Test Challenges in Embedded STT-MRAM Arrays

Insik Yoon, Arijit Raychowdhury

School of Electrical & Computer Engineering,
Georgia Institute of Technology, Atlanta, Georgia
Email: iyoon@gatech.edu, arijit.raychowdhury@ece.gatech.edu

Abstract—Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) is an emerging memory technology which exhibits non-volatility, high density and nanosecond read and write times. These attributes of STT-MRAM make it suitable for last level embedded caches. However, the defects and corresponding fault models of STT-MRAM are not as extensively explored as in SRAM and therefore, there is a growing need for defect and fault analysis. Moreover, stochastic retention failure of STT-MRAM imposes a large burden in testing time. Conventional test schemes for retention of STT-MRAM need to be optimized for testing a large-size embedded STT-MRAM array. This work presents a review of the different defect and fault mechanisms as well as a BIST architecture and circuit to reduce testing time in characterization and manufacturing tests for retention. We address the effect of resistive and capacitive defects and identify retention test setup for measuring worst case retention.

I. INTRODUCTION

A STT-MRAM cell is composed of one access transistor and one magnetic tunnel junction (MTJ) that stores a bit of information as shown in Fig.1(a). [1] An MTJ has two ferromagnetic layers (CoFeB based), which are called fixed and free layer and they are separated by a thin insulator layer(MgO based). The magnetic moment in fixed layer is 'fixed' into one direction and the direction of magnetic moment in free layer can be changed depending on difference and polarity of potential across an MTJ. When a potential difference is applied across MTJ, spin-polarized current passes through an MTJ and it attempts to polarize the current in its preferred direction of magnetic moment. The angular momentum of the electrons in free layer creates a torque that causes a flip in the direction of magnetization inside the free layer of MTJ. Depending on the direction of magnetization in the free layer, resistance of
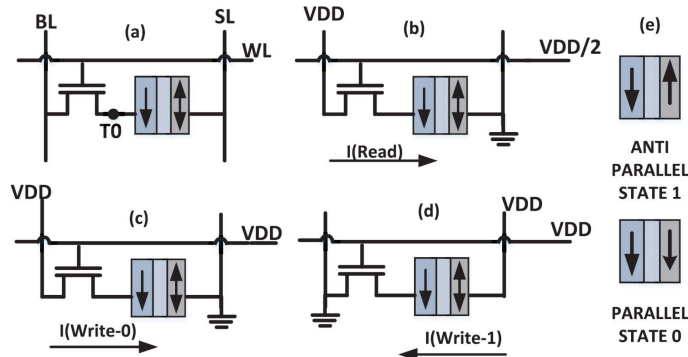

Fig. 1: (a) Basic 1T-1MTJ cell. (b) Bias condition for read. (c) Write 0 bias condition. (d) Write 1 bias condition. (e) States in a MTJ due to orientation of magnetic moments. [1]

MTJ is changed. As shown in Fig.1, when the direction of magnetization in free layer is anti parallel to the magnetization of fixed layer, MTJ has high resistance and when they are in parallel, MTJ has low resistance. Bit 1 and 0 are mapped to the two cases: when MTJ has high resistance and low resistance respectively. The bias conditions on bit line, source line and word line for write and read operation are shown in Fig.1(b)-(d). The write operation is bidirectional, where either the bit-line (BL) or source line (SL) is pulled high and the other one is pulled low depending on the polarity of the write operation. In case of writing 1, the bit line is vdd and source line is ground. When word line is asserted, write current flows from source line to bit line and MTJ will be in anti-parallel state. The bias condition is set to the opposite when writing 0 to a cell and the current from bit line to source line sets MTJ to be in parallel state. The read operation is unidirectional with word line driven to $vdd/2$ and a pre-charged BL voltage discharges through the cell. Depending on the resistance of MTJ, discharging voltage at source line is sensed at sense amplifier. STT-MRAM is non-volatile since a bit is stored in an MTJ as resistance and it is determined by magnetization of free layer. The MTJ can either be an In-plane MTJ (I-MTJ) with magnetic anisotropy in plane due to shape anisotropy or a Perpendicular plane MTJ (P-MTJ) where magnetic anisotropy is aligned out of plane, independent of the shape of the free layer [2]. The relative merits and demerits of the two structures are being extensively studied [2] [3] [4] [5] .

STT-MRAM arrays are expected to suffer from read and write failures which are induced by electrical defects and process variations. In [1] [6], the types of resistive, capacitive and coupling defects are identified. Their manifestation as read and write faults as well as fault activation patterns are analyzed. Apart from read and write faults, STT-MRAMs can also suffer from retention failures, a bit-flip in a cell caused by thermal noise [7], [8]. Since retention time is exponentially proportional to the stored energy (thermal stability), conventional test method for retention described in [7] measures thermal stability of a cell by applying weak write current to a cell. However, it results in prohibitively large test time. In the second part of this paper we review a Memory Built In Self-Test (MBIST) architecture from [9] that can detect the retention failures in a time-efficient manner. It is an efficient MBIST architecture that can perform in-situ read, write and retention (stochastic test) tests on STT-MRAM arrays.

The paper is divided as follows. In section II, we summarize the resistive and capacitive defects and corresponding fault models discussed in [1]. In section III, we discuss conventional retention test that faces a large burden in testing time due to stochastic retention failure of STT-MRAM. Lastly, in section

IV, we introduce the new retention test scheme proposed in [9] that decreases retention test time significantly.

II. DEFECTS AND FAULT ANALYSIS

[6] [1] have identified faults during read and write and how resistive and capacitive defects induce faults. Table 2 summarizes fault models and how they contribute to read/write failures.

Fault Model	Affects	Key Cause
Transition Fault (TF)	WR	Relative Weak WR current due to stray resistive paths
Coupling Fault(CF)	WR	Neighboring cells switching
Stuck At Fault(SF)	WR	T0 , WL stuck at VDD or GND
Incorrect Read Fault (IRF)	RD	Current miscorrelation due to defects affecting WL,BL
Read Disturb Fault (RDF)	RD	Electrical disturbance at T0 node due to larger than normal RD current

Fig. 2: Table I: Defects induced faults [9]

A. Resistive defects

In this section, we discuss the role of resistive defects in 2 by 2 cell array explored in [1] [9]. As shown in Fig.3, the resistive defects are associated with resistive shorts between the node of victim cell and aggressor cells. In the figure, 13 possible resistive defects can affect read/write of the cell.

For example, a case of having $RS_{BL0-SL1}$ short between BL0 and SL1 in Fig.3(a) results in write 0 fault in cell-0 because $RS_{BL0-SL1}$ prohibits BL0 to reach full VDD.

As a result, weak write current across cell-0 causes a transition fault to 0. A read failure can happen due to $RS_{WL0-WL1}$. In case of reading cell-0, word line 0 is weakly asserted due to the nresistive short between wordline 0 and 1. If cell-1 contains 0, a change of incorrect read fault can happen because of slower discharge of bitline 0. [1] analyzed how each resistive short shown in Fig.3 causes faults in either read or write operation.
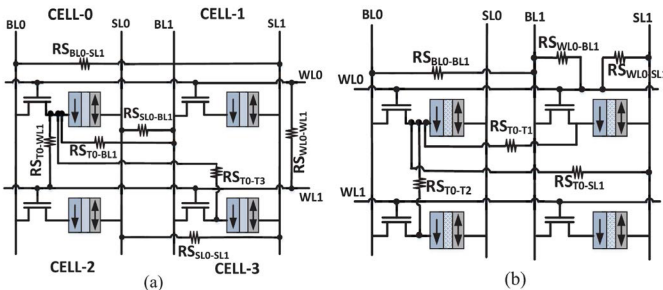

Fig. 3: (a) and (b) Inter-cell resistive defects [1]

B. Capacitive defects

As shown in Fig.4, capacitive defects are associated with capacitive coupling between data storage node, a node between access transistor and MTJ, and word lines. Among capacitive defects, the capacitive coupling between WL and data storage node are the most aggressive defect since unintended writes in neighboring cell can be caused. [9]. For example, $C_{CWL0-T3}$ from Fig.4(a) can cause unintended write to cell-3 in a case of writing to cell-1. Another capacitive defect that cause write failure is shown in Fig.4(b). A coupling capacitor $C_{C\ WL0-WL1}$ weakly asserts WL1 when writing 0 to cell0. If the cell 1 is a weak cell, which means the critical current for write is lower than cell 0, cell 1 may be written to 0 as well. Also, depending on the capacitance ratio between $C_{C\ WL0-WL1}$ and C_{WL1}, write time to cell-1 is varied [9].
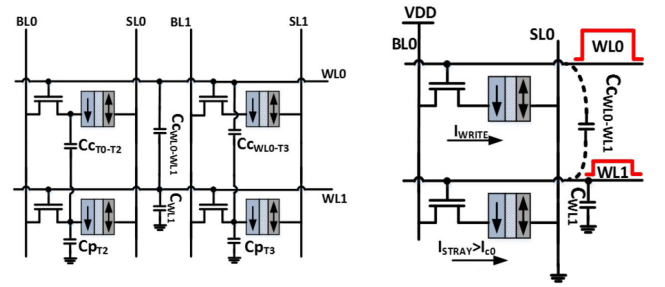

Fig. 4: (a) Most aggressive capacitive defects (b) capacitive coupling between WLs that may cause coupling fault [9]

III. CHALLENGES IN RETENTION AND THERMAL STABILITY TESTS

In order to measure retention, authors in [7] proposed a possible test methodology based on the thermal activation model. In STT-MRAM, retention time is defined as the time it takes for a cell to flip, a stochastic phenomenon, caused by thermal noise [5]. The average retention time is quantified as: $\tau = \tau_0 exp(\Delta)$ and $\Delta = \frac{K_u V}{k_B T} = \frac{H_k M_s V}{2k_B T}$ [5]. In order to ensure system reliability, each cell in an array must have enough thermal stability(Δ = 60 to guarantee 10 years of retention) against stochastic bit flip induced by thermal noise. Therefore, having high Δ allows cells to have long retention but at the same time, it increases write time and current. [10] [11] [12] Since retention of array directly impact system reliability as described above, determining Δ in post-Silicon tests is of utmost importance. However (1) the statistical nature of thermally activated bit-flips, (2) low failure probabilities, (3) large dependence on temperature and process parameters (M_S, H_K, t) and (4) exponential dependence of retention times and retention failure probability on Δ make it a challenging test problem, as has been noted in the Intel publication [7].

[7] uses the thermal activation model [8] [13] [14] and for the test case:

$$\frac{t_p}{\tau_0 exp(\Delta(1 - \frac{I_{WWR}}{I_{c0}}))} << 1 \qquad (1)$$

P_{sw} can be described as follows [8] [7]:

$$ln(P_{sw}) = ln(\frac{t_p}{t_0}) - \Delta(1 - \frac{I_{WWR}}{I_{c0}}) \qquad (2)$$

which results in the final form:

$$P_{sw} = 1 - exp\left(-t/exp\left(\Delta\left(1 - \frac{I_{WWR}}{I_{c0}}\right)\right)\right) \qquad (3)$$

According to equations above, measuring switching probability(P_{SW}) of cells by applying weak write current(I_{WWR}) can yield thermal stability. Once P_{sw} is extrapolated, P_{sw} with no current applied is retrieved and retention is calculated from thermal stability. The relationship between P_{SW} and I_{WWR} and extrapolation of P_{sw} is shown in Fig.5.
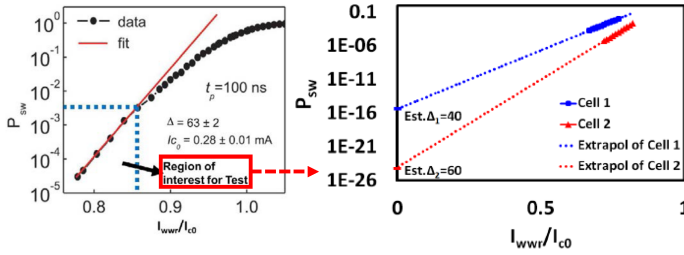

Fig. 5: (a)Experimental data of P_{SW} vs. I_{WWR} [8] (b)Extrapolation of P_{sw} in linear region [9]

However, in order to extrapolate P_{sw}, low weak write current must be applied and it results in low P_{sw} in linear region. Since the thermal activation model is a stochastic model, a large number of successive tests is required to obtain statistically significant results. As a result, retention test time for cells with high thermal stability increases exponentially because P_{sw} with high thermal stability is lower as Fig.5(b) shows. The entire algorithm from [7] is shown below.

Result: Obtain P_{sw} for every cells in an array

initialization;
N_{row} = total number of rows;
Iwwr[N] = array that contains N number of I_{WWR} value;
M = total number of experiments per each I_{WWR};
for *i = 0; i <N_{row}; i++* **do**
    for *j = 0; j <N; j++* **do**
        Write test patterns into the line;
        for *k = 0; k <M; k++* **do**
            Apply current I_{WWR}[j] for t_p;
            Read the line value;
            if *value $\neq$ test pattern* **then**
                error counter of cells with error++;
                rewrite correct value to the row;
            end
        end
    end
end

Algorithm 1: Retention test algorithm with weak WR current [9]

Even though parallelism at a sub-array level can help to reduce retention test time, there is a clear limit in reducing the total retention test time. With increasing size of cell array, the retention test time with this MBIST is not feasible. Therefore, there is a strong need for efficient retention test algorithm which can reduce test time significantly. We address this issue in the next section.

IV. PROPOSED MBIST FOR RETENTION TESTING

In order to solve retention test time problem, [9] proposed a new retention test scheme that perform in-situ, statistical, retention testing of large STT-MRAM arrays. From the retention BIST algorithm [7], a weak write current is applied row by row and the row is read right after to obtain P_{sw}. The major drawbacks from this scheme are; (1) The retention test time increases linearly when the row size of an array increases.
(2) The retention tests have to be carried out in an linear region where P_{SW} is very low. Since P_{sw} is very low in the region that retention test is performed, a bit flip will not happen for most of the iterations; which means most of the read operations after applying current are not necessary.
These two problems are main bottlenecks for improving speed of retention test. The retention test scheme from [9] reduces retention time significantly by:
1. Applying weak write current to multiple rows
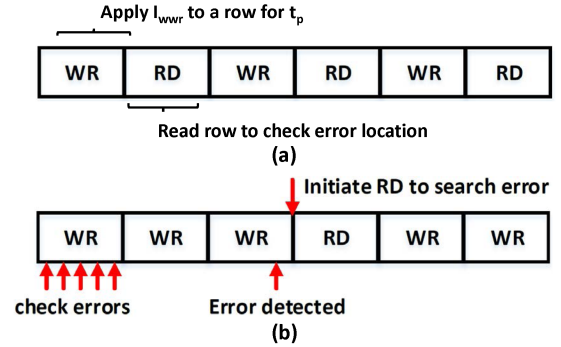2. Avoiding search (reading rows) when error is not detected

Fig. 6: (a) conventional test scheme (b) proposed test scheme

By testing multiple rows in a column at the same time and searching for errors after error detection, retention time testing reduces significantly. As Fig.6 shows, the new test scheme concurrently runs error detection while applying weak write current to multiple rows. Read operation happens only when error is detected within the rows under test.

The retention test is divided into two processes, (1) Error Detection (ED) and (2) Error Search (ES).

A. Error Detection (ED)

The ED architecture detects an error by observing the change in current flowing through a cell since a bit flip in a cell changes resistance ac cross a cell and it results in current change. As shown in Fig. 7, MTJ resistances in multiple cells in a column are connected in parallel when corresponding word-lines turn on simultaneously. When I_{WWR} causes a bit flip in a cell, the current at source line (I_{SL}) changes due to the resistance change, which shows that at least one error is detected within a column.

Fig. 7 shows the scheme for detecting a change in I_{SL} caused by a bit flip of a cell. Current mirror and multi-stage common dran amplifiers amplify the change in I_{SL} and switched capacitors C1 and C2 sample the voltage at the common drain amplifier alternatively based on CLK and
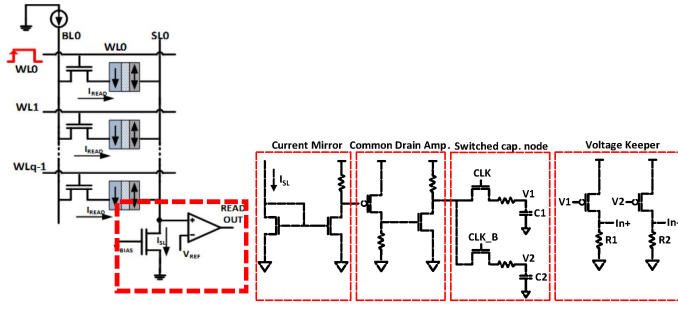
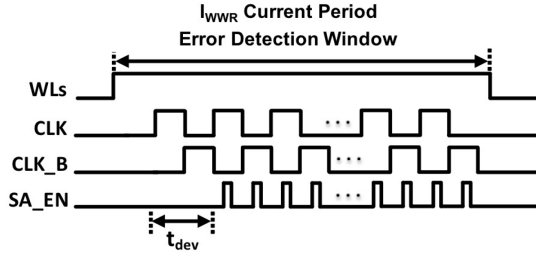Fig. 7: Error Detection circuit for a column with 16 rows [9]

Fig. 8: Timing Diagram illustrating the operation of the MBIST retention test [9]

CLK_B signals. In a case of bit flip, the voltage difference between C1 and C2 is developed and maintained for a half clock cycle. By calibrating value of R1 and R2, in+ port is set to be always 10mV higher than in- to prevent metastability issue in sense amplifier. When sense amplifier enable is on, the sense amplifier fully differentiates the in+ and in- to VDD and GND. Fig. 8 presents waveform of switched capacitor control signals(CLK, CLK_B) and sense amplifier enable.

B. Error Search

Once error is detected from ED stage, The location of error is searched within activated rows in order to obtain P_{sw} and thermal stability of cells. [9] present exhaustive search and temporal locality search algorithms.

1) Exhaustive Search: In exhaustive search, every row in a block of activated rows is read to locate errors. Once error location is identified, it is stored in a error table and original test pattern is written to a row that contains error. Error location stored in a error table is used in temporal locality searches which exploits temporal locality of error locations.

2) Temporal locality search: Temporal locality search can reduce error search time when some cells exhibits relatively low thermal stability due to process variation. Once error table is filled from exhaustive search and an error is detected from error detection, temporal locality search accesses the error location from the error table first to locate errors. Since weak cells tend to cause significantly more number of errors in retention test, they exhibit strong temporal locality in terms of causing errors. If the row specified in the error table contains an error, number of errors associated with the row in a table is updated. When no error is found in the rows from the error table, it switches to exhaustive search to find errors in other rows and add a row that contains the error to the error table. After it finds an error, it reads the block of rows to ensure it corrected all errors.

V. SUMMARY

This paper presents a comprehensive test methodology for STT-MRAM arrays. We identify resistive and capacitive defects that result in read, write failure. The challenges in retention test is also discussed and a new MBIST architecture [9] capable of collecting statistical data in an STT-MRAM subarray to estimate the thermal stability and retention is proposed. The proposed MBIST shows 93.75% improvement in test-time compared to a brute-force approach [7] with less that 5% estimation error.

REFERENCES

[1] A. Chintaluri, H. Naeimi, S. Natarajan, and A. Raychowdhury, "Analysis of defects and variations in embedded spin transfer torque (stt) mram arrays," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. PP, no. 99, pp. 1–11, 2016.

[2] S. A. Wolf, J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The promise of nanomagnetics and spintronics for future logic and universal memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2155–2168, 2010.

[3] G. Jan, L. Thomas, S. Le, Y.-j. Lee, H. Liu, J. Zhu, R.-y. Tong, K. Pi, Y.-J. Wang, D. Shen, R. He, J. Haq, J. Teng, V. Lam, K. Huang, T. Zhong, T. Torng, and P.-k. Wang, "Demonstration of fully functional 8Mb perpendicular STT-MRAM chips with sub-5ns writing for non-volatile embedded memories," *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, vol. 093008, no. 2012, pp. 1–2, 2014.

[4] H. Yoda, S. Fujita, N. Shimomura, E. Kitagawa, K. Abe, K. Nomura, H. Noguchi, and J. Ito, "Progress of STT-MRAM technology and the effect on normally-off computing systems," *Technical Digest - International Electron Devices Meeting, IEDM*, pp. 259–262, 2012.

[5] A. V. Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskii, R. S. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. H. Butler, P. B. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, "Basic principles of STT-MRAM cell operation in memory arrays," *Journal of Physics D: Applied Physics*, vol. 46, no. 13, p. 139601, 2013.

[6] A. Chintaluri, A. Parihar, S. Natarajan, H. Naeimi, and A. Raychowdhury, "A Model Study of Defects and Faults in Embedded Spin Transfer Torque (STT) MRAM Arrays," *2015 IEEE 24th Asian Test Symposium (ATS)*, vol. 1, no. c, pp. 187–192, 2015.

[7] H. Naeimi, C. Augustine, A. Raychowdhury, S.-l. Lu, and J. Tschanz, *Intel Technology Journal, STTRAM Scaling and Retention Failure*, vol. 17. 2013.

[8] R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall, and A. B. Kos, "Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions," *Journal of Applied Physics*, vol. 109, no. 7, 2011.

[9] I. Yoon, A. Chintaluri, and A. Raychowdhury, ""EMACS: Efficient MBIST Architecture for Test and Characterization of STT-MRAM Arrays," *International Test Conference*, 2016.

[10] A. Nigam, C. W. Smullen, V. Mohan, E. Chen, S. Gurumurthi, and M. R. Stan, "Delivering on the promise of universal memory for spin-transfer torque RAM (STT-RAM)," *Proceedings of the International Symposium on Low Power Electronics and Design*, vol. 1, pp. 121–126, 2011.

[11] A. Jog, A. K. Mishra, C. Xu, Y. Xie, V. Narayanan, R. Iyer, and C. R. Das, "Cache revive: Architecting volatile STT-RAM caches for enhanced performance in CMPs," *Proceedings of the 49th Annual Design Automation Conference (DAC)*, pp. 243–252, 2012.

[12] Z. Sun, X. Bi, H. H. Li, W.-F. Wong, Z.-L. Ong, X. Zhu, and W. Wu, "Multi retention level STT-RAM cache designs with a dynamic refresh scheme," *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 329–338, 2011.

[13] A. Driskill-Smith, S. Watts, D. Apalkov, D. Druist, X. Tang, Z. Diao, X. Luo, A. Ong, V. Nikitin, and E. Chen, "Non-volatile spin-transfer torque RAM (STT-RAM): An analysis of chip data, thermal stability and scalability," *2010 IEEE International Memory Workshop, IMW 2010*, vol. 1, no. 408, pp. 5–7, 2010.

[14] M. Pakala, Y. Huai, T. Valet, Y. Ding, and Z. Diao, "Ciritical Current distribution in spin-transfer-switched magnetic tunnel junctions," *Journal of Applied Physics 2005*, vol. 98, no. 5, 2005.