A Doctoral Thesis on

Resistive RAM and Neuromorphic Systems: Role of ions and interface states

(Thesis submitted in the requirements of the partial fulfilment for the degree of Doctor of Philosophy)

Indian Institute of Space Science and Technology,

Thiruvananthapuram

By

Preetam Hazra

(SC15D001)

Under the supervision of

Dr. K.B. Jinesh

(Associate Professor, Dept. of Physics)

DECLARATION

I declare that the Thesis titled ***"Resistive RAM and Neuromorphic systems: Role of ions and interfaces"*** submitted in partial fulfilment for the award of the degree of **Doctor of Philosophy** is a record of the work carried out by me under the supervision of **Dr. K.B. Jinesh (Associate Professor, Dept. of Physics)**, has not formed the basis for the award of any other degree, diploma, associateship, fellowship or other titles in this Institution any other Institution, University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever findings of the others have been cited.

Place: Thiruvananthapuram — Preetam Hazra

Date: December 2020 — (SC15D001)

CERTIFICATE

This is to certify that the thesis entitled ***"Resistive RAM and Neuromorphic systems: Role of ions and interfaces"*** is being submitted by ***Mr. Preetam Hazra (SC15D001)*** to the Indian Institute of Space Science and Technology for the purpose of the partial fulfilment for the award of the degree of Doctor of Philosophy, is a bona fide record of the original work carried out by him under my (***Dr. K.B. Jinesh, Associate Professor Dept. of Physics)***) guidance and none of the results are published or included in any other thesis or publication, except for the ones by the author himself.

Signature of the Supervisor

Dr. K.B. Jinesh

Associate Professor

Dept. of Physics

Signature of the H.O.D

Dr. Umesh R. Kadhane

Associate Professor

Dept. of Physics

Place: Thiruvananthapuram

Date: December 2020

ABSTRACT

This thesis uncovers the working mechanisms and the horizontal and vertical scaling limits of the most promising next-generation non-volatile memory storage scheme; that is Resistive Random Access Memory (ReRAM). By employing a Scanning tunneling microscope (STM) tip as the force electrode we were able to access the grains and grain boundaries of a Zinc Oxide switching layer and after a set of investigations, we concluded the existence of a trade-off between the scaling down and reliability of such devices below $10nm^2$ of electrode area. Next, by using the pre-established physics and mathematics (conductance vs. frequency) of a Metal-oxide-semiconductor device we understood the role of interface traps in promoting the set process of ReRAMs and the detrimental effect of these defects in sub-5nm thick devices. By using the Bias temperature stress (BTS) method and several other current-voltage measurements we recognized the ion diffusion based and oxygen vacancy based switching mechanisms in silver and gold based top electrode devices, respectively, and their subsequent effects in the technology. By depositing different thickness of Aluminium oxide (Al_2O_3) switching layer we discovered the vertical scaling being limited by different tunneling phenomena; thus restricting its lowest possible thickness to 4nm. Further, as an alternative to the conventional von Neumann computing architecture, we proposed an optical synaptic computational method with Molybdenum disulfide (MoS_2) quantum dots. We have also elucidated the nature of PPF response to different stimuli rate and their correspondence with different pass-filter responses. Finally, to evade sneak-in-leakage current in a crossbar based ReRAM organizing scheme, a combination of ReRAM and the selector device was proposed on reduced Graphene oxide (rGO) based three-terminal devices, where the set process and the off-state current was found to be a function of the applied gate voltage.

ACKNOWLEDGMENTS

Pursuing PhD for five long years is a prolonged experience of compelling oneself to not lose the focus from the subject and a constant trial to indulge oneself even through the occasional excruciating phases of the course. I can never be thankful enough to Dr. K.B. Jinesh for supervising my PhD work and skilfully guiding this work into this completion. Throughout these five years, he has been nothing but a steady supporter of my work and a very dear friend of mine. In an environment of limited resources in a relatively new institute, often, pursuing such a course can encounter a lot of unwanted hurdles. But working with Dr. Jinesh has made me realize even in such helpless situations one can always choose to be creative and come up with unconventional experimental ideas and work them to really interesting conclusions.

I was blessed with the craziest friend circle one can ever ask for, lucky for me, I am equally crazy. Spending time with these people hardly made me realize the reciprocation of dire failure my work was facing constantly. To all the future PhDs, if you are surrounded by caricatures like Pramod, Praveen, Shashank, Yogesh, Surya, Darshika, Dayal, Anna, Jeevan, and Aswathy, consider yourself lucky.

Lastly, I would like to thank my parents, my sister, and my wife who have always been the invisible pillars behind all the works I have produced during my tenure in the Indian Institute of Space Science and Technology. These five years have taught me, a PhD is not supposed to be hassle-free and it never will be. But when your family believes in you and becomes the never-ending source of strength for you, trust me, half of the battle is won.

Preetam Hazra

ABBREVIATIONS

CD; Compact disk

DVD; Digital versatile disk

MOS; Metal oxide semiconductor

FG; Floating Gate

MRAM; Magnetoresistive random access memory

PCRAM; phase change random access memory

ReRAM/RRAM; Resistive random access memory

TE; Top electrode

BE; Bottom electrode

HRS; High resistive state

LRS; Low resistive state

I-V; Current-Voltage

C-V; Capacitance-Voltage

G-f; Conductance-Frequency

STM; Scanning tunneling microscope

ALD; Atomic layer deposition

PALD; Plasma assisted atomic layer deposition

TMA; Trimethyl aluminium

MIM; Metal-insulator-metal

STS; Scanning tunneling spectroscopy

G; Grain

GB; Grain boundary

V_{fb}; Flatband voltage

D_{it}; Interface trap charge density

τ_{it}; Trap lifetime

C_{lf}, C_{hf} and C_{ox}; low-frequency, high-frequency and oxide capacitance, respectively

CC; Compliance current

V_i; Oxygen ion

V_o; Oxygen vacancy

BTS; Bias temperature stress

SCLC; Space charge limited current

FN; Fowler-Nordheim tunnelling

PF; Poole-Frenkel Injection

DT; Direct tunnelling

TAT; Trap assisted tunnelling

PPF; Paired-pulse facilation or Pulse-Paired facilation

TFT; Thin film transistor

τ_g; Generation time constant

τ_r; Recombination/decay time constant

GO; Graphene oxide

LDOS; Local density of states

LIST OF FIGURES

CONTENTS

Chapter 1

Introduction

Non-volatile memory is a type of data storage device which stores the data and ensure the existence of the information without the expenditure of any constant supply of electrical power. Non-volatile memory has been guiding and moulding this fast computing era for almost half a century. Today, the digital storage system has embedded itself into every corner and trenches of our daily life. We cannot imagine our life without a smartphone, computer, pc, home theatre, media players, gaming consoles, e-readers, etc. or as a matter of fact any electronics with a storage capability. The amount of storage per unit currency offered to us by the industry has been decreasing ever since. But every technology has its limitation, and so does our current storage scheme, the FLASH technology[1,2]. The constant demand of higher bit density and higher speed while maintaining the reliability has driven the technology to reach at its brink of saturation[3–5]. The very physics which lead FLASH to take over the digital storage industry in the past few decades is now slowing the technology down.

Let us take a look into the history of the non-volatile memory, its current picture, the limitations, future options, and finally why we embarked on the journey of studying Resistive random access memory (ReRAM or RRAM) for 5 long years.

1.1 History

1.1.1 Magnetic core memory

The first-ever successfully commercialized non-volatile memory was a magnetic core memory[6]. In 1947 Frederick Viehi got the patent for a core memory which was later purchased by IBM.

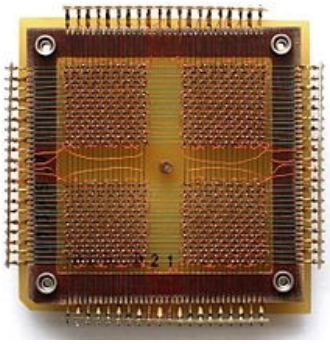


Figure 1.1. A 32 × 32 1 kilobit magnetic core memory unit.

The mode of operation was in crossbar format (please refer figure 1.1). Each of the elements was used to store either a zero or a one. A single bit was semi-hard ferrite core with wires

wound around it like a transformer. The electric pulses applied across such wires would change the magnetic dipole's orientation, thus storing bit "1". To store bit "0" all we needed to do was to change the direction of the current through the wire. The different arrangement of dipoles in the core would change the resistive state of the system. To read the data another wire was employed (which ran through the core) with voltage supply and current read connected to proper terminals.

1.1.2 Magnetic tape memory

First introduced in 1951 in UNIVAC, magnetic tapes[7] was later perfected and popularized by IBM in the 1970s and 80s.

Figure 1.2. Magnetic tape memory in different cartridges.

These tapes were coated with magnetic coatings (figure 1.2), which could be magnetized or demagnetized by an external magnetic field provided by the writer head. Another small magnet (read head) was employed to reach out near the surface of the tapes to read the degree of magnetization on the rotating film and transfer the information to an electronics transducer to convert it into legible data.

1.1.3 Floppy discs

Figure 1.3. Floppy discs with their protective plastic casings

In a floppy disc[8], a circular plastic disk with magnetic material coating was used to store data, much like a magnetic tape. Instead of a separate reader and writer and single header (with

both reader and writer mounted together) used to hover over the spinning disk to record and read the information. The actual disks used to vary in sizes and used to come in protective plastic casings (see figure 1.3).

1.1.4 CDs and DVDs

Compact Discs[9] and Digital Versatile Discs took a different approach for storing and accessing the digital bits.

Figure 1.4. Digital versatile discs

DVDs and CDs (figure 1.4) comprise a protective plastic layer (the back of a disc) followed by a reflective aluminium layer, a phase change (crystalline or amorphous) material, and another transparent polymer. A powerful laser hits the phase change layer and can locally transform it into a crystalline (translucent) or amorphous (opaque) material. A crystalline layer allows light to pass through it and gets reflected from the aluminium surface, on the contrary, the amorphous region prevents the light to pass through. A light detector in the DVD drive can follow this change and thus can assign them to two different logic states.

1.1.5 Magnetic Hard drives

The previous sections made us realize that a substantial number of memory storage schemes were based on the principle of magnetism until the CD and DVD kicked it. However, there is one storage system still exploits the ferromagnetic property of a material to store data in this era. Concentric circles or "tracks" of magnetic material covers a hard disc drive. For writing and reading, a header is connected with an arm is attached with an actuator at one end of the memory assembly (refer figure 1.5). The writer is essentially a tiny solenoid that can fly over the rotating disk at any given location and can magnetize the portion into north or south alignment. The reader in this case is a Magneto Resistive (MR) sensor. The voltage applied across it causes a constant current flow unless altered by a magnetic field present in close

proximity. When the reader flies by the desired location on the surface it can sense the orientation and recognizes north polarity as "1" and south polarity as "0".

Figure 1.5. An uncovered Hard disc drive with the read and write header.

To cope with the growing demand for data density, industries are trying to magnetize the material perpendicularly rather than conventional horizontal fashion; in an attempt to make the most of the allotted space. To increase the bit density even further, companies are working on a relatively newer technology[10] called the heat-assisted magnetic read (H.A.M.R) technology. It allows the writing head to locally heat a portion on the disc (by mounting a laser on the writing head) and magnetize it at a much lower power. The main catch of this technology is it allows the user to magnetize a very small area on the disc.

1.2 Modern era

The FLASH Memory

Shunpei Yamazaki[1] invented and patented FLASH memory in the 1970s. In 1966 Yamazaki observed the C-V hysteresis in a device and speculated the theory behind it. Although, Dr. Kahng (Bell Laboratories) first proposed the idea of a floating gate (FG) MOS structure to store data (1967). Kahng's device was not recognized as a game-changer at that time, because, the floating gate was made with zirconium and the sidewalls were exposed into the air. This did not render very efficient FLASH storage.

Yamakazis's structure consisted of a MOS transistor with a floating gate inside the gate dielectric, completely insulated from the surrounding environment, as depicted in figure 1.6. This non-volatile technology was reliable, scalable, fast, and most importantly, by far the cheapest of them all. Till date, most of our external storage units are FLASH based. Pendrives, solid-state drives, micro SD cards, EEPROMs, M.2 NVME, etc. are some form of

FLASH architecture. Let us take a look into the basic structure of a FLASH transistor and try to understand the mode of its operation.

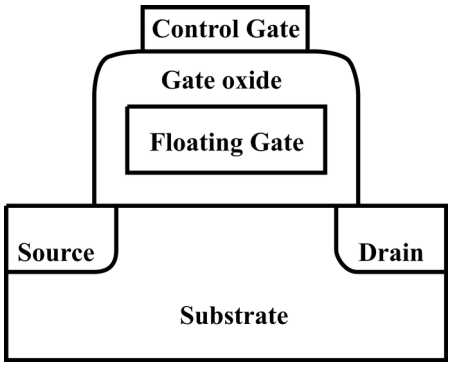


Figure 1.6. An elementary structure of a floating gate MOSFET.

One can easily identify the similarity and difference between a typical MOSFET and a floating gate MOSFET. The additional gate inside the dielectric sets FLASH transistor apart from a conventional one.

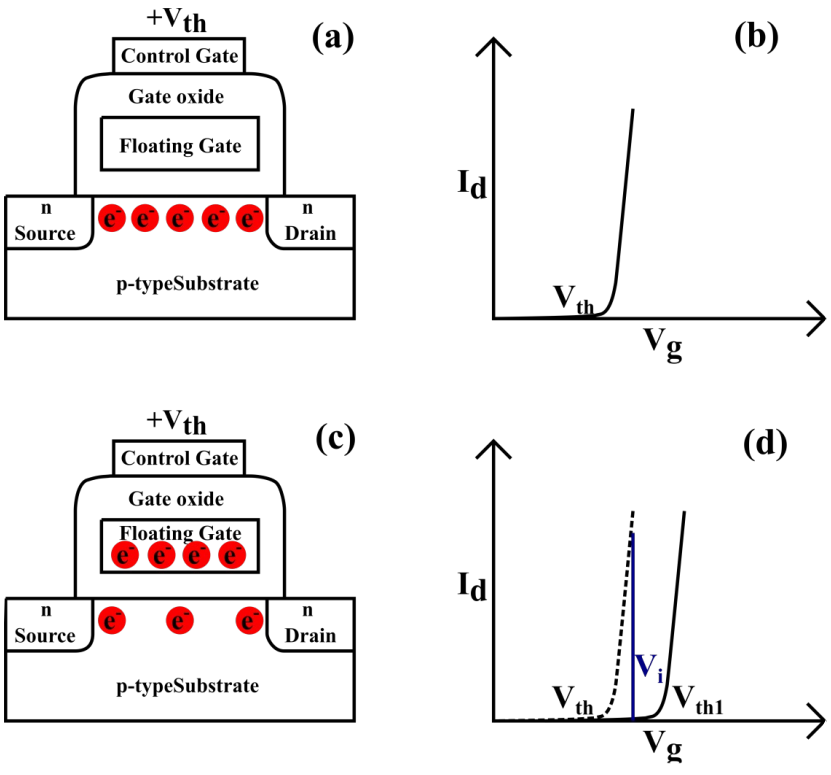


Figure 1.7. (a) An n-channel formation while applying V_{th} on the gate, (b) the corresponding transfer curve, (c) the negatively charged floating gate does not allow for the charge build-up with the same V_{th} and (d) its corresponding transfer curve showing the shift in the threshold voltage to V_{th1}.

In the upcoming sections, we shall be explaining the operation principle of FLASH storage and for that purpose, we shall consider an n-channel MOSFET. When a sufficient positive voltage is applied on the control gate, a substantial amount of negative charges buildup on the substrate, between the source and the drain and shorts them. This voltage is known as the threshold voltage (V_{th}) and the conducting path is called the channel. If we were to introduce some electrons amidst the gate, we could opt for a floating conducting gate in the dielectric and negatively charge it. This charge will induce some positive charge on the channel, thus negating a fraction of the previous build-up. This would make the source and drain open circuit again; now, to grow the channel, one needs to apply a larger voltage on the gate, larger than V_{th}. It is evident that with a V_{th} on the gate and the floating gate being neutral we shall get a high current across the channel (i.e. between source and drain), thus reading “1”, which is also reflected in the transfer curve; see figure 1.7(a) and (b). With the floating gate's participation, the channel remains open and the output current stays low, which reads "0"; see figure 1.7(c) and (d). A more convenient way to read the data is to apply an intermediate voltage (V_i) at the gate; in the first case it shall read out a higher current, assigning it to state “1” and for the state “0” the circuit does not read any current.

Writing data in a transistor is the trickiest part. Here, we exploit the oxide separating the FG and the control gate, which is very thin and so is the portion between the FG and the channel. If we apply a high negative voltage on the control gate the insulator band bending will cause band triangulation, leading to Fowler-Nordheim tunneling of the electrons to the FG. On removal of the electric field the electrons safely reside there; therefore, writing the data. Another way to write the data is by hot-electron injection. It requires a high voltage across the drain and the source with a positive bias on the control gate. The strong electric field builds up high kinetic energy to the electrons and when the electrons gather enough energy to overcome the insulating barrier and reach the FG. The author would like to point out that FN tunneling is far less destructive to the gate insulator than the hot-electron injection method. Erasing is a relatively easier process. A high positive voltage on the control gate can lead to the FN tunneling of the electrons from the FG to the positive terminal and erase the data.

Now, that we covered the reading, writing, and erasing mechanisms of a FLASH memory, we should be able to appreciate its different architectures. The two architectures dominating the FLASH non-volatile industry are the NAND and NOR layouts. Owing to their advantages and disadvantages they find different areas of application.

In NAND[11] layout the devices are connected in series (figure 1.8(a); if you recall NAND = “NOT” series "AND"), whereas in NOR[12] layout devices are in parallel connection (figure 1.8(b); NOR=”NOT” series “OR”).

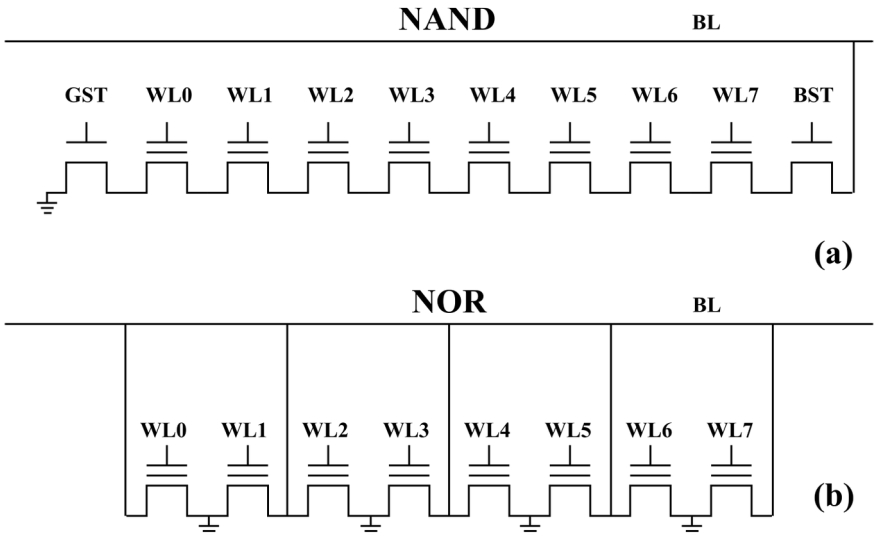


Figure 1.8. (a) NAND layout and (b) NOR layout of the floating gate transistors.

The word lines (WLs) in a NAND arrangement are all connected in series with only one ground connection through the ground select transistor (GST). On the contrary, the devices in the NOR structure require ground connection and a bit line connection for each and every device.

To read data from a specific device in NAND we need to apply V_{it} to the specific word line and set all other devices on (i.e. a voltage higher than V_{th1} in figure 1.7) to read the high or low current. Along with that the GST and BST (block select transistor) should remain on to let the user access the block. However, for a NOR device, the rest of the bits are needed to be in off state (the WL voltage should be below V_{th}), this step is necessary because during reading we always need to make sure the current is only passing through the required device and not accidentally coming from any other cells of the block.

Erasing data from any of these circuits is simple. For both of these layouts, applying a high voltage through the WLs will ensure the floating gate is drained out of the electrons. As we discussed earlier, FN tunneling leads the electron to move from the FG to the control gate and sets all the devices in a block to “1”. Well, since both of these architectures require erasing the entire block altogether in a “FLASH”, the name was popularized.

Writing data ("0") in a specific NOR cell is done by hot-electron injection from the channel. A high voltage is applied at the bit line (BL), and the exact device is fed with a positive voltage on the WL; while the rest remains switched off. NAND devices are set to state "0" by FN tunneling. An application of a high negative field on the word line of a specific device pushes the electrons from the control gate to the FG.

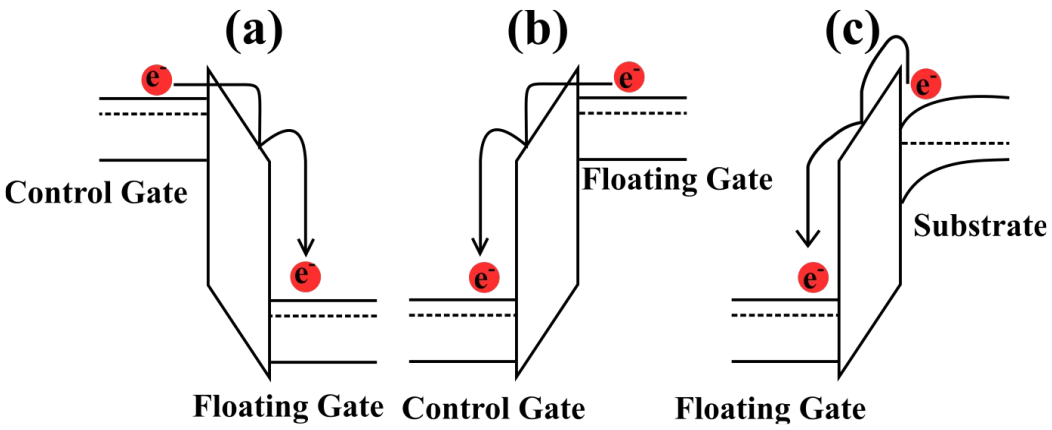


Figure 1.9. (a) Writing ("0") data in the FG by FN tunneling, (b) erasing data by the same principle from the FG, and (c) the alternative writing by hot-electron injection method, by drawing electrons from the conducting channel.

The hot-electron transition is a detrimental process for the gate insulator, with multiple numbers of writing and erasing the devices can get damaged beyond repair. This is why, the average read-write cycle per NOR device ranges within 1k to 10k times, whereas a NAND cell can endure through many more cycles (10k to 100k).

Apart from the writing scheme, the NAND and NOR layouts differ in some trivial aspects. NOR requires separate BL interconnects and ground line and therefore, is characterized with larger footprints than NAND. The lack of excess interconnections from the ground line and WL gives NAND a smaller foot-print and makes it an ideal choice for affordable non-volatile applications. Although, being arranged in series connection, the NAND comes with its inherent disadvantage; the series latency. For applications which need relatively faster non-volatile storage (for example in embedded systems, booting ROMs, etc.) NOR layout is preferred over NAND. The parallel connection in NOR bypasses the device to device propagation delay; thus saving a substantial amount of time in reading. Owing to the less destructive FN tunneling and cost-effective fabrication procedure NAND is always preferred over NOR and is most commonly encountered by us in daily life.

Until now our discussion focused upon FG MOS and how it can be used as a non-volatile memory; where each device serves for a single bit. These devices are known as the SLCs (figure 1.10(a)) or single-level cells. Modern SSDs and M.2 NVMEs offer higher reading and writing speed with higher bit density. This remarkable feat has been achieved because of the introduction of MLCs or multilevel cells[13]. Unlike SLC, MLC does not rely on two states (namely, either electrons are present on the FG or state "0" and FG is neutral or state "1"), rather it can detect the threshold voltage shift in four different levels (neutral FG and 3 successive levels of electron storage; figure 1.10(b)). This allows an MLC device to act like two bits of four combinations from a single cell. But this technology has its downside. More levels mean less is the inter-level threshold voltage shift and lower is the tolerance. MLCs generally offer lesser lifetime than SLCs and a higher probability of data corruption after multiple read-write cycles.

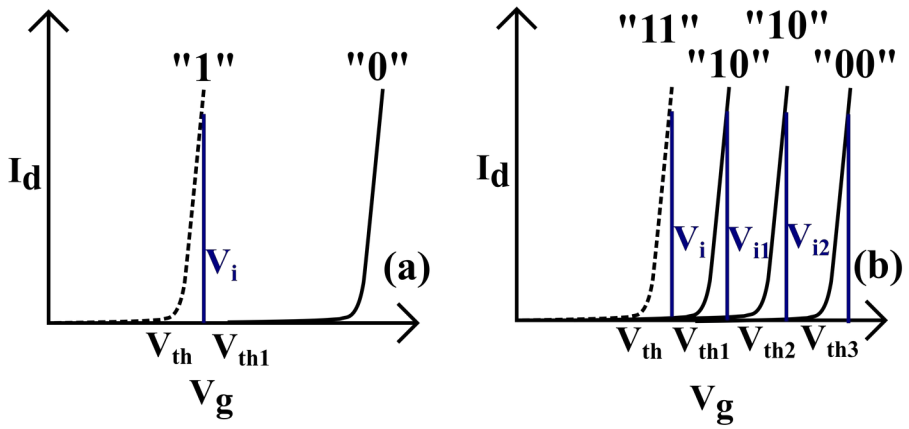


Figure 1.10. (a) Showing two different logical states in an SLC and (b) four different logical states in an MLC floating gate MOSFET.

The above figure depicts the existence of multiple logical states in an MLC, rather than two in SLC. The figure perfectly showcases the noise margin or tolerance degradation in an MLC system.

1.3 The Future

Scaling down in FLASH memory always accompanies short channel effects and complications in fabrication steps. Thinning down the gate dielectric and channel length makes the device more prone to succumb before the direct tunneling effect. Moreover, the metal interconnects for the source, drain, control gate and substrate consumes the majority of

the layout footprint. The deviation of Moore's law itself is a standalone proof that miniaturization is one of the goals today's semiconductor industry is struggling to keep up.

The scientists throughout the globe have started working on alternate methods for permanent data storage solutions. Among them, magnetoresistive RAM (MRAM), phase change RAM (PCRAM or PRAM), atomic switch, Resistive RAM, etc have shown great promises. Let us take a walk through these memory storage schemes one by one.

1.3.1 MRAM

An MRAM[14] exploits the spin-transfer torque property of an electron to store data. Structurally it consists of a ferromagnetic bottom electrode (BE) and a ferromagnetic top electrode (TE) separated by a thin conducting layer (please refer figure 1.11). The BE generally is a Fixed ferromagnetic material or fixed state magnet, while the TE is open to external manipulation and known as a free state magnet.

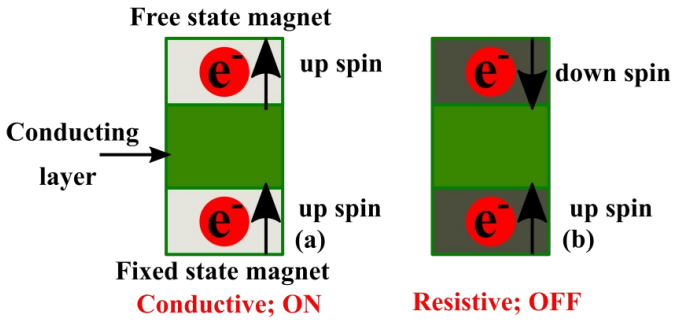


Figure 1.11. Different resistive states of an MRAM depending upon the electron spins.

When both the magnets are in the same spin direction the circuit allows high current to pass through it, rendering a lower resistance (state "1"). Forcing a large current through the free state magnet can change its electron spin direction, exhibiting a high resistive state of the device, which represents state "0".

1.3.2 PCRAM

PCRAM[15] resembles the structure of a capacitor with a phase-changing material embedded between two electrodes. A high voltage pulse (or high-intensity laser pulse) can melt the material because of the joule heating. However, during the cooling down process, a rapid cooling leads the quenching of the material into an amorphous state and a relatively slower cooling (by holding the crystalline temperature for few moments) renders a crystalline phase. Much like MRAM, PCRAM (figure 1.12) also stores data in form of different resistive

states; the amorphous phase corresponds to the highly resistive state, whereas the crystalline/ polycrystalline phase exhibits conducting state.

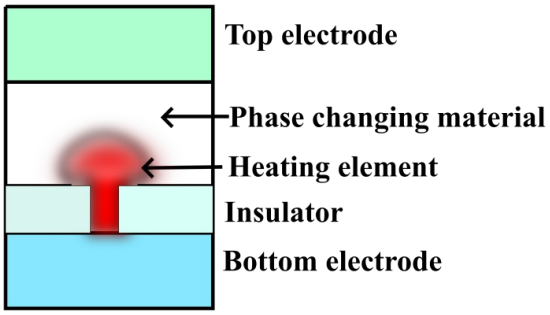


Figure 1.12. A schematic diagram of a PRAM device with the heating element.

1.3.3 Atomic Switch

Atomic switches[16] are 2 terminal (or 3 terminal in case of an additional controlling gate) tunneling junctions; which doubles as a resistive switch.

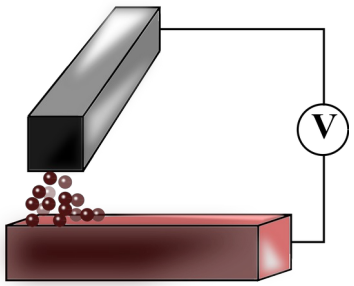


Figure 1.13. A single row-column pair of an atomic switch crossbar array; demonstrating the metallic filament formation in the tunneling gap.

Two electrodes separated by a tunneling gap often can form a metallic bridge under a high electric field (as depicted in figure 1.13). A low electronegative metal (for example copper) upon positive bias tends to lose electrons and becomes a cation; these ions can easily gain electrons from the cathode (by means of direct tunneling) and get reduced to neutral atoms. Eventually, this gives rise to a metallic protrusion from the cathode. This sort of conducting filament (CF) formation and collapse (at negative bias) toggles the resistivity of the tunneling gap and can be accounted for two different states of logic states (namely, "0" and "1"); therefore, can be used as non-volatile memory.

1.3.4 Resistive random access memory

Some insulators[17] and semiconductors[18] are capable of changing their resistive states under voltage stress. These insulators are the heart of a resistive ram. Structurally, a resistive RAM device comprises two electrodes (namely, top and bottom electrode) which are electrically insulated by the switching dielectric. By applying different voltage bias one can easily toggle the resistive state of the device between a high-resistive-state (HRS) (figure 1.14(a)) or off state and low-resistive-state (LRS) (figure 1.14 (b)) or on-state. The following figure is a pictorial depiction of a ReRAM device in action.

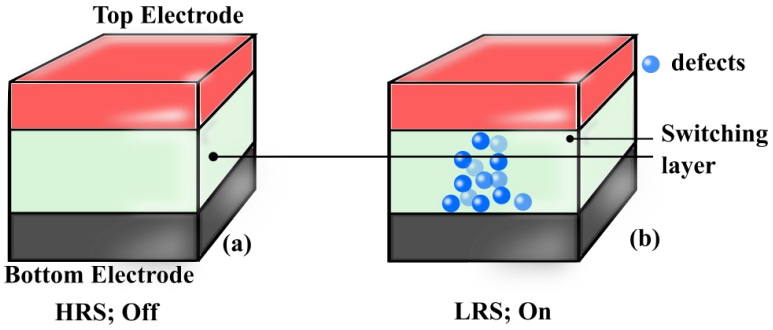


Figure 1.14. (a) A schematic illustration of a ReRAM structure at its HRS state and (b) LRS state.

A certain magnitude of an electric field is capable to induce or create electrically activated defects in the layer. An adequate quantity of these charged defects can form a conducting filament/bridge (CF) between the electrodes (see figure 1.14 (b)). The origin and nature of these defects and how they manipulate the switching behavior will be discussed afterward. Generally, an opposite polarity of voltage reverses the effect and the CF gets dissolved (or partially dissolved); leading to electrical isolation of the Top electrode (TE) and bottom electrode (BE) again. The most interesting fact of these phenomena is, once the state is changed, the dielectric retains the state even without the presence of a field.

Switching behavior

Let us breakdown the switching phenomena in the following steps.

1 At a lower voltage, a resistive switch acts like an ideal capacitor and does not allow any current to pass through it. The only observable current is the leakage current through the thin dielectric and because of other parasitic components. This state is

referred to as high resistive state or HRS. A good resistive RAM is expected to exhibit negligible HRS/off current.

2 An electrical breakdown experienced by the dielectric can be induced by metal ions, dielectric defect generation, electrode-dielectric interface redox, or their cumulative effect. However, to commence the process of breakdown often we need to apply a higher field (higher than the usual set voltage) with a compliance check, to procreate the path which later will guide the breakdown through the device. This is known as the formation process and the voltage at which the breakdown takes place is known as formation voltage. At this stage, the system is at conducting state or low resistive state or LRS or on-state.

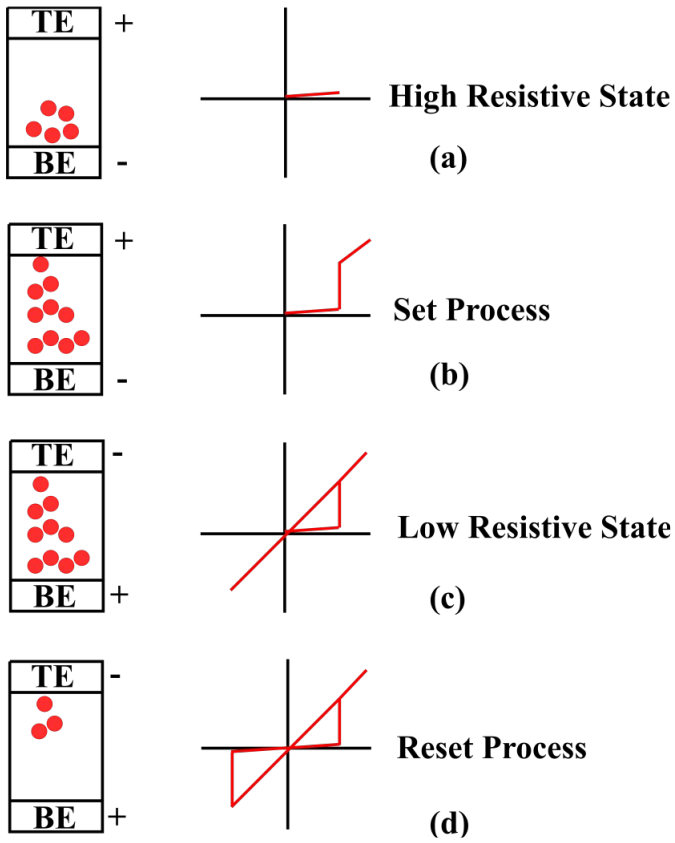


Figure 1.15. Diagram showing (a) the high resistive state, (b) set process, (c) low resistive state and (d) reset process of resistive bipolar switching.

3 Depending upon the compliance check during the formation process, some devices might need switching off process some might not.

4 Assuming the device is back at HRS, now, at a lower voltage (w.r.t formation voltage) the device turns on, this transition voltage is called the set voltage, and this operation is commonly known as the writing process.

5 At a comparable opposite polarity of voltage, the device switches back to its off state; thus erasing the data.

Figure 1.15 perfectly illustrates the different stages of a bipolar switching phenomenon. Several device configurations are known to switch on and off in the same polarity of an applied electric field. This is termed as unipolar switching. In bipolar switching, the set and reset processes are generally caused by a certain phenomenon and its reverse phenomenon, respectively. However, in the case of unipolar switching, even though the set process is commenced due to defect formation, the switching-off due is to the destruction of the filament by joule heating.

Switching mechanisms

Several switching mechanisms have been put forward by various research groups. Among them, the metal ion migration[19], dielectric defect generation[20], and electrode-insulator redox reaction are the most accepted theories.

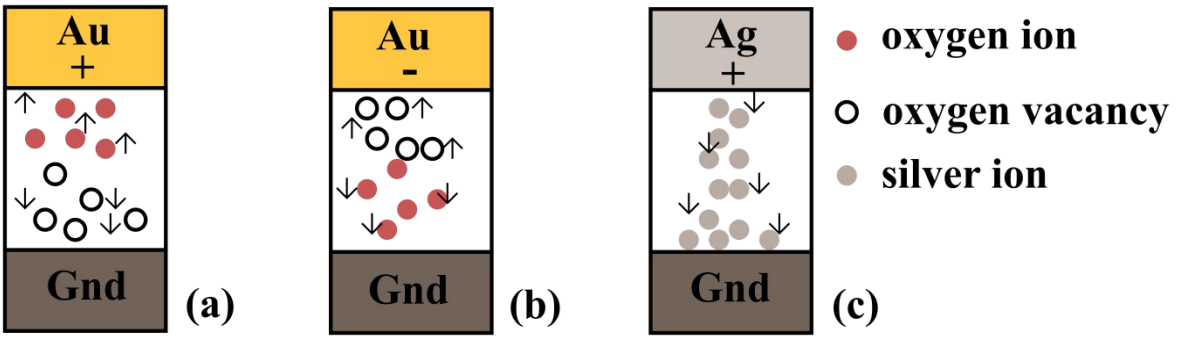


Figure 1.16. (a), (b) Oxygen defect based switching in metal oxide dielectrics in the presence of noble metal as a top electrode and (c) metal ion migration causing filament formation in the switching oxide.

Metal ion migration is mostly observed when the TE is generally a low electronegative metal. With positive bias on the TE, the metal gets oxidized into metal anions. The mutual positive field at the TE tends to repel the ions. An attractive force from the BE causes the reduction of the ions at the BE site. A repetition of such a process subsequently gives rise to a continuous conducting path populated with metal atoms which leads to an electrical breakdown. An

opposite field forces the metal ions back towards the TE, thus dissolving the filament and switching the device back to its original off state.

Noble metals generally have a higher electronegativity and require a substantial amount of field to oxidize them. Devices with noble metallic TE seldom show any metal ion diffusion, rather the electric field has been observed to act as an external parameter to modulate the defects in the switch. Metal oxides as switching layer (with noble TE) are famously known to dissociate into oxygen ions (cations) and oxygen vacancies (anions) under voltage bias. The ions tend to move towards the TE leaving behind a trail of oxygen vacancies through which the electrons can reach from the cathode to the anode.

In some cases, switching relies upon a defect reservoir between the TE and the dielectric. Oxides with aluminum as the top electrode[21] are famously known to form an AlO_x layer at the interface. This layer acts as an oxygen defect reservoir. The metallic aluminum and the AlO_x layer (depending upon the oxidation state) can accept or provide oxygen ions to the switching layer to aid the formation of a defect populated conduction path. Apart from that, the BE-oxide interface can play a vital role in the switching process; the details have been discussed in the 4th chapter.

Crossbar

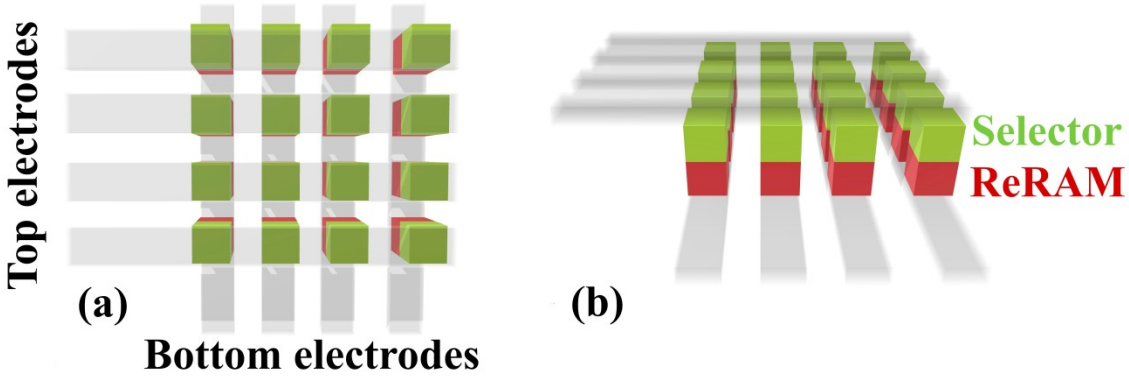


Figure 1.17. A crossbar device arrangement of resistive RAM (a) top view and (b) side view with the ReRAM and selector device embedded together between the two electrodes.

In a real chip scenario, the ReRAM devices are not isolated devices rather arranged in a matrix format[22] (see figure 1.17). A series of parallel metallic lines act as the TEs and another orthogonal set of lines act as the BEs. Any individual targeted device can be accessed (for reading or writing data) by applying the appropriate voltage across the corresponding row

(TE) and column (BE); the way one addresses a matrix element. The TE is biased with the voltage (set/read) while the BE remaining grounded. To avoid accidental switching on of any device the rest of the devices connected with the TE are not grounded, rather biased with V/2 or V/3. In that case, all of those devices suffer from a potential difference of V/ 2 or 2V/3; sufficient to avoid any sort of accidental breakdown.

While being arranged in a crossbar array, the devices suffer from an inherent limitation, the sneak-in-leakage current[23]. During reading data from a single device, additional ghost currents might appear. These currents originate through random set devices, cumulating leakage currents, V/2, and V/3 bias leakage currents, etc. sneak-in-currents give the reading circuitry a false impression of high current even if the targeted device is at HRS. The detrimental effect of sneak-in-current has been demonstrated in the 8th chapter.

To counter this problem an additional non-linear device can be connected with the ReRAM in series, which shall not allow any current at HRS while being conducting after the set voltage regime.

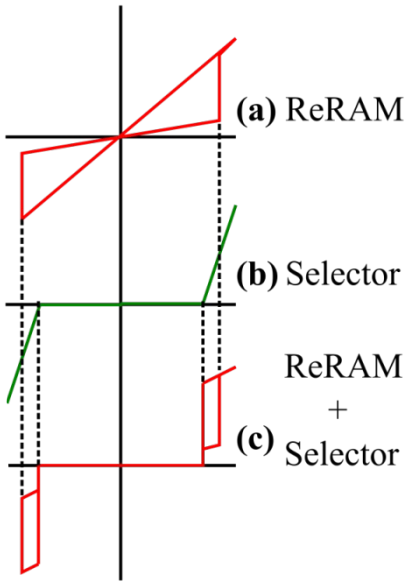


Figure 1.18. Individual current-voltage behavior of (a) a ReRAM and (b) a selector device. The (c) combined effect of these two devices retains the set and the reset process while successfully evading the leakage current through a system.

A selector device[24,25,26,27]can be imagined as a diode whose knee voltage and Zener breakdown voltage is just below the set and reset voltages, respectively. A series combination of a resistive memory and a selector only provides a small window beyond the set and reset process to ensure flawless operation of the system while successfully bypassing the unwanted

leakage current. A countless number of diodes, BJTs, transistors, back-to-back devices have proposed by various groups. Needless to say, tackling sneak in leakage is a major concern in this industry and it leads the selector devices to take off as a different brunch of research altogether.

1.4 Scope of the thesis

The current technology (FLASH) is struggling to keep up with the increasing demands of scalability. Short channel effects, direct tunneling, Fowler-Nordheim tunneling, etc. play a huge role in data corruption in sub- 5nm devices and been disrupting the expected progress in FLASH scaling. Various companies have started shifting their research area from scaling of FLASH to 3D- stacking of FLASH and MLC based FLASH technologies to keep up with the expected demand of bit density. This kind of progress is nevertheless quite promising but comes with a trade-off with device-life-expectancy and manufacturing cost. New promises, such as MRAM and PCRAM can be scaled down to unprecedented dimensions and can meet the growing expectations as well. However, MRAM cannot be packed really close to each other; the inherent magnetic field from the fixed-scale-magnet and free-state-magnet tend to alter the magnetic orientations in the nearby devices, leading to hazardous crosstalk effects. This crosstalk comes with a cost of feature size of such devices, therefore affecting the overall scaling of a memory matrix. PCRAM on the other hand is devoid of such problems but, can fall short when it comes to high-speed operations. To change the switching layer into a crystalline phase the temperature needs to stay at crystalline-temperature for a few moments before it comes to room temperature. This step consumes an unwanted amount of time to write the information, ultimately stretching the overall read-write cycles.

With the current technology (7nm) NAND FLASH can reach a bit density up to 16GB/cm^2 whereas a 10nm×10nm ReRAM (or any other crossbar compatible device like MRAM, PCRAM, etc, assuming a 4F^2 of feature size) device can accommodate ~32GB/cm^2 and with half the length of one electrode, this bit density can climb up to ~128 GB/cm^2. In the light of these problems, an ultrafast memory with sub 5nm scaling promises can only be offered by ReRAM technology.

This thesis is dedicated to uncover the mechanisms, which are involved in governing the switching behaviors and also inspecting their consequences in scaling. In Chapter 3 we have discussed the area scaling limits of a resistive switch for polycrystalline materials. The influence of grains and grain boundaries can modify the electronic structure of such material

and introduce anarchy in the switching phenomena. The chapter also discusses the reliability of sub 100nm^2 devices and how they get undesirably manipulated by the grains/boundaries as the size of a device decreases. Chapter 4 and chapter 5 inspect the role of interface traps and mobile ions in a ReRAM device. Pre and post breakdown studies in these chapters helped us draw some detailed conclusions about the breakdown mechanisms in a resistive memory. Although, chapter 4 has touched the vertical scaling limitations of a resistive capacitor based on the interface trap's role, but a much detailed picture was drawn in chapter 6. We employed an atomic layer deposition system to fabricate devices with different thicknesses of the dielectric. Different conduction mechanisms were studied for these devices and some interesting outcomes were observed. However, the most interesting discovery from this work was the effect of thickness on the on-off ratio, reliability of any device. Finally, the chapter compared all the parameters (to quantify the grade of a device) and concluded with the minimum vertical limit a device can expect depending upon the breakdown mechanisms, without compromising the device quality. Apart from memory storage, a ReRAM device can also be employed to perform complex mathematical computations. By modulating the resistivity of a device between the HRS and LRS by a controlled voltage (or laser pulse) we can artificially perform the operations of a biological synapse. We used single MoS_2 quantum dots (for smaller footprints) and explored the neuromorphic behavior in response to external optical stimulus in chapter 7. This kind of system in a fully-fledged device form can act as a bionic eye. The next chapter (chapter 8) reports a new solution to the selector device and bias mechanism of a resistive memory in a crossbar array. The first part of this chapter dealt with the problems which can be faced in crossbar array without the selector device. The next section illustrates the switching behavior and operation principle of a resistive switch in presence of an external control terminal. The chapter also discusses the possibilities and promises this technology can offer to counter the sneak-in problem and the accidental switching problem as well. The final chapter (chapter 9) is merely a post walkthrough of the thesis while briefly discussing the conclusions and future works of each chapter.

References

1. Yamazaki, S. History and Future Perspective of Nonvolatile Memory. in *2007 Sixteenth IEEE International Symposium on the Applications of Ferroelectrics* 1–3 (IEEE, 2007). doi:10.1109/ISAF.2007.4393148.

2. Zhao, C., Zhao, C., Taylor, S. & Chalker, P. Review on Non-Volatile Memory with High-k Dielectrics: Flash for Generation Beyond 32 nm. *Materials* **7**, 5117–5145 (2014).

3. Youngwoo Park, Jaeduk Lee, Seong Soon Cho, Gyoyoung Jin & Eunseung Jung. Scaling and reliability of NAND flash devices. in *2014 IEEE International Reliability Physics Symposium* 2E.1.1-2E.1.4 (IEEE, 2014). doi:10.1109/IRPS.2014.6860599.

4. Park, B. *et al.* Challenges and limitations of NAND flash memory devices based on floating gates. in *2012 IEEE International Symposium on Circuits and Systems* 420–423 (IEEE, 2012). doi:10.1109/ISCAS.2012.6272053.

5. Compagnoni, C. M. *et al.* First evidence for injection statistics accuracy limitations in NAND Flash constant-current Fowler-Nordheim programming. in *2007 IEEE International Electron Devices Meeting* 165–168 (IEEE, 2007). doi:10.1109/IEDM.2007.4418892.

6. Younker, E. L. A Transistor-Driven Magnetic-Core Memory. *IEEE Trans. Electron. Comput.* **EC-6**, 14–20 (1957).

7. Wilkinson, R. T. & Houghton, D. Portable four-choice reaction time test with magnetic tape memory. *Behavior Research Methods & Instrumentation* **7**, 441–446 (1975).

8. Stability Analysis of a 2 Floppy Disk Drive System and.pdf.

9. Immink, K. A. S. EFMplus: the coding format of the multimedia compact disc. *IEEE Trans. Consumer Electron.* **41**, 491–497 (1995).

10. Kryder, M. H. *et al.* Heat Assisted Magnetic Recording. *Proc. IEEE* **96**, 1810–1835 (2008).

11. Kang, J.-U., Kim, J.-S., Park, C., Park, H. & Lee, J. A multi-channel architecture for high-performance NAND flash-based storage system. *Journal of Systems Architecture* **53**, 644–658 (2007).

12. Campardo, G. *et al.* 40-mm/sup 2/ 3-V-only 50-MHz 64-Mb 2-b/cell CHE NOR flash memory. *IEEE J. Solid-State Circuits* **35**, 1655–1667 (2000).

13. Park, K.-T. *et al.* A Zeroing Cell-to-Cell Interference Page Architecture With Temporary LSB Storing and Parallel MSB Program Scheme for MLC NAND Flash Memories. *IEEE J. Solid-State Circuits* **43**, 919–928 (2008).

14. Mojumder, N. N., Abraham, D. W., Roy, K. & Worledge, D. C. Magnonic Spin-Transfer Torque MRAM With Low Power, High Speed, and Error-Free Switching. *IEEE Trans. Magn.* **48**, 2016–2024 (2012).

15. Lacaita, A. L. Phase change memories: State-of-the-art, challenges and perspectives. *Solid-State Electronics* **50**, 24–31 (2006).

16. Tsuruoka, T., Hasegawa, T., Terabe, K. & Aono, M. Conductance quantization and synaptic behavior in a Ta_2O_5 -based atomic switch. *Nanotechnology* **23**, 435705 (2012).

17. Gale, E. TiO_2 -based memristors and ReRAM: materials, mechanisms and models (a review). *Semicond. Sci. Technol.* **29**, 104004 (2014).

18. Lee, S., Kim, H., Yun, D.-J., Rhee, S.-W. & Yong, K. Resistive switching characteristics of ZnO thin film grown on stainless steel for flexible nonvolatile memory devices. *Appl. Phys. Lett.* **95**, 262113 (2009).

19. Yingtao Li *et al.* Resistive Switching Properties of Au/ZrO2/Ag Structure for Low-Voltage Nonvolatile Memory Applications. *IEEE Electron Device Lett.* **31**, 117–119 (2010).

20. Park, K. & Lee, J.-S. Reliable resistive switching memory based on oxygen-vacancy-controlled bilayer structures. *RSC Adv.* **6**, 21736–21741 (2016).

21. Manjunath, V. J., Rush, A., Barua, A. & Jha, R. Effect of aluminum interfacial layer in a niobium oxide based resistive RAM. *Solid State Electronics Letters* **1**, 52–57 (2019).

22. Sun, P. *et al.* Thermal crosstalk in 3-dimensional RRAM crossbar array. *Sci Rep* **5**, 13504 (2015).

23. Karakulak, E., Mutlu, R. & Uçar, E. Sneak path current equivalent circuits and reading margin analysis of complementary resistive switches based 3D stacking crossbar memories. **44**, 8 (2014).

24. Aluguri, R. & Tseng, T.-Y. Overview of Selector Devices for 3-D Stackable Cross Point RRAM Arrays. *IEEE J. Electron Devices Soc.* **4**, 294–306 (2016).

25. Woo, J. & Yu, S. Impact of Selector Devices in Analog RRAM-Based Crossbar Arrays for Inference and Training of Neuromorphic System. *IEEE Trans. VLSI Syst.* **27**, 2205–2212 (2019).

26. Sungho Kim, Jiantao Zhou & Lu, W. D. Crossbar RRAM Arrays: Selector Device Requirements During Write Operation. *IEEE Trans. Electron Devices* **61**, 2820–2826 (2014).

27. Chen, Y.-C. *et al.* A Novel Resistive Switching Identification Method through Relaxation Characteristics for Sneak-path-constrained Selectorless RRAM application. *Sci Rep* **9**, 12420 (2019).

Chapter 2

Infrastructure and ALD construction

2.1 Introduction

This chapter is dedicated to elaborate on the infrastructure involved in our work. A better part of the instruments was readily available for the work; however, the heart of our fabrication facility, the atomic layer deposition (ALD) unit, was constructed and optimized in house.

All the works in this thesis involved the fabrication of devices and the subsequent set of measurements. The characterization instruments involved "B1500A parametric analyzer" for current-voltage (I-V), capacitance-voltage (C-V), and conductance-frequency (G-f) measurements. Retention, cycling, and endurance measurements were studied with "B2912A SMU". Both of these units were interfaced with separate probe stations and were user interfaced with "EASYEXPERT" and "QuickIV" programs, respectively. The analyzers to probe-station connections were of "triax" type, to avoid noise and distortion as much as possible.

For local density of state and surface electronics studies, scanning tunneling microscopy (STM) was employed. As described in chapter 3, we also performed the I-V characterizations on minuscule grain and grain boundary structures by the STM. The possibility to measure current as a function of time also in STM enabled us in recording the pulsed current output while the sample under inspection was constantly excited with an external optical pulse train.

The thicknesses of the dielectric layers were confirmed by the ellipsometric study. With an accuracy <2% and resolution in the order of angstroms, we were able to optimize the ALD layers with commendable accuracy.

For fabricating devices, a spin coating system, thermal evaporator, and atomic layer depositor were employed. In chapter 8, the solution based active layer (rGO) was deposited on the silicon-silicon dioxide surface with 1000 rpm with acceleration and deceleration parameter of 100rpm/second. Chapter 3 dealt with zinc oxide as the switching material. The 15nm of ZnO layer was deposited by using a thermal evaporator. A major part of this thesis was based on ALD deposited Al_2O_3 layer[1,2]. Chapters 4 to 6 used the ALD system for the

dielectric deposition. Apart from that, the crossbar sneak-in[3] analysis was also done with ALD Al_2O_3 layers. The details of fabrication steps and parameters are mentioned in the respective chapters.

2.2 Atomic layer deposition

ALD[4,5,6] is one of the most sophisticated systems for dielectric, semiconductor, and electrode deposition. The mechanism of an ALD operation and its classification are elaborated in the next section before illustrating the instrumentation and construction of the in house ALD unit.

2.2.1 Operation principle

Assume that a compound AB has to be deposited in an ALD. We start with two different precursors, with ultra-high purity, AY and XB. The system injects a very small amount of AY in the chamber and waits until excess precursor is drained out by the pump. Owing to the surface thermal equilibrium, only a single molecular layer of AY will get stuck on the substrate surface. In the next cycle, XB is injected, which causes a reaction to take place between AY and XB on the surfaces.

$$AY + XB \rightarrow AB + XY$$

AB forms on the target while the XY is a volatile compound, which can be pumped out from the system. Since the reaction is self-limited, a constant growth rate (layer thickness per cycle). This set of alternate cycles continues until the desired thickness of the film has reached. In ALD the thickness per cycle always remains constant and hence, is a perfect way to theoretically estimate the thickness.

2.2.2 Instrumentation

An ALD system generally comprises the following elements[7,8].

1 The chamber: generally stainless steel (SS 306 or 306L) vacuum compatible chamber is required with adequate vacuum compatible openings to handle the target vacuum. This chamber is the arena where the main action takes place.

2 The pump: a rotary or dry pump creates the vacuum and maintains the low pressure inside the chamber during deposition. A good vacuum pump always ensures the evacuation of unwanted contaminants from the chamber or else they might react with the extremely reactive precursors and compromise the deposition process.

3 Fast switching valve: this is the heart of an ALD system. Very high-speed valve (on/off speed ~10ms or less) with high precision controls the precursor intake by the system. These valves are parallelly connected with a constant flow of inert gas (N/Ar) through one channel. The second channel is connected with the precursor and pumps them inside in short pulses.

4 Digital pressure sensor: this component has two functions. The first is to monitor the pressure inside the chamber and the second is to detect the precursor injection and evacuation. An inert gas purges the system constantly, it makes sure (i) the deposition environment is contaminant-free, (ii) the precursor covers the entire surface area, and (iii) the extraction of the excess precursor. If one precursor does not get purged out completely and the following precursor gas kicks in, then there will be a reaction in the gas channel and the chamber. This kind of unwanted reaction not only hinders the continuity of the process but may potentially clog the expensive ALD valves leading to their complete breakdown. The digital pressure sensor monitors records these operations and feds back the required purging time to the user and lets the user set the cycling program with the same (or higher) purging period. A sampling time of the order of ~100ms is considered to suffice the purpose of pressure change reading.

5 Substrate heater: the deposition requires a certain temperature (varies from material to material) to facilitate the film growth and yield maximum growth per cycle. A substrate holder which also doubles as the substrate heater is a bare essential for the system. A good PID temperature controller with ~1% efficiency should ensure proper temperature control of the substrate.

6 System heater: apart from the substrate, the entire chamber along with the pipelines to the chamber and from the chamber needs to remain at an elevated temperature to avoid condensation of the expensive precursors. Most of the systems maintain the chamber temperature around 120^{o}C and the rest of the connection lines at 80-100^{o}C. The deposition process can commence only after all the lines and the substrate have reached their designated temperatures.

7 Electronics: PLC-based controllers are mainly preferred to automate the operations. PLCs are easy to interface with LABVIEW or other open-source alternatives (pylab, my openlab, etc.) through their outputs are pre-installed with high-speed solid-state relays (SSR). Other than PLCs, a combination of any microcontroller with a set of SSRs should serve the purpose as well.

8 Mass flow controller: the mass flow rate of the inert gas is an important parameter. It decides the base pressure of a system and plays a vital role in deciding the purging time of the precursors.

9 Flow lines: the gas flow lines are needed to be vacuum compatible, heat, and corrosion-resistant. Stainless steel (306 or 306L) base with an inner diameter of 6mm is preferred in most systems. VCR, ferrule, or union junction with proper vacuum compatible gaskets are viable options for inter-connections.

10 Trap (optional): an additional chamber can be installed in the pumping line (after the chamber) to trap most of the additional precursors and prevent the pump from being corroded.

11 RF generator and matching network: plasma is added in a PALD system by an inductive coupled method using an RF power generator. For maximum output power (hence, the least loss) a matching network is always recommended to match the input and output impedance. Generally, a power supply with a frequency rating of 13.56 Mhz and a controllable power of 0-300W is referred for such systems.

12 Consumables: needless to say several consumable products like the purging gas (N or Ar), precursors, oxygen gas (in PALD), etc. are required as well.

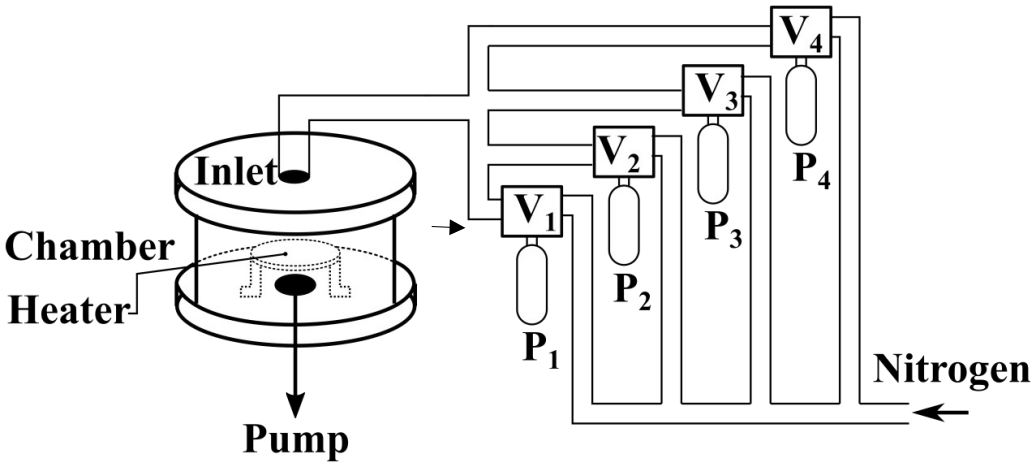


Figure 2.1: A schematic diagram depicting the operation of an ALD system with four fast switching valves (V) and a corresponding precursor (P) containers.

The valve connection configuration (parallel) shown in figure 2.1 ensures no cross-contamination and constant purging of the entire system. Often the showerhead is followed by a diffuser to deliver uniform gas dissipation inside the chamber.

2.3 ALD Construction

Two separate systems were developed from our lab and both of them stand equal importance. The first system was developed in collaboration with Hind High Vacuum (HHV), Bangalore, India. Their existing prototype was modified with a larger deposition chamber suitable for a shower-head model to ensure uniform deposition and to incorporate larger, three-dimensional samples. Such a modification allows the surface to uniformly expose to the precursors, consequently rendering a thin film of the same thickness throughout the substrate. However, a laminar flow is also a deciding factor for the film quality and if not taken care of, it can lead to localized depositions. To counter such a problem, often a diffuser is used at the showerhead.

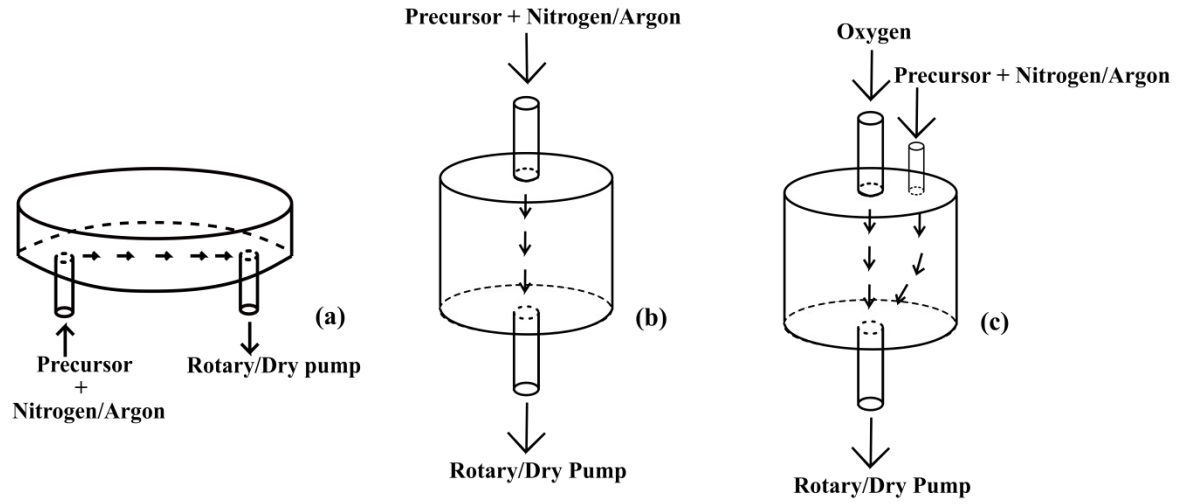


Figure 2.2: (a) Old prototype chamber design, (b) proposed showerhead design and (c) modified showerhead for plasma incorporation.

Although, a necessary modification (see figure 2.2), it came with a cost of additional complexity. The earlier prototype required only two heaters to cover the chamber where the bottom heater served as the bottom plate heater and doubled as the substrate heater as well. The showerhead design, however, insisted a separate substrate heater inside the chamber and additional sidewall heaters to cover the chamber in its entirety. We deposited several layers of Al_2O_3 on a 4-inch silicon wafer in the showerhead chamber and for each of those cases, the thickness error margin measured by an ellipsometer was below 1%. The majority of this thesis' work with Al_2O_3 as the switching medium was carried out in this very system.

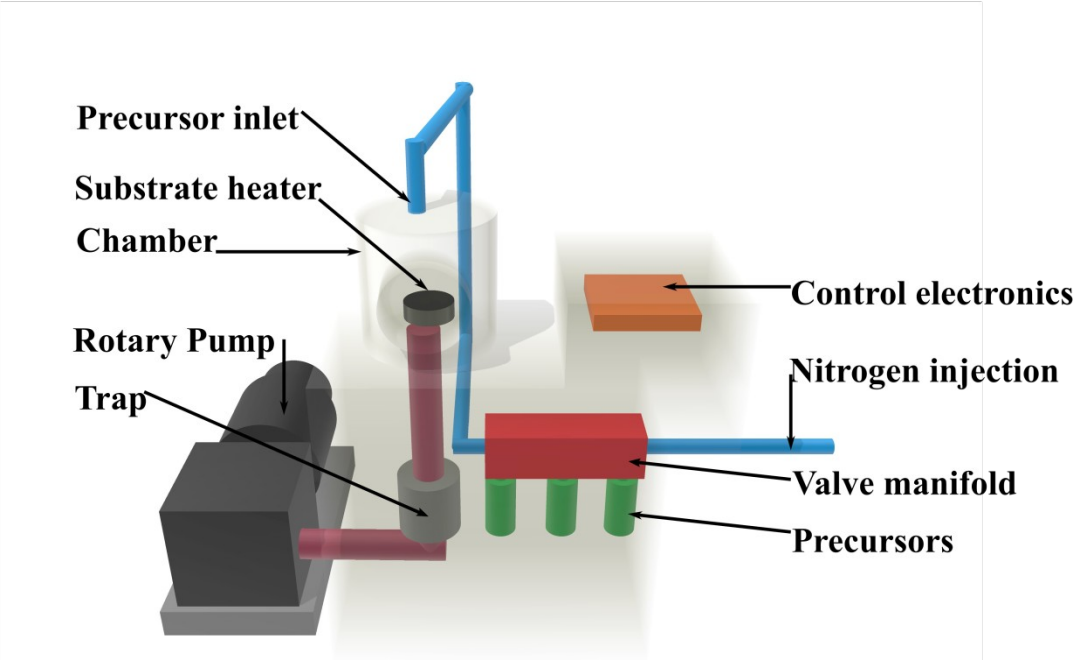


Figure 2.3: A labeled schematic depiction of the commercial atomic layer deposition system.

In this section, we will discuss the in house ALD construction and its working principle. We will go through the necessary components required to build such a system and then revisit each one of them for a brief description. The components' list is as follows.

1 Chamber
2 Rotary Pump
3 Substrate heater
4 Fast switching valves
5 Digital pressure sensor
6 Line heater assembly
7 PLC
8 Mass flow controller
9 Mounting rack

1 Chamber: A cylindrical chamber of inner-height 30cm and inner-diameter of 20cm was constructed with a front-loading dock, and two side viewing ports. Two separate openings at the back of the chamber were connected with the venting knob and digital pressure sensor. The quartz-tube assembly (for PALD) and precursor inlet (via KF-25 port) were installed on the top lid, whereas the pump conjunction was provided to the bottom lid. Please note that the viewports and the loading dock can and have been

replaced with various flanges (fully covered steel flange, high voltage feed-through, etc.) for different purposes.

2 Rotary Pump: It is connected with the bottom lid via a flexible steel pipe.

3 The substrate heater is installed inside the chamber with a PID controller placed outside. The controller can drive the temperature from room temperature to 800°C with 1°C of accuracy.

4 In our earlier system, the fast switching valves were of one plane 3-way connection nature. This type of connection generally has one always-open line and a second line with the control switch. We dedicated the first line for purging gas since it requires uninterrupted flow and the second one was connected with the precursor.

The house built system, on the other hand, had a combination of one 3-way valve and two 2-way valves. To replicate a 3-way valve in a 2-way counterpart, we connected a T-junction at one end, which served as the dedicated line for the purging gas. The other line as usual was used to control the precursor injection in the chamber. Figure 2.4 schematically illustrates the two separate valve manifolds.

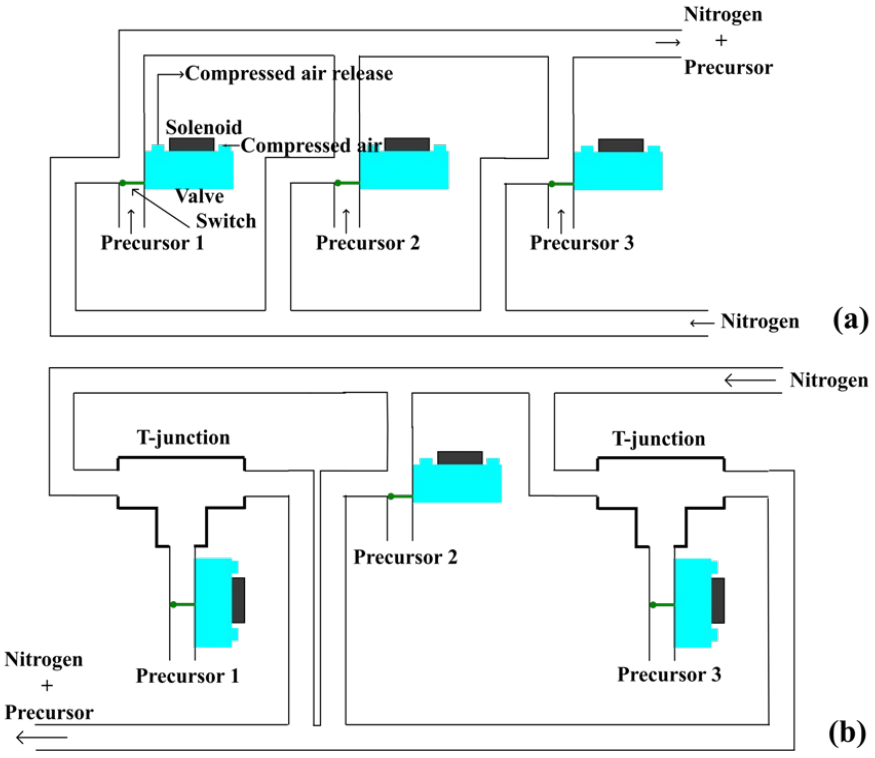


Figure 2.4: A comparative overview of the valve manifold in two separate systems.

5 A pressure sensor was mounted on the dorsal side of the chamber. With a maximum sampling rate of 10/s the device was able to produce enough data points to accurately

read the infusion and extraction of the precursor samples. The data was read via proprietary software.

6 The valve manifold and the steel pipes were heated by heating tapes and controlled by separate PID controllers. To ensure uniform heating, the tapes were further wrapped by several layers of aluminium foils and bound together by kepton tapes.

7 The ALD valves were controlled by a PLC based system. The input power (24V and 100mA) was provided by a separate DC power supply and connected across the output ports of the PLC. The PLC was equipped with SSRs at the outputs with a switching time of 8.5ms and 10.5ms, definitely at par with our ALD valves (5ms). The control program was developed in open-source proprietary software (WPLSOFT).

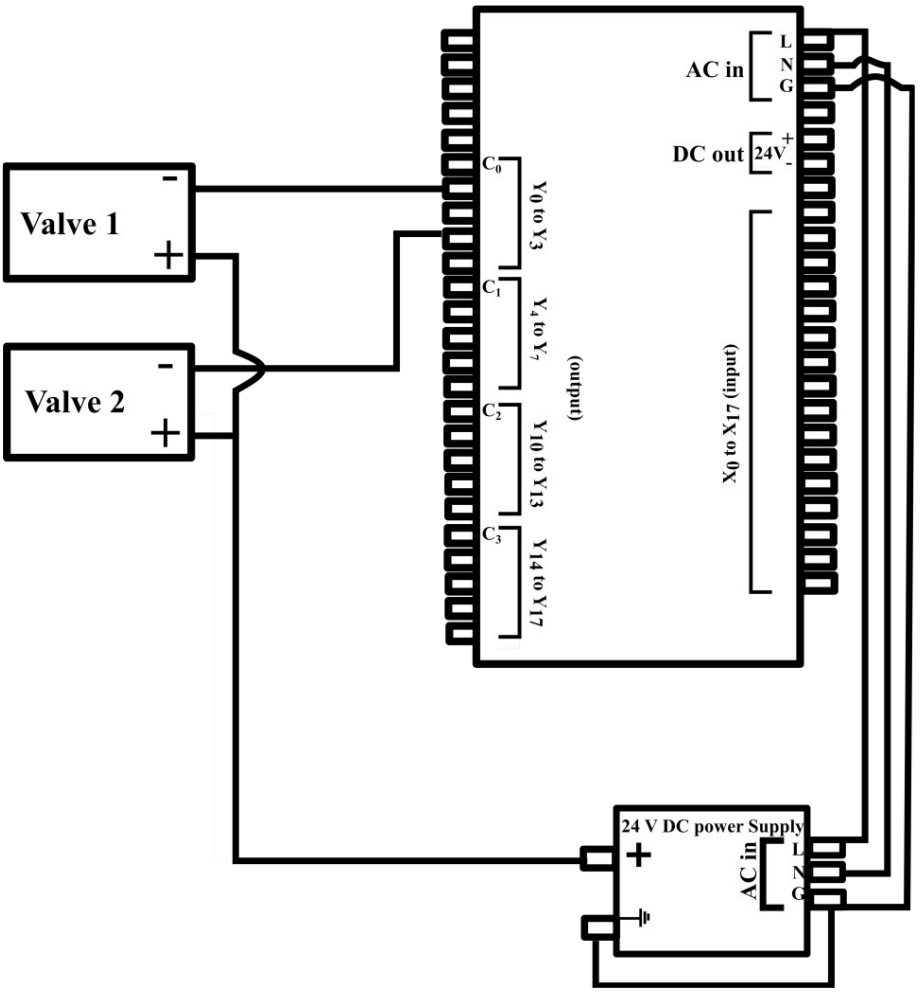


Figure 2.5: Electrical connection of the PLC, power supply, and the ALD valves.

8 Mass flow controller monitors and controls the quantity of purging gas being injected in the chamber. As discussed earlier, it is an essential component of this system and plays the most pivotal role in deciding the base pressure and precursor purging time.

9 The entire setup was stacked on a mounting rack as depicted in the following figure.

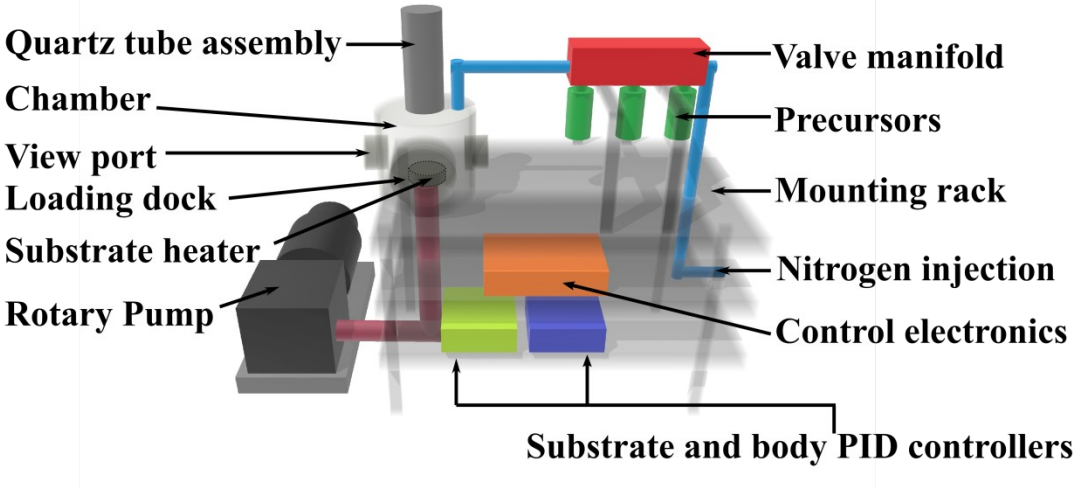


Figure 2.6: A schematic representation of the home-built (EMERALD-ALD) atomic layer deposition system with the labeled components.

The system was tested with TMA and water precursors for Al_2O_3 deposition and diethyl zinc and water combination was tried out for ZnO deposition. Both of the layers showed uniform depositions with the growth rate closely resembling the reported values (1.2Å of Al_2O_3 and 1.9Å for ZnO). With a uniform heating source and showerhead design backed by a diffuser, the precursors were able to cover the 4-inch substrates equally and rendered a highly precise film growth.

2.4 Future works

Currently, the setup is undergoing a transformation to a PALD system. The capacitive coupled plasma was created across the quartz tube mount and one of the viewports (replaced by a high voltage feedthrough). The plasma was created vertically by using two circular metallic meshes as the force and ground electrode. We theorized that the oxygen gas intercepting the plasma path should create enough ionized species to aid faster and more efficient oxidation.

In the next iteration, an RF generator (13.56Mhz, 300W) would replace the high voltage feeds and can directly create oxygen plasma from the inflowing gas.

2.5 Conclusion

This chapter elaborates on the construction of a home-built atomic layer deposition (ALD) system, which was primarily used for the thin film depositions in this thesis. A significant section of the chapter illustrated the operation principle and the component requirement for an ALD system followed by detailed construction details of our own homebuilt thermal ALD system. The authentication of our system was verified from the growth rate and uniformity inspection of the deposited well-reported dielectrics like Al_2O_3 and ZnO.

Figure 2.7 (a) and (b) are the photographs of the in house built ALD system and the system built in collaboration with HHV (Bangalore, India), respectively.

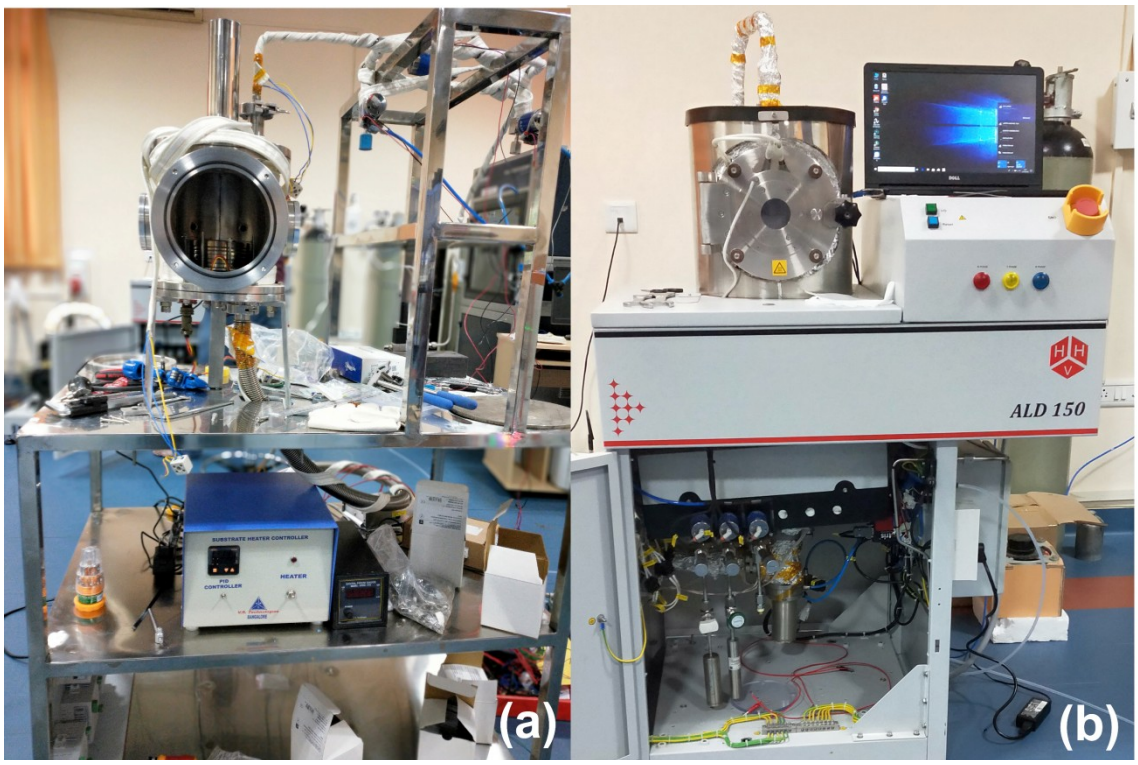


Figure 2.7: (a) The home built Thermal ALD system with the full assembly and (b) the commercial ALD from HHV, Bangalore.

References

1. Groner, M. D., Elam, J. W., Fabreguette, F. H. & George, S. M. Electrical characterization of thin Al2O3 films grown by atomic layer deposition on silicon and various metal substrates. *Thin Solid Films* **413**, 186–197 (2002).

2. Wilson, C. A., Grubbs, R. K. & George, S. M. Nucleation and Growth during Al_2O_3 Atomic Layer Deposition on Polymers. *Chem. Mater.* **17**, 5625–5634 (2005).

3. Karakulak, E., Mutlu, R. & Uçar, E. Sneak path current equivalent circuits and reading margin analysis of complementary resistive switches based 3D stacking crossbar memories. **44**, 8 (2014).

4. George, S. M. Atomic Layer Deposition: An Overview. *Chem. Rev.* **110**, 111–131 (2010).

5. Lisec, T. *et al.* A Novel Fabrication Technique for MEMS Based on Agglomeration of Powder by ALD. *J. Microelectromech. Syst.* **26**, 1093–1098 (2017).

6. Weber, M., Julbe, A., Ayral, A., Miele, P. & Bechelany, M. Atomic Layer Deposition for Membranes: Basics, Challenges, and Opportunities. *Chem. Mater.* **30**, 7368–7390 (2018).

7. Noh, S. J., Lee, S. K., Kim, E. H. & Kong, Y. J. Construction of an atomic layer deposition system for nano-device applications. *Current Applied Physics* **6**, 171–173 (2006).

8. Selvaraj, S. K., Jursich, G. & Takoudis, C. G. Design and implementation of a novel portable atomic layer deposition/chemical vapor deposition hybrid reactor. *Review of Scientific Instruments* **84**, 095109 (2013).

Chapter 3

Effect of Grain and Grain boundary in area scaling

3.1 Introduction

One of the greatest advantages of a resistive RAM device over current FLASH based architecture is the exciting scaling possibilities it can offer. The bit density (bit/In^2) of the current memory technology is limited by the scaling limitations of NAND technology, as well as the fact that their operating speed and processibility are limited only to two dimensions[1,2]. The basic building block of a NAND based FLASH architecture is a MOSFET with an additional floating gate. Configuring multiple 3-terminal (4-terminal, if we consider the substrate) MOSFETs on a surface consumes an enormous amount of space for accommodating different metal contacts, rest aside the channel length. In fact, in the contemporary FLASH technology, the channel length is a very small part of the feature size, and thus scaling down (the channel) does not necessarily promise a huge leap in terms of bit density; although, channel miniaturization delivers higher operating speed and low power consumption. On the other hand, a two-terminal capacitor-like structure incorporated with a crossbar-based arrangement consumes a much smaller footprint for metal contacts, add to that the lack of a lateral conducting channel also contributes in rendering much more efficient feature size.

To cope up with our future data administration needs, several other memory options have emerged as replacement strategies (MRAM, PRAM ,etc.). Among these, Resistive Random Access Memory (ReRAM or RRAM) is one of the most promising candidates because of its simple configuration of a metal-insulator-metal (MIM) capacitor-like structure that allows three-dimensional stacking of devices. This enables higher bit density, large operational speeds, and less programming complexity[3–5] than three-terminal devices. A ReRAM device is simply an insulator stacked between two electrodes which toggle between its intrinsic high-resistance state (HRS) and a low-resistance state (LRS) through a breakdown process depending upon the voltage applied across the electrodes. The proposed breakdown processes involve the formation of conductive filaments by charged vacancies, diffused ions, interfacial oxidation, etc. [6–10].

3.2 Literature survey

A major advantage of the ReRAM architecture is that it allows scaling of the device virtually to sub-nanometer dimensions with the feasibility of multi-layer stacking[11]. The possibility of scaling the ReRAM devices virtually to nano-dimensions has been studied using scanning probe techniques by Hou et al[12] and Kurnia et al[13]. However, as we report here when the device area diminishes, the reliability of the device performance may degrade and the nature of the switching phenomenon may vary from device to device. The performance of the on-chip devices would not be predictable compared to a laboratory-scale standalone, large-area device. ReRAM devices in a chip usually do not have a specific Set or Reset voltage; instead, they have a band of voltages in which they probabilistically toggle between the resistive states[14]. The lesser the reliability of the device, the larger is this spread. This uncertainty can be a consequence of several factors such as process inhomogeneities, stoichiometric anomalies, or thickness nonuniformities in the switching medium, band alignments of electrodes with switching media, or combinations of all these effects. One important parameter that has not been given the importance it deserves so far is the conductivity of individual grains and grain boundaries (GBs) of the switching medium itself. According to previous studies[15,16], different electric conductivities of grains (Gs) and GBs have a trivial influence on the switching behavior, thus inherently introducing certain degrees of unpredictability to the device performance. Adding to that, the random shapes and sizes of grains and GBs and their distribution along the switching surface may cause even more erratic device performance. A few notable works on the explicit contributions of grains and GBs in resistive switching have previously been published, assuming the conducting tip of an atomic force microscope as the top electrode in an attempt to access the Gs and GBs individually[17,18]. Langa et al. have inspected the effect of different conductivities of Gs and GBs due to the abundance of oxygen vacancies in the GB periphery in HfO_2 resistive switches, while Shang et al have reported on the effect of low conducting GB due to oxidation of the oxygen vacancies as a result of annealing WO_3 film.

A comprehensive picture of the influence of the GBs in resistive switching demands analysis of the band diagram of the GBs and their role individually and in combination with grains in a realistic condition. In this work, we re-create the band diagram along the G–GB–G interfaces and simulate the 'different top electrode area' scenario on arbitrarily arranged Gs and GBs to study the reliability of ReRAM devices. A scanning tunneling microscope (STM) gives us sufficient resolution, compared to other scanning probe techniques, to access grain

and GBs and thus sufficient precision over a few Angstroms in the vicinity of the GBs. Figure 3.1(a) is a replica of a memristive device with a polycrystalline switching material, where different possibilities of grains and various types of GBs appear between the electrodes. Figure 3.1(b) shows the experiment conducted to measure the grain and GBs individually using an STM.

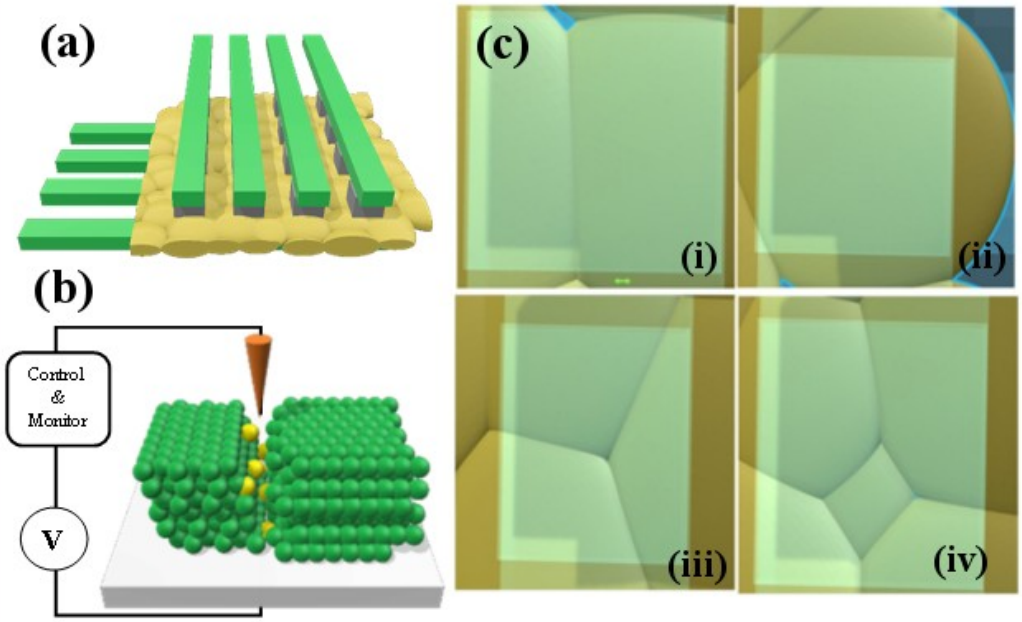


Figure 3.1: (a) A 4×4 crossbar device with polycrystalline film with grains and GBs distributed between the electrodes; (b) a schematic diagram of the experimental setup, where the switching behavior of the grains and grain boundaries are measured using an STM, with the tip as one of the electrodes. The green and yellow spheres represent ZnO molecules and the oxygen vacancies respectively. (c) Different possibilities of electrodes of same area to comprise grains and GBs; (i) is a device with two grains and one doublet GB, (ii) comprise only one grain (iii) with 3 grains 3 doublet GBs and one triplet GB and contains three grains (iv) comprises 5 grains, 7 doublet GBs and 3 triplet GBs.

In this chapter, we show the contribution of individual grains and GBs in the switching characteristics of memory devices and how they eventually influence their reliability statistics. A few instances are demonstrated in figure 3.1(c), where a polycrystalline film is embedded between the crossbar electrodes, which may cover grains and GBs randomly. The following schematics illustrate the way an STM tip was employed to individually access different GBs and Grains on the ZnO surface. In the figure, the green blobs represent individual grain with oxygen vacancies (yellow spheres) populated GBs at their conjunctions.

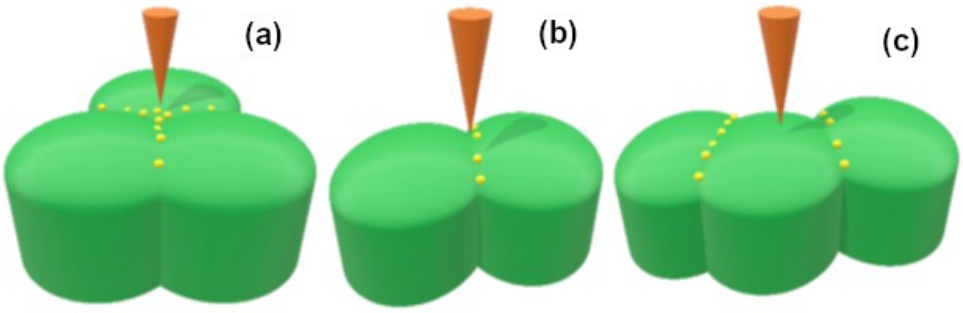


Figure 3.2. (a) A schematic representation of an STM tip (orange cone) scanning (a) a triplet GB, (b) a doublet GB and (c) a grain (G).

The resistive switching properties of ZnO polycrystalline thin films deposited on fluorine-doped tin oxide (FTO) substrate was studied using an STM tip as the top electrode, as schematically shown in figure 3.1(b). Further, by creating gold electrodes of different sizes on the ZnO films, the collective effect of the grains and GBs as an experimental validation of the real case scenario was investigated. The reliability of the devices was quantified using breakdown statistics plotted as Weibull distributions[19] of the Set voltages of the devices. Remarkably, the extrapolation of the Weibull shape function tends to zero as the electrode area scales below 5 nm^2, presaging severe reliability issues at these device dimensions.

3.3 Experiment

ZnO thin films were deposited using physical vapor deposition of ZnO onto cleaned FTO glass plates, at a deposition rate of 0.1nm/s. The thickness of ZnO film used for measurements in this work was 15 nm. The samples were annealed in air at 400 °C for 3 h to ensure maximum oxidation of the sample. Gold electrodes also were thermally evaporated, followed by rapid annealing at 500 °C and rapid cooling in vacuum. The imaging and electrical characterization of the ZnO films were done using a “NanoRev STM” from Quazar Technology Ltd, and the photoluminescence (PL) measurements were carried out using a Horiba Scientific Fluoro Max −4 PL spectrometer.

3.4 Results and Discussions

3.4.1 Current-Voltage characteristics

Figure 3.3(a) shows the STM image of the polycrystalline ZnO thin film, measured at a bias voltage of −1.5 V, applied across the ZnO thin film and Pt/Ir tip. The tunneling current was set at 500 pA. ZnO grains and GBs can be clearly identified in the figure. The GBs between two grains (doublet GBs) and three grains (triplet GBs) are visible in figure 3.3(a).

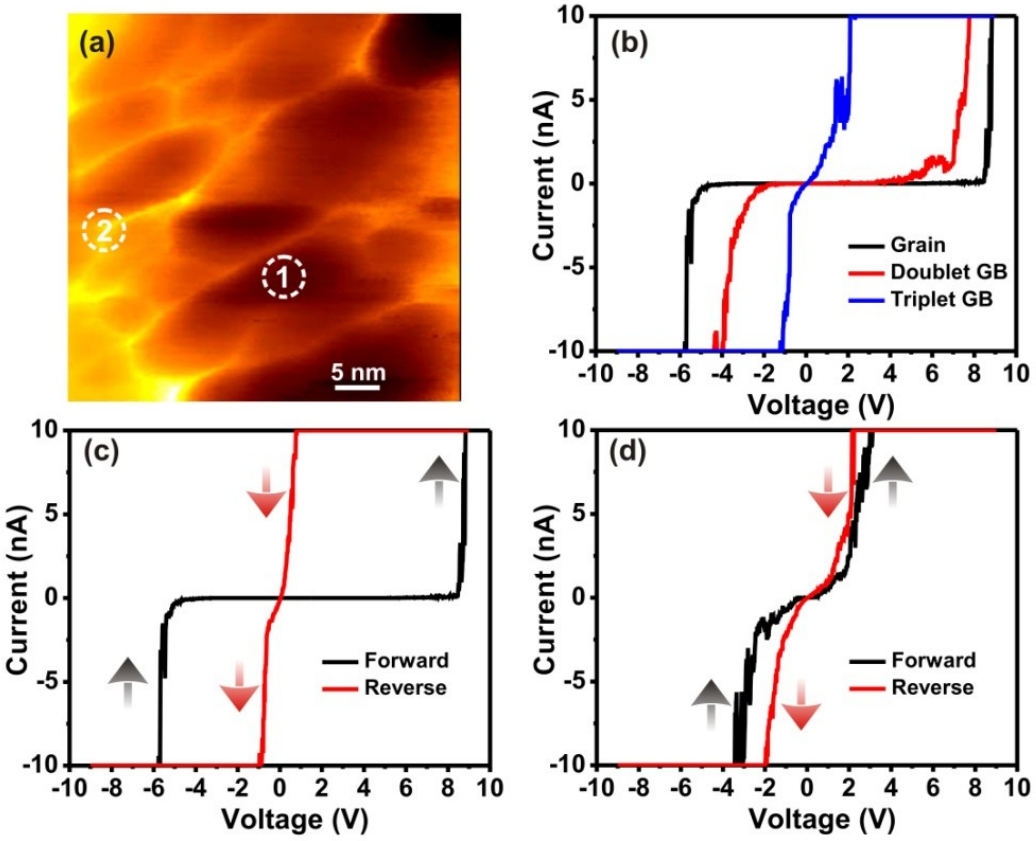


Figure 3.3: (a) STM image of ZnO film showing grains and grain boundaries. Grain boundaries appear brighter than the grains, showing that GBs are more conducting than the grains. This is proved in (b), which shows the tunneling current as a function of the tip-sample bias. Triplet GB's are much more conducting than doublet GB's and grains. (c) The current-voltage hysteresis on a grain (marked 1 in the image) obtained by a voltage sweep -9V to +9V and back to -9V. The SET voltage appears around 8V. (d) The current-voltage characteristics of a triplet grain boundary (marked as 2 in the image), where no explicit switching is observed, rather, appears to be more conducting.

The brightness of the images is proportional to the tunnel current, which suggests that GBs and the area around them are more conducting with respect to the grains. Figure 3.3(b) shows three tunneling current—bias voltage (I-V) characteristics, taken on a single grain, doublet GB, and triplet GB. The resistance of the grain is about 300 GΩ, whereas the resistance of the doublet and triplet GBs are around 5 GΩ and 300 MΩ, respectively, showing that the electronic conduction along the GBs is considerably larger than that through the grains. For grains and doublet GB's a sudden increase in the tunneling current is observed due to the memristive switching in ZnO. Interestingly, GBs do not show current hysteresis, as their conductance is significantly higher than the grains. To measure the hysteresis shown in figures 3.3(c) and (d), a voltage sweep from −9 to +9 V and back to −9 V was applied across the ZnO film and the tip. As shown in figure 3.3(c), the FTO/ZnO/tunnel gap/Pt device is in it's OFF or HRS initially, which switches to the ON or LRS. This transition voltage is taken as the Set voltage of the device. Similar resistive switching events in STM have been reported earlier in metal oxides and sulphides[20], where the switching from HRS to LRS occurs due to the aggregation of metal ions beneath the tip, mediated by the large electric field in the tunnel gap. When these accumulated metal ions form a bridge that short-circuits the tip and sample, they turn into the LRS state. In those experiments, the STM images after switching revealed bright spots corresponding to the accumulated metal ions, while mild or no bright areas could be spotted for ion-driven switching. The absence of such bright spots in the scan images of ZnO after switching indicates that the switching, in this case, is driven by oxygen vacancies[21,22] and toggling of the valence states of oxygen from O^- to O^{2-} [23].

To

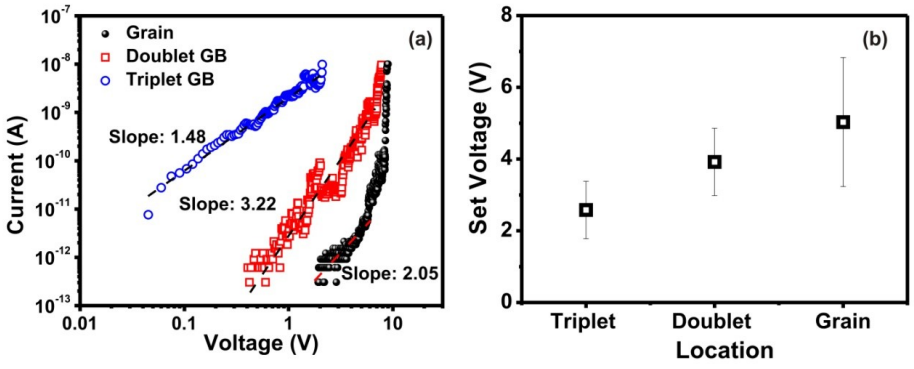


Figure 3.4. (a) Log-log plot of the current-voltage characteristics of the grain, doublet and triplet GBs, and (b) the set voltage at measured on grain doublet and triplet grain boundaries.

understand the charge-transport mechanisms through grains and GBs a log (I)−log (V) plot of the current-voltage characteristics of grain, doublet GB, and triplet GB was plotted as shown in figure 3.4(a). From the slope of the log (I)−log (V) plot, the power law with which the current varies with voltage can be estimated. Before the Set process, grains exhibit a V^2 dependence of current, meaning that the OFF current is dominated by space-charge limited conduction (SCLC), which is normally attributed to the charge injection through oxygen vacancies(Wang et al. 2017; K. Park and Lee, 2016). Interestingly, the doublet GB has a power of 3.2, indicating that the SCLC conduction is assisted by trap ionization and Poole–Frenkel charge transport, whereas the triplet GBs are more conducting, with an SCLC dominated by Ohmic conduction. Therefore, as figure 3.4(b) shows, the triplet GBs exhibit switching at much lower voltages compared to doublet GBs or grains. Apparently, this enhanced conduction through the GBs is due to the abundance of dangling bonds and oxygen vacancies present in ZnO[26,27]. The density of vacancies is much higher along the triplet GBs compared to the doublets. The role of oxygen vacancies in the resistive switching in ZnO has been known for decades; the formation of a conductive filament of oxygen vacancies connected to each other along the electric field lines induces the set process[28].

3.4.2 Quantum Conductance

An additional proof for the hopping of electrons through spatially isolated oxygen vacancies or defects is the signature of quantum conductance with integer multiples of the conductance quantum $G_0=2e^2/h$, where e is the electronic charge and h is the Planck's constant[29]. Figures 3.5(a), (b), and (c) show the quantization observed in conduction through the triplet GBs, doublet GBs, and grains. It's worth noting that the conduction does not occur through a single

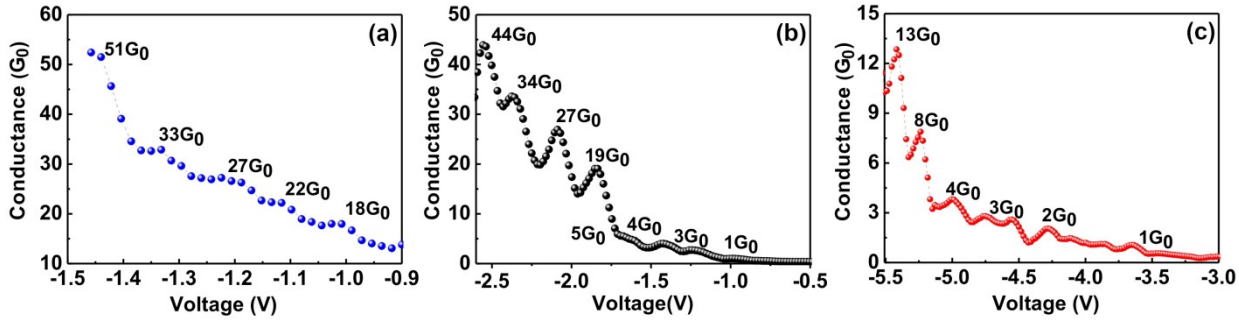


Figure 3.5. The Quantum conductance phenomenon has been recorded along (a) a triplet grain boundary, (b) doublet grain boundary, and (c) a grain. The higher integer values associated with the G_0 in the first two installments clearly suggest the abundance of defects along the boundary periphery.

defect ridden channel (which is highly unlikely), rather through a multiple channelled one; which can be explained by the famous Landuer formula, $G(\mu)=G_0 \sum_n T_n(\mu)$, where G() is the combined conductance, Tn is the transmission eigen value.

Except in very rare instances, quantum conductance was not observed in conductance through grains (figure 3.5 (c)). The large-integer conductance quantum shows that the process is similar to a collective Coulomb blockade phenomenon through GBs whereas the transport through grains shows smaller multiples of G_0 due to high OFF resistance or low OFF conductance of the grains.

3.4.3 G-GB-G Band Diagram

The influence of GBs and their higher conductivity in the switching behavior of ZnO thin films have been previously reported by measuring the conductivity of ZnO thin film and GB enriched nanorod films. The latter was found to exhibit higher conductivity due to the abundance of oxygen vacancies at the interfaces, confirmed from the PL emission spectroscopy measurements (characterized by the green emission). We employed the same scenario to confirm the presence of oxygen vacancies in our samples, also using PL measurement (see figure 3.6 (a)) and LDOS (figure 3.6 (b)) measurement.

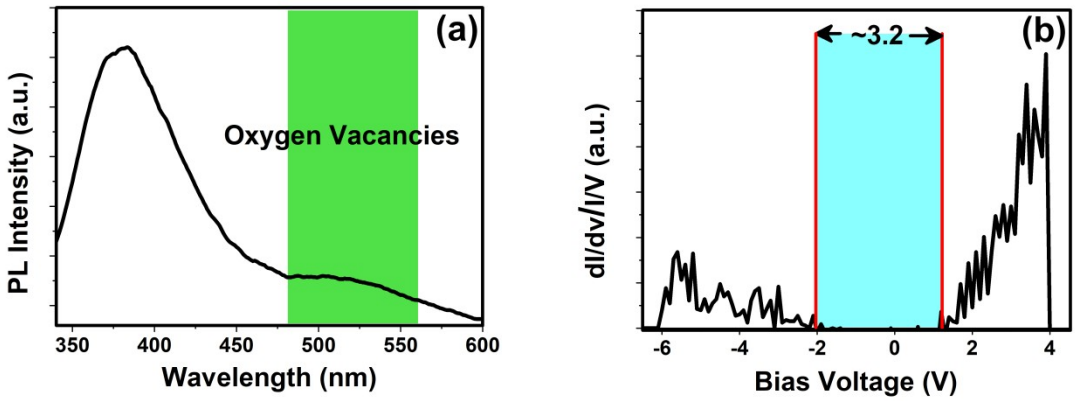


Figure 3.6. (a) The PL analysis of ZnO thin film shows the presence of oxygen vacancies in form of green emission and the semiconductor bandgap calculated from LDOS spectroscopy performed on a ZnO thin film by an STM.

For a better picture of the role of oxygen vacancies in the conduction through GBs, we measured the local work function of grains and GBs individually and reconstructed the band diagram using the tunnel current (I) versus tip-sample distance (z) spectra of individual grains

and GBs. To measure this, first, a doublet GB was located that measured the I-Z spectra at that GB. Subsequently, the I-Z spectra were recorded as a function of the distance from GBs to inner grains. The I-Z spectra generally show typical exponentially decaying characteristics due to the exponential dependence of tunnel current on the tunnel width. The slope of the I-Z spectrum is related to the work function of the substrate through the expression[30].

$$I = I_0 \exp\left(-2\sqrt{2m\varphi/h^2}\,d\right) \tag{3.1}$$

$$\varphi = \frac{h^2}{8m*\pi^2}\left[\frac{d\ln(I_z)}{dz}\right]^2 \tag{3.2}$$

The work function Φ is related to the work function of the sample (Φ_s), tip (Φ_t), and applied bias through the relation $\Phi=(\Phi_s-eV+\Phi_t)/2$, where Φ_s is the surface work function, V is the applied tip-sample bias voltage, and Φt is the tip work function. Figure 3.7(a) shows the I-z plots of grain, doublet GBs, and triplet GBs. Evidently, the slope at the GBs is lower compared to the grains. This shows that the work functions of the GBs are lower compared to the grains. Figure 3.7(b) shows the work function measured as a function of the distance from the selected doublet GBs. A clear dip in the work function is observed at the GBs indicating a different local electronic structure in their vicinities. A reduction of the work function shows that the Fermi level has been shifted upwards, indicating the accumulation of electrons around the GBs. Electron accumulation around the GBs would naturally be due to the positively charged oxygen vacancies distributed along the GBs. This possibility was investigated further by exciting the sample optically. While I-Z spectroscopy was done, the tunnel junction was illuminated with lasers of two different wavelengths, 532 nm (2.33 eV) and 650 nm (1.9 eV). The wavelengths of the laser were chosen such that the red laser (1.9 eV) would not excite the oxygen vacancies while the green laser (2.33 eV) would excite them. The measured work functions in the vicinities of doublet GB's under illumination, with reference to the work function in the dark, are shown together in figure 3.7(c). The work function measured under the illumination of the red laser does not have a significant influence compared to the work function measured in the dark, but there is a systematic reduction in the work function under green laser illumination compared to the work function measured in the dark. This is an indication that the reduction of work function at the GBs is due to oxygen vacancies (V_O), which is the most dominant emission in the fluorescence spectrum of ZnO. V_O holds a positive charge in its both valence (O^+ and O^{2+}) states. This attracts the electrons in the vicinities towards the GBs and forms a negatively charged region.

At thermal equilibrium, the Fermi level is equalized due to the migration of charges, and the conduction bands at the GBs bend downwards due to the net increase in the potential energy of the GBs. The measured band-bending at the GBs is nearly 2.5 eV, well below the Fermi level of the n-type ZnO. This means that the GBs are mostly degenerate because the Fermi level at the GBs crosses the conduction band locally. This explains the high electrical conductivity along the GBs, as shown in figure 3.3(d). Several metal oxides, including ZnO, have been studied in the past; these studies concluded that the higher density of the oxygen vacancies at GBs and their periphery are responsible for their high conductivity [31–33].

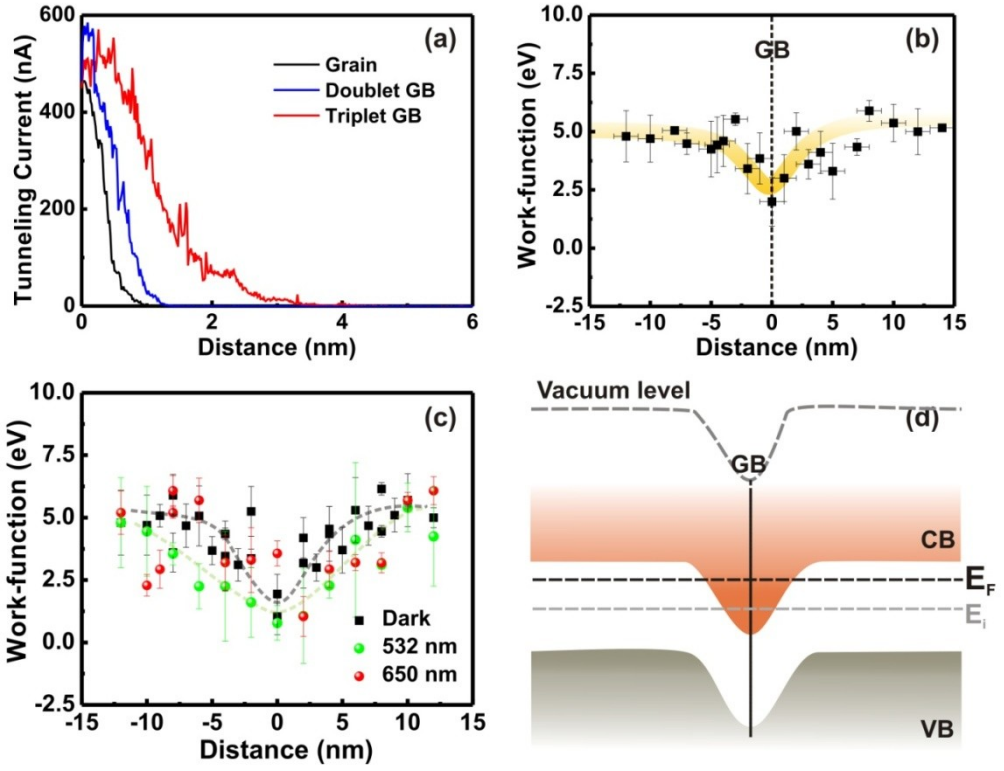


Figure 3.7. ***(a) I-z*** **spectrum of ZnO taken on grain, doublet GB, and triplet GB, from which the work-function was estimated.** ***(b)*** **Work-function estimated as a function of the distance from doublet grain boundaries, measured for several grain boundaries. Distance** ***0 nm*** **represents the position of the doublet grain boundary. The yellow band shown is only a guide to the eye.** ***(c)*** **The work-function of ZnO close to a grain boundary (located at distance 0), measured in dark and under illumination with 532 nm (2.33 eV) and 650 nm (1.9 eV) laser at the tunnel junction. The dotted lines are the average of data points.** ***(d)*** **An approximate band diagram of ZnO at the close vicinity of a grain boundary, indicating the valence band (VB), conduction band (CB), and Fermi level (E_F), reconstructed with the work-function data in** ***(b)*****.**

The influence of oxygen vacancies will be governed by the average bond length as well as the electrostatic potential of the charged vacancies, and they will have a span of 5 Å on either side of the GB. However, from figures 3.7(b) and (c), we see that the bandbending can be extended to the order of 5 nm.

3.4.4 Influence of electrode size

In order to verify the influence of GBs in a more realistic scenario, we simulated the experimental condition using the STM tip on nano-scale electrodes created on polycrystalline ZnO thin films. In order to create sub-20 nm electrodes on ZnO, 10 nm thick gold film was thermally evaporated onto the ZnO thin film followed by annealing at 500 °C. The gold film was agglomerated into isolated clusters as shown in figure 3.8.

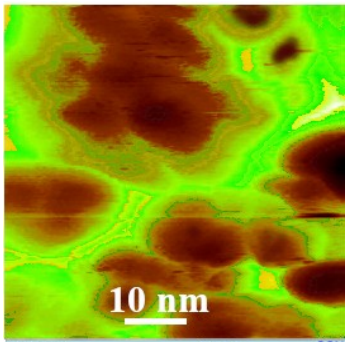


Figure 3.8. An STM image of the gold clusters deposited on the ZnO film. The brownish-golden color represents the gold top electrodes whereas the green portions are assigned to the ZnO film.

We recorded the bandgap of such gold droplets on the oxide thin film to ensure it is

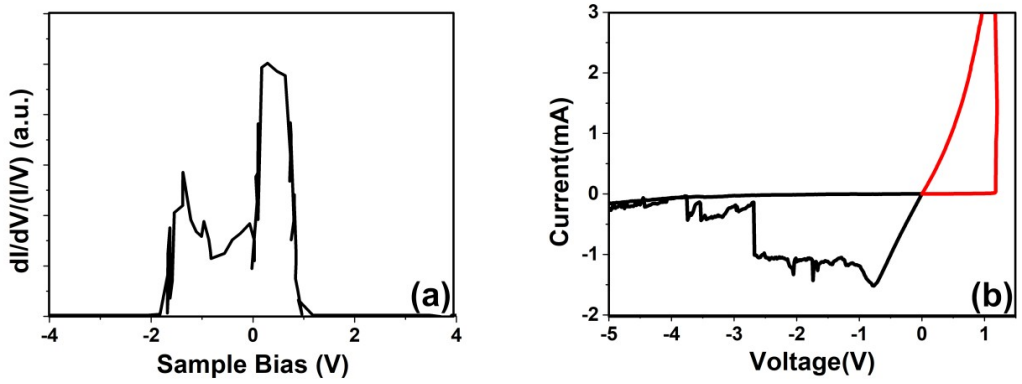


Figure 3.9. (a) The LDOS measurement on the suspected gold electrodes shows the absence of bandgap, thus confirming a metallic presence and (b) I-V characteristics of an Au/ZnO/FTO device clearly showing clear set and reset processes.

highly conducting and indeed do not have a bandgap (see figure 3.9(a)). To confirm the resistive switching capability of such a device configuration, a thicker layer of ZnO was deposited on FTO and gold electrodes were thermally evaporated on them. The resistive

switching characteristics is illustrated in figure 3.9(b). The gold clusters and ZnO grains could be easily identified from the contrast difference in STM images, which was further verified by measuring the local density of states at various locations. Using these gold islands as top electrodes, the switching was measured on device areas ranging from 35 to 800 nm^2. Figure 3.10(b) shows the area dependence of the mean set voltage measured for various gold islands on ZnO. The set voltage drops exponentially with area, but more importantly, the error bars that indicate the randomness of the set process increase with reducing area.

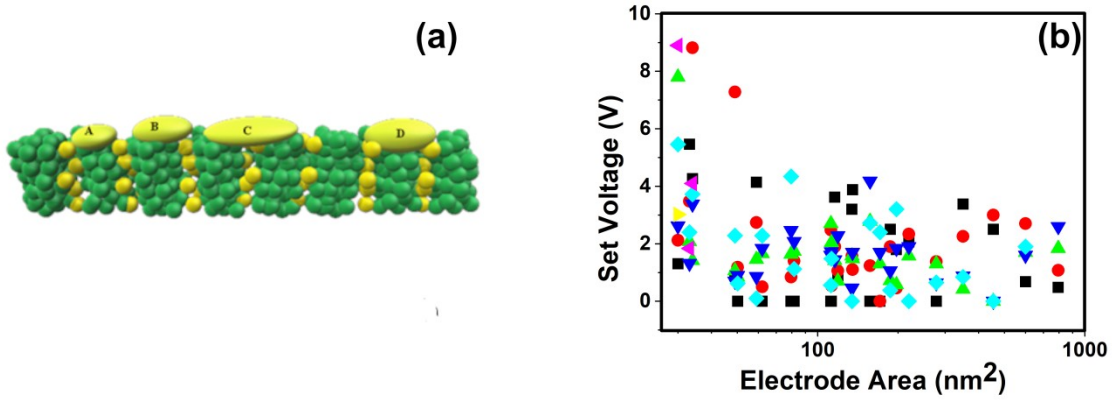


Figure 3.10. (a) A pictorial depiction of how a small gold electrode can randomly sit on a single grain (A), multiple grains (C), single boundary (D) or multiple boundaries, and the set voltage distribution as a function of different gold electrodes.

The random sizes of the grains, GBs, and electrodes lead to an unpredictable arrangement of electrode to G and electrode to GB ratio. Such randomness has been pictorially addressed in figure 3.10(a). The number of Gs and GBs accessed by an individual electrode are completely different in figure 3.10(a); this kind of scenario leads to different HRS conductivity through a particular device. Distribution of set voltages for different sizes of electrodes has been presented in figure 3.10(b); the higher spread at lower size indicates a plummeting of reliability before miniaturization in such technology.

Although this will not be the exact scenario in real crossbar systems, i.e. grains and GBs are unlikely to participate individually in the switching process, they will have a collective effect since the grains and GBs are randomly distributed throughout the architecture. Measuring set voltages of a large set of devices with different areas, the Weibull

distribution of every area was plotted. From these plots, the Weibull slopes (shape parameters) for individual device areas were estimated. Figure 3.11(b) shows the Weibull slopes measured as a function of the device area. Here, the overall reliability of the devices is found to reduce linearly due to the effect of GBs.

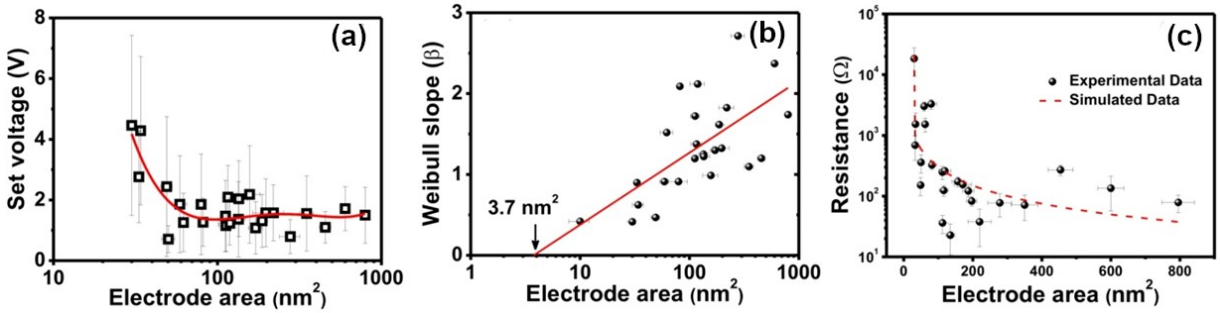


Figure 3.11. (a) is the dependence of the Set voltage on the electrode area. An exponential reduction in the set voltage is observed with increasing area. (b) the Weibull slopes of Set voltages of these FTO/ZnO/Au 'devices' plotted as a function of area, a linear extrapolation shows that the performance of the device will be completely random (β=0) when the area of the device shrinks to 3.7 nm^2; (c) the off-state resistance of the devices as a function of the Au electrode area. The dotted line is a fit using Eq. (3.4).

The exponential-like reduction in set voltage with increasing area shown in figure 3.10(b) illustrates the contribution of GBs in effective switching voltage. The number of GBs increases with increasing electrode area. Therefore, the average set process is dominated by more conducting GBs. The net effect is a lower Set voltage with less randomness. In smaller electrode areas, the grains and GBs dictate the Set voltage individually, thereby elevating the randomness in the set process shown as the large standard deviation of the data. The exponential decrease of the Set voltage can be understood from a simple parallel resistance model comprising the resistances of the grains and GBs. Let us assume that the OFF-state resistances of the grains and GB's are R_G's and R_{GB}'s, respectively. Assuming *n* number of GBs and *m* number of grains underneath the electrode, the effective resistance can be given as

$$\frac{1}{R_{eff}} = \frac{m}{R_G} + \frac{n}{R_{GB}} \tag{3.3}$$

$$R_{eff} = \frac{R_G R_{GB}}{nR_G + mR_{GB}} \tag{3.4}$$

Plotting log (R_{eff}) as a function of electrode area (in nm^2) would give a harmonic decaying nature, as demonstrated in figure 3.11(c). Even though the lowest electrode area is influenced by only a single grain (since its resistance value matches with the average resistance of grains), the resistances of the larger area electrodes are found to be a cumulative effect of both grains and GBs.

3.5 Conclusion

Simulating a close to-real scenario of a nano-scale FTO/ ZnO/Au ReRAM device made of polycrystalline ZnO thin films was probed using an STM tip; in this configuration, we have studied the consequences of downscaling the electrode area in ReRAM devices. We measured the resistive switching characteristics of polycrystalline ZnO and observed that the distribution of crystalline anomalies of the films influences the switching characteristics drastically. The GBs are much more conductive than the grains due to the abundance of oxygen vacancies. To examine this phenomenon, the work function of the grains and GBs were measured by measuring the tunneling current as a function of the tunnel gap or the I-Z spectroscopy. The band-bending at the GBs could be clearly mapped, which indicates that the highly conducting GBs behave like a degenerate semiconductor. After depositing gold nanoclusters onto ZnO, the off-state resistance and the set voltages of nano-scale ReRAM devices was measured as a function of the electrode area. The linear dependence of the Weibull slope on the device area shows that when the electrode area approaches <10 nm^2, the system introduces extreme randomness in the switching behavior. This large resistance distribution at smaller surface areas is related to the possibility of whether the electrode is on top of grain or various types of grain boundaries. In an actual crossbar, ReRAM architecture with the dimension of each device smaller than 100 nm^2, a very similar situation may occur if polycrystalline films of switching materials are employed. That is, if a grain boundary comes between the electrodes, the set voltage of that device will be small or the device will be always in the ON state. A possible solution to tackle this problem is to use amorphous materials as switching media, where grain boundaries can be avoided.

References

1. Pavan, P., Bez, R., Olivo, P. & Zanoni, E. Flash memory cells-an overview. *Proc. IEEE* **85**, 1248–1271 (1997).

2. Zuloaga, S., Liu, R., Chen, P.-Y. & Yu, S. Scaling 2-layer RRAM cross-point array towards 10 nm node: A device-circuit co-design. in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)* 193–196 (IEEE, 2015). doi:10.1109/ISCAS.2015.7168603.

3. Chen, Y.-C., Li, H., Zhang, W. & Pino, R. E. The 3-D Stacking Bipolar RRAM for High Density. *IEEE Trans. Nanotechnol.* **11**, 948–956 (2012).

4. Akinaga, H. & Shima, H. Resistive Random Access Memory (ReRAM) Based on Metal Oxides. *Proc. IEEE* **98**, 2237–2251 (2010).

5. Jo, S. H. & Lu, W. CMOS Compatible Nanoscale Nonvolatile Resistance Switching Memory. *Nano Lett.* **8**, 392–397 (2008).

6. Jo, S. H., Kumar, T., Narayanan, S. & Nazarian, H. Cross-Point Resistive RAM Based on Field-Assisted Superlinear Threshold Selector. *IEEE Trans. Electron Devices* **62**, 3477–3481 (2015).

7. Kamiya, K. *et al.* Generalized mechanism of the resistance switching in binary-oxide-based resistive random-access memories. *Phys. Rev. B* **87**, 155201 (2013).

8. Nauenheim, C. *et al.* Nano-Crossbar Arrays for Nonvolatile Resistive RAM (RRAM) Applications. in *2008 8th IEEE Conference on Nanotechnology* 464–467 (IEEE, 2008). doi:10.1109/NANO.2008.141.

9. Deng, Y. *et al.* RRAM Crossbar Array With Cell Selection Device: A Device and Circuit Interaction Study. *IEEE Trans. Electron Devices* **60**, 719–726 (2013).

10. Mandapati, R. *et al.* The Impact of n-p-n Selector-Based Bipolar RRAM Cross-Point on Array Performance. *IEEE Trans. Electron Devices* **60**, 3385–3392 (2013).

11. Pi, S., Lin, P. & Xia, Q. Cross point arrays of 8 nm × 8 nm memristive devices fabricated with nanoimprint lithography. *J. Vac. Sci. Technol. B Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* **31**, 06FA02 (2013).

12. Hou, Y. *et al.* Multimode resistive switching in nanoscale hafnium oxide stack as studied by atomic force microscopy. *Appl. Phys. Lett.* **109**, 023508 (2016).

13. Kurnia, F. *et al.* Localised nanoscale resistive switching in GaP thin films with low power consumption. *J. Mater. Chem. C* **5**, 2153–2159 (2017).

14. Zhang, M. *et al.* Set statistics in conductive bridge random access memory device with Cu/HfO_2 /Pt structure. *Appl. Phys. Lett.* **105**, 193501 (2014).

15. Huang, C.-Y., Ho, Y.-T., Hung, C.-J. & Tseng, T.-Y. Compact Ga-Doped ZnO Nanorod Thin Film for Making High-Performance Transparent Resistive Switching Memory. *IEEE Trans. Electron Devices* **61**, 3435–3441 (2014).

16. Yoo, E. J., Shin, J. Y., Yoon, T. S., Kang, C. J. & Choi, Y. J. Resistive switching characteristics of a compact ZnO nanorod array grown directly on an Al-doped ZnO substrate. *J. Phys. Appl. Phys.* **49**, 295109 (2016).

17. Lanza, M. *et al.* Resistive switching in hafnium dioxide layers: Local phenomenon at grain boundaries. *Appl. Phys. Lett.* **101**, 193502 (2012).

18. Shang, D.-S., Shi, L., Sun, J.-R. & Shen, B.-G. Local resistance switching at grain and grain boundary surfaces of polycrystalline tungsten oxide films. *Nanotechnology* **22**, 254008 (2011).

19. Mangalam, J., Agarwal, S., Resmi, A. N., Sundararajan, M. & Jinesh, K. B. Resistive switching in polymethyl methacrylate thin films. *Org. Electron.* **29**, 33–38 (2016).

20. Wedig, A. *et al.* Nanoscale cation motion in TaO_x, HfO_x and TiO_x memristive systems. *Nat. Nanotechnol.* **11**, 67–74 (2016).

21. Biju, K. P. *et al.* Parallel memristive filaments model applicable to bipolar and filamentary resistive switching. *Appl. Phys. Lett.* **99**, 113518 (2011).

22. Park, S.-G., Magyari-Kope, B. & Nishi, Y. Impact of Oxygen Vacancy Ordering on the Formation of a Conductive Filament in TiO_2 for Resistive Switching Memory. *IEEE Electron Device Lett.* **32**, 197–199 (2011).

23. El-Atab, N. *et al.* ~3-nm ZnO Nanoislands Deposition and Application in Charge Trapping Memory Grown by Single ALD Step. *Sci. Rep.* **6**, 38712 (2016).

24. Wang, X. *et al.* Influence of metal electrode on the performance of ZnO based resistance switching memories. *J. Appl. Phys.* **122**, 154301 (2017).

25. Park, K. & Lee, J.-S. Reliable resistive switching memory based on oxygen-vacancy-controlled bilayer structures. *RSC. Adv.***00**, 1-3 (2016).

26. Zou, C., Liang, F. & Xue, S. Synthesis and oxygen vacancy-related photocatalytic properties of ZnO nanotubes grown by thermal evaporation. *Res. Chem. Intermed.* **41**, 5167–5176 (2015).

27. Lv, H. *et al.* Thermal Evaporation Synthesis and Properties of ZnO Nano/Microstructures Using Carbon Group Elements as the Reducing Agents. *Nanoscale Res. Lett.* **5**, 620–624 (2010).

28. Cheng-Li Lin *et al.* Unipolar resistive switching and retention of RTA-treated zinc oxide (ZnO) resistive RAM. in *18th IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)* 1–5 (IEEE, 2011). doi:10.1109/IPFA.2011.5992754.

29. Wang, D. *et al.* Quantum conductance in MoS_2 quantum dots-based nonvolatile resistive memory device. *Appl. Phys. Lett.* **110**, 093501 (2017).

30. Altenburg, S. J. & Berndt, R. Local work function and STM tip-induced distortion of graphene on Ir(111). *New J. Phys.* **16**, 053036 (2014).

31. Jose, J. & Abdul Khadar, M. Role of grain boundaries on the electrical conductivity of nanophase zinc oxide. *Mater. Sci. Eng. A* **304–306**, 810–813 (2001).

32. Bai, S.-N. & Tseng, T.-Y. The effect of grain boundaries on the electrical properties of zinc oxide-based varistor. *J. Electron. Mater.* **21**, 1073–1079 (1992).

33. McKenna, K. & Shluger, A. The interaction of oxygen vacancies with grain boundaries in monoclinic HfO2. *Appl. Phys. Lett.* **95**, 222111 (2009).

Chapter 4

Role of the interface state charges in resistive switching

4.1 Introduction

In the first chapter, we have learned the basic structure of a resistive RAM[1], how it toggles between two resistive states (namely, HRS and LRS) and their on-chip arrangement in crossbar architecture. Previous works have described the crossbar architecture as the switching layer and the selector device being embedded between two sets of orthogonally running metal electrodes. However, our current device technology is majorly based on the silicon substrate, also the power and logic circuitry is silicon-based CMOS architectures. Following this argument we reckoned developing silicon-based resistive switches should be an immediate technological priority. Although a handful number of works on MOS based ReRAMs have been reported there is little to no work published which focused upon the nature of silicon-oxide junction and how they can affect the mechanism of switching.

To develop a ReRAM based technique that can be integrated with the current silicon-based CMOS technology, a thorough investigation of the breakdown mechanism of the switching media on silicon substrates is mandatory. Among several factors influencing the switching mechanisms of the switching (dielectric) media, the interface states created by the imperfections of silicon-dielectric interface has been studied very little, despite its gross influence on the operational parameters of resistive memory devices. Generally, this interfacial region is significantly populated with electrically activated defects and traps of various densities of states and resonance time scales. These traps have certain energy levels that resonate at certain frequency bands (around few kHz) and contribute to the device parameters as generation and recombination sites in the device. Such defects eventually influence the off-state current, breakdown of the dielectrics, transconductance, etc. In this chapter, we show that interface trap charges play a vital role in determining the breakdown or SET process of ReRAM devices. For this, ReRAM devices were fabricated on Si <111> and Si <100> surfaces, and the dielectric breakdown behavior together with interface trap density measurements were scrutinized carefully.

The following four sections will gradually unfold this work, starting with providing some basic theoretical backgrounds, then the experimental and measurement procedures,

finally, an elaborate analysis of the capacitance and conductance measurements and their interpretation and conclusion from the results.

4.2 Literature Survey

The nature of silicon <100> and <111> to oxide interface has been extensively studied earlier[2–4]. The silicon atoms are bonded with four other silicon atoms in a tetrahedral configuration in bulk silicon. When oxides are deposited or grown on the silicon surface, the majority of the silicon atoms bond with the available oxygen atoms, the rest is passivated with hydrogen atoms. However, a very small number of atoms fail to form a bond (in some cases they form dangling bonds) with either of them and remain unbonded, giving rise to several unsaturated bonds in such junctions. These broken and dangling bonds on the silicon surface spawn the interfacial defects/charges/traps, which remain embedded between the semiconductor and the dielectric layer. Since <111> surface has a larger atomic density than <100> surface, it is expected to exhibit larger defect density. Moreover, in <111>-oriented silicon, the unsaturated silicon bond is situated at the Si-oxide interface with its unbonded central atomic orbital normal to the interface and in <100>-oriented Si, the four tetrahedral Si-Si directions intersect the interface plane at the same angle, thus forming a weaker trap infested interface than <111>[4]. The following sections of this paper will take these facts into account and will attempt to quantify these defects and consequently will attempt to unveil their role in promoting the conductive bridge formation between the electrodes in resistive random access memory (ReRAM) devices. Let us take a look into the following diagram; this should elucidate the higher concentration of interface trap density (D_{it}) in <111> surface.

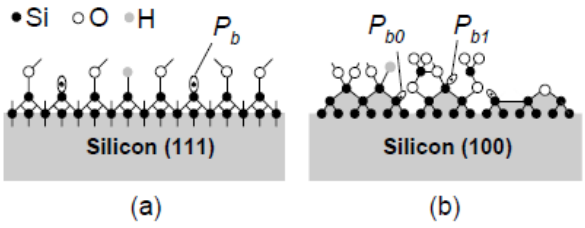


Figure 4.1 The surface defect scenario of (a) silicon <111> and (b) silicon <100> surface (courtesy Schoder). Please refer to the reference for detailed knowhow of the defect orientations and nomenclature.

The effect of the trap charge concentration can influence the flatband voltage (V_{fb}) of a certain MOS device. The mathematical relation can be depicted as follows.

$$V_{fb}=\Phi_{MS}-\frac{Q_f}{C_{ox}}-\frac{Q_{it}(\Phi_s)}{C_{ox}}-\frac{1}{C_{ox}}\int_0^{t_{ox}}\frac{x}{t_{ox}}\rho_m(x)\,dx-\frac{1}{C_{ox}}\int_0^{t_{ox}}\frac{x}{t_{ox}}\rho_{ot}(x)\,dx \qquad (4.1)$$

Where Q_{it} represents the interface trap charges[5]. Since our aim is to inspect the nature of breakdown as a function of the interface trap charges, it is more than necessary to quantify these charges, as precisely as possible. In the next sections, we would be introducing some methods which are popularly practiced by scientists and engineers for such trap investigations.

1 Low frequency versus High frequency capacitance
2 Peak conductance method
3 Conductance versus Frequency method

Low frequency versus High frequency capacitance

By comparing the difference in capacitance in the depletion or inversion region of a particular MOS device[4,6] one can easily calculate the traps' concentration from the following formula.

$$D_{it}=\frac{C_{ox}}{q^2}\left(\frac{\frac{C_{lf}}{C_{ox}}}{1-\frac{C_{lf}}{C_{ox}}}-\frac{\frac{C_{hf}}{C_{ox}}}{1-\frac{C_{hf}}{C_{ox}}}\right) \qquad (4.2)$$

Where q is the single-electron charge in Coulomb and C_{lf}, C_{hf} and C_{ox} are low-frequency capacitance[7], high-frequency capacitance, and oxide capacitance respectively. For further coherence please refer to figure 4.2(a).

Peak conductance method

Another method to compute the D_{it} is from the conductance (G) plot as a function of voltage[4,8]. When voltage is swept across a MOS capacitor at a certain frequency, the trapping and detrapping from the interface defects reaches its peak in the flatband voltage and results in a sudden increase in conductance near V_{fb}. From the peak value of conductance, the D_{it} can be calculated from equation 4.3.

$$D_{it} = \frac{\frac{G_{max}}{\omega}}{\left[\left(\frac{G_{max}}{\omega . C_{ox}}\right)^2 + \left(1 - \frac{C_{max}}{C_{ox}}\right)^2\right]} \cdot \frac{2}{q . A} \left[cm^{-2} e V^{-1}\right] \quad (4.3)$$

Figure 4.2(b) illustrates the capacitance vs. Voltage (C-V) and conductance vs. voltage (G-V) behavior for a low frequency measurement.

Conductance versus Frequency method

This is the most accepted and accurate scheme to quantify the trap density is from the conductance vs. frequency graph. D_{it} can be easily computed from the conductance/angular frequency maxima of the peak taking into account and incorporating it in the next equation.

$$D_{it} \approx \frac{2.5}{q} \left(\frac{G}{\omega}\right)_{max} \quad (4.4)$$

Please note that in equation 4.4 D_{it} is directly proportional to the G/ω value[4,5]; following that argument in figure 4.2(c) we have shown two different G/ω-f plots and pointed out the higher and lower trap density scenario.

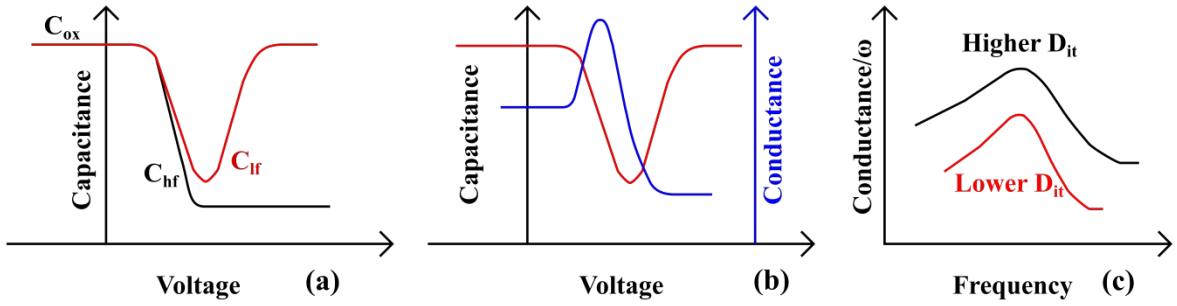


Figure 4.2 (a) The difference in the C_{hf} and C_{lf} can be accounted to calculate the D_{it} of a MOS device, (b) the conductance peak in a G-V sweep near the flatband region and (c) G/ω-f plot showing the G/ω peak at the resonant frequency.

Further explanations of these methods are attempted in the "Results and Discussion" section and their respective advantages and disadvantages shall follow those discussions.

4.3 Sample preparation

The ReRAM devices were fabricated as follows. P-type silicon substrates oriented along <100> and <111> were chosen as substrates and bottom electrodes. The native oxide was removed by a brief HF treatment, followed by a thorough cleaning. 20 nm thick Al_2O_3

was deposited on these wafers using thermal atomic layer deposition (ALD, Hind High Vacuum Ltd.), using trimethyl aluminium (TMA- 99.999% purity, Sigma Aldrich) and water as the precursors. The deposition temperature was 200°C. TMA pulse time was 200 ms, followed by nitrogen purge for 20 seconds. Then water pulsing of 200 ms duration was done followed by 20 seconds nitrogen purging. The growth rate of Al_2O_3 was 1.1Å/cycle. Employing a spectroscopic ellipsometer, we measured the density profiles of the layers and ensured that the densities and refractive indices of Al_2O_3 deposited on both substrate orientations were the same. Gold electrodes were deposited on top of Al_2O_3 films as top electrodes using a shadow mask of 0.33mm^2 opening. The electrical characterizations were performed using Agilent B1500A parametric analyzer connected to a Cascade probe station.

4.4 Results and Discussions

The following figure shows the multiple switching cycles of the fabricated devices for 5nm thick oxide. Please note that in the reset polarity the device acts as a forward biased diode (the bottom electrode being p-type); therefore, accompanied by a substantial amount of off/leakage current; hence the lack in clarity of the reset process.

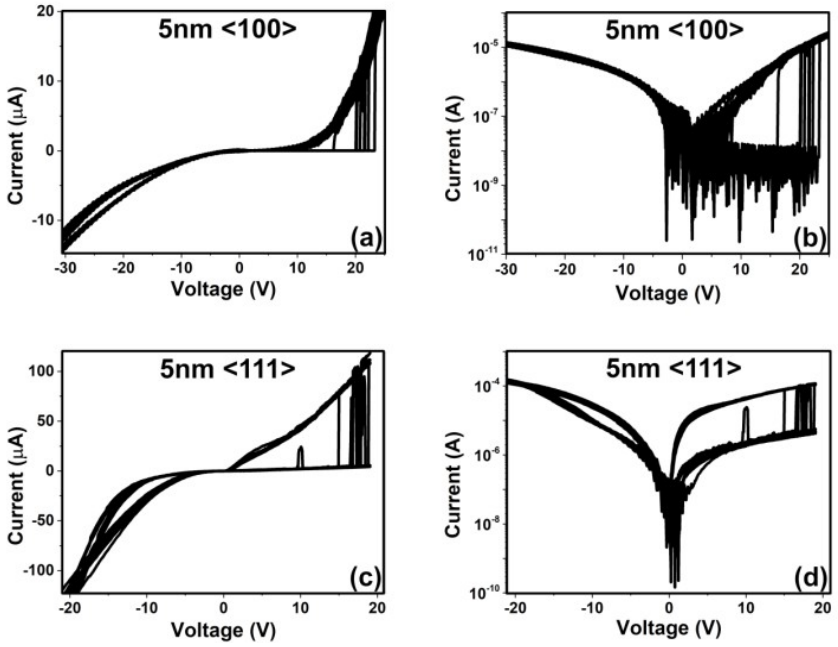


Figure 4.3 (a) and (c) are 10 consecutive switching cycles of 5nm oxides for <100> and <111> substrates, respectively. (b) and (c) are the butterfly plots of the same. The reset behavior is not observed since the devices were fabricated on p-type silicon substrates, which produce a high off-current during the reset event because the device resembles a forward biased diode.

To investigate the set and reset characteristics of the devices on Si<111> and Si<100> orientations, linear voltage sweeps were applied across several devices and their breakdown voltage was monitored. Figure 4.4(a) and (b) show the breakdown process of the devices and their statistical distributions. The large breakdown voltages for 20 nm Al_2O_3 is due to the depletion in the silicon substrate, which acts as an additional resistance in series to the memristor.

4.4.1 Breakdown and C-V characteristics

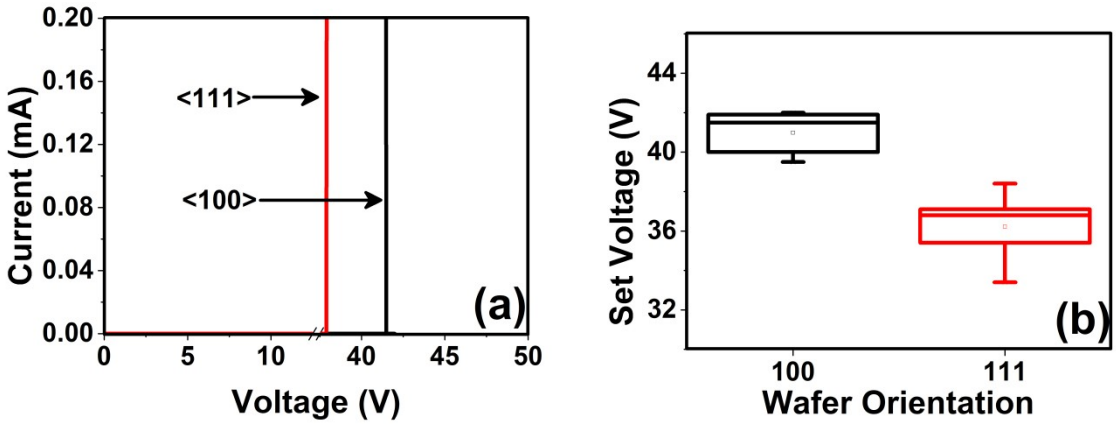


Figure 4.4 (a) Breakdown voltages of 20 nm Al_2O_3 on <100> (black) and <111> (red) silicon, lower breakdown in <111> sample suggests the role of interface traps in switching and (b) statistical distribution of the Set voltages of these devices.

It is clear that even with the same dielectric deposited in identical conditions, <111> shows a noticeably lower breakdown voltage than <100>. Since the Al_2O_3 layer and the top electrodes (gold) are the same, this difference in set voltage indicates the influence of the Si-Al_2O_3 interface. For further inspection, capacitance-voltage (C-V) measurements at low and high frequencies were done on the samples before and after breakdown (Set process). Figure 4.5(a) shows the low frequency (10 kHz) C-V for <100> and <111> devices and a significant difference in the characteristics is evident. Devices fabricated on <111> orientation show the behavior of a heavily leaky capacitor in the inversion region (positive bias) with a characteristic bump, which is not seen in devices on <100> orientation. Since inversion at low frequency is mostly influenced by interface and border defects, the overshooting of the capacitance can be attributed to the leakage due to interface traps. It is well known that a higher interface defect ridden MOS device tends to show the high frequency C-V with a very low slope. The interface defect charges give rise to interface capacitance (C_{it}), which is superimposed to the measured capacitance and the resultant is a depletion capacitance with a

larger spread[4]; the <111> high frequency C-V spans through the voltage axis much more than its counter device as shown in figure 4.5(b), thus inferring a higher population of such defects.

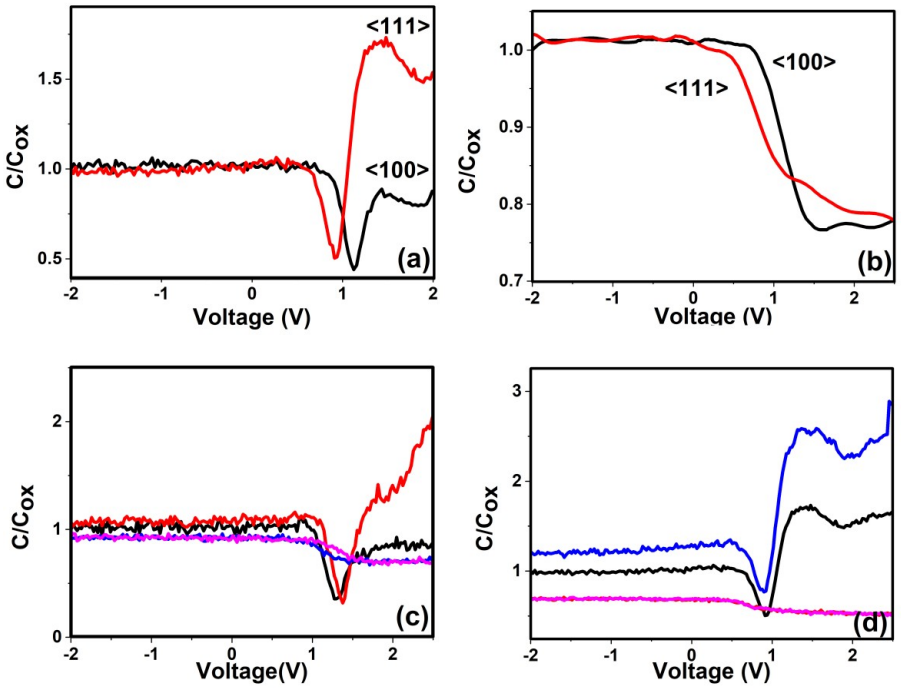


Figure 4.5 (a) Low Frequency C-V measurement of <111> (red) and <100> (black) samples; a higher inversion suggests the presence of interface traps in <111> (b) the broader high frequency C-V showing the presence of interface defects in <111> samples, (c) and (d) showing the shot up nature of capacitance after controlled breakdown for <100> and <111> samples respectively, due to the increase in trap state density.

The concentration of interface charges at the oxide-silicon interface can be quantified from the difference between the low frequency capacitance (C_{lf}) and high frequency capacitance (C_{hf}). Figure 4.5(c) and 4.5(d) are comparative visualizations of LF and HF capacitance for <100> and <111> samples at virgin and post breakdown conditions. An average difference in the inversion capacitance of 50±17 fF and 600±121 fF between C_{lf} and C_{hf} for <100> and <111> respectively was observed and is a clear indication of higher trap density in the latter case. After the breakdown, the devices suffer from an increase of trap states and can be visually verified from the increase in C_{lf} and C_{hf} difference as well; the formula represented as equation 4.2 depicts the quantification of trap charge density (D_{it}) as a function of LF and HF

capacitance. Where C_{ox} is the oxide capacitance, C_{lf} and C_{hf} are the LF capacitance and HF capacitance and q is the electron charge.

4.4.2 D_{it} and τ_{it} estimation

Further, we investigated the trap charge density for <100> and <111> samples at their virgin states and at different levels of controlled break downs to study the evolution of post break down trap density. Besides the C_{lf} and C_{hf} comparison method to find D_{it}, given by equation (4.2), there are alternative ways suggested in the literatures to quantify D_{it}. One of the methods is to calculate it from the peak value of G in C-V measurement at a particular frequency, given by equation 4.3.

However, these methods come with their inherent limitations; it does not necessarily work in the resonant frequency of the traps, rather relies on the user-selected frequencies, thus prone to yield erroneous results. The first method is good enough to quantify trap charge density in the accumulation and depletion region but fails to efficiently estimate it at deep depletion and inversion regions[4]**.** Therefore, estimation of D_{it} from the conductance versus frequency measurements seems to be more appropriate, though it is a tedious task to estimate D_{it} due to different resonance times of traps at various energy levels[5]. However, this method has been proved to be efficient throughout accumulation, depletion, and inversion region with sensitivity in the order of 10^9 and lower[4].

As explained earlier, the trapping and detrapping of charges by these interfacial traps are of lossy nature and this loss reaches its peak at the resonant frequency, which is proportional to the inverse of the lifetime of the carriers in respective traps. Hence, we measured the conductance across the dielectric under certain dc voltage with a superimposed ac voltage (30mV of amplitude) with a linearly increasing frequency. Parallel conductance is the maximum when the applied frequency matches the resonance frequency of the traps, thus, exhibiting a maximum in the G/ω vs. frequency plot. From the apex of the peak of G/ω vs. frequency curve, the trap density can be estimated[4,5,9] from the following formulae,

$$\frac{G}{\omega}=\frac{q\omega\tau_{it}D_{it}}{1+\left(\omega\tau_{it}\right)^2} \qquad (4.5(a))$$

By equating, $$\frac{\partial\left(\frac{G}{\omega}\right)}{\partial\left(\tau\omega\right)}=0$$

we can determine the maximum value of G/ω is reached at

$$\tau_{it} = \frac{1}{\omega} \text{ and} \qquad (4.5(b))$$

$$D_{it} = \frac{2G}{q\omega}. \qquad (4.5(c))$$

However, this relation is valid for only single-level trap energy. In practice, silicon insulator interface posses a band of traps or defects with varying energy levels, thus giving rise to a continuum state of energy distribution represented by

$$\frac{G}{\omega} = \frac{q D_{it}}{2\omega\tau_{it}} \ln\left[1 + \left(\omega\tau_{it}\right)^2\right] \qquad (4.6(a))$$

Adopting a similar approach as the previous equation, G/ω reaches at its maximum magnitude at,

$$\tau_{it} = \frac{1.98}{\omega} \text{ and} \qquad (4.6(b))$$

$$D_{it} \approx \frac{2.5}{q}\left(\frac{G}{\omega}\right)_{max} \qquad (4.6(c))$$

The conductance of memristors fabricated on Si <100> and <111> were measured at inversion voltage (2V in this case) as a function of frequency, ranging from 1khz to 1 Mhz with a resolution of ~1 kHz. Once the resonance was recognized, we selected the specific range of frequencies to achieve a better resolution. Figure 4.6(a) shows the interface trap resonances on both silicon orientations for the virgin devices, and figure 4.6(b) shows the same for the devices after breakdown (Set process). The devices were undergone electrical breakdown (Set) process systematically by controlling the compliance currents as 20nA, 40nA, and 200nA. Figure 4.6(b) has two interesting characteristics. Firstly, the parallel conductance shows a systematic increase when the breakdown process is controlled with a compliance current, and this increase is proportional to the compliance current limit. The only effect of compliance current on memristive switching believed so far is that it restricts the number of filaments forming during the Set process. Interestingly, we find that in addition to the aforementioned reason, the breakdown process significantly increases the parallel conductance as a function of compliance current, which is due to the detrimental increase of interface states. The interface state densities (D_{it}) of <100> and <111> devices before and after controlled breakdowns are plotted in Figure 4.6(c). The D_{it} of devices on Si <111> is at

least one order of magnitude higher than that on Si <100> for every compliance current set for the breakdown process.

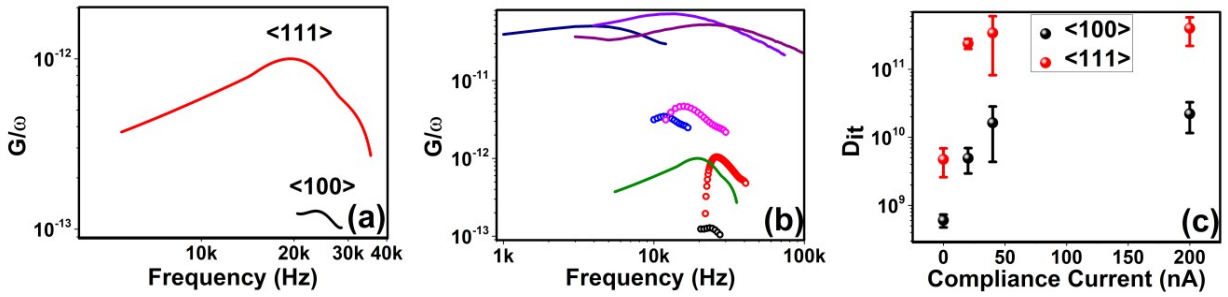


Figure 4.6 (a) A comparative plot of conductance per unit frequency peaks of <100> and <111> sample shows the presence of higher defect density in devices on Si <111>, (b) an increasing defect density from *G/ω* peaks suggests the formation of more defect traps as breakdown becomes stronger; the dotted curves correspond the <100> devices whereas the continuous ones belong to <111> devices and (c) the estimated continuum defect density at virgin and post breakdown states of both samples (zero compliance represents the pristine devices).

Figure 4.7 shows a set of equivalent circuits of the MOS capacitor. The lossy processes involving trapping and detrapping of electrons by these surface defects occur across the R_{it} in the circuit. C_{it} is the capacitance associated with the interface traps which is given by $C_{it}=q^2D_{it}$ and the traps' lifetime is given by $\tau_{it}=R_{it}.C_{it}$. Figure 4.7(b) is a much simpler version with the oxide capacitance (C_{ox}), parallel capacitance (C_p), and the conductance (G_p; equation 4.1).

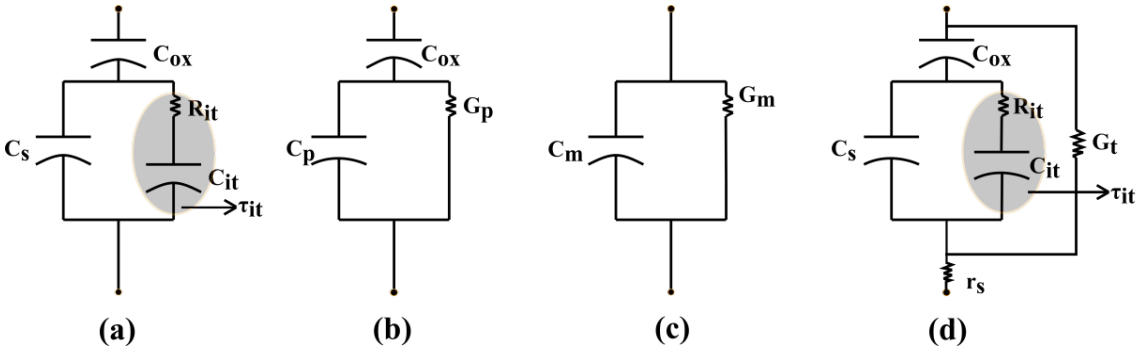


Figure 4.7 (a) Equivalent circuit of a MOS capacitor with dissipation across R_{it} (Resistance offered by interface defects) and the capacitance (C_{it}) offered by the same traps; together they determine the lifetime (τ_{it}) of these defects and (b) the simplified measured circuitry.

Figure 4.7 shows a set of equivalent circuits of the MOS capacitor[4,10]. The trapping and detrapping of electrons by these surface defects are a very lossy process, this loss occurs across the R_{it} in the circuit. C_{it} is the capacitance associated with the interface traps which is given by $C_{it}=q^2D_{it}$ and the traps' lifetime is given by $\tau_{it}=R_{it}.C_{it}$. The next figure (figure 4.7(b)) is a much simpler version with the oxide capacitance (C_{ox}), parallel capacitance (C_p), and the conductance (G_p; equation 4.6(a)). Although, from a data accusation perspective figure 6(c) is much practical and only takes C_m and G_m as measured capacitance and conductance of the whole circuit into account. The last installment of figure 4.7(d) gives better insight into the device's physics when for ultrathin/defect dominated dielectrics tunneling conductance (G_t) and series resistance (r_s) contributes significantly in the measurement.

The second characteristic of the resonance behavior of interface states after the breakdown, as shown in Figure 4.6(b) is that the width (say, full width at half-maximum or FWHM) of the conductance spectra after breakdown has increased considerably. This shows a larger spread in the distribution of the defect states after the breakdown process, presumably due to the multitudes of states newly generated by the Set process. From the FWHM of the conductance spectra, the lifetime or the time constant (τ_{it}) of the interface trap charges can be estimated. To estimate the trap-charge time constants in metal-oxide-semiconductor (MOS) capacitors, the conductance method[4,5] is well received.

This method is a very reliable and direct approach in τ_{it} interpretation. Here we solve from equations 4.5(a) and 4.6(a) and obtain the maximum of G/ω at specific values of τ_{it} as given in equations 4.5(b) and 4.6(b). Since the trap levels after the Set process are much more

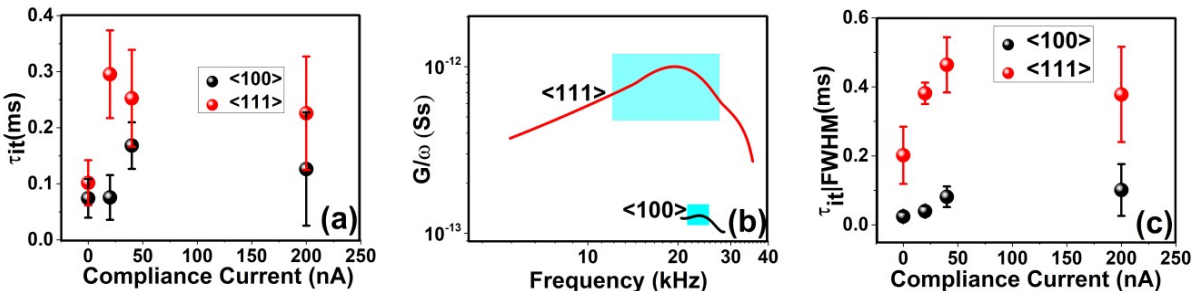


Figure 4.8 (a) lifetime (τ_{it}) measurement shows the generation of more species of interface defects post breakdown, (b) τ_{it} spread is calculated across the FWHM of the *G/ω* peaks, and (c) the same plotted for both the devices before and after the breakdown.

spread over the bandgap of silicon, calculating lifetime at a specific frequency (ω at $(G/\omega)_{max}$) will not be enough; therefore, we calculated the τ_{it} at G/ω peak and its distribution at the either ends of the FWHM of the peaks. Figure 4.8(c) shows the τ_{it} estimated for Si <111> and <100> devices in pre- and post breakdown conditions as a function of the compliance current. All devices appear to show an overall increase in the lifetime values after the breakdown. The increase in the residence time of the charges clearly indicates the increase in the interface states as a function of compliance current. The larger lifetime of the interface defects for <111> samples with respect to <100> devices can be attributed to the presence of more trap charges at the Si<111>/Al_2O_3 interface.

4.4.3 Interface trap and thickness scaling

The immediate consequence of the increased interface states in silicon <111> interface would be the increasing mirror charge at the other side of the Al_2O_3. The presence of these mirror charges has two consequences; first, a larger band bending towards the Au electrodes, and secondly, the larger Schottky barrier lowering (reference), which is the conduction band offsets between Al_2O_3 and silicon. Both these reasons cause enhanced charge flow through the dielectric layer, eventually promoting the onset of the dielectric breakdown.

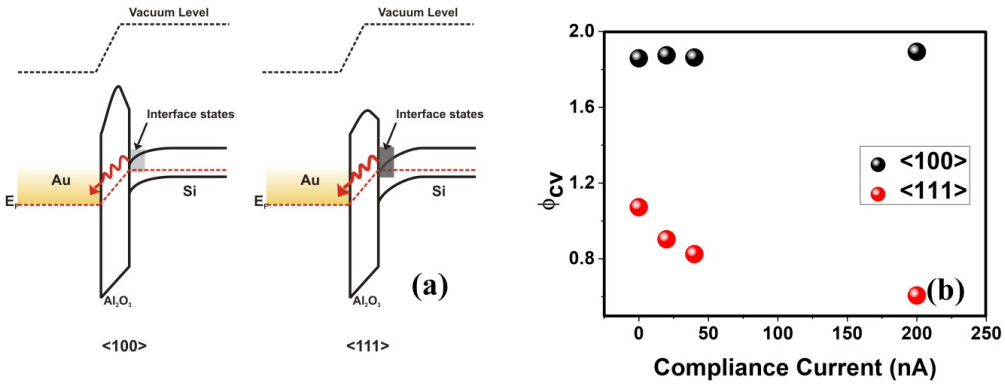


Figure 4.9 (a) Schematic representation of the band diagram of Au/Al_2O_3/Si for Si <100> and Si <111> substrates. The density of interface states in Si <111> is much larger and spread over the bandgap of silicon, which increases the availability of more electrons at the Si/Al_2O_3 interface for <111> orientation and (b) the estimated barrier height as a function of the compliance current. The barrier height for Si <111> / Al_2O_3 falls rapidly with increasing compliance current indicates the Schottky barrier lowering due to the interface traps.

The barrier height (conduction band offset) between Al_2O_3 and silicon can be estimated from the Mott-Schottky plot ($1/C^2$ versus V) measurements[11].

$$\frac{1}{C^2}=\frac{2}{q\varepsilon_0\varepsilon_i N_A A^2}\left(V_D-\frac{kT}{q}-V_R\right) \tag{4.7}$$

$$E_F=\frac{kT}{q}\ln\frac{N_C}{N_D} \tag{4.8}$$

$$c_2\approx\frac{N_D(Experimental)}{N_D(Theoritical)}=\frac{\varepsilon_i}{\varepsilon_i-qt_{ox}D_{it}} \tag{4.9}$$

$$\Phi_{CV}=(c_2V_0+\frac{kT}{q}+E_F) \tag{4.10}$$

Where V_0 is the intercept voltage from Mott-Schottky plot ($V_0=V_D-kT/q$), N_D also can be calculated from the same plot. "c_2" is the relation between experimental and theoretical N_D and finally, Φ_{CV} is the computed effective barrier height.

Figure 4.9(a) is a schematic representation of the $Au/Al_2O_3/Si$ band diagram for Si <100> and <111> orientations. Increased interface states in Si <111> would correspond to a larger Schottky barrier lowering in Al_2O_3, due to the mirror charges accumulated. With the increase of trap defects, the mirror charges also increase and eventually leading to lower barrier height. Figure 4.9(b) shows the estimated barrier heights for <111> and <100> orientations. It is found that, remarkably, there is a drastic reduction in the barrier height for <111> compared to <100> surface, and has a strong dependence on the current compliance set for the breakdown process. However, for <100> devices the low trap density hardly causes any banding and therefore the barrier height remains comparatively unchanged.

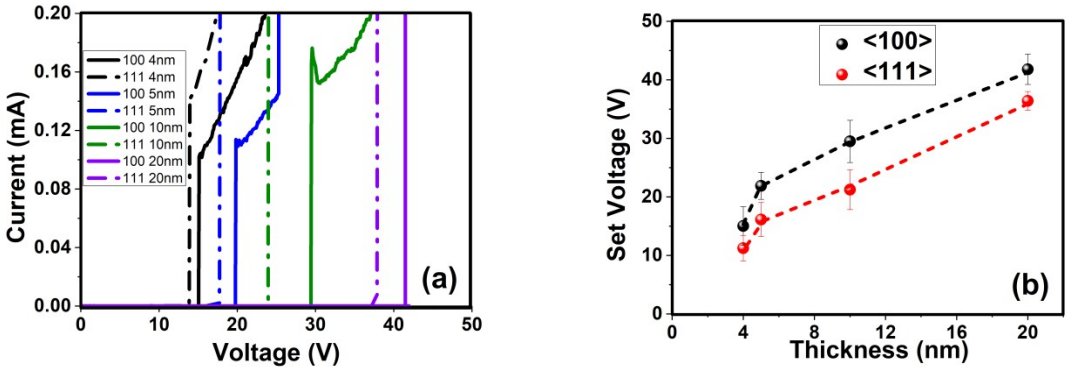


Figure 4.10 (a) Breakdown voltages of <100> (dotted) and <111> (continuous) distribution demonstrated the effect of interface defects in switching and (b) shows the extreme effect of the same in ultrathin dielectrics.

Now, having understood that the interface states have a strong influence on the Schottky barrier lowering, its influence on the thickness scaling of the Set process has been investigated. Employing atomic layer deposition, devices with varying dielectric thickness, ranging from 4 nm to 20 nm (4nm, 5nm, 10nm, and 20 nm) on both <100> and <111> silicon wafers were fabricated with gold electrodes. Figure 4.10(a) shows the breakdown (Set) process of the Al_2O_3 films on both substrates as a function of the film thickness.

The emergence of interface states has been further verified by analyzing the behavior of pre-breakdown current as shown in figure 4.11, where, the devices on Si <111> shows a much stronger space-charge limited current (SCLC) than on Si <100>. Due to the broader distribution of interface charges as mentioned in Figure 4.6 and 4.8, Si <111> show a broader SCLC current region of Child's law ($I \propto V^2$)[12,13] due to charged interface states. It is also noticed that the onset of Fowler-Nordheim tunneling happens much earlier in devices on Si <111> than in Si<100>, illustrating the influence of interfacial traps in breakdown mechanism.

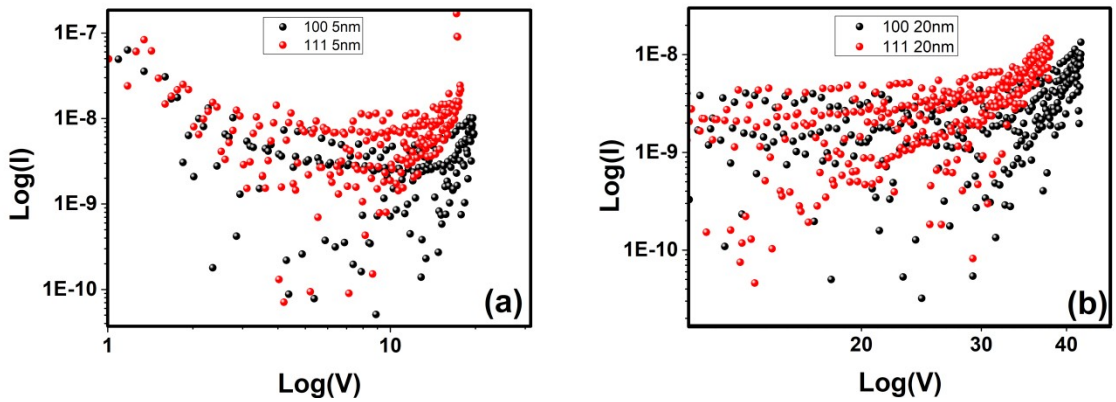


Figure 4.11 The earlier onset of SCLC behavior in <111> devices in (a) 5nm and (b) 20nm devices.

The breakdown characteristics shown in Figure 4.10(a) depicts that the Set voltage for samples on Si <111> substrate is systematically lower than that of the samples on <100> substrate. The density of interface states for a given silicon orientation being the same, reduction in the thickness would cause enhanced hot electron injection from the inversion layer in the silicon substrate to the gold electrodes through the Al_2O_3 layers. Since more interface states are exposed while applying gate bias, Si <111> demonstrates an earlier breakdown than Si <100> devices.

4.5 Conclusion

We have fabricated MOS ReRAMs with atomic layer deposited Al_2O_3 of different thickness on silicon wafers oriented along <100> and <111> to investigate the role of interface traps in dielectric breakdown voltage in general, which is directly employed as the ReRAM Set voltage. By Measuring the interface trap density of both sets of devices, we found that the <111> samples are more prone to generate interface states under voltage stress and thus tend to show a lower breakdown voltage. The controlled breakdown experiments and subsequent D_{it} measurements revealed that the density of interface states gradually increases with increasing compliance current, which gives an additional dimension to the formation process of ReRAM devices. It shows that the compliance current not only influences the conductive filament formation, but it heavily influences the interface trap density as well. This has been further verified by estimating the characteristic time constant associated with the interface traps, which increases with increasing current compliance. This also indicates the presence of deeper interface states generated when the compliance current is larger. The spread of interface states after the Set process also appears to be larger for Si <111> samples. Followed by this, the breakdown characteristics of devices on Si <100> and <111> were measured as a function of the dielectric thickness. This was done to analyze the influence of the interface states on the breakdown characteristics of the ReRAM devices.

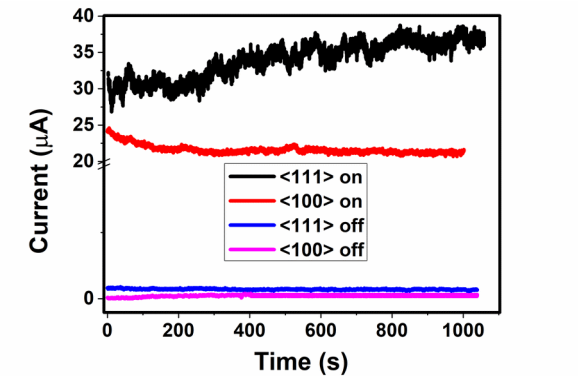


Figure 4.12 The endurance study in <111> sample shows a gradual increment in the current magnitude owing to the larger population of the interface defects. Contrary to that the <100> samples on state suffer a minor decrement in the current value before settling to a constant value.

We find that the breakdown voltage is not linear with thickness; it drops more drastically as the thickness becomes less than 5nm, where electron tunneling from the interface states

becomes more and more prominent. This observation would be very important for the scaling down of ReRAM devices, where reversible dielectric breakdown processes are the basic working mechanism, and the devices are expected to have a cycling capacity over a million times to be technologically feasible. Our observations indicate that the interface states can cause gradual and severe damage to the reliability of the ReRAM when the electrodes or dielectric materials are prone to have interface states.

Another consequence of the defect ridden interface can be observed in the endurance plot (see figure 4.12). At HRS the <111> device is characterized with a slightly higher off-current than its counterpart. Interestingly in the on-state, the <100> devices tend to go through an initial decay but eventually settle to a constant value, while the <111> samples' on-state seems to suffer a gradual increase in the current value; which can be attributed to a larger build-up of interface defects along the silicon-oxide junction.

References

1. Yingtao Li *et al.* Resistive Switching Properties of Au/ZrO_2/Ag Structure for Low-Voltage Nonvolatile Memory Applications. *IEEE Electron Device Lett.* **31**, 117–119 (2010).

2. Kooi, E. *The Surface Properties of Oxidized Silicon*. (Springer Berlin Heidelberg, 1967). doi:10.1007/978-3-662-40210-8.

3. Baccarani, G., Severi, M. & Soncini, G. Electrical Characteristics of MOS Structures on and Oriented N-Type Silicon as Influenced by Use of Hydrogen Chloride during Thermal Oxidation. *J. Electrochem. Soc.* **120**, 1436 (1973).

4. Schroder, D.K., Semiconductor Material and Device Characterization. in *Semiconductor Material and Device Characterization* i–xv (John Wiley & Sons, Inc., 2005). doi:10.1002/0471749095.

5. E H Nicollian_ J R Brews-MOS (metal oxide semiconductor) physics and technology-(Wiley 1982).

6. Viðarsson, A. M. Investigation of interface and near-interface traps in silicon carbide MOS-capacitors using capacitance and conductance techniques. 56 (2018).

7. Kuhn, M. A quasi-static technique for MOS C-V and surface state measurements. *Solid-State Electron.* **13**, 873–885 (1970).

8. Min, K. H. *et al.* Investigation of interface characteristics of Al_2O_3/Si under various O_2 plasma exposure times during the deposition of Al2O3 by PA-ALD. *Curr. Appl. Phys.* **19**, 155–161 (2019).

9. Tataro, A. Analysis of density and time constant of interface states of MIS device by conductance method. *APPL PHYS* **54**, 5 (2016).

10. Yarita, I., Sato, S. & Omura, Y. Proposal of Physics-Based Equivalent Circuit of Pseudo-MOS Capacitor Structure for Impedance Spectroscopy. *IEEE J. Electron Devices Soc.* **4**, 169–173 (2016).

11. Sze S. M., PhysicsOfSemiconductorDevices (John Wiley and Sons INC. 1969).

12. Chen, S.-X., Chang, S.-P., Chang, S.-J., Hsieh, W.-K. & Lin, C.-H. Highly Stable Ultrathin TiO_2 Based Resistive Random Access Memory with Low Operation Voltage. *ECS J. Solid State Sci. Technol.* **7**, Q3183–Q3188 (2018).

13. Borisova, T. M. & Castro, R. A. Mechanism of charge transport in Si/Al_2O_3/Al structures. *J. Phys. Conf. Ser.* **461**, 012017 (2013).

Chapter 5

Evidence of mobile ions in ReRAM devices

5.1 Introduction

Resistive switches[1,2] have been the focus of attention of various research groups around the globe for almost a decade. However, the nature of the switching mechanism is yet a controversial subject. The three widely accepted theories of switching (Set) process are

1. Metal ion migration across the electrodes[3],
2. Defect/vacancy based breakdown[2] and
3. Interface mediated switching[4].

These phenomena are discussed below in detail.

Metal ion migration

When a positive bias is applied on the top electrode (a metal, potentially with low electron affinity), metal atoms lose electrons from their valance band and become positively charged metal ions. With a positive potential on the top electrode (TE) and a negative potential on the bottom electrode (BE) the positively charged metal ions migrate towards the BE, where it accepts electron/s to neutralize itself to a metal atom. This process continues until a continuous metallic path forms between the two electrodes; thus completing the set process in a ReRAM. Upon applying an opposite electric field the metal ions from the BE periphery start moving towards the TE and when the metallic filament retreats far enough to leave with little to no conduction through the switching layer the device resets.

Defect/vacancy based switching

Ionizing noble metals like gold or platinum is difficult due to their high electron affinity. For these systems, a high electric filed creates more defects (say oxygen vacancies in an oxide) and its negative counterpart, like the oxygen ions. With the course of time, the negative ions migrate towards the anode, leaving behind a trail of positively charged vacancies. Once the process is completed, the electrons hop across these sites to reach the anode from the cathode, thus setting the device on. When we apply a reverse field the opposite migration takes place leading to the switching off of the device.

Interface mediated switching

Electrode-dielectric interfaces often contribute majorly in deciding the switching process in a resistive memory device. The top electrode-oxide interfaces are widely known to

act as an oxygen ion storage site, which facilitates switching. Although the bottom electrode and switching layer interface is comparatively less addressed in the role of switching, we have found some correlations between them as addressed in chapter 3.

As we have done a thorough analysis of the role of the silicon-oxide interface in resistive switching and its aftermath in vertical scaling, in this chapter we make an attempt to settle the metal ion versus vacancy switching battle in MOS resistive memories. Like the previous chapter, we would mainly focus on the capacitance-voltage (C-V) measurements of our MOS devices for this inspection. Gold (Au) based and Silver (Ag) based top electrode (TE) MOS devices were fabricated with ALD deposited Al_2O_3 as the switching medium for this study. With Ag ion diffusion we should be able to observe an increase in mobile ions in the system with respect to the Au based devices; the latter being known not to diffuse in the system, rather the switching takes place because of the oxygen ion (V_i) and oxygen vacancy (V_o) generation in the oxide.

5.2 Literature survey

The three inspection schemes, which could detect the presence of possible mobile ions in the switching medium or at the least lead us in the right direction, are discussed in this section.

5.2.1 Oxide charges

Different charges can be trapped inside the oxide[7,8] of a capacitor. Due to the defects/traps present in the oxide or being injected through the gate or channel can also modify the charge content of the dielectric. These are known as oxide charges. The effect of these FO traps is directly reflected in the hysteresis of the C-V measurement. The charge injection can be very much dependent on the barrier lowering of the oxide heavily, which is why these effects can be more pronounced in the thinner oxides.

The expression for the flatband voltage in a MOS structure is given by

$$V_{fb}=\Phi_{MS}-\frac{Q_f}{C_{ox}}-\frac{Q_{it}(\Phi_s)}{C_{ox}}-\frac{1}{C_{ox}}\int_0^{t_{ox}}\frac{x}{t_{ox}}\rho_m(x)\,dx-\frac{1}{C_{ox}}\int_0^{t_{ox}}\frac{x}{t_{ox}}\rho_{ot}(x)\,dx \qquad (5.1)$$

Where ρ_{ot} is the oxide trap area density. While measuring the high frequency capacitance versus voltage (C-V) data by sweeping the voltage from positive to negative and vice-versa, fixed oxide traps can lead to a shift in the C-V measurement resulting in a hysteresis. Often the hysteresis is measured in the form of the charge (Q_{ot}) as a product of the oxide capacitance (C_{ox}) and the change in flatband voltage (ΔV_{fb})[8,9].

$$Q_{ot} = -\Delta V_{fb} C_{ox} \tag{5.2}$$

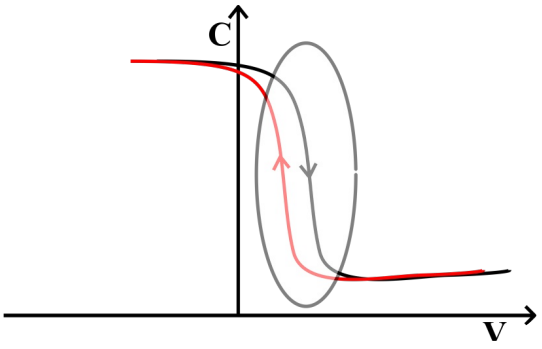


Figure 5.1: Hysteresis arising in C-V plot due to the presence of oxide defects.

With the inclusion of metal ions in the oxide, trap ecology may get modified and the change in hysteresis might follow as well.

5.2.2 Bias Temperature Stress (BTS)

However, to effectively quantify the mobile ions in a MOS dielectric, bias temperature stress[8] study stands out as the ideal candidate. The mobility of mobile ions is known to increase at an elevated temperature. If the system is stressed with a constant voltage (generally much lower than their breakdown voltage) for several minutes in that temperature, then the mobile ions infuse inside the dielectric much vigorously and the effect gets reflected in the huge shift of the C-V data.

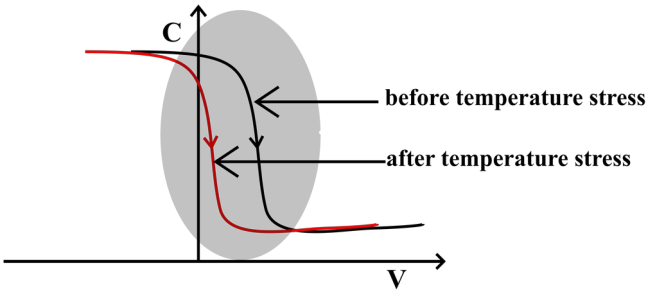


Figure 5.2: A generalized plot of BTS study in a MOS device.

The details of the BTS procedure can be found elsewhere, although we have twitched our BTS study for resistive memory slightly and are described below.

Our intention was to detect the mobile ions' diffusion in the oxide in post-breakdown condition, so we broke the devices at different levels of compliance current to change the concentration of mobile ions in the device. The modified BTS for ReRAM device is as follows: measure C-V in room temperature →break the dielectric under a very low

compliance current (since for a conducting dielectric there will not be a C-V to begin with) → acquire C-V data →heat the sample at moderately high temperature (in this case 120°C) and voltage stress it for few minutes (1 to 3 minutes) → measure C-V data at room temperature → compute the increased hysteresis → repeat this process for several compliance current levels.

5.3 Sample preparation

P-type silicon <100> samples were chosen as the bottom electrode with 5nm of Al_2O_3 as the switching dielectric with gold and silver as top electrodes. Note that <111> silicon was intentionally avoided to negate the effect of interface traps (please refer to chapter 4). The oxide layer was grown on the silicon by atomic layer deposition unit and while the thickness was verified by ellipsometric study. The details of the oxide deposition have been given in the sample preparation section of chapter 4. Au and Ag circular electrodes of area 0.33 mm^2 were deposited by employing a shadow mask by the method of thermal deposition.

5.4 Results and discussions

Initially, we embarked on this inspection with both <100> and <111> silicon substrates as BEs. The mismatch of trapping and detrapping frequency with the probe frequency leads the capacitance to slip into deep depletion domain (while sweeping from accumulation to inversion region); while the capacitances were observed to come back to the equilibrium level eventually when sweeping from the inversion end[8]. Similar behavior has been recorded for our case as well. Devices with silver top electrodes had a tendency to slip in the deep depletion region more often than its counterpart for <111> BE substrates. However, Au TE devices did let us sweep further into the inversion side, owing to its higher breakdown field, and explore the effect in detail.

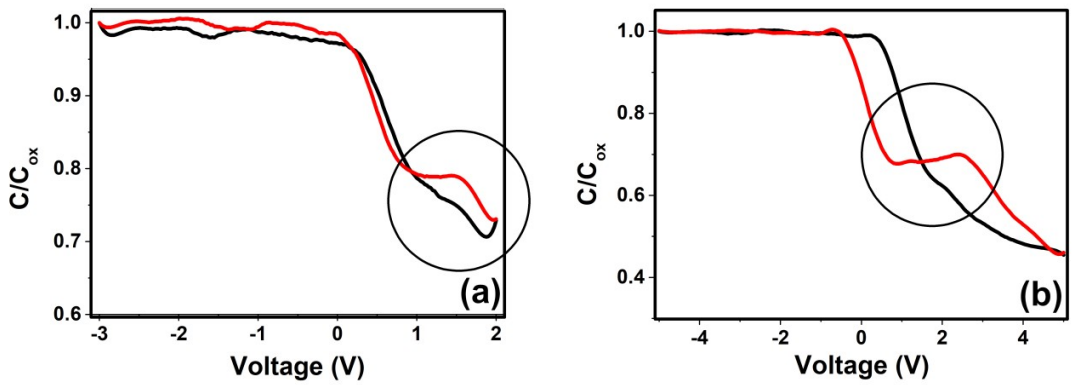


Figure 5.3: Capacitance going into deep depletion towards more positive bias (black) and somewhat recovering into equilibrium while coming back (red) for Ag TE electrode (a) and Au TE electrode (b), respectively.

5.4.1 Mobile ions: C-V hysteresis and BTS

In the following part of the experiment, we systematically broke the $Au/Al_2O_3/Si$ and $Ag/Al_2O_3/Si$ devices at different compliance limits and monitored the C-V sweep (negative → positive → negative) at each of these different levels. The intent behind this experiment was to find any correlation between the compliance current (more compliance current means stronger conducting filament, hence more mobile species)[10,11] and mobile ions. Upon careful scrutiny, little change in the hysteresis in post breakdown condition was observed for both the devices. Although for Ag TE devices the change in hysteresis could be visually confirmed as contrary to the Au counterpart, where the change is too small and it was only brought to light by mathematically calculating the area under the curve.

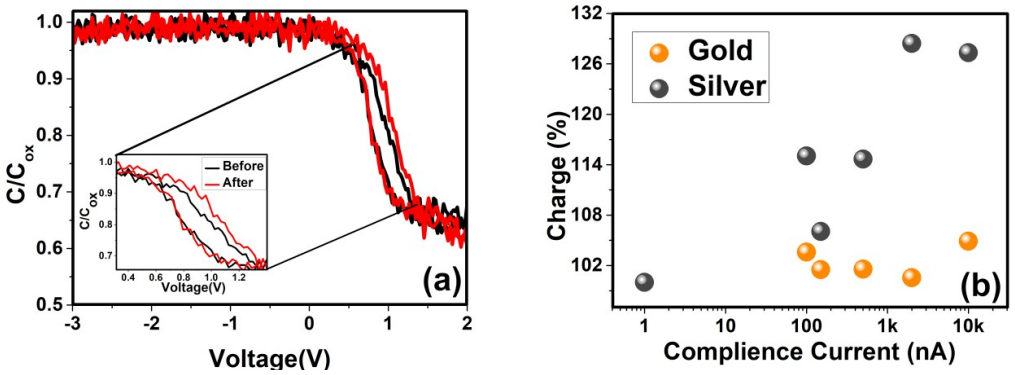


Figure 5.4: Hysteresis increase in Ag TE devices (a) and a comparative hysteresis increase as a function of CC in both devices (b).

From the above set of figures, it is evident that the percentage increase in hysteresis is much enounced in Ag TE devices than the Au ones. Although, a general trend in hysteresis increase can also be observed in Au device as well; which can be attributed to the mobile oxygen vacancies/ions, contrary to the Ag counterpart, where both oxygen related defects and mobile metal ions both are responsible for switching. As discussed in the Literature Survey section, the above set of inspection gives an insight into the oxide defects; nevertheless, they can also be influenced by the charges infused from either electrode. A much more appropriate analysis of mobile species in a MOS dielectric can be performed through the BTS study (the details of the study can be found elsewhere[8]). The high electric field across a 5nm thick oxide rendered a much prominent hysteresis in C-V measurement than the thicker oxides. The MOS resistive switches were broken under a compliance limit of 10μA and stressed at the same CC at a

temperature of 393K for 2minutes; before and after temperature stress the capacitance behavior were monitored and are illustrated in the next figure. The larger shift in V_{fb} after high temperature stress in Ag TE device clearly infers the presence of more mobile ions in those particular devices. One may argue about the shift of V_{fb} in the positive direction, since positive shift should infer more negative charges, whereas the mobile metal ions are of positive in nature. The reason became clear when the low frequency C-V was inspected

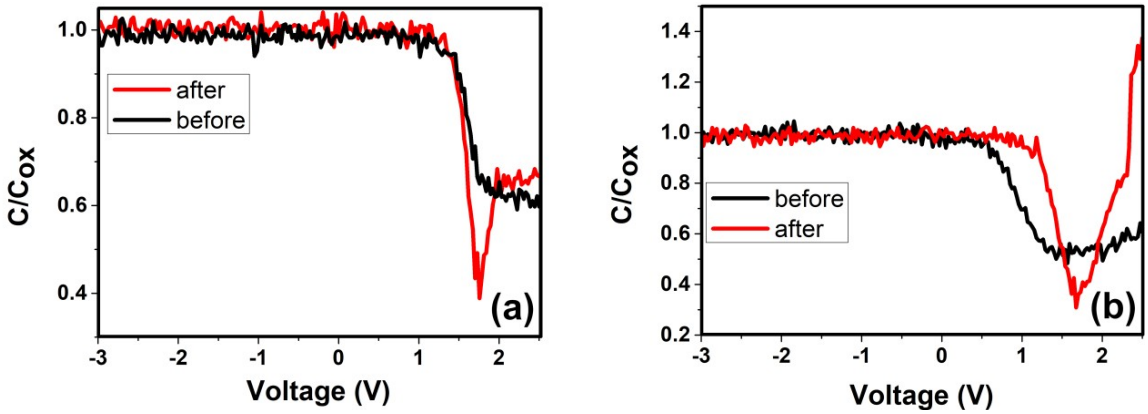


Figure 5.5: The increase in hysteresis (V_{fb} shift) in (b) Ag/Al_2O_3/Si devices is remarkably stronger than that of the (a) Au/Al_2O_3/Si devices.

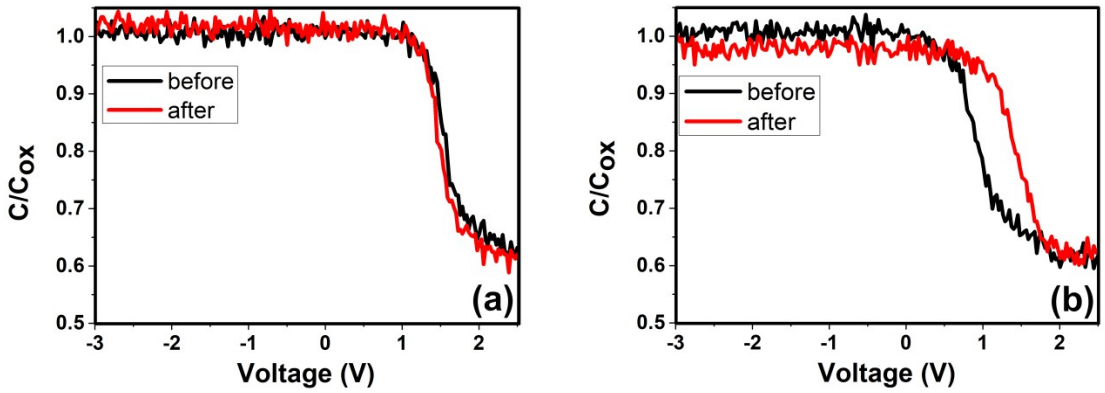


Figure 5.6: The higher jump in the inversion region in post-stress analysis shows accumulation of larger negative charges near the bottom electrode in (b) Ag/Al_2O_3/Si devices than that of the (a) Au/Al_2O_3/Si devices.

before and after the breakdown. From the plotted figures (figure 5.6) it became clear that the excessive positive species originating from the TE attracts more negative charges in the oxide-semiconductor interface and thus causing the positive flatband shift.

Figure 5.4 compares the change in charges (i.e. the area under the hysteresis) for voltage sweep hysteresis and BTS hysteresis (figure 5.5) for both the devices at the highest compliance limit; the unmatched level of charges responsible for such hysteresis from BTS study makes it the ideal candidate to study mobile ions in a MOS capacitor.

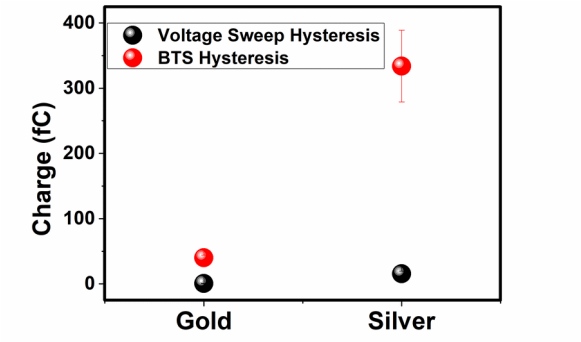


Figure 5.7: Charge responsible for the origin of hysteresis is much larger while studied through BTS method with respect to the voltage sweep study.

5.4.2 Mobile ions: V_{fb} shift

In the next part of this work, we studied the effect of the mobile ions in different thickness of oxides. Owing to the maximum response to mobile ions, only the BTS study was employed for further analysis. Another reason to stick with BTS study alone is the lack of hysteresis occurrence for higher oxide thickness (the reason has been discussed under Fixed oxide trap section). For this part, the shift in flat band voltage has been mainly taken into account as a direct reflection of the mobile ions' presence in the system.

Three different thicknesses of oxides (5nm, 10nm, and 20nm) were used in this experiment. We broke down the devices under three compliance limits (10µA, 20µA, and 30µA) and stressed them (under the same current and equivalent electric fields) at an elevated temperature for three different time limits (1minute, 2minutes and 3 minutes) and the corresponding V_{fb} was calculated. We used Mott-Schottky plot ($1/C^2$ vs. V; the x-intercept from the linear region is the required V_{fb}) for the flatband voltage estimation[12–14]. For the thinnest oxide (5nm) however, we could not perform the entire set of the experiment because of its early breaking down nature.

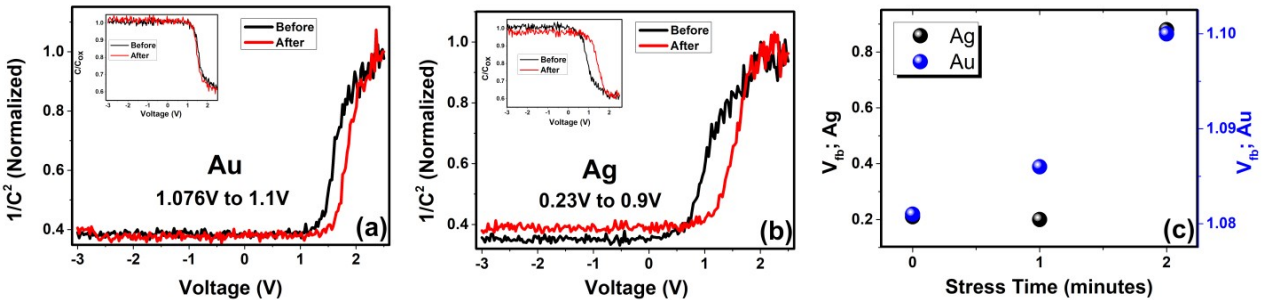


Figure 5.8: Mott-Schottky plots before (black) and after (red) temperature stress in (a) Au TE device, (b) Ag TE device, and (c) the calculated values of V_{fb} as a function of stress time. Please note the original capacitance versus voltage characteristics of the Mott-Schottky plots provided in the corresponding insets.

The shifts in V_{fb} are pretty much self-explanatory from the first two installments of the above figure. Figure 5.8(c) illustrates how small the change in V_{fb} has been accounted for Au TE devices with respect to its counterpart. It should be noted that a 5nm oxide was unable to

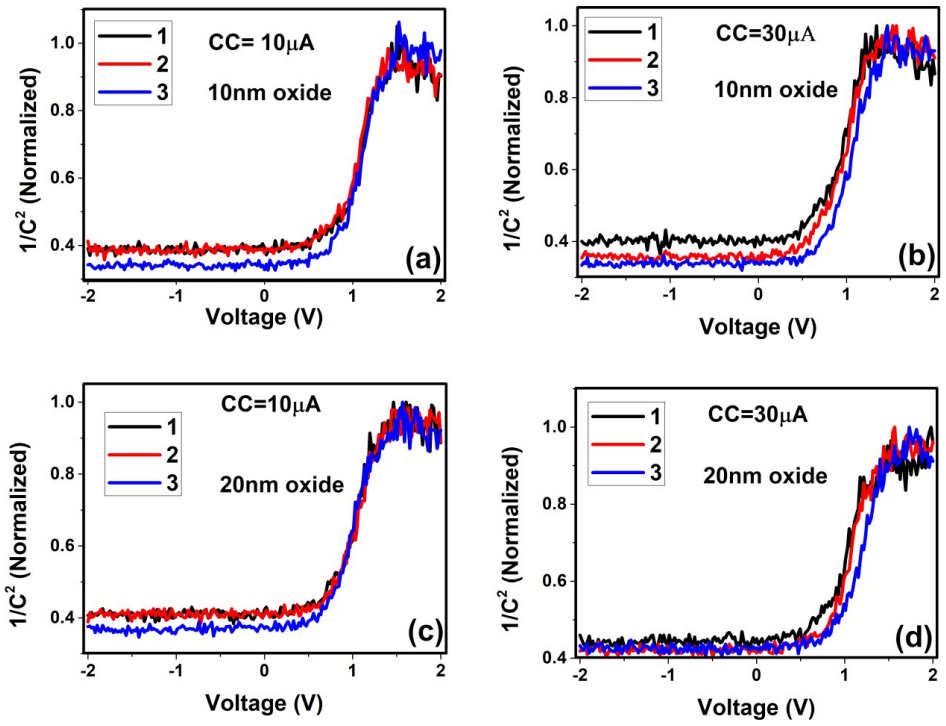


Figure 5.9: (a) and (b) are the Mott-Schottky plot for 10nm devices broke down and stressed at 10μA compliance limit and 30 μA and the same for 20nm devices (c) and (d), respectively.

withstand higher stress times (3 minutes) and higher compliance current stress due to early breakdown, which leads us to restrict our study for this case. Although, the thicker oxides' higher breakdown voltage did allow us to carry out the intended experiment and delivered the

following result. It should be noted that the devices 1, 2, and 3 and stressed for 1 minute, 2 minutes, and 3 minutes respectively. From figure 5.9(a) and (c) we can conclude there is hardly any V_{fb} shift due to temperature stress. It is only at 2 minutes and 3minutes of stress the devices started responding to the stimulus and V_{fb} changed considerably; inferring more ion migration. Another important observation from figure 5.9 (b) and (d) is the more significant mobile ions' effect with higher compliance/stress current, suggesting larger ion diffusion in the oxide. The next set of figures is calculated V_{fb}s for 10nm and 20nm oxides for different compliance current stress and for various time periods as well.

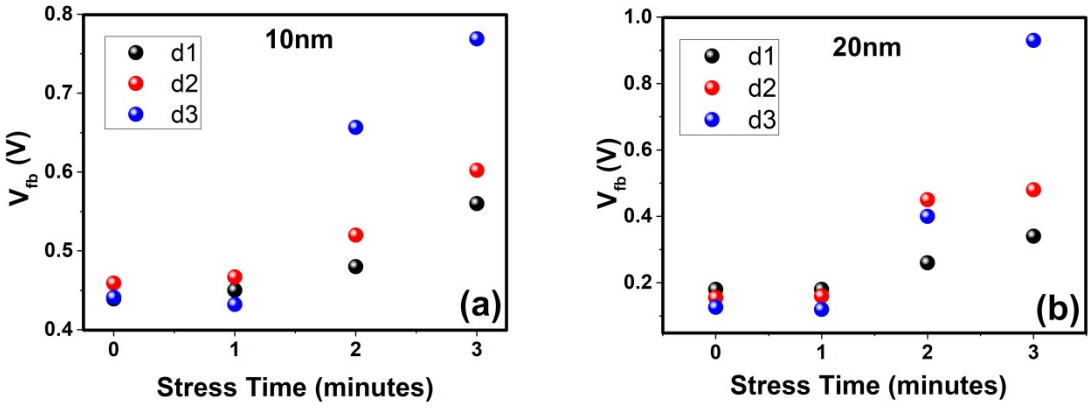


Figure 5.10: The flat band record for MOS resistive switches for different compliance currents and for different stress periods for (a) 10nm oxide layer and (b) 20nm oxide layer.

Like earlier case the devices d1, d2, and d3 represent a stress time periods of 1, 2, and 3 minute/s respectively. As pointed out earlier both larger time stress and larger compliance current ensure more mobile ion diffusion in the dielectric, causing larger V_{fb} displacement.

5.5 Conclusion

Based on the prior published works by other groups, we already had an intuition of metal ion and defect based switching dependency upon the type of top electrodes used in a resistive switch. In this chapter, we replicated similar devices and analyzed the effect on mobile ion-based switching in MOS ReRAMs by C-V methodology. We started with a thinner (5nm) oxide and learned the superiority of the BTS method over voltage sweep method in detecting mobile ions. Gradually the chapter moved towards thicker films for further inspection. Additionally, this work (with thicker oxides) also showcases the increase of positively

charged metal ions in the insulator with increasing compliance monitored stress and also stressing time itself.

It should be noted that C-V measurements are not the only pieces of evidence that hint metal ion based switching in Ag TE devices and defect aided switching in Au electrode devices. The leakage current inspection, Poole-Frenkel analysis, conductivity versus temperature analysis, and electric field direction based switching study also corroborates the same fact. However, those discussions are beyond the limit of this chapter and have been thoroughly studied and explained in the next chapter. The C-V analyses along with the I-V measurements performed on devices with different electrode-based switching together can bring forth a much more comprehensive picture of such resistive switching mechanisms.

References

1. Lee, S., Kim, H., Yun, D.-J., Rhee, S.-W. & Yong, K. Resistive switching characteristics of ZnO thin film grown on stainless steel for flexible nonvolatile memory devices. *Appl. Phys. Lett.* **95**, 262113 (2009).

2. Park, K. & Lee, J.-S. Reliable resistive switching memory based on oxygen-vacancy-controlled bilayer structures. *RSC Adv.* **6**, 21736–21741 (2016).

3. Yingtao Li *et al.* Resistive Switching Properties of Au/ZrO_2/Ag Structure for Low-Voltage Nonvolatile Memory Applications. *IEEE Electron Device Lett.* **31**, 117–119 (2010).

4. Manjunath, V. J., Rush, A., Barua, A. & Jha, R. Effect of aluminum interfacial layer in a niobium oxide based resistive RAM. *Solid State Electron. Lett.* **1**, 52–57 (2019).

5. Amini Moghadam, H., Dimitrijev, S., Han, J. & Haasmann, D. Active defects in MOS devices on 4H-SiC: A critical review. *Microelectron. Reliab.* **60**, 1–9 (2016).

6. Kumar, A., Mondal, S. & Rao, K. S. R. K. DLTS analysis of amphoteric interface defects in high-TiO_2 MOS structures prepared by sol-gel spin-coating. *AIP Adv.* **5**, 117122 (2015).

7. E H Nicollian_ J R Brews-MOS (metal oxide semiconductor) physics and technology-(Wiley 1982).

8. Schroder. D. K., Semiconductor Material and Device Characterization. in *Semiconductor Material and Device Characterization* (John Wiley & Sons, Inc., 2005).

9. Heiman, F. P. & Warfield, G. The effects of oxide traps on the MOS capacitance. *IEEE Trans. Electron Devices* **12**, 167–178 (1965).

10. Napolean, A., Sivamangai, N. M., Samuel, J. & John, V. Overview of Current Compliance Effect on Reliability of Nano Scaled Metal Oxide Resistive Random Access Memory Device. in *2018 4th International Conference on Devices, Circuits and Systems (ICDCS)* 290–296 (IEEE, 2018). doi:10.1109/ICDCSyst.2018.8605178.

11. Sun, B. *et al.* The Effect of Current Compliance on the Resistive Switching Behaviors in TiN/ZrO_2//Pt Memory Device. *Jpn. J. Appl. Phys.* **48**, 04C061 (2009).

12. Patel, K. *et al.* Flat band potential determination of $NbSe_2$ photoelectrode using Mott-Schottky plot. in *AIP Conf.* vol. 2115 030402 (2019).

13. Zhao, Y. L. *et al.* Scaling of flat band potential and dielectric constant as a function of Ta concentration in Ta-TiO_2 epitaxial films. *AIP Adv.* **1**, 022151 (2011).

14. Gelderman, K., Lee, L. & Donne, S. W. Flat-Band potential of a semiconductor: Using the Mott–Schottky Equation.

Chapter 6

Thickness scaling of ReRAM devices

6.1 Introduction

One of the main advantages of crossbar[1] based memory devices is that they can be easily 3D stacked[2] to increase the bit to area ratio. In theory, the number of layers that can be accommodated in a given vertical space can easily be increased by thinning down the individual devices. As long as the lateral resistance of the electrode line does not kill the connection, the thickness of electrodes shall remain at a minimum constant; that leaves us our degree of freedom with the dielectric thickness. In practice, this thickness tuning can be limited by several factors, such as band bending/lowering, direct tunneling, etc. and can significantly compromise the device performance. To learn the effects of these phenomena in a ReRAM device and their consequences in the technology we are compelled to study the switching behavior as a function of thickness.

This chapter is primarily dedicated to identifying the vertical scaling limit of ReRAM devices. To deposit the dielectric, atomic layer deposition (ALD) was naturally the choice of deposition, because of its capability of depositing ultrathin layers (thickness in the order of atomic dimensions), with extreme controllability and uncanny precision. ALD Al_2O_3 has been studied extensively in the past both in terms of the dielectric as well as a ReRAM[3] material. Moreover, Al_2O_3 is a slandered material, which can be easily deposited by an ALD system with alternating pulses of TMA and water and can deliver a commendable resistive switching performance.

We have deposited a series of different dielectric thicknesses namely 20nm, 15nm, 10nm, 7nm, 5nm, 4nm, 3nm, 2nm, and 1nm on FTO-coated glass substrates as the bottom electrode, followed by gold/silver electrode as the top electrodes. An in-depth study was conducted with each of these devices to unravel the variation in the switching nature with dielectric thickness. Following is the list of investigations performed on such devices.

1 Resistive switching characteristics,
   a Switching characteristics; gold versus silver top electrode,
   b Cycling properties, and
   c Reliability study by Weibull distribution plots.
2 Conduction mechanism at different temperatures,
3 Space charge limited current behavior,

4 Fowler-Nordheim tunneling behavior,

5 Poole-Frenkel charge injection,

6 Direct tunneling and

7 Trap assisted tunneling.

The devices with gold (Au) top electrodes (TE) failed to show proper and repetitive switching behavior in most of the cases. On the contrary, silver (Ag) top electrode (TE) devices deliver excellent switching behavior with excellent repeatability. Therefore, in the upcoming sections, we will primarily consider the silver electrode devices as the standard device configuration to investigate the vertical scaling possibilities/limitations with enough statistical backup and gold electrode devices to elucidate the switching phenomena in our devices.

6.2 Literature Survey

6.2.1 Space Charge Limited Current (SCLC)

An insulator or a semiconductor under voltage stress can exhibit different regimes of current conduction mechanism[4] depending on the exponent of the varying voltage mediating the current output. Generally, at lower electric fields the current is mainly driven by the mobile charges intrinsically present in the dielectric and would follow ohm's law ($I \propto V$). At higher electric fields this scenario changes and the current shows a quadratic relationship with the voltage ($I \propto V^2$). This happens because at higher fields the traps/defects are all filled with electrons and this site individually forms a space of opposing electric fields with a radius equal to the Debye length of that material. This effect was first understood by Mott-Gurney in the year of 1940. According to his explanation, when a thin insulator/semiconductor is sandwiched between two Ohmic junctions, the current density at a higher field would be influenced by the charges entering through the electrodes and will follow the following equation.

$$J = \frac{9}{8}\varepsilon\mu\frac{V^2}{L^3} \tag{6.1}$$

where J is the current density, ε is the permittivity, μ is the permeability while a voltage V is applied across an insulator/semiconductor of thickness L.

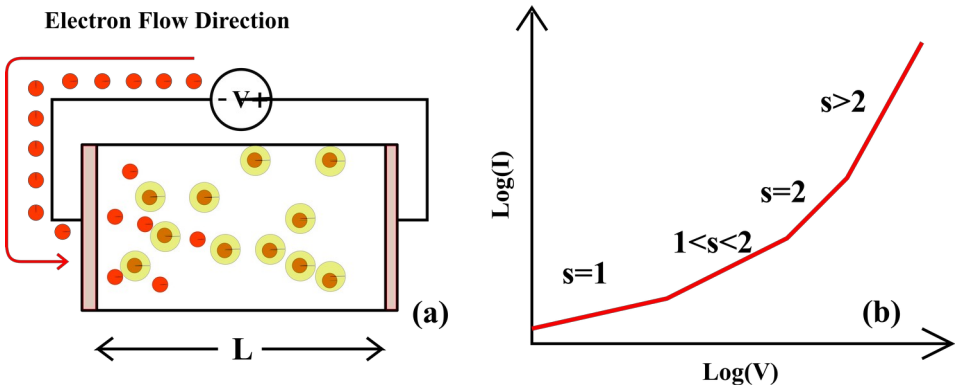


Figure 6.1 (a) A schematic representation of the SCLC current conduction mechanism and (b) the corresponding current vs. voltage characteristics.

The slope "1" in fig 6.1(b) represents the Ohmic region, a slope "2" corresponds to the pure SCLC regime while the intermediate portion between them is the region where SCLC has already commenced with some Ohmic component still remaining. The last part is the leaking /tunneling part, which in practical cases often may suggest an electrical breakdown. Thus, by plotting *log(I) vs. log(V)* plot one can easily find out the slope of different sections of the current plot and can determine the SCLC regime.

6.2.2 Fowler-Nordheim Tunnelling

Electrons when separated by an insulator can reappear on the other side when a certain magnitude of an electric field is applied (without breaking down the oxide). The quantum mechanical phenomenon which allows electrons to travel through an insulator under voltage bias is known as Fowler-Nordheim (FN) tunnelling[5]. Ralph H. Fowler (Britain) and Lothar Wolfgang Nordheim (Germany) were the first to explain this kind of electric field-assisted quantum tunneling in the 1920s. According to them, a negative bias at one side of a capacitor would cause mirror charges to appear on the insulator at the insulator-metal junction and hence causing a band lowering of the insulator. Along with the band lowering the band suffers a voltage induced thinning down, termed as band triangulation, which aids the electrons to direct tunnel through the thin barrier to reach the other terminal of the device. The mathematical formulation of such tunneling is given below.

$$J_{FN} = \frac{A E^2}{\Phi_b} \exp\left(\frac{-B}{E} \Phi_b^{3/2}\right) \qquad (6.2)$$

Where A and B are constants, E is the electric field and Φ_b is the lowered barrier height.

With a given I vs. V data a simple plotting of $Ln(J/E^2)$ vs. $1/E$ will provide the slope of the linear region, which is directly proportional to the lowered barrier height. The next figure should help elucidate this concept further.

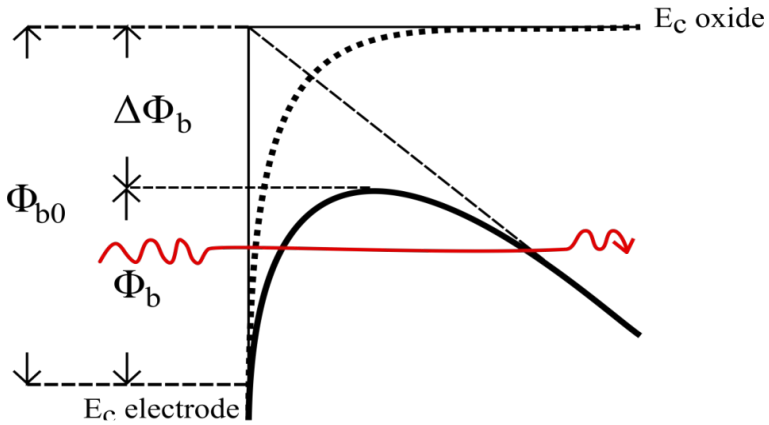


Figure 6.2. The dotted line shows the barrier height at zero voltage bias ($\Delta\Phi_{b0}$), which decreases by $\Delta\Phi_b$ and renders a barrier height of Φ_b (denoted by continuous line; which depicts the band triangulation as well).

FN tunneling is a very common observation for thin and ultrathin dielectric systems since the electric field across them is sufficiently strong and barrier requires lesser thinning down to initiate such process. Although FN tunneling may give rise to unnecessary leakage current, it has been the backbone of non-volatile FLASH memory storage for several decades, and few more to come (details of the operation have been described in the 1st chapter) and has shaped the current world as we know it.

6.2.3 Poole-Frenkel Injection

Poole-Frenkel Injection (PF injection) method[5,6] describes how an electron can travel through an insulator or a semiconductor under the influence of an electric field through the localized states present in the medium. This phenomenon was first recognized by Horace Hewitt Poole (1885-1962) and later published by Yakov Frenkel in 1938. It is well known that if an electron can acquire enough thermal energy it will move to the conduction band from its earlier localized state and can contribute to conduction. When an additional electric field applied to the system the energy required to hop from one localized state to another is partly taken from the field and partly from the thermal fluctuations. The following expression should be much more legible on how the measurement and calculation can be performed in a particular system to calculate the energy barrier height.

$$J_{PF} = CEexp\left(\frac{-\Phi_t + q\sqrt{qE/4\pi\varepsilon_0\varepsilon_r}}{kT}\right) \qquad (6.3)$$

Where C is a constant, E is the electric field, Φ_t is the trap level activation energy, ε_0, ε_r are absolute permittivity and relative permittivity respectively, k is the Boltzmann's constant and T being the temperature in Kelvin scale.

By plotting $Ln(J_{PF}/E)$ *vs.* $E^{1/2}$ one can recognize the linear region and the y-intercept of that best fit is proportional to the Φ_t.

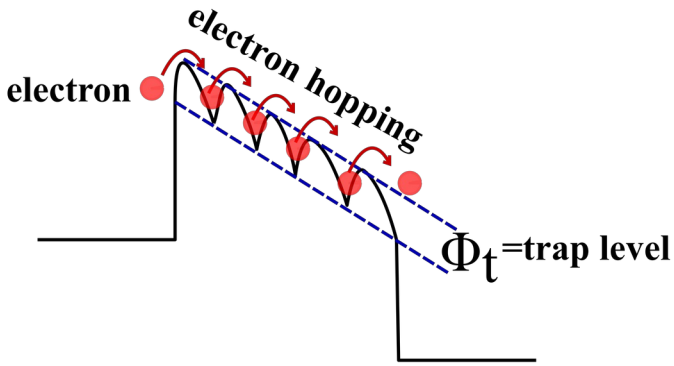


Figure 6.3: A schematic representation of electron hopping through the available traps in an insulator under the influence of an electric field.

From the above diagram, it should be evident that under a proper voltage bias the electrons will gather enough energy to hop across these localized states and the induced band bending shall facilitate the hopping mechanism as well.

6.2.4 Direct Tunnelling

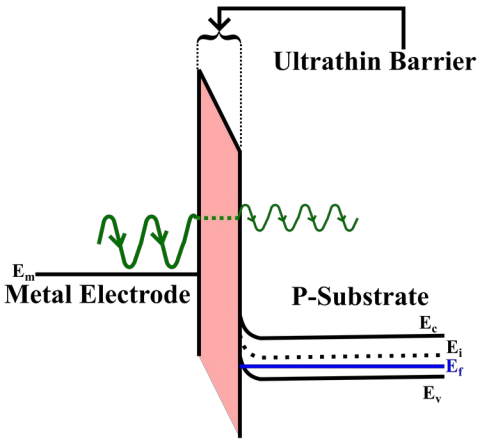


Figure 6.4: The wave probability of electrons (green) crosses through the ultrathin insulating barrier from the metal electrode to the silicon substrate in a MOS capacitor.

Unlike FN tunneling, electron sometimes can overcome a barrier without the applied voltage modifying the energy band diagram. This quantum mechanical tunnelling[7] is termed

as the direct tunneling phenomenon. When two electrodes are separated by an extremely thin insulator, then, the electrons can reappear on the other side of the barrier without any voltage stress, like the way pictorially represented in the following diagram. The next expression is a mathematical representation of the direct tunneling phenomenon.

$$J_{DT} = \frac{4\pi q\, m_x^e kT}{h^3} \int_0^\infty P(E)_X S(E)_x dE_x \qquad (6.4)$$

Where, m^e is the effective mass, h is the Plank's constant, $P(E_x)$ is the probability distribution associated with the electron kinetic energy (E_x) and $S(E_x)$ is the "supply" function of E_x (can be found elsewhere).

6.2.5 Trap Assisted Tunnelling

An insulator infested with traps[8,9] can also form several promising direct tunneling sites and the electrons tend to tunnel from one trap to another, giving rise to high leakage current.

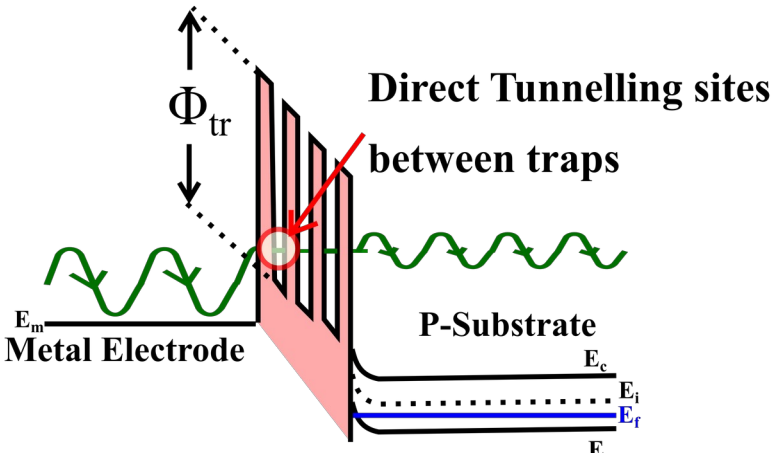


Figure 6.5: Diagram showing the trap assisted tunneling phenomenon across a trap populated insulator in a MOS capacitor.

When the traps lie sufficiently close to each other in an insulator the separations between them are barely able to hold them from directly tunneling from one trap to the next. The following equation states the current to voltage relation in a trap assisted tunneling (TAT) scenario.

$$J_{TAT} \quad \exp\left(\frac{-4\sqrt{2\,q\,m_{ox}}}{3\hbar}\Phi_{tr}^{\frac{3}{2}}/E\right) \qquad (6.5)$$

Where, m_{ox} is the effective mass of electrons in the oxide and Φ_{tr} is the trap energy level below the conduction band of the dielectric.

By plotting $Ln(J_{TAT})$ vs. $1/E$ anyone can recognize the linear region (which would specify the TAT region) and thus, from the slope Φ_{tr} can be calculated.

As this chapter unfolds we will be discussing our results and their association with the previously discussed conduction phenomena. Some of these theories will provide a better insight into the breakdown process and conduction mechanisms through the dielectric, while the others will try to recognize any pattern (if present) with the changing thickness. Ultimately, the chapter will conclude with the effects of these phenomena at lower thickness and if by any means they contribute to limit the vertical scaling of a ReRAM device.

6.3 Sample Preparation

FTO coated glass samples were cleaned first with Acetone and then IPA ultrasonic bath cleaning followed by a DI water bath and nitrogen gas (99.9999% pure) drying. Al_2O_3 layer was grown on the conducting FTO coated side at a temperature of 200°C by thermal ALD system with an alternating TMA and water pulsing. The number of pulses required to deposit the desired thickness of AL_2O_3 was estimated from the thickness/cycle data (1.15 Å per cycle) and was later confirmed by an optical ellipsometry measurement. Later, circular patterned top electrodes were deposited on the dielectric layer by thermally evaporating electrode metals through a shadow mask at ~10^{-6} torr of pressure and for a thickness of 150nm. A schematic methodology of sample fabrication has been represented in the next diagram. A probe station connected to an Agilent B1500A was employed to acquire the related data from the devices.

6.4 Results and Discussion

For a 20nm thick Al_2O_3, we applied a voltage with a current compliance of 10^{-6} amps initially to generate enough defects that would form the filament. This process is known as the formation process[10,11]. One crucial observation was that the devices with the gold electrode required much higher formation voltage (~15V) than the ones with the silver top electrode (~4V). This can be attributed to the migration of silver ions, which facilitates the formation of a conducting path with much ease. On the contrary, in $Au/Al_2O_3/FTO$ devices, the larger voltage was required to create sufficient oxygen vacancies and ions to aid the formation of the conduction path. Post formation scenario required a compliance current (CC) limit for the $Au/Al_2O_3/FTO$ set process; we believe that this prevented the formation of uncontrolled defects (oxygen vacancies/ions) which would eventually lead to a permanent conducting filament formation. However, the $Ag/Al_2O_3/FTO$ systems showed an exceptionally well-controlled set and reset process without any CC check. Let us take a look

at the current-voltage (I-V) characteristics of these two devices in the next diagram (figure 6.6) for a 10nm thick oxide before qualitatively analyzing them.

6.4.1 I-V characteristics

One can easily infer from the following set of figures that the $Ag/Al_2O_3/FTO$ systems show better switching consistency than its Au counterpart. By comparing figure 6.6 (a) and (c) the clear difference in the consistency of the set and reset voltage in both the devices (Ag suppressing Au by a huge margin) is easily conceivable; additionally from figures 6.6 (b) and (d) the firmness of the on-off ratio in $Ag/Al_2O_3/FTO$ devices over $Au/Al_2O_3/FTO$ can be confirmed. In chapter 2 we have learned how Weibull distribution works and how they can help us quantify the reliability of a ReRAM device. We will assume a similar approach for the aforementioned devices as well (later in this chapter) as a measure of device reliability.

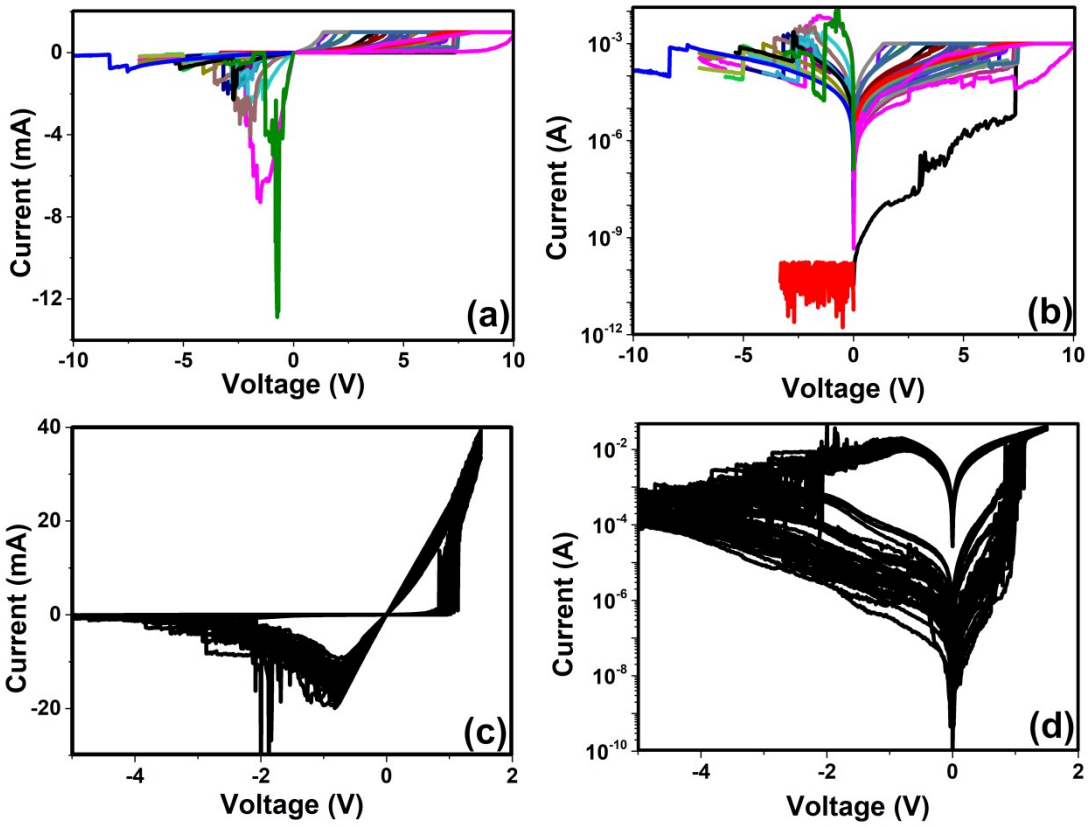


Figure 6.6 (a) and (c) are the I-V switching behavior of $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices respectively and (b) and (d) are their corresponding butterfly representation (current in log scale) accordingly.

We have already postulated the roles of defects and metal ions in the set process of a device in the 5th chapter; the impact of that hypothesis can also be verified from the voltage polarity vs. set occurrence study.

Since the Au electrode does not contribute to metal ion migration (due to high electron affinity of Au), the devices suffer a breakdown because of the oxygen vacancies (V_o) and oxygen ions (V_i) created inside the dielectric matrix by the large electric field. So far for an Au TE device, neither of the electrodes aids the formation of metal ions. Thus, theoretically, the defect formation should be independent of the field polarity and therefore can be switched from any direction. However, Ag ions belonging from a lower stratum of electron affinity chart, are able to lose electrons to ***only*** a positive bias and become positively charged Ag^+ ions. These ions migrate according to the applied electric field and reclaim the electrons from the bottom electrode to re-establish itself as a neutral metallic atom all over again; eventually forming a metallic filament across the oxide. Figure 6.7 perfectly depicts the formation of defect based (figure6.7(b) & (c)) and metal ion (figure6.7(d)) based switching and their dependency on the bias polarity as well. Figure 6.8 (a) illustrates the filament formation by oxide defect

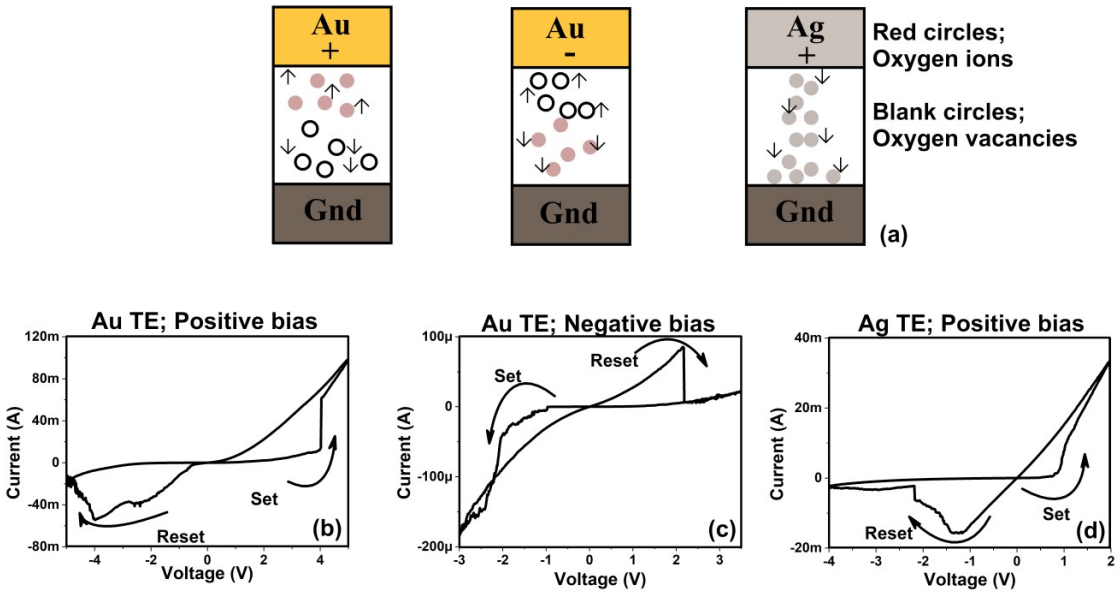


Figure 6.7: (a) Schematics of Resistive switching in Au and Ag based top electrodes, breakdown as a function of voltage bias for positive bias on gold (b), negative bias on gold (c), and positive bias on silver (d).

creation and metal ion migration in $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices respectively. Please note that the first two installments of figure 6.7 (a) depict the oxide defect based filament formation with different polarity of the applied electric field while the last one shows metal ion diffusion mediated dielectric breakdown. The following set of illustrations is of utmost importance as they corroborate the fact that indeed Au electrodes do not contribute to metal ion migration and the switching is purely due to the oxygen vacancies, unlike the Ag electrodes where the filament formation is a function of positively charged Ag ion diffusion.

We performed voltage sweeps ranging from -6 to 2.5 V for the different Al_2O_3 thickness devices to monitor the resistive switching behavior. It should be noted that the voltage cycling did not start directly from -6V (in which case it may and have destroyed several devices with sudden high voltage stress), rather followed 0V→2V→-6V→0V cycling pattern. The following two figures are different switching characteristics of our devices and their corresponding set and reset voltage statistics.

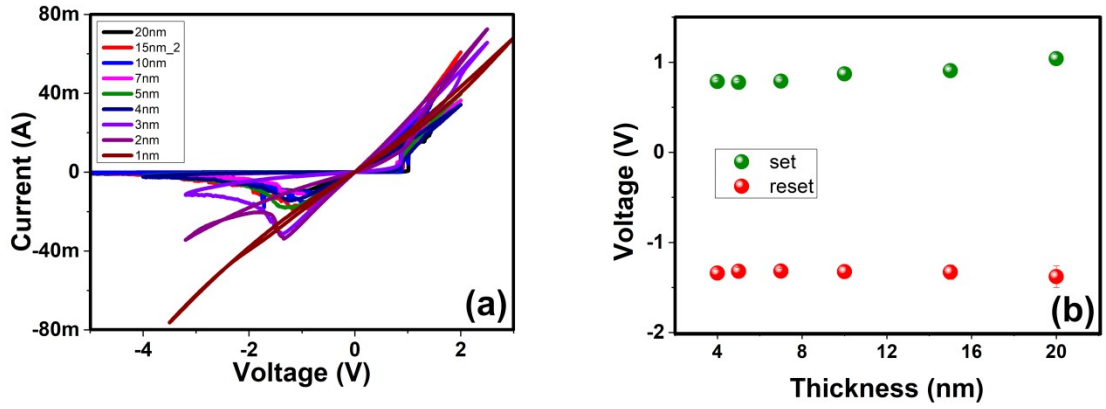


Figure 6.8: (a) Switching characteristics of devices with dielectric thickness ranging from 20nm to 1nm and the set and reset voltage distribution as a function of thickness.

Note that the set and reset voltages are not provided for 3nm to 1nm devices. As clear from the I-V characteristics, we could not see any clear switching characteristics for 1, 2, and 3 nm thick devices, unlike higher thicknesses. Hence determining the exact set or reset was not possible for those devices. One crucial factor that determines the quality of resistive switching device is its on-off ratio i.e. the on current (or conductivity) to off current (or conductivity) at a particular read voltage. Better the ratio better would be the noise margin and hence lesser the probability of data corruption.

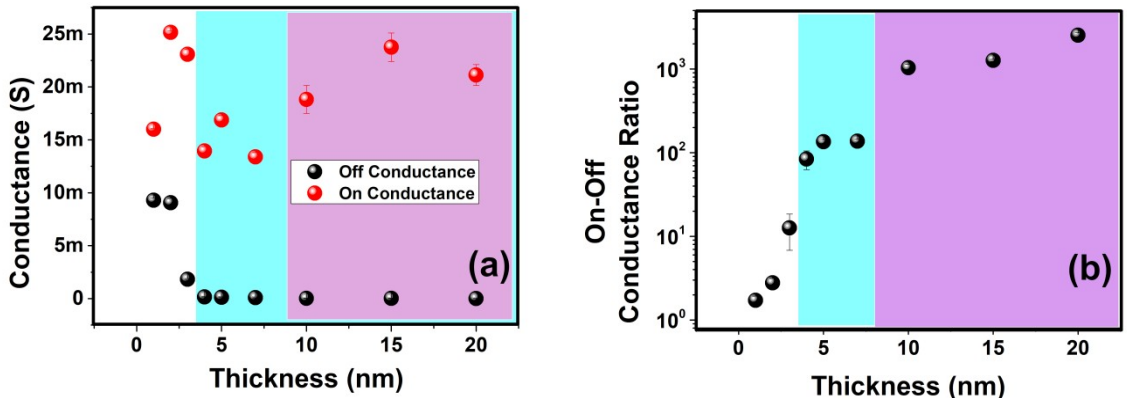


Figure 6.9: (a) on and off conductance for different thickness devices and (b) the corresponding on-off ratio. Different color codes define devices with the ratio in the order of 10^3 (purple), 10^2(sky blue), and 10(white).

As mentioned earlier, in terms of on-off ratio[12,13] the thicker devices will be more celebrated for their superior performance. However, if we take the area of the electrode (0.33×10^{-6} m^2) into account, we can safely assume that, for practical devices (area in the order of 10^{-16} m^2), the on-off will become far better; as the off resistance depends upon the area of the electrode and the on-resistance on the structure of the CF. therefore, from figure 6.9 we can conclude that the devices with thickness 4nm or higher should be able to render sufficient on-off ration which is at par with the industrial standards.

6.4.2 Space Charge Limited Current (SCLC)

Before the onset of the set process, the SCLC regime will appear in the I-V behavior of our resistive switches. As discussed earlier, $Au/Al_2O_3/FTO$ devices are unlikely to switch on and off when the set process is not limited by a compliance current. In a rare occurrence, we were able to do so after the first breakdown in a few devices. The SCLC region was recognized in the first and the very following I-V cycle of an $Au/Al_2O_3/FTO$ device and its corresponding $Ag/Al_2O_3/FTO$ device for a comparative study of the trap charge generation after forming process.

The first crucial information from the following set of pictures would be the increase of off-current of both devices after formation, which is because of the defects (or metal ion diffusion) present in the dielectric. Secondly, even though the off current has increased the

slope in *log(V) vs. log(I)*plot does not seem to vary much. For further elucidation, we calculated the SCLC transition voltages.

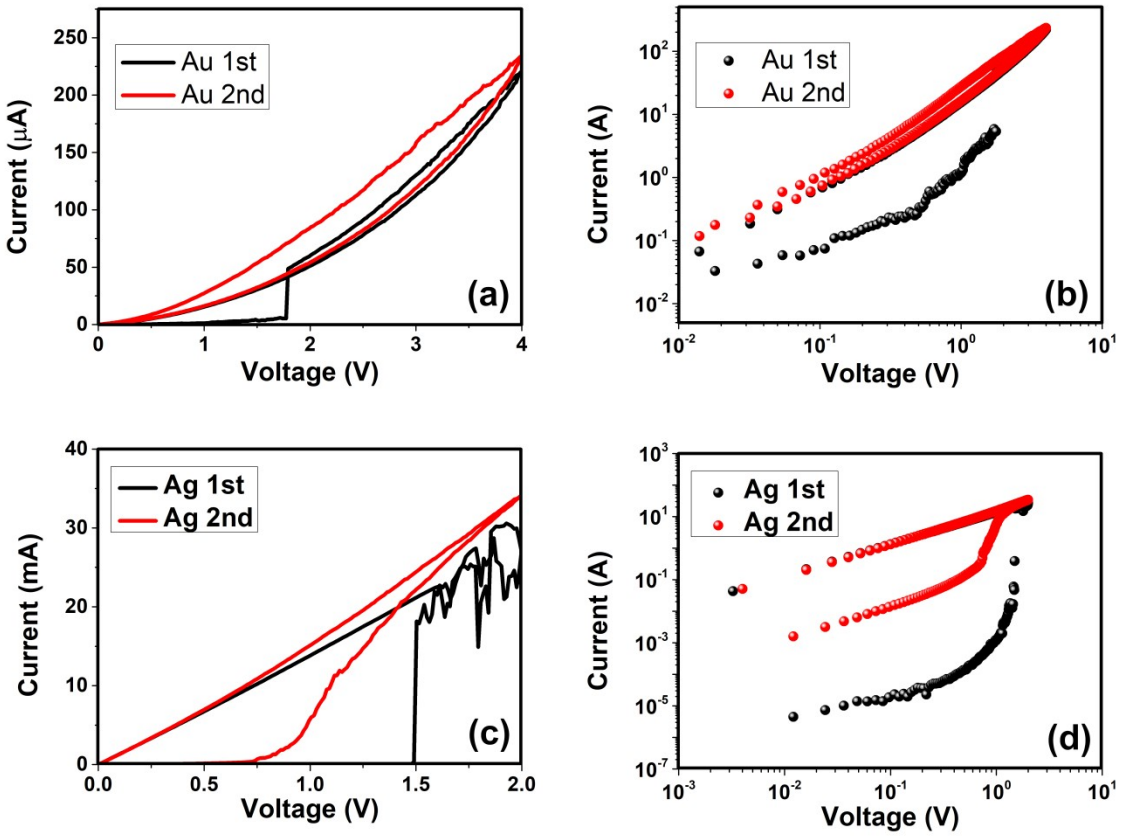


Figure 6.10: (a) and (c) are the first and second cycle I-V data in linear scale and (b) and (c) shows post forming process increase in off (leakage) current because of dielectric defect formation.

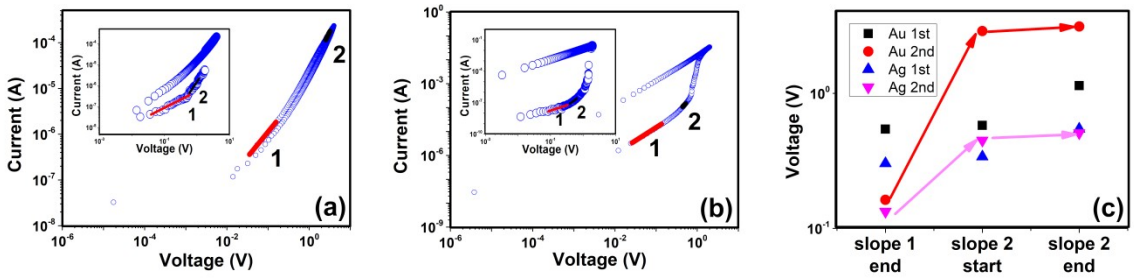


Figure 6.11: (a) and (b) present the slope "1" and "2" regions in Au and Ag TE devices both pre (inset) and post-formation conditions and (c) the transition voltage statistics.

Figure 6.10(b) and (d) clearly illustrate the increase in leakage current after the formation process which is a consequence of excess defect formation in the dielectric matrix. From the

immediate inspection of the transition voltages (see figure 6.11(c)), it could be concluded that the large jump from slope 1 to slope 2 in $Au/Al_2O_3/FTO$ with respect to $Ag/Al_2O_3/FTO$ was an effect of the larger trap generation (post formation). As discussed previously, a larger V_o and V_i generation are responsible for this behavior, on the contrary in Ag TE devices, the electric field does not compel an unchecked generation of V_o and V_i, rather majorly rely upon the Ag ion migration from the TE and their consequent reduction at the BE.

Next, we investigated the SCLC behavior as a function of thickness in the $Ag/Al_2O_3/FTO$ system. We employed the devices with dielectric thickness ranging from 20nm down to 4nm (beyond which, direct tunneling became dominant; will be discussed later in this chapter) and plotted the commencement of SCLC regime (when slope=1 ends, practically SCLC effect kicks in; refer figure 6.1(b)) and pure SCLC regime as a function of dielectric thickness.

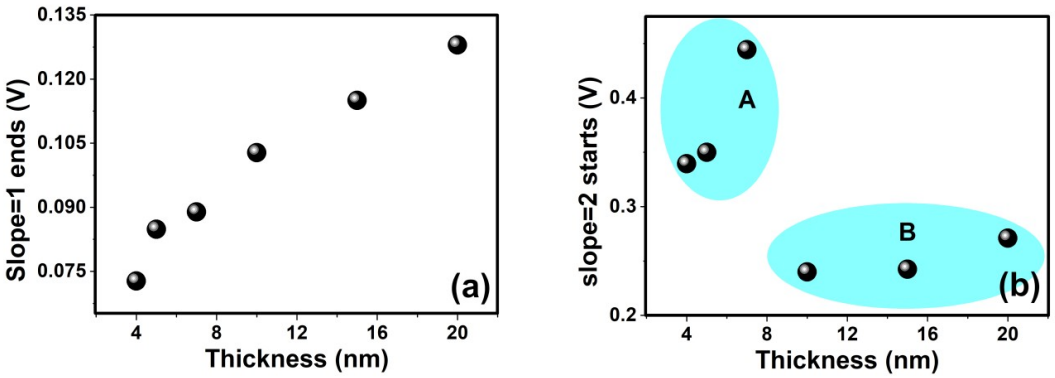


Figure 6.12 (a) the voltage at which the space charge limitation effect starts and (b) pristine SCLC regime as a function of thickness.

With decreasing thickness, we are left with fewer traps and intrinsic mobile ions (proportional to the volume of the dielectric; assuming the mobile ion density is constant) and hence, the early commencement of the SCLC effect. However, figure 6.12 (b) shows the similar trend but in two separate regions, namely "A" (4 to 7nm) and "B" (10 to 20nm). This behavior can be a direct effect of the other tunneling current components' superimposition with the leakage current to give an impression of early occurrence SCLC in the lower half of the thickness chart.

6.4.3 Fowler-Nordheim Tunnelling

The bottom electrode (BE) to the dielectric barrier height can be computed from the slope of the linear region in a $Ln(J/E^2)$ vs. $1/E$ *plot*. The following diagram depicts this plot and showcases the FN region for Au and Ag TE systems respectively. Please note that in the

plots (figure 6.13) both FN as well as DT regimes can be recognized. Although, we will restrict our discussion to FN tunneling in this section and resume with the DT regime in a while.

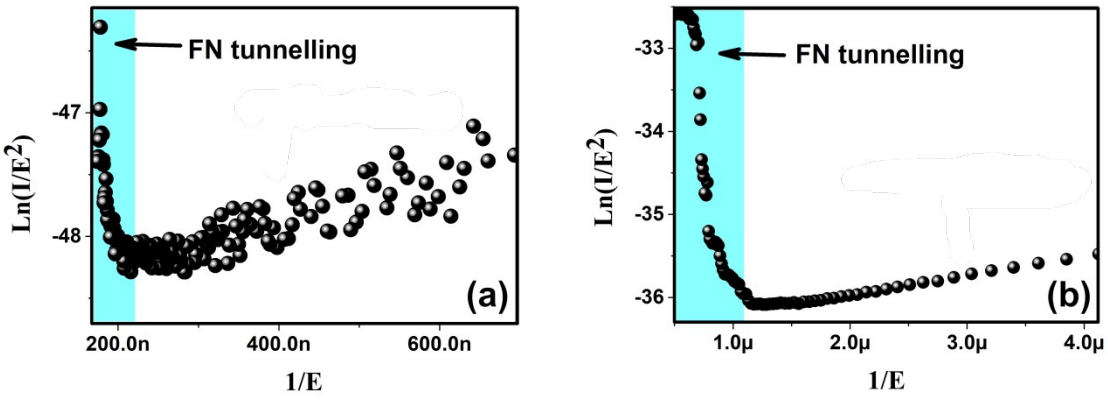


Figure 6.13: Examples of FN tunneling and direct tunneling regions in a TE Au (a) and Ag (b) sample.

With decreasing thickness of the Al_2O_3 layer and the applied voltage remaining constant, a larger electric field exists across the barrier. This caused more severe barrier lowering in sub 7nm devices and ultimately FN tunneling disappears in 1nm devices (see figure 6.15 (d)). The TE hardly participated in this process, as the band lowering in the positive bias was applied only across the Al_2O_3/FTO junction; as a result, the barrier height remained independent of the type of TE (see figure 6.14). This barrier height lowering causes ease of charge flow into the system, which leads to higher off current (hence poor on-off ratio)

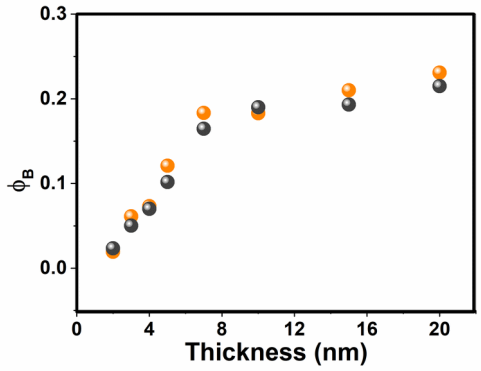


Figure 6.14: The calculated barrier height values for devices with different thickness for gold and silver top electrodes. The golden spheres and silver spheres represent the $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices, respectively.

and early breakdown. Figure 6.15 draws a comparative study of the FN tunneling (*Ln* (J/E^2) vs. *1/E* plot) as a function of thickness in different TE devices. The most pivotal observation from these plots would be the evanescence nature of the FN regime with decreasing thickness; which justifies the extensive barrier lowering in lower thickness systems in figure 6.14.

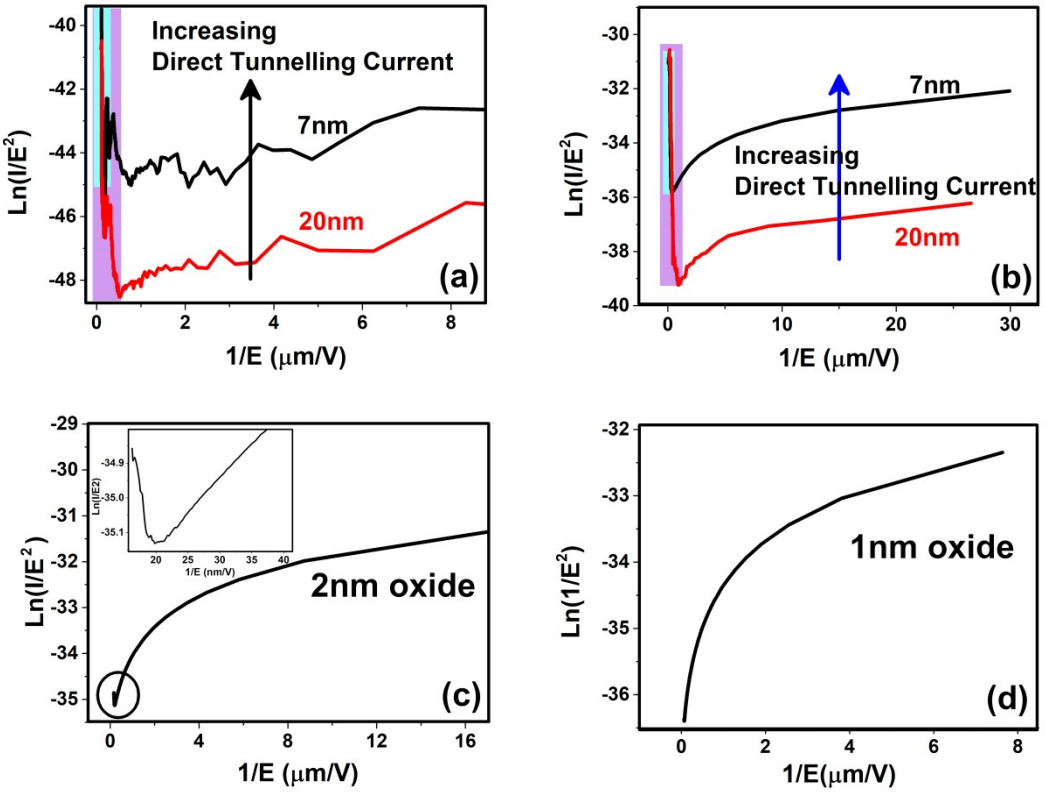


Figure 6.15: *Ln(J/E²)* vs. *1/E* plot for two different thickness in (a) Au and (b) Ag TE devices, (c) a 2nm oxide with Ag TE showing a small percentage of FN tunneling with a majority being Direct tunneling current and (d) a 1nm thick oxide has the leakage current entirely comprising of Direct tunneling.

6.4.4 Poole-Frenkel Injection

From PF carrier injection study we can quantify the trap height and their contribution in the leakage current. With a higher electric field (i.e. lower dielectric thickness) the band bending will be pronounced and so will the trap-barrier bending (Φ_t), leaving the resulting barrier height ($\Delta\Phi_t$) smaller and smaller. The following figure depicts such a scenario; the mathematical explanation follows the figure.

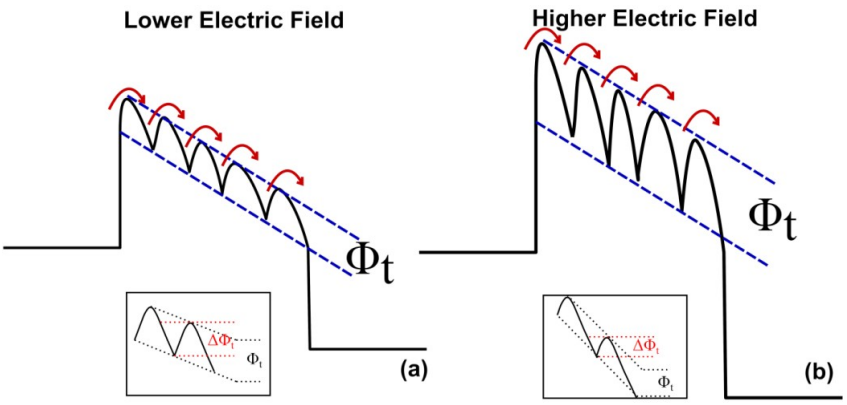


Figure 6.16: A schematic representation of the trap barrier lowering at (a) low and (b) high electric fields; please see the corresponding insets for more elucidation of trap barrier height modification with the electric field.

Continuing from equation 6.3

$$\ln\left(\frac{J_{PF}}{E}\right)=\ln(C)-\Phi_t+m_1E^{1/2} \qquad (6.6)$$

Where m_1 is a constant, which can be computed from equation 6.3. It is evident that - Φ_t is proportional to the y-intercept of the linear plot of $Ln(J_{PF}/E)$ *vs.* $E^{1/2}$.

We studied the PF injection phenomenon in both Au and Ag TE devices and a lower y-intercept for Ag TE devices suggests that the metal ions in the switching layer result in smaller trap barriers than its counterpart; that should explain the higher leakage current in $Ag/AL_2O_3/FTO$ devices.

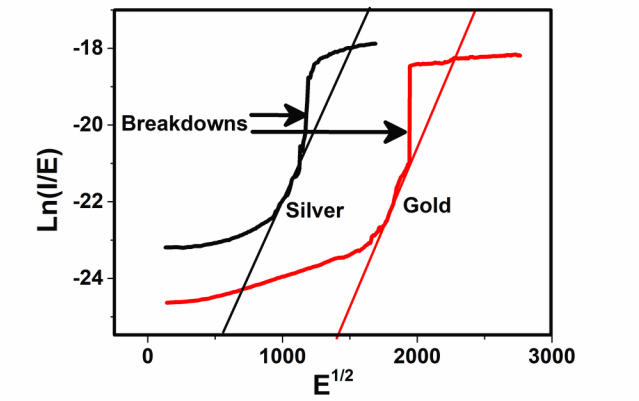


Figure 6.17: PF and breakdown region in Au and Ag TE devices.

Since the electric field in the film increases with decreasing the film thickness, the trap barrier lowering gradually increases as the films become thinner. Therefore, the charge flow increases as the films get thinner. The next diagram shows the y-intercept from PF analysis as a function of thickness. The 1nm dielectric failed to show any linear relation, suggesting the absence of PF injection in that device, where the majority of the leakage current is dominated by direct tunneling.

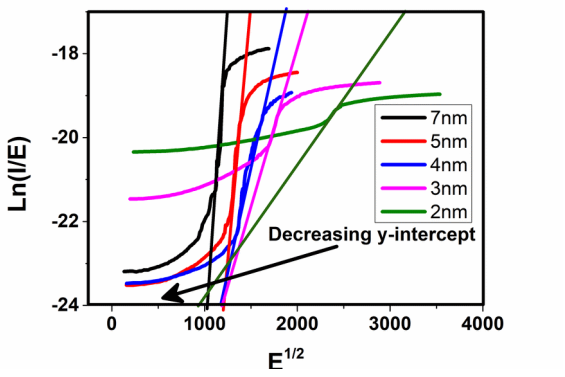


Figure 6.18: The y-intercept decreases with decreasing thickness, implying the trap barrier height lowering.

Like FN tunneling, PF injection also gives rise to higher leakage current (off current), thus compromising the on-off ratio of a resistive switch. Another noticeable observation from figure 6.18 would be the shrinking down of the PF regime (the linear region). This suggests the fact that with lower thickness along with the trap barrier lowering, other forms of tunneling overwhelms, causing higher HRS current.

6.4.5 Trap Assisted Tunnelling

Much like Poole-Frenkel analysis, this section presents a qualitative deduction of TAT's impact on oxide conductance. In PF injection we focused upon the traps located near the conduction band of the dielectric, but, the effect of deep level traps (which can lead to direct tunneling between them) can be studied through TAT analysis. Previous works have demonstrated the extraction of the Φ_{tr} from experimental data on various occasions. Houng *et. al.*[8] and Fleisher's[9] work calculated Φ_{tr} while assuming only a single trap site in the oxide in the middle; although Gushterov and Simeonov's model[14] was based on multilevel traps, some of their assumptions, like the distance of trap from the junctions and number of trap sites are unpredictable for our work (though the approach was suitable for their simulation-based studies), hence we could not deduce the exact value of the Φ_{tr}. It has been discussed in the "literature survey" section that the slope of the linear region of a $Ln(J_{TAT})$ *vs.* $1/E$ plot is directly proportional to the $3/2^{th}$ power of Φ_{tr} and in this work, we will use this slope as an

alternative measure of the trap level. All the devices under scrutiny showed a clear linear region suggesting the presence of trap site based tunneling; interestingly even 1nm thick oxide showed TAT phenomenon. Figure 6.19 (b) is a comparison of the TAT linear region in a 1nm and 7nm oxides respectively. In an ideal condition for a given oxide, Φ_{tr} should not change with the oxide thickness; however, Gushterov and Simeonov's work (and their reference) has stated that a slight variation should arise due to the buildup of the mirror charges on the other side of the oxide. Our investigation also revealed a similar trend as a function of thickness, whilst the values remain almost comparable.

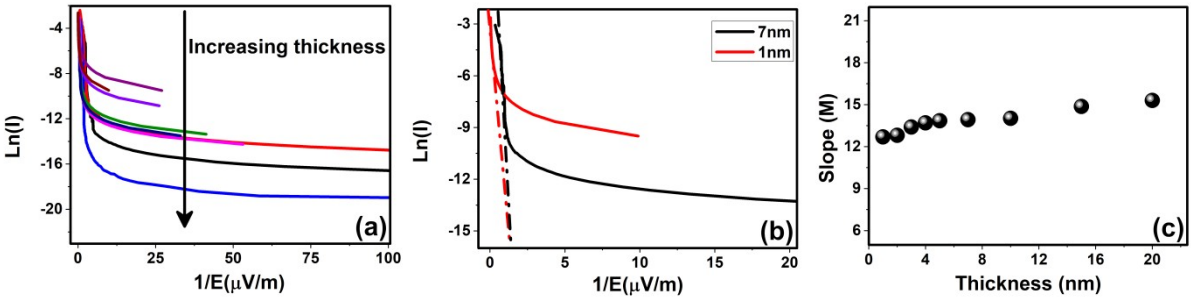


Figure 6.19 (a) Ln(I) vs. 1/E plot for different oxide thickness, (b) a slope comparison between 1nm and 7nm device, and (c) the measured slopes for all the devices.

6.4.6 Direct tunneling

Finally, we inspected the direct tunneling contribution in ReRAM switches. For

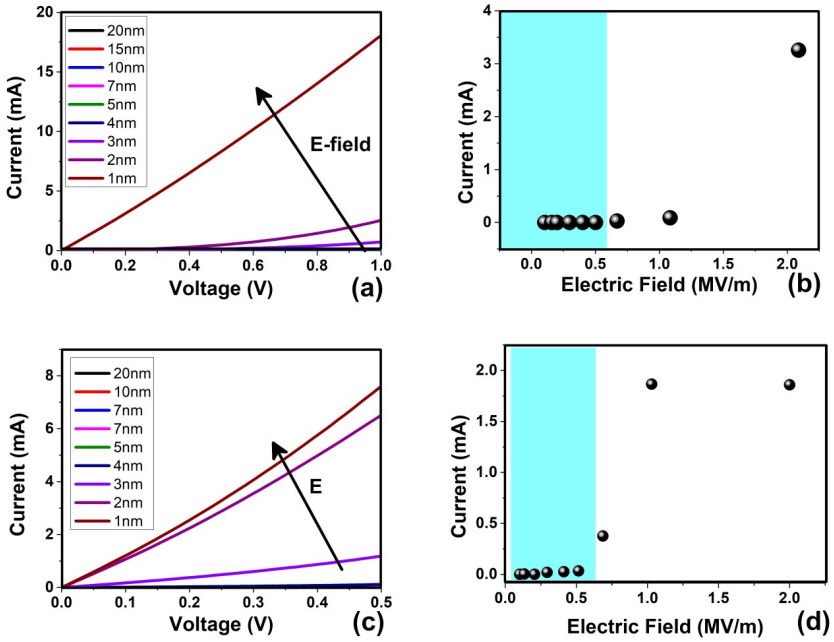


Figure 6.20: Direct tunneling current plots for (a) Au/Al_2O_3/FTO and (c) Ag/Al_2O_3/FTO devices, respectively, and (b) and (d) are tunneling current readings at a low read voltage respectively.

ultrathin oxides, leakage current tends to tunnel through an insulator, causing an enormous leakage current. For devices like resistive switches, this effect can lead to catastrophic degradation of the on-off ratio or even worse, failing to show distinctive HRS and LRS states. One of the distinct characteristics of DT current is its exponential growth as a function of the applied electric field. In the following figure, we have plotted the off current of $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices for different thickness or different electric fields. The exponentially increasing tunneling current seems more drastic for 1, 2, and 3 nm devices with respect to the thicker oxides. This result justifies the I-V behavior of the resistive switches of thinner oxides in figure 6.8 (a).

6.4.7 Conductivity versus temperature

For further elucidation of the conduction mechanism, the currents of the $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices were measured in both on and off conditions. Both of the devices tend to show an increase in off current as a function of temperature, which points towards a hopping based conduction mechanism[15]. However, in LRS all the Au TE devices still continued showing the hopping based conduction but, a significant number of devices with Ag as force electrode behaved in the opposite way, inferring a metallic CF.

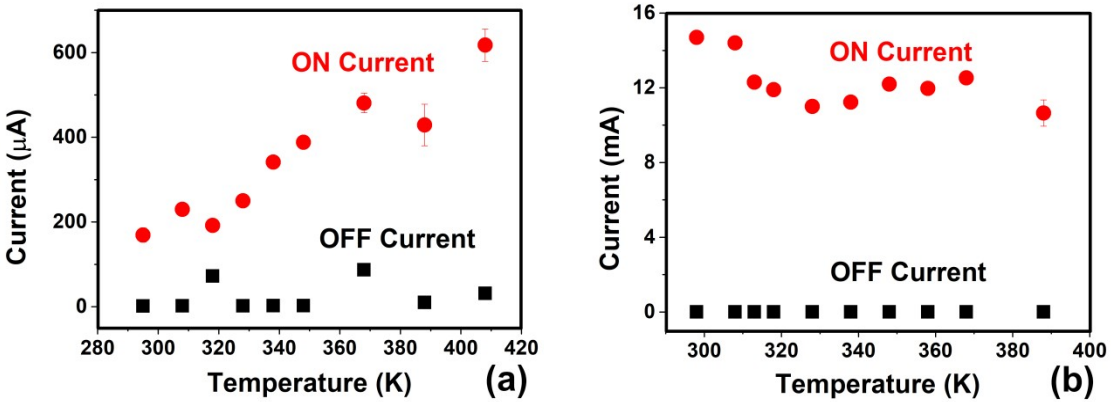


Figure 6.21: (a) and (b) illustrates the output current behavior (at a constant voltage) with temperature of $Au/Al_2O_3/FTO$ and $Ag/Al_2O_3/FTO$ devices, respectively.

One of the pivotal outcomes of this experiment is the LRS current level of the devices. The Ag metallic filament[16] shows a higher conduction value than the hopping path, (i.e. a higher on current) in addition to that the excellent repeatability of these devices makes them perfect resistive switching systems with a very high on-off ratio.

6.4.8 Cycling

Repetitive switching on and switching off is an essential quality check for a ReRAM device. A standard ReRAM device should sustain a train of cycling[17,18] sequence while maintaining the accepted on-off ratio to prove its dependability and agility. The measurement scheme of cycling of a resistive switch can be divided into four consecutive pulses, (a) a pulse slightly higher than the set voltage (V_{s+}) to switch the device on, (b) a read voltage pulse (V_{rd}; usually in millivolt range), (c) a reset pulse with a voltage (V_{r+}) magnitude slightly larger than the reset voltage and (d) a final read voltage of magnitude (see figure 6.21). For quality affirmation generally, a device undergoes a train of such pulses while the output current is monitored.

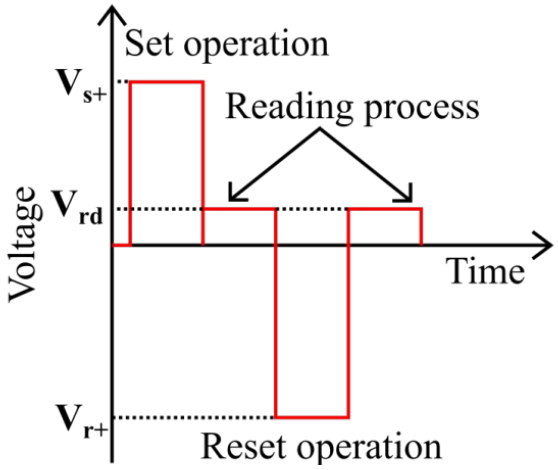


Figure 6.22: Input sequence of a single cycle. V_{s+} and V_{r+} have magnitudes slightly larger than the set and the reset voltages of the devices, respectively. A marginally larger switching voltage always ensures a successful toggling of the states. It does not matter whether the read voltage (V_{rd}) is positive (shown positive in this picture) or negative; as long as it's value is very small (w.r.t V_{s+} or V_{r+}) it would reflect the same result.

We performed this experiment on several Ag/Al_2O_3/FTO and Au/Al_2O_3/FTO devices for 100 cycles. Since the Au/Al_2O_3/FTO repetition is near impossible without compliance check, the devices failed to provide the required output. On the contrary, the Ag/Al_2O_3/FTO devices produced excellent cycling output for each and every cycle while maintaining an on-off ratio in the order of 10^3.

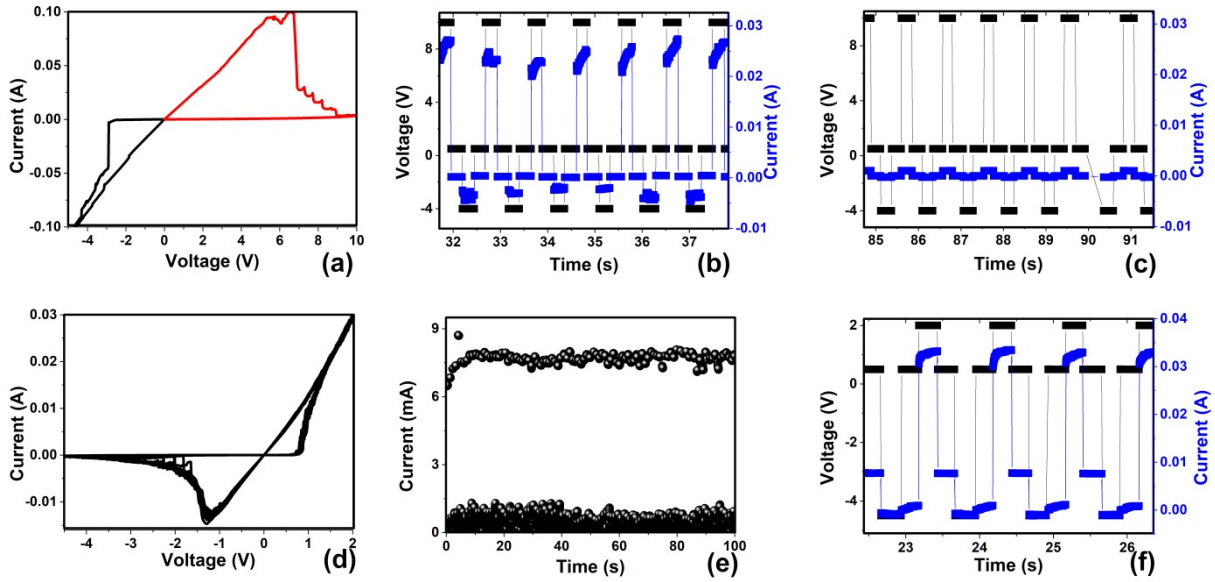


Figure 6.23: (a) and (c) are the I-V characteristics of Au/Al_2O_3/FTO and Ag/Al_2O_3/FTO devices, respectively, (b) and (d) shows the inconsistency of Au/Al_2O_3/FTO cycling behavior; the former depicting a constant ON state of the device and the later demonstrating on and off states with an insignificant on-off ratio and (e) and (f) illustrating the steady cycling output of an Ag/Al_2O_3/FTO device.

From figure 6.23 it can be safely inferred that the Au/Al_2O_3/FTO devices failed to show repetitive switching while the metal ion migration based (Ag/Al_2O_3/FTO) resistive switches are capable of delivering very consistent switching performance while retaining an appreciably good on-off ratio.

6.4.9 Reliability: Weibull distribution

For the different thicknesses of the oxide, we employed Weibull distribution[19] of the set and reset processes, and their slopes analysis to compare the reliability of the resistive switches. From the I-V characteristics of the switches, the SET and RESET statistics were performed and fitted with the formula (chapter 3 equation); the slope thus measured from each of these distributions were later plotted for Ag/Al_2O_3/FTO and Au/Al_2O_3/FTO devices separately as a function of the thickness to draw a proper contrast between them.

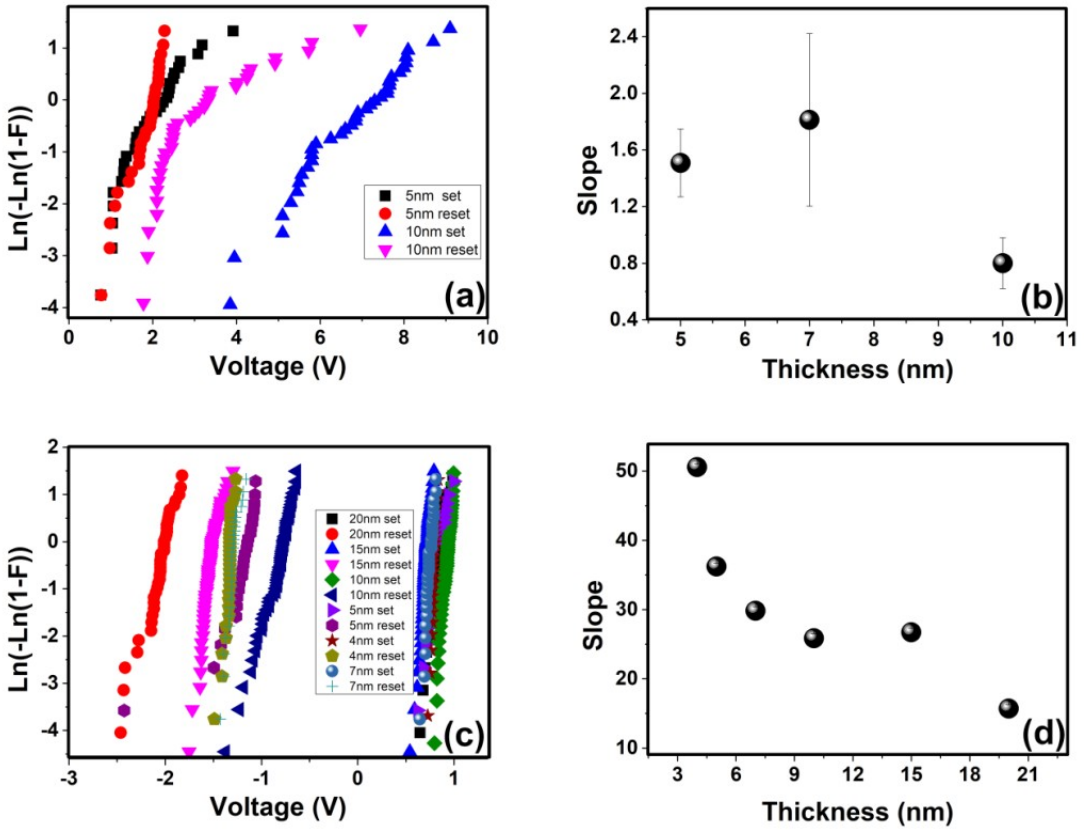


Figure 6.24: (a) and (c) are Weibull distributions of Au and Ag TE devices, respectively, and (b) and (d) are the corresponding slopes for the Set process.

From the above set of data, we can see the transition voltage distributions of Au/Al_2O_3/FTO devices are way inferior to the Ag/Al_2O_3/FTO devices; this claim can be quantitatively confirmed from the measured slopes from figure 6.23 (b) and (d) as well. Since devices with Au TE were extremely difficult to toggle between states repetitively, we could gather enough data only for three different thicknesses for Weibull distribution analysis. Since, devices with Ag TE consistently showed repetition in switching with much more ordered fashion, the Weibull slope is also much larger than its counterpart (see figure 6.23 (d)). Unlike Au TE devices Ag/Al_2O_3/FTO devices present a clear pattern of rising Weibull slope with smaller oxide thicknesses. This proves the fact that smaller the size of a conducting filament (or higher the electric field), easier is the control, thus yielding better packing density of the transition voltages. Another important outcome of this study is the standard deviation of the Weibull slopes itself; the Au/Al_2O_3/FTO device to device has more inconsistency (thus larger deviation) than the Ag/Al_2O_3/FTO ones (much smaller deviation). In nutshell, we can infer

without a doubt that in our case, the metal ion based switching delivered significantly more reliable switching than defect based switching.

6.5 Conclusion

In this chapter, we have fabricated several ReRAM switches with different oxide thicknesses and studied their switching behavior. A steady increase of leakage current (i.e. HRS current) was observed for all the devices as the thickness went down; eventually leading to poor on-off ratios. Further inspections were carried out centering different tunneling and charge injection schemes. FN tunneling study showed a very clear FN and direct tunneling regime and a constant barrier lowering with lowering thickness leading to higher leakage current. PF injection evidence was also spotted in I-V characteristics of our devices. From the y-intercept of the PF plots, it was clear that the trap barrier in the oxide faces more drastic lowering with decreasing thickness, thus contributing to the leakage current. TAT evidence was also quite legible for all these devices, suggesting the trap sites' role as small direct tunneling sites in the oxide. From the leakage current plot, the 3 thinnest oxide layers (1, 2, and 3nm) showed the catastrophic effect of the direct tunneling effect. One crucial observation was FN tunneling and PF injection were dominant in all the devices except the one with 1nm oxide. However, both TAT and DT were evident in 1nm oxide; with DT being negligible at higher thickness and TAT being stronger at thicker dielectrics. This infers the extremely high electric field across 1nm oxide causes an unprecedented barrier lowering (thus no FN tunneling) and trap-barrier lowering (thus no PF injection) leading the charges to mostly tunnel from one electrode to the other and a few to tunnel between the trap sites. Finally, we studied the HRS and LRS current behavior with temperature change. Both the HRS and Au TE devices' LRS hinted hopping based conduction mechanism whereas the highly conducting metallic filament showed decreasing conductivity with an increase in temperature.

The chapter further deals with the cycling and reliability tests of these devices. The resistive switches with Au TE showed little to no success in the cycling study while the Ag TE devices excelled in their performance. Finally, from the Weibull slope measurement, we concluded that Ag/Al_2O_3/FTO devices are superior over Au/Al_2O_3/FTO in terms of reliability and repeatability. Further, with thinner oxide thicknesses the reliability went higher, suggesting that a higher electric field is better for controlling the filament with precession.

In chapter 5 several investigations were carried out to determine the presence of metal ions in a MOS dielectric by C-V methods (especially BTS). This chapter encountered several

pieces of evidence showcasing the difference between the metal ions versus defect mediated switching mechanisms as well. From PF study we learned the metal ions introduce smaller trap barriers (see figure 6.17) to cause the ease of charge transfer, thus leading to higher leakage current and earlier breakdown than its counterpart. Switching based on the direction of the electric field (see figure 6.7) was also possible for $Au/Al_2O_3/FTO$ systems because of the defect based filament formation. The devices with Ag electrodes could switch on at positive electric field since metal ions can migrate only at that particular electric field direction. As correctly pointed out in the previous chapter, the BTS study along with the evidence collected from this work gives us a clear picture of metal ion and oxygen defect based switching mechanism. We have also established (cycling and reliability studies) the fact that Resistive switches with Ag as the top electrode deliver much better switching characteristics in terms of repeatability and reliability.

References

1. Sun, P. *et al.* Thermal crosstalk in a 3-dimensional RRAM crossbar array. *Sci. Rep.* **5**, 13504 (2015).

2. Bai, Y. *et al.* Stacked 3D RRAM Array with Graphene/CNT as Edge Electrodes. *Sci. Rep.* **5**, 13785 (2015).

3. Wu, Y., Lee, B. & Wong, H.-S. P. Al_2O_3-Based RRAM Using Atomic Layer Deposition (ALD) With 1-muA RESET Current. *IEEE Electron Device Lett.* **31**, 1449–1451 (2010).

4. Spahr, H. *et al.* Regimes of leakage current in ALD-processed Al_2O_3 thin-film layers. *J. Phys. Appl. Phys.* **46**, 155302 (2013).

5. Spahr, H. *et al.* Conduction mechanisms in thin atomic layer deposited Al_2O_3 layers. *J. Appl. Phys.* **114**, 183714 (2013).

6. Ramesh, S., Dutta, S., Shankar, B. & Gopalan, S. Identification of current transport mechanism in Al_2O_3 thin films for memory applications. *Appl. Nanosci.* **5**, 115–123 (2015).

7. Patiño, E. J. & Kelkar, N. G. Experimental determination of tunneling characteristics and dwell times from temperature dependence of Al/Al $_2$ O $_3$ /Al junctions. *Appl. Phys. Lett.* **107**, 253502 (2015).

8. Houng, M. P., Wang, Y. H. & Chang, W. J. Current transport mechanism in trapped oxides: A generalized trap-assisted tunneling model. *J. Appl. Phys.* **86**, 1488–1491 (1999).

9. Fleischer, S., Lai, P. T. & Cheng, Y. C. Simplified closed-form trap-assisted tunneling model applied to nitrided oxide dielectric capacitors. *J. Appl. Phys.* **72**, 5711–5715 (1992).

10. Yen, T. J., Gismatulin, A., Volodin, V., Gritsenko, V. & Chin, A. All Nonmetal Resistive Random Access Memory. *Sci. Rep.* **9**, 6144 (2019).

11. Kalantarian, A. *et al.* Controlling uniformity of RRAM characteristics through the forming process. in *2012 IEEE International Reliability Physics Symposium (IRPS)* 6C.4.1-6C.4.5 (IEEE, 2012). doi:10.1109/IRPS.2012.6241874.

12. Bricalli, A. *et al.* SiO_x-based resistive switching memory (RRAM) for crossbar storage/select elements with high on/off ratio. in *2016 IEEE International Electron Devices Meeting (IEDM)* 4.3.1-4.3.4 (IEEE, 2016). doi:10.1109/IEDM.2016.7838344.

13. Hu, C., McDaniel, M. D., Ekerdt, J. G. & Yu, E. T. High ON/OFF Ratio and Quantized Conductance in Resistive Switching of TiO_2 on Silicon. *IEEE Electron Device Lett.* **34**, 1385–1387 (2013).

14. Gushterov, A. & Simeonov, S. Extraction of trap-assisted tunnelling parameters by graphical method in thin n-Si/SiO_2 structures. *J. Optoelectron Adv. M.* **7**, 1389-1393(2005).

15. Zubair Ansari, M. & Khare, N. Thermally activated band conduction and variable range hopping conduction in Cu_2 $ZnSnS_4$ thin films. *J. Appl. Phys.* **117**, 025706 (2015).

16. Pérez, E. *et al.* Impact of temperature on conduction mechanisms and switching parameters in HfO_2-based 1T-1R resistive random access memories devices. *J. Vac. Sci. Technol. B Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* **35**, 01103 (2017).

17. Rodriguez, N. *et al.* Resistive Switching and Charge Transport in Laser-Fabricated Graphene Oxide Memristors: A Time Series and Quantum Point Contact Modeling Approach. *Materials* **12**, 3734 (2019).

18. Rana, A. M. *et al.* Endurance and Cycle-to-cycle Uniformity Improvement in Tri-Layered CeO_2/Ti/CeO_2 Resistive Switching Devices by Changing Top Electrode Material. *Sci. Rep.* **7**, 39539 (2017).

19. Mangalam, J., Agarwal, S., Resmi, A. N., Sundararajan, M. & Jinesh, K. B. Resistive switching in polymethyl methacrylate thin films. *Org. Electron.* **29**, 33–38 (2016).

Chapter 7

Single QD based optical synapse

7.1 Introduction

In 1945 Jon von Neumann[1] proposed digital computer architecture with

1 Processor: containing an arithmetic logic unit (ALU) and processing registers.
2 Control unit: with an instruction register and program counter
3 Internal memory: which contains all the data and the instructions
4 External memory: like hard drive, FLASH drive, solid-state drive, etc.
5 Input and output peripherals: keyboard, mouse, etc.

This design later became famous by the name of von Neumann architecture or the Princeton architecture. However, this design had different sets of buses for instruction fetching and data processing and as a result, the simultaneous occurrences of these two operations were impossible, which costs overall processing time. This generic limitation later became popular as the von Neumann's bottleneck. A slightly advanced version of this architecture was recognized afterward, known as the Harvard architecture; this particular design had dedicated bus lines for instruction read and data processing operations. Although the modern computers apparently appear to be based on von Neumann's architecture (since the instructions/programs and data are stored in the same memory), internally arranged separate bus lines and cache memory partially circumvents the von Neumann's bottleneck. This arrangement is considered to be a modified version of Harvard architecture.

For the past eight decades, these architectures have proven their efficacy and have successfully led us to the fast computing era. But with growing demand in computing speed and with the fast catching up of the artificial intelligence (AI), scientists and engineers have started looking for the basic building blocks which are apt for heavy load computation and parallel processing. The current memory technology like FLASH[2] and the next generation memory schemes such as MRAMs[3], PCRAMs[4], or even ReRAMs[5] are not ideal for these kinds of sophisticated operations. On the other hand, at present, the quantum processors are operable only at sub-zero temperatures. That is why today's supercomputers and AI hardware mainly relies on the software modifications and parallel processing systems constructed with multiple processors/cores.

To get the real essence of parallel computing and ultrafast computing, the challenge was to design the basic building blocks that are ideal for such purposes. It did not take much time to realize that similar computers already exist in the nature in form of neural networks in the brain or the synaptic systems. Our brain does not comprise separate memory and processor, rather relies on the network of the neurons (through the synaptic terminals) to process the huge set of data. In a pursuit to mimic similar operations, the scientific community started to fabricate different artificial synaptic systems like 2-terminal[6], 3-terminal[7], and optical synapses[8]. The most recognizable property of a synaptic junction is the "plasticity"[9] of a particular device. It is defined as the synaptic response to a particular stimulus that becomes stronger (potentiation)[10] or weaker (depression)[11] with successive repetition of the event. The methodology to realize these phenomena in an electronic system has been demonstrated in detail in the next section of this chapter.

7.2 Literature survey

Artificial synapses accept a pulse train (electrical or optical depending upon the type of the synapse) as the input stimuli and produce a similar output with successively increasing

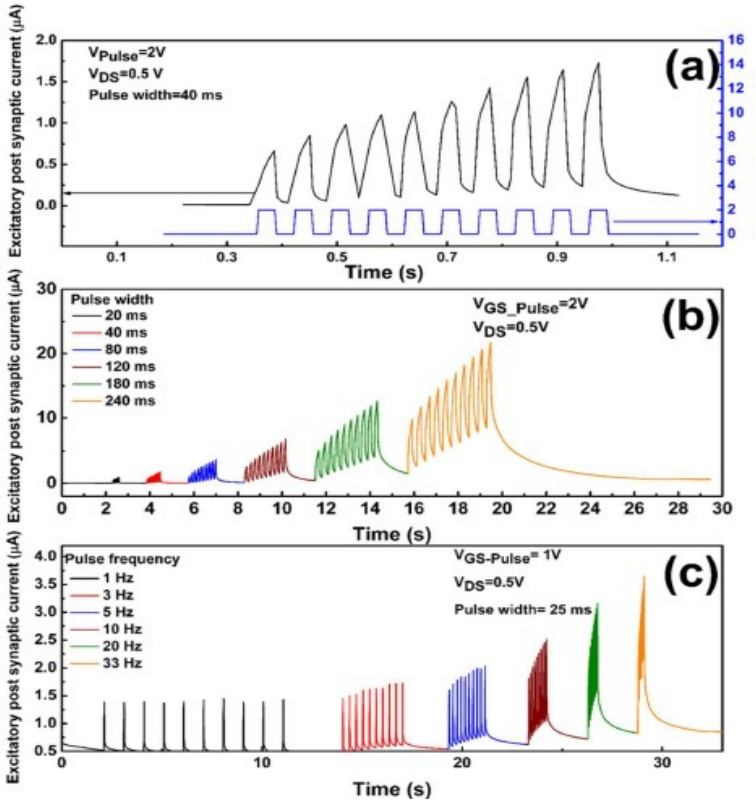


Figure 7.1:(a) A 40ms input pulse train generates a successively rising pulsed output, (b) PPF response gets better as the input pulse width increases, and (c) so does for increasing frequency as well; *courtesy ref* (Balakrishna Pillai and De Souza 2017).

pulse amplitude. The rate of increase of these pulses can be quantified from the ratio of one pulse to its preceding pulse. This ratio is technologically termed as Pulsed-pair facilitation (paired-pulse facilitation) or PPF[12]. The input signal can be fed with a 50% duty cycle and varying frequency or with fixed on/off time and varying duty cycle. Pillai and De Souza's work[13] pictorially depicts the previously mentioned measurement scheme in the figure7.1. Pillai and De Souza fabricated a thin film transistor (TFT) based neuromorphic device with a ZnO active layer and tantalum oxide (Ta_2O_5) as the gate dielectric. In the above diagram, figure 7.1 (a) the increasing height of spikes (black) indicates the learning process of a synaptic system. The second part shows, with an increase in the duty cycle of the input signal how the PPF response becomes much more prominent. However, the last picture comes with the most admiring result, the PPF response or the final spike amplitude increase with the increasing input frequency; it certainly draws an operational similarity with a high pass filter.

Another recent work from Zidan et al. used a system of synaptic two terminal ReRAM devices and by employing a crossbar fashioned arrangement to solve a very complicated set of partial differential equations for analyzing a set of complex mathematical equations[14]. We have already established this far that depending upon the conductivity of the dielectric ReRAMs operate either at state "0" or at state "1", but Zidan and his team modulated the conductivity of ReRAMs in between these two states to emulate the different states of plasticity of a synaptic junction. As compared to Pillai's group this work mainly focuses upon solving a real-life problem by a synaptic network or in other words it shows the massive potential this technology holds for us in the near future. Zidan et. al. formed a set of matrices of differential equations and broke down those matrices into a substantial number of sparse matrices. These sparse matrices were separately mapped onto the crossbar matrix later on to solve the aforementioned problem.

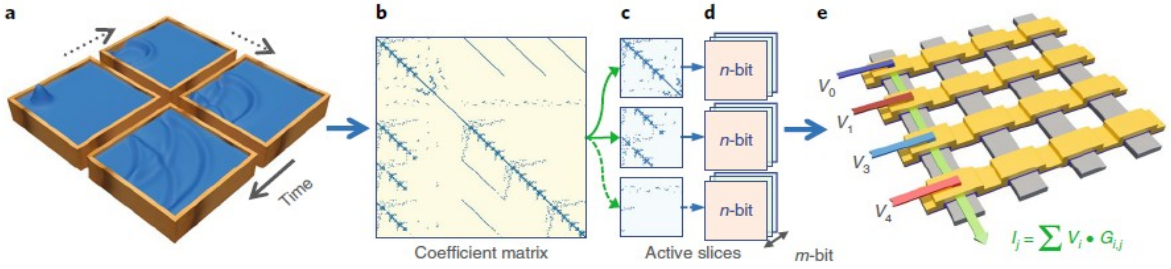


Figure 7.2: (a) shows the time evolution of a water wave, (b), (c), (d) the sparse matrices' arrangement across the crossbar matrix and (d) a section of the crossbar matrix showing one of the sparse matrix's mathematical operation; *courtesy ref (*Zidan et al. *2018).*

S. Seo's work in 2018[15] drew a different perspective in their work. Their entire work was dedicated to a very specific synaptic system, which is abundantly found throughout the fauna kingdom, the optical synaptic junction; the eye. Their system at the very core was a series connection of an optical sensor and an electrical synaptic device. The optical sensors reacted differently to different wavelengths of light and that was recorded in the synaptic network in form of different weights. This collection of sensors and the synapses along with the wavelength-dependent weights was successfully able to detect the difference in the RGB ratio of a particular illumination, thus mimicking a human eye.

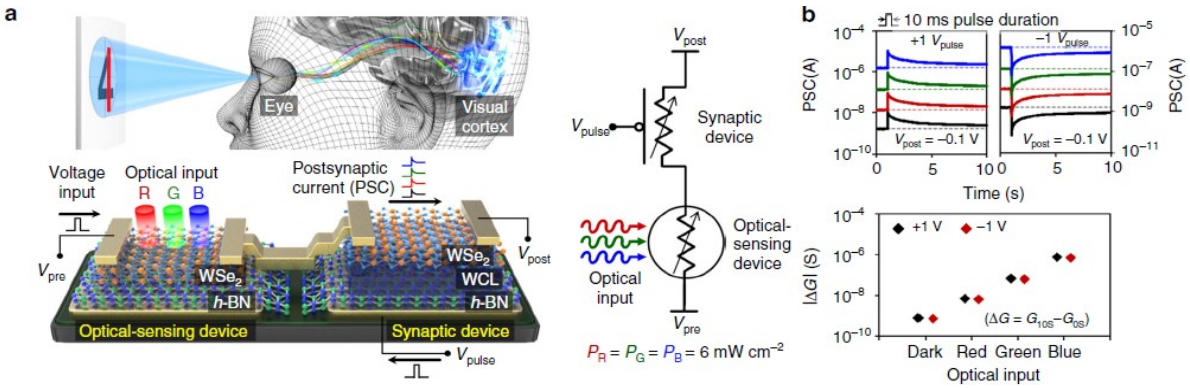


Figure 7.3: (a) schematic of a human optic nervous system with the circuit diagram and device configuration and (b) different conductance response for different wavelength stimuli; *courtesy ref (*Seo et al. 2018).

However, a practical human optical network does not have a separate sensor and a synaptic system built-in; rather, the sensors (rod and cone cells) are the sensing extreme of the network and the synaptic junction of the same neuron in the brain (along with other synapses) form the processing environment. To simulate this behavior we required an optical sensor with synaptic plasticity. This kind of system requires the condition that the output spikes should be characterized by a slower decay time with respect to the rise time. We monitored individual MoS_2 (molybdenum disulfide) quantum dots (QDs)[16–18] in a scanning tunneling microscope (STM) and under certain voltage bias when exposed to ultraviolet (UV) pulse, a sudden rise in tunneling current (conductivity) was observed with slower decay time. The following figure corroborates the aforementioned scenario.

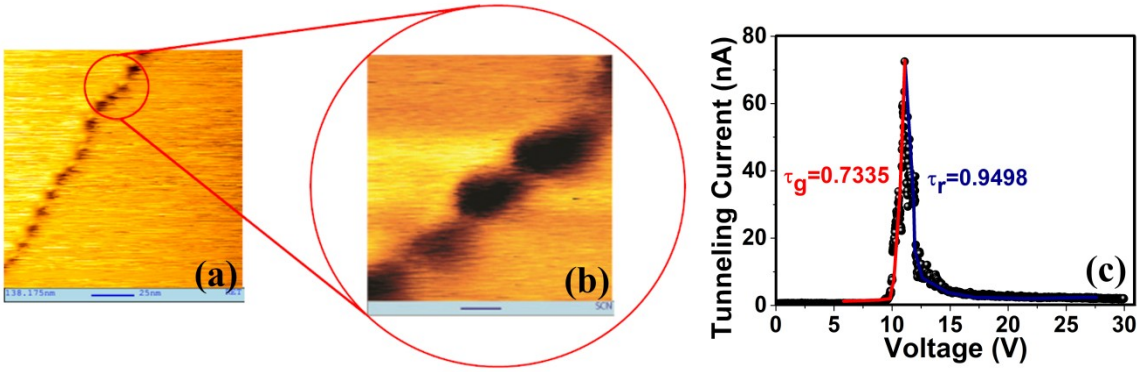


Figure 7.4: (a) STM image of a channel of MoS_2 QDs, (b) a magnified section of the same and (c) a photocurrent spike when the tunnel junction is illuminated with an ultraviolet pulse. The generation time constant (τ_g in red) and recombination time constant (τ_r in blue) are obtained from the exponential fits the data.

Figure 7.4(a) and (b) are STM images of a QD trail and a zoomed-in portion respectively. In figure 7.4(c) we can see the generation time constant (τ_g=0.73s) is significantly lower than the recombination/decay time constant (τ_r=0.95s) for a two-second pulse. Since the recombination process is much slower than the generation process, theoretically these QDs can serve as an ideal optical synapse. Exciton (strongly bound electron-hole pair) generation upon optical excitation in metal dichalcogenides including MoS_2 has been reported earlier[19] and their optoelectronic properties have been quite extensively explored. Although the optical synaptic properties of MoS_2 still remains an uncharted territory and this chapter intends to throw light on it.

7.3 Experiments

The MoS_2 QDs were prepared by liquid-phase exfoliation method. These QDs were dropped cast on HOPG (highly oriented pyrolytic graphene) and mounted inside an STM. The samples were kept at constant voltage stress of 0.5V, while the tunneling current being monitored in a separate window. An Arduino based LED controller was mounted and the light was focused through a convex lens to illuminate the QDs with optical pulses. The current experiment was subdivided into the following four sub-sections.

We measured the PPFs of the QD tunnel junction as a function of

a Energy of the illuminated light,

b Duty cycle of the input pulse train,

c Frequency of the input pulse train and

d QD size.

The following two diagrams schematically illustrate the experimental setup used to perform this work and its promise as an alternative to an intelligent-biological-optical-network respectively.

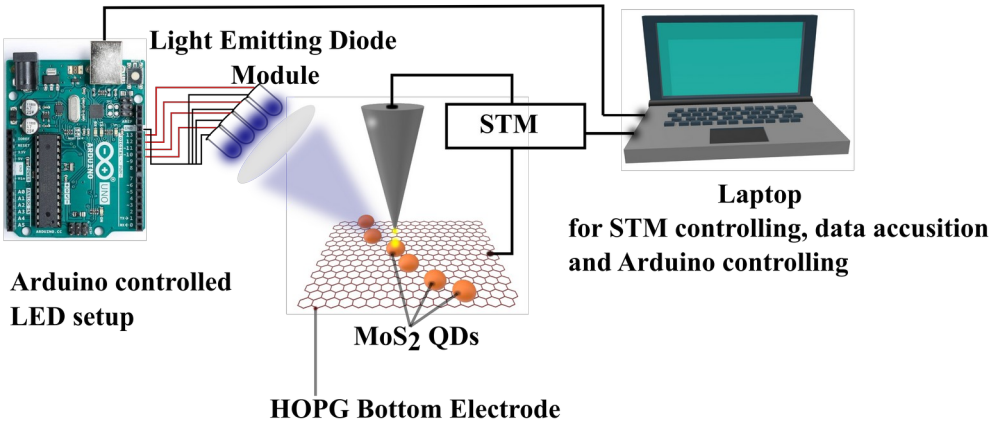


Figure 7.5: Experimental setup; an array of QDs drop cast on a HOPG sample and mounted inside an STM to monitor the output spike sequence under optical pulsed stimuli.

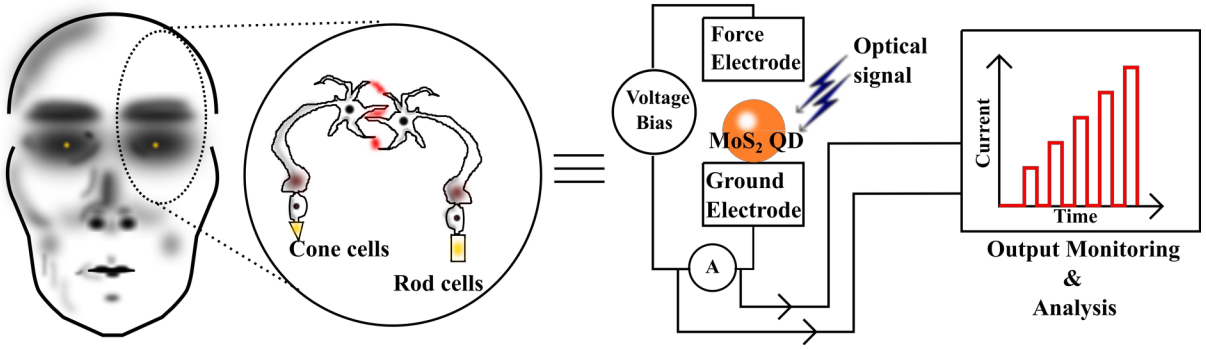


Figure 7.6: MoS_2 QDs' system can potentially be used as an AI alternative to the rod and cone cells of a biological system

7.3.1 PPF as a function of illumination energy

We measured the local density of states (LDOS) of the individual quantum dots employing the differential conductance method. The bandaps of the quantum dots were measured from the LDOS measurement. The following figure shows the density of states and the bandgap of a single QD, of which the bandgap has been calculated to be around 3.2 eV.

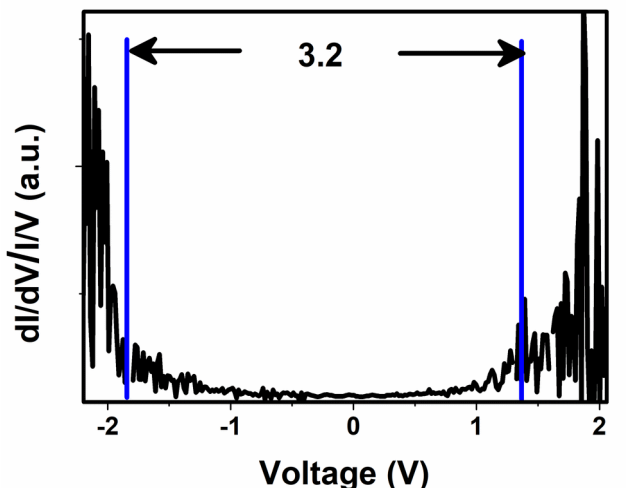


Figure 7.7: Bandgap of single MoS_2 QD estimated from LDOS measurement.

In order to excite the QD optically, it is evident that we need equal to or higher energy illumination that would generate a substantial amount of excitons. To verify this hypothesis, the tunnel junction was illuminated with two different wavelengths, as shown in figure 7.8. A pulse train of 2-second on and 2-second off was used to excite a selected QD. As seen from figure 7.8, the tunneling current does not respond to the illumination by a red photon of wavelength ~670 nm, but it clearly responds to the UV photons of ~365 nm wavelength.

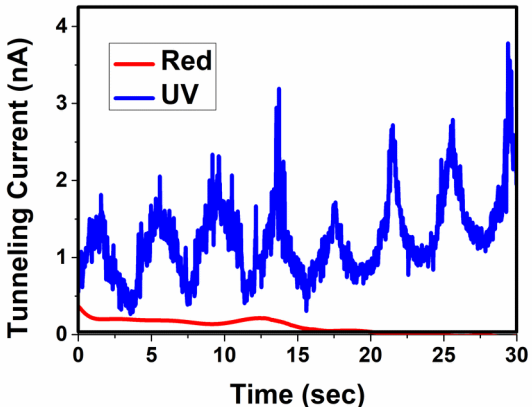


Figure 7.8: A comparative PPF study with UV (~365 nm) and Red (~670 nm) pulses of a MoS_2 QD. The absence of PPF under red illumination is a clear indication of the dependence of QD's synaptic response on its bandgap.

7.3.2 PPF as a function of Duty cycle

It is evident from the figure that the UV pulse train was able to render a positive PPF (blue colored plot) of 1.25±0.10 and red led pulse failed to generate enough excitons to cause any

potentiation plasticity (red-colored plot). It should be noted that for an optical stimulus it is not possible to study the depression plasticity.

Next, the PPF response was measured as a function of the duty cycle of the input optical pulses. Since it is already verified that the recombination is a slower process in this system, by keeping the on-time constant we reckoned if the off time keeps on shrinking, whether the subsequent spikes will be stronger; therefore, delivering a better PPF. In the first part, the on-time was fixed while the off-time durations were varied. The following plot shows the response with the on-time remaining constant, and off time-varying as 5, 3, and 2 seconds. It is clear that the PPF value increases with decreasing off time. With slower recombination, the second input pulse always starts at a higher value, as the entire generated electron-hole had not decayed out yet. A smaller off time ensures a larger head start, which eventually leads to a stronger follow-up spike.

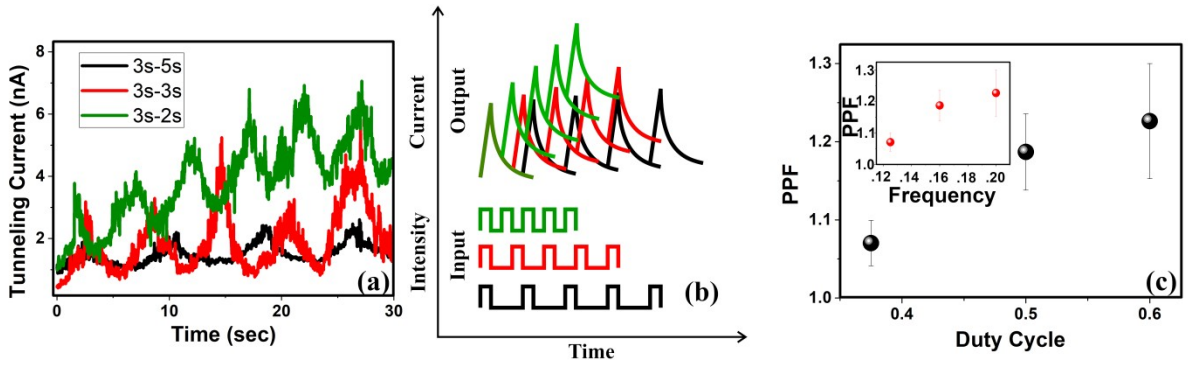


Figure 7.9: (a) PPF response of 3s on and 5s, 3s, 2s off-time pulse trains, (b) a schematic representation of the increasing PPF phenomena, and (c) is the calculated PPF as a function of duty cycle and frequency (inset).

In figure 7.9(b), we have depicted the real scenario of the previous discussion. Correlating the input light pulse with the output sequence, one can easily affirm that a smaller off time leads to a higher starting point, hence producing stronger PPF response.

In the next section of the experiment, we kept the off-time constant and varied the duty cycle by changing the on-time. Although the observed result was apparently similar, i.e. an elevated PPF as a function of duty cycle, the reason was different from the earlier case.

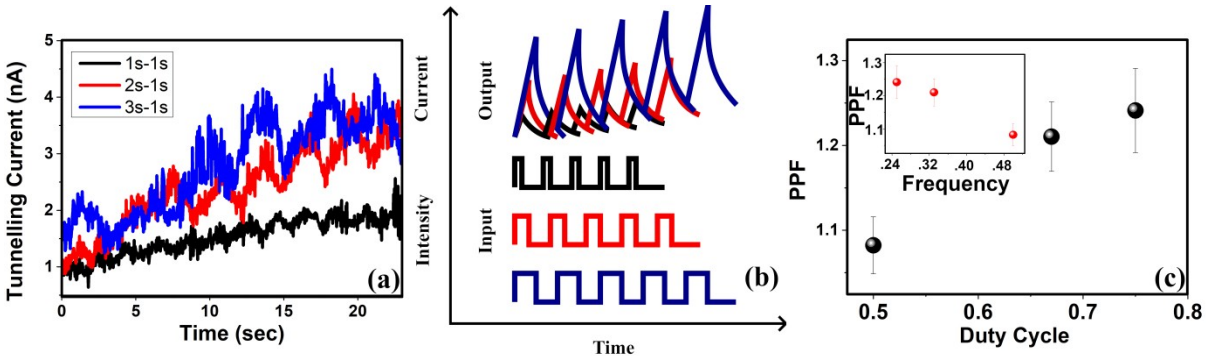


Figure 7.10: (a) PPF response of MoS_2 QD with constant off time and varying on time, (b) a schematic representation of increasing PPF phenomena, and (c) the PPF as a function of duty cycle and frequency (inset).

We measured the generation time constants (τ_{g1} and τ_{g2}) and recombination time constants (τ_{r1} and τ_{r2}) for different on pulses by fitting the rising and falling curve with the following equations.

$$I = I_0 + A_1 e^{(t-t_0)/\tau_{g1}} + A_2 e^{(t-t_0)/\tau_{g2}} \qquad (7.1)$$

$$I = I_0 + A_1 e^{-(t-t_0)/\tau_{r1}} + A_2 e^{-(t-t_0)/\tau_{r2}}, \qquad (7.2)$$

where I is the tunneling current and t is the time.

The obtained result shows a very general trend of decreasing nature of generation time constant with increasing time. Since a lower time constant suggests a faster increase in the current magnitude; this result is a direct reflection of the fact that more exposure to the light causes a proliferation of excitons in our QDs. However, a greater on time does also mean higher output amplitude or more excitons (as the current will keep on increasing as long as the light is on), and based on that, our earlier argument could not be decisive enough. For further insight, we studied the decay/recombination time constant, and from the analysis, it became evident that the recombination process slows down when the system contains more electron-hole pairs; thus providing a stronger starting threshold for the next spike.

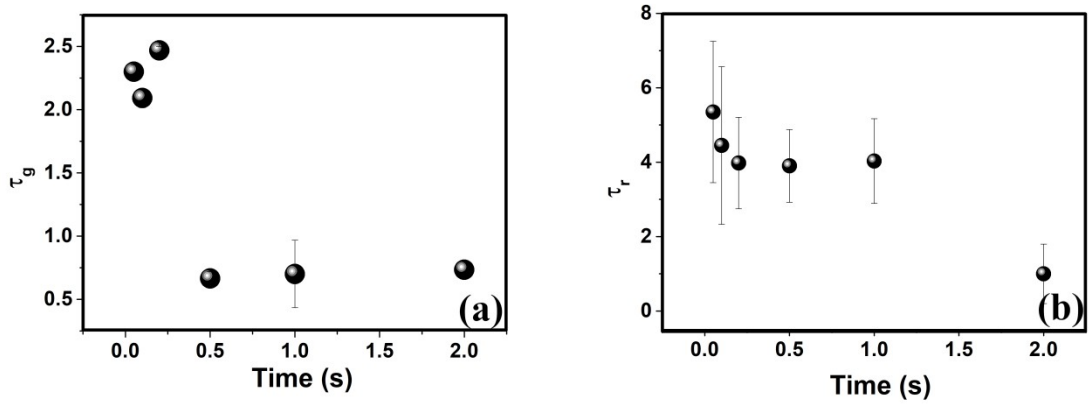


Figure 7.11 (a): the generation time constant shows a vague decreasing trait with the pulse on time and (b) the recombination time showing similar trend with the on-time as well.

Equation (7.1) will show a faster increase in the y value with lower τ_g and vice-versa. Drawing correlation from this logic we can safely assume that figure 7.11(a) concludes that a larger illumination time results in a larger number of exciton pairs, all with a slow decay time-constants. On the other hand, a lower or higher value of τ_r in equation (7.2) will show a faster or slower deviation in the y value respectively. Figure 7.11(b) illustrates τ_r's behavior with on time and clearly suggests a slower rate of recombination when there is already a larger presence of electron-hole pairs in the system.

In the next experiment, we illuminated the QDs for 2 seconds followed by a relaxation period of 8 seconds. The exciton generation in the QDs does reflect in the output data in form of spikes but fails to show any significant PPF. From the earlier discussions, we can safely assume that the larger relaxation period with respect to the on-time provided enough time for the recombination process to take place and hence the result.

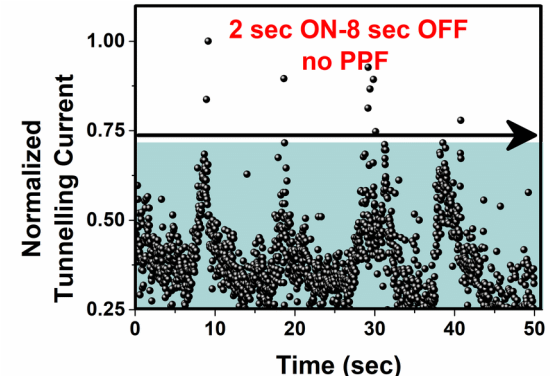


Figure 7.12: With a much larger off time the QDs do not show any plasticity.

In the above figure, a broader off time relaxes all the electro- hole pair which eventually brings down the current amplitude and results in an unsuccessful attempt to generate PPF in the system. In a pursuit to obtain sufficient PPFs, the off-time was shrunk down, starting from 5seconds to 50 milliseconds while sticking to the same on-time duration (2 seconds). The following set of figures illustrates the obtained output and the calculated PPF as a function of the duty cycle.

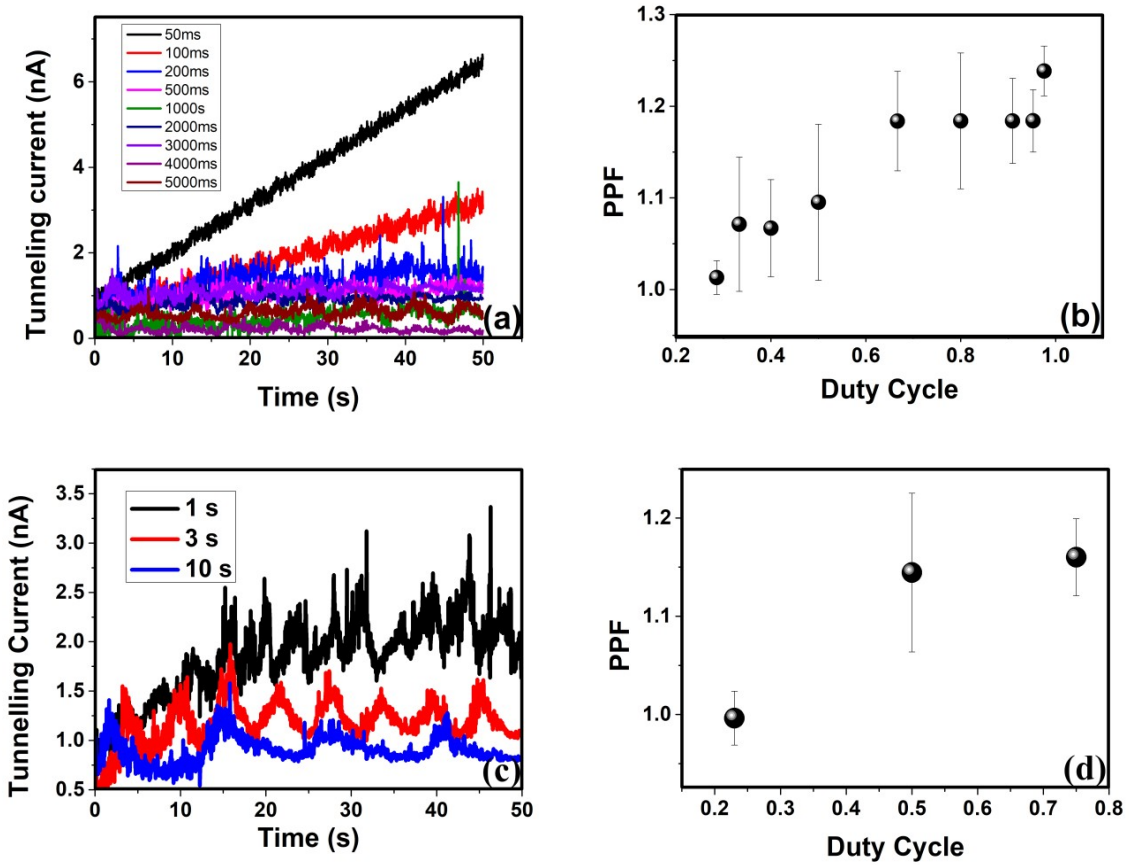


Figure 7.13 (a): The output for different off-time durations for 2 seconds of on-time (b) the calculated PPF as a function of the duty cycle for the same, (c) the output response for different off-time durations for 3 seconds of on time and (d) its PPFs as a function of duty cycle.

The analysis revealed the QDs to show the least PPF (~1.008) when the off-time was at its peak (i.e. 5 seconds) and the PPF itself reached at its best (~1.26) when the relaxation period was the least (50 milliseconds) in figure 7.13(a) and (b). A similar trend has been observed for 3 seconds of on time and with different duty cycles, as demonstrated in figure 7.13(c) and (d).

7.3.3 PPF as a function of frequency

In this section, we had fixed the duty cycle to 50% (i.e. on time and off-time durations are identical) and varied the time period/frequency of the input signal to study the photocurrent response of the MoS_2 QDs. A set of pulse trains of time period ranging from 4 seconds (2 seconds on-2second off) to 40 milliseconds (20 ms on-20 ms off) were fed to the circuit to study the neuromorphic behavior. Since it was difficult to stick to a single QD and perform the entire investigation without drifting the STM-tip off, we decided to cram 2s time period to 40ms time period pulse trains in a single signal and a separate pulse train signal for 4s time period. The following are the pictorial depiction of the acquired data.

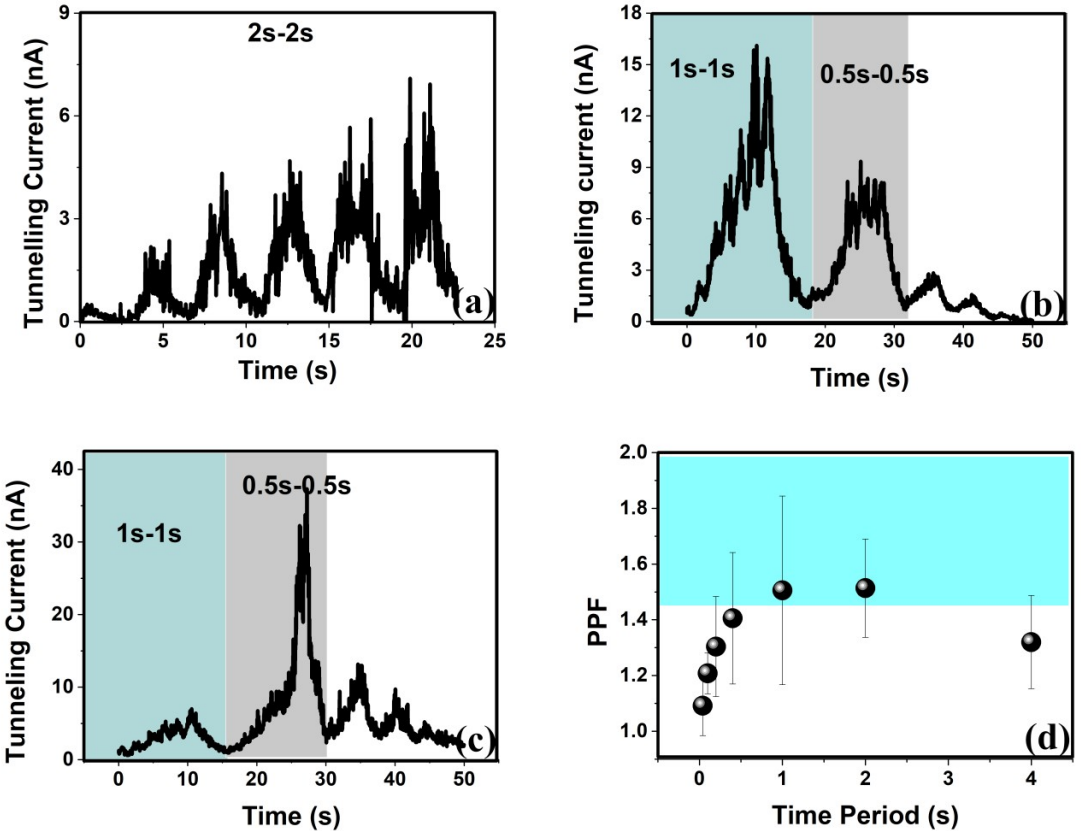


Figure 7.14 (a) output response corresponding to 4s time period pulse train, (b) 2s time period input shows the superior PPF response; most common, (c) 1s time period input shows the superior PPF response; occurs less often and (d) the calculated PPF as plotted as a function of time period.

The first figure 7.14(a) illustrates the synaptic behavior for a 2s-2s input signal with a PPF response of 1.32±0.17. Figure 7.14(b) and (c) show the same for the rest of the pulsed signals and it is evident that the most aggressive outputs were from 1s-1s and 0.5s-0.5s input pulses respectively. It is important to note that, statistically the superior PPF response was more

probable in 1s-1s pulse inputs than the 0.5s-0.5s ones. The rest of the pulsed signals failed to render PPFs as significant as the aforestated ones. Careful observation from the generation and recombination time constants revealed the reason behind this behavior and has been graphically illustrated in the following figures.

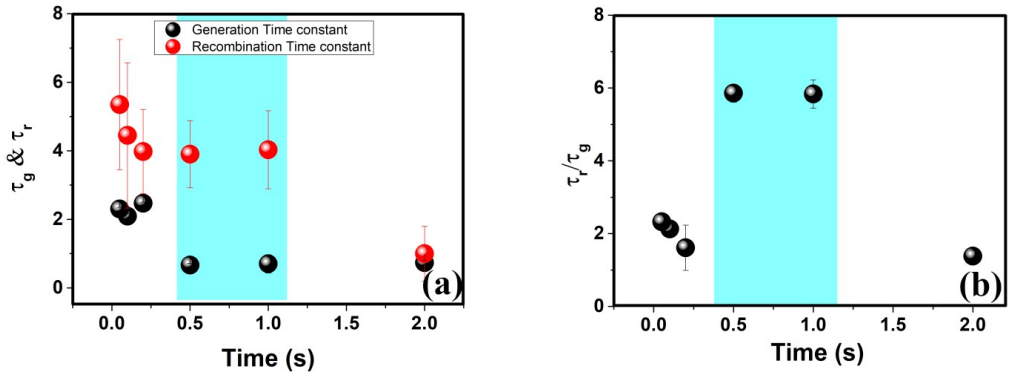


Figure 7.15 (a) a combined representation of rise and decay time constants at different on times and (b) rise to decay time constant ratio showing the maximum PPF occurs when the decaying process is slowest and rising process is the fastest

7.3.4 PPF and quantum confinement

In the earlier discussions, it has already been pointed out that a lower τ_g would suggest a faster generation process and a higher τ_r would suggest a slower decay process, therefore such a combination would yield the best PPF response. By plotting τ_g and τ_r separately (figure 7.15(a)) and as a ratio of τ_r/τ_g (figure 7.15(b)), it became quite clear that at 0.5s-0.5s and 1s-1s pulse inputs, the data set is at the highest strata in the entire τ_r/τ_g spectrum and thus generating the maximum PPFs in figure 7.15(d).

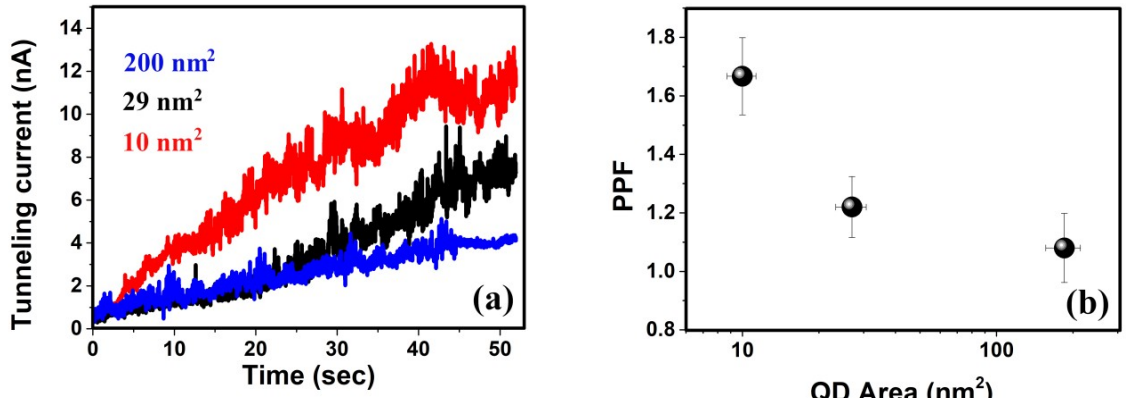


Figure 7.16: (a) The spiked output response of MoS_2 QDs with varying size and (b) the corresponding PPF.

Finally, we investigated the consequence of quantum confinement on the neuromorphic response of the MoS_2 QDs. Three different sizes of QDs were drop-casted on the HOPG sheet and were exposed to a train of 2s-1s pulses (duty cycle=66.7 %). It is well known for QDs to exhibit a larger bandgap due to quantum confinement when the size decreases. We predicted this confinement (due to size variation of the QDs) might slow down the decay time of the output spikes and thus modulating the PPF output. As we have seen earlier a slower decay process can significantly influence the PPF output, the same nature of output can be safely predicted in this scenario as well. In the following set of figure, we have presented the output response and the calculated PPFs for different sized QDs. Figure 7.16 (a) shows a clear difference in the output magnitude and how it is greatly influenced by the size of the QDs. The calculated PPF is plotted in figure 7.16 (b). To study the reason leading to this behavior, the decay time constant was measured from the single pulse response of the QDs'.

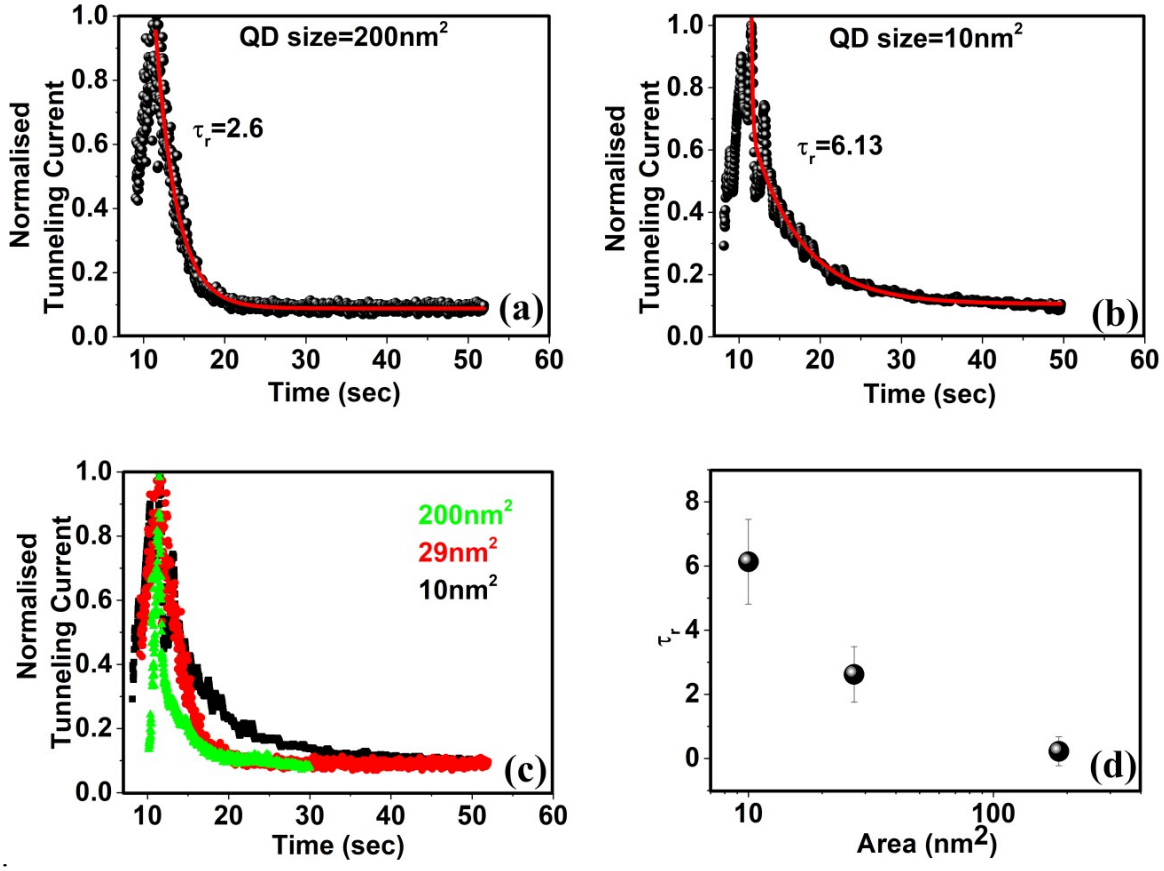


Figure 7.17: (a) and (b) are the calculated decay time constant from the single pulse response of a 200nm² and a 10nm² QD respectively, (c) a set of decay characteristics for different sized QDs and the calculated τ_rs for various sized QDs.

Figure 7.17 (a) and (b) show the decaying nature and their respective calculated τ_rs for two different sizes of QDs. Figure 7.17 (c) is a comparative portrayal of the decay characteristics of different QDs; the calculated τ_rs of the same being plotted in figure 7.17 (d). This demonstration undoubtedly suggests that, as predicted previously the decay time slows down as the QD size decreases; therefore, the improvement in the PPF response.

7.5 Conclusion

In this chapter, we have demonstrated the measurement and analysis of a novel optical neuromorphic system based on quantum dots. To quantify the synaptic behavior of such a device, we measured the PPF as a function of wavelength, duty cycles (on time constant with off-time varying and vice-versa), frequency, and QD size. From the obtained band gap of the QDs (by local density of state measurement) we predicted that the plasticity can be observed in UV periphery and the devices delivered the same while failing to show anything at a lower energy illumination (red light). As a function of duty cycle, we observed more pronounced PPFs at higher duty cycles. From the rise time and decay time calculation, it was concluded that a larger on-time assures a larger exciton generation and hence a higher PPF. However, for different frequencies (with a constant duty cycle of 50%) the most enhanced PPF response was observed at 1s-1s and 0.5s-0.5s input cycles, while showing significantly lower responses in higher and lower frequencies. It should be noted that with on-time remaining constant if the off-time is reduced the input signal frequency increases and for such an input, the observed response resembled a high pass filter response. With the off-time remaining constant and varying on-time would suggest a lower frequency with a higher duty cycle; the PPF continued to climb with the increasing duty cycle, thus rendering a close resemblance with a low pass filter. On the other hand, in case of frequency input study the output spectrum draws more alikeness with a bandpass filter output. The next set of figure schematically shows the expected output spectrum of a high, low, and bandpass filter respectively.

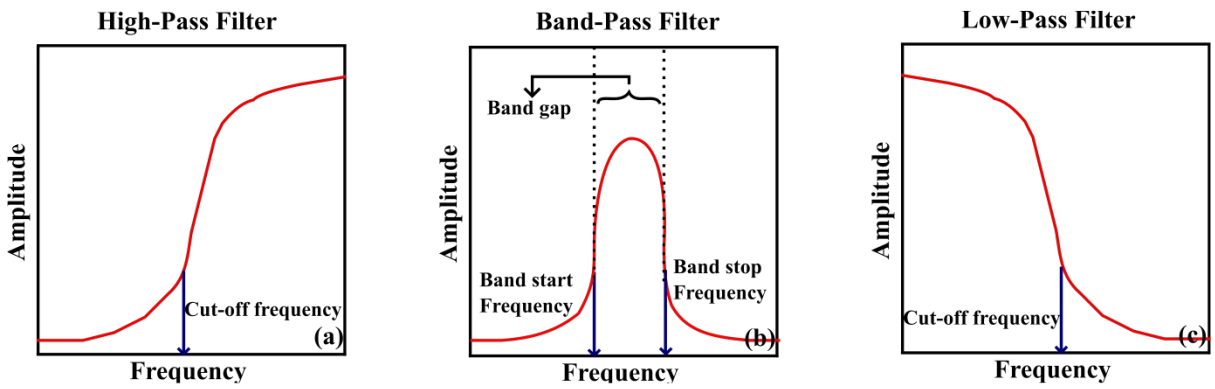


Figure 7.18: a generalized depiction of the output of a (a) high-pass, (b) low-pass, and (c) bandpass filter respectively.

One can easily identify the similarity of these pictures with figure 7.9 (c), 7.10 (c) and 7.14 (d) and verify the above claim. Our interesting finding was the influence of quantum confinement in the output response of our QD based synapses. With a smaller footprint, the quantum confinement became much stronger, causing a slower decay (verified from τ_r measurement) and hence delivering a better PPF.

Referecnces

1. John, V. N. First Draft of a Report on the EDVAC. (1945).

2. Yamazaki, S. History and Future Perspective of Nonvolatile Memory. in *2007 Sixteenth IEEE International Symposium on the Applications of Ferroelectrics* 1–3 (IEEE, 2007). doi:10.1109/ISAF.2007.4393148.

3. Mojumder, N. N., Abraham, D. W., Roy, K. & Worledge, D. C. Magnonic Spin-Transfer Torque MRAM With Low Power, High Speed, and Error-Free Switching. *IEEE Trans. Magn.* **48**, 2016–2024 (2012).

4. Lacaita, A. L. Phase change memories: State-of-the-art, challenges and perspectives. *Solid-State Electron.* **50**, 24–31 (2006).

5. Lee, S., Kim, H., Yun, D.-J., Rhee, S.-W. & Yong, K. Resistive switching characteristics of ZnO thin film grown on stainless steel for flexible nonvolatile memory devices. *Appl. Phys. Lett.* **95**, 262113 (2009).

6. Jiang, L., Lv, F.-C., Yang, R., Hu, D.-C. & Guo, X. Forming-free artificial synapses with Ag point contacts at interface. *J. Materiomics* **5**, 296–302 (2019).

7. Yang, K. *et al.* Tunable flexible artificial synapses: a new path toward a wearable electronic system. *Npj Flex. Electron.* **2**, 20 (2018).

8. Ahmed, T. *et al.* Optically Stimulated Artificial Synapse Based on Layered Black Phosphorus. *Small* **15**, 1900966 (2019).

9. Kim, B.-Y. *et al.* Nanogenerator-induced synaptic plasticity and metaplasticity of bio-realistic artificial synapses. *NPG Asia Mater.* **9**, e381–e381 (2017).

10. Mannan, Z. I. *et al.* Memristive Imitation of Synaptic Transmission and Plasticity. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 3458–3470 (2019).

11. Kim, D. & Lee, J.-S. Liquid-based memory and artificial synapse. *Nanoscale*. **11**, 9726-9732 (2019).

12. Guo, L. Q., Zhu, L. Q., Ding, J. N. & Huang, Y. K. Paired-pulse facilitation achieved in protonic/electronic hybrid indium gallium zinc oxide synaptic transistors. *AIP Adv.* **5**, 087112 (2015).

13. Balakrishna Pillai, P. & De Souza, M. M. Nanoionics-Based Three-Terminal Synaptic Device Using Zinc Oxide. *ACS Appl. Mater. Interfaces* **9**, 1609–1618 (2017).

14. Zidan, M. A. *et al.* A general memristor-based partial differential equation solver. *Nat. Electron.* **1**, 411–420 (2018).

15. Seo, S. *et al.* Artificial optic-neural synapse for colored and color-mixed pattern recognition. *Nat. Commun.* **9**, 5106 (2018).

16. Ha, H. D., Han, D. J., Choi, J. S., Park, M. & Seo, T. S. Dual Role of Blue Luminescent MoS_2 Quantum Dots in Fluorescence Resonance Energy Transfer Phenomenon. *Small* **10**, 3858–3862 (2014).

17. Qiao, W. *et al.* Luminescent monolayer MoS_2 quantum dots produced by multi-exfoliation based on lithium intercalation. *Appl. Surf. Sci.* **359**, 130–136 (2015).

18. Gan, Z. X. *et al.* Quantum confinement effects across two-dimensional planes in MoS_2 quantum dots. *Appl. Phys. Lett.* **106**, 233113 (2015).

19. Mueller, T. & Malic, E. Exciton physics and device application of two-dimensional transition metal dichalcogenide semiconductors. *Npj 2D Mater. Appl.* **2**, 29 (2018).

Chapter 8

An alternative to a selector: Gate dependent switching

8.1 Introduction: Influence of sneak-in-leakage

Despite all the advantages ReRAM has and the intriguing promises it holds for the future, there are still challenges in designing Resistive switches which are delaying this technology entering the market. Limiting the set and reset voltage distribution to a minimum level is a major concern for this technology. Previously published works have shown how drastic the switching behavior can be, depending upon the dielectric, the deposition technique, the doping profile, etc. One interesting finding was reported by Ielmni[1], which showcased the variation of the set/reset voltage with time; the detrimental effect of the metal ion diffusion resulting in this inevitable consequence. In addition to that, not only the switching voltages but also the dependence of the off-current and on-off ratio on several such factors adds another degree of unpredictability in this technology. As we discussed in chapter 6, we observed the change in switching voltage and their distribution as a function of dielectric thickness and how they can affect the reliability of such a device. In chapter 3 we have witnessed how the size of an electrode, the grain, and grain boundary distribution under it can heavily influence the set/reset voltages of a polycrystalline layer; the most pivotal finding from the work was the larger switching voltage distribution with smaller electrode size. This brings us to call for a solution or a counter logic that can minimize this unpredictability. Moreover, this kind of unpredictability (especially on-off ratios and off-currents) can give rise to sneak-in-leakage; which consequently affects the overall reliability of an on-chip ReRAM matrix. We know that unprecedented miniaturization is a salient feature of this technology and exciting possibilities have been reported numerous times. On the other hand, both vertical scaling (chapter 6) and horizontal scaling (chapter 3) were observed to introduce unpredictability in switching behavior. In light of this situation, we reckoned that this is the right time to address this problem and venture for a solution.

To showcase the unpredictability and random behavior resulted from a selector-less switching matrix, we fabricated Al_2O_3 based crossbar[2,3] devices (16×16) and compare the results with its individual counterpart. An FTO coated glass was chosen as the bottom electrode (BE). The FTO substrate was patterned into 16 parallel lines followed by an ALD deposition of 15nm thick Al_2O_3 layer. The top electrode (TE) was deposited through an

appropriate shadow mask in form of 16 parallel lines perpendicular to the BEs. Perhaps the next diagram will walk us through the complete fabrication process.

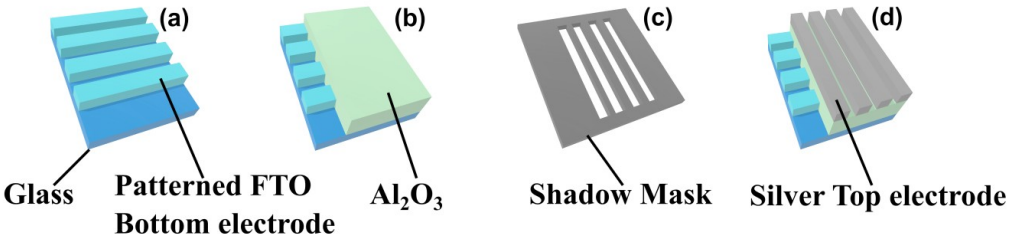


Figure 8.1: The fabrication process was started with (a) a patterned conducting FTO on glass base as the bottom electrode and then (b) the oxide layer was deposited by thermal ALD. The (d) silver top electrodes were deposited by using a (c) shadow mask in a thermal evaporator.

Please note that to simulate a real-life scenario we left a few random switches in on-state in a crossbar and carried out our investigation.

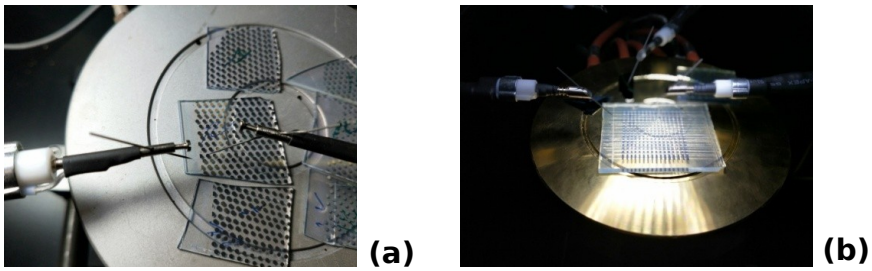


Figure 8.2: (a) Individual ReRAM devices and (b) the crossbar counterpart

Almost all the devices under inspection in individual device structures behaved identically (in terms of switching voltages and on-off ratio). However, devices from the crossbar structure demonstrated different switching behaviors and resistance dynamics at HRS and LRS. Figure 8.3 perfectly illustrates a switching dynamics of 30 consecutive cycles of an individual cell.

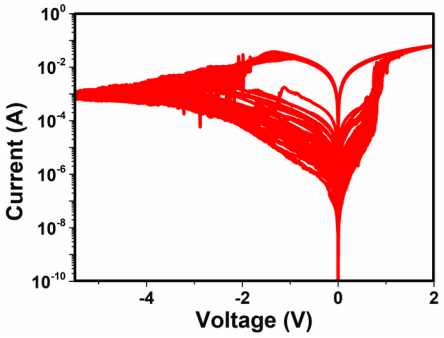


Figure 8.3: Switching characteristics of an individual cell.

The following examples are three different switching cells from the same crossbar structure portraying different switching cycles and on-off distributions.

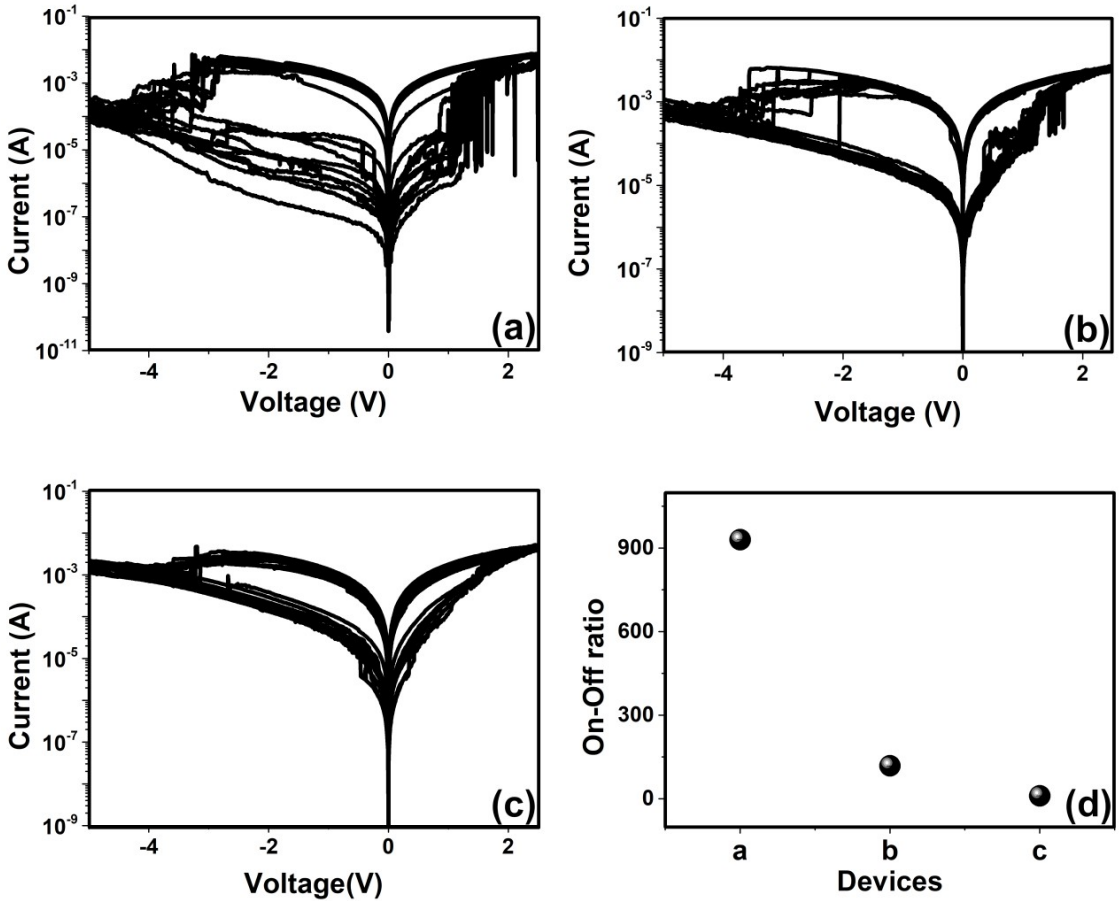


Figure 8.4: (a) 10 consecutive switching cycles for (a), (b), (c) three different devices from a crossbar device, and (d) the measured mean on-off ratio.

The contrast in switching behavior between the individual devices and the crossbar devices is quite pronounced in figure 8.3 and 8.4. Moreover, the individual devices show a consistent on-off ratio of the order of 10^3 for all the devices, on the contrary, this ratio varies throughout the matrix in a crossbar, demonstrating the effect of sneak-in-current[4]. By observing figure 8.4(a) and (c) one can easily identify the difference in switching dynamics as well. A proper ReRAM chip with these limitations will certainly be prone to suffer from varying noise margin (especially the one with low on-off ratio), high sneak-in-current mediated false reading, etc. Another trivial difference between these pictures is the spread of set and reset voltages among them; apart from the afore mention drawbacks, if a crossbar structure also

induces larger spread in switching voltages, then it will cost us the reliability of a device as well.

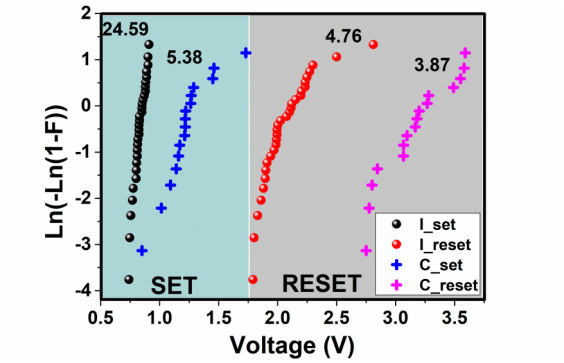


Figure 8.5: The Weibull slope (written at the top) from the set and reset distributions of Individual devices ("I_set" and "I_reset") clearly surpass the slopes from crossbar devices ("C_set" and "C_reset").

The "I" stands for individual device and the "C" stands for crossbar device in the above figure. The smaller Weibull slopes in crossbar devices show the detrimental effect a crossbar can have on the reliability of such technology if not monitored by selector devices. Rather than going with a conventional selector device, we theorized an inbuilt selector option will be much more convenient for this technology. The easiest way to achieve such an alternative is if we can figure out to squeeze a third terminal, which can provide us the opportunity to electrically control the switching dynamics of a ReRAM switch. A thin film transistor (TFT) structure was the closest structure that could hypothetically promise us a switching medium in form of a channel and a third gate terminal to achieve extra control.

In this chapter, we will introduce a reliable technique to control the basic operational parameters of ReRAM devices using an external gate voltage, which gives an additional degree of freedom in the device performance. The basic mechanism behind the set-reset process being the alignment of charges (ions and vacancies) and an external gate voltage can influence the onset of these processes. The advantage of this technique is that regardless of the material/electrode combinations, the device parameters can be externally controlled in a ReRAM device with thin film transistor (TFT) configuration. In addition, the standard CMOS process technology can be employed to fabricate the device arrays in an integrated chip. This study is a proof-of-concept of the gate-controllability of ReRAM set/reset parameters. To demonstrate this, we employed partially reduced graphene oxide (rGO) as the active layer in a thin film transistor configuration, due to its ability to switch at lower electric fields

compared to the conventional highly resistive oxides. Here, we show that the set voltage can be altered very controllably in ReRAM devices. Further, this letter elaborates on the possible mechanisms by which the gate voltages influence the set voltage of the devices.

8.2 Literature Survey

8.2.1 Thin film transistor (TFT)

A conventional MOSFET has atop gate configuration with the substrate serving as the induced channel site and the source and drain (after doping) as well. They generally operate in the inversion region. A TFT[5] however, comprises of metallic source and drain separated by the active channel; layer with a top or bottom gate (in which case the substrate acts as a gate). TFTs mostly operate in the accumulation region. Please see figure 8.8(a) for further clarification.

8.2.2 Graphene vs. Graphene oxide (GO) vs. reduced Graphene oxide (rGO)

Graphene[6] is a well known 2D material with carbon atoms connected to each other by sp^2 bonds to form the familiar hexagonal structure. Graphene can be oxidized to connect several functional groups with the carbon atoms to form a non-conducting ecosystem of sp^3 hybridized carbon atoms. A partial reduction on GO[7] results in the formation of rGO[8], which is a collection of randomly distributed Graphene and GO islands, resulting in an alternating conducting and non-conducting (semi-conducting) layer (see figure 8.12). Grapheme being a conducting layer has more applications in TFT and electrode applications. Go on the other hand acts as a very good insulator and hence, is a favorite in switching devices. However, rGO is semiconducting in nature and can be easily tuned during the reduction process (Modified Hummer's method[9]); therefore, it can be engineered as the application demands.

8.3 Experiment and sample preparation

Our devices were fabricated by spin-coating rGO powder dispersed in chloroform (1 mg in 10 ml) on a p-type Si substrate (resistivity 1–10 Ω cm) with 60 nm thick SiO_2 layer on it. Silver electrodes (200 nm thick) were thermally evaporated through a shadow mask to form TFTs of channel lengths varying from 50 μm to 150 μm. Graphene oxide was derived from graphite powder by standard Hummers' method followed by reduction by hydrazine[10] to get rGO. Raman spectroscopy (RENISHAW InVia Raman Microscope) was employed to study the percentage of GO reduction, and Scanning Tunnelling Microscopy (Quazar Tech. New Delhi, India) was used to map the local tunnel current of rGO. The current-voltage (I–

V) characteristics were measured using a Cascade Microtech four-probe station connected to Agilent B1500A parametric analyzer, and the endurance data were measured with Agilent B2912A source measurement units.

8.4 Results and discussions

8.4.1 STM imaging and LDOS

The Raman spectrum (figure 8.6(a)) of GO shows the typical D (at ~1351 cm^{-1}) and G (at ~1584 cm^{-1}) bands with a 2D peak at ~2708 cm^{-1} .

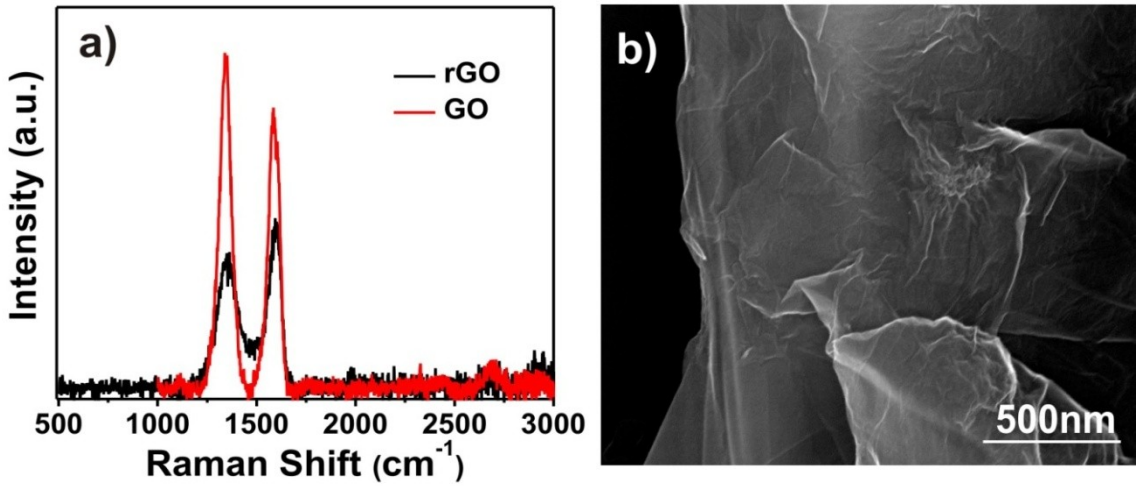


Figure 8.6: (a) Raman spectra of an ensemble of GO (red) and RGO (black) samples. The decrease in the intensity ratio of D and G bands (ID/IG) signifies the proportionate decrease of defects and the increase of sp2 hybridization states. (b) Secondary electron image of the rGO layer showing a continuous single-layer sample.

For GO, the intensity of the D-band is higher compared to the G band, with ID/IG to 1.2. Upon reduction with hydrazene, the intensity ratio reduced to 0.81, indicating that the defect density in the system has been reduced due to partial restoration of sp^2 hybridized graphene in the sample[11–13]. The disappearance of the 2D peak after reduction indicates that agglomeration of the graphene layers might have happened after reducing GO. Figure 8.6(b) shows the secondary electron image of the rGO samples, showing a single-layer graphene. However, agglomeration of the layers occurs in the solution, resulting in multiple stacks of rGO in the powdered sample.

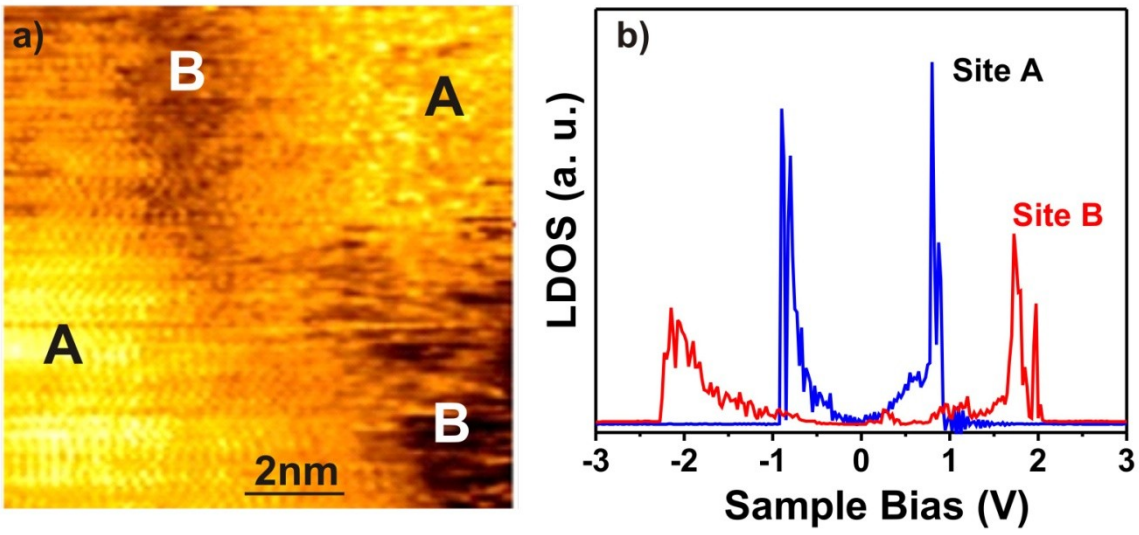
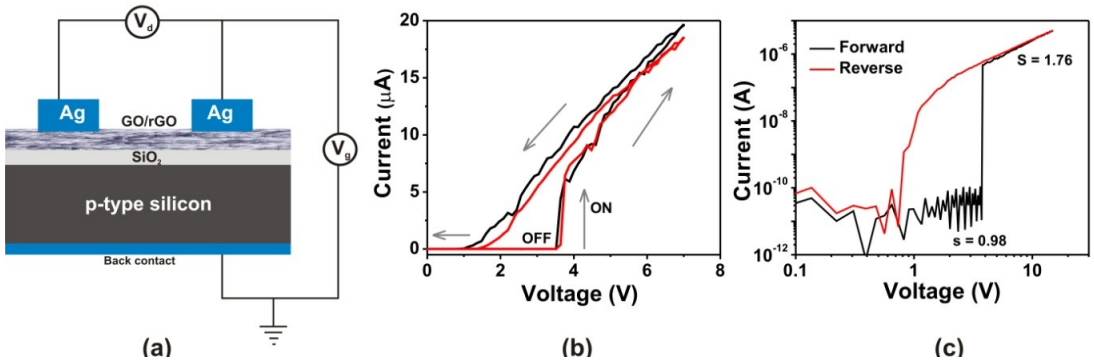


Figure 8.7: (a) STM image of rGO film, where the regions with atomic resolution show the graphene (site A), and the vague area is where the epoxide is present (site B). (b) The local density of states measured with STM at sites A and B, showing the opening of a bandgap at site B indicating the presence of unreduced GO.

The STM image of the rGO (measured at 0.5 V and 1 nA) is shown in figure 8.7(a), which clearly demonstrates regions with and without atomic features (named as sites A and B, respectively). The regions with atomic features are pristine graphene, and the vague area is where epoxy groups (C-O-C, -OH, etc.) are present[14]. The tunnel currents measured at sites A and B are drastically different (not shown here), specifying that site B is less conducting than site A. This indirectly points out the presence of an oxide with a nonzero bandgap. To confirm this, the local densities of states (LDOS) were mapped by measuring dI/dV curve of the sample at the sites A and B. When the tip stands sufficiently far from the substrate, the dI/dV measured will be proportional to the local density of states (LDOS) of the sample[15]. Thus, figure 8.7(b) shows the representative LDOS measured at a frequency of 2.5 kHz. The local band structures evidently differ at locations A and B; there is no bandgap at sites A as expected for pristine graphene and a bandgap of 2.5 eV is evident at site B, confirming that the vague regions are indeed unreduced GO.



8.4.2 Switching and set voltage distributionFigure 8.8: (a) The ReRAM device in the transistor configuration, with GO/rGO active layer; (b) Resistive switching of rGO in the active channel, shown for two different devices. (c) The forward and reverse I-V characteristics in log-log scale, showing regimes of different slopes.

Figure 8.8(a) shows the schematic diagram of the rGO transistor used in this work. During the drain voltage (V_d) sweep, rGO exhibited reproducible resistive switching from the intrinsic high resistance state (HRS) to a low resistance state (LRS) as shown in figure 8.8(b). The on/off ratio of the devices reported here is ~10^4 (with the off-current in 100 pA and on current in micro-Ampere). The switching behavior is not strictly bipolar, because the devices did not require a negative voltage to switch it OFF. Once it was ON, sweeping the voltage down to 0 V across the electrodes was enough to turn the device off (back to HRS). This indicates that certain healing mechanism is acting behind the reset process, which will be discussed later in this chapter. Figure 8.8(c) shows the I–V curve in log–log plot to identify the conduction mechanisms.

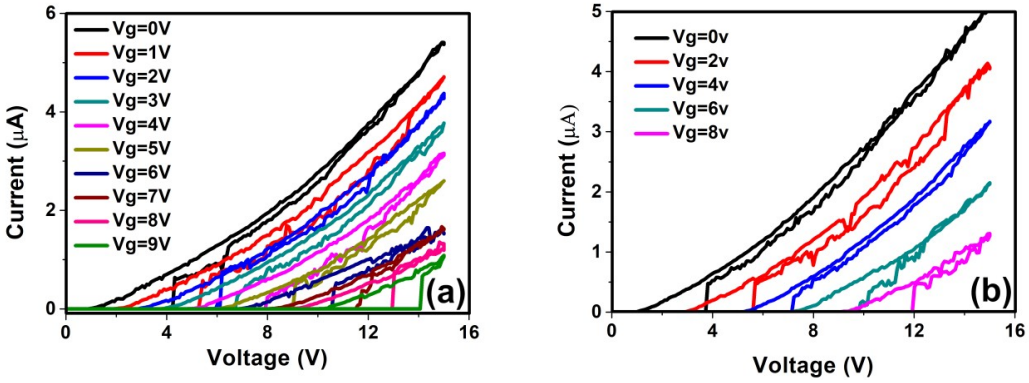


Figure 8.9: (a) Resistive set process along the rGO channel with a gate voltage variation with a step size of 1V and (b) a similar measurement was repeated with a voltage increment step of 2V.

Figure 8.9(a) and (b) shows the resistive switching of rGO under the gate control voltage varying from 0 V to 9 V by a step increment of 1 V and 2 Vs, respectively. Clearly, the set voltage is found to increase with the gate voltage, demonstrating the controllability of the set process with an external bias.

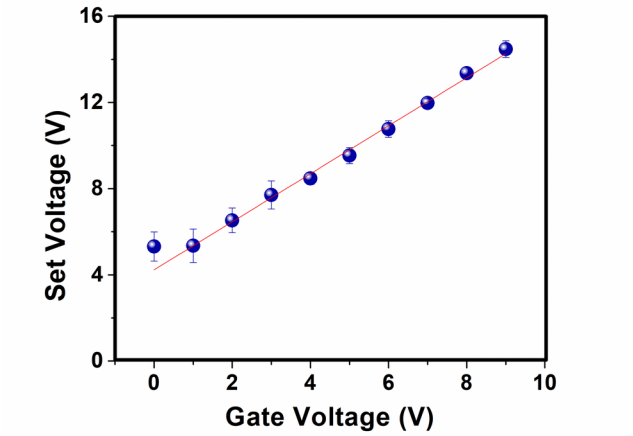


Figure 8.10: The set voltage is plotted as a function of the applied gate bias on the silicon substrate.

Figure 8.10 shows the dependence of the set voltage on the gate bias. In this experiment, the shift in the set voltage was observed only for the positive gate bias, since the formation of a thick depletion region will subdue the gate capacitance in the reverse biasing (negative biasing to the p-type silicon substrate), and therefore, no shift in the set voltage lower to the zero-bias set voltage was observed. This result is very exciting because it not only can evade the time degradation of a device but also can avoid accidental switching. However, often gate leakage current can contribute to these sorts of results and can lead to several misinterpretations. To confirm the legitimacy of these observations the leakage current inspection was mandatory.

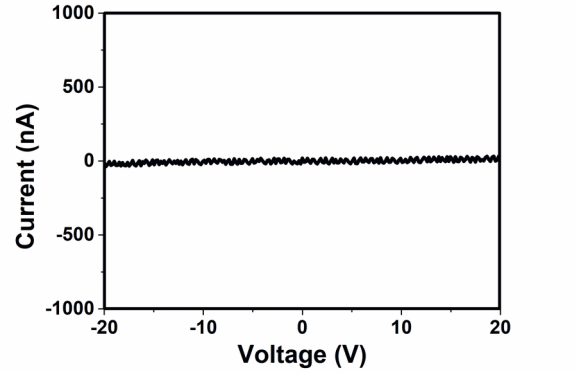


Figure 8.11: The leakage through the gate was more than ~10^3 orders of magnitude lesser than the output current.

From figure 8.11 one can easily affirm the measured output was hardly mediated by the negligible gate current.

The set and reset processes of GO and rGO based memory devices have been widely discussed in the literature, concluding that the underlying mechanism could be either surface oxygen migration[15], absorption/desorption of oxygen-related groups from GO migration of the electrode atoms, or both effects together[16,17]. Among these, the surface oxygen migration or absorption/desorption mechanism of the epoxy groups seems to be a more viable mechanism in the switching observed here. We can justify this with the following arguments: the devices reset naturally by itself when the sweep voltages gradually reduce to zero, indicating that the switching is not due to the migration of the electrode atoms. Interestingly, the log(I)–log(V) plot of the set process shown in figure 8.8(c) shows an Ohmic regime ($I \propto V$) until the breakdown, and no space-charge limited current (SCLC) regime ($I \propto V^2$) is observed in the pre-set biasing. The current remains in few tens of pA until the switching occurs (with a slope 1). The slope of the post-breakdown region in the log–log plot is 1.76, indicating that both Ohmic and SCLC conduction mechanisms are responsible for conduction after switching[15]. The abrupt change from Ohmic conduction to the set process indicates that the set process in this case is due to a spontaneous occurrence of a collective migration or an avalanche effect. If it is a breakdown induced by an electron avalanche, it would not be a reversible process and no reset will be observed. Hence, we can conclude that the set process originates from an electric field induced collective migration or desorption of the epoxy groups from the unreduced regions. The DFT calculations of Li et al. point out that the hopping barrier for an isolated epoxy group is as high as 0.9 eV[18], and the large set voltage we observed must be related to the collective migration of such groups on the graphene surface. The presence of the space-charge limited conduction in the post-breakdown regime shows that the forced migration/desorption of the epoxy groups leaves defects behind, possibly originating from ripping off the sp^3 hybridization of the carbon atoms in the graphene lattice.

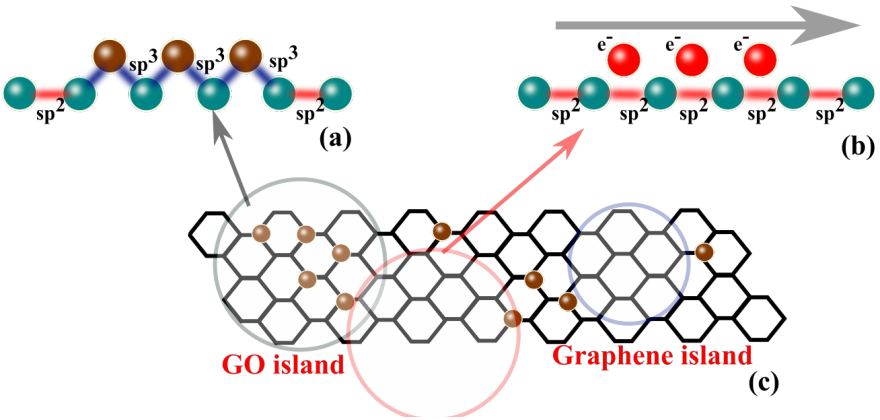


Figure 8.12: A schematic representation of GO and Graphene distribution along an rGO layer and how an electric field influence (a) sp^3 to (b) sp^2 hybridization causes ease of electron transport and hence a conducting path and (c) an rGO layer.

From figure 8.10, the shift of the set voltage on applying a positive gate bias is evident. The epoxy groups are negatively charged and the applied positive bias enhances their hopping barrier. In other words, in the presence of a positive bias from the gate, it takes a larger lateral electric field for the onset of the migration process. The relation between the gate voltage and the switching voltage is nearly exponential with a large exponent; the dotted line in figure 8.10 is a fit with an empirical relation $V_{Set}=V_0 \exp(V_G/\ \Theta)$, where Θ is an exponent with the unit of voltage. Comparing to the lateral electric field (V_d/150μm), the field due to the gate (V_g/60nm) is very large, which explains the large exponent of $\Theta = 14.7 \pm 3.4$V to fit the data. In other words, the migration of the epoxy group is strongly hindered by the gate bias, and the energy required to initiate the collective migration is exponential to the gate voltage.

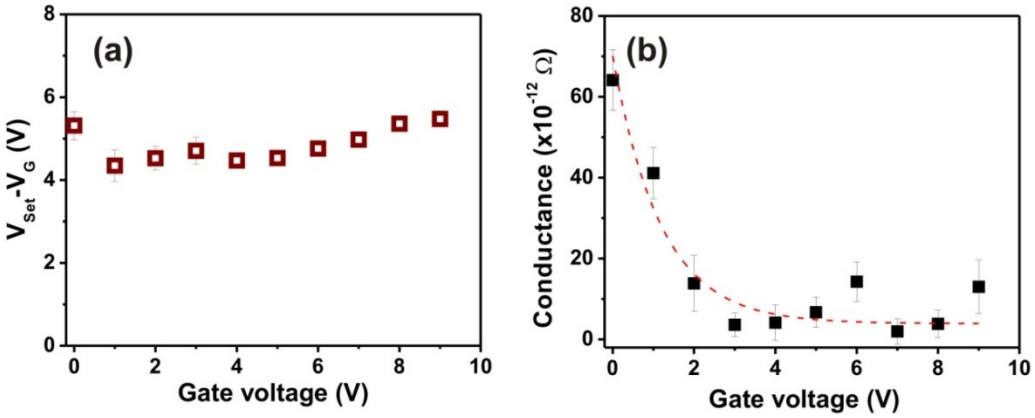


Figure 8.13: (a) The plot between V_{Set}-V_G and the gate voltage V_G. (d) The dependence of off-state conductance on the gate voltage. The dotted line in red is an exponential fit to the data.

Figure 8.12 (a) shows a schematic diagram of the proposed mechanism, where the off-state (HRS) is due to the hindrance of the carriers by the sp^3 hybridized carbon atoms at epoxide sites. As figure 8.12 (b) illustrates, the electric field forces the epoxy groups to migrate to another carbon site. If this migration happens to an energetically unfavorable site, they will migrate back to their original positions when the field is reduced. The plot between V_{Set}-V_G and V_G (assuming voltage linearity with V_G) roughly shows the energy required by the epoxy group to migrate to a neighboring favorable location. Though there is a slight increase in V_{Set}, V_G values, the plot is fairly linear with an average value of 4.84 6 0.41 V. Referring back to the work of Li et al.[18], this would correlate to the energy gain (5.5–4.8 eV) of the epoxy group binding to the hexagonal lattice of graphene resulting in broken epoxy groups. Figure 8.13(b) shows the variation of the Off-state conductance of the devices with increasing gate voltage. The off-state (Ohmic) conductance reduces exponentially with the gate bias, indicating that the hopping barrier for the free electrons exponentially increases with the gate bias. A similar exponential increase in the hopping barrier for the epoxy group can also be anticipated. Also, the current in the reverse sweep does not follow the Ohmic and SCLC behavior, rather demonstrates an abrupt switching down to OFF state near 1 V. These factors point out that a collective electric-field-induced forced migration mechanism of the epoxy groups is the origin of switching in rGO films. The recovery of switching state when the field is reduced in the present experiments shows that the epoxy groups could be migrating to an energetically unfavorable position so that they retrieve their original positions when the field is reduced.

To test the reliability and repeatability, the devices underwent endurance test. Current-voltage characteristics are given in figure 8.14 for two different gate voltages. Endurance of the device was measured between voltage windows of 3 V–9 V for 100 cycles with different gate biases. At both the biases, the device hardly shows any type of degradation, thus confirming its stability.

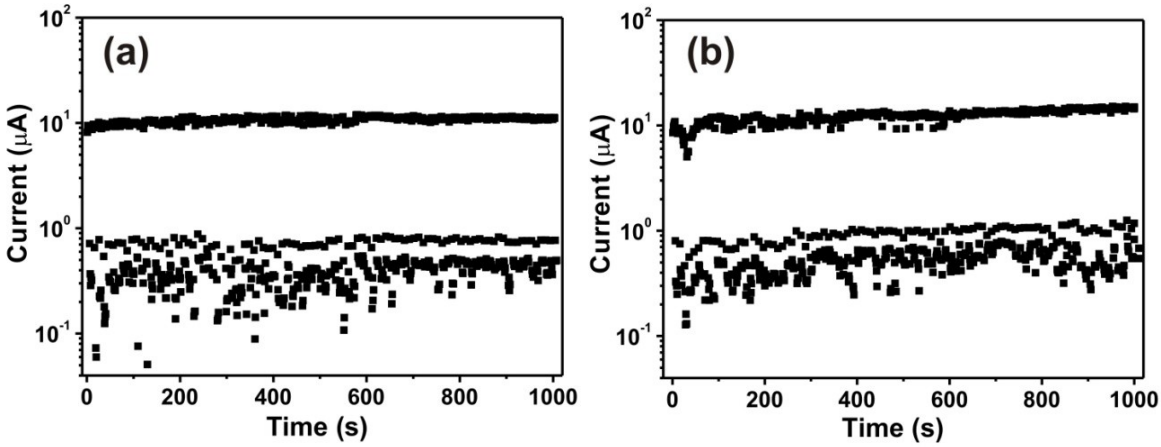


Figure 8.14: Endurance of the device measured at voltage windows of (a) 3V and (b) 9V for 100 cycles with different gate biases (0V and 6V).

8.5 Conclusion

In conclusion, we demonstrate here that using a gate bias, the switching voltage of a ReRAM device can be controlled. In the present scenario, we employed rGO as the active medium for switching, since lateral conduction is easier in this material due to the migration of epoxy groups on graphene. Therefore, a simple TFT configuration was sufficient to demonstrate the gate-controlled switching. The switching voltage appears to vary as a slow exponential of the gate bias with a nearly linear relationship. The influence of the gate bias is modeled on the basis of the increased activation energy of the epoxy groups under the external bias. Our proposed scenario is simple to integrate into the standard CMOS technology, where the TFT and memory element could be integrated together and would help in resolving the challenges related to unifying the operational parameters of devices fabricated with different materials and process technologies. Though the dimensions of the devices reported in this letter (50–150 μm) are considerably large compared to the current CMOS nodes, using the standard CMOS process technology, scaling down the device dimensions to sub100 nm is straightforward.

The two major problems outlined in the introduction section were the set voltage unpredictability and the off leakage's influence on the sneak-in-leakage. A major part of this chapter was dedicated to find a counter solution of the first limitation and from figure 8.9 and 8.10 it can be inferred that we successfully engineered a solution. However, the second problem remained apparently unchecked. Although, a careful deduction of figure 8.13 (b)

should suggest the controllability of the sneak-in-leakage as a function of gate voltage has also been achieved. With an increasing gate bias, the off current seemed to reduce; this gives us a perfect opportunity to manipulate the off-current with a third terminal as well.

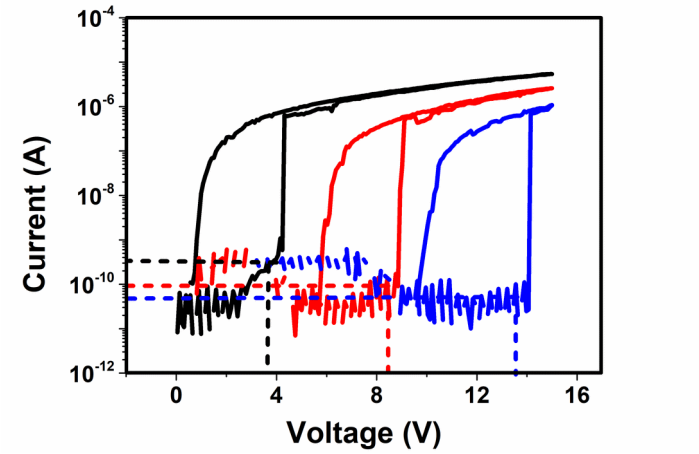


Figure 8.15: The dotted lines for the respective set process show the off-current lowering with higher gate voltage.

This figure shows the decrease in HRS current with an increase in the applied gate bias, as also previously pointed out in figure 8.13(d) in terms of off-conductivity. Therefore, by tuning the gate voltage one can easily keep the sneak-in-leakage under check, especially for the devices which are under V/2 and V/3 bias, in which case the sneak-in-current can increase drastically. The set voltage and off-current tuning option embedded in a single device in form of a third terminal should offer exciting opportunities to this technology (in terms of reliability, protection against accidental switching, and sneak-in-leakage modulation).

References

1. Ielmini, D. Resistive switching memories based on metal oxides: mechanisms, reliability and scaling. *Semicond. Sci. Technol.* **31**, 063002 (2016).

2. Govoreanu, B. *et al.* 10x10nm2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. in *2011 International Electron Devices Meeting* 31.6.1-31.6.4 (IEEE, 2011). doi:10.1109/IEDM.2011.6131652.

3. Deng, Y. *et al.* RRAM Crossbar Array With Cell Selection Device: A Device and Circuit Interaction Study. *IEEE Trans. Electron Devices* **60**, 719–726 (2013).

4. Zhou, J. *et al.* Very Low-Programming-Current RRAM With Self-Rectifying Characteristics. *IEEE Electron Device Lett.* **37**, 404–407 (2016).

5. Carcia, P. F., McLean, R. S., Reilly, M. H. & Nunes, G. Transparent ZnO thin-film transistor fabricated by rf magnetron sputtering. *Appl. Phys. Lett.* **82**, 1117–1119 (2003).

6. Ferrari, A. C. *et al.* Raman Spectrum of Graphene and Graphene Layers. *Phys. Rev. Lett.* **97**, 187401 (2006).

7. Dreyer, D. R., Park, S., Bielawski, C. W. & Ruoff, R. S. The chemistry of graphene oxide. *Chem Soc Rev* **39**, 228–240 (2010).

8. Eda, G., Fanchini, G. & Chhowalla, M. Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material. *Nat. Nanotechnol.* **3**, 270–274 (2008).

9. Zaaba, N. I. *et al.* Synthesis of Graphene Oxide using Modified Hummers Method: Solvent Influence. *Procedia Eng.* **184**, 469–477 (2017).

10. Park, S. *et al.* Hydrazine-reduction of graphite- and graphene oxide. *Carbon* **49**, 3019–3023 (2011).

11. Bracamonte, M. V., Lacconi, G. I., Urreta, S. E. & Foa Torres, L. E. F. On the Nature of Defects in Liquid-Phase Exfoliated Graphene. *J. Phys. Chem. C* **118**, 15455–15459 (2014).

12. Cançado, L. G. *et al.* Quantifying Defects in Graphene via Raman Spectroscopy at Different Excitation Energies. *Nano Lett.* **11**, 3190–3196 (2011).

13. Ferrari, A. C. & Robertson, J. Resonant Raman spectroscopy of disordered, amorphous, and diamondlike carbon. *Phys. Rev. B* **64**, 075414 (2001).

14. Gómez-Navarro, C. *et al.* Electronic Transport Properties of Individual Chemically Reduced Graphene Oxide Sheets. *Nano Lett.* **7**, 3499–3503 (2007).

15. Wang, L.-H. *et al.* The mechanism of the asymmetric SET and RESET speed of graphene oxide based flexible resistive switching memories. *Appl. Phys. Lett.* **100**, 063509 (2012).

16. Ki Hong, S., Eun Kim, J., Kim, S. O. & Jin Cho, B. Analysis on switching mechanism of graphene oxide resistive memory device. *J. Appl. Phys.* **110**, 044506 (2011).

17. He, C. L. *et al.* Nonvolatile resistive switching in graphene oxide thin films. *Appl Phys Lett* 4 (2015).

18. Li, J.-L. *et al.* Oxygen-Driven Unzipping of Graphitic Materials. *Phys. Rev. Lett.* **96**, 176101 (2006).

19. Ielmini, D. Resistive switching memories based on metal oxides: mechanisms, reliability and scaling. *Semicond. Sci. Technol.* **31**, 063002 (2016).

Chapter 9

Conclusion and future works

9.1 Introduction

We have already covered the summary of each and every chapter in their respective "conclusion" sections. In this chapter, we will revisit those summaries one more time, chapter by chapter, and will try to provide an intuition about the work or the set of works that should logically follow our contributions for further insights and development of the technology.

9.2 Revisiting the chapters and future proposals

Chapter 1

Discussion

In this chapter, we have introduced the evolution and the impact of data storage in our daily life. The chapter started with a brief discussion of the early electronic storage solutions and gradually moved through all the iterations before settling down to FLASH storage solutions. A fairly thorough discussion of FLASH based memory, its working scheme, architectural options (NAND and NOR), modern-day problems and attempted solutions covered a significant chunk of this section. This part was really important, because, to understand a next-generation storage solution (like ReRAM), the understanding of drawbacks of the current technology is of utmost importance. In addition to that, different non-volatile memory alternatives such as MRAM, PCRAM, and atomic switch were also given enough limelight. The physics behind their operation and what causes their hindrance to come out as the next memory solution were also covered afterward. After setting up a base, the chapter took a turn to finally introduce Resistive Random Access Memory or ReRAM. The following portions explained the structure of a ReRAM, the switching characteristics or the current-voltage (I-V) characteristics, different types of switching (bipolar and unipolar), and the proposed theory behind them. The subsequent sections were dedicated to illustrate the well-received "crossbar" architecture, and any discussion of crossbar and ReRAM could not be completed without mentioning the inherent limitation of this architecture; the sneak-in-leakage current. To counter the sneak-in-leakage selector devices are introduced in crossbar junctions, hence, a brief discussion of the selectors followed in the next portion. By then it was assumed that we should have had good coherence of the contrasting pictures of FLASH

and ReRAM. In that light, the scope of the thesis was introduced. In this section, we pointed out the limitations of the FLASH technology and how those problems can be solved and in some cases partially evaded by a resistive switching matrix. The last installment of the chapter is dedicated to guide the reader through the thesis by drawing an intelligible logical flow through the different works included in the thesis.

Chapter 2

Discussion

Atomic Layer Deposition system (ALD) is one of those sophisticated film growing technologies which can promise ultrathin film deposition with unparalleled precision; rest aside the purity of the films. We have learned already a major portion of this thesis was based on switching through ALD layers. Thickness scaling, interface trap inspection, mobile ion study, and studying the tunneling current across ultrathin layers would have been impossible without ALD. However, all these advantages of an ALD come with a great cost; any commercial ALD system is far from being cost-effective. As a result, we developed two systems in our lab to meet our requirements. The first system was constructed as a collaborative work with HHV, Bangalore. Even though our part of the design only involved in the chamber configuration and the precursor injection, but that modification indeed added a few advantages in the system, as opposed to a conventional lateral-flow thermal ALD design. A second system was also constructed in the lab with an aim to provide us with much more flexibility in terms of future alterations and experimenting with novel materials. Several depositions of Al_2O_3 layers have proven the worth of the system; the deposited layers' quality, thickness/cycle, electrical characteristics, etc. were at par with the ones from any commercial systems. However, this very system was constructed to serve as a plasma assisted ALD system (PALD). Initial tests have proven a successful generation of capacitive coupled plasma (CCP) in the chamber. Further tests on the quality of CCP assisted layers and future possibilities on inductive coupled plasma (ICP) are expected to follow in the next iterations. Please note that the CCP thus produced was confined between two metallic meshes, namely the anode and cathode respectively. This particular design will let the user to ionize the required gas while it flows through the meshes. Another optional CCP scheme would be to apply the field between a mesh/top-chamber-lid and the substrate holder, but that would hinder the plasma distribution on non conducting substrates; this can also be an issue while growing an insulating layer on a conducting layer.

When it comes to quality and growth rate PALD always has an upper hand over thermal ALD. It is needless to say the next iteration of this current version should be a PALD system. Although, before finalizing the design, we should carefully analyze the pros and cons of CCP, ICP, and ozone-based ALD deposition and only choose the best among them. The defect less switching layers should allow us to study the switching mechanisms in greater detail and help us to carry out much more sophisticated inspection schemes.

Chapter 3

Discussion

Small feature size and drastic scaling is one of the key selling points of ReRAM technology. Polycrystalline material like zinc oxide (ZnO), on a nano-scale, is riddled with numerous grains (G) and grain boundaries (GB). The arbitrary nature of these features in ZnO can heavily affect the switching phenomenon at a smaller scale. We used an STM to individually access Gs and GBs and study their switching behavior. Also from the I-Z spectroscopy, we calculated the work function of a GB-G-GB region and reconstructed the band diagram. The band diagrams were studied under red and green illumination separately. These studies revealed the more conducting GB region is characterized by a higher off-current and lower switching voltages as compared to a G region. From the band diagram, the bending of the conduction band towards the Fermi level in the GB region explained the higher conductivity of such a region. Only the green illumination caused larger band lowering, suggesting the role of oxygen vacancies in promoting the breakdown. We deposited gold electrodes of varying sizes on the oxide and studied the set voltage spread and off-resistance as a function of electrode area. A detailed inspection revealed that for smaller electrodes the unpredictability grows stronger, rendering a less reliable switching system. The off-current study suggested that for larger electrodes more number of GBs are accessed by an electrode hence delivering a much smoother switching behavior. Both the set voltage and off-current exhibited exponentially decaying characteristics with growing electrode sizes.

Even though this experiment gave us a clear direction on what we can anticipate from sub $10nm^2$ electrodes in a polycrystalline resistive switch, this was just a mere real on-chip scenario's simulation. For a better understanding, we need to recreate a similar experiment with proper electrodes of varying sizes fabricated by electron-beam lithography and acquire better statistics. The experiment needs to be repeated for both polycrystalline and amorphous (for e.g. Al_2O_3) material. In theory, the amorphous material should exhibit less reliability

dependency on the electrode size. However, set voltage still can vary as a function of electrode size, because the number of defects participating in dielectric breakdown depends upon the number of defects being affected by the electric field; larger the electrode area, larger is the number of defects under electric field influence. Apart from this, inter-electrode spacing for smaller devices with ultrathin dielectrics is an interesting area which deserves an ample amount of attention. This kind of investigation will identify the lowest limit of inter-electrode spacing to avoid crosstalk effect due to fringing fields; subsequently improving the feature size.

Chapter 4

Discussion

We cannot stress it enough how important it is to develop silicon-based ReRAM technology for a flawless synchronization with the current silicon-based semiconductor industry standards. In such cases, the broken and dangling bond based deformities on the silicon can influence the set process significantly. In the 4th chapter, we studied the effect of these interface traps in silicon-oxide (Al_2O_3) junction. The well-known conduction versus frequency method was employed to compute the interface defect density (D_{it}) in such defect ridden areas. Silicon <100> and silicon <111> wafers were used as bottom electrodes to simulate two different densities of traps (<111> surface has more defects). The I-V cycles clearly showed a significantly lower breakdown voltage in <111> samples. The computed D_{it} from the G-f peaks showed the presence of interfacial defects at least one order of magnitude in <111> samples. We performed a controlled breakdown in all the samples by varying the compliance current; post breakdown analysis also suggests a noticeable increase in D_{it} after the breakdown and resembling an incremental dependence on the compliance limit. More defects suggest more trapping/detrapping time period; this phenomenon was reflected in the trap time constant (τ_{it}) measurement. Further, the mirror charge induced by the defects mediated barrier lowering was calculated from the Mott-Schottky plot; to no surprise, the lowering was more pronounced in <111> samples and seemed to grow stronger with compliance current. Needless to say, this aided Fowler-Nordheim tunneling causing higher leakage current; which was clearly reflected as stronger SCLC regime in log (I) vs. log (V) plot. Finally, a breakdown analysis of different oxide thickness was carried out for different orientations of silicon. The <111> samples continued to show lower breakdown voltages for all the thicknesses with no exception. However, sub 5nm oxides showed a drastic decrease in

the breakdown voltage; thus implying the overwhelming effect of interface defects in lower thickness can heavily hinder further prospects of thickness scaling.

The outcomes from this chapter no doubt gave us a proper insight into the role of oxide-silicon defects in resistive switching. But to properly implement this knowledge in a MOS resistive switching technology we need to develop the system and carry out the aforestated experiment for different interfacial junctions (by changing the oxide) and for p+, p++, n+, and n++ BEs to reproduce a repetitive and full-cycle resistive switching. The detrimental effect of these traps can also be controlled by passivating the defects by annealing then in forming gas; this might allow us to vertically scale down the device without the devastating effect of interfacial traps.

Chapter 5

Discussion

Mobile ions can take a considerable part in the switching process of a resistive RAM. Noble metals like gold, platinum TE based switches mainly rely upon the vacancy/defect generated switching, while, low electronegative material such as silver, copper, etc. can be easily ionized by applying an electric field, causing them to diffuse inside the oxide to form the conducting filament. The presence of similar ions can be easily recognized by the shift in flatband voltage (in other words the capacitance hysteresis) in a MOS capacitor. 5nm oxide-based MOS switches showed an increase of flatband shift with an increase in compliance limit (recall the compliance current's role in the breakdown in the previous chapter); the shift was much larger for silver TEs as compared to the Au TEs, implying diffusion of Ag ions. Bias temperature stress (BTS) study is one of the most effective ways to determine the presence of mobile ions in a MOS system. Flatband inspection pre and post-temperature-stress clearly signifies the abundance of mobile ions in an Ag/oxide/silicon system with respect to its counterpart. Similar experiments were carried out for higher thicknesses as well. The higher breakdown voltage and low leakage current gave us an upper hand; we were able to carry out the experiment at different stress times and different compliance currents at the same time. For both 10nm and 20nm thick oxides, both the compliance current and stress time seemed to inflict more and more silver ions in the dielectric, causing a larger flatband shift.

The immediate step would be to study the effect of vacancy modulated vs. ion diffused switching for different metal electrodes. Studying the reliability and cycling controls

for different metals along with different dielectrics should yield the best combination of metal-dielectric-metal for a resistive switch. Based upon the activation energy and the diffusion coefficient different dielectrics can switch at different speeds. From such a study an optimal combination should arise which delivers an ultrafast switching RAM.

Chapter 6

Discussions

The aim of this chapter was to determine the thinnest possible oxide which can switch reliably without collapsing a good on-off ratio. Throughout this chapter, we have witnessed an Au TE device is hardly able to switch repeatedly without a compliance check during the set process. However, Ag as the primary electrode brings forth very consistent and repetitive switching, as a result, we mostly restricted our studies with the Ag TE devices. A set of varying thicknesses of oxide were fabricated and the I-V characteristics were primarily studied in this chapter. The effect of FN tunneling, PF injection, and direct tunneling became stronger as the thickness decreased; rendering a much leakage ridden switch in sub 3nm switches. Although, a tint of trap assisted tunneling was still observed in a 1nm switch, suggesting the entire leakage is not solely direct tunneling driven. From the Weibull distribution, we learned that with thinner the oxide better is the reliability of the system. It should be noted that Au/oxide/FTO devices hardly showcased multiple switching for all the thickness; only three thicknesses, namely 15nm, 10nm, and 7nm demonstrated several cycles; although, from the Weibull slope analysis they never stood a chance before AG/oxide /FTO devices. A train of set-read-reset-read was fed into the circuitry while monitoring the current output. Devices with Ag as force electrode reliably toggled between the resistive states without ant fail; on the other hand, Au TE devices did not respond to the voltage stimuli and failed within a few cycles. With no role in electrode's ion migration, the Au TE devices could switch on either polarity of the electric field, whereas Ag TE devices could exhibit set process on at positive polarity. From the results of this chapter, there is no question that an ion migration based switching is much more reliable and superior than that of the vacancy influenced ones.

Despite the number of studies carried out in this chapter, the influence of the fringing electric field at lower thickness still remained unscathed. An ample amount of theoretical as well as experimental study is required to uncover the advantage/disadvantage these fringing fields

can have on this technology. Tunneling phenomena like TAT, Direct tunneling, etc deserves much detailed analysis, if possible backed by theoretical results.

Chapter 7

Discussion

Traditional von Neumann architecture suffers from a bottleneck. The speed discrepancy between the memory and processor hinders the overall performance. Recreating a device which can mimic a biological brain seems like an adequate solution for such a problem, where the processor and memory are indistinguishable and works in harmony to bring the best of both. This chapter demonstrated the performance of an optical neuron of possibly the smallest footprint since it is practically a single MoS_2 quantum dot (QD). As a function of duty cycle, the sensitivity of the QDs grew stronger with a successive optical stimulus (measured in terms of PPF). Judging for the bandgap of a QD we speculated and later proved that a UV stimulus should be able to generate enough excitons to deliver a perceivable PPF. By meddling with the frequency, we were able to twitch the system in such a way so that its output can represent a high-pass filter response (fixed on and varying off time of the input train), a low-pass filter (fixed off and varying on time) and a band-pass filter (equal on and off time). But the most important outcome from this experiment was the PPF modulation by varying the size of a QD. A smaller QD ensures stronger confinement thus increasing the decay time constant. We carried our experiment with three different sizes of QDs, among them the strongest PPF was delivered by the smallest QD whereas the least responsive one was the largest of them all. Throughout this chapter, we have observed the importance of the generation and recombination time constants of the output spikes. Often faster generation time and slower recombination time rendered better PPFs. In case of 50% duty cycle study, we, to our great interest learned that during the 1s and 2s time period the decay time was slow enough and generation time was fast enough to create an optimal spot where the PPF response was at its peak.

The optical stimuli time period in our experiment was in the order of seconds, the best performance with the 50% duty cycle was observed with 2s and 1s time periods. The aim of this project was to find an alternative to replace the intelligent/semi-intelligent optical sensor for future AI applications, however, the aforementioned time periods are way off the chart and only good for laboratory demonstrations. A huge amount of work is pending which would develop a faster optical neuron, hopefully, responsive to different wavelengths with

different PPFs and definitely in form of a device rather than just a laboratory-test bench. For better industrial acceptance, silicon-based devices will be much suitable for such studies. Another important parameter designer should take into account in the PPF itself. Better the PPF, better will be the chance to distinguish (by the reading circuitry) among the offered states.

Chapter 8

Discussion

We have already covered the theoretical aspects, the practical application, and the importance of selector devices in the introduction chapter. When we compared the switching characteristics of individual cells and crossbar cells (in this chapter) we first time experienced the catastrophe caused in a crossbar chip in the absence of a selector device. We compared a 15nm thick oxide layer with FTO as the BE and Silver as the TE devices in both individual and crossbar format. The leakage current due to the sneak-in-current was overwhelming throughout the matrix. Different devices showcased different on-off ratios and generally, they were lower than the individual cells; which can potentially decrease the noise margin in a chip. Moreover, the set and reset processes were also all over the place, and from the Weibull distribution revealed a decrement in reliability standards of crossbar cells as well.

We fused the selector and ReRAM device together in a TFT (thin film transistor) format with the rGO channel as the switching medium and the gate being the additional control. A positive potential on the gate made the rGO layer more prone to exhibit stronger sp^3 hybridization with the functional groups, thus making it more difficult to experience a breakdown. As a result, we recorded a slightly exponential increase of the set voltage as a function of gate voltage. In addition to that, the dependency of off-current with the gate has been consistent; a welcome solution to tackle the sneak-in-leakage current. The set voltage modulation along with the off-current modulation could be foreseen as a promising replacement of the conventional selector-ReRAM combination.

We have demonstrated a horizontal switch with a control terminal. However, this configuration will still cost us the feature size. If this sort of technology is at all adopted by the industry, we need to find a way to incorporate the gate terminal in a vertical capacitor like structure. The fabrication steps might be more complex and cumbersome, but we reckon that will no longer compromise the feature size of a memory cell and thus can be made more

scalable friendly. Finally, these results need to be reproduced and studied immensely for conventional switching dielectrics (Al_2O_3, ZnO, HfO_2, etc.) on much smaller footprints.

9.3 Final thought

With the plasma ALD coming up, one should try different other materials in PALD for switching and neuromorphic applications. We have only demonstrated the optical synaptic function in this thesis; however, for proper computational purposes, electronic neurons are yet to be developed. Needless to say, different combinations of TE-dielectric-BE are needed to be analyzed until we arrive at an optimal combination. Multilayer switching and multi-level switching are relatively unscathed in the community. Developing such a system will allow us to write and read multiple bits from a single device; this will result in an increase in reading and writing times. Different doping in ALD layers and their influences deserves immediate attention; by infusing light metal ions in dielectric can lead us to switch the devices at a higher speed. In conclusion, the possibilities are endless; one needs to be really persistent to make sure these technologies guide the semiconductor industry for better.

List of Publications

Journal Publications

1 **"Gate controllable resistive random access memory devices using reduced grapheme oxide"**
Hazra, P., Resmi, A. N. & Jinesh, K. B. Gate controllable resistive random access memory devices using reduced graphene oxide. *Appl. Phys. Lett.* **108**, 153503 (2016).

2 **"Scaling of resistive random access memory devices beyond 100 nm2 : influence of grain boundaries studied using scanning tunnelling microscopy"**
Hazra, P. & Jinesh, K. B. Scaling of resistive random access memory devices beyond 100 nm ^{2}: influence of grain boundaries studied using scanning tunnelling microscopy. *Nanotechnology* **29**, 495202 (2018).

3 **"Hybrid Perovskite Nanoparticles for High Performance Resistive Random Access Memory Devices: Control of Operational Parameters through Chloride Doping"**
Muthu, C. *et al.* Hybrid Perovskite Nanoparticles for High-Performance Resistive Random Access Memory Devices: Control of Operational Parameters through Chloride Doping. *Adv. Mater. Interfaces* **3**, 1600092 (2016).

4 **"Vertical limits of resistive memory scaling: The detrimental influence of interface states"**
P. Hazra and K.B. Jinesh. Vertical limits of resistive memory scaling: The detrimental influence of interface states. *Appl. Phys. Lett.* 116, 173502 (2020).

5 **"A comparison of mobile ion and oxide trap charge based switching mechanisms in ReRAMs"**
P. Hazra and K.B. Jinesh ***(under preparation)***

6 **"Thickness scaling of ALD deposited Al_2O_3 based ReRAMs"**
P. Hazra and K.B. Jinesh ***(under preparation)***

7 **Single MoS_2 quantum dot based artificial optical synapse**
P. Hazra, A. Thomas and K.B. Jinesh ***(under preparation)***

Conferences

8 **"Gate controllable resistive random access memory devices using reduced graphene oxide"**
ICAMPE, M.G. University, Kottyam (2015)—(Oral Presentation)

9 **"Grain boundary conduction in sub-100 nm^2 Resistive Memory devices"**
ICONMAT, CUSAT, Cochin (2019) —(Poster Presentation)