NANOSCALE MEMRISTIVE DEVICES
FOR MEMORY AND LOGIC APPLICATIONS

by

Sung Hyun Jo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2010

Doctoral Committee:

      Assistant Professor Wei Lu, Chair
      Associate Professor L. Jay Guo
      Associate Professor Cagliyan Kurdak
      Associate Professor Jamie D. Phillips

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

NANOSCALE MEMRISTIVE DEVICES
FOR MEMORY AND LOGIC APPLICATIONS

by

Sung Hyun Jo

Chair: Wei Lu

As the building block of semiconductor electronics, field effect transistor (FET), approaches the sub 100 nm regime, a number of fundamental and practical issues start to emerge such as short channel effects that prevent the FET from operating properly and sub-threshold slope non-scaling that leads to increased power dissipation. In terms of nonvolatile memory, it is generally believed that transistor based Flash memory will approach the end of scaling within about a decade. As a result, novel, non-FET based devices and architectures will likely be needed to satisfy the growing demands for high performance memory and logic electronics applications.

In this thesis, we present studies on nanoscale resistance switching devices (memristive devices). The device shows excellent resistance switching properties such as fast switching time (< 50 ns), high on/off ratio ($> 10^6$), good data retention (> 6 years) and programming endurance ($> 10^5$). The studies suggest that the nonvolatile resistance

switching in a nanoscale a-Si resistive switch is caused by the formation of a single conductive filament within 10 nm range near the bottom electrode. New functionalities, such as multi-bit switching with partially formed filaments, can be obtained by controlling the resistance switching process through current programming. As digital memory devices, the devices are ideally suited in the crossbar architecture which offers ultra-high density and intrinsic defect tolerance capability. As an example, a high-density (2 Gbits/cm^2) 1kb crossbar memory was demonstrated with excellent uniformity, high yield (> 92%) and ON/OFF ratio (> 10^3), proving its promising aspects for memory and reconfigurable logic applications.

Furthermore, we demonstrated that properly designed devices can exhibit controlled analog switching behavior and function as flux controlled memristor devices. The analog memristors can be used in biology-inspired neuromorphic circuits in which signal processing efficiency orders of magnitude higher than conventional digital computer systems can be reached. As a prototype illustration, we showed Spike Timing Dependent Plasticity (STDP), one of the key learning rules in biological system, can be realized by CMOS neurons and nanoscale memristor synapses.

CHAPTER 1

Introduction

1. 1 Background

The success of semiconductor devices, observed for decades and mainly driven by intensive transistor down-scaling, is now facing an uncertain future. As the FET (field effect transistor) size approaches sub 100nm regime, a number of fundamental and practical issues start to emerge such as short channel effects that prevent the FET from operating properly and sub-threshold slope non-scaling that leads to increased power dissipation. Hence the transistor scaling is not only slowed down by manufacturing difficulties in making nanometer scale features but also by fundamental performance limits of transistors. In terms of memory applications, it is generally believed that transistor based Flash memory will approach the end of scaling within about a decade. Hence one of the most important challenges in semiconductor industry is the need of a new memory technology which combines the best features of current memories such as high density of DRAM, fast speed of SRAM and nonvolatile property of Flash with a CMOS compatible fabrication technology.

As a result, novel, non-FET based devices and architectures, which do not suffer from the same problems associated with transistor scaling, will be needed to meet the demands for devices with high density and high performances. Several nonvolatile memory device structures such as Ferroelectric RAM (FeRAM), Magneto-resistive RAM

(MRAM), Organic RAM (ORAM) and Phase Change RAM (PRAM) have been proposed [1]. The emerging memory devices can store information and switch into different states (e.g. '0', '1') by several mechanisms such as ferroelectricity of a dielectric layer in a capacitor which affects stored charges in the capacitor [2], switching of magnetization in ferro-magnetic layers [3], conducting paths formation in organic materials [4] and phase change of a material from an amorphous state to a crystalline state [5]. The states of memory devices can then be distinguished by sensing the current flow through the devices at specific conditions, which reflects the state (stored information) of the memory. Unfortunately, none of the approaches seems ideal. For example, even though FeRAM and MRAM possess fast switching (< 20ns) and large programming endurance (> 10^{14} write/erase), these devices are not CMOS compatible and large (>> $10F^2$, where F is the smallest feature size in the device structure) [6]. For PCRAM, the resistance switching involves Joule heating, which inherently imposes power consumption issues. Furthermore, the phase change is a volume effect which poses questions of whether the phase change can be uniformly controlled in ultimately scaled devices in which surface to volume ratio is very high. For ORAM, the device reliability issues and integration of ORAM with conventional CMOS are still significant challenges due to the relatively unstable organic materials used.

In this chapter, we will discuss another type of memory device - resistive random access memory (RRAM), which has received increasing interest from researchers in both industry and academia and may offer all the necessary traits for future memory applications.

1.2 Resistive Random Access Memory (RRAM)

RRAM is a two terminal device that the switching medium is sandwiched between top and bottom electrodes (Fig.1.1) and the resistance of the switching medium can be modulated by applying electrical signal (current or voltage) to the electrodes. Even though large electrical nonvolatile resistance changes are also observed in FeRAM, MRAM and PRAM, the term RRAM typically refers to memory devices in which ferroelectricity, magnetization and phase change of material states (i.e. amorphous to crystalline) are not involved, hence distinguishing RRAM from FERAM, MRAM and PRAM.

Top electrode

Switching medium

Bottom electrode
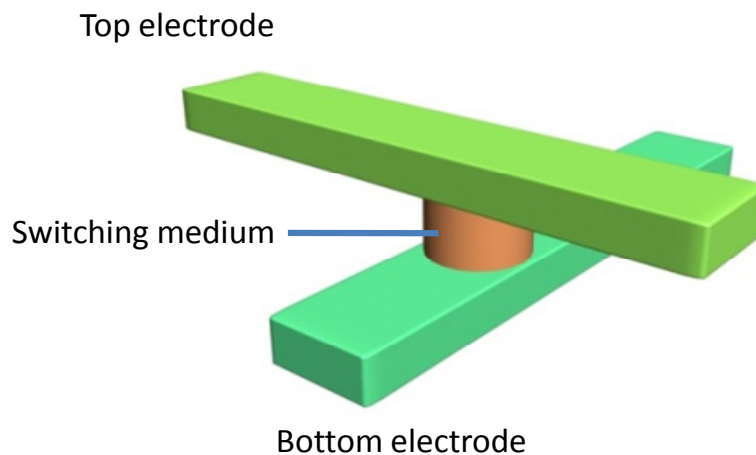
Figure 1.1. Two terminal RRAM structure. The resistance of
a switching medium determines the state of the device.

In fact, the resistance switching effect has been observed in a broad range of materials such as perovskite oxide (e.g. $SrZrO_3$, $LiNbO_3$, $SrTiO_3$) [7-9], binary metal oxide (e.g. NiO, CuO_2, TiO_2, HfO_2) [10-16], solid electrolytes (e.g. AgGeS, CuSiO) [17,

18] and even in some organic materials [19-21]. In general, the type of resistance switching can be divided into bipolar and unipolar behaviors (Fig. 1.2).
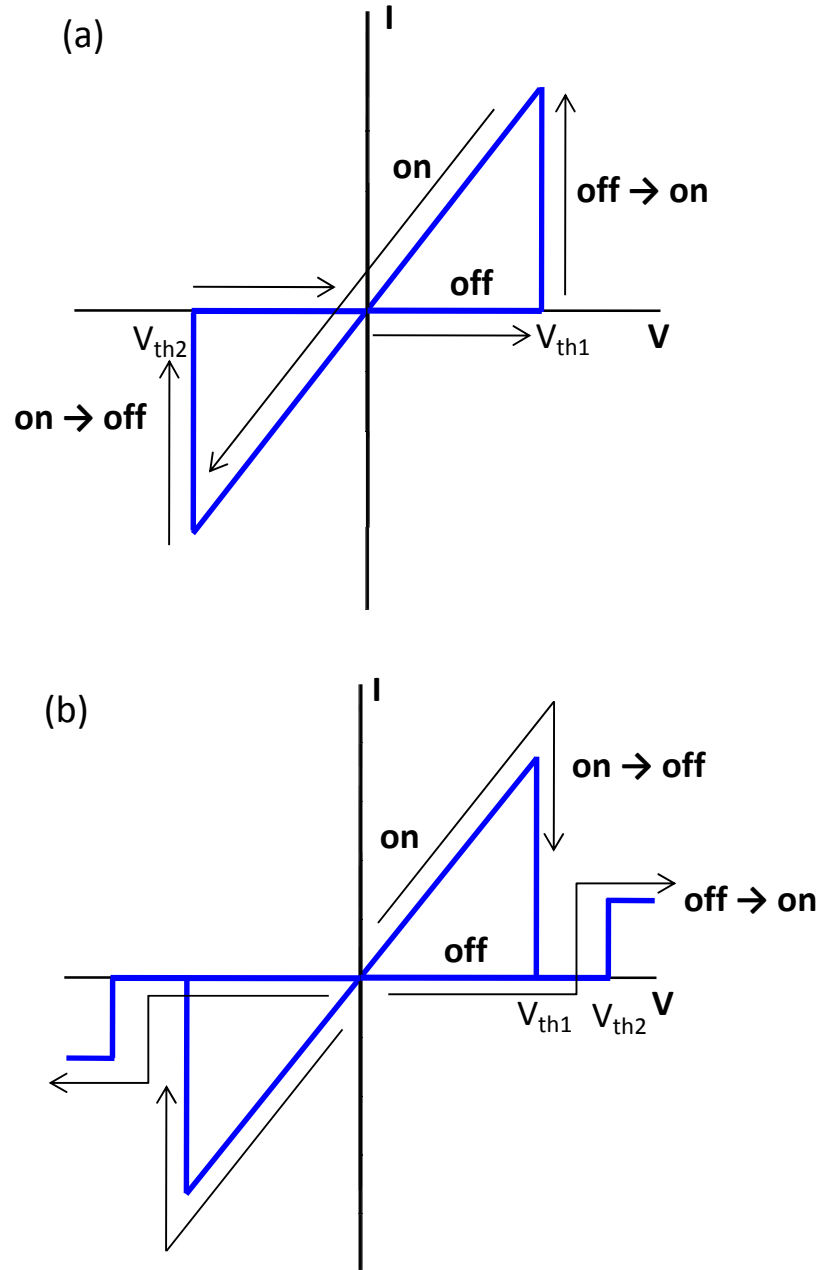
Figure 1.2. Two types of resistance switching. (a) bipolar switching (b) unipolar switching

In bipolar switching, the switching medium changes its resistance depending on the polarity and magnitude of the applied signal. For example, the device can be changed into on-state (low resistance state) when a positive voltage larger than threshold voltage (V_{th1}) is applied to the top electrode, while a negative voltage larger than another threshold voltage (V_{th2}) switches the device back to off-state (high resistance state). The device state is not affected if the applied bias is between two threshold voltage V_{th1} and V_{th2}, enabling the low-voltage read process. Once RRAM is set to specific state, the device can retain the information for a certain period without electrical power, which is termed as the retention time. Because the resistance switching is observed in a broad range of materials, the exact switching mechanism can differ significantly in different materials. In general, since in bipolar switching devices the switching is dependent on the polarity of applied bias the switching is typically associated with electric field inside the device. Widely used switching models are metal ion [16] injection from the electrodes, change of states of trap sites either in bulk or interface of the switching medium such as space charge limited conduction (SCLC) [22] and drift of oxygen vacancies [13].

In case of unipolar devices, switching only depends on the magnitude of the applied bias. Hence switching typically involves Joule heating inside the device [23, 24] and is controlled by adjusting the current levels used to program the device. For examples, to switch off the device, no current compliance is necessary to generate enough heat to rupture the conducting local path inside the device, while the current must be limited to prevent the switching medium from generating excess heat so that the local conducting path is created to turn on the device. In some cases, both bipolar and unipolar switching behaviors are observed in one material [25]. In this case, the switching is related with

5

several mechanisms and one may need to design the device to have one dominant mechanism to control the device performance.

The resistance switching in RRAM can be nonvolatile (data storage > years) and fast (switching speed of a few ns - a few 100ns). In addition, its simple two-terminal structure suggests scaling of RRAM can reach beyond the transistor scaling limit, and allows the fabrication of crossbar memory structure which offers high density, random access and 3D stacking capabilities. As a result, RRAM is considered as one of the most promising approaches for next generation memory technologies.

1.3 Integration of Resistive Random Access Memory (RRAM)

One simple approach to integrate RRAM cells with conventional CMOS circuitry is 1T1R (1 transistor + 1 resistance switching element) structure [12]. As shown in figure 1.3, the switching element is connected with a select transistor in series. The select transistor controls the location of the switching element to be accessed. The 1T1R approach can be readily integrated with CMOS. However, since each switching element requires one transistor in this approach, the storage density will still be limited by transistor scaling and the advantages of the simple two-terminal RRAM devices are not fully utilized.
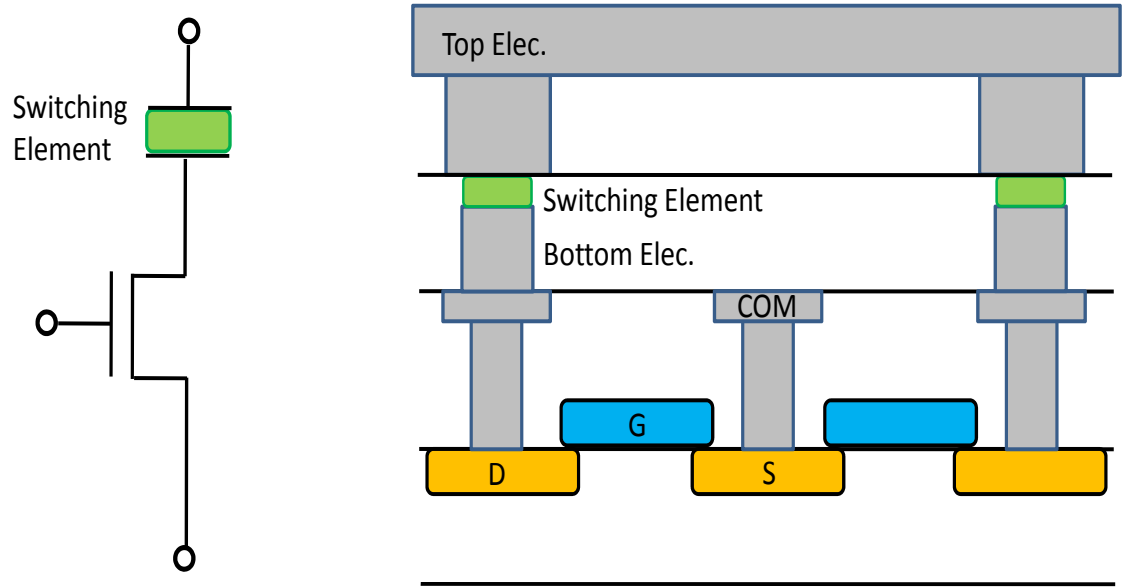
Figure 1.3. A RRAM cell with a select transistor.
The figure was modified from Ref. [12].

The maximum storage density can be achieved in another approach, termed crossbar structure (Fig. 1.4) [26-29]. The crossbar structure consists of an array of parallel bottom nanowire electrodes, an array of parallel top nanowire electrodes with 90° degree with respect to the bottom electrodes and the switching medium between the electrodes. Each cross-point formed at the intersections of the top and bottom electrodes corresponds to an individual RRAM cell. Hence the crossbar cell has a very economical device area of $4F^2$, where F stands for the minimum feature (line) size. If in addition the crossbars can be stacked on top of each other then the effect device area will be $4F^2/n$, where n is the number of stacked crossbar memory layers. Besides offering the highest possible density, the simple, interconnected crossbar network structure offers good defect-tolerance and large connectivity. These features in turn make the crossbar structure not only attractive for high density memory, but also bio-inspired neuromorphic logic

circuits as briefly introduced in chapter 5.
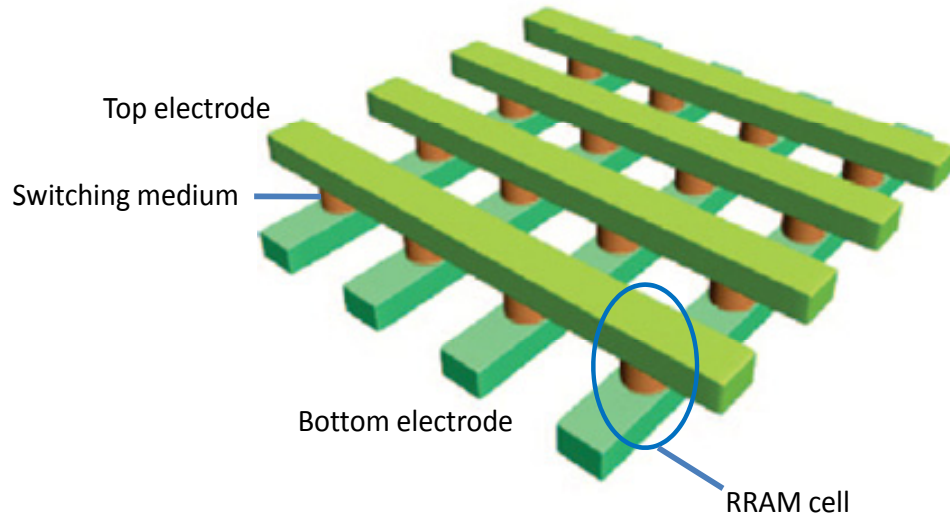
Figure 1.4. The schematic of a crossbar array.

1.4 RRAM and Memristors

The concept of "memristor" was first discussed in 1971 by Prof. Leon Chua at UC Berkeley as the fourth basic circuit element [34]. Similar to the resistor (which relates voltage v and current i), capacitor (which relates charge q and voltage v) and inductor (which relates flux φ and current i), the memristor relates the flux φ with the charge q of the device. Briefly, a memristor is a special type of memristive device in which the change in conductance is linearly proportional to the flux-linkage φ through it, i.e., $i = G(\varphi)i$, where $G(\varphi)$ is the so-called memductance and a linear function of the flux-linkage φ. Mathematically, if q can also be written as a function of φ only, $q = q(\varphi)$, we have

$$\frac{dq}{dt} = \frac{dq}{d\varphi}\frac{d\varphi}{dt} \qquad\qquad (1.1)$$

From $i = \dfrac{dq}{dt}$ and $v = \dfrac{d\varphi}{dt}$, Eq. 1 leads to $G(\varphi) = \dfrac{i}{v} = \dfrac{dq}{d\varphi}$ which shows the memductance.

$G(\varphi)$ indeed relates the charge q with the flux-linkage φ, and such a memristor is termed "flux-controlled memristor". A charge-controlled memristor can be defined in a similar fashion. Since $\varphi = \displaystyle\int_{t=-\infty}^{t} v\,dt$, the conductance of the memristor thus depends not on the current state variables, but on the complete past history of the memristor voltage (or current for a charge-controlled memristor). From a practical point of view, this simple circuit model governing the non-linear memristor characteristics makes it possible to design and predict complex circuits, for example, when implementing the learning rules in neuromorphic circuits discussed later.

In the broadly defined memristive system, the following set of equations needs to be satisfied for a flux-controlled memristive system:

$$i = G(w,v)v \qquad\qquad (1.2)$$

$$\frac{dw}{dt} = f(w,v) \qquad\qquad (1.3)$$

where w is a "state variable" and can be φ, q or other variables related to the device state, and f can be any linear or non-linear function. It is also straightforward to verify that the true flux-controlled memristor satisfies equations 1.2 and 1.3 by letting $w = \varphi$ and $f(\varphi,v) = v$.

Since the conductance of a memristor or a memristive device depends on the

history of the current through it or the voltage applied to it, the device typically exhibit

hysteretic I-V characteristics. In this sense, all RRAM devices fall in the category of the

broadly defined memristive systems since one can always define a state variable w and a

function $f(w,i)$ or $f(w,v)$ to satisfy Equations 1.2-1.3. In this regard, RRAM research and

logic devices based on resistive switches can all be supported by the theoretical

framework developed for memristive devices.

1.5 Current Status of RRAM

The most intriguing RRAM may be devices based on molecules since a single

molecule may be the ultimate functional device. However, fabricating crossbar devices

based on molecules face tremendous challenges of connecting individual molecules to a

pair of electrodes and doing this for a high density array with high yield. So far the

reported RRAM devices based on molecules consist of hundreds of molecules at each

crosspoint that the device size is still limited by the lithography resolution (~ 10 nm)

rather than the size of the molecule itself [21]. In addition, the molecule-based devices

exhibit poor electrical switching properties with low yield, small on/off state current ratio

(<100) and slow switching speed. The exact mechanism of the observed switching effects

is also controversial. Furthermore, polymers or molecules suffer from poor thermal

stability, which adds complexity to controlling the storage medium-electrode interface

and casts questions on the reliability of such devices.

RRAM cells based on perovskite oxides have also been studied and fast (a few

tens of ns) nonvolatile switching and promising multi-level bits per cell have been

reported [7-9]. However, the perovskite based RRAM also suffers from low on/off

current ratio (~100) and short data retention (e.g. 1hour at 150°C). Furthermore, achieving the exact composition ratio of three different elements (in the form of ABO_3) in perovskite oxides has been a key challenge. As a result, nanometer scale perovskite based resistive devices are seldom reported.

Another common material group used for RRAM is transition metal oxide [10-16]. Several promising aspects for memory application such as nanosecond-order switching, multi-level bit storage and scalability have been reported. However, the programming current is typically large (>100μA), resulting in large power consumption and limited write/erase endurance. Moreover, the retention time (data storage period) is typically short (a few months).

1.6 Amorphous Silicon and Amorphous Silicon based RRAM

Amorphous silicon (a-Si) based RRAM has been intensively investigated in the past [30-33]. The resistance switching behavior was explained in terms of metal filament formation (elimination) inside the a-Si matrix at positive (negative) applied voltages. The device fabrication is fully compatible with CMOS processing technology. However, conventional metal/a-Si/metal (M/a-Si/M) based devices require a high voltage (>10V) forming process which is not fully controllable and reduces the device yield and endurance (typically a few thousand programming cycles). It is also not clear whether such devices can stand the destructive forming process as the device scales down to nanoscale. As a result, few studies have been attempted on a-Si resistance switching devices as ultra-high density memories during the past two decades.

In this thesis, we present studies on nanoscale amorphous Si based resistive

switching devices in which the effect of forming is minimized while the advantages of amorphous silicon based devices are retained and amplified. Our studies show that two terminal resistive switching devices based on M/a-Si/p-Si device structure are ideally suited in the crossbar architecture which offers ultra-high density and intrinsic defect tolerance capability. The robust switching characteristics allow us to carefully examine the switching mechanisms and design techniques to control the switching process. Furthermore we demonstrate that such high performance, nonvolatile resistance switching memories may be readily implemented through CMOS compatible, conventional processing processes. Finally, we will demonstrate that well-designed two terminal devices can show behaviors similar to synapses in biological systems. As a prime example, a hybrid CMOS neuron/memristive synapse system has been demonstrated to show STDP (Spike Timing Dependent Plasticity) property which is one of the fundamental learning rules in biology systems.

Amorphous silicon (a-Si), unlike crystalline silicon in which atoms are bonded with four other atoms in tetrahedral direction, has no long-range order and forms a continuous random network. Dangling bonds are presented in amorphous silicon due to the nature of the random network. Engineering the density of dangling bonds (defects) is critical to control the device performance based on amorphous silicon. Amorphous silicon can be deposited by various methods such as evaporation, sputtering and chemical vapor deposition at low temperature (< 200°C for CVDs and as low as room temperature for PVDs), hence a-Si can be deposited on various substrates including plastic substrates.

Typically, amorphous silicon deposited by PVD methods has few applications as a device, since the defect level is too high that reasonable electrical or optical performance

can not be obtained. For CVD methods, hydrogen can be incorporated during the deposition by the decomposition of silane (Si_3H_4) or adding a small amount of hydrogen gas. Hydrogen reduces the defect density and makes defect level engineering possible during the deposition. The lowest defect density can be obtained from the substrate temperature range of 220 ~ 270°C for plasma enhanced chemical vapor deposition. At lower temperature, hydrogen does not have enough energy to form a bond with silicon; while at higher temperatures hydrogen effuses to the outside of amorphous silicon and generating defects.

Unlike applications such as solar cells and thin film transistors in which the optical and electrical properties of amorphous silicon are of importance, for RRAM applications the disordered amorphous silicon network serves as the host matrix for ion storage in resistance switching devices, as will be discussed in chapter 2 - 4.

1. 7 Organization of the Thesis

The organization of the thesis is as follows. In Chapter 2, studies on single device level nanoscale metal/amorphous silicon (a-Si)/crystalline silicon (c-Si) structure for resistive switches will be discussed. In Chapter 3, detailed switching mechanism in a-Si based resistive random access memory (RRAM) will be covered. In Chapter 4, the integration of a-Si RRAM in crossbar structure to form a 1kb memory array will be demonstrated with high device yield and performance uniformity. In Chapter 5, studies on analog memristive devices based on amorphous silicon and tungsten oxide will be presented. Furthermore, synaptic functions (e.g. STDP) will be demonstrated in a hybrid CMOS neuron/memristive synapse system. Finally, in Chapter 6, rectifying switching

characteristics by utilizing the breakdown phenomenon will be discussed as a means to suppress leakage paths in crossbar arrays.

References

[1]   H. Goronkin, Y. Yang, "High-performance emerging solid-state memory technologies," *MRS Bull*. **29**, 805, 2004.

[2]   H. Kanaya, K. Tomioka, T. Matsushita, M. Omura, T. Ozaki *et al*., "A 0.602 μm^2 nestled 'Chain' cell structure formed by one mask etching process for 64 Mbit FeRAM," *VLSl Tech*. *Dig*., 150, 2004.

[3]   A.R. Sitaram , D.W. Abraham, C. Alof, D. Braun, S. Brown *et al*., "A 0.18 μm logic-based MRAM technology for high performance nonvolatile memory applications," *VLSl Tech*. *Dig*., 15, 2003

[4]   R. Sezi, A. Walter, R. Engl, A. Maltenberger, J. Schumann *et al*., "Organic Materials for High-Density Non-Volatile Memory Applications," *IEDM Tech*. *Dig*., 259, 2003.

[5]   S. Lai, "Current status of the phase change memory and its future," *IEDM Tech*. *Dig*., 256, 2003.

[6]   International technology roadmap for semiconductors (ITRS) 2007 Ed.

[7]   Y. Watanabe, J. G. Bednorz, A. Bietsch, Ch. Gerber, D. Widmer *et al*., "Current-driven insulator–conductor transition and nonvolatile memory in chromium-doped SrTiO3 single crystals ," *Appl*. *Phys*. *Lett*. **78**, 3738, 2001.

[8]   C. Y. Liu, P. H. Wu, A. Wang, W. Y. Jang, J. C. Young *et al*., "Bistable resistive switching of a sputter-deposited Cr-doped SrZrO3 memory film," *IEEE Electron Device Lett*. **26**, 351, 2005.

[9]   A. Beck, J. G. Bednorz, C. Gerber, C. Rossel, D. Widmer *et al*., "Reproducible switching effect in thin oxide films for memory applications," *Appl*. *Phys*. *Lett*. **77**, 139, 2000.

[10]   I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo *et al*., "Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," *IEDM Tech*. *Dig*, 587, 2004.

[11]   M. -J Lee, S. Han, S. H. Jeon, B. H. Park, B. S. Kang *et al*., "Electrical Manipulation of Nanofilaments in Transition-Metal Oxides for Resistance-Based Memory," *Nano Lett*. **9**, 1476, 2009.

[12]   A. Chen, S. Haddad, Y.C. Wu, T.N. Fang, Z. Lan *et al*., "Non-volatile resistive switching for advanced memory applications," *IEDM Tech. Dig*, 746, 2005.

[13]   J. J. Yang, M. D. Pickett, X. Li, D. A. A .Ohlberg, D. R. Stewart *et al*., "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nat. Nanotechnol*. **3**, 429, 2008.

[14]   I. -S. Park, K. -R. Kim, S. Lee, J. Ahn, "Resistance Switching Characteristics for Nonvolatile Memory Operation of Binary Metal Oxides," *Jpn. J. Appl. Phys*. **46**, 2172, 2007.

[15]   S. Lee, W. -G. Kim, S. -W. Rhee, K. Yong, "Resistance Switching Behaviors of Hafnium Oxide Films Grown by MOCVD for Nonvolatile Memory Applications," *J. Electrochem. Soc*. **155**, H92, 2008.

[16]   R. Waser, and M. Aono, "Nanoionics-based resistive switching memories,"*Nat. Mater*. **6**, 833, 2007.

[17]   M. N. Kozicki, C. Gopalan, M. Balakrishnan, M. Park, M. Mitkova, "Non-Volatile Memory Based on Solid Electrolytes ," *NVMTS,* 2004

[18]   T. Sakamoto, S. Kaeriyama, H. Sunamura, M. Mizuno, H. Kawaura *et al*., "A nonvolatile programmable solid electrolyte nanometer switch," *ISSCC*, 2004.

[19]   W. L. Kwan, R.J. Tseng, W. Wu, Q. Pei, Y. Yang *et al*., "Stackable Resistive Memory Device Using Photo Cross-linkable Copolymer," *IEDM Tech. Dig*., 237, 2007.

[20]   R. J. Tseng, C. Tsai, L. Ma, J. Ouyang, C. S. Ozkan *et al*., "Digital memory device based on tobacco mosaic virus conjugated with nanoparticles," *Nat. Nanotechnol*. **1**, 72, 2006.

[21]   Y. Chen, G. Y. Jung, D. A. A. Ohlberg, X. M. Li, D. R. Stewart *et al*, "Nanoscale molecular-switch crossbar circuits," *Nanotechnol*. **14**, 462, 2003.

[22]   T. W. Hickmott, "Low-Frequency Negative Resistance in Thin Anodic Oxide Films," *J. Appl. Phys*. **33**, 2669, 1962.

[23]   S. H. Chang, S. C. Chae, S. B. Lee, C. Liu, T. W. Noh *et al*., "Effects of heat

dissipation on unipolar resistance switching in Pt/NiO/Pt capacitors," *Appl. Phys. Lett.* **92**, 183507, 2008.

[24]   D. Lee, S. Baek, M. Ree, O. Kim, "Unipolar resistive switching characteristic of semiconducting poly (o-anthranilic acid) film," *Electronics Lett.* **44**, 596, 2008.

[25]   C. Schindler, S.C.P. Thermadam, R. Waser, M. N. Kozicki, "Bipolar and Unipolar Resistive Switching in Cu-Doped SiO$_2$," *IEEE Trans. Electron. Dev.* **54**, 2762, 2007.

[26]   J. R. Heath, P. J. Kuekes, G. S. Snider, R. S. Williams, "A Defect Tolerant Computer Architecture: Opportunities for Nanotechnology," *Science* **280**, 1716, 1998.

[27]   W. Lu, C. M. Lieber, "Nanoelectronics from the bottom up," *Nat. Mater.* **6**, 841, 2007.

[28]   A. DeHon, "Array-Based Architecture for FET-Based, Nanoscale Electronics," IEEE Trans. on Nanotechnol. **2**, 23, 2003.

[29]   J. C. Scott, "Is There an Immortal Memory?" *Science* **304**, 62, 2004.

[30]   A. E. Owen, P. G. Le Comber, W. E. Spear, J. Hajto, "Memory switching in amorphous silicon devices," *J. Non-Cryst. Solids* **59 & 60**, 1273, 1983.

[31]   J. Kanicki Ed., Amorphous & Microcrystalline Semiconductor Devices Volume II: Materials and Device Physics, Artech House, 1992.

[32]   P. G. LeComber, A. E. Owen, W. E. Spear, J. Hajto, A. J. Snell *et al*, "The switching mechanism in amorphous silicon junctions," *J. Non-Cryst. Solids* **77 & 78**, 1373, 1985.

[33]   M. Jafar, D. Haneman, "Switching in amorphous-silicon devices,"*Phys. Rev. B* **49**, 13611, 1994.

[34]   L. O. Chua, "Memristor-The Missing Circuit Element," *IEEE Trans. on Circuit Theory* **18**, 507-519, 1971.

Chapter 2
Resistance Switching in Amorphous Silicon

2.1 Introduction

Resistance switching behavior has been observed and studied in micrometer scale amorphous silicon (a-Si) devices since the 1980s [1-6]. A typical device consists of a pair of metal electrodes sandwiching an a-Si layer in a so-called Metal/a-Si/Metal (M/a-Si/M) structure, in which the voltage applied across the pair of metal electrodes may induce changes in the a-Si resistance. These conventional M/a-Si/M based nonvolatile switching devices have the advantages of high on/off resistance switching ratio, and can be fabricated with a CMOS compatible simple fabrication process using only common materials. However, before the devices can be used as a switch, they need to go through a high voltage "forming" process (typically > 10 V) which significantly reduces the yield of devices. It is also not clear whether nanoscale devices can stand such destructive forming process as the device is scaled down, since *micrometer scale* metal "filaments" have been typically observed after the forming process.

In this chapter, we report our initial studies on Metal/a-Si/p-Si devices to show that the effect of forming can be minimized by replacing the metal bottom electrode with a heavily doped Si electrode, and that reliable resistance switching can indeed be obtained in nanoscale Metal/a-Si/p-Si device with high yield. The ability to obtain robust

1 7

switching characteristics in nanoscale Metal/a-Si/p-Si devices allows us to perform detailed analysis on the switching mechanism and build high density memory arrays that will be discussed in the subsequent chapters.

2.2 Device Structure and Fabrication

The device consists of a top metal electrode, an active a-Si layer and a heavily doped ($\leq$0.005 $\Omega \cdot$cm) p-type Si substrate as a bottom electrode. The active a-Si layer was grown either by plasma enhanced chemical vapor deposition (PECVD) or by low pressure chemical vapor deposition (LPCVD). During PECVD, a mixture of SiH_4 (45 sccm) and He (500 sccm) was used with a deposition rate of 80 nm per minute (T = 260 °C, P = 600 mTorr, model: GSI UltraDep 2000). During LPCVD, SiH_4 (80 sccm) was used with a deposition rate of 2.8 nm per minute (T = 585 °C, P = 100 mTorr, model: THERMCO TMX-9000). We note that portions of poly-silicon grains may have formed during the LPCVD process, but the deposited Si film will still be referred to as a-Si here for simplicity. Following the a-Si growth, 100 nm thick silicon oxide was deposited by PECVD (figure 2.1 (a)). Polymethyl methacrylate (PMMA) was then coated on the oxide layer, followed by e-beam lithography and reactive ion etching (RIE) which created windows in the oxide layer to define the active device area (figure 2.1 (b, c)). A second e-beam lithography or photolithography process was then performed followed by e-beam evaporation of the electrode material to pattern the top metal electrode to complete the M/a-Si/p-Si device structure (figure 2.1 (d)). More than 40 batches of devices were fabricated and examined for each selected active area. Figure 2.1 (e) shows the cross-sectional SEM image of the fabricated device.
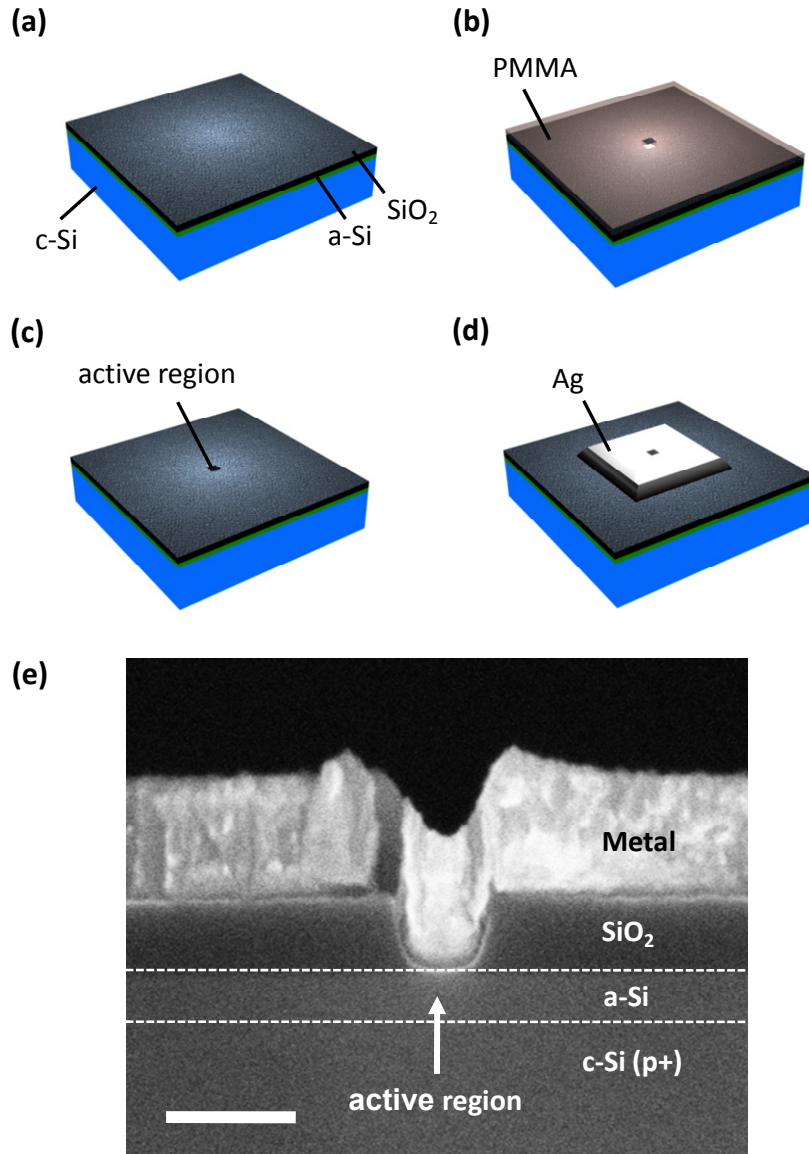
Figure 2.1. The device fabrication flow (a-d) and a cross-sectional SEM image of a fabricated device (e). (a) Amorphous silicon and silicon dioxide deposition on a boron doped crystalline silicon wafer. (b) PMMA coating and e-beam lithography to define active region. (c) Reactive ion etching (RIE) to etch the silicon oxide followed by PMMA removal. (d) Top metal deposition and patterning aligned to the active region. (e) Active region forms Metal/a-Si/c-Si junctions. Scale bar: 100nm.

2.3 Switching Characteristics

The M/a-Si/p-Si devices showed reliable nonvolatile resistive switching and the basic operation characteristics are similar to those of conventional M/a-Si/M devices [1-6]. Due to the highly resistive a-Si layer, the as-fabricated devices show high resistance and negligible current between the top and bottom electrodes. When a positive voltage applied on the top metal electrode is increased to be larger than a certain threshold voltage V_{th1}, which usually ranges from 3~4V depending on the deposition method of the a-Si layer and the type of the top metal, the resistance of the device is suddenly reduced (figure 2.2) and the device is turned on. If a negative voltage is applied on the top metal electrode beyond another threshold V_{th2}, typically ranging from -2V to -4V in magnitude, the device is changed into a high resistance state, i.e., turned off. The off-state resistance is comparable to the resistance of as-fabricated devices. The device state is not affected if the applied voltage is between the V_{th1} and V_{th2}. The resistance ratio between the off and on-states R_{off}/R_{on} ranges from 10^3 to 10^7 depending on the device type.

As shown in figure 2.2 (a), even for the smallest devices tested (active area of 50 nm × 50 nm), only a slightly higher write voltage was needed for the as-fabricated device compared to that for subsequent cycles. Once formed, V_{th1} and V_{th2} show little dependence on the number of write/erase cycles. The reduced forming voltage compared with M/a-Si/M devices may be partly explained by the fact that high-electrical fields can be readily generated inside the a-Si region since the metal/a-Si/p-Si device effectively forms a reverse-biased Schottky junction at positive bias. The reverse-biased Schottky junction may also help limit damage to the device after filament formation in terms of

heating or electromigration, hence improving the device yield. Our control experiments further show that devices fabricated on n-type substrates do not exhibit resistance switching behavior. Since a forward biased Schottky diode will be formed this time, a large leakage current is normally detected and no abrupt switching were observed. By eliminating the high voltage forming process, we were able to obtain a much higher device yield compared to conventional M/a-Si/M structures (of more than three hundred devices examined, only two devices were found not to exhibit resistance switching.) As a comparison, our control experiments showed that M/a-Si/M devices with otherwise identical structures still require forming voltages > 10 V, with a device yield < 40%.
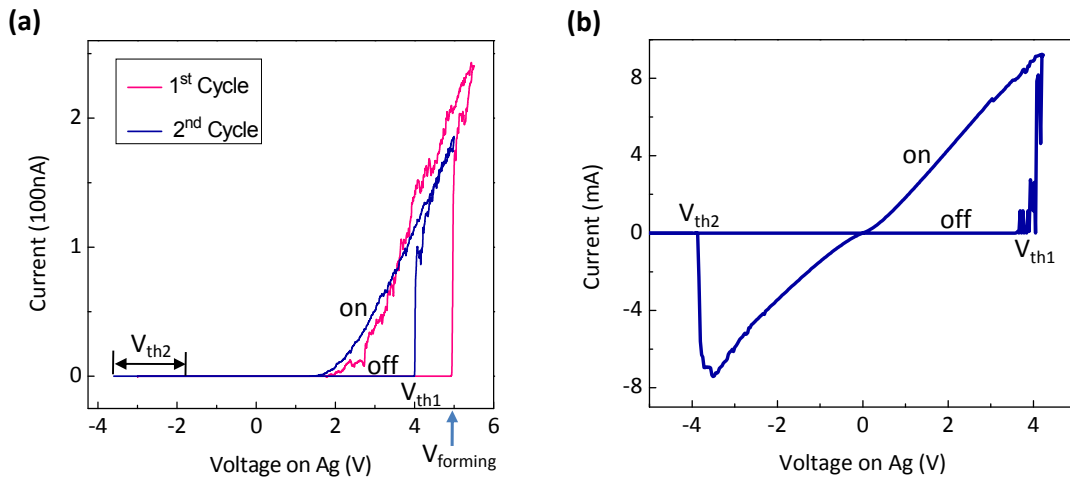
Figure 2.2. Resistance switching characteristics of Ag/a-Si/p-Si devices. (a) Rectifying resistance switching. The device consists of a Ag top electrode, a PECVD deposited 80 nm thick a-Si layer with 50 nm × 50 nm active area. (b) Non-rectifying resistance switching. The LPCVD deposited 60 nm thick a-Si layer was used for a switching medium.

Compared with phase-change memory devices which show no polarity dependence [11] and conductive-bridge devices based on electrochemical reactions that

depend crucially on the electrode material [12, 13], the polarity dependence (i.e., devices can only be written with a positive voltage and erased with a negative voltage) and insensitivity to the choice of electrode material of the M/a-Si/p-Si devices suggest that the observed resistance switching behavior is caused by a filament formation process, as was studied previously in conventional microscale metal/a-Si/metal (M/a-Si/M) devices [1-6]. Briefly, we believe positive voltages on the top metal electrode drives the metal Ag ions into the a-Si matrix and forms conducting filaments in the a-Si layer, resulting in reduced resistance [4]. The Ag ions are retracted from the a-Si layer when a negative voltage beyond V_{th2} is applied, thus recovering the high resistance. The resistance remains unchanged for voltages between V_{th1} and V_{th2}. The detailed switching mechanism will be discussed in Chapter 3 instead.

2.4 Area Dependence

We have tested devices with active areas ranging from 2.5×10^{-3} μm^2 (50 nm × 50 nm) to 1×10^3 μm^2. No sign of device degradation was observed down to 50×50 nm^2, the smallest devices tested. The high device yield at 50×50 nm^2 and continued improvement of I_{on}/I_{off} suggest that the device operation can likely reach the 20 nm range, a limit set by fabrication capabilities of the metal electrodes in the crossbar structure [9]. Figure 2.3 (a) shows the dependence of the on-state resistance on the active device area. Significantly, the on-resistance increases only 2.5 times when the active area is reduced by 6 orders of magnitude in size. This observation supports the filament conduction mechanism and indicates the conducting filaments form locally. Briefly, in the on-state conduction is dominated by the current through the metallic filaments hence are insensitive to the total

device area. On the other hand, in the off state conduction is dominated by leakage through the a-Si layer so R_{off} is inversely proportional to the device area. As a result, improved R_{off}/R_{on} ratio can be achieved in smaller devices, which is an effect that actually favors down scaling of device sizes.
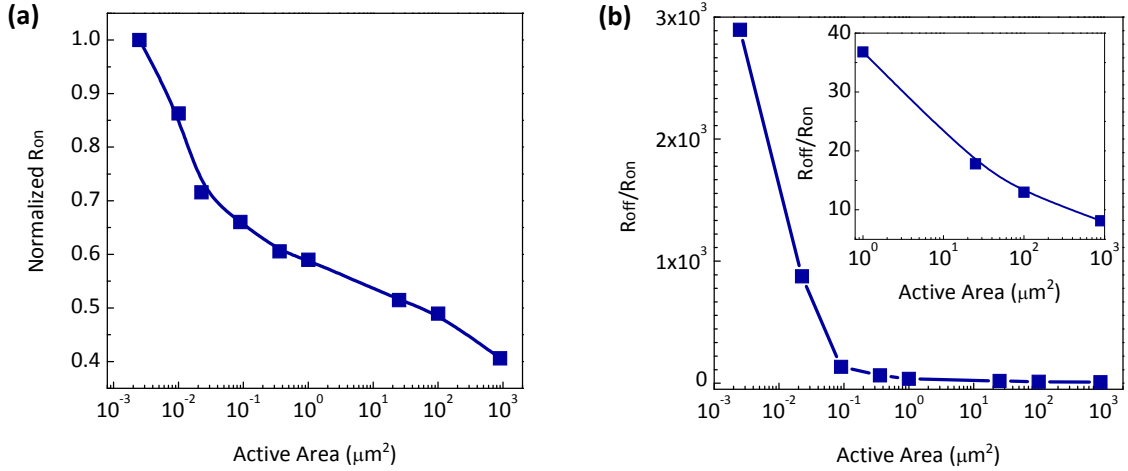
Figure 2.3. Scaling property of devices (a) Dependence of the on resistance on the active device area. The on resistance was normalized to that of the device with the smallest active area. The a-Si layer (120 nm thick) in these devices was deposited by PECVD at 260 °C. (b) Dependence of off/on resistance ratio on the active area for the same devices shown in (a). Inset: the blowup for devices with active area >1 μm^2.

2.5 Rectifying vs. Non-Rectifying Switching

Interestingly, two different switching characteristics, rectifying and non-rectifying I-V curves, were obtained in M/a-Si/p-Si devices in a controllable manner as shown in figure 2-4. As for rectifying resistance switching (figure 2-2 (a)), the device shows a diode-like behavior at on-state, that is, the current in on-state can only flow at positive biases but not at negative biases. Even though no current is detected, the device still remains in the on-state as long as the applied negative bias is small (< ~ -2V in magnitude), as can be verified in subsequent read operations [7]. The device can only be

truly erased at large negative programming voltages ($V < V_{th2} \sim -2V$). The intrinsic rectifying is a desirable property as it eliminates crosstalk in array-level integration and may help the implementation of crossbar-based logic circuits [8-10]. On the other hand, a symmetric I-V was observed for the device in Fig. 2.2(b) in the on-state and the device behaves like a resistor. This behavior is termed the non-rectifying switching case.
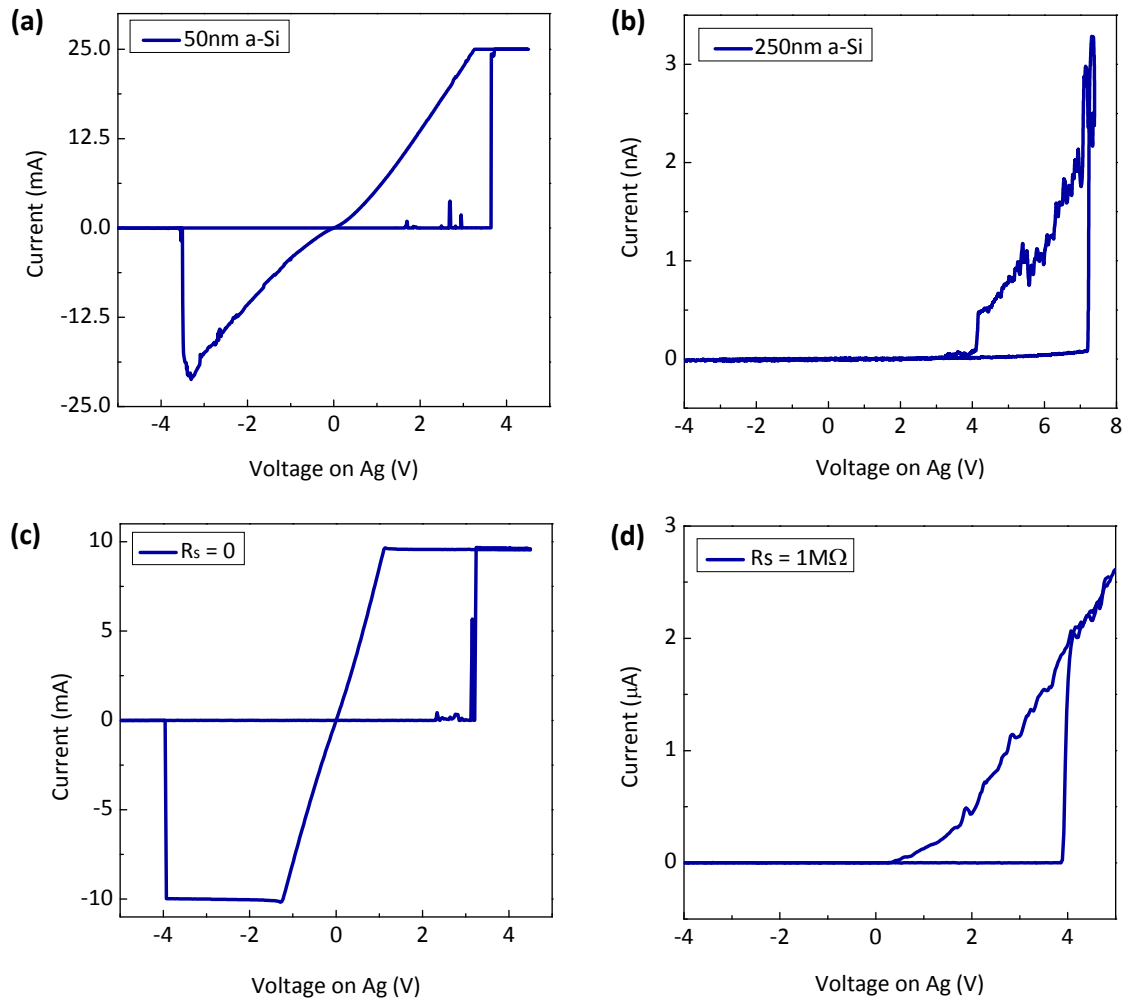
Figure 2.4. Control of the rectifying and non-rectifying behaviors. Switching characteristics of devices with 50 nm thick a-Si (a) and 250 nm thick a-Si (b) deposited by LPCVD. (c) Switching characteristics of another device with 40 nm thick a-Si deposited by LPCVD without series resistors. (d) Switching characteristics of the same device with a 1 MΩ series resistor. Current compliance: (a) 25mA , (c) 10mA .

2 4

In general, we observed that devices fabricated with a-Si grown by the PECVD method showed rectifying switching behavior, and devices with a-Si grown by the LPCVD method showed non-rectifying behavior. The different behaviors correlate well with the different on-state resistances R_{on} in the two types of devices: R_{on} is much higher in the PECVD prepared devices (figure 2.2 (a), the rectifying case) compared to that of the LPCVD prepared devices (figure 2.2 (b), the non-rectifying case). The lower R_{on} for LPCVD devices can be explained by the higher deposition temperature and lower H_2 concentration used in the LPCVD process which likely causes a higher defect density and provides more and closely spaced filaments in the on-state. Further studies confirmed the role of R_{on} in the switching behavior. By adjusting the thickness of the a-Si layer during LPCVD growth, we were able to adjust R_{on} systematically and observed the transition from non-rectifying (figure 2.4 (a)) to rectifying behavior (figure 2.4 (b)) as the thickness of a-Si hence R_{on} was increased. Furthermore, rectifying behavior can be obtained on existing low- R_{on} LPCVD devices by adding a series resistor Rs, as demonstrated in figures 2.4 (c), (d).

We observed that if the on-current is smaller (greater) than ~10 μA (measured at V_{th1} for devices with active area of 50 nm × 50 nm), the switching characteristics is likely rectifying (non-rectifying). The non-rectifying behavior may be related to the more permanent changes in the a-Si matrix at large write currents, likely due to permanent trapping of Ag ions near the a-Si/p-Si interface [7] or closely placed Ag ions. We note that the large range of R_{on} that can be systematically obtained in the a-Si memory devices, which offers the potential to tune the device parameters to suit specific requirements.

2.6 The Potential for Memory Applications

2.6.1 Data Retention

Once the device was written as either on-state or off-state, no degradation of the stored data was observed in both rectifying and non-rectifying switching devices for 150 days at room temperature in air ambient as represented in figure 2.5. For similar devices of M/a-Si/M structure, it has been shown that the device can retain the memory state without any degradation over one year at room temperature [14]. Our additional studies on improved device structures also suggest data retention time > 6 years (Chapter 3). These results ensure that the device has a potential for the nonvolatile memory applications.
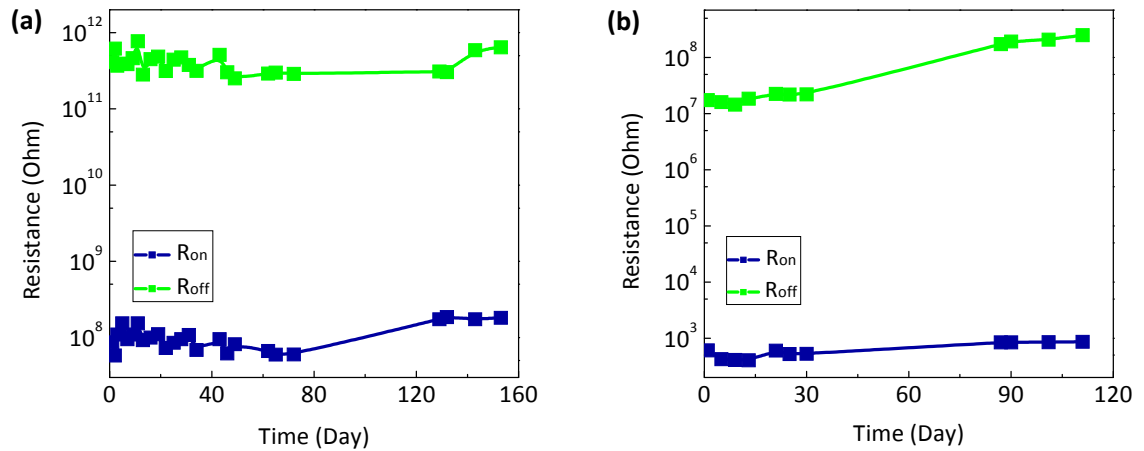
Figure 2.5. Retention of devices. Retention test of a device prepared by (a) PECVD with an active area of 150nm × 150 nm (b) LPCVD with an active area of 300nm × 300 nm. The measurement was performed in air ambient at room temperature.

2.6.2 Programming Endurance

The endurance test was performed on various types of devices. The endurance test results of two extreme cases with programming current levels 10nA and 10mA (corresponding to the rectifying and non-rectifying switching behaviors respectively) are presented here. For devices programmed with a very low programming current, as the number of write and erase cycles is increased, the on-state current is increased and the turn-on threshold voltage V_{th1} tends to be decreased as shown in figure 2.6. No change is detected in the off-state since R_{off} is always larger than the equipment limit. The switching behavior is still reliable even after 10^6 cycles. In addition, if the erase time is increased or the erase voltage is increased (e.g. from -3V to -5V), V_{th1} and R_{off}/R_{on} ratio can be recovered back to their original values. For devices programmed with a high programming current, the general trend is similar to the rectifying switching devices except that the endurance is generally reduced.

For example, in devices programmed with a current of 10 mA, the R_{off}/R_{on} ratio starts to decrease after 10^5 cycles and becomes ~10 after 4×10^5 cycles mainly due to the increased off-current level. This phenomenon may be caused by the residue of metal ions driven into a-Si matrix and the accumulated metal ions contribute to the increased off current. We note the larger than 10^6 cycles of endurance observed in devices with low programming currents are already comparable to conventional flash memory devices.
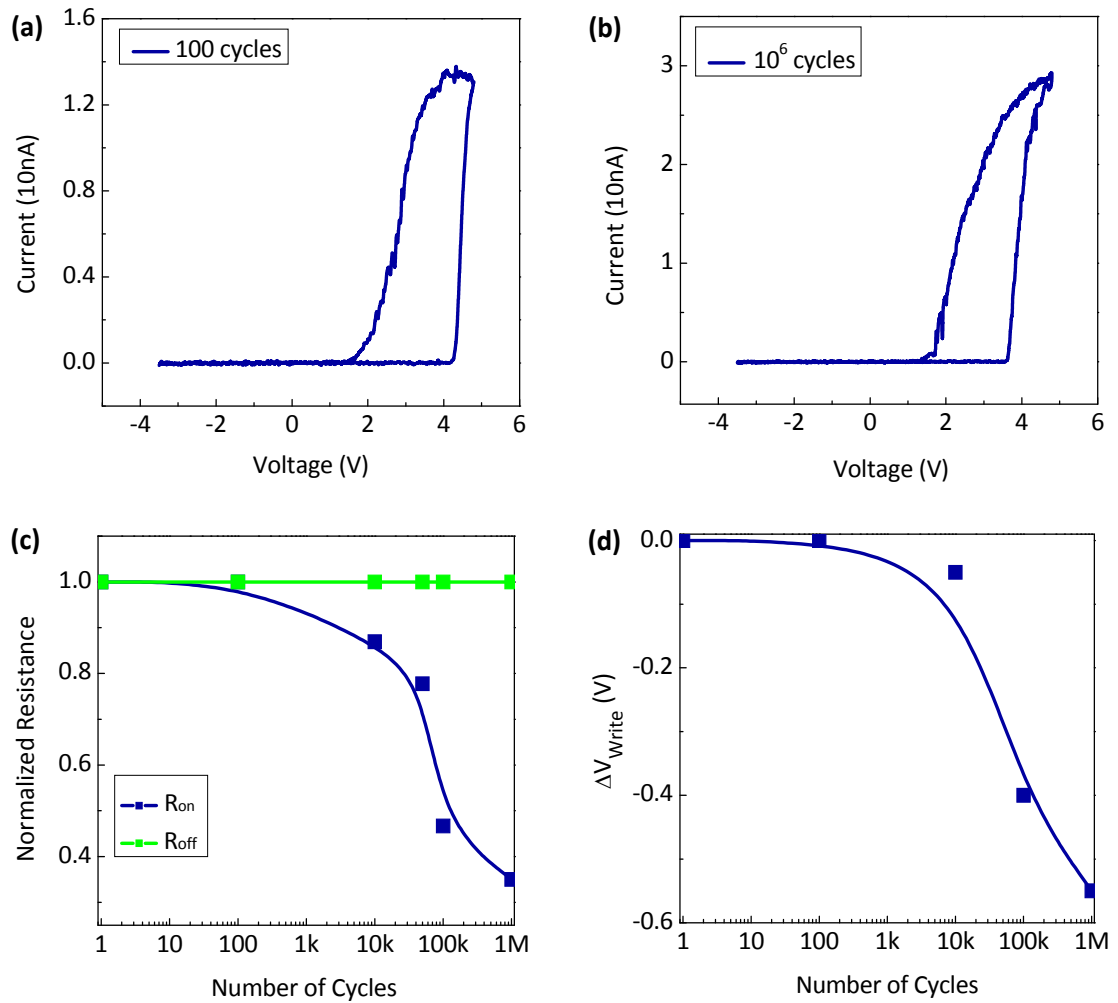
Figure 2.6. Endurance properties of devices. (a-b) Switching characteristics of prepared by PECVD with an active area of 150nm×150nm after 10^2 and 10^6 cycles of write/erase operations, showing excellent switching characteristics and on/off ratio after 10^6 cycles. (c) Dependence of the on and off resistances on the number of programming cycles for a similar device as in (a-b). (d) The shift of threshold voltage V_{th1} for write process versus number of write-erase cycles.

2.6.3 Switching Speed

To examine the speed of the device, very short write and erase pulses were applied and the memory state was probed by measuring the voltage across a resistor in series with the device with a digital oscilloscope. For example, a write/read/erase/read sequence consisting of 5ns duration of 6.5V write pulse, 5ns duration of -6.5V erase pulse and 1.5V read signal was shown in Figure 2.7. At the beginning, the device was turned off. Hence when 1.5V read signal was applied, most of voltage was dropped across the switching device and no voltage drop appeared across the series resistor. When 5 ns duration of 6.5V write pulse was supplied to the switching device, the device switched on and part of the read voltage dropped across the series resistor. The -6.5 V erase pulse turned off the device again and the voltage drop across the series resistor was reduced again to nearly zero. We have routinely observed switching speed faster than 5 ns from the devices with a few mA on-current. This programming speed is the fastest among all proposed RRAM devices and verified that such devices are very promising for high speed application.
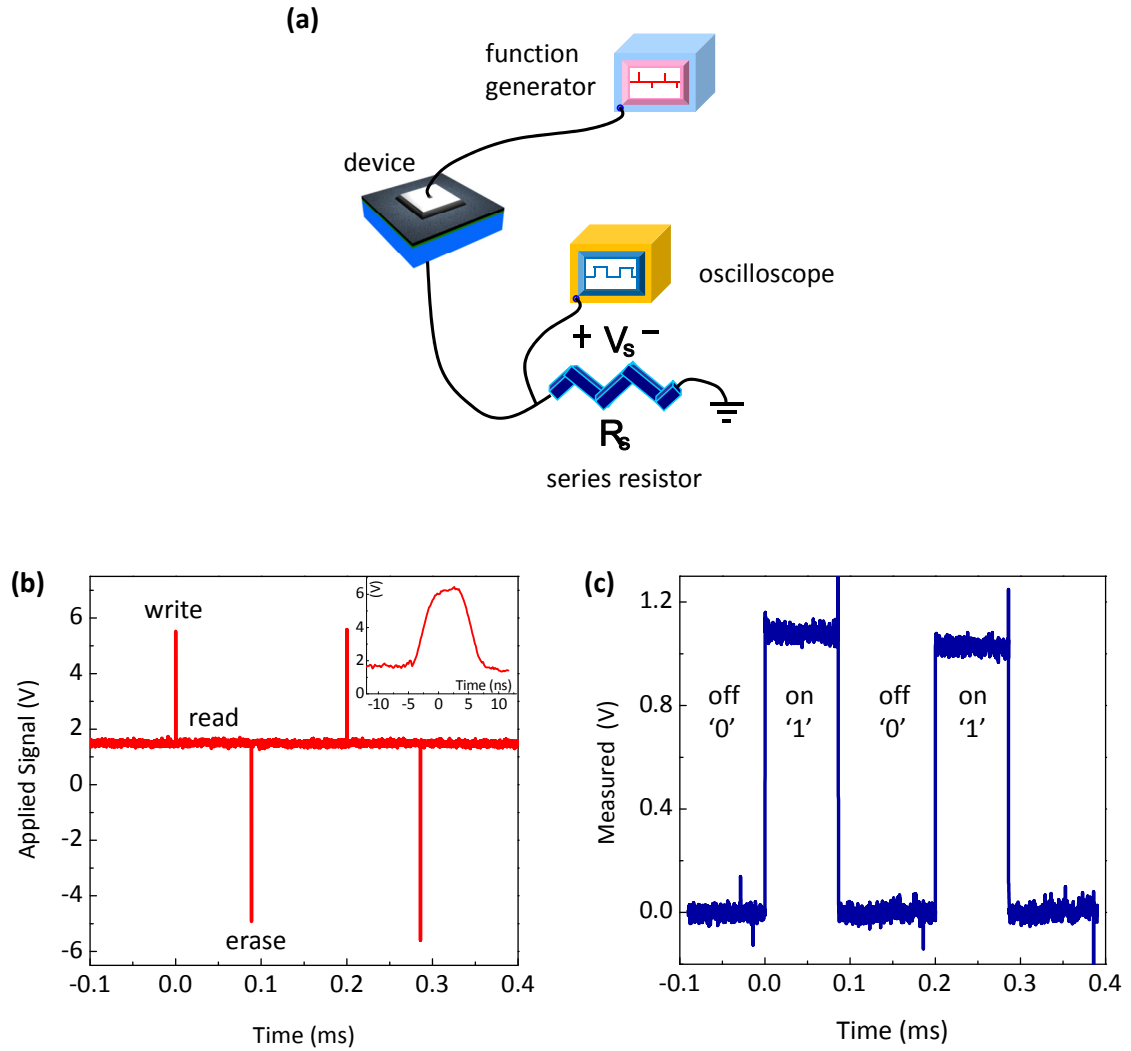
Figure 2.7. Switching speed tests of the M/a-Si/p-Si devices. (a) Schematic figure of the measurement setup. The voltage across a series resistor reflects the state of a device. (b) Representative programming signal used for the switching speed test. Inset: blow-up of the write pulse. (c) The corresponding output signal measured on the sensing resistor R_s. The device in (b) and (c) has a 40 nm thick a-Si layer deposited by LPCVD.

2.6.4 Multi-bit Storage Capability

The capability of storing multiple bit levels in one storage element is another important criterion in accessing emerging memory technologies. We have tested multi-

level bit capability of the M/a-Si/p-Si devices by controlling the maximum write programming current using a series resistor setup. The dependence of the on-state resistance on programming current is shown in figure 2.8 and clearly demonstrates the multi-bit potential of the M/a-Si/p-Si devices. The principle of the multibit storage will be discussed in detail in Chapter 3.
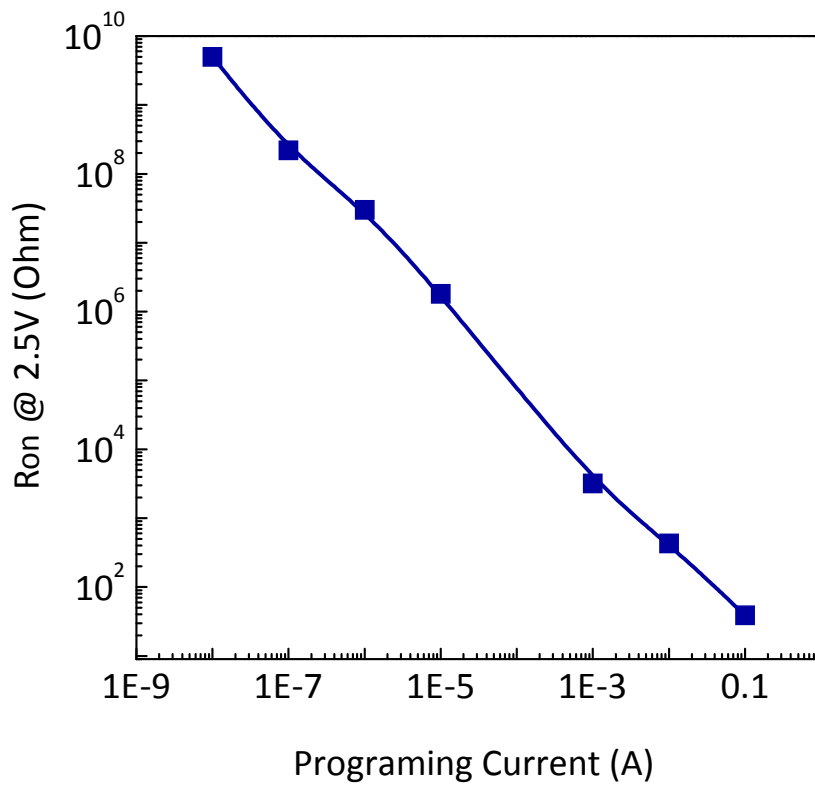
Figure 2.8. Dependence of the on-resistance on the programming current, demonstrating the multi-level bit capability. The device was fabricated with LPCVD with 40 nm thick a-Si.

2.6.5 Controllable Programming Current

Another merit of this device is that the on-current can be fully tunable by controlling amorphous silicon deposition conditions, which allow for the device to be used for either ultra low power consumption (small I_{on}) or high speed (low R_{on}) applications. In figure 2.9, for examples, the device on-state resistance was tuned to extreme cases in which the programming current varied more than 6 orders of magnitude without any external components. The large and controlled parameter space offered by the a-Si based devices offers plenty of room for continued device optimization and suggests the devices, with the large range of performance parameters, can target a number of memory and logic applications.
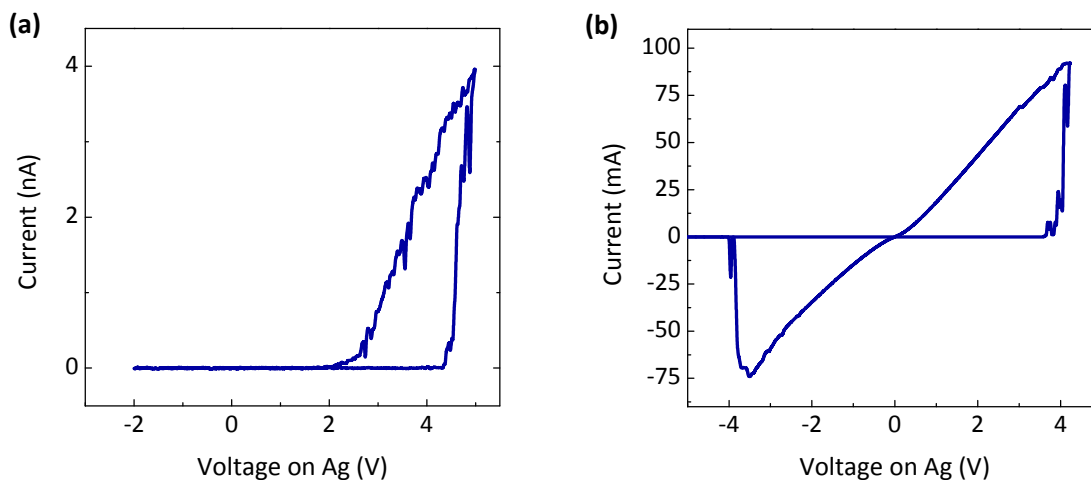
Figure 2.9. Two extremely tuned on state resistance (a) The amorphous silicon was deposited by PECVD at 260°C (c) by LPCVD at 585°C

2.7 Conclusion

In this study, we demonstrated that M/a-Si/p-Si based resistive-switching device structure is a promising candidate for ultrahigh density memory and logic applications. Such devices show excellent scalability, fast switching speed, long retention time and endurance. The elimination of the high voltage forming process allowed very high device yield compared with conventional M/a-Si/M memory devices. The fact that R_{off}/R_{on} ratio increases when the device size is decreased is very attractive for continued downscaling of the devices. The switching characteristic can be further controlled to be either rectifying or non-rectifying through both the fabrication process and circuit implementation. Especially, the intrinsic rectifying behavior is a desirable property in crossbar based devices, and eliminates crosstalk in array-level integration and may help the implementation of crossbar-based logic circuits.

References

[1]   W. de Boer, "Threshold switching in hydrogenated amorphous silicon," *Appl. Phys. Lett*. **40**, 812, 1982.

[2]   P. G. Lecomber, A. E. Owen, W. E. Spear, J. Hajto, A. J. Snell, W. K. Choi, M. J. Rose, S. Reynolds, "The switching mechanism in amorphous silicon junctions," S. *J. Non-crystal*. *Sol*. **77&78**, 1373, 1985.

[3]   A. E. Owen, P. G. Lecomber, J. Hajto, M. J. Rose, A. Snell, "Switching in amorphous devices," *Int. J. Electronics* **73**, 897, 1992.

[4]   M. Jafar, D. Haneman, "Switching in amorphous-silicon devices," *Phys. Rev. B* **49**, 13611, 1994.

[5]   A. Avila, R. Asomoza, R. "Switching in coplanar amorphous hydrogenated silicon devices," *Solid State Electronics* **44**, 17, 2000.

[6]   J. Hu, H. M. Branz, R. S. Crandall, S. Ward, Q. Wang, "Switching and filament formation in hot-wire CVD p-type a-Si:H devices," *Thin Solid Films* **430**, 249, 2003.

[7]   Y. Dong, G. Yu, M. C. McAlpine, W. Lu, C. M. Lieber, "Si/a-Si Core/Shell Nanowires as Nonvolatile Crossbar Switches," *Nano Letters* **8**, 386, 2008.

[8]   A. DeHon, "Array-based architecture for FET-based, nanoscale electronics," *IEEE Trans. on Nanotechnology* **2**, 23, 2003.

[9]   Y. Chen, G. Y. Jung, D. A. A. Ohlberg, X. M. Li, D. R. Stewart, J.O. Jeppesen, K. A. Nielsen, J. F. Stoddart, and R. S. Williams, "Nanoscale molecular-switch crossbar circuits," *Nanotechnology* **14**, 462, 2003.

[10]   J. C. Scott, "Is there an immortal memory?," *Science* **304**, 62, 2004.

[11]   S. Hudgens, B. Johnson, "Overview of phase change chalcogenide nonvolatile memory technology," *MRS. Bull.* **29**, 829, 2004.

[12]   K. Terabe, T. Hasegawa, T. Nakayama, M. Aono, "Quantized conductance atomic switch," *Nature* **433**, 47, 2005.

[13]   M. Kund, G. Beitel, C. -U. Pinnow, T. Rohr, J. Schumann, R. Symanczyk, K. Ufert, G. Muller, "Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20nm," IEDM Tech. Dig. 754, 2005.

[14]   A. E. Owen, P. G. Le Comber, W. E. Spear and J. Hajto, "Memory switching in amorphous silicon devices," *J. Non-Cryst. Solids* **59** & **60**, 1273, 1983.

Chapter 3

Switching Mechanism in Amorphous Silicon

In this chapter, we present results from an improved device structure in which a nanoscale a-Si pillar is used as the switching layer instead of a continuous a-Si film. The improved device structure allows us to carry out detailed studies on the switching mechanism. In particular, we show that in nanoscale two-terminal resistive switches the resistance switching can be dominated by the formation of a single conductive filament. The filament formation (switching process) is in addition stochastic in nature and can be modeled by a Poisson distribution so that the switching process can be controlled by adjusting both the amplitude and the time of the applied programming pulse. These observation leads to new approaches to program the devices. In addition, the improved devices exhibit excellent performance metrics and are well suited for memory or logic operations using conventional or emerging hybrid nano/CMOS architectures.

3.1 Device Structure and Fabrication

The active resistive switching element in this study consists of a nanoscale a-Si pillar embedded in an insulating dielectric, e.g. cured spin-on-glass (SOG) [24, 25]. Fig.

3.1 shows the schematic and top-view of a single-cell device prior to top metal deposition. A typical device is composed of a top Ag electrode, a B-doped poly-silicon (p-Si) bottom electrode and the nanoscale a-Si pillar forming a plug between the two electrodes. Cured spin-on-glass (SOG) is used as the spacer layer that isolates the top and bottom electrodes and provides mechanical support for the a-Si pillar. A single device essentially acts as a stand-alone reconfigurable interconnect or memory bit with its independently controlled top and bottom electrode pairs. Compared with our previous studies in which a continuous a-Si film was used [24], the pillar structure ensures that the active a-Si and filament region is physically restricted to nanoscale sizes. Replacing single crystalline Si [20, 21] with chemical-vapor deposition (CVD) grown poly-silicon as the bottom contact also enables device fabrication on a variety of substrates including existing CMOS structures and the potential for multi-layered 3D structure integration.
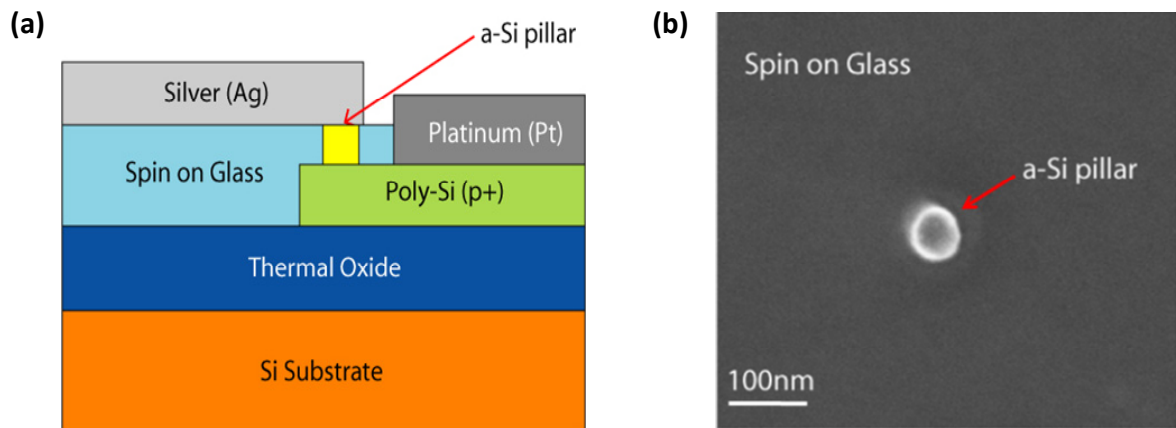
Figure 3.1. The device structure used in this study. (a) Schematic of the nanoscale Ag/a-Si/p-Si pillar structure. (b) Top view of an a-Si pillar embedded in SOG prior to Ag top electrode deposition.

The brief fabrication process for a-Si pillar resistance switch is as follows. The B-doped p-Si bottom electrode layer was deposited by LPCVD (low pressure chemical vapor deposition) on a prime grade silicon substrate with a 200 nm thermal dioxide. The amorphous silicon layer was deposited on top of the B-doped p-Si, followed by two RIE (reactive ion etching) steps to define the a-Si pillar and the p-Si bottom electrode structures. Spin-on-glass (SOG) was then spin-coated on the sample at speed of 3000 RPM and cured at 320 °C for 1 hour. The SOG layer was partially etched away to create a flat surface and expose the surface of the a-Si pillars. Ag electrodes to the a-Si pillars were patterned by e-beam lithography and lift-off. Ohmic contacts to the bottom p-Si layer were achieved by Pt electrodes. Special care was given in the pattern design so that the overlap between the top and bottom electrodes is minimized to keep the direct leakage current through the SOG is low.

Similar to devices based with a-Si films described in chapter 2, the pillar-based nanoscale two-terminal switches show a number of desirable performance metrics in terms of high yield (e.g. > 95% for devices with 60 nm diameter a-Si pillars), ON/OFF ratio (10^4-10^7), fast speed (< 50 ns programming time), and endurance (>10^5 cycles) (Fig. 3.2) that make them well-suited for high-performance memory and logic applications [13-19] based on conventional or hybrid nano/CMOS architectures. In addition, the high yield and high uniformity make it possible to perform detailed studies on the switching mechanism that will be discussed here.
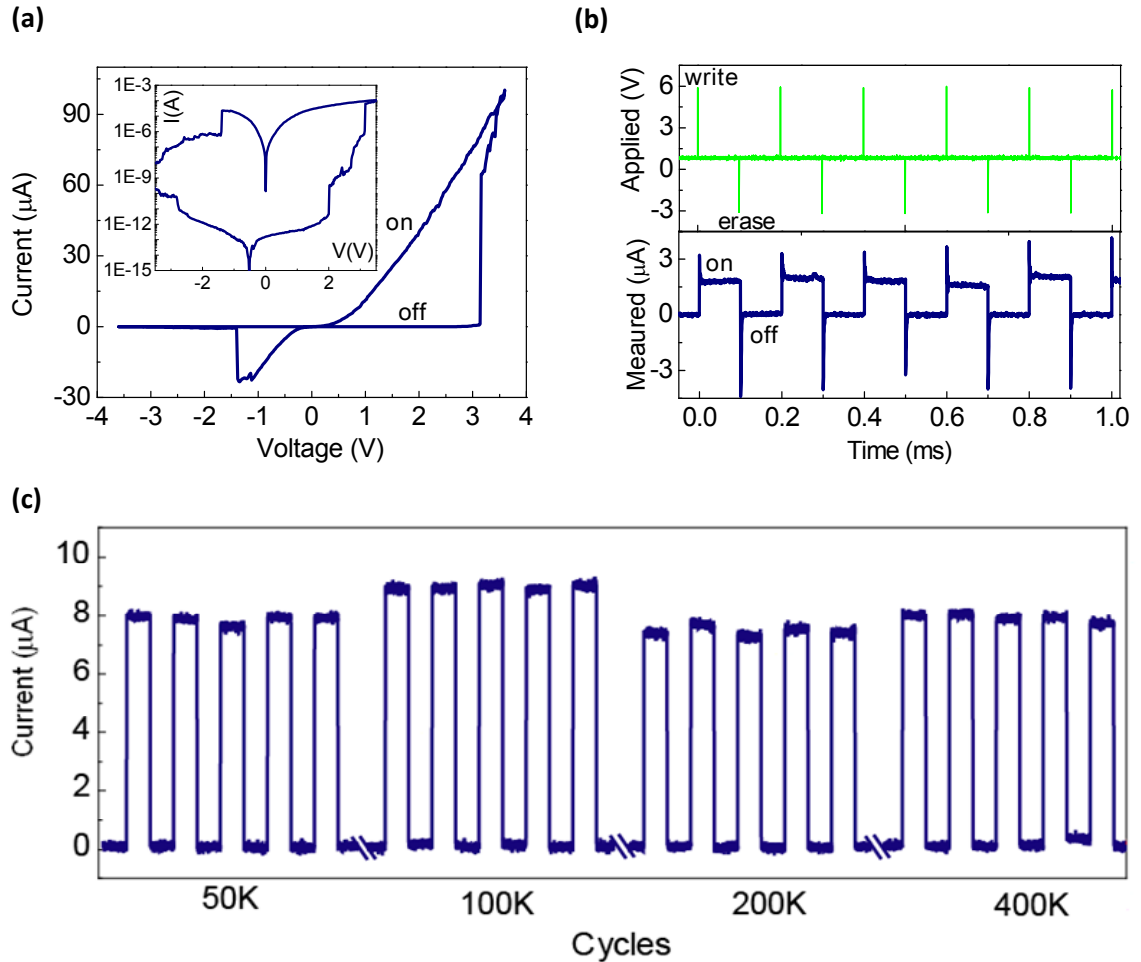
Figure 3.2. The basic device performance. (a) Resistance switching of a typical Ag/a-Si/p-Si pillar switch. Inset: I-V in log scale showing the stepwise transition during the turn-on process. The ON/OFF ratio is 10^7 at a typical read voltage ~1V. (b), A representative write/ read/erase/read pulse sequence with 6 V, 50 ns write, 0.8 V read and -3.2 V, 350 ns erase pulses and the output response from a typical device showing the device can be programmed by 50 ns write/erase pulses with high fidelity. (c) Endurance test of a device by applying repeated write/read/erase/read signals. A typical device with ON-resistance >50 kΩ can survive > 10^5 programming cycles without degradation. Beyond that the ON/OFF ratio starts to decrease due to increased OFF state conductance.

3.2 Resistance Switching Mechanism

In the following we focus our discussions on the programmable switching effects observed in these nanoscale pillar-based devices. Specifically, we show that resistance switching in these devices are dominated by the formation of individual filaments, and how the probabilistic (stochastic) nature of the filament formation process brings new challenges and opportunities to the device operation.

The switching mechanism in a-Si devices has been explained by the formation and retrieval of nanoscale Ag filaments upon the application of the programming voltage, schematically illustrated in figures. 3.3 and 3.8. Prior experimental and theoretical studies have suggested the filament to be in the form of a series of positively charged metal (Ag) particles trapped in defect sites in the a-Si layer [21, 26-28]. The conduction mechanism in the ON state is electron tunneling through the metal particle chain and the device resistance is determined by the tunneling resistance between the last metal particle and the bottom electrode.

In the case of nanoscale resistive switching devices studied here, switching may be dominated by a single chain of metal islands instead of a large number of filaments. This hypothesis is supported by the stepwise increase in current in log scale during the OFF-ON transition (figure 3.2 (a), inset), which is consistent with the fact that an Ag chain grows in a step-by-step fashion when an additional Ag particle hops into a new trapping site (figure 3.3).

off, V = 0

Ag

E_a

off $\rightarrow$ on, V > 0

V/2n

$E_a'\,(V) = E_a - V/2n$

on, V = 0

E_a

Ag $\longleftarrow$ a-Si $\longrightarrow$ p-Si
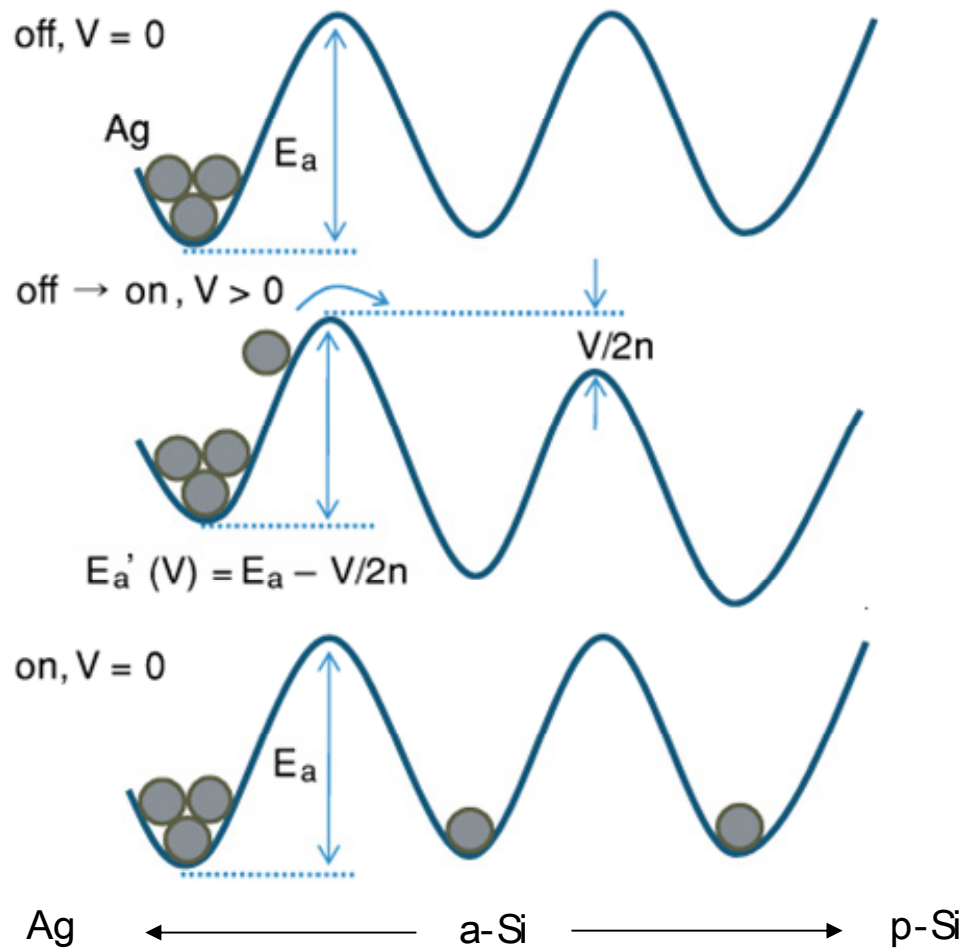
Figure 3.3. Schematics of the device in OFF, during the first transition at a finite bias, and in ON state, respectively.

A direct consequence of this filament formation model is that the switching rate will be strongly bias-dependent. Specifically, the hopping of the Ag particles is a thermally activated process (figure 3.3) and the rate Γ is determined by the bias-dependent activation energy E_a'

$$\Gamma = 1/\tau = \upsilon e^{-E_a'(V)/k_B T} \qquad (3.1)$$

where k_B is Boltzmann's constant, T is the absolute temperature, τ is the characteristic dwell time and ν is the attempt frequency. In addition, the activation energy may be lowered by the application of the bias voltage (figure 3.3), hence resulting in bias-dependent wait time and switching rates.

The ability to observe single-filament formation in the a-Si nano-pillar structure allows us to perform detailed studies to verify these hypothesizes. For example, we studied the wait time for the first transition (i.e., the first current step in figure 3.2 (a)) as a function of bias voltage. The wait time was measured by applying a square pulse with a given voltage magnitude to the device in OFF state and measuring the lapse in time t until the first sharp increase in current. The device was then erased by a negative voltage pulse and the measurement was repeated. Fig. 3.4 shows the histograms of the wait time for the first transition at bias voltages of 2.6 V, 3.2 V and 3.6 V on the same device. Because the stochastic nature of the switching process, the wait time is expected to follow a Poisson distribution [16] and the probability that a switching (the first step jump) occurs within Δt at time t is given by (see Appendix)

$$P(t) = \frac{\Delta t}{\tau} e^{-t/\tau} \qquad (3.2)$$

The histograms in Fig. 3.4 have been fitted to Eq. 3.2 using τ as the only fitting parameter and yielded τ values of 15.3 ms, 1.2 ms and 0.029 ms, respectively. It clearly verifies that τ is a strong function of V and decreases by almost 10^3 when V is increased by only 1 V. Figure 3.5 plots the distribution of the measured τ values at 5 different bias voltages along with a fit assuming exponential decay, treating τ_0 and V_0 as fitting parameters:

$$\tau(V) = \tau_0 e^{-V/V_0} \qquad (3.3)$$
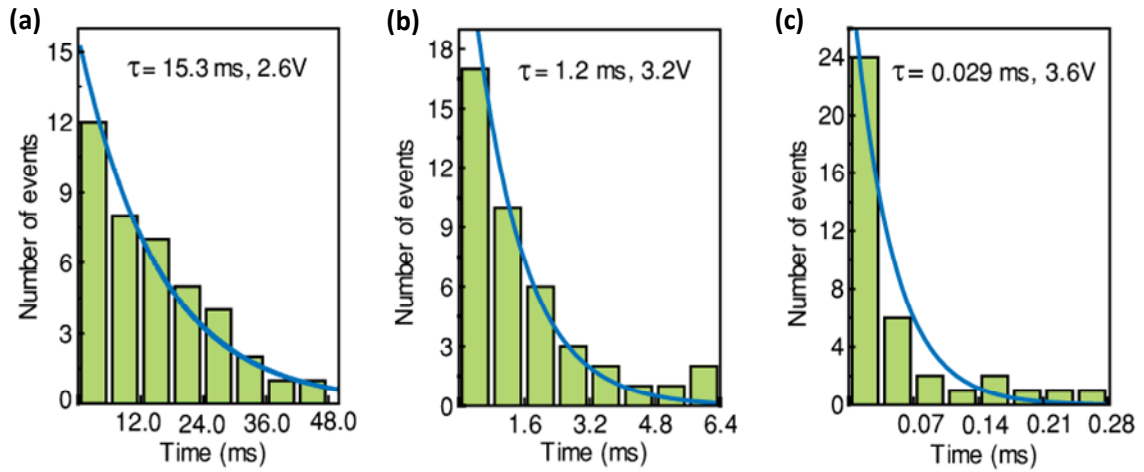
Figure 3.4. Histograms of the wait time for the first switching event at bias voltages of 2.6 V, 3.2 V and 3.6 V, respectively.

The good agreement with Eq. 3.3 clearly suggests that the wait time is strongly bias dependent, and more importantly, can be reduced *exponentially* by increasing the applied bias for high-speed memory operations. It is also interesting to note the physical origin of V_0 in Eq. 3. From figure 3.3 and to first order, $E_a' = E_a - Ed$, where E_a is the activation energy at zero bias, E is the electric field and d is the distance between the

Ag particle and the peak of the barrier. If we assume that most of the voltage is dropped across the Ag chain and the Ag particles are evenly distributed within the chain then to first order (figure 3.3) $E_a'(V) = E_a - V/2n$, where n is the number of the Ag^+ sites. Eq. 3.3 can then be directly derived from Eq. 3.1, with $\tau_0 = 1/\upsilon\ e^{E_a/k_BT}$ and $V_0 = 2nk_BT$. Significantly, the V_0 value of 0.155V inferred from the fitting in figure 3.5 is very close to that predicated by this simple model, $V_0 = 2nk_BT \approx 0.156V$, assuming there are 3 Ag sites in the filament ($n = 3$), as suggested by the number of major current steps in the semi-log I-V plot in figure 3.2 (a).
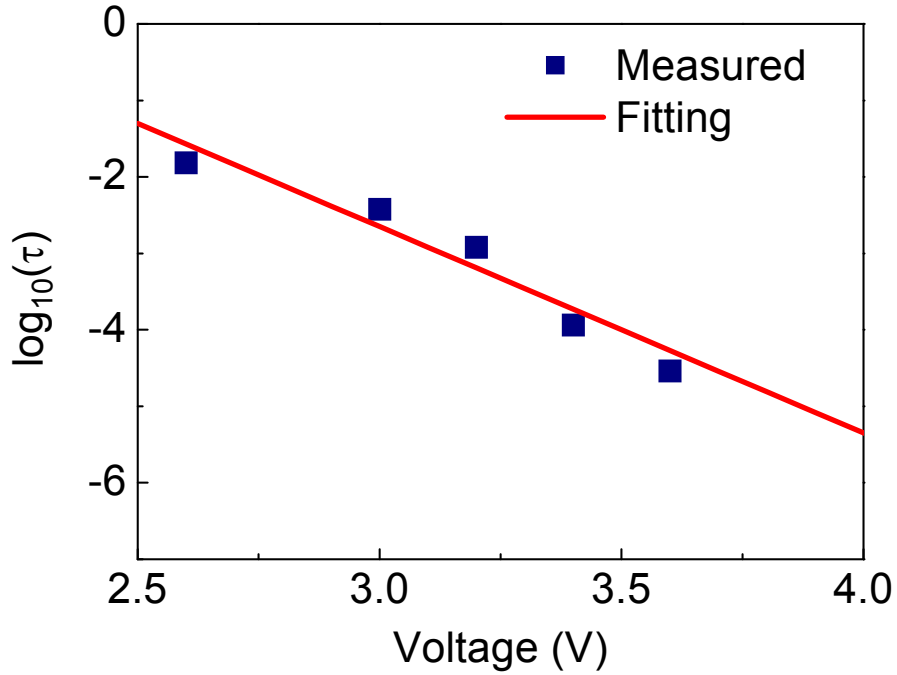
Figure 3.5. Characteristic wait time τ vs. bias voltage V along with an exponential fit according to Eq. (3).

The bias-dependent switching characteristics have important implications on the device operation. *First*, the switching essentially does not have a "hard" threshold voltage even though the switching can be very sharp (e.g. figure 3.2(a)), since there is always a finite probability for switching to occur even at relatively low bias voltages. On the other hand, threshold voltages can be defined for a given programming pulse width. For example, if the threshold is defined as the voltage above which 95% success rate is achieved, then the threshold voltage is 3.3 V for a 1 ms pulse, and 5.1 V for a 10 ns pulse width. This observation is consistent with data in Figure 3.2(b) where a 6 V programming voltage and 50 ns pulse results in ~100% success rate. *Second*, multi-level bit storage can be achieved in these devices by creating partially formed filaments through the control of external circuit parameters. For example, when a series-resistor is attached to the device, the voltage across the device will be reduced after the initial switching event, resulting in significantly longer wait time for subsequent switching events. As a result, partially formed filament can be created if the programming pulse is removed before the subsequent switching events can occur, resulting in intermediate resistance values between the fully programmed ON and OFF states. Figure 3.6(a) shows the final device resistance obtained on a single device using identical programming pulses but with different series resistor values. The $8 = 2^3$ different resistance levels obtained on the device suggest that each device as a memory component can store up to 3 bits of information. The device resistance R correlates well with the resistance R_S of the series resistor, since the voltage divider effect that causes the elongation of the wait time is most pronounced when the device resistance becomes comparable with R_S. For future applications in high-density memories or integrated circuits, the series of control resistors

can be replaced by a single diode whose resistance is a function of the applied bias, so that multi-bit programming of the switch can be achieved by controlling the amplitude of the programming signal instead.

We expect similar arguments may be applicable to other resistive switching devices since many of them involve some sort of activation energy process and involves filament formation, e.g. the diffusion of ions or the redox processes. The bias- and time-dependent switching in these nanoscale switches thus offers new challenges and opportunities in circuit and algorithm development as the device scales down to the level where the formation of a single filament starts to dominate. As an example, we show that the operation of the device as either a digital switch (e.g. a binary device) or an analog switch (e.g. a multi-value device) depends critically on the proper control of the amplitude as well as the duration time of the bias.

For the Poissonian processes discussed above, figure 3.6(b) plots the probability of exactly one switching event occurring during time t while figure 3.6(c) plots the probability of at least one switching event occurring during time t. They (figures 3.6 (b), (c)) correspond to the case with no external series resistance and a single switching rate $1/\tau$ applies to the step-wise filament formation process. It is clear then that the device acts as an excellent digital switch for long-enough programming pulses (e.g., 95% success rate is achieved for $t_{pulse} > 3\tau$). On the other hand, for multi-bit storage or analog operations of the switch, the pulse width has to be optimized. For example, t_{pulse} needs to be centered at τ to achieve the highest probability that only the first switching occurs. Even so the maximum success rate is only $\sim 38\%$. However, the success rate for multi-bit operations can be significantly improved by the addition of the external series resistance

since it reduces the voltage applied across the device after the first switching event and dramatically reduces the subsequent switching rates. In Figure 3.6(d) we plot the probability that only the first switching event will occur in a simplified two-step filament formation process in which two different rates are used (Appendix):

$$P(t) = \frac{\tau_2}{\tau_1 - \tau_2}\left(e^{-t/\tau_1} - e^{-t/\tau_2}\right) \qquad (3.4)$$

where $\tau_1 = 3.36$ μs and $\tau_2 = 1.30$ s corresponding to the switching rates when the voltages across the device are 4V (before the first switching event and R >> R_S) and 2V (after the first switching event and R = R_S) respectively, as a result of the voltage divider effect after the first switching event. A much higher success rate of > 99% can now be achieved for $5\tau_1 < t_{pulse} < 0.01\tau_2$ (corresponding to a 13 ms operation margin at 4 V bias) to limit the switching to the 1st event only.
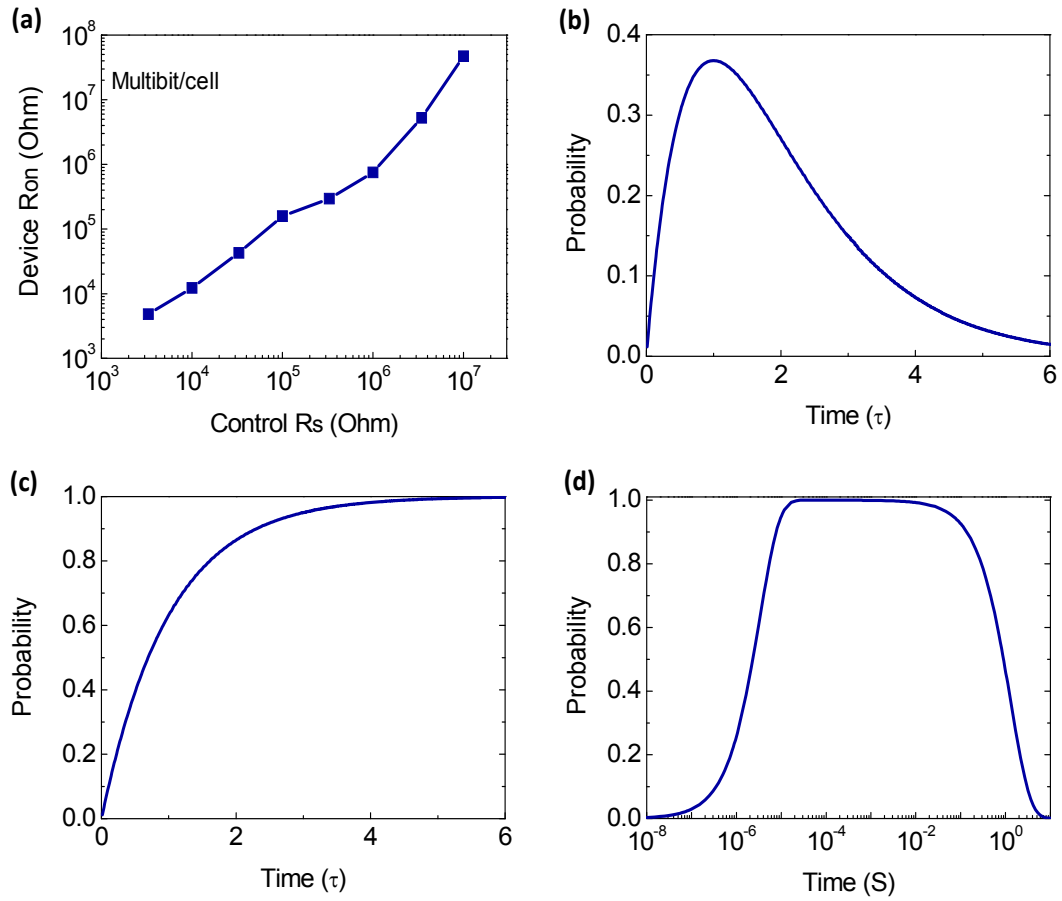
Figure 3.6. Multi-bit and time-dependent switching characteristics. (a) 8 levels of resistance can be obtained in a single switch when programmed with the same write pulse but different series resistance values. (b) The probability of having exactly one switching event. (c) The probability of having at least one switching event during time t without a series resistor. The device works well as a digital switch. (d) The probability of having exactly one switching event during time t with a series resistor $R_S = R$. 99% success rate can be obtained for $5\tau_1 < t_{\mathrm{pulse}} < 0.01\tau_2$.

3.3 Estimation of the Activation Energy

The activation energy of the barriers can be extracted from temperature dependence of the wait time from Eq. 3.1. Figure 3.7 (a) shows the time dependent resistance change at zero-bias at temperatures from 100 ℃ to 150 ℃ for a device originally programmed in the ON state. The sudden transitions to the OFF state

4 7

corresponds to the retrieval of the Ag filament by the thermally activated hopping of the Ag$^+$ particle towards the top electrode from the trapping site nearest to the bottom electrode (figure), as verified by the good fitting in the Arrhenius type plot of the wait time t versus $1/k_BT$ (figure 3.7 (b)). The activation energy for the ON/OFF transition can be extracted to be 0.87eV for this device from the slope of the Arrhenius plot and the retention time at room temperature is estimated to be 6.8 years from extrapolation.
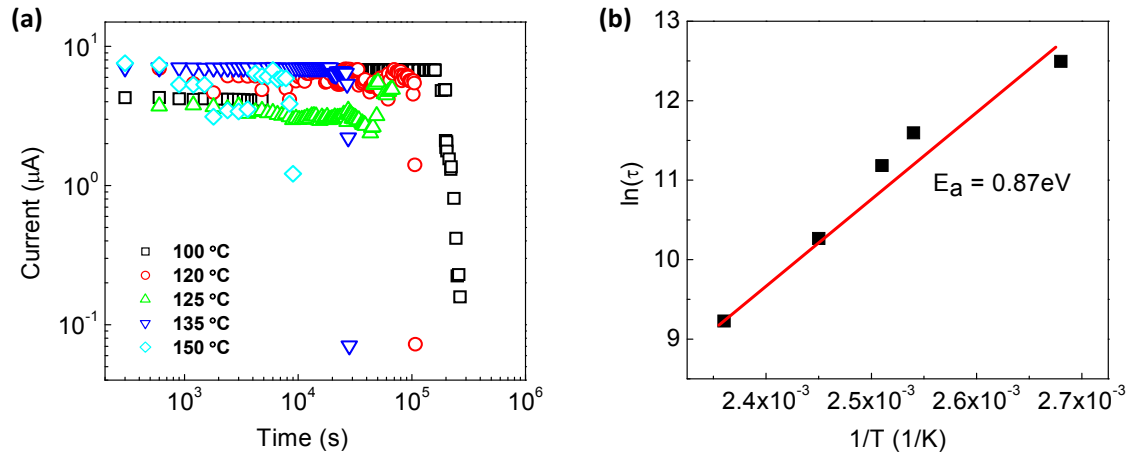
Figure 3.7. Activation energy and retention time measurement. (a) Resistance vs. time measured near zero-bias at temperatures from 100 ºC to 150 ºC plotted for a device originally programmed in the ON state. Elevated temperatures were used in this study to obtain the wait time within a reasonable time frame. (b) Arrhenius type plot of the wait time measured from (a) versus 1/T. Ea was estimated to be 0.87 eV using Eq. (1) and the retention time at room temperature was estimated to be 6.8 years

Transition metals such as Ag, Au, Cu and Pd are known to be fast diffusing impurities in silicon. For crystalline silicon, these metal impurities diffuse by an interstitial mechanism and an diffusion activation energy of ~ 1.15 eV was obtained [29]. In case of amorphous silicon, it was reported that Ag diffuses interstitially and through hydrogen vacancies with activation energies of 1.3 eV and 1.7 eV, respectively [30]. The observed activation

energy of 0.87 eV is roughly consistent but somewhat lower than these literature values. We expect further studies, such as the direct TEM imaging studies discussed in Section 8.7 may help explain the discrepancy by revealing the size and chemical nature of the Ag particles.

3.4 Estimation of Filament Length

The resistance at on-state is dominated by the tunnelling resistance between the last Ag particle and the bottom electrode. When the filament grew by a step length as a Ag particle hops into a new trapping site (figure 3.3), the resistance (current) decreases (increases) exponentially (figure 3.2(a)). The change in current can be calculated within the WKB approximation and is related to the increase in the filament length d:

$$\frac{J(\text{after jump})}{J(\text{before jump})} = e^{-(\sqrt{2m^*\Delta E}/\hbar)\cdot 2\Delta d} \tag{5}$$

where $J \propto T(\Delta E) \propto e^{-\frac{2}{\hbar}\int_0^d \sqrt{2m^*\Delta E}\,dx} \propto e^{-(2\sqrt{2m^*\Delta E}/\hbar)d}$ is the tunnelling current density, $T(\Delta E)$ is the tunnelling rate, ΔE is the barrier height seen by electrons, d is the distance between the last Ag particle and the bottom electrode and decreases by Δd after the jump. The barrier height, to the first order, can be estimated to be 0.31eV from work function difference of Ag and a-Si and $m^* = 0.09\ m_0$ for electrons in a-Si where m_0 is the free electron mass [34]. From Eq. 3.5, we estimated $\Delta d = 2.62$ nm, 2.65 nm and 2.72 nm for the 1st, 2nd and 3rd jumps in figure 3.2 (a), respectively. Hence, the total chain length is order of ~10 nm. It is worthwhile to note that the estimated chain length is shorter than the a-Si layer thickness of 30 nm, suggesting that a large portion of the conduction path

near the top electrode may consists of large volumes of Ag particles or multiple filaments (likely created during the initial forming process) while only the last ~ 10 nm close to the bottom electrode is dominated by a single chain of Ag particles (schematically illustrated in figure 3.8). This hypothesis was also supported by studies on devices with 60 nm a-Si layer, where similar programming voltages and R_{on} were observed and a similar filament length was estimated.
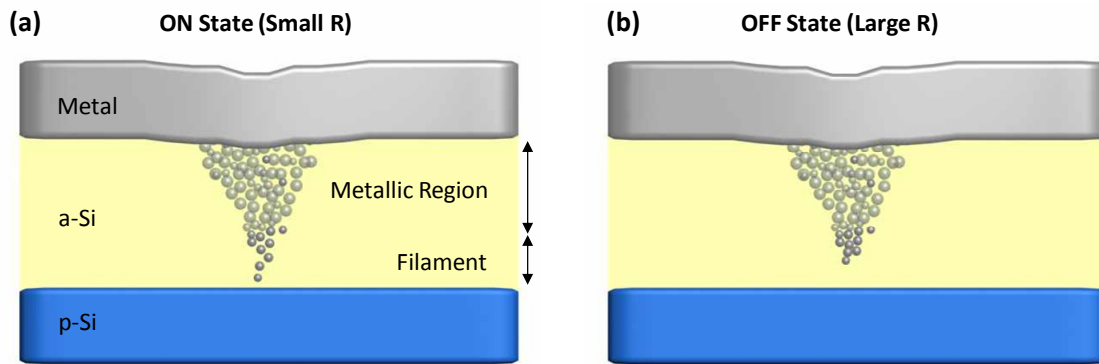
Figure 3.8. Schematic of the conduction path in the a-Si devices. The portion close to the top Ag electrode is composed of a large volume of Ag particles and the device characteristics are dominated by the filament in the ~ 10 nm layer close to the bottom electrode.

3.5 Conclusion

In summary, we have demonstrated a two-terminal reconfigurable switch based on a-Si nano-pillars. The ability to probe single filaments in the nanoscale devices allows us to study the amplitude- and time-dependent switching characteristics and furthermore

allows us to control the filament formation process. This controlled programming process suggests the devices may be used as ultra-high density crossbar memories or as reconfigurable interconnects (e.g. synapses in neuromorphic circuits). The device operation principles developed here may also be applicable to other filament based resistive switches.

Appendix

Spin-on Glass (SOG) Deposition

Spin-on Glass (SOG) is an interlevel dielectric material that is applied to a silicon wafer in liquid form. SOG can fill narrow trenches (< 100nm) and planarize the substrate surface. Hence, the SOG method is an alternative to silicon dioxide (SiO_2) deposited using CVD processes. Multiple SOG coatings can be applied to planarize rough surfaces. Figure A3.1 depicts comparison between CVD (Chemical Vapor Deposition) and SOG techniques. Unlike CVD, SOG can planarize the surface without other techniques such as CMP (chemical mechanical planarization), thus reducing process steps.
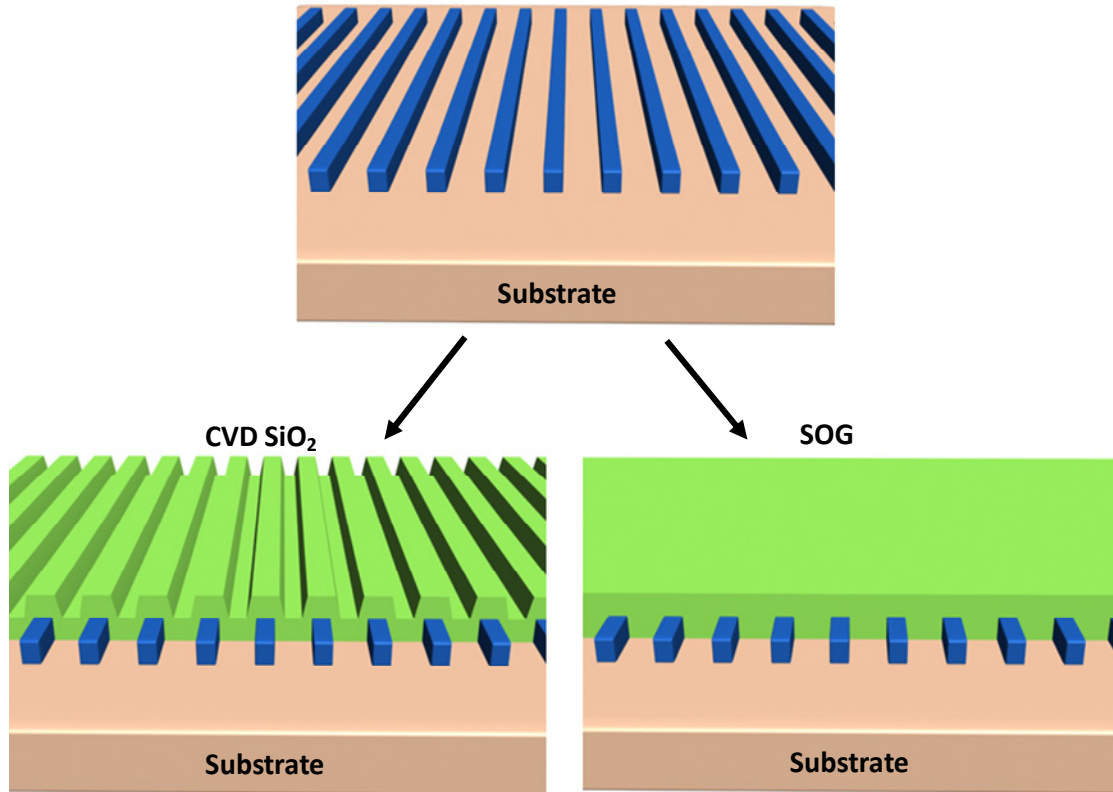
Figure A3.1. CVD deposited SiO_2 and spin-on glass

After SOG application to the substrate, proper baking steps are necessary to remove solvent and to stimulate chemical reaction. Chemically pure SiO_2 films with good electrical properties can be formed after the baking processes. Depending on the purpose, dopants such as boron and phosphorus can also be incorporated in liquid SOG. The film quality depends on the final bake temperature so that higher temperature yields better film quality. Typical bake temperature requires higher than 400°C and the cured SOG has a low dielectric constant and thus provides good electrical insulation. However, poor adhesion to metal, cracking and poisoning of metal are the main challenges in the SOG process. Nevertheless, the excellent gap-filling and planarization properties of SOG still make it an attractive process material. In general, there are two basic types of SOG:

siloxane-based organic SOG and silicate-based inorganic SOG. Baked silicate SOG gives a pure and stable SiO_2 film which does not absorb water. However, silicate SOG shrinks significantly during the bake process so that the film has large stress, limiting the SOG thickness to $<\sim$ 400nm. Adding phosphorus to silicate SOG reduces the film stress, but increases water absorption. Siloxane SOG yields carbon containing SiO_2 so that the film stress can be tailored by controlling carbon contents, enabling the use of thicker SOG for passivation purpose. However, polymer-like siloxane SOG limits the highest fabrication temperature of following processes to 500 °C. Figure A3.2 demonstrates the gap filling capability of SOG.

(a)

SOG

SiO$_2$

(b)

SOG

W

= 15.40 nm

SiO$_2$

Mag = 100.00 K X   200nm   EHT = 20.00 kV   Signal A = InLens
WD = 7 mm   Photo No. = 2389
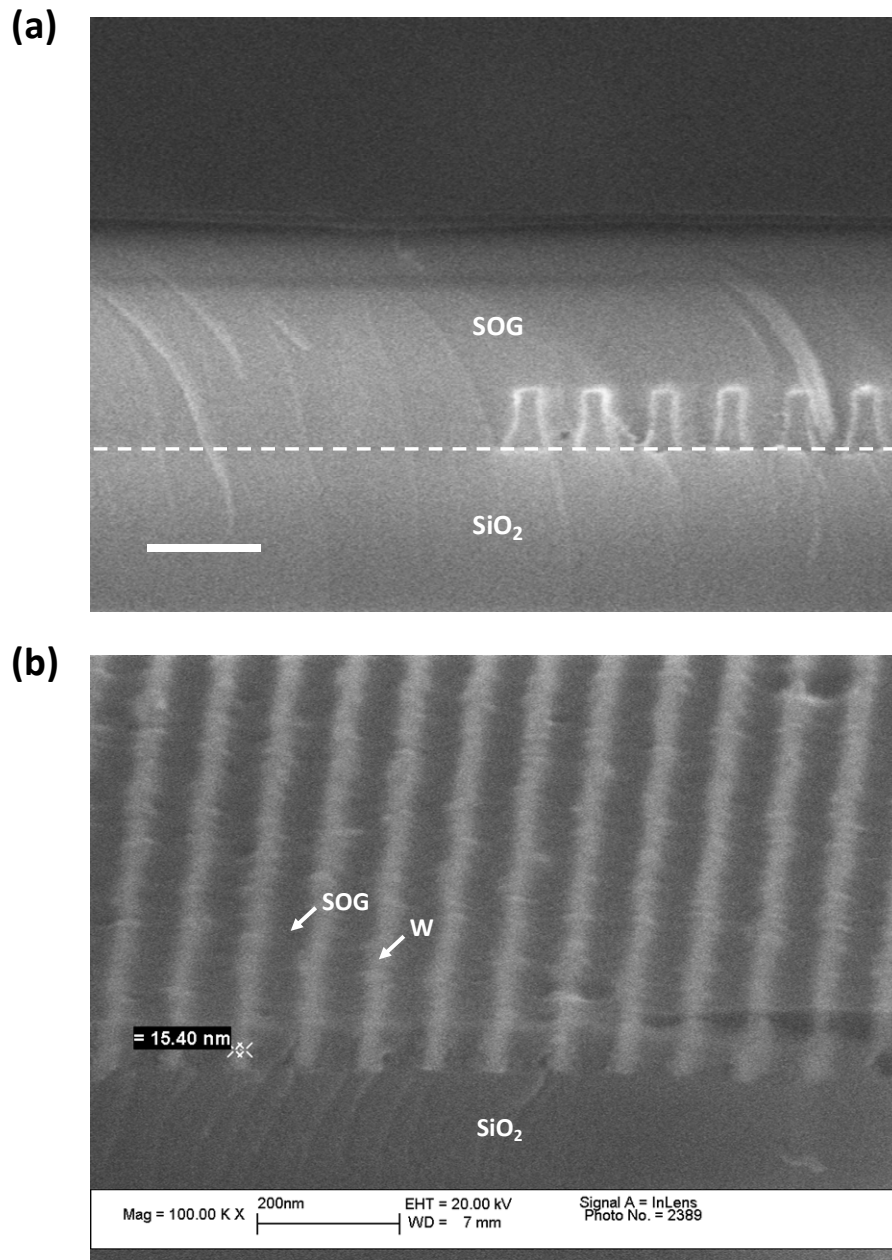
Figure A3.2 Scanning electron microscope (SEM) images of the SOG gap filling process. (a) SOG planarizes and fills gaps between nanostructures so that it forms a smooth surface across the interface between flat bottom structure and periodic nanostructures. Scale bar: 200nm. (b) Another example of SOG gap filling property. After SOG planarization, it was partially etched back to expose the tungsten nanowire electrodes.

Resistance Switching and Poisson Process

	Assumptions in Switching Processes	Assumptions in Poisson Distribution
1	The metal hopping rate is constant at fixed bias and temperature.	Events have a constant average rate.
2	Switching events occur independently, if the erase process produces identical starting conditions after each switching event.	Events occur independently in time since the last event.

Table 3.1. Assumptions used in deriving the expressions
for the resistance switching process.

The resistance switching in nanoscale a-Si based RRAM consists of several abrupt current jumps (Fig. 3-2(a)). Each current jump corresponds to the extension of the conducting filament caused by the hopping of additional metal inclusions. Due to the stochastic nature of hopping, the switching event will follow a Poisson distribution as implied in table 3-1. In a Poisson process, the probability that there are exactly k events within time t is expressed as

$$P(t,k) = e^{-t/\tau} \frac{(t/\tau)^k}{k!} \tag{A1}$$

where $1/\tau$ is the event occurrence rate. Hence, the probability that *the first resistance switching step* will occur within Δt at time t ($t_n - \Delta t < t < t_n$) is (Eq. (2))

$$P_{k=1}(t, \Delta t) = P(t = t_{n-1}, k = 0) \times P(t = \Delta t, k = 1)$$
$$= e^{-(t-\Delta t)/\tau} \times \frac{\Delta t}{\tau} e^{-\Delta t/\tau} = \frac{\Delta t}{\tau} e^{-t/\tau} \tag{A2}$$

which can be also shown by multiplying $P(t, k = 1) = (t/\tau)e^{-t/\tau}$ with $\Delta t/t$, since the events have a constant average rate.

The probability that only the first switching event (the first sudden current jump during resistance switching) will occur within time t can be derived in a simplified two-step filament formation process in which two different hopping rates are used (Eq. (4))

$$P(t) = \lim_{\Delta T \to 0} \sum_{T=0}^{T=t} (\Delta T/\tau_1) e^{-T/\tau_1} \times e^{-(t-T)/\tau_2}$$
$$= \int_0^t \frac{1}{\tau_1} e^{-T/\tau_1} \times e^{-(t-T)/\tau_2} dT = \frac{\tau_2}{\tau_1 - \tau_2} \left(e^{-t/\tau_1} - e^{-t/\tau_2} \right) \qquad (A3)$$

where τ_1 and τ_2 are rates for the first and second steps respectively.

Measurement Setup

A semiconductor parameter analyzer with a low noise pre-amplifier (Keithley 4200) was used for the DC *I-V* measurements to allow high resolution signal detection. For the pulse measurements a setup consisting of the switching device and a series resistor (e.g. 10 kΩ) was used. Write/read/erase/read pulse sequences were and I_{on} and I_{off} was monitored in situ by measuring the voltage drop across the series resistor. The programming pulse signals were generated by an arbitrary function generator (Tektronix AFG 3101) controlled via a Labview program.

References

[1]   International Technology Roadmap for Semiconductors 2007 ed.

[2]   H. Goronkin, Y. Yang, "High-performance emerging solid-state memory technologies," *MRS Bull.* **29**, 805-808, 2004.

[3]   R. Waser, M. Aono, "Nanoionics-based resistive switching memories,"*Nature Mater*. **6**, 833-840, 2007.

[4]   B. J. Choi, D. S. Jeong, S. K. Kim, C. Rohde, S. Choi *et al*., "Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition," *Appl. Phys.* **98**, 033715, 2005.

[5]   K. Terabe, T. Hasegawa, T. Nakayama, M. Aono, "Quantized conductance atomic switch," *Nature* **433**, 47-50, 2005.

[6]   D. C. Kim, S. Seo, S. E. Ahn, D.-S. Suh, M. J. Lee *et al*., "Electrical observations of filamentary conductions for the resistive memory switching in NiO films," *Appl. Phys. Lett.* **88**, 202102, 2006.

[7]   T-N. Fang, S. Kaza, S. Haddad, A. Chen, Y.-C. Wu *et al*., "Erase Mechanism for Copper Oxide Resistive Switching Memory Cells with Nickel Electrode," *IEDM Tech. Dig.* 789-792, 2006.

[8]   D. Lee, D. -J. Seong, I. Jo, F. Xiang, R. Dong *et al*., "Resistance switching of copper doped MoOx films for nonvolatile memory applications," *Appl. Phys. Lett.* **90**, 122104, 2007.

[9]   L. O. Chua, "Memristor - The Missing Circuit Element," *IEEE Trans. Circuit Theory* **18**, 507–519, 1971.

[10]   D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, "The missing memristor found," *Nature* **453**, 80-83, 2008.

[11]   J. J. Yang , M. D. Pickett , X. Li , D. A. A. Ohlberg , D. R. Stewart *et al*., "Memristive switching mechanism for metal/ oxide/ metal nanodevices," *Nature Nanotechnol*. **3**, 429-433, 2008.

[12]   J. R. Heath, P. J. Kuekes, G. S. Snider, R. S. Williams, "A Defect-Tolerant Computer Architecture: Opportunities for Nanotechnology," *Science* **280**, 1716-1721, 1998.

[13]   M. R. Stan, P. D. Franzon, S. C. Goldstein, J. C. Lach, M. M. Ziegler, "Molecular electronics: From devices and interconnect to circuits and architecture," *Proc. IEEE* **91**, 1940-1957, 2003.

[14]   G. S. Snider, R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnol.* **18**, 035204, 2007.

[15]   W. Lu, C. M. Lieber, "Nanoelectronics from the bottom up," *Nature Mater.* **6**, 841-850, 2007.

[16]   D. B. Strukov, K. K. Likharev, "Prospects for terabit-scale nanoelectronic memories," *Nanotechnol.* **16**, 137-148, 2005.

[17]   D. B. Strukov, K. K. Likharev, "CMOL FPGA: A. Reconfigurable. Architecture for Hybrid Digital. Circuits with Two-terminal Nanodevices," *Nanotechnol.* **16**, 888-900, 2005.

[18]   G. S. Snider, R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnol.* **18**, 035204, 2007.

[19]   A. DeHon, "Array-Based Architecture for FET-Based, Nanoscale Electronics," *IEEE Trans Nanotechnol.* **2**, 23-32, 2003.

[20]   S. H. Jo, W. Lu, "CMOS Compatible Nanoscale Nonvolatile Resistance Switching Memory," *Nano Lett.* **8**, 392-397, 2008.

[21]   Y. Dong, G. Yu, M. McAlpine, W. Lu, C. M. Lieber, "Si/a-Si Core/Shell Nanowires as Nonvolatile Crossbar Switches," *Nano Lett.* **8**, 386-391, 2008.

[22]   Z. J. Donhauser, B. A. Mantooth, K. F. Kelly, L. A. Bumm, J. D. Monnell *et al*., "Conductance Switching in Single Molecules Through Conformational Changes," *Science* **292**, 2303-2307, 2001.

[23]   M. A. Reed, J. Chen, A. M. Rawlett, D. W. Price, J. M. Tour, "Molecular random access memory cell," *Appl. Phys. Lett*. **78**, 3735-3737, 2001.

[24]   J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin *et al*., "A 160-kilobit molecular electronic memory patterned at 10^{11} bits per square centimetre," *Nature* **445**, 414-417, 2007.

[25]   H. -C. Liou, J. Pretzer, "Effect of curing temperature on the mechanical properties of hydrogen silsesquioxane thin films," *Thin Solid Films* **335**, 186-191, 1998.

[26] M. Jafar, D. Haneman, "Switching in amorphous-silicon devices," *Phys. Rev. B* **49**, 13611-13615, 1994.

[27] A. E. Owen, P. G. Lecomber, J. Hajto, M. J. Rose, A. J. Snell, "Switching in amorphous devices," *Int. J. Electron*, **73**, 897-906, 1992.

[28] J. Hu, H. M. Branz, R. S. Crandall, S. Ward, Q. Wang, "Switching and filament formation in hot-wire CVD p-type a-Si: H devices," *Thin Solid Films* **430**, 249-252, 2003.

[29] F. Rollert, N. A. Stolwijk, H. Mehrer, "Solubility, diffusion and thermodynamic properties of silver in silicon," *J. Phys. D: Appl. Phys.* **20**, 1148, 1987.

[30] M. S. Ablova, G. S. Kulikov, S. K. Persheyev, and K. K. Khodzhaev, *Sov. Phys. Semicond.* **24**, 1208, 1990.

[31] G. Grimmett, D. Stirzaker, *Probability and Random Processes* 2nd ed. (Oxford Univ. Press, Oxford, 1992).

[32] J. C. Scott, "Is There an Immortal Memory?" *Science* **304**, 62-63, 2004.

[33] H. B. Michaelson, "The work function of the elements and its periodicity," *J. Appl. Phys.* **48**, 4729–4733. 1977.

[34] J. M. Shannon, K. J. B. M. Nieuwesteeg, "Tunneling effective mass in hydrogenated amorphous silicon," *Appl. Phys. Lett.* **62**, 1815–1817, 1993.

Chapter 4

High-Density Crossbar Arrays Based on
Amorphous Silicon Two Terminal Switches

Integration of the a-Si based resistive switching devices into high-density crossbar arrays will be discussed in this chapter. In the crossbar array, a two-terminal resistive switch (memristive device) is formed at each crosspoint and can be addressed through the column and row nanowire electrodes. Relatively large scale (1kb) arrays have been demonstrated with high reliability and reproducibility. The crossbar arrays can be implemented as either random-access-memories or write-once type memories depending on the device configuration. The demonstration of large-scale crossbar arrays with well controlled device parameters also facilitates further studies on hybrid nano/CMOS systems.

4.1 Introduction

As discussed in Chapters 2 and 3, a hysteretic resistive switch device consists of a switching medium sandwiched between two electrodes and exhibits nonlinear *I-V* characteristics that the resistance of the device depends not only on the present voltage (or current) value but also on the history of the device programming (Figure 4.1). These types of hysteretic resistive switches fall in the category of broadly defined memristive system [1-3] which has recently been proposed as a promising candidate for future high-

density, high-performance memory or logic applications [1-6]. In particular, the simple two-terminal structure of memristive devices means that only one dimension (the distance between the two electrodes) needs to be critically controlled that allows for aggressive device scaling. Furthermore, a 2D array of such resistive switches can be readily implemented into the crossbar structure by overlaying two nanowire electrode arrays with 90° angle to each other so that a two-terminal switch is formed at each cross-point [4-8].
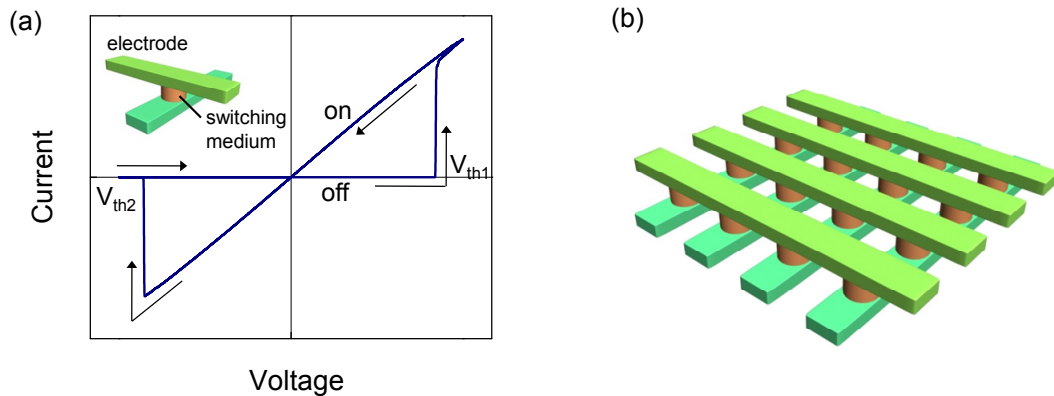
Figure 4.1. (a) A hysteretic resistive switch. Different resistances (ON or OFF) can be obtained at the same voltage depending on whether the programming voltage has passed the write threshold voltage (V_{th1}) or erase threshold voltage (V_{th2}) in the previous operation cycle. (b) Schematic of a crossbar array formed by the two terminal switches and nanowire electrodes.

The crossbar structure possesses many attractive features as it offers the highest possible device density and the simplest interconnect configuration that still allows random access to each nanometer scale device [4, 5, 7]. Since the crossbar arrays can be fabricated on top of a CMOS chip, hybrid nanocrossbar /CMOS architectures have been proposed in which the limited functionality of two-terminal switches can be compensated

by the CMOS components without creating any area overhead [8-10]. It has been shown that hybrid crossbar/CMOS memories using such approaches can offer terabit data storage density and sub-100 ns access time. In addition, hybrid logic circuits based on the crossbar/CMOS structures can offer function density of at least 2 orders of magnitude higher than that of their CMOS counterparts fabricated with the same design rules, at the same power density and comparable logic delay [8-10]. In this chapter, we will demonstrate our work on large-scale (1kb), high-density crossbar arrays based on the a-Si memristive system and show that the nanocrossbar arrays exhibit excellent yield, ON/OFF and uniformity. This demonstration, along with excellent performance achieved at the single-cell level discussed Chapters 2 and 3 in terms of write speed (<10 ns), endurance (>10^5 cycles), retention (~7 years), and scaling potential (<30 nm), suggests that the a-Si memristive system is well positioned to implement the proposed hybrid crossbar/CMOS systems for ultrahigh performance memory and logic applications.

4.2 Device Structure and Fabrication

The crossbar array in our study consists of a parallel array of boron-doped poly silicon nanowires serving as the bottom electrodes, the a-Si active layer, and a parallel array of Ag nanowires serving as the top electrodes as shown in figure 4.2 (a). Spin-on-glass (SOG) was used as a dielectric to isolate the top and bottom electrode arrays outside the cross-points.

The fabrication starts with the deposition of boron doped polycrystalline silicon (p-Si) and amorphous silicon (a-Si) on the thermally oxidized silicon substrate (figure 4.2

(b)). The a-Si and p-Si layers are patterned by lithography (electron beam for the nanometer scale lines and UV photolithography for the larger contact pads) and mask material lift-off processes followed by reactive ion etching (RIE) to form bottom electrodes and the isolated a-Si switching layer (figure 4.2 (c)). To make Ohmic contacts to the p-Si nanowire electrodes, the a-Si layer at the end of each nanowire was etched, followed by Pt metal deposition (figure 4.2 (d)). The device then went through SOG (spin-on-glass) coating, thermal curing and planarization processes. Ellipsometry was used to monitor the thickness of the SOG during the partial etch such that the thickness was controlled to ±10Å. The Ag top nanowire electrodes were patterned by e-beam lithography and lift-off (figure 4.2 (e)). The contact pad patterns were fabricated by a photolithography process and configured to fit a custom made probe card for 1kb array measurement. Essentially, each crosspoint in the array forms a two-terminal Ag/a-Si/p-Si switch discussed earlier. Scanning electron microscope (SEM) images of a 1 kb (32 × 32) crossbar array with density of 2 Gbits/cm^2 (120nm line width) are shown in figure 4.2 (f).
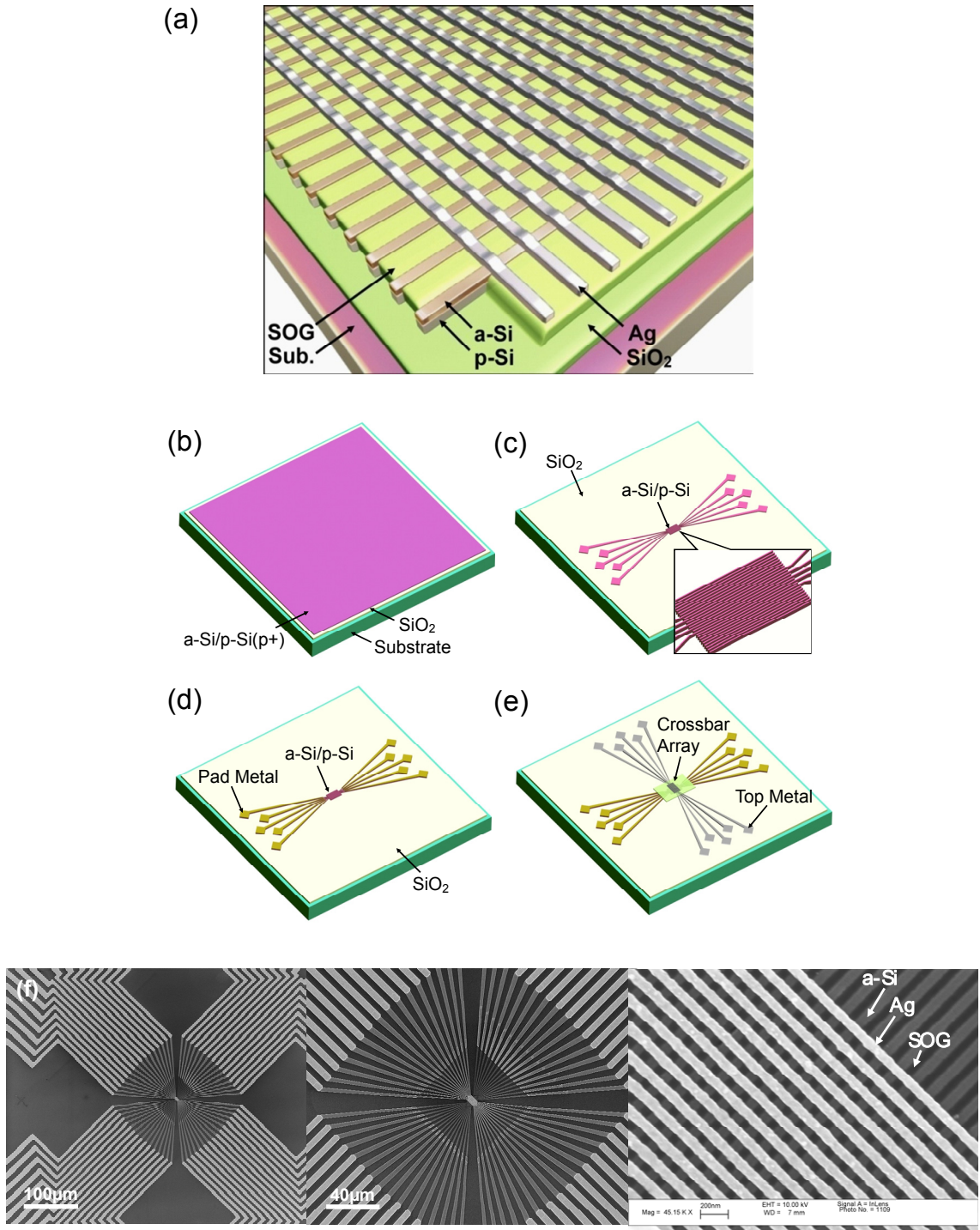
Figure 4.2. Schematic of the fabricated device (a), brief fabrication flow (b-e) and SEM images of a 1kb crossbar array.

4.3 Switching Data in 1kb Crossbar Arrays

Compared with memristive devices based on molecules or metal oxides, the solid-state a-Si used in our system allows robust fabrication and operation. For example, the fabrication yield of the 1 kb crossbar array is higher than 95% and each bit inside the 1 kb array can be addressed automatically with high fidelity using a group of preset write/erase/read programming pulses without having to adjust the programming signals manually or knowing the state of the device. Figure 4.3 (a) shows the yield map of 400 bits in the array tested using the automated program. The bits were written with 200 μS pulses at 5.5 V. The device parametric yield (defined as the percentage of devices showing ON/OFF ratio $>10^3$ after the write/erase pulses) was 93% with most (80%) devices showing ON resistance in the range of 50-150 kΩ. Figure 4.3 (c) shows I_{ON} and I_{OFF} values obtained from the first 400 bits within the array, once again demonstrating the excellent uniformity of the still not fully optimized crossbar array.
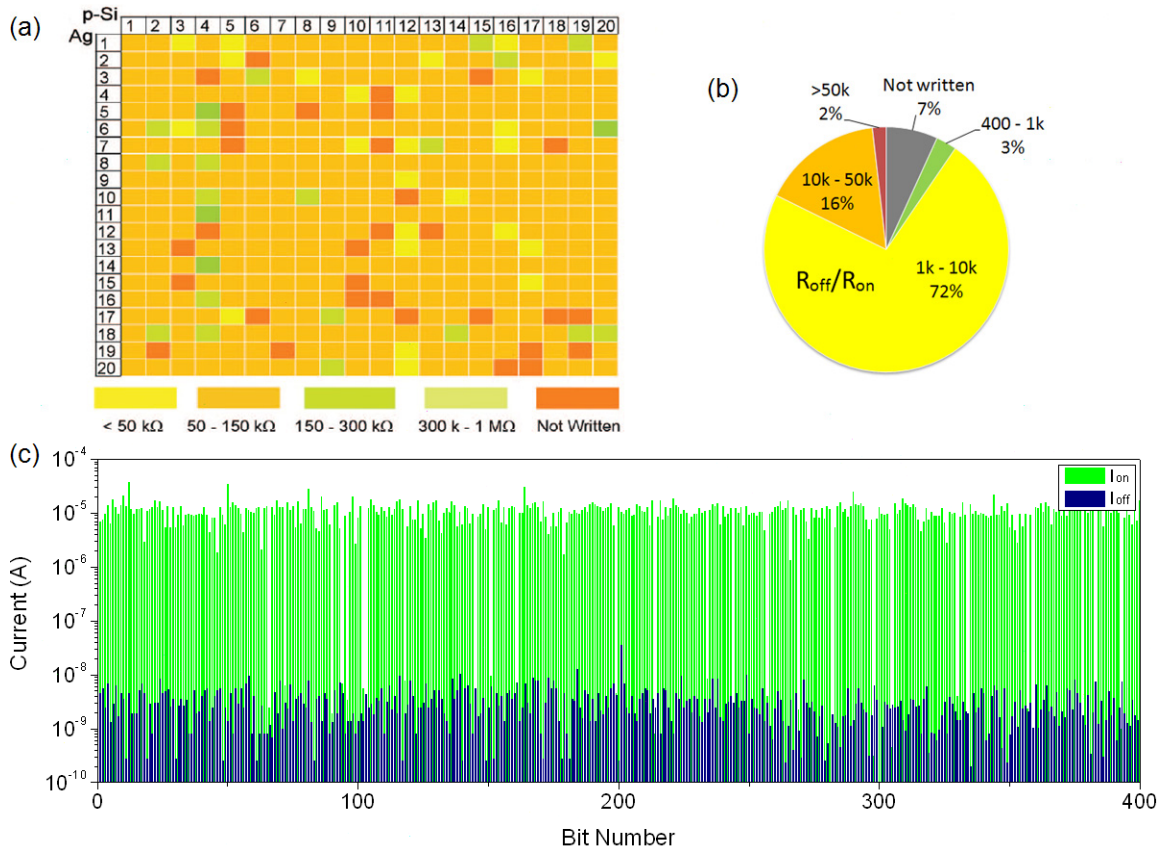
Figure 4.3. Switching characteristics of the first 400 bits within a 1kb crossbar array. Uniform write pulse of 200 μS 5.5V was used. Most of the on-state resistance of the written cells is 50 – 150 kohm and on/off ratio is larger than 10^3 in more than 90% of devices.

A relatively large line width of 120 nm was used in the 1kb array for the bottom p-Si nanowire electrodes to minimize series-resistance associated with the relatively resistive p-Si nanowire electrodes. Even so the resistance associated with the p-Si nanowire electrodes can be up to 30 kΩ that limits fully automated operations of the array. Future improvements will involve the addition of a metal or a silicide layer adjacent to the p-Si layer to reduce the series resistance. The incorporation of the metal layer will in turn allow the use of narrower nanowire electrodes hence resulting in even higher bit density. Cell size smaller than 50 nm × 50 nm has already been demonstrated at the

single-cell level corresponding to a density of 10 Gb/cm^2 limited only by the available lithography technique [16,17]. To demonstrate the feasibility of the crossbar array as high density memory for information storage, a smaller array (8×8) was tested to mitigate the series resistance effect. Figure 4.4 shows that a word "CrossBar" can be stored and retrieved from the 64 bits of 8×8 the crossbar using a fully automated write/read program, where each letter in the word "CrossBar" is represented by an 8-bit ASCII character and written into a single row inside the array.
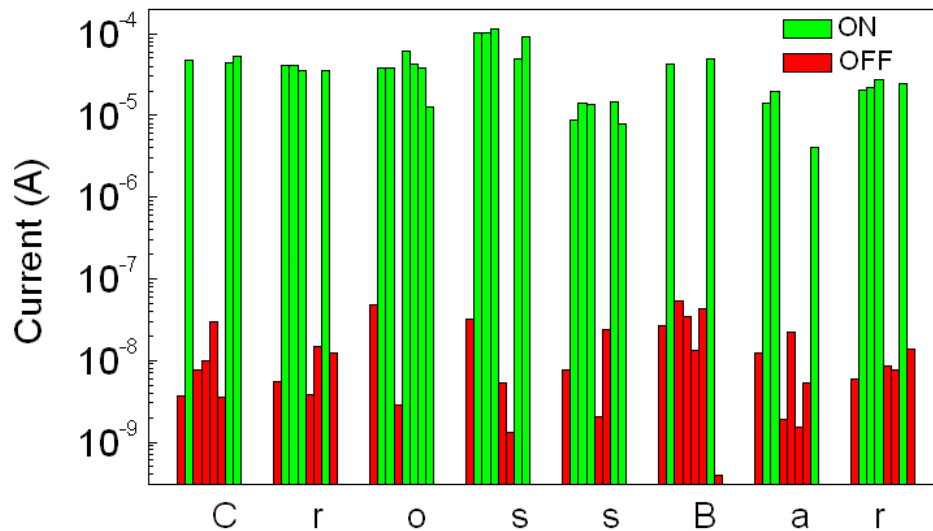
Figure 4.4. A word "CrossBar" stored in a 8×8 array. High on/off ratio allows clearly distinguishable two (on and off) states.
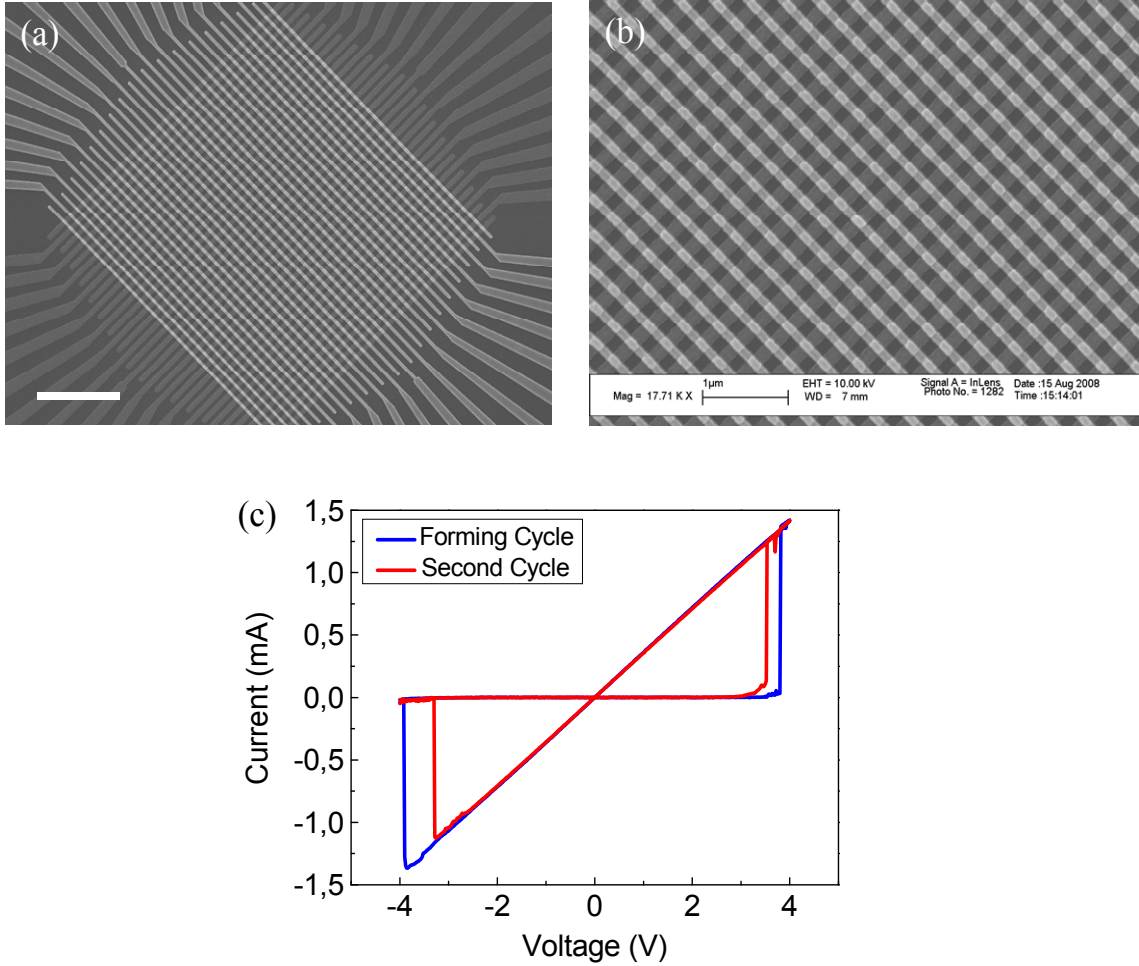
Figure 4.5. The fabricated 1kb Ag/a-Si/Ni crossbar array and its switching characteristic. (a) The SEM image of the device. Scale bar: 2μm. (b) A zoomed-in SEM image presenting RRAM cells formed at each cross-point of top and bottom electrodes. (c) IV characteristic of the device showing virgin and second resistance switching.

Crossbar arrays without the p-Si layer have also been studied to verify the role of the p-Si electrode played in the switching process. Figure 4.5 (a-b) shows the SEM image of a 1kb crossbar memory composed with Ag/a-Si/Ni cross-points. Unlike the Ag/a-Si/p-Si structures in which the ON-resistance can be adjusted by tuning the a-Si growth parameters (e.g., R_{ON} can be controllably varied from the order of 100 MΩ to 10 kΩ by adjusting the deposition temperature) all Ag/a-Si/Ni devices tested have shown low R_{ON}

on the order of 1 kΩ and high programming currents on the order of 1 mA hence undesirably large write energy regardless of the a-Si deposition conditions. In addition, the endurance of the Ag/a-Si/Ni devices is typically only a few hundred cycles, much less than the $>10^5$ cycles obtained on the Ag/a-Si/p-Si devices [16, 17], possibly due to the high stress to the material asserted by the high programming current. We note however that the Ag/a-Si/Ni devices may be suitable as read-only memories (ROMs) or write-once read-many memories since they show excellent retention after the initial programming process.

The different behaviors in the Ag/a-Si/p-Si devices and Ag/a-Si/Ni devices can be qualitatively understood by examining the mechanism of the resistance switching process. As discussed in chapters 2-3, resistance switching in the a-Si devices is explained by the formation of conductive filaments inside the a-Si matrix. The filament has been suggested to be in the form of a series of Ag particles [18-21] (Figure 3.8 in chapter 3) and the resistance in ON state is dominated by the tunneling resistance between the last Ag particle in the filament and the bottom electrode. We postulate that the larger number of defects at the a-Si/metal interface will cause either a larger number of filaments to be formed or more closely spaced Ag particles with the filaments, both of which will lead to a large programming current. This hypothesis is also consistent with the fact that current compliance was needed during the forming process for metal/a-Si/metal devices to prevent device damage due to excess current [19, 20].

4.4 3D Electric Field Simulation of Crossbar Structure

To facilitate the integration and programming of high density crossbar arrays, interference between cells during write and erase processes caused by fringing electric fields need to be minimized. We have performed 3D electric field simulation by using Synopsys 3D device simulator (Davinci V. X-2005.10) to study how the array geometry affects the fringe fields. During the simulation, the value of the barrier height between the top metal and a-Si was adopted from reference for bulk materials [22], in which the barrier height of various metals over undoped discharge-produced amorphous Si was reported.

The simulated device structure is a 50nm half-pitch 8×8 crossbar array with 40nm or 80nm thick a-Si as illustrated in figure 4.6 (a). The protocol to program a cell in the crossbar is as follows: to program the target cell formed by row i and column j in the array, a programming voltage $V > V_{th1}$ will be applied on row i, column j will be grounded ($V = 0$), while all the other rows and columns in the array will be applied with a protective voltage of $V/2 < V_{th1}$ to avoid accidental programming of the unselected cells. The erasing and reading protocols are similar except by replacing V_{th1} with V_{th2} for erasing, and by applying V_{read} to the target row i and $V = 0$ to all other rows and columns for reading. Figures 4.6 (b-e) show the electric field distribution across the selected and half-selected junctions and on the plane near the bottom p-Si electrodes during the programming process. As presented in the figure 4.6, the thinner switching layer (a-Si) allows a larger operation margin which inhibits accidental write/erase, due to enhanced coupling between top and bottom electrodes. Hence, the thickness of a-Si should be scaled down as the device pitch size reduces to ensure proper device operations.
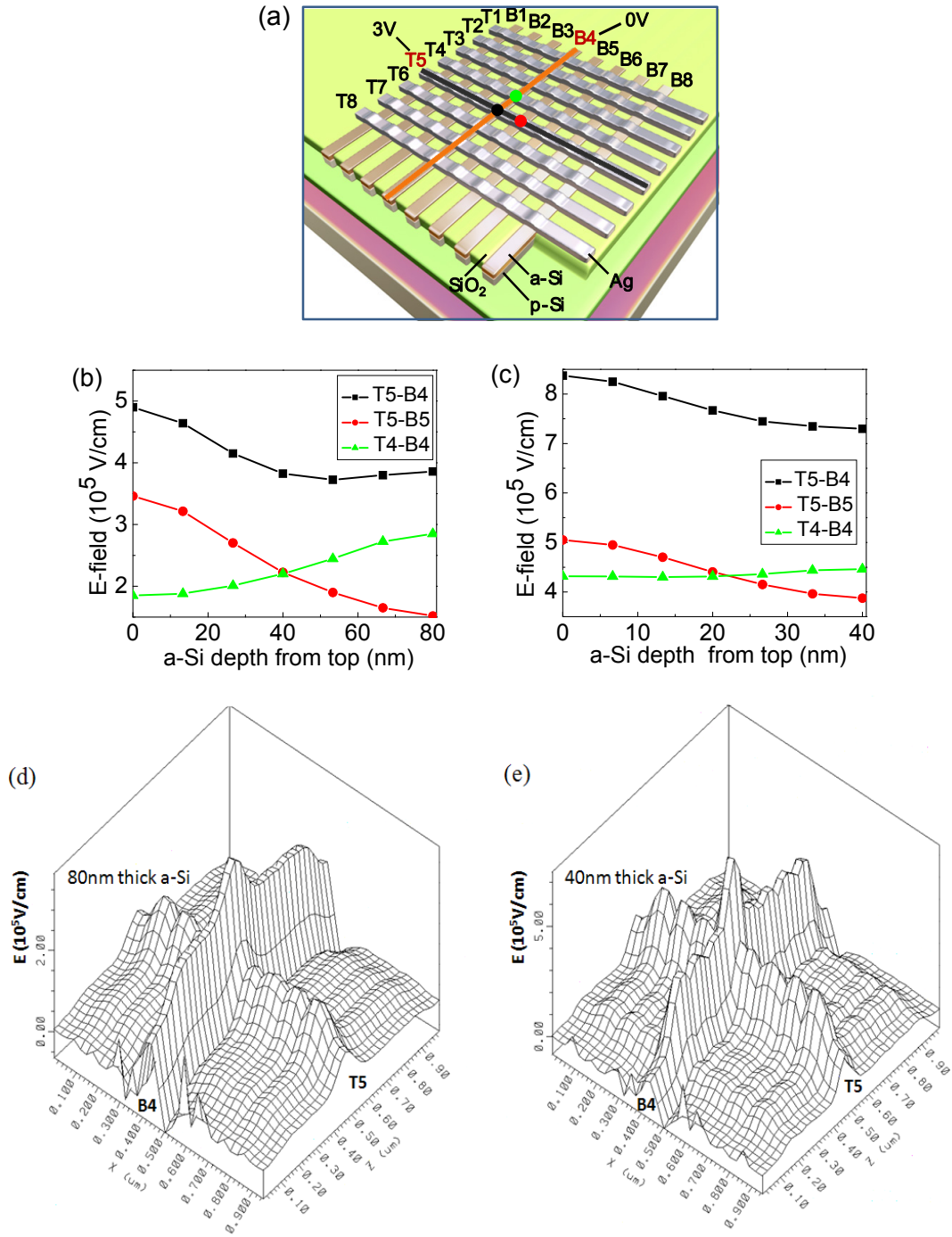
Figure 4.6. 3D electric field simulation. (a) Simulated device structure. 3 V and 0 were applied to a top (T5) silver row electrode and a bottom (B4) p-Si column electrode and all the other electrodes were applied with 1.5 V. (b) E-field distribution across the selected and half-selected junctions within crossbar structure with 80nm thick a-Si and (c) with 40nm thick a-Si. (d) E-field distribution on the plane 5nm above the bottom p-Si at the bias confirmation shown in (a). (e) 2nm above the bottom p-Si.

7 1

Considering the experimental result (chapter 3) that the switching occurs within a 10nm range in a-Si, the a-Si based RRAM would be well positioned for ultra-high density crossbar structure.

4.5 Conclusion

In summary, we demonstrated high density crossbar arrays using a nanoscale a-Si-based memristive system. The excellent yield and performance illustrated by this CMOS compatible approach opens the door for further development and testing of novel electrical circuits based on memristive systems. For example, our recent studies on time- and bias dependent switching characteristics suggest that these devices may be suitable as synapses in solid-state neuromorphic circuits. Additional improvements of the devices may include incorporating metallic electrodes adjacent to the p-Si nanowire electrodes to mitigate series-resistance problem, and incorporating additional current steering elements at each crosspoint to mitigate crosstalk problem. Looking into the future, we expect relatively smooth integration of the a-Si-based crossbar arrays with CMOS components that can lead to a number of high performance hybrid crossbar/CMOS memory and logic systems.

References

[1]   L. O. Chua "Memristor-The Missing Circuit Element," *IEEE Trans. on Circuit Theory* **18**, 507-519, 1971.

[2]   L. O. Chua, S. M. Kang, "Memristive devices and systems," *Proc. IEEE* **64**, 209–223, 1976.

[3]   D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, "The missing memristor found," *Nature 453*, 80–83, 2008.

[4]   J. R. Heath, P. J. Kuekes, G. S. Snider, R. S. Williams, "A defect-tolerant computer architecture: opportunities for nanotechnology," *Science 280*, 1716–1721, 1998.

[5]   K. K. Likharev, "Electronics Below 10 nm," In *Nano and Giga Challenges in Microelectronics*; Greer, J., Korkin, A.;, Labanowski, J. Elsevier: Amsterdam, 27-68, 2003.

[6]   W. Lu, C. M. Lieber, "Nanoelectronics from the bottom up," *Nat. Mater. 6*, 841–850, 2007.

[7]   A. DeHon, "Array-based architecture for FET-based, nanoscale electronics," *IEEE Trans. Nanotechnol. 2*, 23–32, 2003.

[8]   G. S. Snider, R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnology 18*, 035204, 2007.

[9]   D. B. Strukov, K. K. Likharev, "Prospects for terabit-scale nanoelectronic memories," *Nanotechnology 16*, 137–148, 2005.

[10]   D. B. Strukov, K. K. Likharev, "CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices," *Nanotechnology 16*, 888–900, 2005.

[11]   Y. Chen, G. Y. Jung, D. A. Ohlberg, X. Li, D. R. Stewart, J. O. Jeppesen, K. A. Nielsen, J. F. Stoddart, R. S. Williams, "Nanoscale molecular-switch crossbar circuits," *Nanotechnology 14*, 462–468, 2003.

[12]   H. Goronkin, Y. Yang, "High-Performance Emerging Solid-State Memory Technologies," *MRS Bull. 29, 805-808, 2004.*

[13]   J. J. Yang, M. D. Pickett, X. Li, D. A. Ohlberg, D. R. Stewart, R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nat. Nanotechnol. 3*, 429–433, 2008.

[14]   R. Waser, M. Aono, "Nanoionics-based resistive switching memories," *Nat. Mater. 6*, 833–840, 2007.

[15] J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich , E. Johnston-Halperin1*et al.*, "A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimetre," *Nature 445*, 414–417, 2007.

[16] S. H. Jo, K. -H. Kim, W. Lu, "Programmable Resistance Switching in Nanoscale Two-Terminal Devices," *Nano Lett. 9*, 496-500, 2009.

[17] S. H. Jo, W. Lu, "CMOS Compatible Nanoscale Nonvolatile Resistance Switching Memory," *Nano Lett. 8*, 392–397, 2008.

[18] Y. Dong, G. Yu, M. C. McAlpine, W. Lu, Lieber, C. M. "Si/a-Si Core/Shell Nanowires as Nonvolatile Crossbar Switches," *Nano Lett. 8*, 386–391, 2008.

[19] A. E. Owen, P. G. Lecomber, J. Hajto, M. J. Rose, A. J. Snell, "Switching in amorphous devices," *Int. J. Electron. 73*, 897–906, 1992.

[20] M. Jafar, D. Haneman, "Switching in amorphous-silicon devices," *Phys. Rev. B 49*, 13611–13615, 1994.

[21] J. Hu, H. M. Branz, R. S. Crandall, S. Ward, Q. Wang, "Switching and filament formation in hot-wire CVD p-type a-Si:H devices," *Thin Solid Films 430*, 249–252. 2003.

[22] C. R. Wronski, D. E. Carlson, "Surface states and barrier heights of metal-amorphous silicon schottky barriers," *Solid State Commun. 23*, 421, 1977.

Chapter 5

Memristor Synapses for Neuromorphic Systems

5.1 Introduction

In this chapter we discuss how the two-terminal resistive switches may be used in applications such as bio-inspired computing systems. In some sense, modern electronics have already approached the biological systems such as the human cortex in terms of device density and complexity. For examples, the state-of-the art CMOS density can produce similar to neuron density of $10^6 \sim 10^7/cm^2$ in human brain and digital computers can now possess the computing speed and complexity to emulate the brain functionality of animals [1-4]. However, the biological system (e.g. human brain) is still much more efficient by a factor of at least a million than today's intelligent machines in complex environments. For example, to perform certain cortical simulations at the cat scale even at 83 times slower firing rate, the IBM team in Ref. 2 has to employ Blue Gene/P (BG/P), a super computer equipped with 147,456 CPUs and 144 TB of main memory. On the other hand, brains of biological creatures are configured dramatically differently from the von Neumann digital architecture. Key to the high efficiency of biological systems is the large connectivity ($\sim 10^4$ in a mammalian cortex) between neurons that allows efficient parallel analog signal processing [5]. The synaptic weight between two neurons can be precisely adjusted by the ionic flow through them and it is widely believed that the

adaptation of synaptic weights enables the biological systems to learn and function [1, 4, 6-9].

A synapse is essentially a two-terminal device and bears striking resemblance to the memristor [10-11] (memory + resistor). In particular, the high connectivity offered by the crossbar suggests it is possible to perform direct hardware emulation of biological systems with comparable configurable and functional density. As shown in Fig. 5.1, a two-terminal memristor synapse is formed at each crosspoint in the crossbar array and connects CMOS-based pre- and post-synaptic neurons. In this case, every CMOS neuron in the "pre-neuron" layer of the crossbar configuration is directly connected to every neuron in "post-neuron" layer with unique synaptic weights. A high synaptic density of $10^{10}/cm^2$ can also be potentially obtained for crossbar networks with 100 nm pitch, a feature size readily achievable with advanced lithography approaches [12-13].
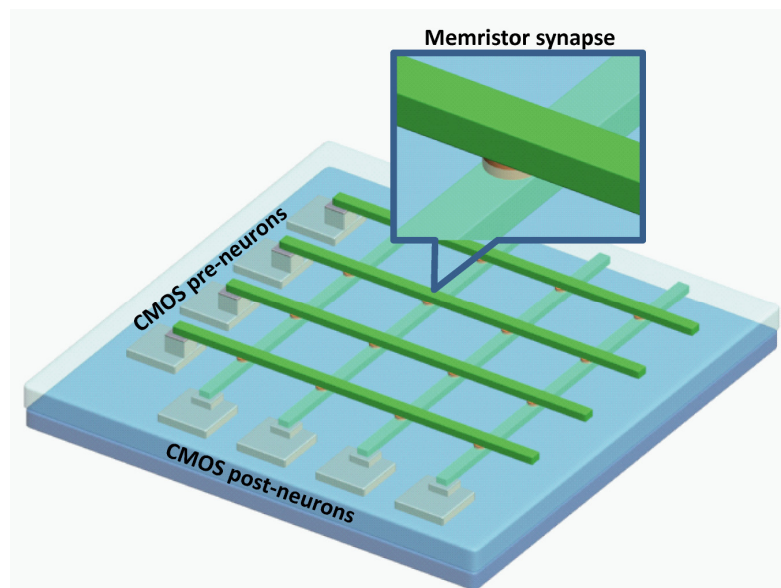
Figure 5.1. Hybrid CMOS neuron/memristor synapse circuit
to emulate the biological neuron system

The hybrid memristor/CMOS circuits discussed here can be fabricated using similar techniques developed for the proposed crossbar-based memory and logic [14-17]. However, unlike abrupt "digital" switching devices used in the memory applications discussed in Chapters 2 ~ 4, analog switches are needed here to change the conductance incrementally by a small amount (a few % change/pulse) depending on the timing and the magnitude of the applied biases. In this chapter, we will discuss our work on the experimental demonstration of synaptic functions in a solid-state memristor device, and show that a hybrid system composed of CMOS neurons and nanoscale memristor synapses can support important learning rules such as Spike Timing Dependent Plasticity (STDP). These results suggest the hybrid CMOS/memristor approach offers a promising path to neuromorphic systems.

5.2 Memristor and Memristive System

Conventional electronic circuit operation is described by four fundament circuit variables, voltage, current, charge and flux (time integration of voltage) and the functional relationships of the circuit variables are defined by passive circuit elements, resistor, capacitor and inductor. However, Prof. Leon Chua at UC Berkeley reasoned in 1971 [10] that there should be a fourth fundamental circuit element (Memristor) which can connect one missing link (Fig. 5.2) between charge and flux, hence completing the functional relationships between the four fundamental circuit variables.

As briefly discussed in Chapter 1, in this theory, the memristance (*M*) determines the relationship between charge and flux, $d\varphi = Mdq$. In the special case the memristance (*M*) is constant (linear element), then the memristance is just resistance. However, if the

memristance is a non-linear element (i.e. $M = f(q)$ for a charge-controlled memristor), the system possesses a set of properties such as hysteretic I-V characteristics that cannot be described by other elements such as resistor, inductor and capacitor [1-3].
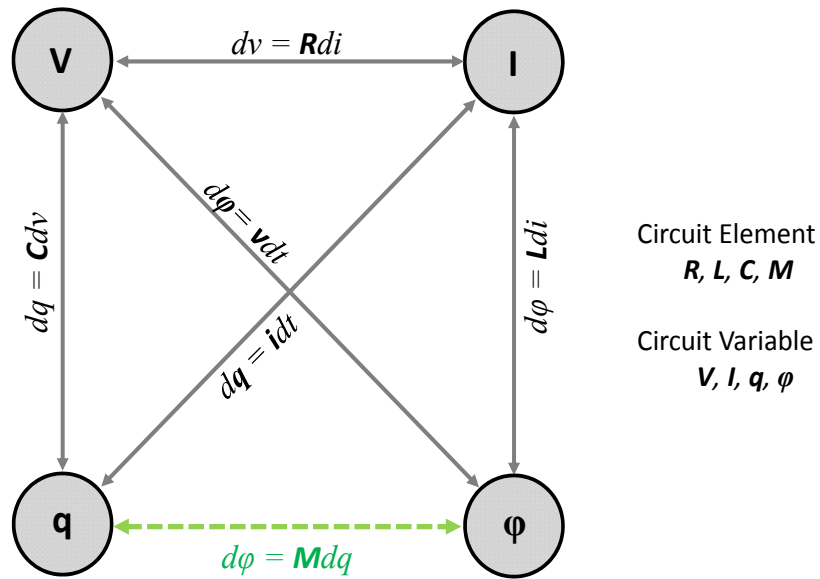
Figure 5.2. Four fundamental circuit variables, voltage (V), current (I), charge (q) and flux (φ). Functional relationships of a par of the variables are defined by the circuit elements, resistance (R), inductance (L), capacitance (C), and memristance (M).

In general, any class of two terminal devices whose resistance depends on an internal state variable (w) of the device can be termed as a memristive system [18] and described by the following set of equations

$$v(t) = M(i, w, t) \cdot i(t)$$

$$\frac{dw}{dt} = f(i, w, t) \qquad \text{(Eq. 5.1)}$$

where w is the state variable. A charge controlled memristor is a special case of Eq. 5.1 where $w = q$ and M is a function of charge (q) only $M=M(q(t))$. Since $q = \int i dt$, the memristance of the device depends not on the current value of the inputs but on the

history of the inputs instead, leading to the observed hysteretic effects. Since q is a continuous function of time, the memristor is in principle an analog device with its memristance incrementally adjusted by the charge flow $\delta M = \dfrac{dM}{dq}\delta q$ (flux flow for a flux-controlled memristor).

The memristive effect in fact has been widely observed in several devices in which the system changes the resistance with respect to the internal state (w) of the device such as spin polarization in spintronic devices [19, 20] and the temperature dependent internal state in thermistor [21]. The memristor theory received significant more attention in 2008 after a group from HP showed that the resistance switching phenomenon observed in nanoscale metal oxide films can be explained by the simple charge-controlled memristor ($M = f(q)$) model [11].

The major properties of a memristive system are (A) memristive systems are passive, in other words, $M(i, w, t) > 0$; (B) whenever $i = 0$ then $v = 0$ (or vice versa) so that the memristive system can not store energy similarly to an inductor and a capacitor; [18], (C) the system shows hysteretic output as response to a periodic input (current or voltage). In particular, (B) and (C) lead to "pinched" hysteretic loops observed the memristor I-V characteristics that are suitable for emerging memory and logic applications [11, 22-24].

5.3 Memristors based on Amorphous Silicon as the Switching Medium

As discussed in Chapters 2 - 4, in "digital" resistance switching devices a localized conducting filament is formed that leads to abrupt resistance change, schematically illustrated in Fig. 5.3 (a). On the other hand, for analog memristors an ion

rich region and an ion-poor region need to be created in the switching layer so the motion of the conductance "front" between the two regions needs to a gradual conductance change, as schematically illustrated in fig. 5.3 (b).
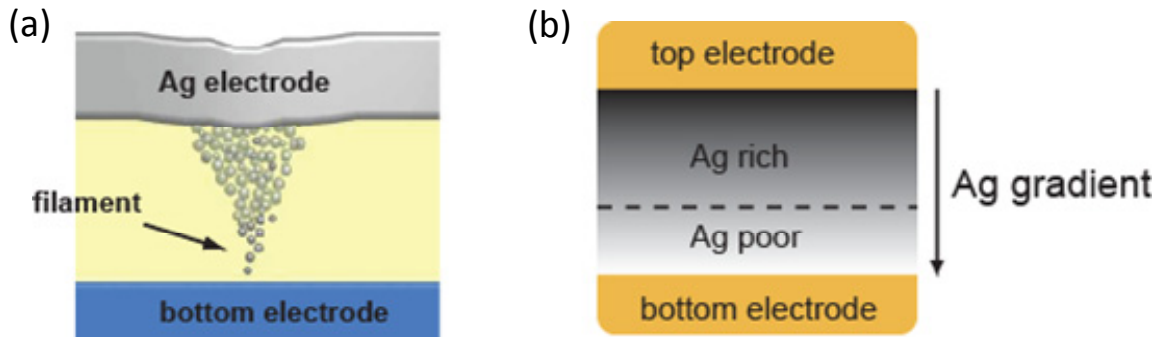
Figure 5.3. Device structures for digital switching (a) and analog switching (b).

We have demonstrated analog memristor devices using a co-sputtered Ag/Si layer as the active layer with a properly designed Ag/Si mixture ratio gradient to create the formation of a Ag-rich (high conductivity) region and a Ag-poor (low conductivity) region (Fig. 5.4 (a)). The memristor device consists of a bottom tungsten nanowire electrode, a sputtered silicon layer (2~4 nm), a PECVD (plasma enhanced chemical vapor deposition) deposited amorphous silicon (a-Si) layer (2.5~4.5 nm), a co-sputtered silver and silicon layer (20~30 nm thick) and a top chrome/platinum nanowire electrode as shown in Fig. 5.4. The bottom tungsten nanowire electrode was fabricated by e-beam lithography, a masking layer lift-off, and RIE (reactive ion etching). Following that, the a-Si deposition, co-sputtering and top metal lift-off processes were performed. Finally, another RIE step was performed to remove the co-sputtered layer outside the cross-point region defined by the of top and bottom electrodes. By co-sputtering of Ag and Si, a

uniform conduction front between the Ag rich and Ag poor regions can be formed, resulting in incremental resistance changes as the front formed by metal particles moves within the a-Si layer. The sputtered Si layer here serves as a "blocking" layer to prevent high currents through the device when the front directly shorts the bottom W electrode.
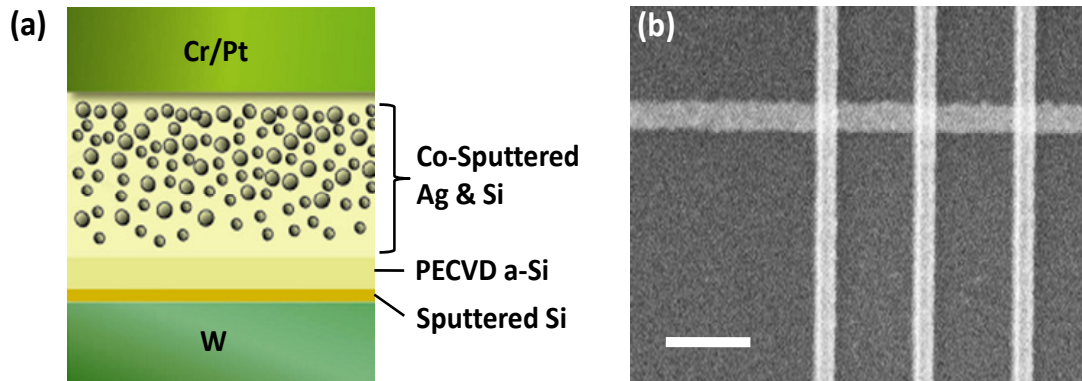
Figure 5.4. (a) Schematic cross-section view of a device. (b) Scanning electron microscope (SEM) image of a 1×3 memristor crossbar array. The top Cr/Pt nanowire electrodes and the bottom W nanowire electrode are clearly visible. Scale bar: 300nm.

Typically, resistance switching devices regardless of switching material being used require an electroforming process during which semi-permanent structural modification occurs inside the switching medium. The forming process creates localized conducting paths (conducting filaments) of which motion typically results in abrupt resistance switching [22-27]. For the co-sputtered device, a uniform conduction front already exists in the as-fabricated device between the metal rich and metal poor regions hence the forming process is no longer necessary. Fig. 5.5 shows the measured device current $i(t)$ (blue lines) as a function of the applied voltage $v(t)$ for five consecutive positive voltage sweeps and five consecutive negative voltage sweeps. Distinct from

digital devices that show abrupt conductance jumps, here the conductance continuously increases (decreases) during the positive (negative) voltage sweeps. In fact, the device *I-V* can be well fitted by a simple memristor circuit model (orange lines, Fig. 5.5):

$$i(t) = \frac{1}{R_{ON}w(t) + R_{OFF}(1 - w(t))}v(t) \qquad \text{(Eq. 5.2)}$$

Here $w(t)$ is the state variable and stands for the normalized position of the conduction front between the Ag-rich and Ag-poor regions within the active layer and has the value between 0 and 1. As $w(t)$ approaches to 0(1), the device reaches the lowest (highest) conductance state with resistance of R_{OFF} (R_{ON}). In this model we further assumed the position of $w(t)$ is a linear function of net flux-linkage $\varphi(t) = \int v(t)dt$ through the device. Eq. 5.2 thus can be rewritten as:

$$i(t) = G(\varphi(t))v(t) \qquad \text{(Eq. 5.3)}$$

This is the equation for a flux-controlled memristor [10, 18] and $G(\varphi(t))$ is the so-called memductance. In addition, for the devices studies here bias voltages of $< V_T = 2.2$ V are not sufficient to drive the Ag ions inside the a-Si matrix and have negligible effect on the memristor resistance. The threshold effect and the value of the threshold voltage VT have been consistently obtained in all the devices tested in this study.
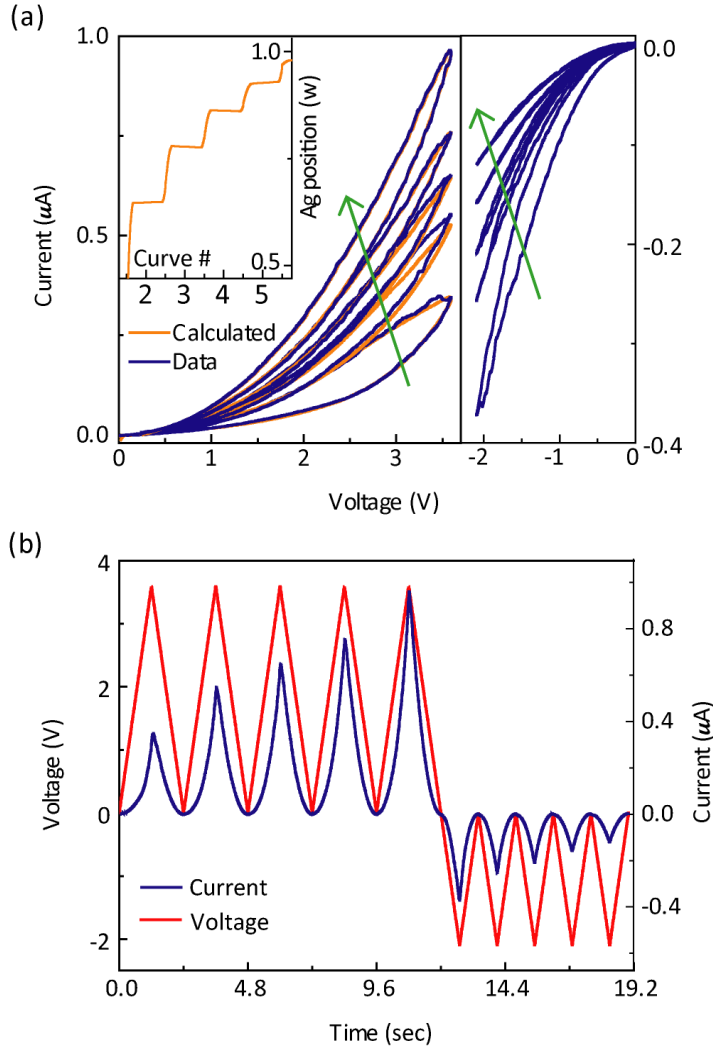
Figure 5.5. Analog behavior of co-sputtered devices. (a) Measured (blue lines) and calculated (orange lines) *I-V* characteristics of the memristor. Device size: 500 nm × 500 nm. Inset: calculated normalized Ag front position *w* during positive DC sweeps. (b) Re-plot of (a) in voltage versus time.

The calculated current values during the voltage sweeps based on the memristor model discussed above were shown as orange lines in Fig. 5.5(a) together with the measured current (blue lines). The inset to Fig. 5.5(a) shows the values of $w(t)$ (orange lines) used to calculate the current during the 5 consecutive positive voltage sweeps by assuming $w(t)$ to be a linear function of the flux-linkage $\varphi(t) = \int v(t)dt$. The relative

good agreements between the calculated and the measured values in $i(t)$ and $w(t)$ verify that the device characteristics above the threshold voltage can indeed be explained by the memristor-model using equations 5.2 and 5.3, where the front position is roughly a linear function of the flux-linkage $\varphi(t) = \int v(t)dt$. However, strictly speaking the device is not a true memristor due to the threshold effect but falls in the more broadly defined memristive device category [18]. On the other hand, the threshold effect makes it possible to perform non-disturbing read of the device state by using read pulses with $V_{read} < V_T$ and can in fact be beneficial in practical applications.

The memristor response can also be observed clearly in pulse programming. From Eq. 5.3, positive voltage pulses potentiate (P) the conductance of the a-Si memristor while negative voltage pulses depress (D) the device conductance gradually, as shown in Fig. 5.6(a). In contrast, the devices without the co-sputtered layer (e.g. with structures similar to those described in Chapters 2 ~ 4) show abrupt resistance changes and quickly reach their final conductance values (Fig. 5.6(b)).
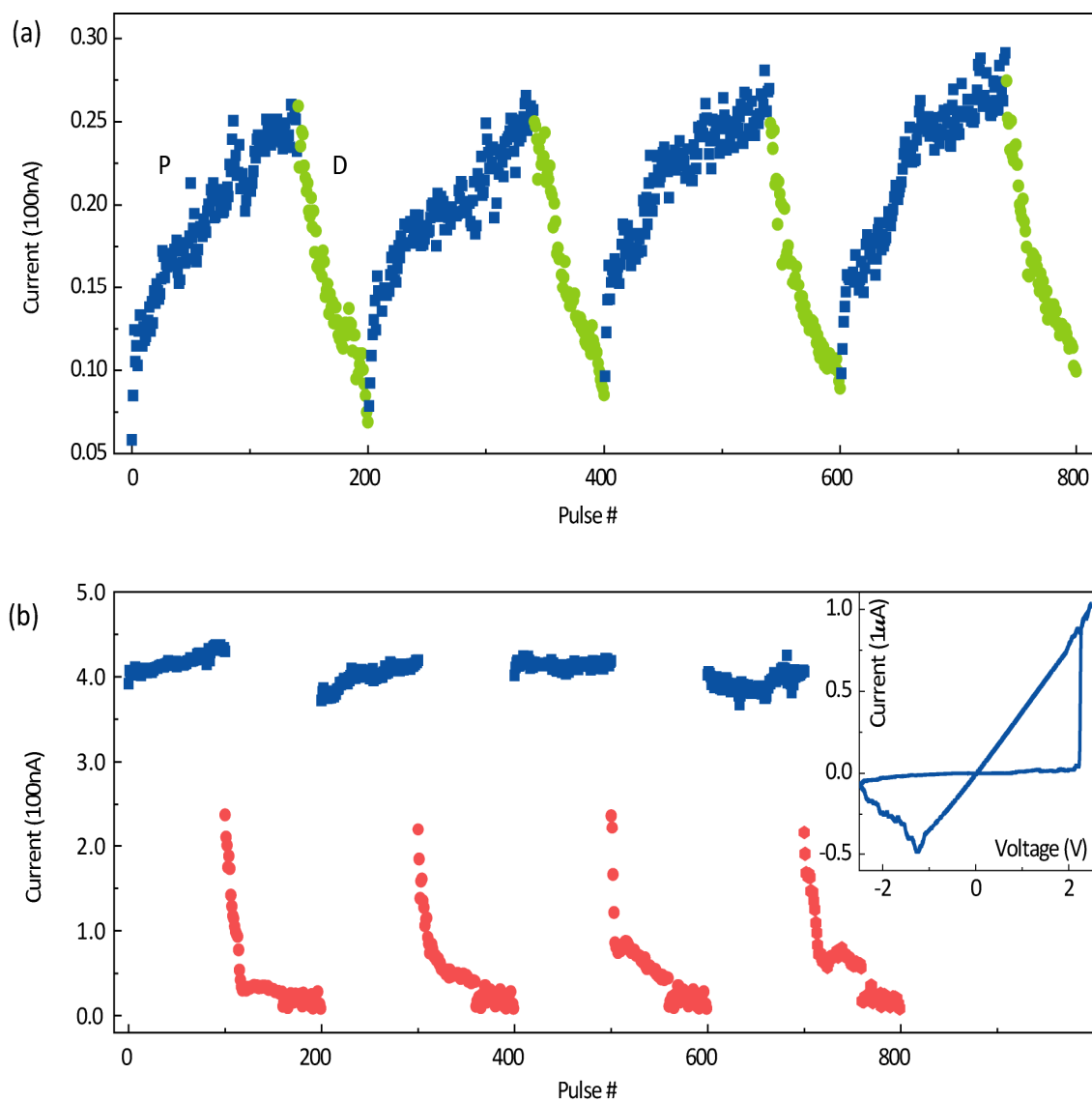
Figure 5.6. Pulse response of the analog and digital switching devices. Alternating 140 potentiation (3V, 300μsec) and 60 depression (-3V, 300μsec) pulses were applied. After each pulse, the device was measured at 1V read pulses and the read current was plotted. (a) Response to the pulses from an analog memristor device. The overall increase in conductance is due to the asymmetrically applied potentiation and depression pulses. (b) Response to the same programming pulses from a digital device. Inset: IV characteristic of the digital resistance switching device.

Though the conduction front cannot be directly observed in the nanoscale vertical device structure as shown Fig. 5.4, the motion of the Ag particles and similar analog switching behavior have been observed in a horizontal device geometry (Fig. 5.7), which supports the hypothesis of uniform motion of a conduction front. In this case, Ag is uniformly distributed between two parallel electrodes in the as-fabricated device. Then the top silicon layer was selectively etched by using a RIE (reactive ion etching) method. During the etching process, nanometer scale Ag particles became larger particles (a few tens to hundreds of nanometer in diameter) due to clustering. The conductance of the virgin device is the highest, since Ag particles are uniformly distributed between the two electrodes. Upon applying voltage biases on the device, analog switching (Fig. 5.7) similar to the vertically stacked device (Fig. 5.4) was observed and the conductance of the device decreases. After several DC sweeps, it can be seen that the Ag particles are repelled from the positively biased electrode, confirming the Ag front-moving hypothesis.
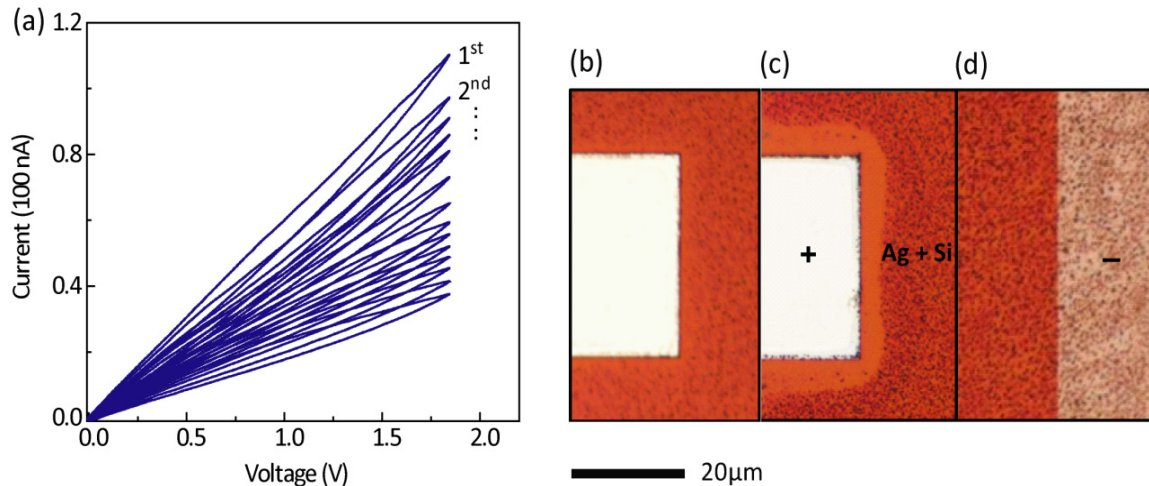
Figure 5.7. Analog switching behavior and Ag redistribution in a planar type device. (a) Successive positive DC sweeps showing continuous conductance change. (b) An as-fabricated electrode. (c) A positively biased electrode during the DC sweeps shown in (a). (d) A negatively biased electrode.

5.4 Memristor Synapse for Neuromorphic Systems

In this section, we discuss the experimental implementation of synaptic functions in a hardware system using nanoscale memristor as synapses and CMOS circuits and neurons, as conceptually shown in Fig. 5.8.
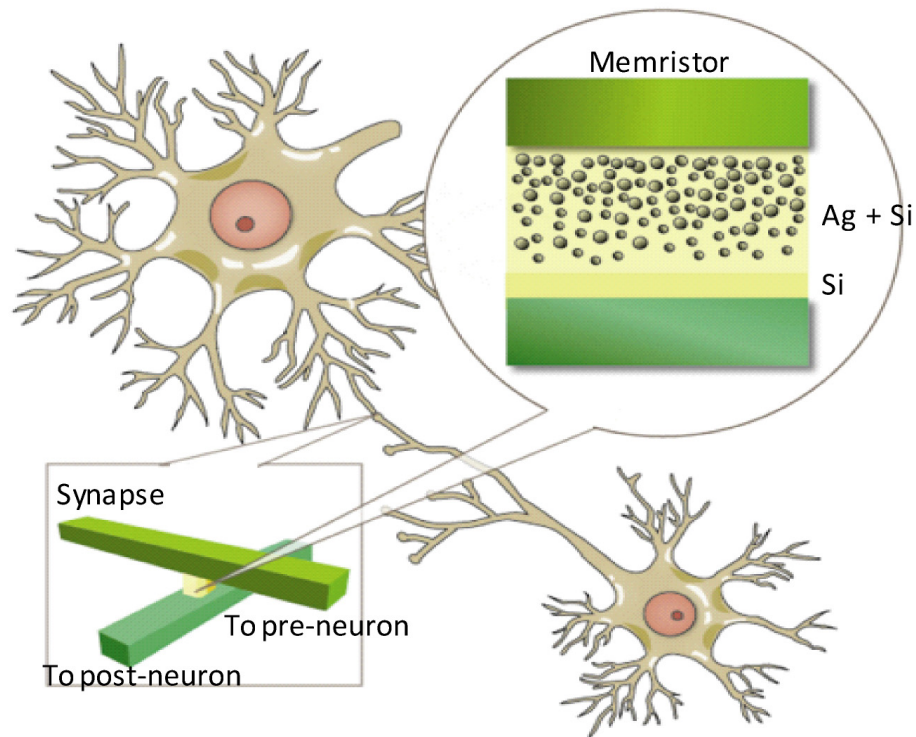
Figure 5.8. Schematic of a memristor synapse approach for neuromorphic systems.

The particular synaptic functionality we are interested in this study is termed spike-timing dependent plasticity (STDP), which has been shown to be an important synaptic modification rule for competitive Hebbian learning [6-8]. Introduced in 1949 by D. Hebb [6], Hebbian rule describes how learning process works in biological neural systems. In general, a learning process involves modification of the synaptic weights between neurons. Hebbian rule is often summarized as:

"cells that fire together, wire together."

even though the above description over-simplifies the actual learning process. The Hebbian rule is widely accepted as one of the fundamental learning mechanisms in neuron system. With recent progresses in measurement technologies, more accurate measurements of the relative timings of the neuron spikes could be achieved [28-32] and it was found that the synaptic connection between two neurons becomes stronger if the presynaptic neuron fires off shortly before the postsynaptic neuron, while the synaptic weight becomes weaker if the presynaptic neuron fires off shortly after the postsynaptic neuron. In addition, the smaller the spike timing difference gives the larger the synaptic weight change with the synaptic weight change roughly an exponential function of the spiking timing difference. In general, this spike timing sensitive synaptic modification is termed the STDP rule. The spike timing window for the synaptic modification is typically an order of a few milliseconds to hundreds of millisecond. Overall STDP is one of the key rules responsible for the efficient learning and stability in biological systems.

The conductance (memductance) of the flux-controlled memristor discussed in the previous section can be incrementally adjusted by tuning the amplitude as well as the duration of the applied programming voltage. Fig. 5.9 (a) shows the results when the device was programmed by a series of 100 identical positive (3.2 V, 300 µs) pulses followed by a series of 100 identical negative voltage pulses (-2.8 V, 300 µs). The device conductance (represented by the measured current at a small read voltage of 1 V) was measured after each programming pulse. As expected from DC response of the device (Fig. 5.5), the application of positive potentiating voltage pulses (P) incrementally increases the memristor conductance, and the application of negative depressing voltage

pulses (D) incrementally decreases the memristor conductance. Different from the digital devices where the programming signals control the final device state, here the flux-linkage φ(t) during each programming pulse controls the relative change of the memristor conductance. This effect was further demonstrated in Fig. 5.9 (b).
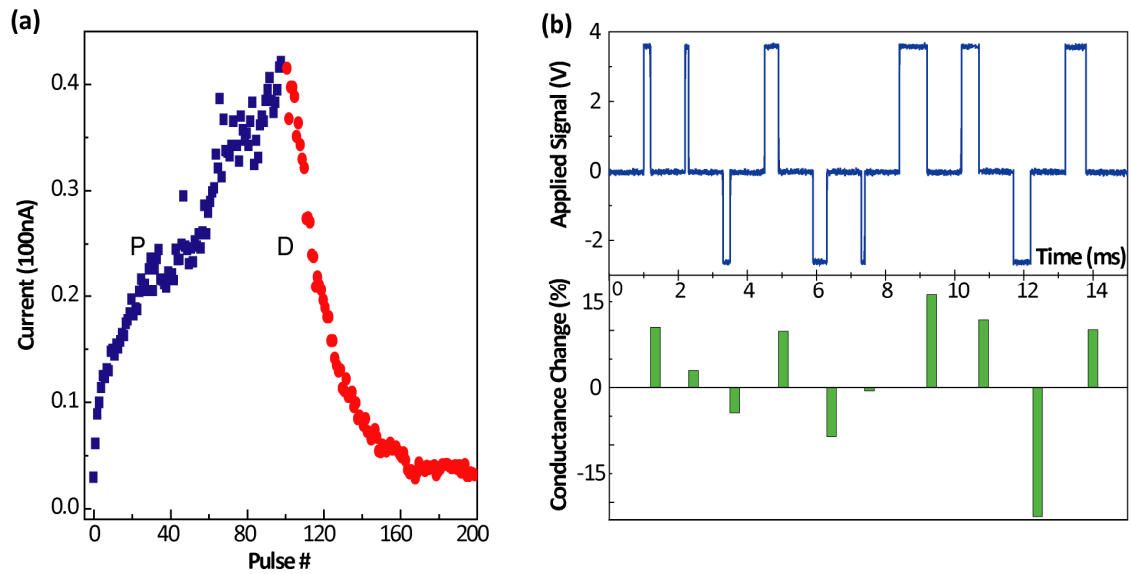
Figure 5.9. Memristor response to programming pulses. (a) The device conductance can be incrementally increased or decreased by consecutive potentiating or depressing pulses. The conductance was measured at 1 V after each pulse and the read current is plotted. P: 3 V, 300 µs, D: -2.8 V, 300 µs. (b) Top: mixed potentiating and depressing pulses with different pulse widths that are used to program the memristor. Bottom: Measured change of the memristor conductance after the application of each pulse. The conductance change was normalized to the maximum memristor conductance.

Here mixed positive (P) and negative (D) voltage pulses with constant pulse height but different pulse widths were applied to the device and the change in memristor conductance ΔG were measured and recorded after each P/D pulse. A clear correlation between ΔG and the pulse width of the applied P/D signals was observed - the application of a longer positive (negative) pulse resulted in a larger increase (decrease) of memristor

conductance, and vice versa. These results suggest the memristor devices are capable to emulate biological synapses with properly designed CMOS neuron components to provide local programming voltages with controlled pulse width and height.

A CMOS neuron circuit was developed in collaboration with Prof. Pinaki Mazumder's group at the University of Michigan to convert the relative timing information of the neuron spikes into pulse width information that will be sensed by the memristor synapse (Appendix). Briefly, the neuron circuit consists of two CMOS based integrate-and-fire neurons. A nanoscale memristor with active device area of 100 nm × 100 nm serves as the synapse between the two CMOS neurons. The neuron circuit involves a mixed analog-digital design and employs a Time Division Multiplexing (TDM) approach with globally synchronized time frames to convert the timing information into a pulse width [33]. Specifically, the neuron circuit generates a potentiating (depressing) pulse across the memristor synapse when the pre-synaptic neuron spikes before (after) the post-synaptic neuron, with the pulse width being an exponentially decaying function of the relative neuron spike timing $\Delta t = t_{pre} - t_{post}$, where t_{pre} (t_{post}) is the time when the pre-synaptic neuron (post-synaptic neuron) spikes.

Fig. 5.10 (a) shows the measured change of the memristor synaptic weight after each neuron spiking event obtained in the hybrid CMOS-neuron/memristor-synapse circuit. As expected, when the pre-neuron spikes before (after) the post-neuron, the memristor synaptic weight increases (decreases). In addition, the change in the synaptic weight vs. the spike timing difference Δt can be well fitted with exponential decay functions, verifying that STDP characteristics similar to that of biological synaptic systems (e.g., Fig. 5.10 (b)) can indeed be obtained in memristor synapses [7, 8].
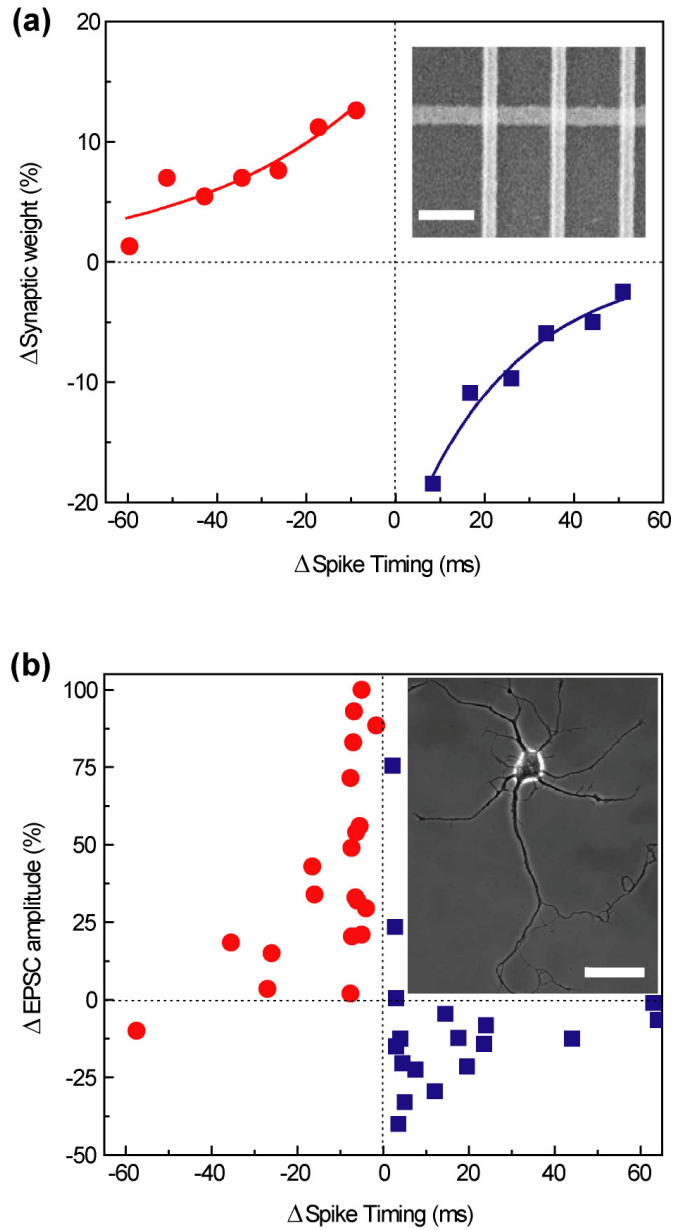
Figure 5.10. Demonstration of STDP in the memristor synapse. Demonstration of STDP in the memristor synapse. (a) The measured change of the memristor synaptic weight vs. the relative timing Δ*t* of the neuron spikes. The synaptic change was normalized to the maximum synaptic weight. Inset: scanning-electron microscope image of a fabricated memristor crossbar array. Scale bar: 300 nm. (b) The change in excitatory postsynaptic current (EPSC) of rat hippocampal neurons after repetitive correlated spiking (60 pulses at 1Hz) vs. relative spike timing. The figure was reconstructed from Ref. 8. Inset: Schematic of neurons. Inset: A phase contrast image of a hippocampal neuron. Scale bar: 50μm. The image was adapted with permission from Ref. 34.

Finally, in figure 5.11 we show the P/D response of the device after 5×10^5, 1×10^6, 1×10^7, and 1.5×10^8 times of potentiation and depression. As shown in figure 5.11 (d), up to 1.5×10^8 times of P/D reasonable conductance modulation can be achieved. If we assume the synapses are updated at a rate of 1 Hz [35, 36], this endurance corresponds to ~5 years of continuous synaptic operation.
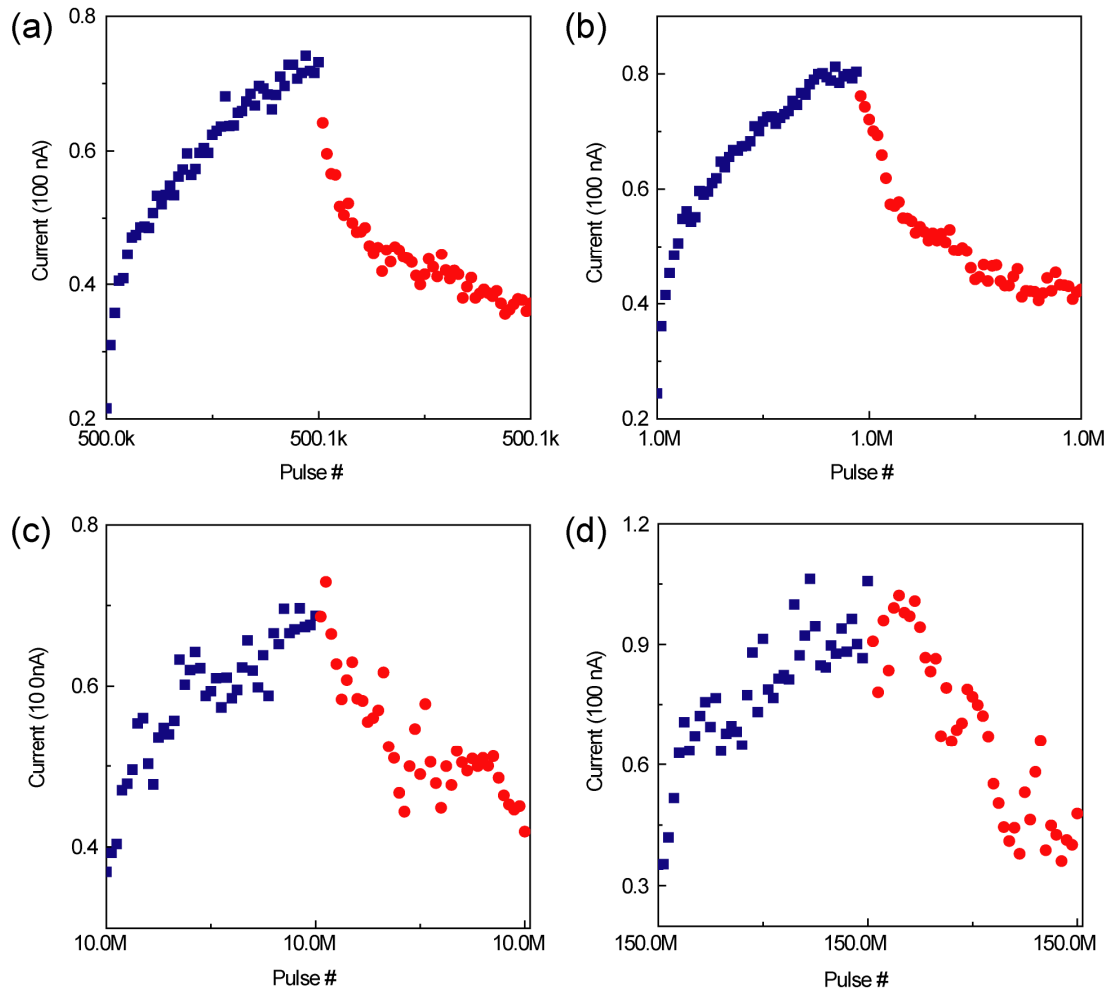
Figure 5.11. The pulse response of the memristor devices. The device was subjected to repeated potentiating (P) and depressing (D) pulses and the device conductance was monitored and plotted after each programming pulse. (a) After 5.0×10^5 P/D pulses. (b) After 1.0×10^6 P/D pulses. (c) After 1.0×10^7 P/D pulses. (d) After 1.5×10^8 P/D pulses. 3.1 V, 800 μs potentiating pulses, -2.9 V, 800 μs depressing pulses and 1 V, 2ms read pulses were used.

5.5 Conclusion

In this chapter, we demonstrated that well designed amorphous silicon based two terminal switches can show analog memristor behavior with long endurance. Furthermore, the memristive devices modulate the resistance incrementally based on the amplitude and width of the voltage pulses applied across it so that the devices can be used to emulate synapses in biological systems. A hybrid CMOS neurons/memristor synapse circuit has been fabricated and was shown to follow STDP (spike timing dependent plasticity) synaptic adaption rule, one of the fundamental learning rules in biological systems, showing promising application of hybrid circuits to neuromorphic systems. We believe these demonstrations, together with the large connectivity and density offered by the two-terminal memristor synapses in the crossbar configuration, make the hybrid CMOS-neuron/memristor-synapse approach promising for hardware implementation of biology-inspired neuromorphic systems.

Appendix

Neuron Circuits

The neuron circuit involves a mixed analog-digital design. Time Division Multiplexing (TDM) – events occur at their prescribed time periods – is used in this design. Timeslots are used to organize communication between the neurons. This system is compatible with digital systems in which a global clock is used to synchronize communication and thereby instill a greater external control over the neurons. Three timeslots are used in this design and are explained further with the aid of Fig. A5.1 which

shows the spiking characteristics of a neuron as well as the use of TDM.
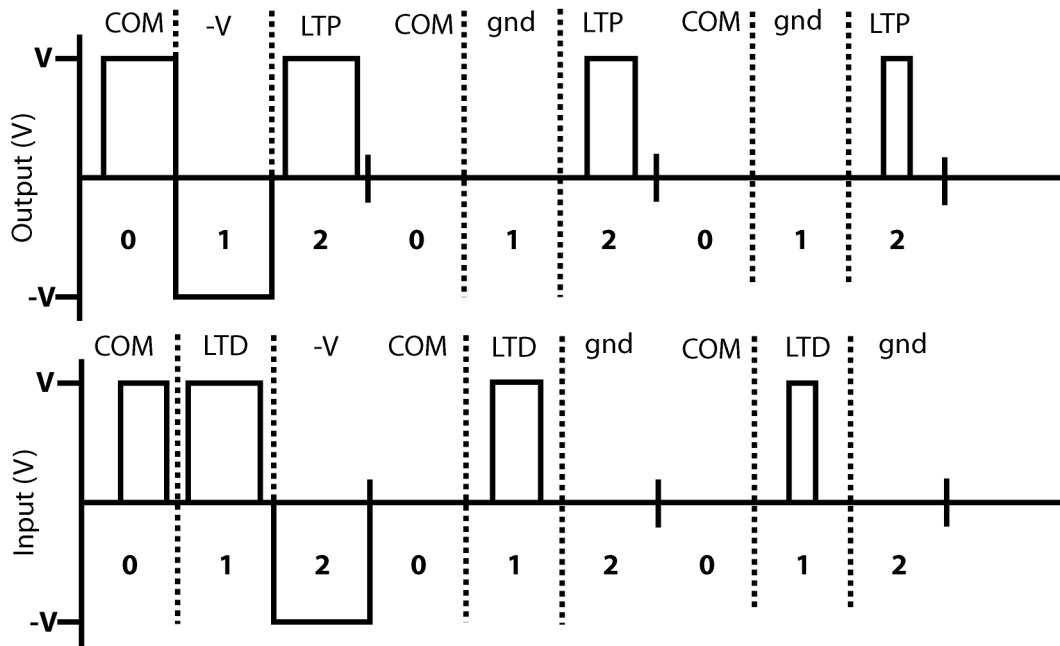
Figure A5.1: A generic neuron's spiking characteristics.

On the output side of the neuron, the three timeslots are 0 (Communication), 1 (Response), and 2 (LTP). In the Communication (COM) timeslot, the neuron can give out a spike signal with a voltage amplitude V. The amplitude V is chosen to be between $V_T/2$ and V_T so that a single LTP or LTD pulse will not change the memristor synaptic weight but a simultaneous application of either an LTP and a $-V$ pulse or an LTD and a V pulse across the memristor will. The COM timeslot signifies the only time when the neuron can spike or declare it has spiked. In the Response (gnd, $-V$) timeslot, if the neuron spikes, then the output of the neuron goes to a voltage of $-V$. If the neuron does not spike, then the output goes to or remains at a ground potential. In the LTP timeslot, the neuron gives pulses of decaying width for the next frames. Each frame is three timeslots long labeled 0,

1, and 2 as depicted in Fig. A5.1. The number of frames of decaying pulses depends on the characteristic of the neuron. On the input side of the neuron, the three timeslots are similar, but 0, 1, and 2 correspond to COM, LTD, and Response. The COM timeslot represents when the input of the neuron can vary based on the current being received through the memristor. LTD and Response behave similarly to LTP and Response on the output side. Combining the output of a pre-neuron to the input of a post-neuron would yield results depicted in Figure A5.2.
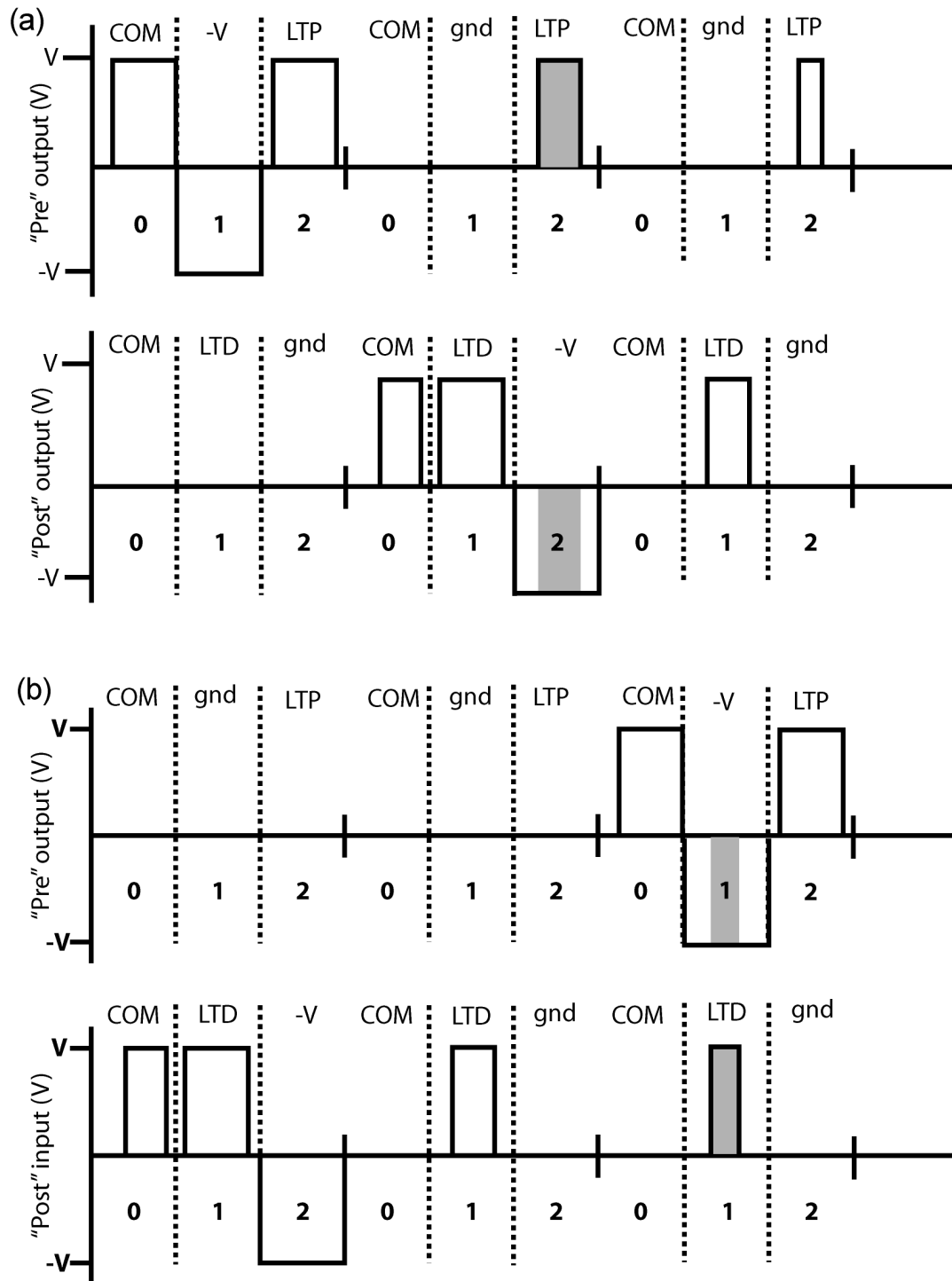
Figure A5.2: Two spiking situations showing the timeslots where synaptic change occurs. (a) Potentiation: the post-synaptic neuron spikes 1 frame after the pre-synaptic neuron. (b) Depression: the post-synaptic neuron spikes 2 frames before the pre-synaptic neuron. The shaded regions represent the time duration where the voltage across the memristor synapse is larger than V_T.

Fig. A5.2(a), the post-synaptic neuron ("Post") spikes a frame after the pre-synapt

ic neuron ("Pre") spikes, so this entails potentiation. During timeslot 2 of this frame, the

"Pre" goes to $+V$ while the "Post" goes to $-V$ so the effective voltage across the

memristor will be $+2V$ which is enough to change the conductance of the memristor. The

time the memristor is exposed to this above threshold voltage is shown in gray in Fig.

A5.2(a). In Fig. A5.2(b), a similar spiking scenario is shown, except this scenario shows

depression when the "Pre" spikes two frames after the "Post". The effective voltage in

gray is $-2V$ and for a shorter amount of time than shown in Fig. A5.2(b). Using the TDM

concept, a neuron circuit was built with discrete components and an FPGA to test the

writing and erasing of a memristor and obtain the STDP curve. Figure A5.3 depicts the

neuron architecture used to obtain the I/O behavior shown in Fig. A5.1.
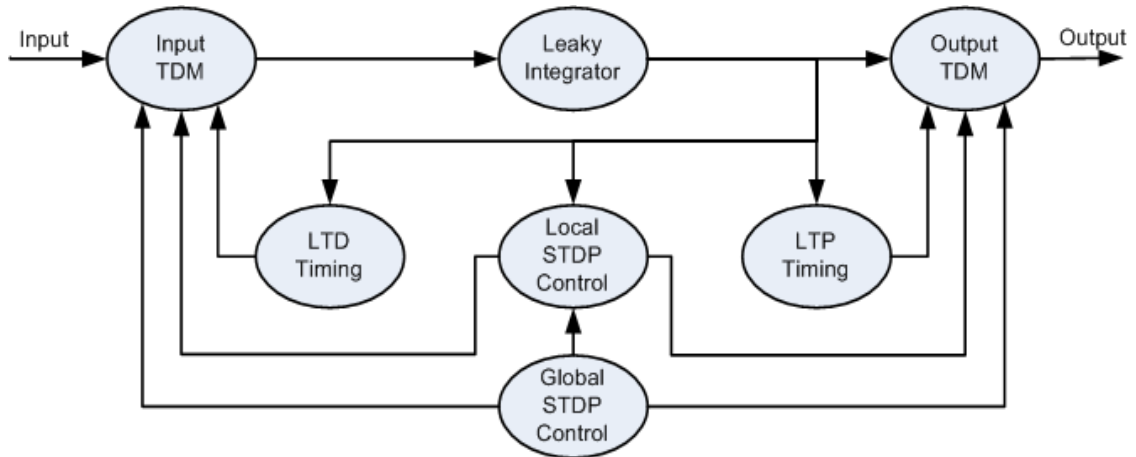
Figure A5.3. Overall neuron architecture grouped by components

The input to the neuron system is a current – assuming a memristor is connected

directly to the input of the neuron. The Input TDM is essentially a demux that allows the

input current, LTD voltage, or the $-V$ voltage to pass at each appropriate timeslot. The

leaky integrator integrates the input current, and once threshold is reached, sends out a

spike signal to activate the LTP/LTD pulses and set the appropriate local controls. The Local STDP control sets the control signal for passing $-V$ at the appropriate timeslot whenever the leaky integrator gives out a spike signal. The Global STDP control sets the individual timeslots for each frame and gives the designer auxiliary inputs to the neuron. Both the Local and Global STDP controls were implemented on an Altera DE1 Development board while the other components were realized on a breadboard using off-the-shelf components. Eight time frames were used during each neuron spike. Figure A5.4 shows the pulse widths of the potentiating and depressing pulses generated by the CMOS neuron circuit that are applied across the memristor synapse as a function of the relative timing Δt of the neuron spikes.
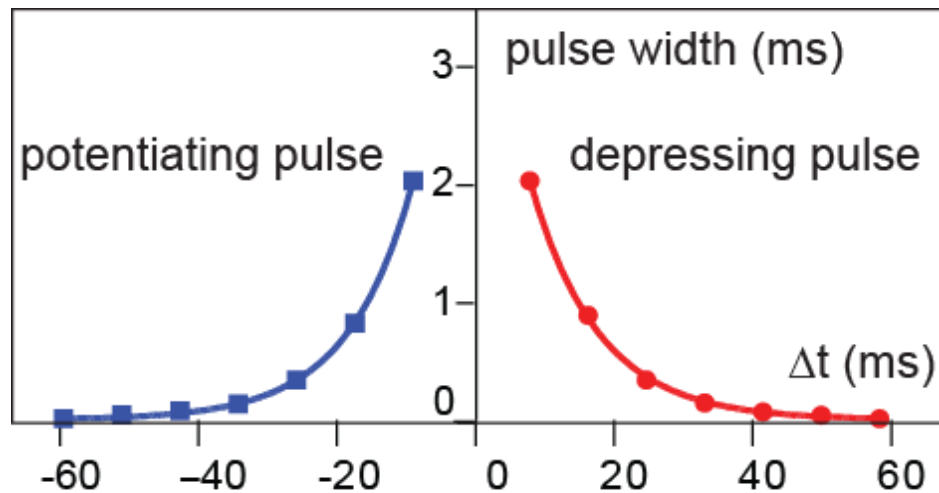
Figure A5.4. Pulse widths of the potentiating and depressing pulses generated by the CMOS neuron circuit.

References

[1]   L. S. Smith, *Handbook of Nature-Inspired and Innovative Computing: Integrating Classical Models with Emerging Technologies*. Springer, 433-475, 2006.

[2]   R. Ananthanarayanan, S. K. Esser, H. D. Simon, D. S. Modha, "The cat is out of the bag: cortical simulations with 109 neurons, 1013 synapses," *Proc. IEEE/ACM Conf. High Perform. Networking Computing*, 2009.

[3]   E. M. Izhikevich, G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *Proc. Natl. Acad. Sci. USA* **105**, 3593-3598, 2008.

[4]   G. Indiveri, E. Chicca, R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike–timing dependent plasticity," *IEEE Trans. on Neural Networks* **17**, 211-221, 2006.

[5]   Y. Tang, J. R. Nyengaard, D. M. G. De Groot, H. J. G. Gundersen, "Total regional and global number of synapses in the human brain neocortex," *Synapse* **41**, 258-273, 2001

[6]   D. O. Hebb, *The organization of behavior. A neuropsychological theory*. John Wiley and Sons Inc., 1949.

[7]   S. Song, K. D. Miller, L. F. Abbott, "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity," *Nature Neurosci.* **3**, 919-926, 2000.

[8]   G. Q. Bi, M. M. Poo, "Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type," *J. Neurosci.* **18**, 10464-10472, 1998.

[9]   R. Douglas, M. Mahowald, C. Mead, "Neuromorphic Analogue VLSI," *Annu. Rev. Neurosci.* **18**, 255-281, 1995

[10]   L. O. Chua, "Memristor--The missing circuit element," *IEEE Trans. on Circuit Theory* **18**, 507-519, 1971.

[11]   D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, "The missing memristor found," *Nature* **453**, 80-83, 2008.

[12]   J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin *et al*., "A 160-kilobit molecular electronic memory patterned at 10^{11} bits per square centimetre," *Nature* **445**, 414-417, 2007.

[13]   G. Y. Jung, E. Johnston-Halperin, W.Wu, Z. Yu, S. Y. Wang *et al*., "Circuit

Fabrication at 17 nm Half-Pitch by Nanoimprint Lithography," *Nano Lett.* **6**, 351-354, 2006.

[14] J. Borghetti, Z. Li, J. Straznicky, X. Li, D. A. A. Ohlberg *et al*., "A hybrid nanomemristor/transistor logic circuit capable of self-programming," *Proc. Natl. Acad. Sci. USA* **106**, 1699-1703, 2009.

[15] D. B. Strukov, and K. K. Likharev, "CMOL FPGA: A cell-based, reconfigurable architecture for hybrid digital circuits using two-terminal nanodevices," *Nanotechnology* **16**, 888–900, 2005.

[16] Likharev, and D. B. Strukov, CMOL: Devices, circuits, and architectures, in G. Cuniberti et al. (eds.), Introducing Molecular Electronics, Ch. 16. Springer, Berlin, 2005.

[17] G. S. Snider, R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnology* **18**, 035204, 2007.

[18] L. O. Chua, S. M. Kang, "Memristive devices and systems," *Proc. IEEE*, **64**, 209-223, 1976.

[19] Y. V. Pershin, M. D. Ventra, "Spin memristive systems: Spin memory effects in semiconductor spintronics," *Phys. Rev. B* **78**, 113309, 2008.

[20] Y. V. Pershin, M. D. Ventra, "Current-voltage characteristics of semiconductor/ ferromagnet junctions in the spin-blockade regime," *Phys. Rev. B* **77**, 073301, 2008.

[21] M. Sapoff, R. M. Oppenheim, "Theory and application of self-heated thermistors," *Proc. IEEE* **51**, 1292-1305, 1963.

[22] S. H. Jo, W. Lu, "CMOS Compatible Nanoscale Nonvolatile Resistance Switching Memory," *Nano Lett*., **8**, 392-397, 2008.

[23] S. H. Jo, K. -H. Kim, and W. Lu, "Programmable Resistance Switching in Nanoscale Two-Terminal Devices," *Nano Lett.* **9**, 496-500, 2009.

[24] J. J. Yang, M. D. Pickett, X. Li, D. A. Ohlberg, D. R. Stewart *et* al., "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nat. Nanotechnol.* **3**, 429–433, 2008.

[25] M. Kund, G. Beitel, C.-U. Pinnow, T. Rohr, J. Schumann *et* al, "Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20nm," IEDM, 754 - 757, 2005.

[26] R. Waser, and M. Aono, "Nanoionics-based resistive switching memories," *Nature Mater.* **6**, 833-840, 2007.

[27] M. Liu, Z. Abid, W. Wang, X. He, Q. Liu *et al.*, "Multilevel resistive switching with ionic and metallic filaments,"*Appl. Phys. Lett.* **94**, 233106, 2009.

[28] D. Debanne, B. H. Gahwiler, S. M. Thompson, "Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus in vitro," *Proc. Natl. Acad. Sci. USA* **91**, 1148–1152, 1994.

[29] H. Markram, J. Lubke, M. Frotscher, B. Sakmann, "Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs," *Science* **275**, 213-215, 1997.

[30] J. C. Magee, D. A. Johnston, "A Synaptically Controlled, Associative Signal for Hebbian Plasticity in Hippocampal Neurons," *Science* **275**, 209-213, 1997.

[31] C. C. Bell, V. Z. Han, Y. Sugawara, K. Grant, "Synaptic plasticity in a cerebellum-like structure depends on temporal order," *Nature* **387**, 278-281, 1997.

[32] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris , M. M. Poo, "A critical window for cooperation and competition among developing retinotectal synapses," *Nature* **395**, 37–44, 1998.

[33] G. S. Snider, "Spike-timing-dependent learning in memristive nanodevices," *IEEE/ACM International Symposium on Nanoscale Architectures*, 85-92, 2008 (NANOARCH 2008).

[34] S. Kaech, G.Banker, "Culturing hippocampal neurons," *Nature Protocols **1***, 2406-2415, 2006.

[35] V. Lev-Ram, S. T. Wong, D. R. Storm, R. Y. Tsien, "A new form of cerebellar long-term potentiation is postsynaptic and depends on nitric oxide but not cAMP," *Proc. Natl. Acad. Sci. USA* **99**, 8389-8393, 2002.

[36] M. F. Bear, "Colloquium Paper: A synaptic basis for memory storage in the cerebral cortex," *Proc. Natl. Acad. Sci. USA* **93**, 13453-13459, 1996.

Chapter 6

Nanoscale Memristor based on Tungsten Oxide

In previous chapters, we discussed the results of amorphous silicon based digital and analog (memristor) resistive devices for memory and neuromorphic applications. The incorporation of a p-Si layer as a local bottom electrode in the a-Si based devices significantly reduces the programming current and improves device yield. However, it may require a high-temperature process to activate the dopants. In this chapter, we present our studies on devices using tungsten oxide as the switching layer, of which only low-temperature processes are used which allows increased flexibility in device design and fabrication.

6.1 Introduction

Tungsten is one of the common metals used in conventional CMOS processes. It is stable at high temperature due to its high meting temperature of >3400°C and can also be produced with various methods such as sputtering and chemical vapor deposition (CVD) at relatively low temperature (<< 400°C), which makes tungsten suitable as a interconnection material in back-end-of-line (BEOL) processing in CMOS integrated circuits.

Tungsten oxide (WO_3), a wide band gap material (> 3.2eV) [1], has been extensively investigated due to its interesting electronic and photo-chromic properties [2-

4]. Tungsten oxide has various valence states and the stoicheometry can be altered by extrinsic metal ions injected in the oxide. Hence the conductivity of tungsten oxide can be varied in various ranges [5-7] even though stoicheometrically pure WO_3 has a wide band gap. Typically tungsten oxide is oxygen deficient and the material can be an n-type semiconductor or even can possess metallic property. With a large concentration of oxygen vacancies, the color of the tungsten oxide is dark-blue, and the color changes toward yellow as the density of vacancy reduces. Tungsten oxide can be formed in various methods such as direct sputtering or evaporation of WO_3 target, chemical vapor deposition (CVD), wet chemical reactions, oxidation of tungsten, sol–gel and spray pyrolysis [8- 12].

Several research groups have reported nonvolatile resistance switching behaviors of tungsten oxide, and explained the effects with chemical reactions in the oxide by treating tungsten oxide as a solid-electrolyte with Cu or Ag top electrodes [13, 14], or with metallic filament formation with the off-state explained by variable-range hoping of electrons from one oxygen vacancy (W^{a+}) to the other oxygen vacancy (W^{b+}) in the oxide [15]. Promising digital nonvolatile resistance switching behaviors such as fast switching, scalability and stable retention at high temperature have been demonstrated. However, analog memristive effects of tungsten oxide have not been reported.

6.2 Resistance Switching in Tungsten Oxide

The device in our study consists of a top palladium electrode, a tungsten oxide switching layer and a bottom tungsten electrode as shown in figure 6.1. The first step of the fabrication process is the thermal oxidation of silicon to grow 100nm thick silicon

dioxide film as an insulating substrate, followed by 60nm thick tungsten deposition by sputtering at room temperature. Then the tungsten film was patterned by e-beam lithography and RIE (reactive ion etching) to form the nanowire (~60nm) bottom electrodes and contact pads. To form the tungsten oxide layer, RTA (Rapid Thermal Anneal) in pure oxygen or plasma oxidation in O_2 plasma was performed at 400°C. During the oxidation, a part of tungsten was consumed (15 ~ 20nm) and 30 ~ 40 nm tungsten oxide (WO_{3-x}) film was formed. The thickness of tungsten oxide film saturates after certain thickness (~40nm) during the annealing process or plasma oxidation due to finite oxygen diffusion at 400°C. Finally the top palladium (Pd) nanowire electrode was formed by e-beam lithography and Pd lift-off followed by tungsten oxide etch outside the cross-points using RIE. The final device has a $Pd/WO_{3-x}/W$ structure with the device size of 80nm × 80nm.

In figure 6.2, typical resistance switching behaviors observed in $Pd/WO_{3-x}/W$ are presented. Specifically, tungsten oxide was prepared either by rapid thermal anneal in pure O_2 at 400°C for 2~3min or by plasma oxidation by using O_2 for 9min at 400°C. Interestingly, a transition of switching behavior from rectifying-like to non-rectifying was observed as the oxidation time increases, accompanied also with reduced programming voltage. Note that the switching behavior is quite different from those observed in Cu (or Ag)/WO_3/W (or Pt) structures [8, 9] in which the resistance switching effects are based on the solid-electrolyte property of tungsten oxide in chemical redox processes involving Cu (or Ag) ions, resulting abrupt on to off (or off to on) resistance change. Here due to the inert nature of the Pd top electrode and the elimination of foreign ions the redox process may not be the dominating switching mechanism.
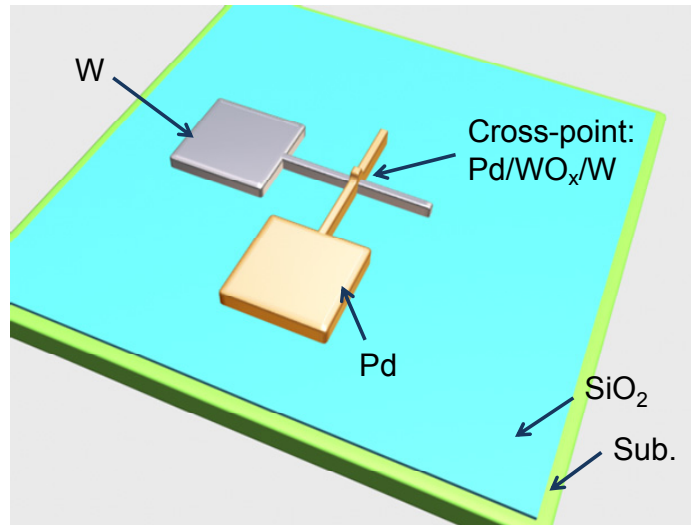
Figure 6.1. Schematic of tungsten oxide resistance switching device.
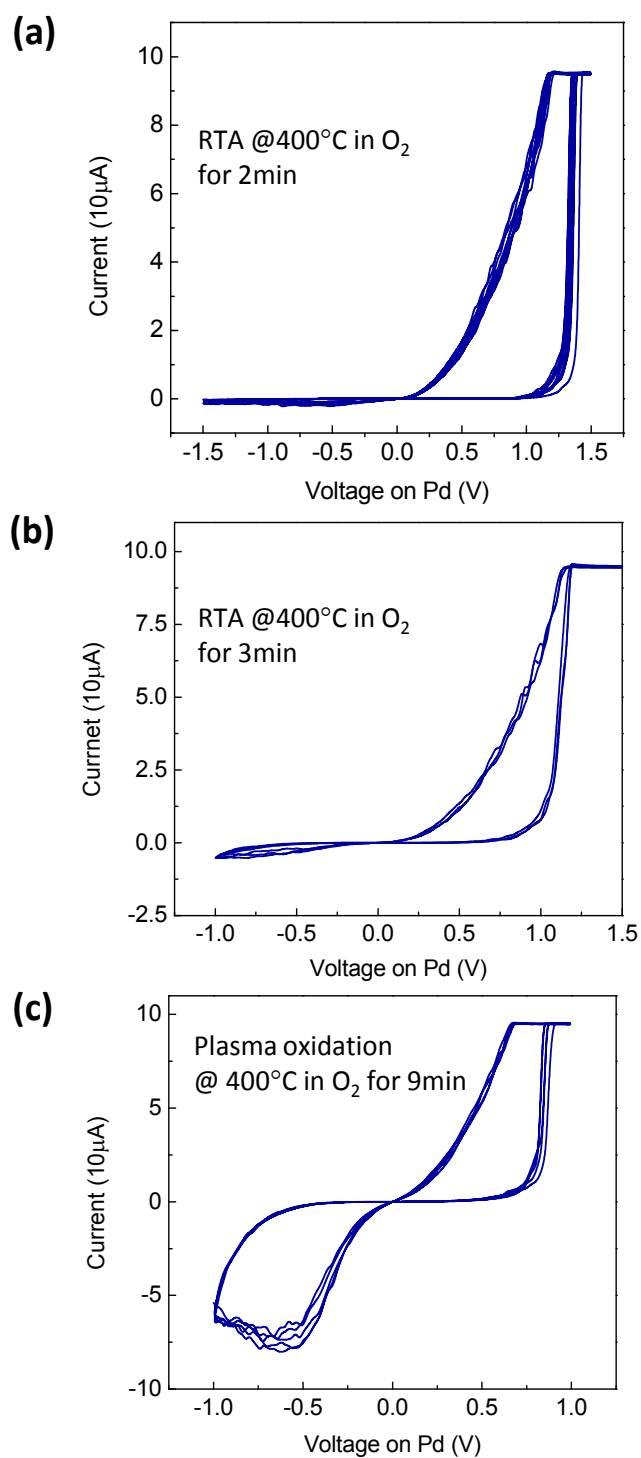
Figure 6.2. Typical resistance switching characteristics of Pd/WO$_{3-x}$/W devices. Tungsten oxide was formed by (a) rapid thermal anneal in pure O$_2$ for 2min at 400°C, (b) for 3min and (c) by plasma oxidation by using O$_2$ for 9min at 400°C.

The resistance switching behavior may be more directly related with the stoichiometry of the tungsten oxide layer at given fabrication process. Even though the exact mechanism still needs to be investigated, some clues can be found in the literature. It has been reported that longer oxidation of tungsten at temperatures of 300°C or higher temperature yields larger W^{5+} to W^{6+} ratio [16]. In other words, the tungsten oxide formed at higher temperature (or longer oxidation time) is not stoichiometric but with high oxygen deficiency. The increased W^{5+} to W^{6+} ratio (hence increased oxygen deficiency) with oxidation time or temperature may be explained by the fact that the as-grown amorphous phase of tungsten oxide is changed to a (poly) crystalline phase in which oxygen vacancies segregate at the grain boundaries [16, 17], while oxygen vacancies within grains may be reduced. Hence the resistance switching may be governed by the grain boundary where resistance is smaller than the bulk and the drift of oxygen vacancies is accelerated as compared to the bulk tungsten oxide.

During the tungsten oxidation at a temperature above 300°C, the density of oxygen vacancies in grain boundaries near the surface may be larger than that near the interface between tungsten oxide and the bottom tungsten electrode. The reason is that the surface of patterned W bottom electrode was cleaned by RIE etching before the oxidation process, which can cause surface roughness and damage to tungsten surface so that the oxidation speed is even accelerated at the beginning of oxidation and the oxidation slows down as the thickness of oxide grows and the non-damaged inner tungsten parts involves in the oxidation. Furthermore, the surface of tungsten oxide experienced the longest annealing time, since the surface of tungsten oxide is the part that grown first and the inner part of the tungsten oxide layer is close to its as-grown state.

Hence the density of defects (oxygen vacancies) in grain boundaries at surface may be higher than that of the inside part and this generates a oxygen vacancy gradient distribution along the tungsten oxide depth direction, enabling bipolar switching. Note that applying positive (negative) bias on the top electrode reduces (increases) resistance. During programming, the oxygen vacancies will move to the bottom electrode side under positive bias on the top Pd (inert) electrode, increasing the overall conductance. Negative bias, on contrary, would cause the oxygen vacancies to move back to the surface region, resulting in the high resistance state. Hence unlike the chemical redox process based tungsten oxide switches which are caused by the motion of metal ions, the resistance switching here is caused by the motion of oxygen vacancies, and the requirement of special top electrode materials such as Cu and Ag for the redox process is no longer needed.

The oxygen vacancy gradient and the absolute density of oxygen vacancies would be higher at longer oxidation time which should lead to more stable bipolar switching. This hypothesis is supported by the experimental results shown in figure 6.2 (c). Low oxygen vacancy density and gradient would yield weak local conducting path which can be interrupted at small negative bias. In this picture, different transition behaviors from off- to on-state and on- to off-state can be also explained for figure 6.2 (c), where oxygen vacancies may be densely packed. For off to on transition, oxygen vacancies can instantly move from top to bottom electrodes freely under positive bias. However for the on to off transition, oxygen vacancies in upper tungsten oxide moves further toward the top electrode followed by movement of oxygen vacancies in lower tungsten oxide toward the top electrode (sequential movement of oxygen vacancies hence involving transition

delay). Hence the write process can occur abruptly, but the erase process involves delay. However, more studies are still necessary to elucidate the switching mechanism, since the switching may be also explained by a Schottky diode effect.

In general, the grain boundaries are considered as short circuits in metal oxides due to fast diffusion of charges along them [18], hence they may govern diffusion related effects in materials such as ionic conductions. For example, in nickel oxide (NiO), vacancies migrate most of the time along the grain boundaries by jumping to the nearest-neighbor atomic place [18]. As a result, nanoscale filament formation was observed in Ni oxide based resistance switches [19].

Similarly, oxygen vacancies are expected to move continuously in atomic scale in the grain boundaries of tungsten oxide and the length of the conductive region can be controlled by controlling the applied programming pulse. For example, short programming pulses would results in small (continuous) conductance change (the memristor behavior). As shown in figure 6.3, conductance modulation in an analog manner was observed when short pulses are applied. It is expected that the overall conductance change depends on the time interval of the applied pulses so that average conductance is unchanged in figure 6.3 (a) for 1.4V/10μs write and -1.3V/10μs erase pulse sequences while the overall conductance increases when erase pulses was reduced to -1.2V/10μs (figure 6.3(b)).
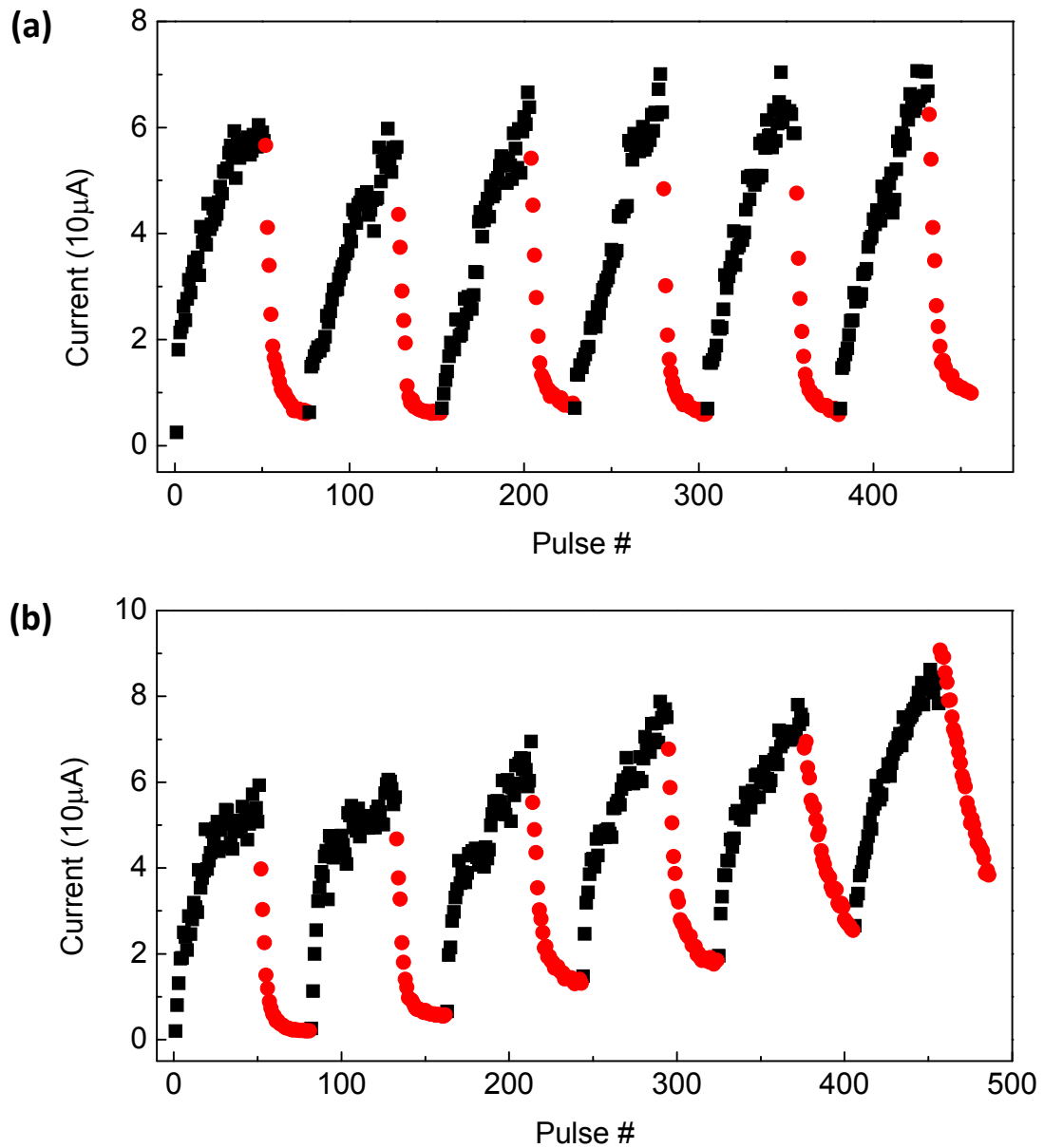
Figure 6.3. Pulse response of a tungsten oxide resistive device of which tungsten oxide was formed by RTP in O_2 for 3min at 400°C. The device was read at 0.4V after each positive (P: black- quadrangle) or negative (N: red-circle) pulse. The device changes the conductance continuously after each pulse. (a) P: 1.4V/10µs, N: -1.3V/10µs. (b) P: 1.4V/10µs, N: -1.2V/10µs. Decreasing the magnitude of the negative pulse results in an overall conductance increase of the device. Device size: 80nm×80nm.

Note that, if applied bias is large, even short pulse can results in sudden changes in the resistance due to enhanced drift (Figure 6.4) at high electric field.
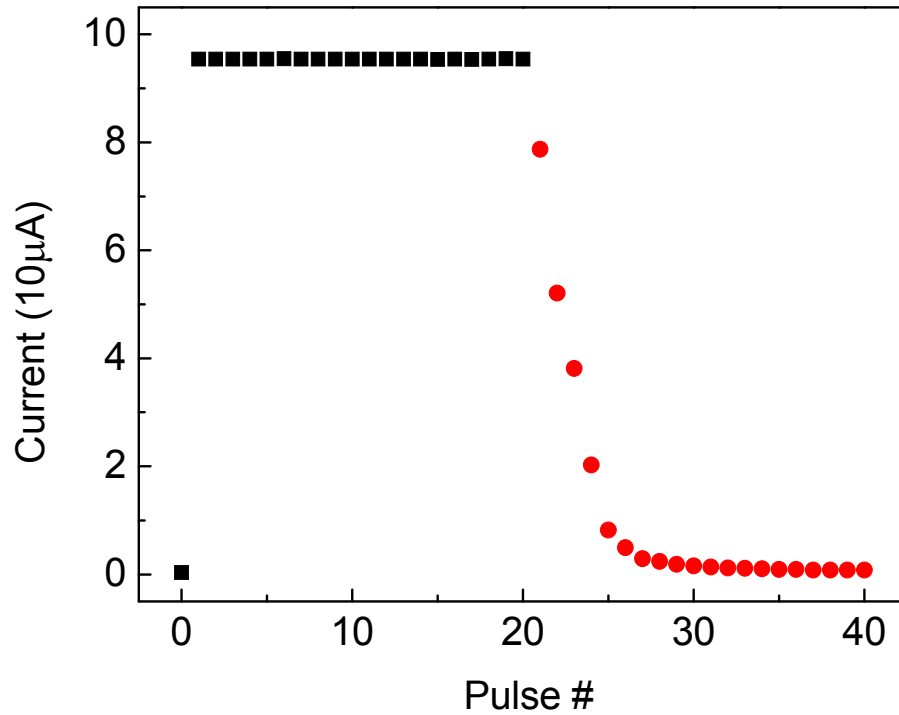
Figure 6.4 Pulse response of the device used in figure 6.3. The device was read at 0.4V after each positive (P: black- quadrangle) or negative (N: red-circle) pulse. P: 1.5V/30μs, N: -1.5V/30μs. The pulse sequence '0' is the initial device state.

The stability of the device was also studied. Figure 6.5 shows the retention property of the device. The device was initially written into the highest conductance state and the device resistance was monitored periodically. Interestingly, the decrease in current (information lost) saturates over the time after an initial fast decay (Note that the Y axis is in linear and the X axis is in log scale). The device may be useful in a neuromorphic system in which learning process happens periodically with short time

interval. Even though more extensive study is still required to elucidate the exact nature of the switching behavior, the device operation can be reasonably explained by the drift of oxygen vacancies in tungsten oxide.
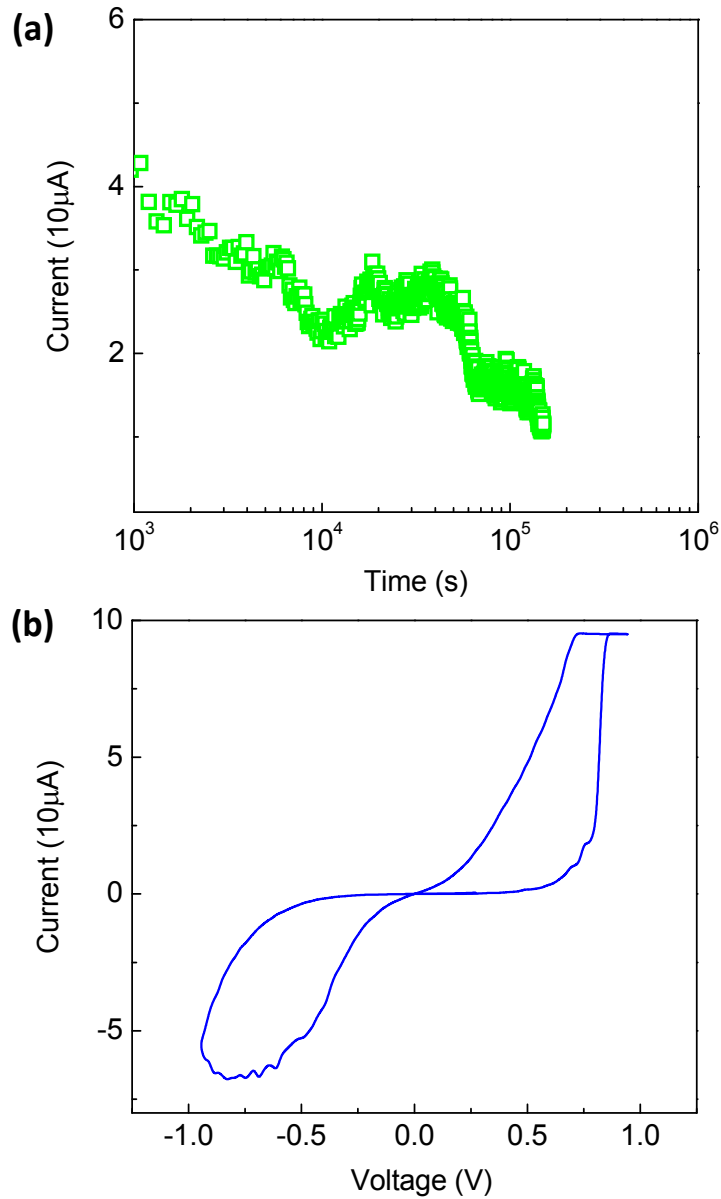
Figure 6.5. (a) The retention property of the tungsten oxide (plasma oxidation 9min at 400°C) device. The current was measured at 0.4V. (b) IV characteristic of the device used in the retention test. Device size: 80nm x 80nm.

6.3 Conclusion

We have shown resistance switching properties of tungsten oxide based nanoscale switches. Potentially, the device can offer both analog and digital switching behaviors by controlling the magnitude and duration of programming pulses. The switching may be explained by oxygen vacancy drift under applied electric field and the device does not require specific top electrodes such as Ag or Cu for redox process, thus giving flexibility in fabrication. Both rectifying and non-rectifying switching behaviors can be obtained. However, unlike a-Si RRAM whose rectifying switching is robust, the rectifying switching in tungsten oxide has limited data retention. The on-state resistance is generally lower than that of a-Si RRAM so that the device may consume larger power than a-Si devices. However, the tungsten oxide based switches can in potential be readily implemented in CMOS circuits with a few additional process steps (e.g. oxidation of tungsten vias) and the state-of-the-art transistors can be used to operate the device due to the low programming voltage (~1V), hence minimizing the device area for the 1T1R or CMOL structure.

References

[1]   K. Miyake, H. Kaneko, M. Sano, N. Suedomi, "Physical and electrochromic properties of the amorphous and crystalline tungsten oxide thick films prepared under reducing atmosphere," *J. Appl. Phys.* **55**, 2747–2753, 1984.

[2]   E. Ozkan, S.-H. Lee, C. E. Tracy, J. R. Pitts, S. K. Deb, "Comparison of electrochromic amorphous and crystalline tungsten oxide films," *Solar Energy Mater. Solar Cells* **79**, 439–448, 2003.

[3]   C. O. Avellaneda, L. O. S. Bulhoes, "Intercalation in WO_3 and WO_3:Li films," *Solid State Ionics.* **165**, 59–64, 2003.

[4]   A. Antonaia, M. C. Santoro, G. Fameli, T. Polichetti, "Transport mechanism and IR structural characterization of evaporated amorphous WO_3 films," *Thin Solid Films* **426**, 281–287, 2003.

[5]   M. Gillet, K. Aguir, C. Lemire, E. Gillet, K. Schierbaum, "The structure and electrical conductivity of vacuum-annealed WO_3 thin films," *Thin Solid Films* **467**, 239–246, 2004.

[6]   C. Bechinger, S. Herminghaus, P. Leiderer, "Photoinduced doping of thin amorphous WO_3 films," *Thin Solid Films* **239**, 156–160, 1994.

[7]   R. C. Agrawal, M. L. Verma, R. K. Gupta, "Electrical and electrochemical properties of a new silver tungstate glass system: x[0.75AgI: 0.25AgCl]: (1−x)[Ag_2O: WO_3]," *Solid State Ionics* **171**, 199–205, 2004.

[8]   A. Monteiro, M. F. Costa, B. Almeida, V. Teixeira, J. Gago, E. Roman, "Structural and optical characterization of WO_3 deposited on glass and ITO," *Vacuum* **64**, 287-291, 2002.

[9]   L. Lozzi, L. Ottaviano, H. Passacantando, S. Santucci, C. Cantalini, "The influence of air and vacuum thermal treatments on the NO_2 gas sensitivity of WO_3 thin films prepared by thermal evaporation," *Thin Solid Films* **391**, 224-228, 2001.

[10]   A.D. Kuypers, C.I.M.A. Spee, J.L. Linden, G. Kirchner, J.F. Forsyth, A. Mackor, "Plasma-enhanced CVD of electrochromic materials," *Surf. Coat. Tech.* **74/75**, 1033-1037, 1995.

[11]   M. Tong, G. Dai, D. Gao, "WO_3 thin film sensor prepared by sol–gel technique and its low-temperature sensing properties to trimethylamine," *Mater. Chem. Phys.* **69**, 176-179, 2001.

[12]   M. Tong, G. Dai, D. Gao, "Electrical and optical properties of WO_3 thin films," *Thin Solid Films* **397**, 238-243, 2001

[13]   C. Kügeler *et al.*, *IEEE NANO* 2009.

[14]   M. N. Kozicki, C. Gopalan, M. alakrishnan, M. Mitkova, "A Low-Power Nonvolatile Switching Element Based on Copper-Tungsten Oxide Solid Electrolyte," *IEEE Trans. Nanotechnol.* **5**, 535, 2006.

[15] C. Ho E. K. Lai, M. D. Lee, C. L. Pan, Y. D. Yao, K. Y. Hsieh, R. Liu, C. Y. Lu, "A Highly Reliable Self-Aligned Graded Oxide WO_x Resistance Memory: Conduction Mechanisms and Reliability," *IEEE VLSI Tech*. *Digest* 228, 2007.

[16] R. Sohal, C. Walczyk, P. Zaumseil, D. Wolansky, A. Fox, B. Tillack, H.-J. Müssig, T. Schroeder, "Thermal oxidation of chemical vapour deposited tungsten layers on silicon substrates for embedded non-volatile memory application," *Thin Solid Films* **517**, 4534-4539, 2009.

[17] C. Bittencourt, R. Landers, E. Llobet, X. Correig, J. Calderer, "The role of oxygen partial pressure and annealing temperature on the formation of W = O bonds in thin WO_3 films," *Semicond*. *Sci*. *Technol*. **17**, 522, 2002.

[18] T. Karakasidis, M. Meyer, "Grain-boundary diffusion of cation vacancies in nickel oxide:mA molecular-dynamics study," *Phys*. *Rev*. *B*. **55** ,13853-13864, 1997.

[19] M.-J. Lee, S. Han, S. H. Jeon, B. H. Park, B. S. Kang *et al*., "Electrical Manipulation of Nanofilaments in Transition-Metal Oxides for Resistance-Based Memory," *Nano Lett*. **9**, 1476-1481, 2009.

Chapter 7

Rectifying Switching by Utilizing Stacked Breakdown Elements

7.1 Introduction

As discussed in Chapters 4 and 5, the crossbar structure maximizes the density of two terminal resistive switching elements, and can be down scaled beyond the transistor scaling limit. Various materials have been studied as the switching element and shown some promising aspects for nonvolatile memory application in terms of scalability, switching speed, programming endurance and data retention [1-7]. However, even though several promising nonvolatile resistance elements have been demonstrated up to now, very few studies have shown successful demonstrations of the integration of those switching devices in the crossbar form. To date, most studies are still in the "proof-of-concept" stage. Besides reproducibility and reliability issues, a key challenge for crossbar based memory and logic is crosstalk between the two-terminal devices. In other words, the main advantages of the crossbar structure, high device density and large connectivity, ironically impose significant challenges for its practical implementation due to potential leakage current through the large, interconnected network. For example, an incorrect high read current may be obtained when trying to read the state of a target cell due to sneak current paths through three neighboring cells that are in the on-state (low resistance state), as shown in Fig. 7.1. The shortest sneak path contains two forward biased and one reverse biased cells. This problem is most severe if the resistance switching element at

on-state behaves as a linear resistor with the same conductance in both forward and reverse-based cases. To suppress the sneak path, current "steering" effects will have to be incorporated, for example, by suppressing current for the reverse-based cell. One approach is to incorporate diodes at each junction (Fig. 7.2), forming the so-called 1D1R structure (one-diode one-resistive element). It was demonstrated that metal oxide based Schottky diodes or PN diodes can be built in series to each switching elements [8, 9]. However, in principle adding an extrinsic diode is only applicable to devices showing unipolar switching behavior which can be programmed to '1' or '0' with the same polarity (e.g. positive polarity during which the diode remains positively biased during programming). Since the unipolar switching devices are based on Joule heating during the programming they typically require orders of magnitude higher power than bipolar switches whose operation are based on electric field. Unfortunately, the operations of bipolar switches are inherently not compatible with the 1D1R structure. For example, when the resistive device needs to be turned off and a negative biased is applied, the diode will be reverse-biased and will consume most of the applied voltage instead, thus preventing the device from being programmed. In addition, besides performance and integration issues, it has been noted that nanoscale diodes may lose their rectifying behavior below the 50 nm range due to enhanced tunneling current [10].

In Chapter 2 ~ 4, we have shown that in two-terminal switches based on a-Si an interesting intrinsic rectifying behavior (Fig. 7.3) can be observed, in which the on-state devices shows a diode-like behavior, i.e. the current is extremely non-linear and can be greatly suppressed at reverse-bias without the addition of extrinsic current steering elements. The intrinsic rectifying switching can thus break the potential sneak paths and

is very desirable for the integration of two-terminal switches in the crossbar form. However, the intrinsic rectifying behavior is rarely observed in other bipolar switching elements. Here we propose and demonstrate a new switching structure which can greatly suppress the current at small reverse biases and may be applied to any two-terminal switching devices.
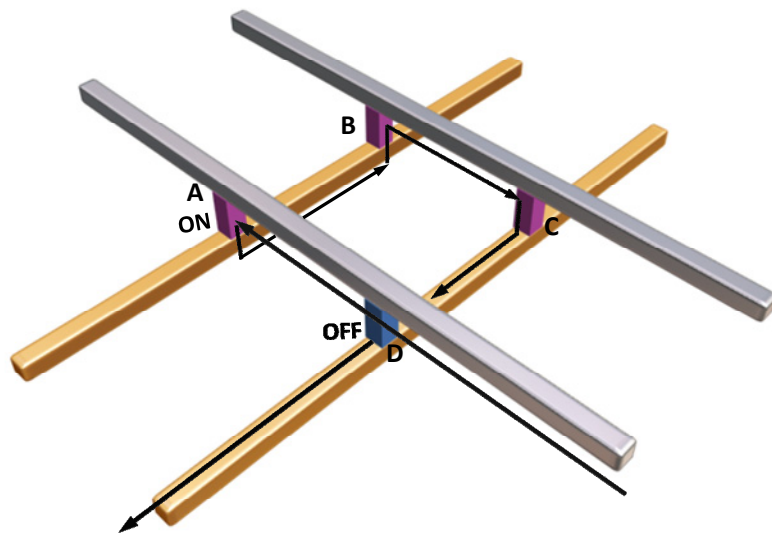
Figure 7.1. An example of the sneak path in a crossbar array through cells A-C (purple) as highlighted by the arrows. As a result, an incorrect high read current is obtained when trying to read the state for cell D (blue).
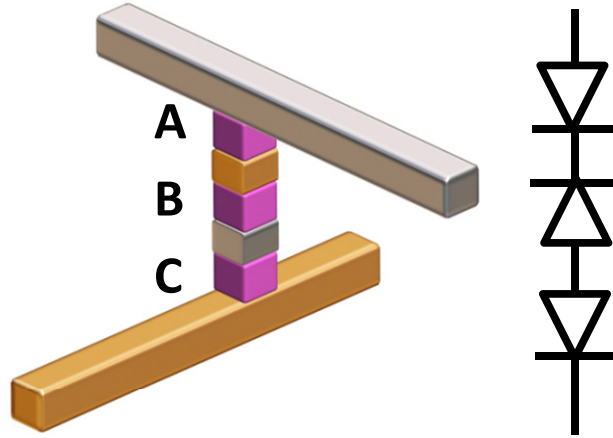
Figure 7.2. The sneak path can be blocked by using rectifying switching elements. A, B and C stand for on-state cells along the sneak path as shown in Fig. 6.1.
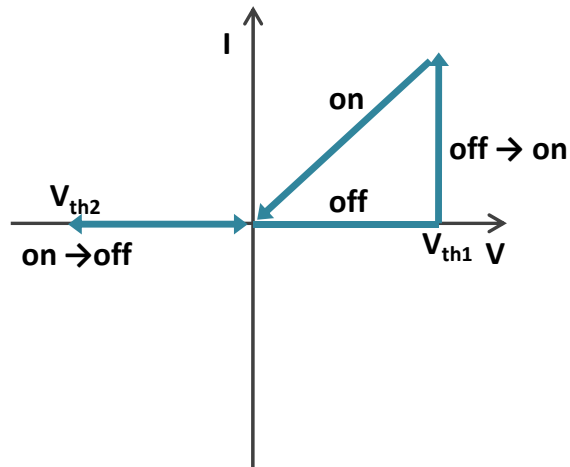
Figure 7.3. Schematic of the intrinsic rectifying switching observed in a-Si switches.

7.2 Current Suppressing with Breakdown

The suppression of current at small voltages in our approach relies on a common phenomenon observed in thin films - (soft) breakdown. In particular, the breakdown (BD)

1 1 9

layer can be stacked directly with the resistance switching (RS) layer, (Fig. 7.4) thus maintaining the high packing density and reducing process complexity.
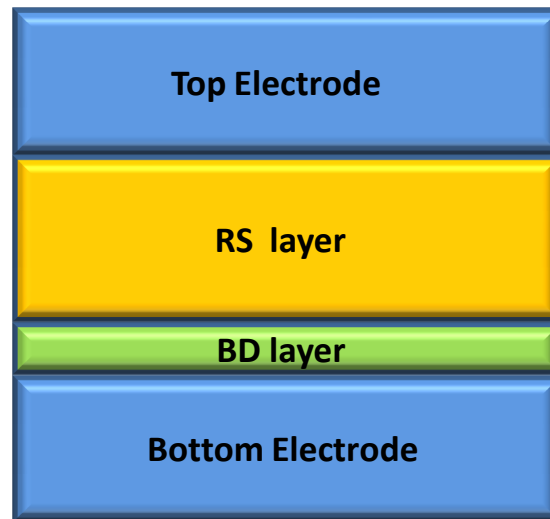
Figure 7.4. Schematic of the stacked device structure with a resistance switching layer and a breakdown layer

Electrical breakdown such as gate dielectric breakdown in CMOS is typically considered as an undesired event and it typically induces damage to the system. However, in some cases such as impact ionization MOFETs (I-MOS) [12-15], electrical breakdown is intentionally generated to boost the device performance. In a RRAM cell, interestingly, we show that if a thin insulating layer is integrated to the switching element, the switching device can show current suppressing capability by utilizing soft breakdown property of the insulator (Figure 7.5).

It is worthwhile to compare breakdown (BD) and nonvolatile resistance switching (RS) before discussing the system that combines these two effects together. In fact, RS share many properties with BD and may in some sense be considered a special case of

BD in which the switching is nonvolatile, since BD is universally observed in all dielectrics, while RS is reported in some materials such as amorphous silicon and metal oxides. In nanometer scale devices, the classical picture of BD (e.g. impact ionization) is not valid due to enhanced effect of defects (e.g. resonant tunneling via traps) in thin dielectrics [16, 17], causing breakdown at lower voltage. Hence, dielectrics without defects are preferred for good insulators whereas moderate defects in a switching medium stimulate RS. It is generally accepted that BD is a localized effect and it was estimated that BD occurs within a few nanometer range in SiO_2 [16, 18, 19]. Similarly, RS is typically a localized effect [20]. Although the exact mechanisms for BD and RS are still a matter of debate, in general, both mechanisms are composed of three steps; degradation (forming) of insulators, onset of breakdown (resistance switching) and thermal effects caused by the breakdown (switching). In BD, the reversibility is critically dependent on the thermal effects [16, 21] such that the (soft) BD would be a reversible process only if current is limited during the BD. Especially, if the oxide is very thin (< 2nm), the breakdown is highly reversible at low voltage (~1V) [22] by reverting the percolation path (conduction path). Similarly, during RS if current compliance is not used the device may be damaged due to runaway thermal effects, and it was demonstrated that smaller current gives large repeatability of RS (programming endurance) [23]. However, a key difference between BD and RS is that it is generally believed that RS involves the motion and storage of atoms (ions) while BD involves the motion and storage of elemental charges (electrons and hole). Since the storage of atoms in defects is much easier than that for electrons or holes due to the heavier mass of atoms, RS effects typically are accompanied with much longer retention times compared with BD effects. A related

difference is that BD processes are typically unipolar, i.e. can occur at both polarities, while many RS processes are bipolar, i.e. requiring positive bias to turn on and negative bias to turn off.

Figure 7.5 (a) shows the I-V characteristics of a typical bipolar resistance switching event and Figure 7.5 (b) presents the I-V of a typical breakdown event of an insulating layer. Unlike resistance switching, (soft) breakdown is volatile that the increased conductance effect vanishes once power is removed from the device or if the voltage across the insulating layer is reduced (far below the breakdown voltage of the insulating layer, V_{th3} and V_{th4}). Assuming the threshold voltage of RS is larger than that of BD ($V_{th1} \approx |V_{th2}| > V_{th3} \approx |V_{th4}|$, a condition that can be matched by proper engineering the BD and RS layers), there are four possible IV characteristics (Fig. 7.5 (c)-(f)) when RS and BD elements are integrated together in series into a device shown in Fig. 7.3. When the RS element is initially at off-state, applying positive voltage larger than RS threshold voltage ($>V_{th1}$) will cause BD and RS (see the next section for details), making the device at on-state (Fig. 7.5 (c)). To perform read, a positive voltage larger than V_{th3} (but smaller than V_{th1}) will cause BD of the insulator, allowing read process for on-state (Fig. 7.5 (e)). When negative bias (with amplitude $> |V_{th3}|$) is applied to the device (currently in the on-state), BD would occur and the voltage will now mostly drop across the RS elements. If the voltage becomes larger than V_{th2}, the device will be turned-off (Fig. 7.5 (f)). In this case, the same read voltage will produce very small current, due to the high resistance of the RS element and the low voltage across the BD dielectric. Applying negative bias with amplitude $< V_{th4}$ will not cause any change in the device state (Fig. 7.5 (d)) since the RS element is now at off-state. The details for programming

and read voltages will be discussed in the next section. Note that the current will be low

when a negative bias smaller than V_{th4} is applied regardless of the state of the RS device,

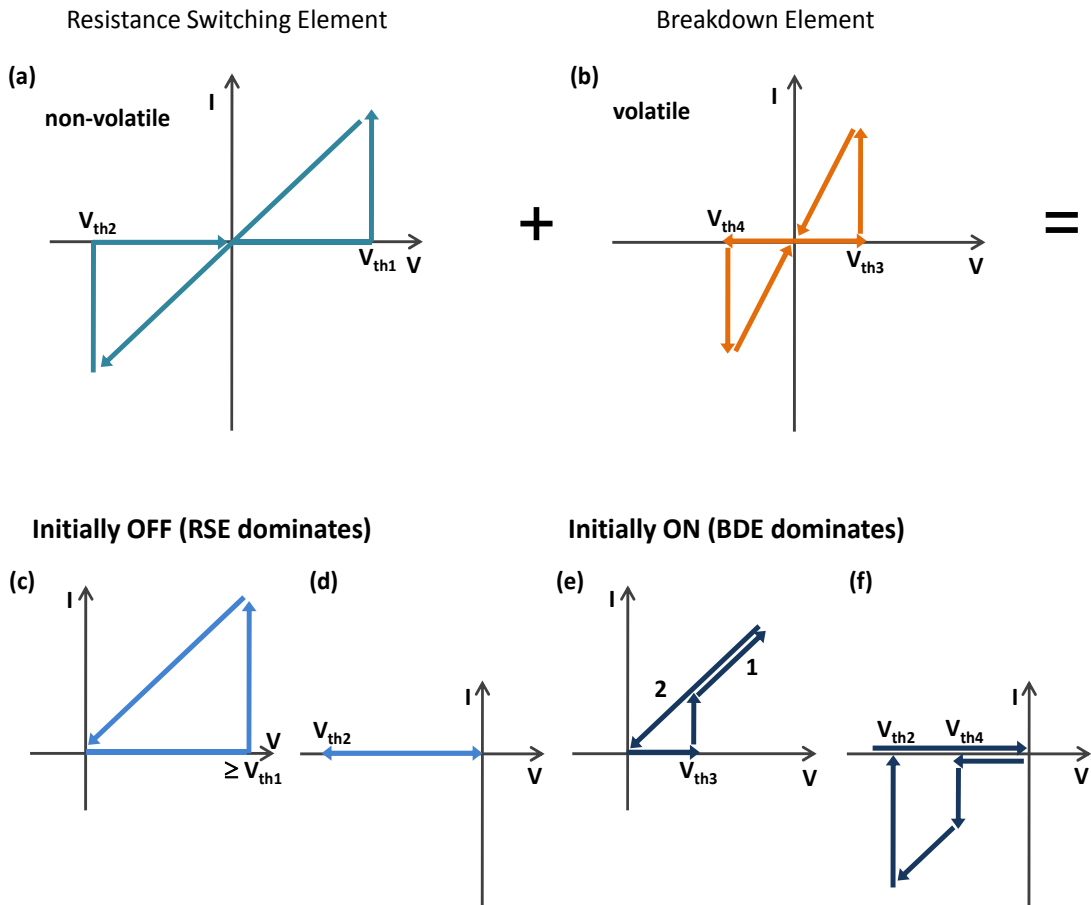thus restricting the sneak current.

Figure 7.5. Combined effects of nonvolatile resistance switching and breakdown. (a) Typical bipolar resistance switching. (b) (Soft) breakdown of an insulator. (c) During write. (d) Off-state. (e) During read. (f) During erase. Note that the leak current is always suppressed at negative bias ($0 - V_{th4}$) which can eliminate the sneak path in the crossbar structure.

To verify this hypothesis, we have fabricated Ag/a-Si/W/HfO$_2$/W devices where Ag, a-Si, HfO$_2$ and W are a top electrode, a switching medium, a breakdown layer and a bottom electrode respectively. The middle W layer is inserted to the device so that the breakdown layer is well defined and isolated. Tungsten was deposited by sputtering at room temperature followed by e-beam lithography and RIE (Reactive Ion Etching) for the bottom electrode formation. The thin HfO$_2$ BD layer was deposited by ALD (Atomic Layer Deposition) at 300°C and the a-Si RS layer was deposited by PECVD (Plasma Enhanced Chemical Vapor Deposition) at 300°C. Finally, another e-beam lithography step and Ag lift-off process completed the device fabrication. The overall device size is 120nm × 120nm. As discuss in Chapters 3 - 5, the resistance of amorphous silicon layer can be altered by the injection metal particles from the top electrode, forming a nanoscale filament. This RS element is now integrated with the HfO$_2$ BD layer. The device schematic and experimentally obtained IV curve are shown in figure 7.6. As expected, the device indeed shows rectifying switching. Four difference cases (write, read, erase and off-state check) follow exactly the hypothesis presented in figure 7.5. In addition, the BD voltages Vth$_3$ and Vth$_4$ are almost identical, consistent with the fact that the BD layer has a symmetric device structure with W top and bottom electrodes.
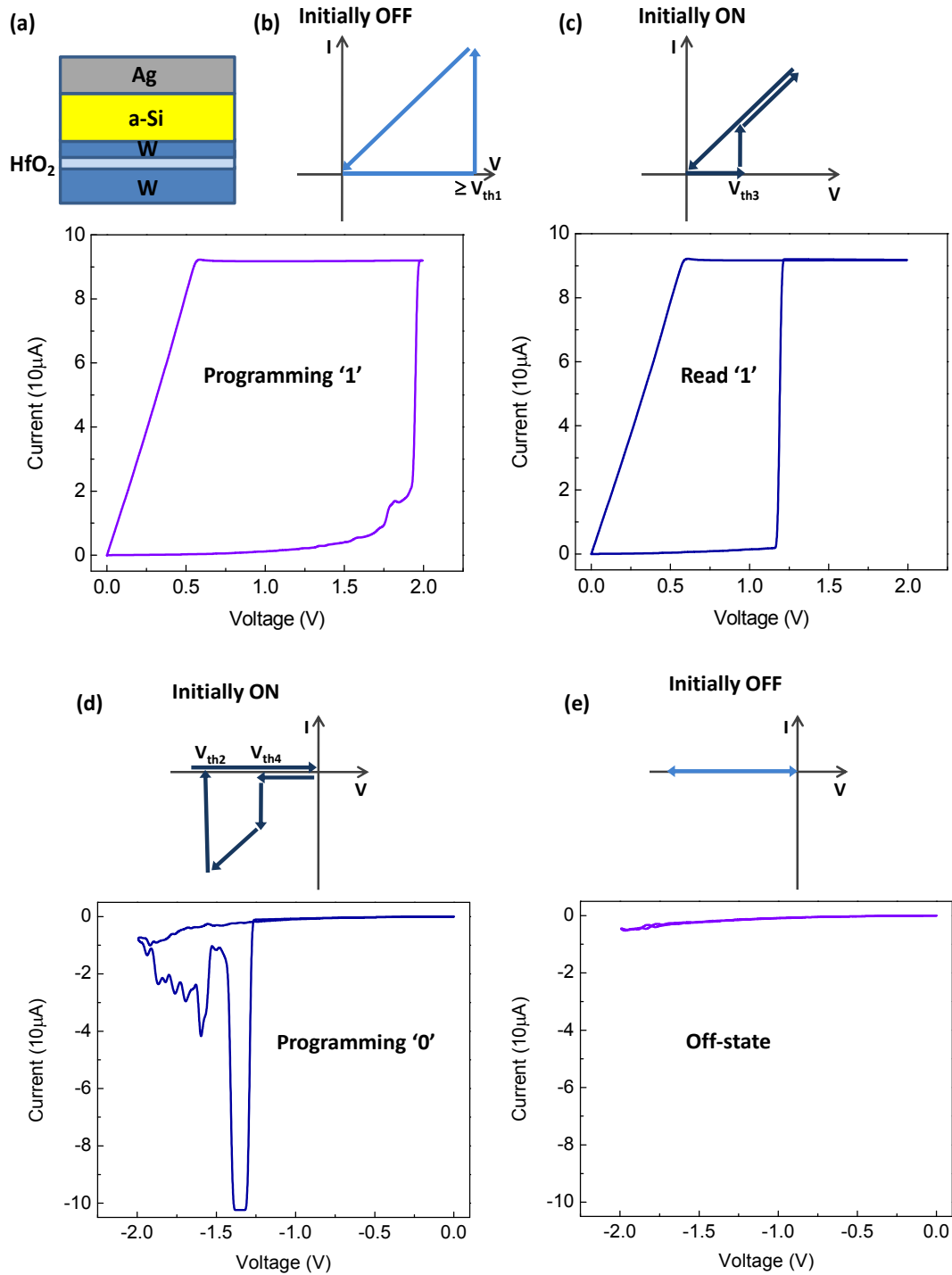
Figure 7.6. Experimentally obtained rectifying switching from the combined structure of resistance switching and breakdown elements. (a) Device structure. (b) Write process. (c) BD enables read process. (d) Erase process. (e) The off-state device does not respond to negative bias.

The high current in these devices are mostly attributed to the low-Ron in the M/a-Si/M RS device. As we have shown earlier (Figure 4.5), in an M/a-Si/M structure the on-state current is typically much higher than that of the M/a-Si/p-Si structure, possible due to high defect density between the a-Si layer and the bottom electrode. To reduce the programming current, control devices were fabricated with the middle tungsten layer removed and all other device fabrication processes remained the same. In this structure, the breakdown device is formed by the HfO_2 layer, the Ag filament inside the a-Si layer as the top electrode (when the RS device is in the on-state) and the bottom W layer. Hence, in this case, write (including the first forming process) voltage need to be small (typically smaller than 3V) to avoid possible direct metal particle injection into the HfO_2 layer during the RS process. As shown in figure 7.7, the current levels can be significantly lowered in the Ag/a-Si/HfO_2/W devices compared to those in the Ag/a-Si/W/HfO_2/W devices. Furthermore, a slight reduced breakdown voltage for read process (at positive bias) and slight increased breakdown voltage for erase process (at negative bias) were observed as compared to the Ag/a-Si/W/HfO_2/W device. This asymmetry for Vth_3 and Vth_4 can be explained by the work function difference of Ag (4.26eV) and W (4.55eV) for the BD layer - the smaller work-function of Ag than that of W would cause a built-in field in the BD layer and leads to smaller BD voltage at positive bias compared to that at negative bias.
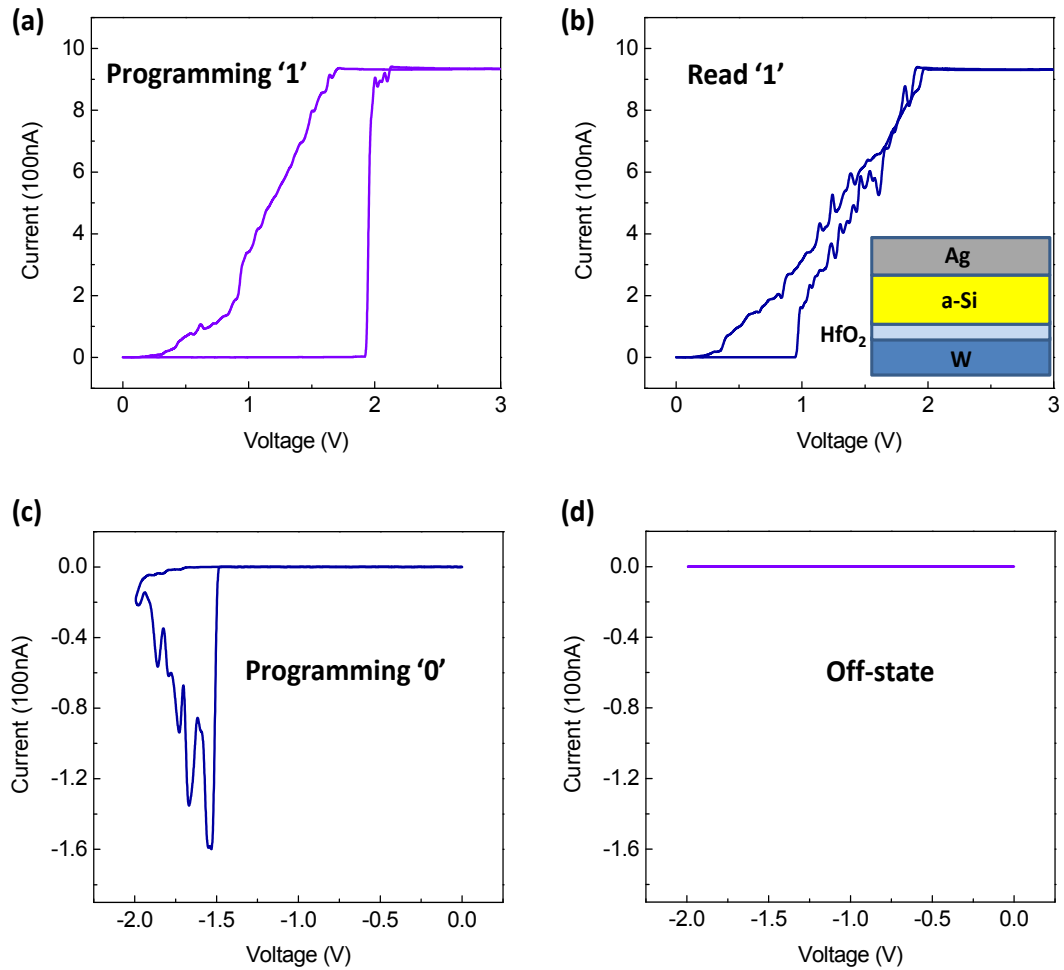
Figure 7.7. Current suppression with a breakdown layer in the Ag/a-Si/HfO2/W system.

7.3 Read/Programming Voltages

The shortest sneak paths in the crossbar structure are composed of three cross-points all of which are at on-state (Fig.7.1 & 7.2). For successful read process, the read voltage should be large enough to generate BD in the target cell, but small enough to not cause BD in the sneak paths or accidental programming of the target cell. In other words,

$$V_{BD} < V_{read} < \min(3V_{BD}, V_{write})$$

where V_{BD} is the breakdown voltage of the insulator and V_{write} stands for the write voltage of the device. Note that the resistance of the sneak path in the worst case scenario (when all the devices along the sneak path are in the on-state) during the read process is now $3R_{insulator}$ instead of $3R_{on}$, which can provide orders of magnitude reduction of the leakage current compared with devices without the BD layers for reasonable BR layers with $R_{insulator} \gg R_{on}$.

As for the programming voltage, adding the BD layer to the RS element almost does not require additional voltage overhead (Table 7-1). This is one of important advantages of using the BD layer for rectification. When the insulator resistance is much larger than off-state of the RS element, applied voltage will drop across the insulator. Hence applying threshold voltage of programming '1' of the RS element will cause the breakdown followed by the programming of the RS element. For the case when $R_{insulator} \approx R_{off}$, voltage dividing effect dictates that programming voltage should not be smaller than $\max(2V_{BD}, V_{RS\text{-}Write})$ for successful programming. An interesting case is when $R_{insulator} \ll R_{off}$, then RS will be initiated earlier than breakdown. However since $R_{on} \ll R_{insulator}$, breakdown will still proceed after (or during) RS programming. Hence programming voltage larger than $\max(2V_{BD}, V_{RS\text{-}Write})$ will always ensure the breakdown of the insulator and complete the writing process. Importantly, if the insulator is engineered to satisfy $2V_{BD} \leq V_{RS\text{-}Write}$, there is no voltage overhead required during the programming process.

For erasing, since $R_{insulator} \gg R_{on}$, BR will occur when the applied voltage $> |V_{BD}|$, so erasing can proceed successfully as long as $|V_{RS\text{-}Erase}| > |V_{BD}|$.

In summary, no programming voltage overhead will be required in the integrated RS/BD device structure if the insulator can be engineered to have $|V_{BD}| <$ min($|V_{RS\text{-}Write}/2|$, $|V_{RS\text{-}Erase}|$).

	Case	Programming Voltage	Note
1	$R_{insulator} \gg R_{off}$	$V_{RS\text{-}Write}$	Breakdown then programming
2	$R_{insulator} \ll R_{off}$	max ($2V_{BD}$, $V_{RS\text{-}Write}$)	Programming then breakdown
3	$R_{insulator} \approx R_{off}$	max ($2V_{BD}$, $V_{RS\text{-}Write}$)	Breakdown then programming
4	$R_{insulator} \gg R_{on}$	max (V_{BD}, $V_{RS\text{-}Erase}$)	Breakdown then erasing

($V_{RS\text{-}Write}$: write threshold voltage for the RS element without the BD layer, $V_{RS\text{-}Erase}$, erase threshold voltage for the RS element without the BD layer, V_{BD}: breakdown voltage of the insulator.)

Table 7-1. Programming voltage of the combined RS and BD device.

Finally, case 1 of the table 7-1 can have another important benefit. Resistance switching devices based on materials such as organics suffer from low off/on resistance ratio (a few tens - hundreds) due to low R_{off}, limiting the switching materials application prospect even though the materials may possess excellent potential for down scaling. However, by utilizing the BD layer, the off-state current during read (or the "computing" phase of memristor-based logic circuits) will be limited by the resistance of the BD layer and the low on-state resistance can still be sensed by generating BD. This approach may thus allow several otherwise non-ideal switching materials for practical crossbar-based memory or logic. A good example is when each cell has a limited connectivity in systems such as CNN (cellular neural networks) [24] which are very useful for high-speed parallel signal processing applications (e.g. image processing).

7.4 Conclusion

Traditionally, breakdown of an insulator and nonvolatile resistance switching have been pursued in opposite directions (e.g. low-defect density, high V_{th} for breakdown and moderate-high defect density, low V_{th} for resistance switching). However, by properly engineering the BD and RS layers and stacking the two types of two-terminal devices together, current suppression can be achieved in the device characteristics to block potential sneak paths in high density crossbars with no increase in cell size, minimal fabrication complexity, and little or no increase in programming voltages.

References

[1]   H. Goronkin, Y. Yang, "High-Performance Emerging Solid-State Memory Technologies," *MRS Bull.* **29**, 805-808, 2004.

[2]   R. Waser, M. Aono, "Nanoionics-based resistive switching memories," *Nature Mater.* **6**, 833-840, 2007.

[3]   B. J. Choi, D. S. Jeong, S. K. Kim, C. J. Rohde, "Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition," *J. Appl. Phys.* **98**, 033715, 2005.

[4]   K. Terabe, T. Hasegawa, T. Nakayama, M. Aono, "Quantized conductance atomic switch," *Nature* **433**, 47-50, 2005.

[5]   D. B. Strukov, G. S. Snider, D. R. Stewart, Williams, R. S. "The missing memristor found," *Nature* **453**, 80-83, 2008.

[6]   J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nature Nanotechnol.* **3**, 429-433, 2008.

[7]   S. H. Jo, W. Lu, "CMOS Compatible Nanoscale Nonvolatile Resistance Switching Memory," *Nano Lett.* **8**, 392-397, 2008.

[8]   M.-J. Lee, S. I. Kim, C. B. Lee, H. Yin, S.-E. Ahn et al., "Low-Temperature-Grown Transition Metal Oxide Based Storage Materials and Oxide Transistors for High-Density Non-volatile Memory," *Adv. Funct. Mater.* **19**, 1587, 2008.

[9]   D. S. Golubović, A. H. Miranda, N. Akil, R. T. F. van Schaijk, and M. J. van Duuren, "Vertical poly-Si select pn-diodes for emerging resistive non-volatile memories," *Microelectron. Eng.* **84**, 2921, 2007.

[10]   G. D. J. Smit, S. Rogge, T. M. Klapwijk, " Scaling of nano-Schottky-diodes," *Appl. Phys. Lett.* **81**, 3852, 2002.

[11]   D. J. Frank *et al.*, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE* **89**, 259-288, 2001.

[12]   M. T. Bjork, O. Hayden, J. Knoch, H. Riel, H. Schmid, W. Riess, "Impact ionization FETs based on silicon nanowires," 65th *Annual Device Research Conference*, 171, 2007.

[13]   U. Abelein et al., *Proc. of the 25th Int. Conf. on Microelectronics (MIEL)* **1**, 127-129, 2006.

[14]   K. Gopalakrishnan, P. B. Griffin, J. D. Plummer, "Impact ionization MOS (I-MOS)-Part I: device and circuit simulations," *IEEE Trans Electron Devices* **52**, 69-76, 2005.

[15]   W. Y. Choi, J. Y. Song, J. D. Lee, Y. J. Park, B.-G. Park, "70-nm impact-ionization metal-oxide-semiconductor (I-MOS) devices integrated with tunneling field-effect transistors (TFETs)," *IEDM Tech. Digest*, 955-958, 2005.

[16]   J. Suñé, E. Farrés, I. Placencia, N. Barniol, F. Martín, X. Aymerich, "Nondestructive multiple breakdown events in very thin SiO_2 films," *Appl. Phys. Lett.* **55**, 128-130, 1989.

[17]   B. Ricco, M. Ya. Azbel, M. H. Brodsky, "Novel Mechanism for Tunneling and Breakdown of Thin SiO_2 Films," *Phys. Rev. Lett.* **51**, 1795-1798, 1983.

[18]   P. Olivo, T. N. Nguyen, B. Riccó, "High-field-induced degradation in ultra-thin SiO_2 films," *IEEE Trans. Electron Devices* **35**, 2259-2267, 1988.

[19]   J. Suñé, I. Placencia, N. Barniol, E. Farrés, F. Martín, X. Aymerich, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films* **185**, 347-362, 1990.

[20]   R. Waser, M, Aono, "Nanoionics-based resistive switching memories," *Nature Mater.* **6**, 833-840, 2007.

[21]   M. Nafría, J. Sué, X. Aymerich, "Exploratory observations of post-breakdown conduction in polycrystalline-silicon and metal-gated thin-oxide metal-oxide-semiconductor capacitors," *J. Appl. Phys.* **73**, 205-215, 1993.

[22]   K. P. Cheung, "Ultrathin gate-oxide breakdown-reversibility at low voltage," *IEEE TRANS. Device and Material Reliability* **6**, 67-74, 2006

[23]   K.-H. Kim, S. H. Jo, S. Gaba, W. Lu. "Nanoscale resistive memory with intrinsic diode characteristics and long endurance," *Appl. Phys. Lett.* **96**, 053106, 2010.

[24]   L. O. Chua, L. Yang, "Cellular neural networks: theory," *IEEE Trans Circuits and Systems* **35**, 1257, 1988.

Chapter 8

Summary and Future Plan

8.1 Amorphous Silicon based Resistive Random Access Memory (RRAM)

In Chapter 2, we reported studies on nanoscale metal/a-Si/p-Si devices in which controlled forming process was achieved by using a heavily doped substrate as the bottom contact material. Devices based on the metal/a-Si/p-Si structure possess a number of desirable performance characteristics such as fast switching speed, large on/off ratio, long retention and endurance. The a-Si layer sandwiched between a metal layer and the p-type silicon substrate can be switched reversibly between the high resistance state and the low resistance state by controlling the magnitude and the polarity of the applied voltage. The devices show improved on/off resistance ratio as the device size is scaled down. Unlike conventional micron-scale metal/a-Si/metal structures, the metal/a-Si/p-Si structure does not require high voltage forming, and high device yield (>98%) can be readily obtained. The switching behavior can be further controlled to be either rectifying or non-rectifying such that at on-state the device can be modeled as either a resistor or a diode, adding functionalities at the circuit level.

8.2 Single Conducting Filament Formation and Probabilistic Switching

In Chapter 3, we demonstrated that the resistance switching in nanoscale two-terminal resistive switches can be dominated by the formation of a single conductive filament. As a result, the switching is stochastic and can be modeled as a Poissonian process. The probabilistic filament formation process strongly affects the device operation principle, and can be programmed to facilitate new functionalities such as multibit switching with partially formed filaments. The device can act as either a digital switch (i.e. memory bits) or an analog switch (i.e. neuromorphic circuit component).

8.3 Crossbar Arrays based on Amorphous Silicon RRAM

In Chapter 4, it was shown that a high-density (2 Gbits/cm^2) 1kb crossbar memory array can be built based on the nanoscale Si two-terminal switches with excellent uniformity, yield (> 92%) and ON/OFF ratio (> 1000). The crossbar array can be implemented as either a resistive random-access-memory (RRAM) or a write-once type memory depending on the device configuration. The thickness of a-Si should be scaled down as the device pitch size is reduced to ensure proper device operations as suggested in 3D electric field simulation. Considering the experimental result that the switching occurs within 10nm range in a-Si, the a-Si based RRAM would be well positioned for ultra-high density crossbar structure.

8.4 Nanoscale Memristors for Synapses in Neuromorphic Systems

In Chapter 5, we demonstrated the experimental implementation of synaptic functions using nanoscale a-Si based memristors. Different from the "digital" devices used in the memory studies, here we built memristors consisting of a co-sputtered Ag and Si layer with a properly designed mixture ratio gradient that forms a front between the Ag-rich (high conductivity) and Ag-poor (low conductivity) regions in the active switch layer. Injection and depletion of ions in the memristor moves the front between the ion rich region and ion-poor region and causes a continuous change in conductance.

The sequential computation nature of modern computer systems represented by von Neumann machines inherently imposes complexity to the system architecture and software design especially where real time computation is required such as in sensory systems. Hence the advance in computer systems has ironically resulted in less efficient architectures as the complexity of environments (inputs) increases. On the other hand, biological systems can perform much more efficient computing for complex tasks due to the largely interconnected neuron network that offers high degree of parallel analog computing capability. In this study we showed that two-terminal memristors can provide the necessary synaptic functions and the connectivity required for hardware implementation of biologically-inspired neuromorphic circuits. In particular, we verified the hybrid memristor/CMOS neuron circuit can exhibit Spike Timing Dependent Plasticity (STDP), an important synaptic adaption rule in which the change of synaptic weight depends on the timing of the pre/post-neurons' spikes. These results suggest a promising approach for hardware neuromorphic circuit implementation based on CMOS neurons and nanoscale memristor synapses.

8.5 Resistance Switching in Nanoscale Tungsten Oxide

In chapter 6, we presented our studies on resistance switching behavior observed in nanoscale tungsten oxide (WO_{3-x}). Tungsten oxide was prepared by either RTA (Rapid Thermal Anneal) in pure oxygen or plasma oxidation in O_2 plasma at 400°C which is still suitable for the back-end-of-line (BEOL) processing in CMOS integrated circuits. The device can be programmed with about ±1V and the switching characteristics are critically related with stoichemetry of tungsten oxide that switching properties (e.g. rectifying or non-rectifying behavior) can be tuned by controlling tungsten oxide preparation methods.

Even though large on-state current and short retention still need to be solved, relatively low process temperature and programming voltage would enable tungsten oxide based devices to be readily implemented with state-of-the-art CMOS circuits in a CMOL form. Furthermore, the device may be useful in a neuromorphic system in which learning process happens periodically with short time interval.

8.6 Rectifying Switching by Utilizing Breakdown

In Chapter 7, we proposed and demonstrated resistive switching elements with a rectifying switching behavior by stacking a resistance switching device with a breakdown device. In particular, unlike the intrinsic rectifying behavior observed in the a-Si devices, the hybrid RS/BD device structure can provide current suppression at small negative bias (current rectifying) with a broad range of materials Using this approach the potential leakage current paths in crossbar arrays can be effectively blocked by since the resistance in the shortest sneak path will be replaced from $3R_{on}$ to $3R_{insulator}$. Furthermore, we

verified that adding a breakdown layer requires little or no voltage overhead, if the insulator is engineered to satisfy $2V_{BD} \leq V_{RS\text{-}Write}$ where V_{BD} is the breakdown voltage of the insulator and $V_{RS\text{-}Write}$ is the write voltage of the resistance switching layer.

8.7 Future Work

8.7.1 Transmission electron microscopy (TEM) Studies of the a-Si Switches

The filament formation picture discussed in Chapters 2-4 consistently explained the behavior of our digital memristive devices. However, it is still based on indirect observation and the filament has not been directly observed in the nanoscale devices. Earlier SEM and TEM studies on microscale devices have only yielded micron scale filaments typically formed after destructive run-away thermal processes [1-4]. In addition to unambiguously confirming the nature of the resistance switching, direct observation of the nanoscale filaments will lead to other valuable information such as the size and chemical state of the Ag particles, the exact location and distribution of the filament inside the a-Si matrix and how the filament is affected by the different material and interfaces.

In this regard, cross-sectional TEM studies would provide direct information about the nature of the Ag filaments and elucidate the resistance switching mechanism in the nanoscale amorphous silicon devices. The key challenge for the TEM study is the sample preparation. Since a single nanoscale filament dominates the device characteristics in the on-state, preparing samples that are thin enough for high-resolution TEM and still contain the filaments will be very challenging. One approach is to perform

focused ion beam (FIB) etching of the pillar based devices. Since the filaments are confined inside the nanoscale pillars whose locations can be pin-pointed with the guide of the electrodes, thinning of the devices to the pillar region will lead to a reasonable possibility to capture the filaments. Alternatively, a cross-sectional device structure can be produced first with FIB, the device can then be programmed to create the filaments. In addition, the programming can even be done in situ inside the TEM chamber to allow direct imaging of the filament formation process. The TEM imaging studies will be carried out with our collaborators (Prof. Xiaoqing Pan's group in Materials Science and Engineering).

8.7.2 Integration of CMOS Circuits and Memristor Arrays

In previous chapters, standalone memristor memory arrays (chapter 4) or CMOS/memristor circuits based on standalone memristors (chapter 5) have been demonstrated. However, to achieve the full potential of memristors for memory and logic applications integrated hybrid CMOS/memristor circuits with large crossbar arrays will have to be fabricated.

Since the crossbar structure is based on passive elements (resistive switches) it can not generate gain and decouple the input from output in the circuit. On the other hand, it provides density and connectivity not matched by CMOS circuits. As a result, hybrid CMOS/nanodevice has been proposed as an approach to take maximal advantage of the crossbar structure while address issue with its limited functionalities. One of the examples is the so-called CMOL (Cmos + Molecular scale device) architecture [5 - 7]. In CMOL structure, the interface between nanodevices and CMOS is provided by pins

which are mounted on top of the CMOS metallization layers as shown in figure 8.1 (a). The nanoscale crossbar array is placed on CMOS circuitry with an angle, $\alpha = arcsin(F_{NANO}/\beta F_{CMOS})$, relative to the square arrays of interface pins (Fig. 8.1 (b)). Here F_{CMOS} and F_{NANO} are pitch size of CMOS and crossbar respectively, and β is a constant larger than 1 which depends on the CMOS design. Since identical CMOS and nanodevice units are distributed uniformly throughout the circuit, precise alignment is not necessary in the CMOL configuration. The vertical integration of nanodevices with CMOS minimizes area overhead for CMOS circuit since 2N CMOS cells can be used to control N^2 nanodevices in a N×N crossbar array. This approach thus effectively addresses the pitch mating problem since all nanodevices can be accessed through the CMOS circuitry even when $F_{NANO} \ll F_{CMOS}$.
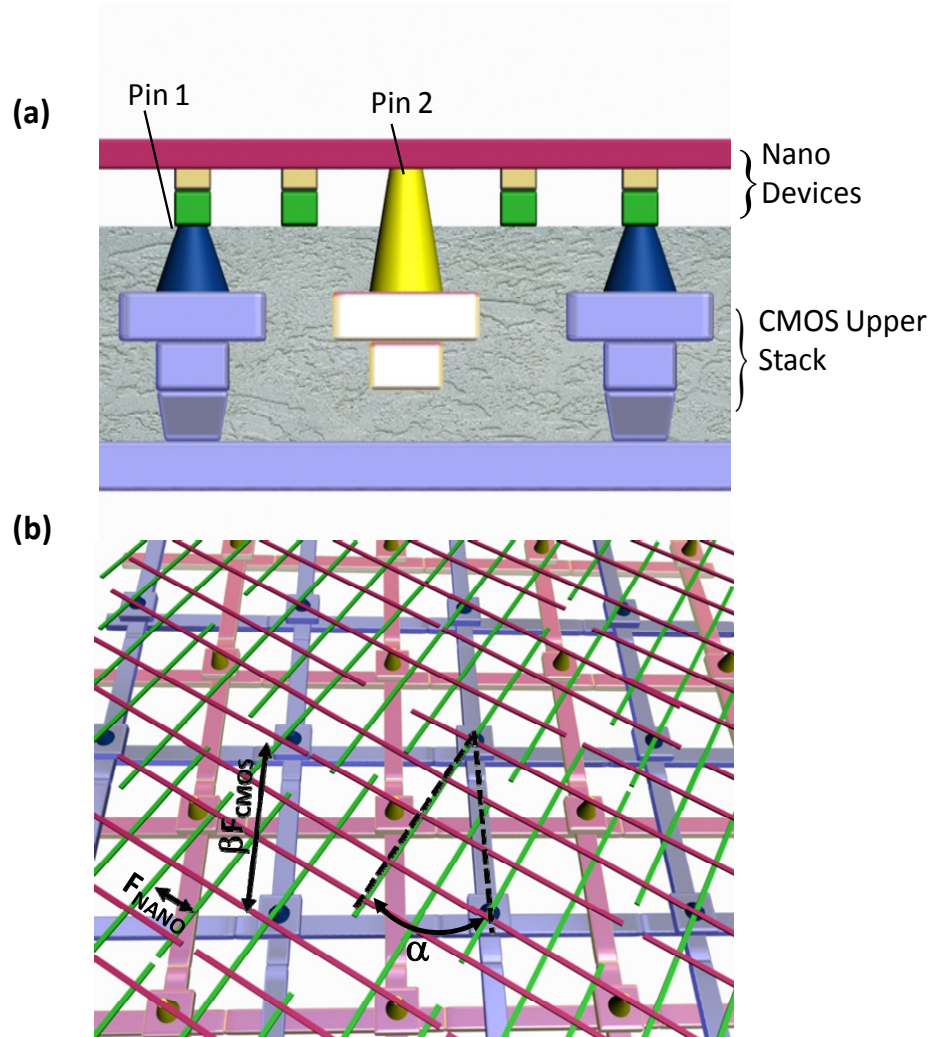
Figure 8.1. Schematic of the CMOL architecture. Nanoscale devices and CMOS components are connected by interface pins mounted on the CMOS upper stack. (a) Cross-section view. (b) Top view.

The CMOL idea can be applied to hybrid neuron circuits/memristor synapses proposed in the previous sections. The integration process of hybrid neuron circuit/memristor in CMOL approach is shown in figure 8.2. The CMOS processing units represent the neurons (Fig. 8.2 (a)) and memristors in the crossbar array act as synapses between neurons (Fig. 8.2 (b)). As discussed, the memristor synapses in the crossbar form

can provide large connectivity (Fig. 8.2 (c)). Similarly to the biological system, in which the system is composed of several neuron layers and synapses are formed between neurons in different layers, the crossbar offers interconnections between different neuron layers (e.g. pre- and post-neurons). Potentially, multi-layer structures can be created by partitioning the CMOS/crossbar system or with stacked crossbar structures. Finally, the crossbar can be further mounted on top of CMOS circuitry, minimizing area overhead caused by conventional CMOS as shown in figure 8.2 (d).

The a-Si based memristive systems offer CMOS compatible materials and processes. The high reliability and reproducibility also suggest this system can be an excellent platform to test the proposed hybrid memristor/CMOS circuit concepts. One potential issue with the vertical integration of a-Si based memristive devices with CMOS circuits is the use of the p-Si local electrodes. In most of our studies the p-Si layer was deposited with LPCVD and a high temperature annealing process ($T \geq 800°C$) was used to activate the dopants in the p-Si layer. This annealing temperature is unfortunately too high for device fabrication on top of the existing CMOS circuits. To address this issue, we have recently tested devices fabricated with PECVD deposited p-Si layer in which the process temperature was controlled to be < 380 °C. Device characteristics very similar to those reported in Chapters 2-4 have been obtained. With these improvements and continued device optimization we expect to fabricate a prototype hybrid memristor/CMOS system in the near future.
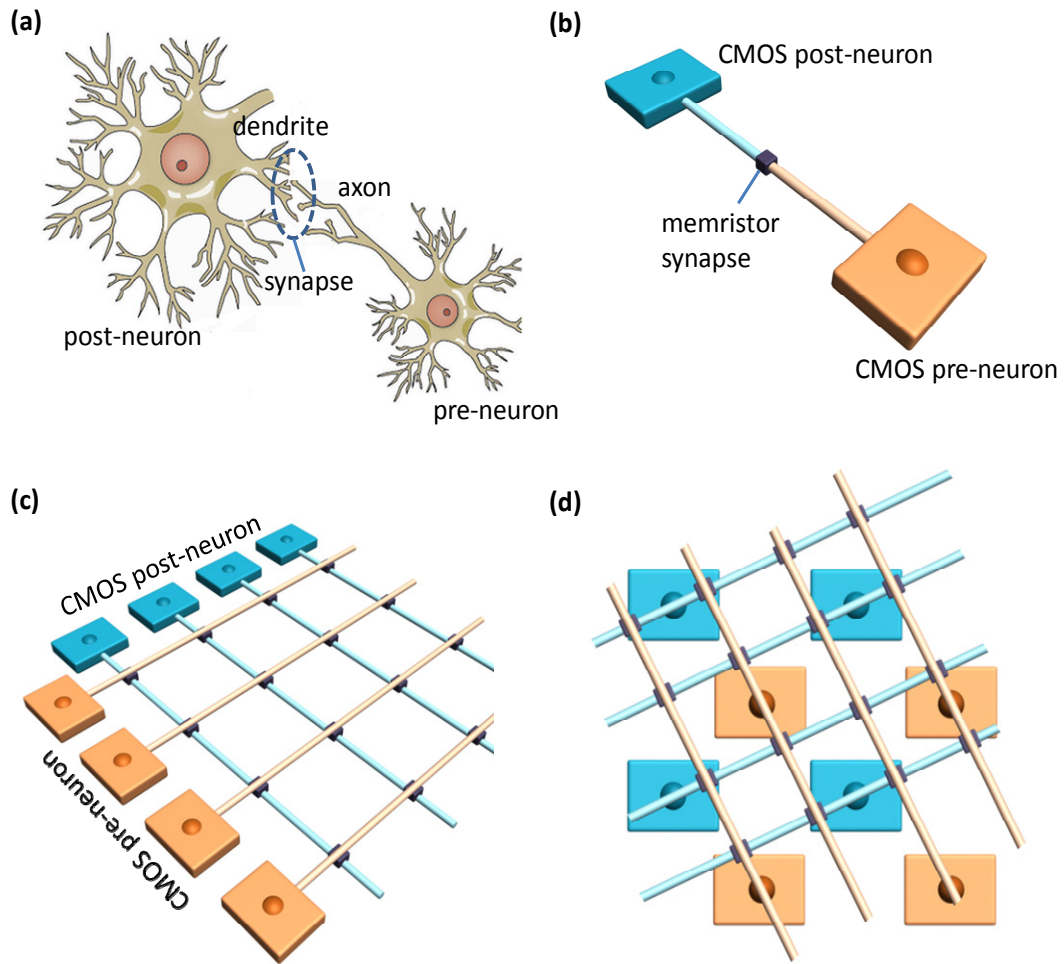
Figure 8.2. Hybrid neuron circuit/memristor synapse. (a) Biological neurons. (b) Representation of the neuron system with CMOS and memristors. (c) Hybrid neuron circuit/memristor synapse in the crossbar form. (d) How the circuit can be achieved in the CMOL structure.

References

[1]   A. E. Owen, P. G. Le Comber, W. E. Spear, and J. Hajto, "Memory switching in amorphous silicon devices," *J. Non-Cryst. Solids* **59 & 60**, 1273, 1983.

[2]   J. Kanicki Ed., Amorphous & Microcrystalline Semiconductor Devices Volume II: Materials and Device Physics, Artech House, 1992.

[3]   P. G. Lecomber, A. E. Owen, W. E. Spear, J. Hajto, A. J. Snell, W. K. Choi, M. J. Rose, "The switching mechanism in amorphous silicon junctions," *J. Non-Cryst. Solids* **77 & 78**, 1373, 1985.

[4]   M. Jafar, D. Haneman, "Switching in amorphous-silicon devices," *Phys. Rev. B* **49**, 13611, 1994.

[5]   D. B. Strukov, K. K. Likharev, "CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices," *Nanotechnology* **16** ,888–900, 2005.

[6]   Likharev, D. B. Strukov, CMOL: Devices, circuits, and architectures, in G. Cuniberti et al. (eds.), Introducing Molecular Electronics, Ch. 16. Springer, Berlin, 2005.

[7]   G. S. Snider, R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnology* **18**, 035204, 2007.