Vertical-SI-nanowire based nonvolatile flash memory for 3-D ultrahigh density application

Sun, Yuan

2011

Sun, Y. (2011). Vertical-SI-nanowire based nonvolatile flash memory for 3-D ultrahigh density application. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50781>

<https://doi.org/10.32657/10356/50781>

**VERTICAL-SI-NANOWIRE BASED
NONVOLATILE FLASH MEMORY FOR 3-D
ULTRAHIGH DENSITY APPLICATION**

SUN YUAN

(B. Eng., Nanyang Technological University)

**A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
DEPARTMENT OF ELECTRICAL & ELECTRONIC ENGINEERING
NANYANG TECHNOLOGICAL UNIVERSITY**

2011

Acknowledgements

First and foremost I offer my most sincere gratitude to my supervisors. Prof. Yu Hong Yu, my main supervisor in the Division of Microelectronics, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (NTU), is the most important person to my research work. I would like to thank him for bringing me into this colorful research world and also giving me inspiration for many ideas. I would also like to thank Dr. Navab Singh, my co-supervisor in the Institute of Microelectronics (IME), a reputable A*STAR research institute, for his guidance and assistance that allowed me to adapt to a new working environment and giving me many opportunities and freedom to pursue my interests. My gratitude is devoted to Dr. Lo Guo Qiang for his insightful comments and discussions on my research work in IME. I am also grateful to Dr. Lap Chan and Mr. Leong Kam Chew, my mentors from GLOBALFOUNDRIES Singapore Pte. Ltd., for giving me the scholarship and helping me in so many ways in completing my research work.

During my PhD life I have been working with a friendly and cheerful group of researchers, seniors and fellow students in IME. I thank Dr. Yu Mingbin, Dr. Li Xiang, Dr. Wang Xinpeng, Dr. Jiang Yu, Ms. Fu Jia, Mr. Shen Nansheng, Mr. Chen Zhixian for giving me so much help, support and encouragement, in many ways I have learnt much from them. I'm thankful to Mr. Lu Weijie, Mr. Fang Zheng, Mr. Li Yida, Mr. Zhang Ling and Mr. Peng Lan for being such companionable classmates and faithful friends. I will value their friendship for lifetime. I would also like to take this opportunity to thank the technical staffs in NTU and IME for their great support and assistance in my PhD study.

Finally, my deepest gratitude goes to my dear parents and my fiancée Ms. Dang Qi whose love makes me go on.

Table of Contents

Acknowledgements	I
Table of Contents	II
Summary	V
List of Tables	VII
List of Figures	VIII
List of Symbols	XV
List of Abbreviations	XVI

Chapter 1 Introduction

1.1 Categories of Semiconductor Memory	1
1.2 Challenges for Flash Scaling in 2-Dimensional (2-D).....	3
1.3 Scope of Project.....	7
1.4 Organization of Thesis.....	8
Reference	10

Chapter 2 Literature Review

2.1 Introduction.....	11
2.2 Flash Memory with Multi-gated Cell Structures.....	11

2.3 3-Dimensional (3-D) Stacked NAND Array.....	14
2.4 Vertical Si Nanowire (SiNW) Transistor Technology	21
Reference	25

Chapter 3 High Performance Junction-based Vertical SiNW SONOS

3.1 Introduction.....	28
3.2 Junction-based Vertical SiNW SONOS Device Fabrication	30
3.3 Results and Discussion	37
3.3.1 Junction-based Vertical SiNW SONOS.....	37
3.3.2 Vertical SiNW Memory with Si-NCs Trap Layer.....	48
3.4 Conclusion	51
Reference	52

Chapter 4 Junction-Less (JL) Stackable SONOS Memory Realized on a Vertical SiNW

4.1 Introduction.....	54
4.2 JL Vertical SiNW SONOS Device Fabrication.....	55
4.3 Results and Discussion	59
4.3.1 JL Vertical SiNW SONOS using F-N tunneling.....	59
4.3.2 JL Vertical SiNW SONOS with 2-bit Storage/Cell.....	68

4.4 Conclusion	75
Reference	76

Chapter 5 Poly-Si Vertical Nanowire (NW) for Thin Film Transistor (TFT) SONOS Memory Application

5.1 Introduction.....	78
5.2 Vertical Poly-Si NW TFT and SONOS Device Fabrication	79
5.3 Results and Discussion	83
5.3.1 Vertical Poly-Si NW TFTs and Inverter.....	83
5.3.2 JL Vertical Poly-Si NW TFT SONOS.....	88
5.4 Conclusion	92
Reference	93

Chapter 6 Demonstration of Memory String with Stacked JL-SONOS

6.1 Introduction.....	94
6.2 Stacked Vertical SiNW SONOS Device Fabrication	95
6.3 Results and Discussion	102
6.4 Conclusion	115
Reference	116

Chapter 7 Conclusions.....

List of Publications

Executive Summary

Nonvolatile memory (NVM) technology is going through a fast evolution amongst the semiconductor technologies in the last decade. To satisfy the increasing demand of flash memory, the memory density has been continuously increased through aggressive scaling of the device dimensions in 2-dimensional (2-D) plane. However, the cell size shrinking in 2-D is getting more and more difficult beyond sub-20-nm technology node due to concerns on lithography, coupling ratio, and crosstalk interference.

In order to overcome the obstacles of flash memory scaling in 2-D planar platform, a large amount of research work has been conducted so far which can be mainly classified into two categories: adoption of novel cell structures and exploration of new materials. Among various cell structures, the Gate-All-Around (GAA) cylindrical nanowire (NW)- field effect transistor (FET) is being considered as one of the strong potential candidates to advance the NVM technology to the extreme miniaturization limits, due to the optimal electrostatic control which translates into nearly ideal subthreshold turn-on and a negligibly small short channel effect (SCE).

This thesis proposes methodologies to resolve issues of memory cell scaling, program/erase (P/E) speed and reliability in the Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) type discrete charge storage NVM based on a novel CMOS compatible single-crystalline vertical silicon nanowire (SiNW) platform.

A novel vertical-SiNW-based GAA SONOS type NVM cell is demonstrated with high P/E speed, good reliability and multi-bit programming capability. The scope of this thesis covers aspects of vertical SONOS from device fabrication, memory

performance characterization, reliability studies and discussion on its applications to build 3-D multi-level stacked high density memory. The underlying physics for memory performance enhancement is discussed and analyzed through simulation. The discussion on vertical SiNW SONOS in this thesis is divided into the following 4 parts: 1) the development of the junction-based vertical SiNW SONOS platform and device performance optimization through Si-nanocrystals (Si-NCs) embedment; 2) the development of stackable junction-less (JL) vertical SiNW SONOS device and discussions on its application for physical multi-bit storage; 3) the demonstration of vertical poly-Si NW thin-film transistors and SONOS memory; 4) demonstration of memory array based on 2-level vertically stacked SiNW SONOS memory.

List of Tables

Table-1.1: Number of electrons required per logic level for an overall V_{th} shift of $3V$.

Table.4.1: Operation conditions of a junction-less vertical-SiNW-based SONOS memory with 2-bit storage.

Table 6.1 Operation bias conditions for stacked NAND memory array based on vertical SiNW SONOS cells.

List of Figures

Fig. 1.1: Categories of semiconductor memory.

Fig. 1.2: Schematics of a 2x2 NAND cell array (a) layout (b) cross-sectional view in the word line (WL) direction formed by cutting plane line AA' and (c) bit line (BL) direction formed by cutting plane line BB'.

Fig. 2.1: 5nm radius twin SiNW GAA MOSFET with V_{th} roll-off, SS, and DIBL remained same along the gate length.

Fig. 2.2: Concept of S-SGT structured cell: (a) left: bird's view of S-SGT structured cell; center: cross-sectional view along (B1)–(B2) of S-SGT structured cell; right: equivalent circuit of S-SGT structured cell and NAND structured cell; (b) ratio of occupied area per bit as function of number of stacked memory cells in one silicon pillar in comparison with the smallest reported cell area per bit of conventional NAND structured cell.

Fig. 2.3: Vertical SEM photographs of the fabricated 3-D stacked NAND cell string, the second active layer is SOI-like perfect single crystal.

Fig. 2.4: (a) Birds-eye view of BiCs flash memory, (b) top down view of BiCs flash memory array.

Fig. 2.5: Bit cost scalability of three-dimensional flash memory.

Fig. 2.6: Schematic views of VSAT and PIPE. It has planar type SLs (GSL and SSL) in the peripheral region, not embedded in the vertical string.

Fig. 2.7: Equivalent circuit of VSAT device with three vertical strings, SLs are separated from vertical strings like a conventional NAND device.

Fig. 2.8: Vertical interconnection (a) is simpler and more cost-effective than stair-like (b) method.

Fig. 2.9: (a) Structure of vertical gate (VG)-NAND, (b) top down view of VG-NAND showing $4F^2$ cell size / layer.

Fig. 2.10: SEM pictures of device fabrication: (a) Vertical NW with diameter of 20 nm. (b) After gate patterning by lithography but before exposing the drain (the tip of the pillar) of the transistor. (c) After the drain (pillar tip) is exposed. Metal contacts are made with such device structure. (d) Vertical NW arrays with pitch of 500 nm. Nanowires are 1 μm tall with a diameter of 20 nm.

Fig. 2.11: XTEM images of a vertical-SiNW memory with $L_g=150$ nm and a SONOS gate stack (with the thickness of $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2=5/5/6$ nm), with the SEM image of a single nanowire as inset (diameter = 50 nm).

Fig. 3.1: Schematic of cross-section of vertical-SiNW-based SONOS memory.

Fig. 3.2: DR-SEM images of vertical SiNW formation using top down approach: (a) SiN hard mask patterning, (b) Si pillar etch, (c) pillar oxidation, (d) after DHF etch.

Fig. 3.3: Planar TEM images of vertical SiNW SONOS with (a) $\Phi=50\text{nm}$ and (b) $\Phi=20\text{nm}$.

Fig. 3.4: Schematics illustrating the process flow used to fabricate vertical GAA SiNW based SONOS.

Fig. 3.5: SEM images for the device with a NW diameter of $\sim 50\text{nm}$ (a) after 1st HDP deposition and etch back; (b) after gate stack deposition and patterning; (c) after 2nd HDP deposition and etch back; (d) after tip poly removal and drain implant; (e) after metallization and metal etch.

Fig. 3.6: (a) Schematic diagrams show the cross section part of nanowire channel SONOS (left) and SONcOS (right) devices. (b) Cross-sectional TEM image of vertical SiNW GAA SONOS: diameter of wire is 50 nm and gate length 150nm. (c) AFM image scanned on $1 \times 1 \mu\text{m}^2$ surface area that reveals the formation of Si-NC with a density of $7.5 \times 10^{10}/\text{cm}^2$.

Fig. 3.7: Depletion-type vertical SiNW (with a diameter of 50nm) SONOS memory ($L_g=150\text{nm}$): (a) Programming characteristics at gate bias of 12V to 18V for $1\mu\text{s}$ to 1s. (b) Erasing characteristics at gate bias of -18V to -12V with a drain bias of 5V for $1\mu\text{s}$ to 1s.

Fig. 3.8: Depletion-type vertical SiNW (with a diameter of 50nm) SONOS memory ($L_g=150\text{nm}$): (a) Retention characteristic at 85°C . (b) Endurance characteristic.

Fig. 3.9: (a) Programming characteristic of vertical SiNW SONOS memory ($L_g \sim 150$ nm and diameter = 20 nm). (b) Erasing characteristic of vertical SiNW SONOS memory ($L_g \sim 150$ nm and diameter = 20 nm).

Fig. 3.10: Comparison of the P/E characteristics of vertical SiNW SONOS memory with different diameters (50 nm versus 20 nm).

Fig. 3.11: Simulation of electrical field distribution in ONO layer after programming at gate bias of 15V for 100ms.

Fig. 3.12: Vertical SiNW (with dimensions of 20 and 50 nm) SONOS memory ($L_g \sim 150$ nm). (a) Retention characteristic at 85 °C. (b) Endurance characteristic up to 104 cycles at 25 °C.

Fig. 3.13: Vertical SiNW (with dimension of 20 nm) SONOS memory ($L_g \sim 150$ nm). (a) Multi bit programming with gate biases of 12, 14, and 16 V for 1 ms. During the first read operation, “00” and “01” states can be distinguished from “10” and “11” states at 4 V. After that, “00” from “01” and “10” from “11” at 2.4 and 5.8 V, respectively, are performed for the second read operation. (b) Cumulative distribution of multi-bit programmed V_{th} with gate biases of 12, 14, and 16 V for 1 ms for randomly selected ten devices.

Fig. 3.14: Transient memory characteristics of SiNW SONcOS device and SONOS control device at +15/-16 V pulse stresses. Devices with a gate length of L_g 120nm and NW diameter 50nm are used for characterization.

Fig. 3.15: Retention characteristic at 85°C of vertical SiNW SONOS and SONcOS memory ($L_g \sim 120$ nm, and diameter=50nm).

Fig. 3.16: P/E endurance characteristic at 85°C of vertical SiNW SONOS and SONcOS memory ($L_g \sim 120$ nm, and diameter=50nm).

Fig. 4.1: (a) Process integration flow of vertical SiNW junction based SONOS memory: (a-1) vertical SiNW formation; (a-2) As+ is implanted to form the source; (a-3) gate stack deposition and gate pad definition; (a-4) tip poly and ONO removal, followed by drain implant; (a-5) metallization. (b) Process integration flow of vertical JL-SONOS: (b-1) bulk implant and annealing; (b-2) vertical SiNW formation; (b-3) gate stack deposition and gate pad definition; (b-4) tip poly and ONO removal; (b-5) metallization.

Fig. 4.2: (a) Cross-sectional TEM image of vertical SiNW GAA JL-SONOS: diameter of wire is 20 nm and gate length 120nm. (b) Cross-sectional TEM image of gate stack showing ONO thickness 5/7/7 nm.

Fig. 4.3: Vertical SiNW JL-SONOS memory ($L_g \sim 120$ nm, channel doping = $1 \times 10^{19}/\text{cm}^{-3}$): I_d - V_g characteristics for devices with different wire diameters from 20 nm to 60 nm (a) after forming gas anneal (415 °C, 45 mins, 10% H₂) and (b) before forming gas anneal.

Fig. 4.4: Typical I_d - V_g curves of JL-SONOS with 20nm SiNW and channel doping of $1 \times 10^{17} \text{cm}^{-3}$ during (a) programming and (b) erasing at ± 16 V with pulse widths from 100nm to 1ms.

Fig. 4.5: (a) Programming and (b) erasing characteristics using F-N tunneling of vertical-SiNW-based GAA JL-SONOS memory device with channel doping of $1 \times 10^{17} \text{cm}^{-3}$ and $1 \times 10^{19} \text{cm}^{-3}$, initial program, 16V/1ms for lightly doped ($10^{17}/\text{cm}^{-3}$) and 15V/1ms for moderately doped ($10^{19}/\text{cm}^{-3}$) devices, was done on fresh devices to get the same programmed V_{th} before erasing. (c) Simulated electric field distribution of the SiNW-based JL-SONOS at $V_g = 4$ V with various channel doping of 10^{17}cm^{-3} / 10^{19}cm^{-3} and a wire diameter of 20nm.

Fig. 4.6: Vertical SiNW (with dimension of 20 nm) JL-SONOS memory ($L_g \sim 120$ nm, channel doping= $1 \times 10^{19}/\text{cm}^{-3}$): multi-bit programming with gate biases of 12, 14, and 16 V for 1 ms.

Fig. 4.7: (a) Retention and (b) endurance characteristic at 85°C of JL-SONOS memory devices with channel doping of $1 \times 10^{17} \text{cm}^{-3}$ and $1 \times 10^{19} \text{cm}^{-3}$ ($L_g \sim 120$ nm, and diameter=20nm).

Fig. 4.8: Schematic of the vertical SiNW GAA JL-SONOS, the charges are separately stored in the nitride layer above the top (T-bit)/bottom (B-bit) portion of the vertical-wire channel.

Fig. 4.9: (a) Erase speed of top bit (T-bit) and bottom bit (B-bit), (b) program speed of top and bottom bits.

Fig. 4.10: I_d - V_g characteristic memory cell in different V_{th} states, SBE indicates the amount of second bit effect.

Fig. 4.11: 2-bit programming characteristics of JL-SONOS Memory.

Fig. 4.12: (a) Retention characteristic: a virgin cell and 10K cycled cell, solid: fresh cell, open: 10K cycled cell. (b) Endurance characteristic: with I-V curves of a programmed cell at various cycles as the inset. Cycle sequences: erase both bits->program B-bit ->program T-bit.

Fig. 5.1: (a) SEM image after pillar trimming, (b) SEM image after gate patterning, (c) SEM image of NFET device, (d) SEM image of an Inverter, (e) TEM image of NMOS device with wire diameter ~90 nm and gate length ~100 nm.

Fig. 5.2: TEM image of junction-less TFT SONOS device with wire diameter ~90 nm and gate length ~100 nm.

Fig. 5.3: (a) $I_{DS}-V_{GS}$ characteristic of individual NMOS and PMOS of the inverter (diameter ~30 nm, $L_g \sim 100$ nm), (b) $I_{DS}-V_{DS}$ characteristic of individual NMOS & PMOS of the inverter (diameter ~30 nm, $L_g \sim 100$ nm), (c) impact of SiNW diameter on device SS and DIBL ($L_g \sim 100$ nm). In total, about 20 devices were measured, and the error bar is estimated by the spread of the data.

Fig. 5.4: (a) Transfer characteristics of the inverter at different V_{DD} with V_{IN} and V_{OUT} plotted interchangeably on X and Y axes, (b) inverter switching current at different V_{DD} values, and (c) transfer characteristic of the inverter at different V_{DD} (ranging from 0.2 V to 1.2 V).

Fig. 5.5: Transfer characteristics of junction-less vertical nanowire TFT SONOS device with 30nm wire channel and 100nm gate length. The inset shows the dependence of SS on the poly-Si nanowire diameter.

Fig. 5.6: The P/E characteristics of nanowire TFT SONOS with wire diameter of 30nm. The gate pulse stress was ranging from 13 V to 16 V and -13 V to -16 V. The inset of (a): comparison of memory window among devices with different nanowire diameters under different stress voltage for a program time of 1 ms

Fig. 5.7: Retention behavior of fresh GAA vertical poly-Si nanowire TFT SONOS memory at room temperature.

Fig. 6.1: Process integration flow for NAND string of eight junction-less SONOS memory cells (M1-M8) on four vertical SiNWs: (a) Bulk implant and annealing; (b) vertical SiNW formation and HDP isolation; (c) 1st level gate stack deposition; (d) 1st level gate pad definition; (e) tip poly removal; (f)

HDP oxide isolation; (g) 2nd level gate pad definition and tip poly-ONO removal; (h) metallization. M1 to M8 are 8 cells on 4 wires.

Fig. 6.2: SEM images of 2-level stacked vertical SiNW SONOS memory with four vertical pillars. Bird's eye view of memory string after vertical pillars with bottom covered by HDP (6.2a); 1st level gate stack deposition (6.2b); 1st level gate pad definition (6.2c); 1st level tip poly removal (6.2d); HDP oxide isolation (6.2e); 2nd level gate pad definition and tip poly-ONO removal (6.2f); metallization. WL1 to WL8 are for 8 cells on 4 wires (6.2g).

Fig. 6.3: Equivalent circuit of (a) conventional and (b) vertical SiNW stacked SONOS memory string with eight cells. SLs can be formed on peripheral region.

Fig. 6.4: Schematic illustration of the footprint of vertical SiNW stacked SONOS device and cell area in NAND array.

Fig. 6.5: Cross-sectional TEM image of 2-level stacked vertical SiNW GAA JL-SONOS: the wire diameter is 120 nm and the gate length is ~100nm, the inset shows ONO thicknesses 5/7/7 nm.

Fig. 6.6: I_d - V_g plot of vertical JL-SONOS with channel diameter=30nm.

Fig. 6.7: (a) Simulated electric field distribution of the SiNW based JL SONO at $V_g=4$ V with various wire diameter of 20nm / 40nm / 50nm / 110nm and a fixed channel doping of 10^{19} cm^{-3} , (b) zoomed electric field near the channel/tunnel oxide interface.

Fig. 6.8: (a) Programming and (b) erasing using F-N tunnelling of stacked SiNW based GAA JL-SONOS memory device with channel doping of $1 \times 10^{19} \text{ cm}^{-3}$ and wire diameter of 30nm. Open: bottom level cells, solid: top level cells.

Fig. 6.9: (a) Multibit programming with gate biases of 13, 15, and 18 V for 100 ms, the memory is able to store 2 bits per cell using four states "00", "01", "10" and "11". (b) Cumulative distribution of multibit programmed V_{th} with gate biases of 13, 15, and 18 V for 100 ms for randomly selected twenty devices.

Fig. 6.10: Room temperature retention characteristic: a virgin cell and 10K cycled cell, solid: fresh cell, open: 10K cycled cell.

Fig. 6.11: Retention at elevated temperatures, extrapolated charge loss is 15% at 125 °C after 10 years.

Fig. 6.12: (a) Endurance characteristic up to 10^5 cycles. (b) I_d - V_g curves of the cell at various cycles.

Fig. 6.13: Schematic illustration of the bias conditions during programming the selected cell. Both V_{pass} disturb and V_{pgm} disturb need to be taken into account.

Fig. 6.14: (a) V_{th} shift after V_{pass} disturb on unselected cells. (b) Disturb free window is a function of BL voltage $V_{INHIBIT}$. (c) V_{pgm} disturb window using ISPP and raised-BL method, the selected device can be programmed to $V_{th} > 4$ V, while unselected devices remain in the erased state.

List of Symbols

C	capacitance (F)
d	thickness
Φ	diameter
E	electrical field (V/cm)
E_g	band gap (eV)
h	Planck's constant (6.626×10^{-34} Js)
I	current (A)
I_d	drain current (A)
J	current density (A/cm ²)
L	channel length (μm)
m^*	effective mass (kg)
Q	charge (C)
Q_r	charges stored in the gate insulator at a distance d_i from the gate (C)
T	temperature ($^{\circ}\text{C}$)
t	time (sec)
t_{ox}	thickness of oxide
V	voltage (V)
V_d	drain voltage (V)
V_g	gate voltage (V)
V_{th}	threshold voltage (V)
ΔE_c	conduction band offset (eV)
ΔE_v	valence band offset (eV)
ΔV_{th}	threshold voltage shift (V)

List of Abbreviations

BL	Bit Line
BTBHHI	Band-To-Band Hot Hole Injection
BiCs	Bit Cost Scalable
BOX	Buried Oxide
CG	Control Gate
CHE	Channel Hot Electron
CT	Charge Trapping
CVD	Chemical Vapor Deposition
DIBL	Drain-Induced-Barrier-Lowering
DG	Double Gate
DRAM	Dynamic Random Access Memory
DRIE	Deep Reactive Ion Etching
DT	Direct Tunneling
EOT	Equivalent Oxide Thickness
FD	Fully depleted
FeRAM	Ferro-electric Random Access Memory
FG	Floating Gate
FinFET	Fin Field Effect Transistor
F-N	Fowler-Nordheim
GAA	Gate-All-Around
GCR	Gate Coupling Ratio
HRTEM	High Resolution Transmission Electron Microscopy

HDP	High Density Plasma
IPD	Interpoly Dielectric
ITRS	International Technology Roadmap for Semiconductors
JL	Junction-less
LPCVD	Low Pressure Chemical Vapor Deposition
MLC	Multi-level Cell
MRAM	Magnetic Random Access Memory
NC	Nanocrystal
NVM	Nonvolatile Memory
PDA	Post Deposition Anneal
PECVD	Plasma-Enhanced Chemical Vapor Deposition
P/E	Program/Erase
Poly-Si	Polycrystalline Silicon
PVD	Physical Vapor Deposition
RAM	Random Access Memory
ROM	Read Only Memories
S/D	Source and Drain
SCE	Short Channel Effect
SEM	Scanning Electron Microscopy
SL	Select Line
SLC	Single-level Cell
SiNW	Silicon nanowire
SOI	Silicon-On-Insulator
SONOS	Si / SiO ₂ / Si ₃ N ₄ / SiO ₂ / Si
SRAM	Static Random Access Memory

SS	Subthreshold Swing
TEM	Transmission Electron Microscopy
VLSI	Very Large-Scale Integration
VSAT	Vertical Stacked Array Transistor
WL	Word Line
2-D	2 Dimensional
3-D	3 Dimensional

Chapter 1

Introduction

1.1 Categories of Semiconductor Memory

Semiconductor memories are electronic data storage devices based on a metal-oxide-semiconductor (MOS) technology. There are various types of semiconductor memories which can be mainly divided into two basic categories: volatile memory that requires power to retain the data content; and nonvolatile memory (NVM), which is able to retain the stored information even when not powered.

The most common volatile memories include dynamic random access memory (DRAM) and static random access memory (SRAM). DRAM has a simple one transistor and a capacitor (1T1C) structure to store one bit. The capacitor can be either charged or discharged which can be considered as two states to represent the two values of a bit either 0 and 1. The simple 1T1C structure enables DRAM to be packed in very high density and low cost-per-bit can be achieved. However, since capacitors leak charge, the information eventually fades and periodical refresh is needed to retain the stored data [1.1]. Because of this refresh requirement, it is a dynamic memory as opposed to SRAM and other static memories. A typical SRAM contains six transistors (6T) to store a memory bit. Each bit is stored on four transistors, while two additional transistors control the access during read and write operations. SRAM is able to retain the information as long as the power is on and no periodical refresh is required. Due to its fast access speed, SRAM is used as a cache

memory in personal computers while DRAM is used for the main memory due to its much lower cost-per-bit.

Fig. 1.1: Categories of semiconductor memory.

On the other hand, in the past decade, memory chips with low power consumption and low cost have drawn more and more attention due to the booming market of portable electronic devices such as mobile phones and digital cameras that require data retention for ten years' time. The booming demand of these portable electronic devices has made NVM a crucial component of overall semiconductor memory market. There are mainly four types of NVM technology: flash memory, Ferro-electric Random Access Memory (FeRAM) [1.2], Magnetic Random Access Memory (MRAM) [1.3], Resistive Random Access Memory (RRAM) [1.4] and phase change memory [1.5].

Flash memory is currently dominating the NVM market mainly due to that it

provides a balance between bit cost and memory performance. The first application of flash memory is using NOR type flash, which provides fast memory read speed and random access to any location, to store program code in cellular phones. Another application is the usage of NAND type flash as data storage medium in devices such as USB memory cards, MP3 music players and PDA's, which plays a significant role in lowering the costs of such devices. As a result, the flash market, especially the NAND type flash, had grown exponentially in the past decade [1.6].

1.2 Challenges for flash scaling in 2-Dimensional (2-D) Device Architecture

The demand of NVM memory is booming due to the increasing demand of portable electronic devices such as mobile phones, tablet PCs and personal media players. To satisfy the increasing demand of flash memory, the memory density has been continuously increased through aggressive scaling of the device dimensions in 2-dimensional or planar style [1.6]. However, the main scaling limits for conventional floating-gate based NAND devices come from the following three perspectives:

- Geometric Limitations: Gate coupling rate (GCR) and oxide thickness,
- Wordline-to-wordline (WL-to-WL) interference for scaling beyond 20nm technology node,
- Running out of number of electrons stored in the cell to represent the program state.

The GCR is the ratio of the floating-gate (FG) capacitance to control-gate capacitance and the overall floating-gate capacitance, as expressed in Eq. (1.1) and (1.2).

$$GCR = \frac{C_{\text{FloatingGate-ControlGateCapacitance}}}{C_{\text{FloatingGate-ControlGateCapacitance}} + C_{\text{FloatingGate-SubstrateCapacitance}}} > 0.6, \quad (1.1)$$

$$V_{\text{Floating-Gate}} = V_{\text{Control-Gate}} \times GCR, \quad (1.2)$$

The ratio must be larger than 0.6 to assure that most of the gate voltage drop would be across the tunnel oxide (see Eq. (1.2)), but not across the inter-poly dielectric (IPD). The floating-gate device is non-planar, so that FG must be high enough to increase $C_{\text{FloatingGate-ControlGateCapacitance}}$ for good GCR. However, beyond 20nm node, there is no space left for the control gate after IPD filling, as shown in Fig. 1.2(a).

Fig. 1.2: Schematics of a 2x2 NAND cell array (a) layout (b) cross-sectional view in the word line (WL) direction formed by cutting plane line AA' and (c) bit line (BL) direction formed by cutting plane line BB'.

Other geometric limitations include the tunnel oxide and IPD thickness. According to the International Technology Roadmap for Semiconductors (ITRS), the tunnel oxide must be thicker than 8nm to keep charges in FG whereas the IPD must be thicker than 12nm. Otherwise, the charges will leakage through by direct tunneling (DT). The DT current model can be expressed as Eq. (1.3):

$$J_{DT} = C \left(\frac{V_{OX}}{T_{OX}} \right)^2 e^{-B \left[1 - \left(\frac{V_{OX}}{\Phi_B} \right)^{\frac{3}{2}} \right] \frac{V_{OX}}{T_{OX}}}, \quad (1.3)$$

where V_{OX} is the voltage drop across the tunneling oxide, T_{OX} is the tunnel oxide thickness, B and C are physical parameters.

On the other hand, as device dimensions and adjacent word-line space being scaled down, the WL-to-WL capacitance between neighboring cells also increases aggressively (see Eq. (1.4)) and thus the associated gate coupling effect during P/E will lead to severe device interference [1.8].

$$C_{WL} = \frac{\epsilon_0 \epsilon_r \times h \times l}{d_{WL}}, \quad (1.4)$$

where h is the word line thickness, l is the word line length, d_{WL} is the word line to word line space, ϵ_0 is the electric constant, ϵ_r is the relative static permittivity of the material between word lines.

The last scaling limit for FG is the small number of storage electrons available in presenting state of the program cell. For a typical 10nm node single level cell (SLC) FG device with GCR = 0.65, tunnel oxide = 8nm, block oxide = 13nm, the number of electrons for a 3V threshold voltage shift (V_{th} Shift) will be around 20 (see Eqs. (1.5) and (1.6)).

$$V_{Tshift} = \frac{Q_{FG}}{C_{Tox}} = \frac{N_e \times q}{\frac{\epsilon_{SiO2} \times W \times L}{T_{OX}}} = 3V, \quad (1.5)$$

$$N_e = \frac{V_{T,shift} \times \epsilon_{SiO2} \times W \times L}{T_{OX} \times q} \approx 20, \quad (1.6)$$

where Q_{FG} represents the charges trapped in the FG, C_{TOX} is the tunneling oxide capacitance, T_{OX} is the tunneling oxide thickness, N_e is the number of trapped electrons, W and L are the gate width and gate length.

For multi-level cell (MLC) operation at about 10nm node, less than ten electrons will distinguish a level. Table 1.1 shows the calculated number of electrons to distinguish a single level for various types of cells and technology nodes.

Node	45nm	32nm	22nm	16nm	11nm	8nm
SLC (2 levels)	400	200	100	50	25	12
MLC (4 levels)	135	67	34	17	8	4
TLC (8 levels)	58	29	15	7	4	2
QLC (16 levels)	32	16	8	4	2	1

Table-1.1: Number of Electrons Required per Logic Level for an overall V_{th} shift of 3V.

In summary, it is obvious that the cell size shrinking in 2-D becomes harder due to aggressive scaling of device dimension and eventually the memory cell will be running out of electrons quickly for any distinguishable programming needs.

The charge trapping (CT) type memory such as silicon-oxide-nitride-oxide-silicon (SONOS) has been developed to replace floating gate device due to its superior scaling capability, as it is immune to the

capacitive-coupling issue and enables the adoption of thin gate stacks for better electrostatic control [1.9]. In CT (SONOS) memory, electrons are stored in discrete trapping sites in the nitride layer and high program speed can be achieved while maintaining good retention. However, SONOS memory also has its own problem such as slow de-trapping rate [1.10] and moreover, the 2-D scaling of CT memory is still limited by the small number of electrons when the volume of charge storage node becomes ultra-small.

1.3 Scope of Project

In this dissertation, we will evaluate new methodologies which enable further scaling of nonvolatile flash memory devices. Due to the scaling advantage of discrete charge trapping type memory devices based on aforementioned discussion, this study will focus on one type of discrete charge trapping memory devices, SONOS-type memory device. We will fabricate and investigate various types of SONOS-type memory based on a novel gate-all-around (GAA) vertical Si nanowire (SiNW) cell structure using CMOS compatible technology. These novel vertical NW devices can be considered as promising candidates for future ultra-high density memory application.

1.4 Organization of Thesis

The whole thesis consists of seven chapters in total which have been arranged as follows.

Chapter 1 provides an introductory overview of the semiconductor memory technology and discusses different categories of semiconductor memory devices. It also presents the current challenges of flash memory device scaling in planar style.

Chapter 2 gives a detailed literature study on key findings in the earlier research work on SONOS-type memory with novel cell structures and compares the advantages of multi-gated memory cells with the planar ones. Various 3-dimensional (3-D) memory array stacking approaches are also reviewed.

In Chapter 3, we demonstrate a junction-based GAA vertical SiNW SONOS memory using a top-down method. The Flash devices realized on highly scaled SiNW channel down to 20 nm show much improved P/E speed and window along with good retention and endurance characteristics as compared to the ones with a large dimension. The performance improvements with scaling of nanowire (NW) dimensions are investigated through device simulation. To further improve the device memory performances, silicon nanocrystals (Si-NCs) were incorporated as charge trapping material into the junction-based memory devices. Engineering of charge trapping layer by incorporating Si-NCs is found to be beneficial for improved SiNW-based memory characteristics.

In Chapter 4, we present junction-less SONOS (JL-SONOS) memory realized on vertical SiNW GAA structures with channel diameter down to 20nm. Using Fowler-Nordheim (F-N) tunneling P/E operation mechanism, the junction-less (JL) devices exhibit comparable electrical characteristics ($SS < 70\text{mV/decade}$, and a memory window

of 3.2V in 1ms P/E) to the junction-based wire SONOS. The absence of junctions also reduces the process complexity and makes this device a suitable platform for vertically stacked ultra-high-density memory applications. It is also found that two physical bits per cell storage scheme can be achieved for JL-SONOS by using band-to-band hot hole (BTBHH) injection programming mechanism. Compared with a planar device, the array cell tuning for physical multi-bit storage is much easier since the channel length is no longer limited by the cell area.

In Chapter 5, we present GAA vertical poly-Si NW thin film transistors (TFTs) and SONOS devices with wire diameter down to 30nm. Both memory and transistor performances are improved with down scaling of the NW diameter, which is owing to the reduced number of grains and grain boundaries in the channel region in addition to the superior gate control provided by the GAA structure. Inverter circuits based on vertical poly-Si NW TFT technology are also presented with well-behaved transfer characteristics under the device dimension of N- to P-TFTs ratio of 1:1.

In Chapter 6, we demonstrate a novel 2-level stacked JL GAA SONOS memory fabricated on vertical SiNW platform with footprint of $6F_2/2$ and a memory string built based on this technology. Without forming junctions on vertical-wires, the stacking process complexity/cost is greatly reduced with promise towards density improvement. Systematic evaluation of the stacked memory cell from the P/E speed, endurance, and retention to the program disturb is presented.

Last but not least, an overall conclusion is given in Chapter 7 to summarize the major results. Possible future work is proposed in this chapter too.

Reference

- [1.1] International Technology Roadmap of Semiconductors (ITRS), Semiconductor Industry Association, San Jose, CA, (2007).
- [1.2] D.J. Jung, "Highly manufacturable 1T1C 4Mb FRAM with novel sensing scheme", pp. 279-282, *IEDM*, (1999).
- [1.3] "Advanced Memory Technology and Architecture", short course, *IEDM*, (2001).
- [1.4] S. Q. Liu, N. J. Wu, and A. Ignatiev, "Electric-pulse induced reversible resistance change effect in magnetoresistive films", *Appl. Phys. Lett.*, Vol. 76, Id 2749, (2000).
- [1.5] S. Lai and T. Lowrey, "A 180nm nonvolatile memory cell element technology for stand alone and embedded applications", pp. 803, *IEDM*, (2001).
- [1.6] A. Niebel, "Business outlook for the non-volatile memory Market", in *IEEE Non-Volatile Semiconductor Memory Workshop (IEEE NVSMW)*, 21st, pp. 6-7 (2006).
- [1.7] C. S. Pan, K. Wu, D. Chin, G. Sery, and J. Kiely, "High-temperature charge loss mechanism in a floating-gate EPROM with an oxide-nitride-oxide (ONO) interpoly stacked dielectric", *IEEE Electron Device Lett.*, Vol. 12, pp. 506-509 (1991).
- [1.8] M. She, "Semiconductor flash memory scaling", PhD Dissertation (University of California, Berkeley) (2003).
- [1.9] J. D. Lee, S. H. Hur, and J. D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation", *IEEE Electron Device Lett.*, Vol. 23, pp. 264-266 (2002).

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

Chapter 2

Literature Review

2.1 Introduction

To advance the down scaling of flash memory devices to its extreme limit, plenty of research work has been done in the last few decades in two main directions: (1) adoption of novel cell structures such as FinFet SONOS and NW-based GAA SONOS; and (2) integration of new material such as silicon nanocrystals and high- κ dielectrics. This chapter reviews the key findings in the earlier research work on SONOS-type memory with novel cell structures and materials. We also review various 3-D NAND flash memory array approaches for multi-level ultra-high density memory application.

2.2 Flash Memory with Multi-gated Cell Structures

From the view point of cell structure, various multi-gate device architectures, that is, double-gate [2.1], tri-gate [2.2], Π -gate [2.3], Ω -gate [2.4] and GAA [2.5] have been explored extensively as planar device scaling approaches the end of the technology

roadmap. These multi-gated cells have the following advantages as compared to the planar cells:

- Enhanced electrostatic control and immunity to short-channel-effects (SCEs),
- Improved gate coupling effect.

Since these multi-gated cells usually employ the fin, cylinder or arch-shaped active profiles along with multi-gate which provide additional suppression of SCEs for sub-100nm CMOS technology with low sub-threshold slope (SS), drain induced barrier lowering (DIBL) and reduced V_{th} roll-off. S. Lombardo *et al.* reported double and tri-gate FinFET SONOS type memory with a large improvement in terms of V_{th} window, which was explained by the distribution of the electric field in the gate dielectric stack. These devices show excellent improvement on the suppression of SCEs even though the overall thickness of the dielectric stack is above 15nm [2.6]. Among various multi-gate devices, the GAA cylindrical NW-FET is being considered as one of the strong potential candidates to advance the CMOS technology to push for the extreme miniaturization limits, due to the optimal electrostatic control which translates into nearly ideal subthreshold turn-on and negligibly small SCEs [11]. Owing to the cylindrical geometry, and the inverse logarithmic dependence of the insulator capacitance on the oxide thickness, the gate length in these devices can be scaled with channel diameter (for instance, close to 10nm scale) without reducing the gate dielectric thickness. Therefore, GAA NW devices represent the most suitable architecture for SONOS-type NVM applications, where the gate dielectric can hardly be scaled due to the severe retention properties (10 years) to be met. S. D. Suk *et al.* reported a high performance 5nm radius twin SiNW MOSFET fabrication on bulk Si wafer [2.8]. Such devices are almost

immune to short channel effect due to the excellent electrostatic control owing to the GAA structure and highly scaled NW channel as shown in Fig. 2.1.

Fig. 2.1: 5nm radius twin SiNW GAA MOSFET with V_{th} roll-off, SS, and DIBL remained same along the gate length [2.8].

Moreover, when the SiNW diameter is scaled to the value that is comparable to the ONO layer thickness (~ 20 nm), it can be well understood that the geometry of the SiNW device will make the E-field more concentrated near the silicon channel at center of the device structure [2.9]. Under the F-N tunneling mechanism (Eq. 2.1), the tunneling current is exponentially proportional to the electrical field, and thus, with the reduction of SiNW dimension, the larger tunneling current will help improve the program speed tremendously.

$$J = \frac{q^3}{8\pi\hbar\Phi_B} \frac{m}{m^*} E_{inj}^2 \exp\left[\frac{-E_C}{E_{inj}}\right], \quad (2.1)$$

$$E_C = 4\sqrt{2m^*} \frac{\Phi_b^{3/2}}{3\hbar q}$$

where q is the electron charge, h is the Planck's constant, m is the electron effective mass in the silicon conduction band, m^* is the electron effective mass in the oxide conduction band, Φ_B is the potential barrier height at the injecting surface, E_{inj} is the electric field at the injecting surface.

J. Fu *et al.* demonstrated a GAA vertical-twin-Si-NW SONOS memory with excellent performance [2.10]. The NW memory device exhibited excellent improvement in terms of P/E speeds as compared to the planar counterpart, and by scaling down the dimension of the wire channel; further improvement of P/E speeds was observed.

2.3 3-Dimensional (3-D) Stacked NAND Array

To overcome the challenges in 2-D scaling, 3-D CT array is considered as a promising candidate for future high density memory due to the following advantages:

- Memory density increased not by the device size scaling but by the 3-D stacking,
- Extended volume of charge storage node and increased the number of electrons,
- Improved device performance due to multi-gated structure.

Through 3-D stacking, the memory density can be increased not by scaling the device dimension but through increasing the number of cells stacked in the vertical direction. Endoh *et al.* reported ultrahigh-density memory with a stacked-surrounding gate transistor (S-SGT) structure cell [2.11]. The S-SGT structure shown in Fig 2.2 (a) which consists of two stacked memory cells as well as the top and bottom select transistors in one silicon pillar achieves a cell area per bit of less than 50% of the smallest reported NAND structured cell. By increasing the number of cells stacked in that pillar, the bit area can be further reduced to $4F^2/N$ per bit, where N is the number of stacked memory cells in one silicon pillar (see Fig 2.2 (b)).

(a)

(b)

Fig. 2.2: Concept of S-SGT structured cell: (a) left: bird's view of S-SGT structured cell; center: cross-sectional view along (B1)–(B2) of S-SGT structured cell; right: equivalent circuit of S-SGT structured cell and NAND structured cell; (b) ratio of occupied area per bit as function of number of stacked memory cells in one silicon pillar in comparison with the smallest reported cell area per bit of conventional NAND structured cell [2.11].

Also as most of the 3-D memories adopt 3-D device structures, volume of charge storage node can be extended in the vertical direction and the number of electrons stored can be increased accordingly and thus, solving the “running out of electrons” issue faced by 2-D devices. In addition, most 3-D cell structures employ the fin, cylinder or arch-shaped active profiles along with a multi-gated device structure, superior electrostatic control can be expected due to better suppression of SCEs for sub-100nm CMOS technology. So far, great efforts have been invested on the development of 3-D stacked NAND flash memory arrays and major 3-D stacking approaches include: Bit-cost scalable (BiCs) [2.12], Vertical-Stacked-Array-Transistor (VSAT) [2.13] and Vertical-gate NAND [2.14].

Jung *et al.* reported a 2-layer stacked memory array by implementing the single-crystal Si layer stacking technology [2.15]. After fabricating the bottom cell array, the top cell array is fabricated on the stacked single-crystal silicon active layer on ILD and interconnected simultaneously with the bottom cell arrays to the peripheral circuits. The NAND cell arrays formed on the ILD as well as on the bulk to double the memory density without increasing the chip size (see Fig. 2.3).

Fig. 2.3: Vertical SEM photographs of the fabricated 3-D stacked NAND cell string. The second active layer is SOI-like perfect single crystal [2.15].

Despite the improved bit density, there are several disadvantages for these stacked planar devices such as:

- Several critical lithography steps needed for each stacked layer,
- The area of driver transistors for control gates is multiplied by the number of layers,
- The cost does not decrease or even increases due to a yield loss and an area penalty if the number of stacked layers is more than three.

To effectively reduce the bit cost, Tanaka *et al.* realized the BiCs flash memory using poly-Si channel, in which holes are punched through after the multi-level gate stack formation of the lower select gate transistor, memory string and upper select gate transistor, and after that poly-Si is plugged into to form the channel [see Fig 2.4]. It is considered as a breakthrough concept as it uses only critical contact drill hole for many layers. Such punch and plug process greatly reduces the process complexity and individual bit cost, especially when the number of memory cells along each vertical channel increases (see Fig. 2.5). However, several potential reliability issues, such as larger SS, leakage, V_{th} distribution and data retention problems, might be correlated with the inherent low grade of poly-Si channel and deposited tunnel oxide, as compared to conventional planar flash devices.

Fig. 2.4: (a) Birds-eye view of BiCS flash memory, (b) top down view of BiCS flash memory array [2.12].

Fig. 2.5: Bit cost scalability of three-dimensional flash memory [2.12].

Kim *et al.* also reported an alternative 3-D NAND array approaches using poly-Si channel called Vertical-Stacked-Array-Transistor (VSAT), as schematically

illustrated in Fig. 2.6 and Fig. 2.7. The major advantages of VSAT as compared to BiCs are:

- Adopt a cost-effective method to connect the array transistors to the peripheral circuits,
- Select Lines are separated from the vertical strings,
- Improvement of off-current.

Fig. 2.6: Schematic views of VSAT and PIPE. It has planar type SLs (GSL and SSL) in the peripheral region, not embedded in the vertical string [2.13].

Fig. 2.7: Equivalent circuit of VSAT device with three vertical strings, SLs are separated from vertical strings like a conventional NAND device [2.13].

In BiCs, stair-like structure is created before making contact holes and connecting memory cells to the periphery circuit. VSAT provides a cost-effective method compared to the stair-like structure, in which all the gate electrodes are exposed on the same plane after a chemical mechanical polish (CMP) process, allowing easy access to the gate electrodes (see Fig. 2.8).

Fig. 2.8: Vertical interconnection (a) is simpler and more cost-effective than stair-like (b) method [2.13].

In addition, BiCs requires select transistors to be fabricated in the vertical string of memory cells. Since select transistors have different gate oxide, gate length, and doping density than the array transistors located along the same string, integrating these select transistors into the vertical string can be a critical challenge. In the VSAT structure, the select transistors are separated from the vertical strings, which are similar to conventional 2-D NAND devices, thus simplifying the procedure to realize the vertical string. Furthermore, the VSAT has double gates connected in series, which effectively extend the channel length to reduce the off-current without sacrificing the memory density.

The major drawbacks of VSAT structure are its lower read speed as well as the X-pitch scaling issue caused by the poly/oxide filling. Another 3-D array design named Vertical Gate NAND (VG-NAND) is proposed by Kim *et al.*. That VG-NAND structure uses vertical gate and horizontal multi-active layers and it solves that problems

encountered by BiCs and VSAT such as WL interconnect, program disturbance, and large channel resistance as the number stacked cells increase (see Fig. 2.9).

Fig. 2.9: (a) Structure of vertical gate (VG)-NAND, (b) top down view of VG-NAND showing $4F^2$ cell size / layer [2.14].

2.4 Vertical Si Nanowire Transistor Technology

In the last section, various approaches of constructing 3-D NAND array were discussed and compared. One similarity among these approaches is that they all adopt a poly-silicon active layer (regardless gate-first or gate-last) and deposited tunnel oxide. Compared with conventional planar SONOS memory using single-crystalline active layer and thermally grown oxide, the inherent low grade of poly channel and deposited tunnel oxide will lead to possible downgrade of device performance and reliability.

An alternative approach to fabricate 3-D memory is proposed by building GAA SONOS memory cells along vertical-SiNW. B. Yang *et al.* in 2008 demonstrated a vertical GAA SiNW transistor on bulk silicon wafer utilizing fully CMOS compatible technology, as shown in Fig. 2.10.

Fig. 2.10: SEM pictures of device fabrication: (a) Vertical NW with diameter of 20 nm. (b) After gate patterning by lithography but before exposing the drain (the tip of the pillar) of the transistor. (c) After the drain (pillar tip) is exposed. Metal contacts are made with such device structure. (d) Vertical NW arrays with pitch of 500 nm. NWs are 1 μm tall with a diameter of 20 nm [2.16].

High aspect ratio (up to 50: 1) vertical NWs with diameter ~ 20 nm are achieved from lithography and dry-etch defined Si-pillars with subsequent oxidation. The surrounding gate length is controlled using etch back of the sacrificial oxide. N-MOS devices thus fabricated with gate length ~ 150 nm showed excellent transistor characteristics with large drive current, high $I_{\text{on}}/I_{\text{off}}$ ratio, good sub-threshold slope and low DIBL [2.16].

In 2009, M. C. Chen *et al.* utilized the same vertical-Si-NW platform to develop the SONOS memory devices that showed excellent memory window and reliability characteristics [2.17].

Fig. 2.11: XTEM images of a vertical-SiNW memory with $L_g=150$ nm and a SONOS gate stack (with the thickness of $\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2=5/5/6$ nm), with the SEM image of a single NW as inset (diameter = 50 nm).

As compared to other 3-D memory cell structures, the GAA vertical Si NW cell shows enhanced device performance and reliability due to the following reasons: single-crystalline Si active and thermal grown tunnel oxide; highly scaled SiNW channel for enhanced memory speed, shown in Fig. 2.11. Using single-crystalline Si channel and high quality thermally grown tunnel oxide, the vertical SiNW based SONOS memory provides a possible solution to solve issues like large V_{th} distribution, SS, and data loss faced by poly-Si active based 3-D cells. Furthermore, in BiCs, the diameters of the poly-Si pillar channel is limited by the resolution limits of lithography equipment as well as the hole filling capability for poly-Si plugs. In vertical SiNW GAA SONOS, a highly scaled vertical SiNW can be achieved through thinning down large silicon pillars using thermal oxide. A highly scaled channel can be beneficial for enhancing the device speed due to the electrical field augment effect [2.18].

Although the vertical SiNW SONOS memory appears as a very attractive approach for 3-D high density storage without encountering issues such as large SS, V_{th} distribution and data loss, the fabrication process of stacking multiple cell can be very challenging, especially when forming junctions on pillars between cells. This makes the JL concept when considering stacking number of cells along the pillar [2.19]. J.P. Colinge *et al.*, in 2010, demonstrated a new type of transistor in which there are no junctions and no doping concentration gradients. These devices have full CMOS functionality and are made using silicon nanowires. They have near-ideal sub-threshold slope, extremely low leakage currents, and less degradation of mobility with gate voltage and temperature than classical transistors [2.20]. Combining the advantages of JL concept and vertical NW platform, without the need to form junctions at the top and bottom of the SiNW, the fabrication complexity and cost for vertical SiNW SONOS memory will be greatly reduced. Being free of doped junctions, device can also become much less sensitive to thermal budget issue than conventional doped devices. Also, the absence of junctions on wire allows easy stacking of multiple cells on a NW, which makes the JL memory device is a suitable platform for vertical stacking of 3-D cells to achieve ultra-high density memory density.

Reference

- [2.1] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, “A folded-channel MOSFET for deep-sub-tenth micron era”, *IEDM Tech. Dig.*, pp. 1032-1034 (1998).
- [2.2] B. S. Doyle, S. Datta, M. Doczy, S. Harelend, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, “High performance fully-depleted tri-gate CMOS transistors”, *IEEE Electron Device Lett.*, Vol. 24, pp. 263-265 (2003).
- [2.3] J.-T. Park, J.-P. Colinge, and C. H. Diaz, “Pi-gate SOI MOSFET”, *IEEE Electron Device Lett.*, Vol. 22, pp. 405-405 (2001).
- [2.4] F.-L. Yang, H.-Y. Chen, F.-C. Cheng, C.-C. Huang, C.-Y. Chang, H.-K. Chiu, C.-C. Lee, H.-T. Huang, C.-J. Chen, H.-J. Tao, Y.-C. Yeo, M.-S. Liang, and C. Hu, “25 nm CMOS omega FETs”, *IEDM Tech. Dig.*, pp. 255-258 (2002).
- [2.5] J. P. Colinge, M. H. Gao, A. R.-Rodriguez, H. Maes, and C. Claeys, “Siliconon-insulator ‘gate-all-around device’”, *IEDM Tech. Dig.*, pp. 595-598 (1990).
- [2.6] S. Lombardo, C. Gerardi, L. Breuil, and C. Jahan, “Advantages of the FinFET architecture in SONOS and Nanocrystal memory devices”, *IEDM Tech. Dig.*, pp. 921-924 (2007).
- [2.7] E. Gnani, S. Reggiani, M. Rudan, and G. Baccarani, “Design considerations and comparative investigation of ultra-thin SOI, double-gate and cylindrical nanowire FETs”, *Solid-State Device Research Conference, Proceeding of the 36th European ESSDERC* (2006), pp. 371-374.
- [2.8] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, and J.-B. Park, “High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET) fabrication on bulk Si wafer, characteristics, and reliability”, *IEDM Tech. Dig.*, pp. 717-720 (2005).
- [2.9] T.-H. Hsu, H.-T. Lue, E.-K. Lai, J.-Y. Hsieh, S.-Y. Wang, L.-W. Yang, Y.-C. King, T. Yang, K.-C. Chen, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, “A high-speed BE-SONOS NAND flash utilizing the field-enhancement effect of FinFET”, *IEDM Tech. Dig.*, pp. 913-916 (2007).

- [2.10] J. Fu, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, "Si-nanowire based gate-all-around nonvolatile SONOS memory cell", *IEEE Electron Device Lett.*, Vol. 29, pp. 518-521 (2008).
- [2.11] T. Endoh, K. Kinoshita, T. Tanigami, Y. Wada, K. Sato, K. Yamada, T. Yokoyama, N. Takeuchi, K. Tanaka, N. Awaya, K. Sakiyama, and F. Masuoka, "Novel ultrahigh-density flash memory with a stacked-surrounding gate transistor (S-SGT) structured cell", *IEEE Trans. on Electron Device*, Vol. 50, pp. 945-951 (2003).
- [2.12] 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," *VLSI Symp. Tech. Dig.*, pp. 14-15 (2007).
- [2.13] J. Kim, A. J. Hong, S. M. Kim, E. B. Song, J. H. Park, J. Han, S. Choi, D Jang, J. T. Moon, and K. L .Wang, "Novel vertical-stacked-array-transistor (VSAT) for ultra-high-density and cost-effective NAND flash memory devices and SSD (Solid State Drive)", *VLSI Tech. Symp. Dig.*, pp. 186-187 (2009).
- [2.14] W. Kim, S. Choi, J. Sung, T. Lee, C. Park, H. Ko, J. Jung, I. Yoo, and Y. Park, "Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage", *VLSI Tech. Symp. Dig.*, pp. 188-189 (2009).
- [2.15] S. Jung, J. Jang, W. Cho, H. Cho, J. Jeong, Y. Chang, J. Kim, Y. Rah, Y. Son, J. Park, M. Song, K. Kim, J. Lim, and K. Kim, "Three dimensionally stacked NAND flash memory technology using stacking single crystal Si layers on ILD and TANOS structure for beyond 30nm node", *IEDM Tech. Dig.*, pp. 37-40 (2006).
- [2.16] B. Yang, K. D. Buddharaju, S. H. G. Teo, N. Singh, G. Q. Lo, and D. L. Kwong, "Vertical silicon-nanowire formation and gate-all-around MOSFET", *IEEE Electron Device Lett.*, Vol. 29, pp. 791-794 (2008).
- [2.17] M. C. Chen, H. Y. Yu, N. Singh, Y. Sun, N. S. Shen, X. H. Yuan, G. Q. Lo and D. L. Kwong, "Vertical Si nanowire SONOS memory for ultra-high density application", *IEEE Electron Devie Lett.*, Vol. 30, pp. 879-881 (2009).

- [2.18] Y. Sun, H. Y. Yu, N. Singh, N. S. Shen, G. Q. Lo, and D. L. Kwong, “Multi-bit programmable flash memory realized on vertical Si nanowire channel”, *IEEE Electron Device Lett.*, Vol. 31, pp. 390-392 (2010).
- [2.19] Y. Sun, H. Y. Yu, N. Singh, E. Gnani, G. Baccarani, K. C. Leong, G. Q. Lo, and D. L. Kwong, “Vertical Si-nanowire based non-volatile memory with improved performance and reduced process complexity”, *IEEE Transaction on Electron Devices*, to be published (2011).

Chapter 3

High Performance Junction-based Vertical Si-Nanowire SONOS

3.1 Introduction

Conventional planar flash memory technology faces serious challenges when approaching sub 20-nm node, due to concerns on lithography, coupling ratio, cross talk interference, and variability [3.1-3.3]. These fundamental limitations make it attractive to explore various 3-D cell architectures that can allow higher integration densities. GAA SONOS flash memory built on vertical SiNW is a promising candidature for future ultra-high density memory, not only because of its GAA structure which provides superior gate control and gate length scalability, but also due to its potential to serve as building blocks for 3-D multi-level flash memory integration [3.4].

In previous chapters, various approaches of constructing 3-D NAND array were discussed and compared. One similarity among these approaches is that they all adopt a poly-silicon active layer (regardless gate-first or gate-last) and deposited tunnel oxide. Compared with conventional planer SONOS memory using single-crystalline active

layer and thermally grown oxide, the inherent low grade of poly channel and deposited tunnel oxide will lead to possible downgrade of device performance and reliability.

An alternative approach to fabricate 3-D memory is to build GAA SONOS memory cells along a vertical SiNW. Chen and Sun et al. report a single-crystalline vertical SiNW based GAA SONOS memory (see Fig 3.1). Using single-crystalline Si channel and high quality thermally grown tunnel oxide, the vertical SiNW based SONOS memory provide a possible solution to solve issues like large V_{th} distribution, SS, and data loss faced by poly-Si active based 3-D cells [3.5].

Fig. 3.1: Schematic of cross-section of vertical-SiNW-based SONOS memory.

In this chapter, using a completely CMOS compatible process technology, we fabricated and studied SONOS memory devices based on a highly scaled SiNW channel (dimension down to 20 nm), with a significant P/E speed performance enhancement as compared to the devices with a SiNW diameter of 50nm. This enhancement is explained by the augment of electric field across the tunneling dielectrics near the silicon core channel due to the scaled dimension as well as the smaller radius of curvatures at the four corners of the wire [3.5], [3.6]. Benefited from the larger program window and good process control, the multi-bit programmable memory enables 2-bit data storage per cell, which can further improve the memory density [3.7]. In addition, to improve the device memory performances, such as P/E speed, retention and endurance, Si-NCs were incorporated as charge trapping material into the junction-based vertical SiNW GAA NVM devices. Nanocrystal-based memory is attractive for its advantage in preventing lateral charge movement. Engineering of charge trapping layer by incorporating Si-NCs in this work has been found to be beneficial for improved SiNW-based memory characteristics.

3.2 Junction-based Vertical SiNW SONOS Device Fabrication

The schematics of the fabrication processes are shown in Fig.3.4. P-type starting substrate was used with a doping level of $5 \times 10^{15} \text{cm}^{-3}$. Si_3N_4 hard mask was firstly patterned with 248nm lithography to form circular dots with diameter ranging from 180nm to 800nm on 8" bulk Si wafer. It was followed by 800nm deep Si etching with SF_6 chemistry under the hard mask to form vertical pillars. Rapid thermal oxidation at 1000°C followed by diluted HF (DHF) etch was carried out to reduce the pillar diameter (down to 20nm in this work), as illustrated in Fig. 3.2. It should be noted that the footing at the bottom of the SiNW was intentionally formed to support long wires.

Fig. 3.2: DR-SEM images of vertical SiNW formation using top down approach: (a) SiN hard mask patterning, (b) Si pillar etch, (c) pillar oxidation, (d) after DHF etch.

Interestingly, at 1000 °C the stress effect was minimal and thus oxidation process delimited in one of the directions, converting the multi-oriented cylindrical wire into a squarish wire with 4 clear facets ($\phi \sim 20\text{nm}$) the thicker wires were converted into less squarish wires ($\phi \sim 50\text{nm}$), as shown in Fig. 3.3.

Fig. 3.3: Planar TEM images of vertical SiNW SONOS with (a) $\Phi=50\text{nm}$ and (b) $\Phi=20\text{nm}$.

After vertical SiNW formation, As⁺ was implanted ($30\text{ KeV}/10^{15}\text{cm}^{-2}$) to form the source n⁺ region at the bottom of the NW, as illustrated in Fig.3.4 (1). The dopants were activated under $1000\text{ }^{\circ}\text{C}$ for 10 sec in N₂. The tip of the NW was protected by the Si₃N₄ hard mask from the implant. Non-conformal deposition of high density plasma (HDP) oxide followed by DHF etch-back was performed to form a layer of 120nm HDP oxide to cover the footing of the NW, as depicted in Fig.3.4 (2). Tunneling oxide (5nm) was thermally grown on SiNW followed by 7 nm low pressure chemical vapor deposition (LPCVD) SiN trapping layer for SONOS flash and Si-NCs for nano-crystal (NC) flash. The device which employs only Si-NCs as the trap layer is named as SOncOS. For Si-NC deposition, silane (SiH₄) gas was used as gas source with 170sccm flow rate, at a deposition temperature of 580°C and pressure of 200mT. After trap layer (SiN or NC) deposition, blocking oxide layer of 7nm were deposited by LPCVD. Gate stack was completed by deposition of gate poly Si. As was implanted ($10\text{ KeV}/10^{15}\text{cm}^{-2}$) and activated under 1000°C for 30 sec in N₂ gate to make the N⁺ poly gate. The gate pad was then patterned and etched, serving as a poly extension for gate contact, as shown Fig.3.4 (3). HDP oxide deposition followed by etch-back was

repeated such that only the top of the poly was exposed while protecting the underlying gate pad. The exposed top poly Si with its underlying ONO stack together with the remaining Si_3N_4 hard mask was isotropically etched using reactive ion etching (RIE), exposing the top of the NW. As^+ ($15 \text{ KeV}/10^{15} \text{ cm}^{-2}$) was implanted from four directions of 90 degree apart at a tilt of 60° to form the n^+ drain region, as it can be seen from Fig.3.4 (5). The dopants were then activated under 1000°C for 5 sec in N_2 . Then HDP deposition followed by etch-back was repeated to expose the Si tip for metal contact. The device fabrication was finished by standard metallization, as shown in Fig.3.4 (6).

Fig. 3.4: Schematics illustrating the process flow used to fabricate vertical GAA SiNW based SONOS.

In Fig.3.5, SiNW SONOS device SEM inspection during several key device fabrication processes are shown, e.g. Fig.3.5 (a) after 1st HDP deposition and etch back, corresponding to Fig.3.4 (2); Fig.3.5 (b) after gate stack deposition and patterning, corresponding to Fig.3.4 (3); Fig.3.5 (c) after 2nd HDP deposition and etch back, corresponding to Fig.3.4 (4); Fig.3.5 (d) after tip poly removal and drain implant, corresponding to Fig.3.4 (5); Fig.3.5 (e) after metallization and metal etch, corresponding to Fig.3.4 (6);

(a)

(b)

(c)

(d)

(e)

Fig. 3.5: SEM images for the device with a NW diameter of $\sim 50\text{nm}$ (a) after 1st HDP deposition and etch back; (b) after gate stack deposition and patterning; (c) after 2nd HDP deposition and etch back; (d) after tip poly removal and drain implant; (e) after metallization and metal etch.

Fig. 3.6: (a) Schematic diagrams show the cross section part of NW channel SONOS (left) and SONcOS (right) devices. (b) Cross-sectional TEM image of vertical SiNW GAA SONOS: diameter of wire is 50 nm and gate length 150nm. (c) AFM image scanned on $1 \times 1 \mu\text{m}^2$ surface area that reveals the formation of Si-NC with a density of $7.5 \times 10^{10} / \text{cm}^2$.

Fig. 3.6 (a) shows the schematic diagrams of cross section part of SiNW channel SONOS (left) and SONcOS (right) devices. Shown in Fig. 3.6 (b) is a cross-sectional transmission electron microscopy (TEM) image of a vertical SiNW GAA SONOS with

wire diameter equals to 50 nm and gate length of ~150nm. All wire diameters were taken near the drain end which is the narrowest part of the wire. Shown in Fig. 3.6 (c) is the atomic force microscopy (AFM) image of an area of $1 \times 1 \mu\text{m}^2$ surface after Si-NCs deposition revealing good uniformity of isolated Si-NCs. The size of NC is about 30 ± 10 nm in diameter and density is around $7.5 \times 10^{10} / \text{cm}^2$.

3.3 Results and Discussion

3.3.1 Junction-based Vertical SiNW SONOS

In this section, junction-based vertical SiNW SONOS devices with gate length $L_g \sim 100$ nm for various wire diameters were used for characterization. All wire diameters are taken at the channel/tunnel oxide interface. The threshold voltage (V_{th}) is obtained from I_d - V_g characteristic with drain current (I_d) equal to 10^{-7} A/ μm .

It is worthy pointing out that the profile of the SiNW sidewalls has crucial impact on the device operation. Before process optimization, the SiNW SONOS fabricated in this work can become depletion type, as the As^+ ions would be doped into the non-vertical channels during source implantation and hence the SONOS cells become depletion-type with an initial negative V_{th} value. Fig.3.7 depicts the transient characteristics of the depletion-type vertical SiNW SONOS device with gate length of 150nm and wire diameter of 50nm for both P/E. A positive pulse (i.e. $V_g > 0$) is used to program the SONOS cell with all other terminals grounded. Due to the nature of GAA structure, the negative pulse (i.e. $V_g < 0$) on the control gate is necessary during erase process. Further considering the n-type channel for the devices, a positive bias (i.e. $V_d=5\text{V}$) to the drain side is applied to move downward the electron potential energy of the channel region to facilitate the erasing process. The gate bias for programming/erasing ranged from

'12/-18V to 18V/-12V' (from 1 μ s to 1s). A memory window of ~ 2.5 V is seen when a P/E time of 1ms was measured at ± 18 V. The difference between the programming and erasing speed is seen, which is believed to be correlated with the additional drain voltage applied during erasing process.

(a)

(b)

Fig. 3.7: Depletion-type vertical SiNW (with a diameter of 50nm) SONOS memory ($L_g=150\text{nm}$): (a) Programming characteristics at gate bias of 12V to 18V for $1\mu\text{s}$ to 1s. (b) Erasing characteristics at gate bias of -18V to -12V with a drain bias of 5V for $1\mu\text{s}$ to 1s.

In Fig.3.8 (a), the depletion-type vertical SiNW SONOS memory shows good high temperature retention characteristics (85°C) after respective programming ($V_g=16\text{V}$ or 18V for 1ms) and erasing ($V_g= -16\text{V}$ or -18V with $V_d=5\text{V}$ for 1ms). In Fig.3.8 (b), the device endurance was done at programming ($V_g=18\text{V}$ for 10ms) and erasing ($V_g= -18\text{V}$ with $V_d= 5\text{V}$ for 10ms). It is noted that the over erasing (i.e. erase V_{th} decreases) occurs after 10^3 cycles, which is believed to be due to the additional drain voltage (i.e. $V_d = 5\text{V}$) applied during erasing. On the other hand, the device does not show any P/E window narrowing after 10^4 cycles.

Fig. 3.8: Depletion-type vertical SiNW (with a diameter of 50nm) SONOS memory (Lg=150nm): (a) Retention characteristic at 85 °C (b) Endurance characteristic.

In the next session, we present electrical characterization results of junction-based vertical SiNW SONOS memory devices with improved NW profile (more straight) and highly scaled squarish channel with dimension down to 20nm in diagonal. Compared with the previously discussed depletion-type memory devices, the optimized vertical SONOS devices with squarish channel have the following advantages:

- With much improved profile (more straight) of the SiNW and larger hard mask overhang (more oxidation), the channel doping with As^+ ions during the source implantation can be significantly reduced and hence, the devices are immune from depletion type. Therefore, F-N tunneling can be used for device erasing,
- By the scaling of the SiNW diameter, the SiNW SONOS memory program window can be greatly enhanced and can thus, can be programmed for multi-level-cell MLC operation,
- For devices with squarish channel, the electric field can be further enhanced at the four corners with smaller radius of curvature, which contributes to further enhanced P/E speeds.

Fig. 3.9 (a) and (b) present the programming and erasing characteristic of vertical SiNW GAA SONOS flash memory after process optimization to reduce the channel doping during source implantation and with a more highly scaled wire channel ($L_g \sim 150\text{nm}$, diameter=20nm). A memory window of $\sim 4.2\text{V}$ is observed when a P/E time of 1ms was measured at +15V and -16V respectively.

Fig. 3.9: (a) Programming characteristic of vertical SiNW SONOS memory ($L_g \sim 150$ nm and diameter = 20 nm). (b) Erasing characteristic of vertical SiNW SONOS memory ($L_g \sim 150$ nm and diameter = 20 nm).

Fig.3.10 compares the P/E characteristics of devices with SiNW dimensions 20 and 50nm. It should be noted that the P/E window is expected to be further improved with the optimization of the gate stack. It is seen that the device with the highly scaled dimension (~20nm) show much enhanced P/E speed than that with a larger dimension (~50nm).

Fig. 3.10: Comparison of the P/E characteristics of vertical SiNW SONOS memory with different diameters (50 nm versus 20 nm).

When the SiNW diameter is scaled to the value that is comparable to the ONO layer thickness (~20nm), it can be imagined that the geometry of SiNW device will make the E-field more concentrated near the silicon channel. In addition, the electric field will also be further enhanced at the four corners with smaller radius of curvature [3.5], [3.6], which will lead to a much enhanced electrical in the tunnel oxide and a corresponding reduced electrical field in the block oxide. Under F-N tunneling

mechanism, the tunneling current is exponentially proportional to the electrical field. And thus with the reduction of SiNW dimension and a more squarish channel, the larger tunneling current will help improve the program speed. In addition, the reduced electrical field in the block oxide will also reduce the tunneling current from trapping layer to the poly-Si control gate, which further contributes to the improvement of the program speed. Vice versa, the smaller dimension of the SiNW channel is also beneficial for the erasing of the devices.

Further from Fig. 3.10, it is seen with the stress time increases, during programming, the V_{th} shift of the device with a diameter of 20nm starts to saturate in the millisecond range, which is believed to be related to the additional electrical field caused by the electrons that are already trapped in the trapping layer during programming process. The saturation causes the V_{th} shift of device with a diameter of 20nm less significant compared to device with a larger diameter (i.e. 50nm) in the millisecond range.

Preliminary device simulation is done to verify this point. The one-dimensional simulation model in cylindrical coordinate uses carrier continuity equation with the Shockley-Reed-Hall recombination theory to describe the trapping mechanism in the nitride. The tunneling and blocking oxide are assumed as pure insulator and hence only Poisson equation is solved throughout the ONO layer. Fig.3.11 compares the simulated electrical field distribution through the ONO layers for devices with different SiNW diameters (20nm, 50nm) by stressing both devices at a gate bias of 15V for 100ms. The device with smaller SiNW diameter (20nm) shows a much enhanced electrical in the tunnel oxide and a corresponding reduced electrical field in the block oxide, which is consistent with the experimental results.

Fig. 3.11: Simulation of electrical field distribution in ONO layer after programming at gate bias of 15V for 100ms.

In Fig. 3.12(a), good high-temperature retention characteristics (85 °C) after respective programming ($V_g = 15$ V for 1 ms) and erasing ($V_g = -16$ V for 1 ms) can be observed for both devices with SiNW dimensions (20 and 50 nm). The P/E window for the device with smaller SiNW dimension degrades faster than that with larger dimension, which can be related to its stronger electrical field at the four corners. Despite that, $\sim 70\%$ of its original P/E window value can still be maintained for ten years. It is also believed that, with the engineering of the trapping layer (e.g., with the incorporation of Si or metal nanocrystal), the retention characteristics can also be significantly improved [3.8]. In Fig. 3.12(b), the device endurance is measured at programming ($V_g = 15$ V for 1 ms) and erasing ($V_g = -16$ V for 1 ms). Both devices with different dimensions can maintain their P/E window after 10^4 cycles at 25 °C.

Fig. 3.12: Vertical SiNW (with dimensions of 20 and 50 nm) SONOS memory ($L_g \sim 150$ nm). (a) Retention characteristic at 85 °C. (b) Endurance characteristic up to 10^4 cycles at 25 °C.

Next, the feasibility on the multi-bit programming for the device with small SiNW dimension (20 nm) is shown in Fig. 3.13. Benefiting from its sufficiently large program window and good process control, the memory cell is able to store 2 bits per cell using four states labeled as “00,” “01,” “10,” and “11,” with each state defined by a different programmed V_{th} . These different programmed V_{th} s were obtained by programming the same device with various voltages at 12, 14, and 16 V for 1 ms. Note that the multi-bit programming condition is yet to be optimized. Fig. 3.13 (b) shows the excellent cumulative distribution of the programmed V_{th} for ten randomly selected devices with the same SiNW dimension (20 nm) and gate length (150 nm) on the same wafer, which indicates the good process control and the reliability for multi-bit data storage [3.9].

Fig. 3.13: Vertical SiNW (with dimension of 20 nm) SONOS memory ($L_g \sim 150$ nm). (a) Multi bit programming with gate biases of 12, 14, and 16 V for 1 ms. During the first read operation, “00” and “01” states can be distinguished from “10” and “11” states at 4 V. After that, “00” from “01” and “10” from “11” at 2.4 and 5.8 V, respectively, are performed for the second read operation. (b) Cumulative distribution of multi-bit programmed V_{th} with gate biases of 12, 14, and 16 V for 1 ms for randomly selected ten devices.

3.3.2 Vertical SiNW Memory with Si-NCs Trap Layer

In this section, the effect of embedding Si-NCs as trapping layer in vertical SONOS type memory is demonstrated and the underlying physics is discussed. Fig. 3.14 compares the P/E speeds of SOnCOS with the control SONOS device. The NW diameter of both devices is kept as 50 nm. To avoid potential impact of trap sites occupation, the two devices were preset to the same V_{th} before the P/E characterization. As shown in Fig. 3.14, the SOnCOS memory shows larger ΔV_{th} for the same P/E time period, indicating a better charge storage capability. The results indicate that the

trapping centers in devices using silicon nanocrystals as charge trapping medium are superior than that in Si_3N_4 , which enhances the trapping efficiency of the devices. On the other hand, as it can be seen in Fig. 3.15, severe window closure is observed for vertical SONOS with diameters ($\Phi=50\text{nm}$) after 10^5 seconds 85°C retention test in which the device was programmed at 15V for $100\mu\text{s}$ and erased at -18V for 1ms . The electrical field formed by the electrons that are already trapped in the trapping layer during programming process will tend to help the trapped electrons gradually leaking out through the tunnel and blocking oxide and hence changes the programmed/erased V_{th} . Incorporating Si-NCs into the trap layer can significantly improve the charge loss by forming isolated energy wells due to conduction band offset between Si and dielectrics [3.10]. As shown in Fig. 3.15, after Si-NCs incorporation, the retention characteristics show significant improvement and negligible memory loss was observed after 10^5 seconds retention at 85°C . Fig. 3.16 shows the endurance properties of the SONOS and SOncOS devices. Negligible V_{th} shifting up and window closure were observed for both devices.

Fig. 3.14: Transient memory characteristics of SiNW SOncOS device and SONOS control device at +15/-16 V pulse stresses. Devices with a gate length of L_g 120nm and NW diameter 50nm are used for characterization.

Fig. 3.15: Retention characteristic at 85°C of vertical SiNW SONOS and SOncOS memory (L_g -120 nm, and diameter=50nm).

Fig. 3.16: P/E endurance characteristic at 85°C of vertical SiNW SONOS and SOncOS memory ($L_g \sim 120$ nm, and diameter=50nm).

3.4 Conclusion

In this chapter, we have demonstrated a multi-bit programmable GAA SONOS flash memory cell using the vertical squarish SiNW channel dimension down to 20 nm. The devices with smaller SiNW dimension show much enhanced P/E speed, which is originated from the increased electric field at the Si channel/SiO₂ interface, as well as the reduced tunneling current through the blocking oxide. On the other hand, the vertical SiNW memory devices with Si nanocrystals (Si-NCs) trap layer show significant performance improvements on P/E speed and retention characteristics over the SONOS devices.

Reference

- [3.1] S. Lai, “Flash memories: Successes and challenges”, *IBM J. Res. Develop.*, Vol. 52, pp. 529–535 (2008).
- [3.2] K. Kim, “Future memory technology: Challenges and opportunities”, *Proc. VLSI-TSA*, pp. 5–9 (2008).
- [3.3] K. Prall, “Scaling non-volatile memory below 30 nm”, *Proc. Non-Volatile Semicond. Memory Workshop*, pp. 5–10 (2007).
- [3.4] M. C. Chen, H. Y. Yu, N. Singh, Y. Sun, N. S. Shen, X. H. Yuan, G. Q. Lo, and D. L. Kwong, “Vertical Si nanowire SONOS memory for ultra-high density application”, *IEEE Electron Device Lett.*, Vol. 30, pp. 879–881 (2009).
- [3.5] T. H. Hsu, H. T. Lue, E. K. Lai, J. Y. Hsieh, S. Y. Wang, L. W. Yang, Y. C. King, T. Yang, K. C. Chen, K. Y. Hsieh, R. Liu, and C. Y. Lu, “A high-speed BE-SONOS NAND flash utilizing field enhancement effect of FinFET”, *IEDM Tech. Dig.*, pp. 913–916 (2007).
- [3.6] J. Fu, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, “Si-nanowire based gate-all-around non-volatile SONOS memory cell”, *IEEE Electron Device Lett.*, Vol. 29, pp. 518–521 (2008).
- [3.7] S. Kim, S. J. Baik, Z. Huo, Y. J. Noh, C. Kim, J. H. Han, I. S. Yeo, U. I. Chung, J. T. Moon, and B. I. Ryu, “Robust multi-bit programmable flash memory using a resonant tunnel barrier”, *IEDM Tech. Dig.*, pp. 861–864 (2005).
- [3.8] J. Sarkar, S. Dey, D. Shahrjerdi, and S. K. Banerjee, “Vertical flash memory cell with nanocrystal floating gate for ultradense integration and good retention”, *IEEE Electron Device Lett.*, Vol. 28, pp. 449–451 (2007).

- [3.9] M. Taub, K. Pangal, C. Abraham, M. Wang, H. Nguyen, J. Coulter, T. Begley, and S. Soss, “90 nm multi-level-cell flash memory technology”, *Proc. IEEE ISSM*, pp. 197–199 (2005).
- [3.10] H. I. Hanafi, S. Tiwari, and I. Khan, “Fast and long retention-time nano-crystal memory”, *IEEE Trans. Electron Device*, Vol. 43, pp. 1553-1558 (1996).

Chapter 4

Junction-Less Stackable SONOS Memory

Realized on a Vertical-Si-Nanowire

4.1 Introduction

Market of the NVM has boomed persistently due to increasing demand of the portable electronic devices. The GAA SiNW FET is being widely investigated for its potential to advance the CMOS technology to the extreme scaling limits. In the last chapter, we demonstrated a GAA SONOS flash memory on a vertical SiNW of diameter of 20 nm as the channel [4.3], the fabricated cell was considered to be an important building block for 3-D multi-level integration for ultra-high density application. Vertical stacking of memory cells, on the other hand, could be very challenging when forming junctions on NWs between the cells. In this chapter, we present a novel junction-less (JL) vertical-SiNW-based SONOS memory to make vertical stacking of memory cells much easier. A systematic study on JL-SONOS memory is presented from the process complexity, device performance and reliability perspective. The effect of channel doping, wire diameter and the underlying physics are discussed and analyzed through

device simulation. Being free of doped junctions, the presented JL memory device makes vertical SiNW a suitable platform for vertical stacking of memory cells to achieve ultra-high density.

On the other hand, physical multi-bit storage in nitride based localized charge trapping flash memory by using channel hot electron (CHE) and band-to-band hot hole injection (BTBHHI) have received much interest for simpler fabrication process, no drain turn on and no erratic bits [4.4]. However, as the channel length shrinks aggressively, severe punch through leakage and interference between the two bits can be found since the required drain voltages for the generation of hot carriers as well as the reverse read voltage cannot scale down accordingly [4.5-4.6]. This chapter further reports a novel JL vertical-wire-based SONOS memory with two physical storage nodes per cell where the charges are separately stored in the nitride layer near the top/bottom portion of the vertical-wire channel. Systematic evaluation of the memory cell from the P/E speed, endurance, and retention to the second bit effect is presented. The vertical channel structure relaxes the channel length scaling limitation and makes this device a highly reliable memory cell for multi-bit storage.

4.2 Junction-less Vertical SiNW SONOS Device Fabrication

Fig. 4.1 compares the process integration flow of junction-based and JL vertical SiNW SONOS memory. Unlike the junction-based device described in Chapter 3, the JL-SONOS process was started with phosphorus implantation on p-type Si slices followed by furnace annealing to create a uniform doping profile with different concentrations in the upper portion of the substrates [Fig.4.1 (b-1)]. After vertical SiNW formation (with various diameters from 100nm to 20nm) using a top-down approach [Fig.4.1 (b-2)], 1st HDP deposition and etch back, the tunneling oxide (5nm), the

LPCVD SiN trap layer (7nm) and the blocking oxide (7nm) were directly grown/deposited. After 80nm LPCVD poly silicon control-gate deposition followed by p-type doping through boron implantation and activation to obtain desirable values for the threshold voltage of the n-channel device, the gate stack was patterned and etched to define the gate contact pad [Fig.4.1 (b-3)]. The gate length was defined by HDP oxide deposition and etch-back process, followed by tip poly-Si removal in Tetramethylammonium hydroxide (TMAH) solution [Fig.4.1 (b-4)]. Another HDP deposition and etch-back was performed to create the gate-to-drain isolation. After the wire top exposition and ONO dry etch, contact holes to the wire bottom and gate pads were created and TaN/Al/TaN was used as the metal contact [Fig.4.1 (b-5)]. Alloying was done at 420 °C in 10% H₂ to conclude the process. According to the JL concept, no source/drain implant was performed, thus reducing process complexity and cost as compared to the junction-based one [4.5]. More than that, the absence of junctions on wire allows easy stacking of multiple cells on a NW. It may be worth mentioning here that this device can be fabricated with very low thermal budget by using a moderately doped wafer to eliminate the 1st implant and by replacing the un-doped gate poly by doped poly or metal gate.

Fig. 4.1: (a) Process integration flow of vertical SiNW junction based SONOS memory: (a-1) vertical SiNW formation; (a-2) As⁺ is implanted to form the source; (a-3) gate stack deposition and gate pad definition; (a-4) tip poly and ONO removal, followed by drain implant; (a-5) metallization. (b) Process integration flow of vertical JL-SONOS: (b-1) bulk implant and annealing; (b-2) vertical SiNW formation; (b-3) gate stack deposition and gate pad definition; (b-4) tip poly and ONO removal; (b-5) metallization.

Shown in Fig. 4.2 (a) is a cross-sectional TEM image of a vertical SiNW GAA JL-SONOS with wire diameter equals to 20 nm and gate length of ~120nm. All wire diameters were taken near the drain end which is the narrowest part of the wire. Fig. 4.2 (b) shows the cross-sectional TEM image of a well-defined gate stack ONO thickness 5/7/7 nm and moderately-doped SiNW channel ($1 \times 10^{19} \text{ cm}^{-3}$).

Fig. 4.2: (a) Cross-sectional TEM image of vertical SiNW GAA JL-SONOS: diameter of wire is 20 nm and gate length 120nm. (b) Cross-sectional TEM image of gate stack showing ONO thickness 5/7/7 nm.

4.3 Results and Discussion

4.3.1 Junction-less Vertical SiNW SONOS using F-N tunneling

JL devices with various wire diameters and dopant concentrations were used for characterization. For all the memory devices characterized in this section, F/N tunnelling, pulse applied on the gate with Source/Drain grounded, is used to program and erase the cells. The threshold voltage (V_{th}) is obtained from I_d - V_g characteristic with drain current (I_d) equal to 10^{-7} A/ μ m. Unless specified otherwise, the data reported in this work is obtained from devices after forming gas anneal.

The measured I_d - V_g characteristics of the memory cell with different NW diameters and fixed channel doping of 10^{19} cm⁻³ are shown in Fig. 4.3 (a). Though the junction-based and JL devices work with different switching mechanism, the JL-SONOS shows comparable I_d - V_g characteristics to junction-based SONOS with sub-threshold swing < 70 mV/decade and off current $< 1 \times 10^{-12}$ A. This is attributed to the scaled wire channel with excellent gate control owing to the GAA structure. The normally OFF behavior of the device is a result of p^+ gate where high work function of the gate poly relative to n-channel depletes the channel electrons. The increase in V_{th} with reduction in channel diameter is a result of deep depletion in narrow wires. Indeed, the potential drop across the cylindrical NW increases quadratically with the NW radius and gives an increasing negative contribution to the threshold voltage [4.7], [4.8]. Further from Fig. 4.3 (a), it can also be inferred that, the switching behavior of JL device is dominated by the narrowest part of the wire. The slight increase in drain current with increasing negative gate voltage is a result of the band-to-band tunnelling resulting from large electrical field across the floating body/drain junction [4.19].

Fig. 4.3: Vertical SiNW JL-SONOS memory ($L_g \sim 120$ nm, channel doping= $1 \times 10^{19}/\text{cm}^{-3}$): I_d - V_g characteristics for devices with different wire diameters from 20 nm to 60 nm (a) after forming gas anneal (415 °C, 45 mins, 10% H_2) and (b) before forming gas anneal.

It may be worthy pointing out that, unlike conventional SONOS memory, as a result of relatively high channel doping (up to 10^{19}cm^{-3}) the JL SONOS is more sensitive to forming gas anneal process as more Si dangling bonds are formed at the Si-SiO₂ interface. Fig. 4.3 (b) depicts the I_d - V_g characteristics of the same memory cells before performing forming gas anneal. Remarkable improvement in SS after forming gas anneal can be observed, which due to the interface passivation.

Figs. 4.4(a) and 4.4(b) show the typical I-V curves during P/E at $\pm 16\text{V}$ of JL-SONOS with 20nm wire and channel doping of $1 \times 10^{17}\text{cm}^{-3}$. It can be seen that all the I-V curves shift in parallel during P/E.

Figs. 4.5(a) and 4.5(b) illustrate the P/E characteristics of JL-SONOS with channel doping of $1 \times 10^{17}\text{cm}^{-3}$ and $1 \times 10^{19}\text{cm}^{-3}$, respectively, for a fixed wire diameter of 20 nm. The measured fresh V_{th} of the two devices are 0.9V for $1 \times 10^{17}\text{cm}^{-3}$ and 0.7V for $1 \times 10^{19}\text{cm}^{-3}$, respectively. Due to the small threshold change, it is reasonable to admit that even "virgin" devices contain a certain amount of trapped charge within the trapping layer, which partially compensates the depleted charge in the channel. For both the doping concentrations the memory window is nearly same, more precisely, 3.2V for low doped ($1 \times 10^{17}\text{cm}^{-3}$) and 2.7 V for moderately doped ($1 \times 10^{19}\text{cm}^{-3}$), when a P/E time of 1 ms was used at $\pm 15\text{V}/\pm 16\text{V}$, respectively. However, there is an impact of channel doping on the P/E speeds: P is faster for high doping while E is faster for low doping. Device simulation was done to understand the doping impact through a home-made code. The Schrödinger-Poisson equations considering quantum-mechanical confinement effects were self-consistently solved within a cylindrical domain for the SiNW-based JL SONOS memory. Shown in Fig. 4.5(c) is the simulated E-field across the device from channel center to gate. It is seen that the E-field across the tunnelling oxide becomes larger with increased channel doping under the same gate voltage. The large vertical

electric field at the Si-SiO₂ interface and depressed electric field in the blocking oxide in addition to high conduction band electron density at the channel surface from the heavily doped channel enhance the carrier-injection probability, and thus increases the programming speed. On the other hand, the slower erase with high doping is a result of the low hole density within the channel and the increased hole tunnelling barrier height [4.10]. In fact, for SONOS memory both hole tunnelling and electron detrapping may contribute to the erasing current depending on the band alignment conditions. The results shown in Fig 4.5 (b) imply that the hole tunnelling mechanism dominates the erase current at these pulse conditions. It is worth pointing out that all the JL-SONOS devices have a p-type control gate, which increases the electron barrier height at the control-gate/blocking-oxide interface during cell erasing, as compared to the case of the n-type control gate, which helps preventing the unwanted gate electron injection and thus improve the erase characteristics [4.11]. It should be also noted that the asymmetry of the P/E process is also believed to partially stem from the lower tunneling probability of holes compared with that of electrons used for programming.

Fig. 4.4: Typical I_d - V_g curves of JL-SONOS with 20nm SiNW and channel doping of $1 \times 10^{17} \text{ cm}^{-3}$ during (a) programming and (b) erasing at ± 16 V with pulse widths from 100ns to 1ms.

Fig. 4.5: (a) Programming and (b) Erasing characteristics using F-N tunneling of vertical-SiNW-based GAA JL-SONOS memory device with channel doping of $1 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{19} \text{ cm}^{-3}$, initial program, 16V/1ms for lightly doped ($10^{17}/\text{cm}^{-3}$) and 15V/1ms for moderately doped ($10^{19}/\text{cm}^{-3}$) devices, was done on fresh devices to get the same programmed V_{th} before erasing. (c) Simulated electric field distribution of the SiNW-based JL-SONOS at $V_g=4 \text{ V}$ with various channel doping of $10^{17} \text{ cm}^{-3} / 10^{19} \text{ cm}^{-3}$ and a wire diameter of 20nm.

The feasibility on the multi-bit programming for the JL-SONON with small SiNW dimension (20nm) is shown in Fig. 4.6. The memory cell is able to store 2 bits per cell using four states labeled as “00,” “01,” “10,” and “11,” with each state defined by a different programmed V_{th} . These different programmed V_{th} s were obtained by programming the same device with various voltages at 12, 14, and 16 V for 1 ms.

Fig. 4.6: Vertical SiNW (with dimension of 20 nm) JL-SONOS memory ($L_g \sim 120$ nm, channel doping= $1 \times 10^{19}/\text{cm}^{-3}$): multi-bit programming with gate biases of 12, 14, and 16 V for 1 ms.

The high temperature retention characteristic at 85°C for JL-SONOS is shown in Fig. 4.7 (a). The JL-SONOS with high channel doping ($1 \times 10^{19} \text{ cm}^{-3}$) exhibits less V_{th} degradation as compared to lightly doped JL-SONOS ($1 \times 10^{17} \text{ cm}^{-3}$), and its memory window can be well maintained up to 10^5 sec. Under retention conditions, direct trap-to-band (TB) tunnelling from a Si_3N_4 related traps to the channel conduction band is the main discharge mechanism. With increased channel doping, the energy states close to the channel conduction band are more likely to be occupied by electrons. This reduces the probability of further electron injection from the trap layer, and thus more reliable charge storage can be expected [4.12]. The cycling characteristic of JL-SONOS is shown Fig. 4.7 (b), in which all devices can maintain the P/E window after 10^5 cycles at 85°C . The excellent endurance properties indicate that the interface trap generation which is caused by high energy electrons and holes in JL-SONOS is

negligible during these P/E cycles.

Fig. 4.7: (a) Retention and (b) Endurance characteristic at 85°C of JL-SONOS memory device with channel doping of $1 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{19} \text{ cm}^{-3}$.

4.3.2 Junction-less Vertical SiNW SONOS with 2-bit Storage/Cell

In this section, we report a novel JL vertical-wire-based SONOS memory with two physical storage nodes per cell where the charges are separately stored in the nitride layer near the top/bottom portion of the vertical-wire channel. Shown in Fig. 4.8 is the schematic of the vertical SiNW GAA JL-SONOS where charges are separately stored in the nitride layer above the top/bottom portion of the vertical-wire channel.

Fig. 4.8: Schematic of the vertical SiNW GAA JL-SONOS, the charges are separately stored in the nitride layer above the top (T-bit)/bottom (B-bit) portion of the vertical-wire channel.

JL-SONOS devices with gate length $L_G = 120$ nm and wire diameters 20nm were used for characterization. The threshold voltage (V_{th}) is defined as the gate voltage when drain current is $0.1\mu A/\mu m$ at a reverse read voltage of 1.6V. Since the aggressive scaling of cell area makes the conventional channel hot electron injection method unsuitable for mass data storage application due to its high power consumption, in this work, a recently developed low power P/E scheme [4.13] is used to operate the device where channel F-N erase is performed to raise the V_{th} while programming is done

through band-to-band hot hole (BTBHH) injection. To read the top (bottom) bit, a source (drain) bias is applied to reduce the channel potential near the bottom (top) bit. The operation biases are summarized in Table 4.1.

		P	E	R
Top Bit	V _g	-5V	15V	-
	V _d	5V	-3V	0V
	V _s	0V	-3V	1.6V
Bottom Bit	V _g	-5V	15V	-
	V _d	0V	-3V	1.6V
	V _s	5V	-3V	0V

Table 4.1: Operation conditions of a junction-less vertical-SiNW-based SONOS memory with 2-bit storage.

Figs. 4.9 (a) and 4.9 (b) illustrate the P/E characteristics of JL-SONOS with channel doping of $1 \times 10^{19} \text{ cm}^{-3}$ and a fixed wire diameter of 20 nm. In Fig. 4.9 (a), F-N electron injection is used to raise the V_{th} of both top bit (T-bit) and bottom bit (B-bit) to high state. As F-N tunneling results in a uniform trapped charge concentration along the vertical channel, no difference in the erase speed is observed for the top and bottom bits under reverse read scheme [4.4]. In Fig. 4.9 (b), program is done by lowering the local V_{th} of the top and bottom bits through edge BTBHH injection. In more details, a negative gate bias is used to invert the channel region of a n-type JL device and forms drain (source) to channel junction at the top (bottom) edge of the poly gate. The channel hot holes are then generated by high lateral electrical field and subsequently trapped locally in the nitride layer with the help of negative V_g . As shown in Fig. 4.9 (b), 3V program window can be achieved for both the top and bottom bit within 5 μsec . However, comparing with the top bit, the bottom bit has a slightly faster program speed

which can be related to the tapered NW channel, leading to asymmetric electrical field distribution when program the top/bottom bit. It is believed that the difference in program speed can be minimized after process optimization to create more straight vertical wires.

Fig. 4.10 depicts the I-V characteristic for the JL-SONOS device in different V_{th} states. The top and bottom bits are first erased to high V_{th} states using F-N tunneling. When only the bottom bit is programmed using edge BTBHH injection, hot holes are injected locally and recombine with those trapped electrons near the source side. Under reserve read scheme, the drain voltage used to read the bottom bit plays the role of maximizing the influence of source-side injected charges on channel current by forming the depletion region on the drain side. Subsequently, the second bit effect (SBE) is observed by reading the top bit. This V_{th} shift due to SBE reflects the influence of source-side trapped charges on the channel current [4.14]. Slight SS degradation is also observed due to the localized charge storage, leading to reduced gate to body capacitance and hence, less control over the channel [4.15].

Fig. 4.9: (a) Erase speed of top bit (T-bit) and bottom bit (B-bit), (b) program speed of top and bottom bits.

Fig. 4.10: I_D - V_G characteristic memory cell in different V_{th} states, SBE indicates the amount of second bit effect.

Fig. 4.11 shows the 2-bit programming characteristics in which the bottom bit is first programmed to low V_{th} state while the top bit is kept at high state. Then the top bit is programmed to low V_{th} state while keeping the bottom bit at low state. The results indicate that one bit almost does not affect the information of the other bit due to the narrow storage region and the reverse read voltage is sufficient to shield the stored charges during forward scanning.

Fig. 4.11: 2-bit programming characteristics of JL-SONOS Memory.

In Fig. 4.12 (a), the retention characteristic of the JL device for both virgin cell (solid) and 10K cycled cell (open) is presented. It is found that the window can be well maintained after 10⁵ sec at 85 °C for both fresh and cycled cells. Fig. 4.12 (b) shows the device endurance characteristic up to 10⁵ cycles. No obvious V_{th} upward shift is observed, which can be further verified through the I-V curves of a programmed cell after various P/E numbers. The non-degraded SS indicates minimal damage or build-up of charges in the ONO layer after cycling.

Fig. 4.12: (a) Retention characteristic: a virgin cell and 10K cycled cell, solid: fresh cell, open:10K cycled cell. (b) endurance characteristic: with IV curves of a programmed cell at various cycles as the inset. Cycle sequences: erase both bits->program B-bit ->program T-bit.

4.4 Conclusion

In this chapter, the JL-SONOS memory realized on vertical SiNW GAA structure with channel dimension down to 20nm have been demonstrated for the first time. Manufacturability aspects, device performance and reliability perspectives are systematically addressed. The vertical SiNW based JL-SONOS with less manufacturing cost and complexity is considered to be a promising candidate towards 3-D multi-level integration for ultra-high density applications. The integration process of the 3-D memory based on vertical SiNW JL-SONOS would be significantly simplified due to the absence of junctions for both the selection transistors and the memory cells.

On the other hand, a low power BBHHI programming scheme is used for the JL-SONOS memory on vertical SiNW GAA structure to realize 2-bit storage/cell. After reviewing the second bit effect, P/E speed, endurance and retention, robust 2-bit/cell storage has been achieved by storing charges locally in the two physical storage nodes near the top/bottom of the vertical SiNW channel. The relaxed channel length limitation due to the vertical structure provides more room for device tuning to overcome the scaling –related reliability issues, and the absence of junctions makes this device highly manufacturable with low cost and thermal budget.

Reference

- [4.1] J. D. Lee, S. H. Hur, and J. D. Choi, “Effects of floating-gate interference on NAND flash memory cell operation”, *IEEE Electron Device Lett.*, Vol. 23, pp. 264-266 (2002).
- [4.2] N. Singh, A. Agarwal, L.K. Bera, T.-Y. Liow, C.H. Tung, R. Kumar, G.Q. Lo, N. Balasubramanian, and D.L. Kwong, “ High performance fully-depleted silicon nanowire (diameter ≤ 5 nm) gate-all-around CMOS devices” , *IEEE Electron Device Lett.*, Vol. 27, pp. 383-386 (2006).
- [4.3] Y. Sun, Y. H. Yu, N. Singh, N. S. Shen, G. Q. Lo and D. L. Kwong, “Multi-bit programmable flash memory realized on vertical Si nanowire with a highly scaled squarish channel”, *IEEE Electron Device Lett.*, Vol. 31, pp. 390-392 (2010).
- [4.4] B. Eitan, P. Pavan, H. Bloom, E. Aloni, A. Frommer, and D. Finzi, “NROM: A novel localized trapping, 2-bit non-volatile memory cell”, *IEEE Electron Device Lett.*, vol. 21, pp. 543-545 (2000).
- [4.5] S. H. Gu, C. W. Hsu, T. Wang, W. P. Lu, Y. H. J. Ku, and C. Y. Lu, “Numerical simulation of bottom oxide thickness effect on charge retention in SONOS flash memory cells”, *IEEE TED*, Vol. 54, pp. 90-97 (2007).
- [4.6] Y. W. Chang, T. C. Lu, S. Pan, and C. Y. Lu, “Modeling for the 2nd-bit effect of a nitride-based trapping storage flash EEPROM cell under two-bit operation”, *IEEE Electron Device Lett.*, Vol. 25, pp. 95-97 (2004).
- [4.7] T. H. Hsu, H. T. Lue, E. K. Lai, J. Y. Hsieh, S. Y. Wang, L. W. Yang, Y. C. King, T. Yang, K. C. Chen, K. Y. Hsieh, R. Liu, and C. Y. Lu, “A high-speed BE-SONOS NAND flash utilizing field enhancement effect of FinFet”, *IEDM Tech. Dig*, pp.913-916 (2007).

- [4.8] J. Fu, N. Singh, K. D. Buddharaju, S. H. G. Teo, C. Shen, Y. Jiang, C. X. Zhu, M. B. Yu, G. Q. Lo, N. Balasubramanian, D. L. Kwong, E. Gnani, and G. Baccarani, "Si-nanowire based gate-all-around nonvolatile SONOS memory cell", *IEEE Electron Device Lett.*, Vol. 29, pp. 518-521 (2008).
- [4.9] P. C. Adell, H. J. Barnaby, R. D. Schrimpf, B. Vermeire, "Band-to-band tunneling (BBT) induced leakage current enhancement in irradiated fully depleted SOI devices", *IEEE TNS*, Vol. 54, pp.2174-2180 (2007).
- [4.10] C. Shen, J. Pu and M. F. Li, "P-type floating gate for retention and P/E window improvement of flash memory devices," *IEEE Trans. Electron Device*, Vol. 54, pp.1910-1916 (2007).
- [4.11] S. Jeon, J. H. Han, J. H. Lee, S. Choi, H. Hwang, and C. Kim, "High work-function metal gate and high-k dielectrics for charge trap flash memory device applications", *IEEE Trans. Electron Device*, Vol. 52, pp. 2654-2658 (2005).
- [4.12] Y. Yang and M. H. White, "Charge retention of scaled SONOS nonvolatile memory devices at elevated temperatures", *Solid State Electronics*, Vol. 44, pp. 949-958 (2000).
- [4.13] C. C. Yeh, W. J. Tsai, M. I. Liu, S. K. Cho, C. J. Lin, T. Wang, S. Pan, and C. Y. Lu, "PHINES : A novel low power program/erase, small pitch, 2-bit per cell flash memory", *IEDM Tech. Dig*, pp.931-934 (2002).
- [4.14] T. H. Hsu, M. H. Lee, J. Y. Wu, H. L. Lung, R. Liu, and C. Y. Lu, "Investigation of maximum current sensing window for two-side operation, four-bit/cell MLC nitride-trapping nonvolatile flash memories", *IEEE Electron Device Lett.*, Vol. 25, pp. 795–797 (2004).
- [4.15] B. Tomar and V. R. Rao, "Sub-threshold swing degradation due to localized charge storage in SONOS memories", *Proceedings of 11th IPFA 2004*, Taiwan.

Chapter 5

Poly-Si Vertical NW for TFT SONOS

Memory Application

5.1 Introduction

Poly-Si thin film transistors (TFTs) have attracted substantial attention because of their wide range of applications in active-matrix liquid crystal displays (AMLCDs) and active-matrix organic light emitting displays (AMOLEDs) [5.1]. The most important feature of TFTs is that they can be built on an insulating substrate, typically glass or fused quartz. The TFT is used as switch to charge and discharge the LCD cell's transparent electrodes during the pixel on and off. To realize the system-on-panel (SOP) technology, besides the pixel switching elements and logic circuits, NVM devices using poly-Si and embedded on glass, are crucial components for storing images and signals to increase the overall functionality and flexibility of circuits associated with the display [5.2]. The most common limitations with the TFT SONOS memories are the relatively low memory speed and poor subthreshold property, which are resulted from the grainy structure of poly-Si and the dangling bonds along the grain boundaries. To improve the performance in TFT logic or memory devices, a lot of efforts have been done to

increase the poly-Si grain size by additional processes such as excimer laser annealing [5.3] or by reducing the cross-sectional dimensions of the poly-Si channel to make those comparable to the grain size, such as NW-like channel devices [5.4].

Poly-Si lateral NW based TFTs have been demonstrated for the benefits of grain-boundary defects reduction and short-channel effects improvement owing to the GAA structure [5.5]. To further improve the bit density of embedded NVM of SOP, various 3-D high-density NAND memories based on multilayer-stack poly-Si thin-film transistors (TFTs) have been proposed. To be implemented into 3-D circuits [5.6], poly-Si vertical NW TFTs are considered to be an ideal candidate owing to its nature of 3-D architecture.

In this chapter, we first demonstrated GAA vertical poly-Si NW TFTs and inverter circuits based on this technology. Inverter circuits show well-behaved transfer characteristics under the device dimension of N- to P-TFTs ratio of 1:1. Next, JL vertical poly-Si NW SONOS devices are also fabricated and electrically characterized. The TFT SONOS memory device with the use of narrow channel body exhibits better device performance as compared to other TFT NVM devices, which is owing to the reduced number of grains and grain boundaries in the channel region.

5.2 Vertical Poly-Si NW TFT and SONOS Device Fabrication

The device fabrication follows our previous study [5.7], and the device SEM inspections at several key fabrication processes are shown in Figs. 5.1(a)–(d). Firstly, ~100 nm thick thermal oxidation was performed on an 8-inch Si substrate to serve as the starting substrate in this work, which mimics the glass substrate in the AMLCD manufacturing. This was followed by ~700 nm poly-Si deposition using low-pressure chemical vapor deposition (LPCVD). After litho patterning, poly-Si vertical wires were

formed using plasma etching in combination with TMAH wet-etch to obtain the wire diameter ranging from 30 nm to 140 nm (Fig. 5.1(a)). It should be noted that the high surface roughness of poly-Si NW after TMAH etch can degrade the device performance and yield, hence further process optimization is needed to fabricate small poly-Si nanowires.

After source implantation (P and BF₂ implants were done alternately for NFET and PFET regions), HDP oxide was deposited and etched back to form the gate-source isolation. In the next step, 4.5-nm gate oxide and 50-nm poly-Si electrode were deposited. HDP oxide deposition and etch back process was used again to expose the wire tip covered with the poly-Si electrode and SiO₂ dielectrics. The exposed poly-Si was then etched using TMAH solution (Fig. 5.1(b)). The second HDP thickness together with gate poly-Si thus defined the gate length. The nitride spacer was then formed surrounding the top of the wires which protected the gate oxide while removing the HDP oxide in DHF and the channel during the following drain implantation process. After gate-extension pad pattern transfer, P and BF₂ implants were again done alternately for the respective NFET and PFET regions, at a tilt of 15° from four orthogonal directions to form drain junctions as well as doping the poly-Si gate. Finally, a layer of HDP oxide was deposited before contact and metal formation using aluminum (Fig. 5.1(c)). The final inverter is shown in Fig. 5.1(d). The cross-sectional TEM image shown in Fig. 5.1(e) was taken from a device with a wire diameter of ~90 nm and a gate length of ~100 nm, showing the thin film average poly-grain size of ~100 nm.

Fig. 5.1: (a) SEM image after pillar trimming, (b) SEM image after gate patterning, (c) SEM image of NFET device, (d) SEM image of an Inverter, (e) TEM image of NMOS device with wire diameter ~90 nm and gate length ~100 nm.

On the other hand, to fabricate the vertical poly-Si NW thin-film SONOS memory devices, we adopted the JL concept as discussed in Chapter 4 on the poly-Si NW platform. The absence of junctions on poly-Si wire makes stacking of multiple cells on a NW much easier and thus, promising for embedded ultra-high density memory application. The JL vertical NW TFT SONOS process was also started with 100 nm thick thermal oxidation on an 8-inch Si substrate to serve as the glass substrate followed by 700 nm amorphous Si deposition using LPCVD. Next, phosphorus was implanted at 90° followed by furnace annealing to create a uniform doping profile within the top Si layer ($1 \times 10^{18} \text{ cm}^{-3}$). The amorphous silicon was crystallized during the annealing process. After that it follows the process flow of single crystalline SiNW JL-SONOS as discussed in Chapter 4 to form the wire channel, gate pad and eventually metallization. Shown in Fig. 5.2 is a cross-sectional TEM image of a JL vertical poly-Si NW TFT SONOS with wire diameter equals to 90 nm and gate length of ~100nm. All wire diameters were taken near the drain end which is the narrowest part of the wire.

Fig. 5.2: TEM image of junction-less TFT SONOS device with wire diameter ~90 nm and gate length ~100 nm.

5.3 Results and Discussion

5.3.1 Demonstration of Vertical Poly-Si NW TFTs and inverter

The I_{DS} - V_{GS} curves of the individual vertical NW poly-Si NFET and PFET (with L_g of ~ 100 nm) measured from a single inverter device are plotted in Fig. 5.3 (a). The linear threshold voltages (V_{th}) of N- and P-TFT were extracted to be 0.45 V and 0.087 V, respectively. Both N- and P-TFT devices exhibit good performance with fast turn on SS of ~ 80 – 100 mV/decade, low DIBL of ~ 20 – 60 mV/V, and high I_{on}/I_{off} ratio of $\sim 10^7$. It might be worth mentioning that under small drain bias ($V_{ds} < 0.4$ V), due to the unoptimized S/D contact resistance of PFET, most of the drain bias is dropped across the S/D area. When increasing the gate bias, the drain current becomes saturated after the channel resistance is comparable to the S/D contact resistance.

Fig. 5.3: (a) I_d - V_g characteristic of individual NMOS and PMOS of the inverter (diameter ~ 30 nm, $L_g \sim 100$ nm), (b) I_d - V_d characteristic of individual NMOS & PMOS of the inverter (diameter ~ 30 nm, $L_g \sim 100$ nm), (c) impact of SiNW diameter on device SS and DIBL ($L_g \sim 100$ nm). In total, about 20 devices were measured, and the error bar is estimated by the spread of the data.

Fig. 5.3 (b) illustrates the device I_{DS} – V_{DS} curves. The measurement is carried out at constant V_g – V_{th} ranging from 0 to 1.2 V with a step size of 0.3 V. The results reveal comparable matched drive current for both N- and P-TFT devices even both devices have a dimension ratio (L_g/L_w) of 1:1. The SiNW diameter impact on device performance (DIBL and SS) is investigated and shown in Fig. 5.3 (c). These measurements are carried out on the devices with the same gate length of ~100 nm. It is seen that DIBL and SS are degraded when the wire diameter is increased, which is believed to stem from the reduced gate electrostatic control over the channel. Overall, the result suggests the benefit of scaled NW dimension to the poly-Si NW TFTs performance.

Fig. 5.4: (a) Transfer characteristics of the inverter at different V_{DD} with V_{IN} and V_{OUT} plotted interchangeably on X and Y axes, (b) inverter switching current at different V_{DD} values, and (c) transfer characteristic of the inverter at different V_{DD} (ranging from 0.2 V to 1.2 V).

Fig. 5.4 (a) plots V_{out} versus V_{in} characteristics of the inverter, where the individual transistor performance is shown in Fig. 5.3 (a) ($L_g \sim 100$ nm). Owing to the matched driving current between N- and P-TFT, a butterfly shape is obtained, which is used to extract the noise margin of the inverter. The noise margins are calculated following the formulas $NM_H = V_{OH} - V_{IH}$ and $NM_L = V_{IL} - V_{OL}$, where the high and low input/output voltages are extracted from the inverter transfer curve. One can see that the NM_H is ~ 0.2 V and NM_L is ~ 0.45 V. At the same time, the inverter switching threshold V_M of 0.75 V is extracted. The result yields reasonable good NM_L , but reveals that NM_H needs to be improved. The low NM_H is likely to be due to the slow switching of N-TFT that caused a slow pull-down of the output voltage.

Switching currents as a function of input voltage are also measured, and plotted against V_{DD} in Fig. 5.4 (b). A reasonably low peak current (~ 280 nA) flows through the inverter during switching from high to low, or vice versa, at a V_{DD} of 1.2 V, which further reduces to ~ 2 nA when the V_{DD} goes down to 0.2 V. This is in contrast with the reported result of vertical SiNW-based CMOS inverter in [5.7, 5.8], which has a 400 nA of peak current at $V_{DD} = 1.2$ V.

The transfer characteristic of the inverter at various V_{DD} (ranging from 0.2 V to 1.2 V) is plotted in Fig. 5.4 (c). Again, owing to the matched driving current between N- and P-TFT, the inverter shows excellent behavior even when the V_{DD} is reduced to a value of 0.2 V. It should be noted that a more symmetric transfer curve can be achieved via the respective N- and P-TFT devices threshold-voltage modulation by incorporating advanced technologies, e.g., high k / metal gate.

5.3.2 Demonstration of JL Vertical Poly-Si NW TFT SONOS

JL vertical poly-Si NW TFT SONOS devices with gate length 100-nm were used for characterization in this section. F-N tunneling mechanism was employed to characterize the transient memory performance of the cell by grounding the source and drain and stressing the gate. The threshold voltage (V_{th}) is obtained from I_d - V_g characteristic with drain current (I_d) equal to 10^{-7} A/ μ m. In typical TFT devices, the subthreshold swing (SS) is usually beyond 200-mV/decade. Fig. 5.5 shows the transfer characteristics of the 30nm poly-Si NW device and low SS value of 129-mV/decade is obtained, which can be attributed to the enhanced gate control from the GAA structure and the reduced number of grain in the NW channel. Since the poly-Si NW width and height are comparable to or even smaller than the poly-Si grain size (50nm to 100nm), the number of grains and grain boundaries in the NW channel are expected to be remarkably reduced. The inset of Fig. 5.5 shows the SS for devices with different wire diameters. In total, about 30 devices were measured, and the error bar is estimated by the spread of the data. A degradation of SS in devices with larger NW width is clearly seen. The degradation of SS could be due to the reduced gate controllability and increased volume of grain boundaries in the channel region of the devices as the NW diameter increases.

Fig. 5.5: Transfer characteristics of junction-less vertical NW TFT SONOS device with 30nm wire channel and 100nm gate length. The inset shows the dependence of SS on the poly-Si NW diameter.

The memory P/E characteristics at various gate stressing voltages as the function of pulse width are presented in Fig. 5.6 (a) and (b). The poly-Si NW memory device exhibits a fast programming speed: a threshold voltage shift (ΔV_{th}) of 5.2 V could be achieved in 1 ms at 16 V gate pulse. In the inset of Fig. 5.6 (a), ΔV_{th} is plotted as a function of stress voltage for 1 ms programming time for the devices with different NW diameters ranging from 30nm to 85nm. It can be seen that the device with the narrowest wire diameter shows the fastest program speed. When the NW diameter is scaled to the value that is comparable to the ONO layer thickness (~ 20 nm), it can be realized that the geometry of the GAA NW device will make the E-field more concentrated near the silicon channel. The enhanced electric field in the tunnel oxide for smaller wire device intensifies the band bending of the tunnel barrier and thus under F-N tunneling mechanism the carriers experience shorter barrier width to tunnel through and faster memory speed is observed.

Fig. 5.6: The P/E characteristics of NW TFT SONOS with wire diameter of 30nm. The gate pulse stress was ranging from 13 V to 16 V and -13 V to -16 V. The inset of (a): comparison of memory window among devices with different NW diameters under different stress voltage for a program time of 1 ms.

The erasing time to achieve the same memory window is about 1 ms at -18 V. The relatively slow erasing speed is because of the higher hole energy barrier (4.6 eV) and larger effective mass of the hole than that of the electron in SiO₂, which results in reduced tunneling of the holes.

Fig. 5.7 shows the room temperature data retention of GAA vertical poly-Si NW SONOS memory. The high and low levels are extracted by applying 16V potential at gate for 100 μ s to program and -18 V for 1 ms to erase the cell. The extrapolated retention behavior shows the memory window degrades from an original 4.2V to 3.3 V at the end of ten years.

Fig. 5.7: Retention behavior of fresh GAA vertical poly-Si NW TFT SONOS memory at room temperature.

5.4 Conclusion

In this chapter, vertical poly-Si NW based TFTs (with wire diameter down to 30 nm and gate length down to 100 nm) and the inverter circuits are firstly demonstrated using a CMOS-compatible process. Respective N- and P-TFT devices show well-behaved performance: high I_{on}/I_{off} ratio, low DIBL and good subthreshold slope. The inverter based on the poly-Si NW TFTs (with dimension ratio between N- and P-TFT of 1:1) exhibits excellent performance even with V_{DD} being reduced to ~ 0.2 V, showing good transfer characteristics and low switching currents.

Next, JL GAA vertical poly-Si NW TFT SONOS memory with minimum wire diameter down to 30nm is demonstrated. Results reveal much improved device subthreshold characteristics and excellent memory performances with down scaling of poly-Si wire diameters. Being fabricated on poly silicon thin film, the GAA vertical NW TFT SONOS is promising for future TFT NVM application to be embedded into actual mobile display panel products for both the SOP and SOC applications.

Reference

- [5.1] S. Zhang, C. Zhu, J. K. O. Sin, J. N. Li, and P. K. T. Mok, “Ultra-thin elevated channel poly-Si TFT technology for fully-integrated AMLCD system on glass”, *IEEE Trans. Electron Device*, Vol. 47, pp. 569-575 (2000).
- [5.2] K. P. Anish Kumar, and J. K. O. Sin, “A simple polysilicon TFT technology for display systems on glass”, *IEDM Tech. Dig.*, pp. 515-518 (1997).
- [5.3] K. Kitahara, Y. Ohashi, and Y. Katoh, “Submicron-scale characterization of poly-Si films crystallized by excimer laser and continuous-wave laser”, *J. Appl. Phys.*, Vol. 95, pp. 7850-7855 (2004).
- [5.4] M. Im, J. W. Han, H. Lee, L. E. Yu, S. Kim, C. H. Kim, S. C. Jeon, K. H. Kim, G. S. Lee, J. S. Oh, Y. C. Park, H. M. Lee, and Y. K. Choi, “Multi-gate CMOS thin-film transistor with polysilicon nanowire”, *IEEE Electron Device Lett.*, Vol. 29, pp. 102-105 (2008).
- [5.5] C. M. Lee and B. Y. Tsui, “A 30nm gate-all-around poly-Si nano wire thin-film transistor”, *VLSI-TSA*, pp. 28-29 (2010).
- [5.6] A. O. Adan, K. Suzuki, H. Shibayama, and R. Miyake, “A half-micron SRAM cell using a double-gated self-aligned polysilicon pMOS thin film transistor (TFT) load”, *VLSI Symp. Tech. Dig.*, pp. 19–20 (1990).
- [5.7] B. Yang, K. D. Buddharaju, S. H. G. Teo, N. Singh, G. Q. Lo, and D. L. Kwong, “Vertical silicon-nanowire formation and gate-all-around MOSFET,” *IEEE Electron Device Lett.*, Vol. 29, pp. 791–794 (2008).
- [5.8] B. Yang, K. D. Buddharaju, S. H. G. Teo, J. Fu, N. Singh, G. Q. Lo, and D. L. Kwong “CMOS compatible gate-all-around vertical silicon-nanowire MOSFETs” ESSDERC 2008.

Chapter 6

Demonstration of NAND Memory Array with Stacked JL-SONOS

6.1 Introduction

The scaling of Flash memory cells in planar technology is becoming increasingly difficult because of concerns on lithography, punch through, reliability, process windows, read current, and isolation requirements between neighboring cells. GAA SONOS Flash memory using the vertical SiNW as the channel is considered as an important candidate to overcome the cell area limitation of $4F^2$ per bit in planar NAND through vertical stacking of cells [6.1]. In this chapter, we design and fabricate stacked NAND test structures based on the developed vertical SiNW platform. Before designing such test NAND array architectures using vertical SiNW SONOS as building blocks, the following requirements were taken into consideration: (1) the designed NAND array test structure must be realized on the vertical SiNW platform; (2) each cell within the designed array must be programmable/erasable using F-N tunneling or CHE injection mechanism; (3) it should make full use of the developed concept and device fabrication flow of unit vertical SiNW SONOS as discussed in previous chapters; (4) the test

structure should be a simplified version real commercial NAND array, but also complex enough to characterize the typical circuit level program/read disturbs that may occur during real NAND array operation; (5) no additional program/read disturbs related to such test architecture; and (6) the test structure must be expandable to realize ultra-high density integration. Considering these objectives, in this chapter, we demonstrate a novel 2-level stacked JL GAA SONOS memory fabricated on vertical-SiNW platform with footprint of $6F^2/2$ and a memory string built based on this technology. The cells in the memory string can be programmed/erased using F-N tunneling mechanism with well-behaved memory performances. In addition, without forming junctions on vertical-wires, the stacking process complexity/cost is greatly reduced with promise towards density improvement [6.2]. Systematic evaluation of the stacked memory cell from the P/E speeds, endurance, and retention to the program disturb is presented. This high performance vertically stacked device structure relaxes the channel length scaling limitation in 2-dimension and makes it a promising candidate for future ultra-high density memory application.

6.2 Stacked Vertical SiNW SONOS Device Fabrication

Fig. 6.1 depicts the 2-level stacked JL-SONOS process integration flow in which a memory string consists of eight memory cells (M1-M8) is fabricated on four vertical SiNWs. Fig. 6.2 shows the corresponding SEM inspection during key process steps. The process was started with phosphorus implant on SOI wafer followed by furnace annealing (1000°C, 3 h) to create uniform dopant profile in the top portion of the wafers [Fig. 6.1 (a)]. After vertical SiNW with various diameters (downscale to 30nm) formation using top-down approach and HDP covering the bottom of the wires [Fig. 6.1 (b) and Fig. 6.2 (a)], tunnelling thermal oxide (5nm) / trapping layer of LPCVD SiN

(7nm) / blocking LPCVD oxide (7nm) were directly formed on the wires. It is worth reiterating that no source/drain implant was performed, thus reducing process cost and complexity as compared to the conventional junction-based SONOS. After 80nm LPCVD poly silicon control gate deposition followed by P⁺ implant and activation [Fig. 6.1 (c) and Fig. 6.2 (b)], the gate stack was patterned and etched to define the bottom level gate pads [Fig. 6.1 (d) and Fig. 6.2 (c)]. The gate length was defined by HDP oxide deposition and etch-back process, followed by tip poly-Si removal in TMAH solution [Fig. 6.1 (e) and Fig. 6.2 (d)]. Once the ONO layer is formed, it can be shared by all the memory cells along each wire. After another HDP oxide deposition and etch-back process to create the isolation between the upper and bottom level cells [Fig. 6.1 (f) and Fig. 6.2 (e)], similar process sequences [Fig. 6.1 (c-e)] were repeated to form the upper level cells [Fig. 6.1 (g) and Fig. 6.2 (f)]. Contact holes for the gate pads were created after the last HDP isolation with only the wire top exposed, and standard metallization was used to conclude the process [Fig. 6.1 (h) and Fig. 6.2 (g)].

Fig. 6.1: Process integration flow for NAND string of eight junction-less SONOS memory cells (M1-M8) on four vertical SiNWs: (a) Bulk implant and annealing; (b) vertical SiNW formation and HDP isolation; (c) 1st level gate stack deposition; (d) 1st level gate pad definition; (e) tip poly removal; (f) HDP oxide isolation; (g) 2nd level gate pad definition and tip poly-ONO removal; (h) metallization. M1 to M8 are 8 cells on 4 wires.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Fig. 6.2: SEM images of 2-level stacked vertical SiNW SONOS memory with four vertical pillars. Bird's eye view of memory string after vertical pillars with bottom covered by HDP (6.2a); 1st level gate stack deposition (6.2b); 1st level gate pad definition (6.2c); 1st level tip poly removal (6.2d); HDP oxide isolation (6.2e); 2nd level gate pad definition and tip poly-ONO removal (6.2f); metallization. WL1 to WL8 are for 8 cells on 4 wires (6.2g).

Fig. 6.3 illustrates the equivalent circuit of conventional and vertical SiNW stacked NAND memory string with eight cells. Select gates can be formed on the peripheral region. Fig. 6.4 shows that the cell area of vertical-SiNW-stacked SONOS in NAND array can be reduced down to $6F^2/N$, where N is the number of stacked cells. Fig. 6.5 is a cross-sectional TEM image of 2-level stacked vertical-SiNW-based JL-SONOS with wire diameter equals to 120 nm and a gate length of ~ 100 nm, and the inset of Fig. 6.5 shows the cross-sectional TEM image of well-defined gate stack ONO thickness 5/7/7 nm and doped SiNW channel ($1 \times 10^{19}/\text{cm}^3$).

Fig. 6.3: Equivalent circuit of (a) conventional and (b) vertical SiNW stacked SONOS memory string with eight cells. SLs can be formed on peripheral region.

Fig. 6.4: Schematic illustration of the footprint of vertical SiNW stacked SONOS device and cell area in NAND array.

Fig. 6.5: Cross-sectional TEM image of 2-level stacked vertical SiNW GAA JL-SONOS: the wire diameter is 120 nm and the gate length is ~100nm, the inset shows ONO thicknesses 5/7/7 nm.

6.3 Results and Discussion

In this section, memory array based on 2-level stacked JL vertical SiNW SONOS with diameters down to 30nm, gate length of ~100nm and wire dopant concentration of $1 \times 10^{19} \text{ cm}^{-3}$ was used for characterization. The threshold voltage (V_{th}) is obtained from I_d - V_g characteristic with drain current (I_d) equal to $10^{-7} \text{ A}/\mu\text{m}$. The operation bias conditions are summarized as follows:

	Program/Erase (F-N)		Read	
	Selected Cell	Unselected Cell	Selected Cell	Unselected Cell
Word Line (WL)	V_{pgm}/V_{ers}	V_{pass}	V_{read}	V_{pass}
Bit Line (BL)	0V	$V_{inhibit}/0V$	V_d	Floating
Upper Select Gate (USG)	V_{pass}		V_{pass}	
Ground Select Line (GSL)	0V		V_{pass}	
Source Line (SL)	Floating		0V	

Table 6.1 Operation bias conditions for stacked NAND memory array based on vertical SiNW SONOS cells.

Fig. 6.6 shows the typical I_d - V_g characteristics of selected JL-SONOS cell with a wire diameter of 30nm, gate length 100nm. A subthreshold slope (SS) of 84mV/decade with I_{off} of 1×10^{-12} A is achieved, thanks to the highly scaled wire channel and thus excellent gate control owing to the GAA structure [6.3].

Fig. 6.6: I_d - V_g plot of vertical JL-SONOS with channel diameter=30nm.

Using a home-made code, Fig. 6.7(a) illustrates that simulated electric field (E-field) across the tunnelling oxide becomes larger with smaller wire diameter under the same gate voltage. When taking a closer look into the E-field distribution along the interface between channel and tunnelling oxide, as in Fig. 6.7(b), the device with smaller wire diameter receives a higher E-field and thus easier to be depleted. It is believed that device SS and I_{off} can be further improved through down-scaling the wire diameter in JL device.

Fig. 6.7: (a) Simulated electric field distribution of the SiNW based JL SONO at $V_g=4$ V with various wire diameter of 20nm / 40nm / 50nm / 110nm and a fixed channel doping of 10^{19} cm^{-3} , (b) zoomed electric field near the channel/tunnel oxide interface.

Fig.6.8 (a) and (b) shows the P/E characteristic of stacked JL-SONOS cell.

The top and bottom cells exhibit similar P/E speed and a 3.4V memory window can be achieved for both top and bottom cells when a P/E time of 10 ms was used at +15V and -18V, respectively. It is also believed that after process optimization to fabricate vertical NW with more straight profile and hence less dimension variation, the difference of P/E performances between the top and bottom cells can be further reduced or even minimized.

Fig. 6.8: (a) Programming and (b) erasing using F-N tunnelling of stacked SiNW based GAA JL-SONOS memory device with channel doping of $1 \times 10^{19} \text{ cm}^{-3}$ and wire diameter of 30nm. open: bottom level cells, solid: top level cells.

The feasibility on the multi-bit programming for the JL-SONOS is shown in Fig. 6.9 (a) and (b). The memory is able to store 2 bits per cell using four states “00”, “01”, “10” and “11” with each state defined a different V_{th} of $> 1V$. These different programmed V_{th} s were obtained by programming the same device with various voltages at 13, 15, and 18 V for 1 ms. Fig. 6.9 (b) shows the excellent cumulative distribution of the programmed V_{th} for ten randomly selected devices with the same SiNW dimension (20 nm) and gate length on the same wafer, which indicates the good process control and the reliability for multi-bit data storage.

Fig. 6.9: (a) Multibit programming with gate biases of 13, 15, and 18 V for 100 ms, the memory is able to store 2 bits per cell using four states “00”, “01”, “10” and “11”. (b) Cumulative distribution of multibit programmed V_{th} with gate biases of 13, 15, and 18 V for 100 ms for randomly selected twenty devices.

In Fig. 6.10, the retention characteristic of the JL device for both virgin cell (solid) and 10K cycled cell (open) is presented. It is found that the window can be well maintained after 10^5 sec at 25 °C for both fresh and cycled cells. Fig. 6.11 shows the extrapolated window loss at elevated temperatures (85 and 125 °C), and it was found 85% of the memory window can be maintained at 125 °C after 10 years. Fig. 6.12 (a) presents the device endurance characteristic up to 10^5 cycles. No obvious V_{th} upward shift is observed, which can be further verified through the IV curves after various P/E numbers. The non-degraded SS shown in Fig.6.12 (b) indicates minimal damage or build-up of charges in the ONO layer after cycling [6.4].

Fig. 6.10: Room temperature retention characteristic: a virgin cell and 10K cycled cell, solid: fresh cell, open: 10K cycled cell.

Fig. 6.11: Retention at elevated temperatures, extrapolated charge loss is 15% at 125 °C after 10 years.

Fig. 6.12: (a) Endurance characteristic up to 10^5 cycles.

Fig. 6.12: (b) I_d - V_g curves of the cell at various cycles.

The program disturb characteristics are shown in Figs. 6.13 and 6.14. As schematically illustrated in Fig. 6.13, V_{pgm} is used to program the selected cell while V_{pass} is applied on unselected cell to bypass the bit line signal. As shown in Fig. 6.13, the CG of cell B experiences the same program voltage V_{pgm} as the selected cell A; Bit Line bias V_{INHIBIT} is applied at its drain to reduce the electrical field in the tunnel oxide and prevent electrons trapping. V_{pass} is applied on other unselected cell to bypass the bit line signal. On the other hand, the CG of cell C experiences V_{pass} while its drain is bias at 0V. Elections will be trapped if V_{pass} is too large. However, if V_{pass} is too small, $(V_{\text{pass}} - V_{\text{th}})$ other than V_{INHIBIT} be transmitted to cell B, which may not be sufficient to prevent electron tapping in cell B.

Fig. 6.13: Schematic illustration of the bias conditions during programming the selected cell. Both V_{pass} disturb and V_{pgm} disturb need to be taken into account.

The V_{pass} disturb is illustrated in Fig. 6.14 (a), and it is found the V_{th} shift for unselected cells is negligibly small with V_{pass} of up to 9V. On the other hand, the V_{pgm}

disturb happens when the CG of unselected cells is biased at V_{pgm} during program as they share the common word line (WL). This V_{pgm} disturb can be suppressed by using a direct raised-BL approach as discussed earlier. As shown in Fig. 6.14 (b), the disturb-free window is a function of applied BL voltage (V_{INHIBIT}) which must be sufficiently large to suppress the charge trapping in unselected cells [6.5].

Fig. 6.14: (a) V_{th} shift after V_{pass} disturb on unselected cells. (b) Disturb free window is a function of BL voltage $V_{INHIBIT}$. (c) V_{pgm} disturb window using ISPP and raised-BL method, the selected device can be programmed to $V_{th} > 4$ V, while unselected devices remain in the erased state.

In Fig. 6.14 (c), the selected cell can be programmed to $V_{th} > 4V$ using the incremental step pulse programming (ISSP), while unselected cells remain in the erased state. Without using junction for the memory cell, the integration process is significantly simplified. It is worth mentioning the difference between the proposed stacked JL memory in this work and BiCS technology [6.2]. In our case, a highly scalable (with diameter down to sub-30nm) single crystalline Si channel with thermally grown oxide is used to build the memory. This could alleviate the potential reliability issues, e.g. SS , leakage, V_{th} distribution and data retention, which might be correlated with the inherent low grade of poly-Si channel and deposited tunnel oxide in BiCS

6.4 Conclusion

In this chapter, the memory string using 2-level stacked JL SONOS realized on vertical-SiNW GAA structure with channel dimension down to 30nm has been demonstrated for the first time. A systematic study on the stacked cell from manufacturability, device performance and reliability perspectives has been presented. The great benefits and opportunities for which the vertical-SiNW platform can bring to the memory applications have been pointed out. The stackable vertical-SiNW-based JL-SONOS with high performance is considered as a promising candidate for future ultra-high density application.

Reference

- [6.1] M. C. Chen, H. Y. Yu, N. Singh, Y. Sun, N. S. Shen, X. H. Yuan, G. Q. Lo and D. L. Kwong, “Vertical Si nanowire SONOS memory for ultra-high density application”, *IEEE Electron Devie Lett.*, Vol. 30, pp 879-881 (2009).
- [6.2] 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,” *VLSI Symp. Tech. Dig.*, pp. 14-15 (2007).
- [6.3] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, and J.-B. Park, “High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET) fabrication on bulk Si wafer, characteristics, and reliability”, *IEDM Tech. Dig.*, pp. 717-720 (2005).
- [6.4] B. Tomar et al., Proceedings of 11th IPFA (2004).
- [6.5] Hang-Ting Lue, Yi-Hsuan Hsiao, Pei-Ying Du, Sheng-Chih Lai, Tzu-Hsuan Hsu, S. P. Hong, M. T. Wu, F. H. Hsu, N. Z. Lien, Chi-Pin Lu, Jung-Yu Hsieh, Ling-Wu Yang, Tahone Yang, Kuang-Chao Chen, Kuang-Yeu Hsieh, Rich Liu, and Chih-Yuan Lu, “A novel buried-channel FinFET BESONOS NAND flash with improved memory window and cycling endurance”, *VLSI Symposium Tech. Dig.*, pp.224 -225 (2009).

Chapter 7

Conclusions

This thesis focuses on the study of flash NVM devices with the integration of vertical NW channel. It started with an review of the semiconductor memory technology and the challenges of memory cell size shrinking faced by both floating gate-based and charge-trap-based devices. The trade-off and future prospects brought by the 3-D nano-wires structures were presented thereafter along various embodiments of the device structure implementations and experiment results.

The gate control of flash devices realized on highly scaled vertical SiNW channel was greatly improved due to the GAA structure, reflected by the SS and DIBL values which were as small as those in MOSFET devices. The devices realized on squarish SiNW down to 20 nm in diagonal showed much improved P/E speed and window along with good retention and endurance characteristics as compared to the ones with a large dimension. The performance improvements with scaling of wire dimensions were attributed to the enhancement of the electric field in tunnel oxide and reduction in blocking oxide as a result of reduced radius of curvatures, particularly on the corners of the squarish wire.

Silicon nanocrystals (Si-NCs) were subsequently incorporated as charge trapping material into the junction-based vertical SiNW memory devices to improve the device memory performances, such as programming/erasing (P/E) speed, retention and endurance. Engineering of charge trapping layer by incorporating Si-NCs in this study was found to be beneficial for improved the NW memory characteristics.

To make vertical SiNW a suitable platform for vertical stacking of memory cells to achieve ultra-high bit density, a novel JL vertical-SiNW SONOS memory was demonstrated which makes vertical stacking of memory cells much easier. A systematic study on JL SONOS memory was presented from the process complexity, device performance and reliability perspective. The effect of channel doping, wire diameter and the underlying physics were discussed and analyzed through device simulation.

Next, a JL SONOS memory realized on vertical SiNW GAA structures with two physical storage nodes per cell was fabricated and electrically evaluated. Comparing with a planar device, the array cell tuning is much easier since the channel length is no longer limited by the cell area. In addition, the absence of junctions reduces the process complexity/cost and makes this device manufacturable with very low thermal budget.

We also present GAA vertical poly-Si NW thin film transistors (TFTs) and SONOS devices with wire diameter down to 30nm. Both memory and transistor performances are improved with down scaling of the NW diameter, which is attributed to the reduced number of grains and grain boundaries in the channel region and the enhanced gate control with down scaling of wire channel dimension.

Finally, a 2-level stacked JL GAA SONOS memory fabricated on vertical-SiNW platform with footprint of $6F^2/2$ and a memory string built based on this technology were demonstrated. Systematic evaluation of the stacked memory cell from the P/E speed, endurance, and retention to the program disturb was presented. The stacked JL-SONOS SiNW is a promising candidate for 3-D ultra-high density memory application.

This study has so far achieved a reliable technique for the top-down vertical NW NVM device fabrication. It ensures a reliable SCE immunity, fast memory transient characteristics and good reliability characteristics. The purpose of

the discussion in these versatile vertical SiNW devices is to demonstrate the vast benefits and opportunities the NW technology can bring to the memory applications. Of course, huge amount of innovation and engineering efforts are required to bring the benefits into the real application with worthy market.

It is recommended that suitable high- κ and metal gate materials could be integrated into vertical NW devices. Although improve P/E speed and vertical stacking were realized in the vertical NW platform discussed in Chapter 6, the charge retention properties, still leave much room for improvement. Hence more research work could be done to further optimize the memory characteristics in the NW device by integrating suitable types of gate stack materials, especially advanced high- κ materials.

List of Publications

Books:

- [1] **Title:** *Nonvolatile Memories: Materials, Devices and Applications* (contribute a chapter: “Ultra-high Density Non-volatile Flash Memory Devices Realized on CMOS Si Nanowires”) to be published by American Scientific Publishers in 2011. Authors: Dim-Lee Kwong, **Yuan Sun**, Yu HongYu, Navab Singh, and Guo-Qiang Lo.

Journal Publications:

- [1] **Y. Sun**, H.Y. Yu, N. Singh, K. C. Leong, G.Q. Lo, and D. L. Kwong, “Junction-less vertical-Si-nanowire channel SONOS memory with 2-bit storage/cell”, *IEEE Electron Device Lett.*, Vol. 32, pp. 725 (2011).
- [2] **Y. Sun**, H.Y. Yu, N. Singh, E. Gnani, G. Baccarani, K. C. Leong, G.Q. Lo, and D. L. Kwong, “Vertical Si-nanowire based non-volatile memory with improved performance and reduced process complexity”, *IEEE Transaction on Electron Devices*, Vol. 58, pp. 1329-1335 (2011).
- [3] **Y. Sun**, H.Y. Yu, N. Singh, K. C. Leong, G.Q. Lo, and D. L. Kwong, “Gate-all-around poly-Si vertical nanowire thin film transistor (TFT) SONOS”, *Electrochemical and Solid-State Letters*, to be submit.
- [4] H. Y. Yu, **Y. Sun**, N. Singh, G. Q. Lo, and D. L. Kwong, “Perspective of flash memory realized on vertical Si nanowires”, *Microelectronics Reliability*, 2011 (Invited).
- [5] D. L. Kwong, X. Li, **Y. Sun**, G. Ramnathan, Z. X. Chen, S. M. Wong, Y. Li, N.S. Shen, K. Buddharaju, Y. H. Yu, S. J. Lee, N. Singh, and G. Q. Lo, “Vertical silicon nanowire platform for low power electronics and clean energy

- applications”, *Journal of Nanotechnology*, Special Issue on Semiconductor Nanowires and Nanotubes: From Fundamentals to Diverse Applications, 2011 (Invited).
- [6] T. T. Le, H. Y. Yu, **Y. Sun**, N. Singh, X. Zhou, N. Shen, G. Q. Lo, and D. L. Kwong, “High performance poly-Si vertical nanowire thin film transistor and the inverter demonstration”, *IEEE Electron Device Lett.*, Vol. 32, pp. 770 (2011).
- [7] **Y. Sun**, Y. H. Yu, N. Singh, N. S. Shen, G. Q. Lo and D. L. Kwong, “Multi-bit programmable flash memory realized on vertical Si nanowire with a highly scaled squarish channel”, *IEEE Electron Device Lett.*, Vol. 31, pp. 390-392 (2010).
- [8] M. C. Chen, H. Y. Yu, N. Singh, **Y. Sun**, N. S. Shen, X. H. Yuan, G. Q. Lo and D. L. Kwong, “Vertical Si nanowire SONOS memory for ultra-high density application”, *IEEE Electron Device Lett.*, Vol. 30, pp 879-881 (2009).

Conference Publications:

- [1] **Y. Sun**, H.Y. Yu, N. Singh, K. C. Leong, G.Q. Lo, and D. L. Kwong, “Demonstration of memory string with stacked junction-less SONOS realized on vertical silicon nanowire”, *IEEE International Electron Devices Meeting 2011*, accepted for oral presentation.
- [2] **Y. Sun**, H.Y. Yu, N. Singh, E. Gnani, G. Baccarani, K. C. Leong, G.Q. Lo, and D. L. Kwong, "Junction-less stackable SONOS memory realized on vertical-Si-nanowire for 3-D application" *IEEE 18th International Symposium on VLSI Technology, Systems and Applications (2011 VLSI-TSA)*, 25 to 27 April 2011, Hsin-Chu, Taiwan.
- [3] **Y. Sun**, Y. H. Yu, N. Singh, N. S. Shen, G. Q. Lo and D. L. Kwong, “Vertical Si nanowire SONOS memory for ultra-high density application” , *The 2010 International Congress on Computer Applications and Computational Science*

(*CACS 2010*), 4-6 December 2010, Singapore.