APPLICATION OF NOVEL GATE MATERIALS FOR PERFORMANCE IMPROVEMENT IN FLASH MEMORY DEVICES

PU JING

NATIONAL UNIVERSITY OF SINGAPORE

2009

APPLICATION OF NOVEL GATE MATERIALS FOR PERFORMANCE IMPROVEMENT IN FLASH MEMORY DEVICES

PU JING

(B. Eng., National University of Singapore)

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2009

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my thesis advisors, namely, Prof. Chan Siu Hung, Daniel and Prof. Cho Byung Jin, for their invaluable guidance, wisdom, and kindness in teaching and encouraging me. I will definitely benefit from the experience and knowledge I have gained from them throughout my life. Thank you for their patience and painstaking efforts devoting in my research as well as the kindness and understanding which accompanied me over the last four years. Hence, my best wishes will go to Prof. Chan and Prof. Cho as I am deeply grateful for their help. I am especially grateful of Prof. Cho's help, who provides me with the opportunity to join his research group in the first place.

In addition, I have had the pleasure of collaborating numerous exceptionally talented graduate students and colleagues over the past four years. Firstly, I would like to thank my colleagues in Prof. Cho's group, including Dr. Shen Chen, Dr. Hwang Wan Sik, Mr. He Wei, and Ms. Zhang Lu, for their useful discussions and kind assistances. Many thanks also go to Mr. Sun Zhi Qiang, Mr, Cheng Jingde, Dr. Tan Kian Ming, Mr. Yang Wei Feng, Mr. Yang Jian Jun, and Mr. Zang Hui for their knowledge and experience which had benefited me, as well as the long lasting friendship. I would also like to extend my appreciation to all other SNDL teaching staffs, technical staffs and graduate students for the good academic environment created.

Last but not least, my deepest love and gratitude will go to my family, my mother, my father, and my husband, for their love, patience and support throughout my postgraduate studies.

SUMMARY

The overall objective of this work is to apply novel gate materials for the performance enhancement of flash memory devices, including both floating gate-type flash memory devices and SONOS-type flash memory devices. These attempts could be of practical value for flash memory devices, especially in improving the operation speed and data retention.

A novel floating gate engineering scheme using carbon doped polysilicon floating gate is proposed to overcome the scaling barrier for floating gate-type flash memory devices. It has been found that incorporating carbon into conventional n^+ polysilicon floating gate will be able to significantly improve the program/erase speed, especially for devices with small coupling ratio (~0.3), which is the bottleneck for sub 30 nm flash memory technology. The data retention of such devices is also improved. All these improved properties originate from the increased conduction band offset of the floating gate caused by the incorporation of carbon. The formation of silicon carbide nano-structure is responsible for the band structure change. Adoption of the carbon doped polysilicon floating gate will result in little process modification to the current technology, and is an effective and simple solution for floating gate-type flash memory scaling.

In the advanced SONOS-type flash memory devices, the application of high dielectric constant materials as the blocking oxide attracts much research interest. The feasibility of a novel rare earth high-κ material, Gd_2O_3, as the potential candidate for the blocking layer application in SONOS-type flash memory devices is evaluated. The material properties of Gd_2O_3, including deposition method, leakage current performance,

crystal information as well as the band structure have been studied systematically. Control of the crystal structure of Gd_2O_3 has been found to be the key point for a high quality dielectric film. SONOS transistors with Gd_2O_3 blocking layer exhibits superior performance over those with Al_2O_3 blocking layer in several aspects such as program/erase speed, room temperature retention, etc.

Experimental results have demonstrated that Gd_2O_3 is a favorable blocking oxide candidate except that the retention after cycling remains problematic. Doping of Al into pure Gd_2O_3 is proposed for the robust data retention after cycling, since the increase in the conduction band offset is always an effective method to block the electron *leakage*, both for room temperature and high temperature retention. The optimized Al concentration needs to be carefully considered to balance all the following factors: dielectric constant, conduction band offset, film morphology as well as memory characteristic. All those questions will be well addressed in Chapter 4. The use of $GdAlO_x$ doped with 35% Al results in superior memory performance over those using Al_2O_3 blocking layers, and this material could be a promising candidate for the future blocking oxide material.

In Chapter 5, structure optimization of SONOS cell with 35% Al incorporated $GdAlO_x$ blocking oxide is discussed. The study focuses on the relationship between the blocking layer thickness and long term retention reliability at room temperature, after program/erase cycles and at elevated temperature. A novel leakage current separation technique will be applied to differentiate the leakage components in SONOS memory in order to improve the retention effectively. Charge leakage mechanisms for SONOS-type flash memory devices will be discussed in this chapter as well.

TABLE OF CONTENTS

CHAPTER 2. CARBON DOPED POLYSILICON FLOATING GATE FLASH MEMORY DEVICES

CHAPTER 3. A FEASIBILITY STUDY OF Gd_2O_3 AS BLOCKING OXIDE IN SONOS-TYPE FLASH MEMORY DEVICES

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)
I	Current (A)
I_d	Drain current (A)
I_g	Gate leakage current (A)
J	Current density (A/cm^2)
L	Channel length (μm)
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
E_g	Band gap

LIST OF ACRONYMS

ALD	Atomic layer Deposition
CMOS	Complimentary Metal Oxide Semiconductor
CHE	Channel Hot Electron
CVD	Chemical Vapor Deposition
C-V, CV	Capacitance versus Voltage
DRAM	Dynamic Random Access Memory
DT	Direct Tunneling
EEPROM	Electrically Erasable and Programmable Read Only Memories
EOT	Equivalent Oxide Thickness
EPROM	Electrically Programmable Read Only Memories
F-N	Fowler-Nordheim
FG	Floating Gate
FGA	Forming Gas Anneal
FET	Field Effect Transistor
IPD	Interpoly Dielectric
ITRS	International Technology Roadmap for Semiconductors
LPCVD	Low Pressure Chemical Vapor Deposition
I-V, IV	Current versus Voltage
MONOS	Metal / Oxide / Nitride / Oxide / Silicon
MOS	Metal Oxide Semiconductor
TANOS	TaN / Aluminum Oxide / Nitride / Oxide / Silicon
PDA	Post Deposition Anneal
PVD	Physical Vapor Deposition
ROM	Read Only Memories
RTA	Rapid Thermal Anneal
SONOS	Silicon / Oxide / Nitride / Oxide / Silicon
SRAM	Static Random Access Memory
S/D	Source and Drain

XPS	X-ray Photoelectron Spectroscopy
XRD	X-ray Diffraction

CHAPTER 1

INTRODUCTION

1. 1 Semiconductor Memory Comparison

Complementary metal-oxide-semiconductor (CMOS) memories, which generate significant profit on investment, are widely used in personal computers, cellular phones, digital cameras, smart-media, networks, automotive systems as well as global positioning systems. The increasing need to access data everywhere at any time is requiring that data processing have higher speed and larger capacity than before. These changes together with the expansion of applications, such as networking devices and mobile products have a direct impact on the semiconductor memory market. In the past few years, semiconductor memory occupied above 20% of the total semiconductor market and this percentage will increase continuously and aim towards 30% in the near future, as shown in Fig. 1.1.1 [1]. It is estimated that on average, each person will use 480 MB memory by the year 2010 [2].

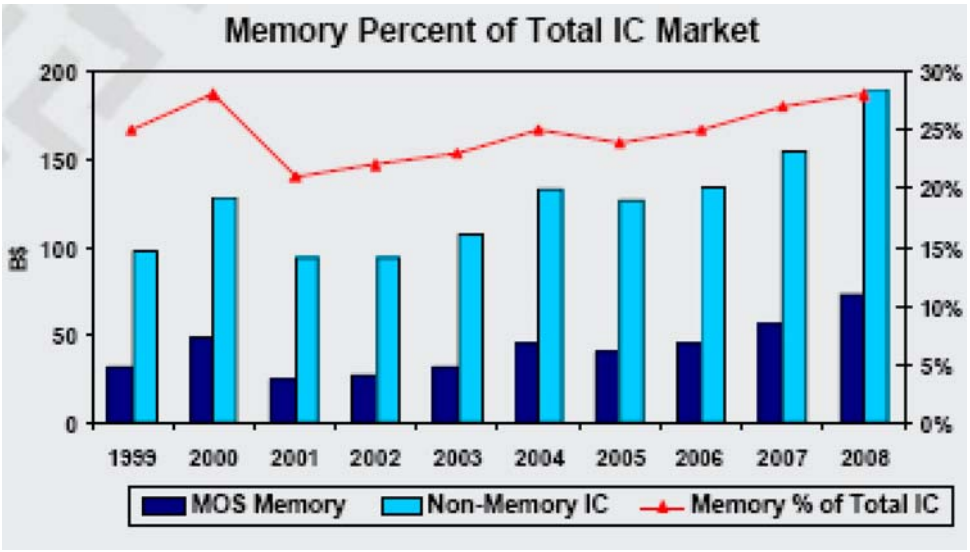


Fig. 1.1.1: Revenues of semiconductor market versus year. The top line is the memory percentage of the total market. The semiconductor memory occupies more than 20% of the total semiconductor market [1].

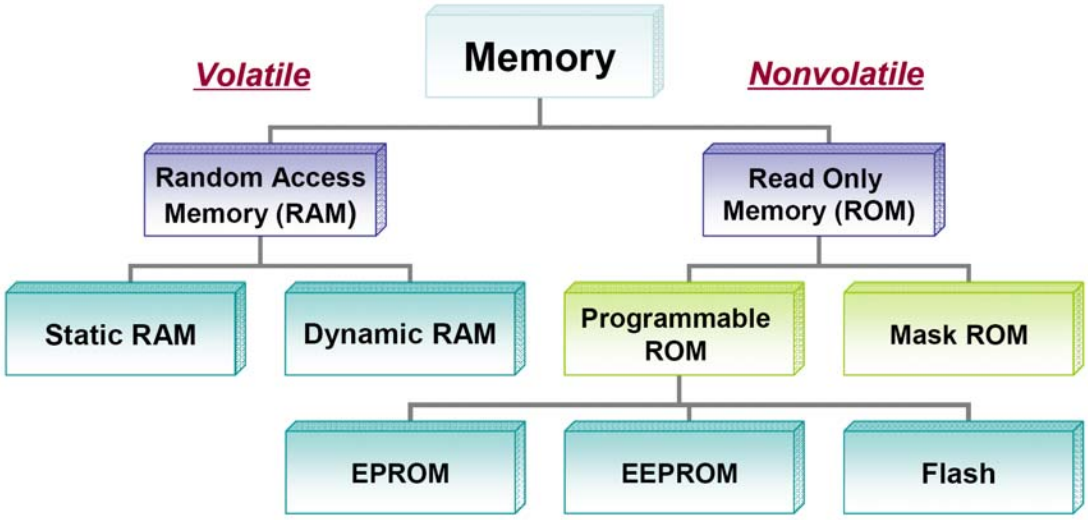


Fig. 1.1.2: Organization of semiconductor memory devices [3].

CMOS memories can be divided into two categories as shown in Fig.1.1.2. One is Random Access Memories (RAMs), which are volatile, i.e., they lose the stored information once the power supply is switched off. The other type is called Read Only Memories (ROMs), which are nonvolatile, i.e., the stored information still remains even after the power supply is switched off. .

The RAM contains two major groups: Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM). One of the most important features for both DRAM and SRAM is low-voltage and low-power application. The DRAM uses a MOS capacitor and a transistor as a memory unit (cell). DRAM has much faster program/read speed with very low operating voltage, while flash memory needs 1 us to 1 ms program time and high program voltage [4]. Unfortunately, DRAM is a volatile memory. The data retention time is about 100 ms in DRAM while it is 10 years in flash memory. The only way that DRAM memory cell does not lose the information is by periodically reading the data and rewriting the same data before the information is completely lost. This operation is called “Refresh” and is an important feature of the DRAM. It poses a tremendous challenge to the goal of lowering the power consumption. Furthermore, the size of a DRAM cell is larger than that of a flash memory cell. Scaling down the DRAM cell size is difficult due to the large capacitor required to store data.

An SRAM cell is a bi-stable transistor flip-flop, or two inverters connected back to back. It is used as a cache memory in personal computers since it offers the fastest write/read (8 ns) speed among all memories. However, a single SRAM cell consists of 6 transistors (6T), so SRAM chip density is very low, although 4T SRAM cell have been demonstrated [5].

The nonvolatile memory, as shown in Fig. 1.1.2, includes Mask Read Only Memory and Programmable Read Only Memory, like Electrically Programmable Read Only Memory (EPROM), Electrically Erasable and Programmable Read Only Memory (EEPROM), or Flash Memory. Those memories are able to keep the data content without a power supply.

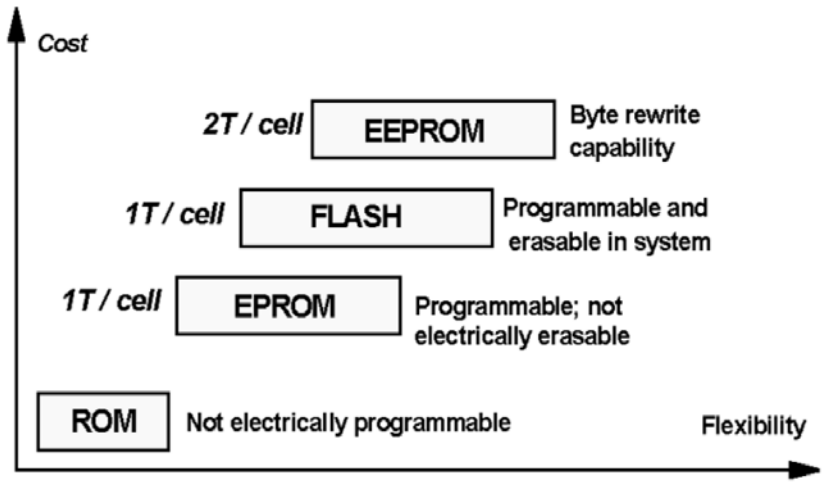


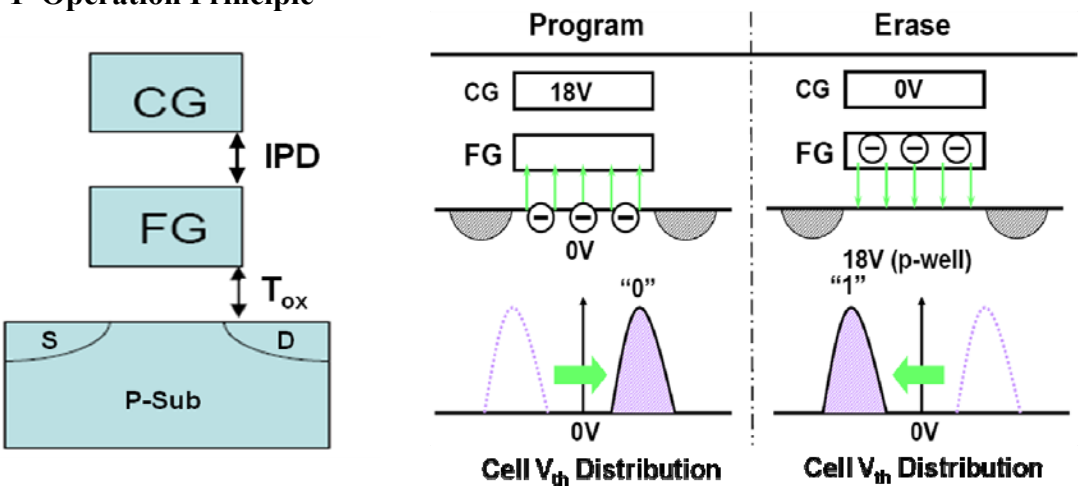
Fig. 1.1.3: The Programmable ROMs qualitative comparison in the flexibility–cost plane. A common feature of Programmable ROMs is to retain the data even without power supply [6].

The Mask ROM is programmed when it is manufactured at the factory with a special mask [3]. The Programmable ROM families can be qualitatively compared in terms of flexibility and cost as shown in Fig. 1.1.3. The EPROM is a reprogrammable memory device can be erased by an ultraviolet light source and programmed by a voltage pulse. It offers a relative low cost; however a low flexibility as well since it is not electrically erasable. On the other hand, although the EEPROM, for which each single

byte is electrically erasable and programmable, has great flexibility, it is also more expensive, since it occupies larger areas. The flash memory device, which is quite similar to EEPROM except that flash memory is erased in blocks instead of individual erase, offers best compromise between cost and complexity. The flash memory cell has the smallest cell size (one transistor per cell) with a very good flexibility (it can be electrically written more than 100, 000 times, with individual byte programming and block erasing) [7]. All those unique features lead to an explosive growth of the flash memory market in recent years, either for high density data storage purpose or with fast random access for code execution.

1. 2 Floating Gate-Type Nonvolatile Memory Devices

1. 2. 1 Operation Principle



(a) **(b)**

Fig. 1.2.1: (a) Schematic cross section of a floating gate–type flash memory transistor. A flash memory transistor is a MOSFET transistor consisting of a tunnel oxide (SiO_2), Floating Gate (n+ polysilicon), Inter-poly Dielectric ($SiO_2/Si_3N_4/SiO_2$), and a Control Gate (n+ polysilicon). (b) Schematic diagram illustrating the program and erase operation of a flash memory cell.

Figure 1.2.1 (a) shows a schematic cross section diagram of a floating gate-type flash memory transistor. The floating gate-type flash memory cell is as simple as a MOSFET transistor except that there is a Floating Gate (FG) inserted between two dielectric layers. It consists of a tunnel oxide (SiO_2), Floating Gate (n+ polysilicon), Inter-poly Dielectric ($SiO_2/Si_3N_4/SiO_2$), and a Control Gate (n+ polysilicon). The potential on the upper gate, Control Gate (CG), controls the current flow between the source and drain of the transistor. A sufficient high gate voltage will enable the electrons to overcome the barrier between the silicon substrate and tunnel oxide and tunnel into the FG. The FG is sandwiched between two dielectric layers, the tunnel oxide and the inter-poly dielectric (IPD), and acts as a potential well for the electrons to stay. Charge on the FG will remain on the gate for a very long time in the absence of external stimuli to remove it.

Figure 1.2.1 (b) illustrates the program/erase operation of a floating gate-type flash memory cell. There are two major mechanisms to charge the floating gate: injection of highly energetic carrier over the barrier (Hot Electron Injection) or tunneling of carriers through a barrier that has been modified by the electric field (Fowler-Nordheim Tunneling). Since the memory devices discussed in this thesis are NAND Flash memories which explore Fowler-Nordheim (F-N) Tunneling program, the program mechanism via F-N tunneling will be studied in detail.

The F-N tunneling is a quantum-mechanical process in which an electron passes from the conduction of one silicon region to that of another silicon region through a triangular energy barrier induced by the high electric field. 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. As shown in Fig. 1.2.1 (b), during the program state, a

large control gate bias (~ 18 – 20 V) is applied, the potential barrier of silicon dioxide is distorted and the barrier to a substrate electron tunneling into the FG conduction band in the oxide is reduced. A large amount of electrons hence are programmed into the FG. The threshold voltage of the MOSFET will change accordingly due to the presence of the excess electrons confined in the FG. The shift in the threshold voltage could be expressed as:

$$\Delta V_T = V_T - V_{T0} = -\Delta Q / C_{FC} \tag{1.1}$$

Where

ΔQ is the amount of charges injected

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

V_T is the threshold voltage when a net charge ΔQ is stored in the FG

C_{FC} is the capacitance between floating gate and control gate [8]

The state of higher threshold voltage after program is referring to the "0"state as shown in Fig. 1.2.1 (b). The erase case could be analyzed in a similar manner. Apply either a sufficient large negative voltage at the gate or a positive gate voltage at the substrate to cause the FG electrons to inject back to the substrate. The threshold voltage of the MOSFET would be reduced to some lower value, which is referring to the "1" state. The read operation of a flash memory cell is achieved by applying a sense voltage which is in-between V_{T0} and V_T. The memory state can be determined by the measured current level, as shown in Fig. 1.2.2 [1]. As the memory technology moves from single bit operation to multi-bit operation, e.g. 2 bit per cell, a large memory window is required to have split V_{th} values to differentiate various states.

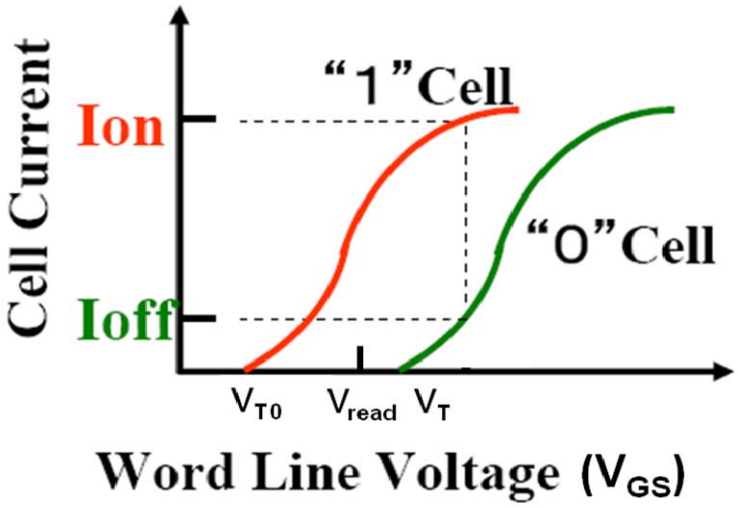
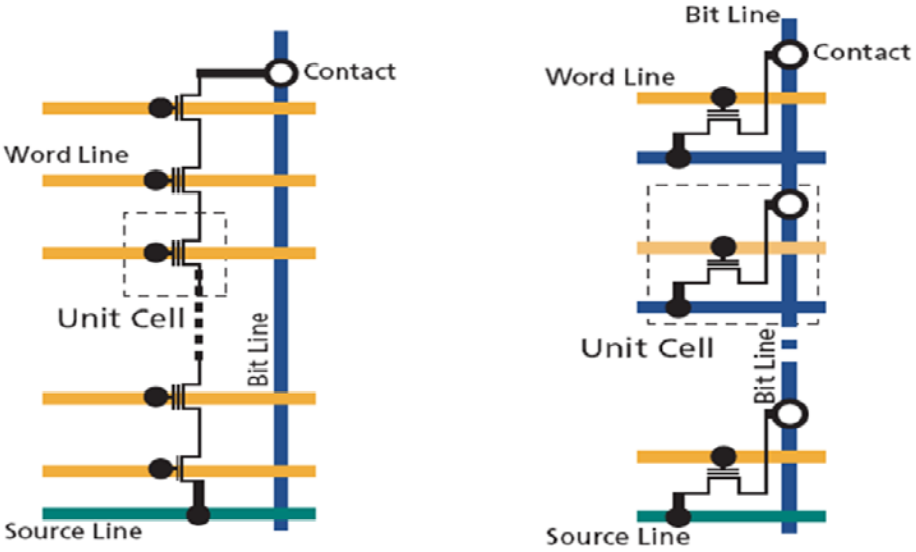


Fig. 1.2.2: I–V curves of a FG device when there is no charge stored in the FG (curve on the left) and when a negative charge Q is stored in the FG (curve on the right). The read operation of a memory cell involves applying a reading voltage in between V_{T0} and V_T.



(a) **(b)**

Fig. 1.2.3: Flash memory architecture of (a) NAND and (b) NOR flash.

When thousands of flash memory cells are connected together, they will form array structure. The flash memory architecture can be clarified into two major groups:

NAND Flash and NOR Flash as shown in Fig. 1.2.3. NAND Flash uses F-N tunneling current for both program and erase, while NOR Flash uses Channel Hot Electron Injection for program and F-N for erase.

As shown in Fig. 1.2.3 (a), the NAND architecture sandwiches a number of memory transistors (16 or 32) in series. These transistors all share a common source and a common drain; hence the effective area for drain and source per memory transistor is significantly reduced. As a result, NAND has the smallest cell size among current semiconductor memories. NAND flash features high cell density, high capacity, fast program and erase rates, and easiness for scaling down, which is mainly used for mass data storage, whereas the NOR flash is for code storage. Since the NOR architecture is having individual connection to each source and drain as shown in Fig. 1.2.3 (b), which gives random access and byte write capability.

1. 2. 2 Floating Gate-Type Flash Memory Scaling

Flash memory application has seen explosive growth in recent years and this trend is likely to continue because new demanding applications are constantly arising. Despite the rapid growth, both NOR and NAND flash memories face steep technology challenges when further scaling into sub-32 nm node. The scaling limitation for NAND flash memories lies in several aspects, namely:

(1) Thickness scaling limitation for tunnel oxide and IPD layer.

(2) Insufficient coupling ratio to maintain the program/erase window.

(3) Increased cell interference through capacitive coupling.

(4) Reduction of the number of stored electrons in the FG.

Those four major challenges limit the continuously scaling of NAND flash memories stated by the Moore's Law. Each of those aspects will be evaluated in detail in the following section.

1. Thickness scaling limitation for tunnel oxide and IPD layer.

The thickness of the tunnel oxide and IPD layer dominate program/erase speed and the amount of read current for a nonvolatile memory cell transistor. A thinner oxide enables faster operation speed, and is also beneficial for multi-level operation. However, the reliability requirement sets the fundamental scaling limit for the tunnel oxide and IPD layer. Many research reports show that tunnel oxide with thickness $\geq$ 8 nm can be relatively leak free because a single defect in the oxide is insufficient to provide a leakage path [9, 10]. However, a tunnel oxide that is less than 8 nm thick is prone to the defect-assisted leakage, which is evidenced by the severely increased SILC leakage current [11, 12].

The limitation factors for the ONO IPD layer are the charge retention capability, especially at elevated temperature, the threshold voltage stability, as well as the charge trapping behavior [13, 14]. Thus, without introducing a high-κ IPD layer, it is only applicable to scale down the effective ONO layer thickness to be around 13 nm for acceptable long term reliability behavior [15-17].

2. Insufficient coupling ratio to maintain the program/erase window.

In flash memory devices, the true gate (control gate, or Word Line) controls the channel indirectly. The degree, which the control gate controls the channel is indicated by the gate coupling ratio, defined as

$$\alpha_g = C_{FC} / C_{FS} \tag{1.2}$$

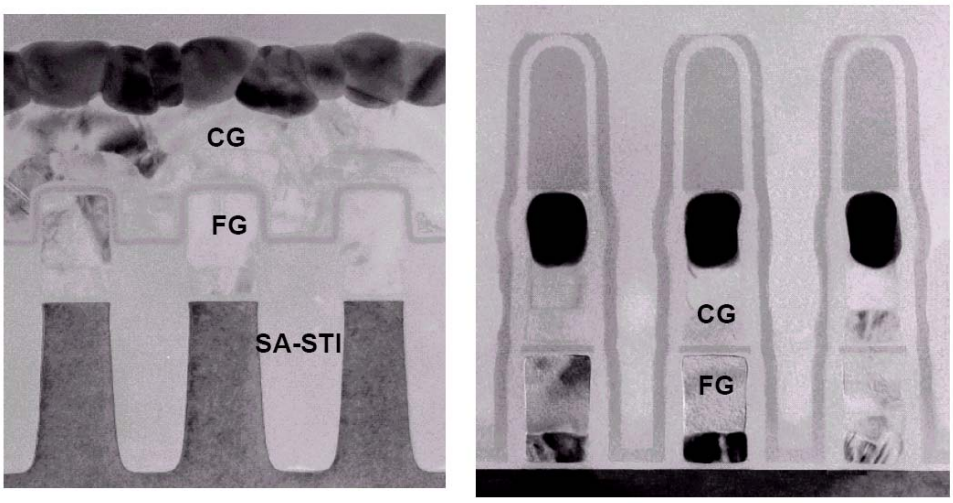
Where

C_{FC} is control gate to floating gate capacitance

C_{FS} is total floating gate to source, drain and substrate capacitance.

α_g represents the ratio of the voltage drop across the tunnel oxide out of the total applied control gate voltage. α_g should be large enough so that the voltage drop across the tunneling oxide is large enough to achieve high charge exchange speed between the floating gate and channel. Meanwhile, α_g can't be too large, considering the impact of charges in floating gate is inversely proportional to C_{FC}, i.e., $\Delta V_T = V_T - V_{T0} = -\Delta Q / C_{FC}$. The range of α_g is kept to be 0.6-0.7 for the optimum program/erase performance.

For the planar device, it is impossible to achieve a coupling ratio which is greater than 0.5 [10]. Therefore, in the word line direction, the control gate wraps around the floating gate to geometrically increase C_{FC}. From 256 Mbit NAND Flash memory onwards, the industry has adopted a Self-Aligned Shallow Trench Isolation (SA-STI) technology [18], as shown in Fig. 1.2.4.



(a) Word-line direction **(b) Bit-line direction**

Fig. 1.2.4: Schematic diagram showing the SA-STI cell in the (a) Word line direction and (b) Bit-line direction [19].

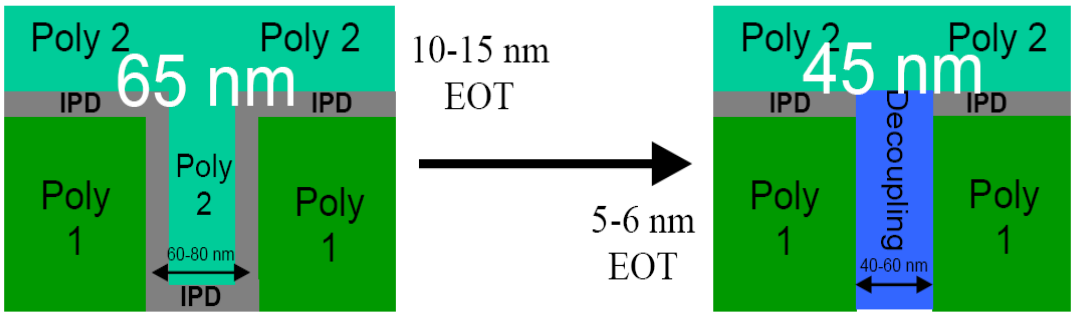


Fig. 1.2.5: Comparison of coupling ratio between the control gate (poly 2) and floating gate (poly 1) between 65 nm and 45nm technology node. As the space between the adjacent cells decreases from 60 - 80 nm to 40 - 60 nm, the gate coupling ratio is decreased accordingly. This is due to decreased overlap area between the floating gate and control gate [20].

The SA-STI technology results in extremely small cell size with high packing density and good cell reliability. In this architecture, part of α_g comes from the overlap area of floating gate and control gate along the sidewalls, as shown in Fig. 1.2.4 [5].

However, this part will become difficult when two adjacent floating gates are too close beyond 45 - 40 nm technology generation as shown in Fig. 1.2.5. This leads to a drastic drop of coupling ratio to about 0.3-0.4 [16, 20, 21], which is a serious drawback for the multi-level operation. Without new innovative structures or introducing new materials, there is little chance to further scale down the NAND flash memory cell.

3. Increased cell interference through capacitive coupling.

In the current flash memory technology, the FG is completely self-aligned to the active area as shown in Fig 1.2.4. The coupling ratio between CG and FG is obtained by the inter-poly ONO layer, covering the top and side wall of the FG. On the other hand, uncovered sidewall of the FG is weekly coupled between adjacent FG-FG, as illustrated in Fig. 1.2.6. As the cell integration density is increased, NAND flash memory cell suffers from increased parasitic capacitance between the cells and it generates serious problems in the multi-level operation. When the adjacent cells are programmed, the V_{th} of the unprogrammed cell will be modulated through capacitive coupling. As the design rule becomes small, the V_{th} modulation grows influentially. To overcome the FG-FG coupling noise, it is important to reduce the FG thickness. Fig. 1.2.7 shows the dependence of the V_{th} modulation on the FG thickness [22]. The thinner the FG, the smaller the V_{th} modulation as the parasitic capacitance decreases.

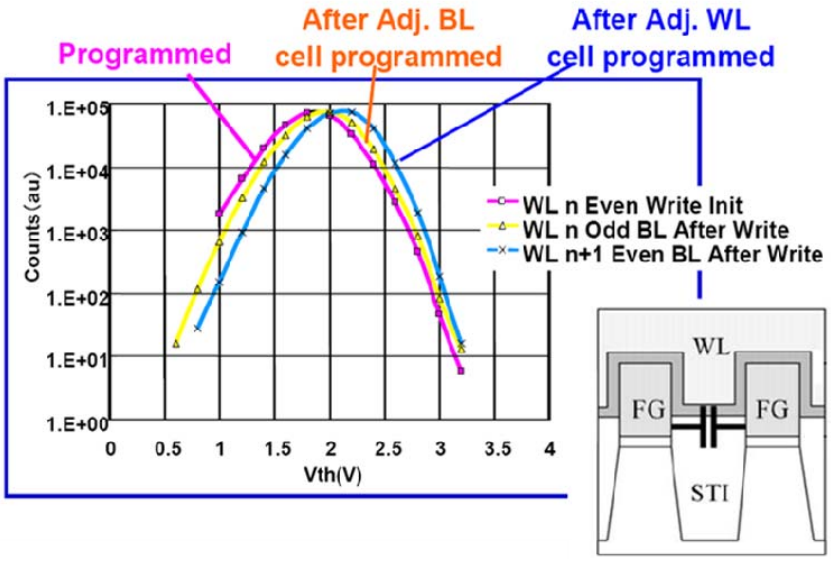


Fig. 1.2.6: V_{th} modulation due to the FG-FG coupling interference. The program state threshold voltage shifts due to the program of the adjacent cell. The insert in the plot shows two adjacent flash memory cells in the world line direction. The parasitic capacitor is clearly shown [23].

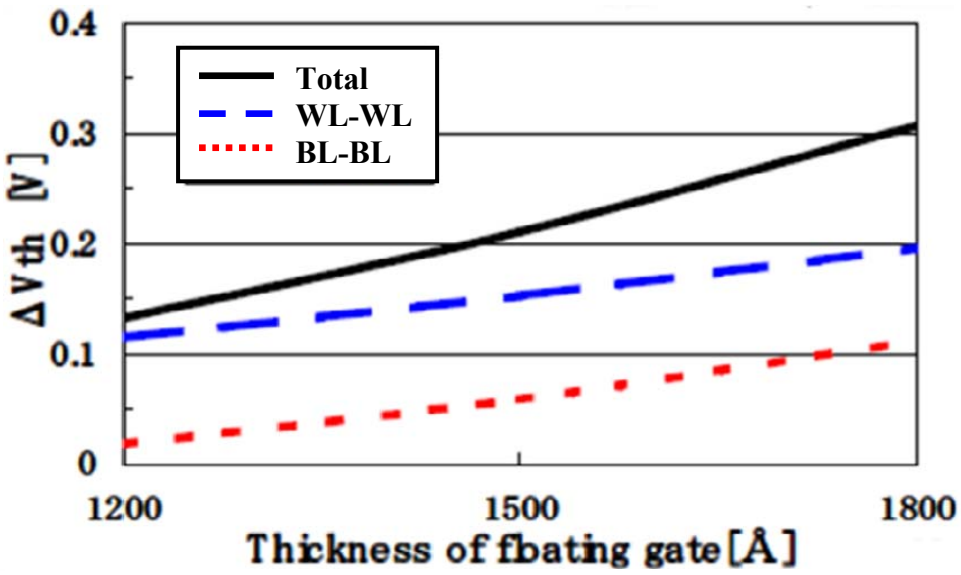


Fig. 1.2.7: The dependence of threshold voltage modulation phenomenon on the thickness of the FG. The ΔV_{th} decreases with the FG height [19].

4. Reduction of the number of stored electrons in the FG

Figure 1.2.8 shows the stored number of electrons in the FG cell, which is estimated as a function of the technology node. As illustrated in this figure, as the dimensions of flash memories are scaled down, the number of electrons representing one bit dramatically reduces for each NAND flash generation [23, 24]. Note that the use of multi-bit cell memory technology will result in an even more reduced number of electrons per bit. For example, FG multi-level operation cell of beyond 35nm node has only less than 50 electrons. In such cases, even the loss of a single electron will cause the V_{th} instability, which puts more stringent requirement on the charge retention. Moreover, in such few electron device, control of the electron location also becomes important. The imbalanced location of the electron will result in V_{th} instability.

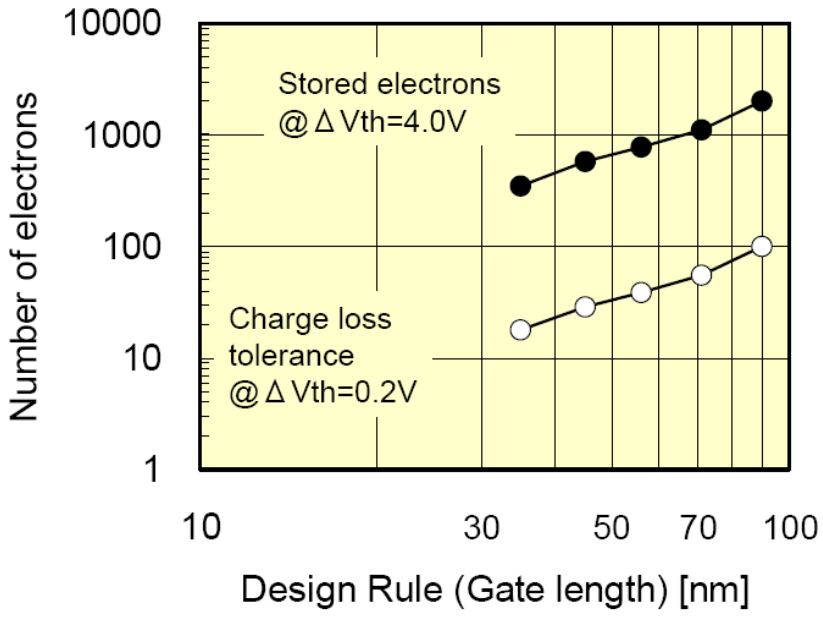


Figure 1.2.8: The stored number of electrons and the charge loss tolerance decreases with the scaling down of flash memory cell [25].

To minimize the parasitic capacitance of the adjacent flash memory cells, it is desirable for the thickness of the FG to be thinned down as shown in Fig. 1.2.7. On the other hand, this will inevitably worsen the few electron effect, and a compromise between these two factors need to be taken into consideration.

According to the aforementioned discussion, the aggressive scaling of floating gate-type flash memory devices almost meets its fundamental limit. Table 1.2.1 shows the ITRS Roadmap 2007 report for the nonvolatile memory technology requirement [26]. As indicated in the ITRS Roadmap, there is no solution yet to utilize the memory performance beyond 30nm technology node. Extensive researches are required on exploration of new materials, new technology and new memory structure to overcome all those scaling barriers.

Table 1.2.1: Nonvolatile memory technology requirement (ITRS 2007) [26].

Table PIDS5a Non-volatile Memory Technology Requirements—Near-term Years

Year of Production	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
DRAM ½ Pitch (nm) (contacted)	*68*	*58*	*50*	*45*	*40*	*36*	*32*	*30*	*25*
NAND Flash poly ½ Pitch (nm)	*51*	*45*	*40*	*36*	*32*	*28*	*25*	*22*	*20*
NAND Flash									
NAND Flash technology – F (nm) [1]	51	45	40	36	32	28	25	22	20
Number of word lines in one NAND string [2]	32	32	64	64	64	64	64	64	64
Cell type (FG, CT, 3D, etc.) [3]	FG	FG	FG	FG/CT	CT	CT	CT-3D	CT-3D	CT-3D
3D NAND number of memory layers	1	1	1	1	1	1	2	2	2
A. Floating Gate NAND Flash									
Cell size – area factor a in multiples of F^2 SLC/MLC [4]	4.0/2.0	4.0/2.0	4.0/1.3	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0	4.0/1.0
Tunnel oxide thickness (nm) [5]	6-7	6-7	6-7	6-7	6-7	6-7	6-7	6-7	6-7
Interpoly dielectric material [6]	ONO	ONO	ONO	ONO	ONO	High-κ	High-κ	High-κ	High-κ
Interpoly dielectric thickness (nm)	10-13	10-13	10-13	10-13	10-13	9-10	9-10	9-10	9-10
Gate coupling ratio (GCR) [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	0.6–0.7	0.6–0.7
Control gate material [8]	n-Poly	n-Poly	n-Poly	n-Poly	n-Poly	Poly/metal	Poly/metal	Poly/metal	Metal
Highest W/E voltage (V) [9]	17-19	17-19	15-17	15-17	15-17	15-17	15-17	15-17	15-17
Endurance (erase/write cycles) [10]	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05
Nonvolatile data retention (years) [11]	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10-20
Maximum number of bits per cell (MLC) [12]	2	2	3	4	4	4	4	4	4

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

1. 3 Charge Trap-Type Nonvolatile Memory Devices

1. 3. 1 Emerge of SONOS-type Nonvolatile Memory

Since floating gate-type flash memory devices almost reaches its scaling limit, intensive researches are carrying on to study new memory structures. Those new memory structures should be capable of easily scaling down as well as simple device structure to have high-density memory array. As summarized in ITRS 2007 as shown in Table 1.3.1, the alternative memory structures which are actively explored include SONOS (Silicon / Oxide / Nitride / Oxide / Silicon), FeRAM (Ferroelectric Random Access Memory), MRAM (Magnetic Random Access Memory) and PCRAM (Phase Change Random Access Memory) [26]. However, the MRAM, FeRAM and PCRAM need one transistor and another component in a cell, which is not suitable for the ultra-high density NAND flash application. In recent years, the development of SONOS memory has been accelerated by its attractiveness for high-density embedded nonvolatile memory applications, mainly because of their potential of aggressive scaling, process compatibility with the CMOS technology, process simplicity and low integration costs.

The SONOS-type nonvolatile memory was developed in the 1970s. The initial device structures were p-channel metal-nitride-oxide-silicon (MNOS) structures with aluminum gate electrodes and thick (i.e., 45 nm) silicon nitride charge storage layer. Program/erase voltages were typically 25-30 V. In the late 1970s and 1980s, scaling moved to n-channel SNOS devices with program/erase voltages of 14-18V [27, 28]. In the late 1980s and early 1990s, n- and p-channel SONOS devices emerged.

Table 1.3.1: Current baseline for the volatile and nonvolatile memory devices as well as emerging research nonvolatile memory devices.

Table ERD3 Current Baseline and Prototypical Memory Technologies

		Baseline Technologies					Prototypical Technologies [A]			
		DRAM		SRAM [C]	Floating Gate [E]		Trapping Charge [G]	FeRAM	MRAM	PCM
		Stand-alone [A]	Embedded [C]		NOR	NAND				
Storage Mechanism		Charge on a capacitor		Inter-locked state of logic gates	Charge on floating gate		Charge trapped in gate insulator	Remnant polarization on a ferroelectric capacitor	Magnetization of ferromagnetic layer	Reversibly changing amorphous and crystalline phases
Cell Elements		1T1C		6T	1T		1T	1T1C	1(2)T1R	1T1R
Feature size F, nm	**2007**	68	90	65	90	90	65	180	90	65
	2022	12	25	13	18	18	10	65	22	18
Cell Area	**2007**	$6F^2$	$12F^2$	140 F^2	10 F^2	5 F^2	$6F^2$	$22F^2$	$20F^2$	$4.8F^2$
	2022	$6F^2$	$12F^2$	140 F^2	10 F^2	5 F^2	$5.5F^2$	$12F^2$	$16F^2$	$4.7F^2$
Read Time	**2007**	<10 ns	1 ns	0.3 ns	10 ns	50 ns	14 ns	45 ns [I]	20 ns [M]	60 ns [P]
	2022	<10 ns	0.2 ns	70 ps	2 ns	10 ns	2.5 ns	<20 ns [J]	<0.5 ns	< 60 ns
W/E Time	**2007**	<10 ns	0.7 ns	0.3 ns	1 µs/ 10 ms	1/0.1ms	20µs/20ms[H]	10 ns [K]	20 ns [M]	50/120ns[P]
	2022	<10 ns	0.2 ns	70 ps	1 µs/ 10 ms	1 ms/ 0.1 ms	~10µs/10ms	1 ns[J]	<0.5 ns [N]	<50 ns
Retention Time	**2007**	64 ms	64 ms	[D]	>10 y	> 10 y	>10 y	>10 y	>10 y	>10 y
	2022	64 ms	64 ms	[D]	>10 y	> 10 y	>10 y	>10 y	>10 y	>10 y
Write Cycles	**2007**	>3E16	>3E16	>3E16	>1E5	>1E5	1E5	1E14	>3E16	1E8
	2022	>3E16	>3E16	>3E16	>1E5	>1E5	1E6	>1E16	>1E16	1E15
Write Operating Voltage (V)	**2007**	2.5	2.5	1.1	12	15	7–9	0.9-3.3	1.5 [M]	3 [P]
	2022	1.5	1.5	0.7	12	15	4-6	0.7–1	<1.5	<3
Read Operating Voltage (V)	**2007**	2	2	1.1	2	2	1.6	0.9–3.3	1.5 [M]	3
	2022	1.5	1.5	0.7	1.1	1.1	1.1	0.7–1	<1.8	<3
Write Energy (J/bit)	**2007**	5E-15 [B]	5E-15	7E-16	>1E-14 [F]	>1E-14 [F]	1E-13 [H]	3E-14 [L]	7E-11 [A]	5E-12 [Q]
	2022	2E-15 [B]	2E-15	2E-17	>1E-15 [F]	>1E-15 [F]	>1E-15	5E-15 [L]	2E-11 [A]	<1E-13 [Q]
Comments					Multiple-bit potential	Multiple-bit potential	Multiple-bit potential	Destructive read-out	Spin-polarized Write has a potential to lower Write current density and energy [O]	Multiple-bit potential

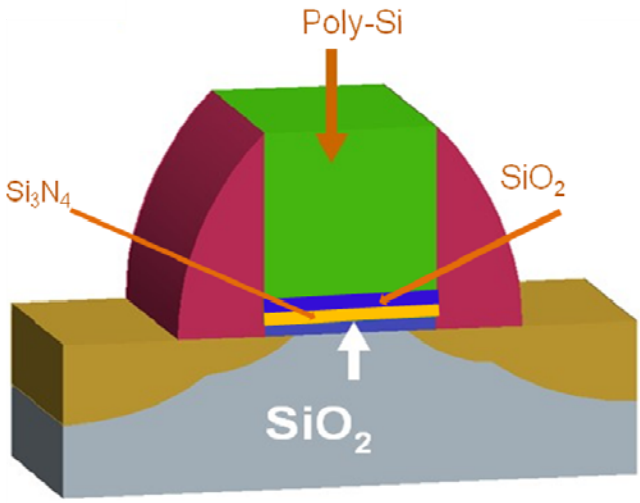


Fig. 1.3.1: Schematic diagram of a SONOS memory. The gate is n+ polysilicon.

When the gate dielectric of a MOSFET transistor is modified to incorporate a layer of silicon nitride, the structure can serve as a memory device with the threshold voltage controlled by the amount of charge stored in the silicon nitride. The structure of a SONOS device is shown in Fig. 1.3.1. The structure has a polysilicon gate and a multi-dielectric gate insulator consisting of an oxide-nitride-oxide (ONO) layer with charge storage in discrete traps in the silicon nitride layer. A net positive or negative charge is stored in the deep traps within the nitride dielectric depending on whether a positive or negative voltage is applied, respectively, to the gate. Fig. 1.3.2 illustrates the program/erase mechanism of a SONOS memory cell with the aid of energy band diagram. At the program state as shown in Fig. 1.3.2 (a), electrons quantum-mechanically tunnel from the substrate silicon inversion layer through an energy barrier height of 3.2 eV into a thin silicon oxide layer. The electrons tunnel through the thin silicon oxide layer into the silicon nitride film and are stored in deep-level traps, which lie about 1 eV below the edge of the nitride conduction band. Those trapped electrons shift the threshold voltage

accordingly and give a "0" state. The electrons, which are not trapped in the nitride film, tunnel through a blocking oxide into the gate electrode. During the erase operation (Fig. 1.3.2 (b)), holes are injected from the substrate into the silicon nitride valence band where they are trapped in a manner similar to electrons, which will shift the threshold voltage back to the initial value giving a "1" state. Electrons may tunnel from the n+ polysilicon gate electrode into the nitride, compensating the injected holes. This limits the shift in erase threshold voltage.

One key advantage of SONOS cell over the conventional floating gate-type flash memory cell is that most of the programmed electrons are trapped by local traps instead of within the conduction band in floating gate flash memory, Therefore, they will be retained even if there is a leakage path in the tunnel oxide and the charge retention of a SONOS cell is less sensitive to the tunnel oxide quality. It allows the use of a thinner tunnel oxide. The program/erase voltage as well as the electric field across the gate stack could also be reduced subsequently, since the charge injection mechanism changes from F-N tunneling to direct tunneling with the scaling down of the tunnel oxide thickness.

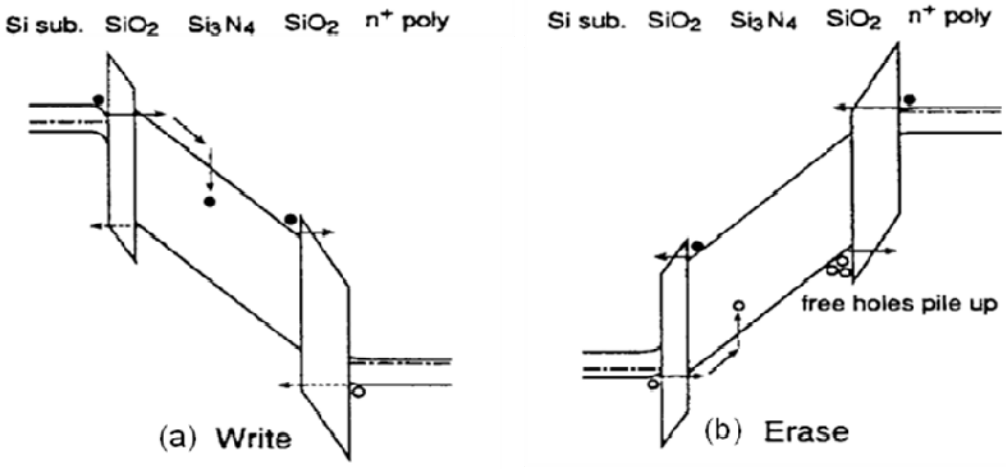


Fig. 1.3.2: Band diagram showing a SONOS flash memory cell at (a) Program sate. (b) Erase state [29, 30].

1. 3. 2 SONOS-type Flash Memory Engineering

To accelerate the commercialization of SONOS memory and enhance its performance, a lot of effort was made on the engineering of the SONOS structure. The engineering involves the gate, charge trapping layer, tunneling oxide as well as blocking oxide.

1. Gate Engineering

In order to enhance the erase speed, a modified SONOS structure with high work function gate instead of the conventional n+ polysilicon gate was proposed. The main reason for this enhancement comes from the reduced back tunneling electrons from the gate under negatively applied gate bias, which will seriously retard the erase speed. The enlarged memory window of the SONOS cell with p+ polysilicon gate was firstly demonstrated by *Reisinger et al* [31] as shown in Fig. 1.3.3. Similar enhancement with mid-gap TaN and high work function IrO_2 metal gate was also observed [32-34] .

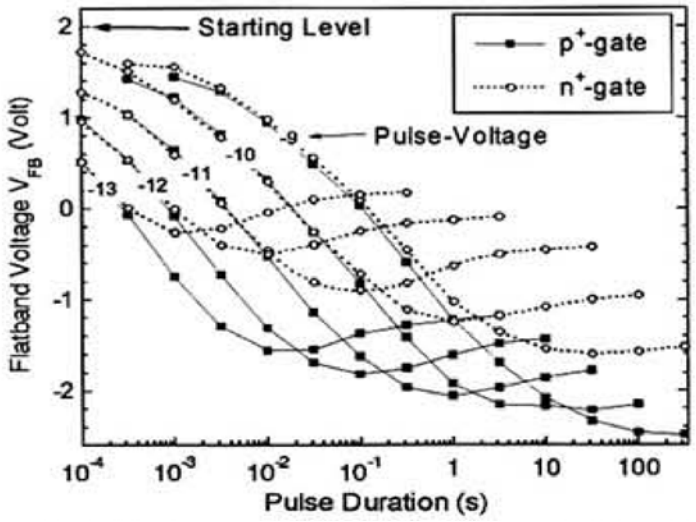


Fig. 1.3.3: Erase characteristics of SONOS MOS capacitors with n+ and p+ gate. Enhanced erase speed of the p+ polysilicon gate is observed [31].

2. Charge Trapping Layer Engineering

The increase in the program speed and lower voltage operation of SONOS devices had been accomplished previously by reducing the tunnel oxide thickness. However, this seriously degrades the retention capability of the device. To overcome such limitations, the so-called SOHOS (polysilicon-oxide-high-κ-oxide-silicon) flash memory structure has been attempted by replacing the nitride charge storage layer with a high-κ dielectric [35, 36]. High-κ material is physically much thicker than Si_3N_4 while having the same EOT. Therefore higher area trap densities can be expected. Additionally, the thicker high-κ increases the scattering possibility, hence increasing the possibilities of charge capturing and charge trapping efficiency [36]. Moreover, most of the high-κ materials have lower conduction band offset than Si_3N_4, and the electrons are expected to be stored in much deeper traps which are difficult to leak out. Therefore the charge retention will also be improved.

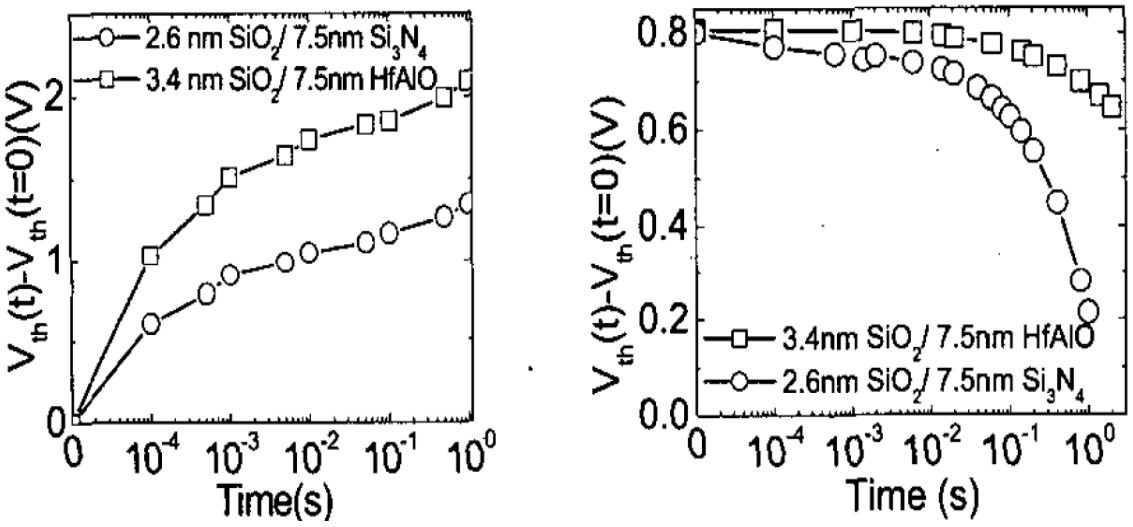


Fig. 1.3.4: Program/erase characteristics of SONOS and SOHOS memory cell with Si_3N_4 and HfAlO charge trapping layer. Enhanced operation speed is observed for high-κ charge trapping layer [37].

The trapping properties of HfAlO were investigated previously by research work, and it was proved to provide better electron trapping capability as shown in Fig. 1.3.4. Other high-κ materials, such as HfO_2, Al_2O_3 were also reported [36, 37] .

3. Tunnel Oxide Engineering

The tunnel oxide engineering mainly focuses on the bandgap engineered SONOS (BE-SONOS) device. This concept is demonstrated by a multilayer structure of $SiO_2/Si_3N_4/SiO_2$ tunnel oxide instead of the single layer SiO_2. The ultra-thin "O/N/O" provides a "modulated tunneling barrier" -- it suppresses direct tunneling at low electric field during retention, while it allows efficient hole tunneling erase at high electric field due to the smaller valence band offset of Si_3N_4. The principle is well explained with the aid of the energy band diagram as shown in Fig. 1.3.5. As shown in the band diagram, BE-SONOS device offers fast hole tunneling erase, while it is immune to the retention problem of the conventional SONOS. A similar idea was initially proposed for the IPD layer engineering in floating gate-type flash memory [38, 39]. Later, the principle was widely applied to SONOS devices to enhance the erase speed by increasing the hole tunneling current. Excellent reliability and performance were demonstrated experimentally [40, 41].

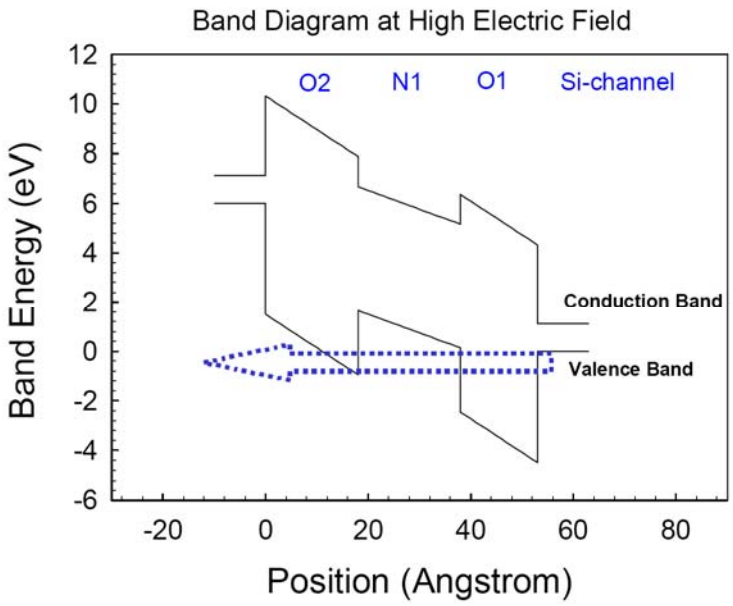


Fig. 1.3.5: Band diagram of the O/N/O tunneling dielectric under negative gate bias. The tunnel barriers of N1 and O2 are almost screened due to the band offset such that hole direct tunneling through O1 can happen [40].

4. Blocking Oxide Engineering

The blocking oxide for advanced SONOS structure is moving from SiO_2 towards high-κ dielectric. When the dielectric constant of the blocking oxide is increased, the electric field on the tunneling oxide will be enhanced, since the voltage drop across a dielectric is inversely proportional to its capacitance. This will improve the program/erase speed. Moreover, the electric field across the blocking layer is decreased with its increased dielectric constant due to the increased physical thickness, which will significantly reduce the electron back tunneling from the gate electrode during erase. In addition, the retention of the memory devices will be improved as well due to the retarded electric field on blocking oxide. For the above mentioned reasons, Lee *et al* made TANOS ($TaN/Al_2O_3/Si_3N_4/SiO_2/Si$) memory devices with Al_2O_3 as blocking oxide, and proved that Al_2O_3 offers better blocking effect than SiO_2 because the erase saturation

of the device is successfully overcomed [32, 42]. Improvement in retention performance of TANOS over the SONOS devices has also been reported [32].

Since Al_2O_3 only has a moderate κ value (~9), materials with even higher κ value were investigated. The requirements for the high-κ materials which are feasible for blocking layer applications are:

(a) High enough dielectric constant such that physically thicker films could be used without any EOT increase.

(b) Relatively large bandgap and conduction band offset to Si, which could reduce the electron back tunneling current from the gate during the erase or prevent electron leaking out at the retention state.

(c) Good thermal stability so that its material property remains unchanged after the high source/drain annealing temperature.

Table 1.3.2: Comparison of dielectric constant, bandgap as well as conduction band offset for various high-κ candidates [43].

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
Y_2O_3	15	5.6	2.3
La_2O_3	30	4.3	2.3
Ta_2O_5	26	4.5	1–1.5
TiO_2	80	3.5	1.2
HfO_2	25	5.7	1.5
ZrO_2	25	7.8	1.4

However, in reality it is difficult to find a material with both sufficient large κ (>25) value as well as large bandgap. For most materials, the dielectric constant and bandgap generally exhibit an inverse relationship as shown in Table 1.3.2 [43, 44]. Therefore, the required permittivity must be balanced, however, against the barrier height for the tunneling process.

1. 4 Organization of Thesis

The overall objective of this work is to apply novel gate materials for the performance enhancement of flash memory devices, including both floating gate-type flash memory devices and SONOS-type flash memory devices. These attempts could be of practical value for the flash memory devices, especially in improving the operation speed and data retention.

In Chapter 2, a novel floating gate engineering scheme, the carbon doped polysilicon floating gate, will be proposed to overcome the scaling barrier for floating gate-type flash memory devices. It has been found that incorporation of carbon into the conventional n+ polysilicon floating will be able to significantly improve the program/erase speed, especially for devices with small coupling ratio (~0.3), which is the bottleneck for sub 30nm flash memory technology as indicated in the ITRS roadmap [26]. The data retention of such devices isimproved as well. All these improved properties originate from the increased conduction band offset of the floating gate itself by the incorporation of carbon and the silicon carbide nano-structure formation is responsible

for the band structure change. By adopting the carbon doped polysilicon floating gate, there will be little process modification to the current technology, which is an effective and simple solution for floating gate-type flash memory scaling.

In the advanced SONOS-type flash memory devices, the application of high dielectric constant materials as the blocking oxide attracts much of the research interest. In Chapter 3, the feasibility of a novel rare earth high-κ material, Gd_2O_3, as the potential candidate for the blocking layer application in SONOS-type flash memory devices will be evaluated. The material properties of Gd_2O_3, including deposition method, leakage current performance, crystal information as well as the band structure have been studied systematically. Control of the crystal structure of Gd_2O_3 will be the key point for high quality film. SONOS transistors with Gd_2O_3 blocking layer exhibits superior performance over Al_2O_3 blocking layer in several aspects, such as program/erase speed, room temperature retention, etc.

Experimental results have demonstrated that Gd_2O_3 is a favorable blocking oxide candidate except that the charge retention after cycling remains problematic. Doping of Al into the pure Gd_2O_3 is proposed for the robust data retention after cycling, since the increase in the conduction band offset is always an effective method to block the electron leakage, both for room temperature and high temperature retention. The optimized Al concentration needs to be carefully considered to balance all the following factors: dielectric constant, conduction band offset, film morphology as well as memory characteristic. All those questions will be well addressed in Chapter 4.

In Chapter 5, structure optimization of the SONOS cell with $GdAlO_x$ blocking oxide will be discussed. The study focuses on the relationship between the blocking layer

thickness and long term retention reliability at room temperature, after program/erase cycles and at elevated temperature. A novel leakage current separation technique will be applied to differentiate the leakage paths in SONOS memory in order to improve the retention effectively. Charge leakage mechanisms for SONOS-type flash memory devices will be discussed in this chapter as well. Finally, the thesis is completed with summary and conclusions in Chapter 6.

References

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

[2] C. Hwang, "Insights on memory technology and the market," *IEEE Circuits and Devices Magazine,* vol. 17, pp. 12-18, 2001.

[3] Y. Q. Wang, "Study on Application of high-k Dielectric Materials for Discrete Charge Storage Memory ": Ph. D dissertation, National Univ. of Singapore, 2006.

[4] M. She, "Semiconductor Flash Memory Scaling," Ph. D dissertation, Univ. of California, Berkeley, CA 2003.

[5] Y. King, "Thin dielectric technology and memory devices," Ph. D dissertation, Univ. of California, Berkeley, CA 1999.

[6] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proceedings of the IEEE,* vol. 91, pp. 489-502, 2003.

[7] C. Chang and S. Sze, *ULSI devices*: Wiley New York, 2000.

[8] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, "Flash memory cells-an overview," *Proceedings of the IEEE,* vol. 85, pp. 1248-1271, 1997.

[9] R. Degraeve, F. Schuler, B. Kaczer, M. Lorenzini, D. Wellekens, P. Hendrickx, M. Van Duuren, G. Dormans, J. Van Houdt, and L. Haspeslagh, "Analytical percolation model for predicting anomalous charge loss in flash memories," *IEEE Trans. Electron Dev.,* vol. 51, pp. 1392-1400, 2004.

[10] L. Chih-Yuan, H. Kuang-Yeu, and R. Liu, "Future challenges of flash memory technologies," *Microelectronic Engineering,* vol. 86, pp. 283-286, 2009.

[11] R. Yamada and T. J. King, "Variable stress-induced leakage current and analysis of anomalous charge loss for flash memory application," *2003 IEEE International Reliability Physics Symposium Proceedings 41st Annual (Cat. No.03CH37400),* pp. 491-496, 2003.

[12] D. Ielmini, A. S. Spinelli, A. L. Lacaita, L. Lacaita, and A. Modelli, "Modeling of anomalous SILC in flash memories based on tunneling at multiple defects," *Solid State Electronics,* vol. 46, pp. 1749-1756, 2002.

[13] K. Lim and C. Ling, "Charge trapping in interpoly ONO film," *1998 IEEE International Conference on Semiconductor Electronics,* pp. 42-46, 1998.

[14] S. Mori, Y. Araki, M. Sato, H. Meguro, H. Tsunoda, E. Kamiya, K. Yoshikawa, N. Arai, and E. Sakagami, "Thickness scaling limitation factors of ONO interpoly dielectricfor nonvolatile memory devices," *IEEE Trans. Electron Dev.,* vol. 43, pp. 47-53, 1996.

[15] S. Mori, Y. Kaneko, N. Arai, Y. Ohshima, H. Araki, K. Narita, E. Sakagami, and K. Yoshikawa, "Reliability study of thin inter-poly dielectrics for non-volatile memory application," *28th Annual Proceedings. Reliability Physics 1990 (Cat. No.90CH2787-0),* pp. 132-144, 1990.

[16] B. Govoreanu, D. Brunco, and J. Van Houdt, "Scaling down the interpoly dielectric for next generation Flash memory: Challenges and opportunities," *Solid State Electronics,* vol. 49, pp. 1841-1848, 2005.

[17] Y. Yamaguchi, E. Sakagami, N. Arai, M. Sato, E. Kamiya, K. Yoshikawa, H. Meguro, H. Tsunoda, and S. Mori, "ONO interpoly dielectric scaling limit for non-volatile memory devices," *Symp.VLSI Tech. Dig.,* pp. 85-86, 1993.

[18] S. Aritome, S. Satoh, T. Maruyama, H. Watanabe, S. Shuto, G. Hemink, R. Shirota, S. Watanabe, and F. Masuoka, "A 0.67um^2 self-aligned shallow trench isolation cell (SA-STI cell) for 3 V-only 256 Mbit NAND EEPROMs," *IEDM Tech. Dig.,* pp. 61-64, 1994.

[19] M. Ichige, Y. Takeuchi, K. Sugimae, A. Sato, M. Matsui, T. Kamigaichi, H. Kutsukake, Y. Ishibashi, M. Saito, S. Mori, H. Meguro, S. Miyazaki, T. Miwa, S. Takahashi, T. Iguchi, N. Kawai, S. Tamon, N. Arai, H. Kamata, T. Minami, H. Iizuka, M. Higashitani, T. Pham, G. Hemink, M. Momodomi, and R. Shirota, "A novel self-aligned shallow trench isolation cell for 90nm 4Gbit NAND flash EEPROM s," *Symp.VLSI Tech. Dig.,* pp. 89-90, 2003.

[20] "International Technology Roadmap for Semiconductors, 2005 update," http://public.itrs.net.

[21] K. Kim, "Technology for sub-50nm DRAM and NAND flash manufacturing," *IEDM Tech. Dig.,* pp. 323-326, 2005.

[22] Fumitaka.Arai, "Future outlook of Floating Gate Flash Memory," *International Conference on Solid State Devices and Materials,* pp. 292-293, 2006.

[23] F.Arai, "Future outlook of Floating Gate Flash Memory," *International Conference on Solid State Devices and Materials,* pp. 292-293, 2006.

[24] G. Molas, D. Deleruyelle, B. De Salvo, G. Ghibaudo, M. Gely, S. Jacob, D. Lafond, and S. Deleonibus, "Impact of few electron phenomena on floating-gate memory reliability," *IEDM Tech. Dig.,* pp. 877-880, 2004.

[25] N. Mielke, "Reliability of flash nonvolatile memories," *International Journal of High Speed Electronics and Systems,* vol. 11, pp. 719-750, 2001.

[26] "International Technology Roadmap for Semiconductors, 2007 update," http://public.itrs.net.

[27] M. H. White, Y. Yang, P. Ansha, and M. L. French, "A low voltage SONOS nonvolatile semiconductor memory technology," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A,* vol. 20, pp. 190-195, 1997.

[28] F. Libsch and M. White, "Nonvolatile semiconductor memory technology," *SONOS Nonvolatile Semiconductor Memories,* pp. 309-357, 1998.

[29] S. Wrazien, Y. Zhao, J. Krayer, and M. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," *Solid State Electronics,* vol. 47, pp. 885-891, 2003.

[30] J. Bu and M. H. White, "Design considerations in scaled SONOS nonvolatile memory devices," *Solid State Electronics,* vol. 45, pp. 113-120, 2001.

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

[32] C. Lee, K. Choi, M. Cho, Y. Song, K. 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," *IEDM Tech. Dig,* pp. 613-616, 2003.

[33] 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," *Symp. VLSI Tech. Dig. ,* pp. 210-211, 2005.

[34] Y. Wang, P. Singh, W. Yoo, Y. Yeo, G. Samudra, A. Chin, W. Hwang, J. Chen, S. Wang, and D. Kwong, "Long retention and low voltage operation using IrO_2

HfAlO/HfSiO/HfalO gate stack for memory application," *IEDM Tech. Dig.,* pp. 162-165, 2005.

[35] 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," *Symp. VLSI Tech.,* pp. 27-28, 2003.

[36] 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 Dev.,* vol. 51, pp. 1143-1147, 2004.

[37] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, T. H. Ng, and B. J. Cho, "High-k HfAlO charge trapping layer in SONOS-type nonvolatile memory device for high speed operation," *IEDM Tech. Dig.,* pp. 889-892, 2004.

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

[39] 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.

[40] 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," *IEDM Tech. Dig,* vol. 2005, pp. 547-550, 2005.

[41] 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, 2005.

[42] S. Jeon, J. Han, J. Lee, C. Choi, S. Choi, H. Hwang, and C. Kim, "High work-function metal gate and high-k dielectrics for charge trap flash memory device applications," *Proceedings of 35th European Solid-State Device Research Conference,* pp. 325-328, 2005.

[43] 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.

[44] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices," *International conference on silicon dielectric interfaces,* vol. 18, pp. 1785-1791, 2000.

CHAPTER 2

CARBON DOPED POLYSILICON FLOATING GATE FLASH MEMORY DEVICES

2. 1 Introduction and Motivation

The ever increasing packing density of flash memory has been driven by both areal scaling and multi-level operation [1]. However, as flash memory scales down beyond the 40 nm technology node, both approaches face significant challenges. As discussed in Chapter 1, the drastic drop in coupling ratio to be about 0.3 - 0.4 [2-4] would lead to inefficient program/erase (P/E) window. It has been suggested that high-κ material should replace the traditional oxide – nitride - oxide stack at the IPD layer to aggressively scale down the EOT of IPD, and thus maintain satisfactory coupling ratio. However, integrating the new material faces both technology and reliability difficulties [3, 5], and is not immediately available. On the other hand, as the volume of the floating gate has been shrinking with the scaling in all three dimensions to minimize interference between adjacent cells [6], the number of electrons stored in the floating gate has been

continuously decreasing [7, 8]. This problem is especially serious for flash memory with multi-level operation, where the window between two different logic levels (ΔV_{th}) is smaller. Therefore, technology and structure improvement for floating gate-type flash memory devices would be highly desirable.

In this chapter, we will assess the possibility of using carbon doped poly-crystalline silicon as the floating gate. The proposed cell structure demonstrates much larger operation P/E window and significantly improved data retention which is important for multi-level operation.

Silicon carbide (SiC-3C) is a wide band gap semiconductor. The basic physical parameters are given in Table. 2.1.1. Comparing to silicon, silicon carbide cubic structure has a larger conduction band offset (3.62 eV) against SiO_2 than polysilicon (3.2 eV) does [9]. If the silicon carbide could be successfully formed within the polysilicon layer by doping certain amount of carbon and subsequently annealing at sufficient high temperature, it will be expected to have better electron storage capability and data retention compared to the conventional polysilicon floating gate, beneficial from the enlarged conduction band offset.

Table 2.1.1: Comparison of physical parameters of silicon carbide and polysilicon. The ΔE_c and ΔE_v value is taking silicon dioxide as the reference.

	E_g	ΔE_c	ΔE_v	Lattice Constant
Silicon Carbide	2.35 eV	3.62 eV	3.05 eV	4.359 Å
Polysilicon	1.1 eV	3.2 eV	4.7 eV	5.431 Å

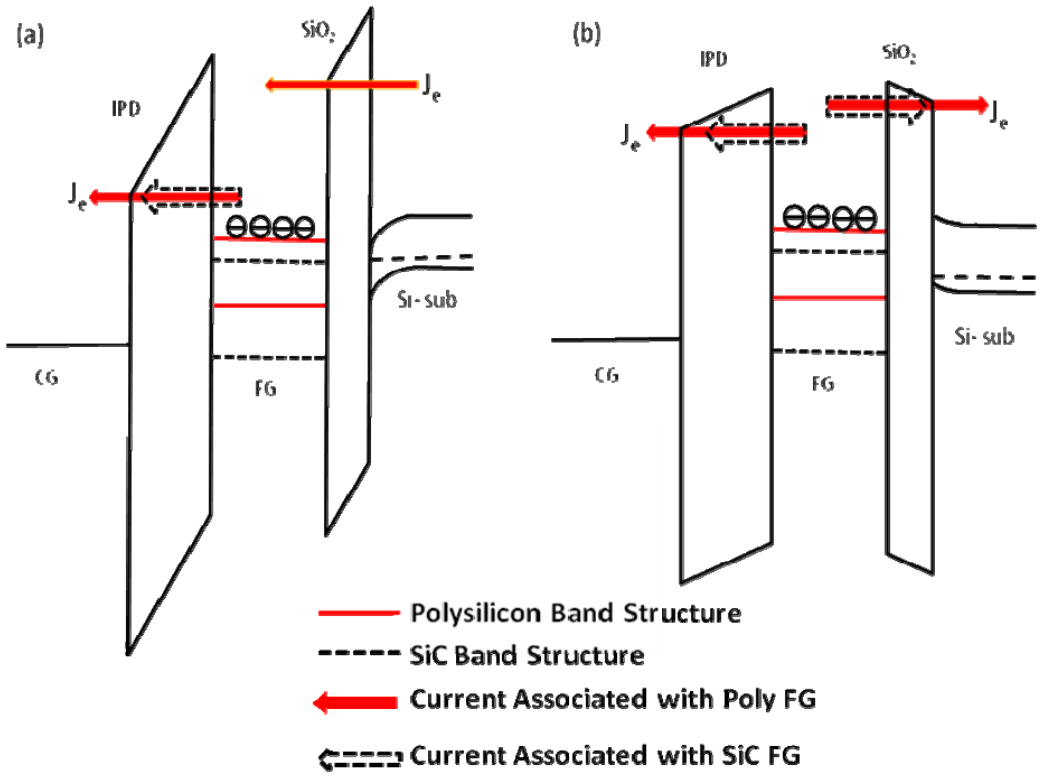


Fig. 2.1.1: Band diagram illustrating the band bending during (a) Program state and (b) Retention state for both of n+ polysilicon FG and carbon doped polysilicon FG.

To explain the reason that higher conduction band offset FG will give faster P/E speed and better retention, the energy band diagrams for both convention n+ polysilicon FG and carbon doped polysilicon FG are drawn in Fig. 2.1.1 for the program and retention cases. Figure 2.1.1 (a) shows the band diagram at program state when a large positive gate bias is applied, while Fig. 2.1.1 (b) is the band structure at retention state. Please note that the depletion layers induced by the gate voltage for both n+ polysilicon FG and carbon doped polysilicon FG are ignored, since all the FGs are heavily doped. As illustrated in Fig. 2.1.1 (a), by doping carbon into the polysilicon FG, although the electric field across the tunnel oxide and IPD layer remains the same, the out tunneling current through the IPD layer reduced significantly due to the increased tunneling barrier

by adopting carbon doped polysilicon FG. The total amount of charge stored in the FG is determined by two factors. One is the tunneling-in current from the tunnel oxide and another is the out-going current through the IPD layer. By reducing the out tunneling current will allow more electrons to be stored. Erasing saturation can be analyzed in a similar argument, which will also result in faster erase speed.

Besides P/E window improvement, enhancing the data retention is another motivation to incorporate carbon into the polysilicon film. As shown in Fig. 2.1.1 (b), during retention, the leakage current through both the tunnel oxide and IPD layer can be suppressed by the deeper tunneling barrier. The charge leakage mechanism for floating gate-type flash memory devices has been systematically studied by many researchers [10-12]. All those modules invoke the direct-tunneling current, trap-assisted tunneling current as well as thermal emission current through the tunnel oxide and IPD layer. Hence, possible ways to improve the charge retention could be either by increasing the dielectric layer thickness, reducing the electric field or increasing the effective tunneling barrier for tunnel oxide and IPD layer. Due to the scaling limitation of tunnel oxide (~ 8 nm) and IPD layer (~ 12 – 13 nm) [13, 14], there is less room left for the thickness scaling. Hence the operation voltage and the electric field across those two layers also have to remain unchanged to fulfill the multi-level operation requirement. As a result, modifying the tunneling barrier is the only practical solution for regulating the retention performance without sacrificing the P/E speed.

Moreover, as silicon carbide is a wide band gap semiconductor (2.35 eV), the hole current will be effectively reduced due to the smaller valence band offset of silicon carbide (1.65 eV smaller) than that of polysilicon against SiO_2 tunnel oxide for negative

gate bias. This may minimize positive charge trapping, which will result in more robust tunnel oxide against the charge leak. Thus, the motivation for this work is to forming silicon carbide phase within the carbon doped polysilicon film. Improvement in the P/E window and retention will be achieved through the floating gate band structure engineering.

2. 2 Deposition Chemistry and Material Property

2. 2. 1 Film Deposition Chemistry

Carbon doped polysilicon was deposited by using Si_2H_6 and SiH_3CH_3 gas chemistry at 650^{o}C. The deposition was carried out in JUSUNG-UHVCVD chamber, which has excellent film thickness and composition uniformity. The deposited film has uniformity of both thickness and carbon concentration better than 2% over 8 inch wafer. Fig. 2.2.1 illustrates the carbon concentration and deposited thickness in polysilicon film as the function of SiH_3CH_3 flow rate. The Si_2H_6 flux is fixed at 20 sccm. The deposition time is 1500s and deposition temperature is fixed at 650^{o}C. By tuning the SiH_3CH_3 gas flux ranging from 30 sccm to 300 sccm, 2% to 10% carbon concentration is achieved. The resulting carbon concentration is confirmed by XPS, while the film thickness is measured by ellipsometer. The film thickness could be precisely controlled in the range of 100 nm to 250 nm. The sheet resistance of the deposited film was measured by Four-Point Probe, and the resistivity was calculated as $\rho = R_s \times X_j$, where R_s is the measured

sheet resistance and X_j is the film thickness. It has been found that once carbon concentration goes up to 10%, there is a significant increase in the film resistivity (×100) comparing to pure polysilicon control film after 950°C, 30s RTP anneal, as shown in Fig. 2.2.2. A similar phenomenon was also reported by other group [15]. Decreasing in effective carrier concentration is believed to be one of the root causes [15].

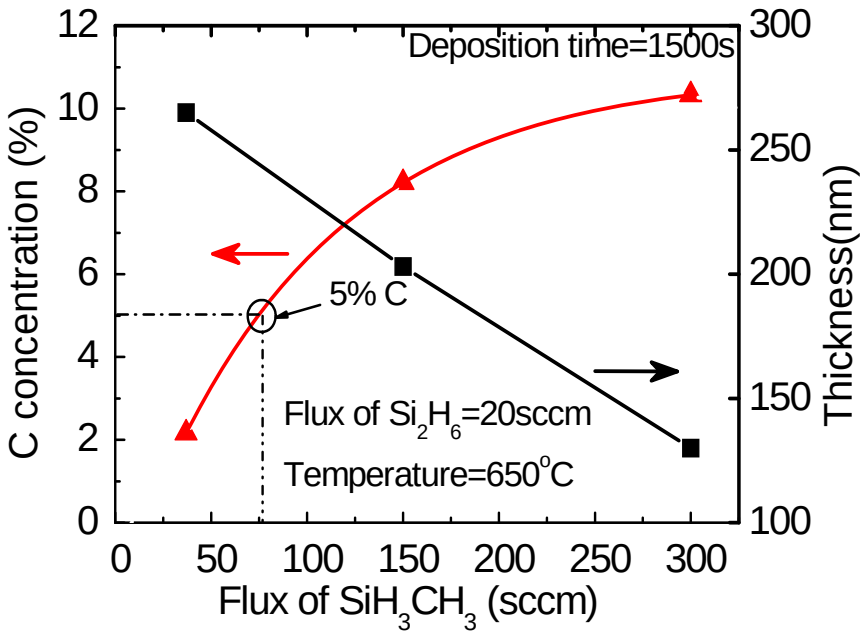


Fig. 2.2.1: Carbon concentration in polysilicon and deposited thickness as a function of SiH_3CH_3 flow rate.

In conventional polysilicon FG flash memory, the conduction band is flat during P/E. Hence, the voltage drop across the floating gate is zero. However, for carbon doped polysilicon FG, if the film resistivity goes too high, decrease in effective carrier concentration will cause the conduction band to be tilted instead of being flat. The voltage drop across the FG is not zero anymore. As a result, the voltage across the tunnel oxide is effectively reduced and the P/E speed will be affected accordingly. Therefore,

compared to polysilicon FG, if carbon concentration goes too high, much higher control gate voltage is required in order to have similar P/E speed, which is not desirable for FG application. For such reasons, 5% carbon doped polysilicon film is selected for electrical characterization.

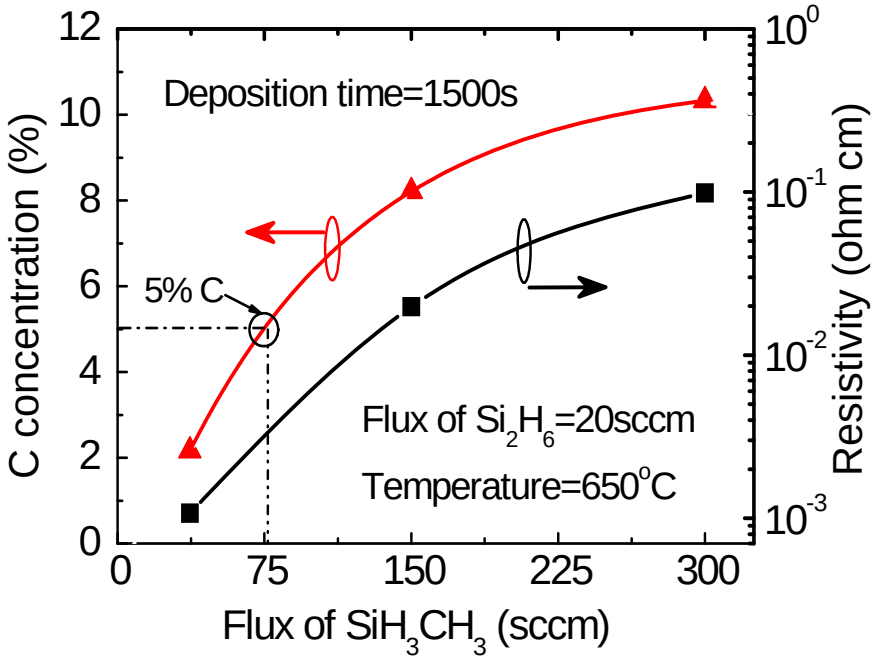


Fig. 2.2.2: Carbon concentration in polysilicon and film resistivity as a function of SiH_3CH_3 flow rate.

2. 2. 2 Chemical State Analysis by XPS

The XPS analysis is performed *ex-situ*. The sample was 100 nm 5% carbon doped polysilicon film, and it was transferred to the XPS chamber immediately after 950°C, 30s RTP anneal to minimize the possible contaminations induced by the exposure to air. The Gaussian-Lorentzian curve fitting method was used for XPS data analysis. Both silicon *2p* and carbon *1s* spectra were analyzed carefully.

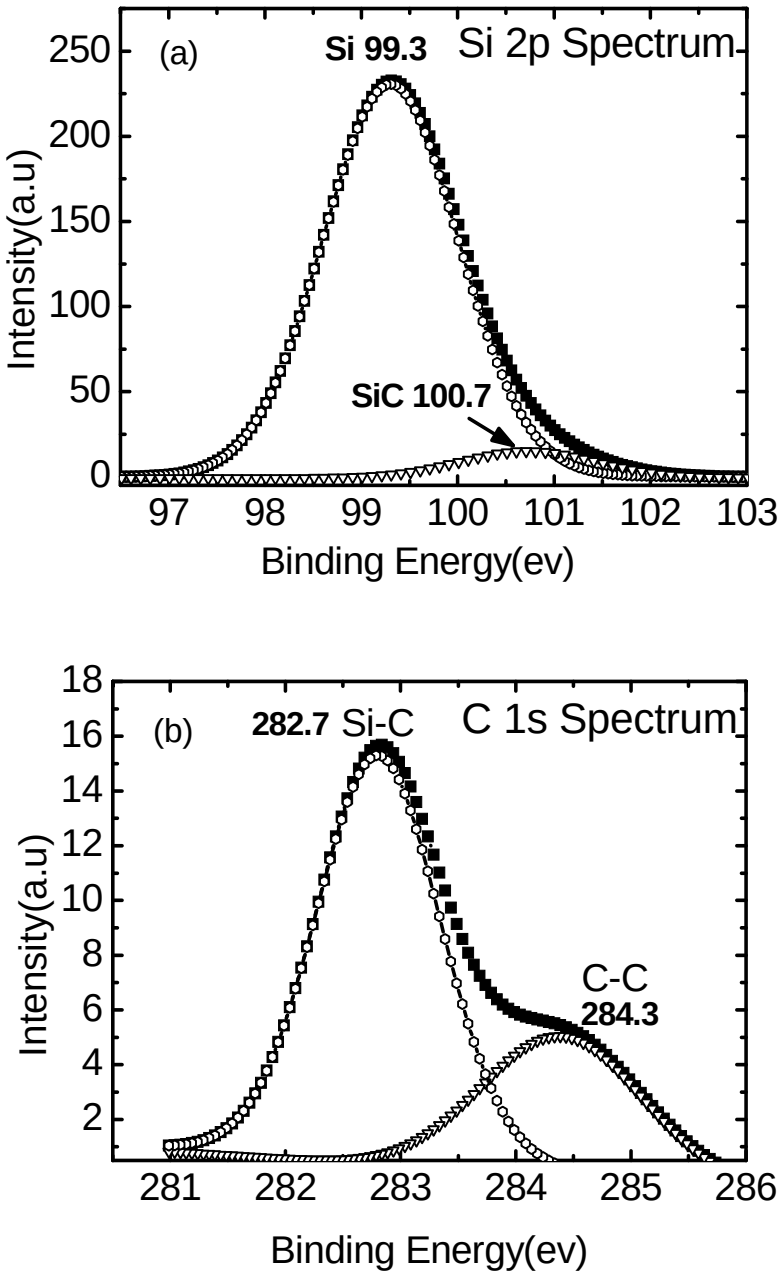


Fig. 2.2.3: XPS spectra for 5% carbon doped polysilicon film. Both the higher binding energy of 100.7 eV in (a) *Si 2p* spectra and the lower binding energy of 282.7 eV in (b) *C 1s* spectra indicate the silicon carbide phase formation in polysilicon after 950^{o}C anneal [16].

The XPS analysis results for silicon *2p* and carbon *1s* spectra are shown in Fig. 2.2.3. The higher binding energy (100.7 eV) in *Si 2p* and lower binding energy (282.7 eV) in *C 1s* spectra suggest the Si-C bond formation. The existence of silicon carbide phase [17] is confirmed. The lower binding energy in *Si 2p* spectra (99.3 eV) refers to Si-Si bond, while the higher binding energy in *C 2s* spectra refers to C-C bond.

2. 2. 3 Film Morphology Analysis by FTIR

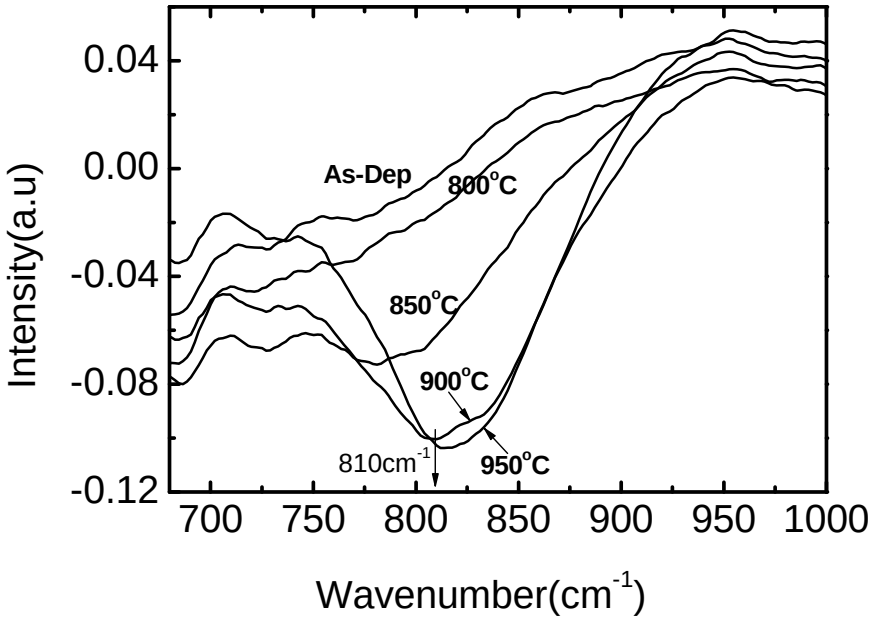


Fig. 2.2.4: FTIR spectra for 5 % carbon doped polysilicon film measured at room temperature. The film annealing temperature is ranging from 800°C to 950°C.

Figure 2.2.4 shows the Fourier Transform Infrared Spectroscopy (FTIR) spectra of the deposited 5% carbon doped polysilicon film at different annealing temperature. The film thickness is 100 nm. The as deposited and 800°C annealed samples show no pronounced peak, which indicates no crystal structure formation. Once the annealing temperature raises to above 850°C, the film morphology starts to change. After carefully

examining the FTIR spectra for carbon doped polysilicon film annealed at 900^{o}C in Fig. 2.2.4, it shows a peak at (810 cm^{-1}), which corresponds to silicon carbide nanostructure [17, 18]. If the annealing temperature is further increased to 950^{o}C, there is a slight shift in peak position to 814 cm^{-1}. This phenomenon is believed to be due to the change in the film strain, as reported by Zhao *et al* [18].

Both the FTIR and XPS analysis results evidence that the formation of the cubic structured silicon carbide phase could be achieved by annealing the deposited film in high temperature, which provides the necessary thermal energy for phase formation. The annealing temperature could not be lower than 900^{o}C, as such a temperature is required to activate the source/drain (S/D) dopants.

2. 2. 4 Compatibility Study with SiO_2 Gate Dielectric

From the FTIR and XPS analysis result, the formation of the cubic structured silicon carbide phase is successfully detected. Another consideration for exploring carbon doped polysilicon FG is whether the presence of carbon will degrade the tunnel oxide quality, which is the critical factor to determine the data retention.

In order to check if carbon will result in poor tunnel oxide (SiO_2) quality, MOS capacitor with SiO_2 gate dielectric, and pure polysilicon or 5% carbon doped polysilicon gate was fabricated. The gate dielectric was 4 nm thermally grown SiO_2 and the gate thickness was 150 nm for both cases. Arsenic was implanted into all types of gate. The implantation condition was $5\times10^{15}/\,cm^2$, 50 keV with 7^{o} tilt. All samples were subjected to 900^{o}C, 30s, RTP anneal for effective dopant activation. After gate patterning and

plasma etching, both the current – voltage (IV) and capacitance – voltage (CV) characteristics were measured. The current - voltage and capacitance - voltage curves are shown in Fig. 2.2.5 (a) and Fig. 2.2.5 (b), respectively. The leakage current performance in Fig. 2.2.5 (a) indicates that the presence of carbon does not cause any significant degradation in the tunnel oxide quality after high temperature anneal. The capacitance - voltage curve in Fig. 2.2.5 (b) shows slight poly-depletion effect under the positive gate bias. As mentioned previously, decreasing in effective carrier concentration is believed to be the root cause [15]. However, this phenomenon has negligible effect on device operation since the percentage of carbon is well controlled at about 5%.

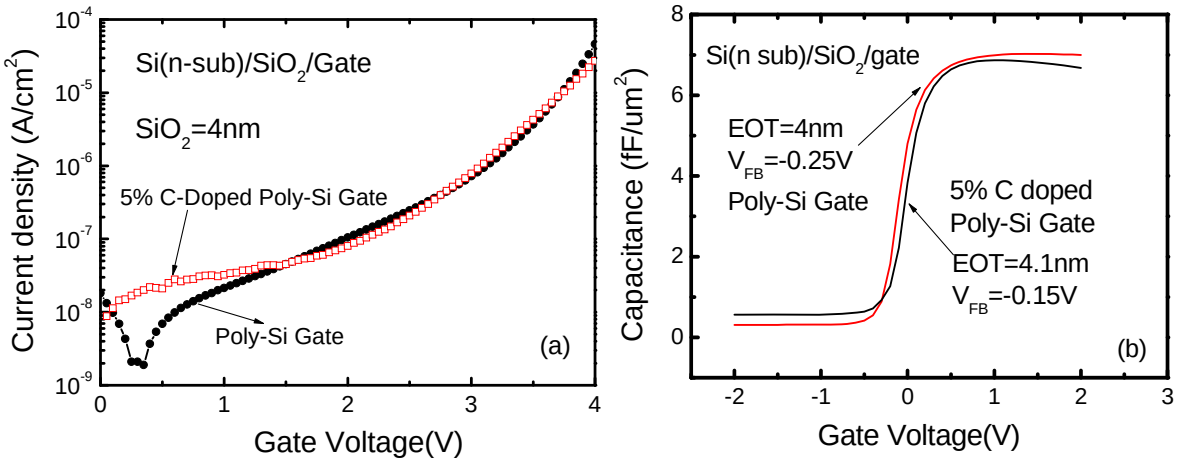


Fig. 2.2.5: Comparison of the (a) current – voltage curve and (b) capacitance - voltage curve of SiO_2 on polysilicon and the 5% carbon doped polysilicon gate. The result shows no degradation of tunnel oxide quality with incorporation of carbon.

2. 3 Experiments and Devices Fabrication

The feasibility study of carbon doped polysilicon film as a potential floating gate material shows promising result. The preferred cubic silicon carbide phase is successfully formed, whereas the presence of carbon has negligible effect on tunnel oxide quality. Hence, flash memory transistors with 5% carbon doped polysilicon FG together with polysilicon FG control sample were fabricated for memory characteristic evaluation. Table 2.3.1 summaries the split conditions, IPD layer thickness as well as the entire gate stack EOT extracted from the CV curve.

Table 2.3.1: Split table for IPD layer thickness and floating gate variations. The EOT of the entire gate stack was extracted from the CV measurement.

Split	IPD Layer Thickness	EOT(nm)
Poly FG	**SiO_2 = 18 nm**	**26.2 nm**
5% Carbon Doped FG	**SiO_2 = 18 nm**	**25.7 nm**
Control Sample	**Tunnel Ox**	**6.2 nm**

P-type Si (100) wafers were used as the starting material. Following the formation of 400 nm field oxide isolation, 6 nm oxide was thermally grown as a tunnel oxide. All floating gate thickness was 100 nm. Arsenic was implanted for n-type doping. The implantation condition was $3\times10^{15}/cm^2$, 40 keV and with 7^o tilt. The inter-poly dielectric is 18 nm SiO_2 deposited by LPCVD. The control gate was PVD deposited 150 nm tantalum nitride. After gate patterning, gate stack etching, S/D implantation and dopant

activation anneal (900^{o}C 30s RTP), 400 nm CVD oxide was deposited as a passivation layer. Finally, aluminum metallization and alloying was performed to provide contacts to the device.

In order to evaluate performance of memory transistors with different coupling ratio, a special set of masks were designed to allow transistors with various coupling ratios to be integrated in a single make layout. The schematic for the mask layout is shown in Fig. 2.3.1. The area ratio (A_{IPD} / $A_{tunnel\ ox}$) of the cells is varied by adjusting the area of poly mask extension over the field oxide isolation. The IPD area, $A_{IPD,}$ equals the area of the drawn poly layer, while the tunnel oxide area, $A_{tunnel\ ox}$, is determined by the intersection of the poly region and the active region. The area ratio is made to vary from 1.4 to 6, corresponding to a coupling ratio (CR) ranging from 0.3 - 0.8 approximately.

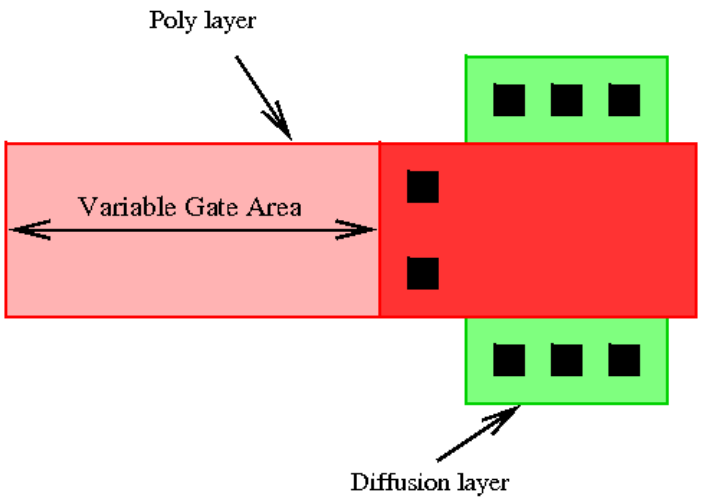


Fig. 2.3.1: Schematic drawing for the transistor mask lay-out. The coupling ratio in the test pattern is varied by stretching the area of poly mask on top of field oxide.

2. 4 Results and Discussion

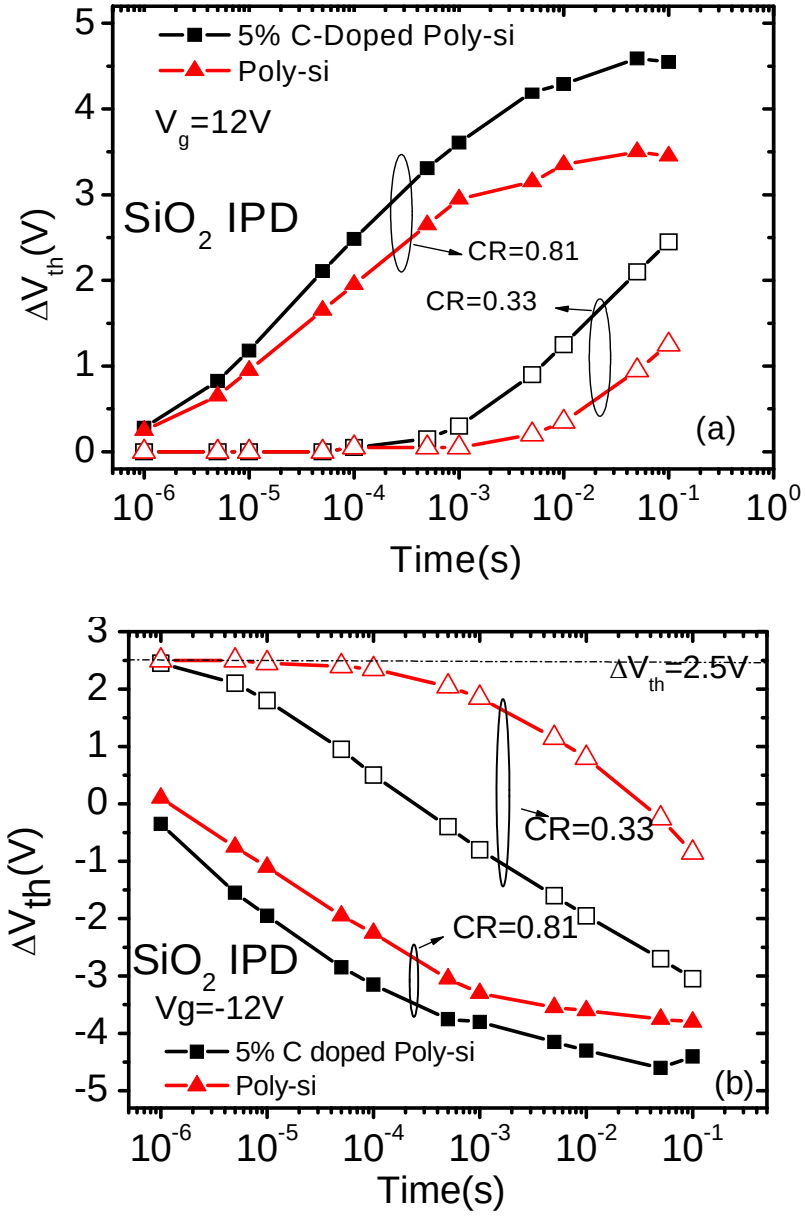


Fig. 2.4.1: (a) Comparison of program speed for 5% carbon doped polysilcion and pure polysilcion FG based on SiO_2 IPD for both large (0.81) and small (0.32) coupling ratio devices. (b) Comparison of erase speed for large and small coupling ratio devices.

The P/E characteristics in Fig. 2.4.1 show that incorporation of carbon can provide faster program and higher program saturation voltage and lower erase saturation voltage. Here, the P/E voltage was fixed at ± 12 V. For 5% carbon doped polysilicon FG, the V_{th} shift after 100 ms program is 4.55 V and the V_{th} shift after 100 ms erasing is -4.4 V for devices with a CR of 0.81. At the same P/E condition but 0.33 CR, the program V_{th} shift is 2.45 V, while the erase V_{th} shift is -3.05 V. The overall V_{th} window reduction is only 39% when the CR drops from 0.81 to 0.33 for carbon doped polysilicon FG. On the other hand, for pure polysilicon FG with the same P/E condition, the program V_{th} shift is 3.45 V and 1.25 V, and the erase V_{th} shift is -3.8 V and -0.85 V, respectively, for devices with CR of 0.81 and 0.33. There is a drastic drop in the P/E window up to 71% for pure polysilicon FG when the CR drops from 0.81 to 0.33. In terms of P/E window, 5% carbon doped polysilicon FG has 23% enhancement compared to pure polysilicon FG when CR = 0.81. This enhancement is as large as 160% when CR = 0.33, which makes this technique a feasible solution for 40 nm technology node and beyond.

The reason for the much more enhanced P/E window, especially for small coupling ratio devices, could be explained as follows. For a simplistic case of single-layer SiO_2 IPD, one first applies Gauss's law to the floating gate surface $\oint_S \varepsilon E \cdot dA = Q_{FG}$ where Q_{FG} is the charge stored in the floating gate and the integration runs over the surface of the floating gate. If the small electric field on the side-walls of the floating gate is ignored, the electric field in IPD and tunnel oxide are then, to the first order, related by

$$A_{IPD} \cdot \varepsilon \cdot E_{IPD} = A_{ox} \cdot \varepsilon \cdot E_{ox} + Q_{FG} \tag{2.1}$$

$$V_{CG} = E_{IPD} \cdot t_{IPD} + E_{ox} \cdot t_{ox} \tag{2.2}$$

where V_{CG} is the control gate voltage; A_{ox}, E_{ox}, and t_{ox} are the area, electric field and thickness of the tunnel oxide, respectively; and similarly A_{IPD}, E_{IPD}, and t_{IPD} for IPD. As program proceeds, more electrons are stored in the floating gate, which increases E_{IPD} and reduces E_{ox}. As a result, the tunneling current through IPD increases and the current through the tunnel oxide decreases. Program saturates when the two currents finally balance. In current technology, A_{IPD} is 2 - 3 times of A_{ox}, which allows a large number of electrons to be programmed. However as the ratio A_{IPD} / A_{ox} decreases towards unity, less electrons can be programmed to the FG, and finally no electron could be stored. When carbon doped polysilicon FG is used, the tunneling current through IPD is suppressed due to lower electron concentration at the interface with IPD. Program saturation thus happens at some high E_{IPD} greater than E_{ox}, which allows more electrons to be stored, even when $A_{IPD} / A_{ox} = 1$. When both program and erase are considered, the P/E window is completely lost in small area ratio device for polysilicon FG memory. On the other hand, P/E window of carbon doped polysilicon FG is less sensitive to the area ratio, and provide larger P/E window at small area ratio. Therefore, the use of carbon doped polysilicon floating gate would be advantageous for cells with small IPD area, which is the case for scaled flash memory beyond 40 nm technology node.

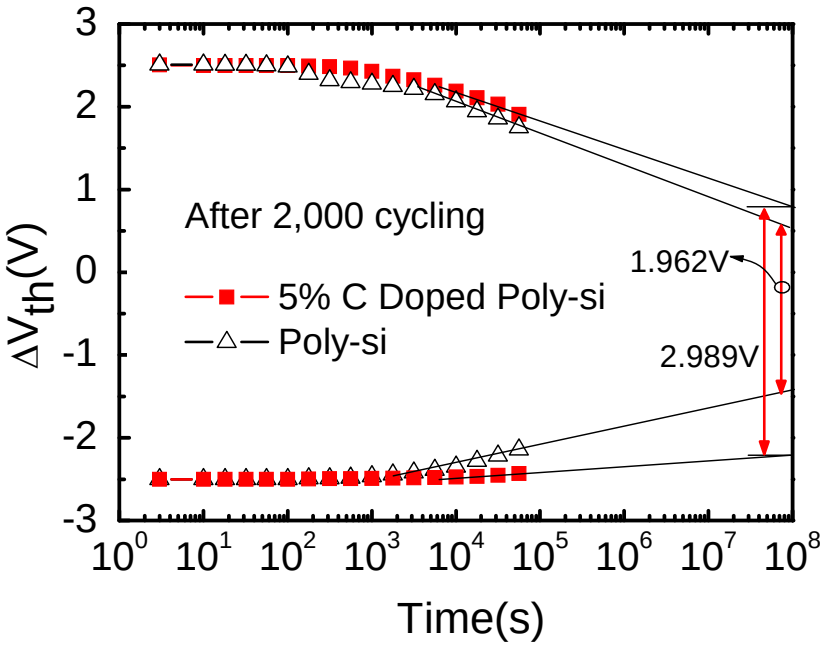


Fig. 2.4.2: Comparison of retention characteristics of 5% carbon doped polysilicon and pure polysilicon FG on SiO_2 IPD after 2,000 P/E cycles. Carbon incorporation into polysilicon FG significantly improves the retention.

Retention characteristic comparison in Fig. 2.4.2 clearly displays the advantage of carbon doped polysilicon FG. The retention measurement was done after 2,000 P/E cycles. The 5 % carbon doped polysilicon has improved the charge retention significantly compared to pure polysilicon FG, which is important for multi-level operation. The bias condition for P/E cycling test was ± 12 V for both polysilicon and carbon doped polysilicon FGs and the pulse duration was adjusted so that they had similar V_{th} after program and erase. As seen in Fig. 2.4.2, after 2,000 P/E cycles, there is 60% charge loss at 10 year retention time for polysilicon FG; however it is only 40% for carbon doped polysilicon FG. The same trend is also found even after 100,000 P/E cycles, as shown in Fig. 2.4.3. This significant improvement in retention time is attributed to the larger

conduction band offset of the carbon doped polysilicon film, resulting in reduced electron tunneling out from the FG during retention state [19, 20].

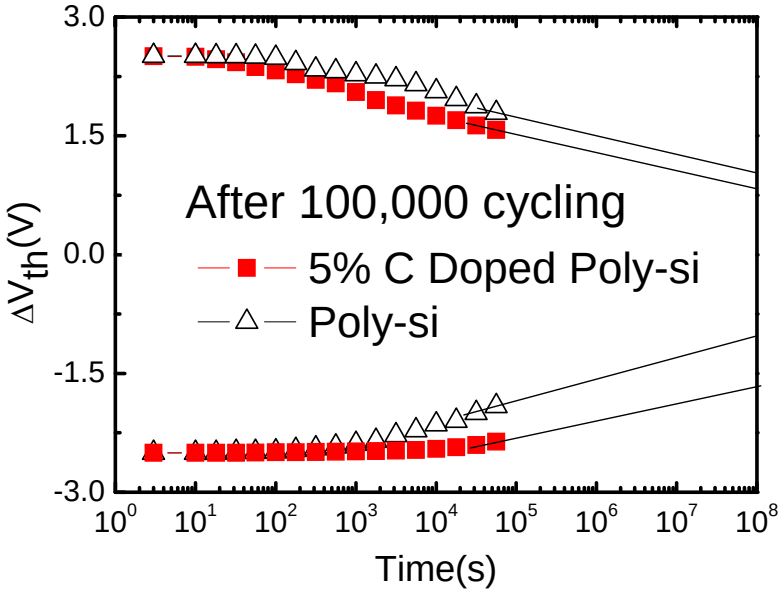


Fig. 2.4.3: Comparison of retention characteristics of 5% carbon doped polysilicon and pure polysilicon FG on SiO_2 IPD after 100,000 P/E cycles.

It is arguable whether 5% carbon could cause such a big difference in the P/E speed and charge retention, since the carbon percentage is low and band structure change is limited. If the carbon is uniformly dispersed in silicon, it is true that 5% carbon in polysilicon will not change the band offset significantly. However, as shown in the FTIR data in Fig. 2.2.3.1, carbon is not uniformly dispersed in polysilicon but segregated and forms silicon carbide phase. In addition, silicon carbide phase formation also will not be uniform within the polysilicon, as carbon prefers to segregate and form clusters at the silicon surface, which has been reported by many other groups [21, 22]. Thus, we can expect a very high concentration of carbon segregated at the FG / SiO_2 interface to cause a local increase of the tunneling barrier, instead of uniformly distributed across the FG.

The improvement in charge retention due to this band structure modification is possible, as local rising of tunneling barrier at the interface is effective in blocking the electrons tunneling out, which could be addressed as the root cause for the superior data retention of 5% carbon doped polysilicon FG.

2. 5 Summary

Carbon doped polysilicon floating gate flash memory is proposed and analyzed in this chapter. The proposed structure exhibits improved P/E window and retention when SiO_2 is used for the IPD layer. The experimental data also suggests that the enhancement is even more significant for small coupling ratio devices, which makes this technique an attractive solution for sub 40 nm technology node. The improvement originates from the modified band structure of the floating gate by incorporating small amount (~ 5%) of carbon. The carbon concentration needs to be well controlled to avoid drastic resistivity increase to decrease the operation speed. Successful formation of silicon carbide phase, which is responsible for the increased band offset of FG, is the critical factor to achieve the performance enhancement. The adoption of carbon doped polysilicon floating gate would introduce minimal process changes to the current floating gate memory technology, and could be considered as a reliable and practical option to ease the urgent need for high-κ IPD layer.

References

[1] G. Hemink, T. Tanaka, T. Endoh, S. Aritome, and R. Shirota, "Fast and accurate programming method for multi-level NAND EEPROMs," *Symp VLSI Tech.,* pp. 129-130, 1995.

[2] "International Technology Roadmap for Semiconductors, 2005 update," http://public.itrs.net.

[3] B. Govoreanu, D. Brunco, and J. Van Houdt, "Scaling down the interpoly dielectric for next generation Flash memory: Challenges and opportunities," *Solid State Electronics,* vol. 49, pp. 1841-1848, 2005.

[4] K. Kim, "Technology for sub-50nm DRAM and NAND flash manufacturing," *IEDM Tech. Dig.,* pp. 323-326, 2005.

[5] E. Cho, C. Lee, T. Kim, S. Sung, B. Cho, C. Lee, H. Cho, Y. Roh, D. Park, and K. Kim, "Hf-silicate inter-poly dielectric technology for sub 70nm body tied FinFET flash memory," *Symp VLSI Tech.,* pp. 208-209, 2005.

[6] J. Lee, S. Hur, and J. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Dev. Lett.,* vol. 23, pp. 264-266, 2002.

[7] N. Mielke and J. Chen, "Reliability of flash nonvolatile memories," *International Journal of High Speed Electronics and Systems,* vol. 11, pp. 719-750, 2000.

[8] G. Molas, D. Deleruyelle, B. De Salvo, G. Ghibaudo, M. Gely, S. Jacob, D. Lafond, and S. Deleonibus, "Impact of few electron phenomena on floating-gate memory reliability," *IEDM Tech. Dig.,* pp. 877-880, 2004.

[9] J. Robertson and B. Falabretti, "Band offsets of high K gate oxides on III-V semiconductors," *J. Appl. Phys.,* vol. 100, pp. 014111 1-8, 2006.

[10] K. Wu, C. Pan, J. Shaw, P. Freiberger, and G. Sery, "A model for EPROM intrinsic charge loss through oxide-nitride-oxide (ONO) interpoly dielectric," *Reliability Physics Symposium,* pp. 145-149, 1990.

[11] J. Kim and J. Choi, "Long-term electron leakage mechanisms through ONO interpoly dielectric in stacked-gate EEPROM cells," *IEEE Trans. Electron Dev.,* vol. 51, pp. 2048-2053, 2004.

[12] K. Lim and C. Ling, "Charge trapping in interpoly ONO film," *1998 IEEE International Conference on Semiconductor Electronics,,* pp. 42-46, 1998.

[13] S. Mori, Y. Araki, M. Sato, H. Meguro, H. Tsunoda, E. Kamiya, K. Yoshikawa, N. Arai, and E. Sakagami, "Thickness scaling limitation factors of ONO interpoly dielectricfor nonvolatile memory devices," *IEEE Trans. Electron Dev.,* vol. 43, pp. 47-53, 1996.

[14] Y. Yamaguchi, E. Sakagami, N. Arai, M. Sato, E. Kamiya, K. Yoshikawa, H. Meguro, H. Tsunoda, and S. Mori, "ONO interpoly dielectric scaling limit for non-volatile memory devices," *Symp VLSI Tech.,* pp. 85-86, 1993.

[15] I. Anteney, G. Parker, P. Ashburn, and H. Kemhadjian, "The role of carbon on the electrical properties of polycrystalline SiC and SiGeC films," *J. Appl. Phys.,* vol. 90, pp. 6182-6189, 2001.

[16] M. Kimura, S. Tam, S. Inoue, and T. Shimoda, "Extraction of trap densities at front and back interfaces in thin-film transistors," *Jpn. J. Appl. Phy.,* vol. 43, pp. 71-76, 2004.

[17] Y. Sun and T. Miyasato, "Infrared absorption properties of nanocrystalline cubic SiC films," *Jpn. J. Appl. Phy. ,* vol. 37, pp. 5485-5489, 1998.

[18] Q. Zhao, B. Wang, J. Li, M. Wang, X. Song, H. Wang, and H. Yan, "Nanocrystalline β–SiC films grown on carbonized Si substrate by Cat-CVD," *Diamond & Related Materials,* vol. 12, pp. 1505 -1509, 2003.

[19] C. Kuo, T. King, and C. Hu, "Bias polarity dependent effects of P+ floating gate EEPROMs," *IEEE Trans. Electron Dev.,* vol. 51, pp. 282-285, 2004.

[20] C. Shen, J. Pu, M. Li, and B. Cho, "P-Type Floating Gate for Retention and P/E Window Improvement of Flash Memory Devices," *IEEE Trans. Electron Dev.,* vol. 54, pp. 1910-1917, 2007.

[21] M. Oehme, M. Bauer, C. Parry, G. Eifler, and E. Kasper, "Carbon segregation in silicon," *Thin Solid Films,* vol. 380, pp. 75-77, 2000.

[22] P. Werner, U. Gosele, H. Gossmann, and D. Jacobson, "Carbon diffusion in silicon," *Appl. Phys. Lett.,* vol. 73, pp. 2465-2467, 1998.

CHAPTER 3

A FEASIBILITY STUDY OF Gd_2O_3 AS BLOCKING OXIDE IN SONOS-TYPE FLASH MEMORY DEVICES

3. 1 Introduction

SONOS (polysilicon - oxide - nitride - oxide - silicon)-type nonvolatile memory devices [1] were proposed to replace floating gate-type flash memory to overcome the scaling issues. The state of art technology for the advanced SONOS memory explores a high-κ blocking layer instead of SiO_2 blocking layer. TANOS-type (TaN - Al_2O_3 - Si_3N_4 - SiO_2 - Si) charge trapping cell has recently shown a good potential [2]. The key element of this concept derived from SONOS is the use of a high-κ dielectric for the top dielectric layer (blocking oxide), which allows the use of a thicker bottom oxide (tunnel oxide) to improve the charge retention without any sacrifice in the program/erase (P/E) speed. The band diagram in Fig. 3.1.1 demonstrates that by using high-κ dielectric for the blocking oxide layer, the electric field across the blocking oxide is proportionally reduced with its dielectric constant. Thus electron injection from the gate during erase can be effectively

suppressed, which will generally enhance the erase speed. It also allows using a physically thicker blocking layer film to minimize the electron leaking out to the gate during retention without any EOT increase. Alternatively, a thicker tunnel oxide could be used if the same P/E speed is preferred. Faster operation speed has been demonstrated by replacing SiO_2 blocking layer with Al_2O_3 [2]. However, relatively low κ value (~ 9) of Al_2O_3 imposes a limitation in further scaling. Materials with higher κ value are intensively studied. The requirements for such high-κ blocking oxides are: high dielectric constant; relatively large bandgap and conduction band offset to Si, which could effectively reduce the tunneling current; good thermal stability such that it would not react with the under layer dielectric.

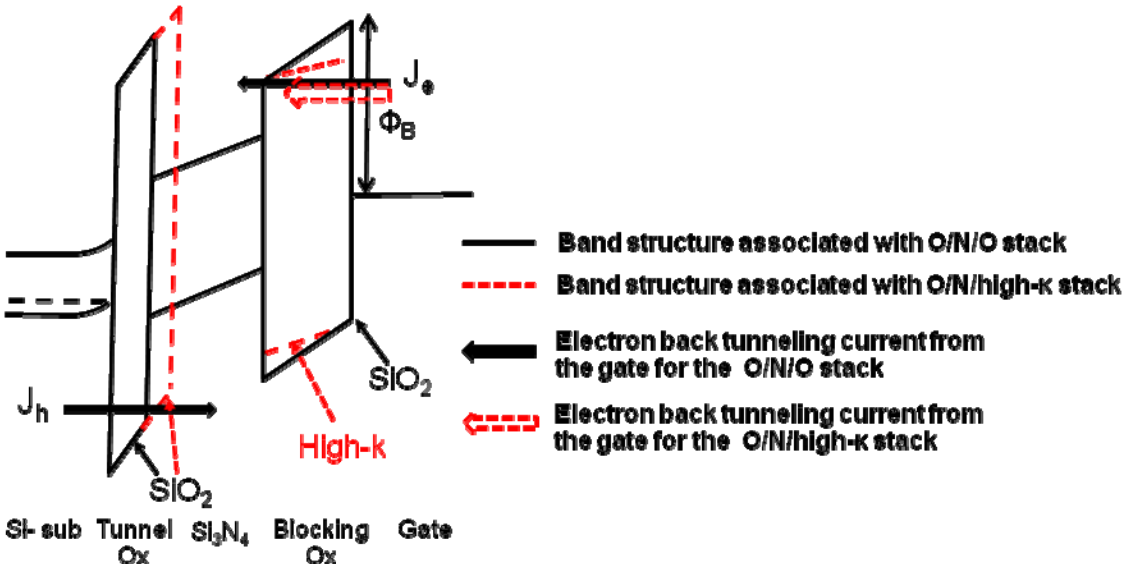


Fig. 3.1.1: Band diagram of a SONOS memory cell during erase case for both TaN / SiO_2 / Si_3N_4 / SiO_2 / Si stack and TaN / high-κ / Si_3N_4 / SiO_2 / Si stack. With the high-κ blocking layer, the back tunneling current from the gate is reduced substantially as indicated by the dotted arrow. A thicker tunnel oxide could be used alternatively.

Table 3.1.1: Basic physical parameters for $LaAlO_3$, Gd_2O_3, La_2O_3, Al_2O_3 and HfO_2. The ΔE_c and ΔE_v value is taking silicon dioxide as the reference.

Material	κ	E_g (eV)	ΔE_c (eV)
$LaAlO_3$	20-27	6.2	1.8
Gd_2O_3	16-18	6.4	2.21
La_2O_3	27-30	5.8	2.3
Al_2O_3	9	8.7	2.8
HfO_2	25	5.7	1.5

Among all the high-κ candidates, rare earth oxides, such as La_2O_3 and Gd_2O_3, are attracting much of the research interest nowadays because of their large energy band gaps of the order of 6 eV [3], which will provide large enough energy barriers for electrons and holes, in addition to relatively large dielectric constants [4]. In this chapter, we will focus on Gd_2O_3 as a potential candidate for the blocking oxide application. Gd_2O_3 has been studied as the gate dielectric for a long time [5]. The basic physical parameters for Gd_2O_3 are listed in Table 3.1.1, together with several other high-κ materials. Gd_2O_3 has a relative high dielectric constant (16-18) [4, 6], a large band gap (5.8-6.4 eV) [7-9], large conduction band offset (2.21 eV) [9] with silicon such that the electron tunneling will be effectively blocked. Comparing to other high-κ candidates listed in Table 3.1.1, Gd_2O_3 offers the best compromise between the dielectric constant and conduction band offset, which always shows a trade-off relationship [10]. Moreover, Gd_2O_3 has very good thermal stability. Thermodynamic calculation suggested that Gd_2O_3 is stable in contact with Si at 1000K [11]. Experimental data also suggests that the Gd_2O_3 has a reasonable

low leakage current with silicon up to 900°C [5]. All these unique properties make it feasible as the blocking oxide. In this chapter, electrical properties of Gd_2O_3 and its feasibility as the blocking layer in SONOS-type flash memory devices will be systematically investigated.

3. 2 Dielectric and Physical Property of Gd_2O_3

3. 2. 1 Deposition Recipe Evaluation

Gd_2O_3 films were deposited by sputtering of using either Gd_2O_3 oxide target in a pure Ar ambient, or Gd metal target in an Ar / O_2 ambient. In both conditions, the chamber pressure is set to be 3 mtorr with 25 sccm Ar flow rate. For PVD deposition, sputtering condition as well as the post deposition annealing conditions (PDA) are the dominant factors for the dielectric quality. In order to study how the oxygen flow rate, the PDA temperature and the sputtering target material (Gd target and Gd_2O_3 target) affect the dielectric property of Gd_2O_3 film, a simple MOS capacitor structure was prepared.

In the MOS capacitor structure, 5 nm SiO_2 was thermally grown on p-type silicon substrate, followed by deposition of 6 nm Si_3N_4 by LPCVD. Then, 16 nm PVD Gd_2O_3 was deposited using two different targets. After the dielectrics were annealed in various PDA temperatures and 150 nm TaN deposition, gate pattern and etch, post metal anneal (PMA) and forming gas anneal (420°C, 10mins, 10 % H_2 + N_2) was performed. The reason to deposit dielectric directly on top of Si_3N_4 instead of bare silicon is to simulate the final SONOS device structure.

Table 3.2.1: Split table of sputtering conditions, PDA and PMA temperatures for Gd_2O_3.

Sputtering Target	Ar	Power	PDA Temp	PMA Temp
Gd_2O_3 (No O_2)	25 sccm	120W (RF)	500°C 100s 5% O_2 600°C 100s 5% O_2	FGA 850°C, Furnace, 30mins
Gd (3sccm O_2)	25 sccm	160W (DC)	500°C 100s 5% O_2 600°C 100s 5% O_2 700°C 60s 5% O_2	FGA 850°C, Furnace, 30mins
Gd (5sccm O_2)	25 sccm	160W (DC)	500°C 100s 5% O_2 600°C 100s 5% O_2 700°C 60s 5% O_2	FGA 850°C, Furnace, 30mins

Table 3.2.1 listed all the split conditions for Gd_2O_3 dielectric quality evaluation. For Gd_2O_3 sputtering using oxide target, Ar flux was fixed at 25 sccm, RF power was fixed at 120 W and the oxygen flow was set to be 0. On the other hand, for Gd_2O_3 sputtered using metal target, the oxygen flow rate during sputtering was varied, from 3 sccm to 5 sccm. The PDA conditions were also varied from 500°C to 700°C, whereas the PMA was done either at forming gas annealing (FGA) condition (420°C) or up to 850°C to simulate the actual S/D activation temperature for the transistor structure.

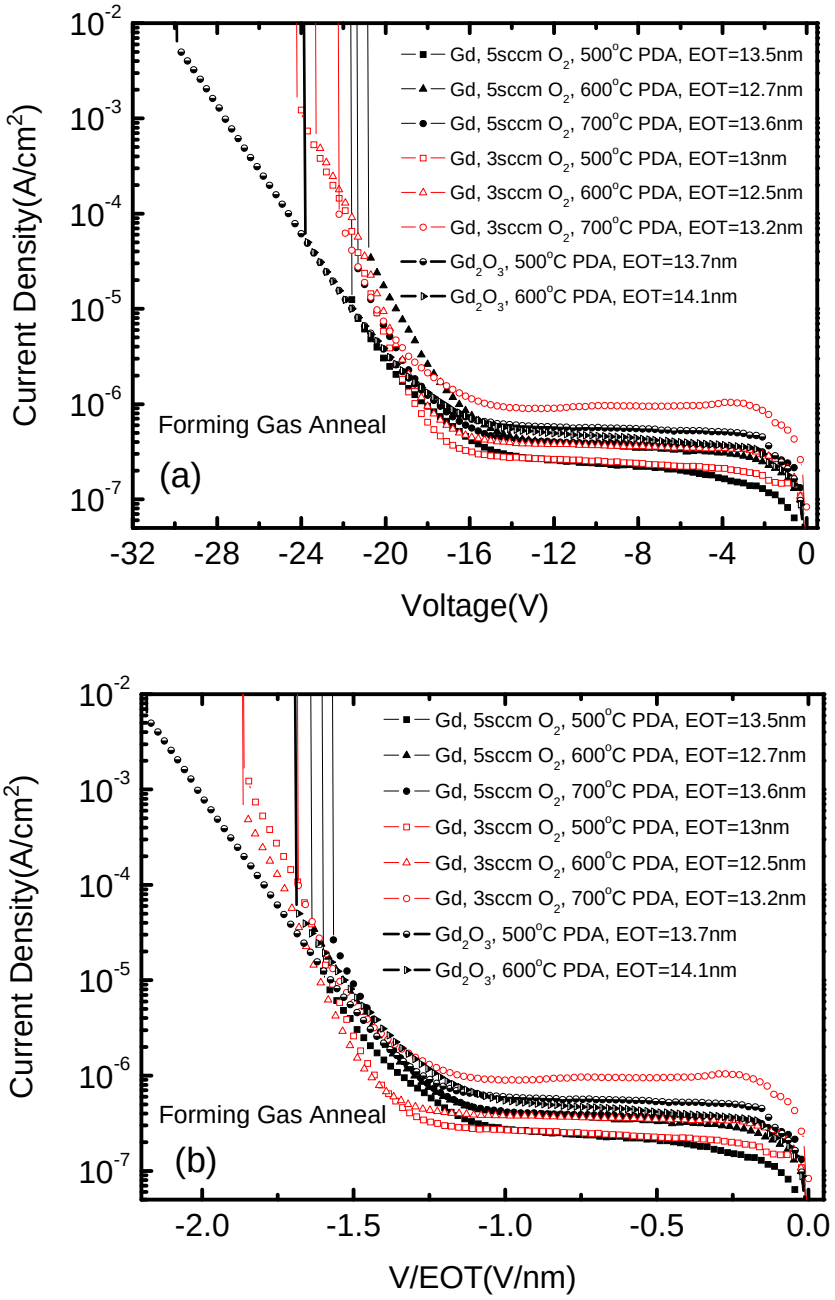


Fig. 3.2.1: Leakage current performance for Gd_2O_3 film sputtered using oxide target and metal target. All the dielectric films are subjected to 500°C, 600°C, 700°C PDA and FGA. Be noted that there is no high temperature process after FGA. (a) Current density vs. voltage curve. (b) Current density normalized to respective EOT.

The electrical characterizations mainly focus on the leakage current performance of Gd_2O_3. Fig. 3.2.1 shows the leakage current of Gd_2O_3 sputtering using either metal target or oxide target after forming gas anneal. The samples were processed with different PDA conditions. Fig. 3.2.1 (a) is the current density vs. voltage curve and Fig. 3.2.1 (b) is normalized with the respective EOT. The figure shows that at low PMA temperature (420^{o}C), Gd_2O_3 sputtered using the oxide target and PDA at 500^{o}C, has the best leakage current performance in terms of its higher breakdown voltage.

For the all the other split conditions, especially for the devices with Gd_2O_3 film sputtered using metal target, they all have rather similar leakage characteristics at both low electrical field regime and high field regime (F-N tunneling regime). Flowing 3 sccm oxygen during the deposition of the Gd_2O_3 film using metal target generally improves the dielectric quality, as the resulting film has a slightly higher breakdown field comparing to Gd_2O_3 film flowing 5 sccm oxygen during sputtering.

Figure 3.2.2 shows the leakage current of Gd_2O_3 sputtered using either metal target or oxide target after high temperature PMA anneal (850^{o}C, 30mins in furnace). The samples were processed with different PDA conditions. Fig. 3.2.2 (a) is the current density vs. voltage curve and Fig. 3.2.2 (b) is normalized with respective EOT. The leakage current characteristics show that at high temperature, the performance of Gd_2O_3 sputtered using oxide target degraded significantly, whereas the Gd_2O_3 film sputtered using metal target has much less performance degradation.

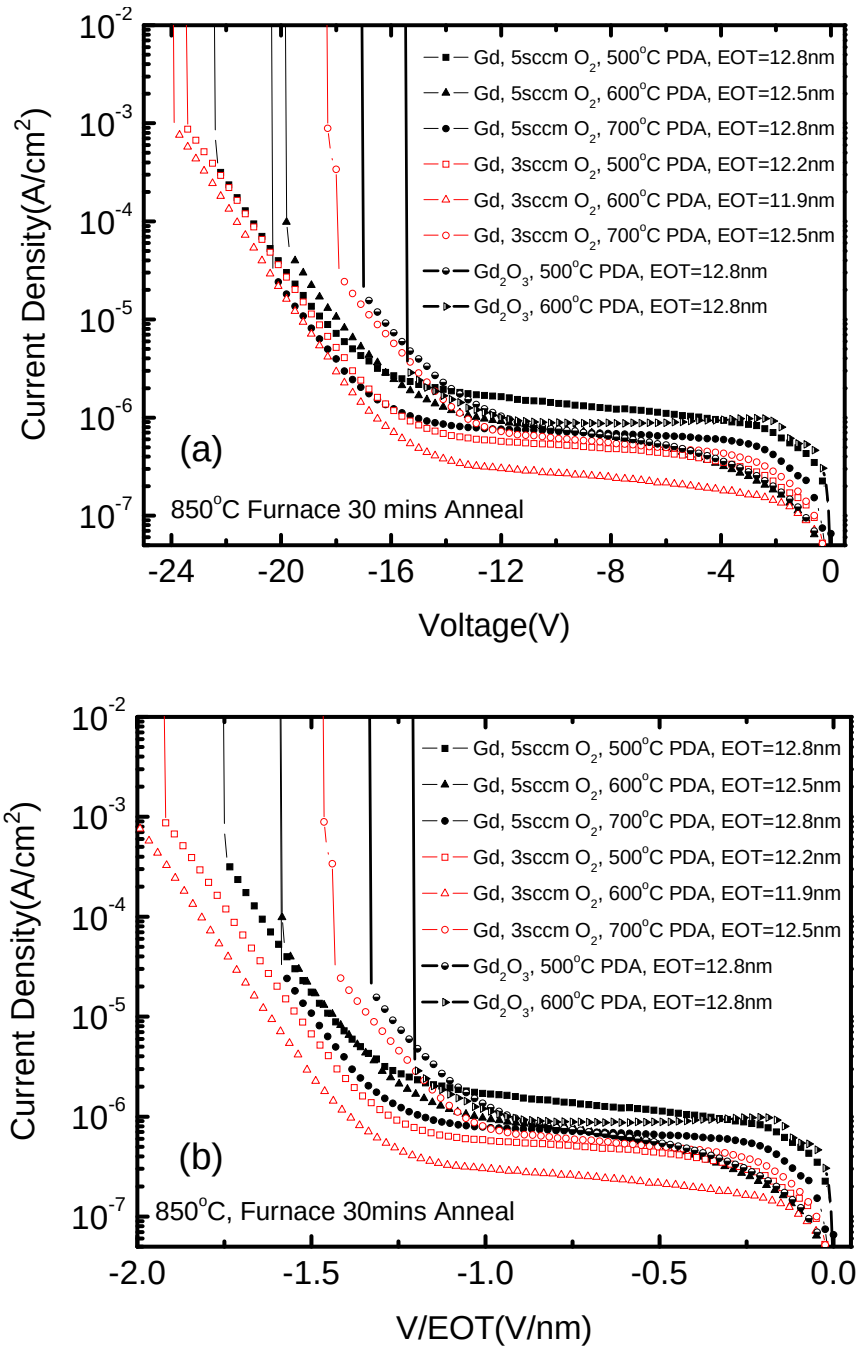


Fig. 3.2.2: Leakage current performance for Gd_2O_3 film sputtered from oxide target and metal target. All the dielectric films are subjected to 500°C, 600°C, 700°C PDA and 850°C PMA. (a) Current density vs. voltage curve. (b) Current density normalized to respective EOT.

Among all the Gd_2O_3 films sputtered using metal target, the film with 3 sccm O_2 flow during the sputtering and subsequently annealed at 600°C slightly suppress the leakage at both low field and high field, which is the optimum process condition for PVD deposited Gd_2O_3 films using the metal target, and hence was set as the process condition for the transistor fabrication.

3. 2. 2 Band Structure Analysis by XPS

High-resolution X-ray photoelectron spectroscopy (XPS) was applied to characterize the electronic structures for Gd_2O_3 film grown on (100) Si substrate. The *ex-situ* XPS measurements were carried out using a VG ESCALAB 220i-XL system equipped with a monochromatized Al K_α source (hv = 51486.6 eV) for the excitation of photoelectrons. Gd 4*d* spectra, O 1*s* core levels spectra, valence band spectra, and O 1*s* energy loss spectra are used to estimate the energy gap (E_g) for Gd_2O_3 , the valence band offset (ΔE_v) and the conduction band offset (ΔE_c) between Gd_2O_3 and (100) Si substrate.

The principle of using high-resolution XPS to obtain both E_g and ΔE_v for the dielectrics was firstly report by Kraut *et al*. [12]. This measurement technique can precisely determine the valence band offset, while minimizing the effect of the interface states. Its accuracy has been practically examined by many other research groups [13].

As described in Kraut*'s* paper, the valence band offset is given by the following equation:

$$\Delta E_v = (E_{cl}^{\ Gd} - E_v^{\ Gd}) - (E_{cl}^{\ Si} - E_v^{\ Si}) - (E_{cl}^{\ Gd}(i) - E_{xl}^{\ Si}(i)) \tag{3.1}$$

The first term in the equation $E_{cl}^{Gd} - E_v^{Gd}$ is the binding energy difference between the Gd core level E_{cl}^{Gd} and valence band maximum E_v^{Gd}. These two values could be extracted from the Gd core level spectra and valence band spectra, which are obtained from 15 nm Gd_2O_3 film prepared using Gd metal target. The film was directly deposited on DHF cleaned Si surface, and subjected to 600°C, 60s, 5 % O_2 PDA anneal. The thickness of the film must be carefully chosen such that the signal from the interface won't affect the accuracy of the measurement. Normally, the detectable range for XPS is around 8 nm. As a result, 15 nm Gd_2O_3 film would be thick enough to block the signal from the interface. Fig. 3.2.3 (a) and (b) shows the Gd *4d* core level spectra and valence band spectra respectively. The resulting E_{cl}^{Gd} is 142.4 eV and E_v^{Gd} is 4.34 eV.

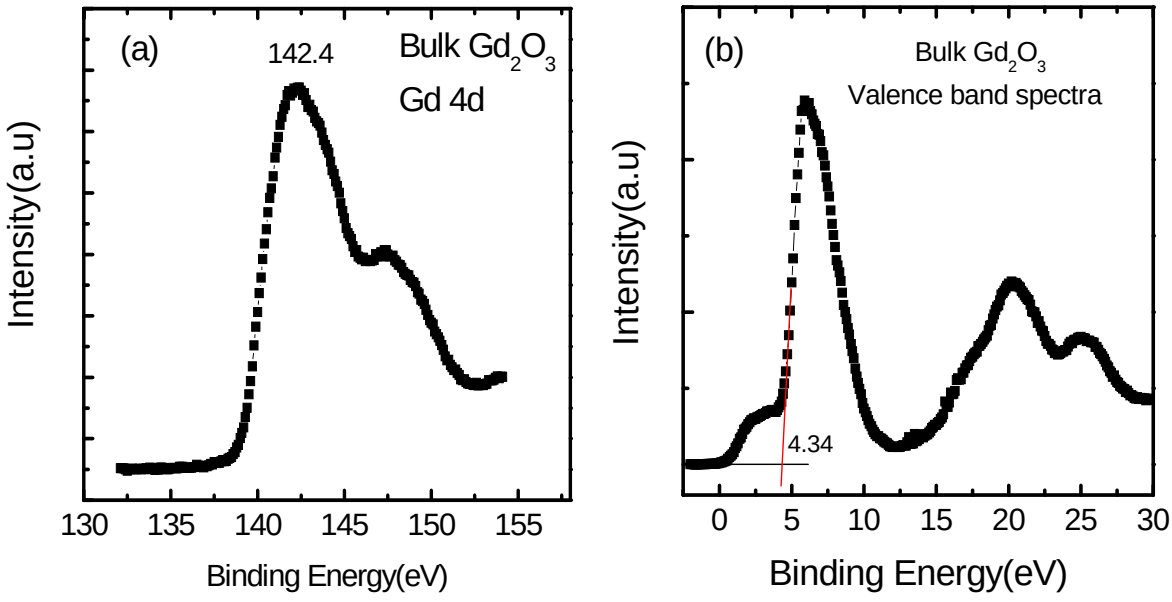


Fig. 3.2.3: Gd *4d* core level and valence band spectra on bulk Gd_2O_3 film (15 nm). (a) Gd *4d* core level spectra. (b) Valence band spectra.

The second term in the equation $E_{cl}^{Si} - E_v^{Si}$ is the binding energy difference between Si core level E_{cl}^{Si} and valence band maximum E_v^{Si}. These two values could be found on Si core level spectra and valence band spectra, which are shown in Fig. 3.2.4 (a) and (b). The silicon wafer was cleaned by SPM ($H_2O_2 + H_2SO_4$) and DHF solutions, and immediately transferred to high vacuum XPS chamber to avoid the native oxide growth. The XPS result gives that E_{cl}^{Si} is 99.2 eV and E_v^{Si} is 0.56 eV, which is quite consistent with the literature reported data [14].

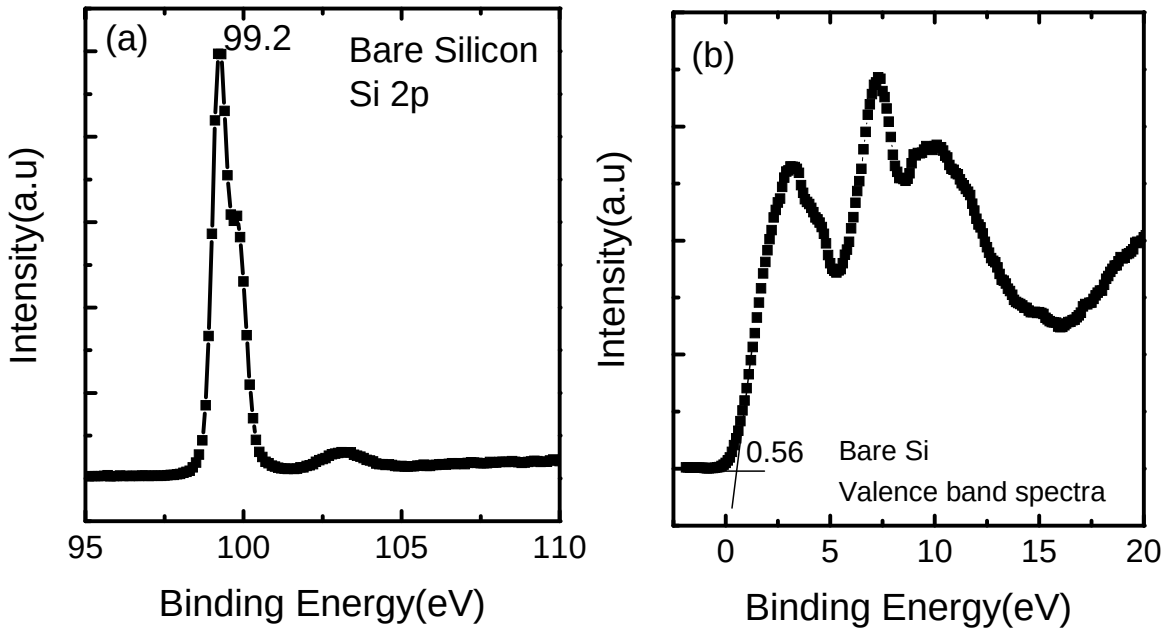


Fig. 3.2.4: Si core level and valence band spectra from bare silicon wafer. (a) Si *2p* core level spectra. (b) Valence band spectra.

The third term in the equation $E_{cl}^{Gd}(i) - E_{xl}^{Si}(i)$ gives the information at the interface. It is calculated by the energy difference between Gd core level and Si core level at the interface. It is measured on thin Gd_2O_3 film (3 nm), such that the signals from

the interface could be correctively detected. The measurement result is given in Fig. 3.2.5. It gives $E_{cl}^{Gd}(i)$ is 141.4 eV and $E_{xl}^{Si}(i)$ is 98.5 eV.

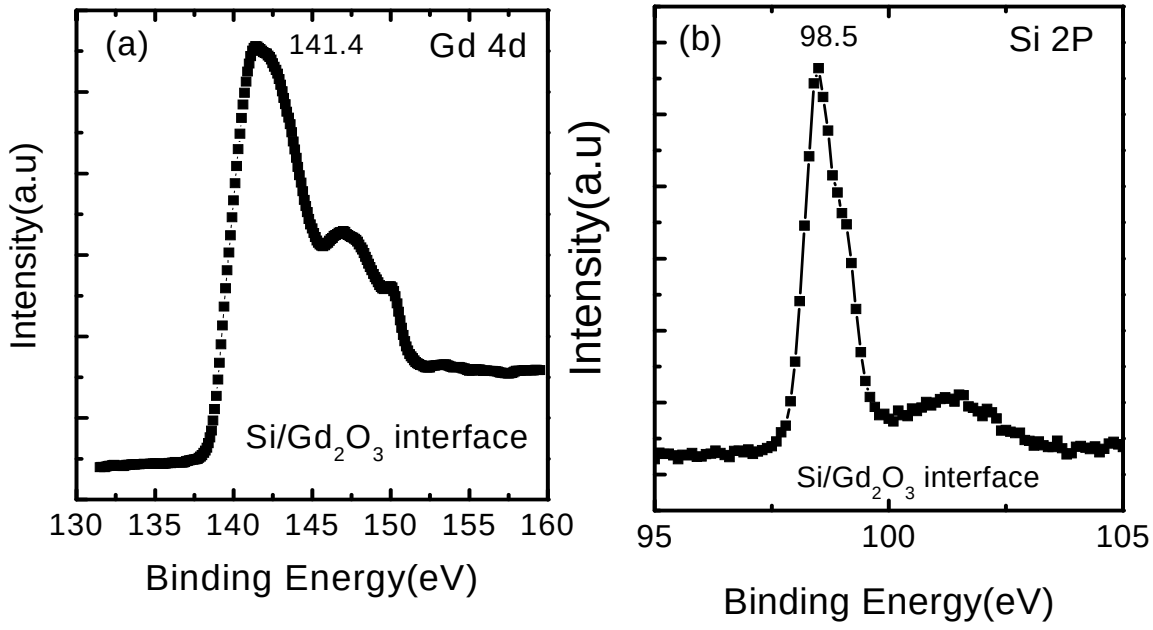


Fig. 3.2.5: Gd core level and Si core spectra from 3 nm Gd2O3 deposited on silicon. (a) Gd 4d core level spectra. (b) Si 2p core level spectra.

After plugging all the terms into the equation, the valence band offset between Gd_2O_3 thin film and silicon substrate is 3.48 eV.

The energy band gap values for the dielectric materials can be determined by the onsets of energy loss from the energy loss spectra [15]. Fig. 3.2.6 (a) gives the O *1s* spectra. Both the Gd - O bond and Gd - OH bond could be clearly identified. This is commonly observed for most of the lanthanide group element, which is due to hydroscopic nature of the lanthanide group oxide itself [16]. Fig. 3.2.6 (b) shows the energy loss spectra derived from the O *1s* spectra. It gives an energy band gap of 6.96 eV.

Knowing the value of valence band offset and energy band gap, the conduction band offset is calculated as $\Delta E_c = E_g - \Delta E_v - E_g(si)$, which is 2.36 eV.

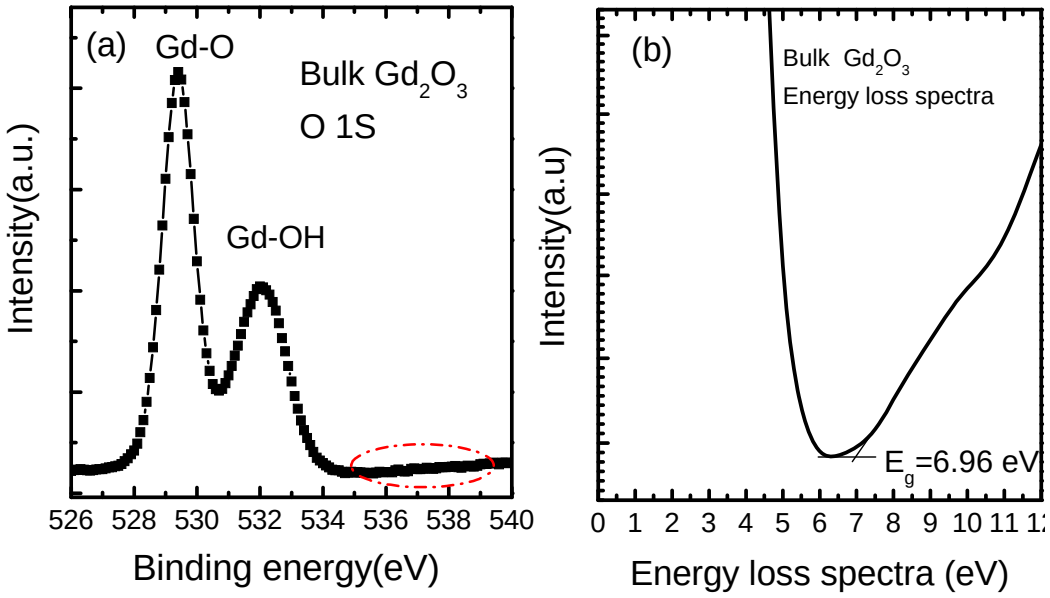


Fig. 3.2.6: (a) O *1s* spectra from bulk Gd_2O_3 film (15 nm). (b) Energy loss spectra. After calculation, the band gap is 6.96 eV.

Table 3.2.2: Comparison of the physical parameters of Gd_2O_3 film measured in this work and other literature reported data. The ΔE_c value is taking SiO_2 as the reference.

E_g (eV)	ΔE_c (eV)	Reference
6.96 ± 0.1	*2.36 ± 0.1*	**This work*
6.5-7.0	------	S. Thakur et al. Optical Materials 27 (2005) 1402–1409 (Optical measurements.) [8]
5.8	2.21	J.Robertson, et al. J. Appl. Phys. 100, (2006) 014111 (CNLs Method) [9]
6.4	3.1	T. Hattori et al. Microelectronic Engineering 72 (2004) 283–287 (RBS and XPS) [7]

The measurement results show that the Gd_2O_3 films prepared using Gd metal target has a band gap of 6.94 eV, valence band offset of 3.48 eV, conduction band offset of 2.36 eV with Si, which is well consistent with other published data for bulk Gd_2O_3 as shown in Table 3.2.2. Comparing to other high-κ materials such as HfO_2 and Al_2O_3, Gd_2O_3 has very suitable properties for the blocking layer in SONOS-type Flash memory devices, considering its large band gap and sufficiently high κ value.

3. 2. 3 Crystal Structure Analysis by XRD

XRD (X-Ray Diffraction) analysis is performed for film morphology study. The measurements were carried out using a GADDS (General Area Detector Diffraction) system equipped with Cu K_α X-Ray (λ = 1.5418 A). Fig. 3.2.7 (a) shows the XRD spectra of Gd_2O_3 film sputtered using Gd_2O_3 oxide target. It shows that the as-deposited Gd_2O_3 film is amorphous when sputtered using the oxide target. Once the film is annealed at a temperature in the range of 500°C - 850°C, it forms a cubic crystal structure. On the other hand, for Gd_2O_3 dielectric films sputtered using Gd metal target, the as-deposited film already has a cubic crystal structure. When the film is annealed, phase transformation from cubic crystal to monoclinic crystal structure is observed once the temperature reaches 500°C to 850°C as shown in Fig. 3.2.7 (b). The two films formed from two different sputter targets show a great difference in electrical properties as well.

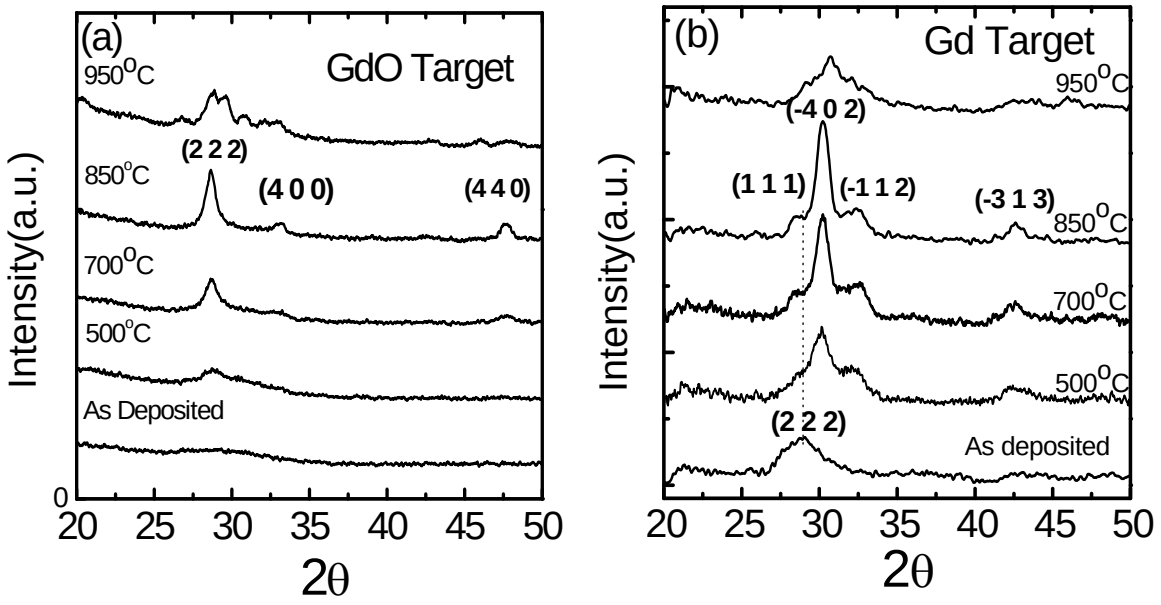


Fig. 3.2.7: XRD spectra of 20 nm Gd_2O_3 film sputtered using (a) Gd_2O_3 oxide target and (b) Gd metal target. The Gd_2O_3 film from Gd_2O_3 oxide target shows a cubic crystal structure while the film from Gd metal target shows a monoclinic crystal structure after high temperature anneal.

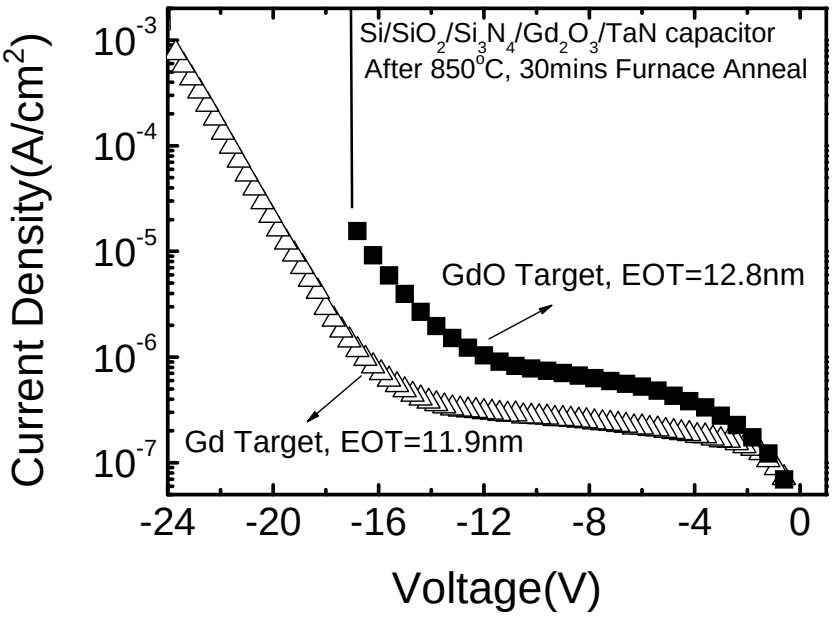


Fig. 3.2.8: Leakage current comparison of the Gd_2O_3 films sputtered using Gd metal target and Gd_2O_3 oxide target. The Gd_2O_3 is deposited on top of Si_3N_4 / SiO_2 / Si stack.

As shown in Fig. 3.2.8, the Gd_2O_3 film sputtered using Gd metal target has one order of magnitude lower in leakage current even though the total EOT of the dielectric stack is thinner. The total EOT value is extracted from the CV measurement of the TaN / Gd_2O_3 / Si_3N_4 / SiO_2 / Si stack after 850^{o}C anneal. The significant difference in electrical properties is believed to be associated with the difference in crystalline structure after high temperature annealing. The difference in the crystal structure will ultimately affect the bonding strength of two neighboring atoms. More importantly, it will affect the atom distribution density significantly. T. S. Böscke *et al*. has also observed the similar phenomenon on HfO_2 based dielectric [17]. The experimental observed data again proves that Gd_2O_3 thin film with monoclinic crystalline structure has better dielectric quality which is favorable for SONOS blocking oxide application.

3. 3 Experiments and Devices Fabrication

SONOS-type flash memory transistors of Gd_2O_3 as the blocking oxide are prepared for the electrical performance evaluation. All the physical parameters and split conditions are listed in Table 3.3.1. The tunnel oxide is 5 nm thick thermally grown SiO_2, and the charge trapping layer is 5.5 nm thick LPCVD Si_3N_4. The blocking oxide layer is 15 nm or 20 nm thick Gd_2O_3 films sputtered using Gd metal target (entire stack EOT is 10.5 nm or 11.5 nm, respectively), and 20 nm thick Gd_2O_3 film sputtered using oxide target (entire stack EOT is 10.7 nm or 11.4 nm, respectively). Control sample with 7.5 nm thick Al_2O_3 blocking layer (entire stack EOT is 11.5 nm) is also prepared for

comparison. 150 nm TaN control gate is deposited by sputtering. After gate stack patterning, S/D implantation, dopant activation (850^{o}C, 30mins at N_2 ambient in furnace) annealing and forming gas annealing (420^{o}C, 10mins) were performed.

In order to evaluate the property of Gd_2O_3 as the blocking layer in SONOS-type nonvolatile memory, the physical thickness of the Gd_2O_3 film was kept to be in the range of 10 - 16 nm, which is the desirable thickness for blocking layer according to the requirement of ITRS Roadmap 2007 [18]. For the Al_2O_3 control sample, it is preferable to keep the same stack EOT as the devices with Gd_2O_3 blocking layer to have a fair comparison for the P/E speed. This is to ensure that the electric field drawn on the tunnel oxide will be the same for all the devices, such that the influence of the different blocking layer material on the P/E speed is becoming easy to be analyzed. This is the reason a rather thinner Al_2O_3 layer thickness is selected to maintain a similar EOT as the Gd_2O_3 blocking layer.

Table 3.3.1: Split conditions for SONOS transistors with Al_2O_3 and Gd_2O_3 blocking layer.

Split	Tunnel Ox (SiO_2)	Charge Trapping (Si_3N_4)	Blocking Ox	Stack EOT
1	5nm	5.6nm	Al_2O_3 7.5nm	11.5nm
2	5nm	5.6nm	Gd_2O_3 15nm (Gd target)	10.5nm
3	5nm	5.6nm	Gd_2O_3 20nm (Gd target)	11.5nm
5	5nm	5.6nm	Gd_2O_3 20nm (Gd_2O_3 target)	11.4nm

3. 4 Memory Characteristic of Gd_2O_3 Blocking Oxide

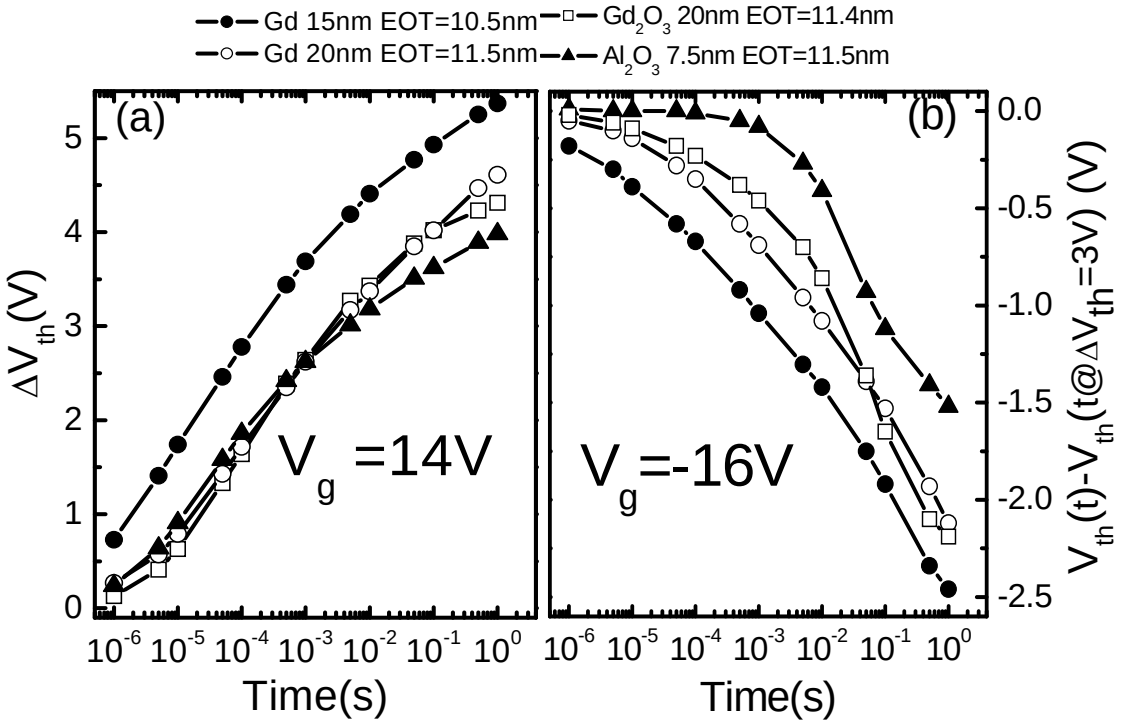


Fig. 3.4.1: (a) Comparison of program speed for Gd_2O_3 and Al_2O_3 blocking layers at program voltage V_g = 14 V. Gd_2O_3 blocking oxide (using both oxide and metal target) shows faster programming. (b) Comparison of erase speed for Gd_2O_3 and Al_2O_3 blocking layer at program voltage V_g = - 16 V. Gd_2O_3 blocking oxide (using both oxide and metal target) shows faster erasing.

The P/E characteristics in Fig. 3.4.1 (a) and (b) show that comparing to Al_2O_3 control sample, Gd_2O_3 blocking oxide layer can provide much faster operation speed at P/E voltage of + 14 V or – 16 V. The P/E voltage is determined to avoid the breakdown of the Al_2O_3 control sample at higher gate voltage. This condition could also be able to fulfill the P/E requirement for 36 nm SONOS flash memory devices as indicated at ITRS road map 2007 [18]. At the program voltage V_g = 14 V, and program time t = 100 ms,

comparing to the Al_2O_3 control sample, 11% and 36% faster program speed (in terms of V_{th} change) is observed for Gd_2O_3 blocking layer samples, sputtered using Gd metal target, with EOT of 11.5 nm and 10.5 nm respectively (Fig. 3.4.1 (a)). Similar enhancement in P/E speed is also found for the samples sputtered using Gd_2O_3 oxide target.

The erase characteristic is always one of the most important criteria for blocking oxide material selection in SONOS memory. If the blocking layer is either poor in dielectric quality or having insufficient electron tunneling barrier, there will be a large amount of back tunneling electrons from the gate during erase, which will slow down the erase speed. However, as shown in Fig. 3.4.1 (b), at the erase time t = 100 ms, the enhancement in the erase speed is as large as 37% and 71% for Gd_2O_3 film sputtered using Gd metal target. Similar enhancement in erase speed is also found for the samples sputtered using Gd_2O_3 oxide target with EOT of 11.4 nm. Such a significant improvement of erase characteristics is attributed to the reduction of electron tunneling current from the gate through the blocking layer under the negative bias condition.

The charge retention characteristic of the devices is shown in Fig. 3.4.2. There is only 6 % charge loss in 10 years retention time for the sample with Gd_2O_3 blocking layer sputtered using Gd metal target, which is almost the same with Al_2O_3 control sample before cycling. After 1,000 P/E cycles, its retention performance degraded a little bit, however still comparable with Al_2O_3 control sample (Fig. 3.4.3). The cycling condition for Fig. 3.4.3 is P/E voltage equals to + 14 V and – 16 V. The cycling time is adjusted to have a similar P/E window for all the devices. The Gd_2O_3 sample, sputtered from Gd metal target with the EOT of 10.5 nm, has a 36% and 57% faster in P/E speed, while

having comparable retention performance with Al_2O_3. Such fast P/E speed also gives a margin to increase the tunnel oxide thickness for better retention if the same P/E speed is required. It makes Gd_2O_3 an attractive candidate for the blocking layer in SONOS-type nonvolatile memory devices. The Gd_2O_3 film sputtered from Gd_2O_3 oxide target shows degradation in retention for both fresh devices and cycled devices. It might be attributed to the higher leakage current as shown in Fig. 3.2.3.2 due to different crystalline structure of the Gd_2O_3.

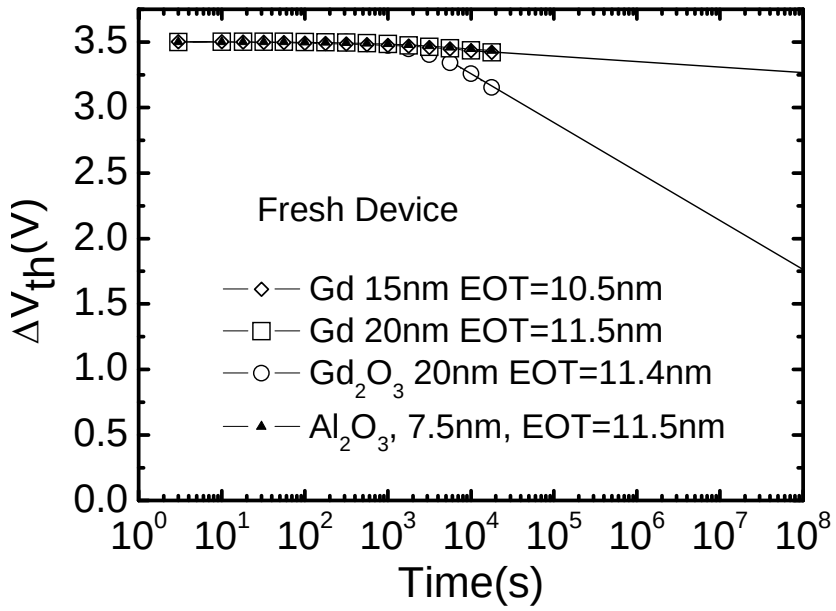


Fig. 3.4.2: Retention characteristics of the memory cells with Gd_2O_3 and Al_2O_3 blocking layer. The devices with Gd_2O_3 sputtered using metal target show almost the same charge retention performance as Al_2O_3 control sample. The devices with Gd_2O_3 sputtered using oxide target show a degraded retention property.

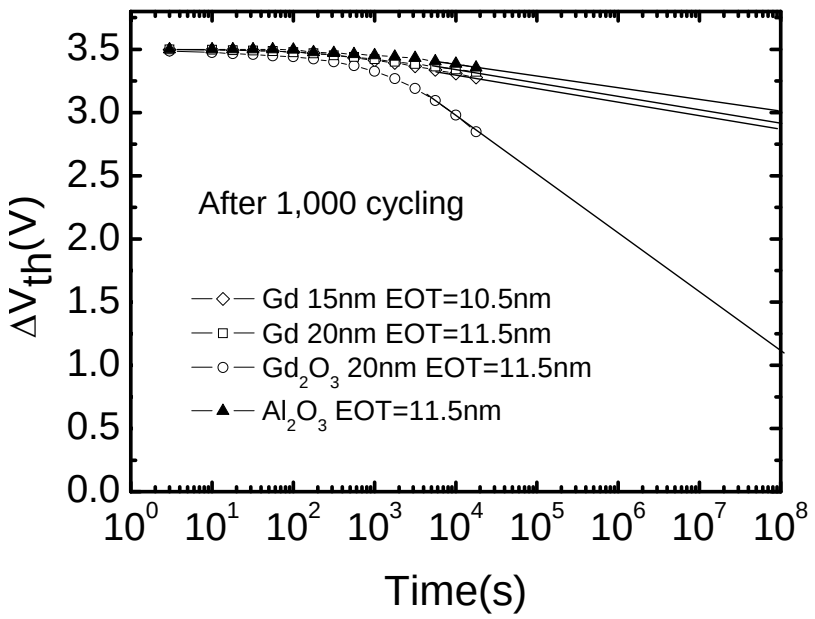


Fig. 3.4.3: Retention characteristics of the memory cells with Gd_2O_3 and Al_2O_3 blocking layer after 1,000 cycles. The devices with Gd_2O_3 sputtered using metal target show slightly degraded retention comparing to Al_2O_3 control sample. The devices with Gd_2O_3 sputtered using oxide target show the worst retention property.

3. 5 Summary

In this report, SONOS-type flash memory devices with Gd_2O_3 blocking layer is demonstrated. The monoclinic structured Gd_2O_3 blocking layer shows a faster P/E speed than the conventional Al_2O_3 blocking layer. The high operation speed, together with satisfactory charge retention characteristics are attractive advantages for flash memory application. The electronic structure of Gd_2O_3 film is measured by XPS for the first time, which has a great scientific value as well as practical application. It is also found that control of the crystal structure is the key factor to get high quality film. The monoclinic crystal structure is favorable for SONOS application because of its low leakage current.

Since the Gd_2O_3 shows promising result as the blocking oxide in SONOS-type nonvolatile memory devices, the suggestion for future work would be engineering the pure Gd_2O_3 film such that its retention performance could be improved further to over come Al_2O_3 after P/E cycles.

References

[1] F. Libsch and M. White, "Nonvolatile semiconductor memory technology," *SONOS Nonvolatile Semiconductor Memories,* pp. 309-357, 1998.

[2] Y. Shin, J. Choi, C. Kang, C. Lee, K. Park, J. Lee, J. Sel, V. Kim, B. Choi, and J. Sim, "A novel NAND-type MONOS memory using 63nm process technology for multi-gigabit flash EEPROMs," *IEDM Tech. Dig.,* pp. 327-330, 2005.

[3] R. Fujitsuka, M. Sakashita, A. Sakai, M. Ogawa, S. Zaima, and Y. Yasuda, "Thermal Stability and Electrical Properties of $(La_2O_3)_{1-x}(Al_2O_3)_x$ Composite Films," *Jpn. J. Appl. Phys.,* vol. 44, pp. 2428-2432, 2005.

[4] R. Shannon, "Dielectric polarizabilities in oxides and fluorides," *J. Appl. Phys.,* vol. 73, pp. 348-366, 1993.

[5] T. Pan, C. Liao, H. Hsu, C. Chen, K. Wang, and J. Wang, "Excellent frequency dispersion of thin gadolinium oxide high-k gate dielectrics," *Appl. Phys. Lett.,* vol. 87, p. 262908, 2005.

[6] J. Gupta, D. Landheer, G. Sproule, J. McCaffrey, M. Graham, K. Yang, Z. Lu, and W. Lennard, "Interfacial layer formation in Gd_2O_3 films deposited directly on Si (0 0 1)," *Appl. Surface Sci.,* vol. 173, pp. 318-326, 2001.

[7] T. Hattori, T. Yoshida, T. Shiraishi, K. Takahashi, H. Nohira, S. Joumori, K. Nakajima, M. Suzuki, K. Kimura, and I. Kashiwagi, "Composition, chemical structure, and electronic band structure of rare earth oxide/Si (100) interfacial transition layer," *Microelectronic Engineering,* vol. 72, pp. 283-287, 2004.

[8] S. Thakur, N. Sahoo, M. Senthilkumar, and R. Tokas, "Optical properties and morphological changes in gadolinia films deposited under ambient substrate temperature conditions," *Optical Materials,* vol. 27, pp. 1402-1409, 2005.

[9] J. Robertson and B. Falabretti, "Band offsets of high K gate oxides on III-V semiconductors," *J. Appl. Phys.,* vol. 100, p. 014111, 2006.

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

[11] X. Wang, M. Li, A. Chin, C. Zhu, J. Shao, W. Lu, X. Shen, X. Yu, R. Chi, and C. Shen, "Physical and electrical characteristics of high-k gate dielectric $Hf_{(1-x)}La_xO_y$," *Solid State Electronics,* vol. 50, pp. 986-991, 2006.

[12] E. Kraut, R. Grant, J. Waldrop, and S. Kowalczyk, "Precise determination of the valence-band edge in x-ray photoemission spectra: Application to measurement of semiconductor interface potentials," *Phys. Review Lett.,* vol. 44, pp. 1620-1623, 1980.

[13] Q. Li, S. Wang, K. Li, A. Huan, J. Chai, J. Pan, and C. Ong, "Photoemission study of energy-band alignment for $RuO_x/HfO_2/Si$ system," *Appl. Phys. Lett.,* vol. 85, pp. 6155-6157, 2004.

[14] A. Tuan, T. Kaspar, T. Droubay, J. Rogers Jr, and S. Chambers, "Band offsets for the epitaxial TiO/SrTiO/Si (001) system," *Appl. Phys. Lett.,* vol. 83, pp. 3734-3736, 2003.

[15] F. Bell and L. Ley, "Photoemission study of $SiO_x(0<x<2)$ alloys," *Phys. Rev. B,* vol. 37, pp. 8383-8393, 1988.

[16] S. Jeon and H. Hwang, "Effect of hygroscopic nature on the electrical characteristics of lanthanide oxides (Pr_2O_3, Sm_2O_3, Gd_2O_3, and Dy_2O_3)," *J. Appl. Phys,* vol. 93, pp. 6393-6395, 2003.

[17] T. Boscke, S. Govindarajan, C. Fachmann, J. Heitmann, A. Avellan, U. Schroder, S. Kudelka, P. Kirsch, C. Krug, and P. Hung, "Tetragonal Phase Stabilization by Doping as an Enabler of Thermally Stable HfO_2 based MIM and MIS Capacitors for sub 50nm Deep Trench DRAM," *IEDM Tech. Dig.,* pp. 1-4, 2006.

[18] "International Technology Roadmap for Semiconductors, 2007 update,," http://public.itrs.net.

CHAPTER 4

ALUMINUM DOPED GD_2O_3 BLOCKING LAYER FOR IMPROVED CHARGE RETENTION IN SONOS-TYPE FLASH MEMORY DEVICES

4. 1 Introduction

In Chapter 3, the feasibility of using pure Gd_2O_3 as the blocking layer in SONOS memory has been investigated. Gd_2O_3 shows good potential as the blocking oxide, since it has almost the same charge retention as the Al_2O_3 blocking layer, but faster program/erase (P/E) speed [1]. However, after 1,000 P/E cycles, its retention performance degrades. The degradation in the charge retention after cycling limits its practical implementation. Further improvement on the retention property, especially after P/E cycles is thus necessary in order to fulfill the requirement for the next generation flash memory devices. In this chapter, engineering of Gd_2O_3 based blocking oxide will be discussed. The discussion will focus on aluminum doped Gd_2O_3 film, which has experimentally demonstrated improved data retention characteristic.

4. 1. 1 Discussion on Charge Leakage Mechanism

The charge loss mechanism must be fully understood first in order to improve the charge retention. The major charge leakage paths for a SONOS memory cell have been illustrated in Fig. 4.1.1. The decay of the charge stored in the nitride layer at room temperature has been modeled by a number of researchers [2-8]. These models involve back tunneling of charge through the tunnel oxide and blocking oxide. An internal field due to trapped charges in the tunnel oxide and blocking oxide also enhances the process of tunneling.

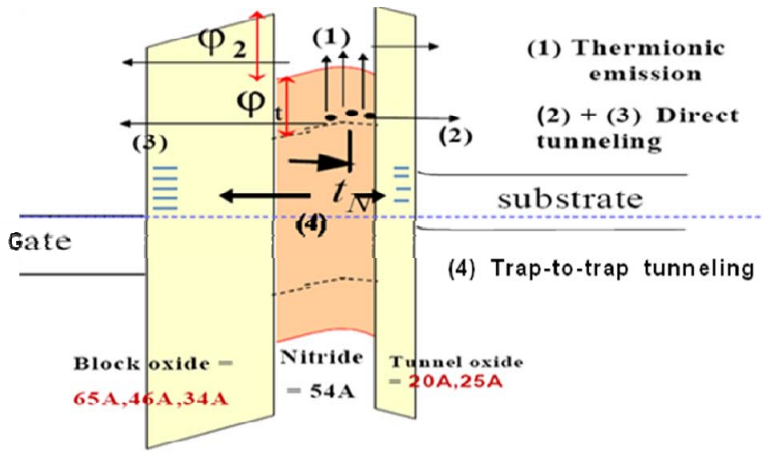


Fig. 4.1.1: The charge loss path and the three dominant leakage components in a SONOS cell. (1): thermionic emission (2) + (3): direct tunneling. (4): trap-to-trap tunneling [9].

The characterization of charge trapped in the nitride at elevated temperatures has also been investigated for SONOS devices [10-12]. Recently, with the advance of scaled SONOS devices, the temperature dependence has been investigated with an amphoteric trap model [13]. The results indicate that the trapped electrons in the nitride layer are

thermally excited at elevated temperatures and tunnel through the blocking oxide and tunneling oxide [14-16].

Hence, tunneling and thermal emission are the two major charge loss mechanisms as discussed in many references [9, 17, 18]. It might be arguable if the trapped charges may leak via other leakage paths such as trap-to-trap tunneling. However, since the trap-to-trap tunneling component will only become important after sufficient number of P/E cycles, it can be neglected for a fresh cell measurement. Therefore, the thermal emission barrier (Φ_B), which is determined by the conduction band offset of the material itself, will be the dominant factor to determine the charge retention, especially at elevated temperature [19].

Based on the discussion above, one effective technique to improve the charge retention of the Gd_2O_3 blocking oxide is to increase its conduction band offset to reduce the thermionic emission current, which could be achieved by doping of other elements into Gd_2O_3. The second stage will be the optimization of the blocking layer thickness to minimize electron leaking out through tunneling, which will be discussed in Chapter 5.

4. 1. 2 Motivation for Doping Al into Gd_2O_3 Dielectric

High-resolution X-ray Photoelectron Spectroscopy (XPS) was applied to characterize the electronic structures for Gd_2O_3 film grown on (100) Si substrate. The measurement results show that the Gd_2O_3 films has a band gap of 6.94 eV, valence band offset of 3.48 eV, conduction band offset of 2.36 eV with Si [1]. Since Al_2O_3 has a conduction band offset of 2.8 eV [20] which is larger than that of Gd_2O_3, incorporation of

Al into Gd_2O_3 film could yield a dielectric film with the conduction band offset to be in between 2.36 eV and 2.8 eV. And it is an effective method to increase the conduction band offset and to improve the charge retention.

Moreover, doping Al could also be able to improve the hydroscopic nature of Gd_2O_3, which is one of the root causes for the reliability degradation. The Al incorporated Gd_2O_3 film is also expected to have enhanced thermal stability as Al_2O_3 is a well-known dielectric material which is thermally stable above 1000^{o}C with silicon [20].

Al incorporated Gd_2O_3 film is capable of increasing the bandgap and conduction band offset as well as improving the thermal and chemical stability of the material. Its application for blocking oxide in SONOS-type nonvolatile memory devices is worthy to be investigated. However, the amount of Al to be incorporated should be carefully determined, as there is a trade-off relationship between the conduction band offset and the κ value. To determine the amount of Al incorporated is to balance these two factors and find an optimum Al percentage to have the best device performance.

4. 2 Deposition Technique and Film Property Evaluation

4. 2. 1 PVD Sputtering Recipe Study

Figure 4.2.1 shows the phase diagram of the $GdO_{3/2}$-$AlO_{3/2}$ system. As indicated in this phase diagram, once the Al mole fraction is less than 35%, the compound oxide will form a GAM structure, which has a monoclinic crystal structure and the chemical formula of $Gd_4Al_2O_9$. If the Al mole fraction increased to 50%, the GAP structure will be formed instead. The GAP crystal has a perovskite-like structure and the chemical formula

of $GdAlO_3$. As a single phase crystal is always preferred for better electrical property, the Al percentage should be carefully controlled to be either less than or equal to 35%, or higher than 50%.

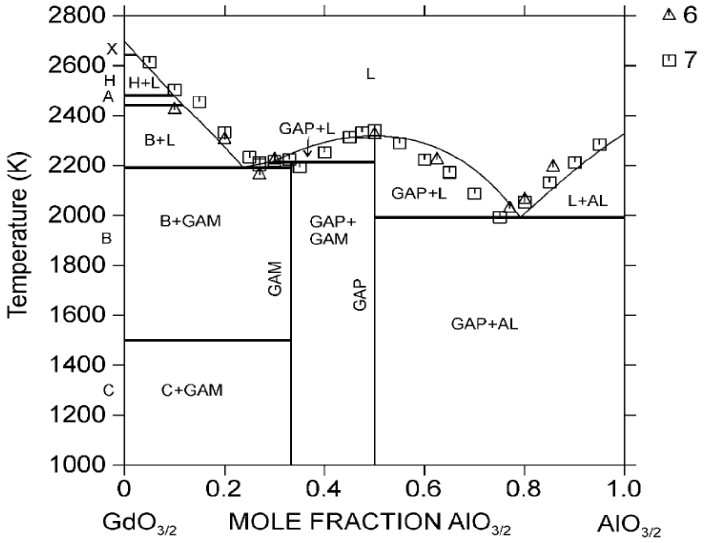


Fig. 4.2.1: Phase diagram of $GdO_{3/2}$-$AlO_{3/2}$ system [21].

The Al doped Gd_2O_3 film is deposited by co-sputtering of Al and Gd target in O_2 and Ar ambient. The Gd sputtering is fixed at 100 W, while the Al sputtering power is varied to get different amount of Al incorporation. Table 4.2.1 lists the sputtering power, Ar flow rate, O_2 flow rate, chamber pressure as well as the Al concentration in the deposited film. The Al percentage is studied by XPS. In order to control the Al percentage to be either below 35% or above 50%, 22%, 35% and 75% Al has been incorporated into the $GdAlO_x$ system respectively.

Table 4.2.1: Sputtering condition and Al concentration for $GdAlO_x$ deposition.

Split	Al Power (W)	Gd Power (W)	Ar (sccm)	O_2(sccm)	Chamber Pressure (mTorr)	Al Conc.
1	155w	100w	25	3	3	22%
2	185w	100w	25	3	3	35%
3	210w	100w	25	3	3	75%

4. 2. 2 Leakage Current Evaluation on MOS Capacitor

The dielectric quality of the PVD deposited Al doped Gd_2O_3 is again studied on TaN / $GdAlO_x$ / Si_3N_4 / SiO_2 / Si MOS capacitor. The capacitor fabrication process is similar as what is described in chapter 3. The bottom two layers were 5 nm thick thermally grown SiO_2 and 5.5 nm thick LPCVD Si_3N_4. The Al doped Gd_2O_3 film with various Al incorporation rate were deposited on top of the Si_3N_4 layer following the sputtering conditions listed in Table 4.2.2. The two different PDA conditions for the dielectric were 500^{o}C for 60s or 600^{o}C for 60s, both in 5% O_2 ambient. A control sample with pure Gd_2O_3 film was also prepared. 1500 nm TaN gate was capped by the sputtering. After gate mask and plasma etch, all the capacitors were annealed in furnace at 850^{o}C for 30 mins to simulate the S/D activation temperature. Table 4.2.2.1 summaries the physical thickness, the entire gate stack EOT as well as the calculated κ value for both Gd_2O_3 and Al doped Gd_2O_3 film. A linearly decreasing κ value with the increase of Al incorporation rate is observed, which is due to the lower κ value of Al_2O_3. For the dielectrics with 22%

Al and 35% Al incorporation, the decrease in the κ is only about 10% and 15%, respectively, comparing to the pure Gd_2O_3 film.

Table 4.2.2: Summarized physical thickness, stack EOT as well as the calculated κ value for the capacitors with Gd_2O_3, and $GdAlO_x$ dielectric

Split	BL layer	Physcial Thickness(nm)	Stack EOT(nm)	Calculated κ Value
1	Gd_2O_3	15.5	11.6	20
2	22% Al-$GdAlO_x$	14.8	11.2	18
3	35% Al-$GdAlO_x$	12.3	10.9	17
4	75% Al-$GdAlO_x$	16	12.7	13

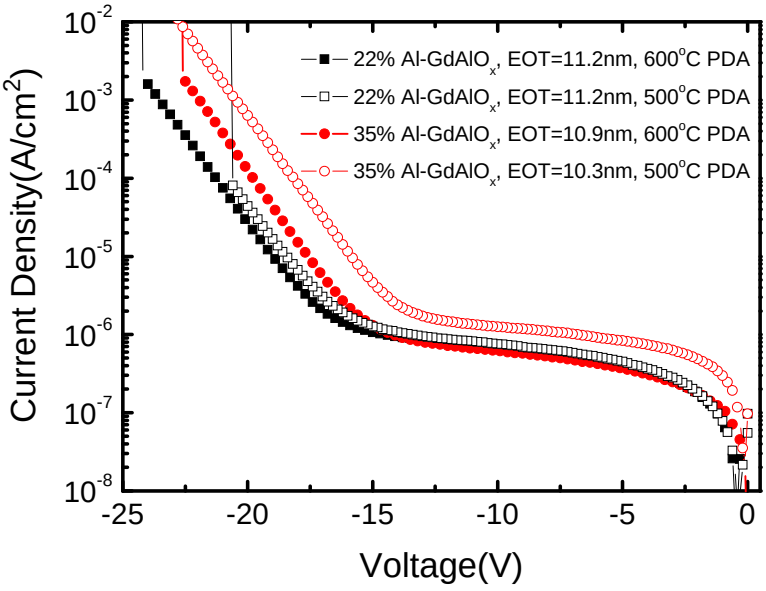


Fig. 4.2.2: The Current density vs. voltage characteristic for $GdAlO_x$ film with 22% and 35% Al incorporation. The dielectric films are subjected to different PDA temperatures (500°C and 600°C). The PMA condition is 850°C, 30 mins in furnace.

The I-V characteristic in Fig. 4.2.2 shows that different PDA temperature will slightly affect the film quality. Comparing the leakage current of $GdAlO_x$ film with 22% and 35% Al incorporation, 600°C 60s PDA with 5% O_2 flow gives a lower leakage current. This PDA condition is set as the standard process condition for the transistor fabrication.

Figure 4.2.3 compares the leakage current performance of pure Gd_2O_3 film with Al incorporated film. Fig. 4.2.3 (a) is the current density vs. voltage curve and Fig. 4.2.3 (b) is normalized to respective EOT. The Al incorporation percentage is 22%, 35% and 75% respectively. It shows that small amount of Al incorporation (22% and 35%) will significantly decrease the leakage current comparing to the pure Gd_2O_3 film. As the Al percentage increases from 22% to 35%, and further up to 75%, the leakage current increases with the increasing Al percentage (Fig. 4.2.3 (b)). The performance of 75% Al – $GdAlO_x$ film is slighter worse than that of pure Gd_2O_3 after normalized to the electrical field as shown in Fig. 4.2.3 (b). This could be explained in the following manner. A large amount of Al incorporation would decrease the dielectric constant of the film which is shown Table 4.2.2. The electric field across the layer is proportionally increased with the decrease of the κ value. Thus the enhanced leakage current will be observed. This is one of the motivations to use the high dielectric constant blocking layer for reduced leakage current.

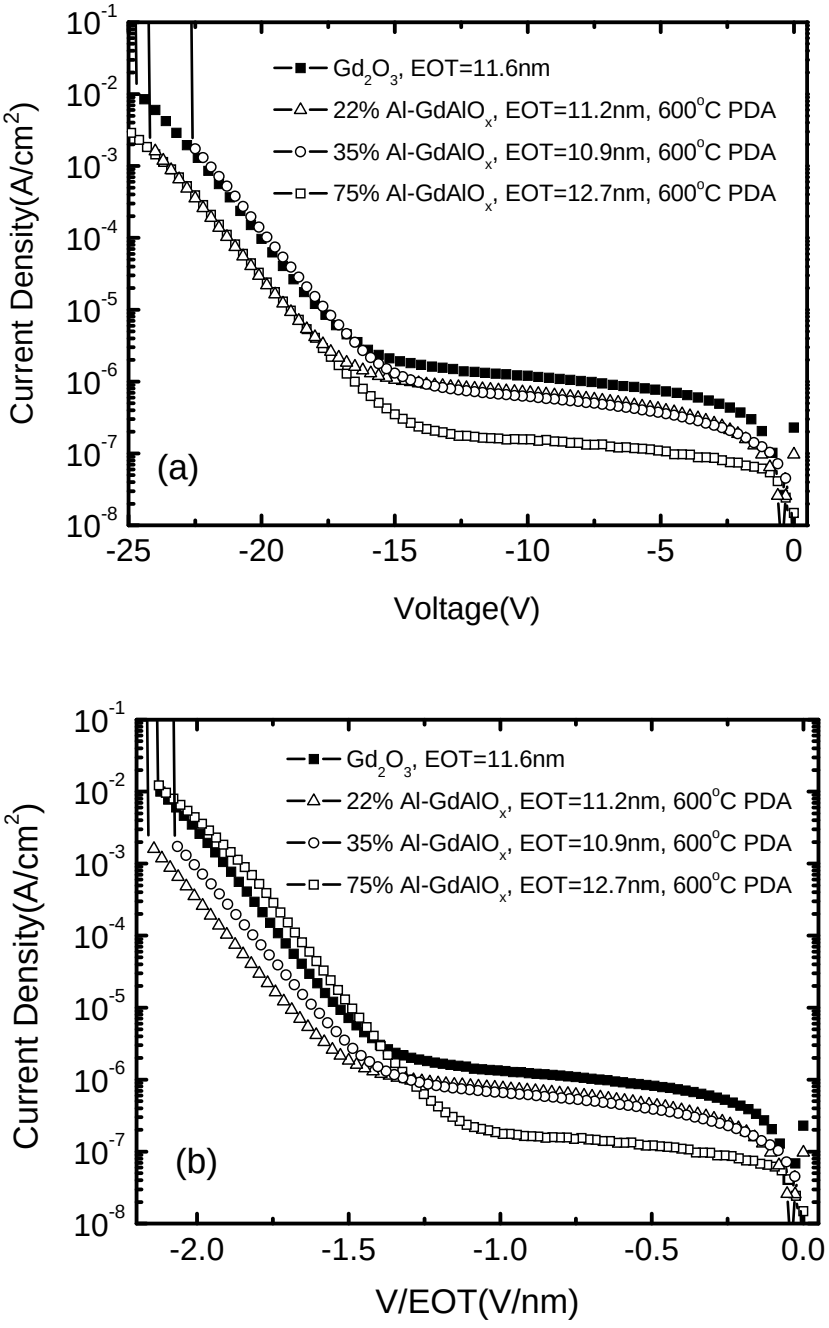


Fig. 4.2.3: Leakage current comparison for pure Gd_2O_3 film and $GdAlO_x$ film with 22%, 35%, and 75% Al incorporation. The PDA temperature is 600°C. The PMA condition is 850°C, 30 mins in furnace. (a) Current density vs. voltage curve. (b) Current density normalized to respective EOT.

4. 2. 3 Composition Analysis by XPS

The *ex-situ* XPS measurements were carried out using a VG ESCALAB 220i-XL system, equipped with a monochromatized Al *Kα* source (*hν* = 1486.6 eV) for the excitation of photoelectrons. All of the high-resolution scans were taken at a photoelectron take-off angle of 90° and a pass energy of 20 keV. The thickness for the all the samples were around 15 nm and the Al percentage is 22%, 35% and 75% respectively. The XPS measurement was carried out immediately after 600°C, 60s PDA anneal in 5% O_2 ambient. All the spectra are calibrated against the C *1s* peak (285.0 eV) of adventitious carbon. The intensity for O *1s* spectra shown in Fig. 4.2.4 (b) here has been normalized for easy comparison.

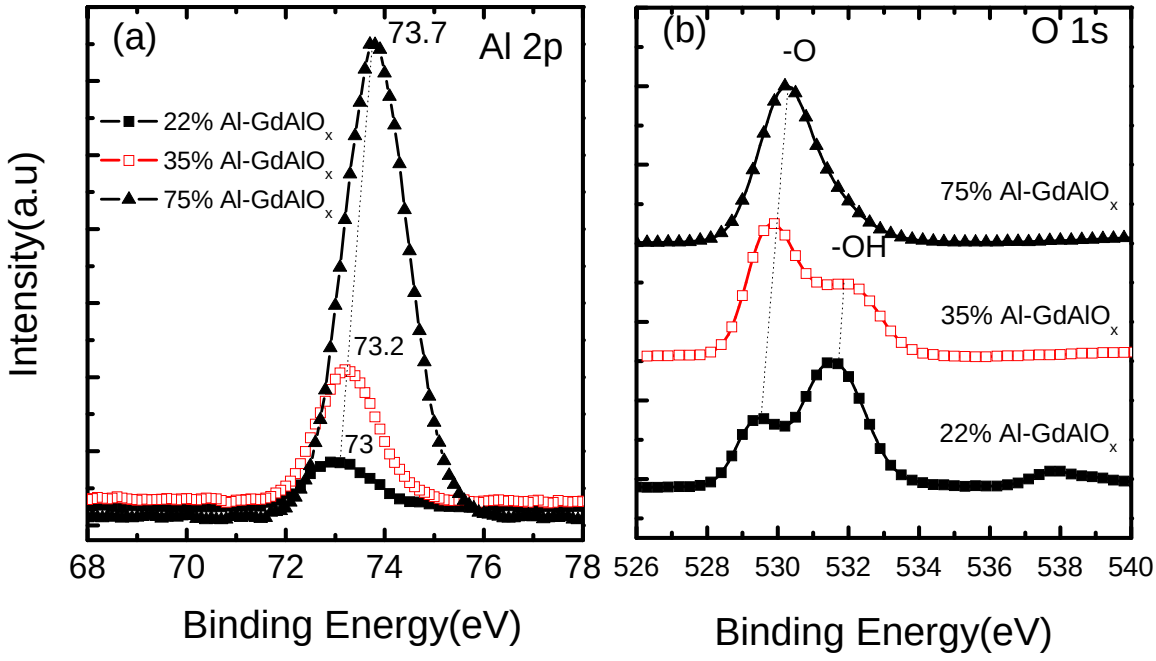


Fig. 4.2.4: XPS spectra for (a) Al *2p* core levels, and (b) O *1s* core level taken from various $GdAlO_x$ samples with Al incorporation rate from 22% - 75%. The core level peak positions of Al *2p* and O *1s* shift continuously towards higher binding energy with increasing Al components.

XPS spectra for Al *2p* and O *1s* core levels are shown in Fig. 4.2.4 (a) and Fig. 4.2.4 (b). It is observed that all the core level peak positions of Al *2p*, and O *1s* experience a shift to higher binding energy with the increase of Al concentration in the $GdAlO_x$ system. These changes are similar to the XPS chemical shifts in $ZrSiO_4$ vs. SiO_2 and ZrO_2 as discussed in *Guittet et al's* work [22]. Similar phenomenon has also been report in *H.Y. Yu et al's* paper for $(HfO_2)_x(Al_2O_3)_{1-x}$ system [23]. The earlier shift is due to the fact that Gd is a more ionic cation than Al in $GdAlO_x$ [24], and thus the charge transfer contribution changes with the increase of Al concentration [22, 24].

For the O *1s* core level spectra, both – O bond and – OH bond could be clearly identified, whereas the peak located at the lower binding energy is contributed from the – O bond and the peak at higher binding energy is corresponding to – OH bond. It is clearly observed in Fig. 4.2.4 (b) that the intensity of the – OH bond dramatic decreases with the increase of the Al percentage. The XPS spectra show evidence that the presence of Al is highly favorable to suppress the hydroscopic nature of Gd_2O_3

4. 2. 4 Crystal Structure Analysis by XRD

The crystal structure of $GdAlO_x$ film was studied by XRD (X-Ray Diffraction) in order to gain a better understanding of the crystal information associated with different Al incorporation percentages. The measurements were carried out using a GADDS (General Area Detector Diffraction) system equipped with Cu Kα X-Ray (λ = 1.5418 Å). The result is shown in Fig. 4.2.4.1. Fig. 4.2.5 (a), (b) and (c) are the XRD spectra for 22% Al-$GdAlO_x$, 35% Al-$GdAlO_x$, and 75% Al-$GdAlO_x$ respectively. The measurements were performed for the as deposited film and the films after 500°C, 700°C, 850°C RTP

anneal. Both 22% Al-$GdAlO_x$ and 35% Al-$GdAlO_x$ crystallize after 850°C anneal, and only 75% Al-$GdAlO_x$ still remains amorphous. This could be addressed to the higher crystallization temperature of Al_2O_3 [20].

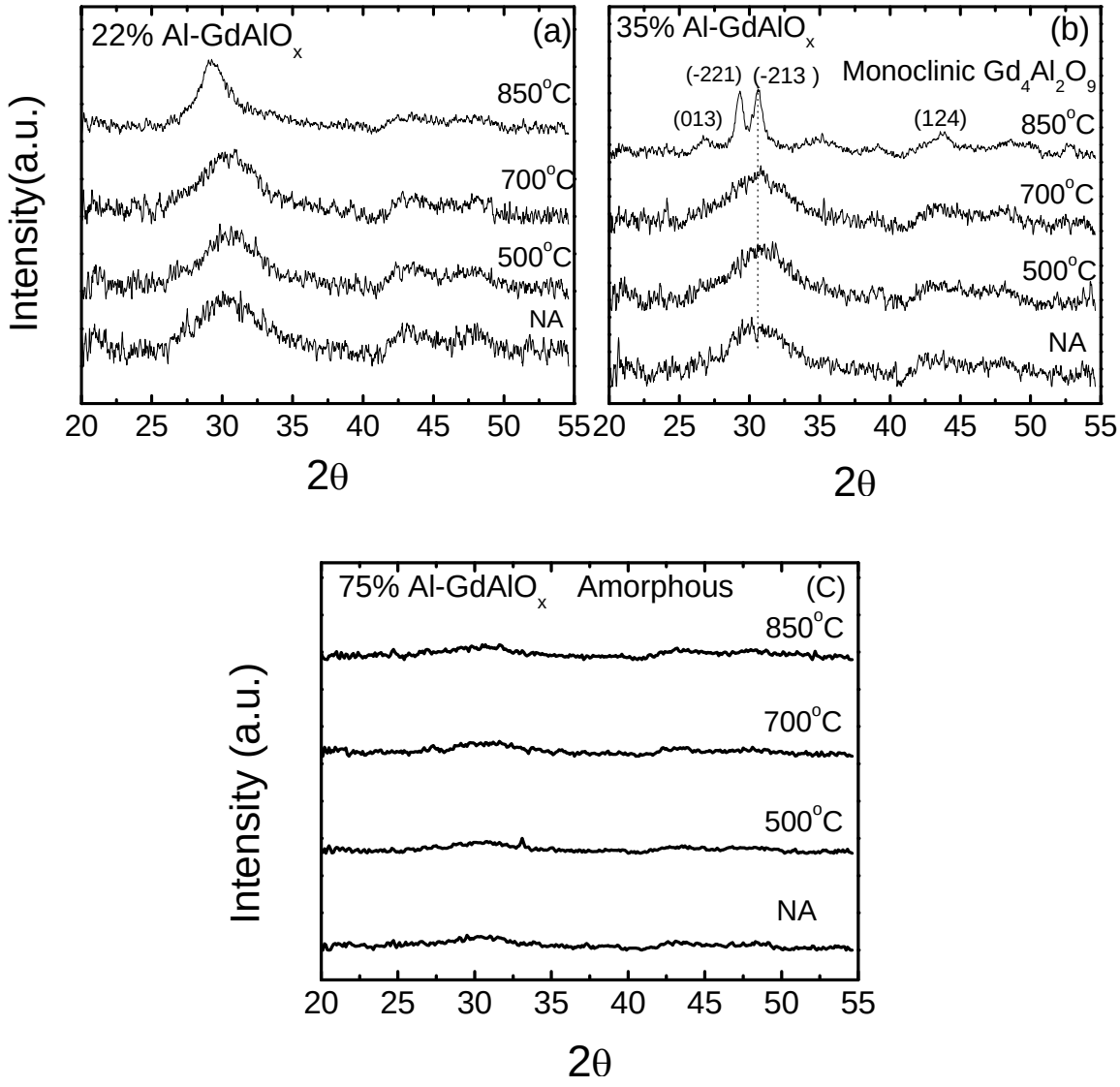


Fig. 4.2.5: XRD spectra of 20 nm $GdAlO_x$ films with various Al percentages. (a) 22% Al-$GdAlO_x$. (b) 35% Al-$GdAlO_x$. Monoclinic $Gd_4Al_2O_9$ crystal structure is clearly observed after annealing at 850°C. (c) 75% Al-$GdAlO_x$. The film remains amorphous.

Among all three $GdAlO_x$ films, only 35% Al-$GdAlO_x$ film forms a monoclinic ($Gd_4Al_2O_9$) crystal structure after annealing at 850°C as shown in Fig. 4.2.5 (b). This could be explained by the Gd to Al relative composition in 35% Al-$GdAlO_x$. The Gd : Al composition in 35% Al-$GdAlO_x$ is close to 2:1, which is the ideal composition for the monoclinic GAM structure whose chemical formula is $Gd_4Al_2O_9$ as indicated in the phase diagram in Figure 4.2.1. Bearing in mind that the crystalline structure of a dielectric material is closely related to its electrical performance, which has been discussed by many researchers [25]. The co-relationship between the crystalline structure and memory performance of $GdAlO_x$ blocking layer will be discussed in detail in the following section.

4. 3 Experiments and Devices Fabrication

SONOS-type flash memory transistors with $GdAlO_x$ blocking layer are prepared for the electrical performance evaluation. The silicon wafer is a p-type (100) substrate with a p-well surface doping concentration of $4\times10^{17}/cm^3$. The tunnel oxide is 5 nm thick thermally grown SiO_2, and the charge trapping layer is 5.5 nm thick LPCVD Si_3N_4. The physical thicknesses as well as the stack EOTs for all the split conditions are listed in Table 4.3.1. The blocking oxide layer is 22%, 35% and 75%-$GdAlO_x$ films (entire stack EOT is 10.8 nm, 10.5 nm and 13 nm, respectively). 15 nm thick Gd_2O_3 blocking oxide (entire stack EOT is 10.5 nm) was also prepared for comparison. A control sample with blocking layer deposited by ALD was also prepared. The thickness for Al_2O_3 was chosen to be 7.5 nm (entire stack EOT is 11.5 nm) such that the Al_2O_3 control sample will have

similar stack EOT as the devices with $GdAlO_x$ blocking layer. 150 nm TaN control gate is deposited by sputtering. After gate stack patterning and S/D implantation, dopant activation anneal was done at 850°C for 30 mins in a N_2 ambient in furnace tube. Forming gas (10 % H_2 in N_2) annealing at 420°C for 10mins was added before wafer fab out.

Table 4.3.1: Summarized physical thickness and stack EOT for SONOS transistors with Gd_2O_3, $GdAlO_x$ and Al_2O_3 blocking layers.

Split	BL layer	Physcial Thickness(nm)	Stack EOT(nm)
1	Gd_2O_3	15	10.5
2	22% Al-$GdAlO_x$	14.5	10.8
3	35% Al-$GdAlO_x$	11.1	10.5
4	75% Al-$GdAlO_x$	14.5	13
5	Al_2O_3	7.5	11.5

4. 4 Results and Discussion

4. 4. 1 Program/Erase Characteristic

The program characteristic in Fig. 4.4.1 (a) shows that $GdAlO_x$ blocking layer device has faster program speed than Al_2O_3 device but slower than pure Gd_2O_3 device. Among $GdAlO_x$ devices, lower percentage of Al shows faster program speed. This is well coincident with the expectation that higher κ value in blocking oxide will result in faster P/E speed, as the decrease of Al percentage in $GdAlO_x$ increases κ value.

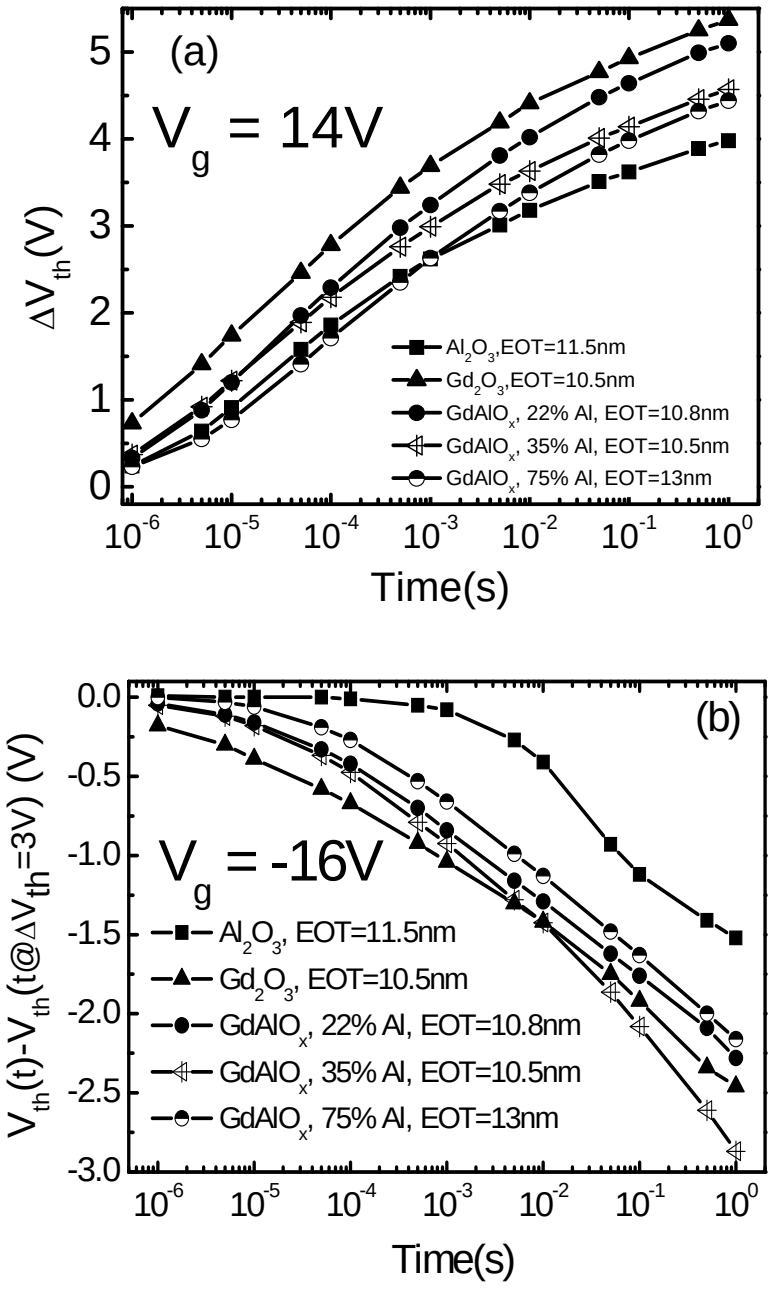


Fig. 4.4.1: (a) Comparison of program speed for $GdAlO_x$, Gd_2O_3 and Al_2O_3 blocking layers at program voltage V_g = 14 V. The Al percentage is varied from 22 % to 75 %. (b) Comparison of erase speed for $GdAlO_x$, Gd_2O_3 and Al_2O_3 blocking layers at erase voltage V_g = -16 V.

The erase characteristic for $GdAlO_x$ blocking layer is shown in Fig. 4.4.1 (b). At erase voltage V_g = -16 V, pure Gd_2O_3 blocking oxide sill has the fastest erase speed, whereas Al_2O_3 has the slowest speed. However, unlike the program situation, 35% Al-$GdAlO_x$ film has the best performance in erase, which is almost comparable with pure Gd_2O_3. The reason for the ultrafast erase speed of 35% Al-$GdAlO_x$ could be addressed to the better quality of the dielectric film, because erase speed is more sensitive to the dielectric quality such as back tunneling current from the gate.

Another phenomenon to be noted is that for the 75% Al-$GdAlO_x$, although its stack EOT is much thicker than the Al_2O_3 control sample, its exhibits much faster erase speed. At erase voltage V_g = -16 V and erase time t = 10 ms, the enhancement is as large as 40 % despite of 15 % increase in the stack EOT. The improvement in the erase obviously cannot be explained by the increase of hole tunneling current from the substrate which is responsible for the erase to happen, since the hole tunneling current is actually decreased for the device with 75% Al-$GdAlO_x$ blocking layer due to the increase in EOT. This phenomenon could only be explained by the reduced back tunneling current from the gate when negative gate voltage is applied. This is the advantage of using $GdAlO_x$ blocking layer over Al_2O_3 blocking oxide.

4. 4. 2 Retention Characteristic

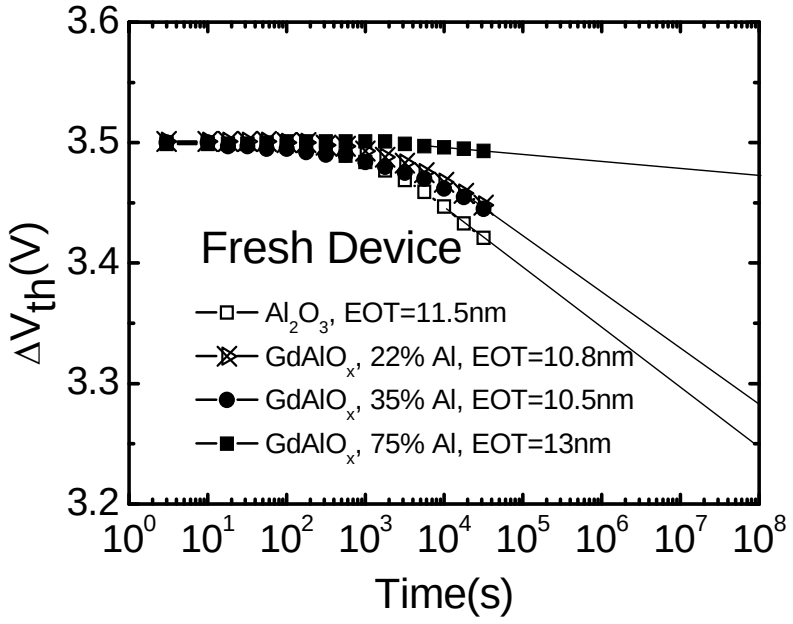


Fig. 4.4.2: Retention characteristics of fresh memory cells with $GdAlO_x$ and Al_2O_3 blocking oxides.

The charge retention characteristic of the fresh devices is shown in Fig. 4.4.2. All the $GdAlO_x$ samples show better retention properties compared to Al_2O_3 control sample, even though 22% and 35% Al-$GdAlO_x$ samples have thinner EOT. The retention property implies that Al doped Gd_2O_3 is a feasible material as the blocking layer in SONOS-type charge trapping flash memory devices.

The improved retention performance of $GdAlO_x$ devices is mainly caused by the increasing physical thickness of the blocking oxide. This is indeed the motivation to use the high-κ dielectric instead of the relatively low κ value Al_2O_3. As it has been explained in the introduction part, using high-κ dielectric for the blocking oxide layer, the electric field across the blocking oxide is proportionally reduced with its dielectric constant, which could be able to suppress the electron leaking from the blocking oxide without any

sacrifice in the P/E speed. For example, as shown in Fig. 4.4.1 and Fig. 4.4.2, for the memory devices with 22% and 35% Al-$GdAlO_x$ blocking layer, their EOT are smaller than the Al_2O_3 reference device, which enables its much faster operation speed. In the meanwhile, its increased physical thickness determines its superior charge retention performance comparing to Al_2O_3 blocking layer, which is shown in Fig. 4.4.2. The advantage of using high-κ blocking layer is well demonstrated in these two figures. Even for the device with 75% Al-$GdAlO_x$ which has a thicker EOT than the Al_2O_3 control sample, the enhancement in the P/E speed is not as large as the device with 22% and 35% Al-$GdAlO_x$ blocking layer, it poses the best retention performance among all the samples due to the smallest electric field across the blocking layer.

4. 4. 3 Charge Trapping Property

The charge retention characteristic of the devices after 1,000 P/E cycles is shown in Fig. 4.4.3. The P/E voltage for the cycling was + 14 V / - 16 V. the cycling time was chosen appropriately such that all the samples would have a program window of 3.5 V. As shown in Fig. 4.4.3, all the devices with $GdAlO_x$ blocking layers are still superior in retention property, compared to Al_2O_3. Different from the fresh devices, 35% Al-$GdAlO_x$ blocking oxide has the best retention after cycling, followed by 22% Al-$GdAlO_x$ and 75% Al-$GdAlO_x$ blocking layer which almost has the same retention performance after cycling.

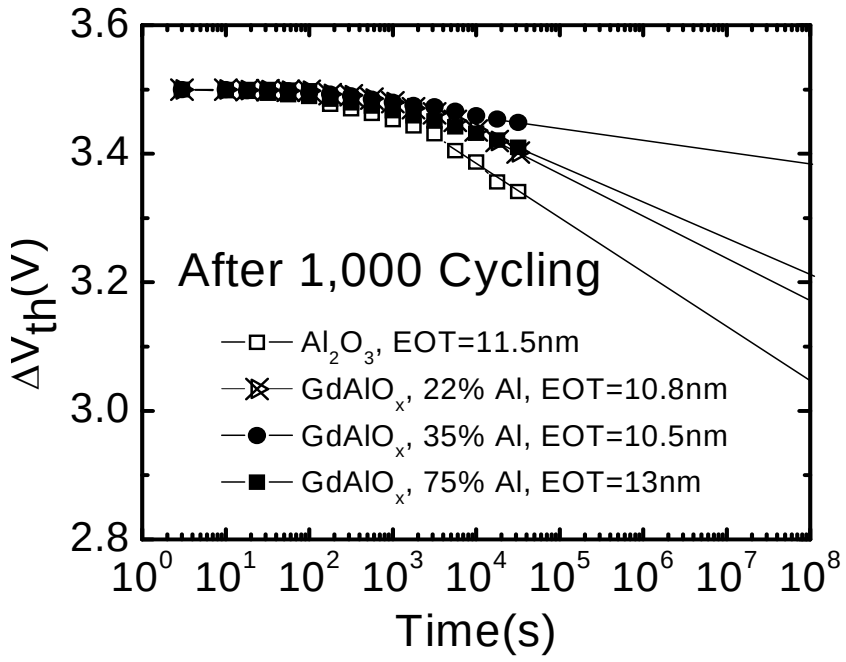


Fig. 4.4.3: Retention characteristics of memory cells with $GdAlO_x$ and Al_2O_3 blocking oxides. The retention is measured after 1,000 P/E cycles.

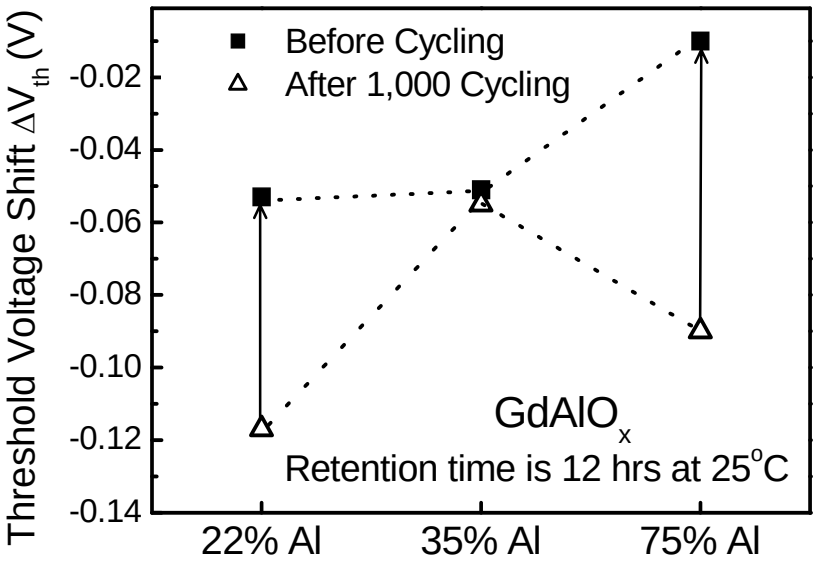


Fig. 4.4.4: Comparison of the charge loss property before and after 1,000 P/E cycles of GdAlOx film with different Al percentages.

It is well known that traps in the oxides are generated during P/E cycles; hence charge loss through trap-to-trap tunneling is enhanced [9, 18]. Comparing the threshold voltage shift before and after cycling of 22%, 35% and 75% Al-$GdAlO_x$ blocking layers shown in Fig. 4.4.4, 35% Al-$GdAlO_x$ has almost negligible trap generation whereas the other two have much larger number of traps generated after cycling. Since 35% Al-$GdAlO_x$ shows a monoclinic structure after 850°C anneal, a highly ordered crystalline structure found in 35% Al-$GdAlO_x$ film is believed to be the reason for small trap generation under electrical stress during P/E.

4. 4. 4 High Temperature Behavior

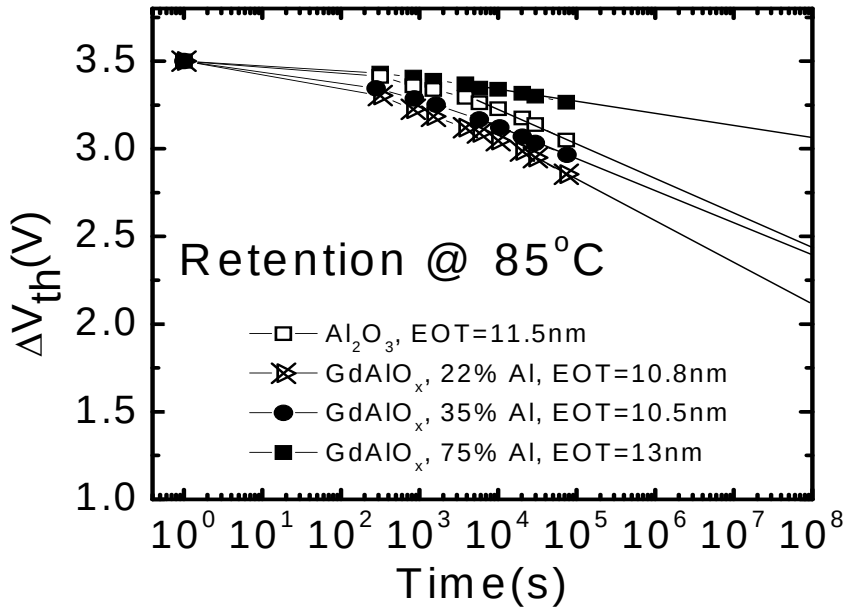


Fig. 4.4.5: Retention characteristic of memory cells with $GdAlO_x$ and Al_2O_3 blocking oxides at 85°C.

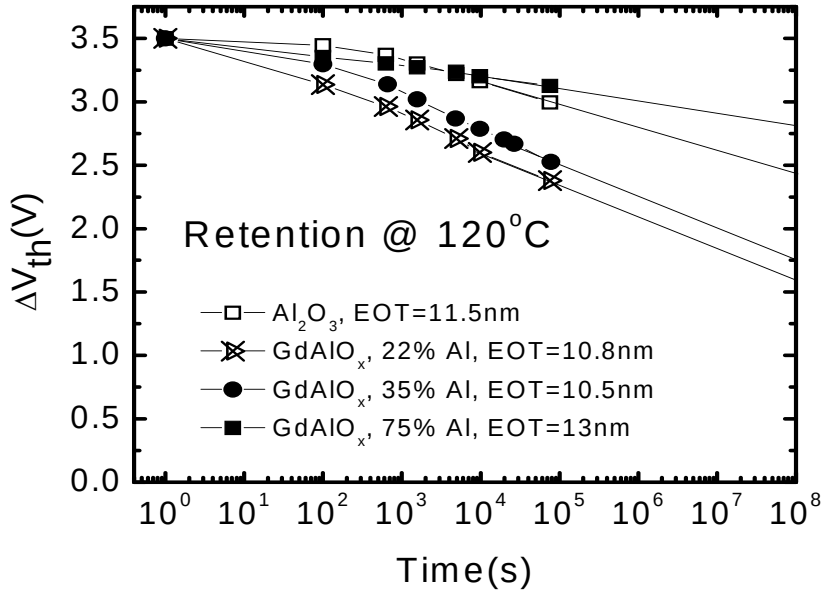


Fig. 4.4.6: Retention characteristic of memory cells with $GdAlO_x$ and Al_2O_3 blocking oxides at 120°C.

The retention characteristic at an elevated temperature has also been evaluated and the results are shown in Fig. 4.4.5 and Fig. 4.4.6. The retention was measured at 85°C and 120°C. Fig. 4.4.5 indicates that at 85°C, 75% Al-$GdAlO_x$ has the best retention performance among all the devices, whereas 35% Al-$GdAlO_x$ has comparable charge retention performance as the Al_2O_3 control device when extrapolated to 10 years retention time. The 22% Al-$GdAlO_x$ has the worst retention. To increase the retention measurement temperature up to 120°C as shown in Fig. 4.4.6, the margin between 75% Al-$GdAlO_x$ and Al_2O_3 control sample is becoming smaller. However, at the normal SONOS memory operation temperature (~100°C) discussed in this thesis, the 75% Al-$GdAlO_x$ blocking layer is still a better choice for blocking layer rather than the Al_2O_3.

The charge retention degradation of 35% Al-$GdAlO_x$ from 85°C to 120°C is also being observed, which makes it worse than the Al_2O_3 control sample at 120°C. The high

temperature retention behavior of $GdAlO_x$ and Al_2O_3 blocking layer shows that at sufficient high temperature, the electric field across the blocking is no longer the only factor to determine the charge retention. Other influencing factors are involved.

In order to find out the dominant mechanism responsible for high temperature charge retention behavior, the threshold voltage shifts of $GdAlO_x$ samples at different temperatures are plotted in Fig. 4.4.7. The measurement range is from room temperature up to 200^{o}C and the threshold voltage is measured after 12 hrs baking at the temperature. It is found in Fig. 4.4.7 that after 12 hrs baking, the ΔV_{th} curve has a maximum slope for the $GdAlO_x$ blocking oxide device with 22% Al incorporation, followed by 35% and 75% Al incorporation. As the conduction band offset of the $GdAlO_x$ film is increasing proportionally with the amount of Al incorporated, the maximum slope actually reflects the blocking oxide material with the smallest barrier, which in this case is 22% Al incorporated film. This is reasonable as Al_2O_3 has a larger conduction band offset than pure Gd_2O_3. Since thermionic emission strongly depends on tunneling barrier, the dominant charge leak mechanism at high temperature is thermionic emission rather than the electric field. It also reveals that one important criterion for blocking oxide in SONOS-type flash memory devices is the tunneling barrier. Materials with high tunneling barrier can effectively reduce the charge loss component through thermionic emission.

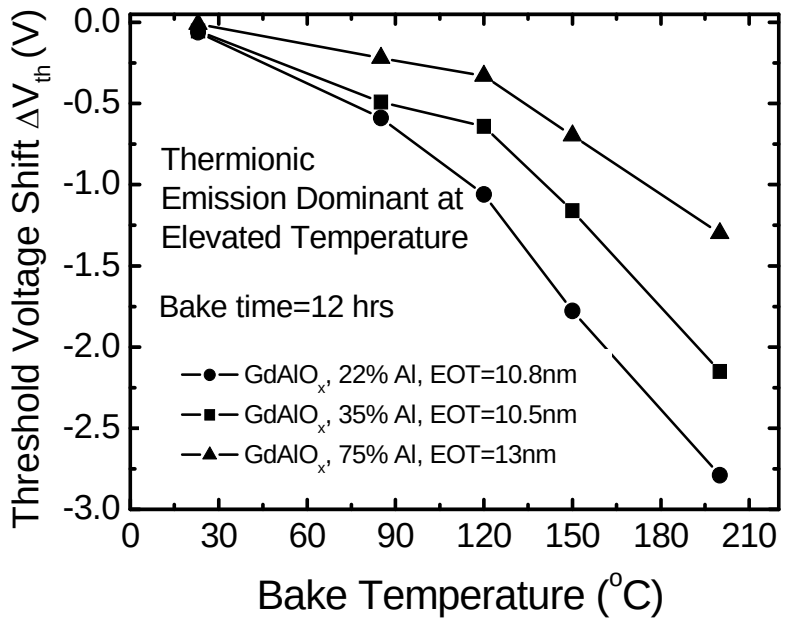


Fig. 4.4.7: ΔV_{th} vs. temperature for $GdAlO_x$ blocking layers. The V_{th} is measured after 12 hrs baking at different temperatures. The program state (@ΔV_{th}=3.5V) is taking as the reference state.

4. 5 Summary

In this chapter, another innovative technique to improve the charge retention in SONOS-type nonvolatile memory devices is proposed. The experimental data has proved that doping of Al into pure Gd_2O_3 film is an effective method to realize SONOS flash memory scaling. The incorporation of Al into the pure Gd_2O_3 film is capable of increasing the conduction band offset, improving the thermal stability as well as suppressing the hydroscopic nature of the rare earth oxide. The flash memory devices with $GdAlO_x$ blocking layer demonstrates improved P/E speed and superior charge retention for fresh devices, after P/E cycling and at elevated temperature up to 120°C. The charge loss properties indicate that the reduced electric field across the blocking

layer is the reason for the improved retention at room temperature, while the increased conduction band offset dominants the dramatic retention improvement at elevated temperature. Considering together with the κ value, P/E speed, charge retention and trap generation rate, 35% Al in $GdAlO_x$ is an optimized Al concentration in $GdAlO_x$ film. The monoclinic structure formation of the 35% Al-$GdAlO_x$ is favorable for the minimized trap generation rate.

References

[1] J. Pu, S. Kim, Y. Kim, and B. Cho, "Evaluation of Gadolinium Oxide as a Blocking Layer of Charge-Trap Flash Memory Cell," *Electrochemical and Solid-State Lett.,* vol. 11, pp. H252-H254, 2008.

[2] M. White and J. Cricchi, "Characterization of thin-oxide MNOS memory transistors," *IEEE Trans. Electron Dev.,* vol. 19, pp. 1280-1288, 1972.

[3] L. Lundkvist, I. Lundstrom, and C. Svensson, "Discharge of MNOS structures," *Solid State Electronics,* vol. 16, pp. 811-823, 1973.

[4] A. Roy and M. White, "Determination of the trapped charge distribution in scaled silicon nitride MONOS nonvolatile memory devices by tunneling spectroscopy," *Solid State Electronics,* vol. 34, pp. 1083-1089, 1991.

[5] Y. Kamigaki, S. Minami, and H. Kato, "A new portrayal of electron and hole traps in amorphous silicon nitride," *J. Appl. Phys.,* vol. 68, pp. 2211-2215, 1990.

[6] Y. Kamigaki, S. Minami, T. Hagiwara, K. Furusawa, T. Furuno, K. Uchida, M. Terasawa, and K. Yamazaki, "Yield and reliability of MNOS EEPROM products," *IEEE Journal of Solid-State Circuits,* vol. 24, pp. 1714-1722, 1989.

[7] Y. Hu and M. White, "Charge retention in scaled SONOS nonvolatile semiconductor memory devices-modeling and characterization," *Solid State Electronics,* vol. 36, pp. 1401-1401, 1993.

[8] S. Gu, C. Hsu, T. Wang, W. Lu, Y. Ku, and C. Lu, "Numerical simulation of bottom oxide thickness effect on charge retention in SONOS flash memory cells," *IEEE Trans. Electron Dev.,* vol. 54, pp. 90-97, 2007.

[9] Y. Lin, P. Chiang, C. Lai, S. Chung, G. Chou, C. Huang, P. Chen, C. Chu, and C. Hsu, "New insights into the charge loss components in a SONOS flash memory cell before and after long term cycling," *Proceeding of 11th IPFA, Taiwan,* pp. 239-242, 2004.

[10] S. Miller, "Effect of temperature on data retention of SNOS transistors," *8th IEEE NVSM Workshop, Vail, CO,* vol. 18, p. 5, 1986.

[11] P. McWhorter, S. Miller, and T. Dellin, "Modeling the memory retention characteristics of silicon-nitride-oxide-silicon nonvolatile transistors in a varying thermal environment," *J. Appl. Phys.,* vol. 68, pp. 1902-1908, 1990.

[12] L. Lundkvist, C. Svensson, and B. Hansson, "Discharge of MNOS structures at elevated temperatures," *Solid State Electronics,* vol. 19, pp. 221-227, 1976.

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

[14] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M. van Duuren, "Long term charge retention dynamics of SONOS cells," *Solid State Electronics,* vol. 52, pp. 1460-1466, 2008.

[15] E. Vianello, F. Driussi, P. Palestri, A. Arreghini, D. Esseni, L. Selmi, N. Akil, M. van Duuren, and D. Golubovic, "Impact of the charge transport in the conduction band on the retention of Si-nitride based memories," *38th European Solid-State Device Research Conference, 2008. ESSDERC 2008. ,* pp. 107-110, 2008.

[16] S. Wrazien, Y. Zhao, J. Krayer, and M. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," *Solid State Electronics,* vol. 47, pp. 885-891, 2003.

[17] A. Furnemont, M. Rosmeulen, J. VanHoudt, K. DeMeyer, H. Maes, and L. IMEC, "Physical modeling of retention in localized trapping nitride memory devices," *IEDM Tech. Dig.,* pp. 1-4, 2006.

[18] S. Chung, P. Chiang, G. Chou, C. Huang, P. Chen, C. Chu, and C. Hsu, "A novel leakage current separation technique in a direct tunneling regime gate oxide SONOS memory cell," *IEDM Tech. Dig.,* pp. 617-620, 2003.

[19] Y. Wang and M. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state," *Solid State Electronics,* vol. 49, pp. 97-107, 2005.

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

[21] S. Lakiza, O. Fabrichnya, Ch.Wang, M.Xinkevich, F.Aldinger, "Phase diagram of the ZrO_2 – Gd_2O_3- Al_2O_3 system", *J. European Ceramic Society*, vol. 26, pp. 233-246, 2006.

[22] M. Guittet, J. Crocombette, and M. Gautier-Soyer, "Bonding and XPS chemical shifts in $ZrSiO_4$ versus SiO_2 and ZrO_2: Charge transfer and electrostatic effects," *Phys. Rev. B,* vol. 63, p. 125117, 2001.

[23] H. Yu, M. Li, B. Cho, C. Yeo, M. Joo, D. Kwong, J. Pan, C. Ang, J. Zheng, and S. Ramanathan, "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.

[24] G. Lucovsky, "Electronic structure of high-k transition-metal and rare-earth gate dielectrics for aggressively-scaled silicon devices," *Extended Abstracts of International Workshop on Gate Insulator. IWGI 2001,* pp. 14-25, 2001.

[25] S. Migita, Y. Watanabe, H. Ota, H. Ito, Y. Kamimuta, T. Nabatame, and A. Toriumi, "Design and demonstration of very high-k (k~50) HfO_2 for ultra-scaled Si CMOS," *Symp. VLSI Tech. Dig.,* pp. 152-3, 2008.

CHAPTER 5

STRUCTURE OPTIMIZATION OF SONOS MEMORY DEVICES WITH 35% Al-$GdAlO_x$ BLOCKING LAYER

5. 1 Introduction

As discussed in Chapter 4, Gd_2O_3 films with Al incorporated shows a great potential to replace Al_2O_3 as the blocking layer in SONOS-type nonvolatile memory devices for performance improvement. 35% is considered as the optimum Al doping concentration to give the best device performance. The 35% Al-$GdAlO_x$ has an acceptable κ value (~17), sufficiently larger bandgap and conduction band offset which could effectively reduce the charge leakage through thermionic emission. In addition, its highly ordered monoclinic crystalline structure minimizes trap generation under high electric field, which is an essential requirement for a blocking oxide to meet the endurance criterion. In this chapter, further investigation to optimize the device structure for the SONOS device with 35% Al-$GdAlO_x$ blocking layer is described. The purpose of this study is to have the 35% Al-$GdAlO_x$ film which is made as thin as possible to give fast program/erase (P/E) speed, while thick enough to block the electrons leaking through

tunneling. The charge loss mechanism through the blocking layer is also systematically studied to provide a guide line for performance enhancement.

5. 2 Experiments and Devices Fabrication

Table 5.2.1: Summarized physical thickness and stack EOT for SONOS transistors with 35% Al- $GdAlO_x$ and Al_2O_3 blocking layers.

Split	BL layer	Physical Thickness(nm)	Stack EOT(nm)
1	35% Al-$GdAlO_x$	10.2	10.2
2	35% Al-$GdAlO_x$	11.4	10.5
3	35% Al-$GdAlO_x$	13.7	11
4	35% Al-$GdAlO_x$	17.7	13
5	Al_2O_3	7.5	11.5

SONOS-type flash memory cell transistors with 35% Al-$GdAlO_x$ blocking layer are prepared for electrical performance evaluation. The silicon wafer is a p-type (100) substrate with a p-well surface doping concentration of $4\times10^{17}/cm^3$. The tunnel oxide is 5 nm thick thermally grown SiO_2, and the charge trapping layer is 5.5 nm thick LPCVD Si_3N_4. The physical thicknesses as well as the stack EOTs for all the split conditions are listed in Table 5.2.1. The 35% Al-$GdAlO_x$ blocking layers have various physical thicknesses of 10.2 nm, 11.4 nm, 13.7 nm and 17.7 nm. The EOTs of the entire gate stack were 10.2 nm, 10.5 nm, 11 nm and 13 nm respectively. A transistor with 7.5 nm

Al_2O_3 blocking layer (entire stack EOT is 11.5 nm) was also fabricated as the control sample. The fabrication procedure is the same as described in chapter 4.

5. 3 Result and Discussion

5. 3. 1 Program/Erase Characteristic

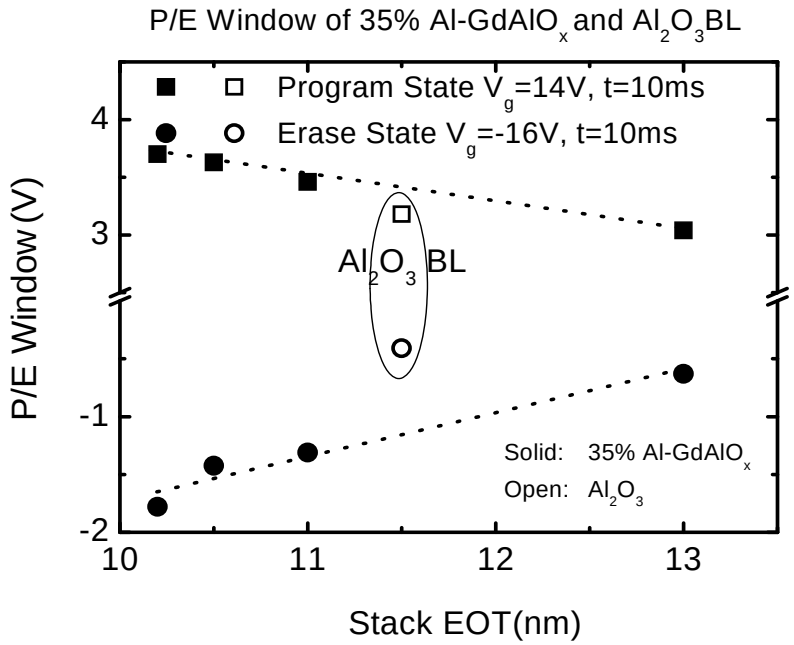


Fig. 5.3.1: Comparison of P/E window of 35% Al-GdAlO$_x$ samples with different thicknesses and Al_2O_3 control sample.

Figure 5.3.1 plots the P/E window as a function of entire stack EOT for both 35% Al-GdAlO$_x$ and Al_2O_3 blocking oxide material. The linear increase of the P/E window has been observed as the thickness of 35% Al-GdAlO$_x$ decreases. This phenomenon could be attributed to the increase of electrical field across the tunnel oxide during

program and erase. 10 % increase in the program window, and 178 % increase in the erase window is observed for 35% Al-GdAlO$_x$ blocking layer comparing to Al_2O_3 blocking layer of the same EOT (11.5 nm). Significant enhancement in operation window, especially in erase window, is an important characteristic of 35% Al-GdAlO$_x$ blocking oxide to fulfill the requirement for multi-bit memory cell operation. The typical P/E window versus the gate stack EOT could be practically useful as the production guideline.

5. 3. 2 Retention Performance Enhancement

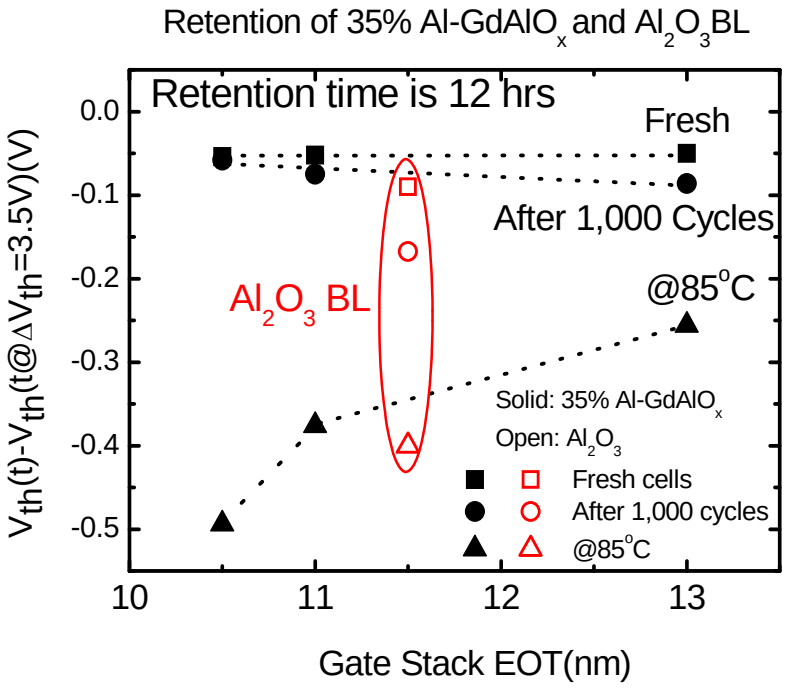


Fig 5.3.2: Comparison of charge retention of 35% Al-GdAlO$_x$ samples with different thicknesses and Al_2O_3 control sample. ΔV_{th} was measured after 12 hrs. The program state (@ΔV_{th} = 3.5 V) is taken as the reference state.

$GdAlO_x$ blocking oxide shows great advantage in retention under usual flash memory operation conditions as shown in Fig. 5.3.2. For all test conditions, including fresh devices, after cycling and at 85°C, 35% Al-$GdAlO_x$ (with a stack EOT of 11 nm and 13nm) has much improved data retention characteristics than the conventional Al_2O_3 blocking oxide. Among these two devices, the sample with 13 nm EOT has much enhanced retention. On the other hand, for the device with 10.5 nm EOT, the advantage in retention becomes obvious only for the fresh devices and after 1000 P/E cycles, but not at 85°C.

It has also been observed in Fig. 5.3.2 that thicker blocking oxide may trap more charges and cause poor retention after cycling. Thicker blocking oxide thickness provides a slightly better retention in fresh cells; however, after 10^3 P/E cycles, device degrades more seriously. This is due to the fact that electrons will experience less collision with the dielectric lattice and thus create fewer traps when tunneling through a thinner dielectric layer during the high voltage cycling process [1-3]. Since the cycling created traps are the most critical factor to determine the charge retention, thinner blocking layer is a better choice rather than a thick one in terms of retention after cycling. The result also suggests that there is a trade-off relationship in retention performance between the fresh and the cycled devices.

The charge retention characteristics of 35% Al-GdAlOx blocking oxide over time at elevated temperature (85°C and 120°C) are shown in Fig. 5.3.3 and Fig. 5.3.4. At 85°C, 35% Al-$GdAlO_x$ with the stack EOT of 13 nm has the best retention, whereas the one with stack EOT of 10.5 nm has the worst retention. This is due to the fact that 35% Al-$GdAlO_x$ (entire stack EOT is 13 nm) has a physically thicker thickness and experience a

smaller electric field, and thus reduced the electron leakage component through direct tunneling. The Al_2O_3 control sample is having the same retention as the device with 35% Al-GdAlO$_x$ blocking layer with a stack EOT of 11 nm, despite of a larger electric field.

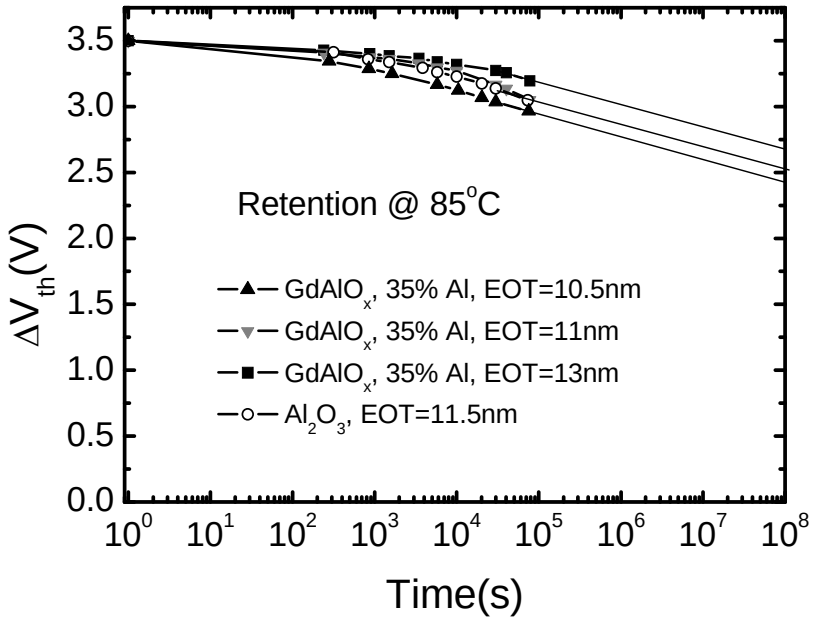


Fig 5.3.3: Comparison of the charge loss property of memory cells with 35% Al-GdAlO$_x$ and Al_2O_3 blocking oxides at 85°C.

The retention behavior at 120°C in Fig. 5.3.4 shows that the performances of the devices with 35% Al-GdAlO$_x$ blocking layer tremendously degrade comparing to their performances at 85°C. However the device with Al_2O_3 blocking oxide almost has the same retention as at 85°C. This makes the Al_2O_3 control sample having the best retention among all the devices at 120°C. To have a better understanding of the degraded retention performance of 35% Al-GdAlO$_x$ blocking layer, the dominant charge leakage mechanisms at different temperature ranges are analyzed in detail.

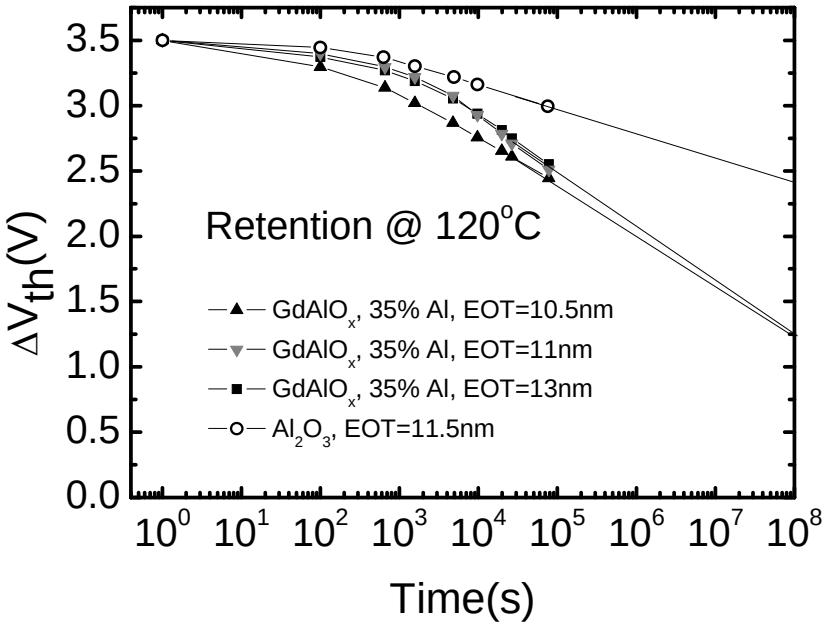


Fig. 5.3.4: Comparison of the charge loss property of memory cells with 35% Al-GdAlO$_x$ and Al$_2$O$_3$ blocking oxides at 120°C.

5. 3. 3 Dominant Charge Leakage Mechanism Study

As discussed in chapter 4, there are two major charge leak components in a SONOS flash memory cell before cycling: direct tunneling and thermionic emission [1, 4-8]. A novel leakage current separation technique has been developed to distinguish the two leakage current components. This technique was firstly proposed by *Steve S. Chung et al* [1] and it is a practically useful and easily implemented method to study the insight of SONOS flash memory charge loss behavior. In a SONOS memory cell, the injected charge ΔQ could be calculated as

$$\Delta Q = \Delta V_{th} \cdot \frac{A\varepsilon_{ox}}{t_{eff} - t_{ox} - t_N \cdot \varepsilon_{ox} / \varepsilon_N}$$

where ΔV_{th} is the shift in the threshold voltage after program, t_N is the distance between the charge centroid and the Si_3N_4 / tunnel oxide interface, A is the gate area, ε_{ox} is the permittivity of the tunnel oxide, t_{eff} is the effective dielectric thickness of the entire gate stack, t_{ox} is the thickness of the tunnel oxide and ε_N is the permittivity of the dielectric film between the charge centroid and the Si_3N_4 / tunnel oxide interface. Assuming the charge centroid is located closer to the Si_3N_4 / tunnel oxide interface after both program and erase, hence, the ΔQ could be obtained by the threshold voltage change multiplying the capacitance of the nitride layer and blocking layer. The charge loss amount ΔQ against different blocking oxide thickness for the devices with various 35% Al-GdAlO$_x$ blocking oxides is plotted in Fig. 5.3.5.

The data is extracted from the retention measurement up to 12 hrs at the temperature of 85^{o}C. As shown in Fig. 5.3.5, in the early stage of retention test at a temperature of 85^{o}C, charge loss is faster in thin blocking oxide cell than in thick blocking oxide cell. However, at a later stage (after ~ 3500 sec in this experiment), the charges are being lost with the same speed regardless of the blocking oxide thickness, which can be observed by the parallel lines in the figure. This indicates that the dominant leakage mechanism is first the thickness-dependent tunneling component and then followed by the thermionic emission component which is not sensitive to the oxide thickness. Therefore, it can be concluded that the charge loss behavior of a SONOS cell is first caused by the direct tunneling via the top and bottom oxides, and then followed by thermionic emission in which the trapped charges are thermally excited to overcome the barrier. Therefore, at the temperature of 85^{o}C, one should pay attention to both

mechanisms – tunneling and thermionic emission – to achieve the good retention performance.

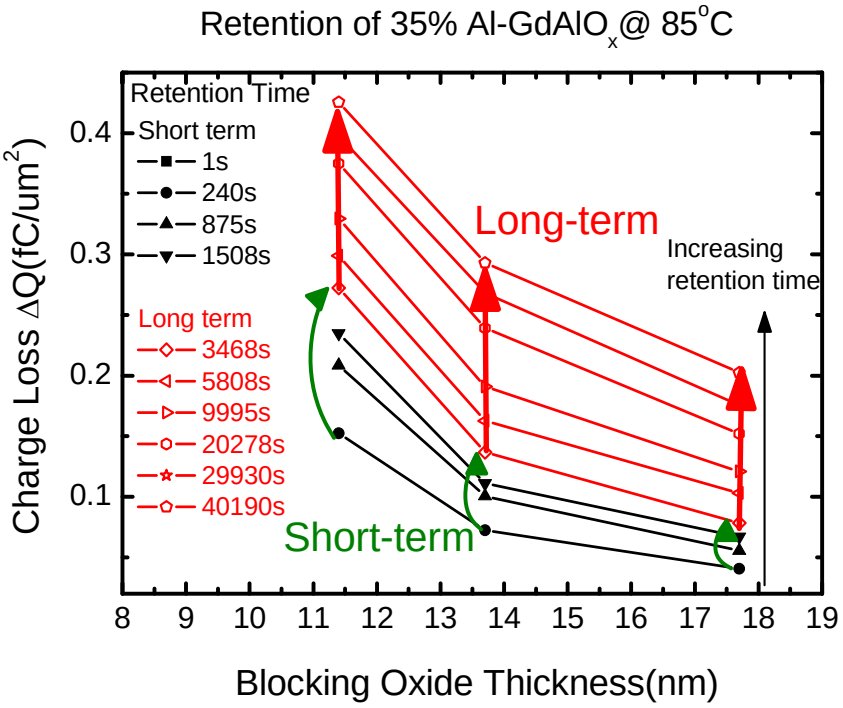


Fig. 5.3.5: At 85°C, charge leaks faster in thin GdAlO$_x$ film than in thick film at initial stage, but eventually leaks at the same rate regardless of the thickness. The program state (@ΔV_{th} = 3.5 V) is taking as the reference state.

When the retention measurement temperature is further increased to 200°C, all the devices exhibit the same retention property regardless of the blocking layer thickness as shown in Fig. 5.3.6. This means that only thermionic emission plays an important role in retention performance at 200°C. Tunneling through the blocking oxide is insensitive to temperature; however, the charge loss through thermionic emission is a strong function of temperature [27]. Therefore thermionic emission becomes the determining charge loss mechanism at 200°C. For devices operated at such a high temperature, the use of thicker blocking oxide would not have any advantage in retention as evidenced in Fig. 5.3.6.

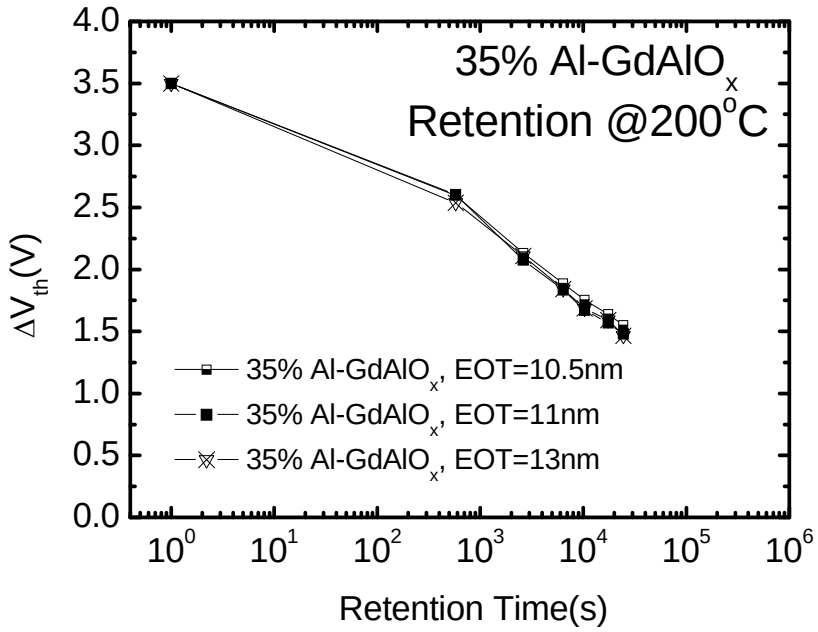


Fig. 5.3.6: Retention characteristic of memory cells with 35% Al-GdAlO$_x$ blocking oxide at 85°C. The V$_{th}$ value of a fresh memory cell is taken as the initial state.

To explain the charge loss mechanism change from 85°C to 200°C, the physical origin to cause tunneling and thermionic emission should first be considered. The tunneling is a function of the tunneling barrier height, oxide thickness as well as the electric field, whereas the thermionic emission is only related to the emission barrier and temperature. At the measurement temperature of 200°C, the amount of charge loss via tunneling through the blocking oxide is exact the same as at 85°C, since tunneling is relatively insensitive to temperature. However, the charge loss through thermionic emission increases tremendously as it is a strong function of temperature. Therefore the electron loss through tunneling only accounts for a very small percentage in the total amount of charge loss and it finally becomes negligible, and thermionic emission becomes the determining charge loss mechanism. Hence, thicker blocking oxide would

not have any advantage in retention at 200°C. It also explains the charge retention behavior in Fig. 5.3.4 while the measurement is done at 120°C which is in between 85°C and 200°C. At this measurement temperature, the retention is still regulated by the direct tunneling and thermonic emission; however, the thermonic emission take a higher portion than that at 85°C. Since Al_2O_3 blocking layer has a higher conduction band offset, it could effectively reduce this component.

5. 3. 4 Endurance Characteristic

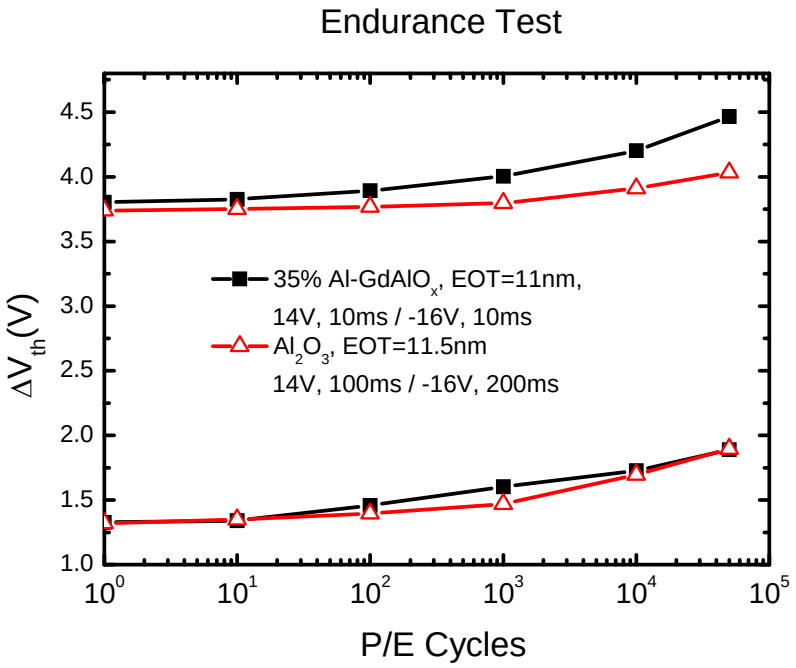


Fig. 5.3.7: Endurance characteristic of 35% Al-GdAlO$_x$ and Al_2O_3 blocking layer.

Figure 5.3.7 shows the endurance of SONOS cells after P/E cycling to have a more complete picture of 35% Al-GdAlO$_x$ blocking layer over Al_2O_3 blocking layer. The two samples have similar EOT and they are cycled at the same P/E voltage, but for different time, to have similar P/E window. Good operation window can be maintained

for up to 10^4 P/E cycles. 35% Al-$GdAlO_x$ blocking oxide is having an acceptable endurance characteristic to satisfy the requirement of SONOS flash memory devices operation.

5. 4 Summary

In this chapter, the performance of SONOS transistor with 35% Al-$GdAlO_x$ blocking layer in various thicknesses is systematically studied. Several aspects of the SONOS nonvolatile memory devices, including P/E characteristic, charge retention characteristic, cycling property as well as the high temperature charge loss behavior were investigated. In summary, a thick blocking oxide has a better retention for fresh devices, but not for P/E cycled devices. Its advantage also diminishes when the operation temperature goes up to 200^{o}C and the charge loss mechanism changes from two dominant components to one dominant component. In addition, the use of thick blocking oxide slows down the P/E speed. On the other hand, although the thin blocking oxide has relatively inferior retention performance at low temperature and for fresh devices, it is more robust to the electrical stress during P/E cycling. The fast operation speed is also an advantage. If the operating temperature is as high as 200^{o}C, the thin blocking oxide has eventually the same retention performance as the thick blocking oxide. Therefore, the optimization of blocking oxide will depend on the application and operation conditions of the devices. With a proper optimization of the structure, it is shown that the $GdAlO_x$ blocking layer can be a good candidate material for the further scaled charge-trap type flash memory devices.

References

[1] S. Chung, P. Chiang, G. Chou, C. Huang, P. Chen, C. Chu, and C. Hsu, "A novel leakage current separation technique in a direct tunneling regime gate oxide SONOS memory cell," *IEDM Tech. Dig.,* pp. 617-620, 2003.

[2] Y. Lin, P. Chiang, C. Lai, S. Chung, G. Chou, C. Huang, P. Chen, C. Chu, and C. Hsu, "New insights into the charge loss components in a SONOS flash memory cell before and after long term cycling," *Proceeding of 11th IPFA, Taiwan,* pp. 239-242, 2004.

[3] Y. Hu and M. White, "Charge retention in scaled SONOS nonvolatile semiconductor memory devices-modeling and characterization," *Solid State Electronics,* vol. 36, pp. 1401-1401, 1993.

[4] S. Miller, "Effect of temperature on data retention of SNOS transistors," *8th IEEE NVSM Workshop, Vail, CO,* vol. 18, p. 5, 1986.

[5] P. McWhorter, S. Miller, and T. Dellin, "Modeling the memory retention characteristics of silicon-nitride-oxide-silicon nonvolatile transistors in a varying thermal environment," *J. Appl. Phys.,* vol. 68, pp. 1902-1908, 1990.

[6] L. Lundkvist, C. Svensson, and B. Hansson, "Discharge of MNOS structures at elevated temperatures," *Solid State Electronics,* vol. 19, pp. 221-227, 1976.

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

[8] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M. van Duuren, "Long term charge retention dynamics of SONOS cells," *Solid State Electronics,* vol. 52, pp. 1460-1466, 2008.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6. 1 Conclusion

This work addressed some of the challenging issues in flash memory technology. Application of novel gate materials in both the floating gate-type flash memory devices as well as in the newly proposed charge trap-type flash memory devices demonstrates enhanced memory performance. In floating gate-type flash memory devices, the use of carbon doped polysilicon as the floating gate is proposed (chapter 2) to enhance the electron storage capability. In charge trap-type flash memory devices, Gd_2O_3 based high-κ blocking oxide (chapter 3, 4, 5) shows promising memory characteristics. Besides the performance enhancement, several process related issues, such as possible process integration scheme and material properties of the proposed novel materials were also discussed to implement the new structure successfully.

6. 1. 1 Study of Carbon Doped Polysilicon Floating Gate Flash Memory

Without implementing innovative memory structures or new materials, the widely commercialized floating gate-type flash memory devices almost meets its fundamental scaling limit. Intensive research to incorporate novel materials to the present memory structures have been carried out over the past decade. In Chapter 2, a novel floating gate engineering scheme, carbon doped polysilicon floating gate, was proposed to overcome the scaling barrier for floating gate-type flash memory devices. By incorporation of carbon into the conventional polysilicon floating gate, the conduction band offset of the floating gate is expected to increase due to the formation of a cubic structured silicon carbide phase, which has a larger conduction band offset than polysilicon. The carbon concentration needs to be well controlled to avoid drastic resistivity increase which decreases the operation speed. Successful formation of silicon carbide phase is the critical factor to achieve the performance enhancement. The proposed structure exhibits improved program/erase window and data retention when SiO_2 is used for the IPD layer. It has been found that incorporating carbon into the conventional n^+ polysilicon floating will result in significantly improved program/erase speed, especially for devices with small coupling ratios (~0.3), which addresses the bottleneck for sub 30 nm flash memory technology as indicated in the ITRS roadmap [26]. The data retention of such devices is improved as well. The adoption of carbon doped polysilicon floating gate needs only minimal process changes to the current floating gate memory technology, and could be considered as a reliable and practical option to ease the urgent need for high-κ IPD layer.

6. 1. 2 Study of Gd_2O_3 based High-κ Material in SONOS Memory

In the advanced SONOS-type flash memory devices, the application of high dielectric constant materials as the blocking oxide attracts much of the research interest. By using high-κ dielectric for the blocking oxide layer, the electric field across the blocking oxide is proportionally reduced with its dielectric constant. Thus electron injection from the gate during erase can be effectively suppressed, which will generally enhance the erase speed. It also allows the use of a physically thicker blocking layer film to minimize the electron leaking out to the gate during retention without comprising the increase in EOT. In Chapter 3, the feasibility of a novel rare earth high-κ material Gd_2O_3 as the potential candidate for the blocking layer application in SONOS-type flash memory devices was evaluated. The electronic structure of Gd_2O_3 film is measured by XPS for the first time, which has a great scientific value as well as practical application. The measurement results show that the Gd_2O_3 film has a band gap of 6.94 eV, valence band offset of 3.48 eV, conduction band offset of 2.36 eV with Si. Considering its large band gap and sufficiently high κ value, Gd_2O_3 has very suitable properties for the blocking layer in SONOS-type flash memory devices. It is also found that control of the crystal structure is the key factor in obtaining a high quality film. The monoclinic crystal structure is favorable for SONOS application as it results in a low leakage current. SONOS transistors with Gd_2O_3 blocking layer exhibits superior performance over Al_2O_3 blocking layer in several aspects, such as program/erase speed, room temperature retention, etc.

Gd_2O_3 shows good potential as the blocking oxide. However, after 1,000 program/erase cycles, its retention performance degrades. The degradation in the charge retention performance after cycling limits its practical implementation. In chapter 4, another innovative technique to improve the charge retention in SONOS-type nonvolatile memory devices is proposed. Doping of Al into the pure Gd_2O_3 has been proved to be an effective technique for the robust data retention after cycling, since the increase in the conduction band offset is always an effective method to block the electron leaking, both for room temperature and high temperature retention. The incorporation of Al into the pure Gd_2O_3 film is capable of increasing the conduction band offset, improving the thermal stability as well as suppressing the hydroscopic nature of the rare earth oxide. Flash memory devices with $GdAlO_x$ blocking layer demonstrates improved program/erase speed and superior charge retention in fresh devices, in devices which have undergone program/erase cycling and at elevated temperature up to 120^{o}C. The charge loss properties indicate that the reduced electric field across the blocking layer is the reason for the improved retention at room temperature, while the increased conduction band offset dominates the dramatic retention improvement at elevated temperature. Considering together with the κ value, program/erase speed, charge retention and trap generation rate, aluminum content of 35% is found to be optimal in $GdAlO_x$ film. The monoclinic structure formation of the 35% Al-$GdAlO_x$ is also found to be favorable for minimizing the trap generation rate.

In Chapter 5, structure optimization of a SONOS cell with 35%-$GdAlO_x$ blocking oxide was discussed. The study focuses on the relationship between the blocking layer thickness and long term retention reliability at room temperature, after program/erase

cycles and at elevated temperature. A novel leakage current separation technique will be applied to differentiate the leakage components in SONOS memory in order to improve the retention effectively. In summary, a thick blocking oxide has better retention in fresh devices, but not for devices that have undergone program/erase cycling. Its advantage also diminishes when the operation temperature goes up to 200^{o}C and the charge loss mechanism changes from two dominant components to one dominant component. On the other hand, although the thin blocking oxide has relatively inferior retention performance at low temperature and in fresh devices, it is more robust to the electrical stress during program/erase cycle. The fast operation speed is also an advantage. If the operating temperature is as high as 200^{o}C, the thin blocking oxide has eventually the same retention performance as the thick blocking oxide. Therefore, the optimization of blocking oxide will depend on the application and operation conditions of the devices.

6. 2 Limitations and Suggestions for Future Work

Even if this work contains a lot of practical and helpful information for advanced memory structures, it should be followed by more detailed investigation and study in order to satisfy the requirements of the coming ITRS road map. Therefore, it is worthy to note the following suggestions for future work.

The limitation of this thesis is that all the transistors fabricated in the experiment are long channel devices with the gate length ranging from 4 μm to 10 μm. The effect of the channel length on the memory performance is not discussed. Besides this, our study only concentrates on the performance of a single memory transistor cell. The reliability issues

related to memory array structures such as threshold voltage distribution, tail bit, and the program disturbance have not been dealt with.

Based on the limitations of this study, we propose the following directions for future research and investigation if the fabrication techniques are available：

1. Although the charge leakage mechanism in a single SONOS memory cell has been investigated in this thesis, further studies on the SONOS flash memory retention model is still necessary to have a more complete picture, such as the charge trap depth, the quantitative description of the charge leakage current and the tunneling barrier height.
2. The short channel effect for SONOS-type flash memory devices should be particularly studied, in order to indicate the scalability of this emerging flash memory structure and its capability of fulfilling the ITRS roadmap requirement.
3. Distribution of the threshold voltage and program disturbance are the important reliability concerns for memory arrays. Memory structures proposed in this thesis should be further studied in array structures if the fabrication facility allows.
4. Despite the attractive scaling capability of SONOS devices as discussed in a previous chapter, the endurance of SONOS is generally not as good as floating gate memory devices. The commercial floating gate memory devices can meet the program/erase endurance requirement of 10^6 cycles. However, normally SONOS type memory can only sustain 10^4cycles. Therefore, improvements in the endurance property of SONOS devices are necessary. The failure mechanism of endurance degradation, e.g. the degradation of the trapping layer or the tunneling oxide, needs to be analyzed theoretically as well as experimentally to find out the solutions for endurance improvement.

APPENDIX: LIST OF PUBLICATIONS

- ***Journal Papers***

1. **J. Pu**, S. J. Kim, S. H. Lee, Y. S. Kim, S. T. Kim, K. J. Choi and B. J. Cho, "Carbon doped polysilicon floating gate for improved data retention and P/E window of Flash memory", ***Electron Device Lett.***, vol. 29, n. 7, 2008, pp.680-690.
2. **J. Pu**, S. J. Kim, Y. S. Kim, and B. J. Cho, "Evaluation of Gadolinium oxide as a blocking layer of charge trap Flash memory cell", ***Electrochemical and Solid-State lett.,*** vol. 11, n. 9, 2008, pp. H252-H254.
3. **J. Pu,** Daniel S.H. Chan, Sun-Jung Kim and Byung Jin Cho, "Aluminum Doped Gadolinium Oxides as Blocking Layer for Improved Charge Retention in Charge-Trap Type Non-Volatile Memory Devices", ***IEEE Transaction Electron Devices,*** vol. 56, n. 11, 2009, pp. 2739–2745.
4. C. Shen, **J. Pu**, M.-F. Li and B. J. Cho, "P-type floating gate for retention and P/E window improvement of Flash memory devices", ***IEEE Transaction Electron Devices***, vol. 54, n. 8, 2007, pp. 1910–1917.
5. W. He, **J. Pu**, Daniel S. H. Chan, and B. J. Cho, "Performance Improvement for SONOS-type Flash Memory Using Lanthanum Based High-κ Dielectric as Blocking Oxide", ***IEEE Transaction Electron Devices,*** vol. 56, n. 11, 2009, pp. 2746–2751.
6. S. Wang, **J. Pu**, Daniel, S. H. Chan, B. J. Cho and K. P. Loh, "Wide Memory Window in Graphene Oxide Charge Storage Nodes", ***Appl. Phys. Lett.,*** vol. 96, n. 14, 2010, pp. 143109 1–3.

- ***Conference Papers***

1. **J. Pu**, S. J. Kim, Y. S. Kim and B. J. Cho, "Rare earth oxide (Gd_2O_3) as a Blocking Layer in SONOS-type Nonvolatile Memory Devices for High Speed Operation", ***The 15th Korean Conference on Semiconductors***, February 2008 in Korea, pp. 615-616.
2. **J. Pu**, S. J. Kim, Y. S. Kim and B. J. Cho, "Gadolinium Oxide (Gd_2O_3) Blocking Layer for Fast Program and Erase Speed in SONOS-Type Flash Memory Devices", Symposium F – Materials Science and Technology for Nonvolatile Memories, ***2008 MRS Spring Meeting,*** March 2008 in San Francisco, USA.
3. **J. Pu**, Daniel S. H. Chan, and B. J. Cho, "A Novel Floating Gate Engineering Technique for Improved Data Retention of Flash Memory Devices". ***The 9th International Conference on Solid-State and Integrated-Circuit Technology,*** Oct 2008 in Beijing, China, pp.285-289.
4. C. Shen, **J. Pu**, M.-F. Li and B. J. Cho, "Doping Optimization of Floating Gate for Retention and V_{th} Window Improvement in Flash Memory Device", ***2nd International Conference on Memory Technology and Design***, May 2007 in Giens, France. pp. 99-101.
5. B. J. Cho, W. He and **J. Pu**, "High-K dielectrics for charge Trap-type Flash Memory Application", ***2008 Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices,*** July 2008 in Sapporo, Japan, pp.37-41.
6. B. J. Cho, C. S. Park, P. W. Lwin, S. Y. Wong, **J. Pu,** W. S. Hwang, L. J. Tang, W. Y. Loh, and D. L. Kwong, "Dual metal gate process scheme for wide range work function modulation and reduced Fermi level pinning". 3rd International Conference

on Materials for Advanced Technologies, Symposium H: Silicon Microelectronics: Processing to Packaging, p. 41, MRS Singapore, July 2005

7. C. S. Park, S. C. Song, G. Versuker, H. N. Alshareef, B. S. Ju, P. Majhi, B. H. Lee, R. Jammy, H. K. Park, M. S. Joo, **J. Pu**, and B. J. Cho, "Demonstration of low V_t NMOSFETs using thin HfLaO in ALD TiN/HfSiO gate stack", 2006 International Conference on Solid State Devices and Materials (SSDM 2006), Yokihama, Japan, September 2006.