The International Journal Of Engineering And Science (IJES)
|| Volume || 5 || Issue || 9 || Pages || PP 61-68|| 2016 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Emerging Memory Technologies

O.D. Alao[1], J.V. Joshua[2], D.O. Kehinde[3], E.O. Ehinlafa[4], M.O. Agbaje[5], J.E.T Akinsola[6]

[1, 2, 5 ,6] Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria
[3] Department of Basic Science, Babcock University, Ilishan-Remo, Nigeria
[4] Department of Physics, University of Ilorin, Ilorin, Nigeria.

ABSTRACT

The processor caches, main memory and storage system is an integral part of any computer system. As information begins to accumulate, higher density and long term storage solutions are necessary. Due to this, computer architects face some level of challenges in developing reliable, energy-efficient and high performance memories. Also, existing storage devises are degrading in performance, cost, and sizes. Power consumption from the factory has increased, as newer codes are written, and server hardware capabilities are not adequate to handle big data of the future. New emerging memories (NEMs) are presently with its properties likely to open doors to innovative memory designs to solve the problems. This paper looks at the features of the emerging memory technologies, and compares incumbent memories types with the expected future memories.

Keywords: *Memory Storage, New Emerging Memory Technologies, Spin-Transfer Torque Magnetic Random-Access Memory (STT-MRAM), Resistive Random Access Memory (ReRAM) and Phase Change Memories PCM.*

Date of Submission: 17 May 2016 Date of Accepted: 20 September 2016

I. INTRODUCTION

The pecking order of memory and storage device is a critical component of various computer systems. Processor caches act as a subset of data and instructions stored in the memory. Data stored in the main memory are stored in large, slow storage devices, such as disks and flash. Data from modern applications such books, maps, photos, audios, videos, references, facts, and conversations rely on both real and offline processing and their dataset can be in gigabytes, terabytes, zettabytes or even larger in size.

Regrettably, the scaling of conventional memory technologies is at risk. Memory technologies, such as SRAM (Static Random Access Memory) and DRAM (Dynamic Random Access Memory), are experiencing scalability challenges as a result to the limitations of their device cell size and power dissipation.

NEMs offer several benefits such as low power (especially low leakage), high density, and the ability to retain the stored data over long time periods (non-volatility) that have made them attractive for use as secondary storage. Flash memory is already widely used in consumer electronics and in solid-state disks due to its low cost and extremely high density [2]

The dynamic and increasing power of DRAM over the power leakage of SRAM is a threat to circuit and architecture designers of future memory hierarchy designs. Energy consumption has become key design limiters as the memory hierarchy continues to contribute a significant fraction of overall system energy and power. The lack of memory technology scaling can make it difficult for the memory hierarchy to achieve high capacity and efficiency at low cost. As a result, it remains a very attractive technology for data archiving, with a sustainable roadmap for the next ten to twenty years, well beyond the anticipated scaling limits of current conventional technology [1].

There is a fundamental trend towards designing entire systems such that they are optimized for particular work-loads, departing from the traditional general-purpose architecture. The typical system, with standard CPUs consisting of a small number of identical cores with a common set of accelerators and relying on a memory and storage hierarchy has reached its limits in terms of delivering competitive performance improvements for an increasingly diverse set of workloads: future systems will be built out of increasingly heterogeneous components. This article examines today’s memory storage requirements, reviews recent research efforts on computer architecture design with New Emerging Memories design.

II. NEW EMERGING MEMORY(NEM) TECHNOLOGIES

NEM technology is brand of computer hardware memory systems being developed with a view to either extending the existing memory technology's capabilities or eventually replacing the whole technology all together. Scalability beyond the incumbent is a critical requirement.

Most NEM technologies have resistive storage elements. Resistive RAM (RRAM or ReRAM) is a type of non-volatile RAM that is highly promising in the next generation of memories for computers of the future. It works by changing the resistance across a dielectric solid-state material often referred to as a Memristor.

2.1. STT-RAM

Spin-Transfer Torque Magnetic Random-Access Memory (STT-MRAM) is the latest design of the magnetic RAM (MRAM). The information carrier of STT-MRAM is a Magnetic Tunnel Junction (MTJ) instead of electric charges which makes the difference between it and the conventional RAM. See Fig 1

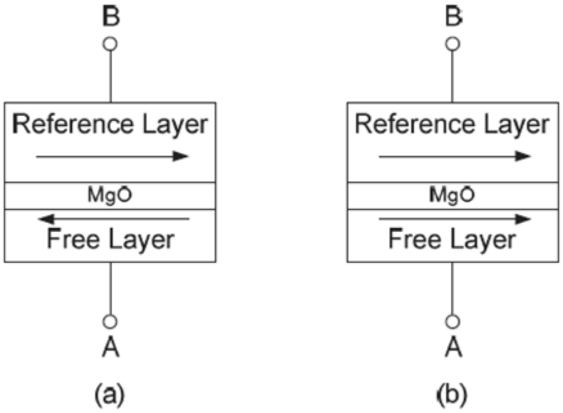

Fig. 1 A conceptual view of MTJ structure. (a) Anti-parallel (high resistance), which indicates "1" state; (b) Parallel (low resistance), which indicates "0" state.

Each MTJ contains two ferromagnetic layers and one tunnel barrier layer. Figure 1 shows a conceptual illustration of an MTJ. One of the ferromagnetic layer (reference layer) has fixed magnetic direction while the other one (free layer) can change its magnetic direction by an external electromagnetic field or a spin-transfer torque [5].

In case, the two ferromagnetic layers have different directions, the MTJ resistance is high, indicating a "1" state (the anti-parallel case in Fig. 1 (a)); if the two layers have the same direction, the MTJ resistance is low, indicating a "0" state (the parallel case in Fig. 1 (a)).

STT-MRAM changes the magnetic direction of the free layer by directly passing a spin-polarized current through the MTJ structure. Comparing to the previous generation of MRAMs that uses external magnetic fields to reverse the MTJ status, STT-MRAMs has the advantage of scalability, which means the threshold current to make the status reversal will decrease as the size of the MTJ becomes smaller.

In the STT-MRAM memory cell design, the most popular structure is composed of one non metal oxide semi-conductor (NMOS) transistor as the access controller and one MTJ as the storage element.

In Fig. 2, the storage element, MTJ, is connected in series with the NMOS transistor. The NMOS transistor is controlled by the word-line (WL) signal. The detailed read and write operations for each MRAM cell is described as follows:

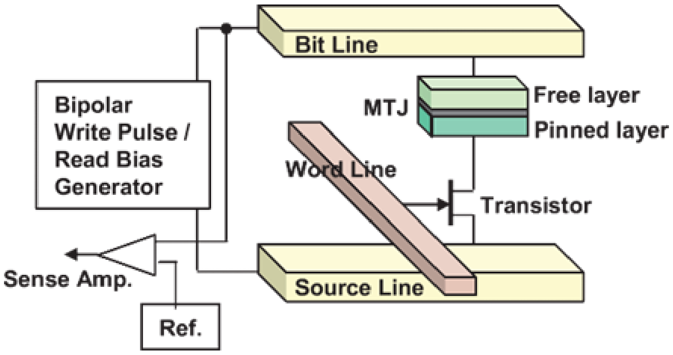

Fig. 2 An illustration of an STT-MRAM cell with read/write circuitry.

There are two operations that happens which are:

i. **Read Operations:** This operation happens when the NMOS is turned on and a voltage is applied between the bit-line (BL) and the source-line (SL). It will be observed that the voltage is negative and usually small which will cause a current passing through the MTJ, but it is not small enough to invoke a disturbed write operation. A sense amplifier compares this current with a reference current and then decides whether a “0” or a “1” is stored in the selected MRAM cell.

ii. **Write Operation:** This operation happens when a positive voltage difference is established between SL and BL for writing for a “0” or a negative voltage difference is established for writing a “1”. The current amplitude required to ensure a successful status reversal is called threshold current. The current is related to the material of the tunnel barrier layer, the writing pulse duration, and the MTJ geometry [5].

2.2. Phase Change Memory(PCM)

PCM is a type of non-volatile memory that exploits the property of chalcogenide glass to switch between two states, amorphous and crystalline, with the application of heat using electrical pulses. The phase change material can be switched from one phase to another reliably, quickly, and a large number of times. The amorphous phase has low optical reflexivity and high electrical resistivity. Whereas, the crystalline phase (or phases) has high reflexivity and low resistance [3].

For storage of information, PCM uses chalcogenide-based material. It has a wide range of resistivity, about three orders of magnitude, and this forms the basis of data storage. The amorphous, high resistance state is used to represent a bit “0,” and the crystalline, low resistance state represents a bit “1.” When germanium-antimony-tellurium (GeSbTe or GST) is heated to a high temperature (normally over 600 ◦C), it gets melted and its chalcogenide crystallinity is lost.

Once cooled, it is frozen into an amorphous and its electrical resistance becomes high. This process is called RESET. A way to achieve the crystalline state is by applying a lower constant-amplitude current pulse for a time longer than the so-called SET pulse. This is called SET process. The time of phase transition is temperature-dependent [5].

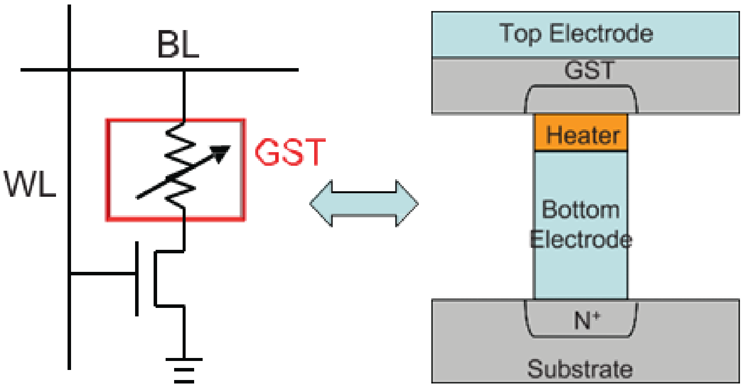

Fig 3: illustration of a PCM cell

There are two operations that happens which are;

i. **Read Operation:** To effectively read data stored in the PCM cells across the GST, a small voltage is applied. The read voltage to invoke detectable current must be strong but low enough to avoid write disturbance. In Fig. 3, every basic cell contains one GST and one NMOS access transistor. This structure has a name of “1T1R” where “T” stands for the NMOS transistor and “R” stands for GST. The GST in each PCM cell is linked to the drain-region of the NMOS in series so that the data stored in the cells can be accessed [4].

ii. **Write Operation:** There are two kinds of write operations,

a. The SET operation that switches the GST into crystalline phase when heated

b. The RESET operation that switches the GST into amorphous phase.

Note: Both operations are controlled by electrical current: high-power pulses for the RESET operation heat the memory cell above the GST melting temperature; moderate power but longer duration pulses for the SET operation heat the cell above the GST crystallization temperature but below the melting temperature. The temperature is controlled by passing through a certain amount of electrical current and generating the required heat.

2.3 Resistive Random Access Memory(ReRAM)

Any memory technology that represent digital information using its variable resistance is what ReRAM is meant for. An insulating dielectric is conducted via conduction path by applying an adequate high voltage. The conduction path can be generated by different mechanisms, including defects, metal migration, etc. The filament may be reset (broken, resulting in high resistance) or set (reformed, resulting in lower resistance) by applying an appropriate voltage [5].

ReRAM structure is a one cell i.e. a one metal oxide layer sandwiched by two metal electrodes - the top electrode (TE) and the bottom electrode (BE), shown in figure 4. A low resistance state (LRS) represents digital "1" while a high resistance state (HRS) represents digital "0."

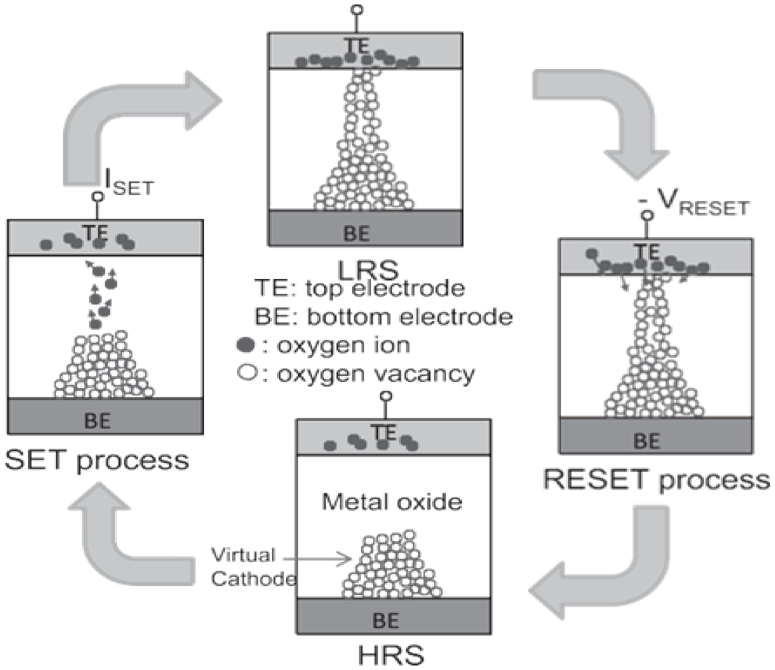

Fig.4: ReRAM structure

III. OTHER EMERGING MEMORY TECHNOLOGIES

3.1. Programmable Metallization Cell (PMC)

According to [7], PMC also known as the Electrochemical Metallization Memory Cells (ECM) rely on electrochemical growth and dissolution of a conducting filament within the insulating layer. They consist of an active Cu or Ag electrode, a cation conducting insulating layer, and an inert counter electrode. PMC cells exhibit multibit data storage capability, scalability almost down to the atomic level, and very low programming power. Moreover, potential backend of line compatible integration has been demonstrated making ECM cells of high interest for future nonvolatile memory.

3.2. Polymer memory

Throughout the last few years, polymers have found growing interest as a result of the rise of a new class of nonvolatile memories. In a polymer memory, a layer consists of molecules and/or nanoparticles in an organic polymer matrix is sandwiched between an array of top and bottom electrodes as illustrated in Figure 5.

Moreover, polymer memory has the advantage of a simple fabrication process and good controllability of materials [9]. Polymer memory could be called digital memory with the latest technology. It is not possible for a silicon-based memory to be established in less space, but it is possible for polymer memory. The non volatileness and other features are inbuilt at the molecular level and offers very high advantages in terms of cost. But turning polymer memory into a commercial product would not be easy.

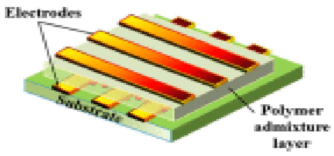

Fig 5: Structure of a polymer memory device.

3.3. Molecular Memory

A molecular memory is a nonvolatile data storage memory technology that uses molecular species as the data storage element, rather than, e.g., circuits, magnetics, inorganic materials, or physical shapes. In a molecular memory, a monolayer of molecules is sandwiched between a cross-point array of top and bottom electrodes as shown in Fig 6. The molecules are packed in a highly ordered way, with one end of the molecule electrically connected to the bottom electrode and the other end of the molecule connected to the top electrode, and this molecular component is described as a molecular switch. Then, regarding the molecular memory operation, by applying a voltage between the electrodes, the conductivity of the molecules is altered, enabling data to be stored in a nonvolatile way. This process can then be reversed, and the data can be erased by applying a voltage to the opposite polarity of the memory cell.

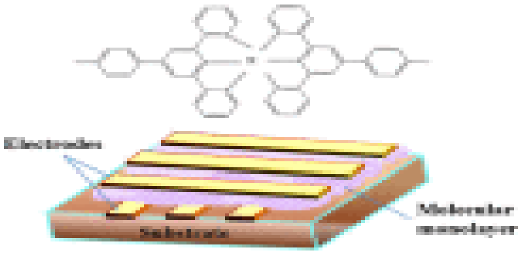

Fig 6: Cell structure of a molecular memory device

3.4. MNW

The molecular nanowire array (MNW) memory is fundamentally different from other semiconductor memories; information storage is achieved through the channel of a nanowire transistor that is functionalized with redox-active molecules rather than through manipulation of small amounts of charge. It is relatively slow and lacks the random access capability, wherein data that can be randomly read and written at every byte are being actively pursued. Figure 7 shows the schematic design of a MNW memory cell.

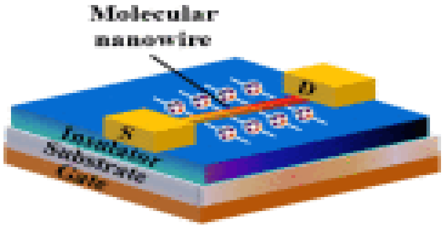

Fig 7: A MNW memory cell structure

3.5. QD (QUANTUM DOTS) Memory

Memory made from tiny islands of semiconductors - known as quantum dots - could fill a gap left by today's computer memory, allowing storage that is fast as well as long lasting. Researchers have shown that they can write information into quantum dot memory in just nanoseconds. New research shows that memory based on quantum dots as shown in figure 8 can provide the best of both: long-term storage with write speeds nearly as fast as DRAM. A tightly packed array of tiny islands, each around 15 nm across, could store 1 terabyte (1,000 GB) of data per square inch, the researchers say. Dieter Bimberg and colleagues at the Technical University of Berlin, Germany, with collaborators at Istanbul University, Turkey, demonstrated that it is possible to write information to the quantum dots in just 6 ns. The key advantages of quantum dot (QD) are the high read/write speed, small size, low operating voltage, and, most importantly, multibit storage per device.

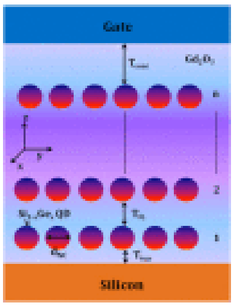

Fig 8 Structure of quantum dot memory [8].

3.6. 1T1R-RRAM

One-transistor one-resistor (1T1R)-RRAM is also one class of emerging memory technology with impressive characteristics. It meets the demands for next-generation memory systems. It is expected that 1T1R-RRAM could be able to meet the demand of high-speed (e.g., performance) memory technology. The 1T1R structure is chosen because the transistor isolates current to cells, which are selected from cells which do not. The basic cell structure of 1T1R is depicted in Figure 9. 1T1R-RRAM consists of an access transistor and a resistor as a storage element. [6] posited that the 1T1R cell structure is similar to that of a DRAM cell in that the data is stored as the resistance of the resistor and the transistor serves as an access switch for reading and writing data. Moreover, the 1T1R structure is more compact and may enable vertically stacking memory layers, ideally suited for mass storage devices. But, in the absence of any transistor, the isolation must be provided by a 'selector' device, such as a diode, in series with the memory element, or by the memory element itself.

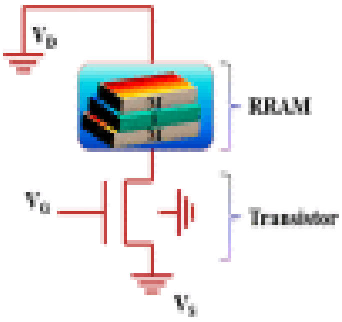

Fig 9: The basic cell structure of 1T1R-RRAM.

IV. COMPARISON BETWEEN THE DIFFERENT NEMS

A comparison between some of the different NEMs as against the main current memory hierarchy components (SRAM, DRAM, disk and flash) is drawn

The following set of criteria is employed;

1. **Maturity:** Whether the technology is currently used in the market or it is in early or later search stages before being commercially mature
2. **Cell size:** the cell size, using standard feature size
3. **Read Latency:** the speed for reading values from memory cell
4. **Write Latency:** the speed for writing values to a memory cell
5. **Endurance:** the number of write cycles that a memory cell endures before eventually wearing out
6. **Energy:** energy spent per bit access. Related to dynamic power
7. **Static power:** whether power needs to be spent while accessing the memory device. This includes refreshing solid memory contents due to energy leakage or keeping disks spinning to achieve lower access latencies.
8. **Non-volatility:** whether the memory technology is volatile or not.

Table 1 compares NEMs with traditional memory and storage technologies, in terms of performance, energy, density, and endurance [5].

FEATURE	SRAM	DRAM	DISK	FLASH	STT-MRAM	PCM	ReRAM
Maturity	Product	Product	Product	Product	Advanced Development	Advanced Development	Advanced Development
Read latency	<10ns	10-60ns	8.5ms	25 µs	<10ns	48ns	<10ns
Write latency	<10ns	10-60ns	9.5ms	200µs	12.5ns	40-150ns	10ns
Energy Per Bit Access	>1pJ	2pJ	100-1.000mJ	10nJ	2pJ	100pJ	0.02pJ
Leakage Power	High	Medium	High	Low	Low	Low	Low
Endurance	>10^{15}	>10^{15}	>10^{15}	104	>10^{15}	$10^5 – 10^9$	$10^5 – 10^{11}$
Non volatility	No	No	Yes	Yes	Yes	Yes	Yes
Scalability	Yes	Yes	Yes	Yes	Yes	Yes	Yes

V. REPLACING TECHNOLOGIES WITH NEMS

Since NEMs are persistent and have similar mechanics (no moving parts) and lower energy consumption, it’s natural to refer to them as persistent storage. NEMs would also perform as a disk replacement.

5.1. PCM as a Disk Replacement

According to [3], when considering PCM as a disk replacement, the following would be considered;

1. Advantage of solid-state implementation (non-moving parts), very low latencies, low cost per bit and physical durability.
2. Poor write endurance (can be attacked through wear- leveling) and scaling down to a point where that can compete with disks (using multi-layers, multi-bit cells, etc.)
3. The prices of NEMs such as PCM will be much higher than disks, but over time it tends to decrease dramatically s market adoption progresses.
4. Its I/O are accessed as a block device.
5. Runtime error detection or correction mechanisms
6. Read or write should be asymmetrical, the read being as a single word.

Table 2 below is a characteristic of NEM replacement PCM

Characteristics of PCM	
Capacity	1 TB
Read or write access time	100 ns
Data rate	> 1 GB/s
Sustained I/O rate	238 000 sio/s (SIO: start I/O)
Sustained bandwidth	975 mb/s
Write endurance	10^{12} write

5.2. PCM or STT-MRAM as Disk Replacement

To evaluate as a persistent storage, five evaluation criteria that should be taken into consideration;

1. **Density:** It is related with cost/GB which is the most important parameter. The cost/GB scales in proportion to the density’ which is cell size divided by the number of bits per cell.
2. **Power efficiency** – important for mobile/embedded devices as well as for datacenters.
3. **Access time:** time interval between write/read request and the writing or availability of the data: Specially important for datacenters
4. **Endurance:** the number of times a bit can be rewritten
5. **Retention:** the length of time a bit remains stable.

Table 3: Comparison of NEM Technologies for Disk Replacement

Device Type	DRAM	Flash	PCM	STT-MRAM
Maturity	Product	Product	Product	Product
Density	83gb/chip	64gb/chip	512mb/chip	2mb/chip
Cell size	$6F^2$	$4F^2$	$5F^2$	$4F^2$
MLC Capacity	No	4 bits/cell	4bits/cell	4biits/cell
Energy	2pJ	10pJ	100pJ	0.02pJ
Access time	10/10ms	200/25ns	100/20ns	10/10ns
Retention	1016/64ms	105/10yr	105/10yr	1016/10yr

From Table 3 above, the technologies with the best opportunity have a small cell size and the capability of storing multiple bits per cell.

Phase Change Memory (PCM) and Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) appear to meet these criteria. It is concluded that the PCM and STT-MRAM are the technologies with higher probabilities of being a feasible replacement for disk in the future since they meet the criteria (small cell size and capability of storing multiple bits per cells).

VI. CONCLUSION

Since the limitations of the traditional technologies threatens the sustainable growth of performance and energy efficiency of computer systems, memory technology should be non-volatile, low-cost, high dense energy efficient, fast and within high endurance.

Therefore, superior density, power, and non-volatility characteristics of emerging NEM technologies provide opportunities to break this traditional system organization.

The problems in performance and strength as compared with traditional memory technologies is as a result of re-architecting processor caches, main memory, and other storage system. By taking the full advantages of the NEM technologies, computer performance can be highly enhanced.

REFERENCES

[1]. Chris H. Kim (2011): *Modeling, Architecture, and Applications for Emerging Memory Technologies.* Pp 44-45

[2]. Clinton W. Smullen, V , Vidyabhushan M, Anurag N, Sudhanva G, Mircea R. Stan.C (2011): Relaxing Non-Volatility for Fast and Energy-Efficient STT-RAM Caches. HPCA'11 Proceedings of the 2011 IEEE 17th International Symposium on High Performance Computer Architecture Pp 50-61

[3]. Qureshi, M, Srinivasan,V and Rivers,J (2009): ''Scalable High Performance Main Memory System Using Phase-Change Memory Technology,'' Proc. 36th Int'l Symp. Computer Architecture (ISCA 09), ACM Press, pp. 24-33.

[4]. Xiaoxia Wu, Jian Li, Lixin Zhang, Evan Speight, Ram Rajamony, Yuan Xie (2009): Hybrid Cache Architecture with Disparate Memory Technologies:ISCA proceedings of the 36th annual international symposium on computer architecture pages 34-45 ACM New York, NY, USA

[5]. Zhao, J; S. Li, D. H. Yoon, Y. Xie, and N. P. Jouppi Kiln (2015): Closing the performance gap between systems with and without persistence support. In Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'13). ACM, New York, NY, 421– 432. DOI

[6]. Zangeneh M, Joshi A. Design and optimization of nonvolatile multibit 1T1R resistive RAM. IEEE Trans VLSI System. 2014;22(8):1815–1822.

[7]. Chen, A., Hutchby, J., Zhirnov, V., and Bourianoff, G. (2015). *Emerging nanoelectronic devises*, First Edition. John Wiley & Sons, Ltd.

[8]. Manna S, Aluguri R, Katiyar A, Das S, Laha A, Osten H, Ray S. MBE-grown Si and $Si_{1-x}Ge_x$ quantum dots embedded within epitaxial Gd_2O_3 on substrate for floating gate memory device. Nanotechnology. 2013;24:505709

[9]. Prakash A, Ouyang J, Lin JL, Yang Y.(2006) *Polymer memory device based on conjugated polymer and gold nanoparticles. Journal of Applied Physics. 2006;100(5):054309–054314.*