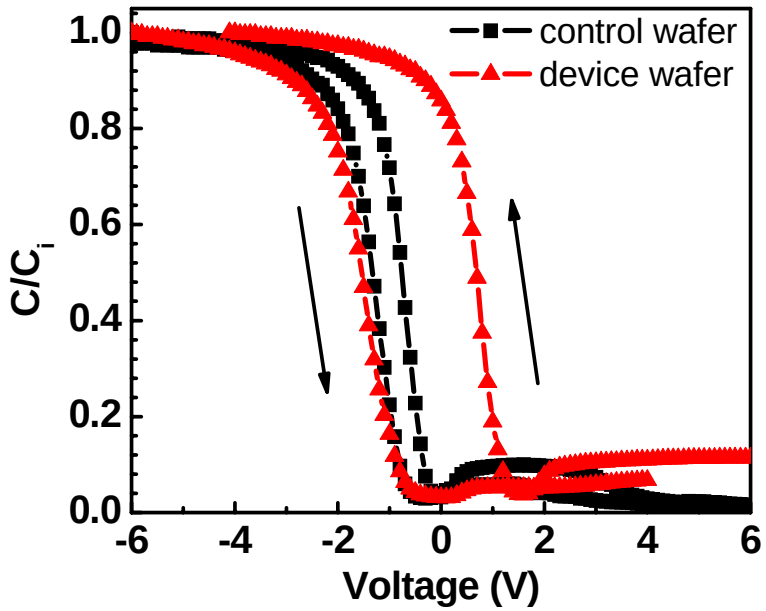


Figure 3.11. *C-V* hysteresis of the control and device capacitors.

3.6 Summary

In summary, we demonstrated the first direct deposition of Ge NCs on HfO_2 and HfAlO by CVD. The dependence of NC size and density on the deposition temperature, time, flow rate and underlying dielectric were investigated. Ge NC formation is evident at a deposition temperature of 600ºC on HfO_2 surface, and provides an obvious memory effect. NCs formation is generally favored on a surface with a higher dielectric constant.

Reference

[1] T. C. Chang, S. T. Yan, P. T. Liu, C.W. Chen, S. H. Lin and S. M. Sze, "A novel approach of fabricating germanium nanocrystals for nonvolatile memory application" *Electrochem. Solid-State Lett.,* vol. 7(1), pp.G17- G19, 2004.

[2] A. Thean and J.-P. Leburton, "Flash memory: towards single- electronics" *IEEE Potentials* 21, pp.35-41, 2002.

[3] T. Usuki, T. Futatsugi, and N. Yokoyama, "A proposal of new floating gate memory storing a small number of electrons with relatively long retention time at low voltage operations", *Microelectron. Eng*., vol. 47, pp. 281-283, 1999.

[4] D. W. Kim, T. Kim, and S. K. Banerjee, "Memory characterization of SiGe quantum dot flash memories with HfO_2 and SiO_2 tunneling dielectrics", *IEEE Trans. Electron Devices,* vol. 50, pp.1823-1829, 2003.

[5] T. Baron, B. Pellissier, L. Perniola, F. Mazen, J. M. Hartmann, and G. Rolland, *Appl. Phys. Lett*. "Chemical vapor deposition of Ge nanocrystals on SiO_2", vol. 83, pp.1444-1446, 2003.

[6] J. J. Lee, X. Wang, W. Bai, N. Lu, J. Liu, and D. L. Kwong, Proc. "Theoretical and experimental investigation of Si nanocrystal memory device with HfO_2 High-k tunneling dielectric", in *Proc. VLSI Technol. Symp*., pp.33-34, 2003.

[7] R. Gupta, W. J. Yoo, Y. Wang, Z. Tan, G. Samudra, S. Lee, D. H. Chan, K. P. Loh, L. K. Bera, N. Balasubramanian, and D. L. Kwong, "Formation of SiGe nanocrystals in HfO_2 using *in situ* chemical vapor deposition for memory applications", *Appl. Phys. Lett.,* vol. 84, pp. 4331-4333, 2004.

[8] S. K. Stanley, S. V. Joshi, S. K. Banerjee, and J. G. Ekerdt, "Ge interactions on HfO_2 surfaces and kinetically driven patterning of Ge nanocrystals on HfO_2", *J. Vac. Sci. Technol.*, vol. A 24, pp.78-83, 2006.

[9] T. Baron, F. Martin, P. Mur, C. Wyon, M. Dupuy, C. Busseret, A. Souifi, G. Guillot, "Low pressure chemical vapor deposition growth of silicon quantum dots on insulator for nanoelectronics devices", *Applied Surface Science*, vol. 164, pp.29-34, 2000.

[10] A. Fernandes, B. DeSalvo, T. Baron, J. F. Damlencourt, A. M. Papon, D. Lafond, D. Mariolle, B. Guillaumot, P. Besson, P. Masson, G. Ghibaudo, G. Panankakis, F. Martin, S. Haukka, "Memory characteristics of Si quantum dot devices with SiO_2/ALD Al_2O_3 tunneling dielectrics**",** in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp.155-158, 2003.

Chapter 4

Ge Nanocrystals Formed by Cosputtering

4.1 Introduction

NC flash memory using a high-*k* dielectric in place of the conventional SiO_2 can achieve significantly improved programming efficiency and data retention [1, 2]. To date, work on the integration of high-*k* dielectric materials with NCs is limited, especially, Ge NCs embedded in Hf-based dielectrics, the most attractive and promising high-*k* dielectrics for CMOS integration.

Chapter 3 investigated the formation of Ge NCs by chemical vapor deposition. In this chapter, another method cosputter is exploited to form Ge NCs. The advantage of cosputter method is that it is not limited by the surface properties of the dielectric materials. More importantly, it is possible to obtain Ge NCs with higher density and smaller size than CVD [3]. Process of formation of Ge NCs in HfAlO is described and its mechanism is analyzed in the next section, followed by the integration of Ge NCs with HfAlO high-*k* dielectric materials to make memory devices. Furthermore, memory device with Ge NCs embedded in HfO_2 is fabricated in the same way as that with Ge NCs in HfAlO, and memory performances especially the retention properties of the HfO_2 based device and the HfAlO based device are compared. Detailed characterization of the HfO_2 / co-sputtered HfO_2+Ge / HfO_2 stack and the HfAlO / co-sputtered HfO_2+Al_2O_3+Ge

/ HfAlO stack are conducted by high resolution transmission electron microscopy (HRTEM), scanning transmission electron microscopy (STEM), and energy dispersive X-ray (EDX) spectrometry to understand the relation between the retention properties of the memory device and their microstructure.

4.2 Formation of Ge NCs in HfAlO by Co-sputtering

A co-sputtered film was fabricated by co-sputtering HfO_2, Al_2O_3 and Ge on a Si substrate in Ar ambient at 3 mTorr with a volume ratio of HfO_2 : Al_2O_3 : Ge = 3 : 1 : 1. The co-sputtered film was then annealed at various temperatures in N_2 ambient at a pressure of 1 atm for 30 s.

Figure 4.1 shows the X-ray photoelectron spectroscopy (XPS) analysis of samples before and after being annealed at 500 ^oC, 700 ^oC, and 950 ^oC. The as-deposited sample reveals the presence of HfO_2, Al_2O_3, Ge, as well as Hf and GeO_2 according to their binding energy [4]. No significant difference is observed between the as-deposited sample and the sample annealed at 500 ^oC. However, XPS intensities associated with Hf and GeO_2 decrease noticeably, while that of Ge increases substantially as the annealing temperature goes up to 700 ^oC. This implies that the phase structure of Hf + GeO_2 changes to Ge + HfO_2 at a temperature of about 700 ^oC. The Al_2O_3 peak also shifts towards higher energy, indicating the presence of stronger oxidation states at 700 ^oC. The 950 ^oC annealed sample has a similar state as the 700 ^oC annealed sample.

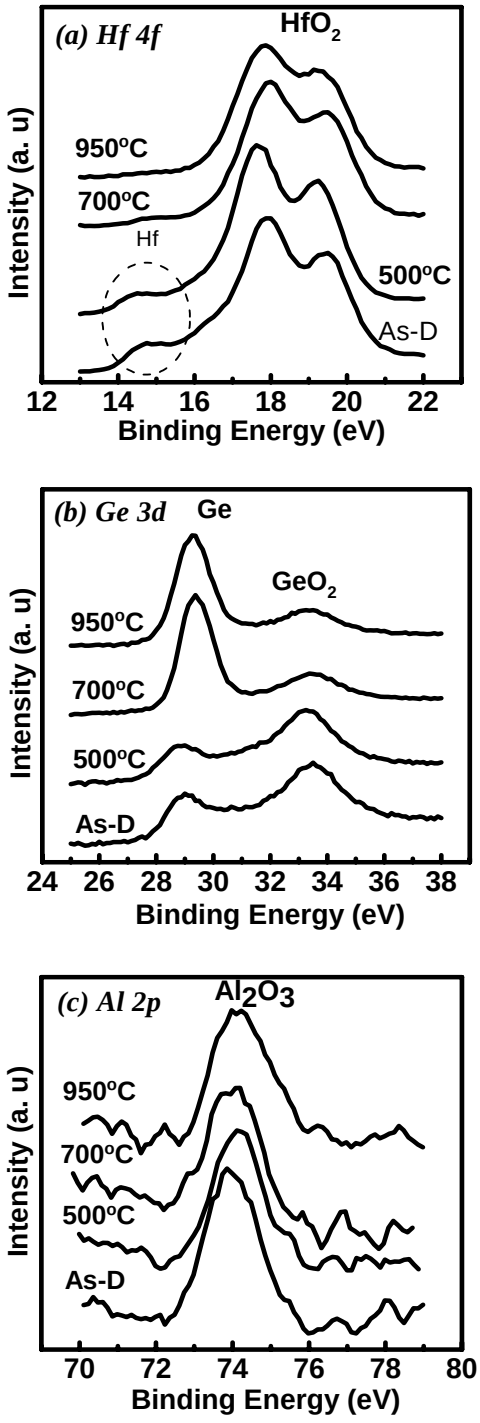


Figure 4.1. XPS spectra of (a) Hf 4f, (b) Ge 3d, and (C) Al 2p. Analysis was performed for the as-deposited (as-D) sample and samples annealed at 500 ^{o}C, 700 ^{o}C, 950 ^{o}C.

To explain the transformation of Hf + GeO_2 to Ge + HfO_2, inference may be made from the oxidation phenomenon of SiGe alloys. Many reports explained the preferential oxidation of Si to Ge to be due to larger negative Gibbs free energy to form SiO_2 than GeO_2.[5] The standard Gibbs free energies of formation ($\Delta_f G^o$) at 298.15K for HfO_2, Al_2O_3 and GeO_2 are -1088.2, -1582.3, and -521.4 kJ mol^{-1}, respectively.[6] Gibbs free energy change (ΔG) of a reaction can be calculated by $\Delta G = \Delta H - T\Delta S$, where T is the reaction temperature, ΔH and ΔS are enthalpy and entropy changes at the reaction temperature, respectively. Since the contribution of entropy term ($T\Delta S$) is small compared to the enthalpy term (ΔH) and ΔH changes little with temperature (for example, the difference between the enthalpy of formation of GeO_2 at room temperature and at 900 ^oC is about 1% [6]), the reaction, GeO_2 (cr) + Hf (cr) → HfO_2 (cr) + Ge (cr), should bring a large negative ΔG in the temperature range used in this work. From the perspective of thermodynamics, GeO_2 is expected to be spontaneously reduced to Ge, while Hf is oxidized to HfO_2. This concurs with our experimental observation.

The reduction of GeO_2 to Ge was observed from the samples annealed at 700 ^oC and 950 ^oC, but not from the 500 ^oC annealed sample. This may be explained by the exponential dependence of reaction kinetics on temperature in which the reaction rate is proportional to the factor $e^{-Ea/kT}$ where E_a is the activation energy of the reaction and k is Boltzman constant. The thermodynamic stability of Ge in the Hf-Al-O matrix can also prevent any traces of oxygen from oxidizing Ge during the NC formation process, resulting in enlargement of the process window.

4.3 Device with Ge Nanocrystals Embedded in HfAlO: Fabrication and Characterization

The Ge NCs formation technique was integrated into a transistor process for the fabrication of memory device. The designed structure of the device and the process flow for the device fabrication is illustrated in Fig. 4.2.

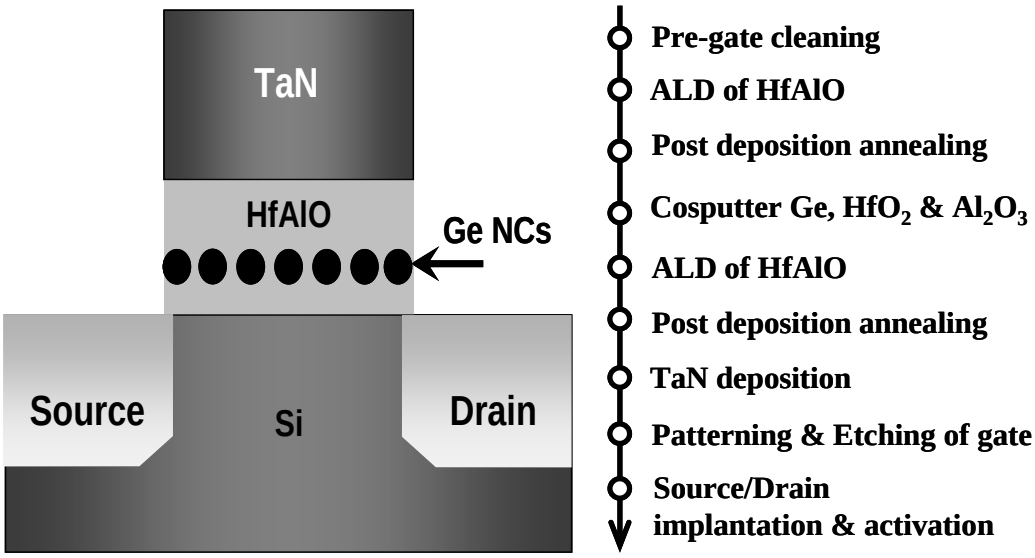


Figure 4.2. Schematic and the process flow of the NC memory device.

After pre-gate cleaning and surface nitridation at 700 ^{o}C for 30 s in NH_3 ambient, a 50 Å thick HfAlO tunnel oxide was deposited at 320 ^{o}C by atomic layer deposition (ALD) and annealed at 700 ^{o}C for 30 s in N_2 ambient. This is followed by co-sputtering of HfO_2, Al_2O_3 and Ge to form a 100 Å thick film. An 80 Å thick ALD HfAlO dielectric annealed at 700 ^{o}C for 30 minutes formed the control oxide. TaN metal gate was subsequently formed by reactive sputtering, gate patterning, and gate etching.

Source/drain regions were formed by As^+ implantation and dopant activation at 950^oC for 30 s. Finally, the device underwent the standard forming gas annealing at 420 ^oC for 1 hour. For comparison, a control sample was fabricated. It underwent the same processes except omitting Ge and introducing O_2 to oxidize Hf in the co-sputtering process.

Figure 4.3 shows a cross-sectional transmission electron microscopy (TEM) image of the fabricated device sample. Images of NCs embedded in HfAlO are clearly observed between tunnel oxide and control oxide. The inset in Fig. 4.3 shows high-resolution TEM of Ge NCs, in which the clear lattice structure extending over a region of 5 - 10 nm is observed. It was observed from TEM plan-view, that the mean size of the Ge NCs is about 10nm, and the density of the NCs is about $10^{11}cm^{-2}$. The size distribution of the Ge NCs is shown in Fig. 4.4.

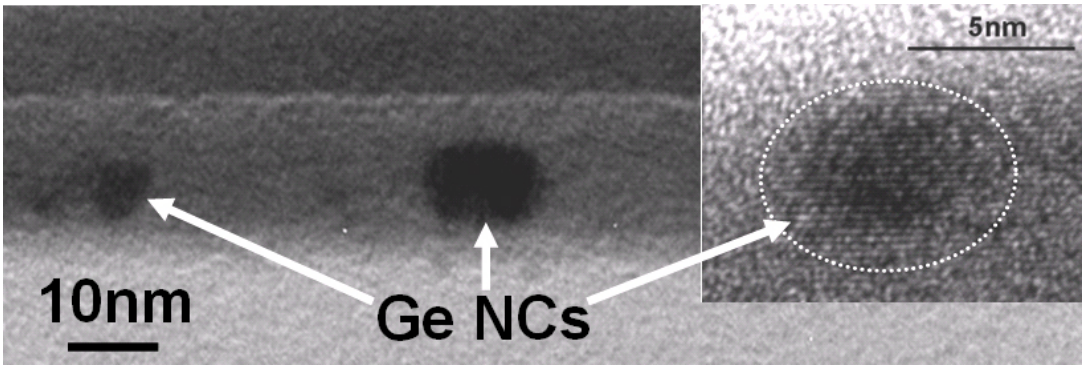


Figure 4.3. Cross-sectional TEM image of Ge NCs embedded in HfAlO dielectric matrix. The inset shows a magnified Ge NC with lattice structure.

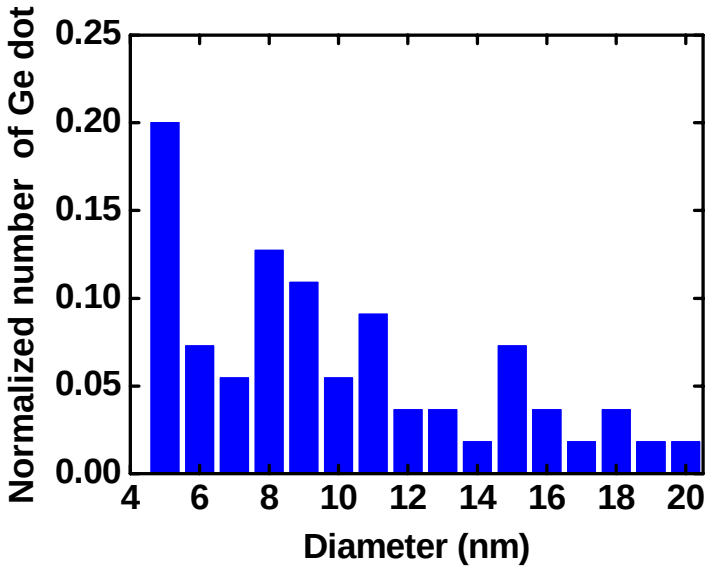


Figure 4.4. Distribution of Ge NCs.

C-V characterization was performed for the Ge NC device and a control sample which only has the same structure as the Ge NC device but no Ge NCs. As seen from Fig. 4.5, flat band voltage shift of 2.2 V was obtained from the device sample after stressing it for 100 ms at 12 V, while very little shift was observed from the control sample. The comparison of the *C-V* results between the device and control samples suggested that Ge NCs, not HfAlO traps, were mainly responsible for the flat band voltage shift. The flat band voltage shift can be estimated by the relation, $\Delta V_{FB} = \frac{d}{\varepsilon_s} Q_t$, where d is the equivalent oxide thickness (EOT) between the trapped charge and the gate electrode, Q_t is the trap charge density and ε_s is the dielectric constant of SiO_2 [7]. The EOT obtained from Fig. 4.5 is 8.1 nm. Referring to TEM and assuming that the Ge NCs are confined at the center of the total gate oxide layer, the 2.2 V flat band voltage shift

corresponds to a charge density of about 10^{13} /cm^2, i.e. 100 electrons in one Ge dot, which can be large enough for the future flash memory applications.

The typical programming and erasing transient response under different voltages are measured and shown in Figs. 4.6 (a) and (b) respectively. The retention properties of the device will be shown in the next section in order to compare with the device with Ge NCs embedded in HfO_2.

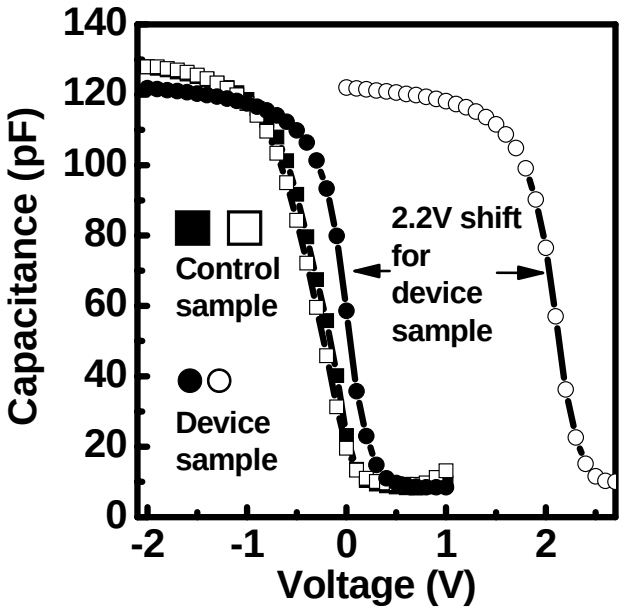


Figure 4.5. Memory effect obtained from *C-V* characterization of Ge NCs embedded in HfAlO memory device.

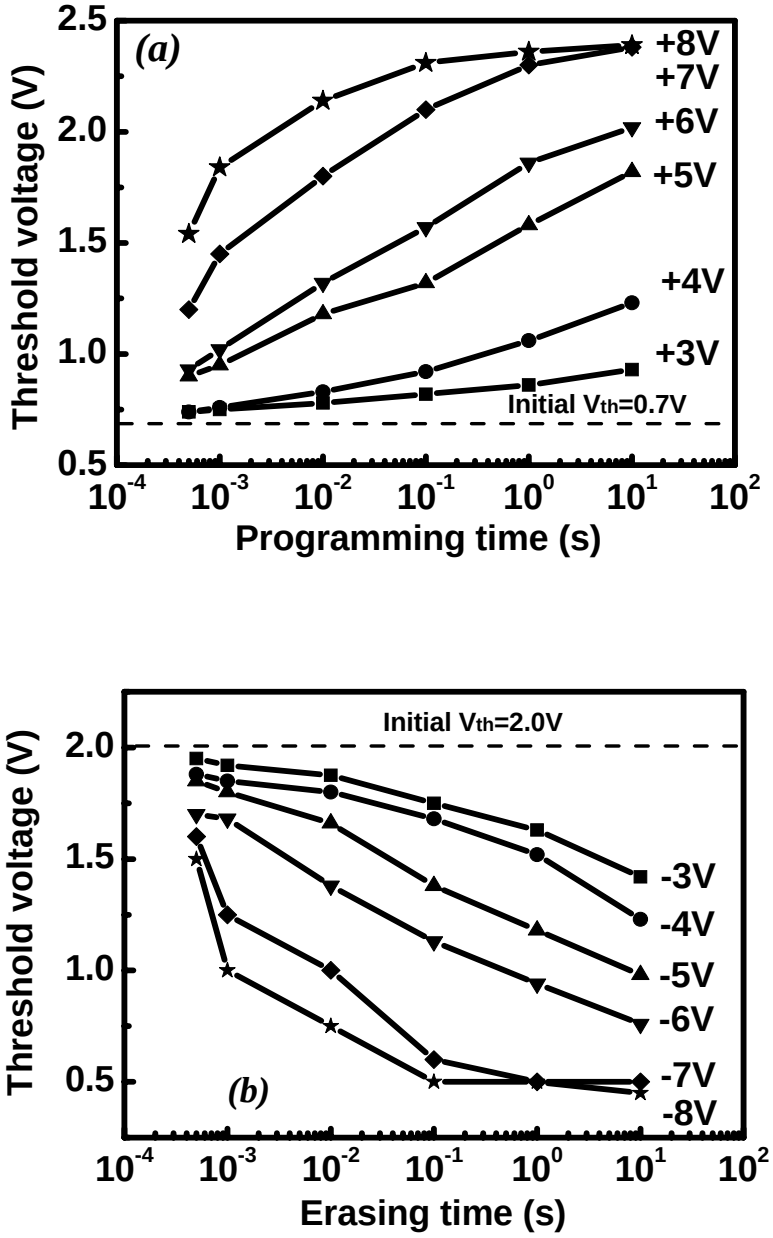


Figure 4.6. Transient characteristics of (a) programming and (b) erasing operations for the transistor device with Ge NCs embedded in HfAlO under various gate voltages and pulse durations.

4.4 Charge Retention Property and Microstructure of Ge NCs Embedded in HfO_2 and HfAlO

The XPS and TEM results in section 4.2 and section 4.3 confirm the formation of Ge NCs in the HfAlO matrix. It is natural to expect that Ge NCs are also formed in HfO_2 matrix due to the same thermodynamic mechanism. In this section, we shall address the difference between charge retention properties of Ge NCs embedded in HfO_2 and in HfAlO and gain an insight into the relation between their microstructure and their retention properties.

Memory devices with Ge NCs embedded in HfO_2 are fabricated following a similar process flow to that for HfAlO based devices. P-type Si wafers were used as starting substrates. After pre-gate cleaning and surface nitridation, a 60 Å thick HfO_2 tunneling oxide was formed by atomic layer deposition (ALD) and a post deposition anneal at 700 ^oC for 30 s in a N_2 ambient. This was followed by co-sputtering of Ge/HfO_2 to form a 100 Å thick film. An 80 Å thick ALD HfO_2 dielectric annealed at 700^oC for 30 minutes formed the control oxide. TaN metal gate was subsequently formed by reactive sputtering, gate patterning, and gate etching. Source/drain regions were formed by As^+ implantation and dopant activation at 950 ^oC for 30 s. Finally, the device underwent the standard forming gas annealing at 420 ^oC for 1 hour.

Figure 4.7 shows the retention properties of the device with Ge NCs embedded in HfO_2, together with the device with Ge NCs embedded in HfAlO. Both devices were programmed at 7 V for 0.5 ms and erased at -7 V for 100 ms. It can be observed that immediately after programming and erasing the window of the HfO_2 based device is

larger than that of the HfAlO based device. The larger memory window may be attributed to the lower barrier of HfO_2 (1.5 eV) than that of HfAlO (2.1 eV), leading to higher injection current through the tunneling oxide [8, 9]. However, the memory window of the HfO_2 based device shrinks quickly, closing in less than 15 minutes (Fig. 4.7). The poor charge retention property of Ge NCs embedded in pure HfO_2 was also reported in Ref.10. In contrast, the HfAlO based device shows very good charge retention property, retaining 75% of memory window for up to 10^5 s.

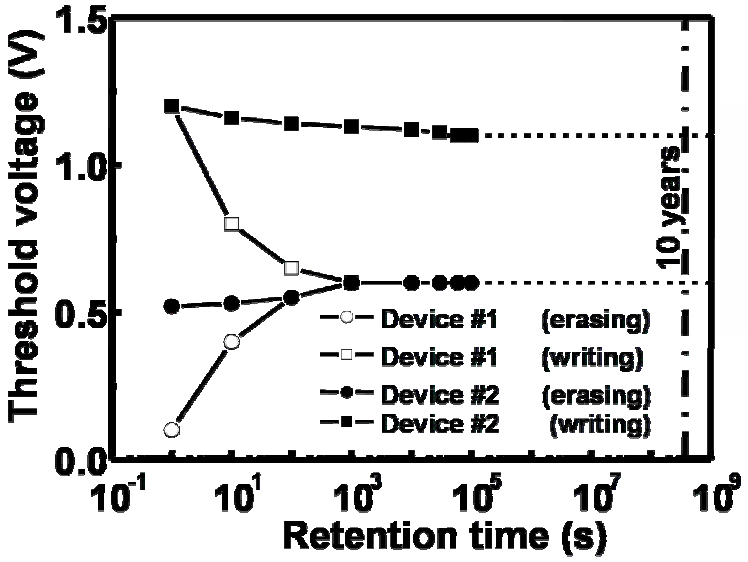
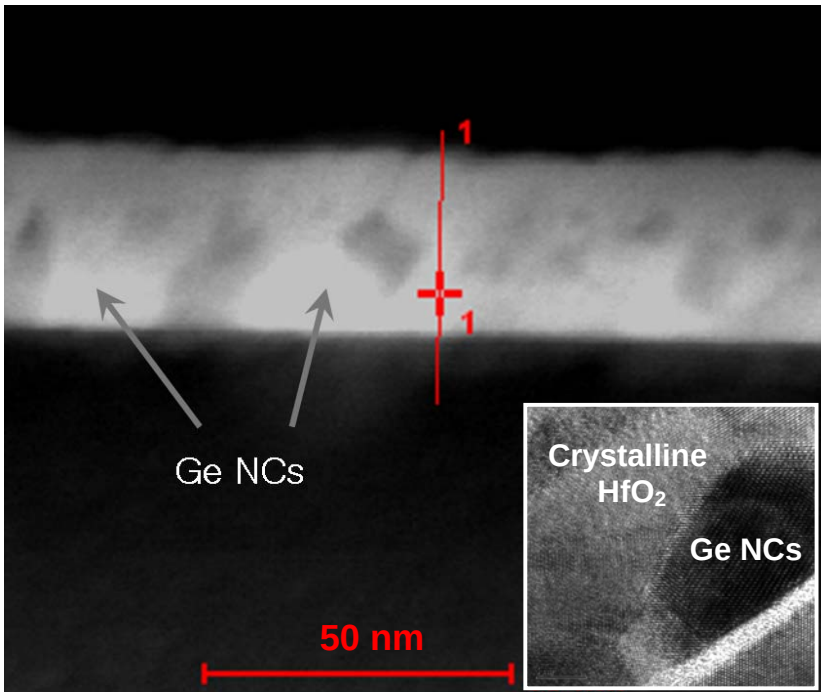


Figure 4.7. Comparison of retention properties between the HfO_2 based device (Device #1) and HfAlO based device (Device #2). The channel lengths of the devices are 10 μm.

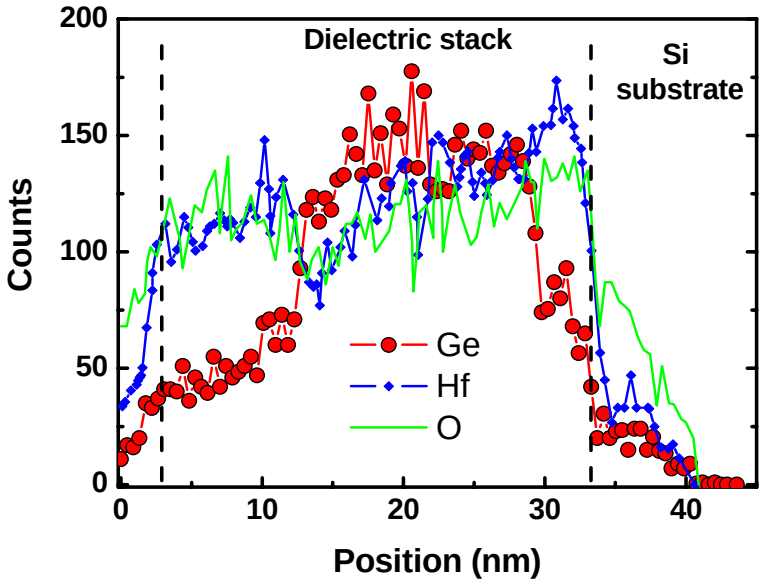
Since the retention is one of the most important characteristics for nonvolatile memory application, it is very crucial to understand the physical mechanism responsible

for the retention properties of devices. The two dielectric stacks, HfO_2 / co-sputtered HfO_2+Ge / HfO_2 stack (referred to as the HfO_2 stack in this chapter) and the HfAlO / co-sputtered HfO_2+Al_2O_3+Ge / HfAlO stack (referred to as the HfAlO stack in this chapter) were examined by STEM and energy dispersive X-ray (EDX) spectrometry. Each layer in the stack has the same thickness as the corresponding device mentioned before and the two stacks had undergone the same thermal process as that during device fabrication. STEM Z-contrast image is so sensitive to the atomic number (Z) that we can utilize it to obtain information such as their location and concentration.

The STEM Z-contrast image of the HfO_2 stack is shown in Fig.4.8 (a). Two types of grains are observed with one embracing the other. This is different from our design scheme. According to the EDX result of Fig. 4.8 (b) which reveals the distribution of Hf, O, and Ge elements across the line highlighted in Fig. 4.8 (a), it is found that Hf and O distribute across the entire dielectric stack; besides, Ge also spreads across the entire layer, and is not confined at the middle of the dielectric stack where it was deposited. The STEM and EDX reveal that Ge redistributes after annealing, and some Ge nanodots situate close to the interfacial layer. The HRTEM analysis also shows clearly that there is Ge NC in contact with the interface layer, as seen from the inset of Fig.4.8 (a)



(a)



(b)

Figure 4.8. (a) STEM Z-contrast image and (b) EDX line scan results of the HfO_2 stack. The EDX were scanned across the line highlighted in (a), and the inset of (a) shows HRTEM image, revealing Ge NCs in contact with the interface layer.

On the other hand, we can observe from the STEM of Fig. 4.9 that, the HfAlO stack consists of three distinctive layers: the top HfAlO, bottom HfAlO, and the middle layer which contains Ge NCs. In Fig. 4.3, the TEM results reveal clearly that Ge NCs are contained in amorphous HfAlO after device fabrication and their size is in the range of 5 – 10 nm. This structure is identical to our design as shown in Fig. 4.2. EDX spectra (not shown here) also confirm that Ge is situated at the middle of the dielectric layer whereas Hf and Al distributes across the entire dielectric layer. In addition, the EDX also reveals that the concentration of Hf in the middle bright layer is higher than those in the top and bottom layers, and Al concentration in the middle layer (layer which was co-sputtered) is lower than those in the tunneling and blocking oxide layers. This is identical to our design scheme, in that we co-sputter the HfO_2, Al_2O_3 and Ge with column ratio 3:1:1 to form the middle layer, and the tunneling and blocking HfAlO are formed by ALD with a ratio of Hf to Al of 1:1.

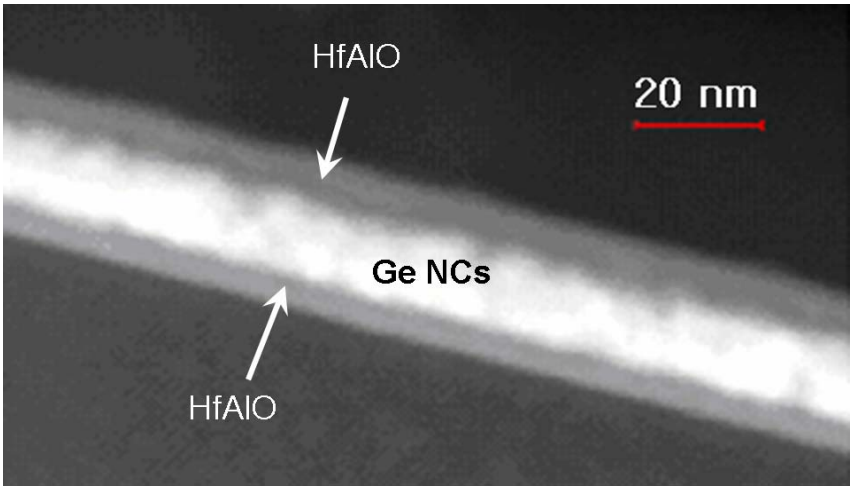


Figure 4.9. STEM Z-contrast image of the Ge+HfAlO stack.

The formation of Ge NCs in HfO_2 can be thermodynamically explained using the Gibbs free energy change of the reaction as aforementioned explanation for Ge NCs in HfAlO,

$$GeO_2\,(cr) + Hf\,(cr) \rightarrow HfO_2\,(cr) + Ge\,(cr),\ (\Delta G_0 \approx -149\ \text{kcal/mol}). \quad (1)$$

As indicated by the reduction of Gibbs free energy in the reaction (1), elemental Ge is formed when the temperature is high enough to drive the reaction [11].

The difference of the distribution of Ge NCs between in HfO_2 and in HfAlO is mainly related to the fact that HfO_2 crystallizes after annealing whereas HfAlO remains amorphous [12]. Therefore, during the crystallization of the Ge NCs, the poly-crystalline HfO_2 may release Ge toward its grain boundaries to result in relocation of Ge NCs, and some Ge NCs end up near the substrate. Since Ge NCs are the charge storage media, their distribution certainly affects the charge retention properties of the memory device. Assuming the direct tunneling (DT) is the main leakage mechanism during retention for our devices, the leakage current under the retention mode can be expressed as $J_{DT} = AE^2 \exp\left(\frac{-B[1-(1-qV_{ox}/\Phi_B)^{1.5}]}{E_{ox}}\right)$ where $A = \frac{q^3(m/m_{ox})}{8\pi h\Phi_B}$ and $B = \frac{8\pi\sqrt{2m_{ox}\Phi_B^3}}{3qh}$ [13].

The equation can be simplified as $J_{DT} \approx AE^2 \exp[(-1.5Bq/\Phi_B)\bullet t]$ by a binomial expansion by neglecting higher order terms, where m_{ox} is the effective carrier mass in the dielectric, Φ_B is the tunneling barrier height, t is the distance between the charge storage position and the channel. It can be understood straightforwardly from the equation that the closer the charges stored to the channel, the higher charge leakage rate. In our experiment, the charges are lost quickly when there are Ge NCs situated close to the

substrate in the device using HfO_2 as the dielectric; while the Ge NCs in the HfAlO device retain the charges efficiently because the charges are well isolated from the substrate by the bottom HfAlO layer to maintain sufficient distance from the channel.

4.5 Summary

In conclusion, we demonstrated formation of Ge NCs in HfAlO high-*k* dielectric. Results show that Ge NCs are thermally stable in HfAlO matrix. A nonvolatile memory device employing Ge NCs embedded in HfAlO high-*k* dielectric was fabricated and showed excellent memory performance. Additionally, it was found NCs formed in HfO_2 is different from that in HfAlO. When NCs were formed in HfAlO, the NCs remained isolated from the substrate and this resulted in the excellent charge retention property. When the Ge NCs were formed in HfO_2, they segregate at grain boundaries of HfO_2. Some of them ended up close to the Si substrate and resulted in fast charge leakage in the memory device.

References

[1] J. J. Lee, X. Wang, W. Bai, N. Lu, J. Liu, and D. L. Kwong, Proc. " Theoretical and experimental investigation of Si nanocrystal memory device with HfO_2 High-k tunneling dielectric", in *Proc. VLSI Technol. Symp*., pp. 33-34, 2003.

[2] D. W. Kim, T. Kim, and S. K. Banerjee, "Memory characterization of SiGe quantum dot flash memories with HfO_2 and SiO_2 tunneling dielectrics", *IEEE Trans. Electron Devices,* vol. 50, pp.1823-1829, 2003.

[3] W. K. Choi, W. K. Chim, C.L.Heng, L. W. Teo, Vincent Ho, V. Ng, D.A. Antoniadis and E.A. Fitzgerald, "Observation of memory effect in germanium nanocrystals embedded in an amorphous silicon oxide matrix of a metal–insulator–semiconductor structure" *Appl. Phys. Lett.*, vol. 80, pp. 2014-2016, 2002.

[4] J. F. Moulder, W. F. Stickle, P. E. Sobol, and K. D. Bomben, Handbook of X-ray Photoelectron Spectroscopy, (Perkin-Elmer, Eden Prairie, MN, 1992).

[5] H. K. Liou, P. Mei, U. Gennser, and E. S. Yang, "Effects of Ge concentration on SiGe oxidation behavior", *Appl. Phys. Lett.*, vol. 59, pp.1200-1202, 1991.

[6] CRC Handbook of Chemistry and Physics, 78th ed., edited by D.R. Lide, H.P.R. Frederiks (CRC, Boca Raton, 1997).

[7] Y. C. King, T. J. King, and C. Hu, "Charge-trap memory device fabrication by oxidation of $Si_{1-x}Ge_x$", *IEEE Trans. Electron Devices,* vol. 48, pp.696-700, 2001.

[8] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices", *J. Vac. Sci. Technol.* B 18, pp.1785-1791, 2000.

[9] J. Chen, Y. Q. Wang, W. J. Yoo, Y. C. Yeo, G. Samudra, D. S. H. Chan, A. Y. Du, and D. L. Kwong, "Nonvolatile flash memory device using Ge nanocrystals

embedded in HfAlO high-k tunneling and control oxides: device fabrication and electrical performance", *IEEE Trans. Electron Devices*, vol. 51, pp.1840-1848, 2004.

[10] S. Wang, W. Liu, Q. Wan, J,Y. Dai, P. F. Lee, S. Luo, Q. Shen, M. zhang, Z. Song, and C. Lin, "Investigation of Ge nanocrytals in a metal-insulator-semiconductor structure with a HfO_2/SiO_2 stack as the tunnel dielectric", *Appl. Phys. Lett.,* vol. 86, pp.113105 1-3, 2004.

[11] Y. Q. Wang, J. H. Chen, W. J. Yoo, Y.-C. Yeo, S. J. Kim, R. Gupta, Z. Y. L. Tan, D. L. Kwong, A. Y. Du, and N. Balasubramanian, "Formation of Ge nanocrystals in HfAlO high-k dielectric and application in memory device", *Appl. Phys. Lett.* vol. 84, pp.5407-5409, 2004.

[12] Y. N. Tan, W. K. Chim, B. J. Cho, and W. K. Choi, "Over-erase phenomenon in SONOS-type Flash memory and its minimization using a hafnium oxide charge storage layer", *IEEE Trans. Electron Devices*, vol. 51, pp.1143-1147, 2004.

[13] D. K. Schroder, *Semiconductor Material and device characterization*, (Wiley, New York, 1998).

Chapter 5

Phase Separated HfSiO as Trapping Layer for MONOS-type Memory Application

5.1 Introduction

The poly-crystalline silicon or metal- oxide-nitride-oxide-silicon (SONOS or MONOS) nonvolatile memory structure has received considerable attention for application in electrically-erasable-programmable-read-only-memory (EEPROM) devices [1-3]. The attractiveness of the MONOS flash memory, compared to the current commercial floating gate memory, lies in its low programming/erasing voltage, immunity to drain-induced turn-on [4], and improved retention and endurance properties because the local defect related charge leakage can be significantly reduced [2,3]. Recently, new materials such as high-permittivity (high-*k*) dielectrics Ta_2O_5, HfO_2 and Al_2O_3 have been explored to improve both the programming and retention properties comparing to silicon nitride [1-3].

Hafnium silicate dielectric film is an attractive trapping material in MONOS memory application due to its process compatibility with conventional CMOS process. Especially, hafnium silicate is known to undergo phase separation [5-8] to form a crystalline hafnium oxide (HfO_2) phase and an amorphous silica rich phase when annealed at temperatures above 900^oC, giving rise to a unique microstructure comprising

HfO_2 nanocrystals that could be suitable for memory applications. Therefore, a memory device with a hafnium silicate trapping layer could combine the benefits of the MONOS structure and the nanocrystal memory. Recently, Lin *et al.* [9] demonstrated memory devices comprising HfO_2 dots formed by phase separation of the hafnium silicate film.

In this chapter, we shall focus on the materials analysis of the dielectric stack in the memory device structure (silicate film deposited on SiO_2 or sandwiched by SiO_2) because the top and bottom SiO_2 may impact the thermal stability of the silicate film. We also provide experimental comparison of the trapping properties of the dual phase HfO_2-$Hf_xSi_{1-x}O_2$ (DPHSO) film with those of the HfO_2 and Si_3N_4 films in MONOS type memories, in order to better evaluate the advantage of each material in MONOS type memory applications. The programming and retention properties are compared among these devices. The relation between the device performance and materials property is analyzed.

5.2 Device Fabrication

Memory devices with the TaN/SiO_2/X/SiO_2/Si structure were fabricated, where X represents the charge trapping layer which can be DPHSO, HfO_2, or Si_3N_4. A schematic of the device structure is shown in Fig. 5.1. Tantalum nitride (TaN) is selected as the gate material, instead of n+ poly-silicon, because the mid-gap work function of TaN has been reported to contribute to an enlarged memory window by reducing electron injection from the gate through the blocking oxide during erase operation [10].

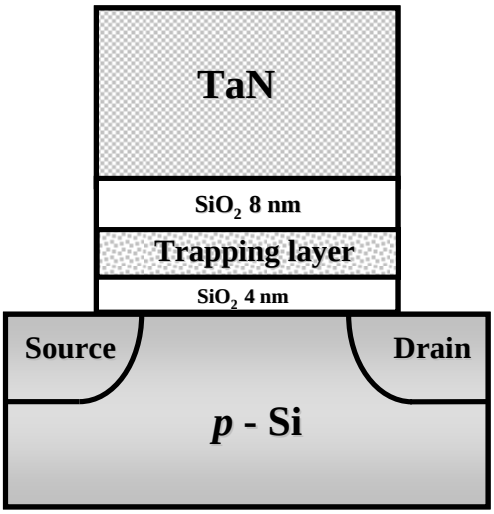


Figure 5.1. Schematic showing the cross-section of the memory device.

P-type Si (100) substrates were used as the starting materials. After a standard pre-gate clean, a 4 nm thick SiO_2 tunnel oxide was grown by dry oxidation at a temperature of 800 ^oC. On one wafer, a $Hf_{0.5}Si_{0.5}O_2$ film was deposited on the SiO_2 tunnel oxide by the co-sputtering of Hf and Si targets in an argon and oxygen (Ar + O_2) ambient at a pressure of 3 mTorr. On the first control wafer, HfO_2 film was formed on the tunnel oxide by reactive sputtering of a Hf target in the Ar + O_2 ambient. On the second control wafer, a Si_3N_4 film was formed on the tunnel oxide by sputtering Si in an ambient containing argon and nitrogen. The thicknesses of the $Hf_{0.5}Si_{0.5}O_2$, HfO_2, and Si_3N_4 films were controlled by adjusting the sputter deposition rate. The physical thicknesses of the $Hf_{0.5}Si_{0.5}O_2$, HfO_2, and Si_3N_4 charge trapping layers were measured by an ellipsometer and a profiler and the results are given in Table 5.1. The thicknesses were different as no optimization was performed to ensure controllability of the sputter deposition rate. Next, all the charge-trapping layers were capped with a 8 nm thick SiO_2 which was formed by

low-pressure chemical vapor deposition (LPCVD) using TEOS ($Si(OC_2H_5)_4$) precursor at a deposition temperature of 700°C. Subsequently, a 150 nm thick TaN metal gate was formed by sputtering a Ta target in an Ar + N_2 ambient with a DC power of 450W at a pressure of 3mTorr, followed by gate patterning and etching. Finally, As^+ implantation and activation anneal at 1000°C were performed to form the source/drain regions.

Table 5.1. Thicknesses of the trapping layer in Fig. 5.1 and the capacitance effective thicknesses (t_{eff}) of the entire dielectric stack which are calculated for the accumulation regime.

	$Hf_{0.5}Si_{0.5}O_2$	HfO_2	Si_3N_4
Trapping layer thickness (nm)	**6**	**8**	**9**
Capacitance effective thickness t_{eff} (nm)	**13.9**	**13.6**	**17.8**

5.3 Materials Characterization

X-ray photoelectron spectroscopy (XPS) and transmission electron microscopy (TEM) were utilized to investigate the material properties of $Hf_{0.5}Si_{0.5}O_2$ film, paying particular attention to phase transformation and distribution.

The $Hf_{0.5}Si_{0.5}O_2$ film employed for XPS characterization was formed by the same process as that used for device fabrication. The $Hf_{0.5}Si_{0.5}O_2$ film was then annealed at various temperatures in an inert ambient, and XPS was used to reveal the change of the

chemical composition of the $Hf_{0.5}Si_{0.5}O_2$ film as the anneal temperature increases. The ratio of Hf and Si in the film was found to be 1:1 from compositional XPS analysis. Figure 5.2 shows the spectra of the O 1s core level of the $Hf_{0.5}Si_{0.5}O_2$ films as-deposited, and after annealing at 900°C and 1000°C.

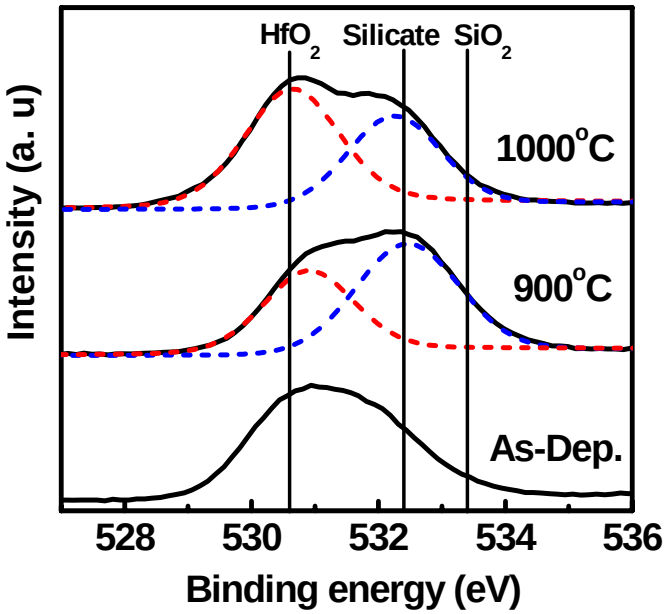


Figure 5.2. XPS spectra showing the O 1s core level of as-deposited (As-Dep.) $Hf_{0.5}Si_{0.5}O_2$ film and $Hf_{0.5}Si_{0.5}O_2$ films after annealing at 900°C and 1000°C. High temperature anneal leads to the formation of two phases, including the HfO_2 phase and the Hf-silicate phase.

The O 1s binding energies of SiO_2 and HfO_2 are 533.4 eV and 530.6 eV respectively [11], as indicated in Fig. 5.2. The O 1s binding energy of an amorphous hafnium silicate with a composition of $Hf_6Si_{29}O_{65}$ is about 532.4 eV [12], and this is also indicated in Fig. 5.2. Considering the fact that Zr and Hf have similar electronic configurations, one may refer to Ref. 13 for an interpretation of XPS spectra. Ref. 13 gives a detailed analysis of the XPS spectra shape and peak positions of Zr-silicate films.

In our experiment, the difference in the XPS spectra of as-deposited samples and samples annealed at high temperatures confirms the phase separation, showing two types of chemical composites in the annealed $Hf_{0.5}Si_{0.5}O_2$ films. By fitting XPS spectra using two Gaussian profiles as shown in Fig. 5.2, it is determined that the two components in the annealed (900^{o}C or 1000^{o}C) samples are HfO_2 and hafnium silicate. The difference between the 900^{o}C and 1000^{o}C annealed samples is that the ratio of O atoms associated with HfO_2 is larger in the 1000^{o}C annealed sample, indicating further phase separation of $Hf_{0.5}Si_{0.5}O_2$ film at 1000^{o}C. Hence, the XPS study reveals HfO_2 and silica rich Hf-silicate phases in the $Hf_{0.5}Si_{0.5}O_2$ film after high temperature annealing.

A cross-sectional TEM picture of a $SiO_2/Hf_{0.5}Si_{0.5}O_2/SiO_2$ sandwich structure annealed at 900^{o}C is shown in Fig. 5.3. Microstructures showing crystalline phases embedded in an amorphous matrix are clearly observed. The size of the crystals is in the range of 5 - 10 nm. Since the kinetic energy barrier to form crystalline hafnium silicate is large, the crystalline silicate phase can never be observed under normal annealing conditions [5-8], thus we believe the crystals in Fig. 5.3 should be HfO_2. In addition, the lattice constant measured from TEM is about 0.31 nm, which is close to the (111) lattice constant of a monoclinic HfO_2 crystal. The amorphous phase in Fig. 5.3 is considered to be a hafnium silicate matrix with a significant brightness contrast from that of the underlying SiO_2. For the sample which was annealed at 1000^{o}C, two phases were also observed in the trapping layer by TEM, but the size of the crystals is larger than that observed in the sample annealed at 900^{o}C, and the brightness constrast between the amorphous phase and the underlying SiO_2 layer is also lower. This indicates that further phase separation had occurred with annealing at a higher temperature.

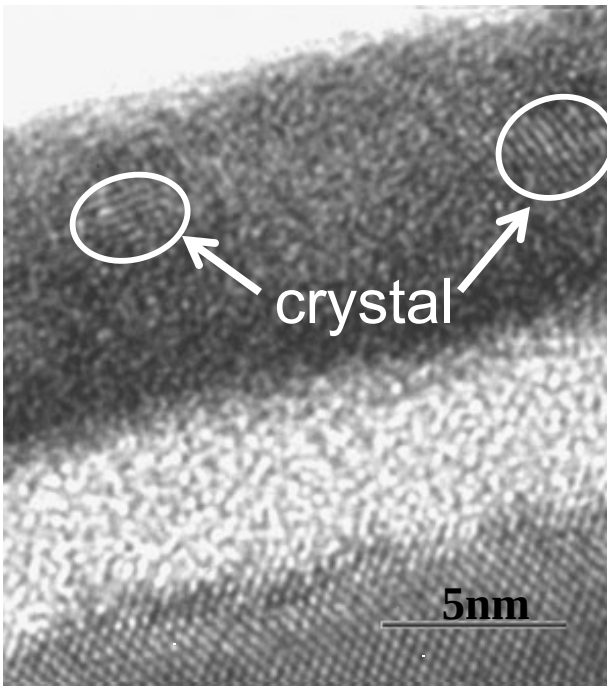


Figure 5.3. TEM image of a $SiO_2/Hf_{0.5}Si_{0.5}O_2/SiO_2$ dielectric stack structure that was annealed at 900°C, revealing microstructure of crystals embedded in an amorphous matrix.

The XPS and TEM analyses consistently indicate that the following reaction occurred when the $Hf_{0.5}Si_{0.5}O_2$ film was annealed at high temperatures:

$$Hf_{0.5}Si_{0.5}O_2 \rightarrow \frac{0.5-x}{1-x} HfO_2 + \frac{0.5}{1-x} Hf_xSi_{1-x}O_2 \quad (1)$$

where $0 < x < 0.5$. Lin *et al*. reported phase separation of $Hf_xSi_{1-x}O_2$ into crystalline HfO_2 and amorphous SiO_2 [9], which can be expressed by $Hf_xSi_{1-x}O_2 \rightarrow x\ HfO_2 + (1-x)\ SiO_2$, where the $Hf_xSi_{1-x}O_2$ film was in contact with the Si substrate. Their different observation from the results of this work could be due to the different initial material compositions or the different layer structures employed in materials characterization.

5.4 Memory Operation and Results Discussion

A. Memory effect and programming characteristics

Figure 5.4 shows the memory effect of a MONOS type device employing $Hf_{0.5}Si_{0.5}O_2$ as trapping layer. Since the devices underwent a source-drain activation anneal at 1000°C, the $Hf_{0.5}Si_{0.5}O_2$ film has a DPHSO structure. After applying a positive voltage pulse to the gate electrode while keeping the source, drain, and substrate terminals grounded, the threshold voltage (V_{th}) shifts to a higher value since net negative charge is injected from the Si substrate. The threshold voltage was determined by the intercept of a tangent of the I_d-V_g curve (measured at V_{ds} = 0.2 V) at the point where the transconductance is the maximum.

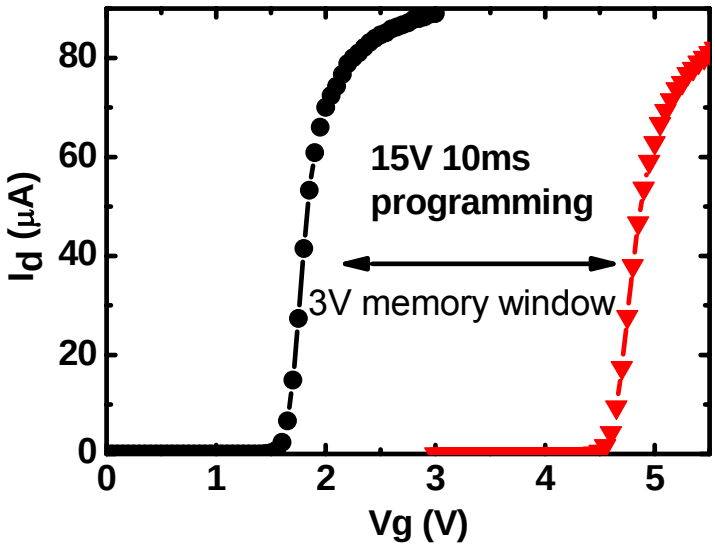


Figure 5.4. Memory window of the device with the dual phase HfO_2-$Hf_xSi_{1-x}O_2$ (DPHSO) trapping layer.

The threshold voltage as a function of stress time is shown in Fig. 5.5. According to the International Technology Roadmap for Semiconductors (ITRS) 2004 [14], the requirement of memory window for flash non-volatile memory is larger than 3V. In our device, 3V can be attained by stressing at 15V for 10 ms.

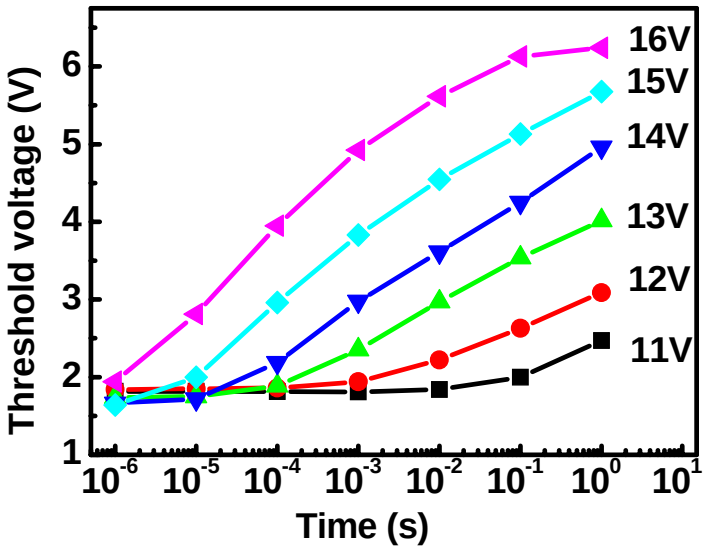


Figure 5.5. Threshold voltage change as a function of programming time and programming voltage of the memory device with the dual phase HfO_2-$Hf_xSi_{1-x}O_2$ (DPHSO) trapping layer.

The programming properties of the device with the DPHSO trapping layer were compared with devices employing HfO_2 and Si_3N_4 trapping layers. It was found that the devices with DPHSO and HfO_2 trapping layers offer higher programming speeds with the lower programming voltages than the device with Si_3N_4 trapping layer.

Table 5.1 also lists the capacitance effective thickness (t_{eff}) of the three devices, which was calculated using $t_{eff} = \frac{A \cdot \varepsilon_{ox}}{C_{eff}}$ [15], where C_{eff} is the capacitance obtained from the capacitance-voltage (*C*-*V*) characteristics in the accumulation region, ε_{ox} is the permittivity of SiO_2, and *A* is the gate area. The t_{eff} of the devices with HfO_2 and DPHSO trapping layers are about the same, but smaller than that of the device with Si_3N_4 trapping layer. This is partly due to the higher dielectric constants of HfO_2 and $Hf_xSi_{1-x}O_2$. The electric field in the tunneling oxide of the fresh device can be estimated by $E_{ox} = \frac{V_g - V_{fb} - \varphi_s}{t_{eff}}$ [15], where E_{ox} is the electric field in tunnel oxide; V_g, V_{fb}, and φ_s are the applied gate voltage, the flat-band voltage, and the surface potential, respectively. Therefore, for a given electric field in the tunneling oxide, the applied gate voltage is smaller for the HfO_2 and DPHSO devices compared to the Si_3N_4 device.

Figure 5.6 compares the programming properties of the three devices, with V_g of 14V applied to the devices with HfO_2 or DPHSO trapping layer, and 18V applied to the devices with Si_3N_4 trapping layers. At this bias condition, the value of E_{ox} is about 10 MV/cm and electrons should be injected from the substrate to the trap layer via FN tunneling, as illustrated in the Fig. 5.7. The FN tunneling current density is given by $J_{FN} = AE_{ox}^2 \exp\left(\frac{-B}{E_{ox}}\right)$, where E_{ox} is the electric field through tunneling oxide, A and B are constants as given by $A = \frac{q^3 (m/m_{ox})}{8\pi h \Phi_B}$ and $B = \frac{8\pi\sqrt{2m_{ox}\Phi_B^3}}{3qh}$ [16]. Hence the

tunneling current is only determined by the electric field and the injection current from substrate to trapping layer should be the same for all the three devices.

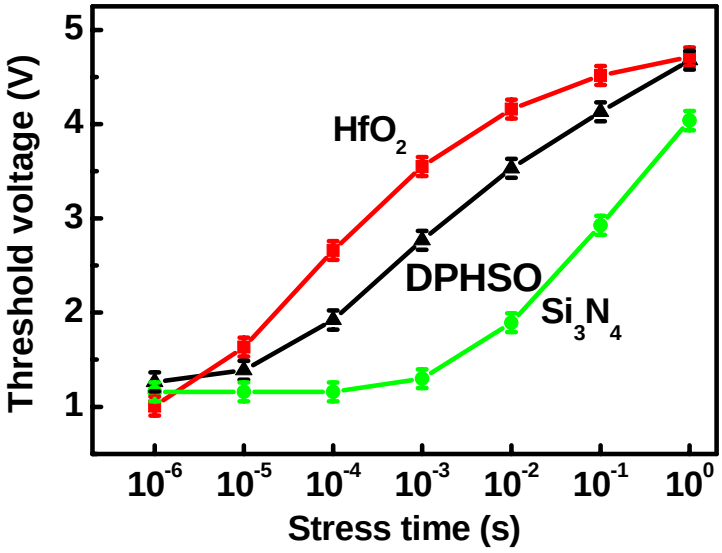


Figure 5.6. Comparison among memory devices with dual phase HfO_2-$Hf_xSi_{1-x}O_2$ (DPHSO), HfO_2 and Si_3N_4 as trapping layer. Electric field E_{ox} of 10 MV/cm was applied across the tunneling oxide in all three devices. Each data point was obtained by measuring 5 devices and the error is within 0.1V across the chip.

The programming speed was determined by $\frac{d\Delta V_{th}}{dt} = \eta \frac{J_{inj}}{C_g}$ [17], where η is capture efficiency, J_{inj} is the injection current, and C_g is the capacitance between a centroid of trapped-charge and the gate and $C_g = \frac{\varepsilon_{ox}}{d}$, where d is the equivalent oxide thickness (EOT) between the trapped charge and the gate electrode. If the centriod of the trapped charge is in the middle of the trapping layer, the values of d for the devices with DPHSO, HfO_2, and Si_3N_4 trapping layers would be 9 nm, 8.8 nm, and 10.9 nm respectively, thus C_g (HfO_2) > C_g (DPHSO) > C_g (Si_3N_4). As stated before, the injection

current of the three devices are nearly the same in Fig. 5.6. Therefore, the faster threshold voltage shift ($\frac{d\Delta V_{th}}{dt}$) of the devices with HfO_2 or DPHSO trapping layer compared with the device with Si_3N_4 trapping layer at the early stage of the programming indicates that the HfO_2 and DPHSO films have a higher capture efficiency (η) than Si_3N_4.

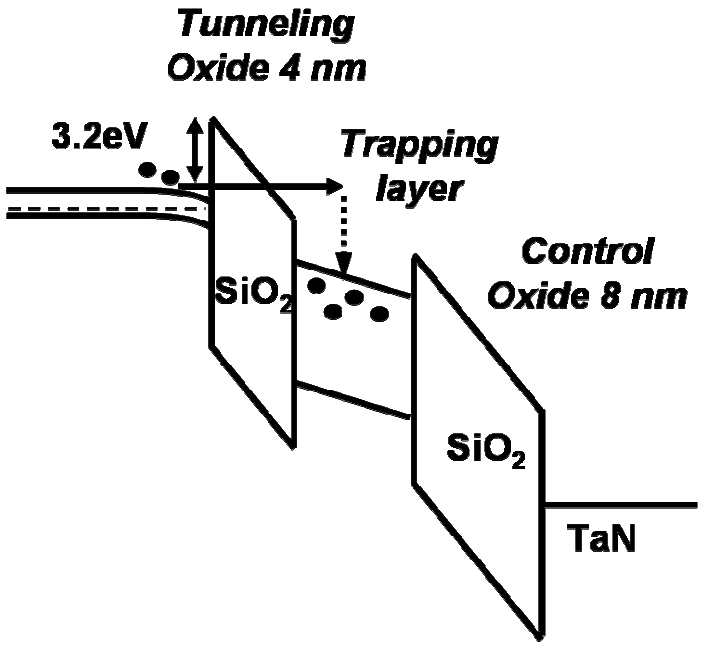


Figure 5.7. Energy band diagram of the MONOS-type device during programming.

B. Retention properties

It is reported that although employing HfO_2 as a trapping layer can lower the programming voltage in comparison with Si_3N_4, the retention property of HfO_2 is poor [3]. The lateral migration of electrons in HfO_2 degrades the retention property, because the grain boundaries can act as lateral conduction paths [3]. In our experiment, we compared the retention properties of the Si_3N_4, HfO_2 and DPHSO films by tracing the flat

band voltage (V_{fb}) shift of the programmed device as a function of time. Figure 5.8 shows the gradual shift of flat band voltage of the three devices with time.

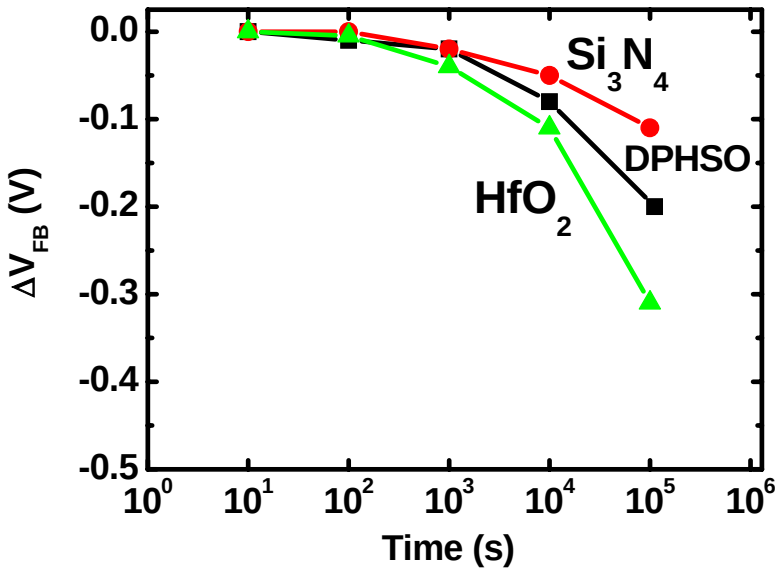


Figure 5.8. Retention characteristics of memory devices with Si_3N_4, dual phase HfO_2-$Hf_xSi_{1-x}O_2$ (DPHSO), and HfO_2 trapping layers.

Si_3N_4 shows the best retention. DPHSO retains the charge better than HfO_2, showing a 0.1 V smaller decay in V_{fb} after 10^5 s. The improvement of DPHSO over HfO_2 is attributed to the presence of the amorphous phase in the film. The amorphous structure does not contain any grain boundaries, and could effectively suppress lateral migration of trapped charges.

5.5 Summary

In summary, the phase separation of the $Hf_{0.5}Si_{0.5}O_2$ film in the $SiO_2/Hf_{0.5}Si_{0.5}O_2/SiO_2$ sandwich dielectric structure was demonstrated. A dual phase structure comprising crystalline HfO_2 and amorphous $Hf_xSi_{1-x}O_2$ was observed after the $Hf_{0.5}Si_{0.5}O_2$ film was annealed at 900^{o}C and 1000^{o}C. The HfO_2-$Hf_xSi_{1-x}O_2$ trapping layer was found to provide a faster programming speed at a lower programming voltage than Si_3N_4, because of its higher dielectric constant and higher trap efficiency. Meanwhile, the HfO_2-$Hf_xSi_{1-x}O_2$ also provided better retention property than HfO_2 because the presence of the amorphous phase suppressed the formation of grain boundary effectively thereby reducing lateral migration. In addition, it was observed that erase of devices using $Hf_{0.5}Si_{0.5}O_2$ and HfO_2 as charge storage media was difficult, and this will be elaborated in Chapter 6 and Chapter 7, respectively. Readers can also find how to improve erase properties of theses devices in Chapter 6 and Chapter 7.

.

References

[1] X. Wang, J. Liu, W. Bai, and D. L. Kwong, " A novel MONOS-type nonvolatile memory using high-k dielectrics for improved data retention and programming speed," *IEEE Trans. Electron Devices*, vol. 51, pp. 597-602, 2004.

[2] T. Sugizaki, M. Kobayashi, M. Ishidao, H. Minakata, M. Yamaguchi, Y. Tamura, Y. Sugiyama, T. Nakanishi, and H. Tanaka, "Novel multi-bit SONOS type flash memory using a high-k charge trapping layer", in *Proc. VLSI Technol. Symp.*, pp.27-28, 2003.

[3] Y. N. Tan, W. K. Chim, B. J. Cho, and W. K. Choi, "Over-erase phenomenon in SONOS-type Flash memory and its minimization using a hafnium oxide charge storage layer," *IEEE Trans. Electron Devices*, vol. 51, pp.1143-1147, 2004.

[4] M. K. Cho and D. Kim, "High performance SONOS memory cells free of drain turn-on and over –erase: compatibility issue with current flash technology," *IEEE Electron Device lett.*, vol. 21, pp.399-402, 2000.

[5] D. A. Neumayer and E. Cartier, "Materials characterization of ZrO_2-SiO_2 and HfO_2 –SiO_2 binary oxides deposited by chemical solution deposition", *J. Appl. Phys.*, vol. 90, pp.1801-1808, 2001.

[6] S. Stemmer, Y. Li, b. Foran, P. S. Lysaght, S. K. Streiffer, P. Fuoss, and S. Seifert, "Grazing-incidence small angle x-ray scattering studies of phase separation in hafnium silicate films", *Appl. Phys. Lett.* 83, pp.3141-3143, 2003.

[7] S. Stemmer, Z. Chen, C. Levi, P. Lysaght, B. Foran, J. Gisby, and J. Taylor, "Application of metastable phase diagrams to silicate thin films for alternative gate dielectrics,", *Jpn. J. Appl. Phys.*, vol. 42, pp.3593-3597, June 2003.

[8] S. Ramanathan, P. C. Mcintyre, J. Luning, P. S. Lysaght, Y. Yang, Z. Chen and S. Stemmer, “Phase separation in hafnium silicated for alternative gate dielectrics” , *J. Electrochem. Soc.*, vol. 150, pp. F173-F177, 2003.

[9] Y. H. Lin, C. H. Chien, C. T. Lin, C. W. Chen, C. Y. Chang, and T. F. Lei, “High performance multi-bit nonvolatile HfO_2 nanocrystal memory using spinodal phase separation of hafnium silicate”, in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp.1080- 1082, 2004.

[10] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, “A novel SONOS structure of $SiO_2/SiN/Al_2O_3$ with TaN metal gate for multi-giga bit flash memories,” in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp. 26.5.1 - 26.5.4, 2003.

[11] S. J. Wang, P. C. Lim, A. C. H. Huan, C.L. Liu, J. W. Chai, S. Y. Chow, J. S. Pan, Q. Li, and C. K. Ong, “Reaction of SiO_2 with hafnium oxide in low oxygen pressure”, *Appl. Phys. Lett.*, vol. 82, pp.2047-2049, 2003.

[12] G. D. Wilk, R. M. Wallace, and J. M. Anthony, “Hafnium an zirconium silicates for advanced gate dielectrics”, *J. Appl. Phys.*, vol. 87, pp.484-492, 2000.

[13] G. B. Rayner, Jr. D. Kang, Y. Zhang, and G. Lucovsky, “Nonlinear composition dependence of x-ray photoelectron spectroscopy and auger electron spectroscopy features in plasma-deposited zirconium silicate alloy thin films,” *J. Vac. Sci. Techno.*, vol. B 20, pp. 1748-1758, 2002.

[14] International Technology Roadmap for Semiconductors. [On-line] http://public.itrs.net/.

[15] S. S. Chung, P. Y. Chiang, G. Chou, C. T. Huang, P. Chen, C. H. Chu, and C. C. H. Hsu, " A novel leakage current separation technique in direct tunneling regime gate oxide SONOS memory cell", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp. 26.6.1 - 26.6.4, 2003.

[16] D. K. Schroder, *Semiconductor Material and device characterization*, (Wiley, New York, 1998).

[17] K. Nomoto, I. Fujiwara, H. Aozasa, T. Terano, and T. Kobayashi, IEDM Tech. Dig., "Analytical model of the programming characteristics of scaled MONOS memories with a variety of trap densities and a proposal of a trap-density-modulated MONS memory", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp.13.5.1-13.5.4, 2001.

Chapter 6

IrO_2/HfAlO/HfSiO/HfAlO Gate Stack for Memory Application

6.1 Introduction

SONOS/MONOS (polysilicon/metal-SiO_2–Si_3N_4–SiO_2–silicon) memory devices are very attractive for electrically erasable programmable read-only memory (EEPROM) application because of its advantages of low program/erase voltage and improved retention properties compared to conventional floating gate memory. Especially, the NAND-type SONOS memories using F-N tunneling programming and erasing are very promising for the mass storage application due to its aggressive scalability and low power consumption [1].

However, there is a trade-off between fast erase and good retention for NAND-type SONOS memory because during erasing, the electron current from the gate electrode across blocking oxide is competed with the current across the tunneling oxide, therefore to achieve higher erasing speed, thinner tunneling oxide must be used and the retention will be degraded. Extensive researches have been done to improve both the properties of erasing and retention of the SONOS type memory, by tailoring trapping layer, tunneling and blocking oxide or the gate electrode.

In Chapter 5, we compared the trapping properties among Si_3N_4, HfO_2 and HfSiO based on MONOS memory structures, and found that HfSiO is a superior trapping material because it provides lower operation voltage than Si_3N_4 and better retention than HfO_2 [2]. However, improvement in electrical performance of nonvolatile memory devices requires extensive optimization of the whole gate stack. Wang *et.al* [3] proposed to use high-*k* HfO_2 as the tunneling and blocking oxide instead of SiO_2 to enable low voltage (4-7 V) operation as well as to attain good retention performance because the low barrier of the HfO_2 can provide high tunneling current with physically larger thickness. However, the memory window of their device is below 1 V. Reisinger *et. al.* [4] and Lee *et. al.* [5] demonstrated improved erasing performance by using p^+ poly-Si gate or TaN gate in place of n^+ poly-Si gate, because the tunneling current from the gate is suppressed by the higher barrier of p^+ poly-Si and TaN gate than that of n^+ poly-Si gate.

In this chapter, we propose an optimized gate stack IrO_2/HfAlO/HfSiO/HfAlO for NAND memory application. By combining the advantages of high-*k* tunneling and blocking oxide (HfAlO), good trapping ability of HfSiO and high work function of IrO_2, it is expected to attain the device properties of low operation voltage, fast erasing speed as well as long retention simultaneously. Here, a unique gate material IrO_2 is selected not only due to its high work function (5.2-5.3 eV) but also its process integration advantage: in contrast to other p-type metal gates such as Pt and Ni, IrO_2 can be plasma-etched and it is also superior to p^+ poly-Si gate in that it doesn't require an additional implantation mask when integrating into NMOS memory cells.

6.2 Experimental Details

Three types of devices were fabricated in the experiment. The cross sectional schematics of the devices are shown in Fig. 6.1. The first device follow the conventional MONOS type structure as a control wafer; the second device uses the high-*k* HfAlO as the tunneling and blocking oxide instead of SiO_2 in the first type of device; and the third wafer uses IrO_2 as the control gate instead of the TaN in the second device.

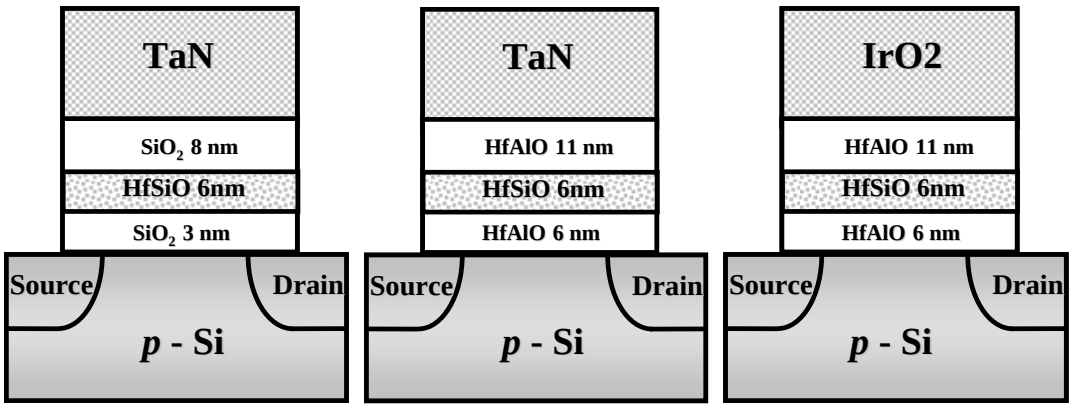


Figure 6.1. Cross sectional schematics of three memory devices fabricated with different gate stacks.

The process flow is shown in Table 6.1. All devices are fabricated using p-type Si substrate. After pregate clean, a thermal oxide of 3nm was grown on one wafer (S1: $TaN/SiO_2/HfSiO/SiO_2$), followed by the co-sputtering of Hf and Si in an O_2/Ar ambient to form HfSiO and post deposition anneal (PDA) at 900^oC for 30 s to form the phase-separated structure. Then an 8nm thick TEOS SiO_2 and a sputtered TaN was deposited. On the other two wafers (S2: TaN/HfAlO/HfSiO/HfAlO and S3: IrO_2/HfAlO/HfSiO/ HfAlO), HfAlO was deposited by co-sputtering of Hf and Al in an O_2/Ar ambient,

followed by PDA at 700°C for 30s. Subsequently, HfSiO was sputtered and annealed to form phase-separated trapping layer. An 11nm thick HfAlO blocking layer was then sputtered and annealed as the blocking oxide. Finally, the metal gates of TaN and IrO_2 were sputtered on the wafers S2 and S3 respectively. After the formation of gate stacks, all three wafers are patterned, implanted by As and undergone activation at 900°C for 30s.

Table 6.1. Process flow of three devices S1, S2 and S3

Device name	S1	S2	S3
Dielectric stack deposition	Thermal oxide 3nm	Sputtering HfAlO 6nm and PDA at 700°C for 30s	
	Sputtering HfSiO 6nm and PDA at 900°C for 30s Phase separated HfSiO film is formed		
	TEOS 8nm oxide	Sputtering HfAlO 11nm and PDA at 700°C for 30s	
Gate deposition	TaN sputtering	TaN sputtering	IrO_2 sputtering
Gate pattern	Gate patterning and etching		
S/D formation	S/D implantation and activation		

6.3 Program/Erase Characteristics of Memory Devices

As stated in section 5.2, HfSiO film separates into two phases HfO_2 and silicate after annealing at 900°C for 30s. Fig. 6.2 shows a plan-view TEM of the HfSiO film after phase separation with dark dots corresponds to the HfO_2 nanocrystals and light amorphous areas correspond to silicate phase, respectively.

Figures 6.3 and 6.4 show the program/erase properties of device S1 ($TaN/SiO_2/HfSiO/SiO_2$). The capacitance equivalent thickness (CET) of the whole stack

estimated from the accumulation capacitance is 13.5nm, and the device can be operated at 13-15V.

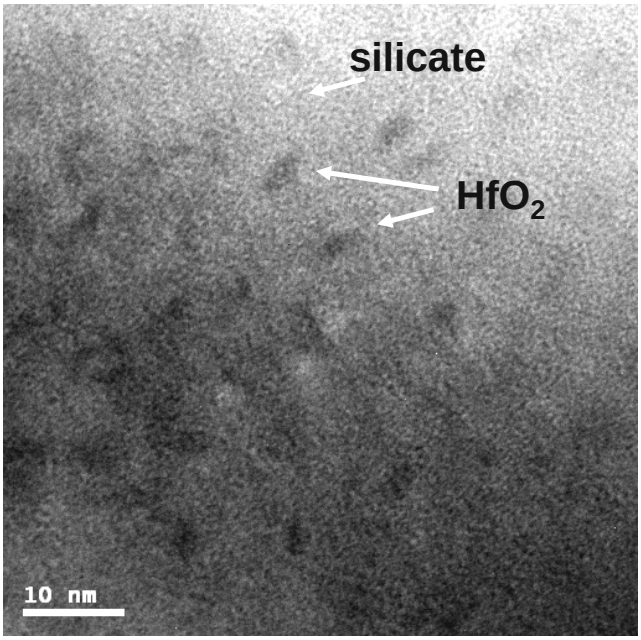


Figure 6.2. TEM plan view of the silicate film after phase separation. The dark dots represent HfO_2 crystal and light amorphous areas represent silicate phases.

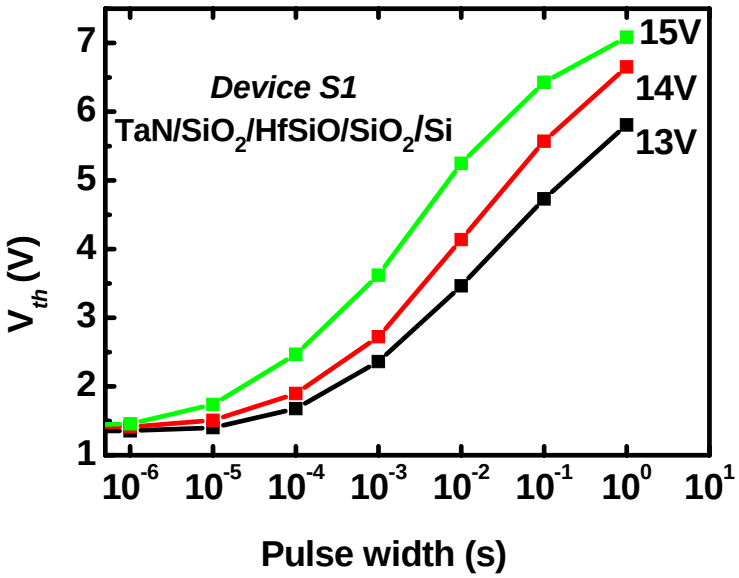


Figure 6.3. Program characteristics of device S1.

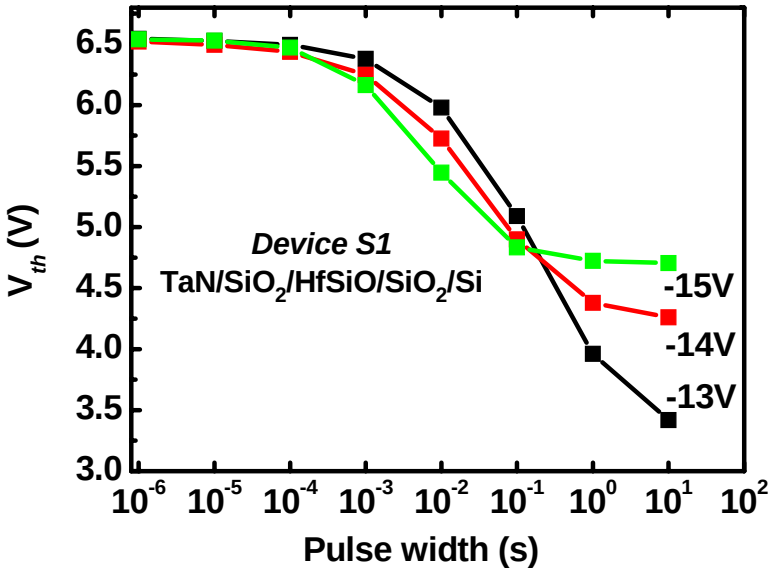


Figure 6.4. Erase characteristics of device S1. Saturation of V_{th} is clearly observed and it increases with increasing erasing voltage.

In comparison, the CET of the whole stack of device S2 (TaN/HfAlO/HfSiO/HfAlO) and S3 (IrO_2/HfAlO/HfSiO/ HfAlO) are about 7.5nm, and the device can be operated at 10-12V, as shown in Fig. 6.5 and Fig. 6.6. Besides the lower operation voltage, S2 and S3 device also offers higher program and erase speeds, as shown in Fig. 6.5 and Fig. 6.6. The lower voltage operation and higher program/erase speed of the device S2 and S3 are attributed to the utilization of the high-*k* HfAlO as tunneling and blocking oxide, which offers lower CET as well as the lower tunneling barrier than SiO_2.

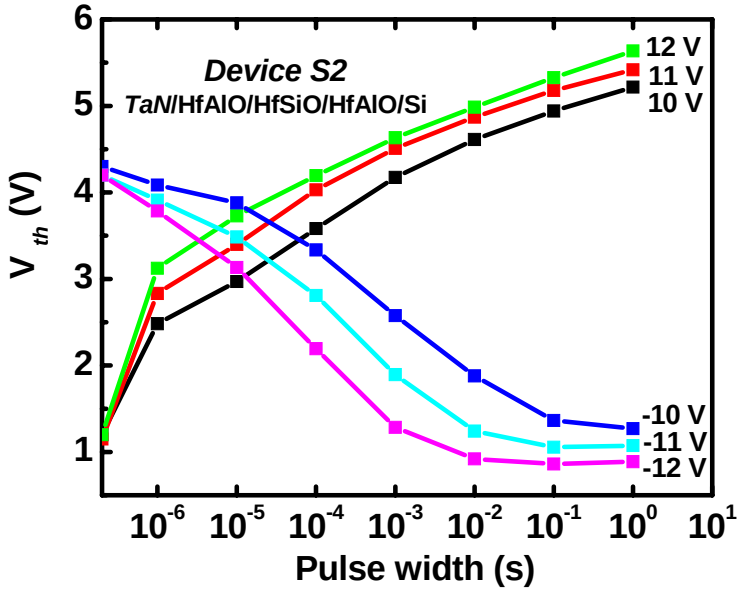


Figure 6.5. Program and erase characteristics of device S2. No increase of saturation V_{th} is observed when increasing erasing voltage.

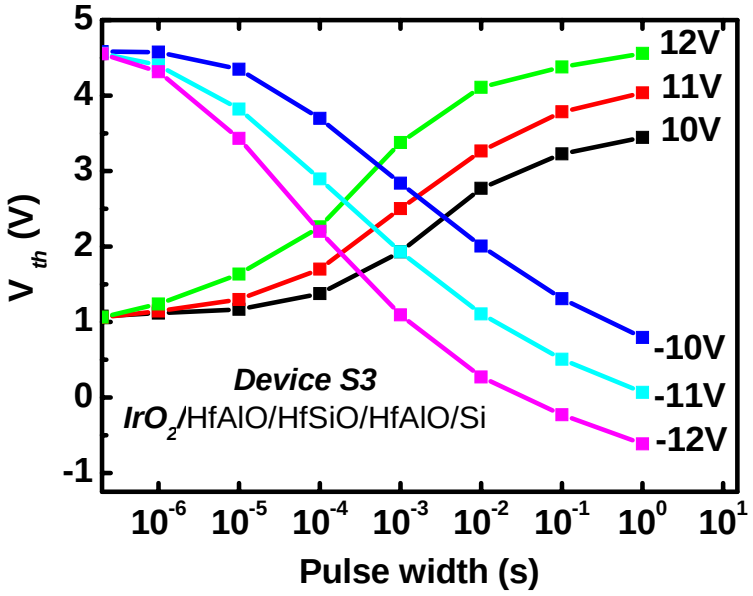


Figure 6.6. Program and erase characteristics of device S3. No erase saturation is observed, which is distinguished from S1 in Fig. 6.4.

6.4 Discussion on the Erase Saturation of Memory Devices

It is interesting to note that erase characteristics among the three devices are different. For S1, V_{th} saturates fast and the saturation level increases with increasing erasing voltage during erasing, which results in a trade-off between erasing speed and memory window. This is a general phenomenon observed for SONOS type memory and is caused by unwanted gate electron injection. In contrast, S2 and S3 do not show the same phenomenon. V_{th} of S2 saturates during erasing, but the saturation level does not increase with increasing of the erasing voltage. S3 presents the best erasing property: no erase saturation is observed and the change of V_{th} is distinct in the V_g range from -10 to -12 V.

Figure 6.7 illustrates the band diagrams of S1 under the steady state of erasing operation. The steady state is a state in which V_{th} saturates because the electron current from the gate electrode across blocking oxide J_e is equal to the holes injection current from Si substrate through the tunneling oxide J_h. It is reported that when erasing, for short times, the erasing is dominated by the trapped electrons back tunneling; for long time, the current across the tunneling oxide at the steady state can be dominated by either electron back tunneling at high electrical field or hole tunneling at mediate electric field [6]. Here, for simplifying the case, we only discuss the hole tunneling which is likely more suitable for our operation voltage range.

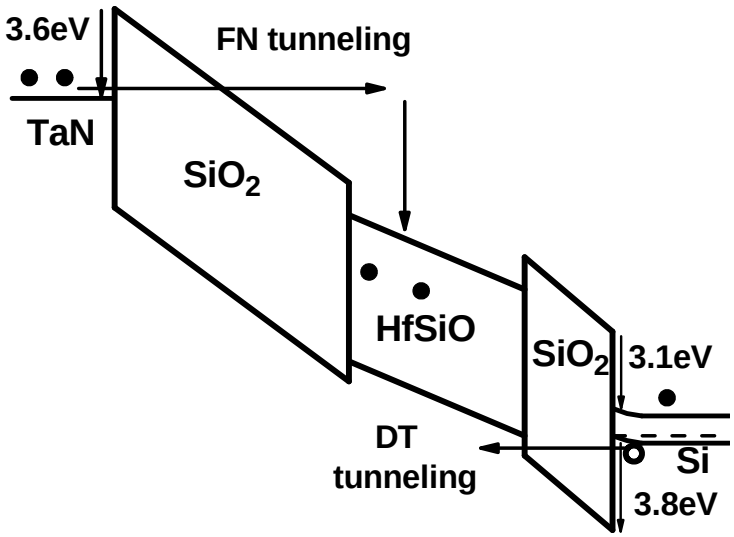


Figure 6.7. Band diagrams of devices S1 in the steady state of erasing operation. Electrons are injected by FN tunneling and holes are injected by direct tunneling (DT).

Figure 6.8 shows the theoretically calculated currents of electrons injected from gate (J_e) and holes injected from Si substrate (J_h) as a function of electric field for S1. In the range of erasing voltage shown in Fig. 6.7, J_e is primarily resulted from FN tunneling whereas J_h is from direct tunneling (DT) at the steady state. FN current is calculated by equation $J_{FN} = AE_{ox}^2 \exp\left(\frac{-B}{E_{ox}}\right)$, $A = \frac{q^3(m/m_{ox})}{8\pi h\Phi_B}$ and $B = \frac{8\pi\sqrt{2m_{ox}\Phi_B^3}}{3qh}$, where the m is the mass of a free electron, q is the electronic charge, $\hbar$ is the reduced Planck's constant, E_{ox} is the electric field in the tunneling SiO_2, d_{ox} is the thickness of SiO_2, m_{ox} is the effective mass of the carrier, and Φ_B is the tunneling barrier. DT current is calculated

according to the equation $J_{DT} = AE_{ox}^2 \exp\left(\frac{-B(1-(1-qE_{ox}d_{ox}/\Phi_B)^{1.5})}{E_{ox}}\right)$. The parameters for the calculation are listed in Table 6.2.

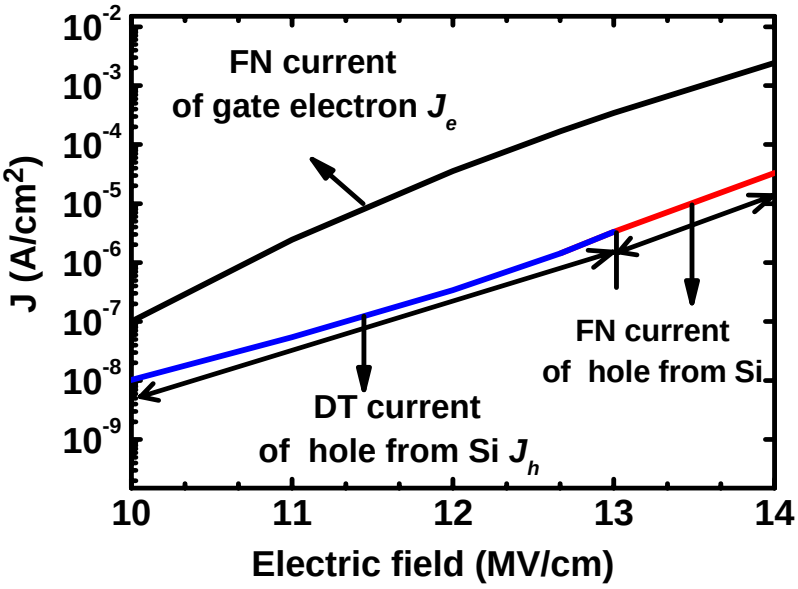


Figure 6.8. Theoretically calculated gate electron current and hole current from substrate as a function of electric field for device S1. The formula and constants for calculation are referred to [4].

Table 6.2. Parameters used for calculating J_e and J_h

	J_e	**J_h**
m_{ox}	**0.5** m_o	**0.6** m_o
Φ_B	**3.6 eV**	**3.8 eV**

The charges stored in the trapping layer at the steady state can be estimated by $Q = \varepsilon(E_b - E_t)$ when erasing, where E_b and E_t are the absolute values of electrical field across the blocking oxide and the tunneling oxide, respectively, if we consider the memory device as two capacitors C_b and C_t in series as shown in Fig. 6.9, where the C_b represents the capacitor in the side of the blocking oxide and C_t represents the capacitor in the side of the tunneling oxide. It is noted from the Fig.6.7 that the increase of the FN current J_e is faster than that of DT current J_h. Therefore, when the erasing gate voltage increases, the saturated current will increase and the $(E_b - E_t)$ decreases (becomes more negative), indicating more electrons or less holes will be stored and the saturation V_{th} will increase and result in decreasing of memory window. In other words, when the erasing voltage increases the increase of J_e dominates, resulting in more electrons injected and stored in the trapping layer to enhance J_h and to suppress J_e to reach a new steady state with higher V_{th} level.

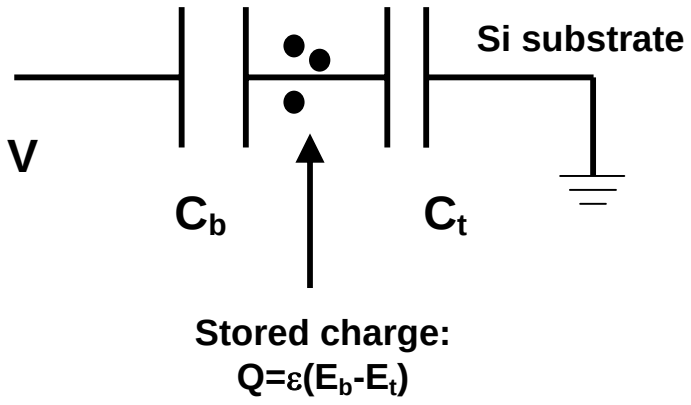


Figure 6.9. Schematically equivalent circuit of a memory device.

Fig. 6.10 illustrates the band diagrams of S2 under the steady state of erasing operation. The conductance and valence band offset of HfAlO versus Si are calculated by ΔE_c=2.37-0.46x (eV) and ΔE_v=3.03-0.81x eV, where x is the atomic ratio of $(HfO_2)_x(Al_2O_3)_{1-x}$ [7]. Taking account of x=0.5, the conduction band and valence band offset of the HfAlO is about 2.14eV and 2.62eV respectively. The work function of TaN in HfAlO is 4.55 eV, therefore the band offset of TaN to HfAlO is about 2.64eV. Due to the reduced barrier of HfAlO compared to the SiO_2, the electrons across the blocking oxide and holes across the tunneling oxide should be resulted from the same mechanism, most likely the trap assistant tunneling or F-N tunneling; therefore, the sensitivity of the two currents with respect to the electric field should be similar. In other words, when the erasing voltage increases, the increase of the current of gate electron injection J_e and substrate hole injection J_h should be similar. A small decrease of saturation V_{th} level in Fig. 6.5 indicates a small amount of net positive charges are injected and stored in the steady state when the erasing voltage increases, and the memory widow is also slightly enlarged.

The erasing properties of S3 become even better when the IrO_2 gate is used. Note that the barrier of the gate electron injection J_e for S3 becomes 3.15 eV (assuming the work function of IrO_2 is 5.2V) which is much larger than the barrier 2.6 eV of hole tunneling current from the substrate J_h; hence the gate electron injection are efficiently suppressed. Fig. 6.11 shows lower gate leakage current observed from S3 (with IrO_2 gate) than that from S2 (with TaN gate), which confirms the higher work function of IrO_2 gate than TaN gate. Thanks to the high work function of the gate electrode, not only its erase speed of device S3 increases with increasing of erase voltage, but also does the memory

window. Additionally, the erased V_{th} level becomes more distinct, which is favorable for the memory application.

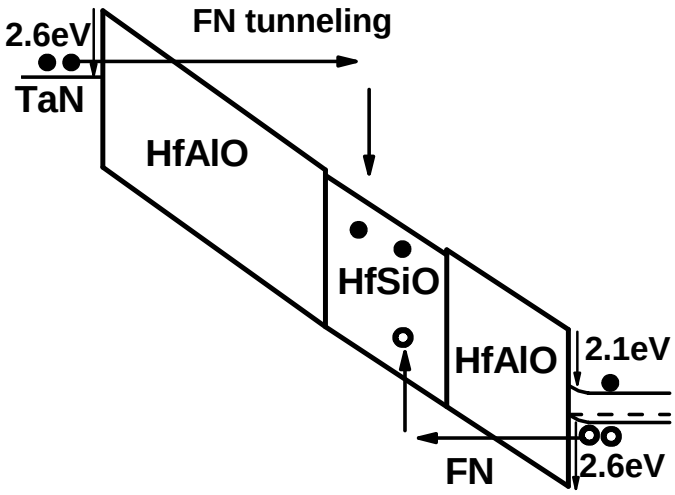


Figure 6.10. Band diagrams of devices S2 in the steady state of erasing operation. Both electrons and holes may be injected by FN tunneling.

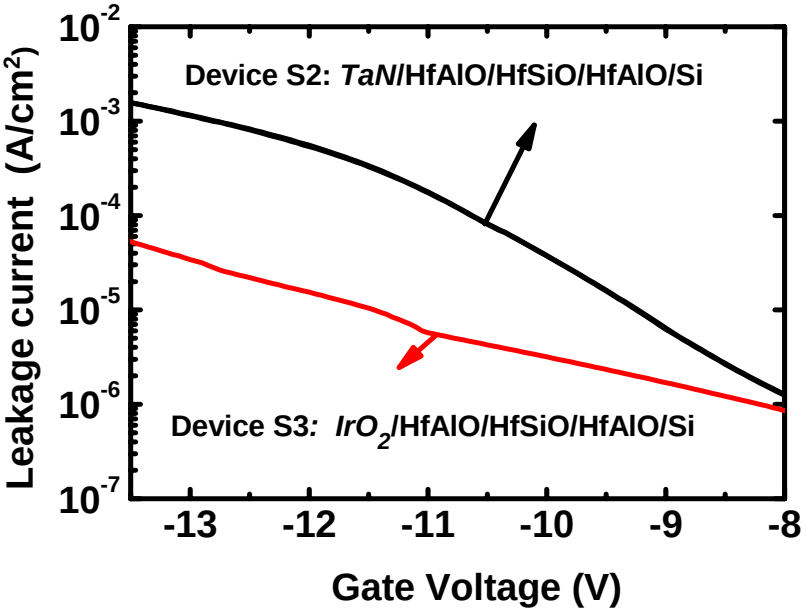


Figure 6.11. Comparison of leakage current between S2 and S3. Lower leakage of S3 is observed due to the higher work function of IrO_2 than that of TaN.

A comparison of programming/erasing properties between S2 and S3 at 12/-12V reveals that the range of operational threshold voltage (V_{th}) of S3 is lower than that of S2 (Fig. 6.12) due to the better erasing ability of S3. This can lead to a lower reading voltage which is favorable to device operation.

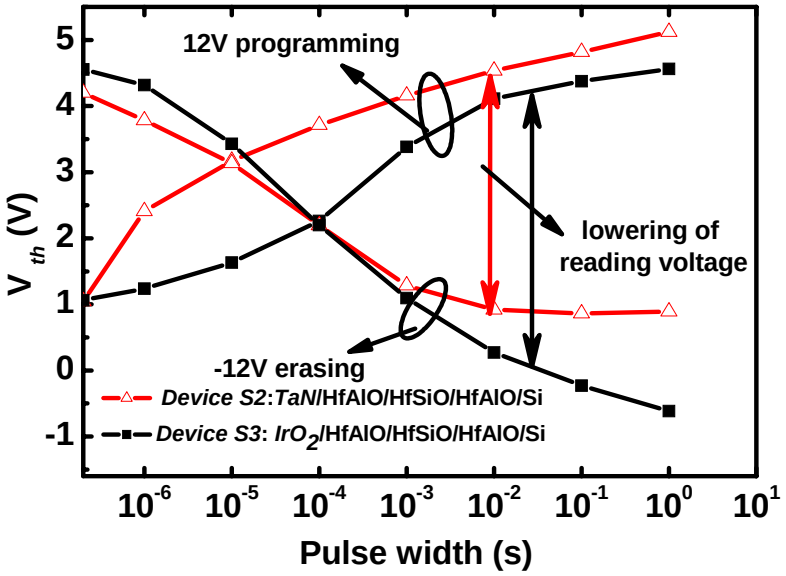


Figure 6.12. Comparison of program/erase properties between device S2 and S3. IrO_2 demonstrates lower V_{th} range than TaN when operating at the same voltage, enabling lowering of reading voltage.

6.5 Retention and Endurance Properties

The retention properties of device S2 and S3 at room temperature are compared in Fig. 6.13. A logarithm degradation of the programming V_{th} with a decay rate 0.2 V/dec was observed from S2, whereas S3 shows very little degradation. The extrapolation of the retention curve up to 10 years for S3 shows reduction in memory window is 18% at room temperature. The reason of the improved retention of device S3 could be that the

high work function gate electrode IrO_2 provides a high initial V_{th} of virgin device; hence fewer charges are required to be stored for a given programmed V_{th} compared to a low work function gate electrode. Figure 6.14 shows the retention properties of devices at 85°C. A degradation of only 28% was found at 85°C upon 10 years extrapolation. This is much better than the data published in recent 2 years, as shown in Table 6.3. According to ITRS, 80% memory window is required to remain after 10 years. The retention of device IrO_2/HfAlO/HfSiO/HfAlO is good enough to meet that requirement.

Figure 6.15 shows endurance properties of S2 and S3. In summary, S3 (IrO_2/HfAlO/HfSiO/HfAlO) demonstrates the best electrical performance overall, considering retention, operating voltage and programming/erasing speed.

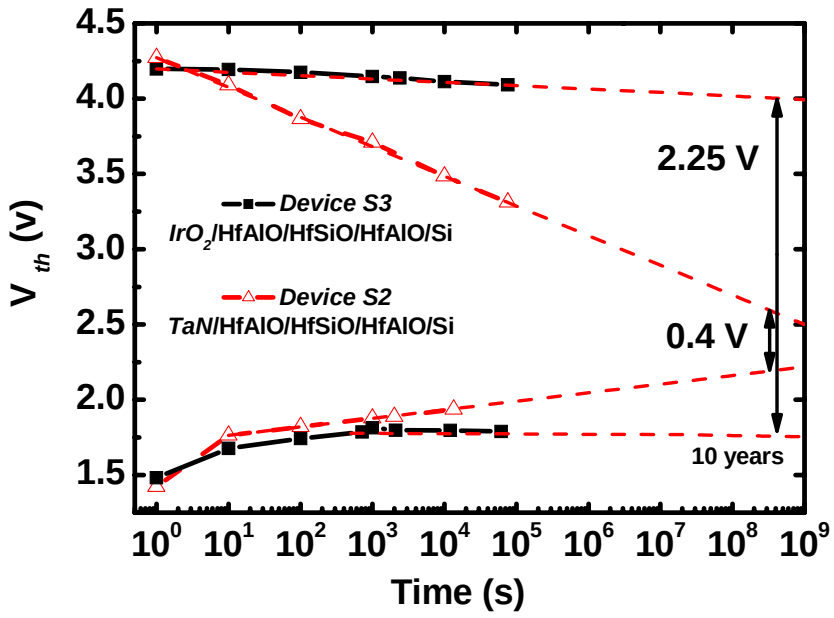


Figure 6.13. Comparison of retention properties between devices S2 and S3 at room temperature. Device S3 shows lower charge loss rate that device S2.

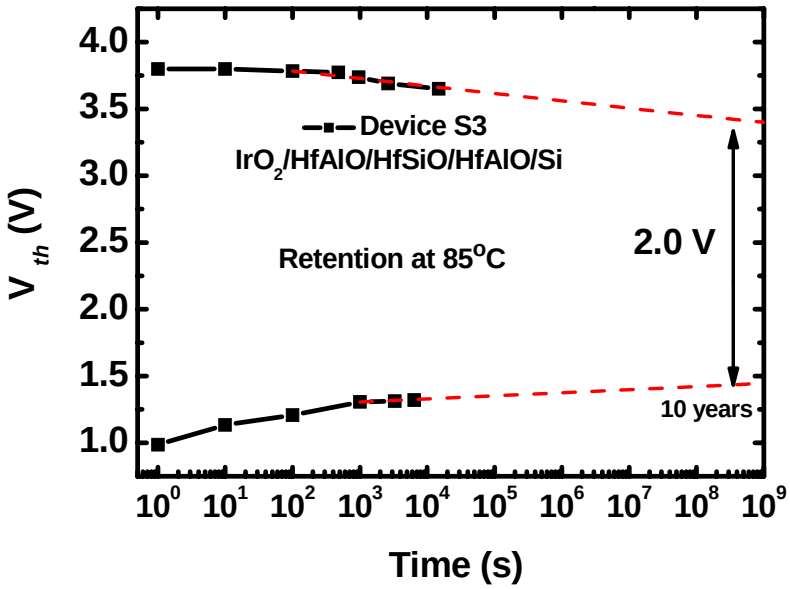


Figure 6.14. Retention of device S3 at temperature of 85ºC. It is predicted that 72% memory window is retained after 10 years.

Table 6.3. Comparison of program, erase and retention properties of this work to other works published in recent 2 years

	Program to gain 3V V_{th} shift		Erase to gain 3V V_{th} shift		85ºC Retention		
	Gate & drain bias	Pulse width	Gate & drain bias	Pulse width	Initial ΔV_{th}	ΔV_{th} after 10 years	***Decay after 10 years***
This work: IrO_2/HfAlO/HfSiO/HfAlO		10ms		0.5ms	2.8V	2V	**28%**
IrO_2 /HfAlO/AlN/SiO_2 [9]	V_g =12V	0.1ms	V_g=-12V	0.1ms	3.7V	1.9V	49%
TaN/Al_2O_3/Si_3N_4/SiO_2 [5]	V_d =0V	0.1ms	V_d=0V	3ms	4.3V	2.07V	52%
FinFET SiO_2/Si_3N_4/SiO_2 [10]		0.2µs		5µs	4.5V	2.4V	47%
Poly/SiO_2/HfSiO/SiO_2 [11]	V_g= 9V V_d= 5V	2ms	V_g = -5V V_d = 10V	2.3V for 1ms	Not shown		

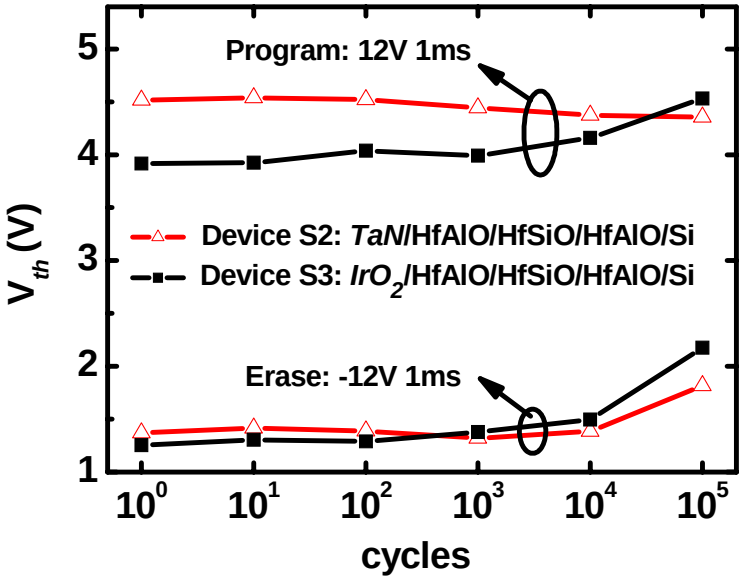


Figure 6.15. Endurance characteristics of devices S2 and S3.

6.6 Summary

In this work, the effects of high-*k* tunneling and blocking oxide and high work function gate electrode on the SONOS type memory device are studied. High-*k* HfAlO tunneling and blocking oxide reduce operation voltage, improve erase saturation compared to the conventional SONOS structure. The device with IrO_2 gate electrode together with HfAlO tunneling and blocking oxide shows no erase saturation at all and additionally, it is found that the implementation of IrO_2 helps to achieve long retention time. Furthermore, the high-*k* stack proposed in this work can be deposited in-situ by one clustered CVD, and this can be a clear advantage over other stacks requiring mixture of various processes, e.g. thermally grown SiO_2 or PVD of high-*k* materials.

References

[1] C. H. Lee, S. H. Hur, Y. C. Shin, J. H. Choi, D. G. Park, and K. Kim, "Charge-trapping device structure of SiO_2/SiN/high-k dielectric Al_2O_3 for high-density flash memory," *Appl. Phy.Lett.*, vol. 86, pp. 152908 1-3, 2005.

[2] Y. Q. Wang, J. H. Chen, W. J. Yoo, Y.-C. Yeo, A. Chin, and A. Y. Du, "Formation of dual-phase HfO_2--$Hf_xSi_{1-x}O_2$ dielectric and its application in memory devices", *J. Appl. Phys*., vol. 98, pp. 013536 1-5, 2005.

[3] X. Wang, J. Liu, W. Bai, D. L. Kwong, "A novel MONOS-type nonvolatile memory using high-k dielectrics for improved data retention and programming speed," *IEEE Trans. Electron Devices,* vol. 51, pp. 597-602, 2004.

[4] H. Reisinger, M. Franosch, B. Hasler, and T. Böhm, "A novel SONOS structure for nonvolatile memories with improved data retention", in *Proc. VLSI Tech. Symp.*, pp. 113-114, 1997.

[5] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp. 26.5.1-26.5.4, 2003.

[6] H. Bachhofer, H.Reisinger, E. Bertagnolli and H. von Philipsborn, "Transient conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures", *J. Appl. Phys.,* vol. 89, pp. 2791-2800, 2001.

[7] H.Y. Yu, M.F. Li, B.J. Cho, C.C. Yeo, M.S. Joo, D.-L. Kwong, J.S. Pan, C.H. Ang and J.Z. Zheng, "Energy gap and band alignment for $(HfO_2)_x(Al_2O_3)_{1-x}$ on (100) Si", *Appl Phys. Lett.,* vol. 81, pp.376-378, 2002.

[8] International Technology Roadmap for Semiconductors. [On-line] http://public.itrs.net/

[9] C.H. Lai, A.C. Chin, K.C. Chiang, W.J. Yoo, C.F. Cheng, S.P. Mcalister, C.C. Chi and P. Wu, "Novel SiO_2/AlN/HfAlO/IrO_2 memory with fast erase, large ΔV_{th} and good retention", in *Proc. VLSI Technol.,* pp.210-211, 2005.

[10] C.W. Oh, S. D. Suk, Y. K. Lee, S. K. Sung, J.-D. Choe, S.-Y. Lee, D. U. Choi, K. H. Yeo, M. S. Kim, S.-M. Kim, M. Li, S. H. Kim, E.-J. Yoon, D.-W. Kim, D. Park, K. Kim, and B.-I. Ryu, "Damascene gate FinFET SONOS memory implemented bulk silicon wafer", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp. 893-896, 2004.

[11] Y.H. Lin, C.H. Chien, C.T. Lin, C.W. Chen, C.Y. Chang and T.F. Lei, "High performance multi-bit nonvolatile HfO_2 nanocrystal memory using spinodal phase separation of hafnium silicate", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig*., pp.1080-1082, 2004.

Chapter 7

Improving Erasing and Reliability of High-*k* Trapping Layer Device Using Si_3N_4/SiO_2 Tunneling Stack

7.1 Introduction

Driven by mobile electronic applications, the market for nonvolatile memory has grown at a very fast rate. As stated in Chapter 1, commercial nonvolatile floating gate flash memory device faces severe scaling challenges beyond the 45 nm technology generation, such as the significantly reduced coupling ratio and the serious floating gate interference [1]. Thus, the metal-oxide-nitride-oxide-silicon (MONOS) type memory device becomes an attractive candidate for further device scaling. The MONOS memory can be fabricated using a simple process, is robust to defect related leakage, and is highly scalable. However, slow erase and poor charge retention could be the main challenges for MONOS devices.

High-*k* materials as a charge storage layer (CSL) for MONOS type memory are advantageous over Si_3N_4 as they provide the same EOT with a larger physical thickness, leading to larger number of traps, higher trapping efficiency as well as faster program speed [2,3]. Recently, engineering of the tunneling barrier has also been explored for MONOS type memory devices [4]. The concept of layered barrier was originally proposed by Likharev for application to floating gate Flash memory [5]. The benefit of

the layered barrier for floating gate Flash memory is that by proper barrier layer design, the current conduction through the barrier can become more sensitive to electric field than that through a single SiO_2 barrier. For example, it was reported that $Si_3N_4/SiO_2/Si_3N_4$ stack can offer higher current at high electric field and lower current at low electric field than a single SiO_2 layer, allowing higher program/erase speed and better retention for floating gate memory devices [6, 7]. Similarly, the use of band-engineered tunnel barrier can also be beneficial for SONOS type memory.

However, no work has been conducted to combine the advantages of the high-*k* trapping layer and the band-engineered tunnel barrier. The exploration of band-engineered tunnel barrier can be especially useful for NAND Flash memories when using high-*k* materials as a trapping layer because use of high-*k* materials as a CSL may faces serious problem in erasing. It was reported that, when using HfO_2 or HfAlO as a CSL and using direct tunneling (DT) mechanism for programming and erasing, the erase of the memory device is suppressed, but this is not the case when Si_3N_4 is used as a trapping layer. The reason is that the valence band offset of HfO_2 or HfAlO with respect to Si, ΔE_v is at least 1 eV larger than that of Si_3N_4 [2]. This reduces hole tunneling current during erasing significantly and results in a limited erasing capability.

The suppressed erase is an advantage for NOR flash operation because the over-erase problem can be avoided, whereas it is not favorable for NAND flash operation. For NAND flash application, the fast erase is desired because fast erase is beneficial to expand the memory window while over-erase is not an issue. In NAND flash memories, cells in a bit line are connected in series. When reading the information of a selected cell, all the unselected cells in the same bit line are applied with a voltage larger than V_{th} of the

programmed state so that they are all turned on. The selected cell is applied with a voltage between V_{th} of programmed state and erased state; thus the status of the bit line is only dependent on the selected cell. The over-erase of any cells does not affect the reading, while the over-program of any unselected cell makes the whole bit line blocked and prevents accurate information from being read from the selected cell.

Considering high-*k* materials as a CSL, most thermally stable high-*k* materials have larger ΔE_v than Si_3N_4; therefore, the limitation of erase can be expected when they are used as a CSL for MONOS type memories. Alternative high-*k* materials with smaller ΔE_v than Si_3N_4 such as TiO_2 may face the process integration challenges due to poorer thermal stability. Consequently, use of high-*k* CSL can be significantly limited; thus the exploration of an engineered tunnel barrier for overcoming the erase problem is necessary.

In this chapter, we explore a layered tunnel barrier (Si_3N_4/SiO_2) for the application of MONOS type memory with a high-*k* HfO_2 CSL. The device has a gate stack of $TaN/Al_2O_3/HfO_2/Si_3N_4/SiO_2/Si$ and we name it the DTL device (**D**ual **T**unneling **L**ayer consisting of Si_3N_4/SiO_2). The program/erase speed and reliability of the DTL device are evaluated in comparison with two types of control devices: one is with the conventional silicon nitride as a charge storage layer, having a $TaN/Al_2O_3/Si_3N_4/SiO_2/Si$ (TANOS) stack structure; and the other is with a high-*k* HfO_2 charge storage layer and a single SiO_2 as a tunneling layer, having a $TaN/Al_2O_3/HfO_2/SiO_2/Si$ (TAHOS) stack structure.

7.2 Theoretical Basis

During erase operation, the magnitude of the hole tunneling current from the Si substrate through the tunneling dielectric is a main factor affecting the erase speed of MONOS type memories. Depending on the vertical electric field, the possible tunneling modes of the hole tunneling is understood to be Fowler-Nordheim (FN) tunneling, direct tunneling (DT), or modified FN (MFN) tunneling, as illustrated in Fig. 7.1 [8, 9].

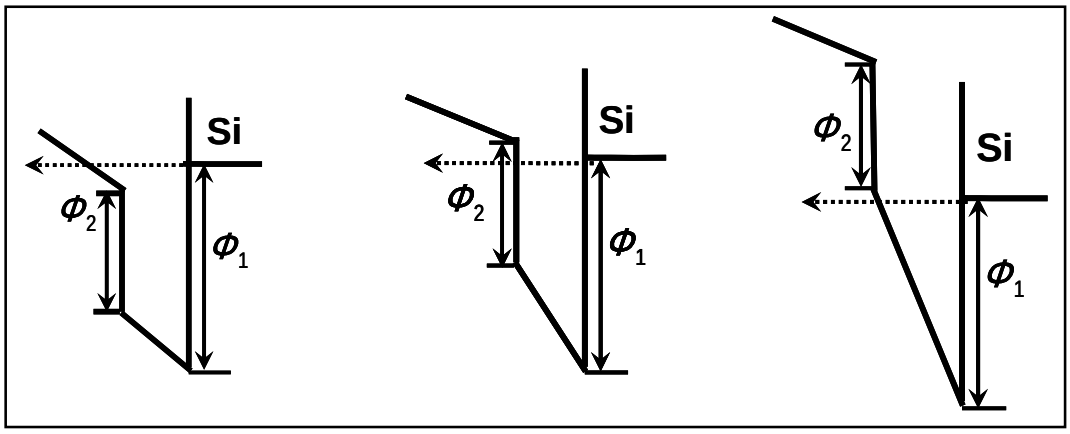


Figure 7.1. Schematics of possible tunneling mechanisms of holes from the substrate. The corresponding electric field range of each tunneling mode is indicated.

The hole tunneling current due to FN tunneling, DT, and MFN tunneling are denoted by J_{FN}, J_{DT}, and J_{MFN}, respectively, and are expressed as follows [8, 9] when we consider the layer on top of Si in Fig. 7.1 is SiO_2:

$$J_{FN} \propto \frac{m_0}{m_{ox}} \frac{q^3}{16\pi^2\hbar} E_{ox}^2 \exp\left(-\frac{4\sqrt{2m_{ox}\phi_1^3}}{3q\hbar E_{ox}} \right) \quad (1)$$

$$J_{DT} \propto \frac{m_0}{m_{ox}} \frac{q^3}{16\pi^2\hbar} E_{ox}^2 \exp\left(-\frac{4\sqrt{2m_{ox}}[(q\phi_1)^{3/2} - (q\phi_1 - qE_{ox}d_{ox}/)^{3/2}]}{3q\hbar E_{ox}} \right) \quad (2)$$

$$J_{MFN} \propto \frac{m_0}{m_{ox}} \frac{q^3}{16\pi^2\hbar} E_{ox}^2 \exp\left(-\frac{4\sqrt{2m_{ox}}[(q\phi_1)^{3/2} - (q\phi_1 - qE_{ox}d_{ox}/)^{3/2}] + r4\sqrt{2m_2}[q\phi_1 - q\phi_2 - qE_{ox}d_{ox}]^{3/2}}{3q\hbar E_{ox}} \right) \quad (3)$$

In the above equations, m_0 is the mass of a free electron, m_{ox} is the hole effective mass in SiO_2, q is the electronic charge, $\hbar$ is the reduced Planck's constant, E_{ox} is the electric field in the tunneling SiO_2, ϕ_1 is the SiO_2–Si barrier height, d_{ox} is the thickness of SiO_2, r is the ratio of the dielectric constant k of the second dielectric which is on top of SiO_2 to that of SiO_2, m_2 is the hole effective mass in the second dielectric, ϕ_2 is the energy difference between the valence band edge of SiO_2 and that of second dielectric. Figure 7.2 shows the calculated current of the stacks Si_3N_4/SiO_2 (4 nm), Si_3N_4/SiO_2 (3 nm), HfO_2/SiO_2 (4 nm) and HfO_2/SiO_2 (3 nm), and Table 7.1 shows the parameters used in the calculation [8, 10]. The calculated result provides an estimate or a qualitative guide for this experimental work, but is not meant to match the experimental current values. A more accurate calculation of the tunneling current requires a more complicated model that considers the influences of traps and the interface scattering on the tunneling, and also requires more accurate values for parameters such as the effective mass of the holes in high-*k* material.

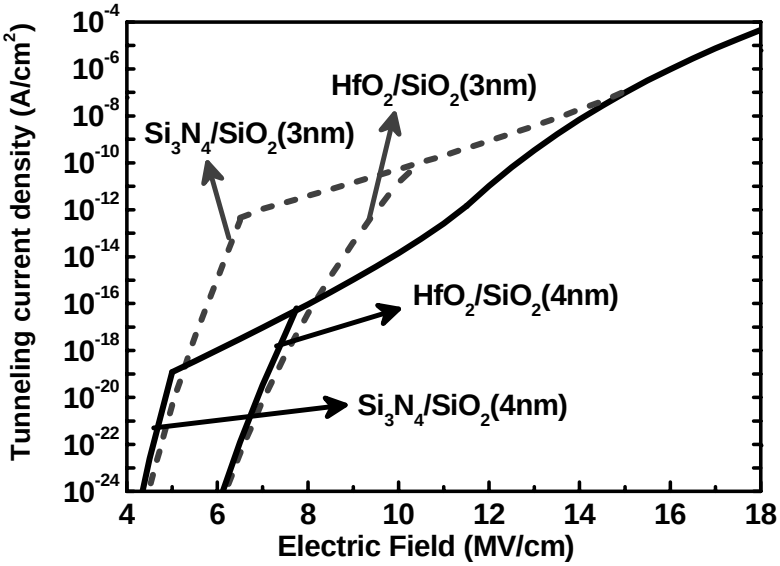


Figure 7.2. Calculated hole tunneling current density through the Si_3N_4/SiO_2 and HfO_2/SiO_2 stacks with different SiO_2 thickness.

Table 7.1. Parameters used for calculation of the current density in Fig.7.2 .

	m_{ox}	ε	ϕ_1
SiO_2	0.6	3.9	4.6
	m_2	ε	ϕ_2
Si_3N_4	0.42	**7**	2.6
HfO_2	0.04	18	1.5

It is understood from Fig. 7.2 that, when comparing two Si_3N_4/SiO_2 dielectric stacks, the one with a thinner SiO_2 layer results in a larger hole current and hence faster erase speed. However, if we simply reduce the oxide thickness to increase the erase speed as well as the program speed, the retention will be adversely affected. As a result, memory device optimization involves more than a simple reduction of the SiO_2 thickness. On the other hand, if we use high-k HfO_2 as a trapping layer instead of Si_3N_4, the hole injection is more suppressed since ΔE_v of HfO_2 is ~1 eV larger than Si_3N_4, despite of the reduction in the tunnel oxide thickness. This leads to the reduction in hole tunneling probability even at high electric fields. This is shown from the calculation results of Fig. 7.2 and is illustrated from energy band diagrams of Fig. 7.3 (a) and (b). Figure 7.3 (a) shows the energy band profile under flat-band condition, while Fig. 7.3 (b) shows the energy band profile during erase. The above analysis indicates that the engineering of the high-k HfO_2 trapping layer or the reduction in the thickness of tunneling oxide alone is insufficient to optimize the memory device. Instead, they have to be engineered together.

A transient layer with a smaller ΔE_v than HfO_2 can be used to replace part of the tunneling SiO_2 as shown in Fig. 7.4 (a) and (b). The barrier for hole tunneling can thus be diminished and the erase speed can be enhanced. At the same time, the retention of the memory device is not compromised since the transient layer provides an additional barrier to block the leakage of charges stored in the high-k material. Si_3N_4 is one of the best material candidates to be used as a transient layer, considering its small ΔE_v, thermal stability, ease of deposition, and familiarity in manufacturing. Note that the other high-k material candidates with lower ΔE_v such as TiO_2 may face serious thermal stability issues.

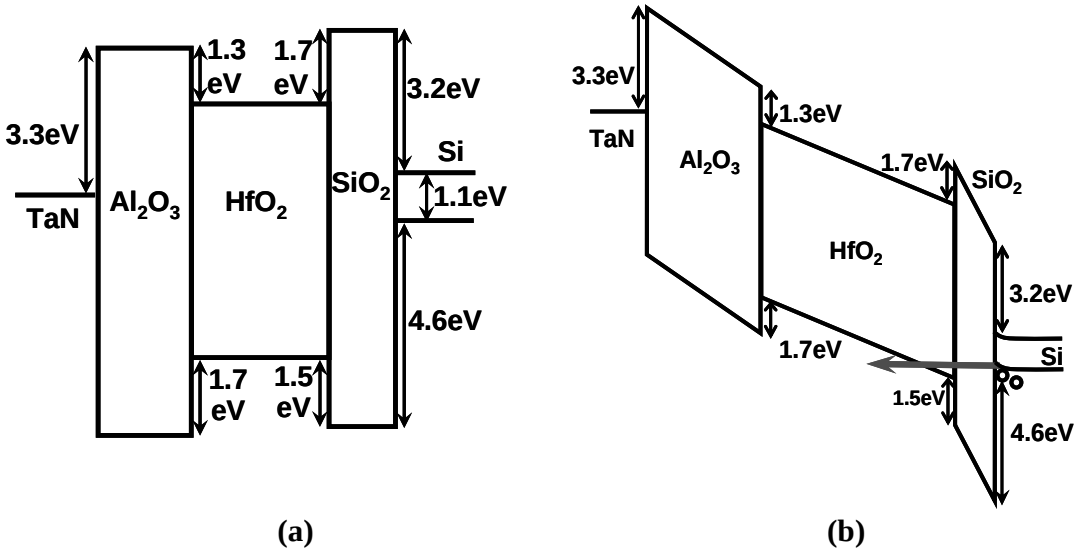


Figure 7.3. Band offsets of TAHOS in the flat band condition (a), and band profiles of TAHOS when erasing (b). The hole tunneling is reduced by the high ΔE_v of HfO_2.

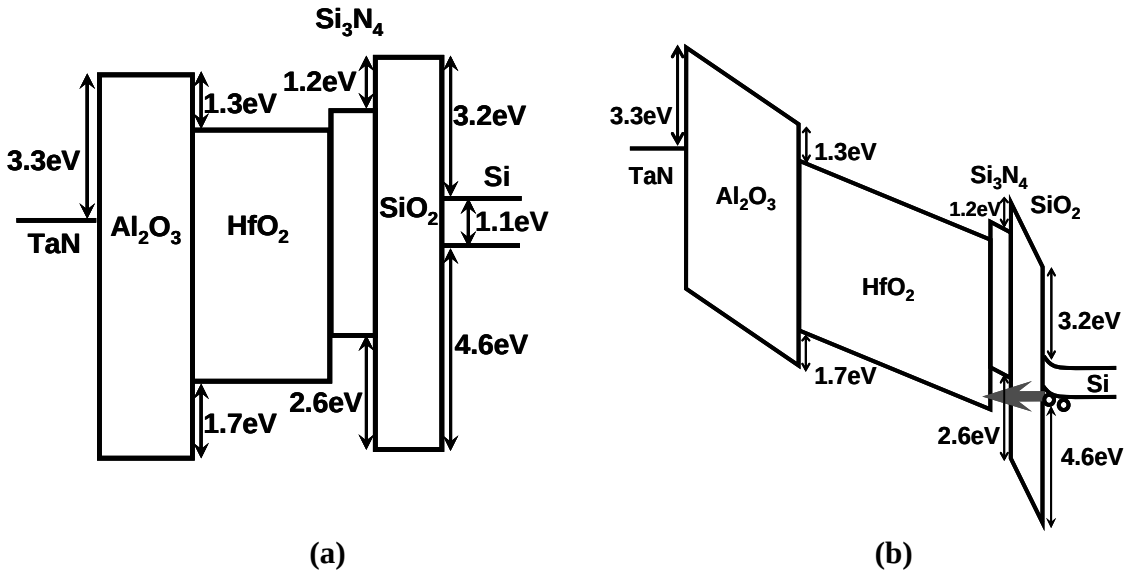


Figure 7.4. Band offsets of DTL in the flat band condition (a), and band profiles of DTL when erasing (b). The hole tunneling in DTL memory is easier than that in TAHOS memory when erasing because of the thinner SiO_2 and the lower ΔE_v of Si_3N_4.

7.3 Experimental Details

Two groups of memory devices were fabricated and details of the device structures including layer materials and thicknesses are documented in Table 7.2. Group 1 has two devices: TAHOS1 and DTL1. They were fabricated to demonstrate the effect of the dual tunneling layer (Si_3N_4/SiO_2). Group 2 has three devices: TAHOS, DTL and TANOS. The tunneling layers of devices in group 2 are thicker than those of devices in group 1. Effect of dual tunneling layer can be confirmed again by the devices in group 2. Moreover, by comparing the performance of DTL, TAHOS with the additional TANOS device, the advantages of DTL can be demonstrated more clearly. In addition, more detail analysis can be made.

Table 7.2. Structures of devices fabricated in the experiment.

	Group 1		**Group 2**		
	TAHOS1	**DTL1**	**TAHOS**	**DTL**	**TANOS**
Gate	TaN				
Blocking Oxide	Al_2O_3 (10.5nm)				
Trapping layer	HfO_2 24nm	HfO_2 16nm	HfO_2 16nm	HfO_2 18nm	Si_3N_4 5nm
Tunneling Oxide	**SiO_2 3.2nm**	**Si_3N_4 1.8nm** / **SiO_2 2.6nm**	**SiO_2 4.1nm**	**Si_3N_4 1.6nm** / **SiO_2 2.9nm**	**SiO_2 4.1nm**
EOT measured by *C-V*	13.5 nm	12.8 nm	13 nm	13.5 nm	12.2 nm

The process flow of the DTL devices is shown in Fig. 7.5. After pre-gate clean of p-type Si substrate, SiO_2 was thermally grown, followed by a thin nitride film deposition by LPCVD. HfO_2 film was then deposited by sputtering Hf in an O_2/Ar ambient followed by a post deposition anneal at 700^{o}C for 30 s. Subsequently, a 10.5 nm thick ALD Al_2O_3 and 150 nm sputtered TaN were deposited. After the formation of the gate stack, the wafers were patterned, implanted by As^+ and underwent activation at 900^{o}C for 30 s. The TAHOS and TANOS wafers were fabricated in a similar way as DTL splits except that the deposition of the tunneling nitride layer was excluded.

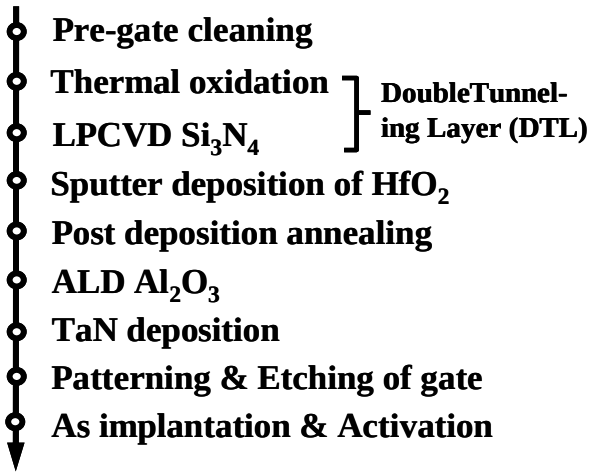


Figure 7.5. Process flow for fabrication of memory device with Dual Tunneling Layer (DTL) Si_3N_4/SiO_2.

7.4 Results and Discussion

This section contains two parts; the first part discusses memory devices in group 1 and the second part focus on memory devices in group 2.

1. Performance of devices in group 1: Ease of erase, improved retention and endurance of DTL1 over TAHOS1

Figure 7.6. shows the program/erase characteristic of TAHOS1. It is observed that the erasing of the memory is slow, demonstrating the difficulty in easing. This is consistent with observation in previous report [2]. In contrast, ease of erasing of DTL1 is observed, as shown in Fig.7.7. The result is although the trapping layer is thicker in TAHOS1 than that in DTL1, DTL1 still shows a larger memory window than TAHOS because of its better erase performance.

It is clear that the improvement of erasing properties of DTL1 over TAHOS can be attributed to two main factors: one is the thinner SiO_2 used for DTL1 than that for TAHOS; the other is ΔE_v of Si_3N_4 is about 1eV smaller than that of HfO_2 as illustrated in Fig. 7.4. These two factors will result in the much higher hole tunneling current for DTL1 than TAHOS1 when erasing, and lead to the better erase performance of DTL1.

Despite of the thinner SiO_2, DTL1 still shows superior charge retention capability compared to TAHOS1, as shown in Fig. 7.8, since the Si_3N_4 provides an additional physical layer to effectively block the charge loss from the HfO_2 trapping layer in the retention state.

Endurance properties of TAHOS1 and DTL1 are compared in Fig. 7.9. The degradation of the erased states is slower for DTL1 than that for TAHOS1, which indicates that the dual-tunneling layer helps to improve the cycling properties.

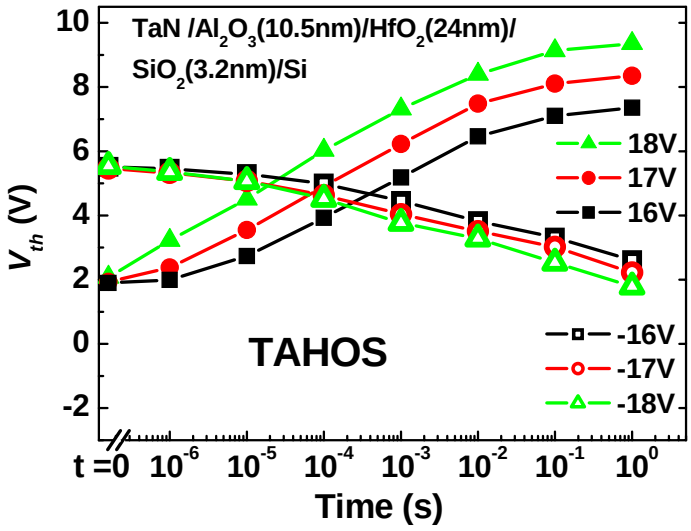


Figure 7.6. Program and erase characteristics of TAHOS1 memory.

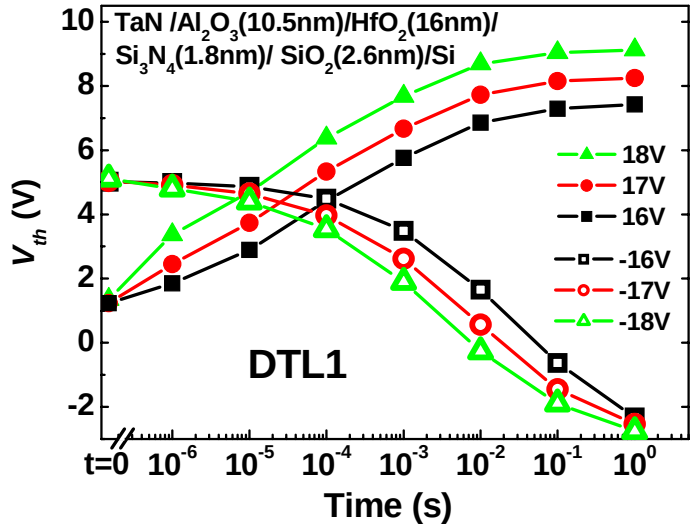


Figure 7.7. Program and erase characteristics of DTL1 memory.

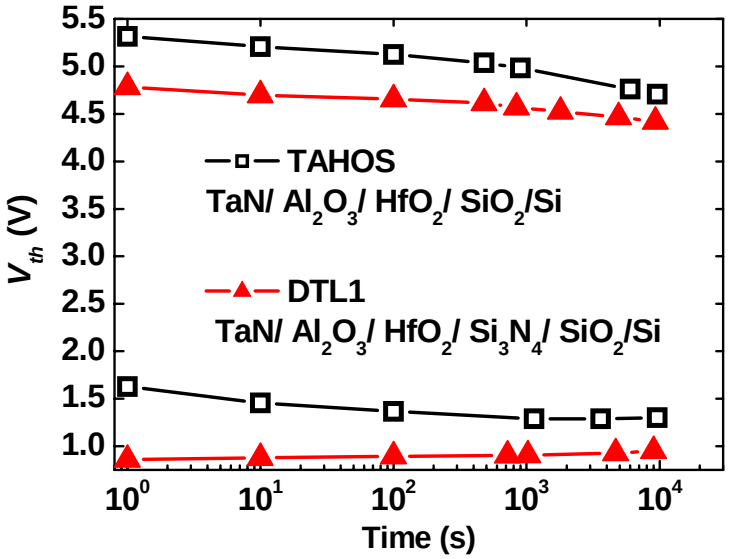


Figure 7.8. Retention properties of TAHOS1 and DTL1 memories at room temperature. The charge retention of DTL1 device is more stable than that of TAHOS1.

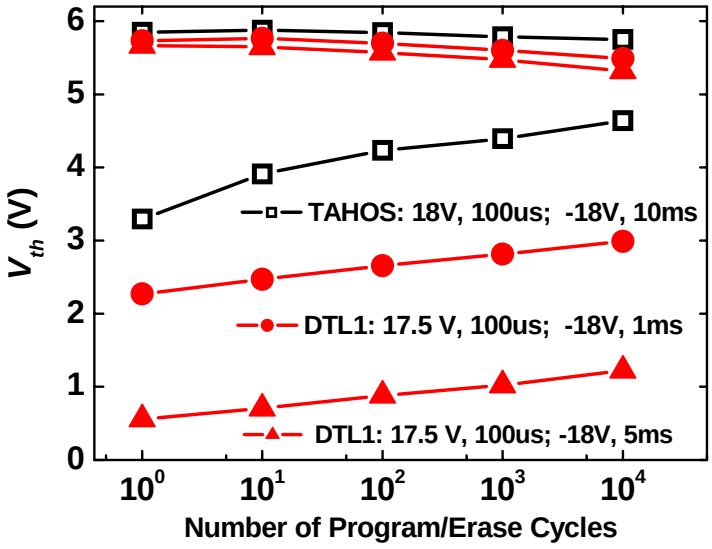


Figure 7.9. Endurance comparison between TAHOS1 and DTL1 memory. Memory window of TAHOS1 deceases faster than that of DTL1 due to the faster degradation of erased state.

2. Performance of devices in group 2: high reliability of DTL over TAHOS and TANOS

Figure 7.10 shows the program and erase characteristics of the TANOS device which uses the conventional silicon nitride as a charge storage layer. The TANOS structure was proposed by Lee *et al* [11]. Another TAHOS device has the same structure as TANOS except the different charge storage material. To compare the trapping properties of Si_3N_4 and HfO_2, the program/erase response of both the TANOS and TAHOS devices at 17 V was measured, as shown in Fig. 7.11. The equivalent oxide thicknesses (EOT) of Si_3N_4 and HfO_2 are nearly the same, but the HfO_2 layer exhibits an obvious advantage in electron trapping. This observation is consistent with Ref. [3]. It is understood that the efficient electron trapping is attributed to the thicker HfO_2 layer which increases the area trap density and also any additional scattering which contributes to higher trapping efficiency. However, the erase performance of the TAHOS device is slower than that of the TANOS device.

Figure 7.12 compares the program/erase characteristics of the TANOS device and the DTL device at 17 V. It is evident that the DTL device offers not only faster program performance but also better erase performance than the TANOS device. The fast program performance is attributed to the high trapping efficiency of HfO_2 and the thinner SiO_2 used in the dual tunneling layer. Besides, although the erase of the DTL and TANOS devices is similar initially, the DTL device performs better than the TANOS device after 10 ms, showing a larger memory window. This can also be attributed to the thinner SiO_2 used in the DTL device which enhances hole injection from the substrate and brings about lower threshold voltage after erasing.

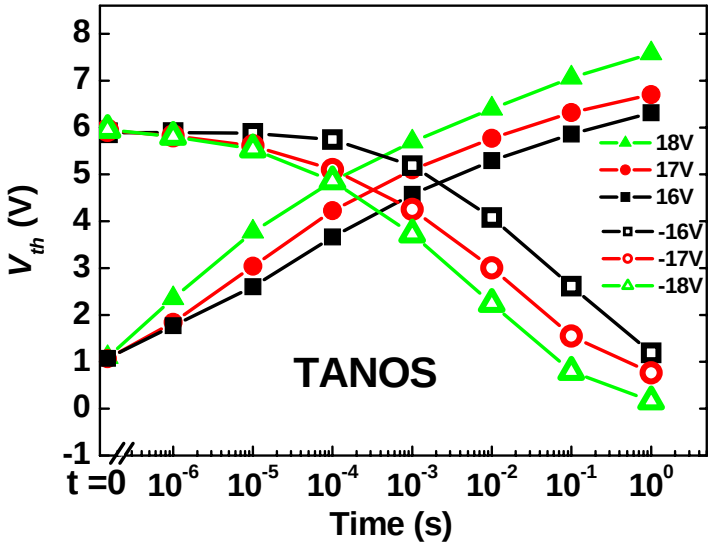


Figure 7.10. Program and erase characteristics of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$) memory.

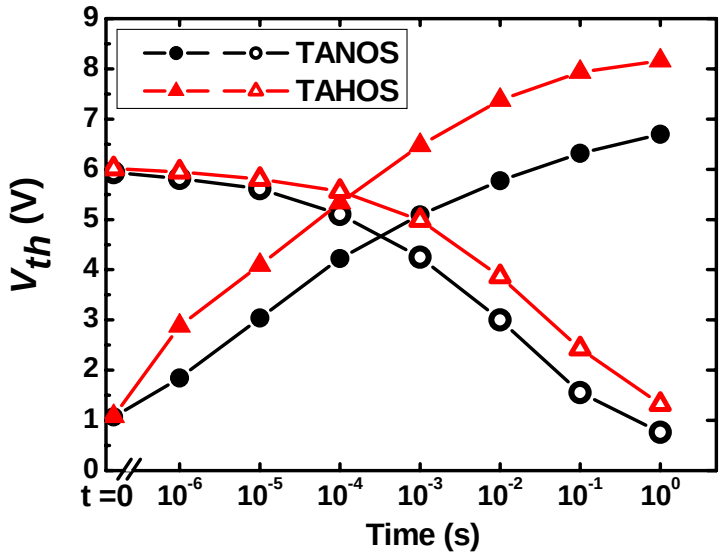


Figure 7.11. Comparison of program and erase characteristics of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$) and TAHOS ($TaN/Al_2O_3/HfO_2/SiO_2/Si$) devices at 17 V.

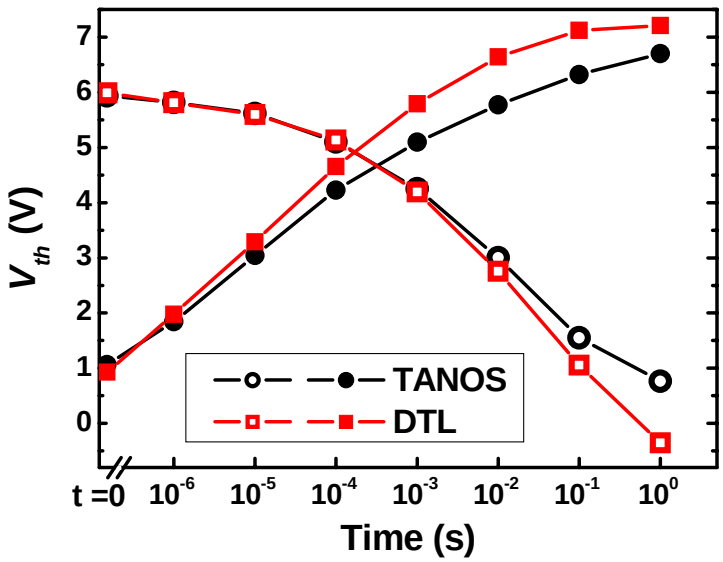


Figure 7.12. Comparison of program and erase characteristics of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$) and DTL ($TaN/Al_2O_3/HfO_2/Si_3N_4/SiO_2/Si$) devices at 17 V.

Figure 7.13 shows the threshold voltage level of the TANOS and DTL devices after programming for 100 μs and erasing for 50 ms at various voltages (>17 V). Difference in V_{th} between the programmed state and the erased state of the DTL device is larger than that of the TANOS device at the same stress voltage, indicating a better charge storage capability of the DTL device. Moreover, the DTL device survives for operation voltages up to 23 V, whereas the TANOS devices break down when the stress voltage exceeds 21 V. The robustness of the DTL device at the high operation voltages is possibly due to its slightly larger EOT as shown in Table 7.2.

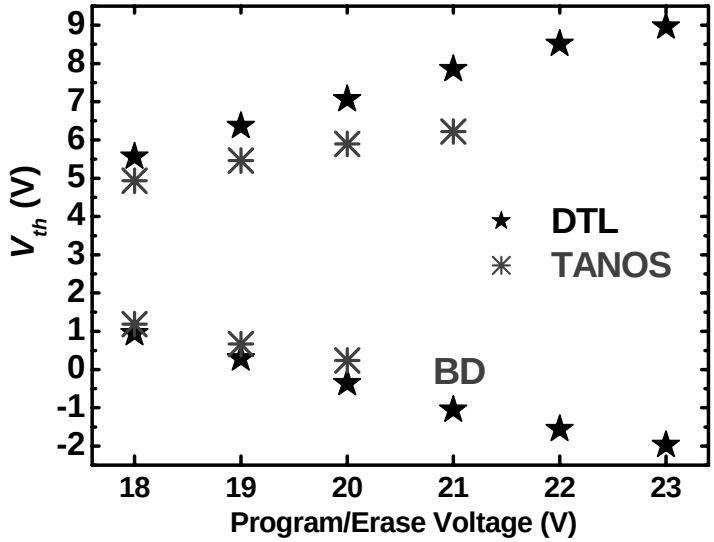


Figure 7.13. Summary of V_{th} level of TANOS and DTL devices after programming for 100 μs and erasing for 50 ms at above 17 V.

Figure 7.14 shows the endurance properties of the three device structures. The TAHOS and DTL devices show the trend of memory window closing, while V_{th} of both the programmed and erased states of the TANOS device shift upward. The observation of the TANOS device is similar to that in a previous report [12] where the upward shift was explained by the accumulation of electrons in deep traps which cannot be easily removed, resulting in an increase in V_{th}. The upward swing of V_{th} of the programmed state with increasing program/erase cycles is undesirable for NAND Flash operation since the over-program is a serious problem as mentioned earlier. On the other hand, the memory window of the TAHOS device closes fast: after 10^4 cycles, only 1.1 V memory window remains. The DTL device exhibits the best endurance property, retaining a 2 V memory window without over-programming after 10^4 cycles.

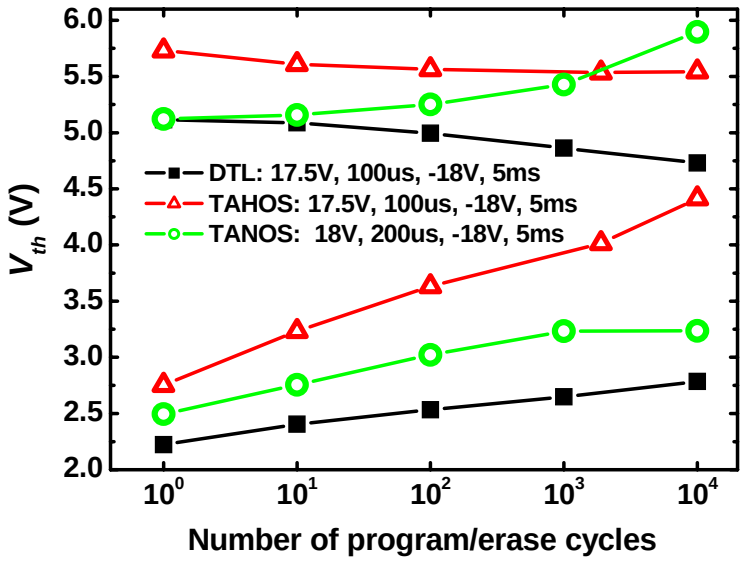


Figure 7.14. Endurance properties of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$), TAHOS ($TaN/Al_2O_3/HfO_2/SiO_2/Si$) and DTL ($TaN/Al_2O_3/HfO_2/Si_3N_4/SiO_2/Si$) memories.

In addition, Fig. 7.15 shows the I_d-V_g characteristics of the DTL, TAHOS and TANOS devices before and after cycling. It can be observed that, the sub-threshold swing (SS) of the DTL device does not degrade even after 10^4 cycles, maintaining low off-leakage current. In contrast, the TAHOS device shows small degradation of SS and the TANOS device shows the largest increase of sub-threshold swing after 10^4 cycles. It was reported that in a floating gate cell, using thicker tunneling SiO_2 enhances the sensitivity to the interface trap generation and hence degrades sub-threshold swing more upon stressing than using the thinner tunneling SiO_2 [13].

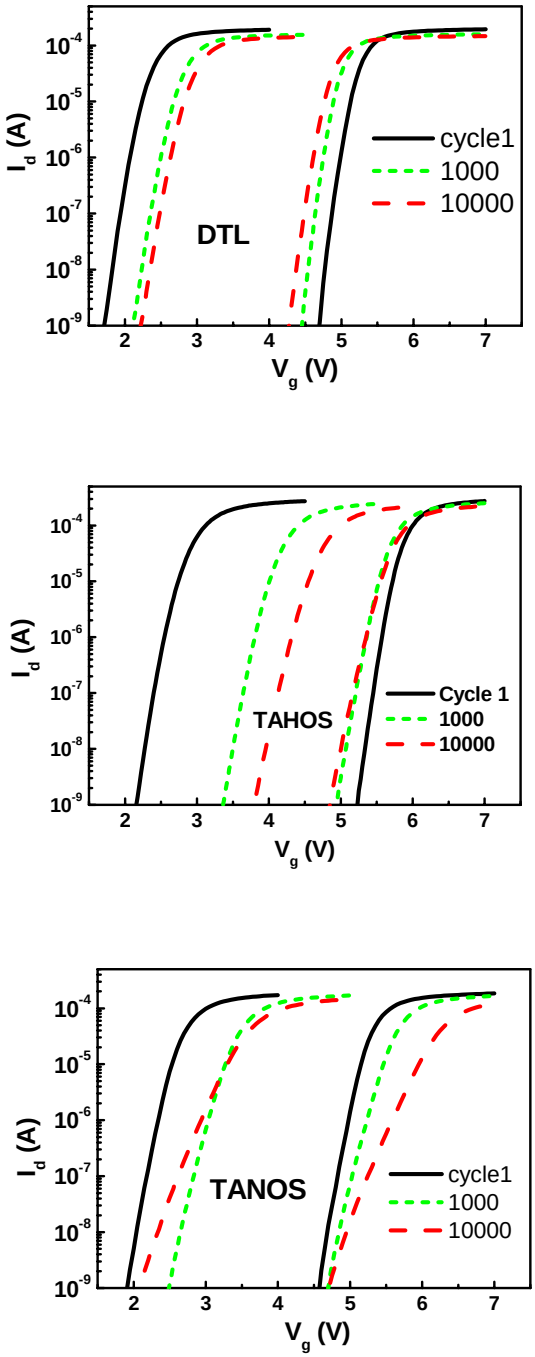


Figure 7.15. I_d-V_g characteristics of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$), TAHOS ($TaN/Al_2O_3/HfO_2/SiO_2/Si$) and DTL ($TaN/Al_2O_3$/ $HfO_2/Si_3N_4/SiO_2/Si$) memories before and after cycling. The best sub-threshold swing after cycling is observed from DTL memory.

The SS degradation should be correlated with the interface trap generation. In the cyclic program/erase characterization, the degradation of the SS is probably caused predominantly by erasing, since the stress time for erase operation is much longer than that for programming. When erasing, some gate electrons tunnel through blocking oxide, traveling into the conduction bands of the trapping layer. Those electrons which are not trapped by the trapping layer will continue to tunnel through tunneling oxide to end up to the Si conduction band, releasing the excess energy there, potentially damaging the Si-SiO_2 interface, and causing the generation of interface states. The electrons in TANOS have higher energy than in TAHOS when they reach the Si-SiO_2 interface due to the higher conduction band of silicon nitride, resulting in larger energy loss and causing more damage. For the DTL device, although the electrons tunnel from silicon nitride conduction band to Si-SiO_2 interface, the nitride conduction band is only 0.5 V higher than HfO_2. However, the DTL device has 1.2 nm thinner SiO_2 than the TAHOS device. Assuming that electric field is 13-15 MV/cm during erase, it is estimated that electrons gain 1.6 - 1.8 V less energy from the electric field of SiO_2 in the DTL device. Therefore, electrons have lower energy when they reach the Si-SiO_2 interface in the DTL device than in the TAHOS device. This may be responsible for the lower extent of degradation of the Si-SiO_2 interface and therefore for the better sub-threshold swing from the DTL device than the TAHOS device.

The comparison of endurance properties of the three devices implies that the shifts of the programmed and erased states are related to the trapping layer; hence TAHOS and DTL devices show the same trend. The presence of the transition layer Si_3N_4 seems to help the improvement of endurance of the DTL device.

Retention properties at 85°C of the three devices are shown in Fig. 7.16. The TANOS device was programmed at 18 V for 200 μs and the TAHOS and DTL devices were programmed at 17.5 V for 100 μs. The TANOS and TAHOS devices were erased at -18 V for 200 ms and the DTL device was erase at -17 V for 100 ms. It is found that the degradation of the three devices is similar each other if the initial memory widow is similar. Although the DTL device has a thinner SiO_2, good retention is still maintained due to the additional Si_3N_4 layer. That is, the DTL device exhibits not only faster speed and lower operation voltage but also excellent retention.

Fig. 7.17 summarizes V_{th} of DTL after undergoing reliability test conditions at various temperatures. The DTL retains 55% memory window after 10^4 cycles and baking for 3 hours at 85 °C.

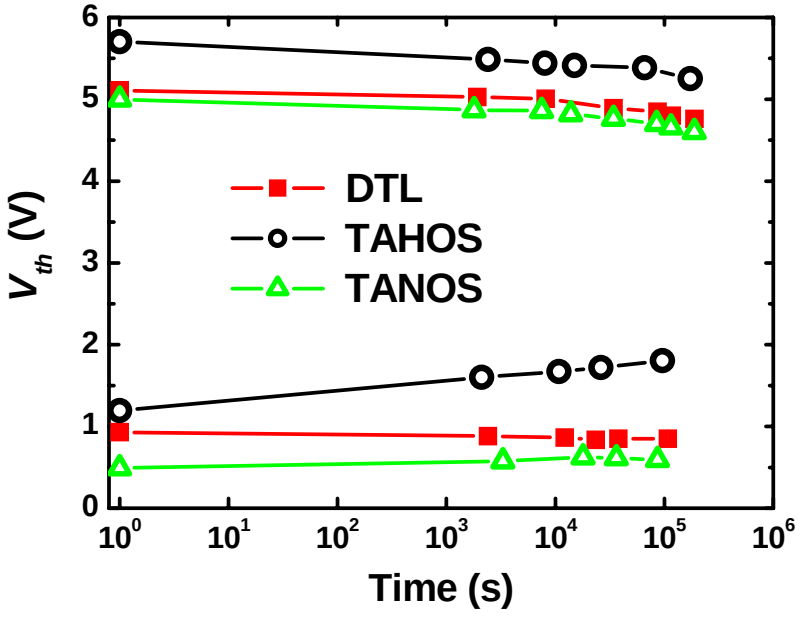


Figure 7.16. Retention comparison of TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$), TAHOS ($TaN/Al_2O_3/HfO_2/SiO_2/Si$) and DTL ($TaN/Al_2O_3/HfO_2/Si_3N_4/SiO_2/Si$) memories.

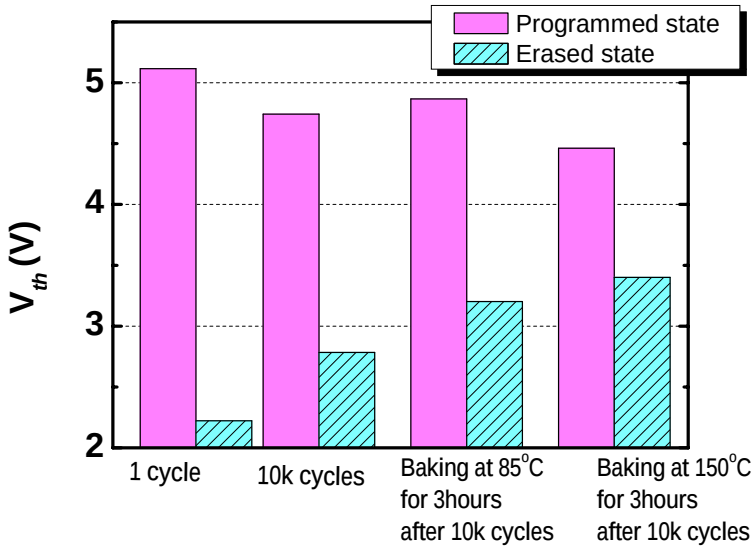


Figure 7.17. Summary of V_{th} level of DTL memory after reliability test. Program was done by 17.5V 100us and erase was done by -18V 5ms.

7.5 Summary

High-*k* HfO_2 provides higher electron trapping capability than Si_3N_4 due to its larger physical thickness for a given EOT. A dual tunneling layer (Si_3N_4/SiO_2) along with a high-*k* HfO_2 charge storage layer was investigated for the optimization of memory device properties for NAND Flash memory application. The dual tunneling layer enhances hole injection from the Si substrate during erase operation because of the low ΔE_v of Si_3N_4, therefore improving the erase speed of the memory device. Also, the dual tunneling layer contributes to excellent retention properties due to its large physical thickness and excellent endurance properties. The DTL concept can also be flexibly applied to other high-*k* charge storage materials with higher *k* value and larger ΔE_v than Si_3N_4.

References

[1] Front end processes, in *International Technology Roadmap for Semiconductors (ITRS)*, pp.51-52, 2005.

[2] Y. N. Tan, W. K. Chim, B. J. Cho, and W. K. Choi, "Over-erase phenomenon in SONOS-type Flash memory and its minimization using a hafnium oxide charge storage layer," *IEEE Trans. Electron Devices*, vol. 51, pp.1143-1147, 2004.

[3] M. She, Ph.D thesis, U.C. Berkeley, 2003.

[4] H. T. Lue, S.-Y. Wang, E.-K. Lai, Y.-H. Shih, S.-C. Lai, L.-W. Yang, K.-C. Chen, J. Ku, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "BE-SONOS: a bandgap engineered SONOS with excellent performance and reliability", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.*, pp.555 – 558, 2005.

[5] K. K. Likharev, "Layered tunnel barriers for nonvolatile memory devices," *Appl. Phy.Lett*., vol. 73, pp.2137-2139, 1998.

[6] S. J. Baik, S. Choi, U.-I. Chung, and J. T. Moon, "High speed and nonvolatile Si nanocrystal memory for scaled Flash technology using highly field-sensitive tunnel barrier", in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.,* pp.22.3.1 - 22.3.4., 2003.

[7] S. H. Hong, J. H. Jang, T. J. Park, D. S. Jeong, M. Kim, C. S. Hwang and J. Y. Won, "Improvement of the current-voltage characteristics of a tunneling dielectric by adopting a $Si_3N_4/SiO_2/Si_3N_4$ multilayer for Flash memory application"*, Appl. Phys. Lett*., vol. 87, pp.152106 1-3, 2005.

[8] H. Bachhofer, H.Reisinger, E. Bertagnolli and H. von Philipsborn, “Transient conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures”, *J. Appl. Phys.,* vol. 89, pp. 2791-2800, 2001.

[9] Y. Yang and M. H. White, “Charge retention of scaled SONOS nonvolatile memory devices at elevated temperatures”, *Solid-State Electronics*, vol. 44, pp. 949-958, 2000.

[10] G. D. Wilk, R. M. Wallace, and J. M. Anthony, “High-k gate dielectrics: current status and materials properties considerations”, *J. Appl. Phys*., vol. 89, pp.5243-5275, 2001.

[11] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, “A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit Flash memories”, in *IEEE int. Electron Device Meeting (IEDM) Tech. Dig.,* pp. 26.5.1-26.5.4, 2003.

[12] K.-H. Wu, H.-C. Chien, C.-C. Chan, T.-S. Chen and C.-H. Kao, “SONOS device with tapered bandgap nitride layer”, *IEEE Trans. Electron Devices*, vol. 52, pp.987-992, 2005.

[13] Y.-B. Park and D. K. Schroder, “Degradation of thin tunnel gate oxide under Constant Fowler–Nordheim current stress for a Flash EEPROM”, *IEEE Trans. Electron Devices*, vol. 45, No. 6, pp.1361-1368, 1998.

Chapter 8

Conclusion

8.1 Conclusions

Scaling of the conventional floating gate memory devices faces unique challenges beyond 45 nm technology node because of the severe capacitance coupling among unrelated floating gate, absence of coupling between the control gate and floating gate along the side wall, short channel effect and the limitation of tunneling oxide scaling in order to assure 10 year retention.

Among the potential candidates to substitute for floating gate memory, discrete charge storage memories including nanocrystal (NC) memory and SONOS type memory are the most promising, thanks to their process compatibility with the conventional semiconductor memory process and their endurance of aggressive scaling.

Extensive researches have been conducted on the NC memory with Si NCs embedded in SiO_2. In our work, we have focused on the formation of Ge NCs and implementation of high-*k* materials as tunneling and blocking oxide. Firstly, self-assembled Ge NCs are formed on HfO_2 and HfAlO by CVD. At the deposition temperature of 600ºC, with 80 sccm gas precursor flow rate and for 30 sec, Ge NCs with density of 10^{11} cm^{-2}, mean diameter of 16 nm and the mean height of 7 nm are achieved on 60 nm thick HfO_2 surface. Ge NCs provide an obvious memory effect. In addition,

we predict that NCs formation is generally favored by a surface of a dielectric material with a higher dielectric constant based on results of our experiment and previous reports, possibly because of the enhanced absorption capability of the high permittivity materials due to their high polarized bonds among atoms.

Ge NCs with diameter about 5-10 nm embedded in HfAlO high-*k* dielectric are obtained by cosputtering method. It is found that Ge NCs are thermally stable in HfAlO matrix. After annealing at source/drain activation temperature 950^{o}C, the NCs remain embedded between the tunneling and blocking HfAlO. A nonvolatile memory device employing Ge NCs embedded in HfAlO dielectric exhibits excellent memory performance. In addition, it is found that the device with Ge NCs in HfO_2 fabricated in the same way as that with Ge NCs in HfAlO shows fast charge leakage in retention mode after annealing at 950^{o}C. This is because the crystallization of HfO_2 matrix during annealing leads to Ge NCs segregated at grain boundaries of HfO_2 and some of them ending up close to the Si substrate.

A simple method to form dielectric HfO_2 NCs is developed by using the self - driven phase separation phenomenon of the hafnium silicate film at high temperature. Dual phase structure comprising crystalline HfO_2 and amorphous $Hf_xSi_{1-x}O_2$ is observed after the $Hf_{0.5}Si_{0.5}O_2$ film is annealed at 900^{o}C and 1000^{o}C. The film containing HfO_2-$Hf_xSi_{1-x}O_2$ dual phase as a trapping layer is found to provide a faster programming speed at a lower programming voltage than Si_3N_4 because of its higher dielectric constant and higher trap efficiency. Meanwhile, the HfO_2-$Hf_xSi_{1-x}O_2$ also provides better retention property than HfO_2 because the presence of the amorphous phase suppressed the formation of grain boundary effectively thereby reducing lateral migration.

Subsequently, the high-k tunneling and blocking oxide HfAlO and high work function gate electrode IrO_2 are integrated with the phase separated HfSiO film. Implementation of high-k HfAlO tunneling and blocking oxide is proved to reduce operation voltage, improve erase saturation as compared with the devices with conventional SiO_2 as tunneling and blocking oxide. The device with IrO_2 gate electrode together with HfAlO tunneling and blocking oxide shows no erase saturation at all and additionally, it is found that the implementation of IrO_2 helps to improve retention properties.

Another method to optimize SONOS type memory structure for NAND Flash application is explored by using the dual tunneling layer (Si_3N_4/SiO_2) along with a high-k HfO_2 charge storage layer. This structure is practically feasible because the materials used in the structure are well developed in the industry. The dual tunneling layer enhances hole injection from the Si substrate during erase operation because of the small ΔE_v of Si_3N_4, leading to increase of erase speed of the memory device. Also, the dual tunneling layer contributes to excellent retention properties because of its large physical thickness. The devices with dual tunneling oxide and high-k HfO_2 trapping layer demonstrates superior endurance properties: without increase of programming V_{th} throughout the cyclic test as compared with devices with conventional Si_3N_4 trapping layer; smaller degradation of the memory window during cycling as compared with memory device with high-k HfO_2 trapping layer but single SiO_2 tunneling oxide. The dual tunneling layer concept can also be flexibly applied to other high-k charge storage materials with higher k value and larger ΔE_v than Si_3N_4.

8.2 Limitation and Future Proposal

The limitation of this thesis work is that all the transistors fabricated in the experiment are long channel device with the gate length ranged from 4 μm to 20 μm. The study does not involve the effect of the channel length on the memory performance which may be particularly important for NC memory. On the other hand, the long gate length also prevents us from studying the channel hot electron injection programming mechanisms. Additionally, our study only focuses on the performance of the single cell. The reliability issues related to arrays of cells such as threshold voltage distribution and the program disturbance have not been dealt with.

Based on the limitations of this study, for future researches and investigation, we proposed the following directions if the fabrication technique are available：

1. The effect of channel length on the memory performance for NCs and SONOS devices need to be studied. Effects of gate length on the memory window of memory device with single layer and double layer Tungsten NCs embedded in HfAlO have been studied by Samanta *et al* [1]. Results show that the memory window decreases when gate length scales for single layer Tungsten NCs device while the memory window increases with the scaling of the gate length for the double layer Tungsten NCs device. Future study is necessary to understand the phenomenon. In addition, it will be essential to study the influence of the gate length on the memory window for SONOS type memory in order to make out the practicability of SONOS as the device scales down.

2. Distribution of the threshold voltage and program disturbance are important reliability concerns when the memory cells are arranged in arrays. Memory structures proposed by this thesis should be further studied in the case of the array fabrication is feasible.
3. Despite the attractive scaling capability of SONOS devices as discussed in previous chapter, generally the endurance of SONOS is not as good as floating gate memory. The commercial floating gate memory can meet program/erase endurance requirement of 10^6 cycles. However, normally SONOS type memory can only sustain 10^4 cycles according to our results and the results reported in the literature. Therefore, the improvement of endurance properties of SONOS devices has to be made. The understanding of the endurance failure mechanism such as the degradation of the trapping layer or the tunneling oxide needs to be investigated furthermore in order to find the solutions to improve the endurance properties.
4. Programming devices using multiple short pulses instead of a single relatively long pulse may be helpful for the improvement of endurance properties of memory devices. The pulsed programming method for floating gate and NCs memory were examined [2, 3] and it is verified that pulse width and heights can be optimized to reduce the trap generation during cycling and thus reduce the degradation after cycling for floating gate devices; meanwhile the programming time can also be reduced. It will be interesting to study in detail the effect of multiple pulses programming on the performance of SONOS type memories. The programming waveform can be varied with respect to the pulse width and pulse height and their impact on memory performance including program and erase time, and subsequential reliabilities such as endurance can be examined.

References

[1] S. K. Samanta, P. K. Singh, W. J. Yoo, G. Samudra, Y.-C. Yeo, L. K. Bera, and N. Balasubramanian, “Enhancement of memory window in short channel non-volatile memory devices using double layer tungsten nanocrystals”, in *IEEE int. Electron Device Meeting (IEDM) Tech.,* pp.177-180, 2005.

[2] F. Irrera, and B. Riccò, “Pulsed tunnel programming of nonvolatile memories”, *IEEE Trans. Electron Devices*, vol. 50, pp.2474-2480, 2003.

[3] G. Puzzilli, D. Caputo, and F. Irrera, “Fast and reliable tunnel programming of nanocrystal nonvolatile memories”, *IEEE Trans. Electron Devices*, vol. 51, pp.1205-1207, 2004.

Publication List

1. Y. Q. Wang, J. H. Chen, W. J. Yoo, Y.-C. Yeo, S. J. Kim, R. Gupta, Z. Y. L. Tan, D. L. Kwong, A. Y. Du, and N. Balasubramanian, "Formation of Ge nanocrystals in HfAlO high-k dielectric and application in memory device", Appl. Phys. Lett. **84**, 5407, (2004).

2. Y. Q. Wang, J. H. Chen, W. J. Yoo, and Y. -C. Yeo, "Chemical vapor deposition of germanium nanocrystals on hafnium oxide for non-volatile memory applications" Materials Research Society Symposium Proceedings, v.830, *Materials and Processes for Nonvolatile Memories*, 2005, p 269-274.

3. Y. Q. Wang, J. H. Chen, W. J. Yoo, Y.-C. Yeo, A. Chin, A. Y. Du, "Formation of dual-phase HfO_2–$Hf_xSi_{1-x}O_2$ dielectric and its application in memory devices" J. Appl. Phys. **98,** 013536 (2005).

4. Y. Q. Wang, P. K. Singh, W. J. Yoo, Y.-C. Yeo, G. Samudra, A. Chin, W. S. Hwang, J. H. Chen, S. J. Wang, D. -L. Kwong, "Long retention and low voltage operation using IrO_2/HfAlO/HfSiO/HfAlO gate stack for memory application", *IEEE International Devices Meeting (IEDM)*, Washington, D.C., 5–7 December 2005 p.169-172.

5. Y. Q. Wang, G. Samudra, D. Y. Gao, C. Shen, W. S. Hwang, G. Zhang, Y.-C. Yeo, and W. J. Yoo, "Fast erasing and highly reliable MONOS type memory with HfO_2 high-k trapping layer and Si_3N_4/SiO_2 tunneling stack", *IEEE International Devices Meeting (IEDM*), San Francisco, 11-13 December 2006 p.971-974.

6. Y. Q. Wang, W. S. Hwang, D. Y. Gao, G. Samudra Y.-C. Yeo, and W. J. Yoo, "Electrical Characteristics of Memory Devices with High-k HfO_2 Trapping Layer and Dual Tunneling Layer Si_3N_4/SiO_2", accepted by IEEE Trans. Electron Devices