Test Methodologies, and, Test Time Analysis and Compression for Emerging Non-Volatile Memory

Mohammad Nasim Imtiaz Khan, *Member, IEEE*, and Swaroop Ghosh, *Senior Member, IEEE*

Abstract*— Emerging Non-Volatile Memories (NVMs) are considered as suitable candidates to replace conventional memories such as Static RAM (SRAM) and Dynamic RAM (DRAM) due to high density, high performance and low (static) power operation. However, NVMs bring new fault issues and call for new tests. For example, NVMs exhibit wide read and write latency distribution, incur high write current (leads to high supply noise), susceptible to external magnetic/thermal field, show high and stochastic retention time, and, prone to endurance and reliability failures. The conventional tests either cannot capture the faults specific to emerging NVMs or they incur significant test time if implemented on emerging NVMs. In this work, we summarize faults models specific to NVMs and explain the related test issues and challenges. We propose new tests along with necessary Design-For-Test (DFT) techniques to characterize the failures. We further summarize NVM tests proposed in prior works and analyze their test time requirement.

Index Terms*— NVM Tests, Test Time Analysis, Read/Write Latency Test, Supply Noise Test, Magnetic Tolerance Test, Thermal Tolerance Test, Retention Test, Endurance/Reliability Test.

I. Introduction

At the end of CMOS scaling, a number of emerging Non-Volatile Memory (NVM) technologies e.g., Spin-Transfer Torque RAM (STTRAM) [1], Magnetic RAM (MRAM) [2], Resistive RAM (RRAM) [3], Phase Change Memory (PCM) [4] and Ferroelectric RAM (FRAM) [5] are considered promising. This is primarily due to high density, high speed, low (static) power operation and the persistence (non-volatility). Discrete chips of these memories have already penetrated the market. Some examples are MRAM/STTRAM by Everspin [6], Conductive Bridging RAM (CBRAM) (a variant of RRAM) [7] by Adesto Technology, PCM (Solid State Drive) by Intel Corporation [8] and FRAM by Cypress [9]. However, the unique characteristics of emerging NVMs bring new test issues and requires new test flows and/or repurposing the conventional test flows. A well-defined test flow which can capture all the NVM-specific faults will certainly ease the wider adoption of these technologies in a variety of applications. The test flow for conventional memories such as, Static RAM (SRAM) and Dynamic RAM (DRAM) is summarized in Fig. 1(a). Die-level tests include functional verification using various March tests. These tests characterize the memory address decoding, read and write operations, and identify various failures. Some example includes coupling faults, stuck-at faults, etc. Die-level tests also identify optimum values for various assist techniques. Some of the advanced assist techniques for SRAM and DRAM are power (supply/ground) gating during idle operation, supply voltage collapse [10] during a write operation, wordline overdrive (WLOV) during a write operation and wordline underdrive (WLUD) during a read operation, and negative bit-line during write operation [11]. For DRAM and embedded DRAM (eDRAM), data retention time is also identified [12].

Existing test flow is not capable of testing emerging NVM-specific features. The rationale is provided below:

Retention time: NVM retention time varies from a few seconds to several years. Special retention tests are performed on DRAM and eDRAM to certify their refresh rate (typically varies from 0.2ms to tens of ms [12]). These retention tests do not increase total test time notably. The same technique is not applicable to NVMs since it can lead to years of retention test time. Furthermore, STTRAM/MRAM retention time varies due to Process Variation and Temperature (PVT) and the same bit can show different retentions if measured multiple times (stochastic in nature) [13]. This makes retention characterization complex and time/energy consuming. Therefore, new techniques are required to characterize retention in short time.

Long read/write latency: SRAM latency is in the order of a fraction of nanosecond [14]. Furthermore, the latency is not a strong function of PVT. DRAM and eDRAM latency could be high but not as sensitive to Process Variation (PV). On the other

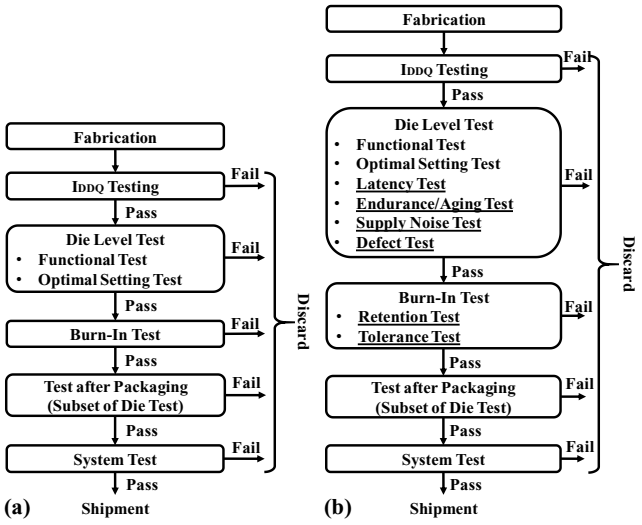

Fig. 1 (a) Conventional memory (such as SRAM, DRAM) test flow; (b) repurposed test flow for emerging NVMs.

May 22, 2019. This work is supported by Semiconductor Research Corporation (#2727.001), National Science Foundation (#CNS-1441757, #CNS-1556301 and #CCF-1718474) and DARPA Young Faculty Award (#D15AP00089).

The authors are with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA 16802, USA (e-mail: muk392@ psu.edu; szg212@psu.edu).

hand, NVM write/read latencies are high (Fig. 2(a)-(b), ~1-3ns) [15]. Furthermore, these latencies also vary significantly due to stochasticity and PVT resulting in a long tail. This can lead to write and read failure [15] if latencies are not characterized well. Therefore, the latencies should be tested under PV within the operating temperature. Hardware features can be added to allocate a few extra clock cycles to read and write operation if needed instead of slowing down the clock.

Supply Noise: NVM write and read current are significantly higher than conventional memories (Fig. 2(c)-(d)). The high current leads to high supply noise such as supply voltage droop and ground bounce. The read/write operation can fail especially when the farthest memory bank is accessed. Therefore, NVM should be tested for supply noise-induced read/write failure.

Endurance and reliability: The resistance of low and high state of NVMs (therefore sense margin) vary with PVT from bit-to-bit. Moreover, the oxide layer in the bitcell for STTRAM, MRAM, and RRAM, and phase change material in PCM can also suffer from a physical breakdown. These issues require special endurance and reliability testing compared to conventional memory technology.

Sensitivities/Tolerances: Spintronic memories store data in terms of the magnetic orientation of a ferromagnetic layer. Therefore, an external magnetic field can corrupt the data [16]. Therefore, spintronic memories should be tested for magnetic tolerance in all operating modes. Note that all NVMs are susceptible to temperature [2], [16-19]. Temperature variation can cause read/write/retention failures. Therefore, all NVMs should be tested for thermal tolerance. RRAM is sensitive to N_2/O_2 gas which requires similar attention [20].

We make the following additional contributions over [21] in this work. We, (i) review existing NVM fault models and test methodologies; (ii) repurpose latency tests; (iii) analyze the required time for different tests, (iv) propose test time reduction technique, (v) propose new procedure to test endurance and reliability test along with the required Design-for-Test (DFT) circuits, (vi) summarize test time of existing test methodologies proposed so far in literature for emerging NVMs. Fig. 1(b) presents a repurposed test flow for emerging NVMs and the proposed new tests are underlined. We restrict our discussion to

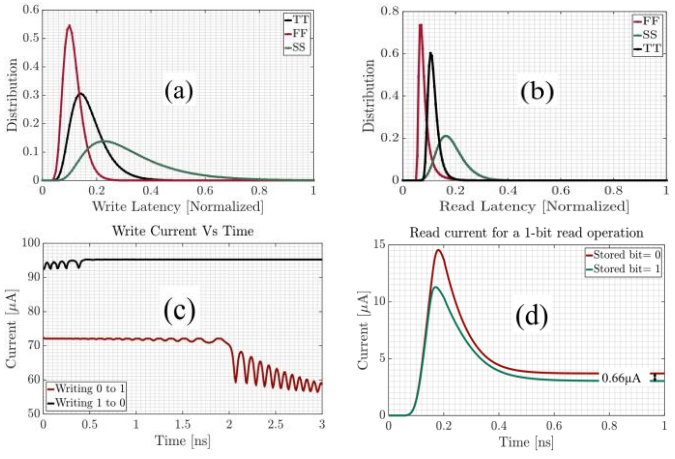

Fig. 2 STTRAM (a) write and (b) read latency variation; (c) assymetric and high write current; (d) assymetric read current. Note that other emerging NVMs also show the similar characteristics.

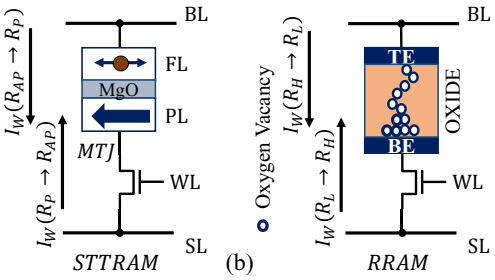

Fig. 3 (a) STTRAM bitcell; and (b) RRAM bitcell.

two flavors of NVMs namely, STTRAM and RRAM for the sake of brevity. We also introduce other NVMs as necessary.

Organization of the rest of the paper is as follows: NVM basics are discussed in Section II; tests for read latency/write latency, supply noise, magnetic/thermal tolerance, retention time and endurance are described in Section III-VII; summary of existing NVM test technique and other fault issues are explained in Section VIII; Conclusions are drawn in Section IX.

II. Emerging NVM Basics

A. *STTRAM*

STTRAM bitcell (Fig. 3(a)) contains one NMOS as an access transistor and one MTJ as a storage element. MTJ contains two ferromagnetic layers known as Free Layer (FL) and Pinned Layer (PL). The equivalent resistance of the MTJ is low/high if FL and PL magnetic orientations are parallel (P)/ anti-parallel (AP) to each other. During write operation, FL magnetic orientation can be toggled from P (data '0') to AP state (data '1') (or vice versa) using current induced Spin-Transfer Torque by passing the write current (> critical current) from Sourceline (SL) to Bitline (BL) (or vice versa). Fig. 2(c) and 2(d) shows STTRAM write/read current with respect to time.

B. *RRAM*

The storage element in RRAM is mainly an oxide material between two electrodes namely, Top Electrode (TE) and Bottom Electrode (BE) (Fig. 3(b)). Filament between the electrodes can be formed or raptured based on the direction and magnitude of the electric field through it. If a filament is formed between the two electrodes, the resistance of the cell is low (Low Resistance State, LRS) and that can be considered as data '0'. However, if the filament is raptured, the resistance of the cell is high (High Resistance State, HRS) and that can be considered as data '1'. Fig. 4(a) and 4(b) shows the RRAM write and read current.

III. Read Latency & Write Latency Test

A. *Test for read/write latency*

In conventional memories, March tests are run at different

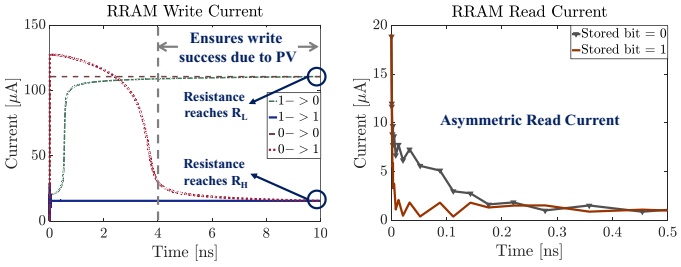

Fig. 4 RRAM (a) write current; and (b) read current.

clock frequencies to search the highest frequency at which the memory passes. However, NVM read/write latencies are strong functions of PVT (shows wide distribution) and thereby, such method will increase the test time significantly. Therefore, we propose a unique latency test for emerging NVMs which extends the number of cycles allocated for read and write operation if failure is observed. This improves the test time and can also be used to mitigate read/write failures. Fig. 5(a) shows the proposed DFT circuit. $Write_{Enable}$ signal can be extended to increase its pulse-width by OR'ing it with phase extension signals (1C4H/2P, 2C4H/4P, 3C4H/6P, and 4C4H/8P). Fig. 5(b) shows the timing waveform of different input phase-extension signals of the DFT circuit shown in Fig. 5(a).

We assume that the memory writes/reads in x/y cycles with a system frequency of fGHz and the maximum targeted cycles for write and read are ($x+x_n$) and ($y+y_n$). Note that x_n and y_n are considered to provide a guard band to the write and read latency respectively. Fig. 6 shows the guard band of x_n cycle that can be added to the write latency (for illustrative purposes). The steps of latency tests are:

Step-1. Write '0's with x clock cycles and read with ($y+y_n$) clock cycles to each address and repeat for entire memory i.e., {w0⇧r0⇩}. Note that more cycles are used to ensure successful read operation and capture only write failures.

- Write fails (even for one address): Increase clock cycle for write operation by 1 (i.e. write cycles = $x + x_0$, where $x_0 = 1$) and repeat Step 1. In case of write failures, keep increasing the clock cycles (i.e. increase x_0 by 1 each time) for write and repeating Step 1 till the maximum targeted cycles for write operation (let's say, $x+x_n$). If write failure is observed with even ($x + x_n$) cycles, discard the chip or slow down the system clock.

- Write succeeds: Minimum write cycle for writing '0' is ($x+x_0$).

Step-2. Read with ($y + y_n$ -1) cycles to verify read with lower cycle.

- Read succeeds: Repeat Step-2 with decreasing number of cycles till the read cycle reaches y.

- Read fails: Minimum read cycle for reading '0' is ($y+y_0$).

Step-3. Write '1's with ($x+x_0$) clock cycles and read with ($y+y_n$) to each address and repeat for entire memory i.e., {w1⇧r1⇩}. Note that more cycles are used to ensure successful read operation and capture only write failures.

- Write fails (even for one address): Repeat Step-3, with ($x+x_1$) cycles (where, $x_1 = x_0 + 1$) and keep increasing if write failure is observed till the maximum targeted cycle for write operation ($x+x_n$). If write failure is observed with even ($x+x_n$) cycles, discard the chip or slow down the clock.

- Write succeeds: Minimum write cycle is ($x+x_1$).

Step-4. Read with ($y+ y_n$ -1) cycles to verify read with a lower cycle.

- Read succeeds: Repeat Step 2 with a decreasing number of cycles till the read cycle reaches y.

- Read fails: Minimum read cycle is ($y+y_1$).

Step-5. Certify the chip with write latency of ($x+x_1$) cycles and read latency of max($y+y_0$, $y+y_1$) cycles.

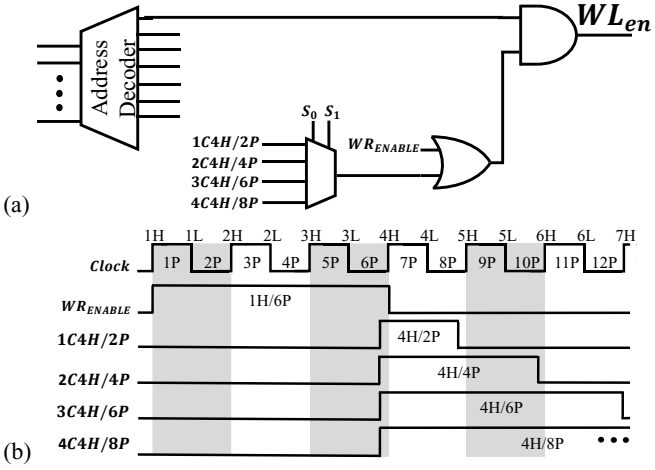

Fig. 5 (a) DFT circuit (proposed in [21]) for latency test; (b) input signals for the DFT circuit.

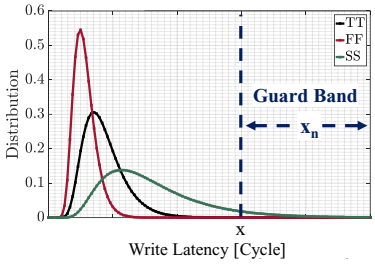

Fig 6: Guard band to write latency (for illustrative purposes only).

Step-6. The test can be repeated for N times (N could be 10-100) to capture the stochastic accuracy. Passing all previous steps with acceptable stochastic accuracy means that the read/write latency for the chip is identified. Next, the following steps are performed with these latencies to detect various faults:

Step-7a. Flush all addresses with '0', write '0' again followed by two reads i.e. {w0⇧w0⇩r0⇧r0⇩}.

- Identify the chips that incur any read/write failure.

Step-7b. Write all addresses with '1', read, write '1' again followed by two reads i.e. {w1⇧r1⇩w1⇩r1⇧r1⇩}.

- Identify the chips that incur any read/write failure.

B. Test Time Compression Techniques

i) Test Worst-Case: Only the worst-case write latency could be checked for test time compaction. For example, writing 0→1 requires more time than 1→0 for both STTRAM (Fig. 2(c)) [22] and RRAM (Fig. 4(a)).

ii) Wordline Overdrive (WL OV): This technique is implemented by increasing the wordline voltage during write operation in test mode. WL OV reduces the access transistor resistance. Therefore, the bits draw more current with the same BL and SL voltage. Fig. 8(a) shows RRAM write current with respect to time for 500mV of WL OV for both 0→1 and 1→0 write operation compared to the case of Fig. 4(a). Fig. 8(a) shows that write current increases for all four writes but latency decreases. Fig. 8(b) shows that all the bits get written within 4ns.

Therefore, test time can be decreased by testing the worst-case write latency and implementing the WL OV technique and ending write operation early during the test mode. However, note that modeling of write latency reduction with respect to WL OV with 100% write success (like Fig. 8(b)) should be done

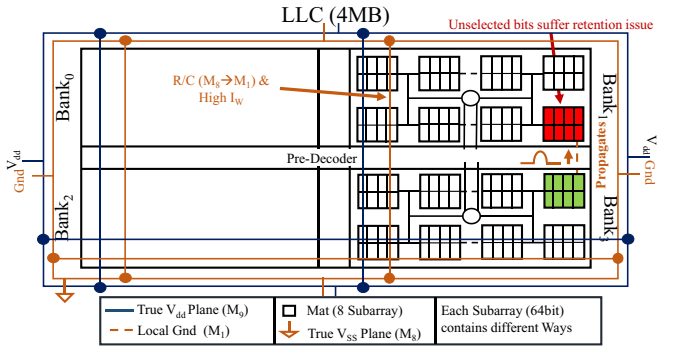

Fig. 7 1T1R-based 4MB LLC containing 4 banks showing worst-case voltage droop and ground bounce. Each bank contains 8 Mats and each Mat contains 8 subarrays each producing 64bits. Each subarray has 8 Ways. Ground bounce generated by write operation propagates to near-by word-line/source-line/bit-line drivers.

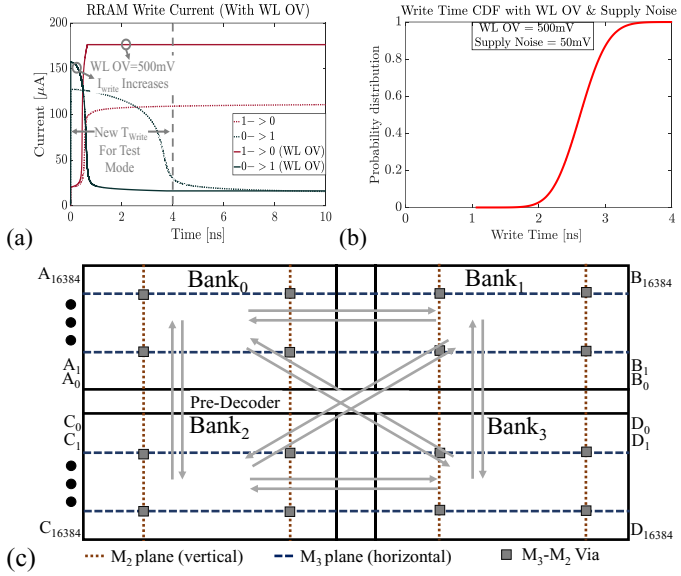

Fig. 8 (a) RRAM write current with 500mV of WL OV for both 0→1 and 1→0 write. Note that the latency reduces as write current increases; (b) write time CDF with 500mV of WL OV and 50mV of additional supply noise; and, (c) diagram showing addresses of $Bank_0$/$Bank_1$ /$Bank_2$/$Bank_3$ as A/B/C/D from 0 to 16K. The grey arrows show different cases for testing the impact of supply noise.

extensively prior to testing a chip.

C. Test Time Analysis

We analyze the latency test time for a 4MB RRAM NVM LLC (Fig. 7) having 4 banks (N_{Banks} = 4) each having 16384 addresses (N_{Add} = 16384) (Fig. 8(c)). We assume that RRAM target base-write-latency is 10ns and latency with 500mV of WL OV is 4ns. RRAM target read latency is 0.5ns. The system clock frequency is 2GHz (f_{clock}) which means that the value of x and y in the test procedure described in Section III.A is 20 and 1, respectively. We consider x_n and y_n to be 5 cycles (assuming 25% of normal write latency) and 2 cycles (assuming 200% of normal read latency), respectively. We also assume that the chip got certified with the highest write/read latency of 25/3cycles to calculate the worst-case test time.

We consider that the chip is certified with N = 100 for stochastic accuracy and considered both polarities (N_p = 2) write/read latency test (0→1 and 1→0).

The latency test time can be approximated by the following equation for the worst-case:

$$T_{Test} = Np * N * (T_{Write} + T_{Read}) + T_{Fault}$$

$$T_{Write} = N_{Banks} * N_{Add} * ((x+5)) * \frac{1}{f_{clock}}$$

$$T_{Read} = N_{Banks} * N_{Add} * ((y+2)) * \frac{1}{f_{clock}}$$

$$T_{Fault} = N_{Banks} * N_{Add} * (4 * (x+5) + 5 * (y+2)) * \frac{1}{f_{clock}}$$

The required test time for the 4MB RRAM considered in this work is 187.3ms. The test time reduces to 87ms which means that a test time reduction of 2.15X can be achieved.

D. Test Energy Analysis

In the proposed test compression method, we incur a higher write current. I_{write} for 1→0, 0→0 and 0→1 increase to 118.16μA, 126.72μA and 90.57μA from 80.08μA, 110.01μA and 84.34μA respectively. Here, the gate current increment with WL OV is ignored since it is insignificant. We considered all bits flipped 0→0 and 0→1 in Steps-1 and Step-3 respectively to calculate the worst-case test energy consumption.

The proposed WL OV test compression technique incurs 16.71μA and 6.23μA of extra write current for writing 0→0 and 0→1 respectively. However, it incurs lower test time. Therefore, the proposed technique saves a total energy of 4.97μJ.

E. DFT Circuit Area and Power Analysis

We have implemented the DFT circuit proposed for the latency test using 22nm PTM technology in HSPICE. The circuit consumes 0.0062μm^2 of area and 1.14nW and 6.82μW of static and dynamic power, respectively. Note that the dynamic power is only consumed during the testing phase and therefore, it can be ignored when the chip in deployed in the market. However, the DFT circuit will consume static power whenever the chip is powered ON even after deployment. Therefore, we can power-gate the DFT circuit with an NMOS switch to further reduce the static power.

F. Considerations for Other NVMs

Read/write latencies of all NVMs show the wide distribution due to PVT. Therefore, latency test should be done for all NVMs. Note that the proposed test method is not memory specific and thereby, can be extended to any NVMs.

IV. Supply Noise Test

A. Supply Noise in NVMs and Write Failure

NVMs incur high write current. Therefore, the write current for a full cache line write operation can be extremely high. For example, total write current for 512bits is 51.2mA considering 100μA/bit (typically STTRAM write current per bit is 512μA/bit). Note that the true V_{dd} and ground of a chip is implemented with upper metal layer (e.g. M_8) and the local V_{dd} and ground is implemented at lower metal layer (e.g. M_1) as shown in Fig. 7. Therefore, there exists a significant parasitic resistance and capacitance between the true V_{dd}/ground and local V_{dd}/ground (Fig. 9) and the bitcells incur a supply voltage droop and ground bounce due to high current during write operation [24]. This means that the bitcell will incur a lower voltage

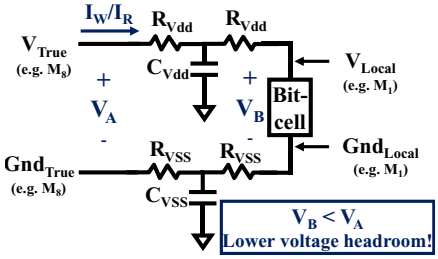

Fig. 9 Cartoon showing parasitic resistance between upper metal layer and lower metal layer.

headroom (Fig. 9) which can lead to increase of write latency and possibly write failure.

B. Parallel Accesses in NVM and Read Failure

Multiple cycles are needed to write NVM since the typical write latency is a few nanoseconds. For example, the required cycles for read and write operations are 2 and 5 respectively considering a system clock of 2GHz and read and write latency of 2.5ns and 1ns. However, the system throughput degrades if only one memory access is allowed. Therefore, parallel read and write operations are performed in independent memory banks in successive cycles. However, the supply noise generated by one operation can propagate and affect another parallel operation. For example, when $Bank_4$ of Fig. 7 is being written:

i) Four parallel read operations can be performed in the next four cycles in other independent banks. This read/write scheme is named 1X write mode. These data are processed in pipeline to maintain high throughput. Note that the parallel read operations might experience read failure if supply noise from the write operation propagates to it. This is true since (a) sense margin reduces at lower read voltage; (b) Tunnel Magneto-Resistance (TMR) ($TMR = \frac{R_{AP}-R_P}{R_P}$) reduces due to higher access transistor resistance at lower word-line voltage.

ii) Multiple writes (n) can be performed with read operations between two writes. This read/write scheme is named nX write mode. Note that multiple parallel writes can worsen the supply noise. Therefore, it can lead to read/write failure.

C. Test for Supply Noise [21][24]

Supply noise test needs to check two issues: i) write failure due to lower voltage headroom, (ii) read failure due to supply noise propagation from a parallel write operation. Note that write operation incurs maximum current when stored data and new write data are both '0'. This, in turn, leads to worst-case supply noise. However, if the stored data and the new write data both are same, write failure cannot be detected. Therefore, we propose a test with a data pattern that causes maximum noise using N_{col}-1 bits of (N_{col} = number of bits in a row) 0→0 writes and one bit of 0→1 write (the remaining bit). Furthermore, read operation needs to be performed in parallel to write operation to test for supply noise-induced read failure. Therefore, the work [21] proposes the following supply noise test (examples of test data pattern are given by considering the write data is 8bit):

Step-1. Memory banks are identified that incur highest supply noise (farthest from the supply regulator). Memory layout can be used to identify these banks.

Step-2. Write 0x00 to all addresses followed by read operation to verify write success. i.e. {w0⇧r0⇩}. Discard the chip if the write operation fails.

Step-3. Write 0x01 (maximum noise with only one 0→1 writing) to all addresses of the bank (identified in Step-1) followed by write success verification i.e. {wN⇧rN⇩}, where N=0x01. After initiating each write in the bank, initiate parallel reads in other independent banks. If only write operation fails decrease the clock frequency or increase V_{dd} by small steps and repeat until 100% write success is achieved. If read fails along with/without write failure, increasing V_{dd} is the only solution. Certify the chip with the maximum clock frequency and the minimum V_{dd}. Discard the chip if target specification is not met.

Step-4. The above-mentioned Step-3 testing only the least significant bit. Repeat Step-2 and Step-3 with other data patterns that have one data '1' (i.e. 0x02, 0x04, 0x08, 0x10, 0x20, 0x30 and 0x40) to test the remaining bits.

Step-5. Step-1-4 can be repeated with modification of Step-3 (performing nX writes) while writing to the farthest bank to validate nX write mode.

In [**24**], a detailed analysis and modeling of supply noise is done for a 4MB RRAM using 65nm technology. The work summarizes the data pattern and cases which should be tested after manufacturing. The grey arrows shown in Fig. 8(c) indicates different cases for testing the impact of supply noise. Furthermore, the work proposes WL OV and ending write operation early to reduce the test time significantly.

In [**21**], each address needs to be tested N times if the address is N-bit long (testing each bit at a time without changing the supply noise profile). Let's call this number N_{repeat}. However, [24] shows that WL OV can help to reduce N_{repeat}. The work summarizes that the average write current for 1→0, 0→0 and 0→1 increases to 118.16μA, 126.72μA and 90.57μA from 80.08μA, 110.01μA and 84.34μA respectively for 500mV of WL OV for the RRAM employed in that work. Therefore, test pattern with combination of the 0→0, 0→1 and 1→0 write patterns can be selected to maintain the same worst-case write current as 512-writes of 0→0 as normal mode (without WL OV). This approach reduces N_{repeat}. A rough estimation shows that each 0→1 reduces total I_{write} (from worst-case) by 26μA whereas one 0→0 and one 0→1 increases total I_{write} by (8+16.7)μA=24.7μA. This means that 170 writes of 0→1, 172 writes of 0→0 and 170 writes of 1→0 (in a 512bit data) yield almost the same total I_{write} as 512-writes of 0→0. Therefore, one address can be tested only 3 times (N_{repeat} = 3) where 170, 171 and 171 bits can be tested each time.

The work in [24] also proposes to end the write operation early since WL OV reduces the write latency. The test technique like [21] can incur a test time of 646.23s whereas [24] incurs only 1.57s. Therefore, [24] provides 410.82X test time compression and saves 79.88J of test energy.

D. Considerations for Other NVMs

All NVMs possess high write/read current and are designed for high density leading to more parasitic R and C. Therefore, other NVMs will also suffer from read/write failure due to voltage droop/ground bounce. Therefore, the proposed test can be extended all NVMs.

V. Magnetic and Thermal Tolerance Test

A. Magnetic Tolerance Test [25]

Spintronic memory such as, STTRAM stores information as the magnetic orientation of the FL of MTJ. An external magnetic field can flip the magnetic orientation and corrupt the information. Therefore, magnetic tolerance of spintronic memory during write, read and retention mode should be tested.

Write Tolerance: Spintronic memories can be certified with write tolerance as the maximum magnetic field under which it can be written successfully at a specified write current with a specified write latency.

Read Tolerance: Spintronic memories can be certified with read tolerance as the maximum magnetic field under which it can be read successfully without causing any disturbance to the bits at a specific read current with a specified read latency. For the STTRAM employed in this work, read tolerance is found to be 370Oe for a read current = 20μA and read latency = 1ns.

Retention Tolerance: Spintronic memories can be certified with retention tolerance as the maximum external magnetic field under which it does not incur any data-corruption for a specified time period during retention mode. For the STTRAM employed in this work, retention tolerance is found to be 319Oe for a 3ns time period during retention mode.

The algorithms to find the write tolerance, read tolerance and retention tolerance are discussed in [25].

Test Time Analysis: Let's analyze the different tolerance test times for a 4MB STTRAM NVM LLC having 4 banks (N_{Banks}=4) each having 16384 addresses (N_{Add}=16384). We assume that STTRAM read/write latency is 1ns/3ns (t_{write}/t_{read}) and retention time is t_{ret}= 3ns and we test both data polarity (N_p = 2) tolerance for write, read and retention tolerance.

The worst-case tolerance test time equations are:

$$T_{write\_tol} = N * N_t * N_{Banks} * N_{Add} * N_p * (t_{write} + t_{read})$$

$$T_{read\_tol} = N * N_t * N_{Banks} * N_{Add} * N_p * (t_{write} + 2 * t_{read})$$

$$T_{ret\_tol} = N * N_t * N_{Banks} * N_{Add} * N_p * (t_{write} + t_{ret} + t_{read})$$

Here, N_t is the number of times the external magnetic field is increased to till the tolerance limit is found, N is the number of times the test is repeated for stochastic accuracy.

Note that we can approximate the tolerance range from design parameters and simulation to lower the value of N_t. For example, our simulation shows that the STTRAM employed in this work has a write tolerance of 86Oe. Therefore, we can start our testing with 80Oe and test till 86Oe. In this case, N_t = 4 considering the test step size for magnetic field strength increment is t = 2Oe. We calculate the test time for write/read/ retention tolerance as 262.14ms/327.68ms/458.72ms respectively, considering N = 100 and N_t = 5.

B. Temperature Tolerance Test

Memory chips should be certified to operate successfully within a target temperature range (typical range: -10°C to 90°C):

(a) At high temperature, energy barrier between two states of the memory reduces. Therefore, the data retention time reduces and can lead to retention failure. Furthermore, read failure can occur since the reduction of resistance difference between two states leads to reduction of sense margin and read disturb can occur since slight disturbance can flip the data at lower energy barrier. Therefore, the manufacturer needs to test retention and read failure/disturb at high temperature.

(b) At low temperature, the energy barrier between two states increases. Therefore, the read/write latencies increase and can lead to read/write failures. Therefore, the manufacturer needs to test read/write failure at low temperature.

Test methodology: Thermal tolerance test can be done using the algorithms proposed for magnetic tolerance test (Section V.A) by applying external thermal field instead of magnetic field. This test can be combined with standard hot-cold test. The highest temperature at which the memory read failure/disturb does not occur and retention time meets the minimum target specification, is the upper limit of thermal tolerance. The lowest temperature at which read/write operation does not fail, is the lower limit of thermal tolerance.

Test Time Analysis: Temperature tolerance test time can be approximated using the equation described for magnetic tolerance test. However, N_t can be large for temperature tolerance test if the entire range is tested with equal step size. For example, temperature tolerance test time could be 10X larger compared to magnetic tolerance test if step size, t = 2°C and the entire target temperature range is considered (-10°C to 90°C). However, we propose to focus the test on the lower range and higher range only. In that case, the temperature tolerance test time could be 2X compared to magnetic tolerance test.

C. Considerations for Other NVMs

The tolerance tests described in this section can be extended to all emerging NVMs. However, note that magnetic tolerance test is needed for spintronic memories only whereas thermal tolerance test is needed for all emerging NMVs.

VI. Retention Time Test

A. Retention Test Challenges

Emerging NVM will incur significant test time if conventional retention tests are used. A rough estimation indicates that STTRAM (with 1yr base retention) retention test time can be ~1.15yr with the eDRAM retention test method since the highest retention time to total test time ratio is 86.5% [12]. However, the retention test time can be improved by ~$1.3x10^8$X if the retention time for each bit is compressed to ~3μs. Note that retention for spintronic memory is stochastic in nature. This means that the same bit shows different retentions when measured multiple times [13]. Therefore, retention test for spintronic memory becomes even more challenging.

B. Temperature-based Compression (For All NVMs)

STTRAM bitcell with base retention of 1yrs at 25°C shows a compressed retention of 24.8s at elevated temperature = 200°C [26]. Therefore, the bitcell should have retention time = 1yrs at 25°C if the cell passes retention test for the rated retention time = 24.8s at 200°C (~$1.3x10^6$X test time reduction) [26].

C. Weak-Write-based Compression (For All NVMs)

NVM retention time reduces if the bitcell incurs a disturb

TABLE II COMPARISON BETWEEN MBI, MBI+BI AND WEAK- WRITE BASED COMPRESSION METHODS

	MBI (220Oe @ 25ºC) [29]	MBI+BI (220Oe @ 125ºC) [29]	Weak-Write (w/ 184µA) [26]	Weak-Write (w/ 184µA) [27]	Weak-Write (w/184µA) [28]
Reduced Retention (µs /bit)	~41.7	~3.03	~162.7	~162.7	~162.7
Test time with LS (s)	~3.42	~0.25	~13.3	-	-
Test time with *EMACS (s)*	-	-	-	~15.6	-
Test time (s)	-	-	-	-	~960
DFT overhead	Nominal		Yes	Yes	Yes
Extra write power	No	No	Yes	Yes	Yes

current through a it. This is especially true for spintronic memories. Therefore, weak-write-based test is proposed for STTRAM retention test time reduction [26-28]. Elevated temperature [26] and efficient DFT [26-27] circuit can be incorporated with disturb current for further reduction.

D. Magnetic Field-based Compression (For Spintronics)

An unique Magnetic Burn-in (MBI) and MBI with elevated temperature (MBI+BI) are proposed in [29] to compress STTRAM retention test time. In this method, an external magnetic field is applied whose direction is antiparallel to the magnetic orientation of MTJ FL. This reduces bitcell retention time and the memory can be certified at lower test time [29].

E. Comparison of Compression Methods

MBI (220Oe @ 25ºC) [29] and MBI+BI (220Oe @ 125ºC) [29] are compared with weak-write-based (with 184µA) compression methods proposed in [26], [27] and [29]. We considered a 64MB STTRAM memory with base retention time = 10yrs (corresponding thermal stability, $\Delta = 60K_bT$) and word size of 512bits. Test time is calculated for MBI [29], MBI+BI [29] and [26] using Linear Search (LS) (proposed in [26]) with resolution = 10ns, read latency = 1ns, write latency = 1ns, rows written in parallel = 128 and number of iterations = 10^3 (for stochastic accuracy). Test time is calculated in [28], by considering 500 rows written in parallel. Test time is calculated in [27] with EMACS, by considering 64 rows are activated for error detection. Table II summarizes the result. Note that MBI+BI achieves the lowest retention test time.

MBI/MBI+BI requires a magnetic chamber where read/write operations can be run on memory. However, these techniques incur minimal DFT changes. The weak-write-based techniques incur significant test power since constant disturb current is passed through all the bitcells and requires major DFT modifications (e.g. modified write driver [26]).

F. Considerations for Other NVMs

All NVM's retention time is a function of temperature [2], [16-19]. Furthermore, retention time of any NVM can be compressed by passing a disturb current through the bitcell with a direction that flips the stored data. Therefore, we conclude that both temperature-based [26] and weak-write-based [26-28] methods can be extended to other NVMs. However, magnetic field-based [29] methods can be extended to spintronic memories only.

VII. ENDURANCE AND RELIABILITY TEST

A. Reliability Degradation Mechanism

NVM performance can degrade over time due to physical breakdown (STTRAM/RRAM/PCM) or resistance drift (RRAM/PCM). STTRAM/MRAM have an oxide layer in their storage element, MTJ, and RRAM has oxide layer between two electrodes in its bitcell. Oxide might breakdown due to high I_{write} leading to function failure. It has also been reported [3] that LRS changes 2X-10X and HRS changes 5X-100X in TaO_2 based RRAM due to variation. In PCM, time-dependent resistance drift in amorphous chalcogenide material is one of the major reliability concerns.

B. Test for Endurance

Endurance test is tricky as we don't want to kill the bit and at the same time, we need to trigger the conditions to cause reliability issue. For example, a memory with 10^6 of endurance means that the memory bits will function correctly for 10^6cycle of operation. Now, if the bits are tested for 10^6cycles, they will no longer be reliable even if they pass the endurance test. This problem can be solved by having some sacrificial bits in the memory. It can be considered that if the sacrificial bits pass the endurance test, the other bits will also pass. However, this naïve approach may incur high test time. For example, recently RRAM has been proposed with an endurance of ~10^{12} [30] which incurs test time of 10^4s with T_{write}=10ns. Therefore, total test time would be too high.

DFT Circuit for Endurance Test: We propose a DFT circuit (Fig. 11(a)) for the endurance test. An n-bit counter generates 2^n cycle pulse (128C1H/256P) (Fig. 11(b)) to stress the bitcell for

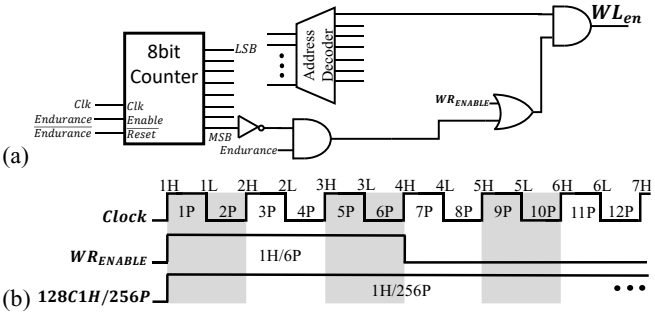

Fig. 11 (a) Proposed DFT circuit for endurance test; (b) inputs waveforms of the proposed DFT circuit (Fig. 11(a)).

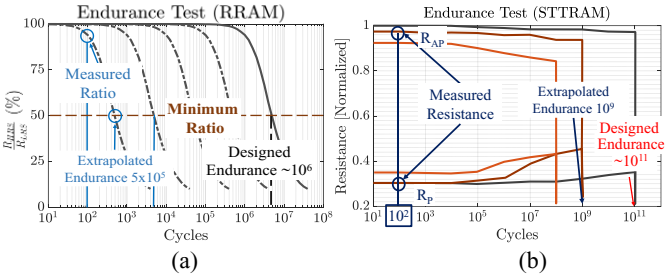

Fig. 12 Proposed endurance test for (a) STTRAM and, (b) RRAM using extrapolation for shorter test time without affecting the reliavility of the bits. The relation is obtained from [31] and [32] respectively.

endurance test. We propose following technique to test RRAM endurance using the DFT circuit in Fig. 11(a).

1. The ratio of R_{HRS}/R_{LRS} decreases as the number of

operation cycle increases [31]. Therefore, a model like Fig. 12(a) can be developed by testing a few chips extensively. Identification of R_{HRS}/R_{LRS} ratio may require a separate test process since direct measurement of the resistance is not practical. One possible approach is to back-calculate the ratio by characterizing the sense margin. The sense margin could be modulated by tightening the read latency (using a DFT or faster clock) or by writing the bit partially. The detailed discussion of R_{HRS}/R_{LRS} characterization is a subject of our future research and beyond the scope of this paper.

2. Keep some sacrificial bitcells in the chip. Let's say that N addresses contain the sacrificial bits. If the cells are tested until the designed endurance, test time would be significantly high. Therefore, stress them for 10^2 cycles and measure R_{HRS} and R_{LRS} and their ratio. The endurance can be extrapolated from this ratio considering that a model like Fig. 12(a) is obtained statistically. A counter-based DFT circuit can be used to generate a 128-cycle write enable signal (Fig 11(a)). This signal can be used to stress the bits for continued 128-cycles.

C. Test Time Analysis

Conventional endurance test for NVMs which requires sacrificial bits in the memory, incurs high test time. For example, let's consider that there are 100 (N) addresses kept for the endurance test, designed endurance is En_{cycle} (=10^6 in our case) and the system clock frequency is 2GHz (T_{clock} = 0.5ns). Read (T_{read})/write (T_{write}) time for them are 10ns/0.5ns for RRAM. We also consider that bank-level parallelism (simultaneously 2-reads or 2-writes can be performed) is implemented. Note that bank-level parallelism reduces the test time. However, the sacrificial bits should be divided into different banks so that parallel operations can be performed. We calculate RRAM test time using the following equation:

$$T_{test} = T_{write} * En_{cycle} * \left(\frac{N}{2}\right) + T_{read} * \left(\frac{N}{2}\right)$$

$$T_{test\_conv} = 10 \times 10^{-9} * 10^6 * \left(\frac{100}{2}\right) + 0.5 \times 10^{-9} * \left(\frac{100}{2}\right)$$

$$T_{test\_conv} = 0.5sec$$

We have ignored the first clock cycle (=0.5ns) after which the second write starts and both writes run in parallel. The above test time is for a single RRAM chip which is too high. However, the proposed test technique for RRAM requires testing each address for 10^2 cycles (En_{cycle} = 10^2) which incurs 50.02µs. Therefore, the proposed test technique achieves 9.99×10^3X test time compression for RRAM. For the proposed endurance test, we effectively reduce the test time while other parameters like operating voltage/current are same. Therefore, the proposed method effectively reduces the test energy consumption by 9.99×10^3X for RRAM.

D. DFT Circuit Area and Power Analysis

We have implemented the DFT circuit proposed for the endurance test using 22nm PTM technology in HSPICE. The circuit consumes 0.019µm^2 of area and 2.18nW and 19.9µW of static and dynamic power, respectively. The DFT circuit can be gated with an NMOS switch to further reduce the static power.

E. Considerations for Other NVMs

The proposed RRAM endurance can be modified for STTRAM and MRAM as follows:

1. R_{AP} (R_P) decreases (increases) as the number of bipolar stress cycle increases [32]. Therefore, by testing a few chips extensively, a model like Fig. 12(b) can be developed.

2. Keep some sacrificial bitcells in the chip. If these cells are tested till the designed endurance, test time would be significantly high. Therefore, stress them for 10^2 cycles and measure R_{AP} and R_P. A counter-based DFT circuit can be used to generate a 128-cycle write enable signal (Fig 11(a)).

This test can be extended to PCM based on its physics.

VIII. Discussion

Environment Sensitivity: Resistance (of LRS and HRS) and the forming voltage of RRAM at which the filament is formed depend on N_2/O_2 gas and, air [20] exposure. Since a chip is usually hermetically sealed, RRAM may not be exposed to them. However, O_2 can be generated from SiO_2 breakdown inside the chip and RRAM bitcell can get exposed to it. Therefore, O_2/N_2 chambers are needed to test such sensitivities.

Testing failures due to sneak-path current: A sneak path testing method is proposed in [33]. The work also proposes an unique DFT circuit which controls the number of sneak paths that are concurrently enabled. This can be further leveraged to detect different faults (such as stuck-at faults and coupling faults) by comparing sneak-path to ideal current.

Fault models captured by the proposed tests: NMVs are susceptible to different faults [33-35]. The latency test proposed in this work can capture the following faults through Step-7:

Stuck-at faults/Transition faults: Latency test checks each address for 1→0 and 0→1 transitions. Therefore, if a cell is stuck-at '0'/'1' or fails to transit 1→0/0→1, it will be captured.

Write destructive faults: Latency test checks each address for 0→0 and 1→1 transitions. Therefore, if any cell flips even though we are writing the same polarity as the stored polarity, it will be captured.

Read destructive faults/Deceptive read destructive faults: Latency test reads each addresses two consecutive times for both data '0' and '1'. Therefore, if both read operation returns same incorrect value, it indicates a read destructive fault. However, if the first read operation is correct followed by an incorrect read, it indicates a deceptive read destructive fault.

Stuck open faults: If the sense amp is designed with a latch after it and an address is stuck open, read operation of that address will return the previous read data [35]. In that case, this can be modeled as stuck at fault [35] and it will be detected by the proposed latency test.

However, the tests proposed in this work cannot detect static coupling faults, transition coupling faults and incorrect read coupling faults.

Summary: Table III summarizes all the test methods proposed for NVMs so far and notes down the required test time and test energy compression. The total test time found for a 4MB emerging memory is around 5.05s considering all the tests with available test time. Note that all tests are not required for each memory type. We approximate that on an average 3 to 4 secs will be required to test each chip. Note that typically test time for each chip is considered as 2-3sec [36]. Therefore, proposed

TABLE III SUMMARY OF DIFFERENT NVM TESTS

Test Name	Application	Proposed Methods	Test Time (Compression)	Test Energy Compression
Retention Test	All NVMs	Weak write-based [26]	13.3s (N/A)	-
		EMACS [27]	15.6s (N/A)	-
		External Temperature-based [21]	-	-
		MBI [29]	3.42s (1.68×10^6X)	-
		MBI+BI [29]	0.25s (4.12×10^7X)	
Latency Test	All NVMs	WL OV + Early Write (this work)	87ms (2.15X)	4.97μJ
Supply Noise Test	All NVMs	[21]	646.23s (N/A)	
		WL OV + Early Write [23]	1.57s (410.82X)	79.88J
Magnetic Tolerance Test	Spintronic Memories	[25]	262.14ms (write) (N/A)	-
			327.68ms (read) (N/A)	
			458.72ms (retention) (N/A)	
Thermal Tolerance Test	All NVMs	This work	524.28ms (write) (5X)	-
			655.36ms (read) (5X)	
			917.44ms (retention) (5X)	
Endurance Test	All NVMs	This work	50.02μs (9.99×10^3X)	9.99×10^3X
Sneak Path Test	All NVMs	[33]	-	-
Sensitivity to Environment	RRAM	-	-	-
Coupling Test	All NVMs	[34]	-	-
Stack at Fault	All NVMs	[34]	-	-

test methods are effective in keeping the test time close to the conventional test time target.

IX. CONCLUSIONS

In this work, we summarized emerging NVMs test challenges and proposed methods to address new emerging NVM-specific faults. Furthermore, we proposed unique methods and the required DFT circuits to facilitate the tests. We analyzed the time required for different tests and summarized the compression achieved by existing techniques.

REFERENCES

[1] 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)," IEEE/ACM International Symposium on Low Power Electronics and Design, Fukuoka, 2011, pp. 121-126.

[2] D. C. Worledge, G. Hu, P. L. Trouilloud, D. W. Abraham, S. Brown, M. C. Gaidis, J. Nowak, E. J. O'Sullivan, R. P. Robertazzi, J. Z. Sun and W. J. Gallagher, "Switching distributions and write reliability of perpendicular spin torque MRAM," 2010 International Electron Devices Meeting, San Francisco, CA, 2010, pp. 12.5.1-12.5.4.

[3] Y. Wu, S. Yu, X. Guan and H. -. P. Wong, "Recent progress of resistive switching random access memory (RRAM)," *2012 IEEE Silicon Nanoelectronics Workshop (SNW)*, Honolulu, HI, 2012, pp. 1-4.

[4] A. Pirovano, A. L. Lacaita, F. Pellizzer, S. A. Kostylev, A. Benvenuti and R. Bez, "Low-field amorphous state resistance and threshold voltage drift in chalcogenide materials," in IEEE Transactions on Electron Devices, vol. 51, no. 5, pp. 714-719, May 2004.

[5] Y. M. Kang and S. Y. Lee, "The challenges and directions for the mass-production of highly-reliable, high-density 1T1C FRAM," 2008 17th IEEE International Symposium on the Applications of Ferroelectrics, Santa Re, NM, USA, 2008, pp. 1-2.

[6] MR4A08BUYS45 MRAM datasheet, [Online], Available: https://www.everspin.com/file/882/download, [Accessed: Oct 16, 2018].

[7] RM24C512C CBRAM Datasheet, [Online], Available: http://www.adestotech.com/wp-content/uploads/DS-RM24C32C_056.pdf, [Accessed: Oct 16, 2018].

[8] Intel® Optane™ Memory Series, [Online], Available: https://ark.intel.com/products/97544/Intel-Optane-Memory-Series-16GB-M_2-80mm-PCIe-3_0-20nm-3D-Xpoint, [Accessed: Oct 16, 2018].

[9] FM28V102A FRAM Datasheet, [Online], Available: http://www.cypress.com/file/140901/download, [Accessed: Oct 16, 2018].

[10] K. Kim, H. Jeong, J. Park and S. Jung, "Transient Cell Supply Voltage Collapse Write Assist Using Charge Redistribution," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 63, no. 10, pp. 964-968, Oct. 2016.

[11] S. Mukhopadhyay, R. M. Rao, J. Kim and C. Chuang, "SRAM Write-Ability Improvement With Transient Negative Bit-Line Voltage," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 19, no. 1, pp. 24-32, Jan. 2011.

[12] H. Yang, C. Chang, M. C. -. Chao, R. Huang and S. Lin, "Testing Methodology of Embedded DRAMs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 20, no. 9, pp. 1715-1728, Sept. 2012.

[13] W. F. Brown, "Thermal fluctuations of a single-domain physics", Physical Review, vol. 130, no. 5, pp 1677-1686, Jun. 1963.

[14] E. Karl, Y. Wang, Y.-G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry and M. Bohr, "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," 2012 IEEE International Solid-State Circuits Conference, San Francisco, CA, 2012, pp. 230-232.

[15] S. Motaman, S. Ghosh and N. Rathi, "Impact of process-variations in STTRAM and adaptive boosting for robustness," 2015 Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, 2015, pp. 1431-1436.

[16] J. Jang, J. Park, S. Ghosh and S. Bhunia, "Self-correcting STTRAM under magnetic field attacks," 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC), San Francisco, CA, 2015, pp. 1-6.

[17] Z. Fang, H.Y. Yu, W.J. Liu, Z. R. Wang, X.A. Tran B. Gao and J. F. Kang, "Temperature Instability of Resistive Switching on HfO_x-Based RRAM Devices," in IEEE Electron Device Letters, vol. 31, no. 5, pp. 476-478, May 2010.

[18] U. Russo, D. Ielmini and A. L. Lacaita, "Analytical Modeling of Chalcogenide Crystallization for PCM Data-Retention Extrapolation," in IEEE Transactions on Electron Devices, vol. 54, no. 10, pp. 2769-2777, Oct. 2007.

[19] J. A. Rodriguez, C. Zhou, T. Graf, R. Bailey, M. Wiegand, T. Wang, M. Bali, H. C. Wen, K. R. Udayakumar, S. Summerfelt, T. San and T. Moise, "High Temperature Data Retention of Ferroelectric Memory on 130nm and 180nm CMOS," 2016 IEEE 8th International Memory Workshop (IMW), Paris, 2016, pp. 1-4.

[20] J. J. Ke, Z.-J. Liu, C.-F. Kang, S.-J. Lin and J.-H. He, "Surface effect on resistive switching behaviors of ZnO", Applied Physics Letter, vol. 99, no. 19, pp. 192106, 2011.
[21] M. N. I. Khan and S. Ghosh, "Test challenges and solutions for emerging non-volatile memories," 2018 IEEE 36th VLSI Test Symposium (VTS), San Francisco, CA, 2018, pp. 1-6.
[22] S. Ghosh, M. N. I. Khan, A. De and J. Jang, "Security and privacy threats to on-chip Non-Volatile Memories and countermeasures," 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Austin, TX, 2016, pp. 1-6.
[23] R. K. Aluru and S. Ghosh, "Droop mitigating last level cache architecture for STTRAM". 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE), Belgium, 2017, pp. 262-265.
[24] M. N. I. Khan and S. Ghosh, "Test for Supply Noise Test for Emerging Non-Volatile Memories," 2018 IEEE International Test Conference (ITC), Phoenix, AZ, 2018.
[25] M. N. I. Khan, A. S. Iyengar and S. Ghosh, "Novel Magnetic Burn-In for Retention and Magnetic Tolerance Testing of STTRAM," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 26, no. 8, pp. 1508-1517, Aug. 2018.
[26] A. Iyengar, S. Ghosh and S. Srinivasan, "Retention Testing Methodology for STTRAM," in *IEEE Design & Test*, vol. 33, no. 5, pp. 7-15, Oct. 2016.
[27] I. Yoon, A. Chintaluri and A. Raychowdhury, "EMACS: Efficient MBIST architecture for test and characterization of STT-MRAM arrays," *2016 IEEE International Test Conference (ITC)*, Fort Worth, TX, 2016, pp. 1-10.
[28] H. Naeimi, C. Augustine, A. Raychowdhury, Shih-Lien Lu, James Tschanz, "STTRAM Scaling and Retention Failure," Intel Technology Journal (ITJ), May 2013.
[29] M. N. I. Khan, A. S. Iyengar and S. Ghosh, "Novel magnetic burn-in for retention testing of STTRAM," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017, Lausanne, 2017, pp. 666-669.
[30] M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U-I. Chung, I.-K. Yoo, and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures," Nature Materials, vo. 10, no. 8, pp. 625-630, 2011.
[31] C. Nail, G. Molas, P. Blaise, G. Piccolboni, B. Sklenard, C. Cagli, M. Bernard, A. Roule, M. Azzaz, E. Vianello, C. Carabasse, R. Berthier, D. Cooper, C. Pelissier, T. Magis, G. Ghibaudo, C. Vallee, D. Bedeau, O. Mosendz, B. D. Salvo, and L. Perniola, "Understanding RRAM endurance, retention and window margin trade-off using experimental results and simulations," 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2016, pp. 4.5.1-4.5.4.
[32] C. M. Choi, H. Sukegawa, S. Mitani and Y. H. Song, "Endurance of magnetic tunnel junctions under dynamic voltage stress," in Electronics Letters, vol. 53, no. 16, pp. 1146-1148, 3 8 2017.
[33] S. Kannan, J. Rajendran, R. Karri and O. Sinanoglu, "Sneak-Path Testing of Crossbar-Based Nonvolatile Random Access Memories," in IEEE Transactions on Nanotechnology, vol. 12, no. 3, pp. 413-426, May 2013.
[34] 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), Mumbai, 2015, pp. 187-192.
[35] Memory Fault Models and Testing, [Online], Available: https://www.edn.com/design/integrated-circuit-design/4439803/Memory-fault-models-and-testing, [Accessed: Mar 29, 2019].
[36] Addressing Test Time Challenges, [Online], Available: https://semiengineering.com/addressing-test-time-challenges/, [Accessed: Oct 16, 2018].

Mohammad Nasim Imtiaz Khan (S'17) is currently pursuing a Ph.D. in the School of Electrical Engineering and Computer Science at The Pennsylvania State University (Penn State). Nasim received his bachelor's from the department of Electrical Engineering of Bangladesh University of Engineering and Technology (BUET), 2014.

Swaroop Ghosh (S'04, SM'13) received his Ph.D. from Purdue in 2008. He is an Assistant Professor at Penn State. Dr. Ghosh was senior research and development engineer in Advanced Design, Intel Corp from 2008 to 2012. His research interests include low-power circuit design and hardware security. He is a senior member of IEEE and NAI.