Novel Technologies for Next Generation Memory

WookHyun Kwon
Tsu-Jae King Liu, Ed.
Vivek Subramanian, Ed.

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-136
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-136.html

July 25, 2013

Acknowledgement

I would like express my sincere gratitude to my advisor, Professor Tsu-Jae King Liu. Prof. King Liu truly inspire me to do better, and I am always looking forward your advices because she always give great feedback.
I am deeply grateful to Professor Vivek Subramanian for his enthusiastic class, especially EE231. He has a superior teaching techniques that lead students to have intuitions on the device. I hope he keeps teaching, during whole life.

Novel Technologies for Next Generation Memory

By

Wookhyun Kwon

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Tsu-Jae King Liu, Chair
Professor Vivek Subramanian
Professor Junqiao Wu

Fall 2013

Novel Technologies for Next Generation Memory

by

Wookhyun Kwon

Abstract

Novel Technologies for Next Generation Memory

by

Wookhyun Kwon

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King Liu, Chair

Historically, our society requires computational memory media to support the development of our civilization. It is likely that our society will keep demanding larger capacity memory. However, conventional memory technologies are facing many challenges such as difficulties of miniaturization and guarantee of good reliability. For this reason, alternate memory device designs are proposed to overcome the conventional memory device technologies.

For DRAM technology, a double-gate array having vertical channel structure (DGVC) with $4F^2$ cell size is proposed, which can be fabricated on a bulk silicon wafer using the conventional memory process flow for stand-alone DRAM application. The operation and scalability of the DGVC cell are demonstrated via TCAD device simulations.

For Flash Memory technology, a new backside charge storage non-volatile memory (BCS-NVM) cell design is proposed. A NAND flash array of the BCS-NVM cells can be fabricated on a modified SOI substrate. TCAD device simulation show that this design allows for a relatively high cell read current and steep sub-threshold slope to enable lower voltage operation in comparison with conventional NAND flash memory devices.

As a new concept of non-volatile memory technology, a nano-electro-mechanical (NEM) diode non-volatile memory cell design is proposed. This design eliminates the need of a selector device to form a cross-point array, by leveraging the gap closing actuator. The electro-mechanical diode cell design can be scaled to 20 nm minimum lateral dimension by following an appropriate scaling methodology in consideration of various practical and fundamental limits. Low-voltage (< 2 V) and high-speed (sub-nanosecond) operation are projected using a calibrated analytical model as well as 3-D FEM simulation. These findings indicate that electro-mechanical diode technology is promising for high density storage beyond the limits of conventional flash memory technology.

To my wife for her devoted love and support,
to my daughters for their sincere smile,
and to my parents for their unbounded love.

Contents

Table of Contents

List of Figures

List of Tables

Acknowledgements

Since I have left school after my master degree 10 years ago, pursuing a Ph. D was a challenge but exciting journey. Taking exam and doing homework were not easy at the beginning but it also is getting used to it. Thus, I have to say that I am grateful for all the changes and encouragements I get during my time at Berkeley.

First of all, I would like express my sincere gratitude to my advisor, Professor Tsu-Jae King Liu. Prof. King Liu truly inspire me to do better, and I am always looking forward your advices because she always give great feedback. In all research meeting, she never failed to advise me with her inspiring perspective, amazing engineering-intuition, tireless enthusiasm, and firm and solid knowledge. There is much about her management style that I would like to bring with me as I go back to professional engineering field, industry. Still I cannot believe how you give so many advices to all your students. You are amazing.

I am deeply grateful to Professor Vivek Subramanian for his enthusiastic class, especially EE231. He has a superior teaching techniques that lead students to have intuitions on the device. I hope he keeps teaching, during whole life.

I also thank Professor Ana Claudia Aria for serving on both my qualifying examination committee and my dissertation committee, and thank Professor Junqiao Wu or serving my qualifying examination committee.

I would like to thank several former and current students with regards to this work. With Dr. Hei Kam, Dr. Rhesa Nathanael, and Professor Jaeseok Jeon, I could establish the foundation of my research. They provided kind assistance and several helpful discussions on electro-mechanical switch devices, their evaluation method and fabrication process method. With Dr. Louis Hutin, Eung Seok Park and Chuang Qian, I had a lot of discussions on process design of the device. They also kindly assisted with many help in the nano-fabrication lab.

In addition to the above, several others have provided indirect assistance. Dr. Min Hee Cho and Dr. Nuo Xu helped me to acquire knowledge of TCA simulation. I am deeply grateful to the former and current members of Berkeley research group for their encouragement and help. Special thanks for Jae Min Park, Sung Hwan Kim, Byron Ho, Peter Matheu, Eung Seok Park, I-Ru “Tim” Chen, Yenhao “Philip” Chen, Jack Yaung, Tae Joon Seok, Hongki Kang, Jaewon and all staffs in our world best NANO Fabrication Lab. I appreciate all of their help.

This work has been funded through a few sources. First of all, my financial stability for my Ph.D. work came from Samsung Electronics Fellowship. Fabrication of a proto-type electro-mechanical diode cell is supported in part by Defense Advanced Research Projects Agency (DARPA) / MTO NEMS program. This work was also supported by the Center for Circuit & System Solutions (C2S2) Focus Center. I would like to express my sincere gratitude for the financial and technical supports, and I will not forget to contribute back to them in future.

Finally, I would not be here today were it not for my family. I can not thank my parents and my younger brother enough for their unbounded love and sacrifices. Their love will always hold a special place in my heart. Lastly, but most certainly not least, I would like to thank my wife, Yeon Ji Kim, for her devoted love and my adorable three daughters, Seungju, Minkyoung (Kelly) and Seunghyun (Elly). All my loves…

Chapter 1

Introduction

1.1 Overview of Computer Memory Technologies

Computer memory is a recording media used to retain digital data. It is a core function and fundamental component of computers. In the early stage of computer technology, computer memory capacity was mostly a few bytes. The first electronic programmable digital computer (ENIAC) used thousands of octal-base radio vacuum tubes. ENIAC could perform simple calculations involving 20 numbers of ten decimal digits which were held in the vacuum tube accumulators. In the early 1960s, Jay Forrester, Jan A. Rajchman and An Wang developed magnetic core memory, which allowed for recall of memory after power loss. Magnetic core memory would become the dominant form of memory until the development of semiconductor memory in the late 1960s [1].

Semiconductor memory is an electronic data storage device implemented on a semiconductor-based integrated circuit (IC). The semiconductor memory has the property of random access so that stored data can be efficiently accessed randomly. Thus, the term "memory" when used with reference to computers generally refers to Random Access Memory or RAM [2]. Since semiconductor memory has much faster access times than other types of data media such as a hard disk or compact disks (CDs), it is used for main computer memory to hold data which the computer is currently working on.

There are two types of semiconductor memory classified accordingly to data retaining ability. Volatile memory loses data when the computer power is off. In general, volatile memory has faster read-out performance, but it has a smaller storage capacity than non-volatile memory. Most common semiconductor volatile memory technologies are Static RAM (SRAM) and dynamic RAM (DRAM). On the other hand, non-volatile memory can retain data even when the computer is not powered. This feature has made these types of

memory devices well suited for portable electronic devices such as a mobile phone, digital camera, and tablet PC etcetera). Non-volatile memory (NVM) are Read-Only memory (ROM), Electrically-Erasable ROM (EPROM), Electrically-Erasable-Programmable ROM (EEPROM) and Flash memory [3].

1.2 Dynamic Random Access Memory (DRAM)

1.2.1 Conventional DRAM and Limitations

A conventional DRAM cell has simple structure, which is composed of one transistor (i.e. access transistor) and one capacitor (1T1C) per bit [4]. Electron charges are stored in the capacitor through the access transistor. Since a real capacitor and transistor leak charge, the stored information (stored charge in the capacitor) eventually fades unless the capacitor charge is refreshed, so the memory cells must be periodically refreshed (rewritten). The DRAM size is smaller than a SRAM cell which has six transistors. This relatively smaller cell size allows DRAM to reach very high density so that DRAM occupies a large portion in the semiconductor industry.

The conventional DRAM technology has evolved over the past 20 years and the cell size has been reduced dramatically as shown in **Figure 1.1**. The advancements in DRAM technology include not only photolithography technology but also $6F^2$ unit cell layout area design and a 3-dimensional transistor channel structure.

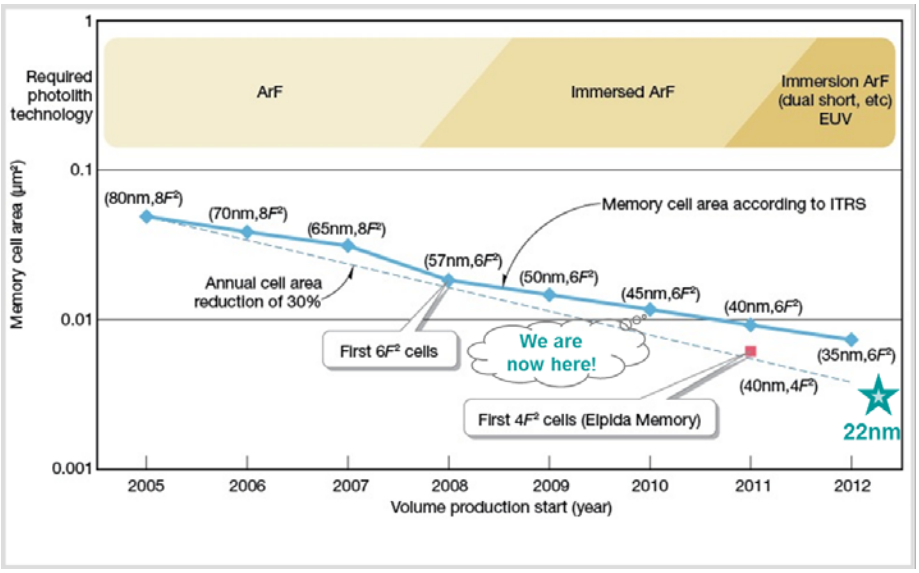
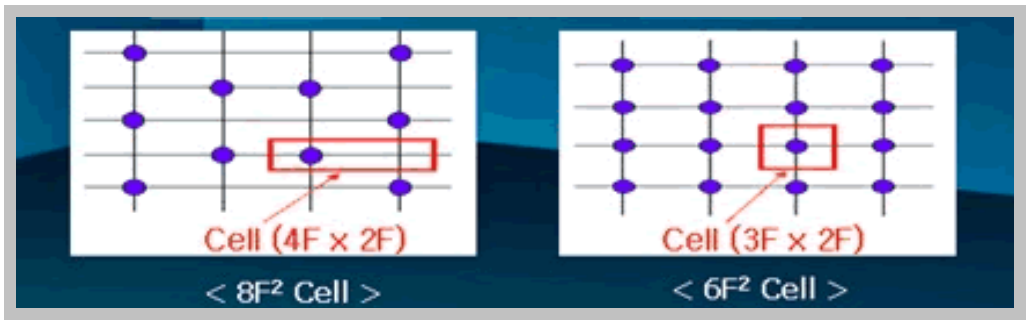
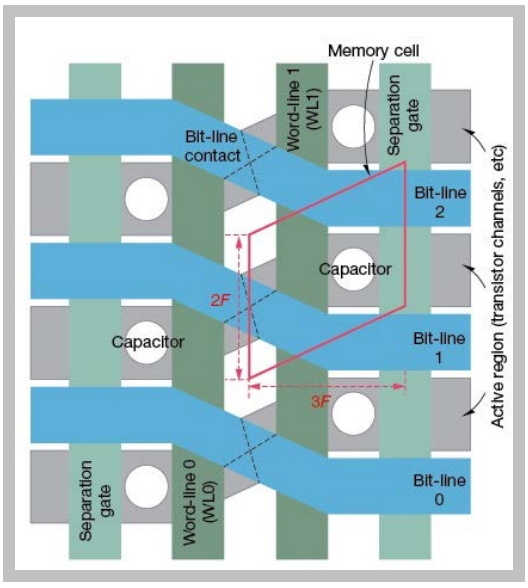


Figure 1.1. Technology trend of conventional DRAM (Ref: *Nikkei Electronics Asia, 2007*)

Figure 1.2 compares the memory cell array layouts for $8F^2$ vs. $6F^2$ cell designs. To achieve the $6F^2$ layout design, the channel length of the access transistor has to be scaled more aggressively. But, a very short channel length of the access transistor causes many reliability issues (such as poor data retention problem and refresh failure). In order to avoid those problems, a 3-dimensional channel structure having recessed-channel is used. **Figure 1.3** shows how recessed-channel transistor (RCAT) evolved from 80nm to 60nm technology generations. The RCAT structure effectively increases the channel length of the access transistor for the same layout dimension. Currently, most DRAM manufacturers use the RCAT structure.



(a)



(b)

Figure 1.2 (a) Comparison between a conventional $8F^2$ and new $6F^2$ layout design. (b) Example of $6F^2$ layout schematic.

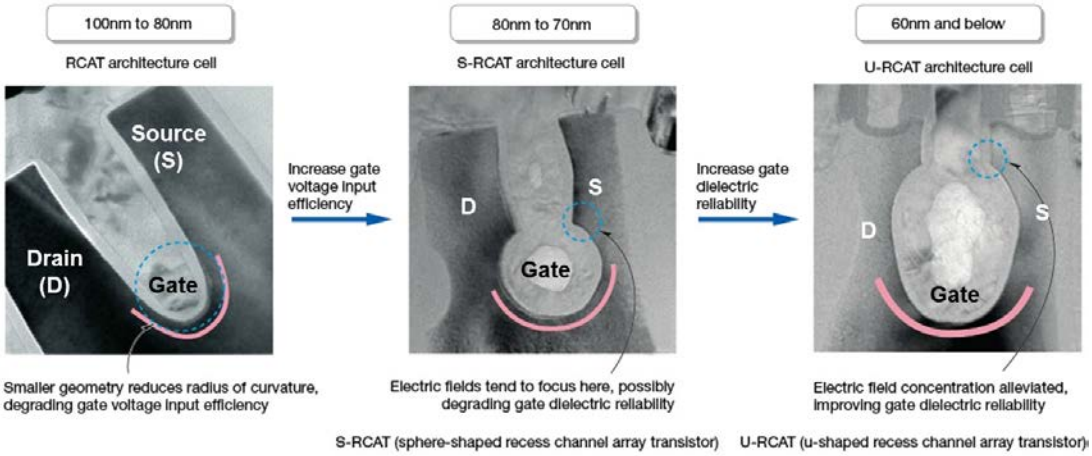


Figure 1.3 Recessed-Channel-Access-Transistor (RCAT) structures in DRAM technology. (Ref: *2005 Symposium on VLSI Technology and 2007 SSDM*)

In spite of thoese technological innovations and DRAM scaling, the worldwide DRAM industry lost over 10 billion US dollars in 2006 ~ 2008. The average price per bit has dropped 26% every year recently and this is likely to continue for a while. That means that in order to survive in today's market, the cell size must be reduced more than 30% every year and this line indicates annual cell area reduction of 30%. However, the nowadays cell size scaling trends cannot achieve the line even if using the $6F^2$ cell design. To achieve a 30% annual reduction, we have to use a $4F^2$ cell design. Since the conventional DRAM cell has one storage capacitor and one transistor, it is really difficult to achieve $4F^2$ cell area. DRAM manufacturers face a tremendous challenge to shrink the cell area for sub-50-nm technologies [5].

1.2.2 Capacitor-less DRAM Technology

In the early 2000s, capacitor-less DRAM memories were proposed to remove the external capacitor from the conventional 1-transistor/1-capacitor DRAM cell [6]. The capacitor-less DRAM cells have a Floating-Body-Structure (FBC) on a partially depleted (PD) SOI as shown in **Figure 1.4** and **Figure 1.5**. This technology has attractive features within a small cell size without the addition of an external storage capacitor.

Although the removal of the external capacitor reduced the area overhead and the process complexities, the fabrication of FBC devices on a SOI wafer has its own challenges for stand-alone memory applications. First, the SOI wafer is still too expensive for the mass production of memory. Second, it is hard to meet basic performance requirements with a very short channel. The basic requirement for a DRAM cell is to maintain a sufficient sensing margin between the two states during the refresh cycle (retention time). When the channel length is scaled down, it is difficult to guarantee sufficient retention time due to short channel effects such as drain-induced barrier lowering (DIBL). To suppress

short channel effects, a very thin body structure is needed. However, it is hard to mass products a thin silicon body SOI wafer with good qualities.

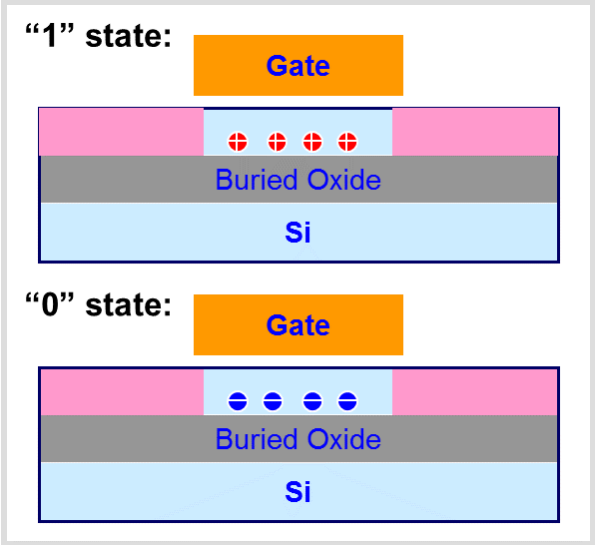


Figure 1.4 Schematic cross-sectional view of a capacitor-less DRAM in "1" and "0" state respectively.

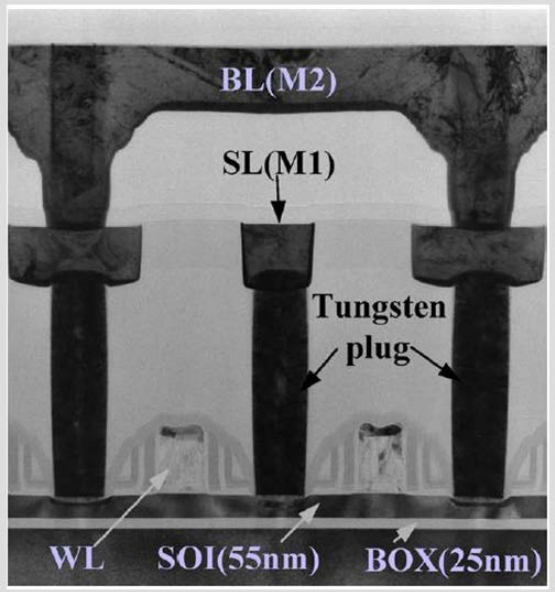


Figure 1.5 Transmission-Electron-Microscopy (TEM) image of a fabricated capacitor-less DRAM cell (Ref: T. Tanaka, E. Yoshida, T. Miyashita, "Scalability study on a capacitorless 1T-DRAM: from single-gate PD-SOI to double-gate FinDRAM," in *IEDM Technical Digest.*, 2004)

1.2.3 Doubly-Gated Capacitor-less DRAM Technology

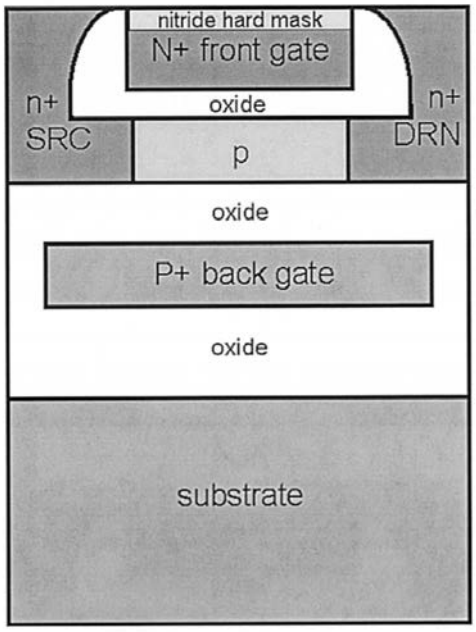


Figure 1.6 Schematic cross-sectional view of a capacitor-less DRAM on SOI

Another concept of capacitor-less DRAM cell was proposed by UC Berkeley researchers [7, 8]. In this design, the back-gate bias electrically induces a potential well at the back gate interface to store charges as shown in **Figure 1.6**. Unlike conventional double-gate logic transistors, the two gates in double-gate DRAM cells are not electrically connected. This structure permits an electrically induced floating body to occur within a thin fully depleted body. Excess holes are collected at the back interface and increase the read current. When a negative back gate bias is applied, it provides a potential-well that permits the retention of holes in the body. Since a double-gate device can terminate field lines from the drain effectively, the double-gate cell is able to suppress short channel effects and maintain sufficient retention time. Similar to a double-gate transistor, it is more scalable to shorter channel length with larger body thickness than PD-SOI device.

Recently, the DG-DRAM cell was successfully demonstrated by Intel (**Figure 1.7**). Even if these cell designs have many advantages, however, the cell sizes are still larger than $4F^2$ and they require a more expensive SOI wafer than bulk-silicon wafer. Another important point is that there are little rooms for scaling previous cell designs.

Figure 1.8 shows the allowed body thickness as the channel length decreases. As the channel length decreases, body thickness must be reduced to suppress the short channel effect such as DIBL. The channel length scaling is limited to about 25nm due to quantum confinement effects in the thin body below 4nm thickness as shown in Figure 1.8.

However, since the conventional DRAM cells are already scaled down near 25nm, the capacitor-less-DRAM cell might not be adventurous.

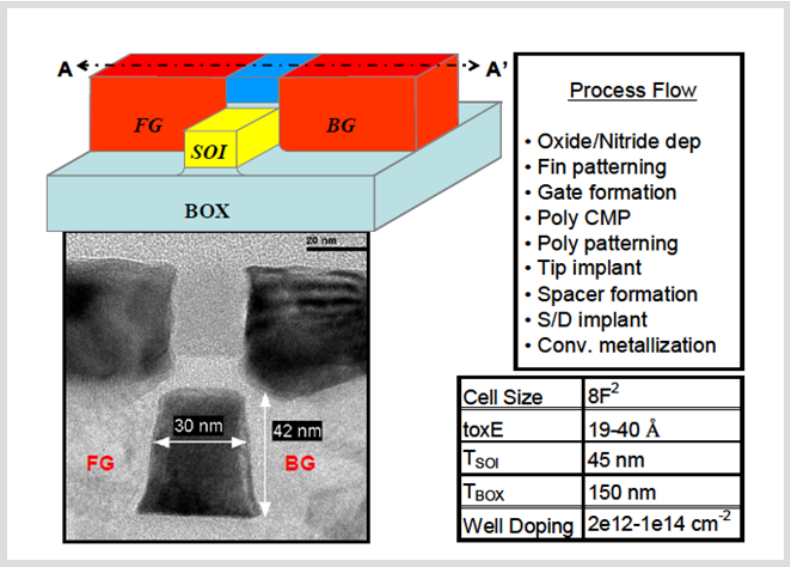


Figure 1.7 Schematic cross-sectional view and TEM image of a typical Independently-Controlled-Double-Gate (IDG) Floating-Body-Cell (FBC) device. (Ref. I. Ban et al, "Floating Body Cell with Independently-Controlled Double Gates for High Density Memory," in *Electron Devices Meeting*, 2006.)

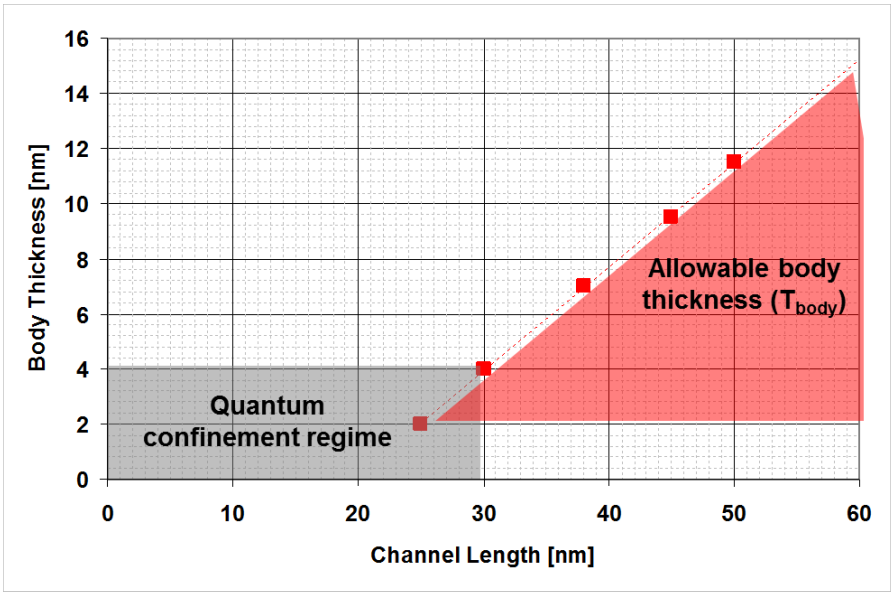


Figure 1.8 Channel length *vs*. Allowable body thickness in a doubly gate capacitor-less DRAM cell design.

1.3 Flash Memory

1.3.1 History of Flash Memory Technology

Flash memory is an electronic non-volatile memory that can be electrically erased and reprogrammed. Figure 1.9 shows a standard Flash memory cell structure. It is similar to the traditional MOS-transistor structure, except that its gate oxide is essentially modified to include a conductive floating-gate. Bit information is stored in this cell through the controlled placement of electrons onto its floating-gate. The cell is electrically erasable and programmable as for EEPROM, except that the erase operation is done by *erase-block* unit. There are two main types of flash memory, which are named after the NAND and NOR logic gates. Those flash memories were invented by Dr. Masuoka in 1980 and 1987. When Dr. Masuoka invented "Flash" memory at Toshiba in the 1980s, the name "Flash" was suggested by Shoji Ariizumi to describe that "*all cells in erase-block are erased same time like a camera Flash*" [9]. The first commercial product of NOR-type Flash memory (**Figure 1.10 (a)**) was released by Intel in 1988 [10]. Toshiba also announced NAND-type Flash memory (**Figure 1.10 (b)**) products (4Mb density) in 1991 [11].

Figure 1.10 (a) illustrates a schematic cross-sectional view of NOR-type Flash memory array along a bit-line. In NOR-type architecture, each cell can be independently accessed or modified so that it has faster (individual cell) read access. Thus, NOR-type Flash memory are mainly used for *executable-code* storage. On the other hand, in NAND-type array architecture, cells belonging to the same bit-line are connected in series between two bit-line select (BLS) transistors, and the gates of all cells in the same row share the same word-line as shown in **Figure 1.10 (b)**. This architecture has a more compact layout but read speed is slower than for NOR-type. For these reasons, NAND-type Flash memory is mainly used for mass data storage applications. **Table 1.1** highlights the major differences between NOR and NAND Flash memory.

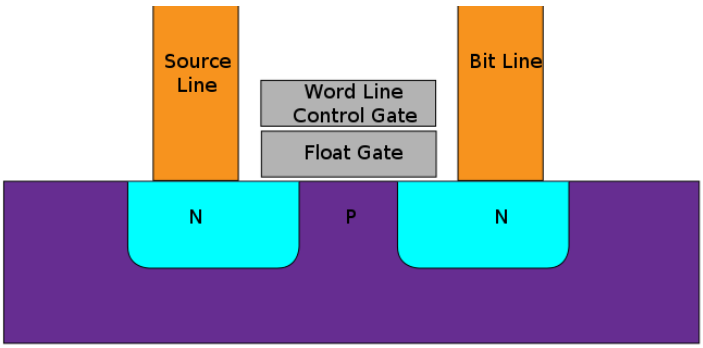
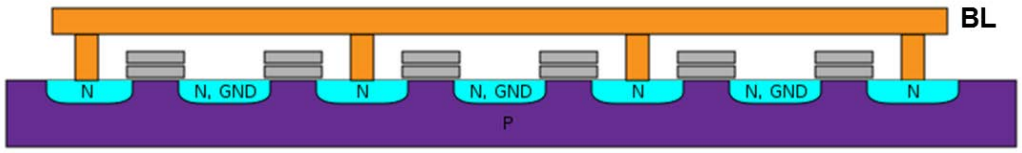
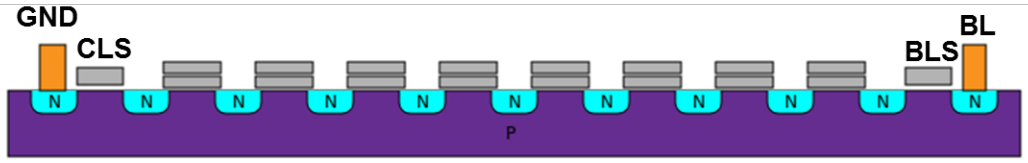


Figure 1.9 Schematic cross-section of the conventional flash memory cell.



(a)



(b)

Figure 1.10 Predominant Flash memory array architectures: (a) Common-Source NOR-type array architecture and (b) NAND-type array architecture.

	NOR	**NAND**
Cell size	$8\sim12F^2$	$4F^2$
Read Performance	Fast (~10usec)	Slow (> 200usec)
Storage Information	Code	Data

TABLE 1.1. Major Differences between NOR and NAND Flash memory.

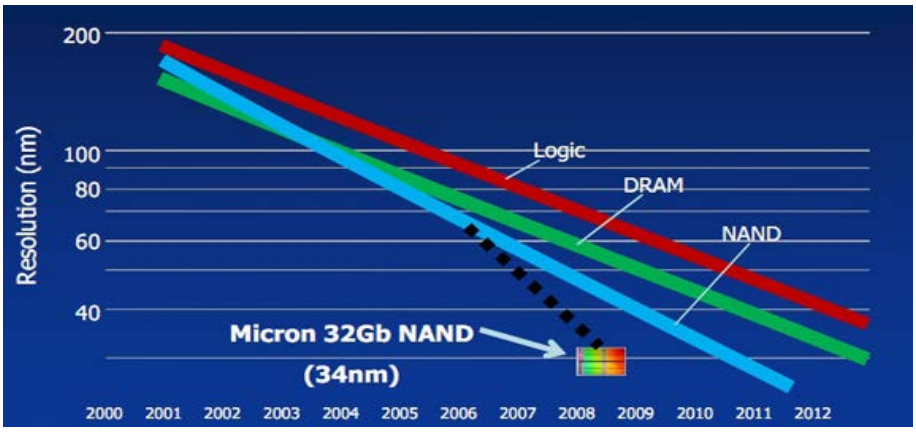


Figure 1.11 Photo-lithographic resolution trends. (Ref: *Semiconductor International*, 2007)

Figure 1.11 shows technology node trends of Logic, DRAM and Flash devices during the past 12 years. NAND Flash memory is the most aggressively scaled technology among electronic devices because the NAND Flash memory cell have relatively simple structure compared with others and there is severe competition among the top manufactures. Also, there were many technological innovations (such as *multi-level cell* and *double patterning lithography*) that resulted in cell size shrinking faster than dictated by Moore's law. This dramatic cell size scaling resulted in a radical price drop of flash memory and booming of NAND flash market. More importantly, the price drop of storage media brought us new innovative electronics. Recently, the NAND Flash memory market has been grown about 40% per year and it has now over 20 billion dollars market as shown in **Figure 1.12**. It is worth noting that when new mobile electronic devices (e.g. digital camera, smart phone, tablet PC and SSD notebook) were released, the market size increased rapidly. Thus, NAND Flash memory market is expected to maintain high growth rate because of continued steady growth of the mobile electronic devices market.

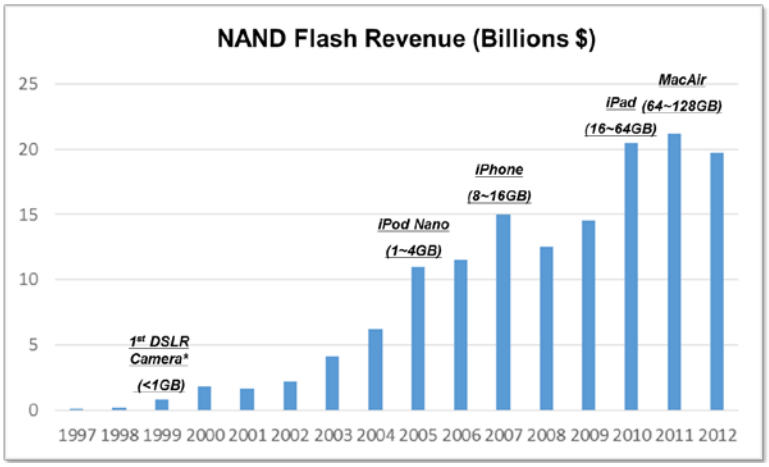


Figure 1.12. NAND Flash memory revenue trend of the past 15 years.

1.3.2 Limitations of conventional Flash memory technology

The conventional Flash memory is facing technological barriers as it scales to sub 20 nm nodes. These barriers are rather fundamental, in view of key cell features. Firstly, further channel length scaling (without gate-oxide scaling) causes poor I_{on}/I_{off} ratio of the cell, which is not enough for read operation. Secondly, no physical space for the control-gate below 20nm technology is expected because NAND Flash cell using 25nm technology has already 10nm space for the control-gate material as shown in **Figure 1.13**. It also results in a significant cell-to-cell interference problem. More importantly, the number of stored electrons in the floating gate decrease as the floating gate size shrinks. **Figure 1.14** shows that the critical (or maximum) number of electrons (to guarantee memory reliability) is only a few tens of electron for the 20nm technology node.

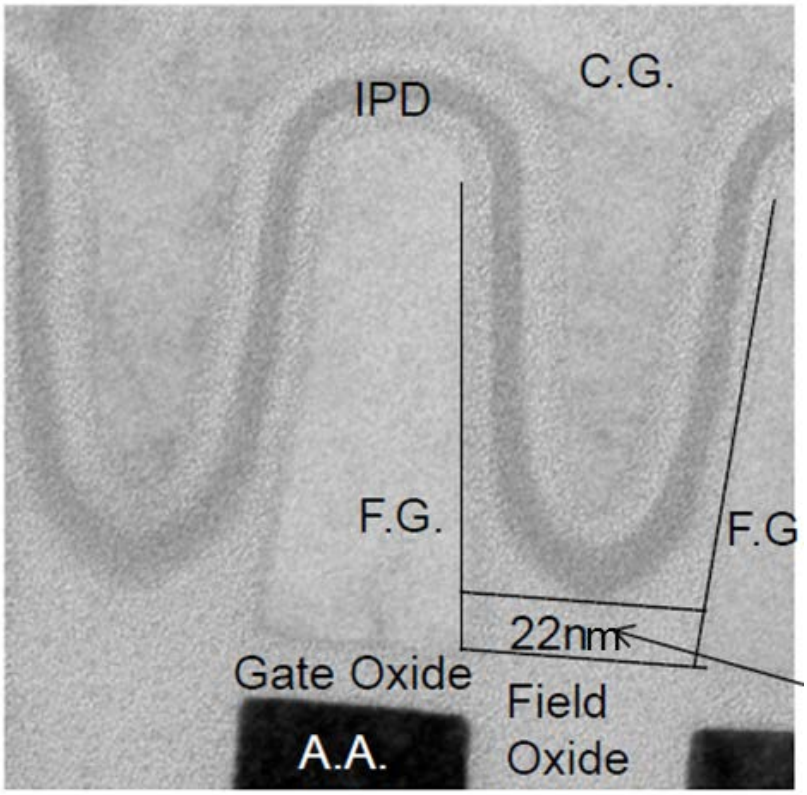


Figure 1.13. Cross-sectional TEM image of 25nm NAND Flash memory cell (Ref: "IMFT 25-nm MLC NAND: technology scaling barriers broken" 3/22/2012 on www.eetimes.com) showing the active Si area (A.A.), tunneling gate oxide, floating gate (FG), inter-poly dielectric (IPD) and control gate (C.G.) structures.

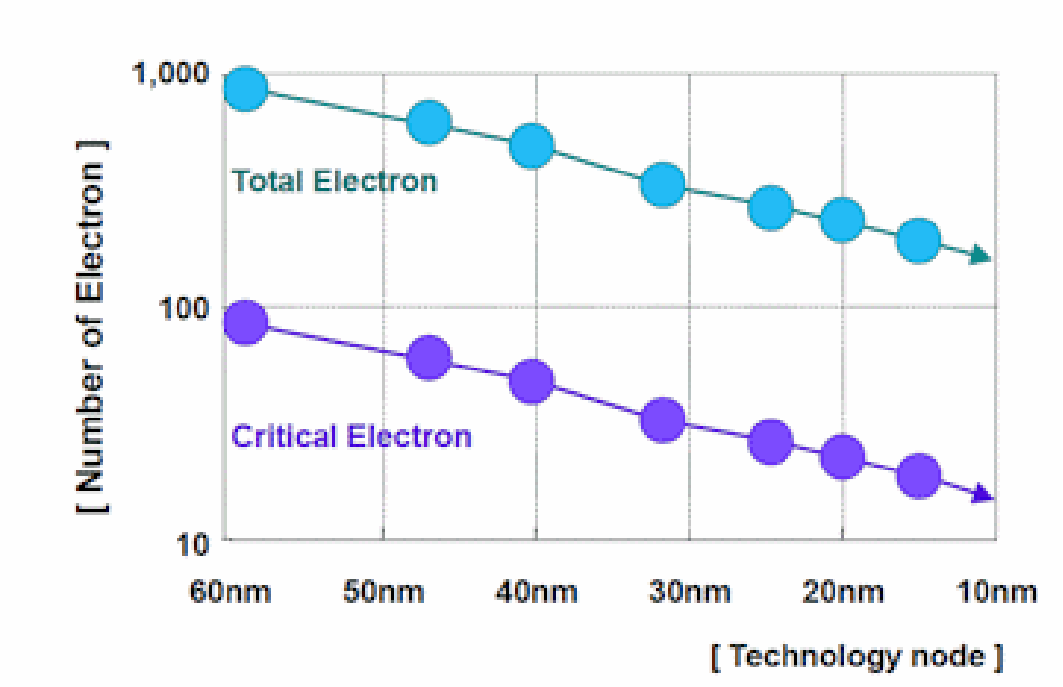


Figure 1.14. The number of electrons stored in the NAND Flash cell's floating gate (Ref: *Flash Memory Summit* 2012)

1.4 Emerging Memory Technology

1.4.1 Cross-point Memory

A lot of alternative memory device designs and new materials have been proposed to overcome the scaling limits for conventional Flash memory technology. Cross-point memory architecture is among the most attractive successors to the current Flash technology because of following reasons: it allows for the most compact storage with $4F^2$ cell layout size (where F is the minimum half-pitch) and also can be fabricated using a relatively simple process that is more amenable to 3-dimensional integration.

Programmable resistance devices such as phase-change memory [12] and resistive RAM [13] have been explored for cross-point memory applications (**Table 1.2**). However, those emerging memory devices generally require a selector device within each memory cell to reduce unwanted leakage current through unselected cells during a read operation; otherwise, the size (number of rows/columns) of the array will be severely limited, resulting in poor memory-array area efficiency [14]. However, the selector devices require additional process steps and can significantly reduce the cell current, resulting in slower read operation.

Single cell @ 45nm node	PCM	STT-RAM (MRAM)	ReRAM
Mechanism	Phase change via thermal heating	Magnetization change by current or E-field	Ox. vacancies control by E-field
Write Power	300 μW	60 μW	50 μW
Switching Time	100 ns	4 ns	5 ns
Endurance	10^{12}	$> 10^{14}$	10^{10}

Table 1.2. Performance of Emerging Memory Devices at the 45nm generation of technology.

1.4.2 Mechanical Switch Memory

A mechanical switch memory is a memory device utilizing mechanical beam actuation. Common mechanical switch devices in our daily life are soft drink cup lid and a toggle switch such as a light switch (**Figure 1.15**). Those devices are utilizing beam bucking to achieve bi-stable states.

(a) **(b)**

Figure 1.15. Example of mechanical switch memory in daily life. (a) Soft drink cup lid having choice button and (b) light toggle-switch.

For electronics device, first mechanical memories was introduced in early 1990s [15]. The device, mechanically bi-stable bridge, uses a same mechanism with the toggle switch as shown in **Figure 1.16**. An electrostatic force actuates the beam deformation and a read-out is done via measuring capacitances of each bi-stable state. But, the memory cell size is too large and it requires extremely high voltage (> 50V) bias for operation.

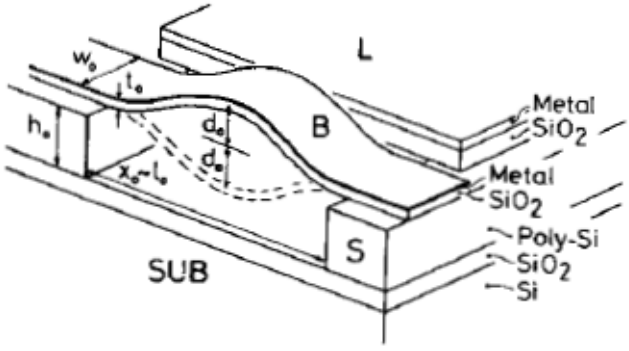
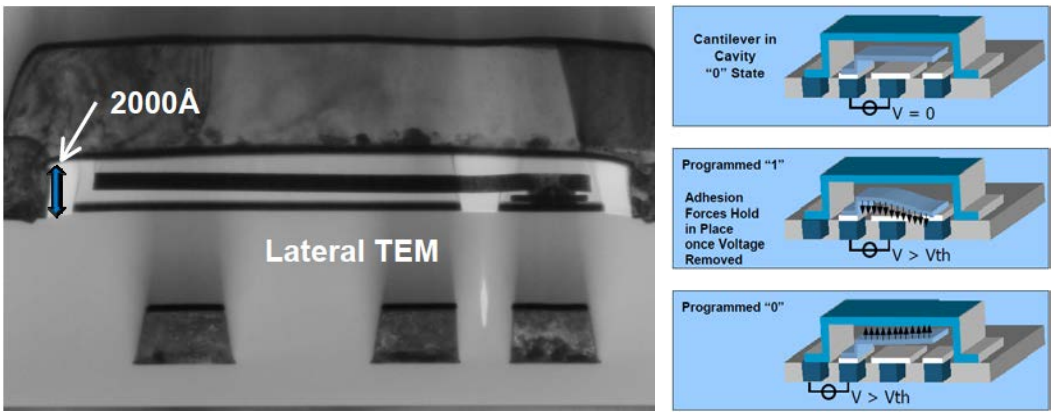


Figure 1.16. Illustration of micro-electro-mechanical non-volatile memory cell in *Ref.* [15].

More recently, Cavendish Kinetics, introduced nano-mechanical NVM technology by using CMOS back-end process [16]. **Figure 1.17 (a)** shows a TEM image of a fabricated nano-mechanical NVM cell and **Figure 1.17 (b)** illustrates the operating principle of the cell; for program ("1"), a bias is applied on the gate so that the electrostatic force pull down the beam. Once the gate voltage is removed, the adhesion force hold the beam in place and makes an electrical connection between drain and source. To erase ("0"), voltage applied to the shell will pull back up the beam.

The nano-mechanical NVM is immune to cosmic-radiation, wide range temperature variation and mechanical shock, so that it can be used for electrical fuse, power switching, metal logic. However, the nano-mechanical NVM technology has significant drawbacks; the device size is too large (~18F^2) and require 4 electrodes. In order to make large capacity cross-point array, this mechanical switching memory cell requires a selector device within each memory cell to reduce unwanted leakage current through unselected cells during a read operation. This additional selector devices make process very difficult.



(a) **(b)**

Figure 1.17. (a) TEM image of Nano-Mechanical NVM cell using CMOS back-end process and (b) Illustration of the operation of Nano-Mechanical NVM cell in *Ref.* [17].

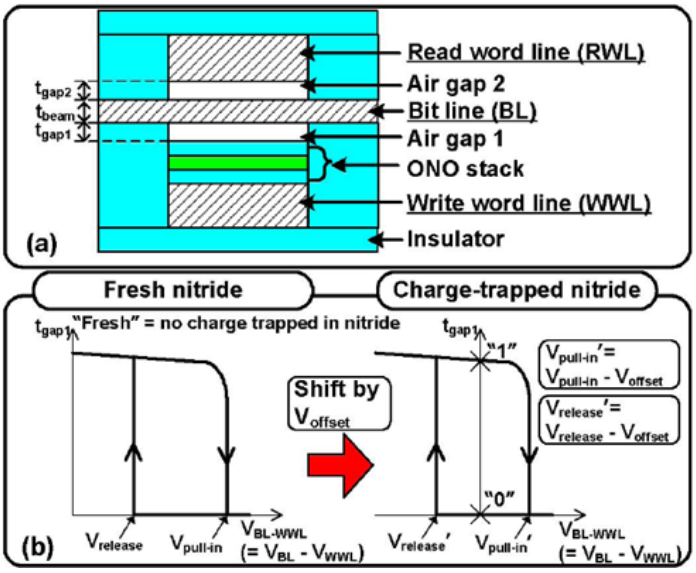


Figure 1.18. (a) Schematic cross-section of NEMory cell and (b) Fundamental theory of NEMory operation in Ref. [18].

Another electro-mechanical non-volatile memory cell design was recently proposed and demonstrated [18]. This design eliminates the need for a selector device, by leveraging the hysteretic behavior of a mechanical gap-closing actuator as shown in **Figure 1.18**. Although this design is well-suited for a cross-point array architecture, it requires separate read and write word-lines, as well as an initial "forming" step (charging of a dielectric layer) to achieve non-volatile operation.

1.5 Research Objectives and Thesis Overview

This dissertation suggest new innovative memory technologies to overcome issues and limitations of previous conventional DRAM and Flash memory technologies. Detailed simulation, theoretical analysis, and/or device fabrication of the new memory designs are performed.

In Chapter 2, a capacitor-less DRAM memory device utilizing a doubly gated vertical channel is proposed. The operations of the doubly-gate-vertical-channel (DGVC) DRAM cell are studied using 3-D device simulations. Scalability down to 22nm half-pitch of DGVC with $4F^2$ cell size is discussed.

In Chapter 3, a new NAND Flash memory cell is proposed. The new NAND Flash cell utilizes a back-gated thin body structure for a read disturb-free and more scalable structure. Through 3-D device simulation work, programming and erase operation are studied. To avoid disturbance problem, a novel negative back-bias scheme is suggested.

In Chapter 4, an electro-mechanical diode NVM cell is proposed. The fabrication process flow for the electro-mechanical diode NVM cell is provided. A proto-type cells are electrically evaluated. The evaluation results show that this new NVM cell design is adventurous for next generation non-volatile memory technology.

Chapter 5 summarizes the contributions of this work and offers directions for further research.

1.6 References

[1] S. W. Miller, "*Memory and Storage Technology*," by AFIPS Press, Montvale, 1977.

[2] S. Graham, "*Random Access Memory*," by French Bread Publication, Carlsbad, 2006.

[3] W.D. Brown and J. Brewer, "Nonvolatile Semiconductor Memory Technology," by Wiley-IEEE Press, 1997.

[4] R.H. Dennard, "Field-Effect Transistor Memory", U.S. patent 3,387,286, 1968.

[5] Kimam Kim, "Technology for sub-50-nm DRAM and NAND Flash Manufacturing", in *IEEE International Device Meeting (IEDM)*, 2005

[6] S.Okhonin, M.Nagoga, J.M.Sallese, P.Fazan, "A SOI capacitor-less 1T-DRAM concept," in *2001 IEEE International SOI Conference*, pp. 153–154, 2001.

[7] H. Wann and C. Hu, "A Capacitorless DRAM Cell on SOI Substrate", in *IEDM*, pp 635-638, 1993.

[8] C. Kuo, T.-J. King, and C. Hu, "A Capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications," in *IEEE Electron Device Letters*, Vol. 23, pp. 345-347, 2002.

[9] F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, and S. Tanaka, "A new flash E2PROM cell using triple polysilicon technology," in *IEDM*, pp. 464-467, 1984.

[10] Intel, "Intel Discloses New NOR Flash Memory Technologies, Sets Stage For 2005", on http://www.intel.com/pressroom/archive/releases/2005/20050302net_a.htm," 2005.

[11] Toshiba, "Flash Memory History," on http://www.toshiba.co.jp/worldwide/about/history.htm," 2008.

[12] H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase change memory," in *Proc. IEEE* Vol. 98, No. 12, pp. 2201-2227, 2010.

[13] H. Akinaga and H. Shima, "Resistive Random Access Memory (ReRAM) based on metal oxides," in *Proc. IEEE* Vol. 98, No. 12, pp. 2237-2251, 2010.

[14] G. W. Burr, B. N. Kurdi, J. C. Scott, C. H. Lam, K. Gopalakrishnan, R. S. Shenoy, "Overview of candidate device technologies for storage-class memory", in *IBM Jounal of Research and Development*, pp. 449-464, 2008.

[15] B. Hälg, "Micro-Electro-Mechanical Non-volatile Memory Cell," in *IEEE Trans. Electron Devices*, Vol. 37, pp. 2230-2236, 1990.

[16] M. A. Beunder et al, "A New Embedded NVM Technology for LowPower, High Temperature, Rad-Hard Applications," in *Non-Volatile Memory Technology Symposium*, 2005.

[17] V. Joshi et al., "A Non Volatile MEMS Switch for Harsh Environment Memory Applications," on *http://www.avsusergroups.org/tfug/tfug2009_10proceedings.htm*

[18] W. Y. Choi, H. Kam, D. Lee, J. Lai and T.-J. K. Liu, "Compact Nano-Electro-Mechanical non-volatile memory (NEMory) for 3D integration" in *IEDM Tech. Dig.*, pp. 603-606, 2007.

Chapter 2

Highly Scalable Capacitor-less DRAM Cell having a Doubly Gated Vertical Channel

2.1 Introduction

Dynamic RAM (DRAM) technology has evolved greatly over the past few decades. In the early 1990s, the highest DRAM capacity was only a few tens of megabits. Currently, DRAM chips have larger capacity than one-gigabits. The scaling of DRAM cell size has played a key role in the amazing progress in DRAM technology. However, since the traditional DRAM cell (1T-1C) has a storage capacitor and an access transistor, DRAM manufacturers face a tremendous challenge to shrink cell area for sub-50-nm technologies [1].

In the early 2000s, capacitor-less DRAM memories were proposed to eliminate the capacitor from the conventional 1-transistor/1-capacitor DRAM cell [2, 11]. Mainstream capacitor-less DRAM cells have a floating body structure (FBC) on a partially depleted (PD) SOI substrate. Although the elimination of the capacitor reduces the cell area overhead and fabrication process complexity, FBC devices have technological challenges for stand-alone memory applications. First, the SOI wafer is too expensive for the mass production of memory. Even though the SOI wafer price has dropped continuously, the SOI wafer cost is still 4 times larger than that of a bulk silicon wafer. Thus, no memory manufacturers are producing of DRAM using SOI wafers. Second, it is hard to meet basic performance requirements with a very short channel. The basic requirement for a DRAM cell is to maintain a sufficient data sensing margin between the two states for the duration of a refresh cycle (retention time). When the channel length is scaled down, it is difficult to guarantee sufficient retention time due to short channel effects such as drain-induced barrier lowering (DIBL). To suppress the short channel effects, a very thin body structure is needed. However, it is hard to mass product thin-SOI wafers with good qualities.

The double-gate DRAM (DG-DRAM) cell was proposed as a more scalable (to short channel length) design for capacitor-less DRAM [3, 4]. Unlike conventional double-gate logic transistors, the two gates in double-gate DRAM cells are not electrically connected. This structure permits an electrically induced floating body to occur within a thin fully

depleted body. Excess holes are collected at the back channel interface and increase the read current. When a negative back gate bias is applied, it creates a potential-well that permits the retention of holes in the body. Since a double-gate device can terminate electric field lines from the drain effectively, the double-gate cell is able to suppress short channel effects and maintain sufficient retention time. An independently controlled double-gate structure enabled the improvement of memory characteristics achieving the best retention characteristics at low voltages for short gate length. Recently, a DG-DRAM cell was successfully demonstrated with 80-nm gate length and 30-nm fin width [5]. Nevertheless, the DG-DRAM has considerable drawbacks due to a relatively large cell size ($8F^2$), difficulties with integration into conventional memory processes, and lack of disturbance-free bias conditions within a cell array.

In order to overcome the deficiencies of previous DG-DRAM technology, a double-gate array having vertical channel structure (DGVC) with $4F^2$ cell size is proposed. This new structure can be fabricated on a bulk silicon wafer using the conventional memory process flow for stand-alone DRAM application. The operation and scalability of the DGVC cell are demonstrated via Sentaurus [6] 3-dimensional device simulations. Read and write disturbances can be avoided by using appropriate biasing conditions for operation within the cell array.

2.2 Device Architecture and Proposed Fabrication

Depictions of the DGVC device are shown in **Figure 2.1**. Each cell has a drain junction on the top of the vertically oriented channels, while all of the cells within the array share a common source junction located at the bottom of the vertical channels. Bit lines connect the drain junctions via contacts, one for each row of cells. Word lines runs orthogonally to the bit lines, and each word line serves to gate the channels located on either side of it. Thus, each cell shares its front gate with an adjacent cell positioned along the same bit line, and each cell shares its back gate with the opposite adjacent cell. The circuit schematic of a DGVC cell is shown in **Figure 2.2**. This array configuration provides the most efficient cell layout area, $4F^2$.

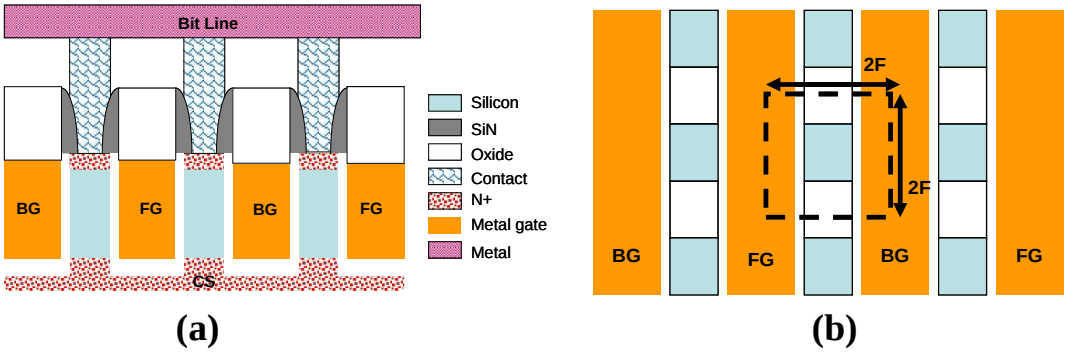


Figure 2.1. (a) Schematic cross-sectional view of DGVC cells (parallel to the bit line). (b) Schematic plan view of DGVC array.

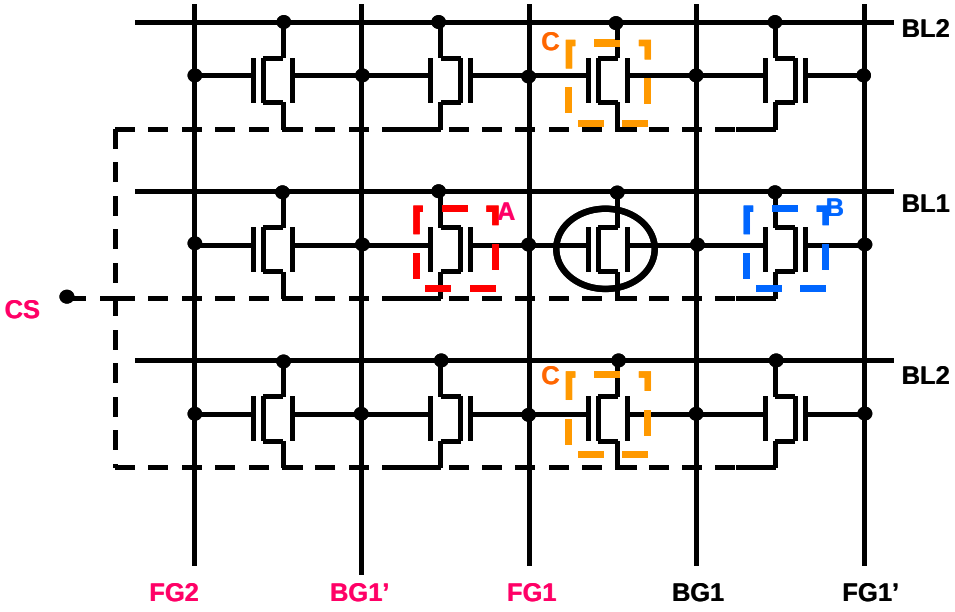


Figure 2.2. Circuit schematic of a DGVC cell array.

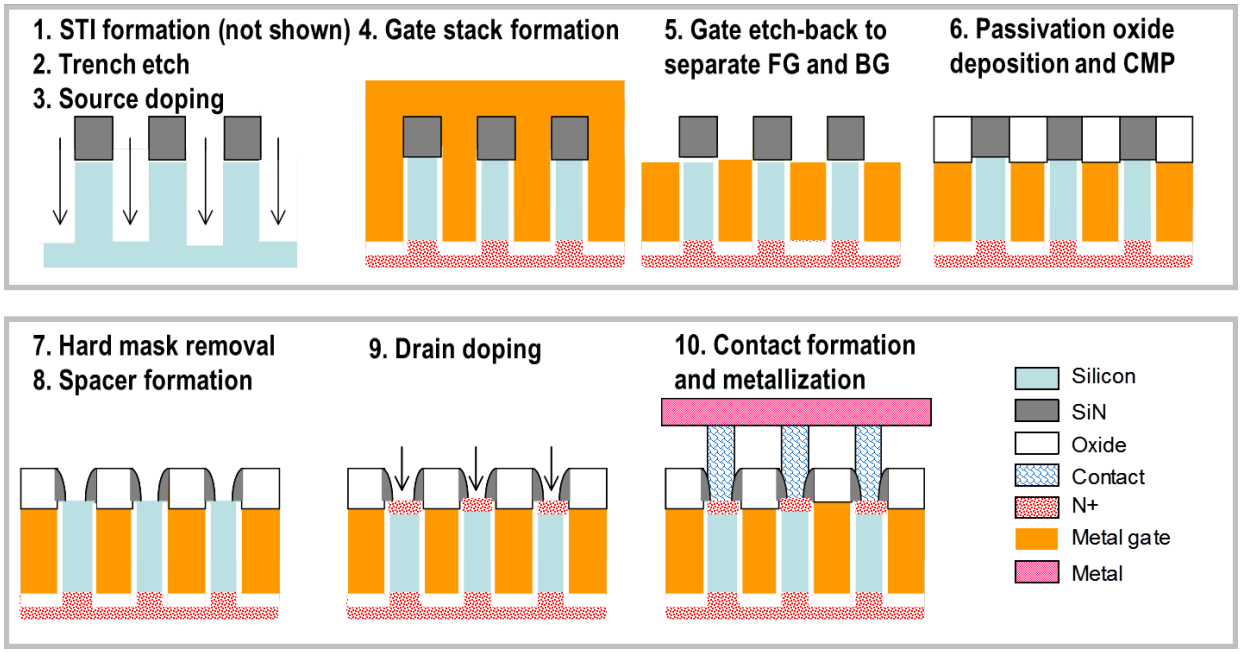


Figure 2.3. Key process steps for DGVC array: 1) STI formation along the bit line direction (not shown here), 2) Trench formation and buried source implantation (not shown), 3) Gate material deposition, 4) Gate etch-back, 5) Capping oxide deposition and CMP, 6) Nitride removal, spacer formation, 7) drain implantation, and 8) Contact formation.

The proposed fabrication steps for the DGVC cell array are illustrated in **Figure 2.3**. First, shallow trench isolation regions are formed to define stripes of silicon along the bit line direction. After deposition of a nitride hard mask and patterning a photoresist mask to form lines running perpendicular to gate lines, the etching process cuts into the oxide and silicon layers to make trenches to form self-aligned vertical channel structures along the word line direction. Gate-electrode material (Poly-silicon or metal) is deposited in the trenches, and the etch-back process of the deposited gate material forms separate front- and back-gate lines. A capping oxide is deposited and chemical mechanical polishing (CMP) of the oxide is undertaken. After removing the nitride hard mask and silicon-nitride spacer formation, drain junctions are formed through ion implantation, and bit line contacts are formed on each drain junction. Recently, a number of cells resembling our structure, except for the DGCV's unique array structure that the cell shares a front and back gate with an adjacent cell, were successfully demonstrated [7, 8]. Experimental data of a double-gate DRAM cell having a vertical channel was also reported in [12].

2.3 Cell Operation

To investigate the operation of a DGVC cell, three-dimensional devices simulations were performed using SENTAURUS for the operation conditions specified in **Table 2.1**. Physical design parameters (**Table 2.2**) were chosen as appropriate for a 22-nm technology cell. All transition times in the simulation are set to 1ns, and light uniform doping is used in the fully depleted body. The source and drain junctions have a typical Gaussian doping profile, and a gate work function of 4.5eV is assumed, as appropriate for a metallic gate material..

	Write "1"	Write "0"	Hold	Read
FG1	1.5V	0.0V	-1.0V	1.5V
BG1	-1.0V	0.0V	-1.0V	-1.0V
BL1	2.0V	0.0V	0.5V	0.7V
CS	0.5V	0.5V	0.5V	0.5V
BL2	0.5V	0.5V	0.5V	0.5V
FG1'	-1.0V	-2.0V	-1.0V	-2.0V
BG1'	-3.0V	-2.0V	-1.0V	-3.0V
FG2	-1.0V	-1.0V	-1.0V	-1.0V
Duration	20ns	20ns	100ms	20ns

Table 2.1. Cell operating conditions.

Write "1" is performed by introducing holes into the fully depleted body. A high drain bias creates electron-hole pairs near the drain junction along the front-gate interface. The generated electrons are rapidly swept out of the cell to the n-type drain. On the other hand, the generated holes are collected in a potential well that is formed by a negatively biased back gate. The concentration of holes in the body increases above the equilibrium value and thereby increases the drain current from its initial state value; the number of accumulated holes determines the state of the cell. Programming times of a few nanoseconds are sufficient to saturate the body with holes.

The accumulated holes must be removed in order to Write "0". When the back-gate is grounded, the potential barrier at the body-drain junction is reduced, allowing holes to diffuse into the drain. With the reduction in the number of stored holes, the source potential barrier height increase until it reaches its equilibrium level. This self-convergent process is very useful to avoid a disturbance of Write "0" operation [4].

Parameters	
Technology node = 22nm	Cell Size = 0.000484 um^2 (4F^2)
N_{body} = 10^{16} cm^{-3} (p-type)	$N_{S/D}$ = 10^{20} cm^{-3} (n-type)
T_{ox} = 40 Å	T_{body} = 200 Å
$L_{channel}$ = 50nm (Si-pillar height)	W_{active} = 22nm (Si-pillar width)
$\tau_{electron}$=1.5 μsec	τ_{hole}=1.5 μsec

Table 2.2. 22nm DGVC cell design parameters

The state of the cell is read by sensing the drain current. If holes are stored in the body, the threshold voltage will be low, and the drain current will be high. On the other hand, if holes are removed from the body, the threshold voltage will be high and the drain current will be low. The current difference between two states is determined by the difference in stored holes within the body.

To hold the state of the cell, the front- and back-gates are biased at negative voltage, and a small positive voltage is applied at the source and the drain to form a potential well to retain holes in the body region. These bias conditions create large hole potential barriers within the body. The previously written state is retained for a refresh cycle, and the retention time of that state is determined by charge-loss mechanisms. These include *Shockley-Read-Hall* recombination/generation (SRH R–G) diffusion to the source/drain junction leakage current, and band-to-band tunneling (BTBT). In the case of the double-gate cell, BTBT does not significantly affect the retention time because the asymmetric gate potential enables the reduction of electric field in the gate overlap region and thereby suppresses BTBT [9]. A minority carrier lifetime of 1.5-μsec is used for the simulations and well-matches measured data [10].

The simulated retention characteristics of a DGVC cell are shown in **Figure 2.4**. The simulation predicts that a measurable difference (> 2μA) in drain current can be sustained for 10sec at 27°C, and for 200msec at 85°C. For stand-alone memory applications, longer retention time (>10sec at 85°C) is required to account for worst case variation in a large density memory array. Thus, further optimization work is needed to achieve longer retention time characteristics.

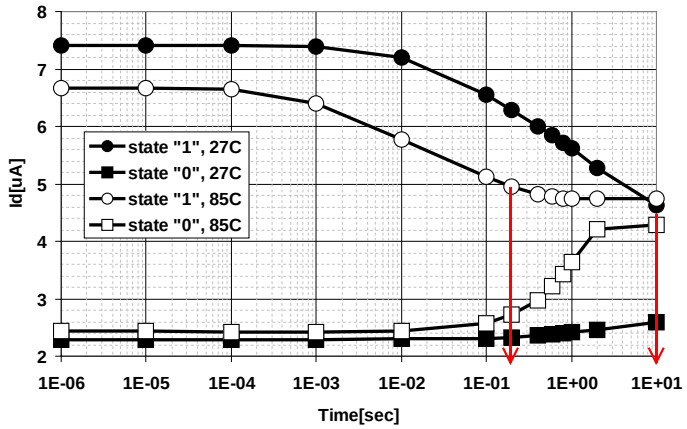


Figure 2.4. Simulated retention characteristics of a DGVC cell, where the read and write times are 20nsec.

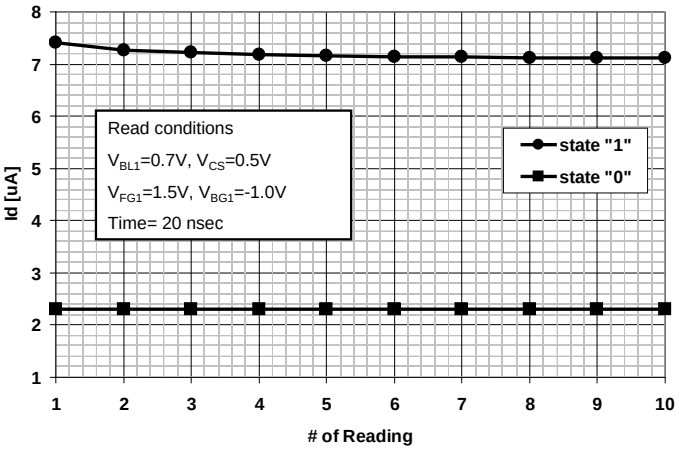


Figure 2.5. Simulated non-destructive readout characteristic of a DGVC cell.

Figure 2.5 shows a *quasi-non-destructive* readout characteristic during repeat read operation. The capacitor-less DRAM cell storing holes in the body shows almost nondestructive readout [13, 14, 15.]. However, in a real device, there are many interface trap sites at Si-SiO_2 interfaces which enhance recombination in the back-interface region. Thus, charge loss in a real capacitor-less DRAM cell is due primarily to recombination at trap sites which eliminates the holes accumulated in the body during repeat readout operation [16]. In this simulation work, the trap assisted recombination model of dangling bonds model is not included due to the lack of proper experimental data, which means that a real DGVC cell is expected to need a refresh operation after a finite number of readout cycles. However, the charge losses of a DGVC cell are much smaller than that of a conventional DRAM cell, which makes it possible to readout data without a refresh function in every cycle.

2.4 Disturbances in the Cell Array

In the previous section, the basic cell operations were described. To form a high capacity array, the cell disturbance characteristics should be investigated. **Figure 2.6** shows simulated disturbance characteristics under various bias conditions. During the writing of "0" state on the selected cell, the unselected cell, "**C**," (**Figure 2.2**) undergoes an undesirable disturbance. In previous research [2, 8], a negative bit line voltage was used to apply forward bias on the drain-body diode to sweep out the accumulated holes during a Write "0" operation. This causes inevitable disturbances in bit line cells connected to a specific bit line. Instead of using a forward diode turn-on mechanism, a self-convergent purge mechanism is proposed to avoid severe disturbance.

When the drain bias and back-gate bias are set to zero (V_d=0 and V_{bg}=0V) while applying a small positive source bias (V_s=0.5), the potential barrier between the drain and the body at the back-channel interface is reduced, and excess holes are released from the back interface to the drain side. As holes are removed from the body, the potential barrier increases until zero hole-current is eventually reached. These bias conditions can provide a sufficient sensing margin within the total disturbance time (5μsec). Disturbances on unselected cells (ref. "**A**" and "**B**" in **Figure 2.2**) along the same bit line can be avoided by applying negative gates biases which allow the potential barriers to retain their holes (**Figure 2.6 (a)**)

During a Write "1" operation, unselected cells connected to the same bit line (ref. "**A**" and "**B**" in **Figure 2.2**) also undergo disturbances. Such disturbances can be avoided by appropriately biasing the front- and back-gate to maintain the unselected cells in the off state (**Figure 2.6 (b)**).

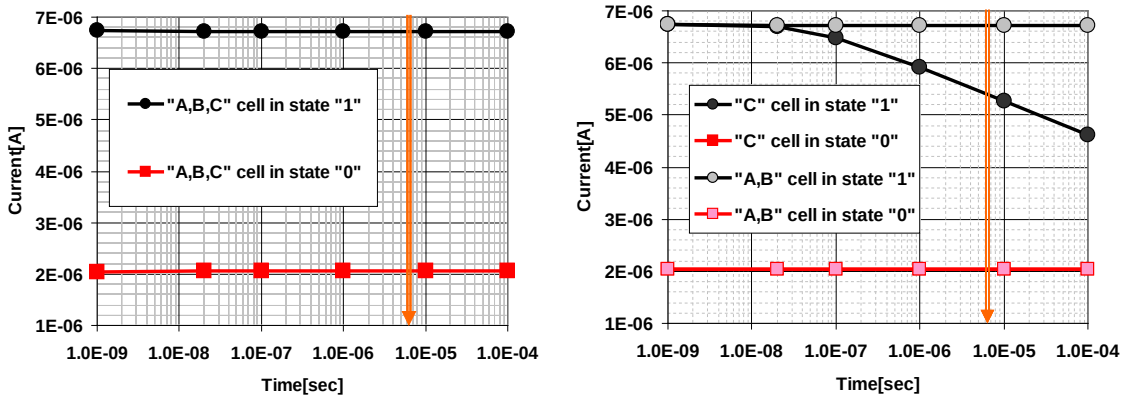


Figure 2.6. (a) Write "0" disturbance on the unselected cells. (b) Write "1" disturbance on the unselected cells. The indicated total disturbance time (5μsec) is derived by assuming 20nsec read time and 128(WL) x 256(BL) array.

2.5 Scalability of DGVC

In order to investigate the scalability of a DGVC cell, read currents were simulated for cells with scaled channel lengths and body thickness. **Figure 2.7 (a)** shows the read current dependency on channel length in each state, and **Figure 2.7 (b)** shows how the sensing margins depend on channel length. The optimal channel lengths are 50-nm, 80-nm and 100-nm for body thickness of 18-nm, 22-nm and 32-nm respectively. These results agree with previous work [9]. As the channel length is scaled, drain-induced-barrier-lowering (DIBL) causes a current increase in the "0" state. On the other hand, as the channel length increases, there is a reduction in the current of cell state "1". To reduce DIBL, a thinner body thickness is desirable. A thin body thickness also increases the body coefficient so that it can enhance sensing margin between the "0" state and the "1" state [4], [9]. As shown in **Figure 2.7**, a reduced body thickness diminishes the DIBL effect and widens the sensing margins. Thus, a thin body thickness is key to achieving shorter optimized channel lengths that maximize the sensing margin.

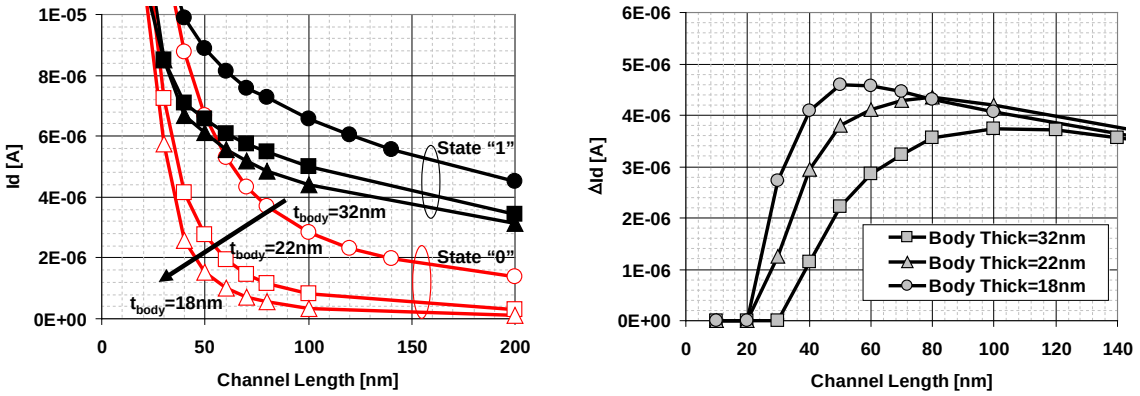


Figure 2.7. (a) Readout currents for State "1" and State "0". (b) Current differences between "1" and "0" states. Thinner body thickness suppresses the DIBL effect and provides for shorter optimal channel length that increases the sensing margin. The read and write times are 20nsec.

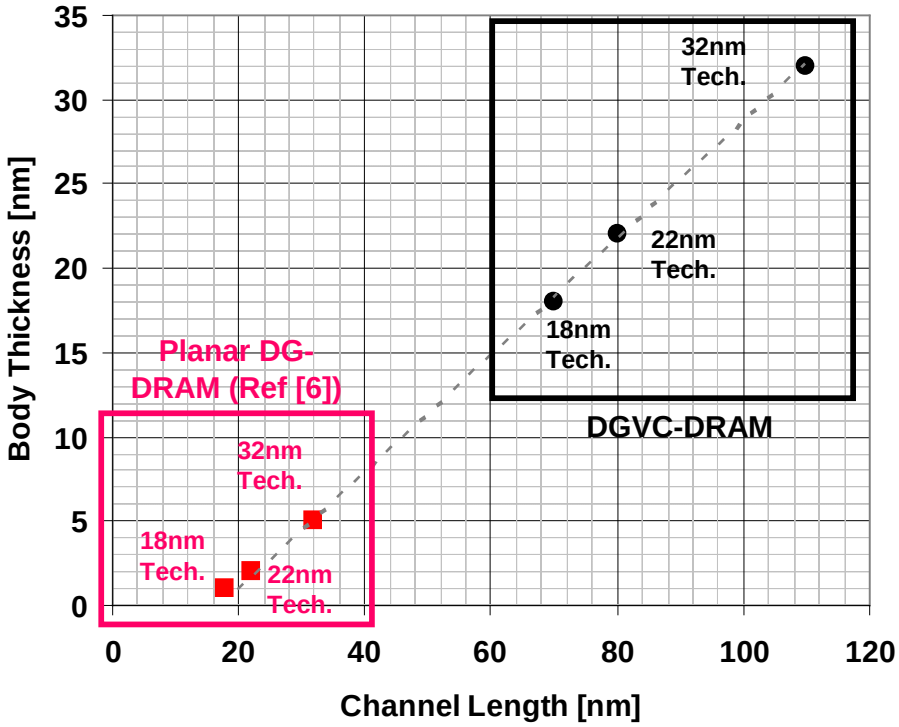


Figure 2.8. The required dimensions of the DGVC cell design vs. the planar DG-DRAM cell design for the same technology nodes.

Figure 2.8 shows the required dimensions of the DGVC cell design vs. the planar DG-DRAM cell design for the same technology nodes. Each point represents a required body thickness for a given channel length. A previous research paper reported that the channel length scaling of a planar DG-DRAM cell is limited to approximately 25-nm due to quantum confinement effects that become significant below 4-nm body thickness [9]. In contrast, the DGVC cells have longer channel lengths and greater body thickness in the same technology nodes. Thus, the DGVC cell design derived from the unique array structure increases the opportunities for DRAM scaling for higher density storage.

2.6 Summary

A doubly gated vertical channel transistor is proposed for highly scalable DRAM memory. This technology allows us for a scalable $4F^2$ cell size design, which can be fabricated with a conventional process flow on silicon- bulk instead of expensive SOI wafer. Cell operations (Read, Write, and Hold) and disturbance characteristics are investigated by 3-dimensional device simulations of 22-nm technology. Disturbance immunity characteristics of a DGVC cell are shown to be adequate. Since the vertical channel design allows for longer channel length, and hence a greater body thickness is allowed, it is possible to scale DRAM memory with a $4F^2$ cell area down to 22-nm minimum half-pitch and beyond.

2.7 References

[1] Kimam Kim, "Technology for sub-50-nm DRAM and NAND Flash Manufacturing", in 2005 IEDM, 2005.

[2] S.Okhonin, M.Nagoga, J.M.Sallese, P.Fazan, "A SOI capacitor-less 1T-DRAM concept," in 2001 IEEE International SOI Conference, pp. 153–154, 2001.

[3] Hsing-jen Wann and Che-nming Hu, "A Capacitorless DRAM Cell on SOI Substrate", in 1993 IEDM, pp 635-638, 1993.

[4] Charles Kuo, Tsu-Jae King, and Che-nming Hu, "A Capacitorless double gate DRAM technology for sub-100--nm embedded and stand-alone memory applications," in IEEE Electron Device Letters, Vol. 23, pp. 345-347, 2002.

[5] Ibrahim Ban, Uygar E. Avci, Uday Shah, Chris E. Barns, David L. Kencke, and Peter Chang, "Floating Body Cell with Independently-Controlled Double Gates for High Density Memory", in IEDM, 2006.

[6] Sentaurus Ver. A-2008.09, SYNOPSYS.

[7] T. Schloesser, F. Jakubowski, J. v. Kluge, A. Graham, S. Slesazeck, M. Popp, P. Baars, K. Muemmler, P. Moll, K. Wilson, A. Buerke, D. Koehler, J. Radecker, E. Erben, U. Zimmermann, T. Vorrath, B. Fischer, G. Aichmayr, R. Agaiby, W. Pamler, T. Schuster, W. Bergner, W. Mueller, " 6F2 buried wordline DRAM cell for 40-nm and beyond", in 2008 International Electron Devices Meeting, 2008.

[8] Hoon Jeong, Yeun Seung Lee, Sangwoo Kang, Il Han Park, Woo Young Choi, Hyungcheol Shin, Jong Duk Lee, and Byung-Gook Park, "Capacitorless DRAM Cell with Highly Scalable Surrounding Gate Structure" in the 2006 International Conference on Solid State Devices and Materials, pp. 574-575, 2006.

[9] Nauman Z. Butt and Muhammad Ashraful Alam, "Scaling Limits of Double-Gate and Surround-Gate Z-RAM Cells" in IEEE ELECTRON DEVICE LETTERS, Vol. 54, pp. 2255-2262, 2003.

[10] Naoki Yasuda, Kenji Taniguchi, Chihiro Hamaguchi, Yasuo Yamaguchi, and Tadashi Nishimura, “New Carrier Lifetime Measurement Method for Fully Depleted SO1 MOSFET’s” in IEEE ELECTRON DEVICE LETTERS, Vol. 39, pp. 1197-1202, 1992.
[11] T. Ohsawa, K. Fujita, T. Higashi, Y. Iwata, T. Kajiyama, Y. Asao, and K. Sunouchi, “Memory design using one-transistor gain cell on SOI,” in Proc. ISSCC 2002, pp. 152–153, 2002.
[12] Ertosun M.G., Hoon Cho, Kapur P., Saraswat K.C., “A Nanoscale Vertical Double-Gate Single-Transistor Capacitorless DRAM”, in Electron Device Letters, IEEE, Vol. 6, P.615 – 617, 2008.
[13] Takashi Ohsawa, Katsuyuki Fujita, Tomoki Higashi, Yoshihisa Iwata, Takeshi Kajiyama, Yoshiyuki Asao, and Kazumasa Sunouchi, “Memory design using a one-transistor gain cell on SOI”, IEEE Journal of Solid-State Circuits, Vol. 37, 2002.
[14] Ranica R., et all, “A one transistor cell on bulk substrate for low-cost and high density eDRAM”, VLSI Technology, 2004.
[15] Kosuke Hatsuda et all, “A 333MHz random cycle DRAM using the floating body cell”, IEEE 2005 Custom Integrated Circuits Conference.
[16] T. Ohsawa et all, "Design of a 128-Mb SOI-DRAM using the floating body cell (FBC)," IEEE Journal of Solid-State Circuits, vol. 41, no. 1, pp. 135-145, 2006.

Chapter 3

A New NAND Flash Memory Cell Design Utilizing a Back Gate Structure

3.1 Introduction

For nonvolatile storage, the concept of a charge trapped memory cell design was first conceived in the 1960s and initial commercial devices were demonstrated in the early 1970s [1], [2]. SONOS, short for "Silicon-Oxide-Nitride-Oxide-Silicon", is a type of charge trapped non-volatile memory cell design which is formed from a standard N-channel MOSFET transistor with the addition of a small sliver of silicon nitride inserted inside the transistor's gate oxide. SONOS devices required up to 30V of operating bias voltage because of a thick tunneling oxide and poor charge trapping efficiency of the nitride layer. After a conventional floating-poly-gate flash memory cell design was introduced in the late 1980s with relatively low operation voltage (< 20V), the popularity of SONOS technology diminished [3]. One of the conventional floating-poly-gate flash memories, NOR-type flash, replaced older read-only-ram (ROM). NAND-type flash has intruded into the data storage media market as a disruptive technology from the late 1990s; the market share of the solid-state-disk (SSD) driver which uses NAND flash memories is about 10 percent of total HDD market now [4]. However, the conventional NAND flash technology faces scaling challenges below 20nm. One of most severe challenges is the inherent cell-to-cell inference caused by capacitive coupling between the floating-poly-gates of adjacent cells.

To overcome scaling challenges in the conventional floating-gate flash technology, the charge trap flash (CTF) technologies are considered again for future NAND flash memory. The CTF technology maintains exactly same array architecture with the NAND flash array, except that it replaces floating-poly gate with a charge trapping layer such as a band-gap engineered silicon-nitride layer. An example of a CTF cell, TANOS (Titanium, Aluminum

oxide, silicon-Nitride, Oxide, Silicon), uses a high-K dielectric layer (i.e. aluminum oxide) to lower the operation voltage and block the leakage current from the silicon-nitride to the gate [5], [6]. Nevertheless, the structure of CTF imposes many restrictions on scaling of the devices because a trapping layer exists between a silicon channel region and gate material of the memory device. It results in a compromise between good electro-static integrity requirements and memory operation requirements; for example, a CTF memory device requires thick tunnel oxide to avoid read disturbance from high pass-gate voltages, but this results in poor electrostatic integrity which lead to low sensing current and large off-leakage currents.

The backside trapping structure was proposed as a more scalable (to short channel length) and read-disturbance-free design since it allows for independently optimizing the gate oxide and ONO stack as shown in **Figure 3.1** [7]. The device utilizes a modified silicon-on-insulator (SOI) substrate where charge is stored in a charge trapping region underneath a thin single-crystal silicon layer employed for the formation of the transistor channel. When it is used in a NAND array, however, it cannot avoid program disturbance problem and also cannot achieve high read currents.

In order to overcome these deficiencies, a new backside charge storage non-volatile memory (BCS-NVM) cell design is proposed [8]. A NAND flash array of the BCS-NVM cells can be fabricated on a modified SOI substrate. This design allows for a relatively high cell read current and steep sub-threshold slope to enable lower voltage operation in comparison with conventional NAND flash memory devices. In this chapter, operation of an 18nm BCS-NVM cell is studied by means of three-dimensional device simulations.

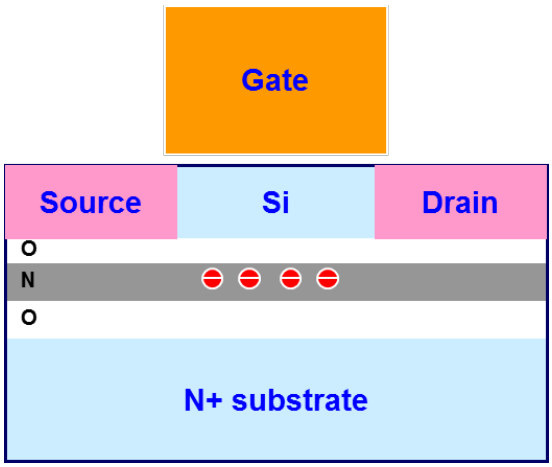
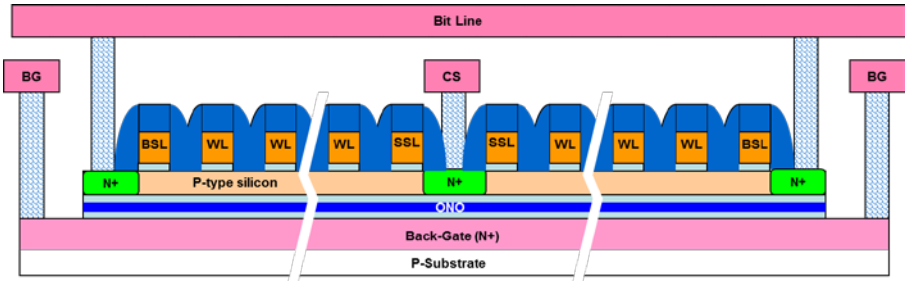


Figure 3.1. Schematic of the backside trapping memory and transistor structure in Ref. [7]. The structure employs an oxide–nitride–oxide (ONO) stack at the back of a thin single-crystal Si channel.

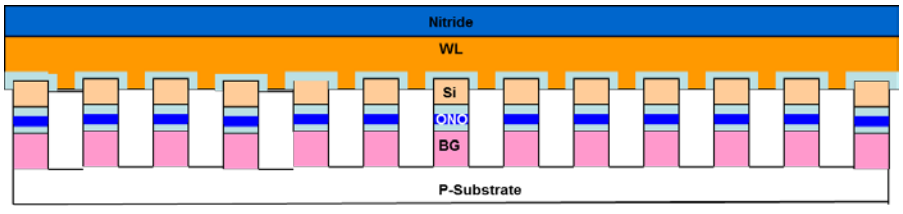
3.2 Device Structure

Schematic cross sections of a BCS-NVM cell array are shown in **Figure 3.2**. The cell has a thin front gate oxide, a thin active silicon channel layer and an underlying ONO charge trapping dielectric stack. The threshold voltage shift is determined by the amount of trapped-charges in the backside ONO dielectric. The effect of trapped charges in the backside ONO are almost same with the substrate bias effects in silicon-on-insulator (SOI) transistors. Thus, the threshold voltage of the BCS-NVM increases when electrons are injected into the nitride layer of the ONO stack. This unique structure allows independent scaling between the front-gate transistor part and the back-gate charge storage part; it allows efficient channel-length scaling of the memory device because the front-gate transistor part is similar to an SOI structure.

To form the NAND array, the cells are connected in series as in a conventional NAND flash memory array, except that there are no p-n junctions. Electrically induced source and drain regions are formed in-between the cells by the neighboring front-gate fringing electric fields during a Read operation [9]. The back-gate electrodes run orthogonally to the front-gate electrodes. They controls the amount of charges trapped in the ONO layer by applying high-voltage. Each back-gate electrode line defines an erase sector since all cells along the back-gate electrode can be erased simultaneously when the back-gate is used to apply the erase voltage. The circuit schematic of a BCS-NVM cell array is shown in **Figure 3.3**. This array configuration achieves the most efficient cell layout area of $4F^2$.



(a)



(b)

Figure 3.2. Schematic cross-sectional view of BS-NVM cells in a NAND array (a) parallel to the bit line, (b) parallel to the word line.

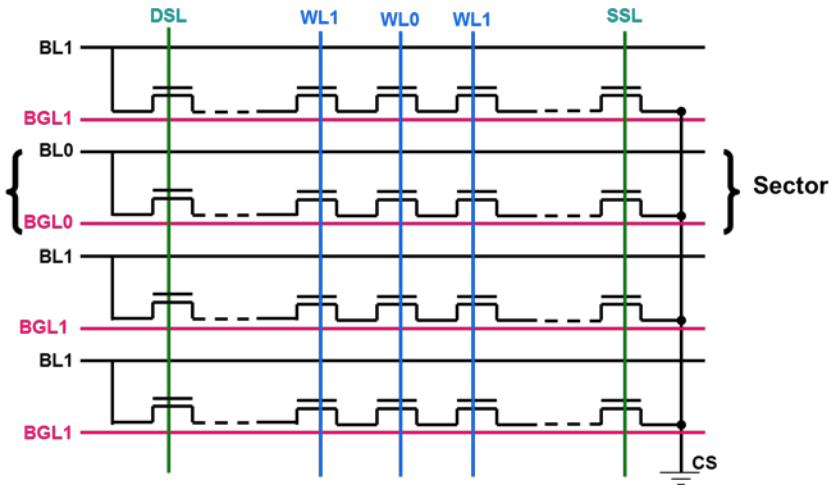


Figure 3.3. Circuit schematic of a BCS-NVM cell array.

3.3 Proposed Fabrication Process

3.3.1 Silicon-On-ONO-Substrate (SOONOS) Formation

The starting substrate for the BCS-NVM device fabrication consists of thin single-crystal silicon layer with an underlying charge trapping region (ONO) on a highly-doped n-type substrate. This silicon on ONO substrate (i.e. SOONOS) can be obtained using direct wafer bonding followed by a hydrogen-induced exfoliation for single-crystal silicon-layer transfer [7], [10]. **Figure 3.4** illustrates the process sequence for SOONOS substrate; the ONO layer is grown on the p-type substrate wafer. The silicon-nitride film serves as the charge trap layer. This "donor-wafer" is then implanted with high dose of H+ which leaves a layer of hydrogen micro-cavities underneath the ONO layer.

The second wafer, "host wafer", is highly doped with n-type impurities which forms an N+ conductive surface layer on the wafer initially. A silicon oxide layer is thermally grown on this wafer. Since the surface of grown oxide is quite smooth, it can form a strong bond between the "donor" and "host" wafers. For bonding the two wafers, a wafer bonder provides a force (over 1000N) at the interface of the wafers under vacuum. The bonding force is strengthened through a low-temperature anneal. Under higher temperature (400C) anneal, the hydrogen micro-cavities expand and cause the donor wafer to cleave. Thus, it leaves single-crystal silicon layer bonded onto ONO layer. This silicon layer can then be thinned and smooth by CMP followed by oxidation.

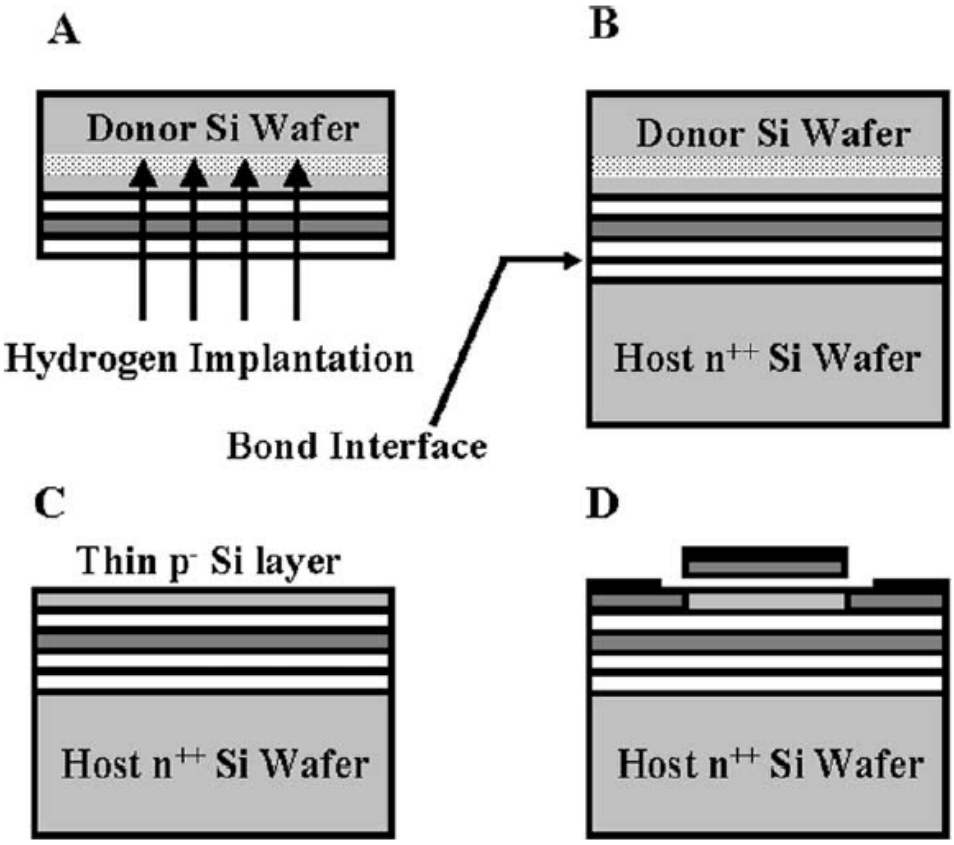


Figure 3.4. Fabrication process sequence for silicon-on-ONO substrate

3.3.2 Memory Device Fabrication

Figure 3.5 illustrates the front-end process steps for the BCS-NVM cell array. A shallow trench etch process is used to define the buried back-gate lines along the bit-line direction; hard mask material is deposited on top of the thin silicon and patterned by photolithography and then reactive ion etch (RIE) process is used to make trenches in-between the back-gate lines. The trenches are filled with gap-filing material followed by planarization process. After selectively removing the hard mask layer, the front-gate oxide is grown on the exposed thin silicon strips and then the front-gate material is deposited. Then the front-gate layer is patterned to form word lines, and gate-sidewall spacers are formed to fill the region in-between the cells. Finally, ion implantation and thermal annealing are used to form n-type regions for the bit-line and common source (CS) contacts.

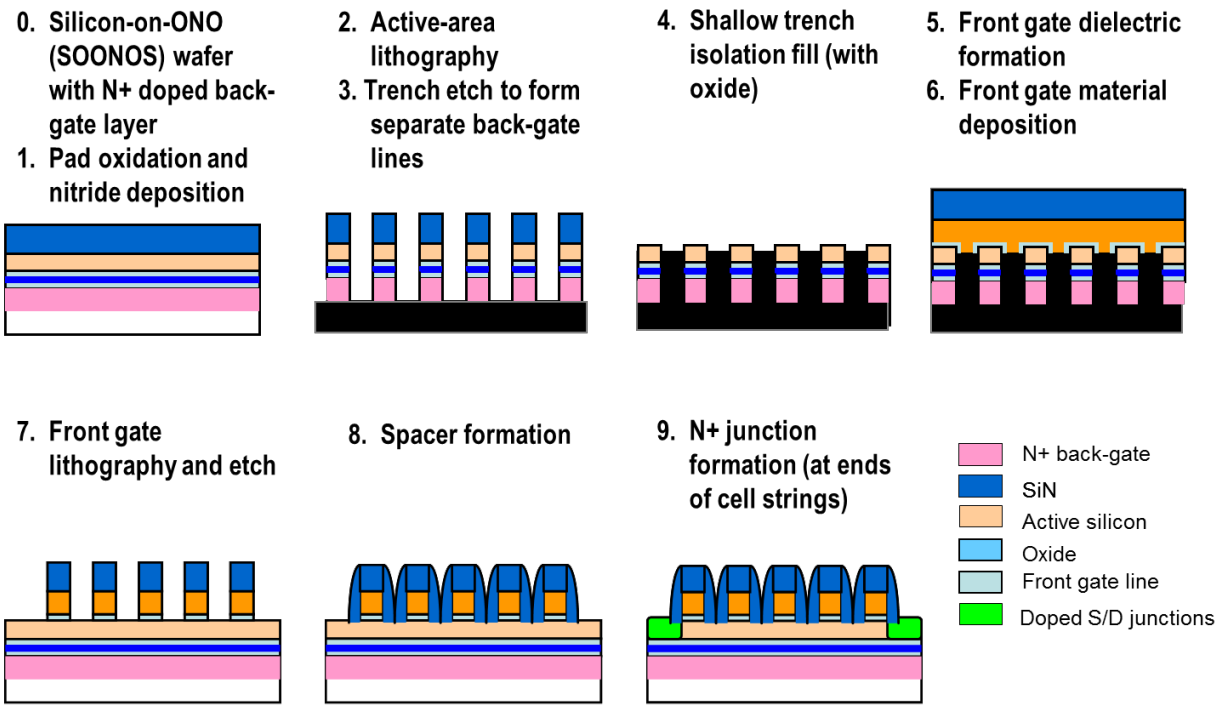


Figure 3.5. Front-end process steps, starting with a SOONO wafer with a heavily doped n-type junction underneath the ONO.

3.4. Device Simulation of the BCS-NVM Cell

The operation of a BCS-NVM cell was studied via 3-D device simulations. The physical design parameters appropriate for 18nm technology are shown in **Table 3.1**. The array bias conditions for Program, Erase, and Read operations are listed in **Table 3.2**. Electron/hole injection into the nitride layer by Fowler-Nordheim (FN) tunneling is used to program/erase a cell.

Parameters (18nm technology node)	
Cell Size = 0.000324 um^2 ($4F^2$)	$N_{body} = 10^{18}$ cm^{-3} (p-type)
T_{Gox} = 2 nm	$T_{silicon}$ = 7 nm
Tunnel oxide = 18 Å Storage nitride = 100 Å Blocking oxide E.O.T = 40 Å	
$L_{channel}$ = 18nm	W_{active} = 18nm

Table 3.1. BCS-NVM cell design parameters.

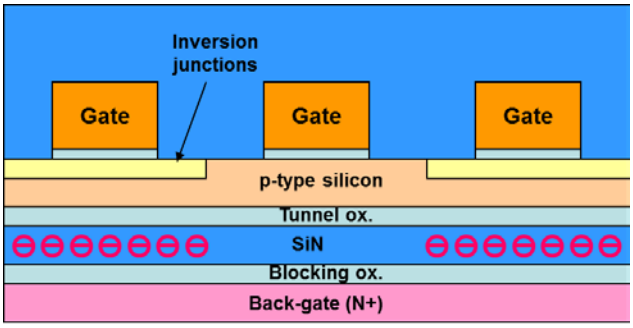
	Program	Erase	Read
WL0	0.0V	0.0V	0.0V
BGL0	12.0V	-12 V	0.0V
BL0	0.0V	0.0V	1.0V
WL1	-2.0V	0.0V	1.8V
BGL1	0.0V	0.0V	0.0V
BL1	0.0V	0.0V	0.0V
DSL	0.0V	0.0V	1.8V
SSL	0.0V	0.0V	1.8V
CS	0.0V	0.0V	0.0V
Time	2 ms	1 ms	< 10µs

Table 3.2. Array biases for Program/Erase/Read of the cell addressed by WL0 and BL0/BGL0 in Figure 3. 2.

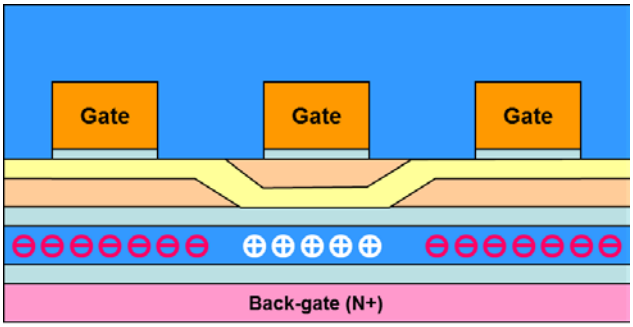
3.4.1 Basic Cell Operation

The state of a cell is determined by sensing the cell current during a Read operation. Figure 3.5 illustrates the Read operation of a BCS-NVM cell in a NAND array. If electrons are stored in the charge storage layer (e.g. nitride) underneath the cell, the cell threshold voltage is higher; when the Read gate voltage is applied, only off-state leakage current flows as shown in **Figure 3.6 (a)**. On the other hand, if no electrons are stored in the storage layer underneath the cell, the cell threshold voltage is lower; when the Read gate voltage is applied, on-state drive current flows as shown in **Figure 3. 6 (b)**.

Figure 3.7 shows the simulated I-V characteristics for a programmed and an erased cell. Each cell shows low subthreshold swing (~80mV/dec) and high on/off current ratio (~ 10^7). Note that the series resistance of the inversion-layer source/drain region (virtual source/drain) in-between the cells is acceptably low when the gate-to-gate spacing is very small (20nm or less) [9]. The maximum threshold voltage shift that can be achieved is determined by material parameters such as the silicon band gap and gate work function [10].



(a)



(b)

Figure 3.6. Illustration of the read operation of a BCS-NVM cell in a NAND array. (a) Programmed cell is off during Read, and (b) Erase cell is on during Read.

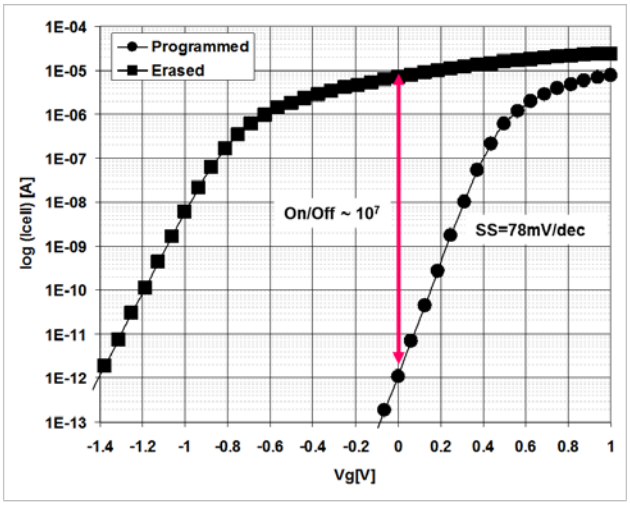
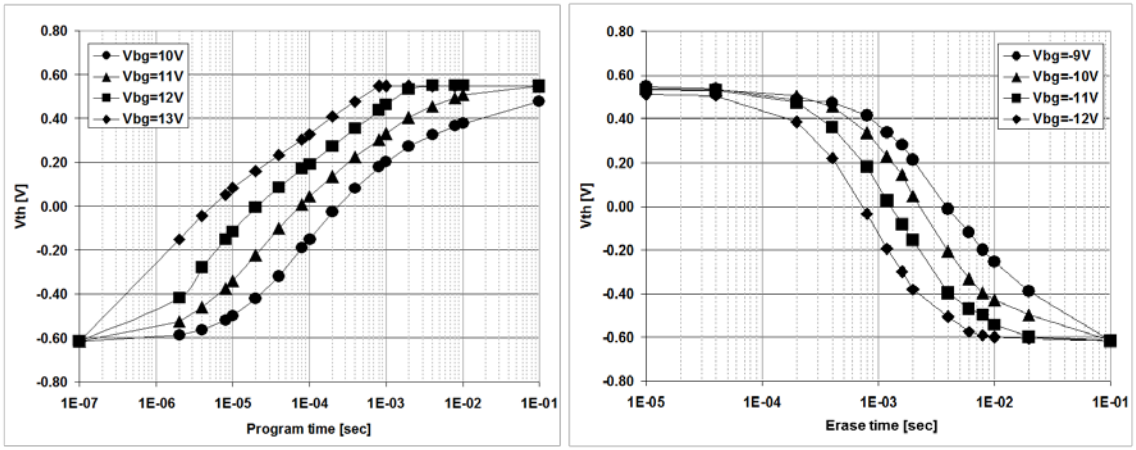


Figure 3.7. Simulated cell current vs. front-gate voltage characteristics for a programmed cell and erased cell

.



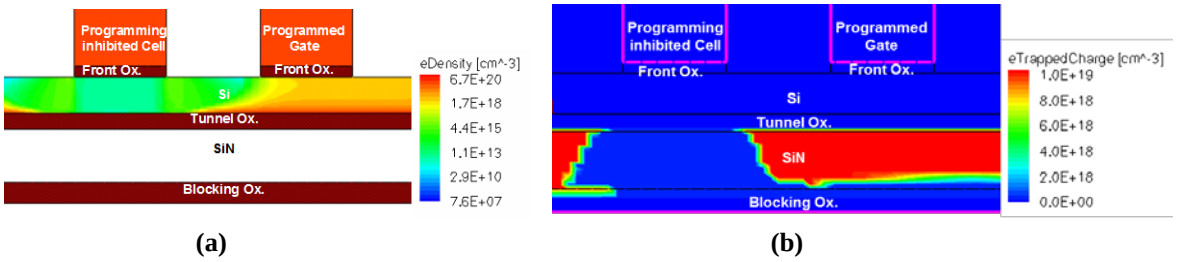
(a) **(b)**

Figure 3.8. Simulated BCS-NVM cell programming (a) and erase (b) behavior for various back-gate operating voltages.

Simulated programming and erase behaviors are shown in **Figure 3.8 (a)** and **(b)** respectively. The operation voltages (around 10V) are lower than those typically used for conventional NAND flash memory because of the relatively large capacitive coupling between the back-gate and the channel region.

3.4.2 Disturbance Immunities

A key requirement for implementation of a NAND cell array is immunity to cell disturbance during programming. A novel charge depletion approach is proposed herein to achieve this requirement. The front-gate is biased at a small negative voltage (V_{fg}=-2V) to deplete the channel region in a non-selected cell shown in **Figure 3.9 (a)**; this avoids electron injections into the charge storage layer despite a large positive voltage (V_{bg}=+12V) applied to the back-gate. **Figure 3.9 (b)** shows the simulated charge trapped density after programming operation on the programming inhibited cell. **Figure 3.10** shows the simulated program-disturbance characteristics of unselected cells along the same back-gate line.



(a) **(b)**

Figure 3.9. Contour plots during programming operation: (a) mobile electron density and (b) trapped electron density.

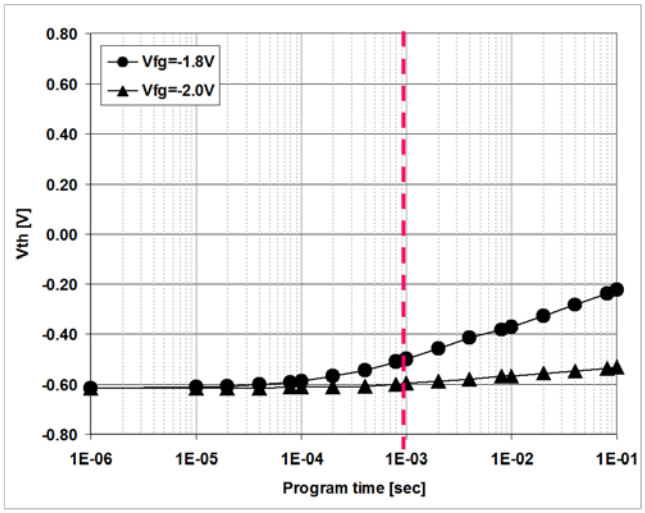


Figure 3.10. Simulated programming disturbance behavior of an unselected cell. The indicated disturbance time (1msec) is the time required to program a cell in the same sector.

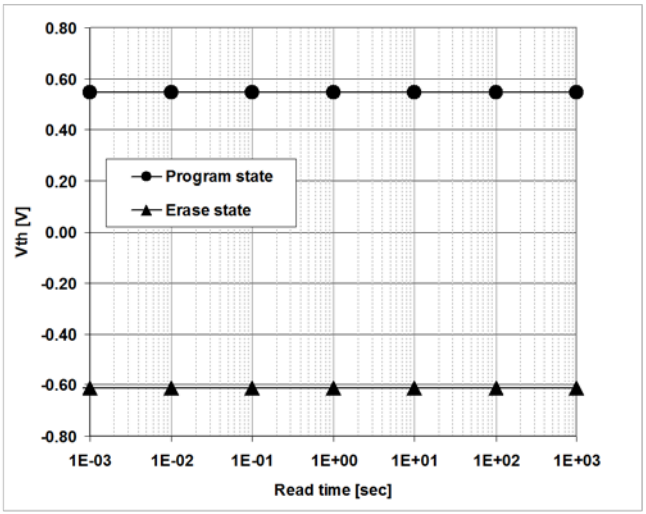


Figure 3.11. Simulated read disturbance behavior of programmed and erased cells. The front-gate is biased at 2V.

The backside charge storage design decouples the read function from the charging mechanism so that a high pass-gate voltage can be applied to the front gate without affecting the charge stored. **Figure 3.11** shows simulated read disturbance behavior of programmed and erased cells.

3.5 Summary

A backside charge storage non-volatile memory cell design is proposed for highly scalable $4F^2$ NAND flash memory. It can be fabricated on a silicon-on-ONO wafer using a conventional process flow sequence. A large read current (~10uA) adequate for high-speed read operation is projected for an 18nm cell (0.000324 μm^2 area). Since the BCS-NVM cell is essentially a back-gated thin-body (fully depleted) transistor, it should be more scalable than a conventional bulk flash memory cell. The thin-body design provides for good electrostatic integrity as compared to a bulk design, to facilitate scaling to well below 20nm gate length.

3.6 References

[1] H.A.R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connell, R. E. Oleksiak and H. Lawrence, "The variable threshold transistor, a new electrically-alterable, non-destructive read-only storage device". International Electron Device Meeting (IEDM) Technical Digest, vol. 13, pp. 70, 1967.

[2] P.C.Y. Chen, "Threshold-alterable Si-gate MOS devices", IEEE Transactions on Electron Devices, Vol. 24, pp. 584, 1977.

[3] Marvin White, "On the Go with SONOS", IEEE Circuits & Devices Magazine, Vol. 16, 2000.

[4] 2012 iSuppli annual report

[5] C.H. Lee, K.I. Choi, M.K. Cho, Y.H. Song, K.C. Park and K. Kim, "A novel SONOS structure of SiO2/SiN/Al2O3 with TaN metal gate for multi-giga bit flash memories", IEDM Technical Digest, pp. 26.5.1, 2003

[6] C.H. Lee, J. Choi, C. Kang, Y. Shin, J. Lee, J. Sel, J. Sim, S. Jeon, B. Choe, D. Bae, K. Park, and K. Kim, "Multi-Level NAND Flash Memory with 63 nm-Node TANOS (Si-Oxide-SiN-Al2O3-TaN) Cell Structure", Symposium on VLSI Technology Digest of Technical Papers, pp. 21, 2006.

[7] H. Silva and S. Tiwari, "A Nanoscale Memory and Transistor Using Backside Trapping", IEEE Transactions on Nanotechnology, Vol. 3, pp. 264-269, 2004.

[8] W. Kwon and T. King Liu, "Compact NAND Flash Memory Cell Design Utilizing Backside Charge Storage," in 2010 IEEE Silicon Nanoelectronics Workshop, 2010.

[9] C.H. Lee, J. Choi, Y. Park, C. Kang, B. Choi, H. Kim, H. Oh, W. Lee, "Highly scalable NAND flash memory with robust immunity to program disturbance using symmetric inversion-type source and drain structure", Symposium on VLSI Technology Digest of Technical Papers, pp. 118, 2008.

[10] J.-P. Colinge, Silicon-on-Insulator Technology. Norwell, MA: Kluwer, 1997, pp. 134–136.

Chapter 4

Electro-Mechanical Diode Non-Volatile Memory Cell

4.1 Introduction

As a non-volatile memory, Flash memory technology has evolved greatly over the past few decades. In the last ten years, lithography enabled a significant size reduction, going down from 100nm to 19nm half pitch and the memory capacity has been doubled every generation. The memory cell size reduction has been the key to increasing storage capacity while lowering cost per bit. To date, 128Gbit density NAND flash memory with 170 mm^2 chip size is available [1] and the price per Giga-Byte (GB) is lower than $1 [2]. The achievements in Flash memory technology have helped to enable a revolution in information technology applications such as smart phones and tablet PCs. However, fundamental scaling limitations for flash memory cell operating voltages and the physical thickness of the tunneling dielectric layer pose a significant challenge for continued scaling in the sub-20-nm regime [3].

A lot of alternative memory device designs and new materials have been proposed to overcome the scaling limits in conventional flash memory technology. Various cross-point memory architectures appear as one of the most attractive successors to the current Flash technology, because they not only allow for the most compact storage with $4F^2$ cell layout size (where F is the minimum half-pitch) but also can be fabricated using a relatively simple process that is more amenable to 3-dimensional integration. Programmable resistance devices such as phase-change memory [4] and resistive RAM [5] have been explored for cross-point memory application. However, there are intrinsic drawbacks of purely passive cross-point architecture, such as parasitic leakage paths in unselected cells and series resistance of the interconnections. These drawbacks hinder the realization of

high density cross-point arrays due to large sensing circuits required, lowering the memory-array efficiency and reducing the maximum allowable array size. To resolve these issues, the incorporation element of selector device, such as a diode or transistor, in series with the memory cell has been investigated. However, this approach demands additional complicated process steps and results in lower sensing current.

Recently, a nano-electro-mechanical (NEM) non-volatile memory cell design that eliminates the need for a selector device, by leveraging the hysteretic behavior of a mechanical gap-closing actuator, was proposed and demonstrated [6]. Although this design is well suited for a cross-point array architecture, it requires separate read and write word-lines as well as an initial "forming" step (charging of a dielectric layer) to achieve non-volatile operation. To overcome these disadvantages, a simpler electro-mechanical memory cell design is proposed and demonstrated [7] [8].

4.2 Device Structure

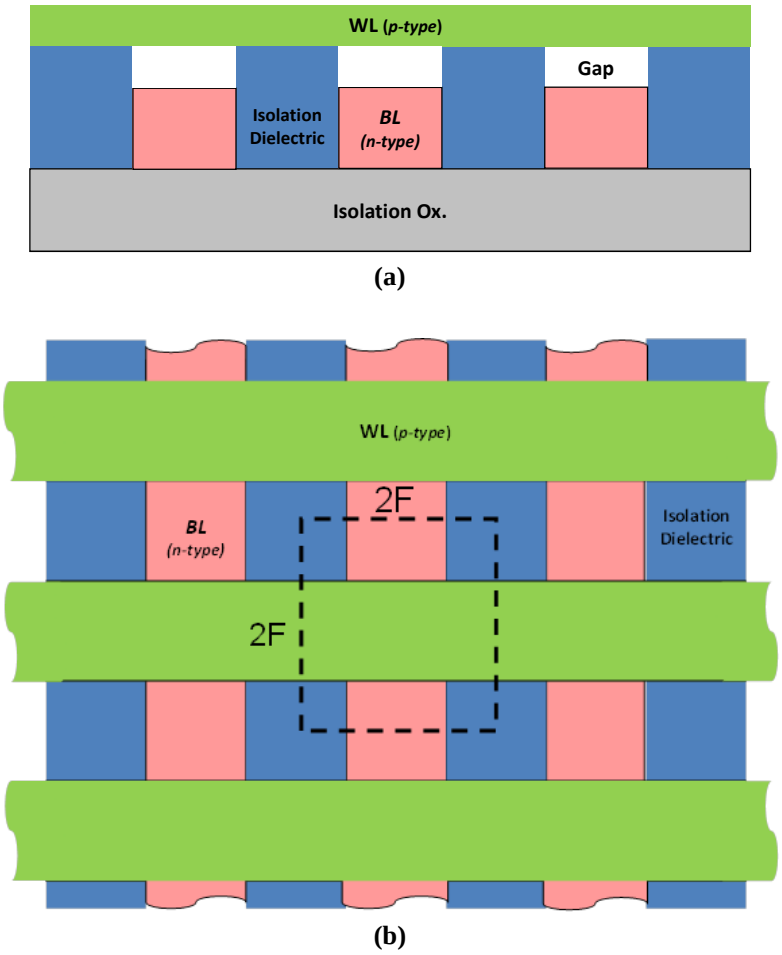


Figure 4.1. (a) Schematic cross-sectional view of the electro-mechanical diode cell (parallel to the word line). (b) Schematic plan view of the electro-mechanical diode memory array.

Figure 4.1 (a) and **(b)** illustrate the cell structure and the array layout of electro-mechanical diode memory. The bit lines (BLs) are formed on the insulating substrate and the word lines (WLs) are formed over the bit lines. The WLs and BLs are formed in different types (n-type or p-type) of semiconductor layer. Initially, the WLs are physically separated from the BLs with an air gap at each cross-point, so that no current can flow between these lines through the cell initially. This array configuration can achieve the most efficient cell layout area of $4F^2$.

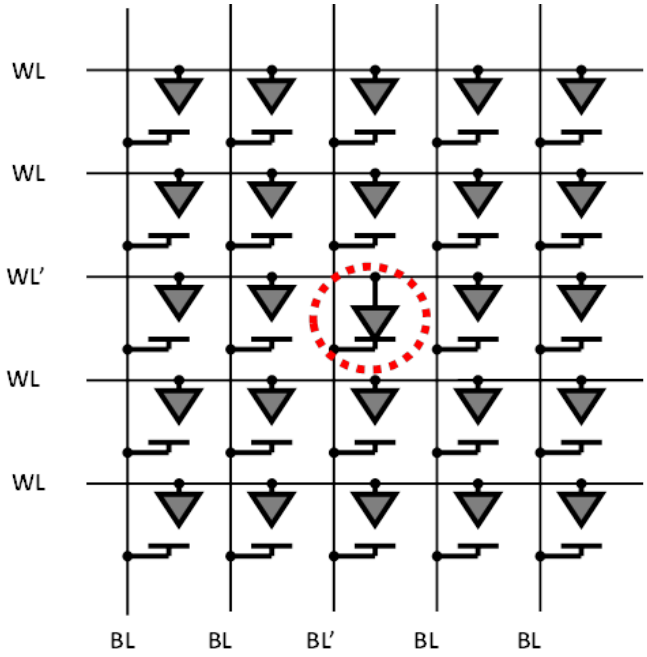
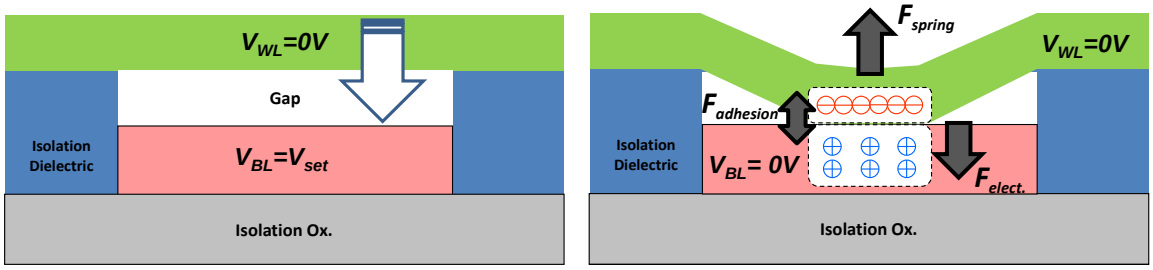


Figure 4.2. Circuit schematic of an electro-mechanical diode memory array.

	Set	Reset	Read
WL' ***(selected)***	**0 V**	**0 V**	**V_{read} >0V**
BL' ***(selected)***	**$V_{Set} > V_{pull\text{-}in}$**	**$V_{Reset} < V_{release}$**	**0 V**
WL ***(unselected)***	**(floating)**	**(floating)**	**0 V**
BL ***(unselected)***	**0 V**	**0 V**	**(floating)**

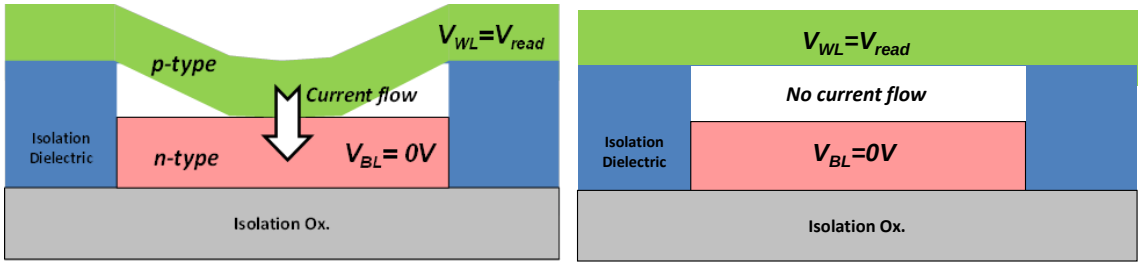
Table 4.1. Exemplary operating voltage conditions for Set, Reset and Read operations.

The circuit schematic of an electro-mechanical diode memory array is shown in **Figure 4.2**. When the selected cell (in the dashed circle) is in the Set state and all other cells are in the Reset state, the Set state cell forms a p-n diode and provides a rectifying current-vs.-voltage characteristic while the Reset state cells are in electrically-open-state. Exemplary operating voltage conditions for Set, Reset and Read operations are listed in **Table 4.1**.



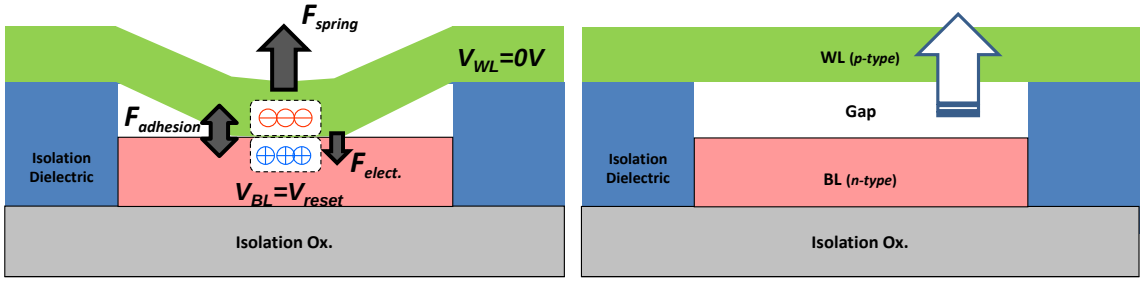
(a) **(b)**

Figure 4.3. (a) Illustration of the Set operation. When V_{set} is applied on the bit line (BL), the word line (WL) is pulled down to the BL by electrostatic force. Circuit schematic of a BCS-NVM cell array. (b) Illustration of the Hold operation. The Set state is sustained by a sum of electrostatic force ($F_{elect.}$) caused by space charges in the depletion region and surface adhesion force ($F_{adhesion}$) between WL and BL materials; $F_{spring} < F_{elect.} + F_{adhesion.}$



(a) **(b)**

Figure 4.4. (a) Illustration of the Read operation in the Set state. The Read current flows through the p-n junction when a forward bias (V_{read}) is applied on the word line. (b) Illustration of the Read operation in the Reset state. No current flows between the bit line (BL) and the word line (WL).



(a) **(b)**

Figure 4.5. (a) Illustration of the Reset operation. As a forward bias (V_{reset}) is applied on the p-n junction, the built-in electrostatic force decreases. When the sum of the electrostatic force and the adhesion force is smaller than the spring restoring force, the spring restoring force pull the WL beam out of the BL. (b) Illustration of the cell state after the Reset operation. The cell state (i.e. Reset state) is maintained due to a sufficiently stiff beam after a Reset operation.

Figure 4.3 (a) illustrates the Set operation; the pull-in voltage ($V_{pull\text{-}in}$) is the minimum voltage for actuating the WL into contact with the BL. Thus, to cause a memory cell to be programmed into the Set state, a voltage pulse (V_{set}) is applied between its WL and its BL to induce an attractive electrostatic force that is larger than the pull-in voltage ($V_{pull\text{-}in}$). In order for the Set state to be retained during a Hold operation (i.e. when no voltage is applied), the built-in electrostatic force ($F_{elec.}$ in **Figure 4.3 (b)**) together with the surface adhesion force ($F_{adhesion}$) in the Set state must be larger than the spring restoring force (F_{spring}) in the Set state.

The state of a cell is determined by sensing the BL current when a Read voltage is applied between its WL and its BL as shown in **Figure 4.4 (a)**. If the cell is in the Reset state, no current flows through it (**Figure 4.4 (b)**); only leakage current (through parasitic paths) flow. If the cell is in the Set state, a much larger forward diode current flows through it.

A Reset operation is shown in **Figure 4.5**. To cause a memory cell to be erased into the Reset state, a voltage pulse (V_{reset}) is applied between its WL and its BL to counteract the built-in electrostatic force of the p-n diode (**Figure 4.5 (a)**). When the sum of $F_{elec.}$ and $F_{adhesion}$ becomes smaller than F_{spring}, the spring restoring force of the WL beam pulls it out of contact with the BL (**Figure 4.5 (b)**). After a Reset operation, the cell state is maintained due to a sufficiently stiff beam.

4.3 Fabrication Process for Proto-type Cell

The process used to fabricate the first prototype electro-mechanical memory array is illustrated in **Figure 4.6**; to form an isolation layer, 100nm Al_2O_3 is coated on the silicon wafer by Atomic Layer Deposition (ALD) at 350°C. An in-situ phosphorus doped poly-Si layer (100 nm, 85Ω/□) is deposited by LPCVD 530°C onto the insulating substrate (Al_2O_3) followed by a sacrificial oxide (LPCVD oxide 30 nm) deposition at 450°C. After making photo resistor patterns of lines, the reactive ion etch (RIE) cuts into oxide and poly-Si layer to make bit lines. A silicon-nitride (SiN_x) layer is deposited over the bit lines, and SiN_x etch process forms silicon-nitride spacers along the BLs; the silicon-nitride spacers provide anchor of the WLs beam and electrical isolation between WLs and BLs also. A in-situ doped boron poly-$Si_{0.4}Ge_{0.6}$ deposition (100nm, 300Ω/□) is deposited as the WLs material. After making photo resist patterns of lines perpendicular to the BLs, the RIE cuts into the poly-$Si_{0.4}Ge_{0.6}$ layer and forms the WLs. Finally, the sacrificial oxide is selectively removed by vapor HF which leaves air-gaps between the BLs and the WLs.

Scanning electron microscopy (SEM) images are shown in **Figure 4.7 (a)** and **(b)**. A birds-eye view shows the 4 μm × 4 μm cells prior to HF vapor release. The cross-sectional view shows the sacrificial oxide layer thickness (12.7 nm) which determine the as-fabricated gap thickness between BL (n-type poly-Si) and WL (p-type poly-$Si_{0.4}Ge_{0.6}$).

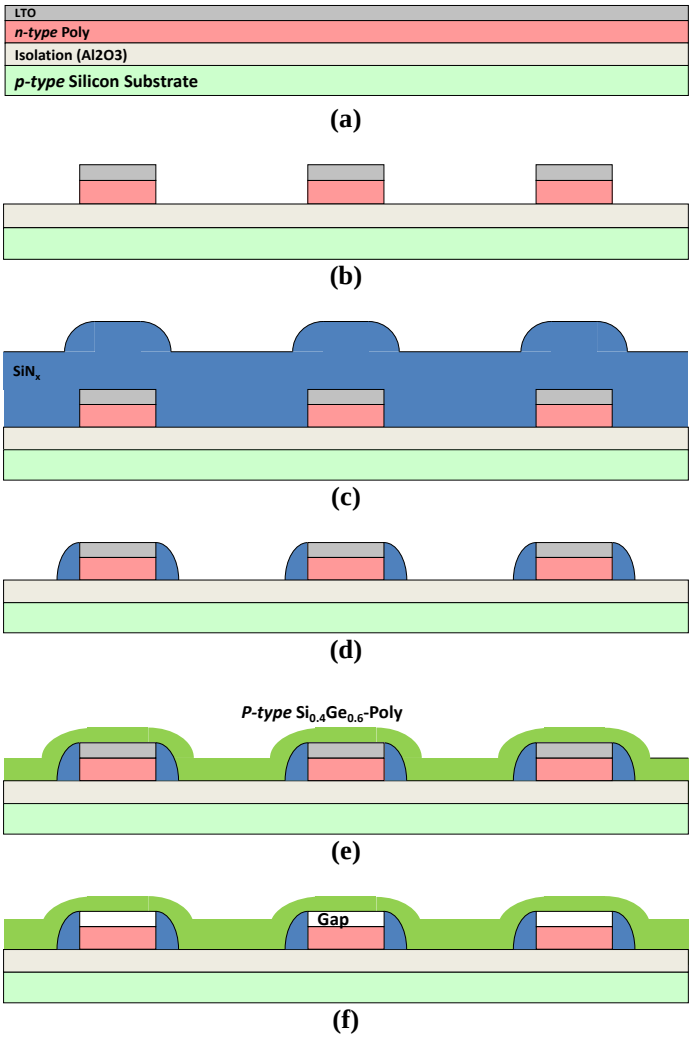
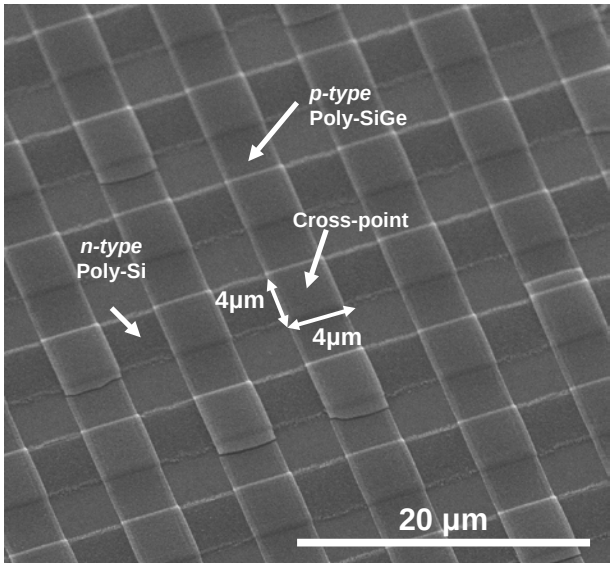
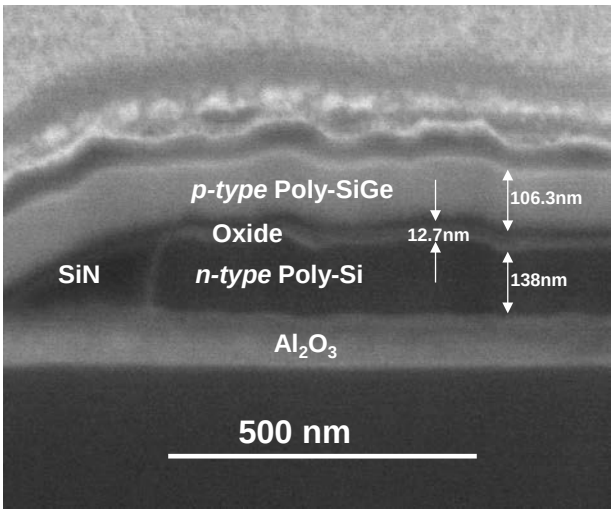


Figure 4.6. Key process steps for fabricating the proto-type electro-mechanical diode memory array; (a) n-type poly-Si and sacrificial oxide (LTO) layers deposition on the isolation oxide (Al_2O_3), (b) Bit lines formation by photo lithography and reactive ion etch (RIE), (c) Silicon nitride (PECVD SiN_x) deposition, (d) Spacer formation, (e) P-type poly-$Si_{0.4}Ge_{0.6}$ deposition and WL formation, (f) Sacrificial oxide removal by vapor HF.



(a)



(b)

Figure 4.7. (a) Bird's-eye SEM view of the fabricated electro-mechanical diode array. The working proto-type memory cell size is 4 μm x 4μm. (b) Cross-sectional SEM image of the fabricated electro-mechanical diode before sacrificial oxide removal by vapor HF. Since the oxide thickness is 12.7nm, the gap thickness after releasing poly-$Si_{0.4}Ge_{0.6}$ in vapor HF is ~ 13nm.

4.4 Electrical Measurement of the Proto-type Cell

Figure 4.8 shows the BL current changes of a proto-type cell during a Set operation. A sudden increase in current at the pull-in voltage ($V_{pull\text{-}in}$ = 6.2 V) is observed when the WL is pulled-in to contact the BL. Since it is an applied voltage (not current) that is required to actuate the WL, the Set current can be lowered by inserting a current-limiting resistance (1MΩ) in series with the BL driver (the red curves in Figure 4.8), to reduce the energy consumed by the Set operation. **Figure 4.9** shows the transient response of the WL voltage during a Set operation. The measured set time for the prototype cell is ~2μs for a Set voltage of 14 V.

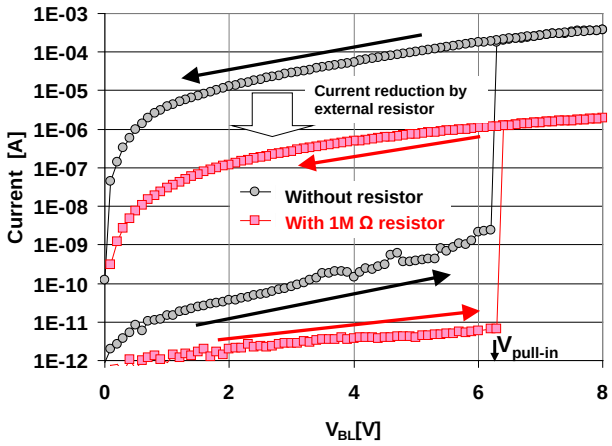


Figure 4.8. Measured Set operation hysteric I-V curves of a proto-type electro-mechanical diode cell as the bit line voltage is swept from 0V to 8V and 8V to 0V. After inserting a current-limiting resistance (1MΩ), the current can be lowed for reducing power consumption of Set operation.

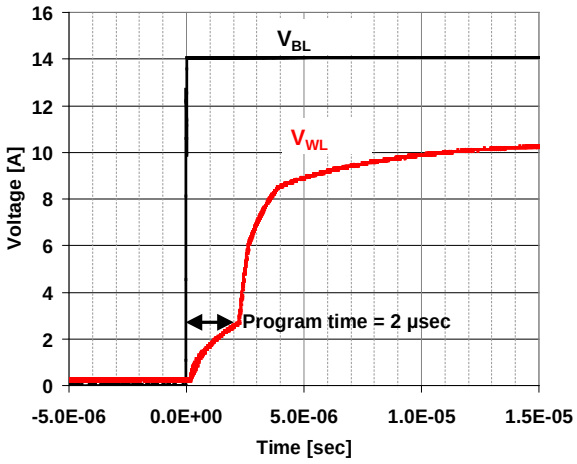


Figure 4.9. Measured the transient voltages of WL and BL during Set operation of a proto-type electro-mechanical diode cell. A step pulse of voltage is applied to the BL and the voltage on WL is monitored by oscilloscope.

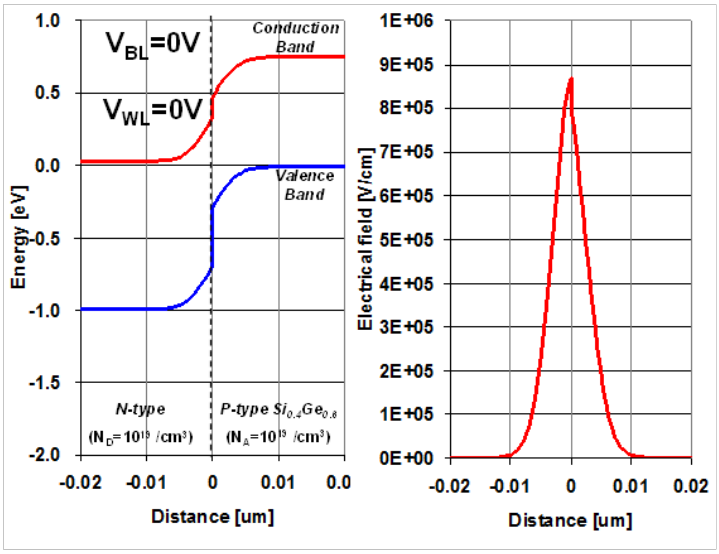


Figure 4.10. Simulated energy-band diagram and electric-field profile within a p-n junction in the Set state with V_{BL}=0V and V_{WL}=0V (Hold state). It is relevant to the proto-type cell's doping profiles.

To "Hold" the Set state without applied voltage, the attraction forces (i.e. the summation of electrostatic force ($F_{elec.}$) and adhesion force ($F_{adhesion}$)) between the WL and the BL has to be larger than the spring restoring force (F_{spring}) of the WL beam. The electrostatic force is caused by built-in space charges in the deletion region of the p-n diode. Finite element method simulation [9] is used for the proto-type cell to calculate $F_{elec.}$ **Figure 4.10** shows the simulated energy band diagram and electric field profile for a p-n junction of the proto-type cell and it gives a value of $F_{elec.}$ 70.5 μN. Since F_{spring} is only 4.22 μN from finite element method simulation [10], the Set state is retained when the external voltage (i.e. the actuation voltage) is removed.

The state of a cell is determined by sensing the BL current when the Read bias voltage is applied between its WL and its BL. If the cell is in the Reset state, no current flows through it; only leakage current flows through the SiN_x spacers in the cell. On the other hand, if the cell is in the Set state, much larger current flows through the forward-biased p-n junction. **Figure 4.11** shows measured I-V curves (linear scale and log scale of currents) in the Set and Reset state; very high Set/Reset current ratio (>10^6) is seen.

Figure 4.12 shows how the BL current of the proto-type cell changes during a Reset operation. An abrupt drop in current at the release voltage ($V_{release}$ = -6.2 V) is seen when the WL is released from the BL and goes back to the Reset state. It should be noted that the release voltage is larger than the built-in voltage (~ 1V in ideal case) of the p-n diode. The large release voltage is caused by the voltage drop on the WL and the BL. Since the thickness of WL beam is thin, the resistance of the WL is too large, so that it results in a large voltage drop on the WL rather than the p-n junction. To reduce these voltage drops, a metallic WL material or a shunting the metal layer can be used.

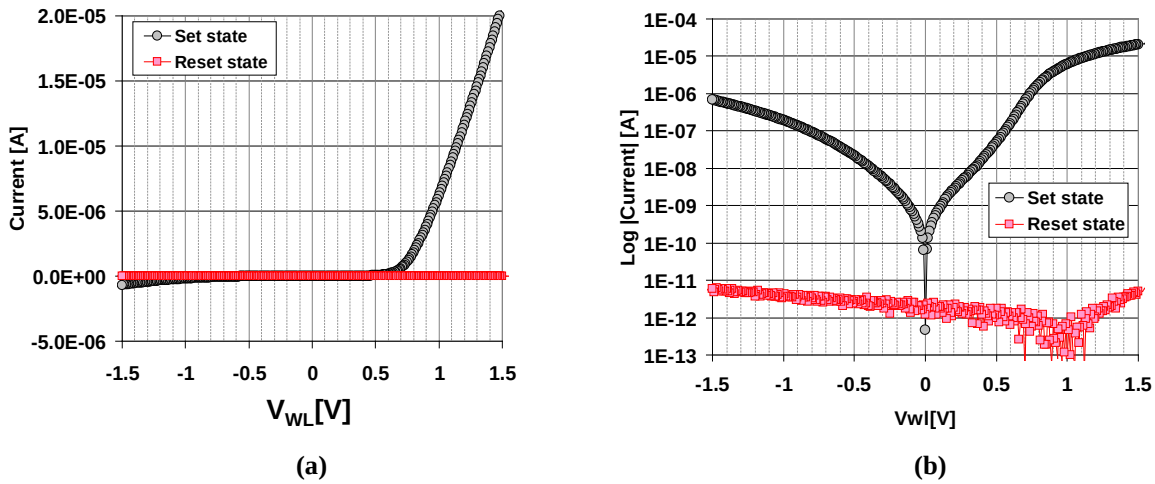


Figure 4.11. (a) Measured I-V curves for Read operation of a proto-type electro-mechanical diode cell. A cell in the Set state shows a typical p-n diode curve (black colored curve), whereas only leakage current is shown in the Reset state cell (red colored curve). (b) Measure log-scaled I-V curves for Read operation of a proto-type electro-mechanical cell.

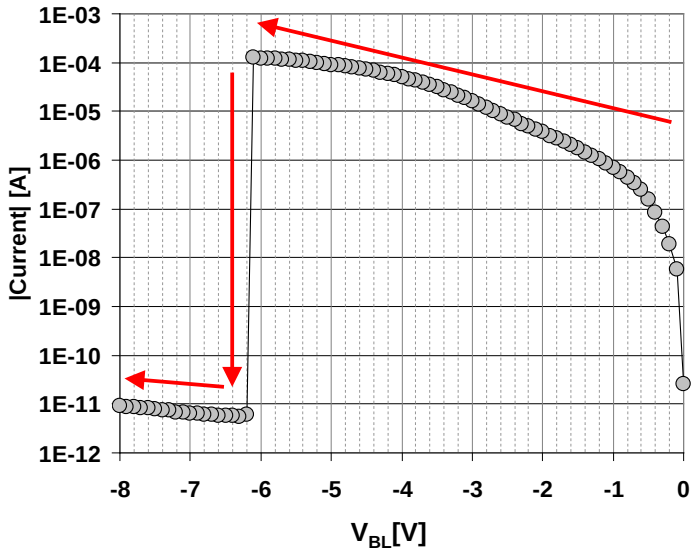


Figure 4.12. Measured I-V curve for Reset operation of a proto-type electro-mechanical diode cell. The bit line voltage is swept from 0V to -8V to increase forward-bias on the p-n junction.

From transient measurements, the reset time for the prototype cell is ~100 ms for a Reset voltage of -7 V (**Figure 4.13**). The longer reset time possibly can be explained by the contact opening model presented in Ref. [11]; the restoring spring force reduces the number of bonds between the contacting surfaces (**Figure 4.14**); only when the number of bonds is sufficiently small, contact opening occur. The contact opening time therefore depends on the number of bonds formed between the contacting surfaces, and can be reduced with scaling (to reduce the contact force and apparent contact area) and/or an appropriate surface treatment.

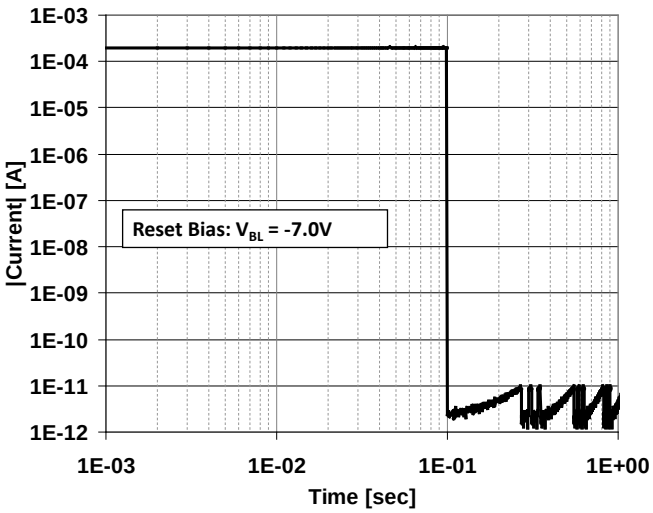


Figure 4.13. Measured Reset time of a proto-type electro-mechanical diode cell, where the bit line voltage is -7V and the word line voltage is 0V (a forward bias condition on the p-n junction).

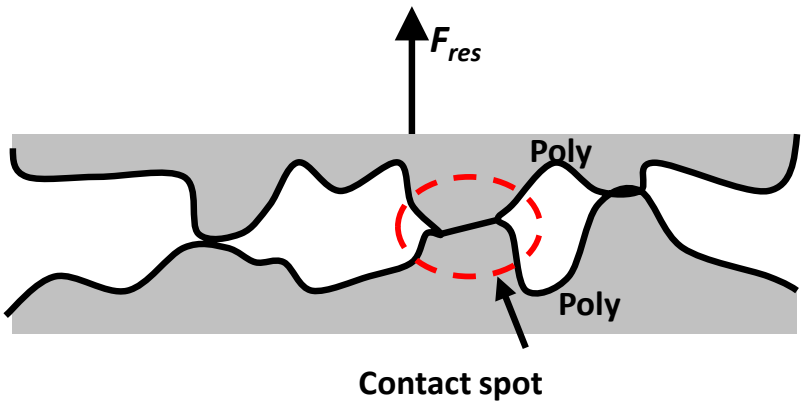


Figure 4.14. Illustration of the p-n junction (contacting surfaces) of the electro-mechanical diode.

4.5 Suppression of Sneak Leakage Currents in the Cell Array

A key requirement for implementation of a cross-point memory array is the suppression of the unwanted "sneak" leakage currents through unselected cells in the array. These leakages can sum to a sufficiently high current to result in an erroneous Read operation [12]. The size of the cross-point array is therefore limited by the ratio of the selected cell current to the summation of the leakage currents of unselected cells in the Set state during a Read operation [13]; a large ratio allows a larger array size. The sneak leakage current through an unselected cell in the Set state is almost equal to the diode current in a reverse-biased condition that is one-half the forward-bias read voltage ($V_{WL,READ}$), for an optimally designed sense amplifier for a resistive cross-point memory array. Thus, the ratio of the forward-bias Set cell current at $V_{WL,READ}$ to the reverse-bias Set cell current at $-V_{WL,READ}/2$, "the rectification ratio", is a key parameter to determine a cross-point array size.

The rectification ratio of the proto-type cell is ~ 300 for $V_{WL,READ}$ =1.2V. It can be enhanced by increasing the work-function difference between the WL and BL materials and by improving the crystalline quality of the semiconductor materials. **Figure 4.15** shows the demonstrated experiment results of p-type poly-Si as the WL material and adding a thermal anneal step. **Table 4.2** also shows the ideality factor and rectification ratio changes for the p-n diode curves of each p-type poly-Si process condition; since p-type poly-Si has a larger work function than p-type poly-$Si_{0.4}Ge_{0.6}$, the rectification ratio is improved nearly three times. Furthermore, the rectification ratio is improved greatly and the diode curve is getting close to the ideal diode curve after thermal annealing step to increase average grain size.

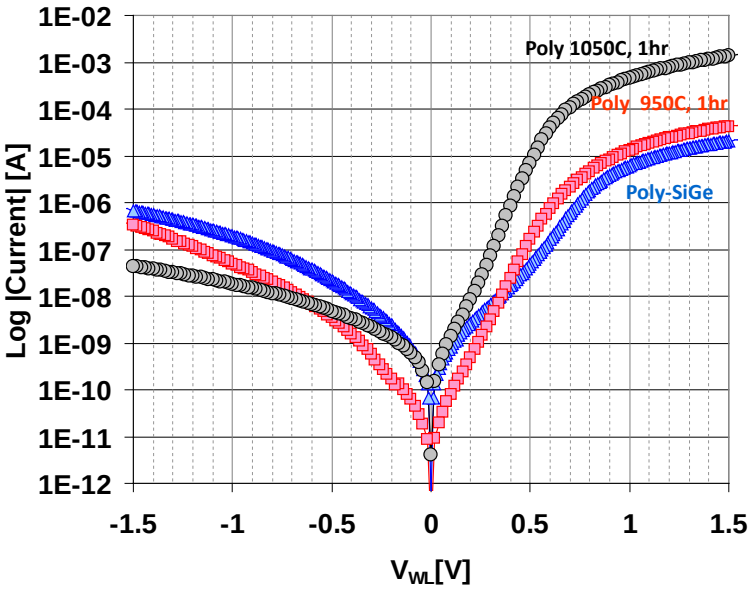


Figure 4.15. Measured diode I-V curves of electro-mechanical diode cell with poly-Si vs. poly-SiGe word line after thermal annealing.

P-type WL Material	Non-ideal Factor	Rectification Ratio
Poly-$Si_{0.4}Ge_{0.6}$ (410℃)	5~8	50
Poly-Si (610℃) with 950℃, 1hr anneal	3.0	10^3
Poly-Si (610℃) with 1050 ℃, 1hr anneal	1.7	10^4

Table 4.2. P-N Parameters Comparison of Electro-Mechanical Diode Cell.

4.6 Reliability

In order to investigate the basic reliability of the proto-type cell, data retention and endurance are measured. Since surface adhesion and electrostatic force hold a diode memory cell in the Set state, the proto-type cell shows very long retention time in **Figure 4.16**; negligible changes in conductance for the Set and Reset states are seen at 200°C, which well exceeds the normal operating temperature range for CMOS. As the electro-mechanical diode relies neither on charge storage in a floating-gate/charge-trap layer nor a material phase change to store data, it has more robust retention behavior than other NVM technologies. Moreover, it opens a possibility for new applications that demand very high (over 150°C) temperature operation.

Figure 4.17 shows a capability of multi-time programming (MTP), with endurance exceeding 10^4 Set/Reset cycles. The endurance of an electro-mechanical device can be very high if it is properly designed. For example, the endurance of a micro-electro-mechanical switch has been shown to exceed 10^9 on/off cycles [14].

A potential reliability concern for mechanical devices is vibration or mechanical shock. But because of the extremely small mass (3.73 picogram for the prototype devices in this work) of the WL beam within a cell, a very large acceleration (10^8 G) would be needed to accidentally Set a cell or to overcome the attraction forces in the Set state (4 μN for the prototype devices in this work) to Reset a cell. This is far beyond the acceleration requirement (~20000 G) for automotive environments [15].

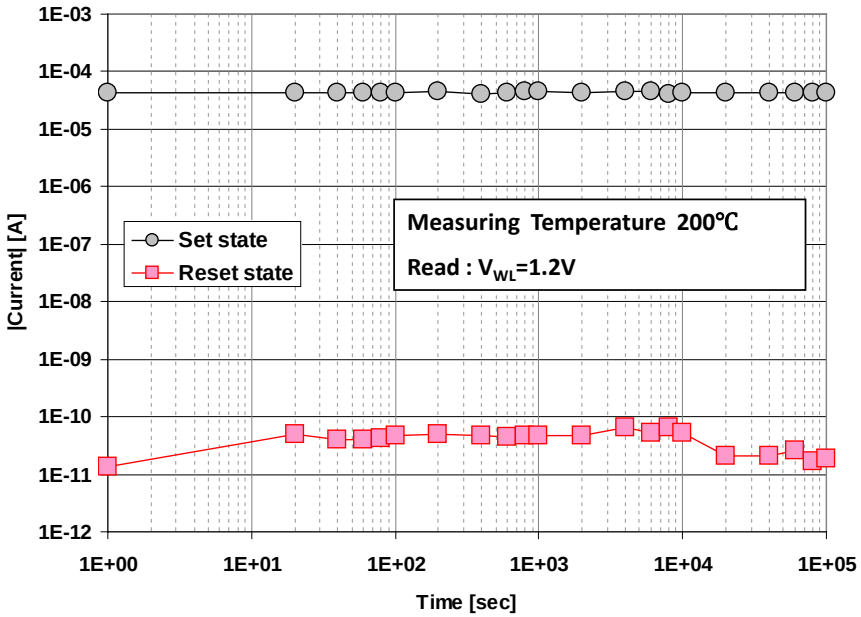


Figure 4.16. Measured retention behavior of a proto-type electro-mechanical diode cell in the Set state (black colored line) and the Reset state (red colored line) at high temperature condition (200 ◦C).

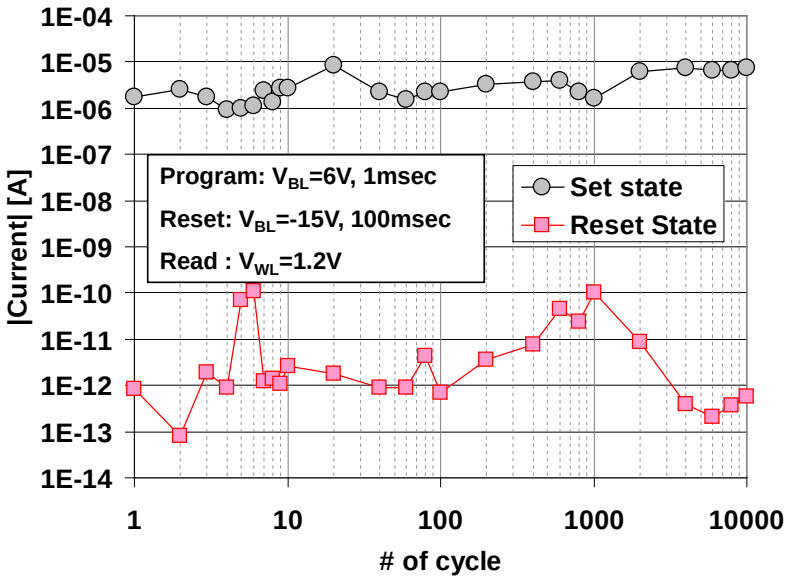


Figure 4.17. Measured endurance characteristics of a proto-type electro-mechanical diode cell in Set state vs. Reset state.

4.7 Modeling and Analysis of Electrical Behavior

In this section, modeling and analysis of the electrical behaviors of the electro-mechanical diode memory cell are discussed. **Figure 4.18** shows measured current *vs.* applied BL voltage (V_{BL}) characteristics of a prototype electro-mechanical diode cell. The lateral dimensions of the WL beam are 2 μm × 2 μm; the as-fabricated gap thickness between the WL beam (~100 nm-thick p-type poly-$Si_{0.4}Ge_{0.6}$) and the underlying BL (n-type poly-Si) is ~17 nm.

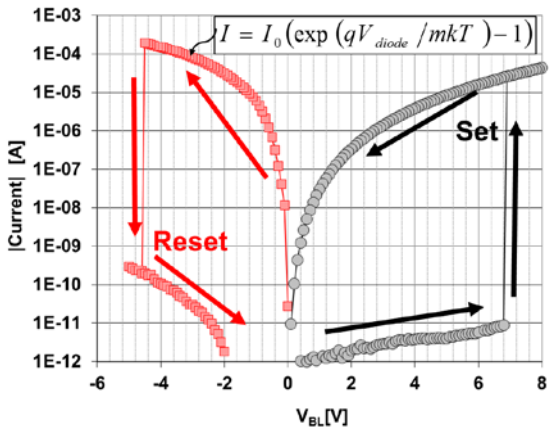


Figure 4.18. Measured current vs. BL voltage characteristics of a prototype electro-mechanical diode memory cell, with WL biased at 0 V: the black circles show Set operation (V_{BL} swept from 0V up to 8V and back down to 0V); the red squares show Reset operation (V_{BL} swept from 0V down to -5V and back up to 0V). The measured forward diode current can be fitted to the equation in the inset, where $V_{diode} = -V_{BL}$, $I_0 = 1.2\times10^{-8}$ and m=4.76.

4.7.1 Set Operation

The WL beam can be modeled as a stiff plate of length l_d and width w suspended by two fixed-guided cantilever beams of length l and width w [16], as shown in **Figure 4. 19**. The effective spring constant of the cantilever beams is

$$k_m = 2Ewt^3/l^3 \qquad (4.1)$$

where E is the Young's modulus and t is the thickness of the cantilever beams. The pull-in voltage ($V_{pull\text{-}in}$) represents the minimum value of V_{Set}:

$$V_{pull-in} = \frac{2g}{3}\sqrt{\frac{k_m}{1.5C_0}} \qquad (4.2)$$

where g is the actuation gap and $C_0 \propto l_d \cdot w$ is the parallel-plate capacitance of the stiff plate, as fabricated. **Figure 4.20** shows the calculated pull-in voltage as a function of g, for l = 1.0 μm and l_d= 2.0 μm (*i.e.* total length of the movable WL L = 4 μm). It can be seen that the simple analytical model given by Eq. (4.2) provides a reasonable estimation of V_{Set}.

The electrostatic force (F_{elec}) and spring-restoring force (F_{res}) in the Set state can be calculated using the following equations:

$$F_{elec} = \int_{depletion\ region} Q(x)\mathcal{E}(x)dx \qquad (4.3)$$

$$F_{res} = k_m \cdot g \qquad (4.4)$$

where $Q(x)$ is the charge density and $\mathcal{E}(x)$ is the electric field within the depletion region of the pn diode. Assuming that the dopant concentrations within the WL and BL are uniform and that the WL and BL are not fully depleted, Eq. (4.3) can be simplified to

$$F_{elec} \approx Q_d \cdot \mathcal{E}_{\max} = (qN_D W_d w l_d)\mathcal{E}_{\max} \qquad (4.5)$$

where N_D is the dopant concentration and W_d is the depth of the depletion region within the BL, and $\mathcal{E}_{max}$ is the peak electric field at the junction ($\mathcal{E}_{max} = - qN_D l_d/\varepsilon_s$). The values of F_{elec} and F_{res} are calculated to be 74.3 μN and 4.46 μN, respectively, for the prototype devices with $N_D = 10^{19}$ cm^{-3}, $W_d = 0.16$ μm, $l_d = 2.0$ μm, $w = 4$ μm and $\mathcal{E}_{max} = 1.77 \times 10^6$ V/cm. Since $F_{elec} > F_{res}$, the Set state is retained with zero applied voltage.

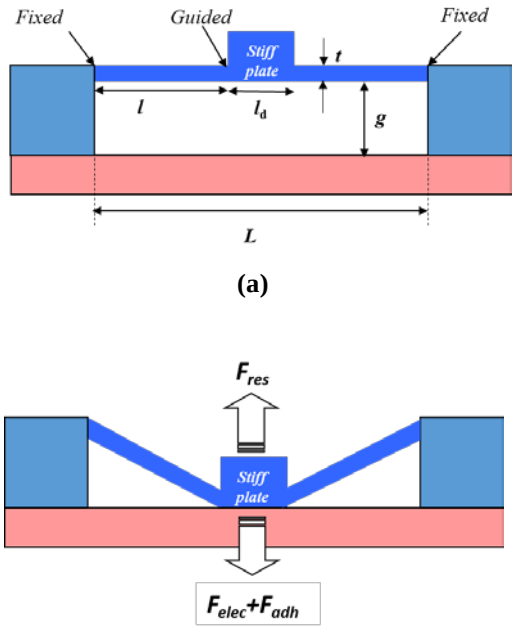


Figure 4.19. The WL beam of an electro-mechanical diode memory cell can be modeled as a stiff parallel-plate capacitor suspended by two fixed-guided cantilever beams.

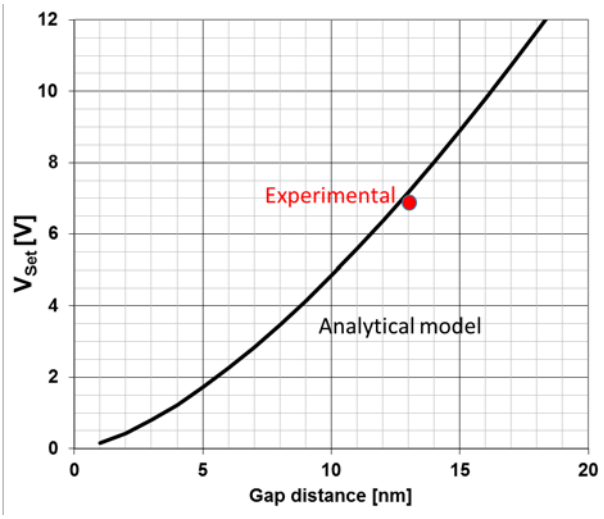


Figure 4.20. Comparison of calculated minimum Set voltage (Eq. 4.2) with experimental measurement.

4.7.2 Reset Operation

To erase the cell into the Reset state, the diode must be forward-biased to reduce F_{elec} so that the sum of F_{elec} and F_{adh} (*i.e.* the total attractive force) is smaller than F_{res}. As can be seen from **Figure 4.18** (red square symbols), as the magnitude of the applied voltage increases above 4.6 V, a sudden decrease in current is seen as the WL comes out of contact with the BL. In principle, the diode voltage (V_{diode}) should not exceed the built-in voltage, ~1.0 V. The applied BL voltage (V_{BL}) can be much larger than this, however, due to voltage drops along the WL and BL:

$$V_{BL} = V_{diode} + V_{ext} = V_{diode} + R_{ext} \cdot I_0 (e^{\frac{qV_{diode}}{mkT}} - 1) \qquad (4.6)$$

where V_{BL} is the applied BL voltage, V_{ext} is the extrinsic voltage dropped along the WL and BL, R_{ext} is the sum of the resistances of the WL and BL (from the memory array periphery to the cell), I_0 is the saturation current and m is the non-ideality factor of the pn diode. **Figure 4.21** shows the equivalent circuit used and **Figure 4.22** plots the calculated total attractive force (with F_{adh} fixed at 3.5 μN [8] and $I_0 = 1.2\times10^{-8}$, m = 4.76 for the pn diode) *vs*. the magnitude of V_{BL}, for various values of R_{ext}. The minimum magnitude of V_{BL} to reset the cell corresponds to the point where the attractive force falls below the spring restoring force (F_{res} = 4.5 μN). For R_{ext} = 100 kΩ (sheet resistance = 105 Ω/□ and number of squares = 100), this occurs at $|V_{BL}|$ = 4.7 V, which is close to the experimentally measured value of 4.6 V. The Reset voltage can be lowered (to be closer to V_{diode}) by reducing the WL and BL sheet resistances, *e.g.* by using a metallic WL material and/or using a thicker BL.

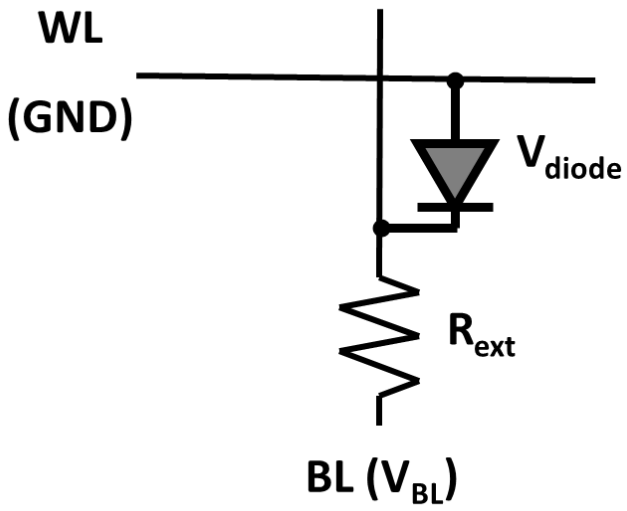


Figure 4.21. Equivalent circuit used to calculate the minimum Reset voltage.

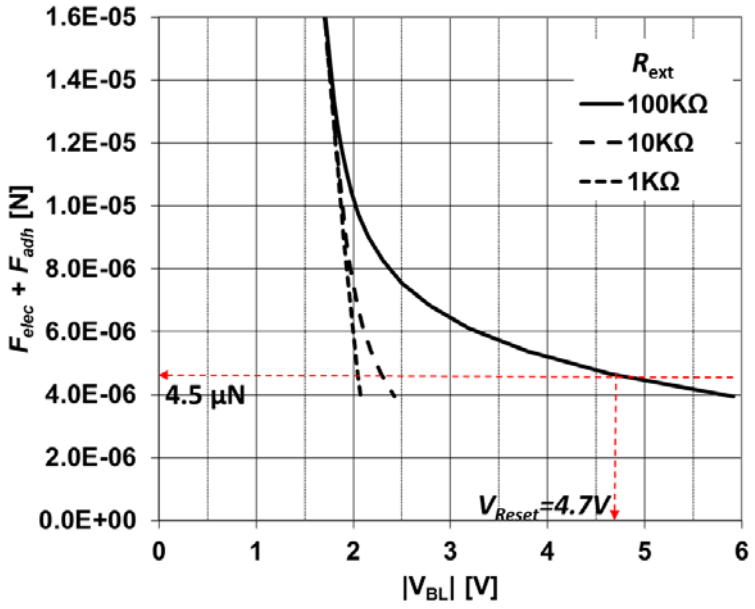


Figure 4.22. Calculated sum of electrostatic force and surface adhesive force as a function of the externally applied BL voltage, for various values of total WL and BL resistance.

4.8 Scaling Theory of Electro-Mechanical Diode Cell

Memory cell miniaturization is necessary to achieve high storage density. Therefore, a methodology for scaling an electro-mechanical diode non-volatile memory cell to nanometer-scale beam dimensions is developed in this section.

As the lateral dimensions of the cell are scaled down, the vertical dimensions (*i.e.* the beam thickness t and actuation gap thickness g) also must be reduced to maintain a low Set voltage, as can be seen from Eqs. (4.1) and (4.2). Since the surface adhesive force F_{adh} is expected to scale with the contact area [17] [18], it is important to maintain the ratio F_{elec}/F_{res} to be greater than 1 in the Set state, to guarantee non-volatile operation. This ratio can be expressed as:

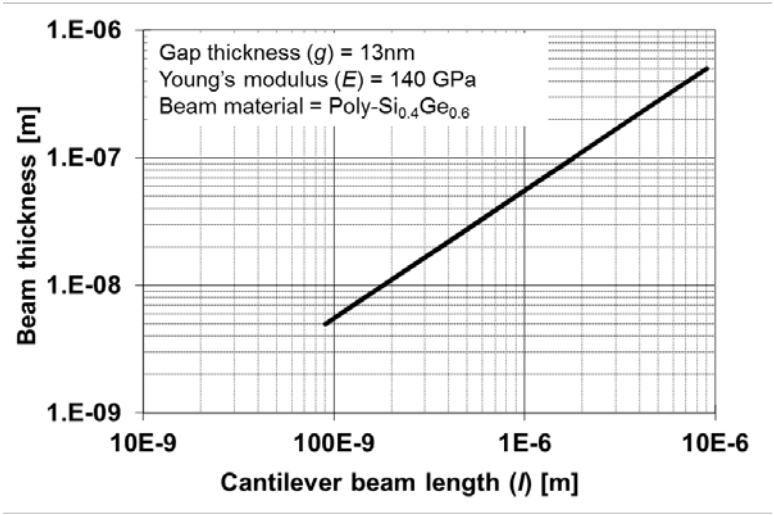
$$\frac{F_{elec}}{F_{res}} \propto \left(\frac{l}{t}\right)^3 \cdot \frac{l_d}{gE} \cdot \mathcal{E}_{\max} \qquad (4.7)$$

Note that it is nominally independent of the beam width (w).

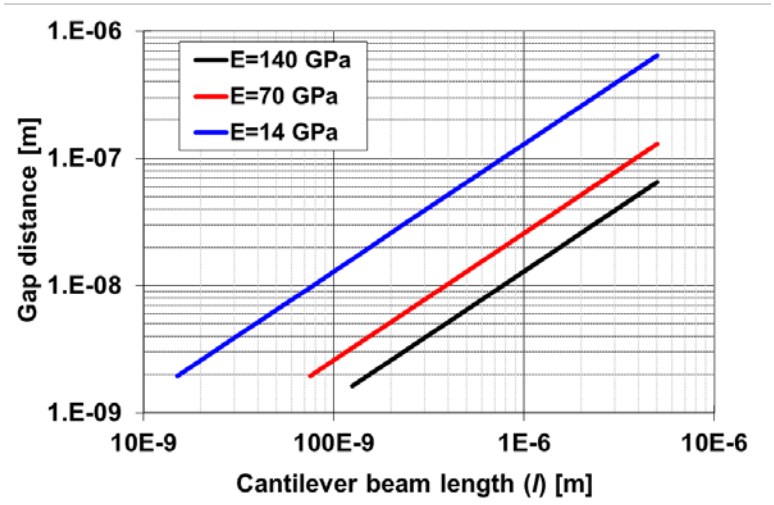
From (4.7) it can be seen that, as the cantilever beam length (l) and diode length (l_d) are scaled down together, the thicknesses of the cantilever beams (t) and actuation gap (g) should be scaled down proportionately in order to maintain this ratio. **Figure 4.23(a)** shows how the beam thickness should scale with beam length. In consideration of manufacturability and the depletion depth (~4 nm), a reasonable lower limit for the beam thickness is ~5 nm. Due to practical limits for thickness scaling, then, a reduction in Young's modulus (E) eventually will be necessary to enable beam-length scaling to well below 100 nm.

Figure 4.23(b) shows how the gap thickness should be scaled down proportionately with beam-length scaling to ensure non-volatile operation. Quantum-mechanical tunneling (resulting in undesirable current flow through a cell in the Reset state) sets a lower limit for the gap thickness, to ~2 nm.

If the beam thickness and/or gap thickness are no longer scaled down together with the length, the peak strain within the beam will increase and hence the material strain limit will become another practical consideration. For a 10 nm-thick polycrystalline-silicon beam (with a strain limit of 0.93%), the minimum beam length is ~80 nm for an actuation gap of 5 nm [19]. Alternative structural materials with higher yield strain will be needed to scale the beams to shorter length. Table I shows that TiNi alloy is a good candidate. Note that a Schottky diode, rather than a pn diode would be formed if the WL beam is metallic and the BL comprises a (heavily doped) semiconductor; such a diode would still have a built-in electric field and exhibit rectifying behavior as desired for cross-point non-volatile memory application.



(a)



(b)

Figure 4.23. (a) Cantilever beam thickness and (b) actuation gap thickness required to maintain F_{elec}/F_{res} ratio with beam length scaling.

The utility of the simple analytical model (Eq. (4.2)) is demonstrated in **Figure 4.24**, which plots V_{Set} as a function of the actuation gap thickness, for various cantilever beam dimensions and Young's modulus values in **Table 4.3**. It can be seen that the analytical model reasonably matches experimental observations and compares well against finite-element-method (FEM) simulation results.

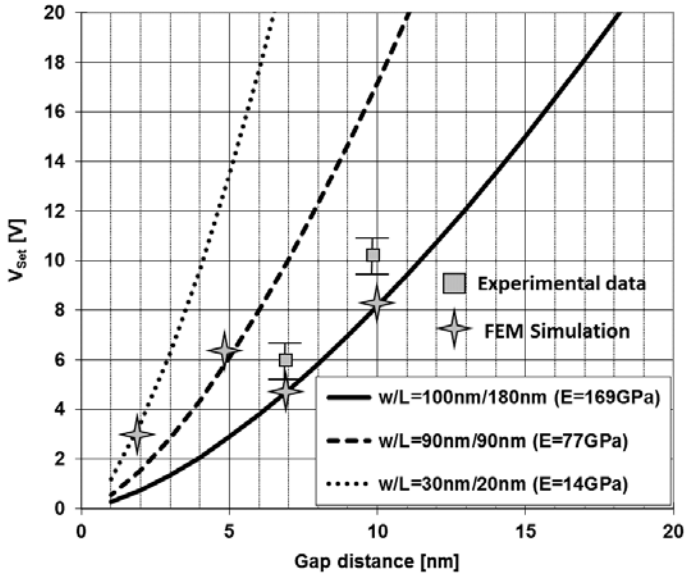


Figure 4.24. Calculated V_{Set} as a function of as-fabricated actuation gap thickness for scaled memory cells, for various values of Young's modulus *E*. Selected measured data and values obtained from finite-element-method simulations are also shown, to verify the accuracy of the analytical model.

Beam Material:	Poly-Si	Aluminum	TiNi
Young's Modulus (*E*)	169 GPa	77 GPa	14 GPa
Thickness (*t*)	58 nm	5 nm	5 nm
Contact length (l_d)	18 nm	9 nm	2 nm
Strain Limit	0.93 %	0.18 %	~ 8%
Built-in Voltage	1.09 V	0.92 V	1.22 V

Table 4.3.

4.9 Exemplary Design for 20nm Electro-Mechanical Diode Cell

In this section, the calibrated analytical model is used to project the performance characteristics of an electro-mechanical diode memory cell with 20 nm minimum layout dimension. TiNi is employed as the beam material and has minimum thickness t = 5 nm; the cantilever beams have minimum width w = 20 nm and total beam length l = 15 nm (i.e. total beam length L = 40nm). **Table 4.4** summarizes the performance parameters.

Beam Material (Young's modulus)	TiNi (14 GPa)
Beam Dimensions ($w \times l \times t$)	20 nm × 15 nm × 5 nm
Actuation Gap Thickness (g)	2 nm
Set Voltage (V_{Set})	1.90 V
Set Time (t_{Set}) (FEM Simulation)	0.3 ns
Reset Voltage (V_{Reset})	1.98 V

Table 4.4.

Three-dimensional (3-D) FEM simulations [10] were conducted to verify the accuracy of the analytical model and also to find the Set time (t_{Set}) for the 20-nm cell. The beam displacement profiles within a 4×4 20-nm cell array are shown in **Figure 4.25**. The maximum strain within the TiNi beam is ~7% in the Set state, which is below the strain limit. The simulated V_{Set} and t_{set} values are 1.9 V and 0.3 ns, respectively, confirming the accuracy of the analytical model. These findings show that a scaled electro-mechanical diode cell can be operated with lower voltage and faster speed than a comparably sized flash memory cell.

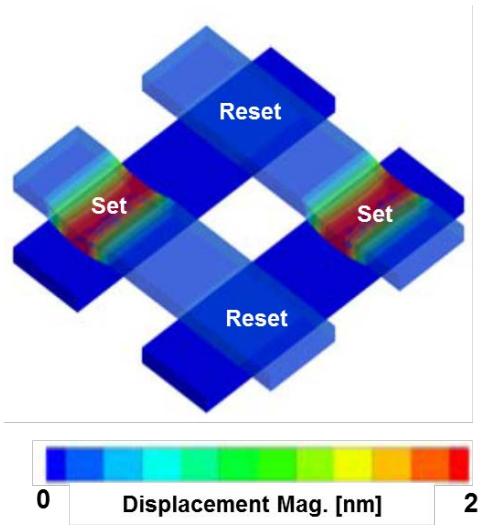


Figure 4.25. Simulated beam displacement profile of 20nm electro-mechanical diode cell design. Two cells are in the Set state within a 4x4 cell array.

4.11 3-Dimensional Integration

As the feature size of flash memory cells reaches the minimum limit, further flash density increases will be driven by possibly 3-dimensional (3-D) stacking of memory cells. Thus, the 3-D stacked NAND Flash memory devices [1]–[6] have caught great attention as a solution. The key of 3-D stacked NAND technology is a multi-stacked array with a low number of critical lithography process steps, without area penalty, to achieve continuous reduction of bit cost [20].

The simplicity of the mechanical diode structure makes it conducive for implementation in a 3-dimensional (3-D) array similar to that described in [21], as illustrated in **Figure 4.26**. The bit lines run horizontally and are formed by patterning a multi-layered stack of poly-Si/nitride films. Sacrificial oxide can be selectively formed by thermal oxidation of the poly-Si, prior to deposition and patterning of the (metallic) word-line layer. In this manner, only two lithography steps are used to form the 3-D array. Note that WL beam actuation occurs in the lateral direction, and that each cell has built-in redundancy (two mechanical diodes per cell). Also, the bit lines can be relatively thick, to maintain adequately low series resistance. This array architecture allows for cost-effective compact implementation for high storage density memory.

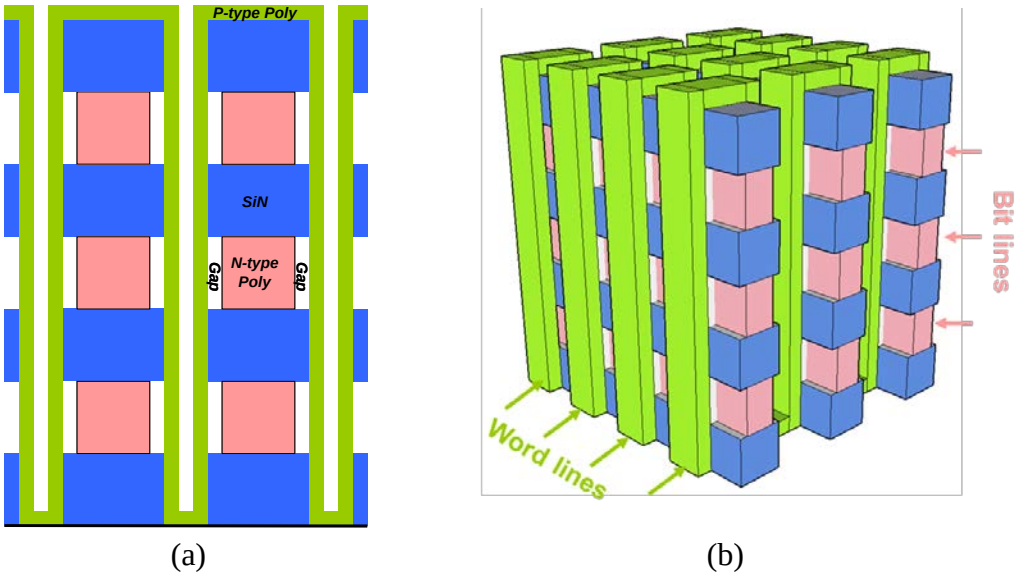


Figure 4.26. Proposed 3-D array architecture for electro-mechanical diode memory. (a) Cross-sectional view along a word line. (b) Bird's-eye view.

4.12 Summary

A new electro-mechanical diode memory cell design is proposed. The first prototype NVM diode cells are demonstrated to operate with relatively low set/reset voltages and excellent retention characteristics, and have endurance exceeding 10^4 cycles.

The electro-mechanical diode cell design can be scaled to 20 nm minimum lateral dimension by following an appropriate scaling methodology in consideration of various practical and fundamental limits. Low-voltage (< 2 V) and high-speed (sub-nanosecond) operation are projected using a calibrated analytical model as well as 3-D FEM simulation. These findings indicate that electro-mechanical diode technology is promising for high density storage beyond the limits of conventional flash memory technology.

4.13 References

[1] Y. Li et all, "128Gb 3b/Cell NAND Flash Memory in 19nm Technology with 18MB/s Write Rate and 400Mb/s Toggle Mode", in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2012.

[2] 64 Giga bit NAND flash price on www.dramexchange.com (1/28/2013)

[3] Y. Koh, "NAND Flash scaling beyond 20nm", in *International Memory Workshop IMW)*, 2009.

[4] H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase change memory", in *Proc. IEEE*, Vol. 98, No. 12, pp. 2201-2227, 2010.

[5] H. Akinaga and H. Shima, "Resistive Random Access Memory (ReRAM) based on metal oxides", in *Proc. IEEE*, Vol. 98, No. 12, pp. 2237-2251, 2010.

[6] W. Y. Choi, H. Kam, D. Lee, J. Lai and T.-J. K. Liu, "Compact Nano-Electro-Mechanical non-volatile memory (NEMory) for 3D integration" in *IEDM Tech. Dig.*, pp. 603-606, 2007.

[7] W. Kwon , J. Jeon, L. Hutin and T. K. Liu, "Electro-Mechanical Diode Memory Cell Design for Cross-Point Memory Arrays", *IEEE Electron Device Letters*, Vol. 33, No. 2, 2012.

[8] W. Kwon, L. Hutin and T.-J. K. Liu, "Electro-Mechanical Diode Performance and Scaling for Cross-Point Non-Volatile Memory Array", in *International Memory Workshop (IMW)*, 2012.

[9] Sentaurus Ver. A-2008.09, SYNOPSYS

[10] CoventorWareTM Ver. 2008

[11] B. D. Jensen, K. Huang, L. L.-W. Chow, and K. Kurabayashi, “Adhesion effects on contact opening dynamics in micromachined switches”, in Journal of Applied Physics, Vol. 97, p. 103535, 2005.
[12] J. Liang and H.-S. P. Wong, “Cross-Point Memory Array without Cell Selectors—Device Characteristics and Data Storage Pattern Dependencies”, IEEE Transaction on Electron Devices, Vol. 57, pp. 2531, 2010.
[13] A. Flocke and T. G. Noll, “Fundamental Analysis of Resistive Nano-Crossbars for the Use in Hybrid Nano/CMOS-Memory”, in International Solid-State Circuits Conference Digest, pp. 328, 2007.
[14] H. Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, and T.-J. K. Liu, “Design and Reliability of a Micro-Relay Technology for Zero-Standby-Power Digital Logic Applications”, IEEE IEDM Technical Digest, 2009.
[15] Automotive Electronic Coulcil, AEC-Q100-Rev-G, May 14, 2007.
[16] C. Liu. “Chapter 4” in *Foundations of MEMS*, 1st ed. Pearson Prentice Hall, 2006.
[17] H. Kam, T.-J. K. Liu, V. Stojanovic, D. Markovic and E. Alon, “Design, Optimization, and Scaling of MEM Relays for Ultra-Low-Power Digital Logic”, in *IEEE Trans. Electron Devices*, Vol. 58, pp. 236, 2011.
[18] J. Yang, L. Hutin and T.-J. K. Liu, “Adhesive Force Characterization for MEM Logic Relays with Sub-micron Contacting Regions,” to up here in *IEEE/ASME J. MEMS*.
[19] D. Lee, H. Tran, B. Ho and T.-J. K. Liu, “Electrical Characterization of Etch Rate for Micro- and Nano-Scale Gap Formation,” in *IEEE/ASME J. MEMS*, vol. 19, pp. 1260-1263, 2010.
[20] H. Tanaka et all, “Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory,” in *Symposia on VLSI Technology*, 2007.
[21] H. T. Lue et al., “A highly scalable 8-layer 3D vertical-gate (VG) TFT NAND Flash using junction-free buried channel BE-SONOS device,” in *VLSI Symposia*, pp. 131-132, 2010.

Chapter 5

Conclusion

5.1 Summary

Humans have invented and used devices to aid computation for thousands of years. The earliest counting devices were fingers and a carving marks on a cave wall. Fingers and marks were actually a kind of memory media to provide a core function for calculating. The sophistication of computational devices has evolved with the development of our civilization. As the amounts of information are getting larger and larger, the computational devices are demanding larger capacity of memory: while an average data size of a computational device in the 1930s (such as IBM punched card Accounting Machines at the U.S. Social Security Administration) was only a few bytes, the average memory capacity of a home computer today is over 2 giga-bytes. It is likely that our society keep demanding larger capacity memory. However, conventional memory technologies are facing many challenges such as difficulties of miniaturization and guaranteed good reliability of memory devices. For this reason, this dissertation investigates alternate memory device designs to overcome the limitation of conventional memory device technologies.

In chapter 2, a doubly gated vertical channel transistor (DGVC) is proposed for highly scalable DRAM. This new structure eliminates the capacitor from a conventional 1-transistor/1-capacitor DRAM memory cell. This technology rpovides a more scalable $4F^2$ cell size design, which can be fabricated with a conventional process flow on silicon- bulk instead of expensive SOI wafer. Cell operations (Read, Write, and Hold) and disturbance characteristics are investigated by 3-dimensional device simulations for 22-nm technology. Disturbance immunity characteristics of a DGVC cell are shown to be adequate. Since the vertical channel design allows for longer channel length, and hence a greater body

thickness is allowed, it is possible to scale DRAM memory with a $4F^2$ cell area down to 22-nm minimum half-pitch and beyond.

In chapter 3, a backside charge storage non-volatile memory cell design is proposed for highly scalable $4F^2$ NAND flash memory. A NAND Flash array of backside charge storage non-volatile memory (BCS-NVM) cells can be fabricated on a silicon-on-ONO wafer using a conventional process flow sequence. A large read current (~10uA) adequate for high-speed read operation is projected for an 18nm cell (0.000324 μm^2 area). Since the BCS-NVM cell is essentially a back-gated thin-body (fully depleted) transistor, it should be more scalable than a conventional bulk flash memory cell. The thin-body design provides for good electrostatic integrity as compared to a bulk design, to facilitate scaling to well below 20nm gate length.

In chapter 4, a nano-electro-mechanical (NEM) non-volatile memory cell design is proposed. The new non-volatile memory cell design leverages the hysteretic behavior of a mechanical gap-closing actuator and eliminates the need for a separate selector device. The first prototype NVM diode cells are demonstrated to operate with relatively low set/reset voltages and excellent retention characteristics, and have endurance exceeding 10^4 cycles. The electro-mechanical diode cell design can be scaled to 20 nm minimum lateral dimension by following an appropriate scaling methodology in consideration of various practical and fundamental limits. Low-voltage (< 2 V) and high-speed (sub-nanosecond) operation are projected using a calibrated analytical model as well as 3-D FEM simulation. These findings indicate that electro-mechanical diode technology is promising for high density storage beyond the limits of conventional Flash memory technology.

In summary, a new cell and array architecture that utilizes new approaches beyond a conventional memory is needed to overcome many obstacles that we are facing now. The new approaches include not only a new material (i.e. SOONOS wafer) but also new type of device (such as a mechanical switch device). Implementation of such devices in a the compact $4F^2$ array will be a serious challenge since it is necessary to meet demands of high capacity and low cost memory devices. Consequently, a widening horizon in memory device study is necessary.

5.2 Contribution of This Work

In this dissertation, possible alternatives for scaling conventional memory have been proposed and demonstrated. First, a doubly gated vertical channel (DGVC) structure is proposed for a new capacitor-less DRAM memory cell. Since the DGVC structure is unique and advantageous for stand-alone DRAM cell, this technology provides a promising cell structure for a mass-production DRAM. Second, as a new type of NAND Flash technology, a backside charge storage non-volatile memory (BCS-NVM) cell is proposed

and studied. Third, an electro-mechanical diode non-volatile memory cell design is firstly proposed for implementation of compact cross-point memory arrays. The first prototype cells are fabricated and demonstrated to operate with relatively low set/reset voltages and excellent retention characteristics. A scaling methodology for this new NVM technology is also developed with the aid of a calibrated analytical model.

5.3 Suggested Future Work

RRAM, resistive random access memory, is of interest for high-density non-volatile embedded memory applications because it has most compact architecture ($4F^2$ cell size) and is stackable on logic circuits. A challenge for implementation of large-capacity cross-point memory array is the existence of sneak leakage current paths which result in an erroneous Read operation. In order to mitigate this problem, a selector device must be incorporated at each cross-point; however, previous solutions incorporating with MOSFET switch or PN diode can increase device area and power dissipation. More importantly, it reduces the current sensing margin between memory cell states. In an initial work, the use of electrostatic gap closing actuators as RRAM selector devices is investigated.

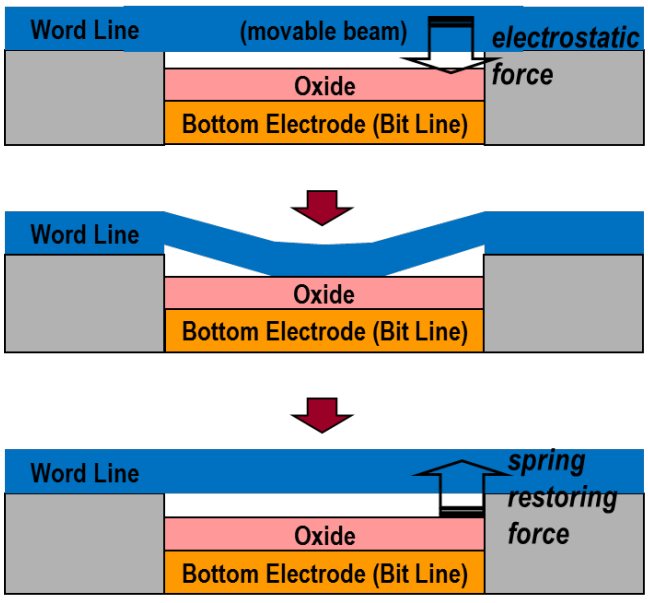


Figure 5.1. Schematic cross-sectional view and exemplary operation of RRAM cell with NEM selector.

Figure 5.1 illustrates a NEM selector incorporated with an oxide material layer of RRAM. Initially, the word line (WL) is physically separated from the bit line (BL) with air gap at each cross-point, so that no current can flow between these lines through the cell. When the pull-in voltage is applied, the electric force actuates the WL into the BL; if the applied voltage is large enough, the electric field in the resistive oxide can change the resistive state of the oxide. On the other hand, the state of RRAM can be determined by sensing the BL current when the applied voltage is not enough to change RRAM state, but enough for gap closing.

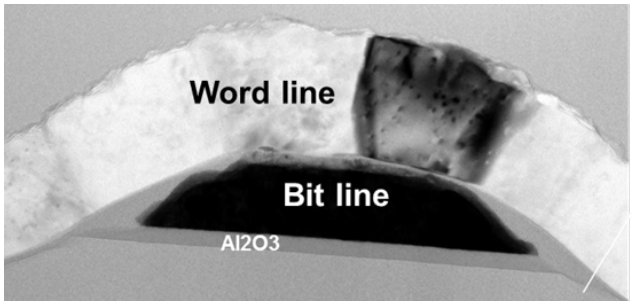


Figure 5.2. Cross-sectional transmission electron microscopy . image of the fabricated proto-type device TEM image of the proto-type device.

A proto-type of NEM selector is fabricated. **Figure 5.2** shows a cross-sectional transmission electron microscopy image of the fabricated proto-type device. The WL is aluminum (Al) with 100nm thickness and the BL is tungsten (W) with 50 nm thickness. The resistive oxide material between the WL and the BL is hafnium oxide (HfO_x) with 5nm thickness. The sacrificial oxide on top of the HfO_x is selectively removed by vapor HF which leaves air-gap.

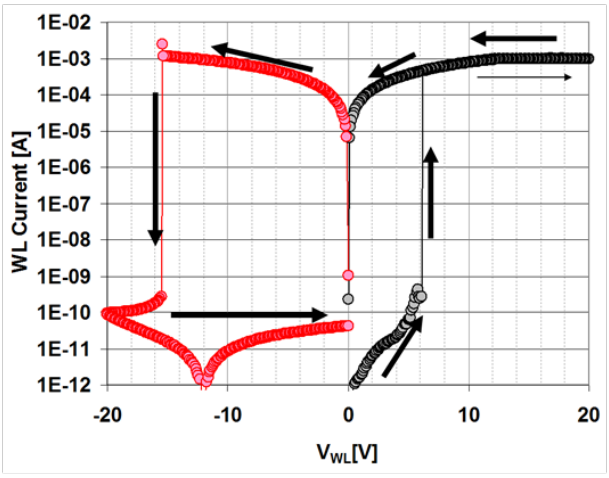


Figure 5.3. Measured hysteric I-V curves of a proto-type RRAM cell with NEM selector.

Figure 5.3 shows measured hysteric I-V curves of a proto-type RRAM cell with NEM selector. The word line voltage is swept from 0V to 20V and 20V to 0V firstly (black symbol); the cell is in a high resistive state (HRS) initially (i.e. small current) and a sudden increase in current is observed when the WL is pulled-in to contact the BL. Since the pull-in voltage is large enough, the electrical field in the resistive oxide causes a transition into a low resistive state (LRS) in the RRAM; the electrical field aligns electron hopping sites so that it makes current conducting filaments in the resistive oxide. To change the LRS into the HRS, the WL voltage is swept from 0V to -20V and -20V to 0V secondly (red symbol); since the applied negative voltage provide opposite direction electrical field, the alignments of electron hopping sites is destroyed and eventually the conducting filaments are removed. It is worth noting that the on/off ratio is very large (over 10^7) since previous RRAM research shows only 10^4. It is not clear, however, why it shows such large on/off ratio, but the unique structure of the NEM selector probably result in this unique hysteric I-V curve. The initial goal is that the WL is supposed to be released when the voltage is smaller than the pull-in voltage. But, the WL does not separate from the BL (as shown in Figure 5.3) due to large surface adhesion force and electrostatic force caused by trapped charges in oxide. This mitigates many advantages of this device.

In this initial study, the concept of NEM selector RRAM is demonstrated by a fabricated proto-type device. It shows a unique hysteric I-V characteristic having larger on/off ratio. To understand I-V mechanisms and improve the device performance, further investigation and work are needed.