Emerging Memory Technologies as the way to better Computing

Subhendu Mondal
Department of Electronics and Communication Engineering,
Techno India University, EM-4, Salt Lake City, Sector V, Kolkata, West Bengal 700091

ABSTRACT

This article will introduce the reader to the emerging non-volatile memory (NVM) technologies, such as MRAM, PCRAM. NVM technologies combine the density of DRAM,the speed of SRAM,and the non-volatility of Flash memory, these technologies are very attractive to future generation universal memories. Emerging NVM cell characteristics are summarized in this article. The ideal characteristics for a memory device include fast write/read speed (<ns), low operation voltage(<1 V), low energy consumption (~fJ/b for write/read), long data retention time (>10 years), long write/read cycling endurance (>1017 cycles), and excellent scalability (<10 nm)[1]. Nevertheless, it is almost impossible to satisfy all of these ideal characteristics in a single "universal" memory device. Several resistance-based emerging NVM technologies have been pursued toward achieving part of these ideal characteristics. The emerging NVM candidates include STT-MRAM [2],PCRAM [3], RRAM [4], N-RAM, SONOS and FRAM.

Keywords: NVM, CB RAM, SONOS, STTMRAM, RRAM, N-RAM, PCM, PRAM, FRAM, MRAM, Van der Waals.

1. Introduction

This article introduces the basics of emerging non-volatile memory (NVM) technologies including conductive bridging RAM (CBRAM), Silicon-Oxide-Nitride- Oxide Silicon SONOS), spin-transfer-torque magnetic random-access memory (STTMRAM), phase-change random access memory (PCRAM), Nano-RAM, Ferroelectric RAM, Magnetic random-access memory (MRAM) and resistive random-access memory (RRAM). The functionality and performance of today's computing system are increasingly dependent on the characteristics of the memory subsystem. The memory subsystem has a well-known memory hierarchy: Today static random-access memory (SRAM), dynamic random-access memory (DRAM), and flash are the mainstream memory technologies serving as cache, main memory, and storage memory. Though, these emerging NVM technologies face challenges from aspects of process compatibility, manufacturing yield, performance variability, and reliability, these different emerging NVM devices have different application spaces in the memory hierarchy due to their unique characteristics. Beyond the conventional memory applications, aerospace electronics applications, embedded applications are also using emerging NVM.

2. Different NVMs in today's world

These emerging NVM technologies have some common features: these are non-volatile two-terminal devices, and the different states are obtained by the switching between a high resistance state (HRS) or "off state" and a low resistance state (LRS) or "on state". The switching from "off state" to "on state" is called "set," and the switching from "on state" to "off state" is called "reset." The transition between the two states can be triggered by an electric voltage or current pulse change. However, the detailed switching physics is different for different NVMs. As for example, STT-MRAM relies on the parallel configuration and antiparallel configuration of two ferromagnetic layers separated by a thin tunnelling insulator layer. The parallel configuration corresponds to LRS and anti-parallel configuration corresponds to HRS. PCRAM relies on chalcogenide materials to switch between the crystalline phaseand the amorphous phase. The crystalline phase corresponds to LRS and the amorphous phase corresponds to HRS, where RRAM relies on the formation and the rupture of conductive filaments in the insulator between two electrodes.

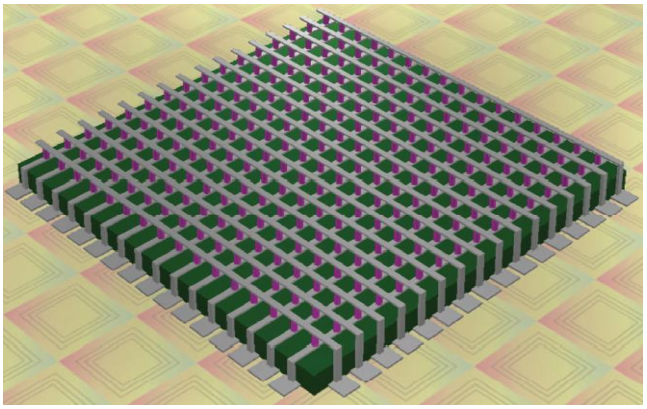

Fig1: A sample pictorial view of a NVM.

As compared to SRAM, STTMRAM has an advantage of a smaller cell area, while STT-MRAM has maintained low programming voltage, fast write/read speed, and long endurance. Thus, STT-MRAM is attractive as a replacement for embedded memories (e.g., SRAM or embedded DRAM) in the last-level cache [5]. As compared to flash, PCRAM/RRAM is attractive due to its lower programming voltage and faster write/read speed. Thus, the PCRAM/RRAM is attractive as a replacement for NOR flash for code storage and, more ambitiously, to replace NAND flash for data storage [6].

Conductive Bridging RAM (CBRAM)

The programmable metallization cell, or PMC is a non-volatile computer memory developed at Arizona State University, Infineon Technologies refers to it as conductive bridging RAM or CBRAM, developed to replace the widely used Flash Memory.

Its characteristics are:

1. PMC/CBRAM is a two terminal resistive memory technology. It is an electrochemical metallization memory that relies on redox reactions to form and dissolve a conductive filament.
2. The states of the device are determined by the resistance across the two terminals. The existence of a filament between the terminals produces low resistance state (LRS) while the absence of a filament results in a high resistance state (HRS).
3. A PMC device is made of two solid electrodes, one relatively inert(eg., tungsten or nickel), the other electronically active (e.g..,silver or copper) with thin film of solid electrolyte between them.

Now let us have a brief overview of the different NVMs introduced above.

Device Operation:

The resistance state of PMC is controlled by the formation (Programming) and dissolution (enraging) of a metallic conductive filament between the two terminals of the cell.

(a) Filament formation: PMC rely on the formation of a metallic conductive filament to transition to a low resistance state (LRS). The filament is created by applying a +ve voltage bias(V) to the anode contact (active metal) while grounding the cathode contact (inert metal). The +ve bias oxidizes the active metal (M):

M --> M+ + e-

The applied bias generates an electric field between the metal contacts. The ionised (oxidized) metal ions migrate along the electric field towards the cathode contact. At Cathode contact, the metal ions are reduced:

M+ + e- --> M

As the active metal deposits on the cathode, the electric field increases between the anode and deposit (E= -V/d). The filament will grow to connect to anode within a few nanoseconds. Once the voltage is removed, the conductive filament will remain, leaving the device in LRS.

Silicon-Oxide-Nitride-Oxide-Silicon (SONOS)

It is a type of non-volatile computer memory (NVM) closely related to Flash RAM. It is one of charge trap flash variant. Where the mainstream flash memory uses polysilicon for charge storage material, SONOS uses of Silicon nitride (Si3N4) for the charge storage material.

A SONOS memory cell is formed from a standard polysilicon N-channel MOSFET transistor with the addition of a small silicon nitride layer inside the transistor gate oxide. The oxide/nitride sandwich typically consists of a 2 nm thick oxide lower layer, a 5 nm thick silicon nitride middle layer and 5-10 nm oxide upper layer.

Table 1: Device characteristics of mainstream and emerging memory technology [7]

TABLE 1. DEVICE CHARACTERISTICS OF MAINSTREAM AND EMERGING MEMORY TECHNOLOGIES.

	MAINSTREAM MEMORIES				EMERGING MEMORIES		
			FLASH				
	SRAM	DRAM	NOR	NAND	STT-MRAM	PCRAM	RRAM
Cell area	>100 F^2	6 F^2	10 F^2	<4F^2 (3D)	6~50F^2	4~30F^2	4~12F^2
Multibit	1	1	2	3	1	2	2
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	<3 V	<3 V
Read time	~1 ns	~10 ns	~50 ns	~10 µs	<10 ns	<10 ns	<10 ns
Write time	~1 ns	~10 ns	10 µs–1 ms	100 µs–1 ms	<10 ns	~50 ns	<10 ns
Retention	N/A	~64 ms	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6~1E12
Write energy (J/bit)	~fJ	~10fJ	~100pJ	~10fJ	~0.1pJ	~10pJ	~0.1 pJ

Notes: F: feature size of the lithography. The energy estimation is on the cell-level (not on the array-level). PCRAM and RRAM can achieve less than 4F^2 through 3D integration. The numbers of this table are representative (not the best or the worst cases).

When the polysilicon control gate is biased positively, electrons from the transistor source and drain regions tunnel though the oxide layer and get trapped in the silicon nitride. This results an energy barrier between the drain and the source, raising the threshold voltage Vt. By applying a negative bias on the control gate, the electrons can be removed again. After storing or removing electrons from the cell, the controller can measure the state of the cell by giving a small voltage across the source-drain nodes; If current is seen in the cell, it must be in the 'NO TRAPPED ELECTRONS" state, which is considered as logical "1". If no current flows, the cell must be in the "TRAPPED ELECTRONS" state, which is considered as logical "0" state. The needed voltages are normally about 2 V for erased state and around 4.5 V for the programmed state.

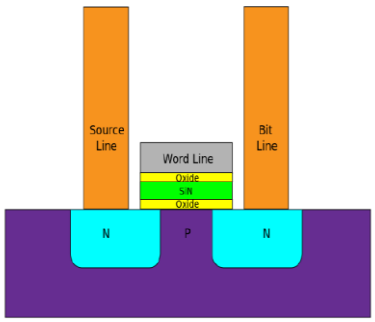

Fig 2: Schematic drawing of a SONOS memory cell

Resistive Random-Access Memory (RRAM)

Resistive random-access memory (RRAM or ReRAM) is a type of non-volatile random-access computer memory (NVM) that works by changing the resistance across a dielectric solid-state material often referred to as a memristor. This technology has some similarities with conductive-bridging RAM (CBRAM) and Phase-change memory (PCM).

CBRAM involves one electrode providing ions that dissolve readily in an electrolyte material, while PCM involves generally sufficient Joule heating to effect amorphous-to-crystalline or crystalline-to-amorphous phase change. On the other hand, RRAM involves generally defects in a thin oxide layer, known as "OXYGEN VACANCIES", which can subsequently change and drift under an electric field. The

motion of oxygen ions and vacancies in the in the oxide would be analogous to the motion of electrons and holes in semiconductor.

A dielectric us used which is normally insulating. Application of a sufficiently high voltage, the dielectric can be made to conduct through a filament or conduction path. The filament (i.e., the conduction path) can arise from different mechanisms including vacancy or metal defect migration. Once the conduction path is formed, it

Nano-RAM/NRAM

Nano-RAM or NRAM is a type of non-volatile random-access computer memory (NVM) based on the position of carbon nanotubes (CNTs) deposited on a chip-like substrate. The small size of the nanotubes. Nantero also refers to it as NRAM.

Initially, NRAM technology was a three-terminal semiconductor device. A third terminal was there to switch the memory cell between memory states. The second generation NRAM technology is a two-terminal memory cell.

The crossed nanotubes can either be touching or slightly separated depending on their position. When touching, the carbon nanotubes are held together by the device's mechanical stiffness (i.e., Van der Waals force).

The NRAM acts as a resistive non-volatile random-access computer memory (RAM) and can be placed in two resistive modes depending on the resistive state of the CNT fabric. When the CNTs are not in contact the resistive state of the fabric is high. This corresponds to an "OFF" or "0" state. When the CNTs comes close, the resistance state of the fabric is low. This corresponds to an "ON" or "1" state. The two resistive states are very stable. In the "0" state, the CNTs are not in contact and remain in a separated state due to the stiffness of the CNTs. This results in a high resistance or low current measurement state between the top and bottom electrodes. In the "1" state, CNTs are in contact and remain contacted due to Van der Waals forces between the CNTs. This results in a low resistance or high current measurement state between the top and bottom electrodes.

To switch the device between different states, a small voltage is applied between top and bottom electrodes. This voltage must be greater than the read voltage. An electrostatic attraction brings the CNTs close if the NRAM cell is in the "0" state and a voltage is applied. This will bring the CNTs close to each other causing a SET operation. After the applied voltage is removed, the CNTs will remain in a "1" or low resistance state due to Van der Waals force.

If the NRAM cell is in the "1" state,the CNT junctions can be separated by applying a voltage greater than the read voltage. The CNTS remain in the "OFF" or high resistance state due to the high mechanical stiffness (Van der Waals forces).

may be reset (broken, causing high resistance in the dielectric) or set (re-formed, causing low resistance in the dielectric) by another voltage.

Phase-change memory/PCM/PRAM

Phase-change memory is a type of non-volatile random-access computer memory (NVM). PRAMs exploit the unique behaviour of chalcogenide glass. In the older generation of PCM, a heating element generally made of TiN, would use. Heat produced by an electric current through this heating element. This either quickly heat or to hold it in its crystallization temperature range for some time. The heat would quench the glass, making it amorphous. When in crystallization temperature range, switching the glass to a crystalline state.

PRAMs switching time and inherent scalability make it most appealing. PRAM's temperature sensitivity is a drawback of the device.

PRAM offers much higher performance in applications where writing quickly is important, because the memory element can be switched more quickly. Also, single bit may be changed to either "1" or "0" and we do not need to erase the cell block. PRAM's high performance, thousands of times faster than conventional hard drives, makes it particularly interesting in non-volatile memory roles that are currently performance-limited by memory access timing.

Ferroelectric RAM

Ferroelectric RAM (FRAM, F-RAM or FeRAM) is a random-access computer memory its construction is very similar to DRAM but FeRAM uses a ferroelectric layer in place of a dielectric layer to attain non-volatility. FeRAM is one of a growing number of alternatives non-volatile random-access memory technologies that offer the same functionality as flash memory.

FeRAM's advantages over flash memory are faster write performance, lower power usage and good read/write endurance. FeRAMs have data retention times of more than 10 years at 850 C (Up to many decades at lower temperature.)

We know that a ferroelectric materials have a nonlinear relationship between the apparent stored charge and the applied electric field. The ferroelectric characteristicsalso have the same form of a hysteresis loop.These characteristics are very similar in shape to the hysteresis loop

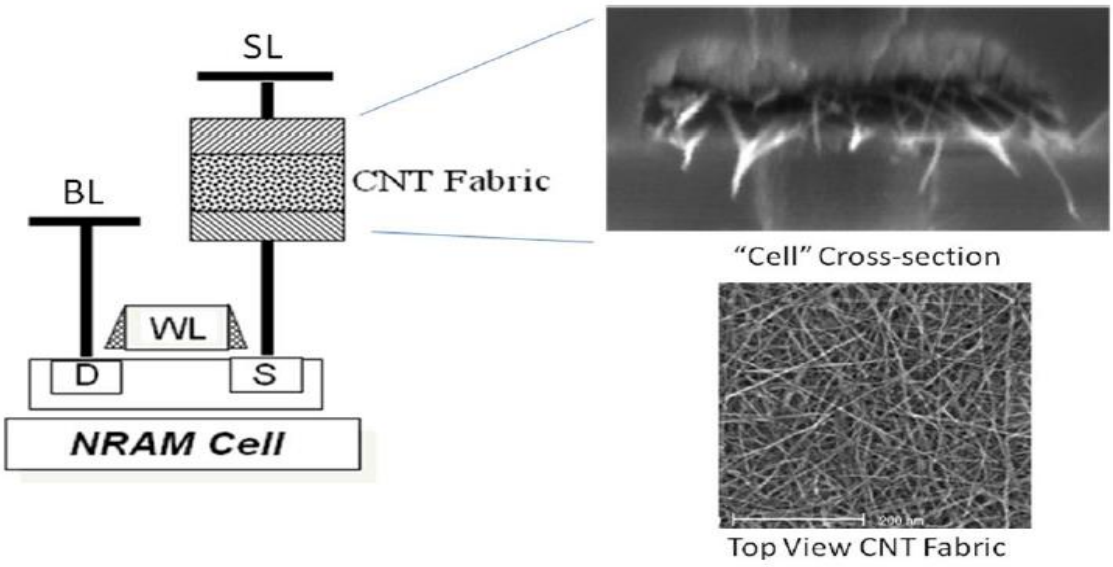

Fig 3: Carbon nanotube fabric

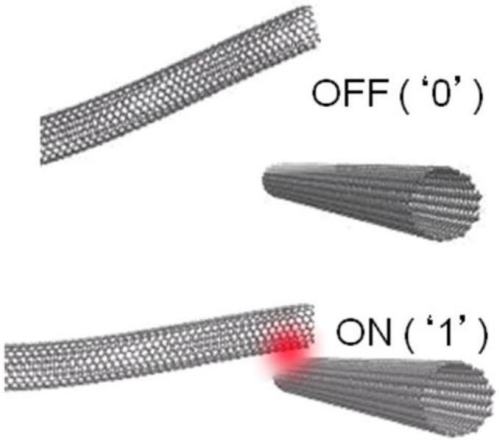

Fig 4: Carbon nanotube contact points

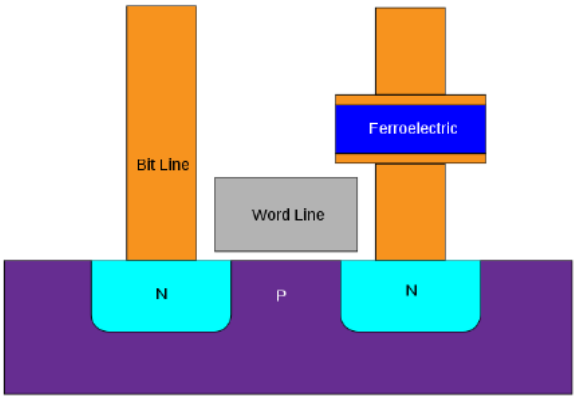

Fig 6: Structure of a FeRAMCell

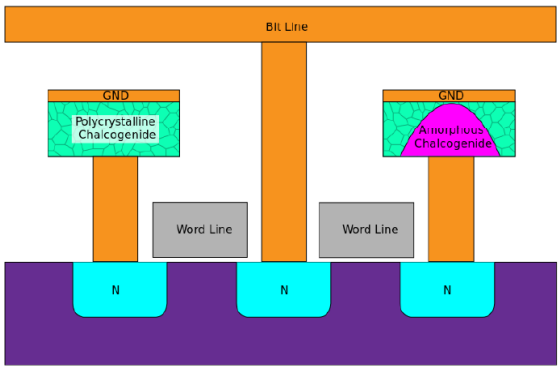

Fig 5 : A cross-section of two PRAM memory cells

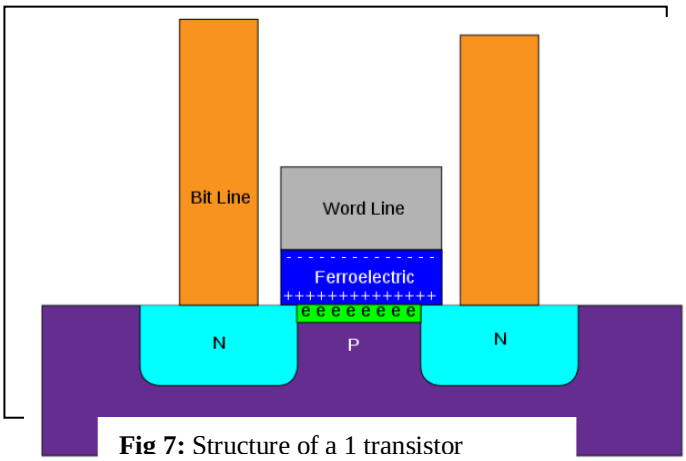

Fig 7: Structure of a 1 transistor

characteristics of ferromagnetic materials. The dielectric constants of a ferroelectric materials arein the much higher range than that of linear dielectrics. The dielectric constants of ferroelectric materials are very high because of the effects of semi-permanent electric dipoles that are formed in the crystal structure of the ferroelectric materials. When across a dielectrican external electric field is applied, the dipoles try to align themselves according to the field direction. This produces small shifts in atomsorientation and in the distributions of electronic charge in the crystal structure. Once thecharge is removed, the dipoles retain their previous polarization state.Thus, the device uses these two possible electric polarization states as two storage data. If one polarization is used for binary "0"s, the other is used for binary "1"s.

Writing is accomplished by applying a field across the Ferroelectric layer by changing the plates on either side of it, forcing the atoms inside into the “up” or “down” orientation.

Spin-Transfer-Torque Magneto Resistive Random-Access Memory (STT-MRAM)

Magneto resistive random-access memory (MRAM) or spin-transfer-torque magneto resistive random-access memory (STT-MRAM) is a non-volatile random-access memory technology available today uses magnetic storage elements as the storage elements.

In conventional RAM chip technologies data are stored as electric charge or current flows, but in MRAM data are stored as magnetic storage elements. Two ferromagnetic plates, separated by a thin insulating layer hold a magnetization, One of the two plates is a permanent magnet of a particular polarity and the other plate's magnetization is changed by the external field to store memory. This type of arrangement is known as a magnetic tunnel junction. This is the simplest structure of an MRAM bit. A memory device can be built from a grid of such "cells".

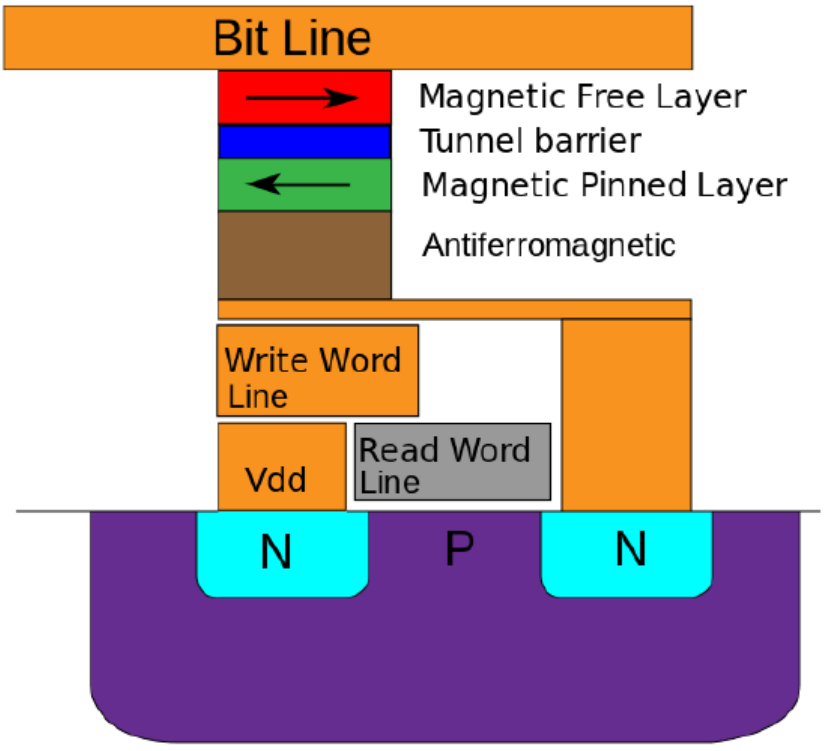

Fig 8: Simplified structure of an MRAM Cell.

Due to the Tunnel magneto resistance, the electrical resistance of the cell changes due to the relative orientation of the magnetization in the two plates. Measuring the resulting current, we can determine the resistance inside any particular cell andalso the magnetization polarity of the writable plate. Generally, if the two plates have the same magnetization alignment, we consider it to mean "1". This is referred to the low resistance state of the memory cell,If the alignment is antiparallel, we consider it to mean "0" as the resistance will be higher, we refer to it as the high resistance state of the memory cell.

CONCLUSION

Moving up the memory hierarchy toward the cache, the memory write/read latency decreases. Moving down the memory hierarchy toward the storage, the memory capacity increases. These mainstream memory technologies are essentially based on the charge storage mechanism: SRAM stores the charges at the storage

nodes of the cross-coupled inverters, DRAM stores the charges at the cell capacitor, and flash stores the charges at the floating gate of the transistor. All these charge-based memories face challenges in scaling down to the 10-nm node and beyond. The easy loss of the stored charges at nanoscale results in the degradation of performance, reliability, and noise margin. In this context, emerging memory technologies that are non-charge based are actively under research and development in the industry, with the hope of revolutionizing the memory hierarchy.[8]

Acknowledgement

I take this opportunity to express my profound gratitude and deep regards to our respected Professor, Dr.AsitBaran Bhattacharya, Professor, Department of Electronics and Communication, Techno India University for his guidance, monitoring and encouragement throughout the course of this paper.

REFERENCES

[1] Wong, H. S. P. and Salahuddin, S."Memory leads the way to better computing," Nat.Nanotechnol., vol. 10, pp. 191–194, Mar.2015.

[2] Zhu,B. J. "Magnetoresistive random access memory: the path to competitiveness and scalability," Proc. IEEE, vol. 96, no. 11, pp. 1786–1798, 2008.

[3] Wong, H. S. P.;Raoux, S.; Kim, S.; Liang,S.;Reifenberg,J. P.; RajendranB.;Asheghi, M. and Goodson, K. E."Phase change memory," roc. IEEE, vol. 98, no. 12, pp. 2201– 2227, 2010.

[4] Wong. H. S. P.; Lee. H. Y. ; Yu. S.; Chen.Y. S.; Wu Y.; Chen. P. S.; Lee. B; Chen. F. T., and Tsai. M. J; "Metal–oxide RRAM," Proc. IEEE, vol. 100, no. 6, pp. 1951–1970, 2012.

[5] Park. S. P.; Gupta. S.; Mojumder. N., Raghunathan. A., and Roy. K, "Future cache design using STT MRAMs for proved energy efficiency: devices, circuits and architecture," in Proc. 49th ACM Design Automation Conf., San Francisco, CA, 2012, pp. 492–497.

[6] Jung. M., Shalf. J., and Kandemir. M., "Design of a large-scale storage-class RRAM system," in Proc. ACM Int. Conf. Supercomputing, 2013, pp. 103–114.

[7] YU. Shimengand Chen. Pai-Yu., "Emerging Memory Technologies: Recent Trends and Prospects," IEEE Solid-State Circuits Magazine, Volume 8, pp. 43-56, Spring 2016.

[8] Freitas. R. F., and Wilcke. W. W., "Storage class memory: The next storage system technology," IBM J. Res. Develop., vol. 52, no. 4.5, pp. 439–447, 2008.