

Multi-Bit Read and Write Methodologies for Diode-MTJ Crossbar Array

Mohammad Nasim Imtiaz Khan and Swaroop Ghosh

School of EECS, The Pennsylvania State University, University Park, PA, USA

Email: muk392@psu.edu and szg212@psu.edu

Abstract— Crossbar arrays using emerging Non-Volatile Memory (NVM) technologies offer high density, fast access speed and low-power. However, the bandwidth of the crossbar is limited to single-bit read/write per access to avoid the selection of undesirable bits. In this work, we propose a technique to perform multi-bit read and write in a diode-MTJ (Magnetic Tunnel Junction) crossbar array. The simulation shows that the biasing voltage of half-selected cells can be adjusted to improve the sense margin during read which in turn, reduces the sneak path through the half-selected cells. Results indicate biasing the half-selected cells by 700mV can enable reading as much as 512bits while sustaining 512x512 crossbar with 2.04 years retention. During write operation, the half-selected cells are biased with a pulse voltage source in addition to V/2 scheme which increases the write latency of these cells and enables writing 2 bits while keeping the half-selected bits undisturbed. The 2bit writing requires pulsing by 50mV to optimize energy.

Keywords— MTJ, STTRAM, crossbar, multi-bit read, multi-bit write, sneak path current.

I. INTRODUCTION

Spin Transfer Torque RAM (STTRAM) with Magnetic Tunnel Junction (MTJ) as storage element offers a multitude of advantages including non-volatility, speed, scalability, endurance and low-power. However, the footprint is limited by three-terminal transistor-based selectors. Selector Diodes (SD) such as Mixed Ionic Electronic Conduction (MIEC) [1], Metal-Insulator-Metal (MIM) tunneling [2], oxide heterojunction p-n, and Schottky diodes [3] for Resistive RAM (RRAM) have been extensively studied [4]. The integration of SD and RRAM has been attempted due to material compatibility and simplistic structure. Unlike RRAM, the MTJ (storage component of STTRAM) does not need electroforming which simplifies the integration of MTJ with SD. However, MTJ faces new challenges such as low

Tunnel Magneto-Resistance (TMR), and, read/write and retention sensitivity to the noise and variations. In [5], a Metal-Insulator-Insulator-Metal (MIIM) SD is proposed to enable 3D integration with MTJ. Although SD increases the switching voltage which is undesirable for cache, it allows high-density and scalability by enabling 3D stacking for new applications e.g., Internet-of-Things (IoTs).

Several emerging Non-Volatile Memory (NVM) technologies have been proposed for crossbar array [6-9]. Techniques such as V/2 and V/3 read/write have been studied to reduce sneak path current in crossbar arrays [10]. However, the crossbar array is bit addressable as shown in Fig. 1 (a) to avoid disturbing the half-selected and unselected cells which limits the bandwidth. This means that for a n -bit cache line, n number of crossbars needs to be activated. Therefore, all the half-selected and unselected cells in all n crossbars incur sneak path current. This increases the power dissipation. Furthermore, peripherals for all n crossbars also increases the power consumption. Multi-bit write for RRAM crossbar has been proposed in [7] [11] where the '1's are written in 1st cycle and '0's are written in the 2nd cycle. This, in turn, increases the write latency. Multi-level cell (MLC) STTRAM has been proposed that can store 2 bit per cell and thus ensure multi-bit read-write [12] [13]. However, the write operation becomes complicated and the read operation requires 4 Sense Amplifiers (SAs) per cell to distinguish 4 states of 2 bits. This increases the circuit design complexity and incurs significant area overhead.

In this paper, we propose techniques to perform multi-bit read and write operation on the diode-MTJ crossbar (Fig. 1(b)). The MIIM diode proposed in [5] is employed as a selector device and optimized to make trade-off among energy, speed, retention time and sense margin. By performing 2bit read and write operation per cycle, we reduce the number of active crossbars per access by half. The read operation exploits the effect of biasing voltage on the half-selected cells to increase the sense margin. During read operation, instead of V/2 scheme, $V/2 + \Delta$ (where Δ is a positive value) is proposed which reduces the sneak path current and improves the sense margin. The write operation exploits the nonlinear dependence of write latency of MTJ on write voltage to write 2bits while keeping the half-selected bits undisturbed. Instead of biasing the half-selected cells with a dc voltage, a pulse voltage source is used which increases the effective write latency of the half-selected cells and gives enough time to write two selected bits simultaneously. The

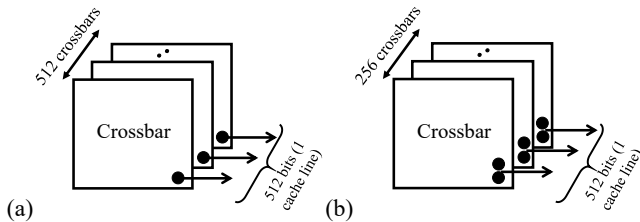


Fig. 1: Crossbar arrays: (a) conventional bit addressable; and, (b) proposed multi-bit array.

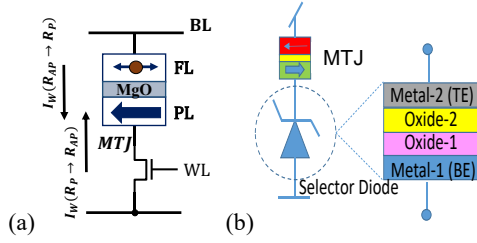


Fig. 2 Schematic of (a) conventional STTMRAM, and (b) MIIM diode-MTJ.

proposed techniques are also applicable to other crossbar arrays such as RRAM and Phase Change Memory (PCM).

In summary, we make the following contributions:

- (a) We propose voltage biasing to enable multi-bit read operation and perform detailed analysis to support large crossbar size. Our analysis shows that a biasing of 0.7V can enable 512bit read in 512x512 crossbar;
- (b) We propose a voltage pulsing to enable multi-bit write operation and perform detailed analysis;
- (c) We propose SD optimization to lower the energy-overhead of multi-bit write;

The paper is organized as follows: Section II provides an overview of MTJ and the MIIM SD [5] that is employed in this study; Section III describes the proposed multi-bit read operation; Section IV presents the multi-bit write operation; Finally, Section V draws the conclusion.

II. BACKGROUND ON MTJ AND MIIM DIODE

In this section, we present a discussion on MTJ and the MIIM diode used in this work.

A. MTJ

Conventional STTMRAM contains one NMOS as an access transistor and one MTJ as a storage element (Fig. 2(a)). MTJ contains two ferromagnetic layers known as Free Layer (FL) and Pinned Layer (PL). The equivalent resistance of the MTJ is low (R_p) if FL and PL magnetic orientations are parallel (P) to each other. The equivalent resistance of the MTJ is high (R_{ap}) if FL and PL magnetic orientations are antiparallel (AP) to each other. Typically, the high resistance state is considered as data '1' and the low resistance state is considered as data '0'. During a write operation, FL magnetic orientation can be toggled from the P \rightarrow AP state (data 0 \rightarrow 1) (or vice versa) by passing the write current ($>$ critical current) from Sourceline (SL) to Bitline (BL) (or vice versa). This means that MTJ requires bidirectional current for switching. If the access transistor is replaced with a two-dimensional SD (to implement a crossbar array [14]), the SD also has to be bidirectional.

B. MIIM Diode

A bidirectional MIIM diode has been studied in [5] (Fig. 2(b)). It turns ON when applied bias is higher than V_{T+} (V_{T-}) in positive (negative) bias and incurs low-reverse saturation current when applied bias is less than $|V_T|$. The operation of

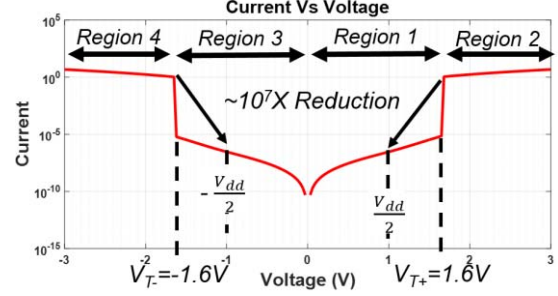


Fig. 3 I-V characteristics of MIIM SD used in this work.

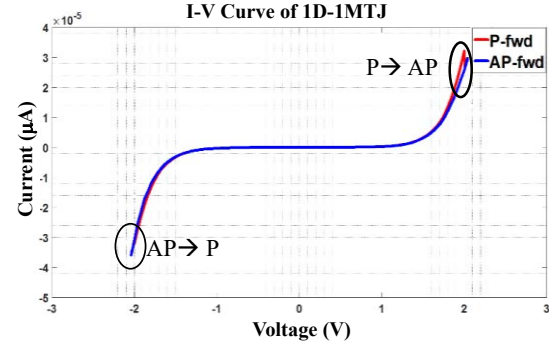


Fig. 4: 1D-1MTJ I-V curve for MTJ ($R_p = 6K$, $R_{ap} = 12K$)

the MIIM diode can be explained as follows: In forward bias (i.e. positive voltage on TE) the electron will experience the thickness of oxide-1 resulting in Direct Tunneling (DT) and when the bias voltage exceeds the positive threshold voltage (V_{T+}) Fowler-Nordheim (FN) current will dominate through the triangular barrier due to the lower barrier height of BE on HfO_2 (Φ_B). This will turn ON the device at V_{T+} .

The work functions of the metal-1 and metal-2, and the bandgap and electron affinities of oxide-1/oxide-2 can be selected to achieve the desired I-V characteristics. For example, Fig. 3 shows the I-V plot for TiN/TaOx/MgO/TiN MIIM diode. This is simulated using the sum of DT and FN currents equations shown below [5]:

$$J_{FN} = A E_{ox}^2 \exp\left(-\frac{B}{E_{ox}}\right) \quad (1)$$

$$A = 1.54e - 6 \left(\frac{1}{m_{ox}/m_e}\right)^{1/2} \left(\frac{1}{\Phi_B}\right) \quad (2)$$

$$B = 6.8327e7 (m_{ox}/m_e)^{1/2} (\Phi_B)^{3/2} \quad (3)$$

$$J_{DT} = A E_{ox}^2 \exp\left[\frac{-B(1 - (1 - \frac{V}{\Phi_B})^{3/2})}{E_{ox}}\right] \quad (4)$$

$$J_{total} = J_{FN} + J_{DT} \quad (5)$$

$$V_T = (\Phi_{TE-MgO} - \Phi_{MgO-HfO2}) \left[\frac{\epsilon_{MgO} t_{HfO2}}{\epsilon_{HfO2} t_{MgO}} + 1 \right] \quad (6)$$

where J_{FN} and J_{DT} are the Fowler-Nordheim and direct tunneling current densities respectively. A and B are diode constants, E_{ox} is the electric field, Φ is the barrier height, ϵ is the di-electric constant, t_{ox} is the oxide thickness and V_T is the threshold voltage.

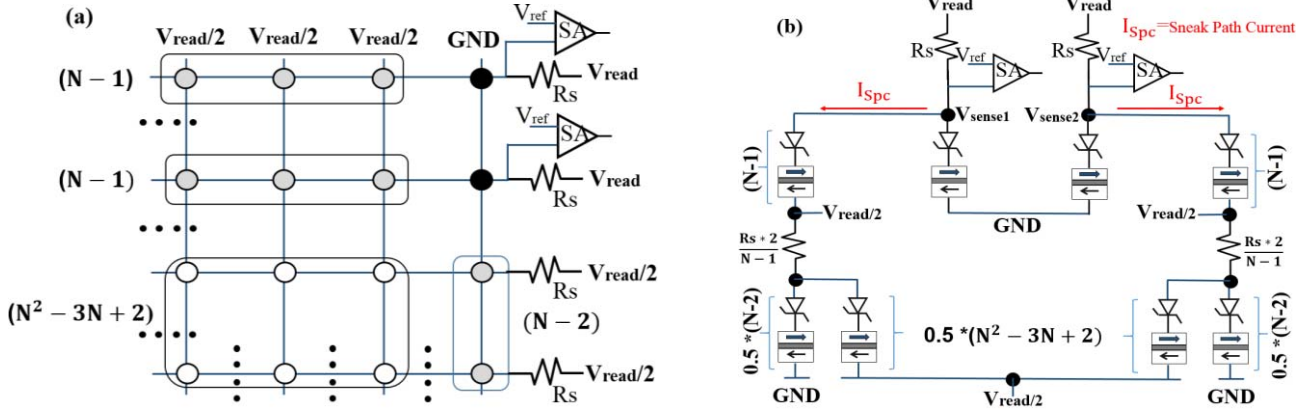


Fig. 5 (a) Proposed circuit for two bit read operation (shown with dark circles); (b) the sneak path equivalent circuit for two bit read. The sense voltages (V_{sense1} & V_{sense2}) are a function of the two MJTs that are being read and the sneak path current through the half-selected bits and unselected bits.

The metal oxide to metal (m_{ox}/m_0) ratio of 0.26 is used in our simulations. Electron affinity of TaOx and MgO of 3.5 eV, and 2.3eV respectively [7], and, the dielectric constant value of TaOx and MgO of 22, and 10 respectively, are used and the work-function of TiN of 4.0 eV is also incorporated. The I-V plot is divided into four regions. Region-I and Region-III are generated assuming tunneling through 1.5nm TaOx, Region-II and Region-IV are generated assuming tunneling through 0.1nm MgO. This unique characteristic is difficult to achieve using a single insulator with TE and BE of different work functions.

In this work, we have assumed an ideal MIIM diode i.e. without bulk/interfacial defects and MTJ with TMR = 2. Fig. 4 shows the I-V curve of 1D-1MTJ for $R_p/R_{\text{ap}} = 6/12\text{K}\Omega$. It is evident that the cell requires a write voltage (V_{write}) of 2.1V.

III. MULTI-BIT READ SCHEME

In this section, we present the proposed circuit and the methodology to read multi-bit data simultaneously.

C. Design and Analysis

The proposed circuit for 2bit read operation (Fig. 5 (a)) shows that when two bits are selected for read in $N \times N$ array, $2(N-1) + (N-2)$ bits are half-selected and $(N^2 - 3N + 2)$ bits are unselected. The design faces a few challenges during read operation. First, while reading two bits, the corresponding bits should be selected without disturbing other bits (half-selected and unselected bits). Even the two selected bits can be

disturbed significantly if the read voltage (therefore, read current) is too high. Note that each SA has to distinguish 2 states (data 0 and 1) for each memory cell. Therefore, these two states per memory cell should have a maximum margin between them for robust sensing. Second, with the increase of crossbar array size, the sense margin shrinks due to sneak path current. Therefore, sneak path current needs to be minimized for a larger array for a robust sense margin.

The voltage at V_{sense1} and V_{sense2} in Fig. 5 (b) is a function of TMR of the selected MJT and the sneak path current through the half-selected and unselected bits. Fig. 6 shows the sense margin with MTJ resistance (R_p). Note that sense margin increases with R_p whereas, higher R_p also requires higher write current i.e. higher write voltage. We have chosen $R_p = 6\text{K}\Omega$ for optimum sense margin and V_{write} .

The sense resistance (R_s) also plays a vital role in determining the sense margin. Fig. 7 shows the sense margin with respect to R_s for 50×50 crossbar array. The sense margin increases till $R_s = 18\text{K}\Omega$ but decreases afterward. Therefore, the design window for R_s is $8\text{K}\Omega$ to $18\text{K}\Omega$. Higher R_s ensures lower read current and lower sneak path current. Therefore, in this work, we have chosen $R_s = 16\text{K}\Omega$.

D. Read and Retention

As mentioned earlier, the sneak path current is a design limitation for the array size increment. To incorporate all the possible paths for sneak current and simulate the read operation for two bits in $N \times N$ crossbar array, an equivalent

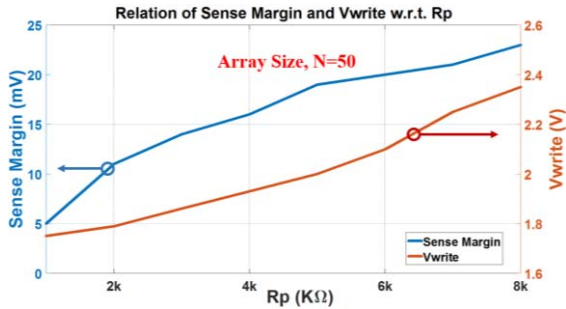


Fig. 6 Relation of 2-bit sense margin and V_{write} vs R_p for $N = 50$.

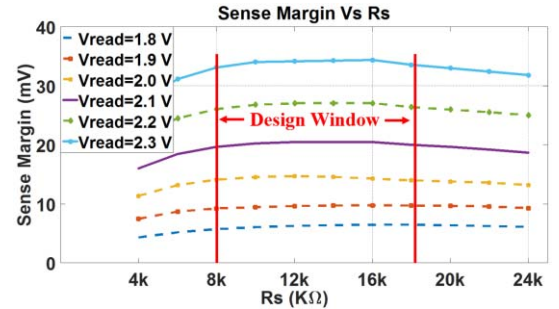


Fig. 7 Sense margin of 2-bit vs R_s for array size, $N = 50$.

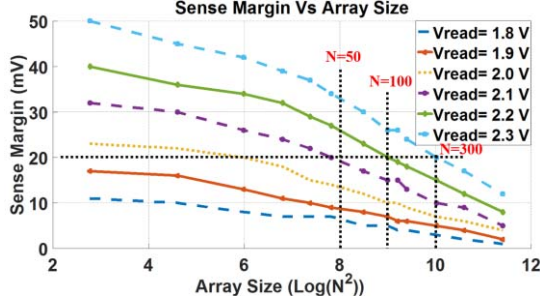


Fig. 8 Sense margin of 2-bit vs array size with $R_s = 16K\Omega$.

circuit has been developed which is shown in Fig. 5(b). For calculating the worst-case sense margin, all the MTJs in the sneak path has been considered with the low resistance (R_p).

The circuit is simulated for different V_{read} values and crossbar array sizes. It is evident from simulation result (Fig. 8) that 20mV sense margin can be achieved for array size $N = 50$ @ $V_{read} = 2.1V$, $N = 100$ @ $V_{read} = 2.2V$ and $N = 300$ @ $V_{read} = 2.3V$. The retention time for selected cells, half selected cells and unselected cells have been presented in Fig. 9. Although higher V_{read} gives a better sense margin for a larger array, the retention time goes down with V_{read} . The thermal energy of the MTJ used in the analysis is $60k_bT$, where k_b is the Boltzmann constant and T is the absolute temperature. The volume of the MTJ is assumed as $2e^{-18} \text{ cm}^3$. In this work, we have chosen $V_{read} = 2.1V$. Although V_{read} is the same as V_{write} , the selected cells get voltage below 1.9V due to drop across R_s and MIIM diode. Therefore, they are not written. At $V_{read} =$

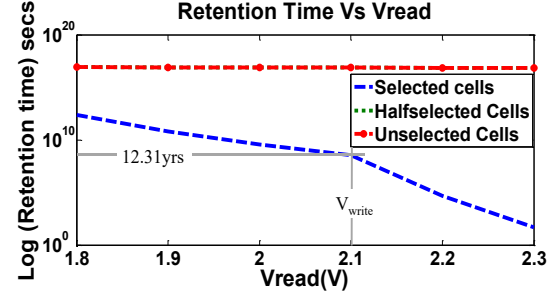


Fig. 9 Retention time of 2-bit vs V_{read} for $R_s = 16K\Omega$ and $N = 50$.

2.1V, the selected cells have ~ 12.31 years of retention time.

The sense margin needs to be improved to increase the crossbar array size and perform read operation smoothly. We propose to use $V_{read}/2 + \Delta$ instead of $V_{read}/2$ to bias all the unselected BLs and Wordlines (WLs), and, improve the sense margin, where Δ is a small positive voltage. Fig. 10 represents an equivalent circuit for n-bit read operation in $N \times N$ array. The current I_h lowers the sense margin for larger array. If Δ is increased the current I_h decreases and the sense margin improves. However, increasing Δ will in turn increase I_{R_s} . Therefore, some half-selected cells will have improved retention time whereas some other half-selected cells will have lower retention time. The retention time of selected and half selected and unselected cells are presented in Fig. 11. Note that retention time for the selected cells has decreased for higher Δ . However, the retention time for $(N-1)$ half-selected has improved and the retention for $(N-2)$ half-selected has decreased. Fig. 12 shows that as Δ increases, the sense margin improves and reaches a saturated value of 32 mV which is the sense margin for 2 bit read in a 4×4 array. The retention time at selected $\Delta = 0.7$ is 2.04 years.

E. Multi-bit Read

Fig. 13 shows the sense margin for multi-bit read operation. For a $N \times N$ array if any n-bits are read (where $n = 2, 3, \dots, N$), the sense margin remains the same. As shown in Fig. 10, the sense margin is mainly affected by the $(N-1)$ half-selected cells that are in the same row of the selected cell in an $N \times N$ crossbar array. The other half-selected and unselected are equally distributed among the selected cells (Fig. 10). With higher Δ , I_h decreases and I_{R_s} increases and since $I_{Rh} = I_{R_s} + I_h$, the additional current is adjusted by I_{Rh} . If N bits are read from $N \times N$ array and $V_{read}/2 + \Delta$ are applied to all BLs and

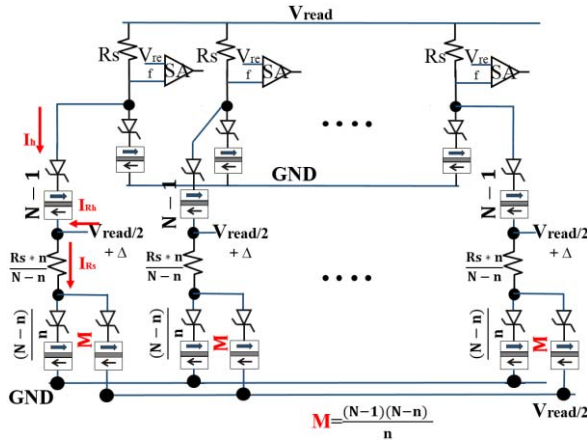


Fig. 10 Equivalent circuit for n-bit read operation for a $N \times N$ array.

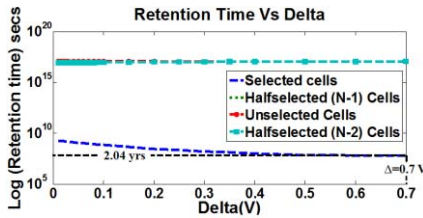


Fig. 11 Retention time vs delta with $V_{read} = 2.1V$ and $N = 50$.

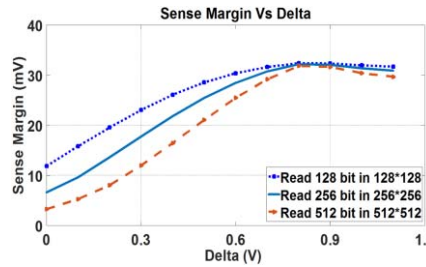


Fig. 12 Sense margin vs delta with $V_{read} = 2.1V$.

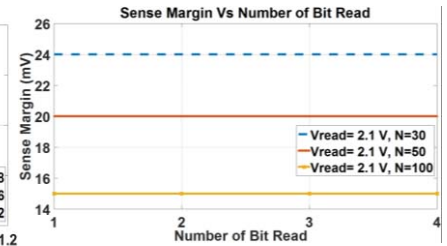


Fig. 13 Sense margin for multiple bit read operation.

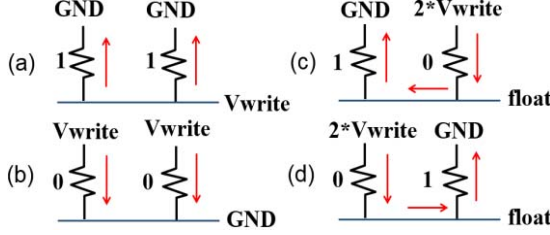


Fig. 14 Two bit writing simultaneously for four combinations.

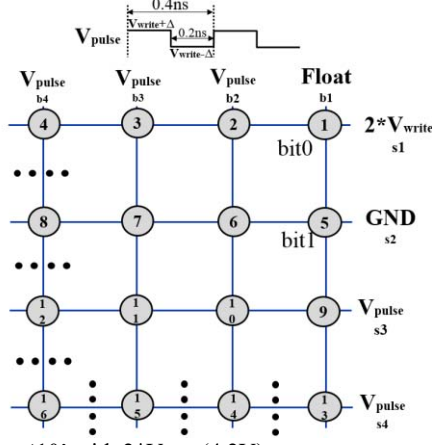


Fig. 15 Writing '10' with $2*V_{write}$ (4.2V).

WLs, the sense margin can be increased up to 32mV (for $\Delta = 0.7V$) and the retention of (N-1) half-selected will also be maximized (Fig. 10). From the analysis, we conclude that it is possible to sustain as much as 512 bit read with $\Delta = 0.7V$ in 512x512 crossbar (Fig. 12) without disturbing any cells.

IV. MULTI-BIT WRITE SCHEME

In this section, we discuss the methodology to write 2bits of memory simultaneously. Writing 2bits together can be categorized into two groups: (a) writing identical bits i.e. '11' or '00'; and, (b) writing complementary bits i.e. '10' or '01'. Both of the cases are explained in this section.

A. Design and Analysis

For writing '11' (Fig. 14 (a)), we apply V_{write} to the BL and ground the WL of the selected bits. For writing '00' (Fig. 14 (b)), we apply V_{write} to the WLs and ground the corresponding BLs. For ensuring the other bits are not disturbed, $V_{write}/2$ is applied to all other WLs and BLs. Therefore, writing '00' and '11' is straight forward. However, the challenge arises for writing complementary bits (Fig. 14 (c) and Fig. 14 (d)) when 2 SDs and 2 MTJs are in series. The write voltage should be $2V_{write}$ to write both bits while the selected BL is floated. The unselected WLs and BLs should be V_{write} . However, this will end up writing the half-selected bits.

From Fig. 15 it is evident that applying $V_{write} + \Delta$ and $V_{write} - \Delta$ on unselected lines will prevent one set of half-selected bits (Cell 2, 3, 4) from getting written whereas, the other set (Cell 6, 7, 8) will be written. To prevent this, we apply a pulse voltage that varies between $V_{write} + \Delta$ and $V_{write} - \Delta$ (Fig. 15) to all unselected BLs and WLs. We note that the write delay

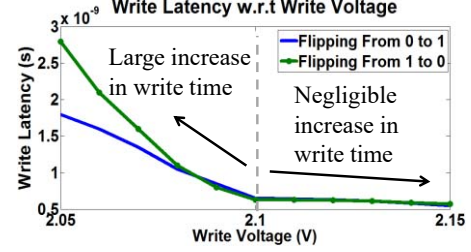


Fig. 16 Write latency w.r.t write voltage.

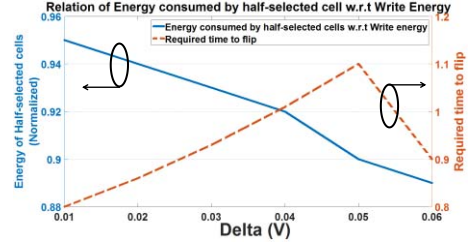


Fig. 17 Energy consumed by half-selected cells and time to flip.

increases non-linearly at lower write voltages whereas, it increases linearly at higher write voltage (Fig. 16). Therefore by choosing V_{write} and Δ carefully we prolong the write latency of half-selected cells (Cells 2, 3, 4 and 6, 7, 8) in a way that they stay undisturbed by the time the selected cells (Cells 1 and 5) are written. Fig. 16 shows the simulation results for write latency with respect to V_{write} . For $V_{write} < 2.1$ V, the write latency reduced rapidly however, for $V_{write} > 2.1$ V, write latency stays almost the same. In this work, Δ is chosen as 0.05V with $V_{write} = 2.1$ V. Although, the half-selected cells suffer significantly in terms of retention ($\sim 15ns$) the 2bits are written successfully.

B. Simulation Results

In Fig.18, magnetization (M_x) with respect to time is shown with V_{pulse} applied to the half-selected cells and $2V_{write}$ to selected cells. It is evident that selected cells require $\sim 0.75ns$ to flip together (cell 1 and 5 in Fig. 15) with $2V_{write} = 4.2V$. However, the half-selected cells go back to their correct states when the biasing voltage becomes zero after 0.8ns. If the system clock speed is considered to be 1.25GHz, 1 clock cycle is needed for this write operation without disturbing the half-selected cells. In this work, the pulse width of 0.4 ns with a 50% duty cycle is applied to unselected BLs and WLs. The pulse toggles between 2.15 V and 2.05 V. Simulation result shows that the half-selected cells with the pulse voltage source require 1.1 ns for $0 \rightarrow 1$ and 1.59ns for $1 \rightarrow 0$ flipping. If the worst case is considered, 1.5 clock cycles are needed to disturb the half-selected cells. Therefore, the same crossbar cannot be accessed back to back. A time gap between two consecutive accesses is required in the crossbar so that the half-selected cells relax and become stable again. This can be taken care of at the bank level by prohibiting back-to-back writes to a bank.

A shortcoming of the proposed write technique is an energy penalty while writing '10' and '01'. If the size of the crossbar array is $N \times N$, the current through each of the $2(N-1)$

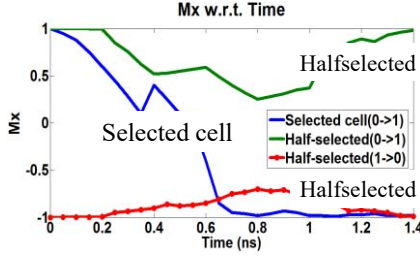


Fig. 18 Magnetization (Mx) of selected and half-selected cells with respect to time.

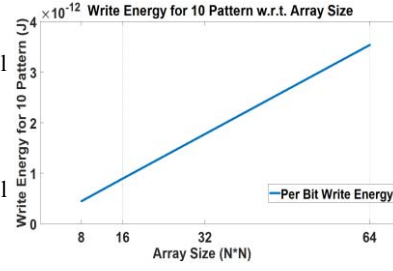


Fig. 19 Write energy for '10' pattern with respect to size of crossbar

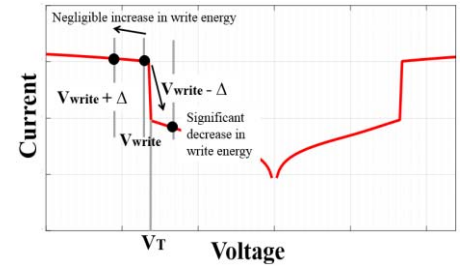


Fig. 20: V_T modulation of SD to reduce leakage current of half-selected bits.

Table 1: Write energy vs bit pattern for 64x64 crossbar

Data	Write energy in pJ (per bit)
00/11/10/01	0.089/0.091/3.53/3.53

half-selected cells will be almost equal to the write current. Therefore, these cells will consume significant energy (Table 1 for 64x64 crossbar array). Note that although the energy overhead of 10/01 pattern is high, a cache line will contain only a fraction of such patterns to partially mitigate the overhead. We propose the following techniques to lower the energy overhead further:

(a) Adjustment of the value of Δ and adjustment of V_T of the MIIM diode: The I-V curve of MIIM diode in Fig. 3 shows that $V_{write}-\Delta$ reduces the current non-linearly compared to the current increment by $V_{write}+\Delta$. Fig. 17 shows the average energy of the half-selected cells with respect to Δ . The consumed energy is reduced to 5% by varying delta.

(b) Reducing the crossbar size: The per bit write energy decreases as the number of half-selected cells (in s1 and s2 row of Fig. 15) decreases (Fig. 19). Therefore, a smaller crossbar can lower the energy overhead.

(c) V_T modulation: If the V_T of the MIIM diode is modulated in a way that the 1D-1MTJ writes at V_{write} but at $V_{write} - \Delta$, the diode turns off (very low current through the cell) and at $V_{write} + \Delta$, current through the cell is not very high compared to I_{write} (Fig. 20), one row of half-selected cells remains off. Therefore, 50% write energy saving for writing 10 or 01 is possible.

(d) Data encoding: By encoding the cache line data, the population of 00/11 compared to 01/10 can be increased. This will, in turn, lower the required write energy.

V. CONCLUSION

In this work, two bit read and write schemes have been presented. We improve the sense margin during read to sustain larger crossbar arrays. We propose voltage biasing technique for multi-bit read operation. We also propose two-bit write by employing pulsed voltage on unselected wordlines and bitlines.

ACKNOWLEDGMENT

This work is supported by the NSF under Award No. CNS –

1722557, CCF – 1718474, DGE – 1723687 and DGE – 1821766, DARPA Young Faculty Award under Award No. D15AP00089 and SRC under Award No. 2847.001. The authors also thank Dr. Jashmi Jha (University of Cincinnati) and Radha Krishna Aluru (Ford Motor Company) for their valuable inputs.

REFERENCES

- [1] Gopalakrishnan, K., et al. "Highly-scalable novel access device based on mixed ionic electronic conduction (MIEC) materials for high density Phase Change Memory (PCM) arrays". VLSIT, 2010.
- [2] Huang, Jiun-Jia, et al. "One selector-one resistor (1S1R) crossbar array for high-density flexible memory applications." IEDM, 2011.
- [3] Seo, Jung Won, et al. "A ZnO cross-bar array resistive random access memory stacked with heterostructure diodes for eliminating the sneak current effect." Appl Phys Lett, 2011.
- [4] Kim, Sungho, et al. "Crossbar RRAM Arrays: Selector Device Requirements During Write Operation.", IEEE Transactions on Electron Devices, 2014.
- [5] S. Ghosh, et al. "Design Space Exploration for Selector Diode-STT RAM Crossbar Arrays," IEEE Transactions on Magnetics, 2018.
- [6] Jiantao Zhou, et al. "Crossbar RRAM Arrays: Selector Device Requirements During Read Operation," TED, 2014.
- [7] Y. Deng et al., "RRAM Crossbar Array With Cell Selection Device: A Device and Circuit Interaction Study," TED, 2013.
- [8] C. Nauenheim, et al. "Nano-Crossbar Arrays for Nonvolatile Resistive RAM (RRAM) Applications", IEEE Nano, 2008.
- [9] Hyein Lim, et al. "ReRAM Crossbar Array: Reduction of Access Time by Reducing the Parasitic Capacitance of the Selector Device," TED, 2016.
- [10] C. Xu, et al. "Design implications of memristor-based RRAM cross-point structures," DATE, 2011.
- [11] C. Xu et al., "Overcoming the challenges of crossbar resistive memory architectures," HPCA, 2015.
- [12] Y. Zhang, et al. "Multi-level cell STT-RAM: Is it realistic or just a dream?," ICCAD, 2012.
- [13] Y. Zhang, et al, "MLC STT-RAM design considering probabilistic and asymmetric MTJ switching," ISCAS, 2013.
- [14] M. N. I. Khan, "Multi-Bit Read Write Methodology For Diode-STT RAM Crossbar Array." M.Sc. dissertation, Pennsylvania State University, 2019.