A Model Study of Defects and Faults in Embedded Spin Transfer Torque (STT) MRAM Arrays

Ashwin Chintaluri, Abhinav Parihar, Suriyaprakash Natarajan[+], Helia Naeimi[+], Arijit Raychowdhury
School of Electrical & Computer Engineering, Georgia Institute of Technology
[+] Intel Corporation
Contact Email: achintaluri3@gatech.edu

Abstract — *There has been a significant interest in Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) as a candidate for emerging memory technology for last-level embedded caches in the recent years. High density (3-4x of SRAM), non-volatility, nano-second Read and Write speeds, and process and voltage compatibility with CMOS are the attractive properties of this technology. A few studies have expounded on the reliability in this technology but various fault manifestations have not been studied in detail in the past. This paper attempts to study the fault models in STT-MRAM under both parametric variations as well as electrical defects (opens and shorts). Sensitivity of Read, Write and Retention to material and lithographic process parameters has been studied. Also electrical defects viz. intra-cell and inter-cell opens and shorts have been considered and the corresponding fault models have been identified and classified.*

Index Terms: **STT-MRAM, defects, variations, testing**

I. INTRODUCTION

With rising demand for larger on-die memory, Spin Torque Transfer (STT)-MRAM has gathered significant interest in the research community owing to its high density, non-volatility and easy integration with the CMOS process [1]. By solving the problem of increased switching currents at lower process nodes in traditional MRAM, it has emerged as a successor to MRAM and is considered a promising alternative to eDRAM and SRAM. The huge potential of STT-MRAM as a viable embedded memory technology at advanced process nodes has been well demonstrated in [2] and [3] at 45nm and 65nm nodes respectively. As the STT-MRAM technology continues to mature, rigorous analysis of fault models in this novel resistive memory and the role of extreme variations and defects in Read (RD), Write (WR) and Retention in failure need to be studied in detail. Some previous works [4][5][6] addressed fault modelling in Memristor arrays by injecting electrical defects and identified possible fault models in memristors, but because of the fundamental differences between the technology in Memristor and STT-MRAM, all the fault models of Memristor arrays [4] are not applicable to STT-MRAM based memory systems. STT-RAM is a truly bistable device and therefore the state of logic stored in the MTJ doesn't fall under any "undefined region" as described for other emerging non-volatile technologies like PCM or Memristor in [4]. Further, Dynamic Write Disturb Fault (dWdF) identified in [5] for Memristor is not present in STT-MRAM because of this bistability. STT-MRAM faces its own unique set of possibilities of failure and faults due to parametric variations and injected defects. This paper provides a comprehensive treatment and classification of the fault models manifesting due to both parametric variations and electrical faults in STT-MRAM memory arrays. We also study sensitivity of both Write and Read failure with the parameter variations. The flow of this paper is as follows. In Section II, we briefly describe the STT-MRAM and an end-to-end and vertically integrated model capable of simulating STT-MRAM from devices to arrays encompassing the spin dynamics used for the study. The role of parametric variations and RD and WR fault mechanisms is described in Section III. Defects (resistive opens and shorts) within a cell (intra-cell) and cell-to-cell (inter-cell) drawing inspiration from the framework used by [4][5] and their effects are discussed in Section IV. Conclusions are drawn in Section V.

II. SIMULATION INFRASTRUCTURE

A. 1T-1MTJ STT-MRAM:

When a spin-polarized current passes through a mono-domain ferromagnet, it attempts to polarize the current in its preferred direction of magnetic moment. As the ferromagnet absorbs some of the angular momentum of the electrons, it creates a torque that causes a flip in the direction of magnetization in the ferromagnet [7]. This is used in magnetic tunneling junction (MTJ) based spin torque transfer (STT) RAM cells where a thin insulator (MgO) is sandwiched between a fixed ferromagnetic layer (polarizer) and the free layer (storage node) [8]. This can be integrated in the metal stack and hence provide high memory density. Depending on the direction of the current flow (perpendicular to these layers in our study), the magnetization of the free layer is switched to a parallel (P: low resistance state or state-'0') or anti-parallel (AP: high resistance state or state-'1') state Fig.1(e). Fig.1 shows the 1T-1MTJ cell and Fig.1(b),(c),(d) show the bias conditions to write different logic states. Thus WR is bidirectional, whereas RD, with an under-driven Word-line (WL) is unidirectional.
A pre-charged Bit-line (BL) voltage is allowed to discharge through the cell; the discharge rate being governed by the resistive state of the cell; and finally it is read out through a sense circuit. When the cell is not accessed, BL, Source-line

978-1-4673-9739-1/15 $31.00 © 2015 IEEE
DOI 10.1109/ATS.2015.39

IEEE computer society

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.

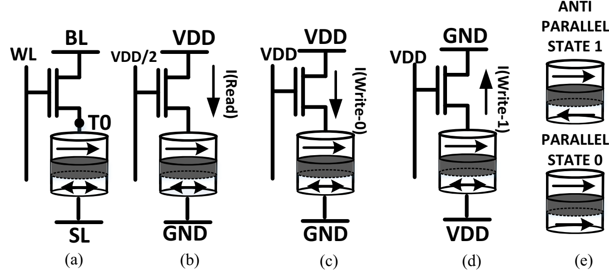

Fig 1. (a) Basic 1T-1MTJ cell (b) bias condition for read (c) write 0 bias condtion (d) write 1 bias (e) States in a MTJ due to orientation of magnetic moments

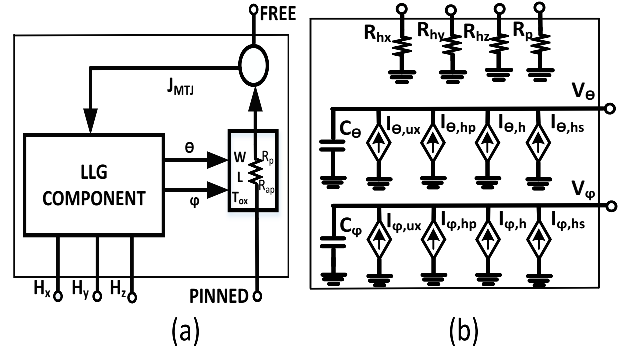

Fig. 2: (a) The MTJ subcircuit (b) The implementation of LLG equation.

(SL) and WL are all turned to '0', enabling a non-volatile (NV) cell [8] . However, true non-volatility is achieved only when the free-layer magnet can store enough internal energy (~60-80kT), which in turn is proportional to its volume ($\Delta \sim 1/2 M_S H_K V$) [9]. With scaling, the volume, and hence Δ decreases giving rise to the notion of quasi-NV cells.

B. STT-MRAM device model

The simulation model is based on the macrospin assumption of the free-layer nanomagnet [9][7] that the MTJ is comprised of. The linearized LLG equation (1) [7] predicts the switching dynamics of the free-layer magnetic moment *m(t)* in presence of the torque experienced because of uniaxial anisotropy field (T_U), easy plane anisotropy field (T_K), external magnetic field (T_H), and spin torque from electrons (T_S). Solving a linearized form of this LLG equation in polar coordinates (2) gives the trajectory of the magnetization vector for the free layer in terms of angles θ and ϕ in the 3-D space. The LLG equation models the spin dynamics under a current driven spin-torque and can be expressed as [7]:

$$\frac{d\vec{m}(t)}{\mathrm{dt}} + \alpha\left(\vec{M}(t) \times \frac{d\vec{m}(t)}{\mathrm{dt}}\right) = \gamma\,\vec{T} \tag{1}$$

$$\frac{1+a^2}{\gamma H_k}\begin{bmatrix} \frac{d\theta}{\mathrm{dt}} \\ \frac{d\phi}{\mathrm{dt}} \end{bmatrix} = \vec{T_U} + \vec{T_K} + \vec{T_H} + \vec{T_S} \tag{2}$$

where α is the LLG damping coefficient γ is the gyromagnetic ratio , *Hk* is the uniaxial anisotropy field.The resistance of the MTJ stack as a function of the angle θ between the fixed and free layers and the operating conditions can be expressed as [10]:

$$R(\theta, V, T) = \frac{\sin(\mathrm{cT})}{\mathrm{cT}}\left[P_3\theta^3 + P_2\theta^2 + P_1\theta + R_P\right] \cdot (1 - \mathrm{abs}(V)/\mathrm{Slope}) \cdot 10^{s(T_{\mathrm{ox}} - T_{\mathrm{ox},0})} \tag{3}$$

where T_{ox} is oxide thickness, R_p is the tunneling resistance in parallel mode, V is applied voltage , c is a material constant and Slope determines voltage dependece of R_{AP}.

Earlier work by the authors [9][10][11] have demonstrated how such magnetic dynamics can be incorporated in an HSPICE environment through explicit functional definitions of controlled voltage and current sources. Fig.2 depicts the circuit equivalent SPICE model of the MTJ which is modelled using in-built voltage dependent current sources and capacitors to capture the dynamics predicted by the LLG equation providing a physical realization of the switching and an equivalent electrical resistance for the transport component.

The MTJ model is fully parameterized using device and material parameters discussed in the next section, which allows comprehensive variation analysis. This device model is incorporated in a bit-cell with a 2-Fin series selector transistor from a 14nm process [12] and the design has been scaled up to an array with peripherals similar to conventional memory systems similar to organization presented in [4]. This model features advanced simulation capabilities including: (a) simultaneous WR on different BL, (b) back-to-back RD/WR, (c) evaluation of sneak current paths through inter-cell bridges (d) smart Monte-Carlo techniques with in-built response surface analysis for statistical data collection. We use the developed end-to-end simulation environment to study key material, device and circuit parameters and their roles in different failure mechanisms in the array.

III. Parametric Variations And Fault Models

Like every other memory technology, we expect STT-MRAM to also face severe process induced variations.We look at the three major types of failure mechanisms in STT-MRAM, namely the Read, Write and Retention failures [13][9], and identify fault models that result in these failures.The main sources of parametric variations in STT-MRAM can be categorized as [14] : **MTJ Material Parameters:** (a) normally distributed localized fluctuation of magnetic anisotropy,H_K [10], (b) Saturation Magnetization (M_S) (c) Tunnel Magneto-Resistance ratio (TMR) which is the ratio of difference between high and low resistances to the low resistance of MTJ, **Transistor Electrical Parameter:** (a) normally distributed threshold voltage (V_t) with σ~10%. **Lithographic Variation:** (a) normally distributed variation

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.

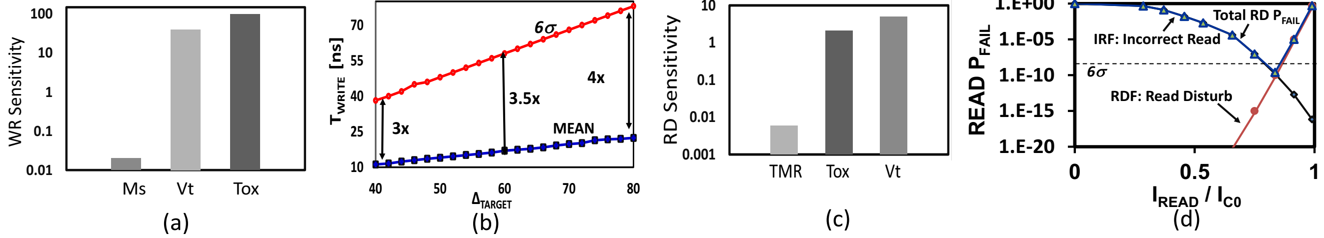

Fig. 3: (a) Sensitivity of WR time with process parameters, (b) WR time as a function of a target Δ., (c) Sensitivity of RD time with process parameters, and (d) RD Failure probability as a function of RD current, showing the different failure mechanisms.

of planar dimensions with σ~10%, and (b) normally distributed variation of MgO thickness with (μ=1.1nm and σ=0.1nm). The μ and σ of the other distributions (not mentioned here) have been derived from [9] through model calibration. All these sources of variation lead to variations in RD, WR and Retention . Enough guard-bands are provided in designs for a target failure probability (P_{FAIL}), typically for a 6σ corner (P_{FAIL}~10^{-9}). Under extreme variations and defects (to be studied in Sec III), a particular bit-cell may fail (in RD, WR or retention) even when design margins up to 6σ guard-bands are used. Such a failure will manifest as a fault. Hence we need to: (a) understand how large
parameters shows large dependence on the transistor threshold voltage V_t and the T_{OX} of the MTJ (Fig. 3(a)). For the guard-bands are, and (b) categorize the Fault Primitives and provide corresponding 'Fault Models'. Let us first analyze the process of WR under parametric variation. A sensitivity of WR for a parameter, p is defined as $S=(\partial T_{WR}/T_{WR})/(\partial p/p)$. The sensitivity analysis of WR time with respect to key process
a target storage energy Δ, it is observed that the 6σ values of T_{WRITE} are 3x-4x larger than the mean (Fig. 3(b)), which is significantly larger than competing memory technologies. Similar analysis of RD has been shown in Fig. 3(b),(c). The trade-off between WR and retention times has been shown in Fig. 4, where large σ/μ are noted. For 7yr retention for a 6σ cell, Δ_{TARGET}~60 is required. ***Extreme parametric variations and/or defects during high-volume manufacturing can exceed RD, WR and Retention guard-bands, and are modeled as faults.***

WR Fault Model because of parametric variation: If a 3X T_{WRITE} margin is provided for the worst-case cell (nominal T_{WRITE} =20ns for Δ_{TARGET}=60, Fig. 3(b)) any cell with T_{WRITE}>60ns is deemed un-writable. We characterize this as a 0→1 or 1→0 **Transition Fault** (**TF1** or **TF0**).

RD Fault Models because of parametric variation: The RD is characterized by two failure modes. (a) **Incorrect Read Fault (IRF):** The inability of the cell to distinguish between a '0' and '1' due to low READ current and/or low TMR (Fig. 3(d)) and (b) **Read Disturb Faults (RDF)**: The read current for a cell is so high that the value in the cell flips during RD (Fig. 3(d)). In STT-MRAM, RDF can either cause a faulty read in the same cycle or manifest itself in the subsequent read.

Retention Fault because of parametric variation: Finally a bit-cell can lose its state due to thermal noise, a problem more prominent in scaled bit-cells with decreasing Δ. Such a fault primitive is called **Retention Fault (RTF)**. Fig. 4 shows the average retention time in a nominal and a 6σ cell for varying Δ_{TARGET}. The key fault models and parametric variations leading to these faults are summarized in table I.

TABLE I: FAULTS DUE TO EXTREME PARAMETRIC VARIATIONS

Fault Model	Affects	Key Cause
Transition Fault (TF)	WR	WR Time > 6σ of nominal
Incorrect Read Fault (IRF)	RD	Low TMR, low READ current
Read Disturb Fault (RDF)	RD	High RD current due to low transistor V_t, causes bit-flip
Retention Fault (RTF)	Retention	$\Delta < \Delta_{TARGET}$ with guard-band

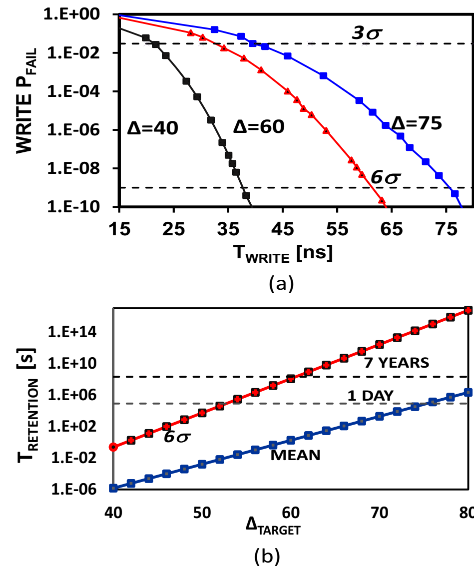

Fig. 4: (a) WR Failure probability distribution w.r.t. WR time shows a long tail which increases with Δ. (b) Average retention time shows a large variation owing to its exponential dependence on Δ.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.

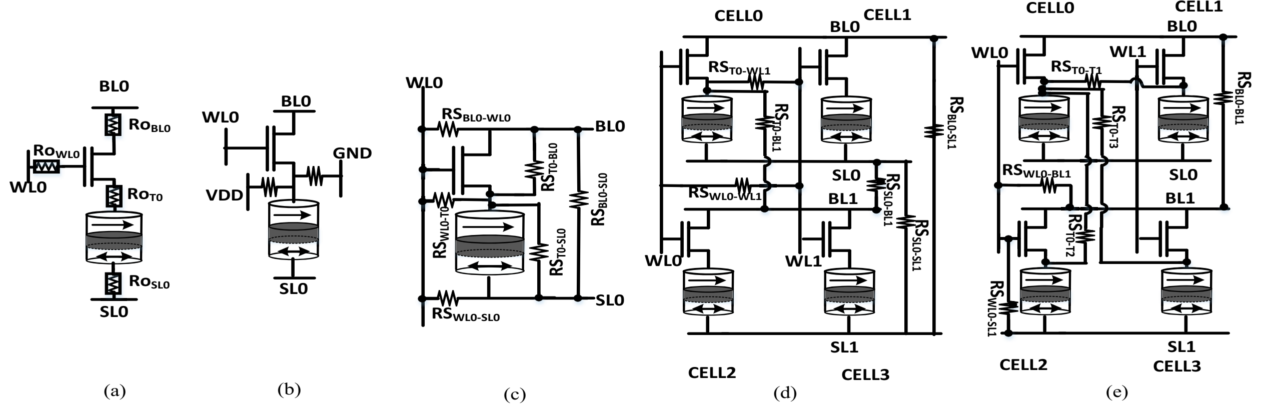

Fig. 5: Comprehensive defect models: (a)-(c) Intra or within cell defects, (d)-(e) inter or between cell defects.

IV. Defects And Fault Models

Defects in a hybrid CMOS memory cell can manifest in the form of opens and shorts between various terminals [4]. These defects may form during the BEOL process. To comprehensively study all the defect models we categorize them as intra-cell (within a cell) and inter-cell (cell-to-cell) defects and study their manifestation as faults. We have studied RC faults where resistive bridges are ac-coupled with parallel capacitors. However, our analysis reveals that even high capacitance values (~fF) have negligible effect on the fault model in STT-MRAMs. Hence in the rest of the paper we will only discuss resistive defects. In addition to the Write faults listed in Table I, defects manifest other traditional fault models[15] : **(a) Stuck At Fault (SF0 or SAF1):** Here resistive bridges short WL or node T0 (between transistor and MTJ) to either VDD (SF1) or GND (SF0). **(b) Coupling Fault (CF):** Here the process of WR on a neighboring cell can disturb the value in the victim. The fault models excited by defects and their key causes are summarized in Table II. RTF is not induced by resistive defects.

Intra-Cell Defects and Faults Models: The four terminals of the cell (BL, SL, WL and T0, the internal node) are considered and defects and bridges are injected covering all the nodes as shown in Fig. 5(a)-(c). The opens and shorts are modeled as resistors (open: 1kohm to 1Meg ohm and short: 10 ohm to 10kohm). The defects and the WR and RD fault models they excite are shown in Table III. Here, xWy refers to a cell whose original value is x and we are trying to write y. Rx refers to reading a value of x from a cell. $x, y \in \{0,1\}$. xW0/xW1 refers to writing 0/1 independent of the stored value. xWx refers to any WR process on the cell.

TABLE II: DEFECT INDUCED FAULTS

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

Key Observations: For intra-cell opens, any WL open sensitizes the TF even for relatively small values of the defect resistance (Fig. 6). Correspondingly any short at node T0 causes TF or SAF (if the short is to VDD/GND). On the other hand, shorts across the MTJ decreases RD margin (activates IRF) and across the transistor increases the RD current (causes RDF). Fig. 7 illustrates their corresponding sensitivities. Intra-cell opens increase the RD time by decreasing the RD current and cause IRF as shown in Fig. 7(a).

Inter-Cell Defects and Faults Models: Inter-cell defects are associated with resistive shorts between the nodes of the

TABLE III: DEFECT & FAULT MODELS WITH INTRA-CELL DEFECTS

Defect Type	Location	Write Fault model	Write Data Pattern (Victim)	Read Fault model	Read Data Pattern (Victim)
Open	BL0	TF0,TF1	xWx	IRF0	R0
	WL0	TF0,TF1	xWx	IRF0	R0
	SL0	TF0,TF1	xWx	IRF0	R0
	T0	TF0,TF1	xWx	IRF0	R0
Shorts	BL0 - T0			IRF1, IRF0, RDF	R1
	T0 - SL0	TF0,TF1	xWx	IRF1	R1
	WL0- BL0	TF1	xW1	RDF	R0
	WL0-T0	TF0	xW0	RDF	R0
	WL0-SL0	TF0	xW0	IRF1	R1
	BL0-SL0	TF0	xW0	IRF1	R1
	T0-VCC	SF1	xWx	IRF0	R0
	T0-GND	SF0	xWx	IRF1	R1

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.

TABLE IV: DEFECT & FAULT MODELS WITH INTER-CELL DEFECTS

Location	Agressor Cell	Write Fault model	Read Fault model	Read Data Pattern Victim
BL0-SL1	1	TF0	IRF0	r0
SL0-SL1	1	TF0	No effect	NA
BL1-SL0	1	TF1	IRF0	r0
T0-WL1	1,2	SA1F,CF	IRF0	r0
T0-SL1	1	SA1F	IRF1	r1
T0-BL1	1	SA1F	IRF1	r1
T0-T1	1	SA1F,CF	RDF	r0
T0-T2	2	SA1F,CF	RDF	r0
T0-T3	3	SA1F,CF	IRF1	r1
BL0-BL1	1	TF1	IRF1	r1
WL0-BL1	1	SA1F	RDF	r0
WL0-SL1	1	SA1F	IRF0	r0
WL0-WL1	1,3	SA0F,SAF1 CF	IRF0	r0

victim cell and those of an aggressor cell. To study the defect and fault models, we consider a 2x2 cell array, as shown in Fig. 5(d). We observe the presence of 13 possible defects that

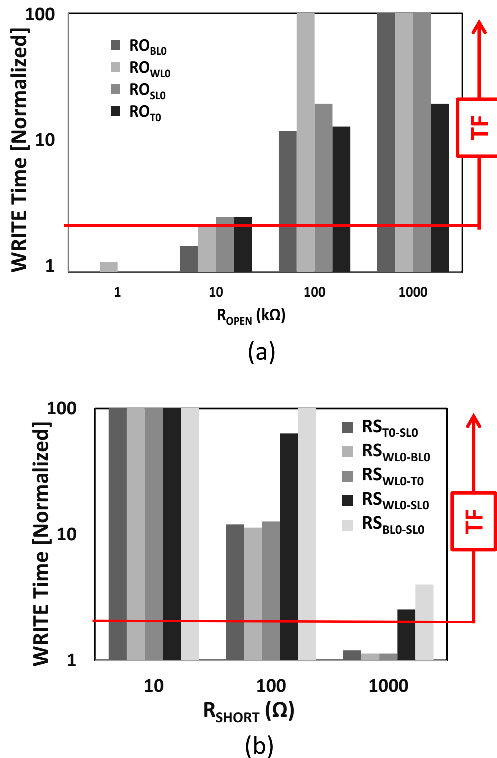

Fig. 6: Role of resistive: (a) opens from Fig. 6(a) on WR time and (b) shorts from Fig. 6(c) on WR time. The horizontal line shows the 6σ margin, above we see WR Transition Faults (TF).

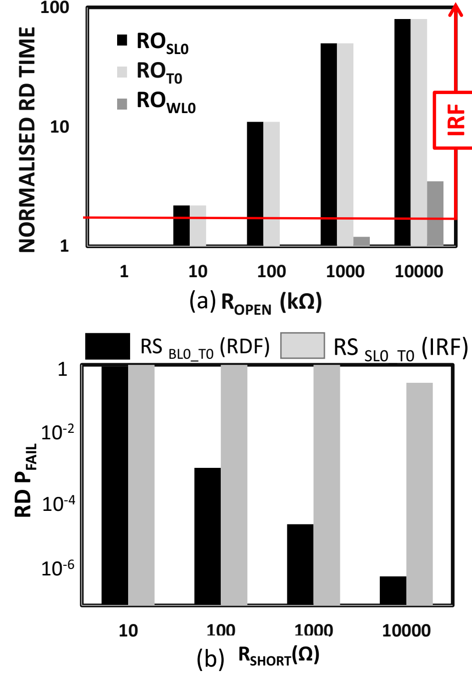

Fig. 7: (a) Resistive intra-cell opens cause IRF where any cell whose RD time is over the 6σ margin (horizontal line) has IRF. (b) Shorts across the MTJ can cause IRF due to degraded margin whereas shorts across the transistor can cause RDF due to high current. Corresponding P_{FAIL} for RD for different values of short is shown.

can affect cell RD/WR. The victim cell considered is Cell0 and the aggressors are cells 1, 2 or 3. In the inter-cell defects, apart from static CF's we observed dynamic fault models occurring due to data dependant CF's which have not been studied in [4][5]. These are faults that get activated when certain pattern is being written into the aggressor and victim cell simultaneously causing the bias voltages to the cells to get shorted amongst one another. These faults are observable clearly when the analysis is performed at a word level such that neighborhood cells when written together effect each other's writability.

Key Observations: The critical fault model is the ***data-dependent*** CF. As noted above, these arise in hybrid CMOS memory arrays because of the different bias conditions used for writing logic 1 and 0 and in few cases these conditions lead to undesired writes. For example, when writing 0 to both Cell0 and Cell1, if there is a bridge between BL0 and SL1 (Fig 5(d)) this leads to weakening of BL0 possibly leading to a TF0 (transition to 0 fault).The defects and fault models activated with these defects faults are shown in Table IV. The data patterns sensitizing these faults are not studied in this work. Shorts between T0 and WL1 lead to static coupling faults where if Cell2 is being read or written (WL1 is high), the short drives current through MTJ0 possibly switching its state inadvertently. In addition, a short between WL0 and WL1 can also cause both WL being simultaneously turned on (last row of Table IV) causing an inadvertent WR on cell-0.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.

Most of the inter cell defects activate IRFs. Inter cell defects occurring at T0 can potentially lead to RDF when the neighboring cell is being read or written as seen in Table IV.

V. CONCLUSIONS

This paper presents a comprehensive analysis of variations and defects in STT-MRAM. Fault models corresponding to the defect models have been discussed. The results and observations will enable test pattern generation for target fault coverage. It is observed that both Write and Read operations have high sensitivity to variations in threshold voltage of transistor Vt and the thickness of the dielectric Tox. Defects in WL have a very huge effect on the Write failure probability of the cell. Future work will involve identification of patterns sensitizing the inter cell defects and test pattern generation for the faults discussed. Temperature dependence of the parametric variation and interplay of defects and variations on faults are other interesting aspects to look at as an extension of this study.

Acknowledgements: Ashwin Chintaluri and Abhinav Parihar were supported by the Semiconductor Research Corporation (TASK 2493.001).

REFERENCES

[1] S. A. Wolf, et al. “The promise of nanomagnetics and spintronics for future logic and universal memory,” *Proc. IEEE*, vol. 98, no. 12, pp. 2155–2168, Dec. 2010.

[2] Kim, J.P et al. "A 45nm 1Mb embedded STT-MRAM with design techniques to minimize read-disturbance," *VLSI Circuits (VLSIC), 2011 Symposium on* , vol., no., pp.296,297, 15-17 June 2011.

[3] Noguchi, H et al. "7.5 A 3.3ns-access-time 71.2μW/MHz 1Mb embedded STT-MRAM using physically eliminated read-disturb scheme and normally-off memory architecture," *Solid-State Circuits Conference - (ISSCC), 2015 IEEE International* , vol., no., pp.1,3, 22-26 Feb. 2015.

[4] N. Z. Haron and S. Hamdioui , “On defect-oriented testing for hybrid CMOS/memristor memory,” in IEEE Asian Test Symp. (ATS), Nov. 2011.

[5] Chen, Yong-Xiao; Li, Jin-Fu, "Fault modeling and testing of 1T1R memristor memories," *VLSI Test Symposium (VTS), 2015 IEEE 33rd* , vol., no., pp.1,6, 27-29 April 2015.

[6] O. Ginez, J. Portal, and C. Muller, “Design and test challenges in resistive switching RAM (ReRAM): An electrical model for defect injections,” in IEEE European Test Symp. (ETS), May 2009, pp. 61–66.

[7] J.Z. Sun, “Spin-current interaction with a monodomain magnetic body:A model study,” *Phys. Rev. B*, vol. 62, no. 1, pp. 570–578, Jul. 2000.

[8] A V Khvalkovskiy *et al* 2013 *J. Phys. D: Appl. Phys.* **46** 139601

[9] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, “Design space and scalability exploration of 1T-1STT MTJ memory arrays in the presence of variability and disturbances,” in *Proc. IEEE IEDM*,Dec. 2009, pp. 1–4.

[10] Augustine, C et al."Numerical analysis of typical STT-MTJ stacks for 1T-1R memory arrays," *Electron Devices Meeting (IEDM), 2010 IEEE International* , vol., no., pp.22.7.1,22.7.4, 6-8 Dec. 2010.

[11] Panagopoulos, G.D.; Augustine, C.; Roy, K., "Physics-Based SPICE-Compatible Compact Model for Simulating Hybrid MTJ/CMOS Circuits," *Electron Devices, IEEE Transactions on* , vol.60, no.9, pp.2808,2814, Sept. 2013.

[12] Predictive Technology Model: http://ptm.asu.edu/latest.html.

[13] Jing Li; Augustine, C.; Salahuddin, Sayeef; Roy, K., "Modeling of failure probability and statistical design of Spin-Torque Transfer Magnetic Random Access Memory (STT MRAM) array for yield enhancement," *Design Automation Conference, 2008.*

[14] Yaojun Zhang; Xiaobin Wang; Yiran Chen, "STT-RAM cell design optimization for persistent and non-persistent error rate reduction: A statistical design view," *Computer-Aided Design (ICCAD), 2011 IEEE/ACM International Conference on* , vol., no., pp.471,477, 7-10 Nov. 2011.

[15] A. J. van de Goor and Z. Al-Ars, “Functional memory faults: a formal notation and a taxonomy,” in Proc. IEEE VLSI Test Symp. (VTS), 2000

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on July 04,2021 at 02:50:28 UTC from IEEE Xplore. Restrictions apply.