TOP-DOWN SI NANOWIRE TECHNOLOGY IN DISCRETE CHARGE STORAGE NONVOLATILE MEMORY APPLICATION

FU JIA

(B. Eng., Xi'an Jiaotong University)

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

I have been very fortunate to be a member of an active and open research lab in the National University of Singapore (NUS) and the Institute of Microelectronics (IME), a reputable A*STAR research institute. Prof. Zhu Chun Xiang, my main supervisor in the Silicon Nano Device Laboratory, NUS, is the most important person to my research work. I would like to thank him for bringing me into this colorful research world and also giving me many opportunities and freedom to pursue my interests. I would also like to thank Dr. Yu Ming Bin, my co-supervisor in IME, for his guidance and assistance that allowed me to adapt to a new working environment quickly. My sincere gratitude goes to Dr. Navab Singh and Dr. Lo Guo Qiang, who played a significant role in my research work in IME and gave me inspirations for many ideas. All of them are great mentors who have provided me with their patient guidance.

PhD life has been an enjoyable journey. I will always remember the interesting and precious memories I had of lectures by inspiring lecturers Prof. Cho Byung Jin, Prof. Ganesh Samudra, Prof. Yoo Wong Jong, Prof. Lee Sung Joo and my own professor in my initial two years in NUS. I also had a joyful time collaborating with my lab fellows. I would like to thank my seniors and colleagues in Prof. Zhu's group, such as Dr. Yu Xiongfei, Dr. Wu Nan, Dr. Zhang Qingchun, Dr. Huang Jidong, Dr. Song Yan, Chunfu, Jianjun and Ruilong for their guidance and encouragement. I would also like to express my gratitude to all my SNDL friends, such as Dr Ren Chi, Dr. Wang Yingqian, Dr. Gao Fei, Dr. Wang Xinpeng, Dr. Tan Kian Ming, Dr. Shen Chen, Chen Jingde, Pu Jing, Jiang Yu, Zhang Lu, Zhao Hui, Zang Hui, He Wei, Yang

Weifeng, Peng Jianwei, Wang Jian and Lee Wayne Yong, for their guidance and close friendship which I will always treasure. I would like to take this opportunity to thank the technical staffs in SNDL and IME for their great support and assistance in my PhD study.

My deepest gratitude goes to my dear parents. Your strong confidence towards me could drive any scares away from me over these years. My special thanks here go to you, Mr. Loh Guo Pei. Your knowing love makes me feel that I could never be lonely in any difficult circumstance.

Table of Contents

Summary

The commercial flash memory, which currently uses a polysilicon floating gate as the charge storage material, has faced issues of non-scalability of the tunnel oxide and interpoly dielectric in the course of scaling, due to the significantly reduced coupling ratio and serious gate interference. Due to scaling limitations of the conventional floating-gate nonvolatile flash memory cells, another type of nonvolatile memory based on discrete charge trapping is currently being considered as a promising alternative. The discrete charge storage nonvolatile memories are immune to local defect related leakage due to isolated charge storage nodes, providing larger scaling capability than floating gate devices.

This thesis proposes methodologies to resolve issues of gate stack scaling and voltage scaling in the SONOS type discrete charge storage nonvolatile memory in order to increase the possibility of it being employed in future semiconductor nonvolatile memory application. This thesis discusses solutions to scale the discrete trapped charge-storage nonvolatile memory based on a state-of-the-art non-traditional structure – a gate-all-around nanowire channel structure – whose fabrication method completely follows the CMOS-compatible rule in order to increase industrial adaptability of this novel technology.

A high-speed SONOS nonvolatile memory cell with a gate-all-around (GAA) Si-nanowire architecture is discussed in detail. The method of fabricating vertically stacked top-down nanowires with 5-nm diameter is highlighted. The nanowire SONOS device exhibits evident improvements in low voltage programming and fast programming and erasing speeds with regards to the planar control device. The

performance enhancement mechanism shall be explained by device modeling which investigates electron energy distribution, potential energy profile and electric field along each layer surrounding the nanowire channel. The gate-all-around nanowire channel structure is introduced into the poly-Si memory as a promising methodology to resolve issues of poor device subthreshold performance, low memory speed and inferior device uniformity in low temperature polycrystalline Si TFT memory devices, which can be integrated in future system-on-panel and system-on-chip applications. A strategy of optimizing SONOS-type memory characteristics is illustrated and discussed by integrating high-κ dielectric materials and metal gate electrode. The application of high-κ materials and TaN metal gate electrode, used to replace the conventional material used in nitride-based SONOS devices, exhibits improvement of memory erasing characteristics and causes of the performance enhancement will be investigated.

This thesis discusses several strategies to overcome challenges that SONOS-type discrete charge storage nonvolatile memory currently faces. In conclusion, novel device structures, in addition to new materials such as high-κ dielectrics and high work function metal gates, are promising candidates that can potentially be integrated into memory devices. Devices with the nanowire channel structure show promise for future nonvolatile applications due to their improved performance.

List of Tables

List of Figures

List of Symbols

C	capacitance (F)
d	thickness
d_{nw}	diameter of nanowire
E	electrical field (V/cm)
E_g	band gap (eV)
E_{inj}	electric field at the injecting surface
h	Planck's constant (6.626×10^{-34} Js)
I	current (A)
I_d	drain current (A)
I_g	gate leakage current (A)
J	current density (A/cm2)
L	channel length (μm)
m^*	effective mass (kg)
Q	charge (C)
Q_T	charges stored in the gate insulator at a distance d_1 from the gate (C)
T	temperature (°C)
t	time (sec)
t_{ox}	thickness of oxide
V	voltage (V)
V_{cg}	control gate potential (V)
V_d	drain voltage (V)
V_g	gate voltage (V)
V_{fb}	flatband voltage (V)

V_{th}	threshold voltage (V)
Φ_b	barrier height (eV)
Φ_M	work function of metal (eV)
κ_{Si3N4}	dielectric constant of Si_3N_4 (relative permittivity)
κ_{SiO2}	dielectric constant of SiO_2 (relative permittivity)
ΔE_c	conduction band offset (eV)
ΔE_v	valence band offset (eV)
ΔV_{th}	threshold voltage shift (V)

List of Abbreviations

ALD	Atomic layer Deposition
BBHH	Band-To-Band Tunneling Hot Hole Injection
BOX	Buried Oxide
CG	Control Gate
CHE	Channel Hot Electron
CVD	Chemical Vapor Deposition
DIBL	Drain-Induced-Barrier-Lowing
DG	Double Gate
DRAM	Dynamic Random Access Memory
DRIE	Deep Reactive Ion Etching
DT	Direct Tunneling
EEPROM	Electrically Erasable and Programmable Read Only Memory
EOT	Equivalent Oxide Thickness
EPROM	Electrically Programmable Read Only Memory
FG	Floating Gate
F-N	Fowler-Nordheim
GAA	Gate-All-Around
HRTEM	High Resolution Transmission Electron Microscopy
IPD	Interpoly Dielectric
ITRS	International Technology Roadmap for Semiconductors
LPCVD	Low Pressure Chemical Vapor Deposition
MONOS	Metal/ Oxide / Nitride / Oxide / Silicon

NC	Nanocrystal
NVM	Nonvolatile Memory
PDA	Post Deposition Anneal
PECVD	Plasma-Enhanced Chemical Vapor Deposition
Poly-Si	Polycrystalline Silicon
PVD	Physical Vapor Deposition
RAM	Random Access Memory
ROM	Read Only Memories
S/D	Source and Drain
SCE	Short Channel Effect
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectroscopy
SOI	Silicon-On-Insulator
SONOS	$Si / SiO_2 / Si_3N_4 / SiO_2 / Si$
STI	Shallow Trench Isolation
SRAM	Static Random Access Memory
SS	Subthreshold Swing
TAHOS	$TaN / Al_2O_3 / HfO_2 / SiO_2 / Si$
TEOS	Tetraethyl Orthosilicate
TEM	Transmission Electron Microscopy
UTB	Ultra-Thin Body
VLS	Vapor-Liquid-Solid
VLSI	Very Large-Scale Integration

Chapter 1

Introduction

1.1 Introduction of Semiconductor Memory Technology

1.1.1 Semiconductor Memory Categories

Despite some unpredictable fluctuations, the semiconductor market has been increasing steadily over the years, and the growing trend is expected to continue in the future. Memory components have been an important part of the semiconductor market and are projected to account for more than 20% of the IC market, making them the second largest category of IC's overall behind logic components [1.1], as shown in Fig. 1.1.

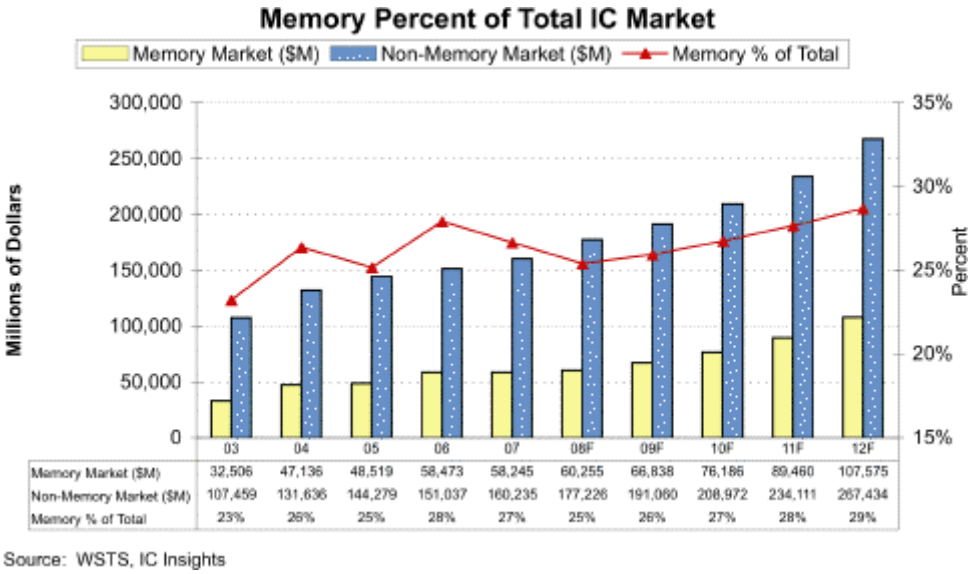


Fig. 1.1: Revenues of semiconductor memory market versus year.

Semiconductor memories are based on a metal-oxide-semiconductor (MOS) technology. As shown in Fig. 1.2, there are various types of semiconductor memories, which are fundamental to the architecture of computers. There are two basic categories of semiconductor memories: volatile memory, which requires power to maintain the data content; and nonvolatile memory, which is able to maintain the data content without any power supply. Most forms of random access memory (RAM) are of the volatile type. Random access means that locations in the memory can be written to or read from in any order. All data on the computer is stored on the hard drive, but in order for the Central Processing Unit (CPU) to work, the data is written into the RAM chips. There are two different types of RAM: dynamic random access memory (DRAM) and static random access memory (SRAM), which are different in the technology they use to store data.

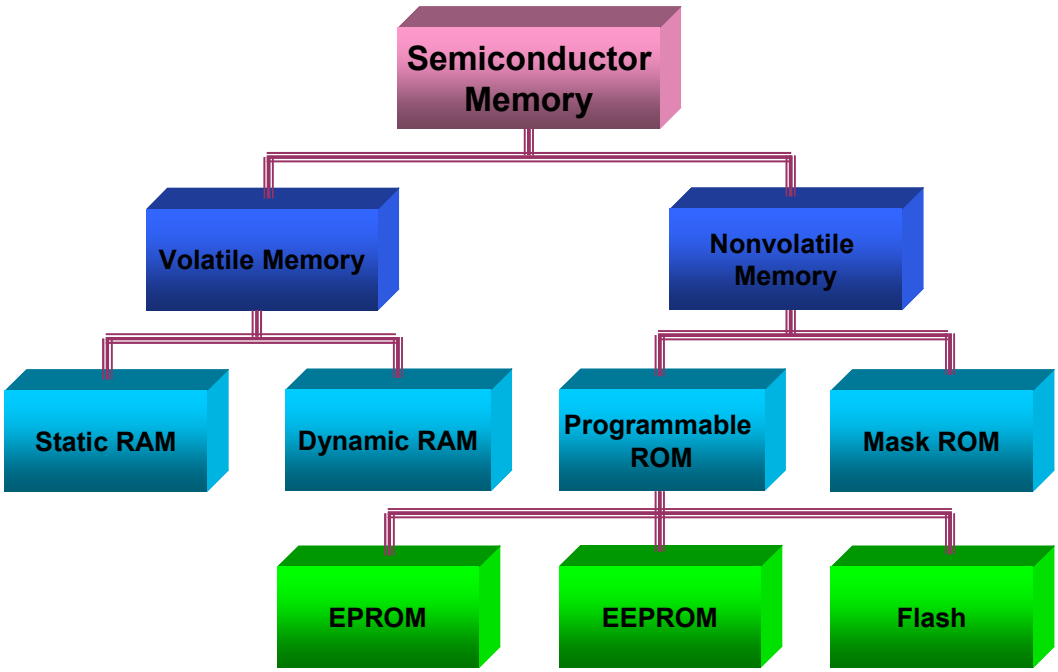


Fig. 1.2: Semiconductor memory family tree

Being the more common type, DRAM has a simple structure: only one transistor and a capacitor are required for each bit. This enables DRAM to be packed with a high density. Since capacitors leak charge, data stored in DRAM has an extremely short storage time, typically about 100 ms. Information is stored for DRAM by refreshing the capacitor charge periodically. DRAM offers access times of about 60 ns [1.2].

A typical SRAM contains six transistors (6T) to store a memory bit. Each bit is stored on four transistors, while two additional transistors control the access during read and write operations. Unlike DRAM, SRAM does not need to be refreshed periodically. Therefore, SRAM is faster as it can give as low as 10 ns access time. Despite being faster, SRAM is not as commonly used as DRAM due to the much higher cost-per-bit. Hence, SRAM is used as a memory cache in powerful microprocessors which requires fast speed to access data frequently, while slower DRAM is used for main memory for its attractive low cost per byte.

On the other hand, nonvolatile memories are typically used as secondary storage in commercial electronic products. In the last decade, memory chips with low power consumption and low cost have attracted more and more attention due to the increasing popularity of portable electronic devices. Nonvolatile memories have become a very important category of semiconductor memory ever since the first Erasable Programmable Read Only Memory (EPROM) was invented [1.3]. Almost all electronic systems require the storage of some information in a permanent way. Thanks to the nonvolatile aspect, nonvolatile memories have enhanced the development of multimedia applications and personal consumer appliances such as digital cameras and USB Flash drives. The market share of nonvolatile memories has increased exponentially over the past few years, and is projected to grow further.

Nonvolatile memories include mask-programmed ROM and reprogrammable memories such as EPROM, Electrically Erasable and Programmable Read Only Memory (EEPROM) and flash. The mask-programmed ROM chips physically encode the data when they are manufactured at the factory with a special mask; however, they are not allowed to change the content. EPROM has a one-transistor memory cell and can provide high density and cost effectiveness, but it provides the opportunity to reprogram the device after the data are erased by exposure under strong ultraviolet light for a long time. The erasure ability of EPROM enables it to be reused and makes it an important invention in the development of semiconductor memory. EEPROM is based on a structure similar to EPROM, but it differentiates itself from EPROM by its electrical programming and erasing ability. EEPROM can write and erase each bit separately, and the data stored can be maintained for as long as required. Thus, EEPROM has the features of both RAM and ROM in that the EEPROM can be accessed per single bit like RAM, and at the same time, it can keep the contents when it loses electrical power like ROM. However, EEPROM's are manufactured for specific applications, due to the larger area and higher cost per cell.

Flexibility and cost are usually the two aspects to be compared between different nonvolatile memories [1.4]: flexibility shows the robustness of the device, while cost represents process complexity for a specific cell size. The flash memories presented by Toshiba in the 1984 [1.5] turned out to be the best compromise of the two parameters. Although flash is a specific EEPROM, flash is erased by blocks of different size while a regular EEPROM is erased bit by bit. Flash memory also costs much less than EERPOM and therefore became a dominant nonvolatile memory technology. Two main applications have opened up and driven the development of the current flash market [1.6]. The first application of flash memory is using NOR type

flash, which provides fast memory read speed and random access to any location, to store program code in cellular phones. Another application is the usage of NAND type flash as data storage medium in devices such as USB memory cards, MP3 music players and PDAs, which plays a significant role in lowering the costs of such devices. As a result, the flash market, especially the NAND type flash, had grown exponentially in the past decade. It is projected that the flash market will generate tens of billions of dollars in revenue and reach the size of DRAM market by 2010 [1.7].

1.1.2 Structure and Operation Mechanism of Flash Memory

A common flash memory cell consists of one floating-gate transistor. The transistor in flash memory is similar to a standard MOSFET, except there are two gates instead of one in MOSFET. The schematic cross section of a floating gate device is shown in Fig. 1.3. The first gate is referred to as a control gate, which acts as the external gate. The second gate is a floating gate (FG) completely surrounded by dielectric layers, tunnel oxide and interpoly dielectric (IPD). Being electrically isolated, the FG is able to charge and hold carriers for the memory cell.

The basic operation principle of flash memory devices is the storage of charges in the floating gate, as illustrated in Fig. 1.3. If the charges are injected into the gate insulator layer, the threshold voltage of the device can be changed between two distinct values. From the theory of the MOS transistor, the shift of threshold voltage can be expressed by equation (1) [1.8]:

$$\Delta V_T = V_T - V_{T0} = -\frac{Q_T}{\varepsilon_I} d_I \tag{1}$$

where

Q_T = the charges stored in the FG at a distance d_I from the control gate

ε_I = the dielectric constant of the insulator

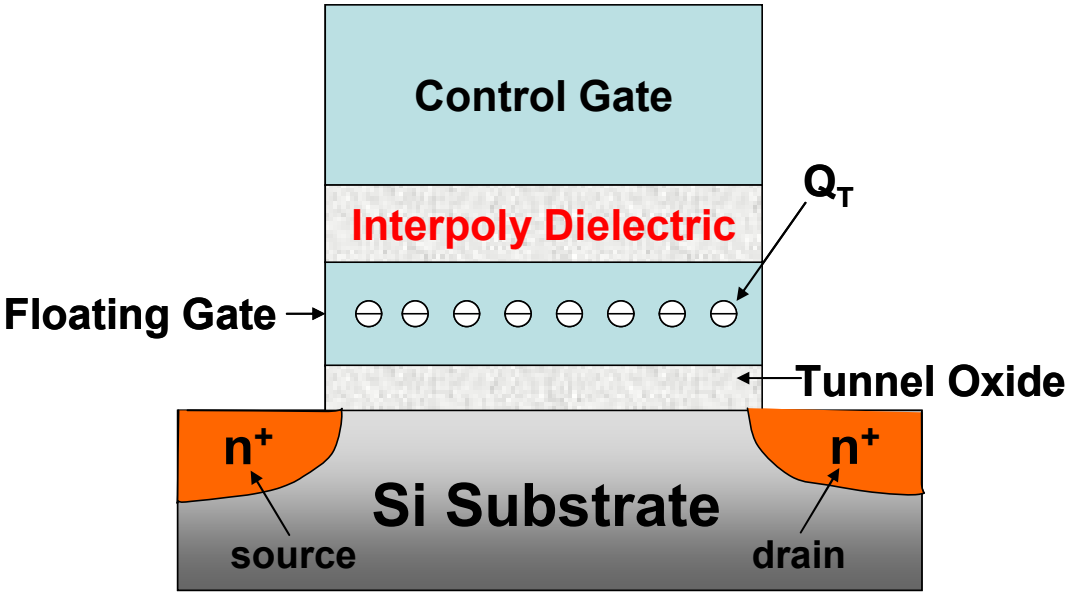


Fig. 1.3: A schematic cross-section of a single floating-gate transistor. FG is surrounded by dielectric layers and isolated from channel and IPD. Taking the n-type memory cell as an example, electrons are injected from substrate by applying a positive voltage stress at the gate. Electrons will remain trapped in the FG even after the power is removed from the gate.

The information on the device is detected by reading the drain current using a gate voltage with a value between two possible threshold voltages. Conventionally, the state with high read current and lower threshold voltage is recognized as logic "1", as the transistor is conducting a current; while the other state with low read current and higher threshold voltage corresponds to logic "0" state, since the transistor is cut off at this state.

Depending on the material of the storage medium in the transistor, the flash-type nonvolatile memory can be divided into two classes. The first class of devices contains a conducting or semiconducting layer to store the charges, which are trapped and electrically isolated by the surrounding dielectric layers; this type of devices is usually referred as a floating gate structure as shown in Fig. 1.3. In the second class of devices, the charges are stored in discrete trapping centers of a dielectric layer, such as Si_3N_4 [1.9] or HfAlO [1.10]. This type of devices is usually referred to as

charge-trapping memories. SONOS (Silicon/Oxide/Nitride/Oxide/Silicon) memory is the most typical charge-trapping memory and it has been considered as a promising candidate to replace floating-gate memory, due to the scaling advantage of the SONOS structure compared to the floating-gate device structure.

In both classes of nonvolatile memories, the charges are stored in the trapping materials and the programming of the memory devices is realized by modifying the threshold voltage. In order to inject the charges into the trapping material through the tunnel oxide layer, there are two mechanisms commonly used nowadays: Fowler-Nordheim (F-N) tunneling mechanism, which is usually used in devices with relatively thin tunnel oxides; and channel hot electron (CHE) injection, which is employed for device with relatively thick tunnel oxides. The details of the two mechanisms are introduced below.

1.1.2.1 F-N Tunneling

F-N tunneling mechanism is based on the quantum mechanical tunneling mechanism induced by a high electric field; therefore it is also a field-assisted electron tunneling mechanism. When a voltage is dropped across the Si-SiO_2 structure, as shown in Fig. 1.4, the band structure will be influenced. The barrier height for electron tunneling is kept unchanged since it is determined by the conduction band offset between the two different materials. However, electrons in the silicon conduction band are able to tunnel through the triangular energy barrier as the barrier width has been greatly reduced. The probability of the tunneling through the oxide layer is dependent on the magnitude of the applied electric field.

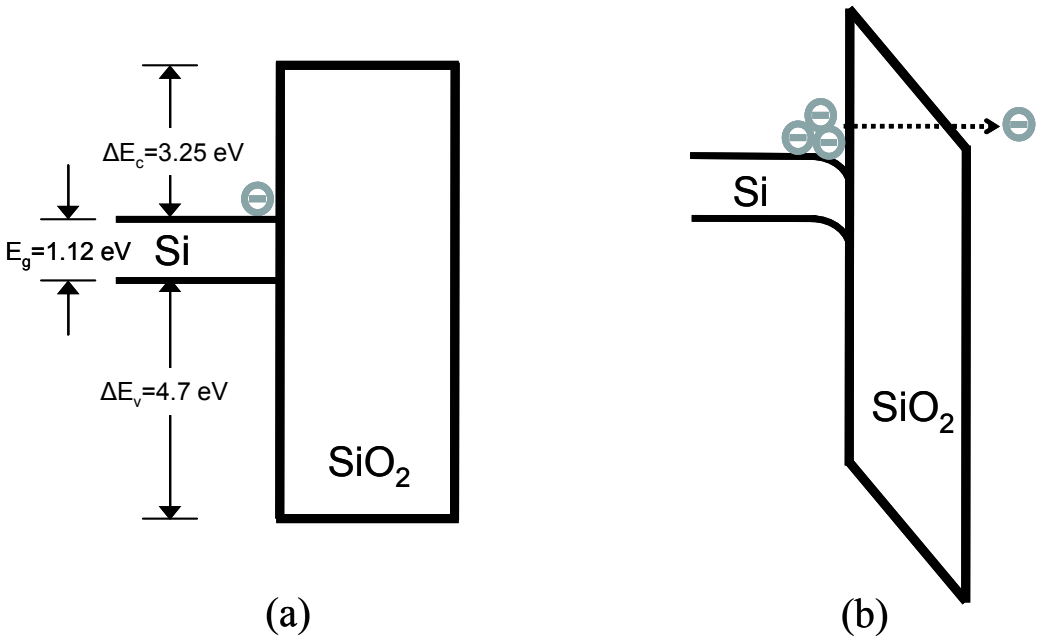


Fig. 1.4: Si and SiO_2 energy band diagram system (a) without applying any voltage and (b) with applying a positive voltage at SiO_2 side. Electrons are able to tunnel through the thick SiO_2 layer by F-N tunneling due to a strong electric field reduces the barrier width. Conduction and valence band offset (ΔE_c and ΔE_v) keeps unchanged during the process.

In Fig. 1.4, ΔE_c indicates the Si-SiO_2 energy barrier height (3.25 eV for electrons and 4.7 eV for the holes). The applied voltage at the control gate induces the electric field at the injecting surface (E_{cg}), resulting in a potential barrier with a width dependent on the applied voltage. The electrons collected at the floating gate leads to a tunneling current density which is given by equation (2) [1.11]:

$$J = \alpha E_{cg}^2 \exp\left(-\frac{\beta}{E_{cg}}\right) \tag{2}$$

with

$$\alpha = \frac{q^3}{8\pi h \Delta E_c}\frac{m}{m^*},\ \beta = 4\sqrt{2m^*}\frac{\Delta E_c^{\frac{3}{2}}}{3hq}$$

1.1.2.2 CHE Tunneling

Nonvolatile memory can also be programmed by hot-carrier injection

mechanism [1.12]. The hot electron injection is for n-type NVM on a p-substrate. Hot carriers generally refer to the particles which attain the kinetic energy from a high electric field. The hot-electrons get their energy from the drain voltage, and are further accelerated by the lateral electric field along the channel. Once they obtain sufficient energy, these hot electrons will surmount the SiO_2/Si barrier and tunnel into the floating gate to program the cell. Potential at the control gate V_{cg} affects the charges tunneling into the floating gate while the potential at the drain V_d plays a role in determining the speed of programming. Fig. 1.5 (a) shows a cross section of a NVM with hot electron injection programming. The change of band structure during the hot electron programming process is shown in Fig. 1.5 (b).

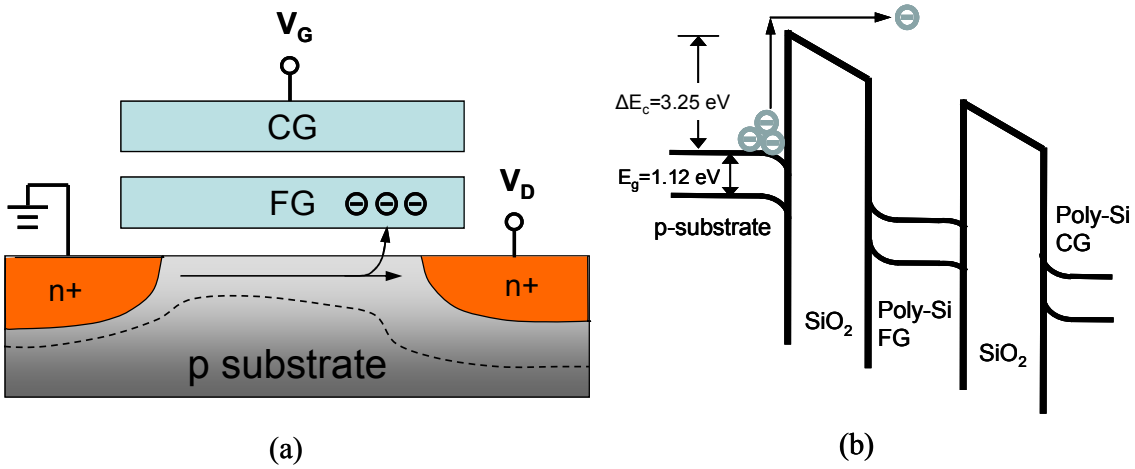


Fig. 1.5: (a) At CHE stress condition, electrons gain enough energy while drifting across the channel and are injected through the tunnel oxide, causing a gate current. (b) Energy band diagram of a floating-gate memory cell during programming by hot-carrier injection.

1.1.3 Challenges of Semiconductor Flash Memory Scaling

The tunneling mechanisms are commonly known to be closely related to the architectures of flash memory circuits. NOR flash and NAND flash are the two main

categories which dominate the nonvolatile memory market today. In the internal circuit configuration of NOR as shown in Fig. 1.6, the individual floating-gate n-type transistors share a bit line, while the sources are connected together by sharing a source line on the other side. Programming of NOR flash is usually done by CHE tunneling in 10 μs, and erasing is conducted by F-N tunneling with a relatively longer time (~0.5 s) [1.6]. The array of NAND flash allows a much smaller configuration size, as a string of 16 or 32 cells are connected between a common bit line and source line. Its simpler design translates to a small cell size and low cost-per-bit. Moreover, the cell is programmed and erased by F-N tunneling as a block, enabling fast memory erase speed. Despite its complicated design, NOR flash architecture allows for fast random read access, which is required for code execution. In contrast, NAND is more suitable for high capacity data storage applications.

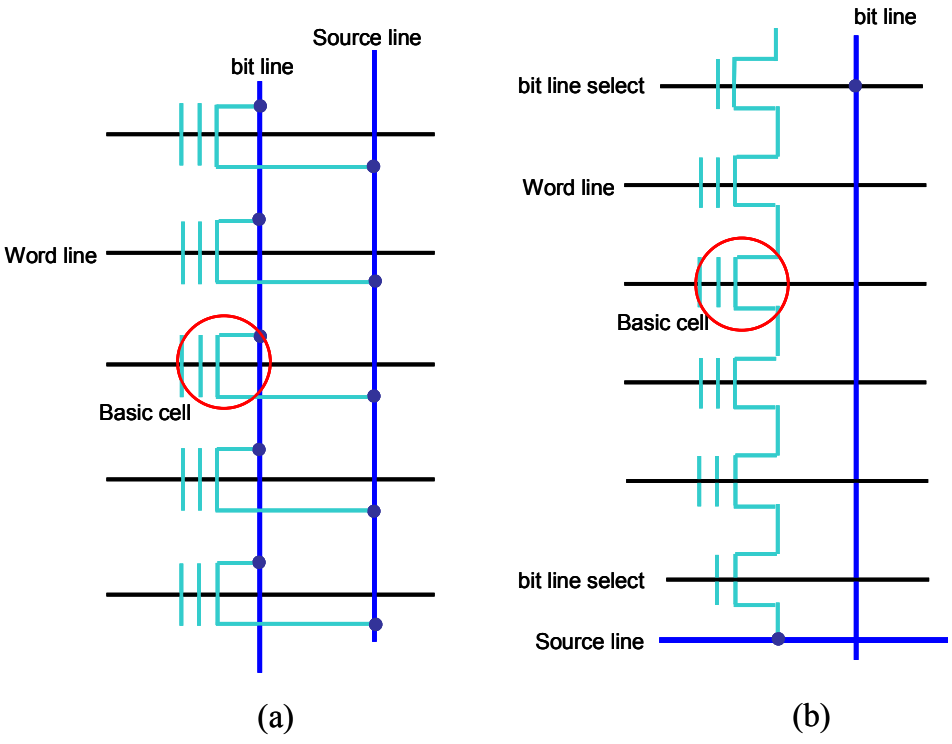


Fig. 1.6: Comparison of NOR and NAND flash architectures (a) NOR-type with shared bit line and source line. (b) NAND-type with a common bit line and a common source line, showing concise structure advantage.

Nonvolatile memory technology is going through a fast evolution amongst the semiconductor technologies in the last decade. The development of NAND-type flash memory technology has led to higher density memory designs and Giga-bits commercial products are also available now [1.13]. In the current NAND technology trend, 60-nm node has been completed in the R& D process [1.14]. However, the vast majority of flash memory devices will face significant challenges when the technology further scales down NAND-type flash memory to the 40-nm node [1.15].

It is important to scale the electrical oxide thickness (EOT) of the gate stack to achieve a small memory cell density and to reduce utilization power. A common issue when scaling the gate stack is the non-scalability of the tunnel oxide thickness. Tunnel oxide needs to be scaled to enable channel length scaling. There is always a tradeoff between the memory transient performance and the charge retention characteristics. A thin tunnel oxide is desirable to achieve increased memory speed. However, if the tunnel oxide is reduced to too much, it will degrade the memory nonvolatile properties to a large extent, since electrons stored in the floating gate are prone to tunnel out through the Frenkel-Poole mechanism when there is a defect in the tunnel oxide [1.16]. Moreover, due to the conductive poly-Si floating gate, a single leakage path resulting from endurance cycles would lead to loss of all stored charges. In order to maintain data retention for 10 years at an elevated temperature after 10^5-10^6 endurance cycling without program/read disturbance, the lower limit of tunnel oxide thickness must be greater than 7~8-nm for floating-gate devices due to stress induced low field tunneling.

One key issue is maintaining adequate coupling of the control gate to the floating gate. The control gate controls the channel indirectly. The control degree is indicated by the gate coupling ratio which is achieved by $C_{(\text{control gate to floating gate}}$

$_{\text{capacitance)}}/C_{\text{(total floating gate capacitance)}}$. The gate coupling ratio should be larger than 0.6. In the flash memory circuit, the control gate wraps around the FG to increase the CG to FG capacitance, and the thickness of IPD layer is approximately 20-nm. The reduced coupling ratio results in the increased programming and erasing voltage to maintain memory speed, which degrades the memory reliability. As the scale down proceeds, the space between neighboring FG becomes too narrow to be filled with two IPD layers and control gate poly-Si, as shown in Fig. 1.7. Therefore, at below 40-nm node, it is difficult to maintain the gate coupling ratio at a value of larger than 0.6.

During the rapid reduction of the memory array density, the poly-Si word lines would be patterned very closely to each other. Thus, the poly-Si word line would suffer serious capacitance interference from the adjacent word lines as illustrated in Fig. 1.7 (b), especially when the word line is very long. The cross coupling effect resulting from the capacitance interference can sometimes cause erroneous programming [1.17].

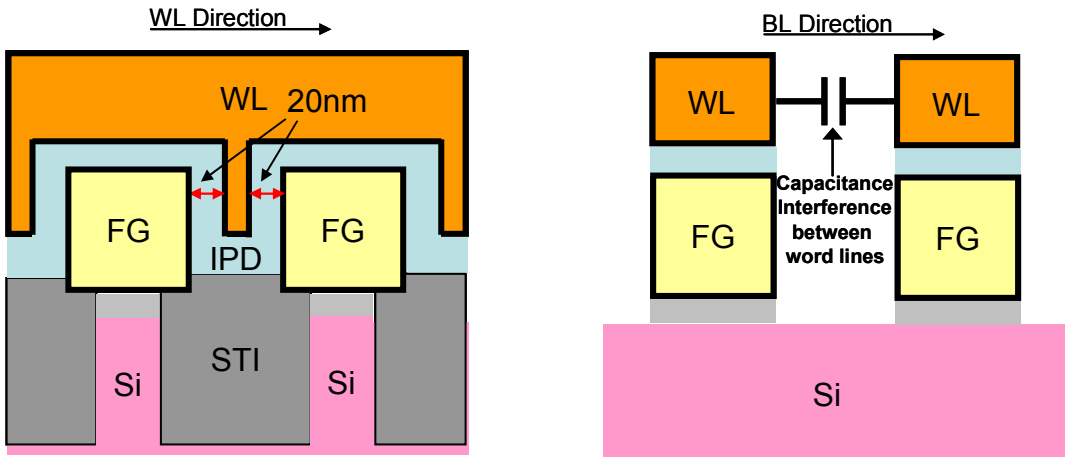


Fig. 1.7: Schematic cross section of a floating-gate cell in a (a) word line direction and (b) bit line direction. (a) Space between neighboring FG becomes too narrow to be filled with two IPD layers and control gate poly-Si. (b) V_{th} of an unselected cell can be programmed mistakenly due to the capacitance interference of the adjacent charge.

Based on the previous discussion, it is understandable why the nonvolatile memory research community is seeking innovative technologies to replace the current floating-gate flash memory. One of the main challenges is that planar floating gate memory device scaling is approaching the physical limits. A large amount of research work is being conducted on replacing the floating-gate flash with alternatives, such as FeRAM (Ferro-electric RAM), MRAM (Magnetic RAM) and phase change memory. It turns out that the discrete charge trapping type flash memory is the more suitable choice for nonvolatile memory applications compared to other candidates. Discrete charge trapping type flash memories have a similar device structure as the floating-gate flash, while the other candidates consume a larger area. A FeRAM cell consists of one transistor and one capacitor [1.18], while a MRAM is composed of a magnetic tunnel junction [1.19]. The continued scaling of stack capacitor is quite challenging for a FeRAM, and both FeRAM and MRAM have the drawback of sensitivity to IC processing temperatures and conditions [1.2]. Although phase change memory is considered a promising technology, the biggest challenge is the cost disadvantage and the contact between the hot phase-change region and the adjacent dielectric. [1.20].

On the other hand, the discrete charge trapping type flash memory cell such as SONOS, MONOS (Metal/AlO/SiN/Oxide/Si) [1.21] and nanocrystal memory [1.22] can achieve the highest chip density. It has been widely investigated recently to use charge trapping materials such as silicon nitride, high permittivity (high-κ) dielectric and nanocrystals to replace the conductive poly-Si floating gate. The discrete charge trapping nonvolatile memory has inherent advantages over traditional floating-gate devices, such as natural immunity to capacitance interference, simple process, lower applied voltage and robust tolerance to dielectric defects. In a floating-gate device, all

stored charges would be able to leak through a single defect chain in the thin tunnel oxide. The requirements for advanced technology nodes in NAND flash memory in the near term are summarized in Table 1.1, according to International Technology Roadmap of Semiconductors (ITRS) 2007 Edition [1.2]. As illustrated in the table, SONOS-type flash memory has an advantage in device density scaling.

Table 1.1: Flash Nonvolatile memory technology requirements

Year of Production	2009	2010	2011	2012
A. Floating gate NAND Flash				
Tunnel oxide thickness (nm)	6-7	6-7	6-7	6-7
Interpoly dielectric material	ONO	ONO	ONO	High- κ
Interpoly dielectric thickness	10-13	10-13	10-13	9-10
Highest P/E voltage (V)	15-17	15-17	15-17	15-17
B. Charge trapping NAND Flash (MANOS)				
Tunnel oxide thickness (nm)	3-4	3-4	3-4	3-4
Blocking oxide thickness (nm)	6-8	6-8	6-8	6-8
Trapping layer thickness(nm)	5-7	5-7	5-7	5-7
Highest P/E voltage (V)	15-17	15-17	15-17	15-17

However, the planar oxide-nitride-oxide (ONO) gate stack thickness is not easily scalable in the long term due to the data retention concerns. With a typical equivalent oxide thickness (EOT) of more than 10nm used in the device, it is easier to suffer the problems induced by severe short channel effects (SCE) in memory devices as compared to CMOS logic devices. The low on-off ratio and unwanted large subthreshold swing (SS) could also potentially trigger the reading error of a memory cell [1.23].

1.2 Scope of Project

From the view point of device structure, multi-gate device structures, that is, double-gate, trigate, Π-gate, Ω-gate and gate-all-around, have been explored extensively as planar device scaling approaches the end of the technology roadmap. The ultra thin body SOI field effect transistor structure and the multi-gate fully depleted FinFET structure have been proposed to suppress SCE for sub-100-nm CMOS technologies. The thin layer of silicon channel on SOI wafer can eliminate subsurface punch-through observed in bulk-Si devices, while the multi-gate structure features a narrow channel body which controls the channel potential better than the single gate structure. It reflects better SCE suppression effect by lower SS, DIBL and reduced V_{th} roll-off in device performance. Among the many types of multi-gate structure, devices with a nanowire channel are being widely investigated for their potential to advance the CMOS and nonvolatile memory technologies to extreme scaling limits.

In this dissertation, new methodologies which enable further scaling of nonvolatile memory devices will be evaluated. Due to the scaling advantage of discrete charge trapping type memory devices based on the aforementioned discussion, this study will focus on one type of discrete charge trapping memory devices, SONOS-type memory device, which is built on non-traditional device architecture. Considering the industrial compatibility, the top-down nanowire channel structure is fabricated and investigated by using CMOS technology. The novel nanowire device can potentially be a strategy to be exploited for the next technology node.

1.3 Organization of Thesis

This thesis addresses the issues of gate stack scaling and voltage scaling for future semiconductor nonvolatile memory device and discusses solutions to scale the discrete trapped charge-storage nonvolatile memory by using the advanced nanowire structure. The whole thesis consists of six chapters which have been arranged as follows.

Chapter 1 provides an introductory overview of the semiconductor nonvolatile memory technology and discusses the advantages of SONOS-type based discrete charge trapping nonvolatile memory. It also presents the current critical challenges existing in the memory device scaling progress.

Chapter 2 gives a detailed literature study on key findings in the earlier research work on SONOS-type memory, including the advantages and disadvantages considering its employment in next generation commercial memory products. Memory devices with different types of state-of-the-art non-traditional structures are also reviewed.

In chapter 3, we focus on the fabrication process of gate-all-around (GAA) Si-nanowire which is being integrated for the nonvolatile SONOS memory application. The introduced GAA structure, controlling the conducting channel from all directions, may provide the device much better gate controllability and hence influence the memory characteristics. The performance of high-speed SONOS device and the reason for enhancement will be examined in detail.

In Chapter 4, we introduce the nanowire structure into the poly-Si TFT memory as a promising candidate to be integrated in system-on-panel or system-on-chip applications. The GAA structure has an advantage of small channel cross section,

which may reduce the negative effect brought about by poly-Si grains and the boundaries between them. The performances will be presented and the causes for the performance improvement will also be discussed.

In Chapter 5, by integrating high-κ dielectric materials and metal gate electrode, a strategy of optimizing nanowire SONOS-type memory characteristics is illustrated. The application of high-κ materials and high work function metal gate electrode is used to replace the conventional materials, and have the potential to solve the limits existing in nitride-based SONOS devices. The performance enhancement in the new device will be investigated.

Last but not least, an overall conclusion is given in Chapter 6 to summarize the major results. Possible future work is also proposed in this chapter too.

Reference

[1.1] P. Cappelletti, in *IEDM Tech. Dig.,* 2004 Short Course: Memory devices.

[1.2] *International Technology Roadmap of Semiconductors (ITRS)*, Semiconductor Industry Association, Banjoes, CA, 2007.

[1.4] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, in *Proceedings of the IEEE,* vol. 91, no. 4, April, 2003.

[1.5] F. Masuoka, M. Asano, H. Ivahashi, T. Komuro, S. Tanaka, "A new flash E^2PROM cell using triple polysilicon technology", in *IEDM Tech. Dig.*, pp. 464-467, 1984.

[1.6] Jan Van. Houdt, "Flash memory: a challenged memory technology", in *IEEE International Conference on Integrated Circ. Design and Tech.,* pp. 1-4, 2006.

[1.7] Alan Niebel, "Business outlook for the non-volatile memory Market", in *IEEE Non-Volatile Semiconductor Memory Workshop,* pp. 6-7, 2006.

[1.8] William D. Brown, Joe E. Brewer, "Nonvolatile semiconductor memory technology: A comprehensive guide to understanding and using NVSM devices" *Wiley-IEEE Press*, pp. 5, October 1997.

[1.9] M. H. White, Y. L. Yang, A. Purwar, and M. L. French, "A low voltage SONOS nonvolatile semiconductor memory technology", in *IEEE Trans. on Components, Packing, and Manufacturing Tech.,* vol. 20, no. 2, pp. 190-195, 1997.

[1.10] Yan Ny Tan, Wai Kin Chim, Wee Kiong Choi, Moon Sig Joo, Tsu Hau Ng, and Byung Jin Cho, "High-k HfAlO charge trapping layer in SONOS-type nonvolatile memory device for high speed operation", in *IEDM Tech. Dig.*, pp. 889-892, 2004.

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

[1.12] Daniel C. Guterman, Isam H. Rimawi, Te-Long. Chiu, Richard D. Halvorson, and David J. McElroy, "An electrically alterable nonvolatile memory cell using a floating gate structure," *IEEE Tran. Electron Devices,* vol. 26, no. 4, p. 576-586, 1979.

[1.13] K. Kim, J. H. Choi, J. Choi, H.-S. Jeong, "The future prospect of nonvolatile memory", in *IEEE VLSI-TSA-TECH, Proceedings of Technical Papers*, pp. 88-94, 2005.

[1.14] J. H. Park, S. H. Hur, J. H. Lee, J. T. Park, J. S. Sel, J. W. Kim, S. B. Song, J. Y. Lee, J. H. Lee, S. J. Son, Y .S. Kim, M. C. Park, S. J. Chai, J. D. Choi, U. I. Chung, J. T. Moon, K. T. Kim, K. Kim, and B. I. Rye, "8Gb MLC (Multi-Level Cell) NAND flash memory using 63nm process technology", in *IEDM Tech. Dig.*, pp. 873-876, 2004.

[1.15] C. Y. Lu, K. Y. Hsieh, R. Liu, "Future challenges of flash memory technologies", in *Microelectronics Eng.*, pp. 283-286, 2009.

[1.16] C. S. Pan, K. Wu, D. Chin, G. Sery, J. Kiely, "High-temperature charge loss mechanism in a floating-gate EPROM with an oxide-nitride-oxide (ONO) interpoly stacked dielectric", in *IEEE Electron Device Lett.*, vol. 12, no. 9, pp. 506-509, 1991.

[1.17] She Min, "Semiconductor flash memroy scaling", PhD dissertation, 2003.

[1.18] R. E. Jones, Jr., P. D. Maniar, R. Moazzami, P. Zurcher, J. Z. Witowski, Y. T. Lii, P. Chu, and S. J. Gillespie, "Ferroelectric non-volatile memories for low-voltage, low-power applications", in *Thin Solid Films*, vol. 270, no. 1-2, pp. 584-588, 1995.

[1.19] S. Tehrani, J. M. Slaughter, E. Chen, M. Durlam, J. Shi, M. Deherren, "Progress and outlook for MRAM technology", in *IEEE Trans. on Magnetics,* vol. 35, no. 5, pp. 2814-2819, 1999.

[1.20] Http://en.wikipedia.org/wiki/Phase-change_memory.

[1.21] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEDM Tech. Dig.*, pp. 613-616, 2003.

[1.22] S. K. Samanta, P. K. Singh, Won Jong Yoo, Ganesh Samudra, Yee-Chia Yeo, L. K. Bera, and N. Balasubramanian, "Enhancement of memory window in short channel nonvolatile meomry devices using double layer tungsten nanocrystals", in *IEDM Tech. Dig.*, pp. 170-173, 2005.

[1.23] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS", *IEEE Circuits and Devices*, Vol.16, pp. 22-31, 2000.

Chapter 2

Literature Review

2.1 Introduction

Applications for nonvolatile memory have experienced exponential growth in recent years due to the need for low power solid-state storage and rapidly declining prices of nonvolatile memories. Conventional floating-gate flash faces significant challenges below 45-nm node. Charge trapping devices are proposed to continue the scaling of nonvolatile memories, and innovative gate stack engineering or three-dimensional structure integrated on charge trapping device may provide the ultimate solution.

2.2 Gate Stack Engineering

2.2.1 Nanocrystal Memory

Nanocrystal based charge trapping memory has shown scaling advantages over nitride based charge trapping memory, despite the fact that nanocrystal memory is relatively new and that nitride memory has been studied for quite a long time [2.1]. The ordering of the trap sites can be controlled in nanocrystal memory, while the ordering of trap sites is randomly distributed in nitride memory. Nanocrystal memory is also capable of multi-bit storage application by using Coulomb blockade. A

significant amount of research work has been done on nanocrystal based flash memory structures using semiconductor nanocrystals like Si and Ge. Several Si nanocrystal fabrication processes have been proposed, the most common of which are the ion implantation method and the direct growth method. Nanocrystal features, such as size, size distribution, lateral isolation and uniformity, generally require good process control in the fabrication process [2.2]. The ion implantation method uses an implantation step to implant excess silicon through the poly-silicon gate of the memory transistor and into the gate oxide layer; it will then be followed by a thermal anneal step to precipitate the Si atoms to form distinct Si nanocrystal dots [2.3]. In the direct growth method, low pressure chemical vapor deposition (LPCVD) is used to deposit small amounts of poly-Si on top of the tunnel oxide surface [2.4]. The deposition process is stopped before a continuous layer is formed, resulting in hemispherical Si islands on the tunnel oxide. A further thermal oxidation step can be added to enhance the lateral isolation between the nanocrystals [2.5]. The direct growth method is a more promising technique in terms of process compatibility. Compared with Si, Ge has a narrower bandgap and a similar electron affinity [2.6]. It has been reported that the Ge nanocrystal was integrated into flash memory devices by thermal oxidation techniques such as $Si_{1-x}Ge_x$ [2.7], co-sputtering and annealing [2.8]. Metal nanocrystals have also been proposed [2.9], aiming to engineer the potential well depth of the storage nodes and achieve longer data retention.

Nanocrystal memories have demonstrated robust memory behaviors. Threshold voltage (V_{th}) windows were increased from only several hundreds of milli-volts in the pioneering works [2.10], to a 6V memory window as demonstrated by Singh *et al.* in Au nanocrystal devices [2.11]. Excellent memory endurance behavior on nanocrystal memories was also observed, displaying only limited V_{th} window

closure after more than 10^5 write/erase cycles [2.12]. Retention data that have been measured also suggest better retention behavior in nanocrystal memory than floating-gate memory devices [2.12] [2.13]. However, as statistical information has yet to demonstrate that nanocrystal memory meets the requirements for nonvolatile memory, the claims that nanocrystal memory is nonvolatile remains unverified.

2.2.2 Bandgap Engineering Memory

An ideal nonvolatile memory device should possess long data retention as well as fast operation speed. However, there is a tradeoff between these two characteristics: the adoption of a thin tunnel oxide to achieve high memory speed would sacrifice the data retention time, while growing a thick tunnel oxide to ensure long data retention time will significantly reduce the memory speed. The reduction is specially obvious for the erase speed, as the hole experiences a higher barrier height at Si/SiO_2 interface than the electron does. BE-SONOS (Bandgap Engineered SONOS) with bandgap engineering of the tunnel oxide shows the potential to maximally suppress the tradeoff and improve both characteristics [2.14]. In the bandgap engineering, the traditional SiO_2 tunnel oxide is replaced by a triple layer of $SiO_2/SiN/SiO_2$ (ONO), since ONO demonstrates non-trapping ability with an ultra-thin thickness. As shown in Fig. 2.1, due to the lower hole-tunneling barrier in SiN (~1.9 eV) as compared to SiO_2 (~4.5 eV), a large electron or hole tunneling current could tunnel through the ultra-thin ONO tunneling layer under a high electric field. On the other hand, in the case of data retention, hole direct tunneling is blocked by the total thickness of the ONO barrier under a low electric field. Another bandgap engineering example, as demonstrated by Y. Q. Wang *et al.* in 2006, shows that a simple Si_3N_4/SiO_2 double tunneling layer for MONOS type memory [2.15] based on the concept of layered tunnel barrier [2.16] is beneficial for enhancing hole tunneling efficiency.

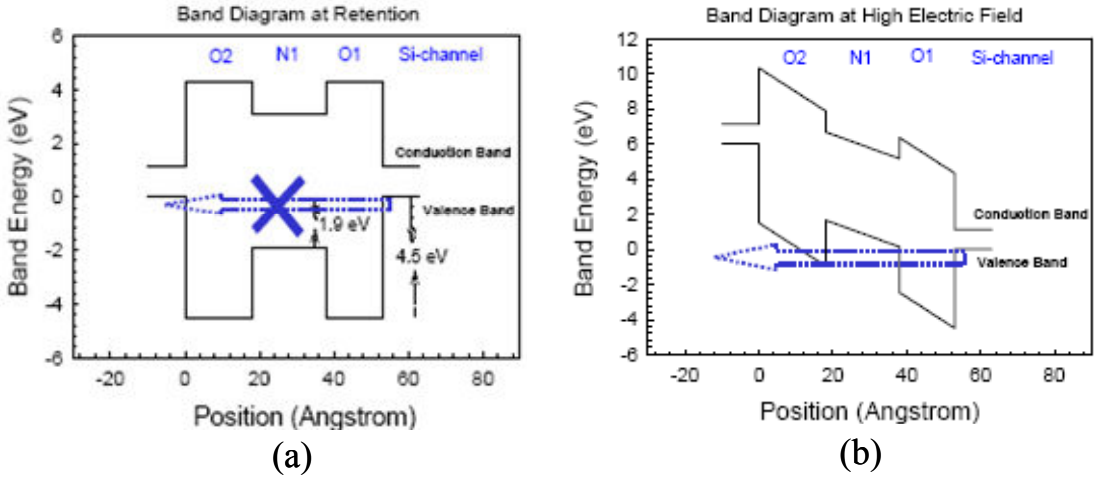


Fig. 2.1: The band diagram of BE-SONOS under different electric fields, showing only the tunnel oxide part. (a) At retention mode, direct tunneling is prohibited since the barrier width carriers experience is the whole physical thickness of ONO layer. (b) At program/erase mode or under high electric field, carriers could only see the thickness of a thin layer of oxide; hence the speed can be significantly enhanced due to reduced barrier width.

2.2.3 High-κ-based MONOS Memory

SONOS devices have received increasing interest due to their better endurance, smaller chip size and lower power consumption than floating-gate devices. However, charge retention and erase speed remain major challenges for SONOS to overcome in order to enter into the commercial mass storage flash memory market. Since the erase speed for SONOS type memory is determined by the competition between the direct tunneling current of holes through the tunnel oxide and the unwanted F-N tunneling current of electrons through the blocking oxide, the tunnel oxide thickness must be very thin in order to establish a stable erase operation. The charge loss through a very thin tunnel oxide is also a significant challenge to overcome in order to secure data storage. Therefore, aggressive scaling of CMOS logic and mass storage devices has triggered the investigation on high permittivity dielectric materials (high-κ) as an alternative to SiO_2 and Si_3N_4.

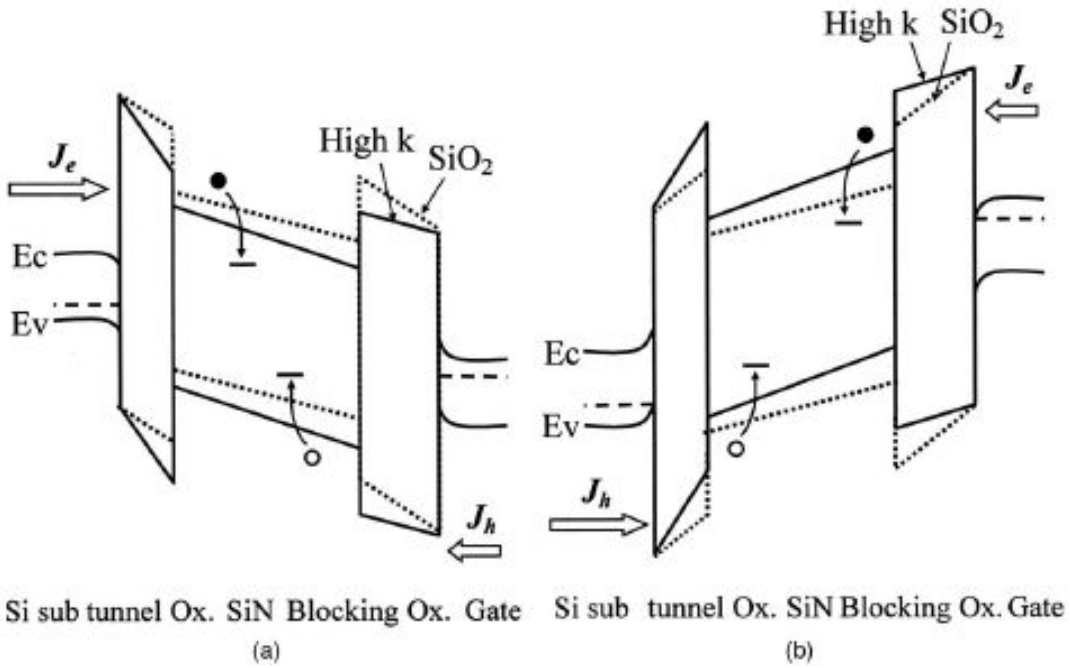


Fig. 2.2: Schematic band diagram for the memory device using nitride (dashed line) and high-κ (solid line) as blocking layer under electric field at the progam mode (a) and erase mode (b). The electric field across the blocking oxide is decreased and transferred to tunnel oxide, hence the carriers tunneling efficiency is improved compared with ONO device.

In a conventional SONOS device, the electric field in blocking oxide (SiO_2) is generally about two times larger than in Si_3N_4 for devices with comparable thickness of nitride and top oxide, due to the lower permittivity of SiO_2 ($\kappa_{SiO_2} \sim 3.9$) as compared to Si_3N_4 ($\kappa_{Si_3N_4} \sim 7.5$). A substantial part of applied voltage would drop across the top oxide and cause the unwanted electron back tunneling. By replacing SiO_2 with high-κ dielectric as the blocking oxide, the electric field across the blocking oxide would be reduced, while the electric field across the tunnel oxide would be proportionally increased, as shown in Fig. 2.2 [2.17]. A thicker tunnel oxide can thus reduce the program and erase voltage for SONOS devices without losing any operation speed. The simulation results given by Gritsenko *et al.* shows that by using Al_2O_3 or ZrO_2 as the top dielectric, the programming voltage amplitude decreases or

the programming time reduces from 1 ms to 10 μs [2.18]. C. H. Lee *et al.* developed a TANOS device with the structure of SiO_2/SiN/Al_2O_3/TaN and the memory cell demonstrated efficient erase with a thick tunnel oxide [2.19]. It also demonstrated that the TaN metal gate blocks the gate electron current more effectively than the poly-Si gate, resulting in a significant decrease of the saturation level of V_{th} at the erasing state. The data retention of the TANOS device was investigated by J. S. Lee *et al.* and was found to be excellent [2.20]. With an original memory window of 6 V, such a device is expected to take more than 40 years to have less than 0.5 V charge loss at 85°C. A 4Gb TAHOS NAND flash memory has been successfully developed with the 63-nm design rule [2.21]. There are also remarkable research works on SOHOS structure devices with high-κ material as the charge storage layer. SOHOS structure memory devices can minimize the over-erase phenomenon found in nitride based memory device, due to the difference in band offset and the degree of crystallization of the high-κ material [2.22].

2.3 Novel Structure Nonvolatile Memory Devices

The scaling of planar bulk devices is facing significant challenges due to limited electrostatic control of the channel potential and leading to SCEs: 1) V_{th} roll-off, large DIBL and poor subthreshold slope; 2) increased gate and junction leakage; and 3) increased source/drain resistance. N. Collaert *et al.* predicted that planar devices will not make any further scaling after the 32-nm node [2.23]. The research community is pursuing new routes for further scaling. One route is by employing new materials discussed previously – 45-nm CMOS with high-κ and metal gate electrode has been put into production by Intel in 2007 [2.24]. The other route is by adopting a new device structure, with the objective of improving electrostatics of a single MOSFET or memory cell, and accommodating the integration needs of new materials.

Fully depleted ultra-thin body (UTB) silicon-on-insulator (SOI) MOS device structure consisting of a very thin ($t_{Si} \leq$ 10-nm) has drawn much interest a decade ago. K. Ishii *et al.* presented suppressed SCE on a 40-nm gate length thin channel (11-nm) MOSFET [2.25]. M. K. Kim *et al.* demonstrated good memory behavior as well as excellent SCE immunity on a minimum 30-nm gate length SONOS charge trapping memory [2.26]. The ratio of the channel length to the channel thickness typically requires $\geq$ 3 to maintain full substrate depletion under gate control [2.27], which means a Si channel as thin as 4-nm is required to scale device to the 22-nm node. However, at a gate length of 25-nm, the thickness-induced V_{th} variation will start to happen when the body thickness reduced to 7-nm [2.28]. Too much thin channel body also raises issues of mobility degradation and high source and drain resistance in logic devices.

A variety of multiple-gate nontraditional structures have been proposed in recent years. The increased number of gates provides improved electrostatic control to the channel, which alleviates the stringent requirement of body thickness scaling in UTB SOI structure, and the multi-gate device structure can be used in conjunction with UTB SOI to further improve the gate control of the channel. Fig. 2.3 shows the evolution of the multi-gate structure schematically. Double gate (DG) SOI devices have two symmetrical gates electrically interconnected, and can be achieved with many techniques such as the fin process [2.29], tunnel epitaxy [2.30], etc. Numerical simulation results suggest that the DG SONOS device can be scaled down to 50-nm, and in addition to the improved scalability of the memory cell, the two conducting channels provide a 2-bit per cell storage [2.31]. Tri-gate structures [2.32], such as the Π-gate [2.33] and omega-gate structures [2.34], feature similar processing technique like the bulk-Si. The multiple gates control the fin on three geometric sides,

suggesting that excellent short-channel device performance can be achieved. Xuan *et al.* demonstrated the first SONOS device with three conducting channel surfaces, and the SONOS can be scaled down to 40-nm gate length with good subthreshold properties [2.35]. A window of 2 V is achieved with a 5 ms/10V program pulse and a 10 ms/-11V erase pulse in such a device. M. Specht *et al.* demonstrated a tri-gate SONOS memory cell which features a 20-nm gate length and a sub 10-nm fin width [2.36]. Such a device could be operated at high speed and in multi-level mode despite the thick tunnel oxide it contains.

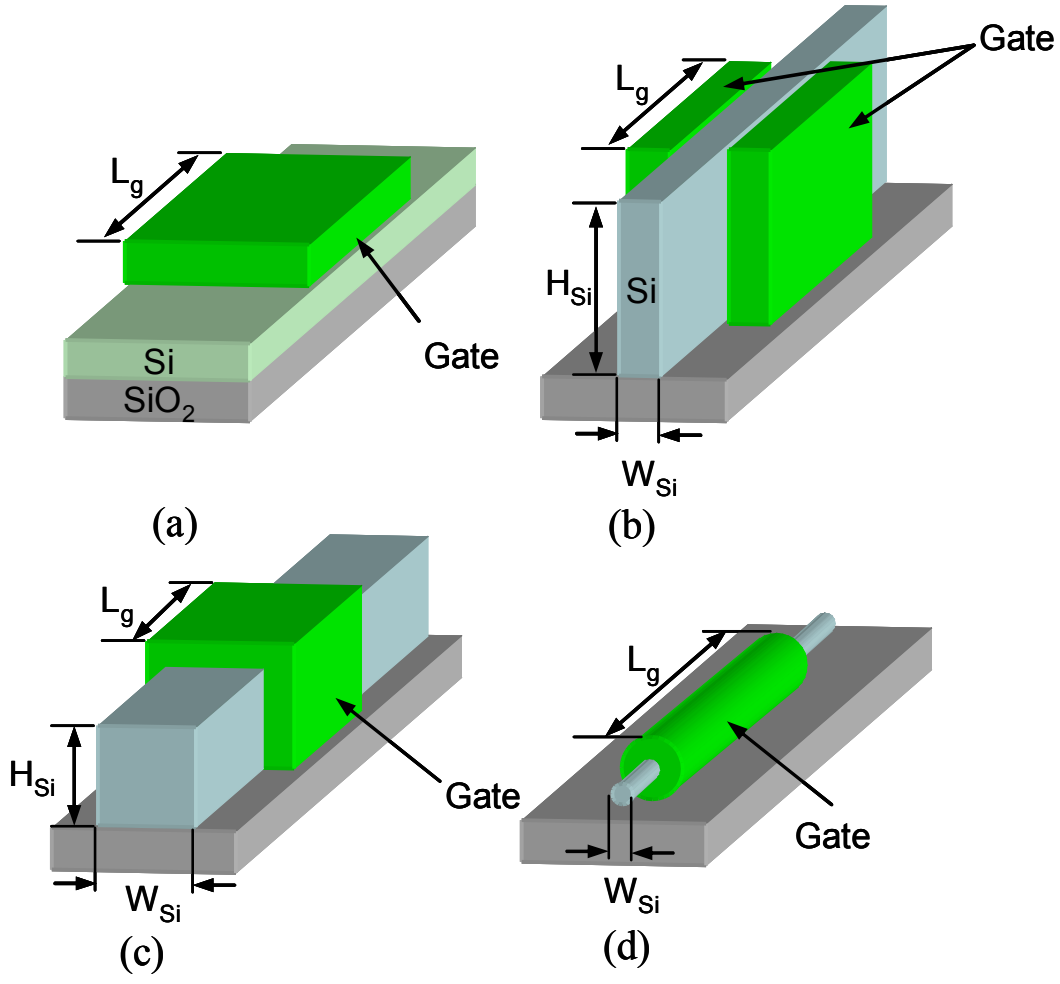


Fig. 2.3: Progression of device structure from a single-gated planer on SOI to a fully GAA nanowire channel, with the number of gates increasing. (a) Single-gate structure. (b) Double gate structure, with a tall fin and two symmetrical gates electrically interconnected. (c) Tri-gate or FinFET structure, where gate electrode controls the channel on three surfaces. (d) Gate-all-around structure with a nanowire channel.

The gate-all-around (GAA) structure is the most promising among all emerging device structures for the ultimate device scaling. As the name suggests, the GAA structure possesses a gate which is fully surrounds the channel body and thus provides the best electrostatic control [2.37]. The reduction in channel width and thickness further increases the effectiveness of the gate control. As the simulation indicates, the gate length can be scaled to 5-nm with the scaling of the channel to nanowire [2.38]. Therefore, the GAA structure, combined with the ultra-narrow body or nanowire body, shows the most promise as a candidate for extreme device scaling.

2.4 Si Nanowire Technology

2.4.1 Bottom-up Approach

One dimensional semiconductor materials, such as nanotubes and nanowires, offer the possibility of ballistic transport since scattering is much suppressed [2.39]. There is an increasing interest in the synthesis and assembly of nanomaterials which may serve as the building blocks for the next generation of electronic and optical devices. Most of the reported works use the bottom-up approach to grow nanowires, and gold is the most popular catalyst for synthesizing Si nanowires via the vapor-liquid-solid (VLS) mechanism, which was first reported and studied in detail by Wagner *et al.* [2.40]. In the VLS mechanism, Si nanowire can be synthesized by coating a Si substrate with a thin gold film and annealing the substrate to obtain Au-Si catalyst particles. When a vapor phase Si precursor is introduced at a substrate temperature, which is the Au-Si mixture melting temperature, the catalyst particles become supersaturated with Si. Crystalline Si nanowires will then nucleate and grow. Since the Au-Si catalyst particle located at the tip of the nanowire is in a liquid state during growth, the method was named the VLS method. The liquid particle serves as

a preferential site for absorption of Si from the vapor phase precursor.

However, the catalyst metal gold could trap electrons and holes in Si, which pose a serious contamination problem for CMOS technology. Although there are some reports on the use of other metal catalysts for Si nanowire growth, such as Pt, Fe, and Ga, either the growth temperatures are too high, or the metals are not compatible in modern semiconductor manufacturing processes. The control of diameter and length thus is critical in the nanowire growth process as the electrical and optical properties of Si nanowires are strongly dependent on their size. The "bottom up" synthesized nanowires typically show poor control of size and dimension, and the positions of the nanowires are hard to predict as the nanowires are usually distributed on the wafer surface arbitrarily [2.41]. More complicated processes are hence required for further device fabrication on such nanowires. Other bottom-up techniques reported for nanowire preparation include laser ablation [2.42], liquid suspension, fluid flow, etc. All these techniques have an issue with the control of repeatability and scalability. In addition, these techniques have very low throughput that do not result in significant economic advantages.

2.4.2 Top-down Approach

The bottom-up approach creates integration issues that have hindered the widespread application of Si nanowires. On the other hand, the top-down approach could provide better control by using precise lithography and etching processes. The top-down approach is CMOS compatible; hence nanowires fabricated using the top-down process could be more easily embedded into circuits and other micro-systems. A top-down Si nanowire field effect transistor with channel width down to 20-nm has been fabricated using electron-beam lithography on SOI wafer [2.43]. The results show that the carrier mobility of the nanowire device is 1.3 to 2.2 times higher than

that of the reference device with larger dimension, which could be attributed to the possible suppression of inter-valley phonon scattering. S. D. Suk *et al.* has demonstrated a twin Si nanowire transistor on bulk wafer using a complicated self-aligned damascene-gate process [2.44], with the damascene-gate process schematics shown in Fig. 2.4. No roll-off of V_{th}, small SS and Drain Induced Barrier Lowering (DIBL) can be observed down to 30-nm gate length for both n- and p-type MOSFETs with a 10-nm diameter. GAA Si nanowire SONOS nonvolatile memory has been developed by the same research group using the same process technique [2.45]. By using the CHE and hot hole injection mechanisms for program and erase, program speed of 1 μs at V_d=2 V, V_g=6 V and erase speed of 1 ms at V_d=4.5 V, V_g=-6 V could be achieved with a 2~3-nm nanowire and a 30-nm gate.

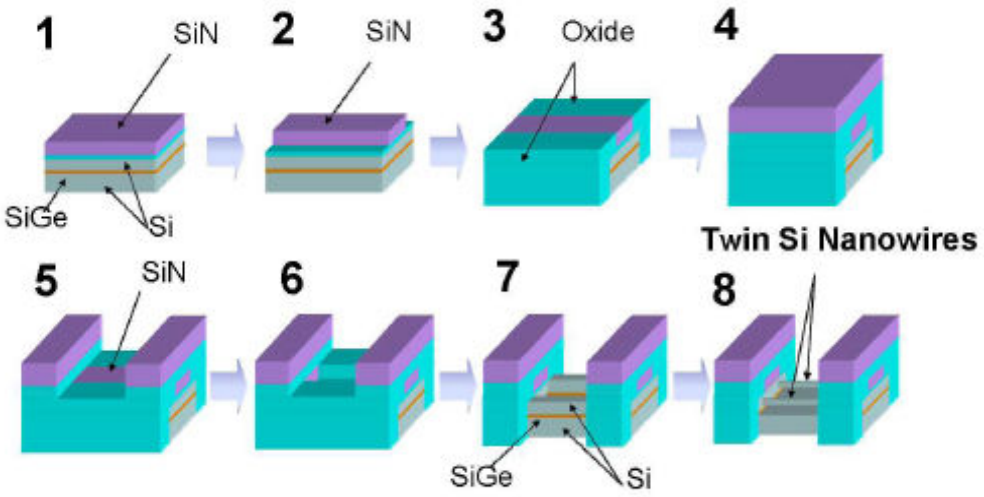


Fig. 2.4: Progress flow of the damascene-gate nanowire device fabrication used by Samsung's group (1) SiGe / Si growth and shallow trench isolation (STI) (2) hard mask SiN trimming (3) oxide fill in STI and CMP (4) damascene gate stack deposition (5-6) 1st and 2nd damascene gate etch (7) field oxide recess (8) SiGe removal and H_2 anneal and (9) gate oxide and gate material deposition.

Nanowire MOSFET and nonvolatile memory devices with GAA device structure are able to achieve excellent performance in a scaled size and be fabricated

for a mass production purpose due to the modern process technology used. However, the complicated damascene-gate technique that Samsung used to fabricate nanowires is difficult to adopt in the industrial environment. Hence, it is necessary to find a new technique to realize the potential of nanowires.

Previously, some research work reported that thermal oxidation of small silicon wires shows a self-limiting effect due to high stresses [4.46]. Sub-10-nm diameter vertical Si nanowire column was reported using a combination of electron beam lithography, reactive ion etching and self-limiting thermal oxidation. Below a critical temperature of 950°C, thermal oxidation proceeds for a while, but then a self-limiting effect starts, in which after a few hours of oxidation the Si reduction and the oxide growth become extremely slow compared to the planar oxide growth rate. Moreover, they appear to saturate to different asymptotic values depending on the starting Si volume. The time to achieve the saturation decreases with increasing temperature. The limiting core (nanowire) diameter depends on the outside oxide shell thickness, and the temperature dependence of limiting core diameters becomes weaker at higher temperature [4.47]. Liu *et al.* postulated that this self-limiting phenomenon is due to the increasing activation energy of oxidant diffusivity in a highly stressed oxide, as newly formed SiO_2 has to expand more to accommodate a given increase in volume for Si-O reactions. The smaller the core, the greater the oxide layer has to expand, which causes a greater stress and retards the process of further oxidation.

Reference

[2.1] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS", in *IEEE Circuits and Devices*, Vol.16, pp. 22-31, 2000.

[2.2] Jan De Blauwe, "Nanocrystal nonvolatile memory devices", in *IEEE Trans. on Nanotechnology*, vol. 1, pp. 72-77, 2002.

[2.3] P. Dimitrakis, E. Kapetanakis, D. Tsoukalas, D. Slarlatos, C. Bonafos, G. B. Asssayag, and P. Normand, "Silicon nanocrystal memory devices obtained by ultra-low-energy ion-beam synthesis", in *Solid-State Electronics*, vol. 48, pp. 1511-1517, 2004.

[2.4] V. Ioannou-Sougleridis, and A. G. Nassiopoulou, "Investigation of charging phenomena in silicon nanocrystal metal-oxide-semiconductor capacitors using ramp current-voltage measurements", in *J. Appl. Phys.*, vol. 94, pp. 4084-4087, 2003.

[2.5] D. N. Kouvatsos, V. Ioannou-Sougleridis, and A. G. Nassiopoulou, "Charging effects in silicon nanocrystals within SiO_2 layers, fabricated by chemical vapor deposition, oxidation, and annealing", in *Appl. Phys. Lett.*, vol. 82, pp. 397, 2003.

[2.6] S. M. Sze, *Physics of Semiconductor Devices*, 2^{nd} ed., New York: Wiley, pp. 849-850, 1981.

[2.7] Y.-C. King, T.-J King, and C. Hu, "MOS memory using germanium nanocrystals formed by thermal oxidation of $Si_{1-x}Ge_x$", in *IEDM Tech. Dig.*, pp. 115-118, 1998.

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

nanocrystals embedded in HfAlO high-k tunneling and control oxides: Device fabrication and electrical perofrmance" in *IEEE Tran. Electron Devices*, vol. 51, no. 11, pp. 1840-1848, 2004.

[2.9] J. J. Lee, and D.-L. Kwong, "Metal nanocrystal memory with high-k tunneling barrier for improved data retention" in *IEEE Tran. Electron Devices*, vol. 52, no. 4, pp. 507-511, 2005.

[2.10] I. Crupi, D. Corso, S. Lombardo, C. Gerardi, G. Ammendola, G. Nicotra, C. Spinella, E. Rimini, and M. Melanotte, "Memory effects in MOS devices based on Si quantum dots" in *Materials Science and Engineering*, vol. 23, pp. 33-36, 2003.

[2.11] P. K. Singh, K. K. Singh, R. Hofmann, K. Armstrong, N. Krishna, and S. Mahapatra, "Au nanocrystal flash memory reliability and failure analysis", in *IEEE International Symp. Phys. and Failure Analysis of Integrated Circuit,* pp. 7-11, 2008.

[2.12] H. I. Hanafi, S. Tiwari, I. Khan, "Fast and long retention-time nano-crystal memory" in *IEEE Tran. Electron Devices*, vol. 43, no. 9, pp. 1553-1558, 1996.

[2.13] T. Z. Lu, M. Alexe, R. Scholz, V. Talelaev, and M. Zacharias, "Multilevel charge storage in silicon nanocrystal multilayers", in *Appl. Phys. Lett.*, vol. 87, pp. 202110, 2005.

[2.14] H.-T. Lue, S.-Y. Wang, E.-K. Lai, Y.-H. Shih, S.-C. Lai, L.-W. Yang, K.-C. Chen, J. Ku, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "BE-SONOS: A bandgap engineered SONOS with excellent performance and reliability", in *IEDM Tech. Dig.*, pp. 547-550, 2005.

[2.15] Y. Q. Wang, D. Y. Gao, W. S. Hwang, C. Shen, G. Samudra, Y.-C. Yeo, and W. J. Yoo, "Fast erasing and highly reliable MONOS type memory with HfO_2

high-k trapping layer and Si_3N_4/SiO_2 tunneling stack", in *IEDM Tech. Dig.*, pp. 971-974, 2006.

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

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

[2.18] V. A. Gritsenko, K. A. Nasyrov, Y. N. Novikow, A. L. Aseev, S. Y. Yoon, J.-W. Lee, E.-H. Lee, and C. W. Kim, "A new low voltage fast SONOS memory with high-k dielectric", in *Solid-State Electronics*, vol. 47, pp. 1651-1656, 2003.

[2.19] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEDM Tech. Dig.*, pp. 613-616, 2003.

[2.20] J.-S. Lee, C.-S. Kang, Y.-C. Shin, C.-H. Lee, K.-T. Park, J.-S. Sel, V. Kim, B.-I. Choi, J.-S. Sim, J. Choi, and K. Kim, "Data retention characteristics of nitride-based charge trap memory devices with high-k dielectrics and high-work-function metal gates for multi-gigabit flash memory", in *Japanese J. of App. Phys.*, vol. 45, no. 4B, pp. 3213-3216, 2006.

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

[2.22] 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" in *IEEE Tran. Electron Devices*, vol. 51, no. 7, pp. 1143-1147, 2004.

[2.23] N. Collaert, A. Keersgieter, A. Dixit, I. Ferain, L. Lai, D. Lenoble, A. Mercha, A. Nackaerts, B. Pawlak, R. Rooyackers, T. Schulz, K. San, N. Son, M. Dal, P. Verheyen, K. Arnim, L. Witters, K. Meyer, S. Biesemans, M. Jurczak, "Multi-gate device for the 32 nm technology node and beyond", in *Solid-State Electronics*, vol. 52, pp. 1291-1296, 2008.

[2.24] M. T. Bohr, R. S. Chau, T. Ghani, and K. Mistry, "The high-k solution" in *IEEE Spectrum*, vol. 44, no. 10, pp. 29-35, 2007.

[2.25] K. Ishii, E. Suzuki, S. Kanemaru, T. Maeda, K. Nagai, and T. Sekigawa, "Suppressed threshold voltage roll-off characteristic of 40 nm gate length ultrathin SOI MOSFET", in *IEEE Electron Lett.*, vol. 34, no. 21, pp. 2069-2070, 1998.

[2.26] M. K. Kim, S. D. Chae, H. S. Chae, J. H. Kim, Y. S. Jeong, L. W. Lee, H. Silva, S. Tiwari, C. W. Kim, "Ultrashort SONOS memories", in *IEEE Trans. on Nanotechnology*, vol. 3, no. 4, pp. 417-424, 2004.

[2.27] T. Skotnicki, J. A. Hutchby, T.-J. King, H.-S. P. Wong, and R. Boeuf, "The end of CMOS scaling" in *IEEE Circuit and Device*, vol. 21, no. 1, pp. 16-26, 2005.

[2.28] K. Takeuchi, R. Koh, and T. Mogami, "A study of the threshold voltage variation for ultra-small bulk and SOI CMOS" in *IEEE Tran. Electron Devices*, vol. 48, no. 9, pp. 1995-2001, 2001.

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

[2.30] H.-S. P. Wong, K. K. Chan, and Y. Taur, "Self-aligned (top and bottom)

double-gate MOSFET with a 25 nm thick silicon channel", in *IEDM Tech. Dig.*, pp. 427-430, 1997.

[2.31] H. Y. Kam, Y. M. Tsz, M. Chan, "A highly scalable 2-bit asymmetric double-gate MOSFET nonvolatile memory", in *IEEE Electron Devices and Solid-State Circ.*, pp. 59-62, 2003.

[2.32] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted tri-gate CMOS transistors", in *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263-265, 2003.

[2.33] J.-T. Park, J.-P. Colinge, and C. H. Diaz, "Pi-gate SOI MOSFET", in *IEEE Electron Device Lett.*, vol. 22, no. 8, pp. 405-405, 2001.

[2.34] F.-L. Yang, H.-Y. Chen, F.-C. Cheng, C.-C. Huang, C.-Y. Chang, H.-K. Chiu, C.-C. Lee, H.-T. Huang, C.-J. Chen, H.-J. Tao, Y.-C. Yeo, M.-S. Liang and C. Hu, "25 nm CMOS omega FETs", in *IEDM Tech. Dig.*, pp. 255-258, 2002.

[2.35] P. Xuan, M. She, B. Hartenech, A. Liddle, J. Bokor, and T.-J. King, "FinFET SONOS flash memory for embedded applications", in *IEDM Tech. Dig.*, pp. 609-612, 2003.

[2.36] M. Specht, U. Dorda, L. Dreeskornfeld, J. Kretz, F. Hofmann, M. Stadele R. J. Luyken, W. Rosner, H. Reisinger, E. Landgraf, T. Schulz, J. Hartwich, R. Kommling, and L. Risch, "20 nm tri-gate SONOS memory cells with multi-level operation", in *IEDM Tech. Dig.*, pp. 1083-1085, 2004.

[2.37] J. P. Colinge, M. H. Gao, A. R.-Rodriguez, H. Maes, and C. Claeys, "Silicon-on-insulator 'gate-all-around device'", in *IEDM Tech. Dig.*, pp. 595-598, 1990.

[2.38] E. Gnani, S. Reggiani, M. Rudan, and G. Baccarani, "Design considerations and comparative investigation of ultra-thin SOI, double-gate and cylindrical

nanowire FETs", in *IEEE ESSDERC*, pp. 371-374, 2006.

[2.39] H.-S. P. Wong, "Beyond the conventional transistor", in *Solid-State Electronics*, vol. 49, pp. 755-762, 2005.

[2.40] R. S. Wagner, and W. C. Ellis, "Vapor-liquid-solid mechanism of single crystal growth", in *Appl. Phys. Lett.*, vol. 4, pp. 89-90, 1964.

[2.41] Y. Cui, L. J. Lauhon, M. S. Gudiksen, J. Wang, "Diameter-controlled synthesis of single-crystal silicon nanowires", in *Appl. Phys. Lett.*, vol. 78, pp. 2214-2216, 2001.

[2.42] N. Wang, Y. H. Tang, Y. F. Zhang, C. S. Lee, and S. T. Lee, "Nucleation and growth of Si nanowires from silicon oxide", in *Phys. Review B*, vol. 58, no. 24, pp. R16024- R16026, 1998.

[2.43] S.-M. Koo, A. Fujiwara, J.-P. Han, E. M. Vogel, C. A. Richter, and J. E. Bonevich, "High inversion current in silicon nanowire field effect transistors", in *Nano Lett.*, vol. 3, no. 11, pp. 2197-2201, 2004.

[2.44] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, and J.-B. Park, "High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET) fabrication on bulk Si wafer, characteristics, and raliability", in *IEDM Tech. Dig.*, pp. 717-720, 2005.

[2.45] S. D. Suk, K. H. Yeo, K. H. Cho, M. Li, Y. Y. Yeoh, K. H. Hong, S. H. Kim, Y. H. Koh, S. G. Jung, W. J. Jang, D. W. Kim, D. G. Park, and B. I. Ryu, "Gate-all-around twin silicon nanowire SONOS memory", in *VLSI Technology Symp.*, pp. 142-143, 2007.

[2.46] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, and R. F. W. Pease, "Self-limiting oxidation of Si nanowires", in *J .Vac. Sci. Technology B*, vol. 11,

pp. 2532-2537, 1993.

[2.47] H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson, and R. F. W. Pease, "Self-limiting oxidation for fabricating sub-5 nm silicon nanowire", in *Appl. Phys. Lett.*, vol. 64, no. 11, pp. 1383-1385, 1993.

Chapter 3

Gate-All-Around Si Nanowire SONOS Memory

3.1 Introduction

The market for nonvolatile memory (NVM) has been growing rapidly in recent years, due to the increasing demand for portable electronic devices. The commercial flash memory, which currently uses a polysilicon (poly-Si) floating gate as the charge storage material, is facing increasingly rigorous challenges in the course of scaling beyond the 45-nm technology node, due to short channel effects (SCEs) which significantly reduced coupling ratio and serious gate interference [3.1]. The polysilicon-oxide-nitride-oxide-silicon (SONOS) structure has been widely investigated recently, in which discrete charge trapping materials such as silicon nitride, high permittivity (high-κ) dielectrics or nanocrystals are used to replace the conductive poly-Si floating gate. This is because the SONOS-type nonvolatile memory has inherent advantages, such as simpler process, lower programming voltages and more robust tolerances to defects in the thin tunnel oxide, over traditional floating gate devices [3.2-3.7]. However, the oxide-nitride-oxide (ONO) gate stack thickness is difficult to scale due to data retention concerns when the tunnel oxide reaches only 3~4-nm thickness. It is also more difficult to avoid concomitant

problems due to severe SCEs in memory devices that use a typical gate stack of more than 10-nm equivalent oxide thickness (EOT) in a single SONOS-type memory cell, as compared to CMOS transistor logic devices. The low on-off ratio and large unwanted SS can easily trigger reading error of a memory cell [3.8].

Various non-conventional device structures have been explored as planar device scaling approaches the end of the technology roadmap. Extensive research work is being performed on the multi-gate structural memory which features a narrow channel body and gate electrode controlling channel electrostatic potential from more than one direction. In double-gate or tri-gate memory devices, enhanced gate controllability enables memory devices to achieve higher read current, with suppressed SCEs and improved memory speed [3.9, 3.10]. There is increasing interest in the gate-all-around (GAA) nanowire (NW)-FET in the past few years because of its potential to advance the CMOS technology to extreme scaling limits. Best gate control is reflected by nearly ideal SS, negligible DIBL and V_{th} roll-off. Although most of the research work conducted on Si nanowires focuses on the bottom-up technique to fabricate the nanowires [3.11], the top-down technique shows better control and repeatability over the nanowire fabrication process. Suk *et al.* has demonstrated high-performance GAA nanowire MOSFETs on bulk silicon wafer by a CMOS compatible top-down process, which makes it more attractive for industrial applications [3.12]. Nanowire MOSFETs using the self-aligned damascene-gate process displays high drive current and well controlled SCEs in 30-nm gate length and 10-nm diameter nanowire channel. In 2006, the 3-nm diameter cylindrical nanowire CMOS proposed by Singh *et al.* exhibited near-ideal SS and low DIBL with 9-nm gate oxide, suggesting the benefit of relaxing the requirement for ultra-thin gate oxide thickness [3.13]. It indicates that the GAA nanowire structure could supply the

most effective gate control and overcome the negative impact from the physically thick gate dielectric. Hence GAA Si nanowire architecture shows promise that it can be integrated in the SONOS-type nonvolatile memory application, where thick tunnel oxide is favored for long retention time. In 2007, Suk *et al.* demonstrated GAA lateral twin nanowire SONOS with excellent memory characteristics by using the same self-aligned damascene-gate process developed in 2005 [3.14].

In this chapter, using a completely CMOS compatible process technology, nanowire-based GAA SONOS nonvolatile memory was fabricated and studied. The vertically aligned cylindrical twin silicon nanowire channel with only 5-nm diameter was formed on silicon-on-insulator (SOI) wafer with a self-limiting dry oxidation scheme. The two vertically stacked nanowire channels have the advantage of space saving per memory cell as compared to the published lateral two wires [3.14]. The nanowire SONOS device exhibits evident improvements in low voltage programming and fast programming/erasing speed (P/E) in comparison to the planar device. The performance enhancement mechanism will be explained by device modeling, investigating the electron energy distribution, potential energy profile and electric field along each layer surrounding the nanowire channel.

3.2 Nanowire and Nanowire Memory Device Fabrication

A standard CMOS fabrication process was used in this experiment. In our nanowire fabrication process, photolithography and self-limiting oxidation are the key processes to fabricate a “top-down” nanowire. The 8-inch starting p-type SOI wafer specifications are as follows: top silicon layer of 120-nm and buried oxide (BOX) of 150-nm with initial doping concentration of boron ~10^{15} cm^{-3}, as shown in Fig. 3.1 (a). The Si fins were first patterned on the wafer by using the KrF alternating phase shift mask lithography. The thinnest width of the photoresist fin pattern could be further

narrowed down to 40-nm with an additional resist trimming process.

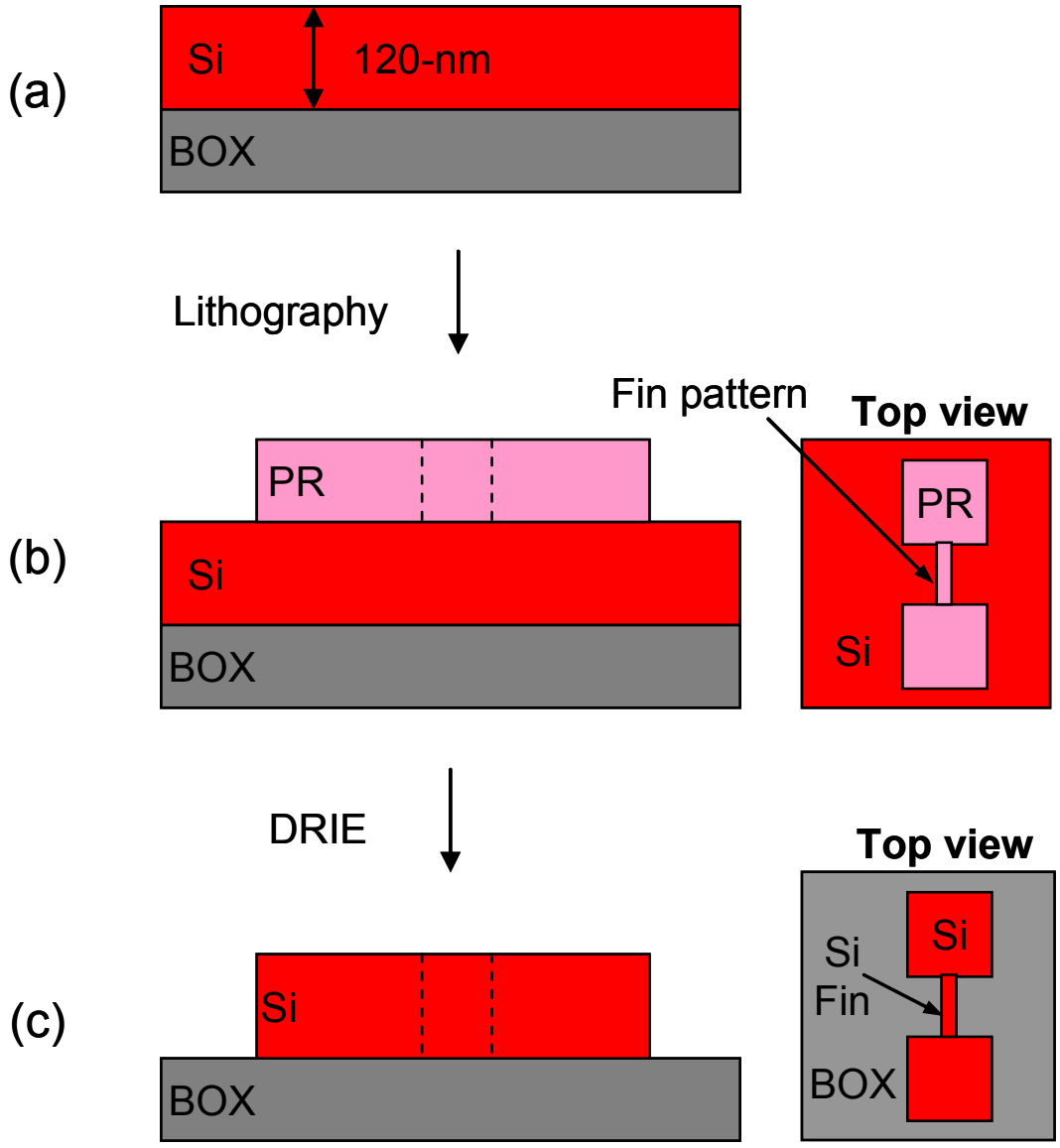


Fig. 3.1: Schematic of the Si fin fabrication process. (a) Starting wafer consisting of 120-nm Si layer (b) has undergone lithography (c) and the DRIE process to achieve the Si fin. The smallest defined fin width is 40-nm after the resist trimming process. The dash lines in (b) and (c) indicate the unseen part of Si fin.

Fig. 3.1 (b) shows the starting wafer with lithographic patterning of the active area. Each device contains a thin fin connected to a wider source and drain (S/D) extension region. The trimming process was followed by deep reactive ion etching

(DRIE) of the top Si layer down to the BOX to transfer the photoresist pattern onto the Si layer and to achieve a Si fin with a height of 120-nm. After photoresist stripping by oxygen and wafer cleaning by SPM (H_2SO_4 : H_2O_2 mixture in a ratio of 1 : 4), the wafer was then sent into furnace and subjected to dry oxidation for four hours at 875°C to form the nanowire. Fin oxidation time is dependent on the fin width: for instance, oxidation time is reduced if the fin width is less than 40-nm. The Si fin was almost fully oxidized in this step, leaving two vertical and self-aligned Si nanowires. The formation of vertically stacked nanowires is actually the consequence of retarded oxidation rate at the curved Si surfaces [3.15], which is, the top and bottom sides of the fin. In this experiment, we noticed that the oxidation process could be reliable in fabricating two vertical staked wires out of a fin if the top Si thickness of the SOI wafer is larger than 70-nm. In Fig. 3.2 (a), the transmission electron microscopy (TEM) image shows a sample prepared before oxidation is fulfilled, as the fin has not been totally consumed in the middle to become two nanowires. Such phenomenon indicates that two nanowires could be achieved from fin oxidation when the aspect ratio of the Si fin cross section is more than 7:4; otherwise there would only be one single nanowire formed as the cross section of the fin is close to a square. The purpose of simultaneously forming two nanowire channels in one memory cell is to obtain higher read current as compared to single nanowire channel memory devices. After the oxidation step, the Si nanowires were then released by removing the external thermal SiO_2 shell using 1:25 diluted HF solution. In Fig. 3.2 (b), the scanning electron microscopy (SEM) image shows an example of the vertically stacked nanowires. Fig. 3.3 shows the CMOS compatible nanowire fabrication process using self-limiting oxidation. The Si nanowire is fabricated using self-limiting oxidation at a temperature below 950°C.

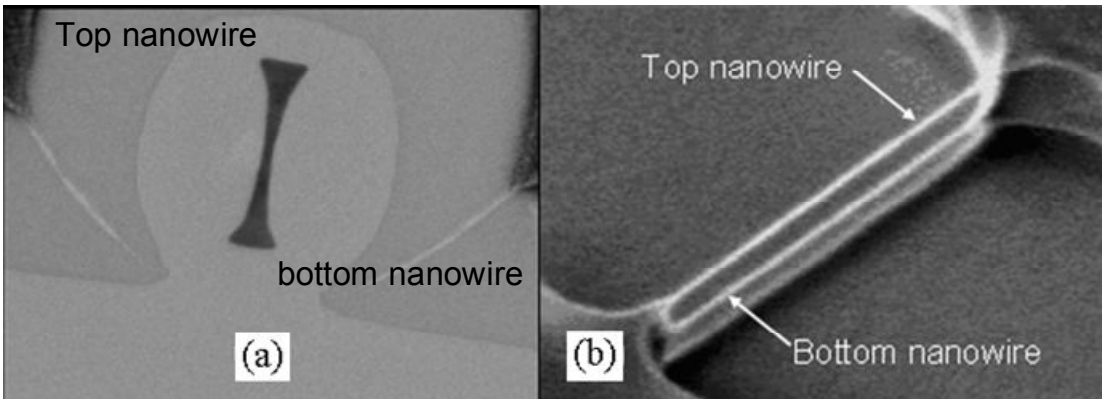


Fig. 3.2: (a) Sample prepared before the oxidation process is finished, TEM image shows the phenomenon where two nanowires are formed if the Si fin has a high aspect ratio. (b) Tilted SEM image shows the two vertically staked Si nanowire channels connecting S/D pads.

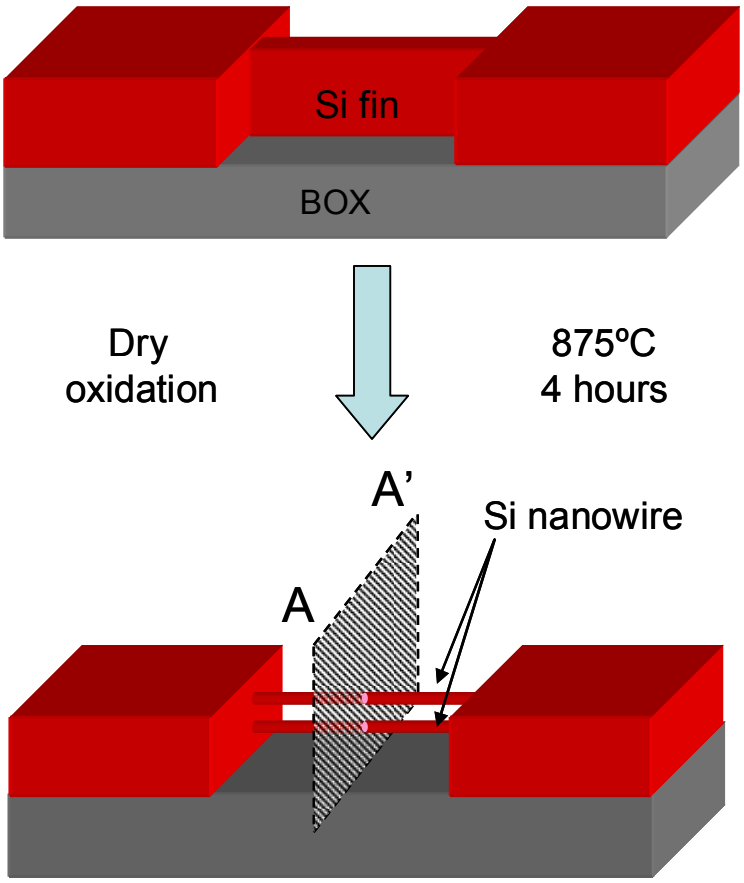


Fig. 3.3: Schematics of the Si nanowire fabrication process. The Si nanowire is fabricated using self-limiting oxidation at a temperature below 950°C. A high aspect ratio of the fin ensures there will be two vertically stacked nanowires. The vertically stacked nanowires have the advantage of saving space as compared to laterally packed nanowires.

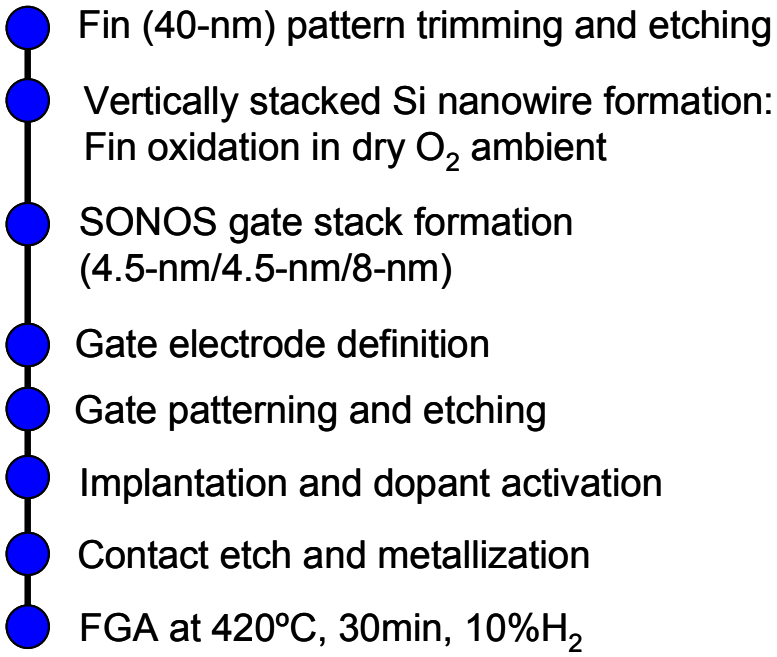
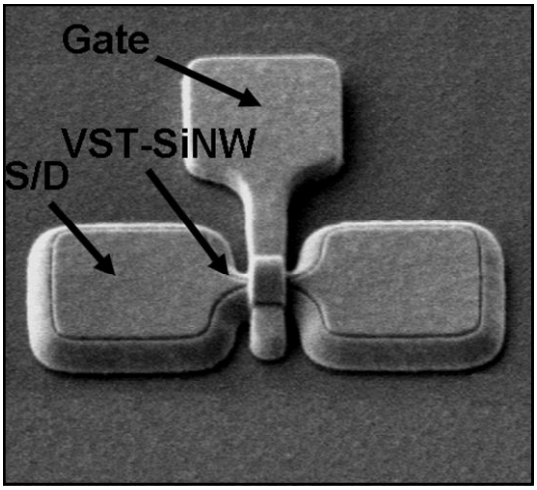


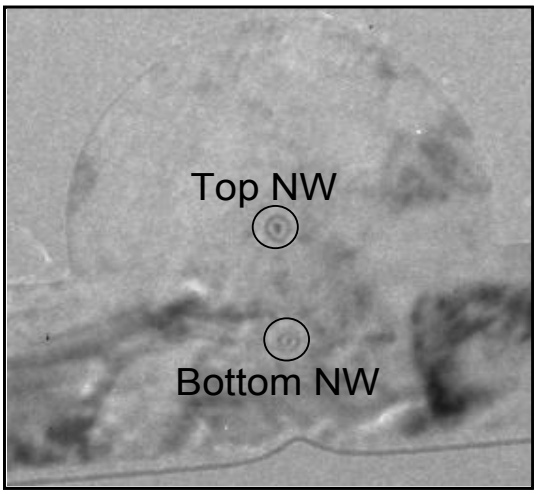
Fig. 3.4: Process flow depicting the formation of vertically stacked twin Si nanowire and GAA Si nanowire nonvolatile memory device.

As illustrated in the process flow shown in Fig. 3.4, the SiO_2, Si_3N_4 and SiO_2 (ONO) tunneling layer, charge trapping layer and blocking layer with thickness of 4.5-nm/4.5-nm/8-nm were deposited sequentially after the formation of the nanowire, at a temperature of 650ºC for the oxides and 720ºC for the nitride, in the low pressure chemical vapor deposition (LPCVD). Tetraethyl orthosilicate (TEOS oxide) was chosen as the tunneling oxide instead of the thermally grown oxide in order to avoid Si consumption in the small nanowire channel during oxidation. Poly-Si of 130-nm was deposited as the gate material in the LPCVD at a temperature of 550ºC. After patterning and etching the poly-Si gate to form the gate electrode, the wafer was implanted with 4×10^{15} cm^{-2} of phosphorus at 30-keV for S/D and gate, followed by dopant activation at 1000ºC for 5 seconds in N_2. TaN/Al/TaN layers were deposited as interconnecting metal lines after the device was passivated by 400-nm SiO_2, which

was deposited in the plasma-enhanced chemical vapor deposition (PECVD) machine tool. Subsequent annealing in forming gas (10% H_2 in N_2) at 420 ºC for 30 minutes was to complete the process. The titled SEM image of the nanowire memory device, taken before the passivation SiO_2 was deposited, is shown in Fig. 3.5 (a). Fig. 3.5 (b) shows the cross-sectional TEM image of twin nanowires surrounded by ONO stack and poly-Si gate. In the high resolution TEM image shown in Fig. 3.6, a nanowire with 5-nm diameter and ONO layers could be seen.



(a)



(b)

Fig. 3.5: (a) The titled SEM image of actual device taken before the passivation SiO_2 was deposited. (b)Vertically stacked two Si nanowire (VST-SiNW as indicated) channels were surrounded by ONO and poly-Si gate electrode.

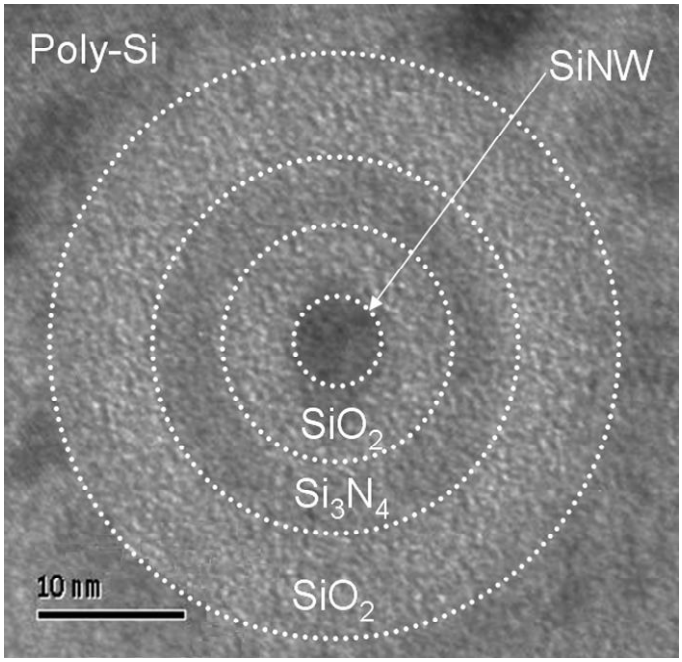


Fig. 3.6: The high resolution TEM picture shows the cross section part of one of the two nanowire channels of a fully processed nanowire SONOS device. The Si nanowire with surrounding ONO layers followed by poly-Si gate could be seen clearly. Diameter of wire is about 5-nm and the thickness of each layer of ONO gate stack is 4.5-nm/4.5-nm/8-nm.

3.3 Results and Discussion

In this study, a nanowire device with a gate length of 850-nm is used for the analysis of electrical and memory transient characteristics. The transfer characteristics of nanowire SONOS memory cells, with gate length of 850-nm and diameter of 5-nm, is first characterized and shown in Fig. 3.7. The SS is 75-mV/dec and the DIBL is 15-mV/V. Such subthreshold properties are comparable to those of high performance MOSFET devices, demonstrating the potential of GAA nanowire devices to be employed in the next generation of MOSFET or memory devices. The subthreshold property of such devices is superior considering that the EOT can be as thick as 15-nm. The excellent SCEs immunity in nanowire memory devices can be attributed to

the outstanding gate controllability of the GAA structure.

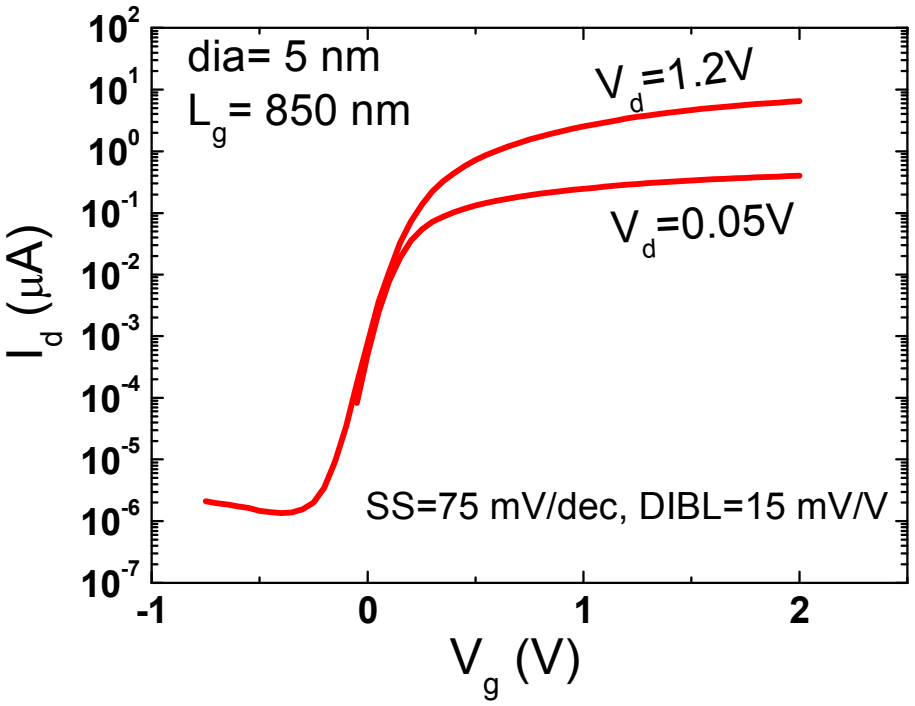


Fig. 3.7: The transfer characteristics of GAA nanowire memory devices with nanowire diameter of 5-nm and gate length of 850-nm shows good electrostatic behavior.

In addition to the advantages of the GAA structure with nanowire channel in device electrostatic property, the GAA device in this work also exhibits superiority in memory transient characteristics. GAA nanowire memory with 5-nm nanowire channel and planar channel memory device perform P/E operations using the Fowler-Nordheim (F-N) tunneling mechanism, in which positive and negative pulses were applied onto the gate while keeping the source and drain grounded. Their threshold voltage shifts (ΔV_{th}) were compared at ± 11 V pulse stresses in Fig. 3.8. The planar SONOS device with 5-μm channel width has the same gate length and ONO thickness as the nanowire SONOS device, as we are only focusing on what the different gate structures induce. Although the gate dielectrics have the same thickness, the GAA nanowire device exhibits visibly fast P/E speed. The ΔV_{th} of the Si nanowire SONOS

device could reach 2.61 V by applying a pulse of only 1 μs, while the ΔV_{th} in the planar device could hardly be observed within the same pulse duration. The 11 V stress magnitude is too small to open the memory window for such a thick EOT used in the conventional planar SONOS device. However, the threshold voltage shift starts from 1 μs and extremely large ΔV_{th} could be observed in SONOS devices with the GAA nanowire structure.

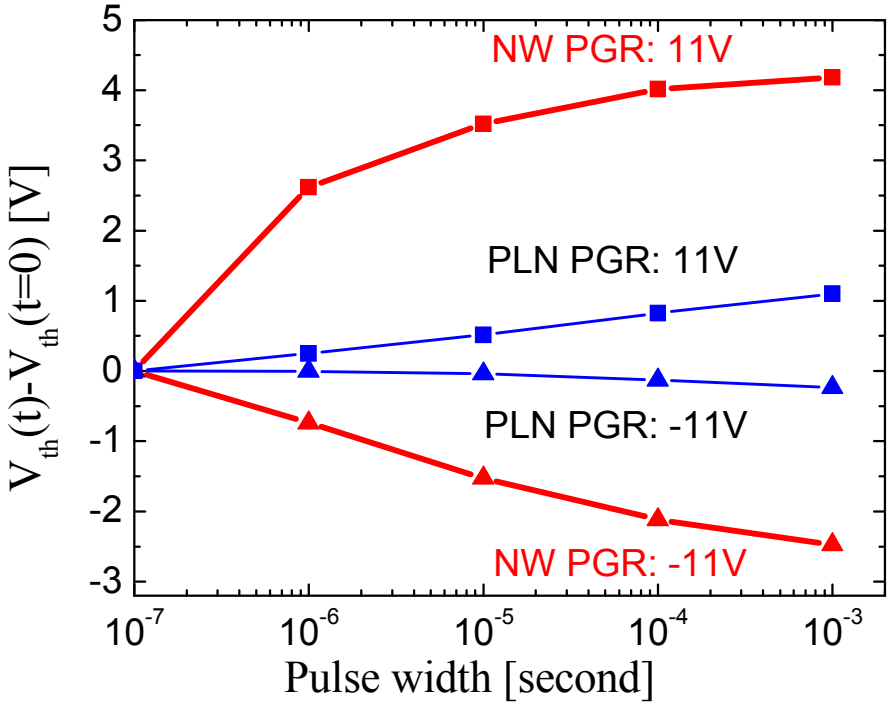


Fig. 3.8: Transient memory characteristics of nanowire SONOS (NW) device and planar (PLN) control device at ± 11 V pulse stresses. Devices with a gate length of L_g 850-nm are used for characterization. Channel widths are 5-nm for NW device (diameter) and 5-μm for PLN device respectively.

For the reasons above, it would be interesting to investigate the reason for speed enhancement related to such cylindrical GAA nanowire structure. The Schrödinger-Poisson equations, which take into account quantum-mechanical confinement effects, were consistently solved within a cylindrical domain through device simulations. Taking the programming behavior of the devices as an example,

the simulated electric field across the gate dielectrics of the GAA nanowire and planar SONOS devices with pulse stress at 11 V, just before programming commences (no charge inside the nitride), are plotted in Fig. 3.9 (a). The thickness of the ONO layers of the two devices in the simulation is identical to that used in the experiment. As shown in Fig. 3.9 (a), the electric field at the Si-SiO_2 interface is nearly three times higher for the Si nanowire SONOS device as compared to the planar SONOS device with the same dielectric. The electric field of the GAA structure is not constant in the oxide region but shows the $1/x$ trend given by the cylindrical symmetry. As the same voltage is applied on the two devices, the electric field across the blocking oxide, which is far away from the cylinder axis, would be much smaller than in the planar device. The larger electric field across the tunnel oxide in the nanowire SONOS device enhances the carrier-injection probability, increasing both P/E speeds. A prevalent method used in planar structural devices is reducing the thickness of the tunnel oxide to obtain large vertical electric field and high memory speed. However, it is evident that the scaling of planar devices has come to the end of the roadmap as excessive reduction of the tunnel oxide is detrimental to retention-time. High-κ dielectric materials have been used as blocking oxide [3.16] in order to reduce the electric field in the blocking oxide and increase it in the tunnel oxide. With a cylindrical geometry, a high electric field could be generated at the silicon-oxide interface even when a thick tunnel oxide is used, thus electron tunneling could be greatly enhanced across the oxide without degrading the retention time under reading conditions. Gate electron tunneling is prevented during erasing as the electric field is also reduced at the blocking oxide. Hence the GAA structure plays the same role as the high-k does.

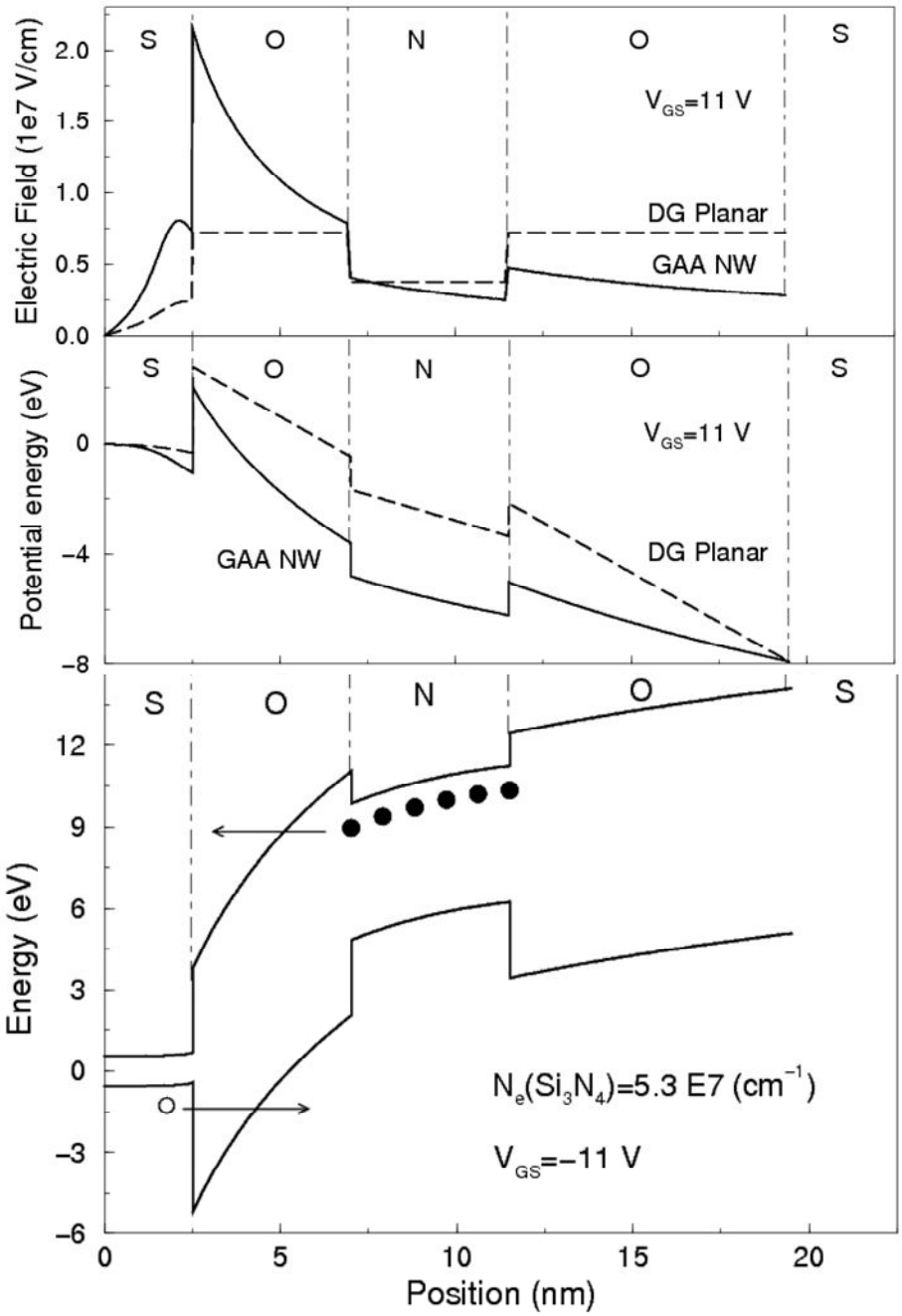


Fig. 3.9: Simulated (a) electric field distribution and (b) potential energy profile of GAA nanowire and planar structures at gate stress 11 V. For the GAA device (solid line), the electric field at the Si-SiO_2 interface is almost three times larger as compared to the planar device (dashed line). The effective barrier width of the GAA nanowire device is also less than half the oxide physical thickness. (c) Potential energy profile at V_{GS} = –11 V prior to the start of the erasing process for the GAA nanowire device. An electron concentration of $5.3x10^7$ cm^{-1}, corresponding to a shift of 2.6V in V_{th}, is assumed in the nitride layer. The barrier width for holes tunneling from the channel to the oxide is reduced due to the cylindrical architecture, thus increasing the erase speed.

Fig. 3.9 (b) also shows the potential profile at 11 V gate pulse. As the voltage almost drops across the entire gate oxide, electrons in the channel are mainly localized at energy levels close to 0 eV, corresponding to the value of the Fermi level. The barrier height remains the same since it is determined by the band offset between SiO_2 and Si; however, the barrier width experienced by electrons tunneling at zero energy decreases to nearly 2-nm from the physical 4.5-nm. For the planar device with the same gate voltage, the tunneling length remains equal to the oxide thickness at 4.5-nm, which is more than twice as large when compared to the GAA structure. Since the tunneling probability decreases exponentially with the tunneling length, the cylindrical NVM cell exhibits a much larger programming efficiency than that of the planar device.

Fig. 3.9 (c) shows the potential energy profile at V_{GS}= -11 V prior to the start of the erasing process. An electron concentration of 5.3×10^{7} cm^{-1}, corresponding to a shift of 2.6 V in V_{th}, is assumed in the nitride layer. Similar to the case of programming, the barrier width for holes tunneling from the channel to the oxide is reduced due to the cylindrical architecture, thus increasing the erase speed.

Fig. 3.10 and Fig. 3.11 present typical P/E characteristics of the two nanowire devices with nanowire diameters of 8-nm and 5-nm, and at various gate stressing voltages as a function of pulse width. Positive pulses with heights ranging from 6 to 11V were applied for programming, while negative pulses with heights ranging from – 6 V to – 11 V were applied to the gate for erasing. Fast P/E speed could be observed for the same device when it was measured for I_d-V_g transfer characteristics. Despite a thick EOT of 15-nm, the ΔV_{th} in the nanowire SONOS device could reach as high as 1.13 V and 2.61 V at a stress pulse of 11 V and with a programming time of only 1 μs.

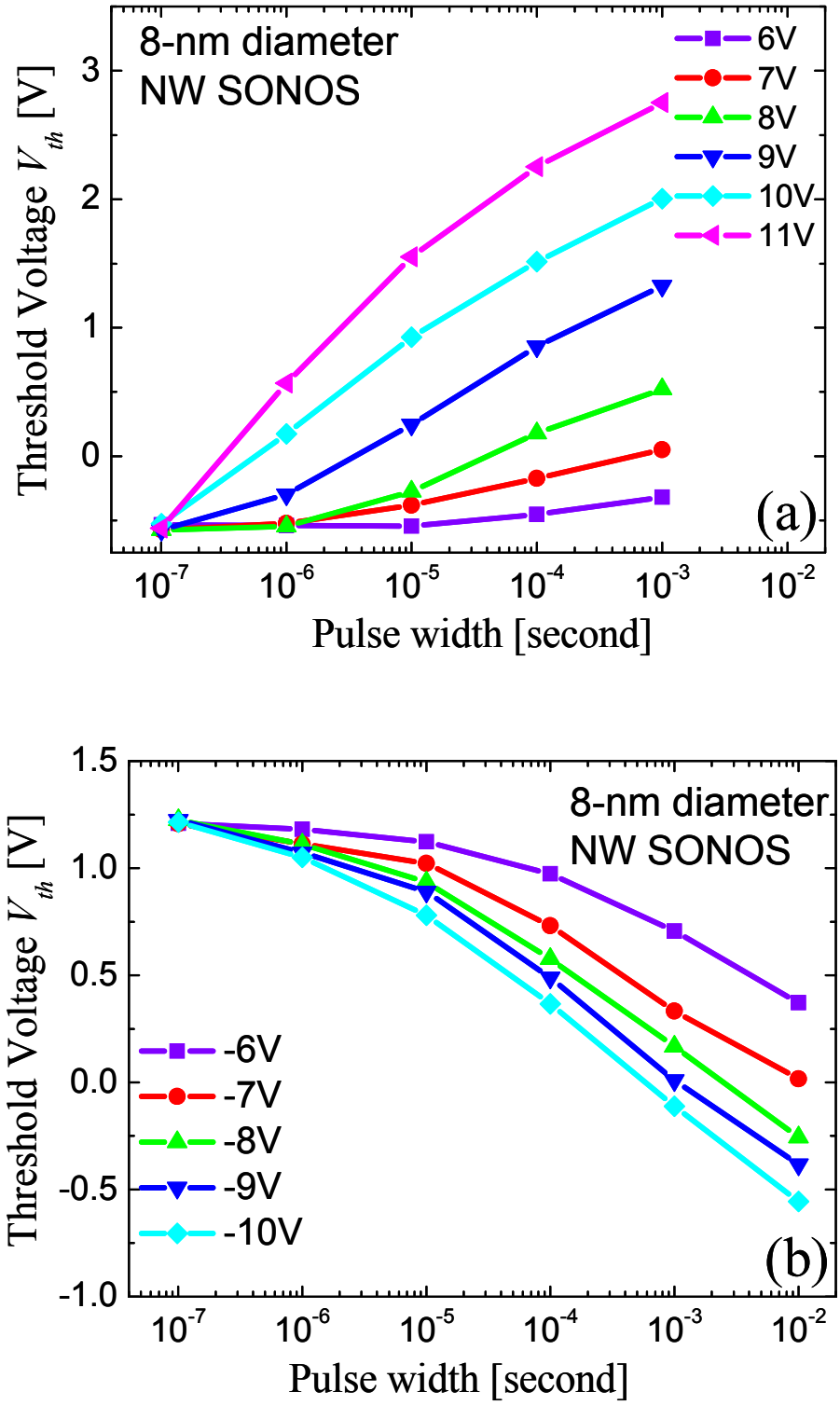


Fig. 3.10: (a) Programming and (b) erasing characteristics of Si nanowire SONOS cell (NW diameter ~ 8-nm) of gate length of 850-nm. Based on the programming characteristics, this device exhibits a V_{th} shift of 1.13 V in 1 μs using a pulse of +11 V on the gate.

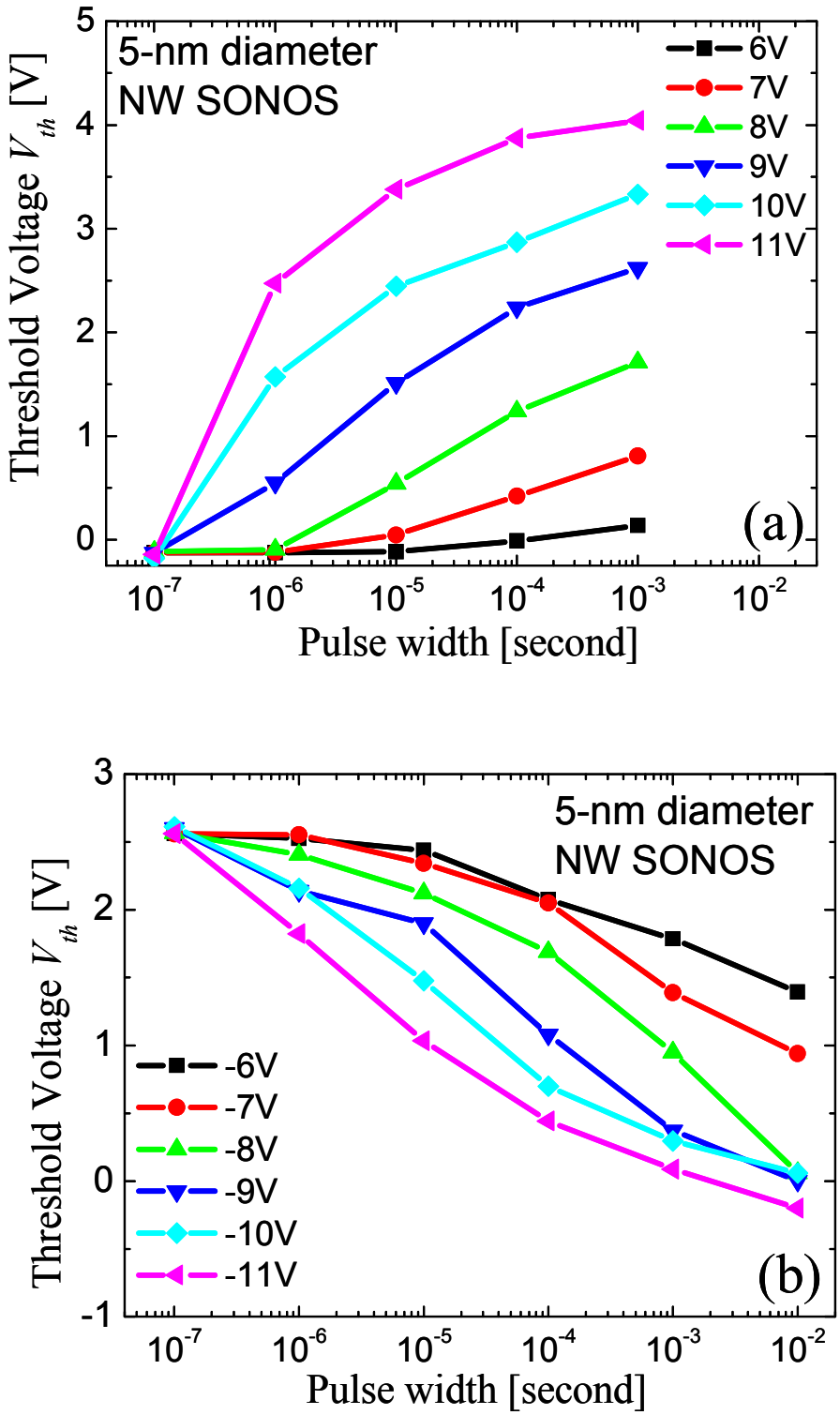


Fig. 3.11: (a) Programming and (b) erasing characteristics of Si nanowire SONOS cell (NW diameter ~ 5-nm) of gate length of 850-nm. Based on the programming characteristics, this device exhibits a V_{th} shift of 2.61 V in 1 μs using a pulse of +11 V on the gate.

The fast program performance could be attributed to the much enhanced electric field across the tunnel oxide as discussed before. However, the P/E speed of the nanowire device with 8-nm diameter is found to be slower as compared to the device with 5-nm nanowire diameter, which shows the evident dependence of memory speed on nanowire diameter in GAA structure devices. It should be noted that the nanowire of 8-nm diameter is obtained from a fin which has a width larger than 40-nm before thermal oxidation, rather than from a variation of the nanowire fabrication process. A previous research work reported that the vertical electric field is in a proportional relationship of $2/\{d_{nw}\ln(1+2t_{ox}/d_{nw})\}$, suggesting that the electric field at the tunnel oxide is larger for memory devices with smaller wire diameter [3.14]. The same ΔV_{th} during erase takes more than 1 ms at -11 V gate pulse. The slower erase speed can be attributed to an energy barrier as large as 4.6 eV at the Si/SiO_2 interface for hole as compared to a 3.2 eV energy barrier for electrons, which influences the F-N tunneling current density due to their exponential term [3.17].

The investigation of room-temperature retention property for the nanowire devices is performed by programming both devices at 8 V for 100 μs and erasing at -9 V for 1 ms without cycling the devices before retention. Although the measurement time could not be performed for ten years, as required for nonvolatile memory retention, preliminary room temperature characterization of the un-cycled nanowire SONOS memory cell (shown in Fig. 3.12) exhibits excellent retention. It is found that these devices can keep the stored charge well during the measured time, and the projected memory window is 2.36 V for ten years, during which the charge loss is only about 10%. Since relatively similar retention property can be achieved by planar devices, good retention can also be attributed to the physically thick tunnel oxide (4.5-nm), other than the GAA structure.

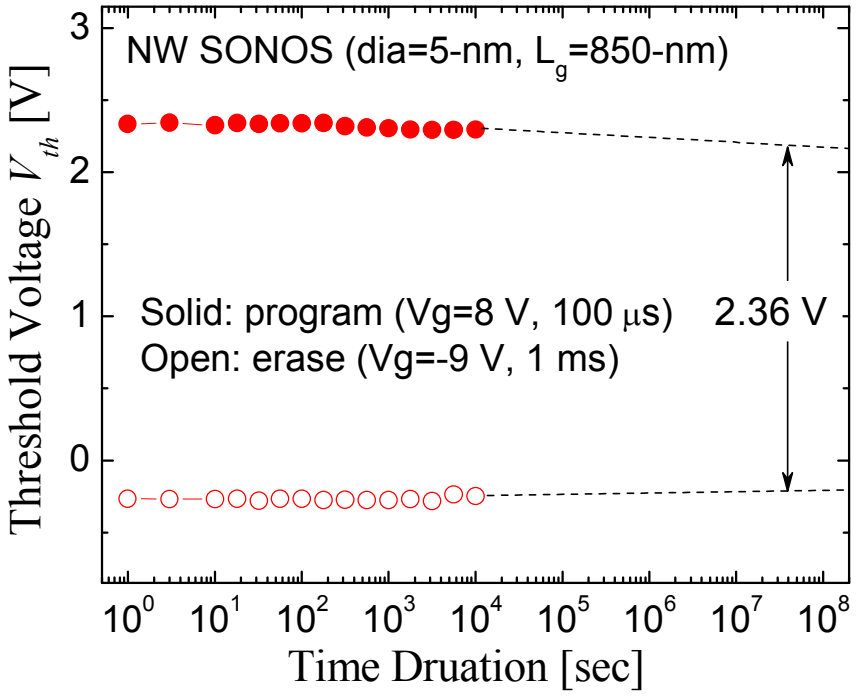


Fig. 3.12: Room-temperature data retention properties for nanowire devices. The stored charge could be kept well for the measured 10^4 seconds.

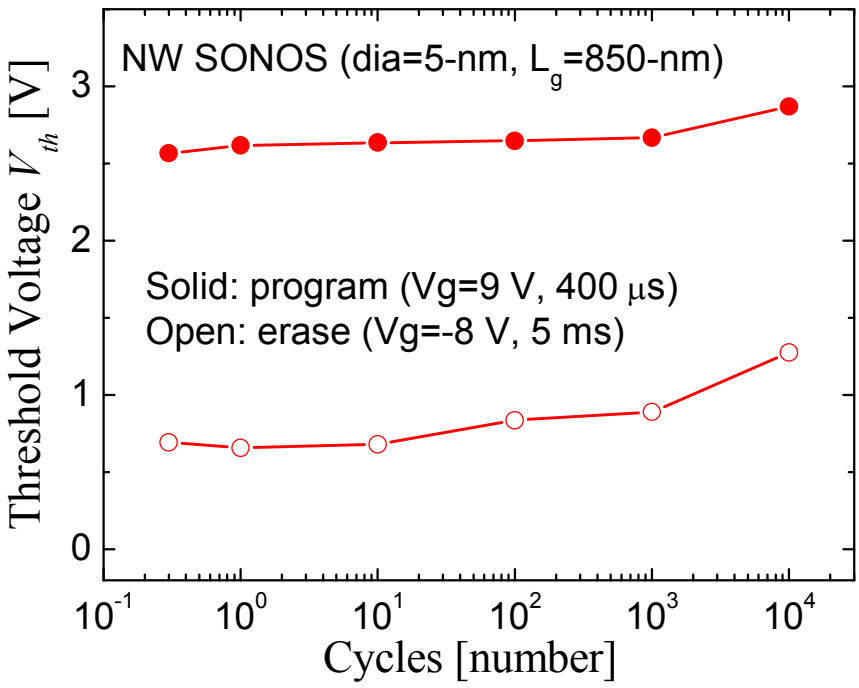


Fig. 3.13: Endurance characteristics of the Si nanowire SONOS device.

Fig. 3.13 shows the endurance properties of the same device. It is observed that the trends of V_{th} for both programmed and erased states shift upwards, and that the device displays a little window closing. The increase of interface traps between the relatively inferior deposited SiO_2 and the channel also result in the V_{th} shifting up, since some high energy electrons and holes could generate interface traps when the devices are under P/E cycling.

3.4 Conclusion

A vertically-stacked twin Si-nanowire GAA SONOS NVM cell was presented using a CMOS-compatible process technology, with enhanced P/E speed, and good retention characteristics. The simulation results reveal that the enhanced P/E speed and functioning at low voltage is a result of the increased vertical electric field at the channel-to-gate dielectric interface, and the reduced electron tunneling length. This novel NW-NVM with fast P/E speed, large window under low voltage, good data retention and process compatibility is very promising for future non-volatile flash-memory applications.

Reference

[3.1] J. D. Lee, S. H. Hur, and J. D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation", in *IEEE Electron Device Lett*., vol. 23, no. 5, pp. 264-266, 2002.

[3.2] *International Technology Roadmap of Semiconductors (ITRS)*, Semiconductor Industry Association, Banjoes, CA, 2007.

[3.3] Y. Zhao, X. Wang, H. Shang, and M. H. White, "A low voltage SANOS nonvolatile semiconductor memory (NVSM) device", in *Solid-State Electron*, vol. 50, pp. 1667-1669, 2006.

[3.4] H. H. Hsu, I. Y. Chang, and J. Y. Lee, "Metal-oxide-high-κ dielectric-oxide-semiconductor (MOHOS) capacitors and field-effect transistors for memory application", in *IEEE Electron Device Lett.*, vol. 28, no.11, pp. 964-966, 2007.

[3.5] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, and B. J. Cho, "Hafnium aluminum oxide as charge storage and blocking-oxide layers in SONOS-type nonvolatile memory for high-speed operation", in *IEEE Trans. Electron Device*, vol. 53, no.4, pp. 654-662, 2006.

[3.6] S. Jeon, J. H. Han, J. H. Lee, S. Choi, H. Hwang, and C. Kim, "High work-function metal gate and high-κ dielectrics for charge trap flash memory device applications", in *IEEE Trans. Electron Device*, vol. 52, no.12, pp. 2654-2659, 2005.

[3.7] Y. H. Shih, H. T. Lue, K. Y. Hsieh, R. Liu, and C. Y. Lu, "A novel 2-bit/cell nitride storage flash memory with greater than 1M P/E-cycle endurance", in *IEDM Tech. Dig*., pp. 881-884, 2004.

[3.8] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS", in *IEEE Circuits and Devices*, vol.16, pp. 22-31, 2000.

[3.9] C. Friederich, M. Specht, T. Lutz, F. Hofmann, L. Dreeskornfeld, W. Weber, J. Kretz, T. Melde, W. Rosner, E. Landgraf, J. Hartwich, M. Stadele, L. Risch, D. Richter, "Multi-level p+ tri-gate SONOS NAND string arrays", in *IEDM Tech. Dig*., pp. 708-711, 2006.

[3.10] P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, and T. J. King, "FinFET SONOS flash memory for embedded applications", in *IEDM Tech. Dig*., pp. 609-612, 2003.

[3.11] Y. Cui, X. Duan, J. Hu, and C. M. Lieber, "Doping and electrical transport in silicon nanowires", *J. Phys. Chem.*, B 104, pp. 5213-5216, 2000.

[3.12] S. D. Suk, S. Y. Lee, S. M. Kim, E. J. Yoon, M. S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D. S. Shin, K. H. Lee, H. S. Park, J. N. Han, C. J. Park, J. B. Park, D. W. Kim, D. G. Park, and B. I. Ryu, "High performance 5nm radius twin silicon nanowire MOSFET(TSNWFET): Fabrication on bulk Si wafer, characteristics, and reliability", in *IEDM Tech. Dig*., pp. 717-720, 2005.

[3.13] N. Singh, F. Y. Lim, W. W. Fang, S. C. Rustagi, L. K. Bera, A. Agarwal, C. H. Tung, K. M. Hoe, S. R. Omampuliyur, D. Tripathi, A. O. Adeyeye, G. Q. Lo, N. Balasubramanian, and D. L. Kwong, "Ultra-narrow silicon nanowire gate-all-around CMOS devices: Impact of diameter, channel-orientation and low temperature on device performance", in *IEDM Tech. Dig*., pp. 547-550, 2006.

[3.14] S. D. Suk, K. H. Yeo, K. H. Cho, M. Li, Y. Y. Yeoh, K. H. Hong, S. H. Kim, Y. H. Koh, S. G. Jung, W. J. Jang, D. W. Kim, D. G. Park, and B. I. Ryu, "Gate-all-around twin silicon nanowire SONOS memory", in *VLSI Technology Symp*., pp. 142-143, 2007.

[3.15] D.-B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-dimensional silicon oxidation experiments and theory", in *IEDM Tech. Dig*., pp. 388-391, 1985.

[3.16] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEDM Tech. Dig.*, pp. 613-616, 2003.

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

Chapter 4

GAA Nanowire for TFT SONOS Multi-Level-Cell Memory Application

4.1 Introduction

Polycrystalline silicon (Poly-Si) thin films have been a key material in silicon integrated circuit technology since the poly-Si gate MOS very large-scale integration (VLSI) technology was developed in the mid 1970s. The common application of poly-Si films is as a gate electrode material with doped impurity atoms. They can also be used as a semiconductor in field-effect devices. Low temperature poly-Si (LTPS) thin-film transistors (TFTs) technology appears to be one of the most promising applications, for the ultimate goal of building fully-integrated active-matrix liquid crystal display (AMLCD) system on glass [4.1].

TFTs have a history almost as long as that of MOSFETs [4.2]. As a special kind of field-effect transistor, they use a semiconductor thin film which is deposited on an insulating substrate. Various semiconductor materials have been studied for use in TFTs, including CdS, CdSe, Se, Ge, SiC and Si. Only two types of TFTs that use Si thin films, amorphous Si thin films and poly-Si thin films, have been used in practical applications so far, since TFTs using compound semiconductor materials have never

achieved reproducibility and acceptable reliability for commercial production.

The most important feature of TFTs is that they can be built on an insulating substrate, typically glass or fused quartz. This allows optical transparency of the device structure, which is essential in most applications. A lot of new applications are becoming feasible by fabricating active devices in poly-Si thin films. In addition, it allows the use of much larger substrate sizes than those possible with Si wafers. Typical mobility of poly-Si TFTs is usually of the order of 10-$cm^2/(V \cdot s)$ for n-channel, which is one order of magnitude higher than that of amorphous Si TFTs. The large mobility makes the realization of large area and high-resolution displays possible. Because of their performance advantages over amorphous Si TFTs, namely, higher driving current, in addition to lower thermal budget and process costs than that of conventional MOSFETs, intensive research has been carried out to apply poly-Si TFT in AMLCD technology in which a TFT is placed at each picture cell (pixel) [4.3]. The TFT is used as a switch to charge and discharge the LCD cell's transparent electrodes during the pixel on and off. Nonvolatile memory based on poly-Si TFT has attracted much attention very recently for use in system-on-panel (SOP) because of its high performance and ease of integration [4.4-4.6]. Low-end circuit integration by poly-Si is currently being applied to actual mobile display panel products [4.7]. Applications of SOP, such as low-bit central processing unit (CPU) integration [4.8] and different embedded memory structure on panel based on some advanced LTPS TFT technologies, have also been demonstrated [4.9].

These nonvolatile memory devices, using poly-Si and embedded on glass, can increase the function and flexibility of circuits associated with the display [4.10]. However, a number of technical challenges have to be overcome in order to fully integrate advanced and complicated circuits on panel products. The most critical of

which are lowering operation voltage to save power consumption, scaling down device size for better integration efficiency, and improving process uniformity for higher integration level. The most common limitations with the TFT SONOS memories are the relatively low memory speed and poor subthreshold property, which are due to the grainy structure of poly-Si and the dangling bonds along the grain boundaries. To improve performance in TFT logic or memory devices, additional processes such as excimer laser annealing to increase the grain size have been utilized [4.11]. Yamauchi *et al.* [4.12] have shown significant device performance enhancement by increasing the grain size to be comparable to the channel dimension, thereby reducing the negative effects of poly-Si grain boundary. The poly-Si device characteristics have also been improved significantly by reducing the cross-sectional dimensions of the poly-Si channel to make these comparable to the grain size, such as fin-like poly-Si channel [4.13] or nanowire-like channel devices [4.14]. The devices with such architectures offer alleviated aforementioned issues, as well as enhanced gate controllability through the advanced gate and channel structures.

In this chapter, we extensively investigate the structure and electrical characteristics of an n-type poly-Si TFT SONOS memory with GAA nanowire. By following the developed method and by improving the process conditions, the poly-Si nanowire channel was realized using well controllable top-down CMOS process technology without any additional non-compatible process techniques. The TFT SONOS memory device, with the use of nanowire channel, exhibits better device property at a much lower operating voltage, which is attributed to the reduced number of grain boundaries in the channel region, since the fabricated device takes advantage of the narrow channel body and exhibits a much enhanced memory speed and improved SS as compared to conventional TFT memory devices. The decreased SS is

comparable over that of conventional single crystalline MOSFET devices.

4.2 Poly-Si Nanowire TFT Memory Device Fabrication

A layer of 400-nm thermal SiO_2 was first grown on Si bulk wafer to simulate the glass substrate, followed by a layer of 60-nm amorphous silicon film deposited at a temperature of 550ºC in LPCVD. The active regions, in the form of narrow fin connecting to wider source and drain pads (as shown in Fig 4.1), were then patterned by alternating phase shift mask lithography using a 248-nm exposure wavelength scanner followed by a Si dry etch process. Resist and hard mask trimming was used to achieve fins of even smaller critical dimensions. The fins were then converted into nanowires by steam oxidation at a temperature of 730ºC for 30 minutes – the highest thermal budget used in this work and was limited by the lowest temperature setting of our oxidation furnace equipment. The simulation work showed that the steam oxidation could be conducted below 600 ºC which is accepted for TFT device fabrication. We observed from the TEM analysis that during this step, the amorphous-Si crystallized into poly-Si grains of size ranging from 15-nm to 30-nm. The poly-Si nanowires were then released in DHF, followed by the deposition of tunnel oxide, charge trapping layer nitride and blocking oxide (ONO: 5-nm/15-nm/11-nm) using PECVD, all at a temperature of 400ºC. The 130-nm amorphous Si was then deposited, patterned and etched to form the gate electrode. Phosphorous (4×10^{15}-cm^{-2}, 30keV) was subsequently implanted and activated at 580ºC for 24 hours, followed by standard steps of contact, metallization and sintering to complete the fabrication. Fig. 4.2 shows a cross-sectional TEM image of the GAA nanowire TFT SONOS device, with a minimum poly-Si nanowire width and height of 23-nm and 36-nm respectively achieved in this work. The wire cross section exhibits a relatively rectangular shape that reflects the original shape of the fin, indicating that the oxidation is insufficient to

show the self-limiting behavior. The ONO layer was not evenly grown along the nanowire channel due to the inferior isotropic property of the PECVD. In this work, the words width/wide are used to describe the horizontal nanowire dimension across the current flow and height/high are used to describe the vertical dimension.

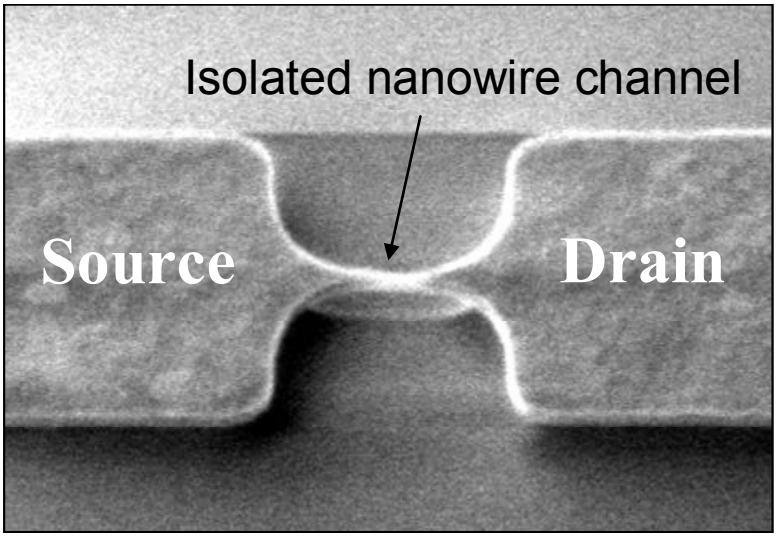


Fig. 4.1: A tilted SEM image of isolated nanowire channel with source and drain pads located on SiO_2 after the nanowire was released in DHF. It can be seen that the amorphous Si has converted to poly-Si after the steam oxidation step.

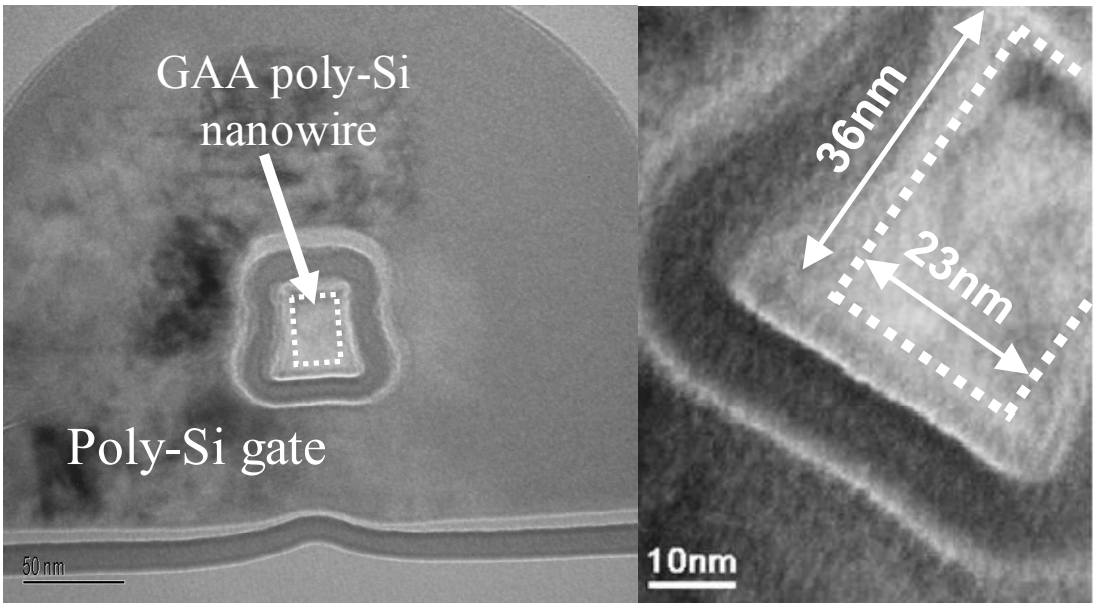


Fig. 4.2: TEM cross-section through the poly-Si nanowire of a fabricated device. The cross-section was cut across the nanowire channel. The nanowire channel width and height are 23-nm and 36-nm respectively as shown by the arrows.

4.3 Results and Discussion

Devices with gate length of 350-nm were used for characterization in this work. Fig. 4.3 shows the transfer characteristics of the 23-nm wide nanowire device with a gate length of 350-nm. The SS is usually larger than 200-mV/dec in typical TFT devices. Along with a high ON/OFF ratio, a low SS value of 122-mV/dec is obtained. However, the SS of the GAA nanowire TFT SONOS device is significantly reduced as compared to other reported TFT devices. Enhanced gate controllability from the GAA structure is one reason that such good SS is obtained, and the reduced number of grains in the nanowire channel also accounts for the good device performance.

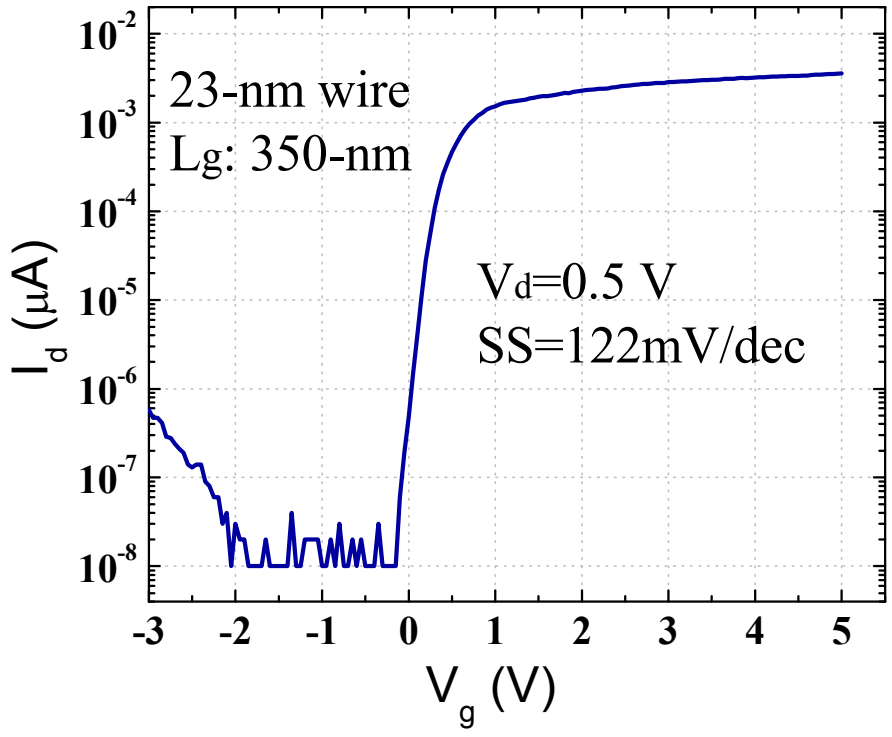


Fig. 4.3: Transfer characteristics of nanowire TFT SONOS device with 23-nm width and 350-nm gate length. Improved SS is exhibited as compared to other reported TFT SONOS memory devices.

The cumulative distribution of SS for three nanowire devices with different wire-widths is plotted in Fig. 4.4. Improved SS can be achieved as the wire width decreases, indicating that smaller wire width is favored in nanowire memory devices. Degradation of SS in devices with wider nanowire width is clearly seen. Degradation of SS could be due to reduced gate controllability and increased volume of grain boundaries in the channel region of the devices with wider nanowire channel.

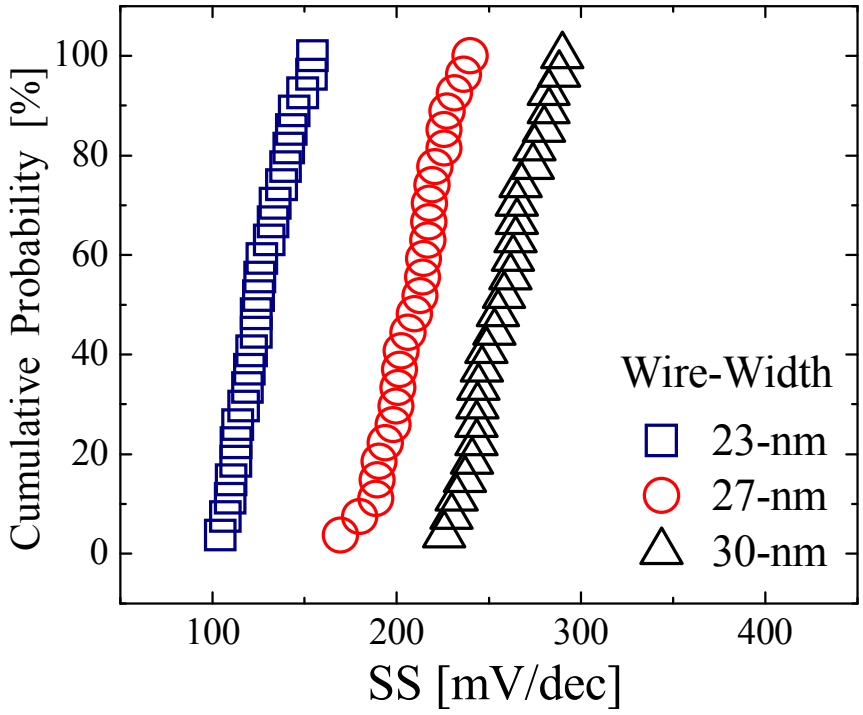


Fig. 4.4: The dependence of SS on poly-Si nanowire width. It can be seen that the device with smaller wire width achieves better SS, which is due to reduced gate controllability and increased volume of grain boundaries in the wider nanowire channel.

The width of all fabricated nanowires in this work are around 20-nm to 50-nm, which are in the order of grain size of the poly-Si film — 50-nm to 100-nm [4.15]. If one starts the nanowire process with poly-Si, it is possible that the grain boundary can be anywhere. However, we started the fabrication process using the amorphous silicon layer, which was crystallized during nanowire oxidation and S/D activation processes after the active area pattern transfer — nanowire channels connected to S/D pads. The

formed poly-Si grain is confined in the nanowire when the amorphous silicon was crystallized during the steam oxidation process (730ºC, 30 minutes). Since the poly-Si nanowire width and height are comparable to or even smaller than the poly-Si grain size, the number of grains and grain boundaries in the nanowire channel are expected to be significantly reduced. In the TEM image, the color of the nanowire is quite even as observed in the poly-Si nanowire region. Based on such reasons, it could be deduced that there are few grain boundaries in the nanowire channel itself. Hence, the SS is improved in a TFT transistor device with comparing channel and grain size, and the improvement is greater in thinner wire devices.

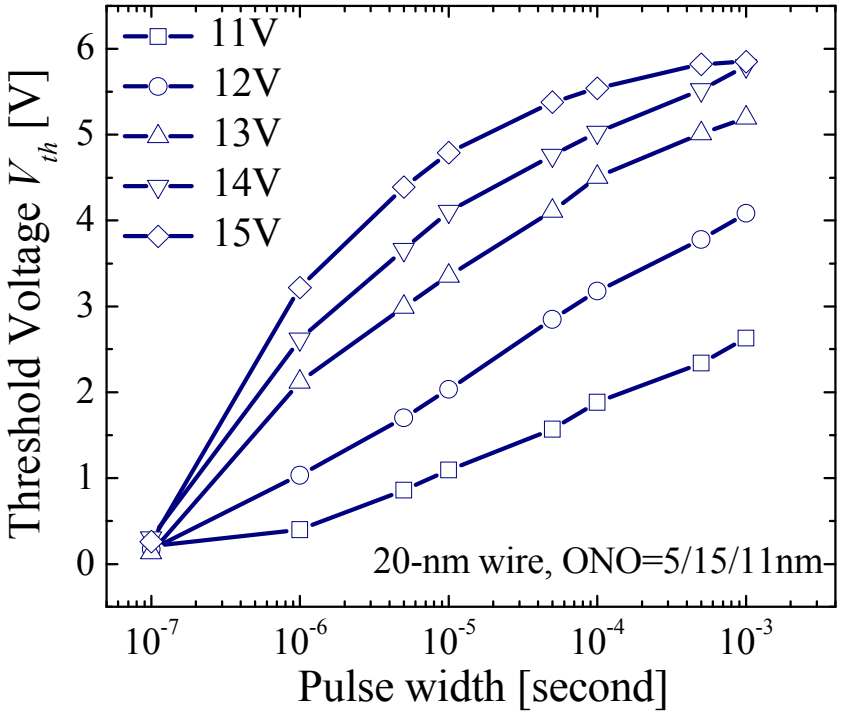
The memory P/E characteristics at various gate stressing voltages as a function of pulse width are presented in Fig. 4.5. F-N tunneling mechanism was employed to characterize the transient memory performance of the cell by grounding the source and drain and stressing the gate. The poly-Si nanowire memory device exhibits a fast programming speed: a threshold voltage shift (ΔV_{th}) of 2.96 V could be achieved in 1 μs at 15 V gate pulse. The programming speed improvement is a result of the crowding and convergence of electric field lines in the tunnel oxide at the corners of the rectangular channel structure. The wire corners in our device have a very small radius of curvature due to the use of low temperature oxidation of the fin in the fabrication process, and the injection of electrons/holes is dramatically boosted especially when the corner radius is small [4.16]. The erasing time to achieve the same memory window is about 1 ms at -16 V. The relatively slow erasing speed is due to the higher hole energy barrier (4.6 eV) and larger effective mass of the hole than that of the electron in SiO_2, which results in reduced tunneling of the holes. As we know, there is certain imbalance between the programming and erasing speed in most of the SONOS memory devices. In fact, the erasing speed remains slower than

programming speed at the first hundreds of microseconds despite improvements made in the erasing speed using appropriate band engineering [4.17]. Factors contributing to the slowness of erasing speed can be found in the F-N current density definition given by equation (3) [4.18]:

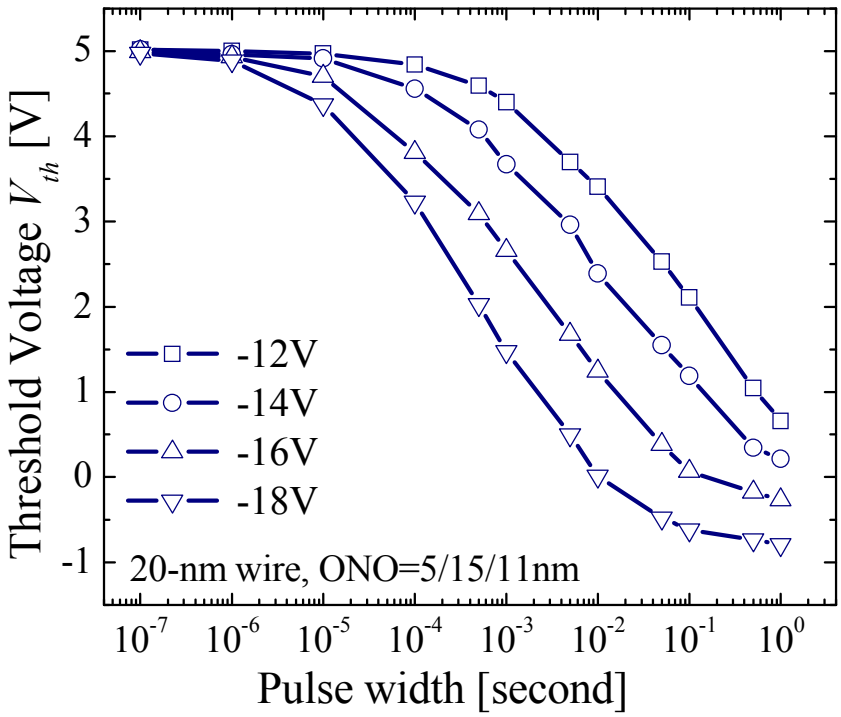
$$J = \frac{q^3}{8\pi h \Phi_b} \frac{m}{m^*} E_{inj}^2 \exp[\frac{-E_c}{E_{inj}}] \tag{3}$$

with $E_c = 4\sqrt{2m^*}\,\frac{\Phi_b^{3/2}}{3\mathrm{h}q}$,

where Φ_b is the energy barrier at the SiO_2/Si interface (3.2 eV for electron and 4.6 eV for hole), E_{inj} is the electric field at the injecting surface and m^* is the effective mass of carriers in the SiO_2. It implies that there is a reverse relationship between the current density *J* and the effective mass m^*. The effective mass of a hole is larger than that of an electron in most of semiconductors as well as SiO_2. The effective mass and the energy barrier also have a negative exponential effect on the tunnel current density, which further decrease the injecting current density. In addition, the energy barrier height affects the tunnel speed exponentially. Both the higher effective mass and large barrier height of holes than that of electrons contribute to the slow erasing speed.



(a)



(b)

Fig. 4.5: The P/E characteristics of nanowire TFT SONOS with wire width 23-nm. The gate pulse stress ranged from 11 V to 15 V and -12 V to -18 V. ΔV_{th} of 2.96 V could be achieved in 1 μs at 15 V gate pulse during program.

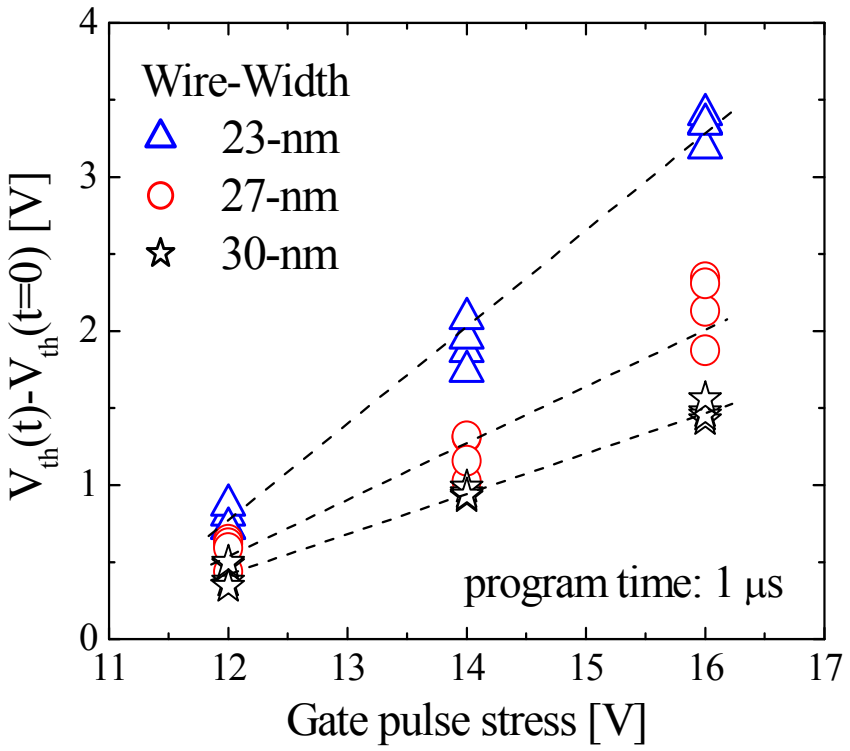


Fig. 4.6: Comparison of memory window among devices with three different nanowire widths under different stress voltage for a program time of 1 μs. The device with the narrowest wire width shows the fastest speed.

Next, we present the dependence of memory transient property on the width of the nanowire. The ΔV_{th} is plotted as a function of stress for 1 μs programming time in Fig. 4.6 for the same three kinds of nanowire devices used for testing the cumulative distribution of SS earlier. It can be noticed that the device with the narrowest wire width shows the fastest speed. Although the electric field in the tunnel oxide is non-uniform due to the rectangular shaped channel, the enhanced ΔV_{th} in the smaller nanowire width device could still be attributed to the increased vertical electric field in the tunnel oxide due to the shortened distance between the channel interface and the channel centre, as it was in the circular nanowire device [4.19]. The enhanced electric field intensifies the bending of the conduction band when the device is being stressed, as Fig. 4.7 shows the programming process. With the larger band bending,

the carrier experiences a shorter barrier width to tunnel through and thus faster memory speed is observed. The increased electric field in the device with smaller wire size promotes more carriers injecting into the charge trapping layer and leads to a larger ΔV_{th}. Due to such effects, a smaller cross section of nanowire is preferred in GAA nanowire channel devices.

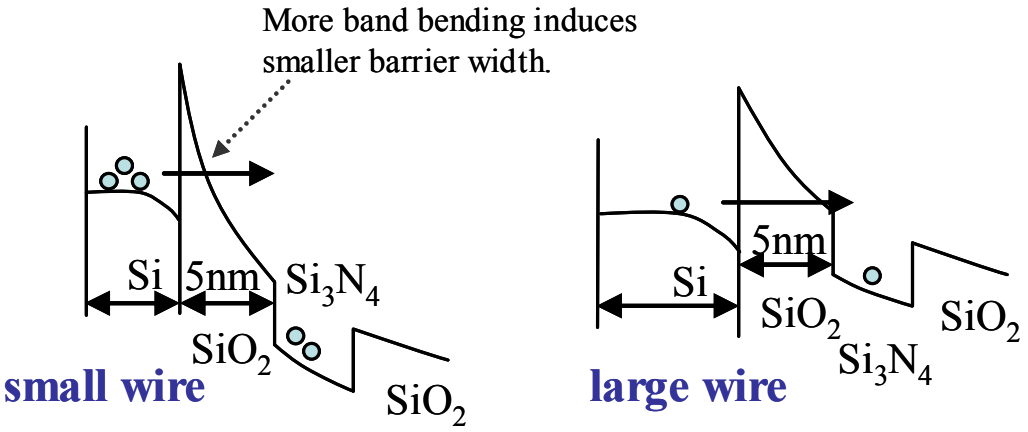


Fig. 4.7: The band diagram during program in devices with smaller wire width (left) and larger wire width (right). Stressed at the same voltage, the electric field across the tunnel oxide of the device with smaller wire width is enhanced as compared to that of the counterpart, due to the particular GAA structure.

The nanowire devices, including those published so far from different groups, are sensitive to the channel diameter and hence a good control in the fabrication process is very important. However, that should not be a problem for top-down etched nanowire as advanced lithography tools are able to control the critical dimensions to within 2 to 3-nm in this work. Furthermore, the oxidation process at low temperature slows down with size reduction of the nanowire and that reduces incoming variation. We would like to clarify that devices with different wire widths measured in this work originate from different starting fin widths, which were designed with different critical dimensions on the mask for the purpose of studying the impact of width. Devices from the same designed width display the same memory speed.

Process technology has allowed significant memory cell density increases and lower cost-per-byte since flash memory was introduced. The expanding need for additional nonvolatile data storage by applications creates demand for higher and higher density flash memory products. A concept called multilevel cell (MLC) technology was invented to build on process scaling. This involves storing multiple bits of information on a single memory transistor, and storing two bits per cell could instantly double the density. In the MLC application, stress voltage would be divided into multiple levels. For two-bit per cell, voltages levels are defined to program/erase the cell to the "11" to "10", "01", or "00" states. MLC devices should be able to place and sense the charge precisely. As illustrated in Fig. 4.6, poly-Si nanowire TFT memory devices of the same dimension display relatively the same V_{th} shift, which allows them to be capable for the MLC application. To study the feasibility of MLC application, we plot the I_d-V_g at four states in Fig. 4.8.

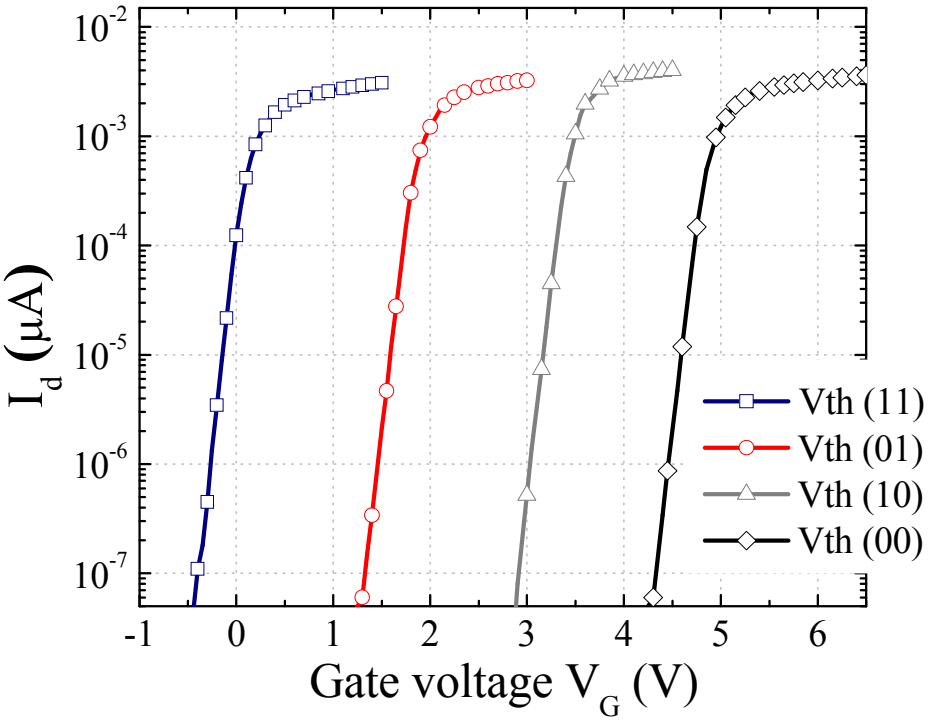


Fig. 4.8: The I_d-V_g at four states shows the feasibility of the MLC application for TFT SONOS. Around 1.5-V memory window is set between the different states.

As the threshold voltage decay logarithmically in time for the SONOS device [4.20], the room temperature data retention property of the GAA poly-Si TFT SONOS MLC device is extrapolated for each V_{th} level and illustrated in Fig. 4.9, with the most severe degradation of only 12% of the ~1.5 V original memory window after 10 years. To differentiate the memory state, either between "0" and "1" or among the four levels for multi-level purpose, around 1 V memory window is considered acceptable [4.21]. Based on that, we made the comment that ~1.5 V window is safe for state determination. Table 4.1 compares the GAA nanowire SONOS presented in this work with most recently published poly-Si SONOS memory devices of other different device structures, in terms of electrostatic behavior, memory transient speed and retention property. The merits of the GAA poly-Si nanowire memory in improving device characteristics can be observed clearly here.

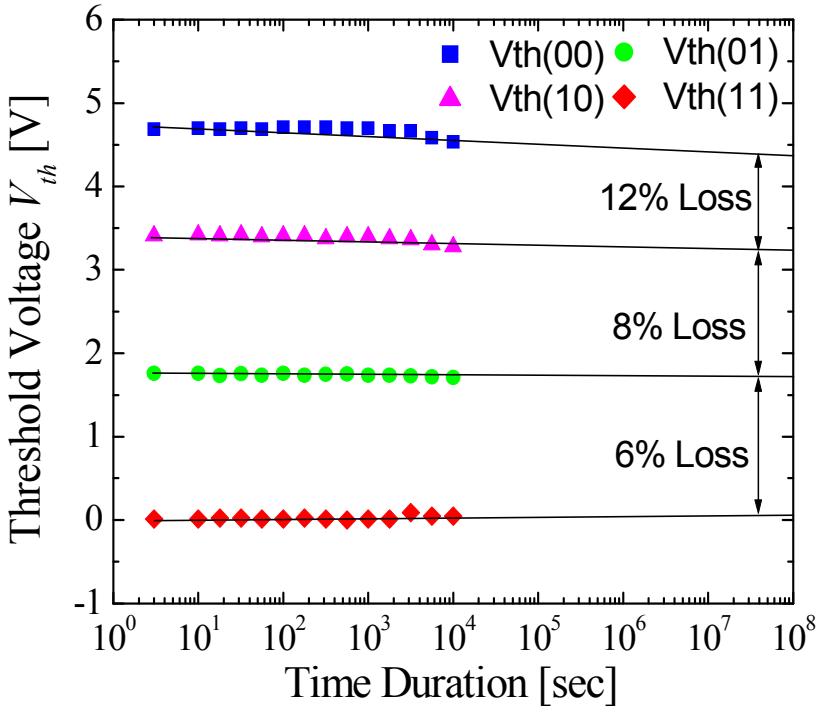


Fig. 4.9: The room temperature retention properties measured at different data states. The most severe degradation is 12% charge loss after 10 years.

Table 4.1 Comparison of memory characteristics with reported TFT SONOS-type memory devices. The GAA nanowire TFT memory in this work displays the advantages in both electrostatic and transient characteristics.

	This work	Hsieh *et al.*, [4.9]	Walker *et al.*, [4.22]	Chen *et al.*, [4.23]	Lin *et al.*, [4.24]
ONO material	$SiO_2/SiN/SiO_2$	$SiO_2/SiN/SiO_2$	$SiO_2/SiN/SiO_2$	$SiO_2/SiN/SiO_2$	$SiO_2/ZrSi/SiO_2$
Thickness (nm)	5/15/11	15/25/30	2.5/7/5	15/20/25	9/13.9/33
L/W	0.35μm/23nm	3μm/3.5μm	0.25 μm technology	1μm/65nm	--
SS	122mV/dec	--	260mV/dec	720mV/dec	~500mV/dec
P/E mechanism	FN/FN	CHE/BBHH	FN/FN	FN/FN	CHE/BBHH
P/E voltage	Vg=14V/ Vg= -18V	V_g=18V, V_d=12V/ V_g=-8V, V_d=20V	Vg=15V/ Vg= -12.5V	Vg=25V/ Vg= -30V	V_g=12V, V_d=12V/ V_g=-10V, V_d=10V
P/E speed (ΔV_{th}=3V)	1×10^{-6} sec/ 4×10^{-4} sec	2×10^{-3} sec/ 5×10^{-2} sec	1×10^{-5} sec/ 1×10^{-2} sec	2×10^{-2} sec/ 1×10^{-1} sec	1×10^{-3} sec/ 1×10^{-2} sec

4.4 Conclusion

In summary, GAA poly-Si nanowire TFT SONOS nonvolatile memory with minimum wire width of 23-nm demonstrated a much improved device subthreshold characteristics and excellent device uniformity. Fast memory speeds are achieved with lower power consumption due to the gate-all-around structure. Improvement of memory speeds is a result of the crowding and convergence of electric field lines in the tunnel oxide at the corners of the rectangular channel structure, as the injection of electrons/holes is dramatically boosted especially when corner radius is small. The demonstrated large memory window enables its success for MLC applications. Both memory and transistor performances are improved with the decrease of the nanowire

width.

The fabricated memory device is useful for both the SOP and SOC applications. To our understanding, the SOP process start from thin films on glass panel and hence memory and other electronics remain as thin film devices. Our device, being fabricated on poly silicon thin film, fits well into that application. The fin oxidation temperature remains a concern, which can be reduced by using a furnace with low temperature capability. Therefore, the GAA nanowire TFT SONOS holds great promise for future TFT nonvolatile memory applications to be embedded into actual mobile display panel products.

Reference

[4.1] S. Zhang, C. Zhu, J. K. O. Sin, J. N. Li, and P. K. T. Mok, "Ultra-thin elevated channel poly-Si TFT technology for fully-integrated AMLCD system on glass", in *IEEE Trans. Electron Device*, vol. 47, no.3, pp. 569-575, 2000.

[4.2] P. K. Wiemer, "The TFT−a new thin-film transistor" in *Proc. IRE*, vol. 50, pp. 1462-1469, 1462.

[4.3] A. G. Fische, "Design considerations for a future electroluminescent TV panel", in *IEEE Trans. Electron Device*, vol. 18, no.9, pp. 802-804, 1971.

[4.4] M. Cao, T. Zhao, K. C. Saraswat, and J. D. Plummer, "A simple EEPROM cell using twin polysilicon thin film transistors", in *IEEE Electron Device Lett.*, vol. 15, no. 8, pp. 304-306, 1994.

[4.5] N. D. Young, G. Harkin, R. M. Bunn, D. J. McCulloch, and J. D. French, "The fabrication and characterization of EEPROM arrays on glass using a low-temperature poly-Si TFT process", in *IEEE Trans. Electron Device*, vol. 43, no.11, pp. 1930-1936, 1996.

[4.6] J. W. Lee, N. I. Lee, H. J. Chung, and C. H. Han, "Improved stability of polysilion thin-film transistors under self-heating and high endurance EEPROM cells for systems-on-panel", in *IEDM Tech. Dig*., pp. 265-268, 1998.

[4.7] C. S. Tan, W. T. Sun, S. H. Lu, C. H. Kuo, L. T. Chang, S. H. Yeh, C. C. Chen, L. Liu, Y. C. Lin and C. S. Yang, "A simple architecture for fully integrated poly-Si TFT-LCD", in *Proc. SID*, pp. 336-339, 2005.

[4.8] B. Lee, Y. Hirayama, Y. Kubota, S. Imail, A. Imaya, M. Katayama, K. Kato, A. Ishikawa, T. I. Ikeda, Y. Kurokawa, T. Ozaki, K. Mutaguch, and S.

Yamazaki, "A CPU on a glass substrate using CG-silicon TFTs", in *ISSCC Tech. Dig.*, pp. 164-165, 2003.

[4.9] S.-I. Hsieh, H.-T. Chen, Y.-C. Chen, C.-L. Chen, and Y.-C. King, "MONOS memory in sequential laterally solidified low-temperature poly-Si TFTs", in *IEEE Electron Device Lett.*, vol. 27, no. 6, pp. 272-274, 1999.

[4.10] K. P. Anish Kumar, and J. K. O. Sin, "A simple polysilicon TFT technolgy for display systems on glass", in *IEDM Tech. Dig*., pp. 515-518, 1997.

[4.11] K. Kitahara, Y. Ohashi, and Y. Katoh, "Submicron-scale characterization of poly-Si flims crystallized by excimer laser and continuous-wave laser", in *J. Appl. Phys*. Vol. 95, pp. 7850-7855, 2004.

[4.12] N. Yamauchi, J. J. J. Hajjar, and R. Reif, "Polysilicon thin-film transistors wich channel length and width comparable to or smaller than the grain size of the thin film", in *IEEE Trans. Electron Devices*, vol. 38, no. 1, pp. 55-60, 1991.

[4.13] H. Yin, W. Xianyu, A. Tikhonovshy, and Y. S. Park, "Scalable 3-D fin-like poly-Si TFT and its nonvolatile memory application", in *IEEE Trans. Electron Devices*, vol. 55, no. 2, pp. 578-584, 2008.

[4.14] M. Im, J. W. Han, H. Lee, L. E. Yu, S. Kim, C. H. Kim, S. C. Jeon, K. H. Kim, G. S. Lee, J. S. Oh, Y. C. Park, H. M. Lee, and Y. K. Choi, "Multi-gate CMOS thin-film transistor with polysilicon nanowire", in *IEEE Electron Device Lett.*, vol. 29, no. 1, pp. 102-105, 2008.

[4.15] N. Yamauchi, J.-J. J. Hajjar, and R. Reif, "Polysilicon thin-film transistors wich channel length and width comparable to or smaller than the grain size of the thin film", in *IEEE Trans. Electron Devices*, Vol. 38, No. 1, pp. 55-60, 1991.

[4.16] T.-H. Hsu, H.-T. Lue, E.-K. Lai, J.-Y. Hsieh, S.-Y. Wang, L.-W. Yang, Y.-C. King, T. Yang, K.-C. Chen, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "A high-speed BE-SONOS NAND flash utilizing the field-enhancement effect of FinFET", in *IEDM Tech. Dig*., pp. 913-916, 2007.

[4.17] Z. Huo, J. Yang, S. Lim, S. Baik, J. Lee, J. Han, I.-S. Yeo, U.-I. Chung, J. T. Moon, and B.-I. Ryu, "Band engineered charge trap layer for highly reliable MLC flash memory", in *VLSI Technology Symp*., pp. 138-139, 2007.

[4.18] M. Lenzlinger, and E. H. Snow, "Fowler-Nordheim tunneling in thermally grown SiO_2", in *J. Appl. Phys.*, vol. 40, pp. 278, 1969.

[4.19] J. Fu, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, "Si-nanowire based gate-all-around nonvolatile SONOS memory cell", in *IEEE Electron Device Lett.*, vol. 29, no. 5, pp. 518-521, 2008.

[4.20] William D. Brown, Joe E. Brewer, "Nonvolatile semiconductor memory technology: A comprehensive guide to understanding and using NVSM devices" *Wiley-IEEE Press*, pp. 74, October 1997.

[4.21] S. Kim, W. Kim. J. Hyun, S. Byun, J. Koo, J. Lee, K. Cho, S. Lim, J. Park, I.-K. Yoo, C.-H. Lee, D. Park, and Y. Park, "Paired FinFET charge trap flash memory for vertical high density storage", in *VLSI Technology Symp*., pp. 84-85, 2006.

[4.22] A. J. Walker, S. Nallamothu, E.-H. Chen, M. Mahajani, S. B. Herner, M. Clark, J. M. Cleeves, S. V. Dunton, V. L. Eckert, J. Gu, S. Hu, J. Knall, M. Konevechi, C. Petti, S. Radigan, U. Raghuram, J. Vienna, and M. A. Vyvoda, "3-D TFT-SONOS memory cell for ultra-high density file storage applications", in *VLSI Technology Symp*., pp. 29-30, 2003.

[4.23] S. C. Chen, T. C. Chang, P. T. Liu, Y. C. Wu, C. C. Ko, S. Yang, L. W. Feng, S. M. Sze, C. Y. Chnag, and C. H. Lien, "Pi-shape gate polycrystalline silicon thin-film transistor for nonvolatile memory applications", in *Appl. Phys. Lett.*, vol. 91, pp. 213101, 2007.

[4.24] Y.-H. Lin, C.-H. Chien, T.-H. Chou, T.-S. Chao, C.-Y. Chang, and T.-F. Lei, "2-bit poly-si-TFT nonvolatile memory using hafnium oxide, hafnium silicate and zirconium silicate", in *IEDM Tech. Dig*., pp. 927-930, 2005.

Chapter 5

GAA Nanowire MONOS for High Speed Memory Application

5.1 Introduction

Nonvolatile memories are being employed in almost all computing devices and systems, ranging from mobile devices such as cell phones, personal digital assistants (PDAs) to laptops. The scaling of NAND-type flash memory has been aggressively pursued in recent years [5.1]. Its low power consumption, low cost, and large memory capacity have triggered its application in numerous new consumer products. Nevertheless, the current floating gate NAND flash faces a critical scaling challenge beyond 45-nm technology node because of the interfloating gate coupling issue [5.2]. The SONOS-type memory has been investigated as a potential alternative due to its immunity to the floating gate coupling interference.

Silicon-nitride discrete charge trapping material based SONOS memory devices have been studied extensively, and the key dielectrics SiO_2 and Si_3N_4 are widely investigated. The SONOS memory, however, still faces an issue in scaling due to the non-scalable gate dielectric stack. An ultra-thin tunnel oxide (~2-nm), which is preferred for F-N tunneling, is not reliable to prevent direct tunneling of stored

carriers, causing a retention problem in these devices [5.3]. On the other hand, SONOS devices with a thick tunnel oxide (~5-nm) have very slow memory speed especially for erasing. Although band-to-band tunneling hot hole injection (BBHH) is effective for erasing SONOS devices with thick tunnel oxide, it is not suitable for NAND-type Flash as it makes the device vulnerable to the hot-hole related issues [5.4]. Hence metal-oxide-nitride-oxide-silicon (MONOS) memory devices, which use high-κ as dielectric materials and metal as gate electrode, become an attractive candidate for further device scaling. A substantial amount of research work have been performed on MONOS-type memory devices, involving engineering of the gate electrode, blocking oxide, trapping layer, even the tunneling oxide. As high-κ films possess discrete traps and trap carriers in the spatially isolated deep energy level, it has been reported that the use of high-κ such as HfO_2 as a trapping layer in MONOS could improve P/E speed, vertical scaling and charge retention properties [5.5]. The use of high work function metal gates suppresses gate electron injection and hence improves the erase speed. From this point, metal gate electrodes, with a work function higher than poly-Si, were widely employed [5.6-5.7].

Innovative device structures are also required for the new generation NAND-type flash memory. SONOS memory cells with novel GAA single crystalline Si nanowire channel and relatively thick tunnel oxide have recently demonstrated fast memory speed using the top-down process [5.8], since a larger portion of the electric field was dropped the tunneling layer [5.9]. However, much slower erasing speed as compared to programming speed could still be observed in these nanowire SONOS devices due to unwanted backward gate electron injection and higher hole tunneling barrier (4.6 eV) as compared to that of electron (3.2 eV) at the channel/tunnel oxide interface [5.9-5.10].

To the best of our knowledge, high-κ materials have not yet been applied in memory products with nanowire devices even though they have been extensively investigated in nonvolatile memory devices with planar architecture. No work has been conducted to combine the advantages of the high-κ/metal gate and the GAA nanowire device structure. In this chapter, charge trapping memory with $Al_2O_3/HfO_2/SiO_2$ stack and metal gate fabricated on GAA nanowire structure is proposed and investigated as the next generation nonvolatile memory technology. The application of high-κ materials reduces the electric field in the blocking layer. Besides trap layer engineering, the purpose of high-κ dielectric introduction is to reduce the EOT – a need in scaling. Hence we have intentionally retained the same physical thickness and allowed the EOT to decrease in order to make the study meaningful for the purpose of application. In addition, the metal as the gate electrode could further prevent unwanted backward tunneling of electrons significantly due to its higher work function. As a result, the memory speed, especially for erasing, is improved.

5.2 TAHOS Nanowire Memory Device Fabrication

Fig. 5.1 shows the schematics of the main process steps along with process fabrication flow. The fabrication started on p-type SOI wafer with 150-nm BOX, and the top silicon thickness of SOI wafer was first thinned down to 40-nm by thermal oxidation. The active region (fin with source and drain pads) was then defined by utilizing conventional KrF DUV alternating phase shift mask lithography and dry etching process. After the photo-resist strip and wet clean, the patterned Si-fins were oxidized in dry O_2 ambient as described before. The oxidation process converted the fins into Si nanowires wrapped by the SiO_2. In this work, single nanowire was formed from each fin and studied again. Shown in Fig. 5.2 is the SEM image at 45º tilted view of a single Si nanowire connecting the S/D pads, after the thermal oxide shell

outside the nanowire was removed in 1:25 diluted DHF. We kept the wire dimension relatively thick (25 ~ 30-nm) in order to tolerate the large compressive stress existing in the deposited tantalum nitride (TaN) layer, which could bend the very thin nanowire [5.11].

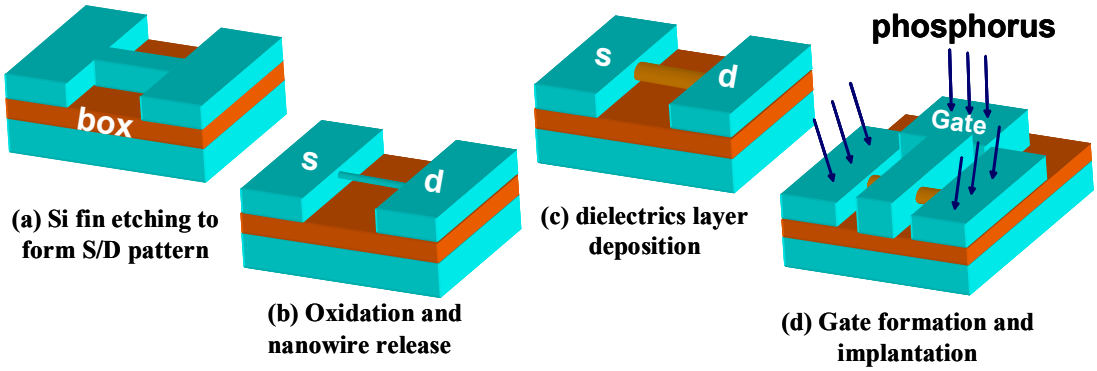


Fig. 5.1: The schematics of the main process steps along with the process fabrication flow for fabricating the GAA nanowire TAHOS memory cell.

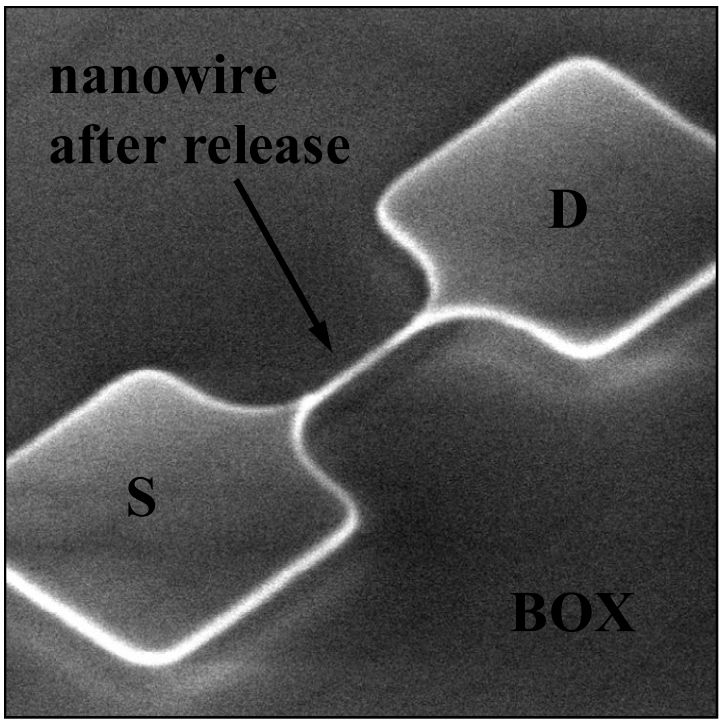


Fig. 5.2: A tilted SEM image of the nanowire with source and drain pads located on BOX after the nanowire was released in DHF.

After the nanowire channel was formed, 5-nm TEOS SiO_2 as the tunnel oxide layer was deposited in LPCVD, followed by subsequent deposition of 7-nm HfO_2 as the charge trapping layer and 10-nm Al_2O_3 as the blocking oxide layer, using the atomic layer deposition (ALD) machine tool at a temperature of 300ºC. The ALD process was performed in a Genus Lynx2 ALD reactor. In this work, tetrakis, $Hf(NEtMe)_4$, was used as the Hf precursor and TMA was used as precursor for the growth of Al_2O_3. The chamber pressure was set to 200-mTorr during deposition. N_2 was used as the source carrier gas. A deposition cycle consists of a pulse for the metal precursor supply and a pulse for the oxidant supply. Argon gas keeps flowing for 8 seconds to purge precursor residuals and reactant molecules in between the pulses. Post deposition annealing (PDA) in N_2 ambient at 700ºC for 30 seconds was performed to dehydrate and condense the ALD grown high-κ dielectrics in order to remove residual H_2O, which was used as an oxidant in the ALD process. TaN with 200-nm thickness was sputter deposited, patterned and etched to fabricate the gate electrode. The fabricated nanowire memory device with $Al_2O_3/HfO_2/SiO_2$ gate stack and TaN metal gate is abbreviated as TAHOS for short. For the purpose of comparing device performances, control device nanowire SONOS memory featuring the LPCVD $SiO_2/Si_3N_4/SiO_2$ (5-nm/7-nm/10-nm) and 130-nm poly-Si gate was fabricated at the same time. All wafers then underwent the same implantation (phosphorus, 30 keV, 4×10^{15} cm^{-2}), thermal activation (900ºC, 30 seconds) and contact metallization to conclude the process.

Thick gate dielectrics cannot be deposited on the nanowire in a GAA structure due to the limited space between the nanowire and the BOX. Hence two kinds of devices with the same physical oxide thicknesses, rather than the same electrical oxide thicknesses, for gate stacks were fabricated and compared in this work, since

the objective of this work was to study the feasibility of high-κ metal gate technology in GAA nanowire device architecture for nonvolatile memory application. From the view point of application, we kept the dielectric physical thicknesses to the minimum possible and hence did not match the EOT. In addition, the purpose of growing equal physical thicknesses of the gate dielectrics was to enhance the electric field in the tunneling oxide and reduce the electric field in the blocking oxide in the nanowire TAHOS devices, as a consequence of using high dielectric constant (κ) materials.

5.3 Results and Discussion

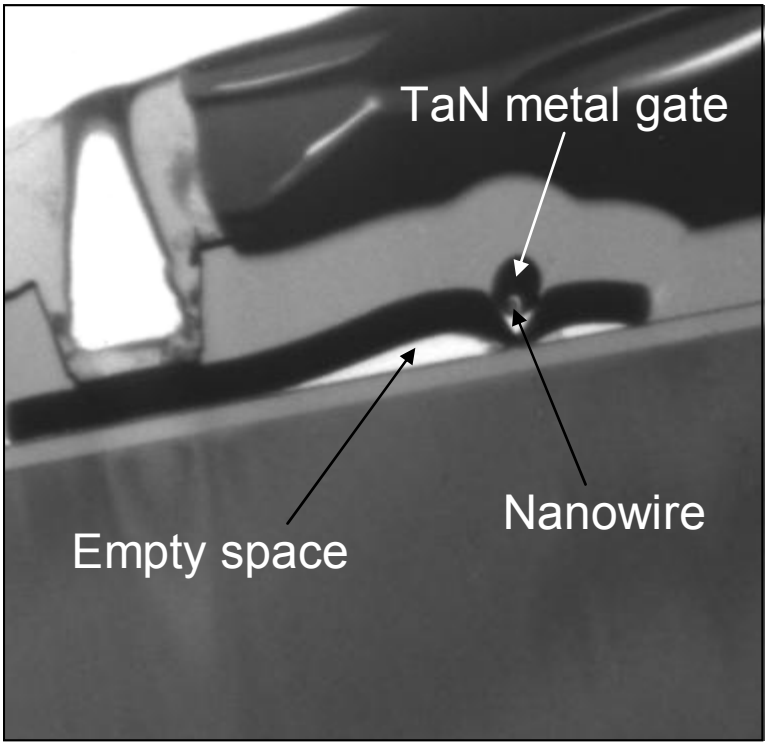


Fig. 5.3: TEM image showing the cross-section of fabricated GAA nanowire TAHOS device. The nanowire can be observed to be surrounded by TaN metal gate and high-κ oxide. Empty spaces are attributed to high stress formed at the interface TaN layer.

Fig. 5.3 shows the TEM cross section of the fabricated device perpendicular to the nanowire channel. Different layers in the gate stack could not be imaged clearly due to poor contrast in the high-κ and TaN metal gate layers. The nanowire is also

clearly seen in the low magnification images to be almost surrounded (leaving some empty space below due to the directional nature of the TaN PVD deposition process) with the metal. Although our fabrication method is capable of producing nanowire down to 5-nm in diameter, the minimum diameter of the nanowire used in this work was limited to ~ 25 to 30-nm in order to cater for the large stress in the metal. In fact, the metal layer had such high stress that the TaN layer near the nanowire channel has lifted as a result, which can be seen clearly.

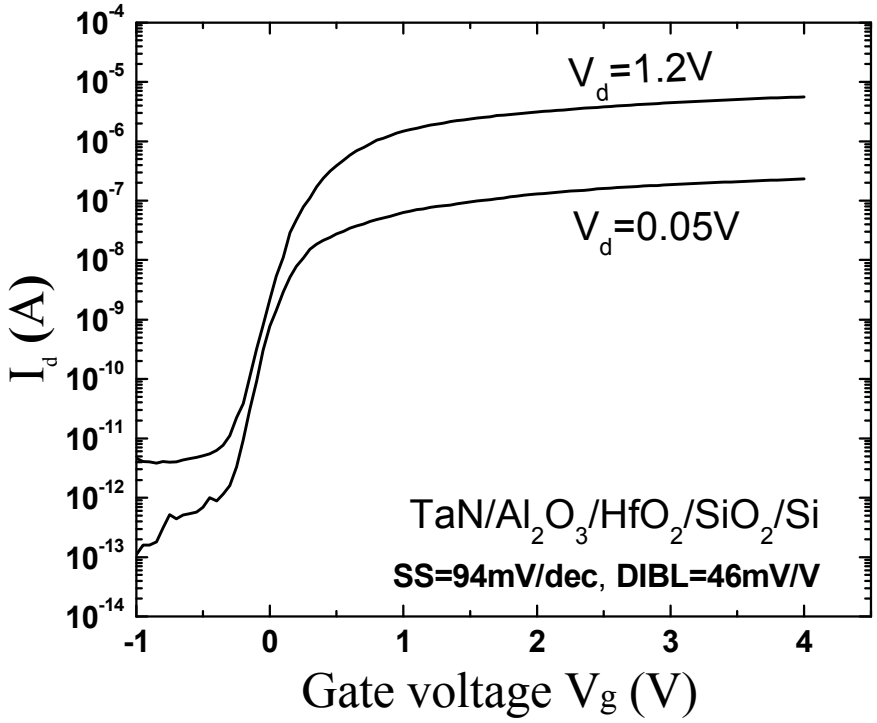


Fig. 5.4: The I_d-V_g curve measured from the nanowire TAHOS device shows good subthreshold and electrostatic property, despite a thick EOT used in memory.

We first characterized the memory cells for transfer properties of the TAHOS nanowire memory cell with channel length of 300-nm. Both the TAHOS nanowire device and SONOS nanowire device were then characterized for P/E operations using the F-N tunneling mechanism by applying voltage stress on the gate and keeping the source and drain grounded. As measured, TAHOS nanowire device I_d-V_g characteristics are shown in Fig. 5.4. DIBL of <50 mV/V and sub-threshold swing of

<100 mV/dec are achieved respectively. Despite the 22-nm physical thickness of the gate dielectric (κ value of HfO_2 and Al_2O_3 are considered as 25 and 9 [5.12]) and a single implant step, these values are as small as those of MOSFET transistors. This is attributable to the advanced GAA structure which leads to superior gate controllability.

Fig. 5.5 depicts the transient memory characteristics of the nanowire TAHOS device as well as the nanowire SONOS control device at the onset of the P/E. Memory windows of both devices are kept closed until ± 7 V stress pulse was applied, if we define the memory window open as ΔV_{th} equals to 0.5 V. It shows that the programming window starts to open for the nanowire TAHOS device at +7 V pulse height for a 3 ms pulse width. However, the pulse duration to open the same window for the nanowire SONOS device is longer than 10 ms. Furthermore, the erasing process did not start at -7 V pulse for the nanowire SONOS memory device.

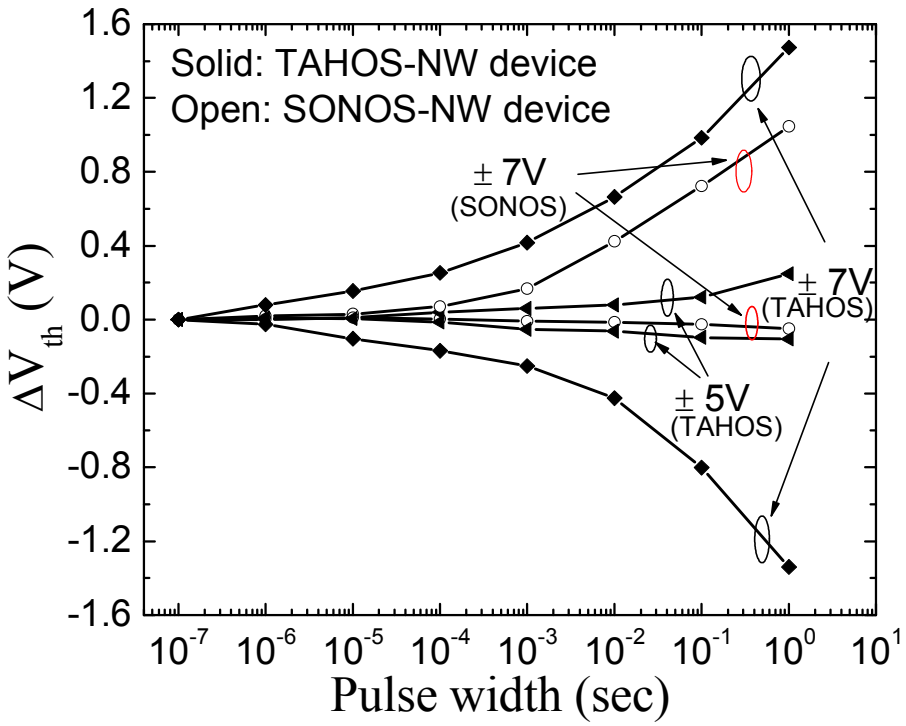


Fig. 5.5: Transient memory characteristics of the nanowire TAHOS (NW-TAHOS) device as well as the nanowire SONOS (NW-SONOS) control device at the onset of the P/E. The TAHOS device shows equivalent program and erase speed while the SONOS device does not.

The program/erase transient characteristics of a single TAHOS memory cell are shown in Fig. 5.6. Large threshold voltage shift (V_{th} shift) of 3.8 V could be achieved by the nanowire TAHOS memory in as short as 10 μs at a programming stress of 13 V and in 200 μs at an erasing stress of – 12 V. Fig. 5.7 shows the transient speed characteristics of the nanowire SONOS device as a comparison. It could be observed that the TAHOS device displays much faster memory speed, especially for erasing, as compared to its counterpart. It is already known that large speed imbalance exists during program and erase at the first hundreds of micro-second which seriously retards the fast cycling of SONOS devices. The reduced electrical oxide thickness (EOT) in the TAHOS device is one of the factors which causes the increase of memory speed. An overall electric field enhancement of 17% could be achieved in TAHOS when the same pulse magnitude is applied due to the higher κ value of the charge trapping layer and blocking layer, Smaller EOT of the charge trapping layer and blocking layer also results in a larger portion of electric field dropping across the tunnel oxide, further enhancing the carrier injection from the nanowire channel. Hence the employment of high-κ stack could reduce the voltage required to write and erase the memory cell. It is noted the utilization of high-κ dielectric materials and TaN metal gate can greatly improve the erase characteristics. The smaller electric field at the blocking oxide effectively reduces the electron current from the gate. Higher work function of TaN (4.5~4.6eV) relative to n+ poly-Si (4.15~4.2eV) could also effectively increase the tunneling barrier height to suppress the gate electron injection when the negative pulse is applied on the gate [5.13].

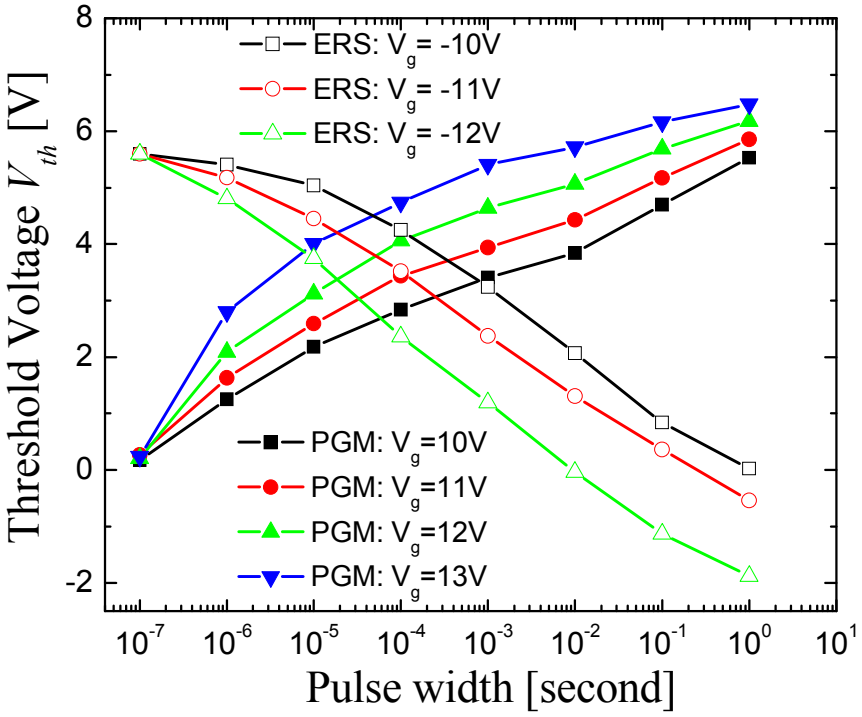


Fig. 5.6: Program and erase characteristics of nanowire TAHOS memory (TAHOS-NW in the figure). Large memory window and especially enhanced erase speed can be seen.

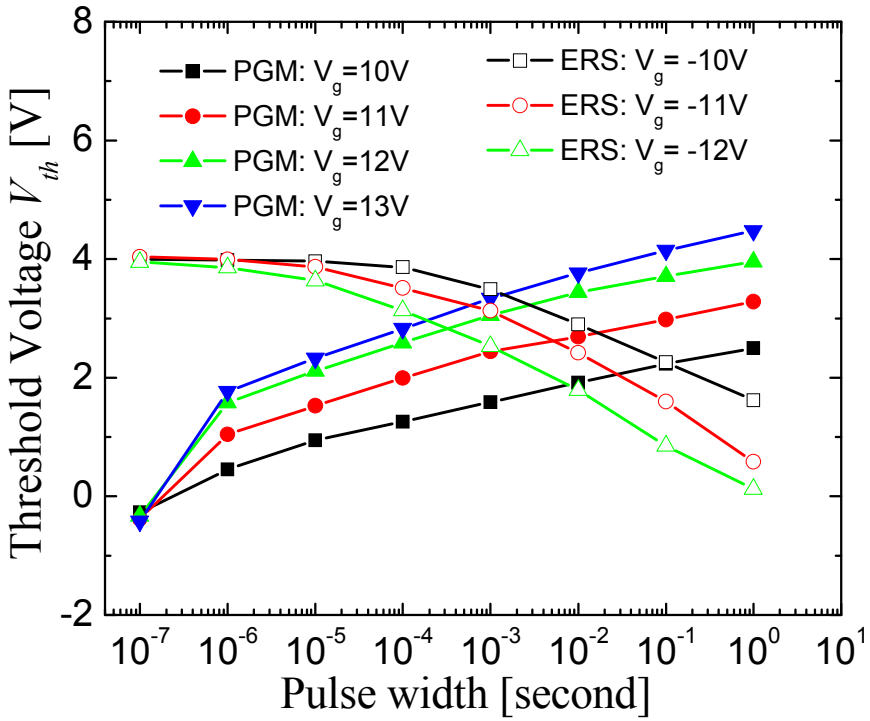


Fig. 5.7: Program and erase characteristics of nanowire SONOS memory device (SONOS-NW in the figure). The erasing speed is much less than the programming speed which greatly retards the cycling of SONOS.

We have not measured the work function of TaN for nanowire memory devices due to inaccuracies in measurement of Capacitance of nanowire devices on SOI wafer. We used previous experimental data in which the TaN was deposited using the same PVD tool and same process recipe to estimate the work function of the TaN metal gate [5.14]. The work function of TaN is 4.5~4.6eV according to that work. Compared to 4.15~4.2eV n+ poly-Si gate work function, the barrier height for gate electron tunneling during erasing for the TaN gate device can be confirmed to be much higher than the one with n+ poly-Si gate. Hence it can be concluded that part of the reason for improved erasing characteristics could come from the employment of the TaN gate electrode.

Besides the advantages of reduced EOT, enhanced electric field in the tunnel oxide as well as the higher gate work function for nanowire TAHOS devices, a proper conduction and valence band offset (ΔE_c and ΔE_v) change with respect to SONOS is also favorable for P/E characteristics. The band diagrams of TAHOS and SONOS during the program and erase mode are compared and shown in Fig. 5.8, in which the solid line indicates the TAHOS device while the dash line is used to represent the SONOS device, and the band offset data are from published literature [5.7, 5.12]. The Fermi-level of the gate position of both devices should be put at the same level since we used the same magnitude of voltage stress during characterization. The 1.7 eV ΔE_c between HfO_2 and SiO_2 is more favorable as compared to 1.1 eV ΔE_c between Si_3N_4 and SiO_2, since the electron could tunnel through the band gap and into the conduction band of the charge trapping layer earlier. It should be mentioned here that plotting of the energy band does not take into account of the influence of the nanowire structure due to two reasons. Firstly, the nanowire we obtained in this work shows a relatively trapezoidal cross section rather than a circular one, and we did not calculate

the band diagram quantitatively for the trapezoidal nanowire. Secondly, enhancement of the transient property could only be related to the different gate dielectric materials used since nanowires for both devices are the same. Hence we only compared the band diagrams qualitatively without taking into account the effect of GAA nanowire curvature.

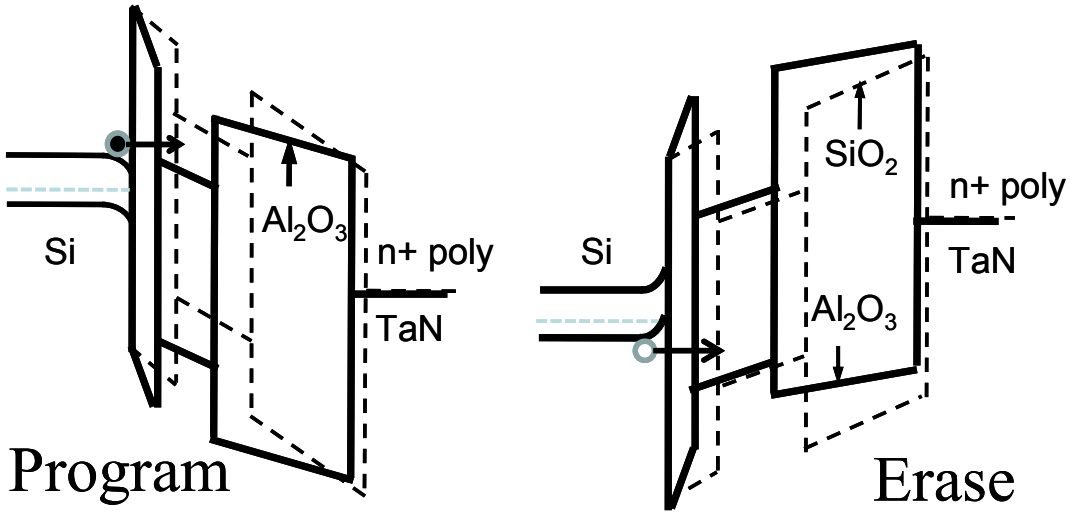


Fig. 5.8: Band diagram of nanowire TAHOS (solid line) and nanowire SONOS (dash line).

Fig. 5.9 depicts the endurance characteristics of the nanowire TAHOS memory device. The memory window of 1.45 V remains constant when the device was cycled at a smaller stress voltage. The memory window of 2.67 V was achieved when the device was under ± 11V stress voltage cycling conditions and it shows a slight upward shift. The phenomenon of the slight upward shift of the memory window could be explained by the generation of oxide traps at the interface between the channel and the deposited TEOS oxide.

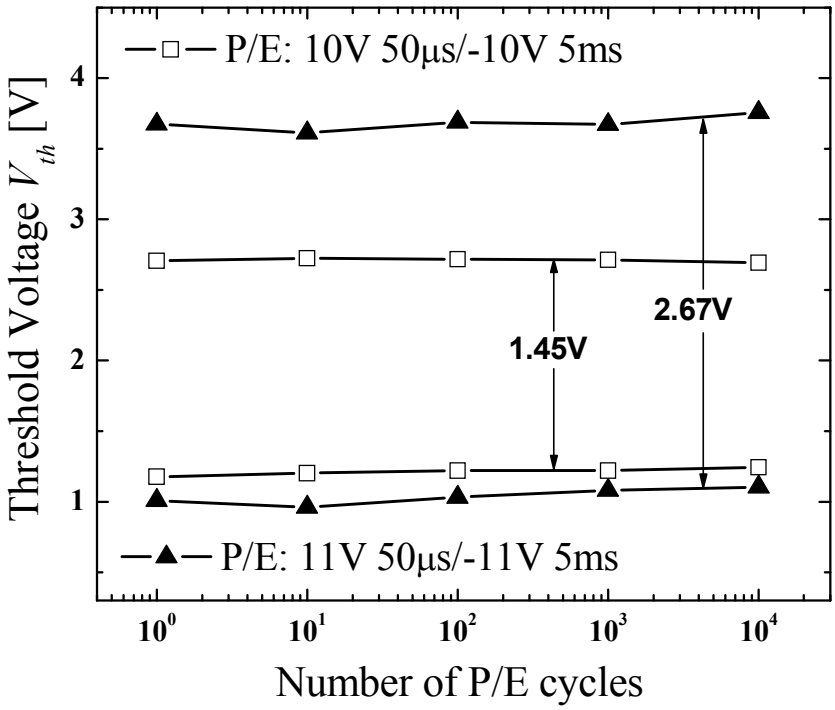


Fig. 5.9: P/E endurance of nanowire TAHOS device at two cycling conditions. The larger memory window with 2.67 V magnitude was achieved when the device was under ±11V stress voltage cycling conditions with a slight upward shift.

Room-temperature retention for the nanowire TAHOS nonvolatile memory device is also tested and shown in Fig. 5.10. A large initial window of 3.73 V is obtained under 13 V/100 μs and -12 V /5 ms retention conditions, which decreases to 3.51 V after 10^4 seconds. The decay rate is 50mV/dec for the measured time period. The nanowire SONOS device achieved an initial window of 3.45 V under 13 V/1 ms and -12 V /50 ms retention conditions, which dropped to a magnitude of 3.33 V after 10^4 seconds. As can be seen, retention of the nanowire TAHOS nonvolatile memory device is slightly degraded as compared to retention of the nanowire SONOS device – the magnitude of the memory window is reduced by 0.22 V for TAHOS and 0.12 V for SONOS in 10^4 seconds from almost the same initial window, and the charge loss are about 5.8% and 3.5% respectively. We suspect that degradation of the memory

retention behavior in the nanowire TAHOS device may come from the states at interface between the high-κ/oxide or high-κ/metal gate since both devices possess an identical tunnel oxide thickness. However, technical merit of the nanowire TAHOS device still remains since larger memory window could be achieved with much lesser power consumption by the nanowire TAHOS device.

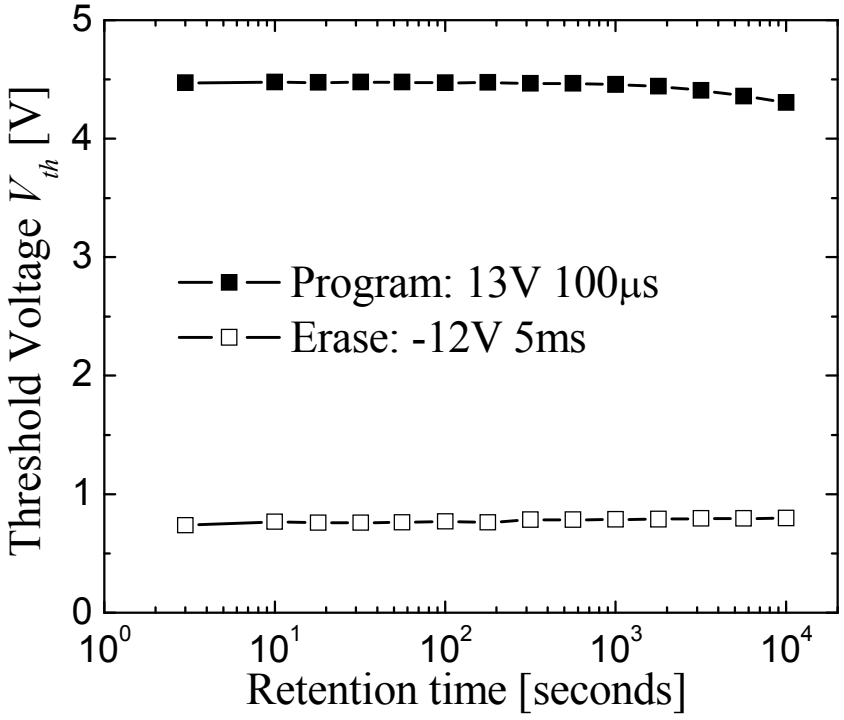


Fig. 5.10: Room temperature data retention of the nanowire TAHOS device. The same magnitude of memory window can be achieved by a smaller P/E voltage as compared to its counterpart, despite slightly larger charge loss at the same measurement time period.

5.4 Conclusion

We have proposed a novel MONOS memory device fabricated on the GAA nanowire structure for high density nonvolatile memory application, using a CMOS-compatible process technology. The GAA nanowire structure is currently one of the best scalable device structures for the VLSI memory circuit design. Isotropic and conformal ALD machine tool ensures good growth of dielectrics surrounding the

nanowire, although some empty space exists below the nanowire channel due to the directional nature of PVD deposition. Results reveal that both programming and erasing transient properties have been enhanced as compared to nitride-based SONOS nonvolatile memory. The erasing speed is also further improved, which is attributed to decreased electric field across the blocking oxide as well as reduced electron injection from the gate electrode. Hence the nanowire TAHOS charge trapping memory with fast P/E speed, reduced P/E and CMOS-compatible fabrication method shows promise as a potential alternative for next generation nonvolatile flash memory.

Reference

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

[5.2] J. D. Lee, S. H. Hur, and J. D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation", in *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264-266, 2002.

[5.3] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS", in *IEEE Circuits and Devices*, vol.16, pp. 22-31, 2000.

[5.4] Y.-H. Hsiao, H.-T. Lue, M. Y. Lee, S.-C. Huang, T.-Y. Chou, S.-Y. Wang, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "A study of SONOS charge loss mechanism after hot-hole stressing using trap-layer engineering and electrical re-fill methods", in *Reliability Phys. Symp.*, 2008, pp. 695-696.

[5.5] H.-C. You, T.-H. Hsu, F.-H. Ko, J.-W. Huang, L.-H. Yang, and T.-F. Lei, "SONOS-type flash memory using an HfO_2 as a charge trapping layer deposited by the sol-gel spin-coating method", in *IEEE Electron Device Lett.*, vol. 27, no. 8, pp. 653-655, 2006.

[5.6] Y. Q. Wang, G. Samudra, D. Y. Gao, C. Shen, W. S. Hwang, G. Zhang, Y.-C. Yeo, and W. J. Yoo, "Fast erasing and highly reliable MONOS type memory with HfO_2 high-k trapping layer and Si_3N_4/SiO_2 tunneling stack", in *IEDM Tech. Dig.*, pp. 971-974, 2006.

[5.7] W. Wang, D.-L. Kwnag, "A novle high-k SONOS using TaN/ Al_2O_3/ Ta_2O_5/ HfO_2/Si structure for fast speed and long retention operation", in *IEEE Trans.*

Electron Devices, vol. 53, no. 1, pp. 78-82, 2006.

[5.8] S. D. Suk, K. H. Yeo, K. H. Cho, M. Li, Y. Y. Yeoh, K. H. Hong, S. H. Kim, Y. H. Koh, S. G. Jung, W. J. Jang, D. W. Kim, D. G. Park, and B. I. Ryu, "Gate-all-around twin silicon nanowire SONOS memory", in *VLSI Technology Symp*., 2007, pp. 142-143.

[5.9] J. Fu, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, "Si-nanowire based gate-all-around nonvolatile SONOS memory cell", in *IEEE Electron Device Lett.*, vol. 29, no. 5, pp. 518-521, 2008.

[5.10] J. Fu, Y. Jiang, N. Singh, C. X. Zhu, G. Q. Lo, and D. L. Kwong, "Polycrystalline Si nanowire SONOS nonvolatile memory cell fabricated on a gate-all-around (GAA) channel architecture", in *IEEE Electron Device Lett.*, vol. 30, no. 3, pp. 246-249, 2009.

[5.11] C.-T. Wei, and H.-P. D. Shieh, "Effects of processing variables on tantalum nitride by reactive-ion-assisted magnetron sputtering deposition", in *Jpn. J. Appl. Phys.*, vol. 45, no. 8A, pp. 6405-6410, 2006.

[5.12] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, and B. J. Cho, "Hafnium aluminum oxide as charge storage and blocking –oxide layers in SONOS-type nonvolatile memory for high-speed operation", in *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 654-662, 2006.

[5.13] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories", in *IEDM Tech. Dig.*, 2003, pp. 613-616.

[5.14] C. Ren, D. S. H. Chan, W. Y. Loh, S. Balakumar, A. Y. Du, C. H. Tung, G. Q. Lo, R. Kumar, N. Balasubramanian, and D. -L. Kwong, "Work-Function

tuning of TaN by high-temperature metal intermisxing technique for gate-first CMOS process", in *IEEE Electron Devices Lett.*, vol. 27, no. 10, pp. 811-813, 2006.

Chapter 6

Conclusions

This thesis focuses on the study of nonvolatile memory devices with the integration of ultra-thin nanowire channel. The nanowire channel was specially developed and fabricated using the CMOS compatible top-down process technology.

This study started with the development of top-down nanowire fabrication method. The nanowire fabrication process was initiated on SOI wafer and took advantage of the advanced lithography technology and Si oxidation process. SOI wafer was used not only because of the thin layer of silicon which could eliminate the subsurface punch-through observed in bulk-Si devices into the nonvolatile memory, but also because of the feasibility of obtaining one or two thin nanowires from a fin on BOX. The nanowire could not be found using the same process technology on bulk wafer after the oxidation step due to many uncontrollable parameters. The diameter of the nanowire could reach 5-nm or even smaller with this method, which requires a much longer oxidation time of more than 4 hours as the oxidation rate is almost zero when the oxidation comes to the self-limiting stage. The precise lithographic tool and the long and slow oxidation step at a moderate temperature ensure good uniformity among nanowires with the same design, while the CMOS-compatible fabrication method guarantees a sound repeatability of the nanowire process.

Nanowire SONOS discrete charge storage nonvolatile memory was fabricated using the developed nanowire fabrication. Dielectrics layers and gate material were

deposited using LPCVD. The good isotropic property of LPCVD ensures a good coverage of those materials on the nanowire channel, realizing the possibility of a gate-all-around structure. The gate controllability was greatly upgraded due to the gate-all-around structure, reflected by the SS and DIBL values which were as small as those in MOSFET devices. It was found that Si-nanowire SONOS devices with the same dielectrics thickness demonstrated a much enhanced memory transient speed as compared to planar structure SONOS devices. The theoretical study indicated that the electric field and potential-energy profiles across the dielectric layers exhibited many differences as compared to planar structure SONOS devices, through device simulations which took quantum-mechanical confinement effects into account. The cylindrical geometry greatly enhanced the electric field at the Si-SiO_2 interface, which increased the carrier-injection probability and reduced the electron tunneling length, increasing the memory speeds in the cylindrical nanowire memory device. Hence fast memory speed could be achieved with low voltage stresses, and the low voltage cycling conditions could benefit the reliability of the memory device.

The nanowire structure was subsequently integrated into thin film nonvolatile memory devices as a promising way to resolve issues that exist in conventional TFT SONOS memory devices. Previously, additional processes which could increase the grain size of poly-Si film were typically used to reduce the negative effect induced by the poly-Si grains. Neither laser annealing nor MILC is applicable in the CMOS technology. The integration of the nanowire structure could reduce the negative effect from poly-Si grains in a contrary way. This method was used to reduce the channel size besides increasing the grain size of the poly-Si film. The nanowire channel obtained an ultra-thin channel cross-sectional area in which a few grains could be formed. Thus the poly-Si nanowire TFT SONOS device exhibited much improved

subthreshold characteristics and memory properties than other reported TFT SONOS devices.

The SONOS memory still faces issues of a non-scalable gate dielectric and lower erasing speed. However, the MONOS-type nonvolatile memory has the advantages of smaller EOT and reduced electric field located at the blocking oxide. Hence high-κ and metal gate materials were used in the nanowire structure nonvolatile device with the purpose of resolving the aforementioned issues. The first demonstrated nanowire TAHOS memory cell exhibited increased erasing memory speed due to the reduced EOT and enhanced electric field at the tunnel oxide. The use of high work function metal gates suppresses gate electron injection and also improves the erase speed to the equivalent of the programming speed, which realizes a good P/E speed balance.

This study has thus far achieved a reliable technique for the top-down nanowire nonvolatile memory device fabrication. It ensures a reliable SCE immunity and fast memory transient characteristics with a relatively low stress voltage usage. Further investigations should also be conducted in other CMOS-compatible technical environments as the top-down technique can be easily carried out. Several recommendations are listed below based on the limitations we faced in this study for further possible studies:

(1) Due to the non-availability of high temperature measurement facility, the high temperature retention was not tested in all our nanowire memory devices. There is however a way of extracting the trap density distribution, which is to carry out the retention measurement at different temperatures up to about 200°C, because the thermal emission process from traps is dominant with respect to tunnel mechanisms at high temperatures. Either a SONOS or TAHOS device could be used to carry out such

measurement, which is useful in studying the physics related to the trap density distribution in the nanowire memory cell.

(2) More suitable high-κ and metal gate materials could be integrated into the nanowire devices. Although improved P/E speed and vertical scaling was realized in the nanowire TAHOS device discussed in Chapter 5, the charge retention properties was slightly degraded as compared to the referenced nanowire SONOS device, reasons for which are still unknown at this moment. Hence more research work could be done to improve the charge retention characteristics in the nanowire MONOS-type device by integrating more suitable types of gate stack materials, especially advanced high-κ materials.

(3) The vertical Si pillar nanowire was successfully fabricated recently. However, the vertical Si nanowire SONOS memory is still under the development stage. Vertical Si nanowire has the advantage of high device density as compared to lateral nanowire. The application of vertical nanowire in memory is especially promising due of its capability of fabricating three-dimensional multilevel memory structures.

List of Publications

Journal:

[1] **J. Fu**, N. Singh, C. X. Zhu, G. Q. Lo, D. L. Kwong, "Integration of high-κ dielectrics and metal gate on gate-all-around Si-nanowire based architecture for high speed non-volatile charge trapping memory", in ***IEEE Electron Device Letter***, vol. 30, no. 6, pp. 662-664, June 2009.

[2] **J. Fu**, Y. Jiang, N. Singh, C. X. Zhu, G. Q. Lo, D. L. Kwong, "Polycrystalline Si nanowire SONOS nonvolatile memory cell fabricated on a gate-all-around (GAA) channel architecture", in ***IEEE Electron Device Letter***, vol. 30, no. 3, pp. 246-249, March 2009.

[3] **J. Fu**, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, "Si-nanowire based gate-all-around nonvolatile SONOS memory cell", in ***IEEE Electron Device Letter***, vol. 29, no. 5, pp. 518-521, May 2008.

Conference:

[1] **J. Fu**, Y. Jiang, N. Singh, C. X. Zhu, G. Q. Lo, N. Balasubramanian,D. L. Kwong, "Low temperature GAA poly-Si nanowire TFT SONOS memory for MLC application", in ***International Conference on Solid State Devices and Materials (SSDM-2008)***, Tsukuba, Japan, Dec 2008, pp. 822-823.

[2] **J. Fu**, N. Singh, B. Yang, C. X. Zhu, G. Q. Lo, D. L. Kwong, "Si-nanowire TAHOS (TaN/Al_2O_3/HfO_2/SiO_2/Si) nonvolatile memory cell", in ***European Solid-State Device Research Conference (ESSDERC-2008)*** in Edinburgh, Scotland, Sep 2008, pp. 115-118.

[3] **J. Fu**, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, "Trap layer engineering gate-all-around vertically stacked twin Si-nanowire nonvolatile memory", in ***International Electron Device Meeting (IEDM-2007)***, Washington DC, USA,

Dec 2007, pp. 79-82.

[4] E. Gnani, S. Reggiani, A. Gnudi, G. Baccarani, **J. Fu**, K. D. Buddharaju, N. Singh, D. L. Kwong, "Modeling of nonvolatile gate-all-around charge-trapping SONOS memory cells", accepted in ***European Solid-State Device Research Conference (ESSDERC-2009)*** in Athens, Greece, Sep 2009.

[5] B. Yang, K. D. Buddharaju, S. H. G. Teo, **J. Fu**, N. Singh, G. Q. Lo, D. L. Kwong, "CMOS compatible gate-all-around vertical silicon-nanowire MOSFETs", in ***European Solid-State Device Research Conference (ESSDERC-2008)*** in Edinburgh, Scotland, Sep 2008, pp. 318-321.