ADVANCED RRAM AND FUTURE OF MEMORY

Selva Lakshman Murali

***Abstract* - Memory is one of the fundamental components of computing that can be defined as the unit which enables data to be stored, retrieved, processed, and manipulated. It is visibly seen that in recent years the demand for compact, large volume, reliable, and high-speed memories have been enormous added to the huge dependence on data for all decision making and enablement of artificial intelligence. It is to be noted that a modern CPU is composed of more than fifty percent of memory blocks along with processing cores where this memory element is a key factor influencing the performance of the processor[1][2]. While there are metrics such as bandwidth, latency, energy efficiency, material selection, architecture, and operating conditions hindsight is the process that delivers the top-class memories. In this paper, we discuss concepts of emerging, future memories achieved through innovation, research, and development and focus more on the RRAM[3], its working, and targeted applications which is strongly being considered the memory element for high-performance computing (HPC) devices.**

***Keywords* - Bandwidth, Latency, Process, RRAM, Energy efficiency, Material selection, Architecture, High-Performance Computing, Memory wall**

I. INTRODUCTION

Emerging nonvolatile memories, such as ferroelectric random access memory (FRAM), magnetic random access memory (MRAM), phase-change random access memory (PRAM), and resistive random access memory (RRAM), have received a lot of attention in recent years. RRAM, which switches between a high and low resistance state via resistive switching, has the most advantages in terms of easy fabrication, simple structure, excellent scalability, fast switching, high integration density, and compatibility with current complementary metal oxide semiconductor (CMOS) technology. As a result, these devices are receiving increasing attention. Resistance ratio[3][4], or the ratio of resistance at HRS (High Resistance State) to resistance at LRS (Low Resistance State), is significant in RRAM, as shown in Fig.1 applications because it has a direct impact on programming and erasing accuracies. To identify the two resistance states in circuit design, a resistance ratio larger than 10 is usually necessary[3]. On the side note, currently RRAMs are being evaluated for applications in biosensors, space applications and biological studies as it is expected to produce better inline results.

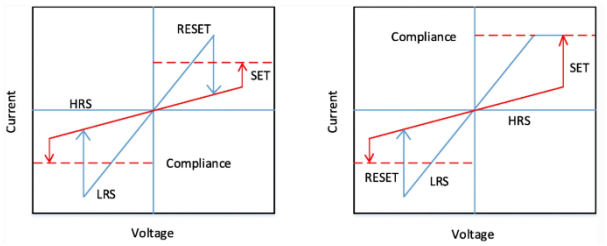

Fig. 1. I-V Curves of RRAM with unipolar and bipolar switching

A. WORKING OF RRAM

The metal-insulator-metal (MIM) structure is made up of an insulating or resistive material (I) sandwiched between two electron-conductive electrodes (M) in a resistive-switching memory cell of an RRAM. The MIM cell may be switched between two states at the right voltage. A high-resistance state (HRS) and a low-resistance state (LRS)[3] are two types of states. These two states can be used to represent the logic values 1 and 0. The resistive-switching behavior of an RRAM[5] device is characterized as unipolar or bipolar depending on the voltage polarity. Resistive switching is generated by a voltage with the same polarity but a different magnitude for unipolar switching. One polarity is utilized to switch from HRS to LRS, while the opposing polarity is used to convert back to HRS in bipolar switching. Nonpolar switching is a third resistive-switching behavior in which positive or negative voltage is used to transition from LRS to HRS (the reset process) and from HRS to LRS (the set process). To avoid a device's harsh breakdown, the setting current is normally capped, regardless of the kind of switching.[3]

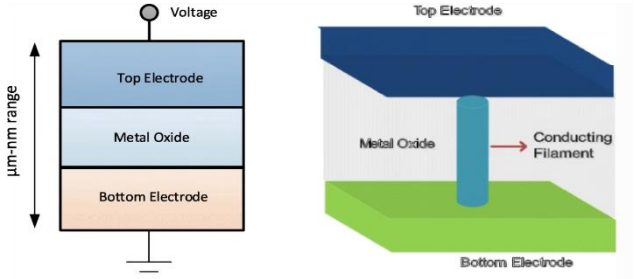

Fig. 2. Schematic of metal-insulator-metal structure for RRAM and Cross-sectional view of RRAM

B. PERFORMANCE METRICS OF RRAM

There are a set of performance metrics that define the quality of the RRAM when put under operation.

Endurance: Transitions between a high resistance state (HRS) and a low resistance state (LRS) are common in resistive random-access memory (LRS). Each switch between the resistive states has the potential to inflict lasting damage and degrade RRAM performance. The number of times an RRAM device can be switched between the HRS and the LRS while maintaining a reasonably identifiable ratio between them is described as endurance.[6][3]

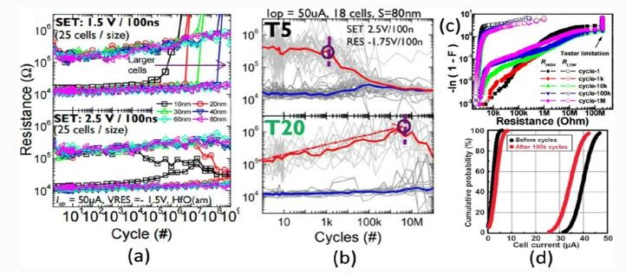

Fig. 3. Plot of endurance versus resistance on multiple cycles

Operating Speed: The shortest time for programming or deleting a device cell is known as operating speed. This device's switching speed is as quick as 5 nanoseconds, which is the fastest speed ever documented. RRAM devices' optimal operating speeds are now between 5 and 100 nanoseconds.[3-6]

Retention Time: The retention period of a memory cell after programming or erasing reveals the memory cell's inherent capacity to maintain its material. Most commercially available devices are guaranteed to store data for at least 10 years, whether the power is on or off[3-6].

Device Yield: Due to the unpredictability of the oxygen content, nonstoichiometric oxide devices have a fatal fault of poor device yield. Certain suggestions have used stoichiometric ZrO2 instead of nonstoichiometric ZrO2 for nonvolatile RRAM devices to tackle the problem of low device yield for nonstoichiometric oxides[3-6]. Other strategies for improving device yield have also been proposed, such as employing a suitable electrode material, using doped metal oxide, and purposefully inserting metal nanocrystals in oxide.

C. APPLICATIONS OF RRAM

Because of its high speed, non-volatile data store capabilities, increased storage density, and logic processing function, RRAM is viewed as one of the outstanding prospects among developing memory technologies that has the potential to reorganize the memory hierarchy. Some cutting-edge applications include non-volatile logic development, neuromorphic computing, non-volatile SRAM and security intensive applications while utilizing the unique physical unclonable function (PUF)[6][3] based on RRAM. RRAM has seen unique application areas which have been mentioned below in the chart.

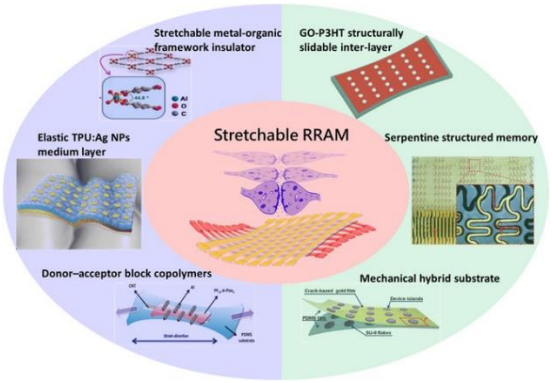

Fig. 4. Novel applications of RRAM

D. CHALLENGES AND FUTURE OF RRAM

Because of its simple structure, compatibility with existing CMOS technology, high switching speed, and ability to scale to the tiniest size, RRAM is one of the most promising memory technologies. In reality, Flash memory technology is now having difficulty shrinking to smaller dimensions, and as a result, RRAM is developing as a viable substitute, particularly for quick operation and medium storage density memory applications. [12][6]The dependability of RRAM is one of the most important features that has to be extensively examined. A system must be created to guarantee that the device's functioning failure is detected. I'm also a circuit designer. [12][5]The 1D1R operation is required to accomplish high-density memory operation in RRAM. By using the RRAM device in unipolar mode, this may be accomplished. However, as compared to bipolar operation, the unipolar operation requires more current for the reset procedure. Random telegraph noise (RTN)[12][3] is another factor that affects the performance of RRAM.

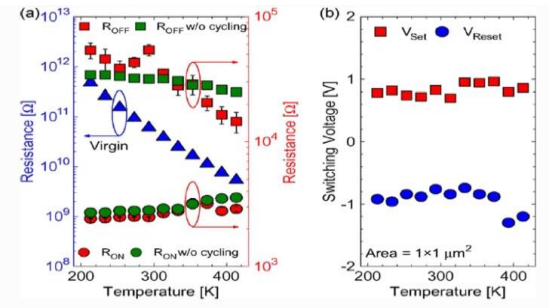

Fig. 5. Plot representing the effect of temperature on resistance and switching voltage in RRAM

Although RRAM[6] is appealing for use in neuromorphic computations, the capacity to address variability concerns, not only under nominal operating settings but also at high temperatures, is the major obstacle in industrializing RRAM before it can be employed in a wide range of applications.

II. FUTURE OF MEMORIES

In the digital era, there are many technological revolutions, the amount of data is growing exponentially. By 2025 about 175 million terabytes will be manufactured, which is ten times more than the volume manufactured in 2015. This trend comes with a huge increase in the usage of the various applications of memories, networks and video on demand. While the computational engines have continuously benefitted from the scaling effect developed by Moore's law and it is well known that future enhancements of the system performance is linearly dependent on the memories and it is well-known as a memory wall.

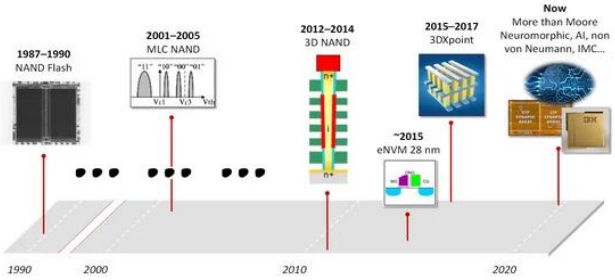

Fig. 6. Growth of memories over the decades

The disruptive technologies are expected to balance the memory hierarchy and interconnect to catch-up and match the scale requirements of the performance, energy-efficient and tremendously scalable in storage capacity, performance, and energy-efficient.

A. EMERGING NON-VOLATILE MEMORIES (NVM)

NVM technologies are Magnetic Memory, Sing Torque Transfer RAM (STT-RAM), Phase change RAM(PCRAM). These are illustrating the great demands in the replacement of ongoing SRAM/DRAM technologies. Technology scaling for SRAM and DRAM are highly constrained by technology limits. In particular, increasing leakage power and refreshing dynamic power for DRAM have posed challenges for circuit/architecture designers for future memory hierarchy design.

Magnetic Memory MRAM is a type of memory that stores data in magnetic material that is integrated with silicon circuitry. It utilizes the magnetism of electron spin to provide non-volatility. MRAM not only has good durability for DRAM and SRAM applications, but it has also recently shown improved stability for embedded applications. The term "stack complexity" is frequently used to describe a device constraint. The next step will be to improve the system's scalability and capacity.

STT-RAM It is a new type of Magnetic RAM (MRAM). The storage element of STT-RAM is the magnetic tunneling junction, in which a thin tunneling dielectric is sandwiched by two ferromagnetic layers.

PCRAM The technology is based on a chalcogenide alloy which is typically Ge2-Sb2-Te5, (GST) material, the data storage capability is achieved from resistance differences between an amorphous (high-resistance) and a crystalline(low-resistance).

B. COMPARISION OF MEMORIES

Current trend for volatile memory power requirements, especially static power have been increased with smaller semiconductor lithographic features. And, of all these memories MRAM is promising for SRAM and DRAM alternative and working at lower power.

	SRAM (Static) for cache	DRAM (Dynamic) for main memory	PCM (Phase change)	RRAM (Resistive)	MRAM (Magneto-resistive)
Latency	300 ps ~ ns	10 ~ 30 ns	~50 ns	< 10 ns	few ns or faster
Endurance	$> 10^{16}$	$> 10^{15}$	$10^{8} – 10^{12}$	$10^{6} – 10^{12}$	$> 10^{15}$
Retention	volatile	volatile	> 10 years	> 10 years	> 10 years

Fig. 7. Performance comparison of various emerging memories

C. FACTORS AFFECTING MEMORY GROWTH

Latency / Energy overhead problems This is mainly because of non-volatility nature of memories. The high-density is the main reason for latency, driving more cells will require more time and energy and on a positive note the NVM has a good stability compared to volatile memory it is an intrinsic characteristic of NVM. There are many ways to exploit this drawback: Hybrid Cache/Memory Architecture, Read-Preemptive Novel Buffer Architecture, Redundant Write Elimination.

Lifetime reliability The memory cells experiencing frequent write and read operations will experience this and this will be aggravated with worse skew between the cache and main memory. Wear leveling techniques are applicable to work around the limitation of write endurance by arranging the read so that write will be evenly spread across storage cells. Wear level techniques would include Row Shifting, word-line Remapping and Bit-line shifting, and segment swapping.

III. INDUSTRY INSIGHTS

There is a continuous movement at the industry level through collaboration with academic researchers and scientists including companies such as Micron Technology, Western Digital, Kioxia, Intel Corporation and at the foundry level companies such as ASML, KLA-Tencor are also actively engaged in going beyond the exiting memory architectures to increase yield, access times and process improvements. For instance, recently Micron Technology launched the 1-Alpha DRAM process and the Quad-level SSD based on NAND technology[8].

Additionally, companies such as Tesla which depend on data to model its vehicles for better autonomous operation depend heavily on high speed, high bandwidth memories which provide safety while enabling driver with best of artificial intelligence.

IV. CONCLUSION

In summary, this paper discusses about Resistive Random Access Memory along with its components such as performance, applications, advantages, disadvantages, and future exploration. On a parallel note, it is evident that there is continuous research leading to futuristic memories which drive the growth of computing power while being supported by the process industry working at the nanoscale level optimizations. This paper also addresses the gap between the current state of art memory with the next generation of memories. It is clear high performance computing demands can be achieved only through innovation in memories. At the industry level, application such as telecommunication, datacenter operations, autonomous driving, augmented reality is a function of the memory architecture being used.

REFERENCES:

1. Coughlin, Tom. “IEEE Roadmaps Guide Future Memories & Applications.” Forbes, Forbes Magazine, 29 Nov. 2018, https://www.forbes.com/sites /tomcoughlin/2018/11/29/ieee-roadmaps-guide-future-memories-applications/?sh=2bd651f9689f
2. Mark Neisser, "International Roadmap for Devices and Systems lithography roadmap," J. Micro/Nanopattern. Mats. Metro. 20(4) 044601 (4-October,2021) https://doi.org/10.1117/1.JMM.20.4.044601
3. Im, In Hyuk, et al. “Memristive Devices for New Computing Paradigms.” Advanced Intelligent Systems, vol. 2, no. 11, 2020, p. 2000105., https://doi.org/10.1002/aisy.202000105.
4. Li Y T, Long S B, Liu Q, et al. An overview of resistive random access memory devices. Chinese Sci Bull, 2011, 56: 3072–3078, doi: 10.1007/ s11434-011-4671-0.
5. Zahoor, F., Azni Zulkifli, T.Z. & Khanday, F.A. Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications. Nanoscale Res Lett 15, 90 (2020). https://doi.org/10.1186/s11671-020-03299-9
6. J. Yu, H. A. D. Nguyen, L. Xie, M. Taouil and S. Hamdioui, "Memristive devices for computation-in-memory," 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2018, pp. 1646-1651, doi: 10.23919/DATE.2018.8342278.
7. Upadhyay, Navnidhi K., et al. “Emerging Memory Devices for Neuromorphic Computing.” Advanced Materials Technologies, vol. 4, no. 4, 2019, p. 1800589., https://doi.org/10.1002/admt.201800589.
8. “Micron: About Our Innovation.” Micron Lettered in White, https://www.micron.com/about/our-innovation
9. Molas, G.; Nowak, E. Advances in Emerging Memory Technologies: From Data Storage to Artificial Intelligence. Appl. Sci. 2021, 11, 11254. https://doi.org/10.3390/app112311254

10. Slotnick, S. (2017). The Future of Memory Research. In Cognitive Neuroscience of Memory (Cambridge Fundamentals of Neuroscience in Psychology, pp. 219-237). Cambridge: Cambridge University Press. doi:10.1017/9781316026687.012
11. P.Y. Chen et al., "Compact modeling of rram devices and its applications in 1t1r and 1s1r array design", IEEE Transactions on Electron Devices, vol. 62, pp. 4022-4028, Dec 2015.
12. S. Hamdioui, "Memristor based computation-in-memory architecture for data-intensive applications", DATE'15. EDA Consortium, pp. 1718-1725, 2015.
13. Wong, HS., Salahuddin, S. Memory leads the way to better computing. Nature Nanotech 10, 191–194(2015). https://doi.org/10.1038/nnano.2015.29
14. L. Xie et al., "Scouting logic: A novel memristor-based logic design for resistive computing", VLSI (ISVLSI) 2017 IEEE Computer Society Annual Symposium on, pp. 176-181, 2017.
15. Xie, Y. Future memory and Interconnect Technologies - seal.ece.ucsb.edu. https://seal.ece.ucsb.edu/sites/default/files/publications/2013-DATE-invited-Xie.pdf