



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Ph.D. THESIS

Channel-Stacked NAND Flash Memory with High- κ Charge Trapping Layer for High Scalability

집적도 향상을 위하여 고유전체 전하저장층을 적용한
채널 적층형 낸드 플래시 메모리

BY

JOO YUN SEO

February 2016

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

ABSTRACT

Exploding demands for mobile devices induce the drastic expansion of the market of NAND flash memory as high density storage devices. Three-dimensional (3D) NAND flash memory paved a new way of increasing the memory capacity by stacking cells in three-dimension. For stacked NAND flash memory, the thickness of ONO (memory dielectric layers) will be a roadblock in scaling-down of the minimum feature size, while channel diameter can be scaled down to < 20 nm. It is challenging to reduce the thickness of oxide-nitride-oxide (ONO) layer, since the charge trapping properties degrade when the Si_3N_4 is made thinner.

.In this thesis, the channel stacked NAND flash memory array (CSTAR) with high- κ charge trapping layer for high scalability is proposed. To adopt high- κ layer into 3D NAND, its memory characteristics were evaluated with capacitors and gate-all-around flash memory devices. Finally, 4-layer channel stacked memory with high- κ layer was successfully fabricated and

characterized.

Recent trend of nonvolatile memories are introduced and the overview of 3D stacked NAND flash memory technology is presented in Chapter 1 and 2. In Chapter 3, the memory characteristics of high- κ layer were evaluated with fabricated capacitors and flash memory devices. In Chapter 4, fabrication process and electrical characteristics of CSTAR with high- κ are shown. With the comparison with previous works using ONO layer, CSTAR with high- κ is evaluated. In Chapter 5, the novel operation method of CSTAR is presented. Using TCAD and measurement, a newly designed operation method is verified.

Keywords: 3-D NAND flash memory, stacked array (STAR), charge trap flash (CTF), high- κ charge trap memory

Student number: 2012-30938

CONTENTS

Abstract.....i

List of Figures.....vii

Chapter 1

Introduction.....1

1.1 Flash Memory Technology.....1

1.2 Flash Memory Unit Cell and Array Structure6

1.3 NAND Cell Operation13

1.4 Charge Trap Flash Memory25

Chapter 2

3-D Stacked NAND Flash Memory.....28

2.2 Examination of Previous 3-D Stacked NAND Flash.....	28
2.2.1 Gate Stack Type 3-D NAND Flash Memory.....	28
2.2.2 Channel Stack Type 3-D NAND Flash Memory.....	36
2.2.3 Comparison between the Gate Stack Type and the Channel Stack Type.....	45

Chapter 3

Channel Stacked NAND Flash Memory with high- κ

Charge Trapping Layer.....48

3.1 Introduction.....	48
-----------------------	----

3.2 Memory Characteristics of HfO ₂	52
--	----

3.3 Fabrication of Nanowire Memory Devices with high- κ Dielectric Layer.....	56
---	----

Chapter 4

Fabrication of Channel Stacked NAND Flash Memory

with High- κ66

4.1 Introduction66

4.2 Fabrication Method67

4.3 Key Process Steps of CSTAR with high- κ 72

4.3.1 Single Crystalline Silicon Channel72

4.3.2 Fin Patterning74

4.3.3 Stacked Nanowires76

4.4 Measurement Results81

4.5 Comparison with Previous Works88

Chapter 5

Novel Program Operation in CSTAR92

5.1 Introduction.....92

5.2 Simulation Results.....93

5.3 Measurement Results103

Chapter 6

Conclusions.....106

Bibliography.....108

Abstract in Korean.....116

List of Figures

Chapter 1

Fig. 1.1. Classification of semiconductor memories	2
Fig. 1.2. Flash memory surpasses DRAM market	4
Fig. 1.3. Flash memory application demand.....	5
Fig. 1.4. Unit cell structure of FG NAND flash memory.....	6
Fig. 1.5. (a) Cross-sectional view of NOR flash memory. (b) Equivalent circuit schematic of NOR flash memory.....	8
Fig. 1.6. (a) Cross-sectional view of NAND flash memory. (b) Equivalent circuit schematic of NAND flash memory.....	9
Fig. 1.7. Comparison between NOR and NAND flash memory...	10
Fig. 1.8. NAND flash memory array schematic	12
Fig. 1.9. Fowler-Nordheim tunneling for program operation mechanism.....	13
Fig. 1.10. Program operation scheme of NAND flash memory array	16
Fig. 1.11. Self-boosted program inhibit scheme and channel potential of strings	16
Fig. 1.12. Erase operation scheme of NAND flash memory array	17

Fig. 1.13. Read operation of NAND flash memory	20
Fig. 1.14. Basic sensing scheme and the related timing diagram..	20
Fig. 1.15. Sensing circuit using parasitic BL capacitance (C_{BL}) and its timing diagram	23
Fig. 1.16. Bit line parasitic capacitors	24
Fig. 1.17. Sensing circuit of ABL architecture and its timing diagram	24
Fig. 1.18. Comparison between FG and SONOS related with band diagram and material properties	25

Chapter 2

Fig. 2.1. Classification of 3-D NAND flash memories according to the stacking type. (a) Gate stack type (b) Channel stack type	30
Fig. 2.2. BiCS architecture and its string structure	31
Fig. 2.3. Equivalent circuit of BiCS flash memory	31
Fig. 2.4. (a) Schematic of P-BiCS flash memory (b) Fabrication of process of P-BiCS	33
Fig. 2.5. Bird's eye view and cross sectional view of TCAT flash memory	35
Fig. 2.6. (a) Equivalent circuit of TCAT (b) 'gate replacement' process method	35
Fig. 2.7. (a) Structure of VG-NAND (b) Top view of active array	37

Fig. 2.8. (a) Equivalent circuit of VG-NAND (b) Schematic of SSLs and its operation table for a layer selection	37
Fig. 2.9. Structure of 3D VG NAND	38
Fig. 2.10. Structure of the pn diode decoded 3D VG NAND architecture	39
Fig. 2.11. 3D array of single crystal Si nanowires introduced by CEA-LETI	41
Fig. 2.12. (a) Bird's eye view of the STAR NAND flash memory. (b) The fabricated array having two SSL (SST_{dep} and SST_{enh}) for layer selection.....	42
Fig. 2.13. The bird's eye view of the conventional CSTAR, its equivalent circuit, and operation scheme	43
Fig. 2.14. The bird's eye view of CSTAR with LSM and its layout top view	43
Fig. 2.15. Summary of gate stack type and channel stack type ...	47
Fig. 2.16. Trend of bit line current of 3D NAND as a function of the number stacked layers and the hole diameter	47

Chapter 3

Fig. 3.1. (a) $C-V$ curve of the sample with HfO_2 charge trapping layer, $V_{pgm} = 13$ V (b) The amount of the shifted flatband voltage as the program pulse is applied repeatedly	54
---	----

Fig. 3.2. (a) Program speed with respect to the thickness of HfO_2	
---	--

trapping layer (b) ISPP slope with respect to the thickness of Al ₂ O ₃ blocking oxide	55
Fig. 3.3. A schematic view and transmission electron microscopic (TEM) images of the cross-section of single crystalline silicon nanowire	58
Fig. 3.4 (a) Program and erase characteristics of planar SAHOS device ($W_{ch} = 1 \mu\text{m}$. $L_G = 1 \mu\text{m}$) (b) Program and erase characteristics of gate-all-around SAHOS device	59
Fig. 3.5. Program and erase speed of gate-all-around SAHOS device	62
Fig. 3.6. (a) Retention characteristics of SAHOS device at 85°C. (b) Endurance characteristics of SAHOS device. (Inset) Transfer characteristics after program/erase cycles	64

Chapter 4

Fig. 4.1. Fabrication process of channel stacked array with high- κ	71
Fig. 4.2. A TEM image of SiGe/Si/... epitaxially grown layers ...	73
Fig. 4.3. An SEM image of silicon fins with steep etch slope	75
Fig. 4.4. An SEM image showing the result of selective SiGe etch	77
Fig. 4.5. An SEM image showing the result of selective SiGe etch using APM solution	78

Fig. 4.6. A schematic of stacked nanowires and a TEM image of cross-section view of stacked Si nanowires surrounded by SiO ₂ , HfO ₂ , Al ₂ O ₃ , and Ti	80
Fig. 4.7. CSTAR array and its fabricated 4-type devices	81
Fig. 4.8. Microscope top view of fabricated device	82
Fig. 4.9. TEM image of cross sectional view of fabricated array	83
Fig. 4.10. Bit line current vs. cell gate voltage	84
Fig. 4. 11. Program and erase characteristics of fabricated array ..	85
Fig. 4. 12. Program and erase speed of fabricated array	86
Fig. 4. 13. Reliability of fabricated array	87
Fig. 4. 14. Comparison of program/erase speed. (a) this work (b) ONO devices	90
Fig. 4. 15. Energy band diagram in erase operation (a) planar devices (b) nanowire devices	91

Chapter 5

Fig. 5.1. ISPP wave form and I-V curve of circular channel	94
Fig. 5.2. ISPP characteristics of circular channel and elliptic channel	94
Fig. 5. 3. Electron trapped charge density distributions for the elliptic channel	96
Fig. 5.4. Electron current density distributions for the elliptic channel	96

Fig. 5.5. Comparison of ISPP for various starting program voltages of circular channel.....	100
Fig. 5.6. Comparison of ISPP for various BL voltages of circular channel	100
Fig. 5.7. CSTAR structure and its equivalent circuit diagram and BL biasing of new programming method	101
Fig. 5.8. Comparison of ISPP for various BL voltages of circular channel	101
Fig. 5.9. Comparison of the required number of program pulses in the conventional and new programming methods	102
Fig. 5.10. (a) Proposed program method. (b) Incremental step pulse program (ISPP) operation of fabricated device. (c) Program speed variation between fast cell (a=30 nm) and slow cell (a=60 nm) when conventional program method is applied.	104
Fig. 5.11. (a) We can adjust program speed by varying V_{BL} and starting V_{pgm} . (b) By applying V_{BL} to a fast cell, we can reduce the program speed variation	105

Chapter 1

Introduction

1.1 Flash Memory Technology

Semiconductor memory is an essential part of modern information processors, and it has been more or less growing in density and performance in accordance with Moore's law. Semiconductor nonvolatile memory (NVM) is a major subset of solid-state memory. Nonvolatility means that the contents of the memory are retained when power is removed. NVM family includes electrically programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), and flash memory [1]. EPROM is erased by exposure to UV light and programmed electrically.

EEPROM can be both erased and programmed electrically. The use of “EE” implies block erasure. E²PROM can also be both erased and programmed electrically, but the difference is its byte alterability rather than block erasable. Flash memory is an EEPROM that can have a whole memory block erased, thus the name “flash.”

Toshiba Corporation firstly invented NOR-type and NAND-type flash technology in the 1980’s. And it is the Intel engineers to put the idea into

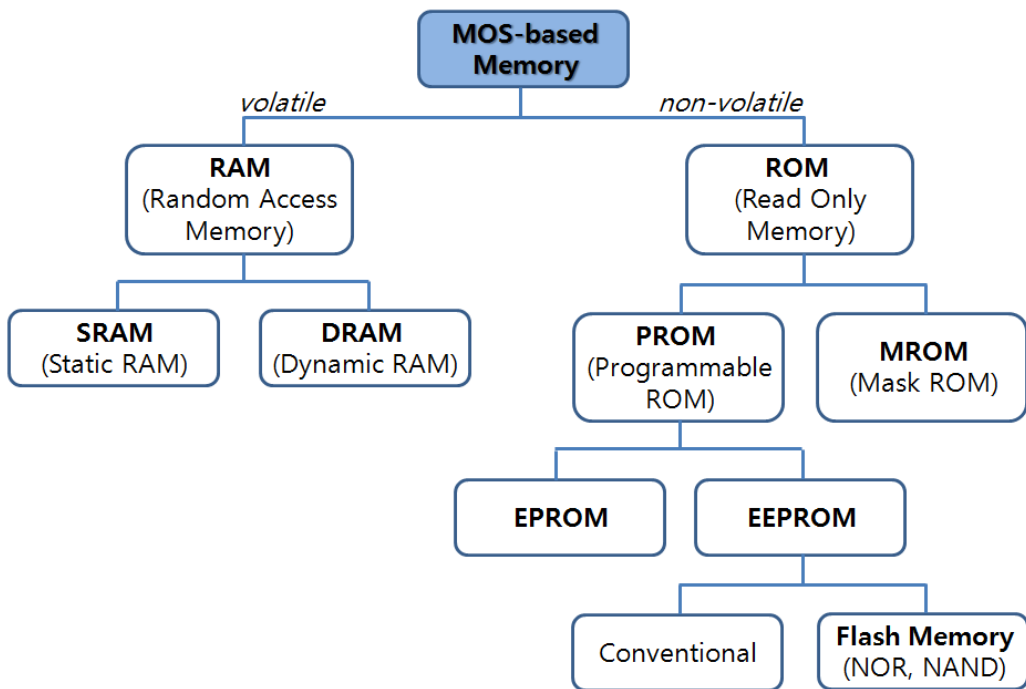


Fig. 1.1. Classification of semiconductor memories

practice for the first time. They called it “Flash EPROMs” to emphasize it can be erased “in a flash” electrically and without UV light. Because of its roots as an extension of EPROMs, Flash was easily adopted as easier-to-use EPROMs. NOR flash was the basis for the first Flash memory cards and non-volatile solid state drives in the 1990s. NAND-type flash came later and has remained primarily a data storage memory in the removable-cards arena. flash memory has been the dominant form of NVM for a long period of time, since it has high density and high scalability. The strong demand for digital data storage is driving the technology developments of extremely dense storage media. Driven by continued demand for smartphones, tablet PCs, and other personal media devices, the total flash memory market (NAND and NOR) is forecast to grow 2% to \$30.4 billion in 2012, surpassing the \$28.0 billion DRAM market in sales for the first time. With the exception of 2010, the DRAM and flash memory markets have been growing closer in size to each other for several years but demand for flash used in portable media devices, coupled with two years of weaker demand and price erosion for commodity DRAM used in

personal computers will finally be enough to push total flash sales beyond those for DRAM (Fig.1.2).

NAND flash sales are forecast to increase 14% annually from 2012-2017, growing to \$53.2 billion at the end of the forecast period while the DRAM market is forecast to grow 9% annually over this same time.

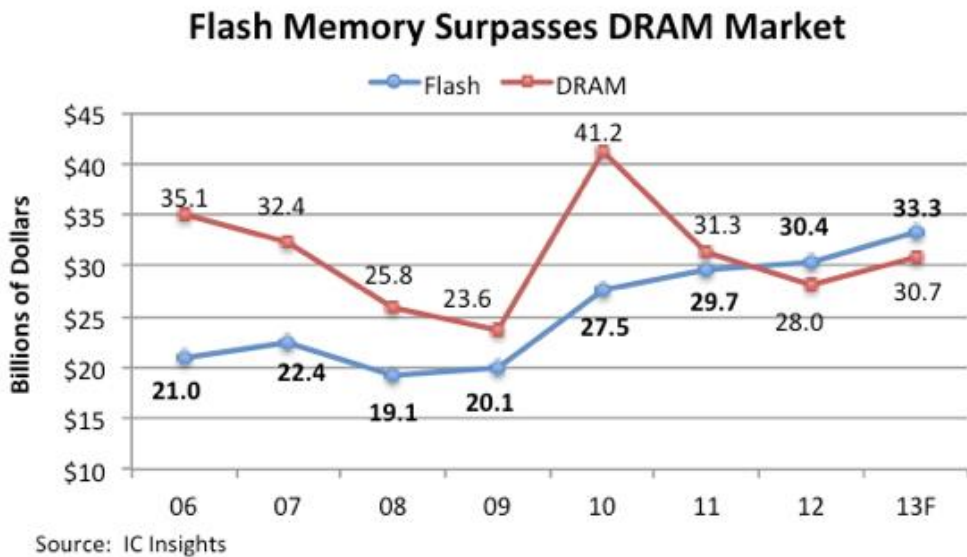


Fig. 1.2. Flash memory surpasses DRAM market (source:IC insights)

NAND flash is used for data storage of mobile devices, media card, and digital consumer electronics. Furthermore, NAND flash is drawing attention as a replacement of hard disk drives recently. Flash memory does not have the mechanical limitations and latencies of hard drives, so a solid-state drive (SSD) is attractive when considering speed, noise, power consumption, and reliability. NAND flash-based SSDs are expected to become a new driver of NAND flash bit consumption beyond 2012 driven by adoption of SSDs in mainstream consumer PCs.

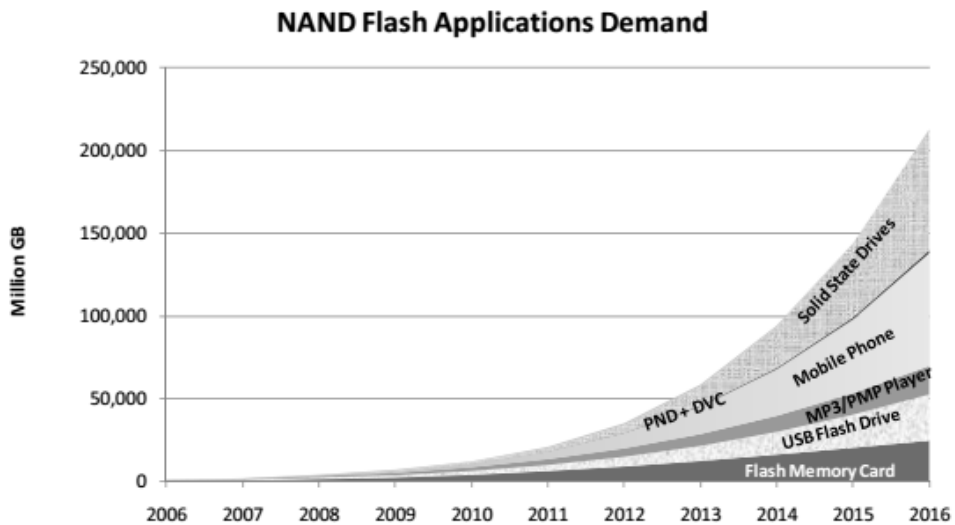


Fig. 1.3. Flash memory application demand (source:IC insights)

1.2 Flash Memory Unit Cell and Array Structure

The most popular NAND memory cell is based on the *Floating Gate* (FG) technology, whose structure is shown in Fig. 1.4. The floating gate usually consists of a polysilicon layer formed within the gate insulator of a field-effect transistor between the normal gate electrode (the control gate) and the channel. The operations performed to inject and remove electrons from the isolated floating gate are called program and erase, respectively. These operations modify the threshold voltage of MOS transistor. Applying a fixed read voltage to control gate, it is then possible to discriminate two information levels: when the read gate voltage is higher than the cell's threshold voltage, the cell is on ("1"), otherwise it is off ("0").

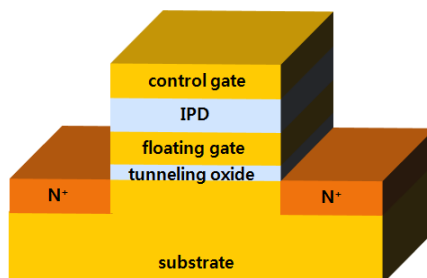
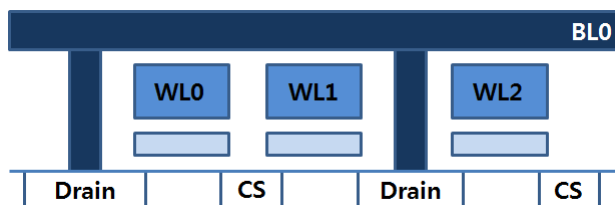


Fig. 1.4. Unit cell structure of FG NAND flash memory.

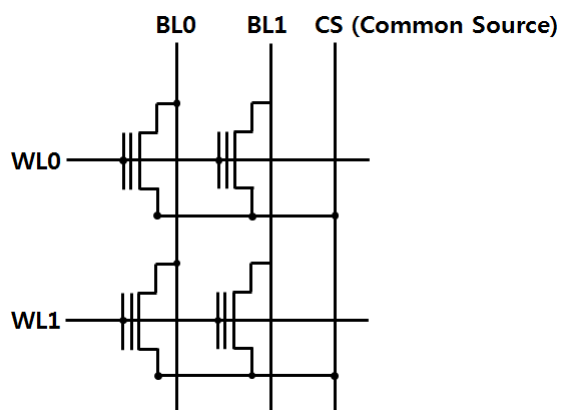
The most widely adopted flash memory array architectures can be grouped into two categories: NOR type and NAND type. NOR-type flash has inherited the array structure used in early ultraviolet (UV) EPROM architectures. Figure 1.5 shows the basic layout outline of a NOR-type flash array. NOR-type flash is characterized by high-speed random access and is very well suited to program storage applications. Cell programming is usually achieved by hot-electron injection and electrical erase by Fowler-Nordheim (FN) tunneling.

The most common use of NOR flash memory is for code and data storage in embedded applications since its random-access reads are fast and reliable. However, its programming and erasing speed is slow and requires larger areas allocated to a single bit compared to NAND flash memory.

On the other hand, NAND flash memory cells share resources, and this technology achieves a small cell area (Fig. 1.6). The NAND flash memory uses a uniform FN tunneling between the substrate and floating gate for both erasing and programming.

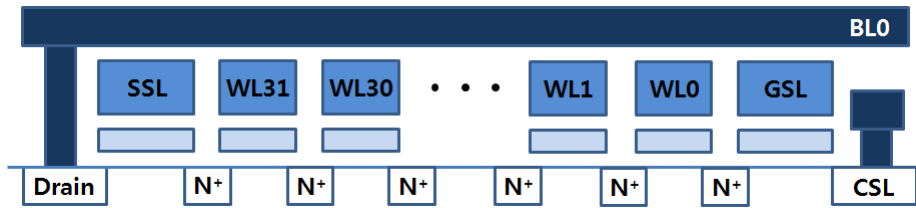


(a)

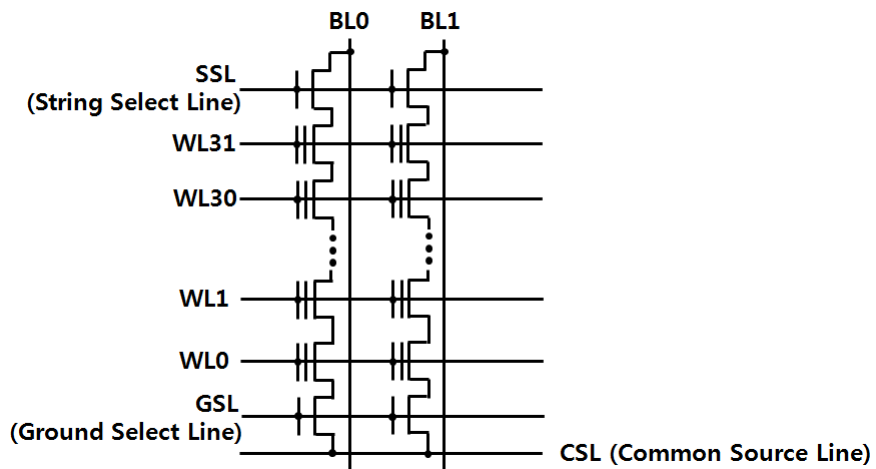


(b)

Fig. 1.5. (a) Cross-sectional view of NOR flash memory. (b) Equivalent circuit schematic of NOR flash memory.



(a)



(b)

Fig. 1.6. (a) Cross-sectional view of NAND flash memory. (b) Equivalent circuit

schematic of NAND flash memory.

The power consumption for programming does not significantly increase even when the number of memory cells to be programmed is increased.

The strengths of NAND flash memories are fast writes and high-density arrays, which lower die costs at the expense of random-access read capability.

Therefore, NAND is strongly suited to data-only applications such as audio storage, video storage, or data logging where fast data writes are required and the data will be accessed sequentially. Figure 1.7 summarizes the characteristics of two main flash memory architectures: NOR and NAND flash memories.

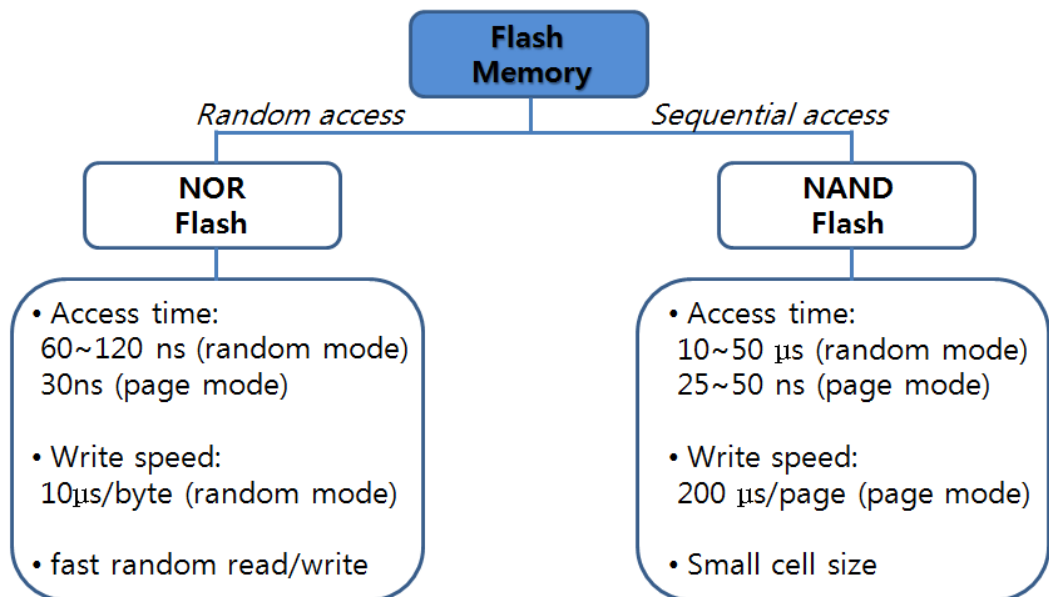


Fig. 1.7. Comparison between NOR and NAND flash memory.

The NAND flash memory array is grouped into a series of blocks, which are the smallest erasable units. One block is usually composed of 16, 32, or 64 pages. The read and program operation takes place on a page basis. The pages consist of cells belonging to the same word lines. The page size is proportional to the number of cells per word line and it is determined by the word line length limited by its RC delay and the bit line pitch limited by lithography. By increasing the page size, the number of cells operated in parallel program or read performance can be increased. All the NAND strings sharing the same group of word lines form a block. The block size is proportional to the length of string and it is limited by the circuit ability of handling its series resistance related with current drivability during the read operation. By increasing the block size, the array efficiency is increased, since more cells share both the

select transistors (SSL Tr. and GSL Tr.) and contact area.

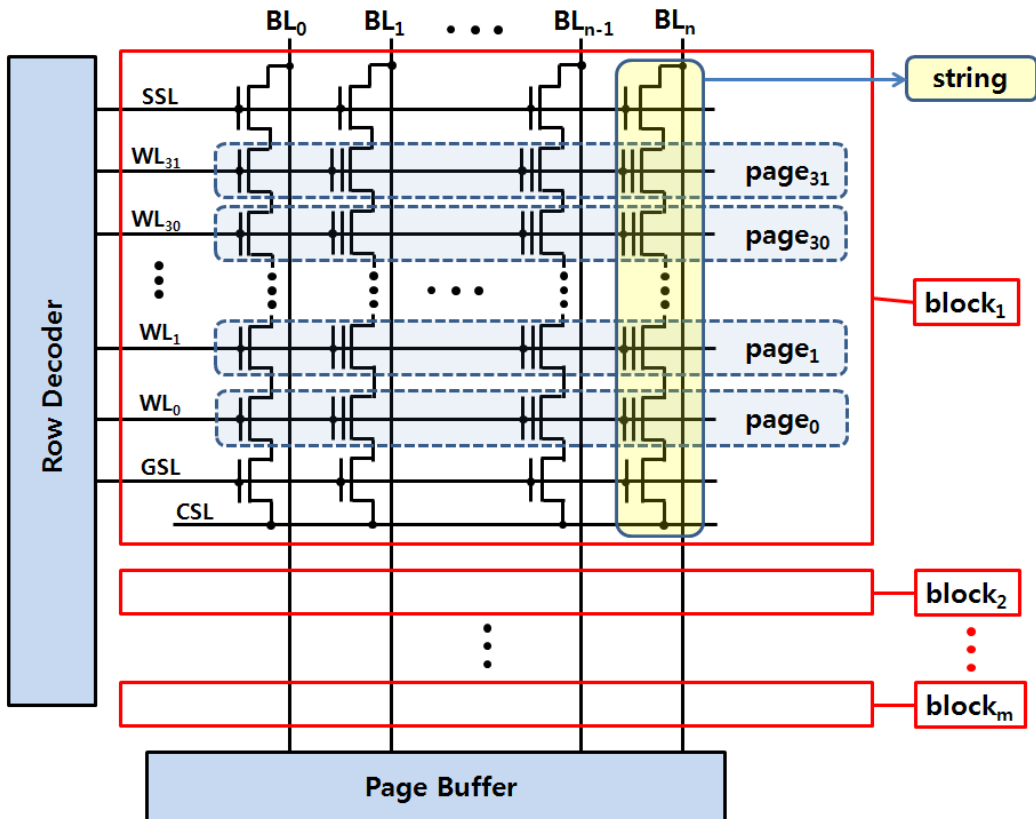


Fig. 1.8. NAND flash memory array schematic

1.3. NAND Cell Operation

The program and erase operation in the flash memory can be done by charge transport through gate dielectrics. The Fowler-Nordheim (FN) tunneling is the most widely used transport mechanism in the NAND flash memory. The FN tunneling occurs at sufficiently high gate voltage conditions. Electrons in the conduction band of the channel can tunnel through the triangular energy barrier of oxide as depicted in Fig. 1.9.

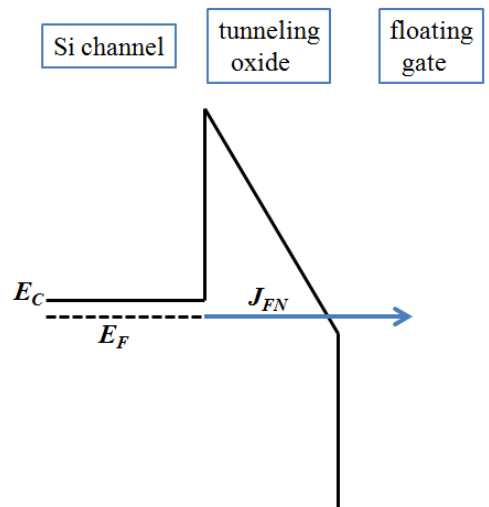


Fig. 1.9. Fowler-Nordheim tunneling for program operation mechanism.

The program operation in the NAND flash memory is performed by a page unit. The programming voltage (V_{pgm}) is applied to the word line of the selected page. For constant FN tunneling current, the gate voltage of programming pulse (V_{pp}) is increased by a constant value V_{step} . This increases the charges in the floating gate by an amount which shift the cell's V_{th} by a constant increment ΔV_{th} . Using the self limitation of the tunneling current, the cell's threshold voltage shift (ΔV_{th}) is equal to V_{step} . When a floating gate cell is intended to be programmed to a certain V_{th} state, this is typically called Incremental Step Pulse Programming (ISPP) [2]. To reach a targeted cell threshold voltage, programming pulses are applied with increasing pulse amplitude (Fig. 1. 10). Each programming step is followed by a sense operation to evaluate whether the target threshold voltage has been reached. Therefore, the program step voltage directly affects the cell V_{th} distribution width in a memory array with large numbers of cells.

When a large number of cells need to be operated in the NAND array, it has to be taken into account that one cell is located at every crossing point of bit

lines and word lines. To distinguish between selected string and unselected (inhibited) string, the self-booster program inhibit scheme [2] is used. By applying V_{cc} to the SSL and BL, the select transistor of SSL is turned off since the channel potential is higher than $V_{cc} - V_{th, SSL-TR}$. At this point, the channel of the inhibit string becomes floating node. When program voltage (V_{pgm}) and pass voltage (V_{pass}) are applied to the word lines, the channel potential is self-boostered by the series coupling capacitance through the control gate, floating gate, channel and bulk. Therefore, a highly boosted channel potential can reduce the electric field across the tunneling oxide of unselected cell, and it prevents FN tunneling into the floating gate. On the other hand, the channel potential of selected string is 0 V since the BL voltage of selected string is set to 0 V. As a consequence, the large potential difference between the floating gate and channel causes FN tunneling current for the selected cell.

The channel potential diagrams of selected string and inhibited string are shown in Fig. 1.11.

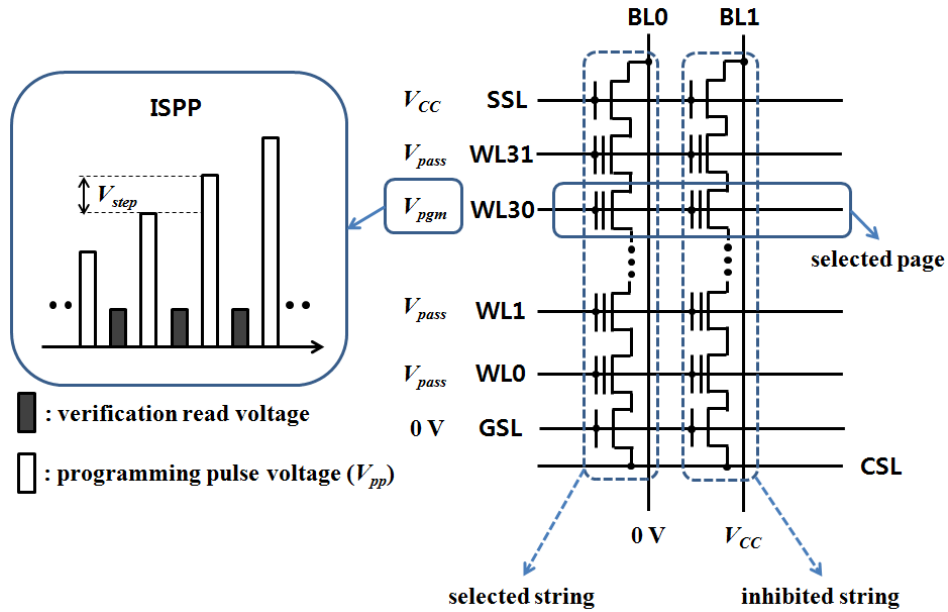


Fig. 1.10. Program operation scheme of NAND flash memory array. Using the ISPP

scheme can obtain smaller V_{th} distribution rather than the one pulse programming.

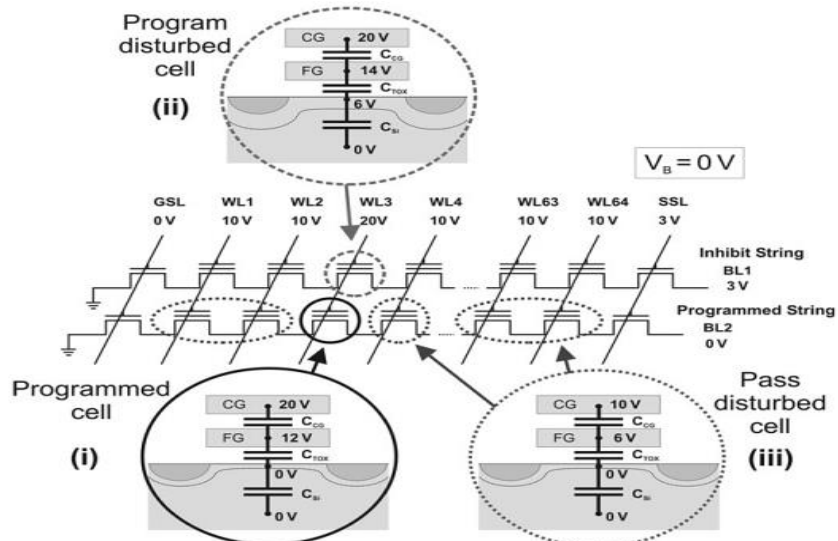


Fig. 1.11. Self-biased program inhibit scheme and channel potential of strings [3].

The erase operation resets the data of all cells in one block simultaneously [4]. The control gate voltage is biased at 0 V and the p-type substrate of selected block is biased to a high positive voltage (~20 V). This causes FN tunneling of electrons off the floating gate in the selected block and into the substrate as shown in Fig. 1.12. The erase operation could also be realized by applying negative voltages to word lines, as is the case with silicon on insulator (SOI) arrays or several 3D stacked NAND flash memories where no substrate exists. However, in this case the row decoders which are linked with word lines would have to make both high positive voltage (for programming) and high negative voltage (for erase), which means an increased peripheral circuit area consumption.

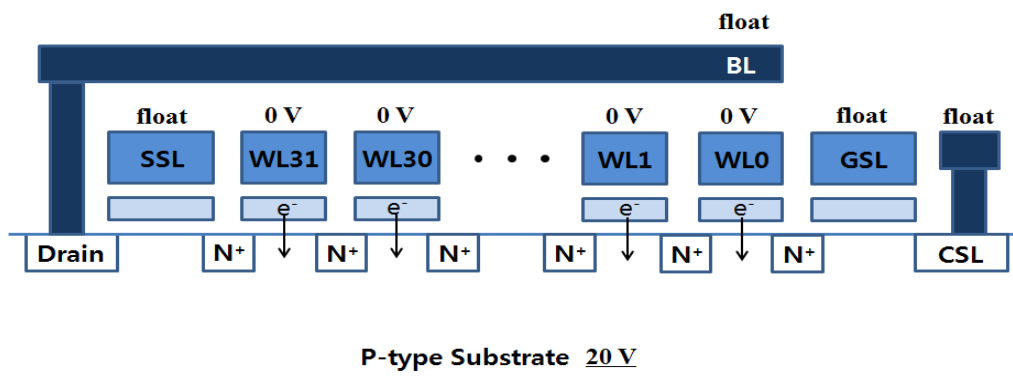


Fig. 1.12. Erase operation scheme of NAND flash memory array.

The reading method of NAND flash memory is shown in Fig. 1.13. To connect strings to the bit lines, unselected transistors have to turn on during program or read operation. In other words, the unselected cells must behave as pass transistors so that their threshold voltages do not affect the current of the string. As a result, the pass voltage (V_{pass}) should be higher than the maximum possible programmed threshold voltage ($V_{th0,max}$). And, the read voltage (V_{read}) has to be set between the erased threshold voltage (V_{th1}) and the programmed threshold voltage (V_{th0}). If the selected cell is erased, current flows from the bit line to common source line. Conversely, current does not flow in the case where the selected cell is programmed. The information (programmed or erased) of the cell is determined by a sense amplifier, which sense whether bit line current flows or not.

As mentioned previously, NAND flash memory is used for applications that require fast bulk data transfer. As a consequence, it uses page-based operation for read operation as well as program operation. For parallel operation, each bitline has its sensing circuit, usually called Page Buffer (PB) as sketched in

Fig. 1.14.

The basic principle of sensing circuit is indicated in Fig. 1.14 [5]. During the precharge step, the C_X capacitor is charged up to V_{PRE} and then it is left floating. At the end of the precharge phase, the capacitor node is connected to NAND flash string. Then, the evaluation step starts. If the cell has a threshold voltage higher than V_{read} , no current flows and the capacitor maintains its precharged potential value (V_{PRE}). As a result, the output node of the sense amplifier having latch circuit component is set to “0”. Otherwise, if the cell has a threshold voltage lower than V_{read} , the current flows and the potential of capacitor discharges. Consequently, the output node of sense amplifier is set to “1”.

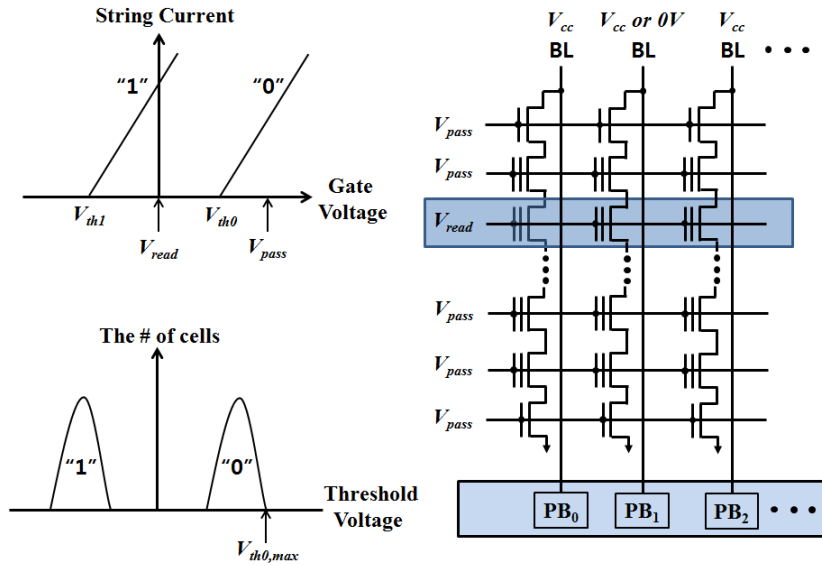


Fig. 1.13. Read operation of NAND flash memory.

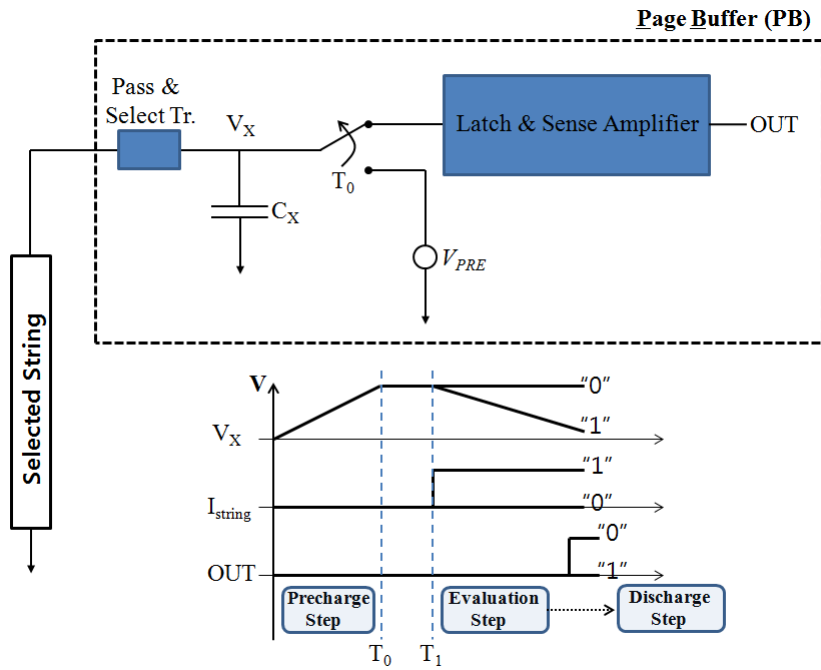


Fig. 1.14. Basic sensing scheme and the related timing diagram

Historically, the first reading technique used the parasitic capacitor of the bit line, where the cell state is detected by sensing a voltage of bitline capacitor (BL(n)) as shown in Fig. 1.15. However, there is an issue related with coupling of adjacent bit lines. Figure 1.16 shows of cut plane along the selected page. Let's suppose the BL(n+1) and BL(n-1) bit lines are discharged to $V_{PRE}-V_{SEN}$ which is voltage drop when the selected cell is in erased state. On the coupled BL(n), the following voltage variation is induced [[5], [6]]:

$$\Delta V_{BL} = -\frac{C_C + C_C}{C_{BL}} (V_{PRE} - V_{SEN})$$

This unwanted coupling noise can lead to sensing error. Moreover, the coupling capacitance between BLs increases with the continuous bitline pitch scaling-down.

To solve this issue, the interleaving architecture is introduced. While the even (or odd) bit lines are read, the odd (or even) bit lines are forced to a fixed voltage (usually ground), acting as an electrical shield. The read (or program)

operation for selected page is performed by two steps: one for even BLs' cells and the other for odd BLs' cell.

Another solution to the bitline coupling issue is the All Bit Line (ABL) architecture illustrated in Fig. 1.17 [7]. In the ABL architecture, the voltage of SO node is used for sensing instead of BL node. And, the bit line voltage is maintained during the evaluation step so that the constant BL current flows (I_{READH} for erased state or zero current for programmed state). Whether BL current flows or not is detected at SO node connected with C_{SO} and sense amplifier. Thus, ABL architecture is called by the “current sensing” as opposed to “voltage sensing” used by interleaving architecture. As all bit lines of selected page are read (or programmed) simultaneously, ABL architecture can realize maximum parallelism. The ABL architecture has the following advantages over interleaving architecture: evaluation time reduction, lower bit line voltage, simultaneous data process of even/odd cells, and reduction of floating gate interference [8]. However, the ABL architecture needs one sense amplifier for each bitline while one sense amplifier is connected per two bit

lines. Therefore, the number of reading circuit is doubled compared to the interleaving architecture.

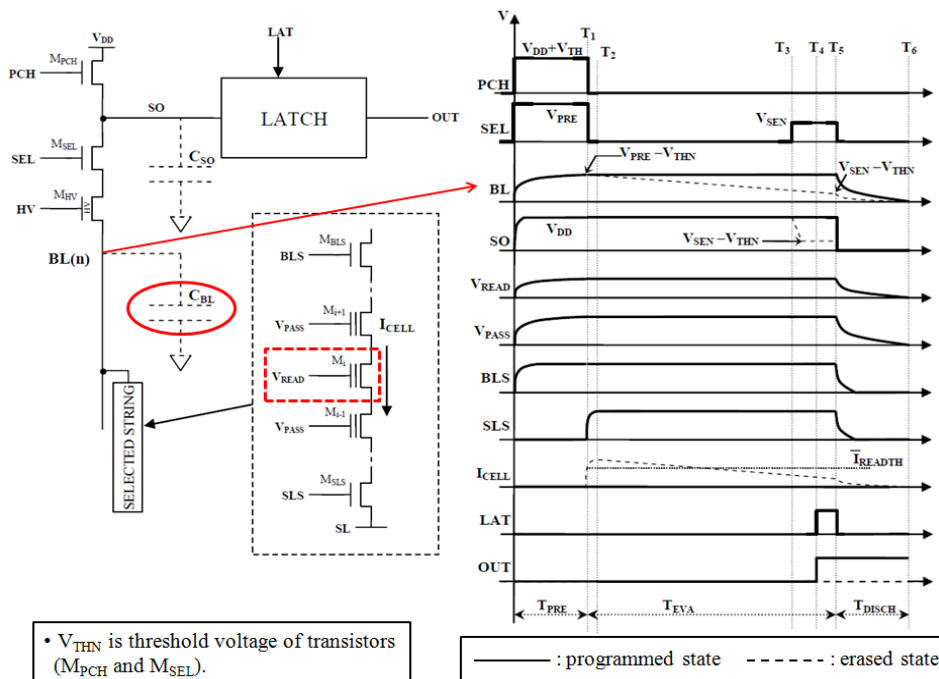
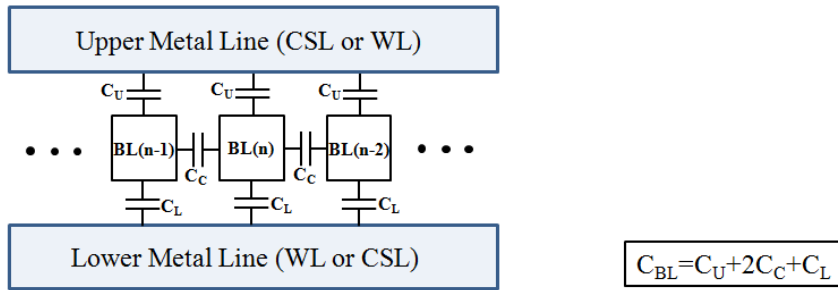


Fig. 1.15. Sensing circuit using parasitic BL capacitance (C_{BL}) and its timing diagram



- C_C is the parasitic capacitance between neighboring bit lines.
- C_U is the parasitic capacitance between the bit line and the upper metal line.
- C_L is the parasitic capacitance between the bit line and the lower metal line.

Fig. 1.16. Bit line parasitic capacitors

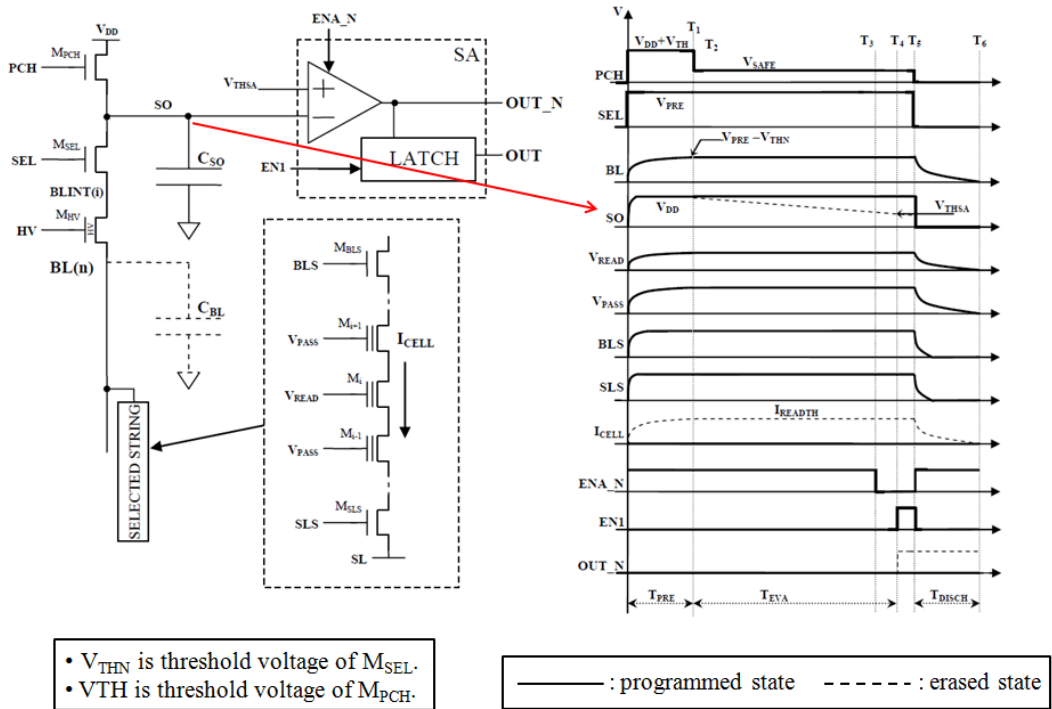


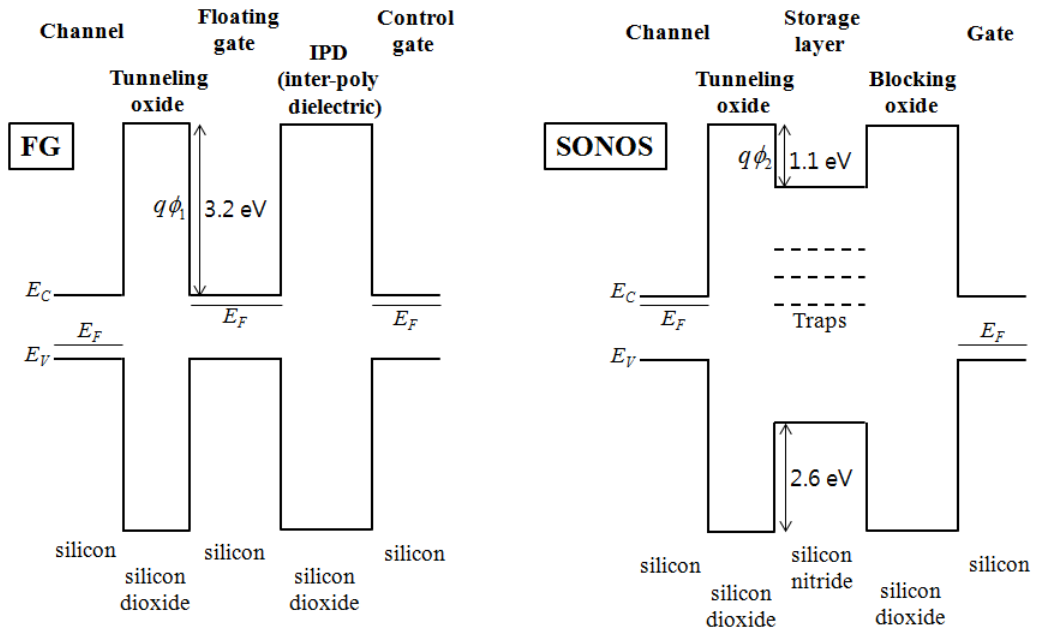
Fig. 1.17. Sensing circuit of ABL architecture and its timing diagram.

1.4 Charge Trap Flash Memory

Floating-gate flash memory has been successfully developed in the last few decades with continuous down-scaling of cell dimensions to obtain high data-storage density, high program/erase speeds, low operating voltage and low power consumption [9]. Smaller and smaller devices have driven drastic increase of fabrication density of flash memory. However, some intrinsic limitations make this type of memory rapidly approach the scaling limit. As flash memory device scales down beyond the 2x nm technology node, these approaches face significant challenges. A relatively thick tunneling oxide and inter-poly dielectric layer have to be used in the floating-gate memory to maintain acceptable reliability, limiting further down-scaling of the cell size in the vertical direction [10]. In addition, maintaining a high gate coupling ratio is still one main bottleneck in down-scaling the floating-gate devices [11]. Moreover, as the spacing between adjacent devices is down-scaled, this parasitic capacitance plays an increasingly dominant role in the device

performance since data stored in the adjacent cells can interfere with each other by capacitive coupling [12]. Additionally, a limited margin poses a great challenge on the reliability of the floating-gate memory devices, as the number of electrons stored in floating-gate significantly decreases with continual down-scaling of the cell size.

One of the most important innovation proposed to overcome the scaling limits of FG devices while maintaining its basic architecture was the introduction of an insulating charge trapping layer, typically made of nitride, to replace the poly-silicon FG [13]. This concept, which dates back to the late 1960 [14], has been exploited by several devices proposed in the last decades. Fig.1.18 depicts band diagrams of FG NAND cell and SONOS cell. Basic concept of the two memories is the same: energy inside storage materials is lower than bottom of conduction bands of tunnel and blocking dielectrics preventing diffusion of trapped charge. For both cells, write operations mean the modulation of the number of stored electrons by applying write voltage to gate.



	Silicon	Silicon dioxide	Silicon nitride
Band gap (eV)	1.12	8.9	5.1
Relative dielectric constant	11.9	3.9	7.5

Fig. 1.18. Comparison between FG and SONOS related with band diagram and material properties.

By adopting charge trap NAND architecture, some issues that can limit the future scalability of FG NAND technology can be solved. Most important improvements are huge electrostatic interference reduction, higher immunity to active oxide defects, and easier process integration.

Chapter 2

3-D Stacked NAND Flash Memory

2.1 Examination of Previous 3-D Stacked NAND Flash

2.1.1 Gate Stack Type 3-D NAND Flash Memory

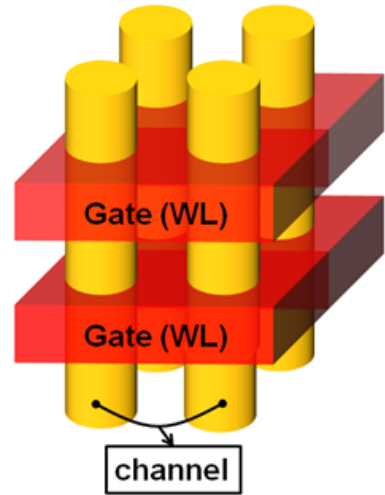
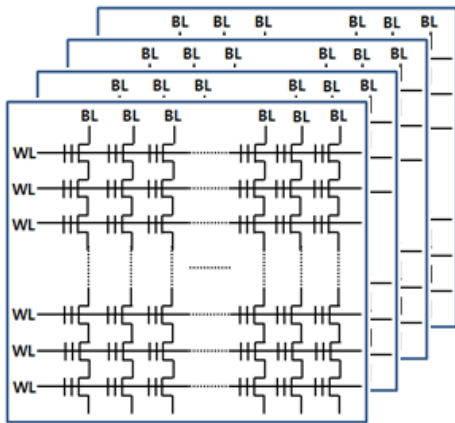
The 3-D stacked NAND flash memories can be divided into two categories as depicted in Fig. 2.1. (the gate stack type and the channel stack type) In this section, previous gate stack type 3-D NAND flash memories are examined.

In 2007, Toshiba Corp. devised the bit-cost scalable (BiCS) 3D cell technology that held costs down even as the number of layers increased. Since Toshiba reported the BiCS, several gate stack type 3D NAND flash memory architectures have been proposed, and TCAT (Terabit Cell Array Transistor) by Samsung is one of the most representative architectures.

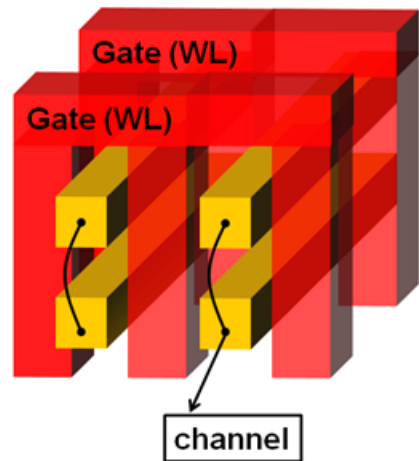
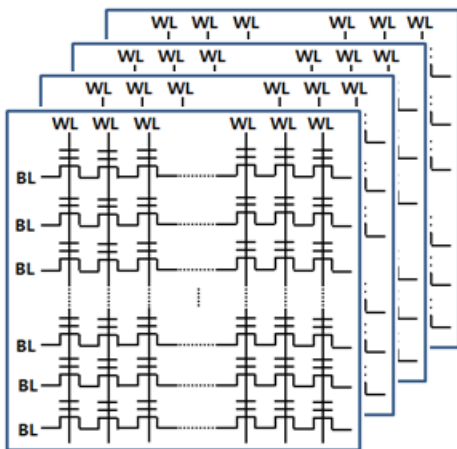
The BiCS structure was shown in Fig. 2.2 [15]. In this structure, "punch and plug" process method is used. String of the cells is on the plugs located vertically in the holes punched through whole stack of the gate plates. Since the channel should be deposited inside the holes, poly-crystalline channel is utilized. Consequently, a unit cell of BiCS structure is based on poly-crystalline Si thin-film transistor (TFT) device having vertical channel and GAA structure.

The equivalent circuit of BiCS structure is indicated in Fig. 2.3. In BiCS structure, a single cell is accessed by control gate on the string which is selected by a bit line and row select line.

Except for erase operation, the operation method of BiCS is similar to the conventional planar NAND flash. The body of vertical channel is not directly connected to p-type substrate. Instead of the conventional bulk erase method, the body potential is raised by holes generated by gate-induced drain leakage (GIDL) at the edge of lower select gate.



(a)



(b)

Fig. 2.1. Classification of 3-D NAND flash memories according to the stacking type.

(a) Gate stack type (b) Channel stack type

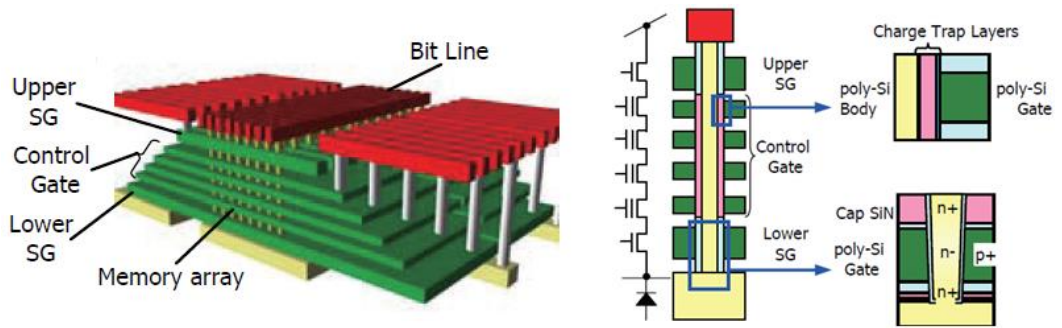


Fig. 2.2. BiCS architecture and its string structure [16].

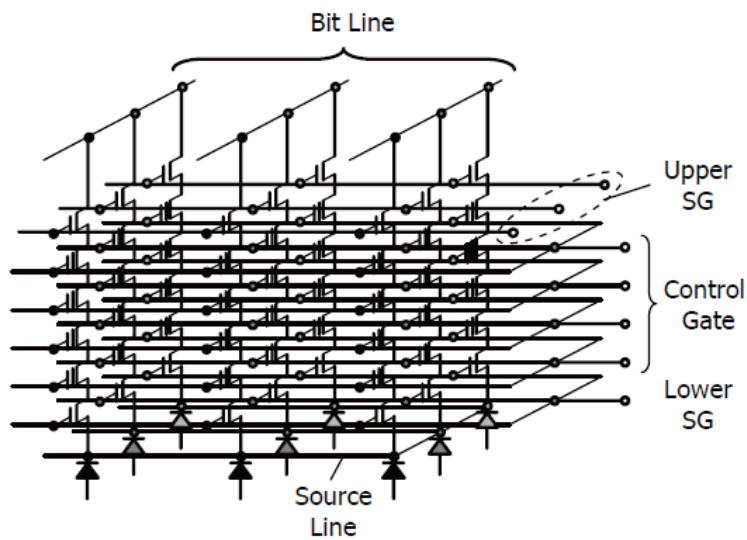


Fig. 2.3. Equivalent circuit of BiCS flash memory [15].

After the first introduction of BiCS flash memory, Toshiba has announced various modified versions. To solve poor sub-threshold characteristics of polysilicon TFT due to trap density at grain boundary, “macaroni” body vertical

FET having thinner body poly-silicon than the depletion width was adopted [17]

BiCS flash memory having macaroni body FET has better sub-threshold characteristics with increased current and uniform V_{th} distribution. However, there are still several issues to be resolved. As the bottom source region is formed on the substrate, the removal (HF treatment) of bottom gate dielectrics at the bottom of hole before plugging channel is inevitable. Consequently, SiN-based tunnel film is needed to minimize the damage by diluted-HF treatment for the “gate-first” process [18]. Pipe-shaped BiCS (P-BiCS) was devised [17] to solve some critical problems on BiCS flash such as poor reliability of memory cells, cut-off characteristics of the lower select gate and high resistance of source line. By the change of the vertical NAND string from straight shape to U-shape as shown in Fig. 2.4., P-BiCS realizes good data retention and wide threshold voltage window because of less process damage on tunnel oxide during the fabrication process. Low resistance of source line by introducing metal wiring and tightly controlled cut-off characteristics of

select gate enable good functionality of memory array.

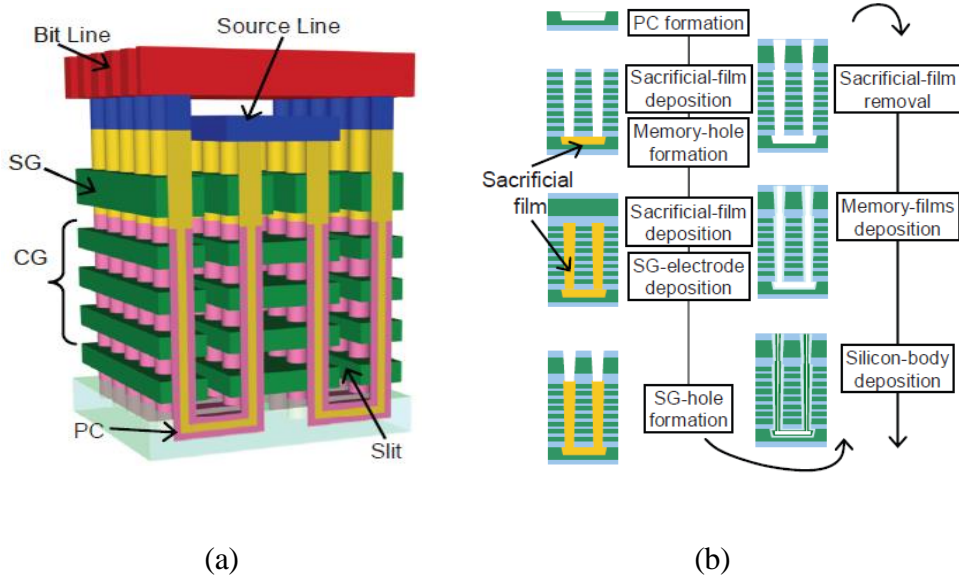


Fig. 2.4. (a) Schematic of P-BiCS flash memory (b) Fabrication of process of P-BiCS [15].

Samsung proposed vertical NAND flash memory cell array using TCAT (Terabit Cell Array Transistor) technology to improve two major drawbacks of the BiCS technology [19]. BiCS flash memory is almost impossible to implement metal gate structure because it is very difficult to etch metal/oxide

multilayer simultaneously. Consequently, BiCS flash memory cannot utilize various advantages of metal gate SONOS cell structure, for example, faster erase speed, wider V_{th} margin, and better retention characteristics. Another concern is GIDL erase of BiCS flash. An extensive periphery circuit change may be necessary to apply negative bias on word line during erase operation. Figure 2.5 shows the structure of TCAT flash memory. Metal gate structure and bulk erase operation are key features of TCAT flash. By embedding the concept of the process flow named with 'gate replacement', metal gate could be utilized as shown in Fig. 2.6. Also, the channel of TCAT is directly connected to p-type substrate, so that conventional bulk erase method can be used. Therefore, the NAND string can be operated without any major peripheral circuit change.

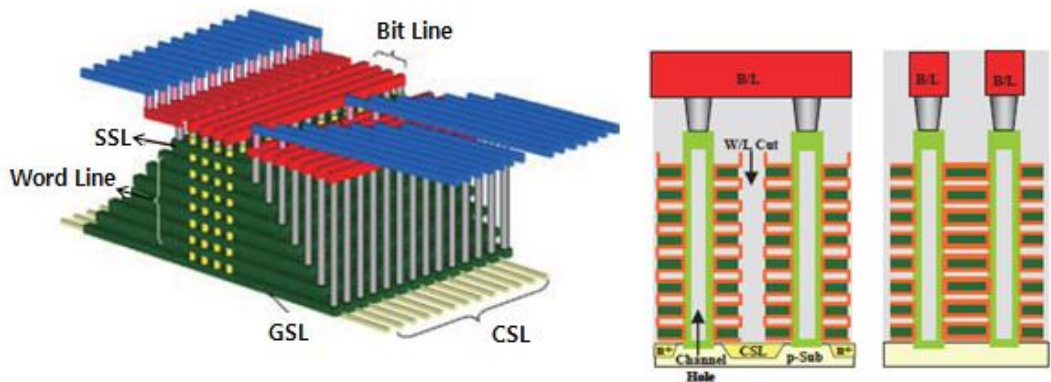


Fig. 2.5. Bird's eye view and cross sectional view of TCAT flash memory

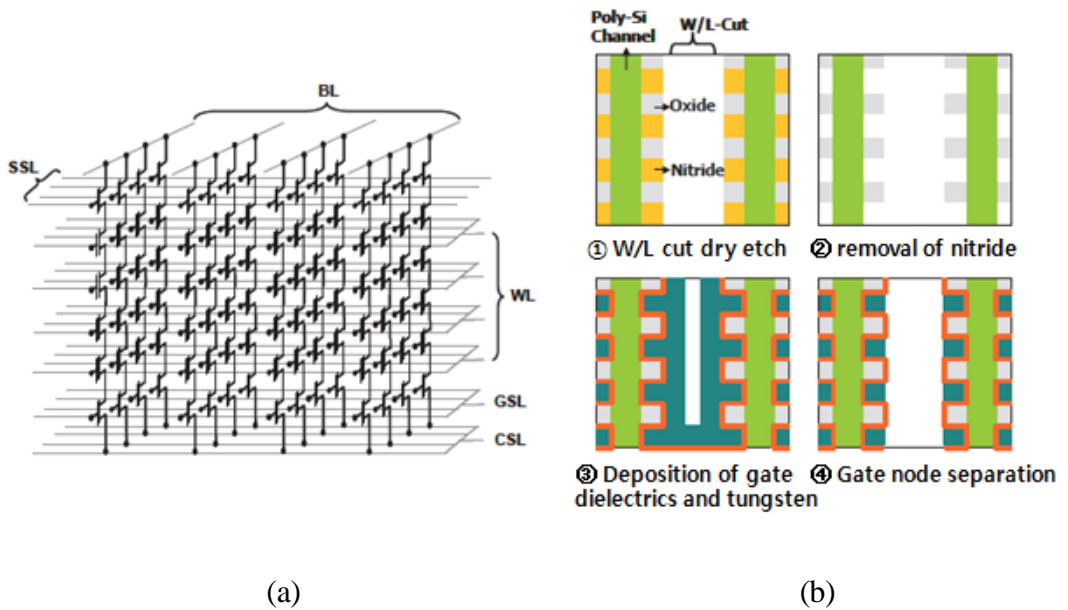


Fig. 2.6. (a) Equivalent circuit of TCAT (b) 'gate replacement' process method

2.2.2 Channel Stack Type 3-D NAND Flash Memory

A number of channel stack type 3-D NAND flash memories have been reported. They have horizontal stacked channels and the gate is vertically shared. Firstly, Samsung devised channel stack type 3-D NAND flash memory, which was called VG-NAND as depicted in Fig. 2.7 [20]. The VG-NAND is comprised of horizontal stacked string, vertical gate for WL and SSL, and CSL tied with source and active body for bulk erase operation. For layer decoding, VG-NAND has multiple SSLs which are connected in series. SSL transistors are either depletion mode or enhancement mode, and proper bias is applied to transistors to distinguish stacked strings.

Although VG-NAND can solve the problems of gate stack 3-D NAND flash memory's problems such as difficulty of WL interconnect, program disturbance, and channel resistance, there is a burden of making several SSL transistors which require additional die area.

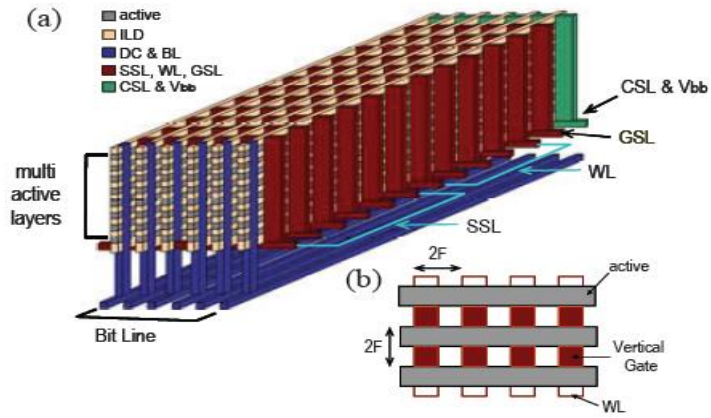


Fig. 2.7. (a) Structure of VG-NAND (b) Top view of active array [19]

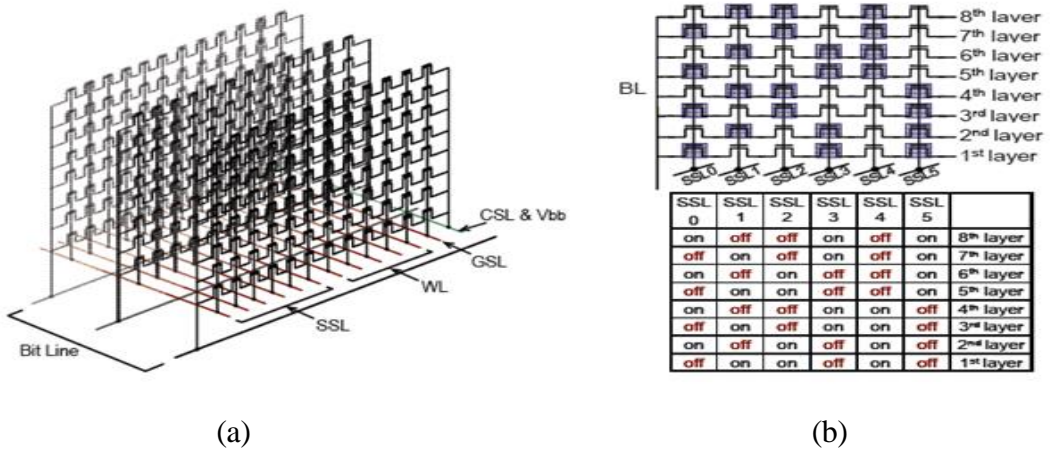


Fig. 2.8. (a) Equivalent circuit of VG-NAND (b) Schematic of SSLs and its operation table for a layer selection

The second example of channel stack type 3-D NAND flash memory is 3D VG NAND of Macronix as shown in Fig. 2.9 [21]. It has stacked poly-Si channels and the vertical gate (VG) NAND array with a much smaller half pitch (75 nm) than VG-NAND of Samsung. However, memory array design for VG NAND is much more difficult. Since the BL is horizontal and parallel to each other, it is difficult to decode the multi-layer BL's in any simple way. For a simple decoding method, Macronix introduced a PN diode decoding structure to overcome drawbacks of VG NAND. The new version of 3D VG NAND is shown in Fig. 2.10 [22]

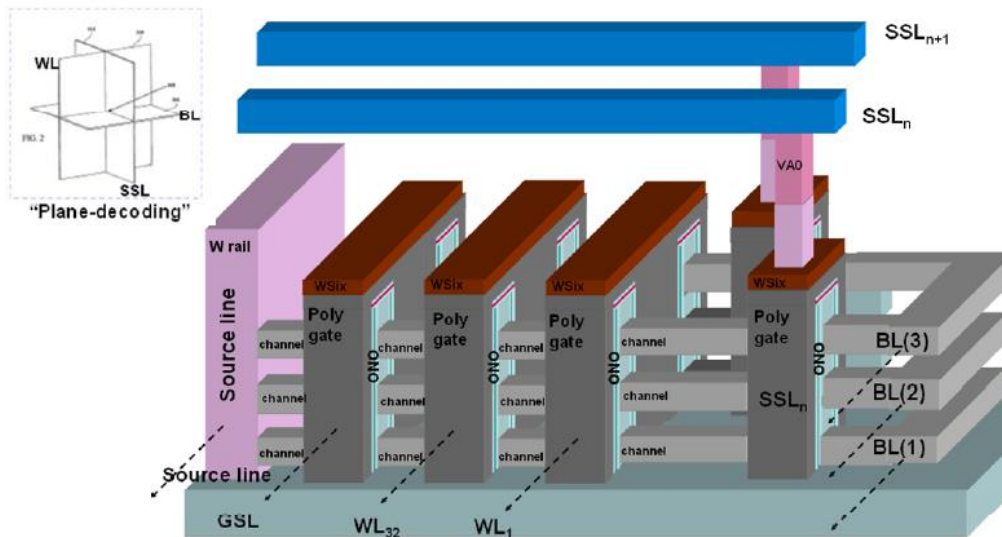


Fig. 2.9. Structure of 3D VG NAND [20]

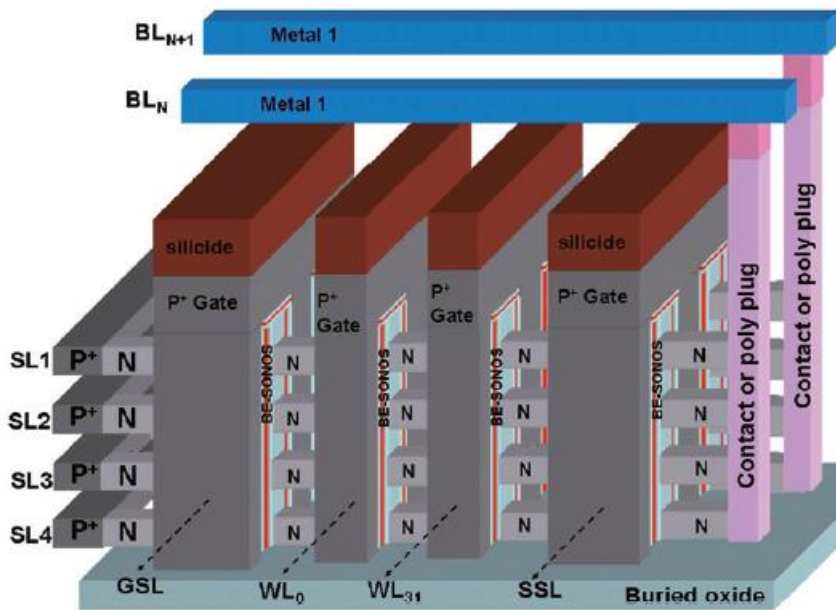


Fig. 2.10. Structure of the pn diode decoded 3D VG NAND architecture. [21]

The PN diode decoded 3DVG does not use plural SSL in each block, but instead separates the source lines (SL) of different memory layers. The memory cell is then selected by the intercept of WL, BL and SL. However, this type of decoding has sneak paths between memory layers and this prohibits correct program and read selection. A PN diode is thus introduced at the source side to block the sneak paths during reading. For the program-inhibit operation, a three-step pulse is introduced during self-boosting programming. The

operation method of pn diode decoded VG NAND is somewhat different from the conventional NAND operation. During read operation, the source-side sensing [23] is conducted where a positive voltage is applied at the p-type selected source line (SLs) because the pn diode is connected at the source side.

Contrary to the previous 3D VG NAND approaches, there is no need to fabricate plural string select (SSL) transistors inside the array of pn diode decoded VG 3D NAND, thus enabling a highly symmetrical and scalable cell structure. However, a good PN diode with very low leakage is important. More metal routing is also needed to minimize the source resistance. Due to the issues of polycrystalline Si channel (lower mobility, non-uniform characteristics, and poor reliability), single crystalline Si nanowire stacking was investigated [24]. SiGe/Si layers are epitaxially grown on a single crystal Si wafer for stacked-nanowire formation. However, 3-D stacked array was not fabricated, but just the concept was proposed in Ref. [24] as shown in Fig. 2.11. Moreover, this proposed array structure is not suitable for realistic NAND flash memory because the number of bit-lines of the 3D array is limited by the

number of stacked layers.

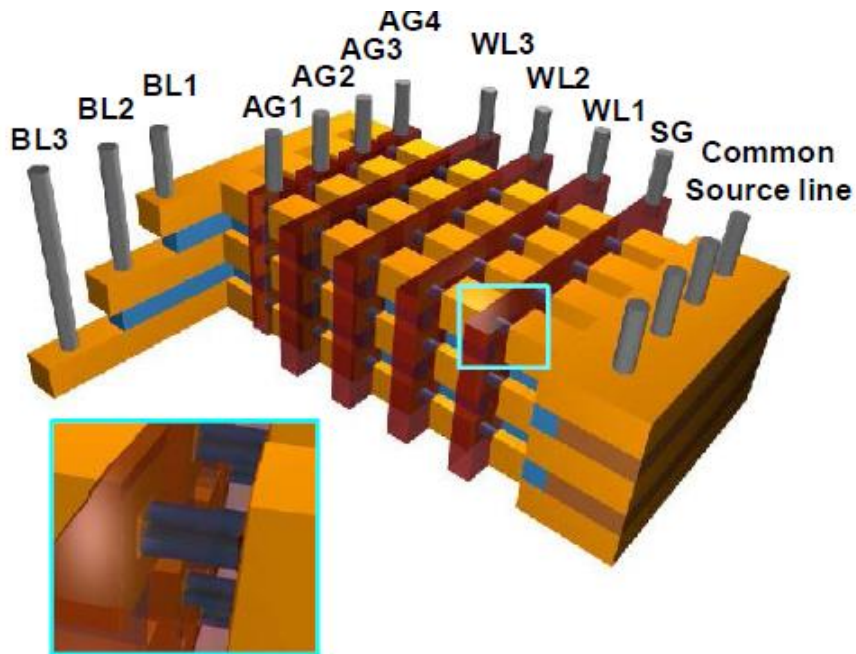


Fig. 2.11. 3D array of single crystal Si nanowires introduced by CEA-LETI [24].

In our previous research, “STAR” (STacked ARray) having single-crystal Si channels and gate-all-around (GAA) structure was investigated for channel stack type 3-D NAND flash memory as shown in Fig. 2.12 [25]. The cells in STAR having a single-crystal Si channel have uniform distribution of cell performance because of the absence of the defects associated with grain boundaries. The subthreshold characteristics of poly-Si FETs are dependent on

trap density at grain boundaries in poly-Si. Thus, STAR can have better performance with a uniform and stable characteristic. However, channel stacked array NAND flash memory requires the effective layer selection schemes and architecture, so the optimized architecture needs to be devised. 3D NAND flash memory architecture should satisfy several conditions; the easiness of fabrication, the compatibility with the conventional peripheral circuits, and the acceptable area for contact formation.

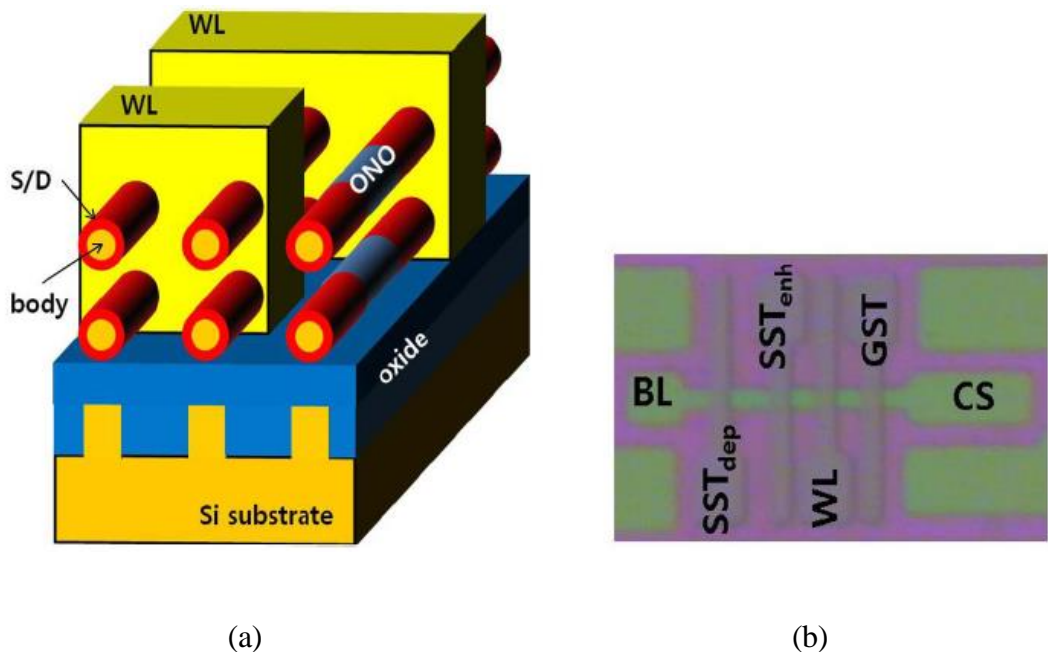


Fig. 2.12. (a) Bird's eye view of the STAR NAND flash memory. (b) The fabricated array having two SSL (SST_{dep} and SST_{enh}) for layer selection.

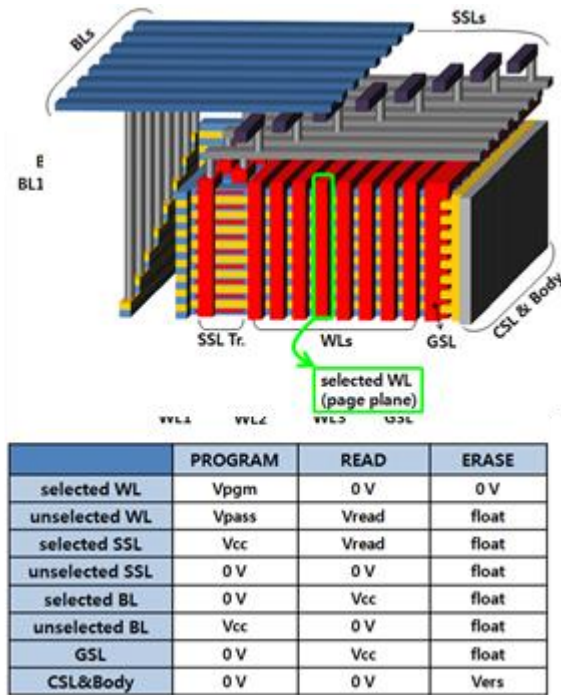


Fig. 2.13. Bird's eye view of the conventional CSTAR, its equivalent circuit, and operation scheme [26].

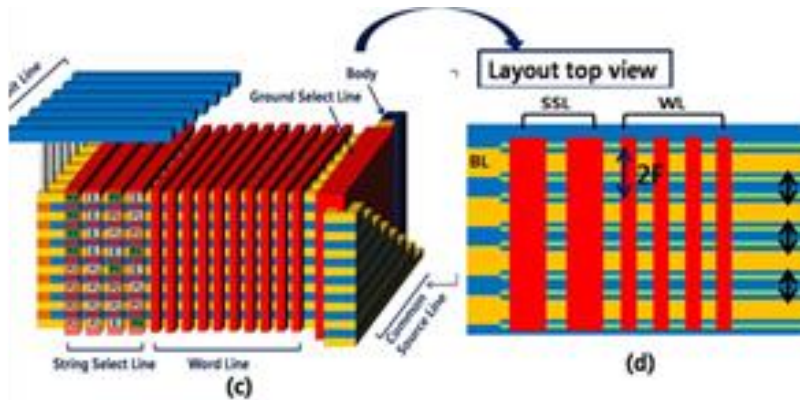


Fig. 2.14. Bird's eye view of CSTAR with LSM and its layout top view [27].

Fig. 2.13 shows the conventional CSTAR architectures [26]. In the case of the conventional CSTAR array, each bit-line is connected to the different layer. Since the bit line is common in a layer, string selection in the selected layer is done by using *the island type* string selection line (SSL). The strength of the conventional CSTAR is that it can use 2D planar NAND operation scheme without degradation of performance throughput. However, the formation of the island type SSLs is challenging. To overcome this issue, CSTAR with layer selection by multi-level operation (LSM) was designed (Fig. 3, [27]). In the case of layer selection by multi-level operation, the SSLs in a layer have different levels of threshold voltage (V_T). Applying different bias to each SSL can distinguish the different strings in the same layer. Since there is no need of the island type SSLs, the bit line pitch can be scaled down, which increases cell density. However, the effective method to initialize V_T of SSLs should be optimized

2.2.3 Comparison between the Gate Stack Type and the Channel Stack Type

Figure 2.15 summarizes the comparison between the gate stack type and the channel stack type. The minimum unit cell size for the gate stack type (P-BiCS and TCAT) should be around $6F^2$ because of the “word line (WL)-cut” process [17], [19]. Each drilled hole must be located inside one SSL line for decoding. The drilled hole size is limited by the minimal ONO thickness ($\sim 15\text{nm}$) and the channel diameter ($\sim 20\text{nm}$). The minimal channel diameter is constrained by the read current and the tolerable field enhancement (FE) effect. There are also other lithography constraints such as the overlay between the drill hole and SSL, and the minimal SSL/WL cut space. The estimated minimal cell size is possibly around $4X\text{nm}$ $1/2$ pitch with $6F^2$ [28].

The channel stack-type NAND flash architecture has a large memory density with smaller stacked layers than gate stack. As the number of stacks increases, more notable differences between the gate stack type and the channel stack

type can be found. Therefore, we can conclude that the channel stack type can attain terabit density more quickly. A certain memory density may be achieved by any array architecture but with different numbers of stacking layers. A smaller pitch allows the achieving of high density at reasonable number of stacked memory layers (< 32) and thus potentially offers lower cost.

Figure 2.14 shows the trend of bit line current of 3D NAND as a function of the number of stacked layers and the hole diameter for gate stacked type and channel stacked type. For this simulation, we use poly-Si channel with macaroni structure for gate stacked NAND, and single-crystalline silicon nanowire channel for channel stacked NAND. GSTAR shows bit line current degradation as the number of stacked layers increases, and this degradation should be compensated by increasing the hole diameter. Therefore, CSTAR has an advantage in scaling-down the minimum feature size. The key in scaling-down the feature size of channel stacked NAND is how to reduce the thickness of trapping layer and blocking layer, and adopting high- κ dielectric layer can solve this issue.

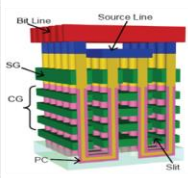
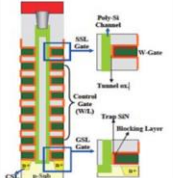
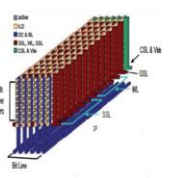
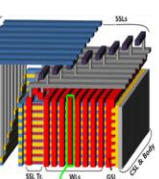
	GATE STACK TYPE		CHANNEL STACK TYPE	
	P-BICS	TCAT	VG NAND	CSTAR
STRUCTURE				
CELL STRUCTURE	Macaroni & GAA	Macaroni & GAA	Double gate	Gate-All-Around
CHANNEL	Poly Si channel	Poly Si channel	Poly Si channel	Single crystalline Si channel
KEY ISSUE	Low read current Reliability pitch scaling	Low read current Reliability pitch scaling	Number of lithography, implantation	Layer decoding Wiring complexity

Fig. 2.15. Summary of gate stack type and channel stack type.

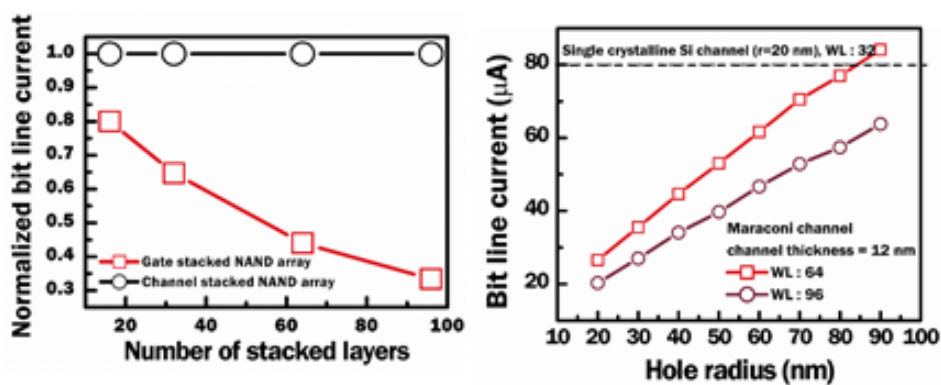


Fig. 2.16. Trend of bit line current of 3D NAND as a function of the number stacked layers and the hole diameter.

Chapter 3

Channel Stacked NAND Flash Memory with high- κ

Charge Trapping Layer

3.1 Introduction

For all 3D NAND structures, the lateral scalability is limited by the thickness of memory dielectric stack (blocking, trapping, and tunneling layer) and that of channel [29]. If the channel thickness is scaled down to <20 nm, the total dielectric layer thickness becomes a roadblock in scaling-down of the minimum feature size. Although the conventional Si_3N_4 trap layer is the most widely used charge trap layer, it is challenging to reduce the thickness of oxide-nitride-oxide (ONO) layer, since the charge trapping properties degrade when the Si_3N_4 is made thinner. On the other hand, high- κ based trapping layer

has advantages for scaling-down of the thickness due to its higher dielectric constant and higher trap density than the conventional Si_3N_4 charge trapping layer. The charge trapping and tunneling characteristics of HfO_2 layer that replaces Si_3N_4 layer as trap layer were investigated by You *et al.*[30]. The electrical properties of metal-hafnium-oxide-semiconductor (MHOS) capacitors were compared with those of metal-nitride-oxide-semiconductor (MNOS) capacitors. MHOS capacitors showed a larger memory window than the MNOS capacitors at the same trap layer thickness, because the HfO_2 layer has better charge trapping efficiency than the Si_3N_4 layer. Whereas the Si_3N_4 trap layer whose thickness is below 4 nm shows enormous increase of the tunneling current, the HfO_2 trap layer does not show the degradation of trapping efficiency when its thickness is reduced below 4 nm. It has been found that an HfO_2 trapping layer has greater conduction band offset (~ 1.65 eV) than a Si_3N_4 trapping layer (~ 1.03 eV). Moreover, the trap density and trap energy is tunable by engineering the deposition methods.

The leading candidate as an alternative blocking oxide layer is Al_2O_3 . The

dielectric constant of Al_2O_3 (~9) is higher than that of SiO_2 (~3.9). The main advantage of Al_2O_3 for blocking oxide is decreased back tunneling current during erase operation. When negative bias is applied for the erase operation, a blocking dielectric with a higher dielectric constant can reduce the electric field, and electron injection from the gate during erase operation can be effectively suppressed [31]. Table 3.1 summarizes the material properties of interesting high- κ trapping and blocking layer materials as well as of Si_3N_4 [32]–[34].

With the adoption of high- κ trapping layer and blocking layer, wider memory window with smaller physical thickness can be obtained. While many studies have reported the electrical properties of memory devices with high- κ based charge trapping layer, most studies are limited to large capacitor devices or single flash memory devices. In this work, we demonstrate high performance high- κ based charge trap NAND flash array that is applicable to 3D stacked memory.

Table 3.1 Material properties of high-κ materials

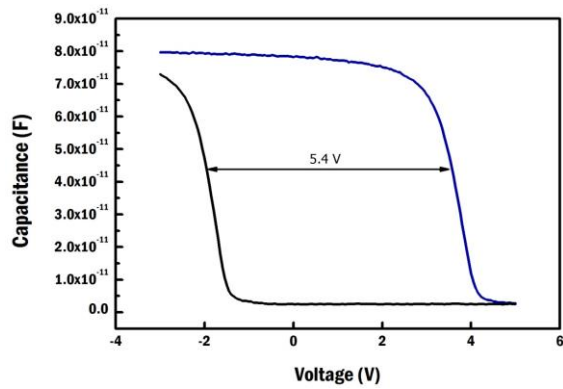
	Si ₃ N ₄	HfO ₂	Al ₂ O ₃
Conduction band offset respect to Si (eV)	2.12-2.4	1.5	2.8
Dielectric constant	7.5	24	7
Trap depth (eV)	0.8-1.1	~0.5-2.45	~2.35
Trap capture cross section (cm ²)	~10 ⁻¹⁶	~2x10 ⁻¹⁸	-

3.2 Memory Characteristics of HfO₂

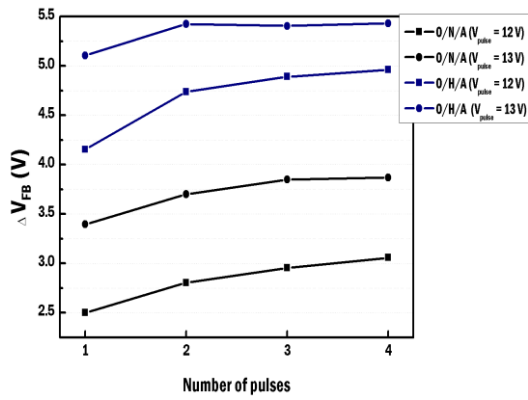
To investigate the electrical characteristics of HfO₂ as a charge trapping layer, capacitors composed of SiO₂/HfO₂/Al₂O₃ (3.5/8/15 nm) and SiO₂/Si₃N₄/Al₂O₃ (3.5/8/15 nm) were fabricated. The 3.5 nm SiO₂ film was thermally grown at 800 °C, and HfO₂ and Al₂O₃ layers were deposited using atomic layer deposition (ALD). Hf[N(CH₃)(C₂H₅)]₄ was used as a precursor for HfO₂, and Al(CH₃)₃ was used as a precursor for Al₂O₃. Process temperature was 300 °C. Si₃N₄ was deposited using low pressure chemical vapor deposition (LPCVD) at 785 °C. The capacitance-voltage (C-V) curves were measured with Agilent B1500A semiconductor characterization system. As shown in Fig. 1 (a), when the program voltage (13 V) is applied to the gate electrode, the sample with HfO₂ charge trapping layer shows 5.4 V of flat band shift. Fig. 1(b) indicates the amount of the shifted flat band voltage when the program pulse is applied repeatedly. The sample with HfO₂ shows wider flat band voltage shift compared with the sample with Si₃N₄ charge trapping layer. It can be

concluded that HfO₂ trapping layer shows superior program characteristics to the Si₃N₄ trapping layer with the same physical thickness.

By varying the thickness of HfO₂ charge trap layer and Al₂O₃ blocking layer, its memory characteristics are investigated to ensure their limitation of thickness scaling. Fig. 3. 2 indicates that the thickness of HfO₂ is reduced to 2 nm, there is a degradation of trapping efficiency. In the case of Al₂O₃ blocking layer, its thickness should be larger than 9 nm to achieve stable program characteristics. When its thickness is reduced below 9 nm, leakage components are increased. Based on these experiments, we can conclude that HfO₂ trapping layer should be thicker than 4 nm, and Al₂O₃ blocking layer should be thicker than 9 nm.

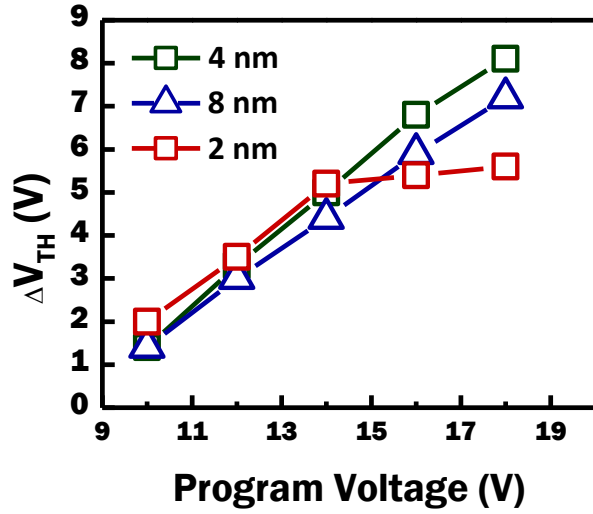


(a)

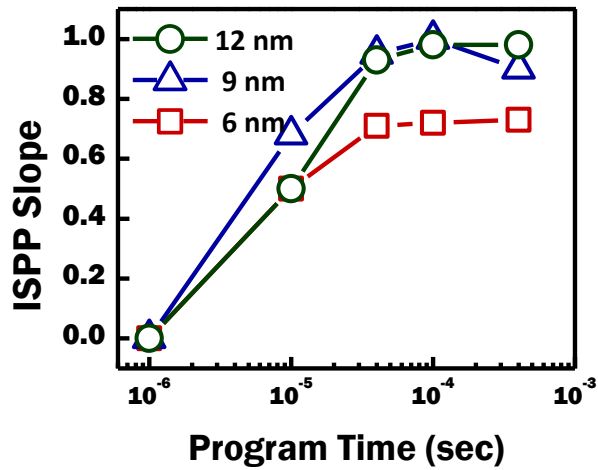


(b)

Fig. 3.1. (a) C - V curve of the sample with HfO_2 charge trapping layer, $V_{pgm} = 13$ V (b) The amount of the shifted flatband voltage as the program pulse is applied repeatedly.



(a)



(b)

Fig. 3.2. (a) Program speed with respect to the thickness of HfO₂ trapping layer (b) ISPP slope with respect to the thickness of Al₂O₃ blocking oxide (experiments were conducted by Yonsei University)

3.3 Fabrication of Nanowire Memory Devices with high- κ

Dielectric Layer

To demonstrate the scalability of CTF memory devices with high- κ material, we fabricated single crystalline silicon nanowire flash memory devices with HfO₂ and Al₂O₃. For the nanowire devices, the total thickness of dielectric layer can be a bottleneck for scaling-down of the cell size. So, we implement 4.5-nm-thick HfO₂ layer and 12-nm-thick Al₂O₃ layer to nanowire memory devices. The fabrication started on silicon-on-insulator (SOI) wafers. Ion implantation with BF₃ at the dose of $1 \times 10^{12} \text{ cm}^{-2}$ was performed to adjust the threshold voltage. Then, the active region was defined with mix-and-match process of e-beam and photolithography followed by reactive ion etch (RIE). For the corner-rounding of Si nanowires, wet etch with NH₄OH/H₂O₂/de-ionized H₂O mixture (APM) at 65°C was performed, and buffered HF solution was used to remove the buried oxide (BOX) beneath Si nanowires. A

3.5-nm-thick oxide layer was thermally grown for a tunneling layer, and a 4.5-nm-thick HfO₂ layer and a 12-nm-thick Al₂O₃ layer were deposited with atomic layer deposition (ALD) process. After that, phosphorus-doped polysilicon was deposited and word-line patterning was performed with mix-and-match process followed by reactive ion etch (RIE). Subsequently, an ion implantation with arsenic was performed to form n⁺ region, and contact hole was formed followed by metal (Ti/TiN/Al/TiN) patterning as the back-end-of-line (BEOL) process.

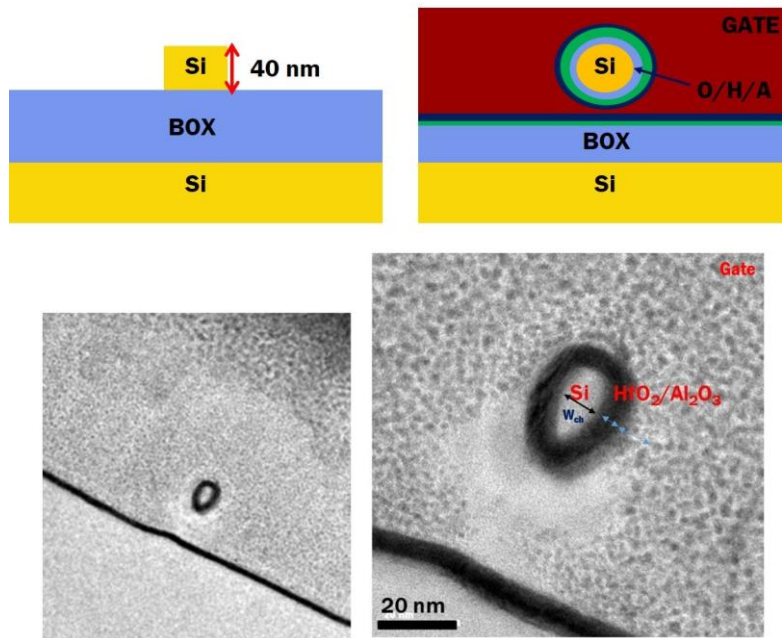
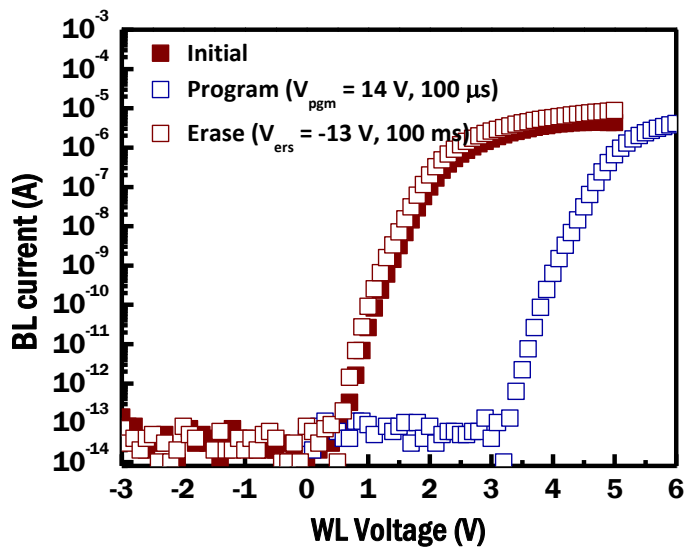
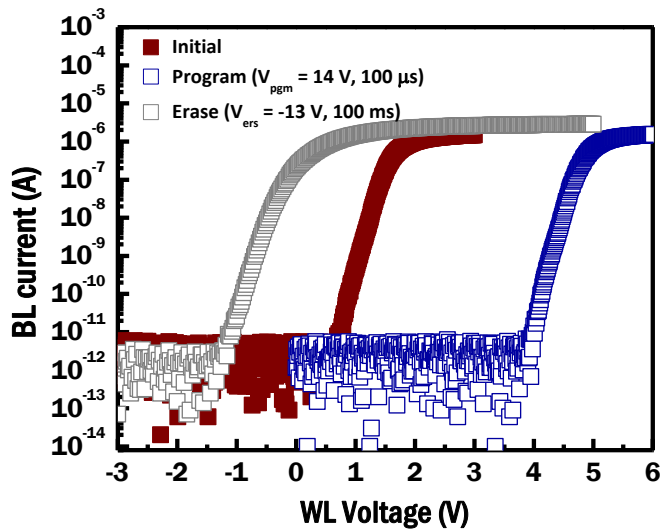


Fig. 3.3. A schematic view and transmission electron microscopic (TEM) images of the cross-section of single crystalline silicon nanowire.



(a)



(b)

Fig. 3.4 (a) Program and erase characteristics of planar SAHOS device ($W_{ch} = 1 \mu\text{m}$, $L_G = 1 \mu\text{m}$) (b) Program and erase characteristics of gate-all-around SAHOS device ($W_{ch} = 15 \text{ nm}$, $L_G = 200 \text{ nm}$)

Fabricated NAND flash array consists of a string select transistor, three word-lines, and a ground select transistor. Figure 3.3 is the cross-section view of the single crystalline Si nanowire channel surrounded by SiO₂, HfO₂, and Al₂O₃. The transmission electron microscopic (TEM) image shows uniformly deposited dielectric layer with atomic layer deposition (ALD) process. The channel width (W_{ch}) is defined by e-beam lithography, and W_{ch} of the fabricated device is about 15 nm. For comparison, planar structure devices with 1 μm of W_{ch} were also fabricated. Figure 3.4 (a) shows program and erase characteristics of the planar device with 1 μm of W_{ch} , 1 μm of cell gate length (L_g) and 1 μm of word line gap. Program/erase voltage is applied to the selected cell gate and pass voltage (6 V) is applied to unselected cell gates. With the program pulse of 14 V for 100 μs and the erase pulse of -13 V for 100 ms, about 2.5 V of memory window can be obtained. With the help of the gate-all-around structure, memory characteristics are improved and memory window is increased to 3.5 V as shown in Fig. 3.4 (b). Measured devices have a nanowire channel with 15 nm of W_{ch} , 200 nm L_g and 100 nm of word line

gap. Better gate controllability induces lower subthreshold swing as well. In the cylindrical geometry, the electric field at the Si-SiO₂ is enhanced while the electric field applied to the blocking oxide is depressed [9]. Therefore, program and erase efficiency can be enhanced.

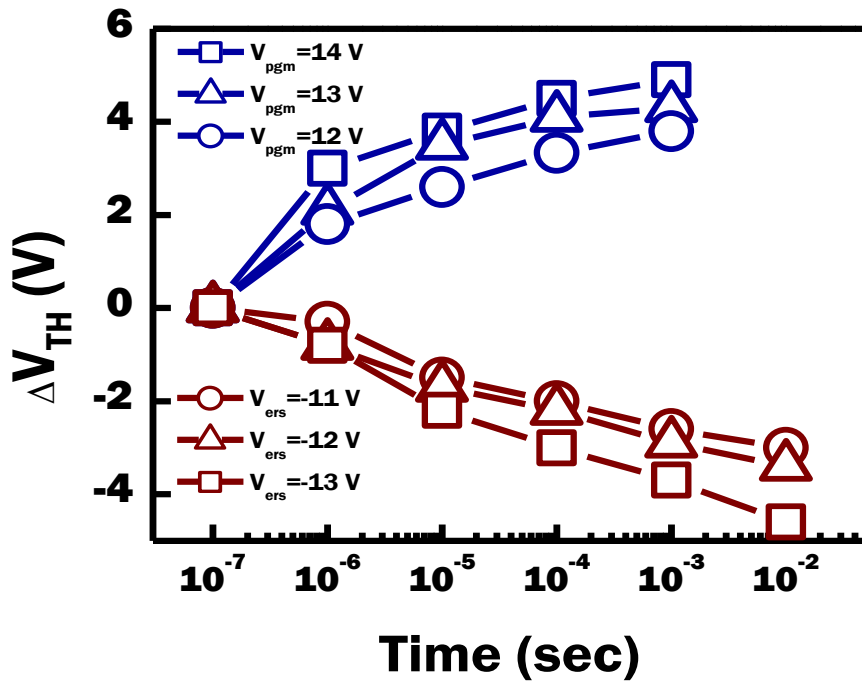
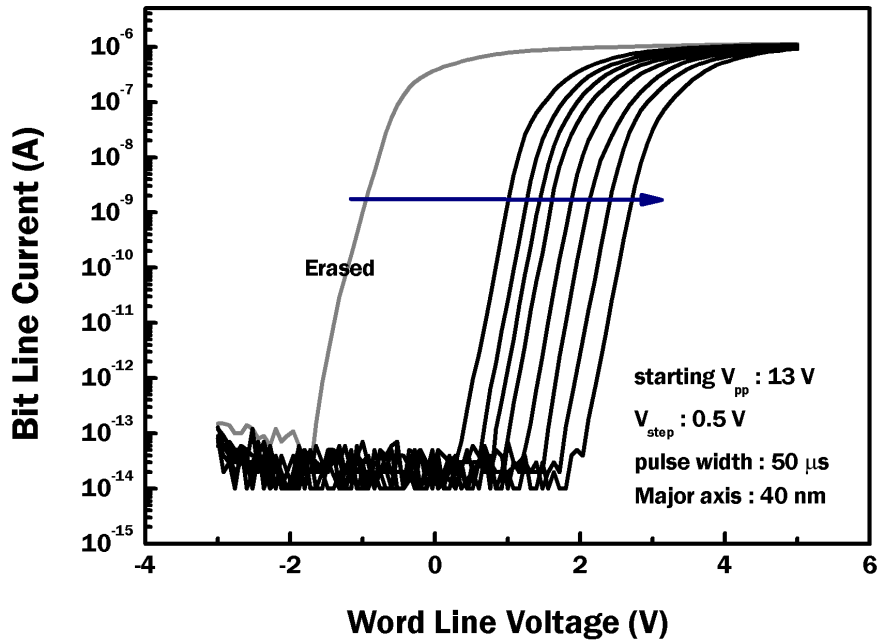
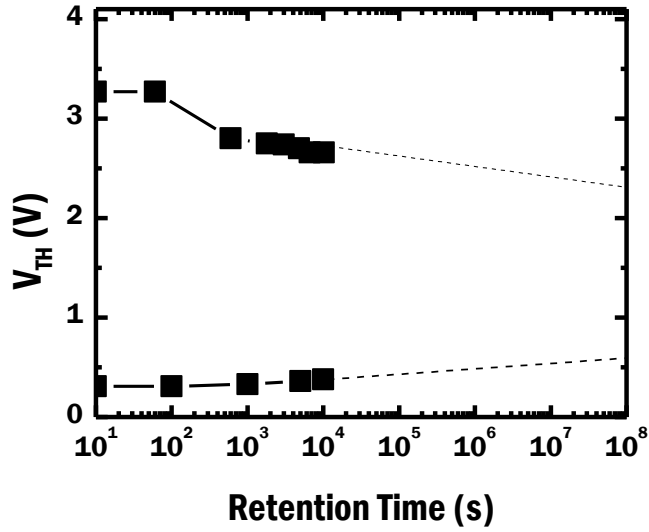
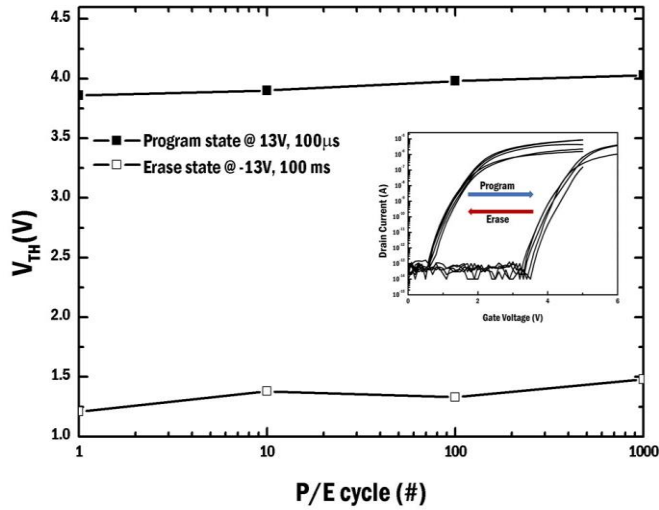


Fig. 3.5. Program and erase speed of gate-all-around SAHOS device ($W_{ch} = 15$ nm, $L_G = 200$ nm)

Figure 3.5 is the program/erase characteristics of the gate-all-around (GAA) nanowire devices. In this work, threshold voltage is defined as the value of the cell gate voltage, corresponding to 10 nA of bit-line current. Note that erase saturation does not occur, which is a usual phenomenon in SONOS devices [36]. The main cause of the erase saturation is electron current injected from the gate to the charge trapping layer, when high electric field is applied in erase operation. When Al_2O_3 ($\kappa \sim 9.3$) replaces SiO_2 ($\kappa \sim 3.9$) as a blocking oxide, the electric field applied to the blocking oxide is reduced due to its higher permittivity. In other words, the tunneling barrier width that electrons should pass through gets longer, and the chance of electrons tunneling through the blocking oxide is reduced significantly. Consequently, gate-all-around structure and high- κ dielectric contribute to the high program/erase speed and wide memory window.



(a)



(b)

Fig. 3.6. (a) Retention characteristics of SAHOS device at 85°C. (b) Endurance characteristics of SAHOS device. (Inset) Transfer characteristics after program/erase cycles ($W_{ch} = 1 \mu\text{m}$, $L_G = 1 \mu\text{m}$)

Figure 3.6(a) illustrates the retention characteristics of nanowire devices. We can expect that there will be > 30 % charge loss after 3 years at 85 °C. Most of the charge loss occurs in a few minutes. The initial charge loss is mainly due to shallow traps in the trapping layer. This mechanism should be studied further.

The endurance characteristic is shown in Fig. 3.6(b). After 1000's program and erase cycles, memory window remains at 2.76 V.

Chapter 4

Fabrication of Channel Stacked NAND Flash Memory with High-k

4.1 Introduction

In this chapter, fabrication process and measurement results of channel stacked NAND flash memory with high- κ are presented. By adopting high- κ materials as charge trapping and blocking layer, the total thickness of memory dielectric can be reduced to ~18 nm. Comparable memory window and retention characteristics to our previous works can be achieved.

4.2 Fabrication Method

In order to demonstrate the feasibility of high- κ charge trapping layer and blocking layer into channel stacked NAND flash memory, fabrication was carried out. The fabrication method is similar to our previous works which use ONO memory dielectric stack. Fabrication process is described in Fig. 4.1.

Fig. 4.1. (a) SiGe/Si/SiGe/Si --- layers are epitaxially grown on the silicon substrate by ultra-high vacuum chemical vapor deposition (UHVCVD) to make the stacked single crystalline silicon channel. Each layer is 40 nm as the optimum thickness to avoid lattice dislocation, and in-situ doped with boron whose concentration is $5 \times 10^{17}/\text{cm}^3$. This process was conducted by IQE silicon corp. in UK.

Fig. 4.1. (b) After initial cleaning (SC1, SC2, and HF cleaning), hydrogen silsesquioxane (HSQ)-based negative electron-beam lithography (EBL) and i-

line photolithography were carried out. After that, reactive ion etch (RIE) was performed to form active region.

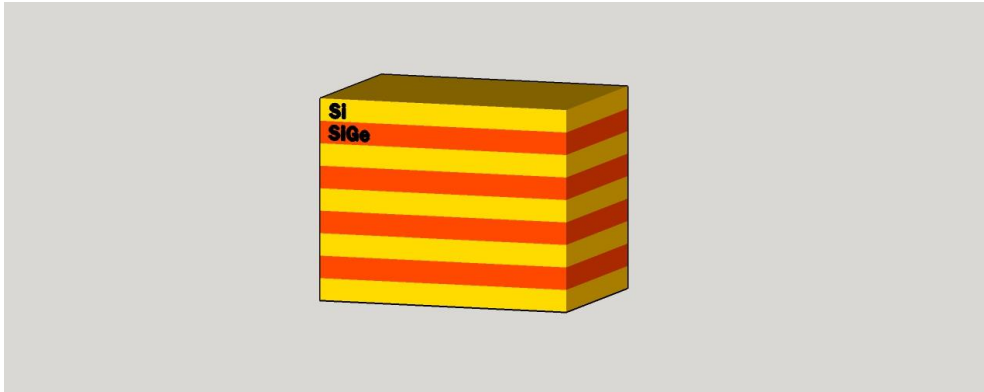
Fig. 4.1. (c) To remove polymer residual and hard mask, SC-1 cleaning and HF cleaning steps are carried out. To detach silicon channel layer, SiGe between silicon layers is selectively removed by chemical dry etch (CDE) and wet etch. $\text{NH}_4\text{OH}:\text{H}_2\text{O}_2$ solution was used as an etchant. During these isotropic etch process, the fin is rounded as well.

Fig. 4.1. (d) Tunneling oxide is thermally grown, and HfO_2 and Al_2O_3 layers are deposited with atomic layer deposition (ALD) process as a charge trapping layer and a blocking layer, respectively. After that doped poly silicon for gate material is deposited and chemical mechanical polishing (CMP) is performed for planarization.

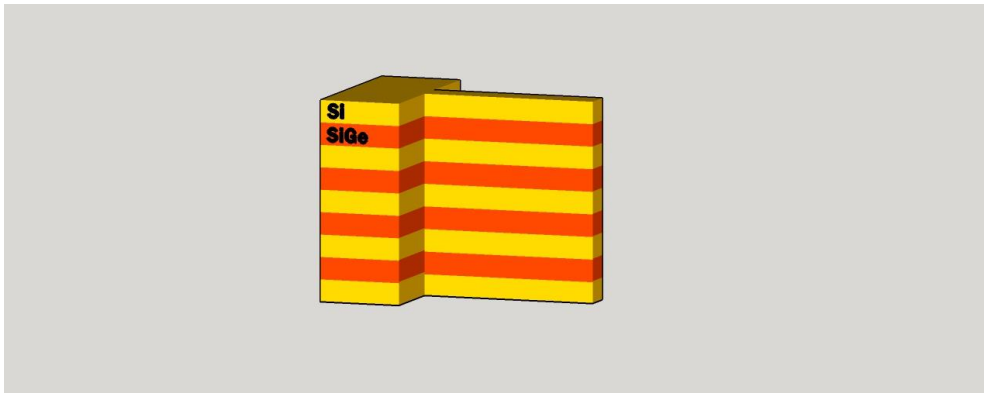
Fig. 4.1. (e) For patterning WLs, SSLs, and GSL, e-beam lithography and photolithography using HSQ hard mask is carried out, followed by RIE. Main etchant gas is HBr.

Fig. 4.1. (f) To remove residual poly-Si, isotropic etch using CF_4 plasma is performed. After that, n^+ and p^+ region is defined with ion implantation.

Fig. 4.1. (g) Inter-layer dielectric (ILD) is deposited using plasma enhanced chemical vapor deposition (PECVD) process, followed by contact etch and metallization. For metallization, Ti/TiN/Al/TiN stack is deposited using sputter.



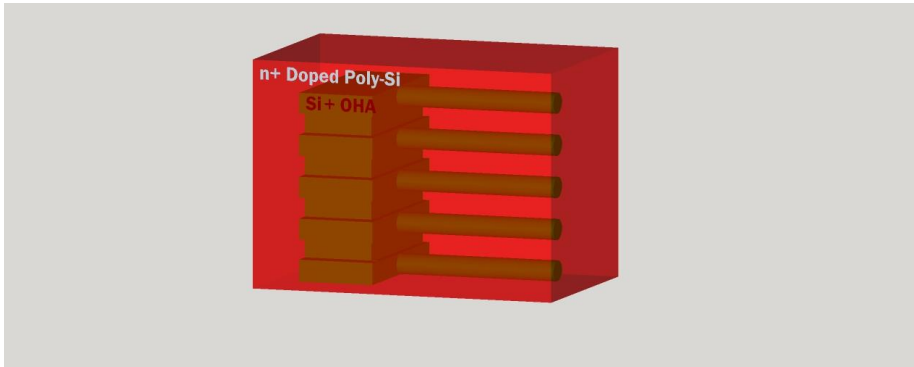
(a)



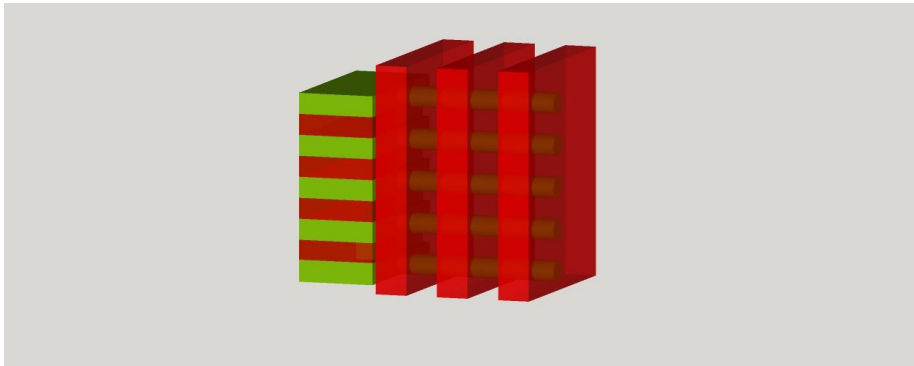
(b)



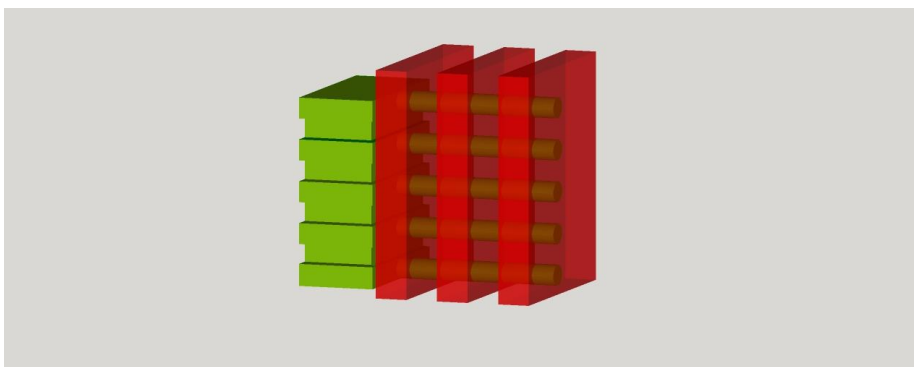
(c)



(d)



(e)



(f)

Fig. 4. 1. Fabrication process of channel stacked array with high- κ

4.3 Key Process Steps of CSTAR with high- κ

4.3.1. Single Crystalline Silicon Channel

The biggest strength of channel stacked array (CSTAR) is that it can use single crystalline silicon channel. Whereas vertical channel should be deposited after forming the hole through stacked gates, lateral channel can be grown epitaxially at the first step of process. Single crystalline silicon channel has advantages over poly-Si channel with respect to mobility, uniformity, and reliability.

Single crystalline channel for 3D NAND is grown using $\text{Si}_{1-x}\text{Ge}_x$ alloy as a sacrificial layer. The important point to be considered is a critical thickness in epitaxial growth of heterostructures. In case of $\text{Si}_{1-x}\text{Ge}_x$ alloy, its lattice constant is larger than that of Si (5.530 Å) and smaller than that of Ge (5.658 Å). It means that $\text{Si}_{1-x}\text{Ge}_x$ layer on the Si substrate is under compressive strain [37]. This stress can cause misfit dislocation if its thickness is larger than the critical thickness. Since the critical thickness is inversely proportional to the

Ge fraction of SiGe, the Ge fraction should be optimized to grow single crystalline Si channel without dislocation.

Fig. 4.2 is a transmission electron microscope (TEM) image of Si/SiGe stacked layers. Si/SiGe layers were grown by ultra-high vacuum chemical vapor deposition (UHVCVD). Ge fraction of SiGe layer is ~30%, and each layer is 40 nm thick. Single crystalline silicon channel is successfully grown without dislocation, as shown in Fig. 4.2.

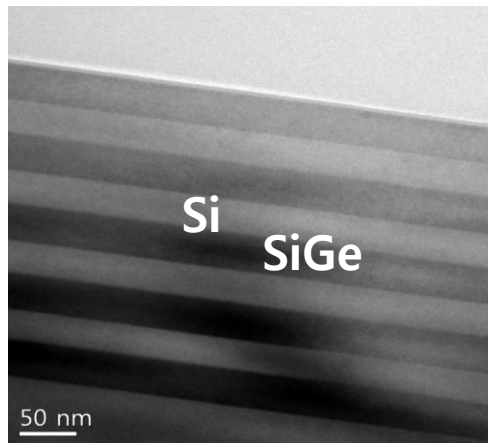


Fig. 4.2. A TEM image of SiGe/Si/SiGe... epitaxially grown layers.

4.3.2. Fin Patterning

For CSTAR, the etch slope in active fin patterning is closely related to the channel diameter. The difference in channel diameter induced by the etch slope causes the variation of memory performance among cells in different layers. Similarly, a slight deviance from a right angle of the etch slope results in drastic change of the dimensions among the gates. It is important to achieve a steep etch slope in highly stacked 3D NAND.

In fabrication process of CSTAR, active patterning and gate patterning are carried out using e-beam lithography and photolithography followed by dry etch process. In process of dry etching, HSQ e-beam resist slope is transferred to Si etch slope. Therefore, the optimization of etch recipe is required to minimize for HSQ being etched while etching Si fin.

There are two main etch processes; removal of native oxide by using Cl_2 gas and Si etch by using HBr gas. The etch time of Cl_2 etch process and platen power in HBr etch process is optimized and Table 4.1 shows the conditions of

dry etch process. With these conditions, silicon fins with steep etch slope can be obtained as shown in Fig. 4.3.

Table 4.1 Silicon etch conditions

	Coil Power [W]	Platen power [W]	HBr flow rate [sccm]	O ₂ flow rate [sccm]	Pressure [mTorr]	Cl ₂ etch time [s]
Standard	900	80	40	2	3	7
New	900	60	40	2	3	5

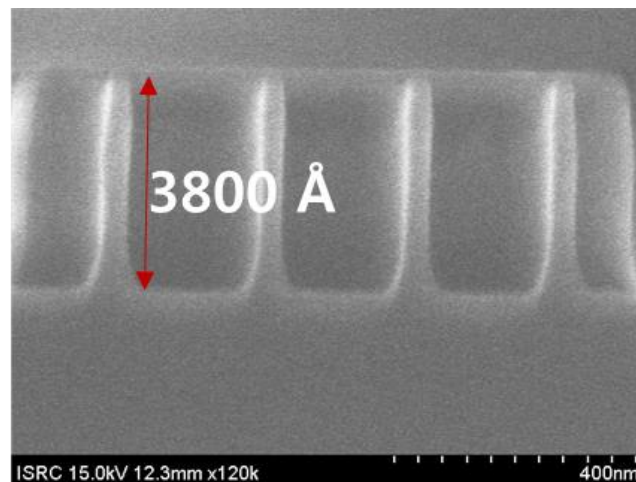


Fig. 4.3. An SEM image of silicon fins with steep etch slope

4.3.3. Stacked Nanowires

Formation of stacked nanowire channels is the critical process in fabrication of CSTAR. Methods of forming stacked channels have been developed and the key process is to remove SiGe sacrificial layers between stacked channels. To selectively remove SiGe layers, both chemical dry etch (CDE) and wet etch are used. Combining CDE which is relatively fast and selective process and wet etch process, SiGe can be removed successfully and Si channel can be rounded as well.

CDE utilizes the plasma generated in a radical generator. Radicals chemically reactive are distributed equally in the process chamber and attached to the surface without any directionality, which results in isotropic etching. Reaction between radicals and exposed surface yields volatile materials and continues further reaction. Table 4.2 shows the process condition of SiGe isotropic etch using CDE. The main etchant of this reaction is fluorine. The selectivity between Si and SiGe is resulted from the difference in binding energy of Si-Si

and Si-Ge. The binding energy of Si-Ge is 2.12 eV, which is smaller than that of Si-Si (2.31 eV). Si-Ge bonds are broken easily and make Si-F or Ge-F bonds easily compared with Si-Si bonds. Continuous reaction between fluorine atoms and Si- or Ge- yields volatile SiF_x and GeF_y , which results in complete removal of SiGe layers [38], [39]. With this mechanism, SiGe can be successfully removed with selectivity of 20:1 to Si (Fig. 4.4).

Table 4.2. Process condition of SiGe selective etch using CDE

Power [W]	Pressure [mTorr]	CF ₄ [sccm]	O ₂ [sccm]	N ₂ [sccm]
700	350	80	12	12

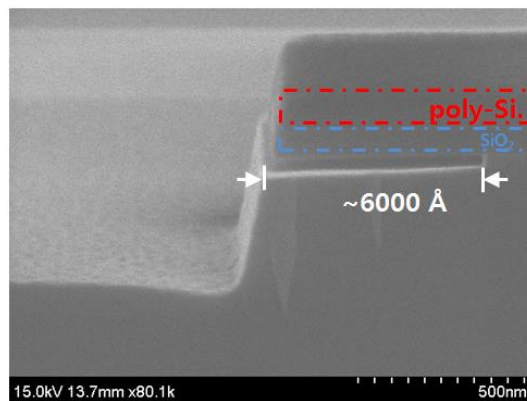


Fig. 4.4. An SEM image showing the result of selective SiGe etch

Another method used in selective etching of SiGe is wet etch process using ammonia-peroxide mixture (APM). The $\text{NH}_4\text{OH}:\text{H}_2\text{O}_2:\text{deionized water}$ ratio is 1:8:64 and process temperature is 65°C . In this mixture, H_2O_2 plays a role as an oxidizer, and NH_4OH as an oxide etchant. The oxidation speed is greater in case of Ge than Si. Oxidized Ge can be etched by NH_4OH more rapidly so that selective etch with selectivity of $>5:1$ could be achieved (Fig. 4.5). Moreover, APM solution contributes to corner-rounding of Si nanowire, since it etches the corner region faster than other regions.

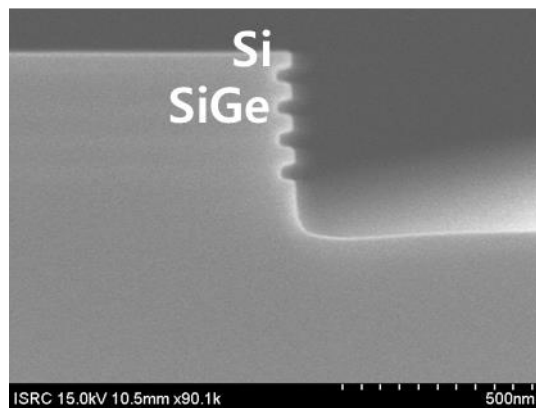


Fig. 4.5. An SEM image showing the result of selective SiGe etch using APM solution

After stacked Si channels being formed, HfO₂ and Al₂O₃ layers were deposited using atomic layer deposition (ALD). Hf[N(CH₃)(C₂H₅)]₄ was used as a precursor for HfO₂, and Al(CH₃)₃ was used as a precursor for Al₂O₃. Process temperature was 300°C. With the great uniformity of ALD process, HfO₂ and Al₂O₃ layers are deposited uniformly around nanowires (Fig. 4.6)

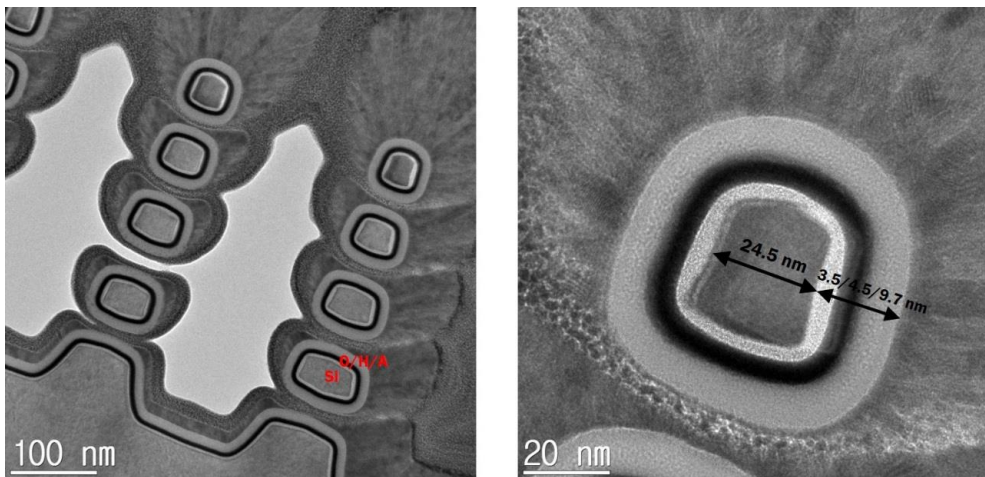
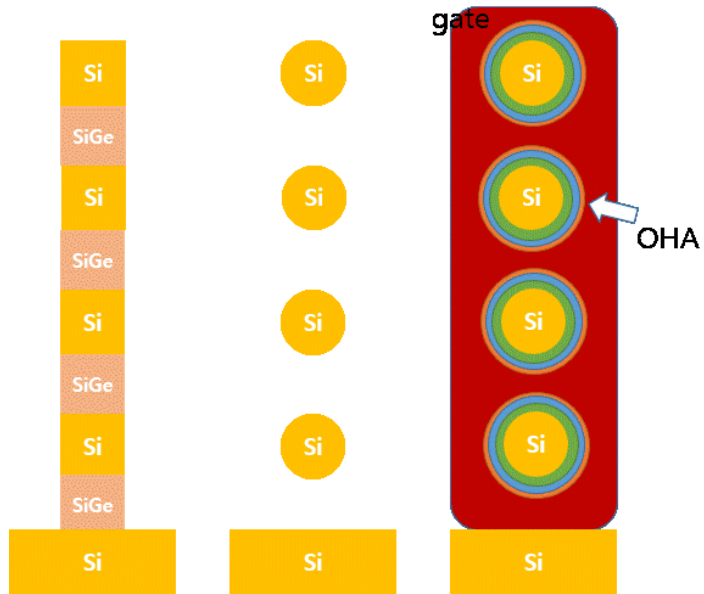


Fig. 4.6. A schematic of stacked nanowires and a TEM image of cross-section view of stacked Si nanowires surrounded by SiO₂, HfO₂, Al₂O₃, and Ti. Ti is the capping layer to protect nanowires in preparation for TEM inspection.

4.3 Measurement Results

Using the fabricated devices, various characteristics are examined. Fig. 4.7 depicts the equivalent circuit of the fabricated devices. The fabricated array has 4 stacked layers, and each layer has three word-lines and four bit-lines. String select transistors and ground select transistors have the same dielectric composition as that of memory cells.

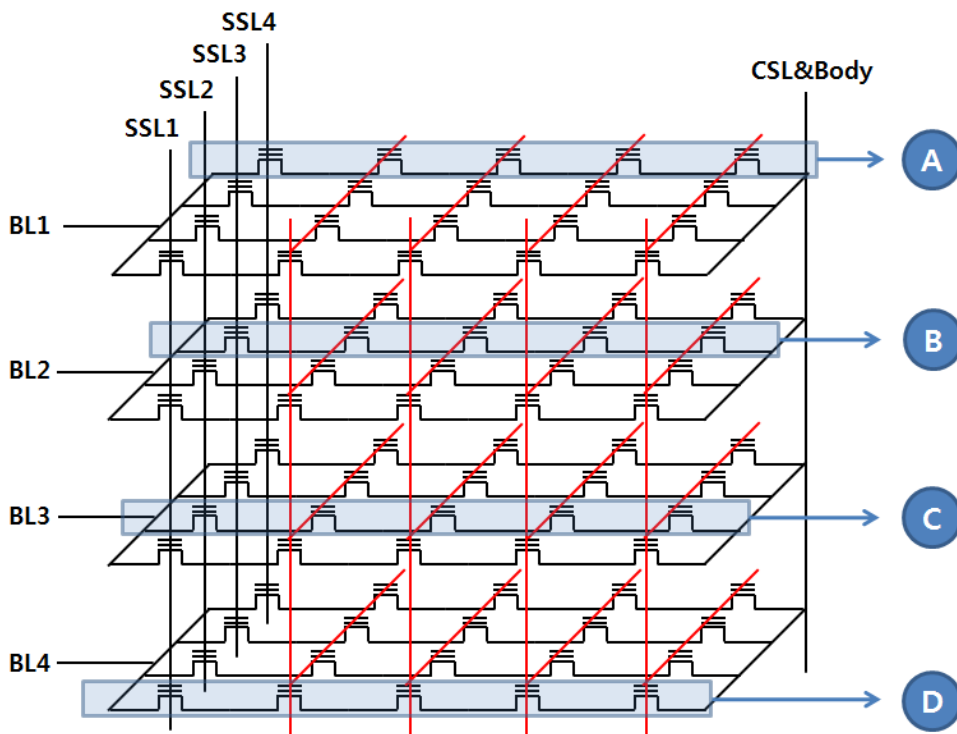


Fig. 4.7. CSTAR array and its fabricated 4-type devices

Figure 4.8 is a microscope top view of fabricated array, and Fig. 4.9 is a cross-sectional TEM image of the fabricated array. Nanowire diameter is < 20 nm, and the diameter difference among cells in different layers is very small thanks to the optimized active etch process.

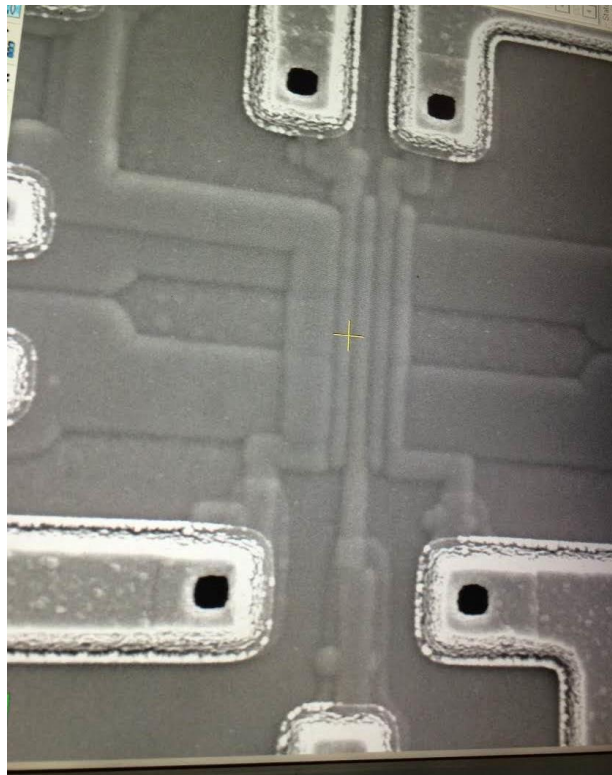


Fig. 4.8. Microscope top view of fabricated device

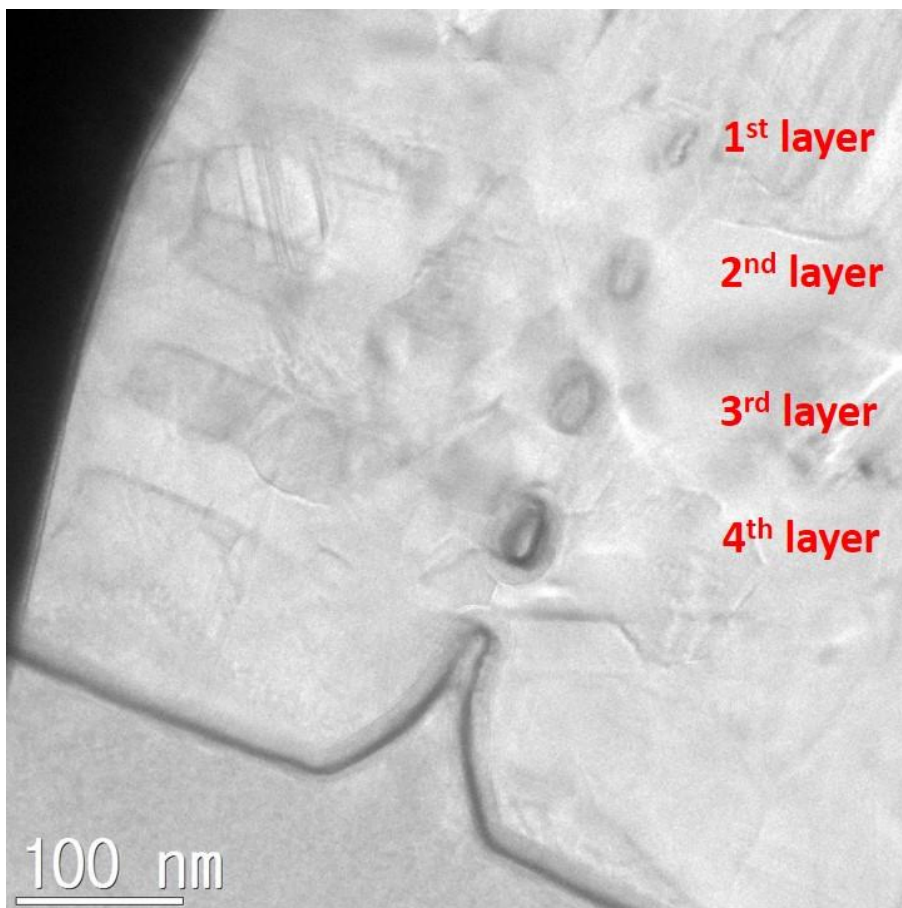


Fig. 4.9. Cross-sectional TEM image of the fabricated array.

Layer selection and program/erase operation with the fabricated array is verified as shown in Fig. 4.10. In CSTAR, a layer is selected with corresponding bit-lines, and a string in a layer is selected with string selection line (SSL) transistors. By turning on the corresponding SSL transistor, the string can be selected as verified in Fig. 4.10.

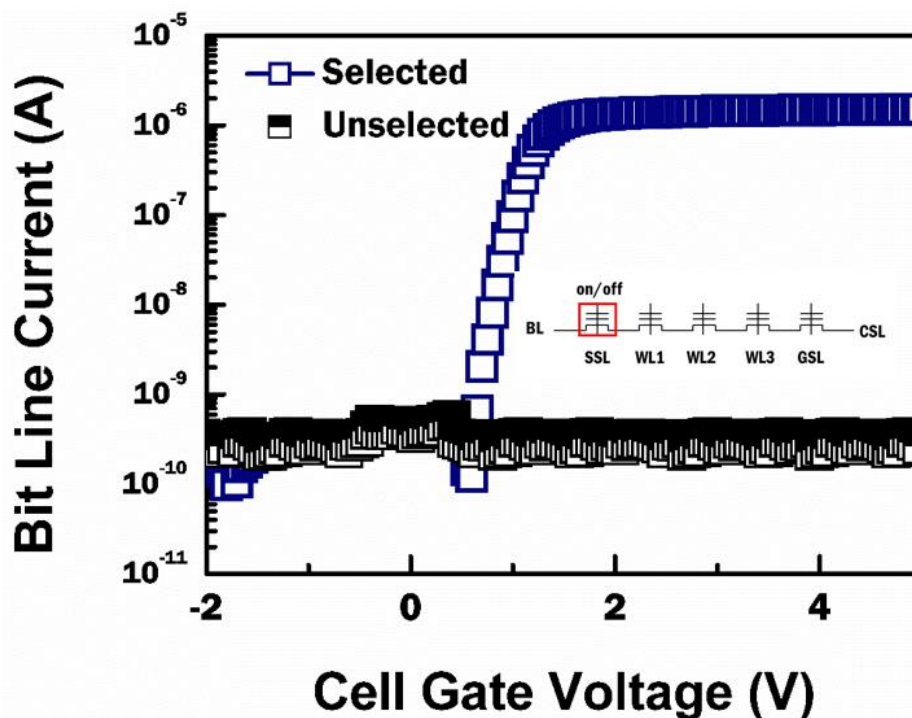


Fig.4.10. Bit line current vs. cell gate voltage. By applying V_{cc} to SSL transistor, we can select a string in the layer.

Fig. 4.11 shows the program/erase characteristics of selected cell. Its cell gate length is 90 nm, and space between gates is 70 nm. Threshold voltage is defined as the value of the cell gate voltage, corresponding to 10 nA of bit-line current. 4 V of V_{TH} can be achieved when $V_{pgm} = 13$ V and $V_{ers} = -14$ V. Fig. 4.12 is program and erase speed of the fabricated array.

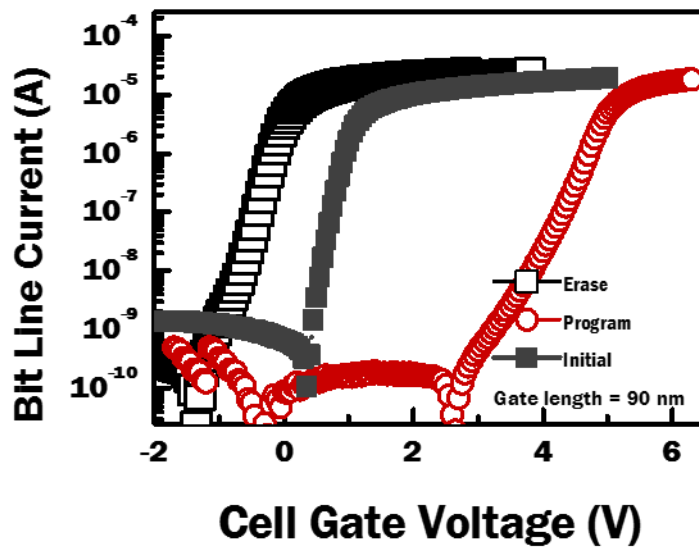


Fig. 4. 11. Program and erase characteristics of fabricated array. 4 V of V_{TH} window can be achieved with $V_{pgm} = 13$ V and $V_{ers} = -14$ V.

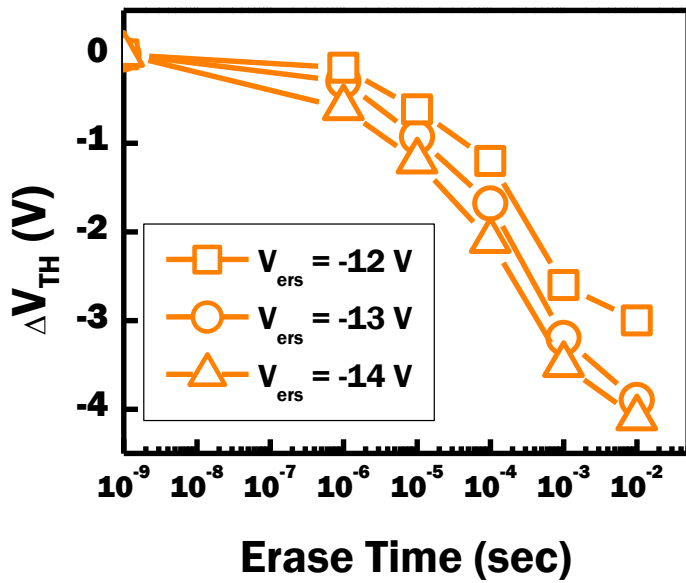
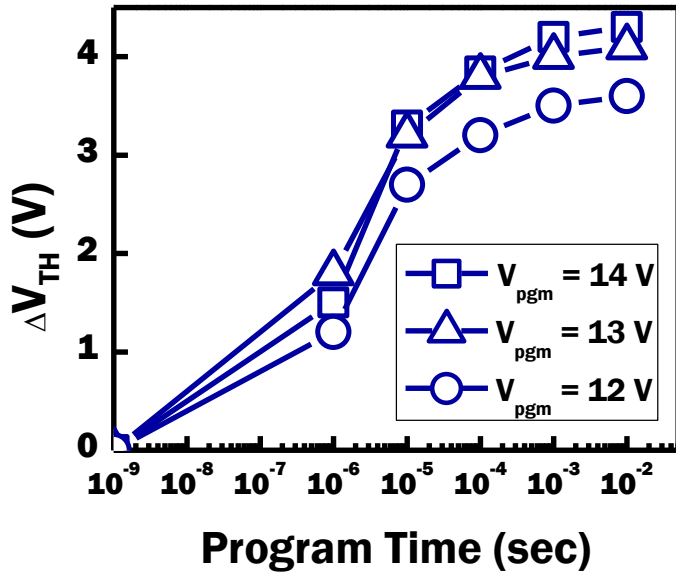


Fig. 4. 12. Program and erase speed of fabricated array.

Reliability characteristics are measured as shown in Fig. 4.13. It is expected that V_{TH} window will be about 2 V after 10 years at room temperature. The large amount of charge loss occurred in 1000s, and it is attributed to electrons in shallow traps. In the case of endurance characteristic, after 1000's program and erase cycles, V_{TH} window remains at 2.5 V. Although its retention characteristics are poorer than SONOS devices, but it can be improved with optimized deposition technique [40].

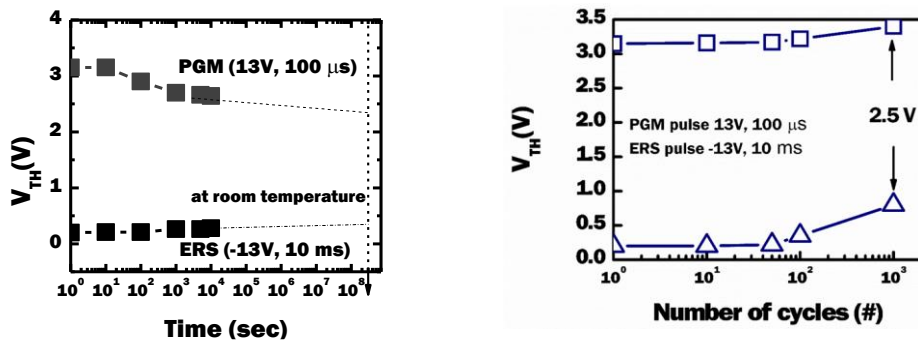


Fig. 4. 13. Reliability of fabricated array.

4.5 Comparison with Previous Works

To evaluate the memory performance of this work, memory characteristics are compared with previous works. Figure 4.14 shows program and erase speed of ONO devices fabricated in 2013 [27]. Devices were also fabricated in ISRC and overall fabrication process is similar to this work besides ONO dielectric layers. As shown in Fig. 4. 14, program speeds are comparable, and erase speed is better in case of high- κ dielectric layer. It is because high- κ dielectric layer effectively reduce the injection of electrons from the gate to the channel. In case of ONO devices, erase bias cannot be larger than 13 V due to back tunneling issue, but there is no back tunneling even 14 V is used as erase bias in case of AHO devices.

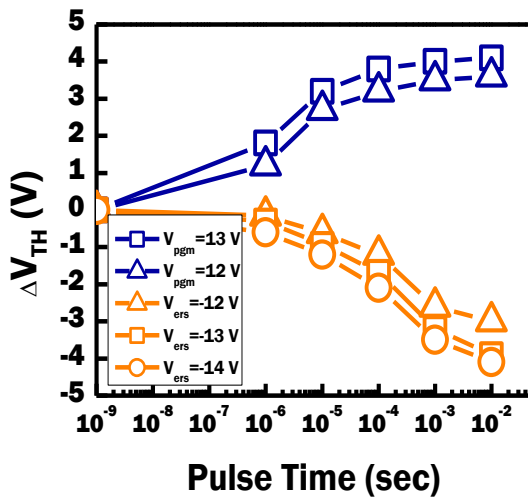
Table 4.3 summarizes the comparison of memory performance with previous works that use ONO dielectrics.

In this work, the total dielectric thickness can be reduced to 17.7 nm using

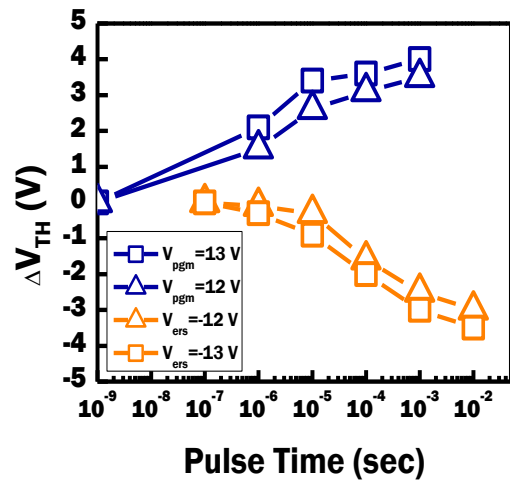
4.5-nm-thick HfO_2 charge trapping layer. While SONOS devices have back tunneling issue in erase operation, SAHOS devices can show fast erase speed in erase bias larger than 13 V. Therefore, wider memory window can be achieved. Also, it has been widely known that the erase speed of HfO_2 based charge trap memory is slower than that of Si_3N_4 based devices, because of a larger valence-band offset with respect to the SiO_2 [41]. However, our fabricated devices show comparable erase speed compared to ONO devices, and this can be explained by the use of high- κ trapping and blocking layers allows higher electric field to be applied to the tunneling layer as depicted in Fig. 4. 15. Fig. 4. 15 (a) is energy band diagram in planar devices under erase operation. Higher electric field compensates the difference of valence band offset and holes in valence band can tunnel into the trapping layer. Even in nanowire structure, high- κ trapping layer and blocking layer can be helpful.

Table 4.3. Comparison of memory performance with previous works

	Dielectric thickness	Memory window	Retention characteristics	Endurance	Main issue
CSTAR with SONOS	20 nm	3.5	32 % charge loss	~5 % decrease	Back tunneling in large erase bias
CSTAR with high-κ	17.7 nm	3.9	42 %	~10 % decrease	Retention characteristics



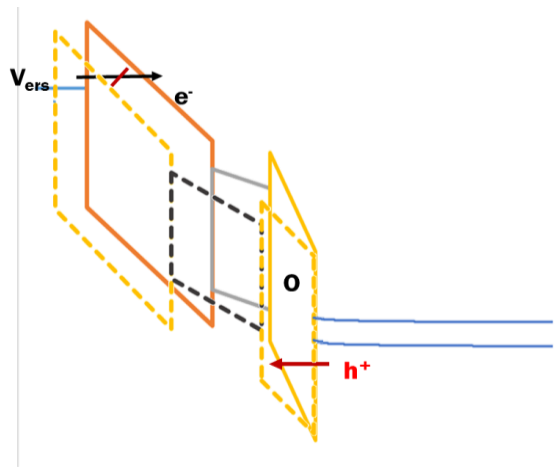
(a)



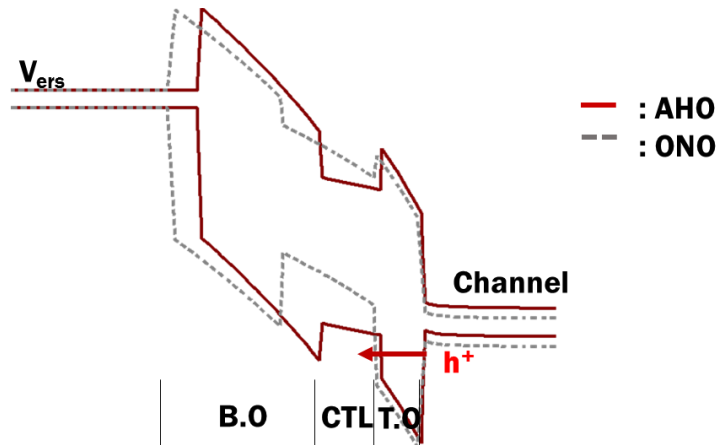
(b)

Fig. 4. 14. Comparison of program/erase speed. (a) this work (b) ONO devices

fabricated in 2013.



(a)



(b)

Fig. 4. 15. Energy band diagram in erase operation (a) planar devices (b) nanowire devices (channel radius = 10 nm). AHO thickness : 9.7/4.5/3.5 nm (dash lines), ONO thickness : 9/7/4 nm (dotted lines)

Chapter 5

Novel Program Operation in CSTAR

5.1. Introduction

In channel stacked array (CSTAR), the cross sections of the stacked channels are not circular but elliptical with different geometry eccentricity due to the declined etch slope. Even though etch slope can be improved with advanced etch technology, it is inevitable to avoid geometric eccentricity in highly stacked 3D NAND.

To analyze the characteristic of elliptical channel stacked NAND flash memory, 3D TCAD simulation (Synopsys Sentaurus) is used. The devices are designed as follows. The p-type body is lightly doped with a concentration of $1 \times 10^{15} \text{ cm}^{-3}$, and n^+ polysilicon gate is used. Tunneling oxide/nitride/blocking

oxide (O/N/O) thickness is 4/8/8 nm, and the gate length is 50 nm. For nitride trapping, the uniform trap distribution $1 \times 10^{20} \text{ cm}^{-3}$ concentration and trap energy level of 1.0 eV from conduction band are assumed.

5.2. Simulation Results

The program operation of NAND flash memory is composed of several program pulse steps. For FN tunneling used as programming mechanism, the gate voltage of the program pulse V_{pp} is increased by a constant value V_{step} after each program step as shown in Fig. 5.1. Therefore, this programming scheme is called Incremental Step Pulse Programming (ISPP). Basically, the cell's V_{TH} shift (ΔV_{TH}) by one programming pulse is equal to V_{step} due to the self limitation of the tunneling current. Therefore, ISPP slope in the graph of V_{TH} vs. pulse is 1 as indicated in Fig. 5.1. Figure 5.2 shows ISPP characteristic of circular and elliptic channels. The ISPP slope of circular channel is ~ 1 . However, the ISPP characteristic of elliptic channel is different from that of

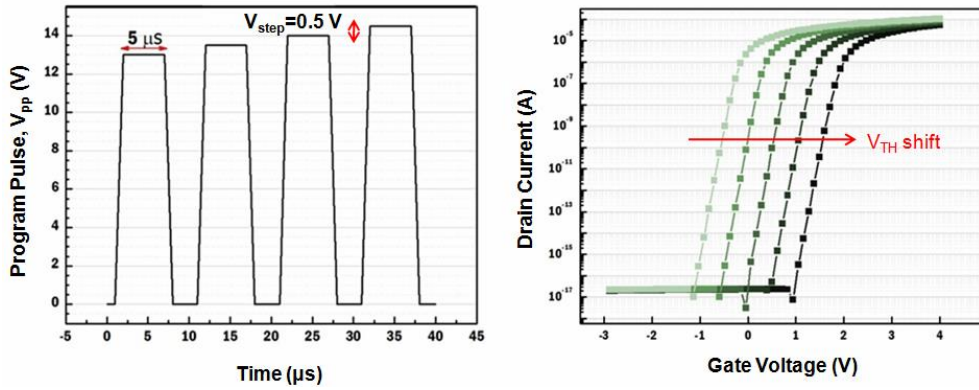


Fig. 5. 1. ISPP wave form and I-V curve of circular channel.

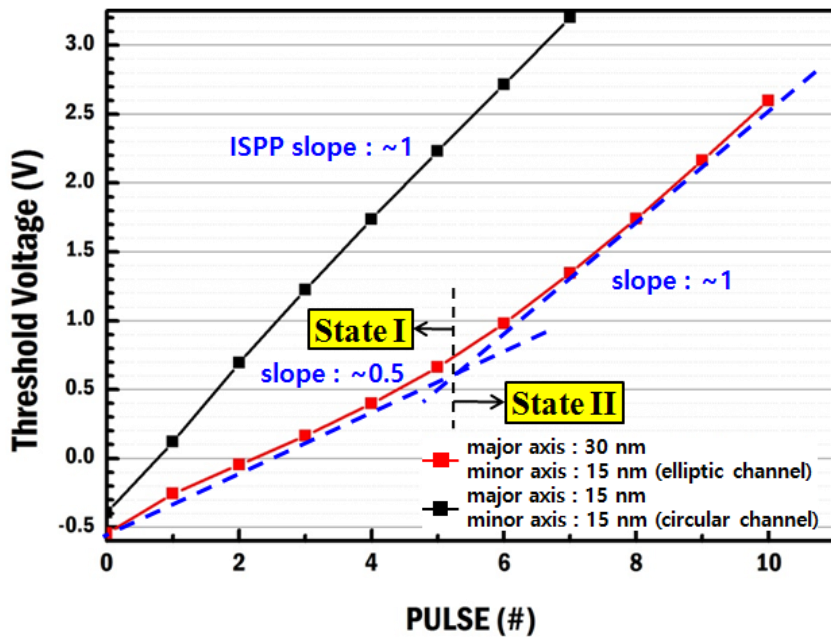


Fig. 5. 2. ISPP characteristics of circular channel and elliptic channel

circular channel. In the early state (State I), the ISPP slope is significantly degraded (~ 0.5). After 5~6 programming pulses, the ISPP efficiency recovers with ISPP slope of ~ 1 (State II). The ISPP slope degradation of elliptical channel becomes more severe as the eccentricity of elliptical channel increases. In order to understand the degradation of ISPP slope in the elliptic channel, the electron trapped charge distribution and channel current behavior are analyzed as shown in Figs. 5.3 and 5.4.

In the early ISPP state (State I), the trapped electron distribution along the nitride layer is asymmetric owing to the different electric field concentrations near the major and minor axes. The curvature of the nitride trapping layer near the minor axis is smaller than that of the layer near the major axis. Consequently, more electrons are trapped in the nitride layer near the major axis due to the field concentration. In this situation, the device is controlled by the local minimum V_{TH} regions near the minor axis where less localized trapped charge makes channel inversion easier than the region higher V_{TH} regions near the major axis. Therefore, when the elliptical channel flash device

is turned on, the channel current congregates in the vicinity of the minor axis as shown in Fig. 5.4.

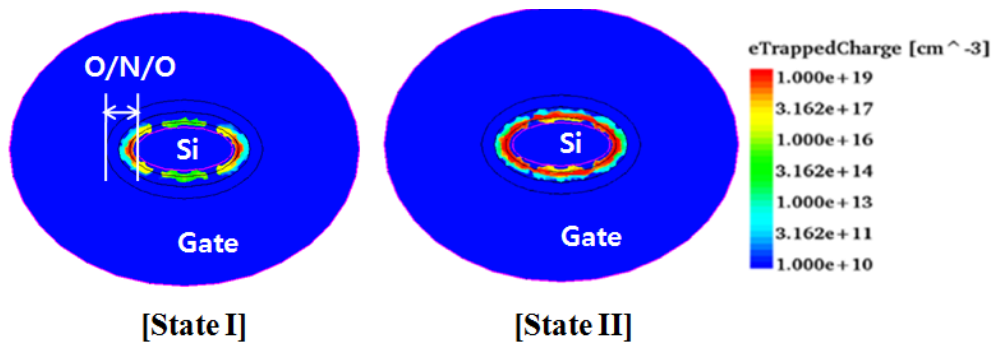


Fig. 5. 3. Electron trapped charge density distributions for the elliptic channel

with major/minor axis=30/15 nm

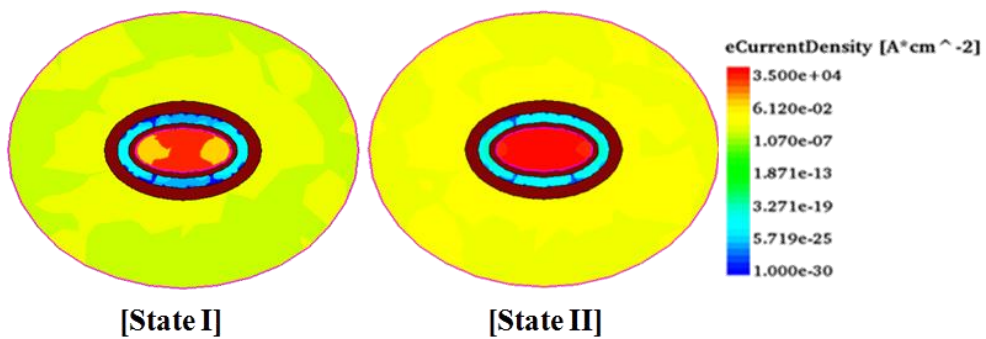


Fig. 5.4. Electron current density distributions for the elliptic channel with

major/minor axis=30/15 nm

The charge of trapped electrons by ISPP cannot influence the entire channel region, which leads to the ISPP slope degradation. For larger eccentricity, the ISPP slope degradation is more significant since the localized electron trapping and channel current distribution becomes more intensified. In the later ISPP state (State II), the already trapped charge decreases the electric field between the gate and the channel and the difference of electric field concentrations originated from the curvature difference disappears. As shown in Fig. 5.3, the electrons are trapped in the entire nitride region. Thus, bit line current flows through the entire channel region. When this self-leveling of local electric field is complete, the ISPP slope recovers to ~ 1 .

In order to control the programmed V_{TH} of the NAND flash memory cell, a bit-by-bit program verify algorithm is used. The ISPP operation is split into several program pulse steps with a V_{TH} verification operation in between. If the V_{TH} of a cell is detected above a target verify level, further ISPP of this single cell will be prohibited by setting a program inhibit condition. And, ISPP with verification algorithm continues until all selected cells have their target V_{TH}

level. Therefore, the total program time of NAND flash array is determined by the slowest cell. In 3D NAND flash memory with channel stacked array, the cell on the ground floor with the highest eccentricity will be the slowest cell due to the most severe ISPP slope degradation. As a consequence, ISPP slope degradation of elliptical cell affects the program speed most significantly. Also, additional programming pulses for programming of slowest cell aggravate program disturbance on the already programmed cells. To solve this disadvantage of channel stacked 3D NAND flash memory, a new programming method which can reduce the program time of the slowest elliptical cell is needed.

The total V_{TH} shift of ISPP algorithm can be expressed by

$$\Delta V_{TH}(\text{total}) = \Delta V_{TH}(\text{first-pulse}) + \sum_{i=1}^N (\text{ISPP Slope})_i \times V_{step}$$

, where $\Delta V_{TH}(\text{first-pulse})$ is the V_{TH} shift by first ISPP pulse, and N is the total number of ISPP pulses. According to the equation, the program time to reach the target V_{TH} level can be reduced by increasing the starting program voltage

or ISPP slope. As the starting program voltage is higher, the ΔV_{TH} (first-pulse) increases as shown in Fig. 5.5. However, ISPP slope is constant regardless of the starting program voltage. Because ΔV_{TH} (first-pulse) is determined by electric field between gate and channel potential, we can also modulate ΔV_{TH} (first-pulse) by applying BL voltage to channel as indicated in Fig. 5.6.

During the program operation in conventional planar NAND flash array, all BLs of selected strings are biased at 0 V [2]. On the other hand, each stacked channel in our proposed Channel STacked ARray (CSTAR) structure [26] is connected with different BL as depicted in Fig. 5.7. We can compensate the program speed variation by applying adaptive BL voltage to each layer with high start program voltage as indicated in Fig. 9(b)

The start program voltage is usually determined to prevent the fastest cell (the upper cell) exceeding the target V_{TH} level by just one ISPP pulse. To maintain the program speed of the upper cell, the highest BL voltage is applied to BL4.

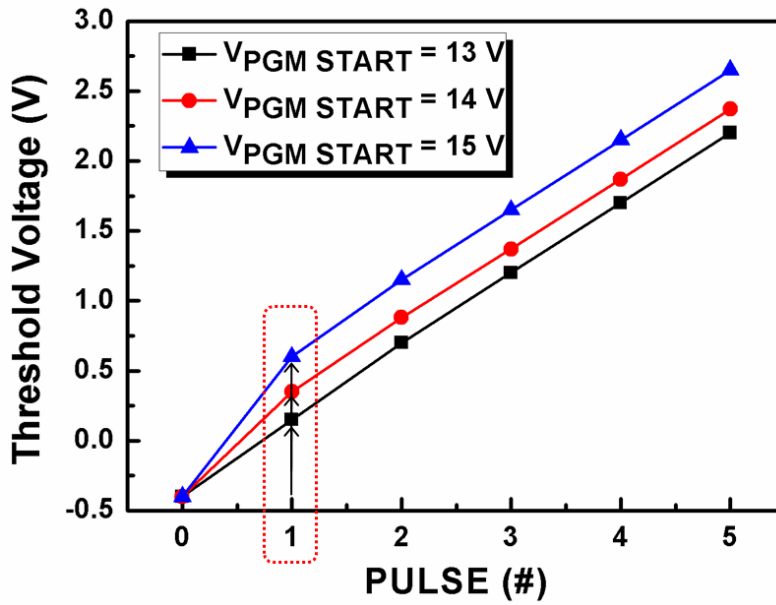


Fig. 5.5. Comparison of ISPP for various starting program voltages of circular channel with diameter=15 nm.

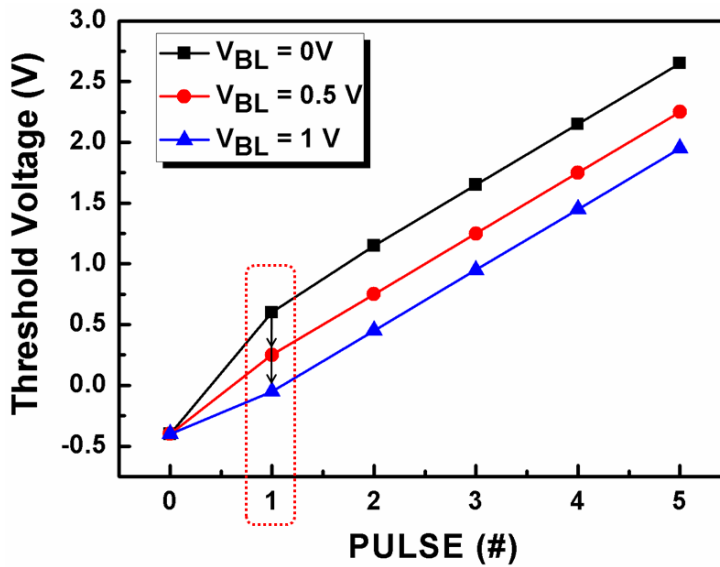


Fig. 5.6. Comparison of ISPP for various BL voltages of circular channel with diameter = 15 nm.

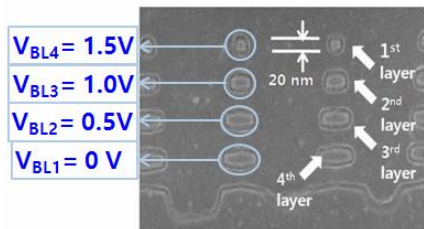
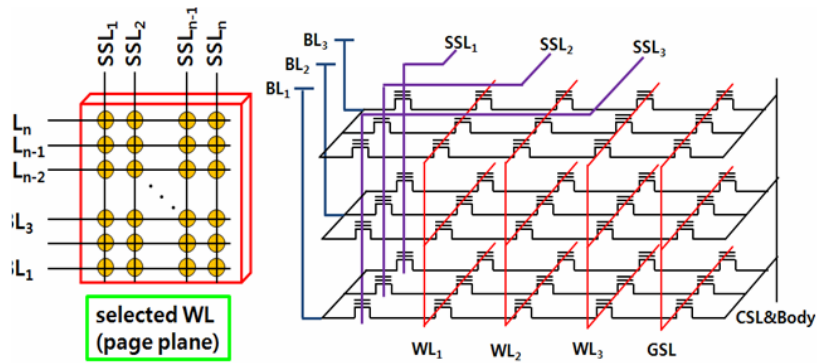


Fig. 5.7. CSTAR structure and its equivalent circuit diagram and BL biasing of new programming method.

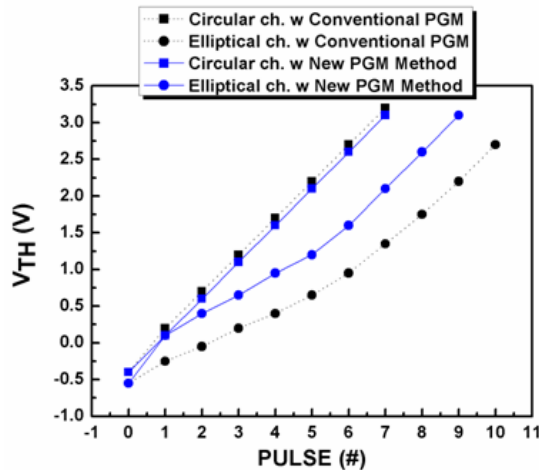


Fig. 5.8. Comparison of ISPP for various BL voltages of circular channel with diameter = 15 nm.

Using this new programming method, we can reduce the program speed variation as shown in Fig. 5.8. Also, we can obtain better program speed performance by reducing the total program pulses (~25%) as indicated in Fig. 5.9.

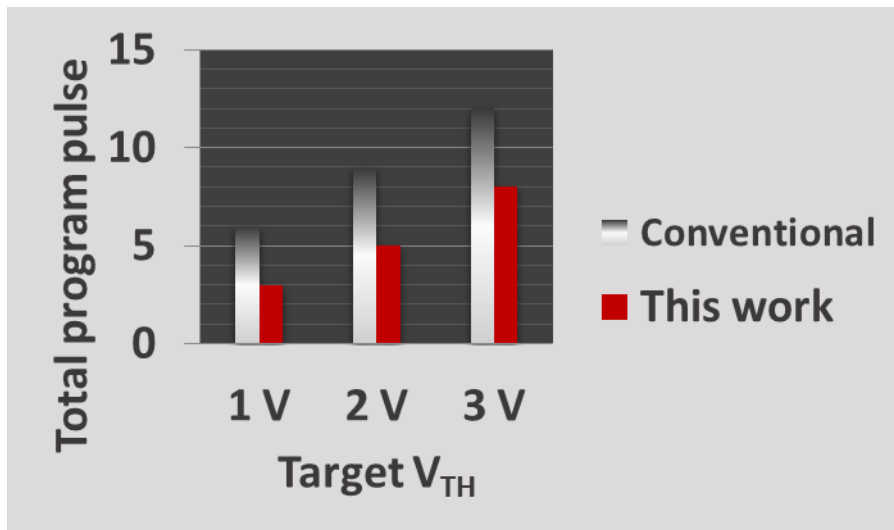
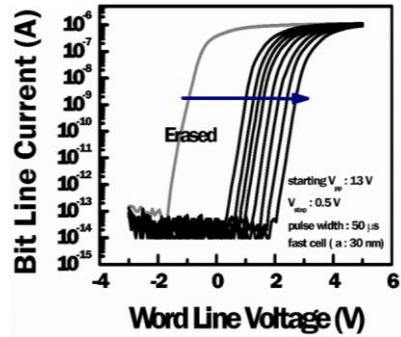
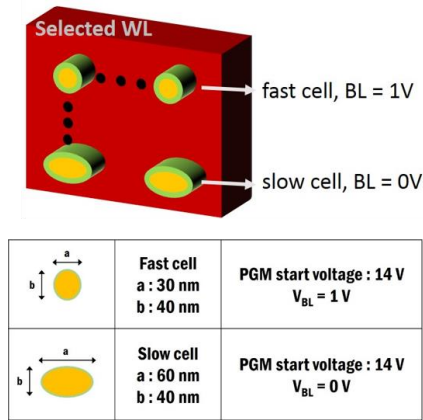


Fig. 5.9. Comparison of the required number of program pulses in the conventional and new programming methods.

5.3. Measurement Results

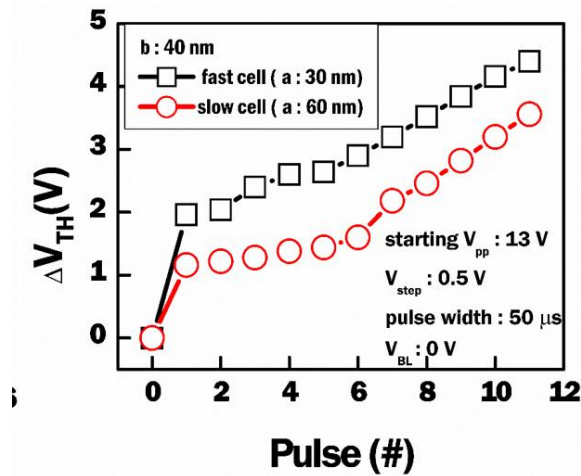
Newly designed programming method to reduce the total program time was verified with measurement. All cells have the same the length of minor axis, but their lengths of major axis are varied from 30 nm to 60 nm. Fig. 5.10 shows proposed program method and the variation of program speed. By applying different V_{BL} , we can compensate the variation of program speed among cells in different layers. We used incremental step pulse program (ISPP) operation of fabricated device. There was program speed variation between fast cell ($a=30$ nm) and slow cell ($a=60$ nm) when conventional program method is applied. When the major axis is longer, the program speed degraded, and this result is consistent with the simulation results. We can reduce the speed difference by applying 1 V to the fastest cell, and increase the overall speed by increasing the starting program voltage as shown in Fig.5.11.

With this method, the number of the total pulses to reach the target program V_{TH} level can be reduced.



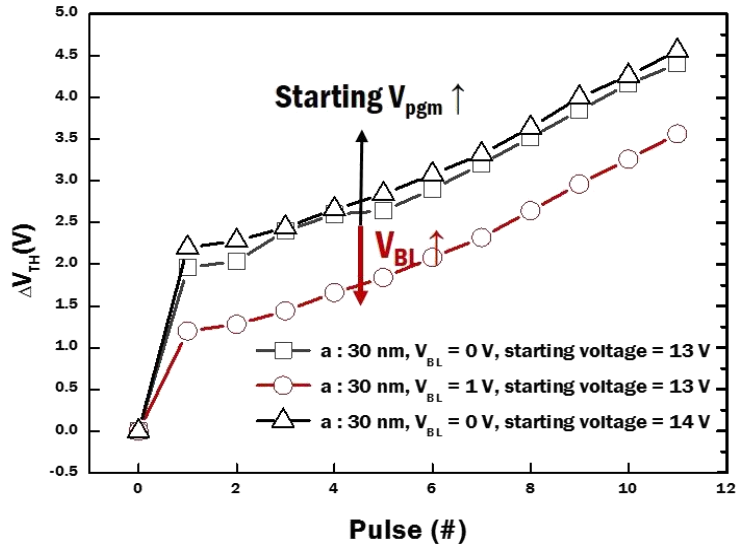
(a)

(b)

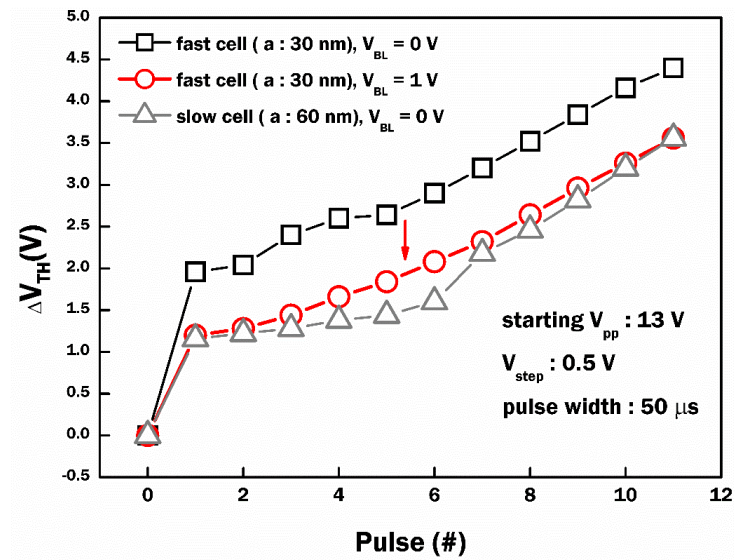


(c)

Fig. 5.10. (a) Proposed program method. (b) Incremental step pulse program (ISPP) operation of fabricated device. (c) Program speed variation between fast cell ($a=30$ nm) and slow cell ($a=60$ nm) when conventional program method is applied.



(a)



(b)

Fig. 5.11. (a) We can adjust program speed by varying V_{BL} and starting V_{pgm} . (b) By

applying V_{BL} to a fast cell, we can reduce the program speed variation

Chapter 6

Conclusions

In this thesis, channel stacked NAND flash memory (CSTAR) with high- κ charge trapping layer is proposed, fabricated, and characterized. For stacked NAND flash memory, the thickness of ONO (memory dielectric layers) is a roadblock in scaling-down of the minimum feature size, because channel diameter can be scaled down to < 20 nm. However, it is challenging to reduce the thickness of oxide-nitride-oxide (ONO) layer, since the charge trapping properties degrade when the Si_3N_4 is made thinner. Adopting HfO_2 as a charge trapping layer enables us to reduce the thickness of charge trapping layer due to its large trap density compared with that of Si_3N_4 .

To demonstrate the feasibility of HfO_2 as a charge trapping layer, CSTAR having 4-layer stacked single-crystalline silicon nanowire channel with GAA structure was fabricated and measured. Stacked nanowire formation, SiGe

selective etch, and uniform deposition of high- κ layer were demonstrated. Consequently, memory characteristics comparable to ONO dielectric flash devices can be achieved, when using HfO₂ trapping layer with more than 35 % reduced thickness.

For the optimization of CSTAR operation, a new programming method was designed. This method was verified with TCAD simulation and measurement using fabricated arrays. With this method, the number of the total pulses to reach the target program V_{TH} level can be reduced.

Bibliography

- [1] “IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays,” *IEEE Std 1005-1998*, p. i, 1998.
- [2] Kang-Deog Suh, Byung-Hoon Suh, Young-Ho Lim, Jin-Ki Kim, Young-Joon Choi, Yong-Nam Koh, Sung-Soo Lee, Suk-Chon Kwon, Byung-Soon Choi, Jin-Sun Yum, Jung-Hyuk Choi, Jang-Rae Kim, and Hyung-Kyu Lim, “A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme,” *Solid-State Circuits IEEE J. Of*, vol. 30, no. 11, pp. 1149–1156, 1995.
- [3] R. Michelsoni, *Inside Solid State Drives (SSDs)*. .
- [4] Jin-Ki Kim, Koji Sakui, Sung-Soo Lee, J. Itoh, Suk-Chon Kwon, K. Kanazawa, Ji-Jun Lee, H. Nakamura, Kang-Young Kim, T. Himeno, Jang-Rae Kim, K. Kanda, Tae-Sung Jung, Y. Oshima, Kang-Deog Suh, K. Hashimoto, Sung-Tae Ahn, and J. Miyamoto, “A 120mm² 64Mb NAND flash memory achieving 180ns/byte effective program speed,” *VLSI Circuits 1996 Dig. Tech. Pap. 1996 Symp. On*, pp. 168–169, Jun. 1996.
- [5] T. Tanaka, Y. Tanaka, H. Nakamura, K. Sakui, H. Oodaira, R. Shirota, K. Ohuchi, F. Masuoka, and H. Hara, “A quick intelligent page-programming architecture and a shielded bitline sensing method for 3 V-only NAND flash memory,” *Solid-State Circuits IEEE J. Of*, vol. 29, no. 11, pp. 1366–1373, Nov. 1994.
- [6] Tae-Sung Jung, Young-Joon Choi, Kang-Deog Suh, Byung-Hoon Suh, Jin-Ki Kim, Young-Ho Lim, Yong-Nam Koh, Jong-Wook Park, Ki-Jong

- Lee, Jung-Hoon Park, Kee-Tae Park, Jang-Rae Kim, Jeong-Hyong Lee, and Hyung-Kyu Lim, "A 3.3 V 128 Mb multi-level NAND flash memory for mass storage applications," *Solid-State Circuits Conf. 1996 Dig. Tech. Pap. 42nd ISSCC 1996 IEEE Int.*, pp. 32–33, Feb. 1996.
- [7] R.-A. Cernea, Long Pham, F. Moogat, Siu Chan, Binh Le, Yan Li, Shouchang Tsao, Tai-Yuan Tseng, Khanh Nguyen, J. Li, Jayson Hu, Jong Hak Yuh, C. Hsu, Fanglin Zhang, T. Kamei, H. Nasu, P. Kliza, Khin Htoo, J. Lutze, Yingda Dong, M. Higashitani, Junnhui Yang, Hung-Szu Lin, V. Sakhamuri, A. Li, Feng Pan, S. Yadala, S. Taigor, K. Pradhan, J. Lan, James Chan, T. Abe, Y. Fukuda, H. Mukai, K. Kawakami, C. Liang, T. Ip, Shu-Fen Chang, J. Lakshmipathi, S. Huynh, D. Pantelakis, M. Mofidi, and K. Quader, "A 34 MB/s MLC Write Throughput 16 Gb NAND With All Bit Line Architecture on 56 nm Technology," *Solid-State Circuits IEEE J. Of*, vol. 44, no. 1, pp. 186–194, 2009.
- [8] Jae-Duk Lee, Sung-Hoi Hur, and Jung-Dal Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *Electron Device Lett. IEEE*, vol. 23, no. 5, pp. 264–266, 2002.
- [9] T. P. Ma and J.-P. Han, "Why is nonvolatile ferroelectric memory field-effect transistor still elusive?," *Electron Device Lett. IEEE*, vol. 23, no. 7, pp. 386–388, Jul. 2002.
- [10] Gang Zhang, L.-K. Wang, Won Jong Yoo, and Ming-Fu Li, "Spatial Distribution of Charge Traps in a SONOS-Type Flash Memory Using a High- k Trapping Layer," *Electron Devices IEEE Trans. On*, vol. 54, no. 12, pp. 3317–3324, Dec. 2007.

- [11] Y. Yamauchi, Y. Kamakura, and T. Matsuoka, "Scalable Virtual-Ground Multilevel-Cell Floating-Gate Flash Memory," *Electron Devices IEEE Trans. On*, vol. 60, no. 8, pp. 2518–2524, Aug. 2013.
- [12] C.-Y. Lu, K.-Y. Hsieh, and R. Liu, "Future challenges of flash memory technologies," *Fourth IEEE Int. Symp. Adv. Gate Stack Technology ISAGST 2007*, vol. 86, no. 3, pp. 283–286, 2009.
- [13] L. Larcher and A. Padovani, "High- κ related reliability issues in advanced non-volatile memories," *21st Eur. Symp. Reliab. Electron Devices Fail. Phys. Anal.*, vol. 50, no. 9–11, pp. 1251–1258, 2010.
- [14] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connell, R. E. Oleksiak, and H. Lawrence, "The variable threshold transistor, a new electrically-alterable, non-destructive read-only storage device," *Electron Devices Meet. 1967 Int.*, vol. 13, pp. 70–70, 1967.
- [15] A. Nitayama and H. Aochi, "Bit Cost Scalable (BiCS) flash technology for future ultra high density storage devices," in *Proceedings of 2010 International Symposium on VLSI Technology, System and Application, VLSI-TSA 2010*, Hsin Chu, 2010, pp. 130–131.
- [16] H. Tanaka, M. Kido, K. Yahashi, M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, M. Sato, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi, and A. Nitayama, "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," in *VLSI Technology, 2007 IEEE Symposium on*, 2007, pp. 14–15.
- [17] R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, Y. Komori, M. Ishiduki, J. Matsunami, T. Fujiwara, Y. Nagata, Li Zhang, Y. Iwata, R.

- Kirisawa, H. Aochi, and A. Nitayama, "Pipe-shaped BiCS flash memory with 16 stacked layers and multi-level-cell operation for ultra high density storage devices," *VLSI Technol. 2009 Symp. On*, pp. 136–137, Jun. 2009.
- [18] Y. Komori, M. Kido, M. Kito, R. Katsumata, Y. Fukuzumi, H. Tanaka, Y. Nagata, M. Ishiduki, H. Aochi, and A. Nitayama, "Disturbless flash memory due to high boost efficiency on BiCS structure and optimal memory film stack for ultra high density storage device," in *Technical Digest - International Electron Devices Meeting, IEDM*, San Francisco, CA, 2008.
- [19] Jaehoon Jang, Han-Soo Kim, Wonseok Cho, Hoosung Cho, Jinho Kim, Sun Il Shim, Younggoan Jang, Jae-Hun Jeong, Byoung-Keun Son, Dong Woo Kim, Kihyun, Jae-Joo Shim, Jin Soo Lim, Kyoung-Hoon Kim, Su Youn Yi, Ju-Young Lim, Dewill Chung, Hui-Chang Moon, Sungmin Hwang, Jong-Wook Lee, Yong-Hoon Son, U-In Chung, and Won-Seong Lee, "Vertical cell array using TCAT(Terabit Cell Array Transistor) technology for ultra high density NAND flash memory," in *VLSI Technology, 2009 Symposium on*, 2009, pp. 192–193.
- [20] Wonjoo Kim, Sangmoo Choi, Junghun Sung, Taehee Lee, Chulmin Park, Hyungsoo Ko, Juhwan Jung, Inkyong Yoo, and Yoondong Park, "Multi-layered Vertical Gate NAND Flash overcoming stacking limit for terabit density storage," in *VLSI Technology, 2009 Symposium on*, 2009, pp. 188–189.
- [21] H.-T. Lue, T.-H. Hsu, Y.-H. Hsiao, S. P. Hong, M. T. Wu, F. H. Hsu, N. Z.

- Lien, S.-Y. Wang, J.-Y. Hsieh, L.-W. Yang, T. Yang, K.-C. Chen, K.-Y. Hsieh, and C.-Y. Lu, "A highly scalable 8-layer 3D vertical-gate (VG) TFT NAND flash using junction-free buried channel BE-SONOS," in *Digest of Technical Papers - Symposium on VLSI Technology*, Honolulu, HI, 2010, pp. 131–132.
- [22] Chun-Hsiung Hung, Hang-Ting Lue, Kuo-Pin Chang, Chih-Ping Chen, Yi-Hsuan Hsiao, Shih-Hung Chen, Yen-Hao Shih, Kuang-Yeu Hsieh, M. Yang, J. Lee, Szu-Yu Wang, Tahone Yang, Kuang-Chao Chen, and Chih-Yuan Lu, "A highly scalable vertical gate (VG) 3D NAND Flash with robust program disturb immunity using a novel PN diode decoding structure," *VLSI Technol. VLSIT 2011 Symp. On*, pp. 68–69, Jun. 2011.
- [23] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, "Negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories," in *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, 1998, pp. 234–235.
- [24] A. Hubert, E. Nowak, K. Tachi, V. Maffini-Alvaro, C. Vizioz, C. Arvet, J.-P. Colonna, J.-M. J.-M. Hartmann, V. Loup, L. Baud, S. Pauliac, V. Delaye, C. Carabasse, G. Molas, G. Ghibaudo, B. De Salvo, O. Faynot, and T. Ernst, "A stacked SONOS technology, up to 4 levels and 6nm crystalline nanowires, with gate-all-around or independent gates (ϕ -Flash), suitable for full 3D integration," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2009, pp. 27.6.1–27.6.4.
- [25] J.-G. Yun, G. Kim, J.-E. Lee, Y. Kim, W. B. Shim, J.-H. Lee, H. Shin, J. D. Lee, and B.-G. Park, "Single-crystalline Si STacked ARray (STAR)

- NAND flash memory,” *IEEE Trans. Electron Devices*, vol. 58, no. 4, pp. 1006–1014, 2011.
- [26] Y. Kim, M. Kang, S. H. Park, and B.-G. Park, “Three-dimensional NAND flash memory based on single-crystalline channel stacked array,” *IEEE Electron Device Lett.*, vol. 34, no. 8, pp. 990–992, 2013.
- [27] Wandong Kim, Joo Yun Seo, Yoon Kim, Se Hwan Park, Sang Ho Lee, Myung Hyun Baek, Jong-Ho Lee, and Byung-Gook Park, “Channel-stacked NAND flash memory with layer selection by multi-level operation (LSM),” *Electron Devices Meet. IEDM 2013 IEEE Int.*, pp. 3.8.1–3.8.4, Dec. 2013.
- [28] H.-T. Lue, S.-H. Chen, Y.-H. Shih, K.-Y. Hsieh, and C.-Y. Lu, “Overview of 3D NAND Flash and progress of vertical gate (VG) architecture,” in *Solid-State and Integrated Circuit Technology (ICSICT), 2012 IEEE 11th International Conference on*, 2012, pp. 1–4.
- [29] Yi-Hsuan Hsiao, Hang-Ting Lue, Tzu-Hsuan Hsu, Kuang-Yeu Hsieh, and Chih-Yuan Lu, “A critical examination of 3D stackable NAND Flash memory architectures by simulation study of the scaling capability,” in *Memory Workshop (IMW), 2010 IEEE International*, 2010, pp. 1–4.
- [30] H.-W. You and W.-J. Cho, “Charge trapping properties of the HfO₂ layer with various thicknesses for charge trap flash memory applications,” *Appl. Phys. Lett.*, vol. 96, no. 9, p. 093506, 2010.
- [31] A. Padovani, A. Arreghini, L. Vandelli, L. Larcher, G. Van den bosch, P. Pavan, and J. Van Houdt, “A Comprehensive Understanding of the Erase of TANOS Memories Through Charge Separation Experiments and

- Simulations,” *Electron Devices IEEE Trans. On*, vol. 58, no. 9, pp. 3147–3155, Sep. 2011.
- [32] M. She, “Semiconductor Flash Memory Scaling,” Ph.D. Thesis, University of California, Berkeley, 2003.
- [33] T. Sugizaki, M. Kobayashi, H. Minakata, M. Yamaguchi, Y. Tamura, Y. Sugiyama, H. Tanaka, T. Nakanishi, and Y. Nara, “New 2-bit/Tr MONOS type flash memory using Al₂O₃ as charge trapping layer,” *Fujitsu Lab. Ltd Date Unkn.*, 2003.
- [34] H. Y. Yu, M. F. Li, B. J. Cho, C. C. Yeo, M. S. Joo, D.-L. Kwong, J. S. Pan, C. H. Ang, J. Z. Zheng, and S. Ramanathan, “Energy gap and band alignment for (HfO₂)_x(Al₂O₃)_{1-x} on (100) Si,” *Appl. Phys. Lett.*, vol. 81, no. 2, pp. 376–378, 2002.
- [35] E. Gnani, S. Reggiani, A. Gnudi, G. Baccarani, J. Fu, N. Singh, G. Q. Lo, and D. L. Kwong, “Modeling of gate-all-around charge trapping SONOS memory cells,” *Solid-State Electron.*, vol. 54, no. 9, pp. 997–1002, 2010.
- [36] Sheng-Chih Lai, Hang-Ting Lue, Ming-Jui Yang, Jung-Yu Hsieh, Szu-Yu Wang, T.-B. Wu, G.-L. Luo, C.-H. Chien, Erh-Kun Lai, Kuang-Yeu Hsieh, Rich Liu, and Chih-Yuan Lu, “MA BE-SONOS: A Bandgap Engineered SONOS using Metal Gate and Al₂O₃ Blocking Layer to Overcome Erase Saturation,” in *Non-Volatile Semiconductor Memory Workshop, 2007 22nd IEEE*, 2007, pp. 88–89.
- [37] R. People and J. C. Bean, “Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures,” *Appl. Phys. Lett.*, vol. 47, no. 3, pp. 322–324, 1985.

- [38] S. Borel, C. Arvet, J. Bilde, S. Harrison, and D. Louis, "Isotropic etching of SiGe alloys with high selectivity to similar materials," *Micro Nano Eng.* 2003, vol. 73–74, pp. 301–305, 2004.
- [39] F. Johnson, D. Miles, D. Grider, and J. Wortman, "Selective chemical etching of polycrystalline SiGe alloys with respect to Si and SiO₂," *J. Electron. Mater.*, vol. 21, no. 8, pp. 805–810, Aug. 1992.
- [40] W. Zhu, "HfO₂ and HfAlO for CMOS: Thermal Stability and Current Transport," *IEDM Tech Dig 2001*, 2001.
- [41] 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," *Electron Devices IEEE Trans. On*, vol. 51, no. 7, pp. 1143–1147, 2004.

초 록

플로팅 게이트를 기반으로 하는 2 차원 낸드 플래시 메모리는 소자 크기 축소화를 통한 집적도 향상이 한계에 부딪히게 되었다. 이러한 문제를 해결하기 위하여, 3 차원으로 적층된 낸드 플래시 메모리가 그 대안으로 떠오르고 있다. 적층형 낸드 플래시 메모리의 면적을 축소함에 있어서 전자 저장층의 두께가 걸림돌이 되고 있다. 가장 널리 쓰이는 질화막은 두께가 감소함에 따라 전자 저장 능력이 저하되기 때문에 두께를 줄이는 데 어려움이 있다. 이에 본 논문에서는 고유전체를 전자 저장층으로 사용하여 채널 적층 어레이를 기반으로 하는 3 차원 낸드 플래시의 집적도 향상을 위한 연구를 수행하였다.

1, 2 장에서는 최근 비휘발성 메모리의 연구 동향과 3 차원 적층형 낸드

플래시 메모리의 기술 발전을 소개하였다. 다음으로 고유전체의 메모리 능력을 평가하기 위하여 캐패시터 및 단층 플래시 메모리 소자를 제작하여 측정결과를 제시하였다. 4 장에서는 적층된 실리콘 나노와이어에 고유전체 메모리 절연막을 적용하여 메모리 어레이를 성공적으로 제작하여 그 결과를 보여주었다. 기존 질화막 기반으로 제작하였던 소자와의 성능을 비교함으로써 본 연구에서 제작한 소자를 평가하였다. 5 장에서는 채널적층형 메모리에서의 프로그램 속도를 향상 시킬 수 있는 새로운 동작 방법을 제시하였으며, 3D 시뮬레이션과 측정을 통해 검증한 결과를 제시하였다.

주요어: 3 차원 낸드 플래시 메모리, 적층형 어레이 (STAR), 전하 트랩형 메모리, 고유전체 전하 저장형 메모리

학 번: 2012-30938

List of Publications

Journal

- [1] Dae Woong Kwon, Wandong Kim, Do-Bin Kim, Sang-Ho Lee, **Joo Yun Seo**, Myung-Hyun Baek, Ji-Ho Park, Eunseok Choi, Gyu Seong Cho, and Byung-Gook Park, "Multi-Level threshold voltage setting method of String select transistors for layer selection in Channel stacked NAND flash memory," *IEEE Electron Device Letters*, Vol. 36, No. 12, pp. 1318-1320, Dec. 2015
- [2] Yoon Kim, **Joo Yun Seo**, Sang-Ho Lee, and Byung-Gook Park, "A new programming method to alleviate the program speed variation in three-dimensional stacked array NAND flash memory," *Journal of Semiconductor Technology and Science*, Vol. 14, No. 5, pp. 566-571, Oct. 2014
- [3] **Joo Yun Seo**, Yoon kim, and Byung-Gook Park, "New program inhibition scheme for high boosting efficiency in three-dimensional NAND array," *Japanese Journal of Applied Physics: Rapid Communications*, Vol. 53, No. 7, pp. 070304-1-070304-3, Jun. 2014
- [4] **Joo Yun Seo**, Sang-Ho Lee, Se Hwan Park, Wandong Kim, Do-Bin Kim, and Byung-Gook Park, "Characteristics of gate-all-around polycrystalline silicon channel SONOS flash memory," *International Journal of*

Nanotechnology, Vol. 11, No. 12, pp. 116-125, Mar. 2014

- [5] Se Hwan Park, Yoon Kim, Wandong Kim, **Joo Yun Seo**, and Byung-Gook Park, "Vertical-Channel STacked ARray (VCSTAR) for 3D NAND flash memory," *Solid-State Electronics*, Vol. 78, No. , pp. 34-38, Dec. 2012

Conference

- [1] **Joo Yun Seo**, Sang-Ho Lee, Myung-Hyun Baek, and Byung-Gook Park, "Overview of channel-stacked NAND flash memory," *International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 237-238, Jun. 2015
- [2] **Joo Yun Seo**, Dae Woong Kwon, Sang-Ho Lee, and Byung-Gook Park, "Investigation of the effect of gate space length variation on the retention characteristics in NAND flash array," *International Conference on Electronics, Information and Communication(ICEIC)*, pp. 234-235, Jan. 2015
- [3] **Joo Yun Seo**, Yoon Kim, Do-Bin Kim, and Byung-Gook Park, "A new programming method to alleviate the program speed variation for three-dimensional channel stacked array architecture," *Korean Conference on Semiconductors*, pp. 80-80, Feb. 2014
- [4] Sang-Ho Lee, **Joo Yun Seo**, Wandong Kim, Se Hwan Park, Do-Bin Kim,

and Byung-Gook Park, "Investigation of SANOS memory device for suppressing gate injection current," *International Conference on Electronics, Information and Communication(ICEIC)*, pp. 489-490, Jan. 2014

[5] Wandong Kim, **Joo Yun Seo**, Yoon Kim, Se Hwan Park, Sang-Ho Lee, Myung Hyun Baek, Jong-Ho Lee, and Byung-Gook Park, "Channel-Stacked NAND flash memory with layer selection by multi-level operation (LSM)," *IEEE International Electron Devices Meeting*, pp. 3.8.1-3.8.4, Dec. 2013

[6] Myung Hyun Baek, Se Hwan Park, **Joo Yun Seo**, and Byung-Gook Park, "포획된 전하량 분석을 통한 planar 구조와 gate-all-around (GAA) 구조의 SONOS 낸드 플래시 메모리특성 비교," *대한전자공학회 추계학술대회*, pp. 90-93, Nov. 2013

[7] Sang-Ho Lee, **Joo Yun Seo**, Wandong Kim, Se Hwan Park, Do-Bin Kim, and Byung-Gook Park, "Investigation of programmed state threshold voltage variation caused by unequal cell size in gate-all-around NAND flash memory," *International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 330-331, Jul. 2013

[8] Wandong Kim, Yoon Kim, Se Hwan Park, **Joo Yun Seo**, Do-Bin Kim, Sang-Ho Lee, and Byung-Gook Park, "Channel selection by multi level operation of string select line in 3D channel stacked NAND flash memory array," *한국반도체학술대회*, pp. WE3-K-3-, Feb. 2013

- [9] **Joo Yun Seo**, Sang-Ho Lee, Yoon Kim, Se Hwan Park, Wandong Kim, Do-Bin Kim, and Byung-Gook Park, "A study on gate-all-around (GAA) polycrystalline silicon channel SONOS flash memory," *IEEE International NanoElectronics Conference*, pp. 69-71, Jan. 2013
- [10] Hyun Woo Kim, Min-Chul Sun, Sang Wan Kim, **Joo Yun Seo**, Garam Kim, Jang Hyun Kim, and Byung-Gook Park, "Investigation on effects of changing body doping concentration in short-channel junctionless transistor," *International Microprocesses and Nanotechnology Conference*, pp. 1P-7-41-, Nov. 2012
- [11] **Joo Yun Seo**, Yoon Kim, Se Hwan Park, Wandong Kim, Do-Bin Kim, Jong-Ho Lee, Hyungcheol Shin, and Byung-Gook Park, "Investigation into the effect of the variation of gate dimensions on program characteristics in 3D NAND flash array," *Silicon Nanoelectronics Workshop*, pp. 19-20, Jun. 2012

Patent

- [1] 박병국, 서주연, 김필중, 손선익, “비트라인의 커패시턴스 차이를 줄이기 위한 3 차원 채널 적층형 낸드 플래시 메모리 어레이”,
 Korean Patent filed 10-2014-0039277, April 2, 2014

Korean Patent No. 10-1528806, June 9, 2015

[2] 박병국, 김윤, 서주연 “수직 적층된 SSL 을 갖는 스타구조 낸드 플래시 메모리 어레이 및 그 제조방법,”

- Korean Patent filed 10-2011-0106525, October 18, 2011
- **Korean Patent No. 10-1263313, May 6, 2013**