International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.4., 2015
(July-Aug)

RESEARCH ARTICLE

ISSN: 2321-7758

EVALUATING A NOVEL STT RAM AS AN ENERGY EFFICIENT MEMORY ALTERNATIVE USING READ AND WRITE OPTIMIZATION TECHNIQUE: A STATISTICAL VIEW

LAKSHAY SACHDEVA[1], GHANSHYAM [2]

[1]M.Tech Scholar (E&C + VLSI) from Suresh Gyan Vihar University

[2]Assistant professor, Gyan Vihar School of Engineering and Technology

LAKSHAY SACHDEVA

GHANSHYAM

ABSTRACT

There is growing interest in emerging non-volatile memory technologies such as Phase-Change Memory, Memristors, and Spin-Transfer Torque RAM (STT-RAM). STT-RAM, in particular, is experiencing rapid development that can be difficult for memory systems researchers to take advantage of. What is needed are techniques that enable designers to explore the potential of recent STT-RAM designs and adjust the performance without needing a detailed understanding of the physics. In this paper, we present the STT-RAM optimization System to assist memory systems researchers. After providing background on the operation of STT-RAM magnetic tunnel junctions (MTJs), we demonstrate how to optimize different published model and compare their characteristics with respect to common metrics. The high-speed switching behavior of the designs is evaluated using macro magnetic simulations. We have also added a first order model for STT-RAM memory arrays to the CACTI memory modeling tool, which we then use to evaluate the performance of latency for: (i) a write performance cache (ii) a read performance cache.

Keywords— STT-RAM, Magnetic Tunnel Junction, spin torque, spin transfer switching

I. INTRODUCTION

Spin-Torque Transfer Random Access Memory (STT- RAM) is a rising memory technology with the expectation to become a true universal memory. It have the quality of all preceding memories, this can be summarized like density of DRAM, the speed of SRAM, and the non- volatility of flash. STT-RAM process with Magnetic Tunnel Junction (MTJ) devices that can be called non volatile magnetic memory element used for storage. It uses the technology discovered recently called spin torque to switch magnetic states. In this process, the basic quantum mechanical nature of the MTJ is worked to develop a highly depth physics principle model of its spintronic operation. Special design-space techniques are used to investigate existing and proposed STT-RAM structure. Each chip has a more than two times greater memory density and a more than 10 times greater read/write performance when it is compare to other architecture.

Theoretical and observed scaling trends show ash-like densities, with SRAM-equivalent access times, while using 10 times less energy in more advanced technology nodes(below 32nm). STT-RAM is an emerging memory technology that exploits the recently discovered phenomena of spin-torque transfer (STT) in MTJs. This chapter provides a brief motivation for STT-RAM, as well as outlines the rest of the thesis. STT-RAM has the main strength is that it do not need any kind of charging to maintain its

orientation that's why they are counted in the types of non volatile memory [11].

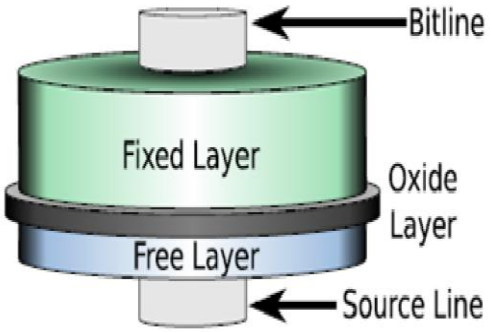

Figure 1.1 Structure of MTJ

II. Background and Related Works

STT RAM uses different process named spin transfer torque effect to switch the free layer from its one state to another which requires electric current pass directly to MTJ. This switching takes place accordingly to a stochastic process that is thermally controlled and will studied on later Section.

We will discuss the spintronic operation later. But for beginning we can say that MTJ is the pair of ferromagnets and a thin insulating layer is present between them. The two possible states are used; the parallel combination of the two layers (Fig.1.3 (a) and the anti parallel combination (Fig.1.3 (b)). The parallel configuration gives a low resistive state (RP), while the Antiparallel configuration gives a high resistive state (RAP).

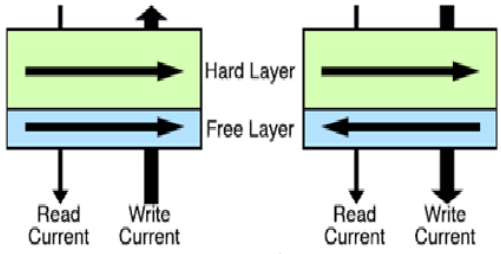

Figure 1.2: MTJ ferromagnetic layers in parallel and anti-parallel configurations.

This effect is important because of this effect the basic concept of advancement of RAM is possible.STT RAM has the ability to be scaled down below 65nm with the help of reducing the writing current with a hundred factors. Due to very lesser in size and very large write current, Power density in write operation of MTJ is very high. Simulation also proved that consecutive write Operation produces a 9-15°C increment in the temperature. A high No. of writes followed by a read leads to disturbed sensing margin. Self induced Heating is destroyed in the writing mechanism introduced by thermal Assisted Switching MRAM's. However, In STT-MRAMs lower write current are generally used to avoid self induced heating

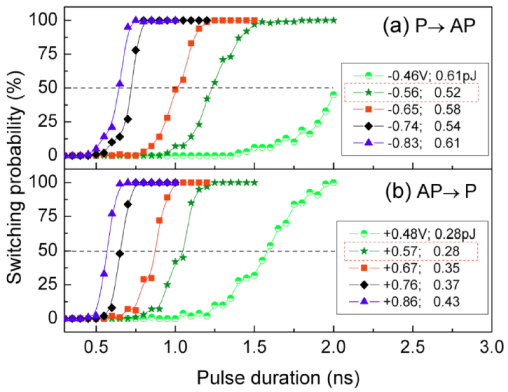

Figure 1.3: Graph Plotted in between Switching probability and Pulse duration

Advances and Future Prospects of Spin-Transfer Torque Random Access Memory/ IEEE transactions on magnetic, Vol. 46, No. 6, June 2010; E. Chen, D. Apalkov, Z. Diao worked on the process of Spin Transfer Torque (STT) switching consist of spin polarized current that is directly used for the purpose in switching the magnetization of a nanomagnet. At origin it was explained and demonstrated in the year 1996 and these theory uses the concept of metallic spin valve thin films with critical switching current density JCo>10^7-10^8 A/cm^2 which is not practically fitted for any memory device means it has very high value[5]. A organization named Grandis Reported the very new technology in 2006 related with STT RAM its named was Magnetic Tunnel Junction (MTJ) which have JCo less than 1.12 X 10^7 A/cm^2.Therefore we can say that MTJ shows higher resistance as respect to the previous technology (spin valve). It is compatible with semiconductor CMOS transistor and MTJ. It have one more advantage that it have higher tunnel its free layer is present in between the two tunnel barriers and in this paper they proved that the value of JCo is considerably reduced to the value 1-2 MA/cm^2 range in 90nm x 180nm ellipse devices. In the anti parallel state to parallel state shifting, performance of BMTJ exceeds DMTJ. And it Parallel to anti parallel

switching, BMTJ switches has lesser performance than DMTJ. The TMR of BMTJ devices is around 100-115% and 84% for DMTJ. Analysis of STT-RAM Cell Design with Multiple MTJs Per Access/ IEEE transaction 2011; Henry Park, Richard Dorrance, Amr Amin worked on the cell design process. This paper explained the connection of access transistor with MTJ cache unit. It described the structure of a cell which is present in common with all MTJ memory structure. Achievability and disadvantages of the cell structure was also explained at very high depth in the concern of reading and writing circuitry. The practical and simulation result shown in this paper is a proof that in MTJ cell structure a small amount of sharing is possible and in this paper it was maintained that MTJ must have less read and write current capacity and after the result shown we can easily reach to a conclusion that performance of the multi cell of MTJ that are connected with a single access transistor depends on the characteristics of MTJ. The area of cell was also reduced on the reduction of access device [6].The reduction in device can be calculated by maintaining the bit-line voltages to satisfy the targets. The highest parasitic current of non accessed MTJ is tested under various case conditions. This paper focused the affect of multiple MTJs which remains common in the single access transistor. There are some problem arises in the read and write circuitry stability because of the resistive nature of the MTJ. There result is very efficient in the memory array method but this is applicable to the small no. of MTJs per column size. Low-Leakage SRAM, Low Write-Energy STT-RAM, and Refresh-Optimized eDRAM/ DAC 2012, June 11-14, 2013 in Porto Allege Mu-Tien Chang, Paul Rosenfeld, Shih-Lien Lu, and Bruce Jacob has proposed the work for the purpose in comparison of caches made with STT RAM,SRAM and eDRAM. Earlier that it was found that they have long and large write operation, for this purpose a counter controlled dynamic refresh method is proposed. It's important work is used to validate the time delay and save refresh energy more than 80% compared to previous technology and in cache formation it reduces 30-70% of the total energy compared to previous works [17]. They worked on the STT RAM Write performance and proposed the result for write performance and energy consumption by 18% and 60% respectively. Variable-Energy Write STT-RAM Architecture with Bit-Wise Write-Completion Monitoring Tinaho Zheng, Jaeyoung Park, Michael Orshansky and Mattan Erezto/ IEEE transaction 2013; worked on the procedure for moving away the traditional worst-case approach, they made per-cell write process that is continuously monitored and get termination as soon as each cell's state matched with the written state. They proved the duration of average write is always far smaller than the duration of worst case, the architecture they proposed has significantly reduced average write energy by large factor [16]. They developed a small circuit for rapidly changing state detection and bit-line shutdown and equalize it, using a compact STT-RAM model considering an implementation in a 16nm technology node. There analysis indicates that at the required write-error rate the proposed technique reduces write energy by 87.3%-99:5% depending on the right direction and on average achieves 96:5% write energy saving in comparison to previous design. Low-Current Probabilistic Writes for Power-Efficient STT-RAM Caches/ IEEE transaction 2013; Nikolaos Strikos, Vasileios Kontorinis and Xiangyu Dong, performed the work in the field of STT RAM caches. This paper discussed the problem by giving a new term named low current probabilistic write that is written as LPCW. However after the reduction it felt less successes in bit write operations, they made and explained a different method for the reduction in consumed power in comparison to the MRAM architecture. In beginning, they explained their techniques and procedure after the application of LPCW to their multi thread circuits. In order to their calculation, they worked on the CACTI and SMTSIM simulator. They have modeled a single system made of only 1 core consisting of 3 levels of hybrid cache and private L2 MRAM cache and L2 SRAM cache. There execution starts with 4 billion instructions warm the model for 60 million instructions and then again stimulate for 2 billion instructions.

III. Design Consideration

Our designed work is model is based on the Non Uniform Cache Access (NUCA).This model is based on the following concept that is explained as follow:

Bank- it is kind of data memory arrangement that involves combination of data and a tag array. A STT RAM cache is distributed into many banks and their bandwidth is assumed as equal therefore they can be used at the same moment. As the cache model will change then there adopting network topology will also get changed.

Sub-Arrays – A tag or big array is sub divided into number of sub arrays. This structure is preferred because it will reduce the delay of wordline and bitline its operation is not similar as banks because they work only one at a time.

Mat - A matrix of 2x2 sub array that involves a single predecoder. Its comprehensive search starts from a minimum element of at least one mat.

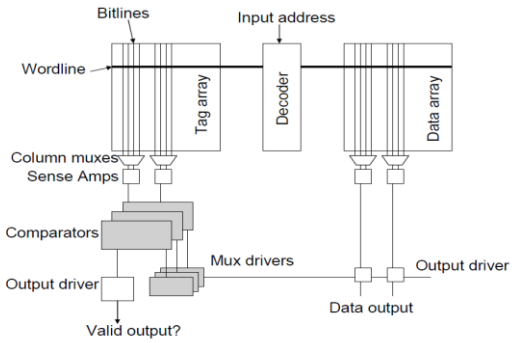

Figure 1.4: Proposed Modeling of STT RAM using CACTI 6.0.

Sub-bank - In a cache designing, a cache block is distributed into many sub-arrays to increase the trustworthiness of a cache. CACTI takes all cache blocks in a distributed manner that is it takes cache divided into row made up of mats. Then row address can also be fetched out on the basis of block address. Each row in an array is also referred as sub-bank. Then multiple tags come under in the working process which has work to read out the tag array in comparison to the detected input address. If one of the paths of the array contains the data element then multiplexor drives the data element with the help of comparator logic to read out port of data array which will be return back to the requesting processor. The tag array and data array are very large in size therefore it is not sufficient to implement it in a single unit as a very large structure. Hence CACTI divide each storage array in the horizontal and vertical dimension to produce very smaller sub arrays which also help in reduction of wordline and bitline delays.

The bitline is divided into distinct segments named Ndbl; the bitline is divided into distinct segments named Ndwl and further. Every sub-array has its separate decoder but some common is pre-decoding is desired for the purpose to find the request of right sub array.

CACTI also made a comprehensive search in different sub arrays counts for the calculation of total time delay cache arrangement. Bottom sub arrays are actually the mirror images of subarrays present in top direction and subarrays present in left direction are the mirror images of subarrays present in right side. All address, data in, data out signal are taken as to move the mat in the middle direction

A mat is arrangement of 4 subarrays and also includes one predecoding/decoding logic which is present at the mid of the mat. The predecoding circuit is present for all subarrays. Here some other subarrays are also present and they are mirror images of previous sub arrays.

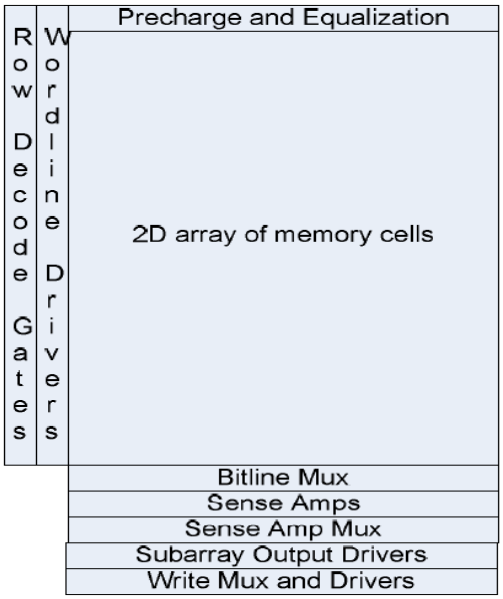

Figure1.5: Representation of memory cell

IV. Optimization Result

The peripheral organization is largely depend between reads and writes, the write latency can be calculated equals to the read latency in addition with the MTJ write time. The horizontal dashed lines displaying the latency for a SRAM cache

with High Performance, and the continuous vertical line at 1.35 ns showing the STT-RAM design calculation that equates the value of SRAM read latency at the time while we are minimizing the write latency. By the increment of the Value of MTJ write time beyond this threshold value will give us values of faster reads than SRAM at the cost of even the value of slower writes, while reducing this gives us very faster writes at the charge of slower reads.

Thus, to match the performance of SRAM in Read cycle circuiting design, it requires only the cycle-based read latency must remain the same. Here below, two methods are explained that meet or better the read performance of SRAM while incrementing the write performance. However, the method described below is applicable to any clock speed.

A. Proposed Write Optimization Method:

The write performance of STT RAM is first incremented while matching the SRAM read performance (with the help of cycle-based latencies). The first motive is to increment the value of read latency that will reduce the write latency itself, without deviate the reads performance, as showing in Figure 4.6 The vertical dashed line show us the Previous design choice while the vertical continuous line shows us the design chosen by our first step, with the continuous arrow representing us the traveling direction and we named this as second step, in which the write time of MTJ was further increased in the search of the Pareto optimal point.

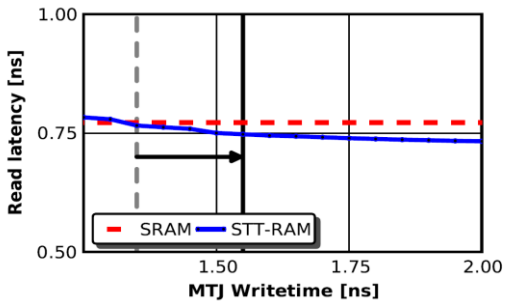

Figure 1.6 Result of Read Latency using the Proposed Write Optimization Method

Thus we have achieved the reduction of the read energy by 6% and the write energy by 4% while without changing the write and Read performance. Here, the vertical dashed lines representing the previous designs, while the continuous vertical line representing us the write-optimized design that is after the Application of the optimization chosen by following arrow., whilst maintaining the exactly equal no. of effective read latency of 4 cycles.

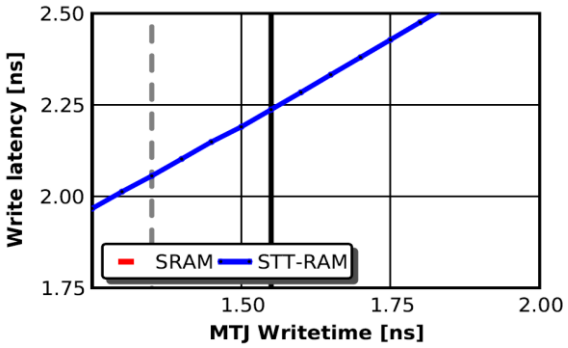

Figure 1.7: Result of write Latency using the Proposed Write Optimization Method

B. Proposed Read Optimization Method:

The write optimization method of the last page achieved to reduce the non matching behavior in moving to STT-RAM by relating the SRAM as near as possible. Secondly, it is a chance that we desire sometimes to provide better read performance by eliminating the write performance. As previous, the dashed vertical line represent us basic design selected, and the arrow represent us the traveling direction for the purpose of optimization method, and the continuous vertical line is the transitional design marked in beginning step. Even if more decrement to the read latency of MTJ are possible, they will not involve in any extra growth for the read performance of modeled here.

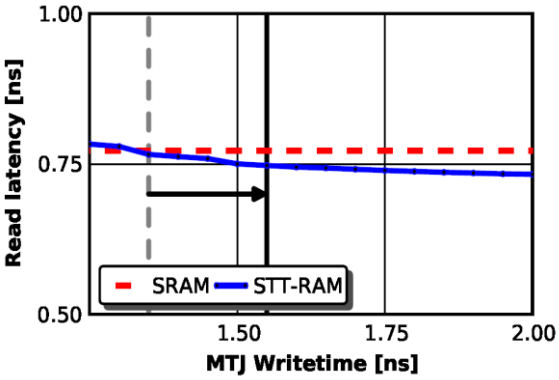

Figure 1.8: Result of Read Latency using the Proposed Read Optimization Method

The next step is similar as performed in write optimization method that is it will continue to decrement the write time of MTJ in the search of Pareto Optimal design point. In this example, no

further reductions are possible, which is not represented by any kind of arrow. This design method has exactly one less cycle of read latency in MTJ than the basic previous write optimized method and therefore we can say that in reality it takes the complete nine cycles of write latency of previous design.

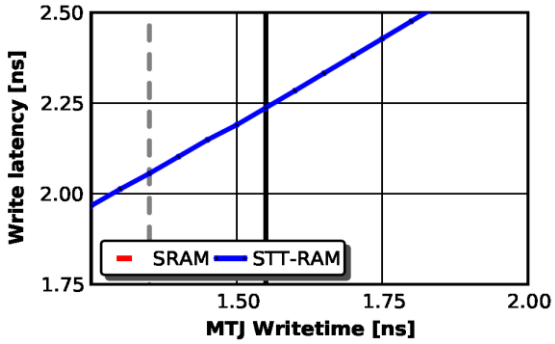

Figure 1.9: Result of write Latency using the Proposed Read Optimization Method

Hence we can say it takes three cycles more than the write optimized design method. Thus, after representing the waveform results of the proposed method and the results of the optimization between proposed and previous Optimization, let's study the differences between the previous method and the proposed method.

V. CONCLUSION

In this work, we have studied the performance of the STT RAM for the cache formation. In this work the conventional and the proposed STT RAM have been designed using e Verification language. Looking again at high-performance caches, the cause of the low performance of STT-RAM (both reads and writes) is the high write energy. The high write currents required for fast switching require larger access transistors within the cell and larger peripheral circuitry without, both of which lower the density, increase the energy consumption, and reduce performance the latency of read and write cycle of the proposed STT RAM is more efficient in comparison to the conventional STT RAM. We have I MB memory array encoder in terms of its Write and Read latency, which deals with the new optimization technique. Modified STT RAM increases the speed of operation by reducing latency cycle. The proposed STT RAM provides improved write and read latency. Here in this proposed design following few terms have been observed:

- Earlier the obtained result for Write and Read Latency was 9 and 6 cycles respectively and for our proposed design it comes for Write and Read Latency is 6 and 3 cycles respectively. The write latency is reduced by 3 cycles i.e. 33.33%. And the read latency is reduced by 1 cycle i.e. 25.00%.

REFERENCES

[1]. A. Agarwal, B.C. Paul, S. Mukhopadhyay, and K. Roy "Process variation in embedded memories failure analysis and variation aware architecture" Solid-State Circuits, IEEE Journal of, 40(9):1804–1814, 2005.

[2]. Xiaobin Wang, Yuankai Zheng, Haiwen Xi, and Dimitar Dimitrov "Thermal fluctuation effects on spin torque induced switching: Mean and Variations" Journal of Applied Physics, 103(3):034507–034507, 2008

[3]. Ashish K Singh, Ku He, Constantine Caramanis, and Michael Orshansky "Mitigation of intra-array sram variability using adaptive voltage Architecture" In ICCAD, pages 637–644 ACM, 2009.

[4]. DC Worledge, G Hu, PL Trouilloud, DW Abraham, S Brown, MC Gaidis, J Nowak, EJ O'Sullivan, RP Robertazzi, JZ Sun, et al. Switching distributions and write reliability of perpendicular spin torque mram" In IEDM, 2010.

[5]. E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S. A. Wolf, A. W. Ghosh, J.W. Lu, S. J. Poon, M. Stan, W. H. Butler, S. Gupta, C. K. A. Mewes, Tim Mewes, and P. B. Visscher, "Advances and Future Prospects of Spin-Transfer Torque Random Access Memory" IEEE transaction on magnetic, vol. 46, no. 6, June 2010

[6]. Henry Park, Richard Dorrance, Amr Amin, Fengbo Ren, Dejan Markovi, and C.K. Ken Yang, "Analysis of STT-RAM Cell Design with Multiple MTJs per Access" on IEEE 2011

[7]. Zhenyu Sun, Weng-Fai Wong, Xiaochun Zhu "Multi Retention Level STT-RAM Cache Designs with a Dynamic Refresh Scheme"

MICRO'11, December 3-7, 2011, Porto Alegre, Brazil Copyright 2011 ACM 978-1-4503-1053-6/11/12

[8]. Asit K. Mishra, Xiangyu Dong, Guangyu Sun,Yuan Xie, N. Vijaykrishnan, Chita R. Das "Architecting On-Chip Interconnects for Stacked 3D STT-RAM Caches in CMPs" ISCA'11, June 4–8, 2011, San Jose, California, USA. Copyright 2011 ACM 978-1-4503-0472-6/11/06

[9]. Woojin Kim, JH Jeong, Y Kim, WC Lim, JH Kim, JH Park, HJ Shin, YS Park, KS Kim, SH Park, et al. Extended scalability of perpendicular stt-mram towards sub-20nm mtj node. In IEDM, 2011

[10]. W. Xu, H. Sun, X. Wang, Y. Chen, and T. Zhang, "Design of last-level on-chip cache using spin-torque transfer RAMS (STT RAM)" on IEEE TVLSI, 2011

[11]. R. Gabrys, E. Yaakobi, L. Grupp, S. Swanson, and L. Dolecek "Tackling Intracell variability in tlc flash through tensor product codes" In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium On, pages 1000–1004, 2012.

[12]. Yue Zhang, Weisheng Zhao, Yahya Lakys, J Klein, Joo-Von Kim, Dafin´e Ravelosona, and Claude Chappert "Compact modeling of perpendicular anisotropy cofeb/mgo magnetic tunnel junctions" IEEE Trans. Electron Devices, 59(3):819–826, 2012.

[13]. Adwait Jog, Asit k.Mishra, Cong Xu, Yuan Xie "Cache revive: Architecting volatile STT-RAM caches for enhanced performance in CMPs" DAC 2012, June 3-7, 2012, San Francisco, California, USA;

[14]. Nikolaos Strikos, Vasileios Kontorinis, Xiangyu Dong, Houman Homayoun, Dean Tullsen, "Low-Current Probabilistic Writes for Power-Efficient STT-RAM Caches" on IEEE transaction 2013