2018 IEEE 36th International Conference on Computer Design

Analysis of Row Hammer Attack on STTRAM

Mohammad Nasim Imtiaz Khan
School of Electrical Engineering and Computer Science
The Pennsylvania State University
University Park, U.S.A.
Email: muk392@psu.edu

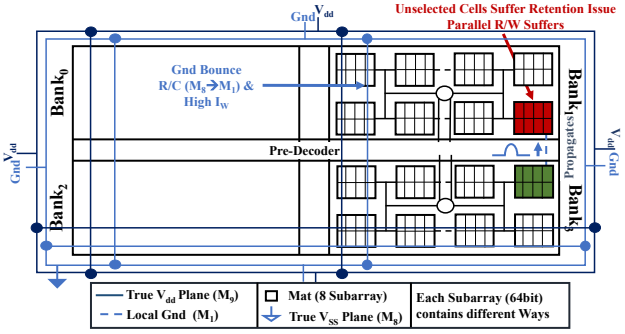
Swaroop Ghosh
School of Electrical Engineering and Computer Science
The Pennsylvania State University
University Park, U.S.A.
Email: szg212@psu.edu

***Abstract*—In this paper, we model and investigate the impact of Row Hammering (RH) on Spin-Transfer Torque RAM (STTRAM) by exploiting its write operation. STTRAM suffers from high write current and long write latency which can result in ground bounce. The magnitude of the bounce depends on the old data and the new data that is being written. The bounce can propagate to the nearest word-line drivers and partially turn ON the access transistors making weak current flow through the memory bitcells and reducing their thermal energy barrier. Therefore, continuous write at a particular location can force the massive number of unselected bits to suffer from degraded thermal barrier due to weak RH current. Reduced thermal barrier may lead to retention failures and make the bits sensitive to stray magnetic field/thermal noise. Those bits can also suffer from read disturb if they are read. These issues could be even worse for Short Retention NVM (SRNVM) which is suitable for Last Level Cache (LLC) and has a base retention of only few seconds. The ground bounce can also propagate to bit-line/source-line drivers and the selected cells will experience lower headroom voltage. This will lead to read failure (due to degraded sense margin) and write failure (due to increased write latency). Simulation result indicates that RH attack can flip the bits in just 30.84secs for STTRAM with base retention of 1 min. In presence of elevated temperature, the retention time can be further reduced to 2.46secs and 0.19secs for T=50C and T=75C respectively. RH attack can increase read disturb by 2.09X for bitcell with 1min base retention at T=25C. Simulation result also indicates that RH attack can cause read/write failure if the bitcell being read/written experience 306mV (for data 0)/110mV (for writing $0 \rightarrow 1$) of bounce. To the best of our knowledge, this is the first RH attack study for STTRAM-based cache.**

***Index Terms*—Spin-Transfer Torque RAM, Row Hammer, Non-Volatile Memory, Security**

I. Introduction

Spin-Transfer Torque RAM (STTRAM) [1] offers numerous benefits such as high-density, non-volatility, high-speed, low-leakage-power and CMOS compatibility, and therefore it is considered as a promising memory technology to replace Static RAM (SRAM) cache. Furthermore, STTRAM enables low-power computation and novel architecture [2]–[4]. However, it is susceptible to various security threats. It has been shown that sensitivity of STTRAM to magnetic fields can be exploited by the adversary to launch Denial-of-Service (DoS) attacks [5]. Another work shows that a mild application of magnetic field and/or temperature can also trigger soft performance failures by increasing write latency or degrading sense margin [6]. Apart from launching data integrity attacks, the adversary can also steal sensitive data from STTRAM cache (data privacy attacks) [7] by exploiting memory persistence. Temperature can be exploited as added means to enhance the persistence to recover stored data before power down [8]. The persistent user data in Non-Volatile Memory (NVM) cache can also be compromised by launching unauthorized read and write operations and probing the data buses after the authentic user has logged off [9]. In [10]–[12], it has been pointed out that STTRAM suffers from high write current and long latency which can be exploited to launch a side channel attack. In [13] and [14], it has been shown that high write current of emerging NVM can cause supply noise which can be leveraged to launch fault injection and information leakage attacks respectively. However, Row Hammer (RH) attack on STTRAM or other emerging NVMs has not been demonstrated.



STTRAM LLC (4MB)

Fig. 1: 1T1R-based 4MB LLC (containing 4 banks) showing ground bounce. Each bank contains 8 Mats and each Mat contains 8 subarrays each producing 64bits. Each subarray has 8 Ways. Parallel read/write in $Bank_1$ (red) suffers due to propagation of bounce from $Bank_3$ (green) (or vice versa). Similarly, unselected cells suffer from retention issue.

RH on traditional memories have been studied in the past. In [15], the authors investigate RH attack on Dynamic RAM (DRAM). Their investigation shows that it is possible to corrupt the data in nearby addresses by repeatedly reading from the same address. The authors demonstrate this phenomenon on Intel and AMD systems using a malicious program that generates many DRAM accesses. With new memory alterna-

This work is supported by SRC 2018-TS-2847, NSF CNS-1722557, CNS-1814710, CCF-1718474, DGE-1723687, DGE- 1821766 and DARPA Young Faculty Award [D15AP00089].

2576-6996/18/$31.00 ©2018 IEEE
DOI 10.1109/ICCD.2018.00021

tives, such as STTRAM, it is necessary to investigate their resistance to such known threats.

Fig. 1 shows a simplified diagram of STTRAM memory array with data bus width of 512-bit. Let's consider I_{write} = 100μA/bit. Therefore, the total write current is 51.2mA. If we consider the worst-case, all of the 512-bits will be written to one bank (green marked) and the total write current will be dumped into the local ground (which is routed in metal-1, M_1) of that particular bank. The local ground will experience a voltage bounce since there is a parasitic resistance between the local ground of that bank and true ground (routed on upper metal layer e.g. M_8) of the chip. This bounce will propagate to the word-line/source-line/bit-line drivers of the neighboring bits. If the bounce propagates to word-line drivers, the unselected bits sharing the same bit-line/source-line drives will partially turn the access transistor ON and a disturb current will pass through them. These bitcells will experience retention failure and read disturb. Furthermore, if the bounce propagates to source-line/bit-line drivers, the bit-cells will experience lower voltage headroom during read/write operation. Therefore, the operations may fail. In summary, the unselected (selected) bits will suffer from following issues if they experience a continuous disturb current (lower voltage headroom) during retention (read/write) operations:

Retention Failure: NVM retention can range from a few seconds (designed for LLC) to years (designed for storage alternatives). However, the disturb current through the bitcell reduces the thermal stability and subsequently, the retention time. Therefore, the time period for which the stored information in the bitcell is stored reliably (known as retention time) goes down. The issue is severe for Short Retention NVM (SRNVM) where the retention time is in the order of seconds [16].

Read Disturb: If a bitcell is accidentally flipped (written) during a read, it is known as read disturb. In the voltage sensing technique [17], a current is passed through the NVM bitcell to read its content. The read current is chosen in a way that does not disturb the content and at the same time achieves a good sense margin. However, the bits can flip during a read due to a lower thermal energy barrier (resulting from disturb current due to RH attack) creating the read disturb failure.

Read Failure: If a bitcell is read incorrectly, it is known as read failure. Bitcells experience lower voltage headroom if the bounce generated from a parallel operation propagates to the corresponding source-line drivers. Lower voltage headroom reduces the sense margin and may lead to read failure.

Write Failure: Bitcells experience lower headroom voltage if the bounce generated from a parallel operation propagates to the corresponding source-line drivers (for writing $1 \rightarrow 0$) or bit-line drivers (for writing $0 \rightarrow 1$). Lower voltage headroom increases the write latency and may lead to write failure.

Interestingly, the magnitude of ground bounce generated due to a write operation depends on the present state of the memory bit as well as the new data being written since I_{write} for $0 \rightarrow 0$, $0 \rightarrow 1$, $1 \rightarrow 0$ and $1 \rightarrow 1$ are different. Therefore, the adversary can employ worst-case patterns to create maximum bounce and accelerate the RH. In this work, first we model local ground bounce due to high write current, long write latency, and parasitic capacitance/resistance from local ground to the true ground. Bank-level parallelism (i.e., read/write to independent banks in parallel) is employed in LLC to achieve high bandwidth since single read/write takes multiple clock cycles. However, parallel access in STTRAM-based LLC can lead to elevated levels of bounce (worsen the RH of unselected bits). Furthermore, the bounce generated from one access can propagate to another access location and affect the corresponding read/write operation of those selected bits.

We investigate the impact of high parasitic capacitance/ resistance if the ground rail is designed same as conventional embedded memories such as, SRAM and eDRAM. Next, we discuss the impact of the ground bounce on the retention time of the bitcell of the adjacent rows to launch RH attack. To the best of our knowledge, this is the first effort in this direction.

Following contributions are made in this paper:

(a) We show that high write current of STTRAM leads to ground bounce. This can be exploited to launch RH attack;
(b) We model the ground bounce for STTRAM during write operation and perform detailed analysis of RH with respect to ground rail parasitic, such as resistance and capacitance;
(c) We also consider the impact of bank-level parallelism and ambient temperature on the effectiveness of RH attack;

The rest of the paper is organized as follows: Section II describes basics of STTRAM and simulation details; Section III describes the attack model, ground bounce model and corresponding simulation result; Section IV describes the simulation result and analysis; Section V draws conclusion.

II. Basics of STTRAM and Simulation Environment

In this section, we discuss the basics of STTRAM. We also present the simulation environment used in this work.

A. Basics of STTRAM

Fig. 2a shows the STTRAM cell schematic with Magnetic Tunnel Junction (MTJ) as the storage element. The MTJ contains a free (FL) and a pinned (PL) magnetic layer. The resistance of the MTJ stack is high (low) if FL magnetic orientation is anti-parallel (parallel) compared to the PL. The MTJ can be toggled from parallel (P) (data '0') to anti-parallel (AP) (data '1') (or vice versa) using current induced Spin Torque Transfer by passing the appropriate magnitude of write current from source-line to bitline (or vice versa). For a successful write, the write current must be greater than the critical current (I_{co}). During read operation, the access transistor is turned ON, and a small voltage is applied on the bit-line while keeping the voltage of source-line zero.

The two stable magnetization states of FL are separated by an energy barrier, E_0 (Fig. 2b). The thermal stability factor of the MTJ is denoted by Δ_0, and the relation between E_0 and Δ_0 is given by (1):

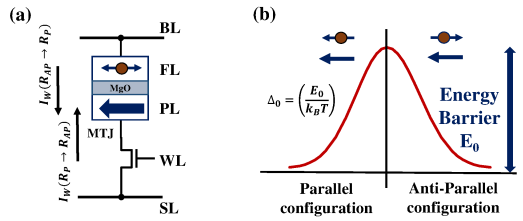
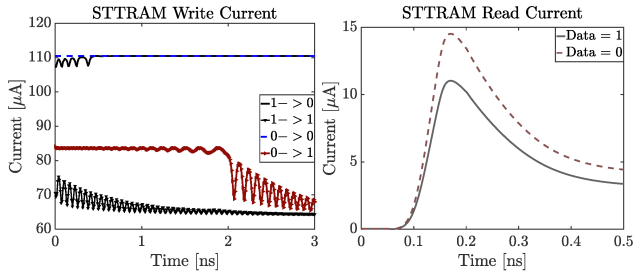


Fig. 2: Schematic of STTRAM bitcell and, (b) energy barrier, E_0 separating the two MTJ states.



(a) Write current (b) Read current

Fig. 3: STTRAM (a) write current; and (b) read current.

$$\Delta_0 = \frac{E_0}{k_B T} = \frac{H_k M_s V}{2 k_B T} \tag{1}$$

where H_k is the effective anisotropy field including magneto-crystalline anisotropy and shape anisotropy of the FL of MTJ [18], M_s is the saturation magnetization, V is the volume of the MTJ FL, k_B is the Boltzmann's constant, and T is the temperature.

In presence of current through MTJ, the thermal stability factor is given by (2) [19] and the retention time of the STTRAM bitcell is given by (3) [20]:

$$\Delta_I = \Delta_0 * (1 - \frac{I}{I_{co}}) \tag{2}$$

$$t_{ret} = t_0 * exp[\Delta_I] \tag{3}$$

where Δ_I is the thermal stability factor in presence of the current through MTJ, I_{co} is the critical current of MTJ, I is the current flowing in the direction which tries to flip the current state, t_{ret} is the retention time in presence of the current and t_0 is attempt time (~1ns).

B. High Write/Read Current

Write/read current of STTRAM, and the corresponding latencies are high (Fig. 3a-3b). Typically, the write current is 100-200μA/bit and the read current is ~5X less compared to write current. Although the read current is also high compared to conventional memories, it is not significant enough to cause high ground bounce which might partially turn ON the access transistor. Therefore, in this work, we investigate the RH attack on STTRAM exploiting the write operation only.

C. Simulation Environment

The STTRAM cell used in this work consists of a MOSFET of 65nm technology and an MTJ based on [21]. The simulations are performed in HSPICE. The parameters used in the simulation are provided in Table I. The critical current of the MTJ cell is found to be ~97μA. Therefore, 110μA/bit of average write current is considered for modeling RH attack.

TABLE I: Parameters Used for the Simulation

Parameter	Value
Threshold Voltage of Access Transistor (V_T)	0.423V
Reduction of V_T with Temperature	-2mV/K
Volume of MTJ free layer	1.04x10^{-17}cm^3
Uniaxial Anisotropy of MTJ, K_u	150150erg/cc
Saturation magnetization of MTJ, M_s	790Oe
Anisotropy Magnetic field of MTJ, H_k	380Oe
Thermal Barrier of MTJ, Δ	37.99
Tunnel magnetoresistance (TMR)	119%
Read latency/Write latency of STTRAM	1ns/3ns

III. Attack Model

In this section, we discuss RH attack modeling for STTRAM. We present basics of ground bounce and describe its modeling. We also present the basics of bank-parallelism in LLC for high system throughput.

A. Modeling RH on STTRAM

RH attack is launched on DRAM by reading one particular row many times. This causes a disturbance to the bits of the neighboring rows. When reading one address many times, the corresponding word-line voltage is repeatedly toggled. As a result, cells sharing the same column leak charge much faster [15]. Those cells cannot retain charge till the time interval at which they are refreshed. Therefore, the cells lose the stored data. As STTRAM stores data through magnetic orientation, traditional RH needs to be modified (described below) to be effective.

RH Attack by Exploiting Write Operation: One particular address is written multiple times to launch RH attack by exploiting write operation of STTRAM. The total write current gets dumped into the ground rail. This leads to a large ground bounce if the ground rail is designed conventionally (i.e., without consideration to high write current). Note that the bounce will be present even with a carefully designed

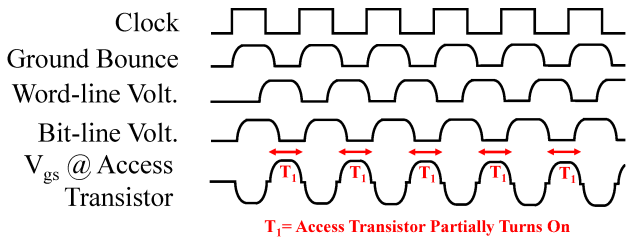


Fig. 4: Ground bounce propagates with a delay to the farther word-line drivers which results in a phase shift between bounce of word-line and source-line. Access transistor partially turns ON for time, T_1.

ground rail. The bounce propagates to the peripherals such as, word-line, bit-line, and source-line drivers. These can cause the following issues:

i) Bounce propagates to word-line and source-line or bit-line drivers (retention failure and read disturb): The nearest unselected bits, (in case of writing $1 \rightarrow 0$) whose source-line (or bit-line in case of writing $0 \rightarrow 1$) drivers share the same supply rails as source-line (bit-line) drivers of the selected cells, have zero V_{GS} at the corresponding access transistor since word-line and source-line (or bit-line) bounce together. However, the bounce propagates with a delay to the farther word-line drivers (due to different path delay) which results in a phase shift between the bounce of word-line and source-line. Therefore, those access transistors will experience a brief period when they will weakly turn ON i.e., the V_{GS} will be greater than 0V (Fig. 4). This will introduce disturb current through those unselected cells, and they will eventually be written to either '0' or '1' (depending on the direction of disturb current) if the disturb current flows for duration longer than the reduced retention time. Furthermore, if the attack takes place at an elevated chip temperatures, the threshold voltage of the access transistor and the retention time of the MTJ (3) will be lowered, and the current through the MTJ will be higher, leading to faster corruption of the bits and more effective attack. Therefore, by writing a particular address repeatedly, the massive number of unselected bits whose bit-line or source-line drivers share the same supply rails as bit-line or source-line drivers of the selected cells can be written/flipped.

It should be noted that the disturb current lowers the thermal barrier. Therefore, if the partially selected bits in other independent banks are read, the probability of read disturb of these bits will increase.

Note that process variation can amplify these issues further since the weak unselected bits (with lower thermal stability) and low access transistor threshold can easily get corrupted.

ii) Bounce propagates to source-line only (read failure): Let's assume that adversary is writing in a bank (i.e. generating ground bounce) and the victim is reading data from another independent bank. Therefore, the bitcells that are being read by the victim, will have a zero source-line, and a non-zero word-line and bit-line voltage (Fig. 5a). If the bounce generated by the adversary reaches the bitcells that are being read, the read operation will incur a lower sense margin due to a lower voltage headroom (since source-line voltage bounces) (Fig. 5a). Therefore, read failures may occur if the decreased sense margin is too low.

iii) Bounce propagates to source-line or bit-line (write failure): Let's assume that adversary is writing in a bank (i.e. generating ground bounce) and victim is writing in another independent bank. Therefore, the bitcells that are being written by the victim will have a zero source-line (for $1 \rightarrow 0$ writing) or bitline (for writing $0 \rightarrow 1$) (Fig. 5a). If the ground bounce generated by the adversary reaches those bitcells, the write operation will incur a longer write latency due to lower voltage headroom (since source-line or bitline voltage bounces) (Fig. 5a). Therefore, write failures may occur if the increased write latency is greater than the design target (3ns in this work).

B. Modeling the Ground bounce

Fig. 5b shows a simplified illustration of STTRAM memory array. One global column is connected to eight local columns and a write driver. All the write drivers share a local ground through metal-1 (M_1). True ground of the chip, which is implemented using metal-8 (M_8), connects to the local ground (M_1) via M_2, M_3, M_4, M_5, M_6 and M_7 metal layers. The connection from M_1 to M_8 is repeated only after a specific distance to ensure high density of memory array. Therefore, parasitic capacitance and resistance exists between the local ground and true ground. When the total write current is dumped into the local ground, it creates a bounce which propagates to the peripherals such as, word-line, bit-line, and source-line drivers as shown in Fig. 5c.

We modeled the resistance and capacitance of path M_1 to M_8 (R_1, C_1 and C_2) and the path between the local ground to the gate of access transistors (R_2, R_3 and C_3) of the unselected bit (Fig. 5d). We have designed the metal plan for the 4MB 1T1R LLC organization shown in Fig. 1. It is a 4-way set associated cache. All the Ways of each Mat are accessed simultaneously and buffered at the edge of each Mat resulting in total of 512-bit accesses. The figure also shows the upper layer metal plan. V_{dd} plane is in M_9 and gnd plane is in M_8. Both V_{dd} and gnd is implemented from M_7 to M_1 where M_7, M_5, M_3 and M_1 are horizontal and M_6, M_4, M_2 are vertical. The total area of the chip is $4970\lambda \times 3950\lambda$ where each bank occupies $2046\lambda \times 1536\lambda$ and the remaining is occupied by the peripheral circuitry (e.g. pre-decoder etc.). Note that λ is the feature size. Below we present ground bounce modeling:

Ground Bounce: The total write current is dumped to the local ground (implemented in M_1) causing a ground bounce which can propagate to nearest banks though metal M_1 via metal M_2, and then down to M_1 again. We modeled the resistance of path M_1 to M_8 by R_1. Fig. 6 shows the connection of true ground (M_8) with the local ground (M_1) of a Mat. We modeled the equivalent resistance using 65nm layout parameters (Table II) [22], [23]. We divided 512-bits to 4 groups (only two of them shown in Fig. 6 for simplicity). Each metal layer R/C and via between metal layers are also given in Table II. Our estimation shows that R_1 is ~25Ω (Fig. 6). The magnitude of ground bounce depends on this value. Fig. 7a shows that as R_1 increases, bounce increases. Capacitance calculation is omitted for the sake of brevity.

The impact of I_{write} (per bit) and the total number of bits writing are shown in Fig. 7b and Fig. 7c respectively. Fig. 7b shows that as I_{write} (per bit) increases, ground bounce increases. Typically, STTRAM I_{write} varies from 100μA to 200μA. Fig. 7c shows that as the number of bits writing simultaneously in a memory array increases, ground bounce also increases. Average write current for a full cache line is divided into four constant Current Sources (CS). Therefore, current magnitude of CS, XmA is equal to $I_{Total}/4$ (for e.g.

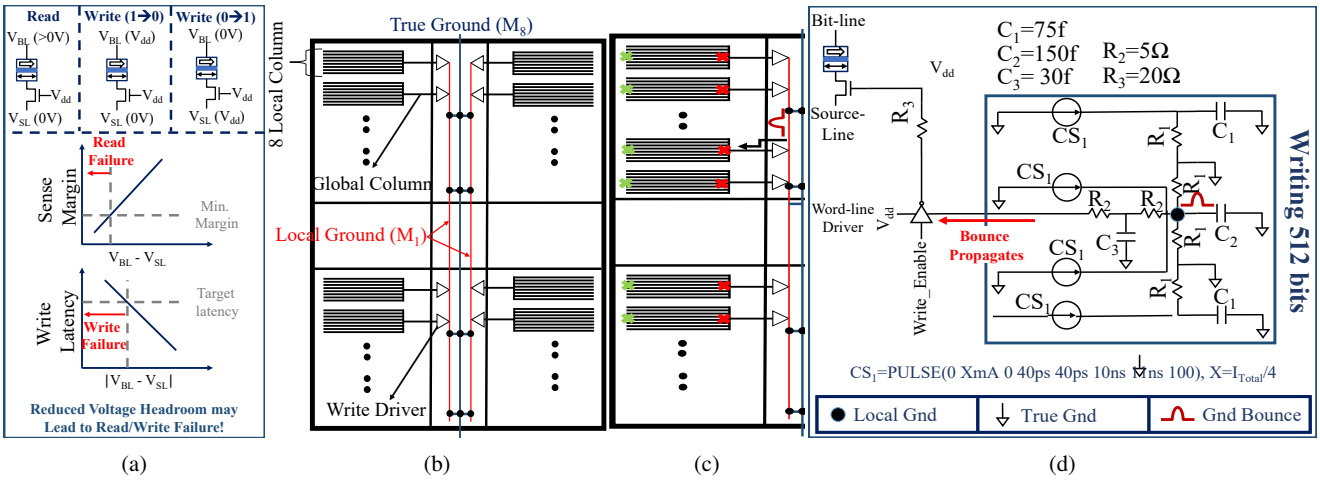


Fig. 5: Reduced voltage headroom decreases sense margin and increases write latency and may lead to read/write failure; (b) illustration of memory array. The local ground is connected to the true ground of the chip with periodic gap; (c) generated ground bounce due to writing the selected bits (green cross) affects the unselected bits (red cross); (d) modeling parasitic R and C of the ground rail and the ground voltage bounce due to high write current. The figure also shows that the generated bounce propagates to the word-line drivers.

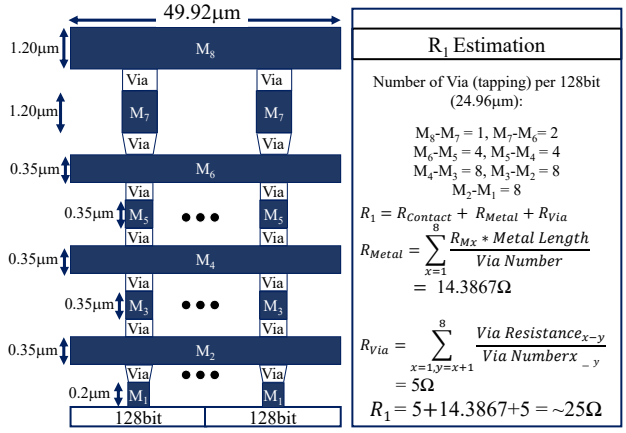


Fig. 6: Estimation of R_1 (of Fig. 5d) for ground bounce modeling.

512-bit of STTRAM $0 \rightarrow 0$ writing, I_{Total} = 56.32mA and X = 14.08) and each one presents write current for 128-bit.

TABLE II: Parameters used for Ground Bounce Modeling

Parameter	Value
Resistance (Ω/µm) $M_1/M_2/M_3/M_4/M_5/M_6/M_7/M_8$	0.91/0.41/0.41/0.41/0.41/ 0.41/0.04/0.04 [22], [23]
Capacitance (fF/µm) $M_1/M_2/M_3/M_4/M_5/M_6/M_7/M_8$	0.13/0.17/0.17/0.17/ 0.17/0.17/0.19/0.19 [22], [23]
Miller Coupling Factor (MCF)	1.5
Via Resistance (Ω) M_{1-2}/ $M_{2-3}/M_{3-4}/M_{4-5}/M_{5-6}/M_{6-7}/M_{7-8}$	6/5/5/3/3/1/1 (CVD Tungsten-based) [24]
Di-electric Constant for Cap. Calculation (C_{plate}/C_{side})	2.2/2.79 [23]
Res. between M_1 to Source/ Drain Contact, $R_{Contact}$ (Ω)	~5 [25]

In summary, we assume 512-bit writing with an average write current of 110μA/bit, a write latency of 3ns and R_1 = 25Ω for further analysis. The current source mimics the write current drawn by STTRAM. The period and ON time (i.e., write latency of STTRAM) of the current source is assumed to be 6ns and 3ns respectively. A period of 6ns (twice compared to ON time) is used to consider a gap of 3ns in between two consecutive writes. The following assumptions are made to model the worst-case retention failure:

(i) The gate leakage of the access transistor (unselected cell) is negligible;
(ii) The phase difference of ground bounce propagating to word-line and source-line driver is maximum (3ns).

C. Parallel Read/Write Operation

STTRAM write latency requires multiple clock cycles. For example, the required number of clock cycles are 5 (1) with a clock frequency of 2GHz and write (read) latency of 2.5ns (0.5ns). However, the throughput will degrade if memory access is completely stopped during 5 or 1 cycles (Fig. 8a-8b). In practice, STTRAM write latency is even higher (3ns, i.e. 6 cycle for this work). Therefore, parallelism is used to perform write/read simultaneously to different independent banks and increase system throughput. The parallel access can take following forms:

1X Write: Read can be initiated in the next 4 cycles in other independent banks (Fig. 8a) when a write has been initiated in a bank. These data are processed in a pipeline to maintain high throughput. We call this write/read scheme 1X Write.

nX Write: Multiple (n) writes can be initiated with read. For example, one write along with 3 consecutive reads can be initiated in the next four clock cycles in other independent banks (Fig. 8a) when a write has been initiated in a bank. The local ground will bounce due to second write along with multiple reads and propagate to unselected bits. We call this write/read scheme 2X (n = 2) Write.

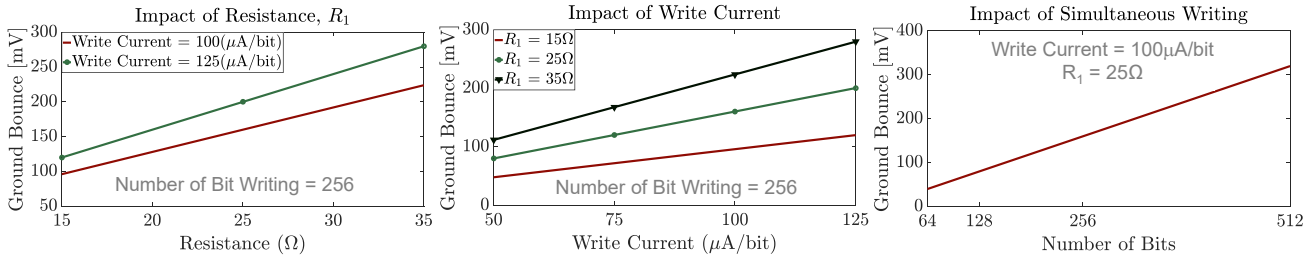


Fig. 7: Impact of (a) resistance (R_1), (b) I_{write} (per bit) and, (c) number of bits writing on the local ground voltage.

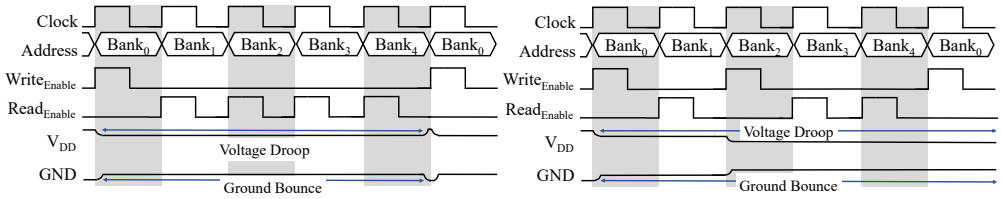


Fig. 8: (a) Four reads are initiated between two writes. We call it 1X mode; (b) three reads and one write are initiated between two writes. We call it 2X mode (n = 2).

We describe four types of failure in STTRAM due to ground bounce, namely retention failure, read disturb, read failure and write failure. We have assumed that parallel operations are performed on different independent banks. Parallel accesses can worsen retention failure since the adversary can exploit it to increase the magnitude of ground bounce which can propagate to more bits and/or increase the disturb current triggering faster retention failures.

IV. Simulation Results and Analysis

A. Retention Failure

At first, we consider an MTJ with Δ_0 = 37.99 (base retention = 1 year) and volume = 1.04×10^{-17}cm^3. The ground bounce is ~352.46mV with a write current of 110μA/bit and a cache line of 512-bit. The disturb current is found to be 1.88μA considering this voltage (ground bounce) propagates to the gate of access transistor of the unselected bits whose source-line/bit-line drivers share the same supply rails as source-line/bit-line drivers of the selected cells. The retention time of those bits becomes 174.94 days with a disturb current of 1.88μA. Therefore, an adversary has to keep writing and generating ground bounce so that the unselected cells experience the disturb current for 174.94days continuously and gets flipped. This may not be a feasible attack scenario.

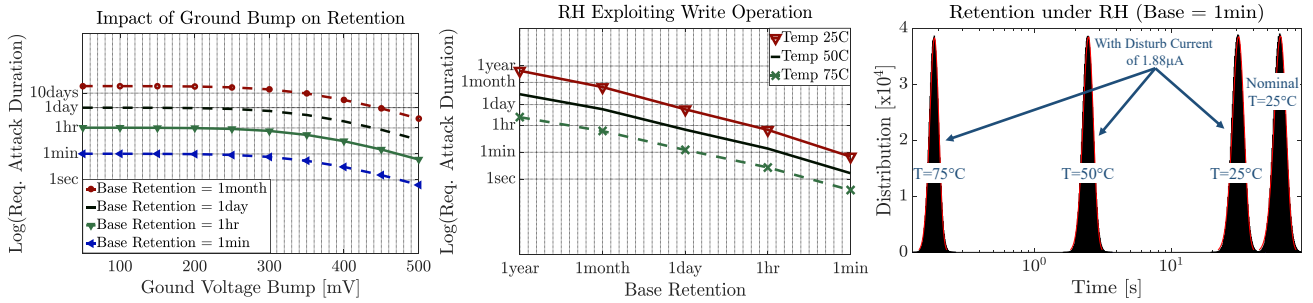
Next, we investigate the impact of RH attack on the write operation of STTRAM with different base retention time (Table III) to account different target design such as SRNVM. The result is plotted in Fig. 9a. It can be seen that the retention time of the cells suffer more as the base retention time reduces. RH attack can flip the bits in 30.84secs if the base retention is 1 min. Furthermore, the required attack duration reduces (shown in Fig. 9b) if high external temperature is applied during the attack. The attack duration for base retention of 1min can be 2.46secs and 0.19secs at T = 50°C and T = 75°C respectively.

We further analyze the impact of RH attack on SRNVM STTRAM (base retention 1min) under process variation. A 1-million-point Monte-Carlo analysis is conducted with 3σ of 2% of MTJ thermal stability factor, Δ_0 and with a mean of Δ_0 = 24.85 (corresponding mean retention 1min). The result is shown in Fig. 9c. It is evident that RH attack is more impactful on the weaker bits under process variation as the minimum required attack duration reduced to 19.41secs, 1.63secs and 0.13secs at T = 25°C, T = 50°C and T = 75°C respectively.

TABLE III: Volume and Thermal Stability of MTJs with Different Base Retention Time

Volume (cm^3)	Thermal Stability, Δ_0	Base Retention Time
1.041×10^{-17}	37.99	~1year
0.973×10^{-17}	35.50	~1month
0.845×10^{-17}	32.10	~1day
0.758×10^{-17}	28.95	~1hr
0.681×10^{-17}	24.85	~1min

Read Disturb: A small read current is passed through the bitcells during read operation. Higher read current gives better sense margin but increases read disturb probability. Therefore, the read current is selected in a way that does not flip the bit as well as yields good sense margin. However, disturb current due to ground bounce lowers the thermal barrier of the bitcell. If the bitcell is read at lower thermal barrier, switching probability during read operation (read disturb) increases [27]. Simulation result indicates that the switching probability increases from 1.01×10^{-09} to 2.06×10^{-09} (2.06X increment) for mean base retention of 1min (corresponding thermal barrier Δ_0 = 24.85) at T=25°C. A 1-million-point Monte-Carlo analysis is conducted with 3σ of 2% of MTJ thermal stability factor, Δ_0 with a mean of Δ_0 = 24.85 (corresponding mean retention time 1min). The worst-case



(a) (b) (c)

Fig. 9: (a) Impact of ground voltage bounce on retention time of unselected bits (base retention = 1month); (b) impact of RH attack on STTRAM write operation for different base retention time; and, (c) retention distribution under RH attack for base retention = 1min. A 1-million-point Monte-Carlo analysis is conducted with 3σ of 2% of MTJ thermal stability factor, Δ_0 with a mean of Δ_0 = 24.85 (corresponding retention time ~1min).

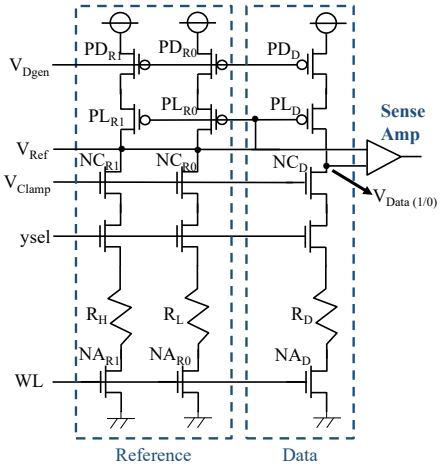


Fig. 10: Non-destructive read circuitry used it this work [26].

switching probability is found to be 2.9x10^{-07}. Furthermore, higher temperature further increases the switching probability.

Read Failure: Read failure may occur if the bitcell being read experience ground bounce (generated by a parallel access in another independent bank) in its source-line. Fig. 10 shows the single-ended read circuitry used it this work [26]. We analyze the sense margin for a bitcell with a volume of 1.041×10^{-17}cm3 (corresponding thermal stability, Δ_0 = 37.99 and base retention time 1year). The low/high resistance of the cell is 1.864KΩ/4.077KΩ which represents data '0'/'1' respectively. Two parallel branches with one low and one high resistance cell is used in the read circuit to generate the reference voltage. $V_{Data(1/0)}$ is compared with this reference voltage to sense the data using a sense amplifier. Fig. 11a shows the sense margin for both data '0' and '1' with respect to ground bounce. We have considered read error if sense margin is below 150mV. It is evident that as the ground bounce experienced by a bitcell during a read operation increases, sense margin for data '1' reduces whereas sense margin for data '0' stays relatively constant. This is because the reference voltage also increases with ground bounce (Fig. 11a). We can conclude that if the bitcell incurs ground bounce > 306mV during read, the operation reads '1' incorrectly. However, read '0' does not fail as it is insensitive to ground bounce.

Write Failure: Write failure may occur if the bitcell being written experience ground bounce (generated by a parallel access in another independent bank) in its source-line (for writing $1 \rightarrow 0$) or bit-line (for writing $0 \rightarrow 1$). Fig. 11b and 11c represents the impact of ground bounce on $0 \rightarrow 1$ and $1 \rightarrow 0$ writing respectively. It is evident that $0 \rightarrow 1$ writing fails if the bitcell experience 110mV of ground bounce as the magnetic orientation (M_x) is not reaching -1 (anti-parallel state). However, $1 \rightarrow 0$ write failure might not be possible as even with 400mV of ground bounce the magnetic orientation (M_x) successfully reaches 1 (parallel state). Therefore, $1 \rightarrow 0$ write failure requires very high ground bounce.

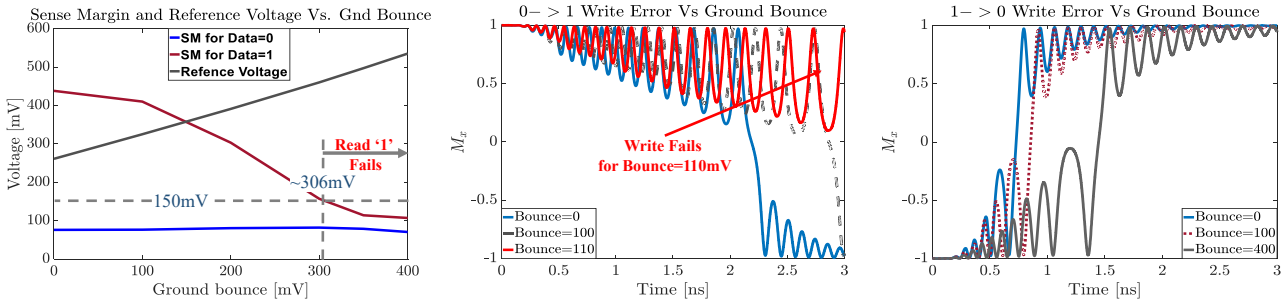
All the aforementioned impacts of ground bounce can be further enhanced if parallel write operations are performed in independent bank which can evidently double the ground bounce (for two parallel write operation). Furthermore, the weak bits considering the process variation suffers the most. At first glance, RH attack on STTRAM might not seem severe compared to DRAM. However, in contrasts to RH attack on DRAM which only cause data corruption, the RH attack on STTRAM can cause data corruption (retention failure, read disturb) and fault injection (read/write failure).

V. Conclusion

We studied the impact of RH attack on STTRAM by exploiting high write current. Although RH attack on STTRAM is not as severe as DRAM, the attack can create different types of failures and affect more bitecells. The weak bits due to process variation may suffer the most. RH attack can cause retention issues as well as read disturb if read operation is performed while the cells incur disturb current due to ground bounce. The attack can also cause read/write failure. Furthermore, the attack can be worse for SRNVM.

References

[1] T. Ohsawa, H. Koike, S. Miura, H. Honjo, K. Tokutome, S. Ikeda, T. Hanyu, H. Ohno, and T. Endoh, "1mb 4t-2mtj nonvolatile stt-ram



(a) (b) (c)

Fig. 11: (a) Sense margin degradation as the bitcell being read experience higher ground bounce; write latency for (b) $0 \rightarrow 1$ and (c) $1 \rightarrow 0$ increases as the bitcell being written experience higher ground bounce.

for embedded memories using 32b fine-grained power gating technique with 1.0ns/200ps wake-up/power-off times," in *2012 Symposium on VLSI Circuits (VLSIC)*, pp. 46–47, June 2012.

[2] A. Chen, "A review of emerging non-volatile memory (nvm) technologies and applications," *Solid-State Electronics*, vol. 125, pp. 25 – 38, 2016. Extended papers selected from ESSDERC 2015.

[3] S. Motaman, M. N. I. Khan, and S. Ghosh, "Novel application of spintronics in computing, sensing, storage and cybersecurity," in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 125–130, March 2018.

[4] C. J. Xue, G. Sun, Y. Zhang, J. J. Yang, Y. Chen, and H. Li, "Emerging non-volatile memories: Opportunities and challenges," in *2011 Proceedings of the Ninth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, pp. 325–334, Oct 2011.

[5] J.-W. Jang, J. Park, S. Ghosh, and S. Bhunia, "Self-correcting STTRAM under magnetic field attacks," in *Proceedings of the 52Nd Annual Design Automation Conference*, DAC '15, (New York, NY, USA), pp. 77:1–77:6, ACM, 2015.

[6] J.-W. Jang and S. Ghosh, "Performance impact of magnetic and thermal attack on sttram and low-overhead mitigation techniques," in *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*, ISLPED '16, pp. 136–141, ACM, 2016.

[7] N. Rathi, S. Ghosh, A. Iyengar, and H. Naeimi, "Data privacy in non-volatile cache: Challenges, attack models and solutions," in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 348–353, Jan 2016.

[8] X. Bi, H. Li, and J. Kim, "Analysis and optimization of thermal effect on stt-ram based 3-d stacked cache design," in *2012 IEEE Computer Society Annual Symposium on VLSI*, pp. 374–379, Aug 2012.

[9] I. Jacobs and C. Bean, *Fine Particles, Thin Films, and Exchange Anisotropy: (effects of Finite Dimensions and Interfaces on the Basic Properties of Ferromagnets)*. General Electric. Research Laboratory. Tecnical Information Series, Research Information Section, The knolls, 1963.

[10] Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen, L.-C. Wang, and Y. Huai, "Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory," *Journal of Physics: Condensed Matter*, vol. 19, no. 16, p. 165209, 2007.

[11] K. Shamsi and Y. Jin, "Security of emerging non-volatile memories: Attacks and defenses," in *2016 IEEE 34th VLSI Test Symposium (VTS)*, pp. 1–4, April 2016.

[12] M. N. I. Khan, S. Bhasin, A. Yuan, A. Chattopadhyay, and S. Ghosh, "Side-channel attack on STTRAM based cache for cryptographic application," in *2017 IEEE International Conference on Computer Design (ICCD)*, pp. 33–40, Nov 2017.

[13] M. N. I. Khan and S. Ghosh, "Fault injection attacks on emerging non-volatile memory and countermeasures," in *Proceedings of the 7th International Workshop on Hardware and Architectural Support for Security and Privacy*, HASP '18, (New York, NY, USA), pp. 10:1–10:8, ACM, 2018.

[14] M. N. I. Khan and S. Ghosh, "Information leakage attacks on emerging non-volatile memory and countermeasures," in *Proceedings of the International Symposium on Low Power Electronics and Design*, ISLPED '18, (New York, NY, USA), pp. 25:1–25:6, ACM, 2018.

[15] Y. Kim, R. Daly, J. Kim, C. Fallin, J. H. Lee, D. Lee, C. Wilkerson, K. Lai, and O. Mutlu, "Flipping bits in memory without accessing them: An experimental study of dram disturbance errors," in *2014 ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)*, pp. 361–372, June 2014.

[16] C. W. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, and M. R. Stan, "Relaxing non-volatility for fast and energy-efficient stt-ram caches," in *2011 IEEE 17th International Symposium on High Performance Computer Architecture*, pp. 50–61, Feb 2011.

[17] Z. Sun, H. Li, Y. Chen, and X. Wang, "Voltage driven nondestructive self-reference sensing scheme of spin-transfer torque memory," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, pp. 2020–2030, Nov 2012.

[18] Z. Sun, X. Bi, H. Li, W. Wong, Z. Ong, X. Zhu, and W. Wu, "Multi retention level stt-ram cache designs with a dynamic refresh scheme," in *2011 44th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pp. 329–338, Dec 2011.

[19] M. N. I. Khan, A. S. Iyengar, and S. Ghosh, "Novel magnetic burn-in for retention and magnetic tolerance testing of sttram," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, pp. 1508–1517, Aug 2018.

[20] 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 *2009 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–4, Dec 2009.

[21] S. Srinivasan, *All spin logic: Modeling multi-magnet networks interacting via spin currents*. PhD thesis, 2012. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2016-03-11.

[22] "Interconnect: Capacitance and Resistance for 65nm technology." http://ptm.asu.edu/, 2005. [Online; accessed May-03-2018].

[23] "Wire Capacitance and Resistance Calculator for 65nm." http://users.ece.utexas.edu/~mcdermot/vlsi-2/Wire_Capacitance_and_Resistance_65nm.xls, 2008. [Online; accessed May-03-2018].

[24] I. Shao, J. M. Cotte, B. Haran, A. W. Topol, E. E. Simonyi, C. Cabral, and H. Deligianni, "An alternative low resistance MOL technology with electroplated rhodium as contact plugs for 32nm CMOS and beyond," in *2007 IEEE International Interconnect Technology Conferencee*, pp. 102–104, June 2007.

[25] X. Li, W. Zhao, Y. Cao, Z. Zhu, J. Song, D. Bang, C. C. Wang, S. H. Kang, J. Wang, M. Nowak, and N. Yu, "Pathfinding for 22nm CMOS designs using Predictive Technology Models," in *2009 IEEE Custom Integrated Circuits Conference*, pp. 227–230, Sept 2009.

[26] J.-H. Song, J. Kim, S. H. Kang, S.-S. Yoon, and S.-O. Jung, "Sensing margin trend with technology scaling in mram," *International Journal of Circuit Theory and Applications*, vol. 39, no. 3, pp. 313–325.

[27] T. Zheng, J. Park, M. Orshansky, and M. Erez, "Variable-energy write stt-ram architecture with bit-wise write-completion monitoring," in *International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 229–234, Sept 2013.