PDS: Pseudo-Differential Sensing Scheme for STT-MRAM

Wang Kang[1], Tingting Pang[1], Bi Wu[1], Weifeng Lv[1], Youguang Zhang[1], Guanyu Sun[2], Weisheng Zhao[1]

[1] Spintronics Interdisciplinary Center, Beihang University, Beijing, 100191, China

[2] Center for Energy-efficient Computing and Applications, Peking University, Beijing, 100871, China

{wang.kang, tingting.pang, bi.wu, zyg, weisheng.zhao}@buaa.edu.cn

lwf@nlsde.buaa.edu.cn; gsun@pku.edu.cn

ABSTRACT

STT-MRAM has been considered as one of the most promising nonvolatile memory candidates in the next-generation of computer architecture. However, the read reliability and dynamic write power concerns greatly hinder its practical application. In this paper, we propose a synergistic solution, namely pseudo-differential sensing (PDS), to jointly address these two concerns. Three techniques, including cell cluster, asymmetric sensing amplifier (ASA) and self-error-detection-correction (SEDC), are proposed to implement the PDS concept. Our experimental results show that the PDS scheme with the 3T3MTJ cell cluster can reduce the area (~21.7%) and write power (~25.6%) of the differential sensing (DS) scheme while improve the read reliability (read margin, ~35.6%) of the typical sensing (TS) scheme for a 16 Mbit cache. Furthermore, the PDS scheme with the 1T3MTJ cell cluster can outperform both the TS and DS schemes in terms of area (~40.0%, ~66.1%), read latency (~16.6%, ~32.1%), read power (~16.7%, ~37.1%), write latency (~5.4%, 16.3%) and write power (~18.6%, ~43.4%).

CCS Concepts

• Hardware~Non-volatile memory • Hardware~Emerging technologies • Software and its engineering~Contextual software domains

Keywords

STT-MRAM, Asymmetric sensing, Reliability, Write power, Error detection and correction

1. INTRODUCTION

With continuous technology scaling, the increasing leakage power and reliability concerns have become two critical challenges for conventional CMOS-based memories, such as SRAM, DRAM and Flash. Recently, emerging nonvolatile memory technologies, including phase-change memory (PCM), magnetic random access memory (MRAM) and resistive random access memory (ReRAM), have attracted much attention and been extensively studied [1-3]. In particular, spin-transfer torque MRAM (STT-MRAM), which employs a bi-directional current to write data into a memory cell instead of utilizing a magnetic field, has shown great potentials in embedded or cache applications. In addition to the nonvolatility like Flash, STT-MRAM provides high integration density close to DRAM, fast access speed and practically unlimited endurance comparable to SRAM [4-6]. These advantageous features enable STT-MRAM to be a promising nonvolatile memory candidate.

Although attractive, STT-MRAM also suffers from challenges before wide application. The first challenge of STT-MRAM is the high write power consumption [7,8]. As we all known that, STT-MRAM utilizes a bi-directional current to write data into a memory cell (or magnetic tunnel junction, MTJ) through STT mechanism. First, the STT-driven MTJ switching requires a relatively large current to overcome the energy barrier between the two resistance states of the MTJ device. Second, the STT switching mechanism is intrinsically stochastic and the actual time to complete the write operation varies dramatically among all the memory cells of a chip, thereby a long write current pulse is required for a reliable write operation taking into consideration the worst-case corner [7]. It has shown that the write power per bit of STT-MRAM is one order of magnitude higher than SRAM. The second challenge of STT-MRAM is the poor read reliability. This can be explained from two perspectives: First, the tunnel magneto-resistance (TMR, typically < 300% at room temperature) ratio of the MTJ is generally limited by the device material and structure, thus resulting in limited read margin (RM). Second, the process-voltage-temperature (PVT) variations of both the MTJs and CMOS transistors further degrades the RM [9-10]. Increasing the read current can improve the RM to some extent, but it should be noted that large current leads to high read disturbance probability, again affecting the read reliability.

To reduce the write power of STT-MRAM, two strategies can be employed. The first one is to reduce the required write current amplitude and pulse duration [11,12] from the device design point of view. This strategy is direct and effective, but it is generally limited by the device material and structure. The second one is to reduce the write operation frequency through circuit or architecture level design techniques [13, 14]. For improving the read reliability, the self-reference sensing (SRS) scheme, including destructive or non-destructive SRS schemes, can reduce the PVT variations by eliminating the reference cell. However the destructive SRS needs to restore the data after each read operation, wasting much power, while the non-destructive one has limited RM [15,16]. Differential sensing (DS) is an effective solution to improve the RM, however, it requires two complementary memory cells to store only one bit of data, introducing large area and power overheads [17, 18]. In this work, we propose a synergistic pseudo-differential sensing (PDS) scheme for STT-MRAM to jointly address the read reliability and write power concerns. In order to implement the PDS concept, we propose three design techniques, including the cell cluster structure, asymmetric sensing amplifier (ASA), and self-error-detection-correction (SEDC). Our simulation results show that the proposed PDS scheme can achieve good performance in terms of area, power, latency as well as reliability. In summary, the main contributions of this paper can be summarized as follows,

- We propose a synergistic PDS scheme to jointly address the read reliability and write power concerns of STT-MRAM from cross-level design point of view.
- We propose a cell cluster structure to re-organize the data bits. This cell cluster structure adds redundancy to improve the read reliability and to reduce write operations.

- We propose a novel ASA to readout the data stored in the cell cluster with improved RM.
- We propose a SEDC module combined with the cell cluster structure and the ASA to utilize the redundancy for detecting or correcting possible errors.

The rest of the paper is organized as follows. In Section II, we introduce the backgrounds of STT-MRAM and our motivation. Section III presents the concept, implementation and discussion of our proposed PDS scheme. Our evaluation results are presented in Section V. Finally Section VI concludes the paper.

2. BACKGROUND AND MOTIVATION
2.1 STT-MRAM Basics

As shown in Fig. 1, a typical STT-MRAM bit-cell consists of a MTJ connected in series with a NMOS transistor, named 1T1MTJ cell structure [18]. The MTJ is the core device of STT-MRAM for storing data bit information while the NMOS transistor acts as an access device. The memory cell can be randomly accessed by controlling the bit-line (BL), word-line (WL) and source-line (SL). An MTJ is mainly composed of three ultra-thin layers: one oxide barrier layer (e.g., MgO) sandwiched between two ferromagnetic (FM) layers (e.g., CoFeB). The magnetization orientation of one FM layer is fixed (named pinned layer, PL) while the other FM layer is free to change (named free layer, FL). An MTJ presents two stable resistance states (i.e., low resistance, R_P and high resistance, R_{AP}) depending on the relative magnetization orientation (parallel (P) or anti-parallel (AP)) of the two FM layers. Therefore each MTJ can store one bit of data. The resistance difference between the two stable resistance states can be characterized by the TMR ratio, i.e., $TMR = (R_{AP} - R_P)/R_P$. To write data bit into the MTJ, only a bi-directional current (through STT mechanism [19]) is required. Due to the TMR ratio, the data stored in the STT-MRAM bit-cell can be readout by distinguishing the two different resistance states of the MTJ utilizing voltage or current sensing techniques.
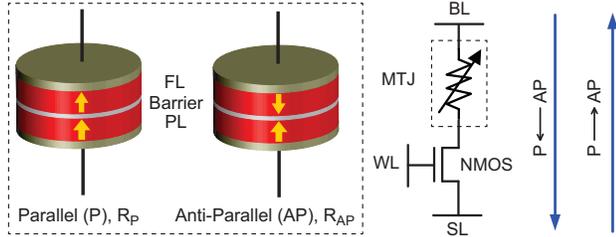
Fig. 1. Typical 1T1MTJ bit-cell structure and writing method [18].

2.2 Write Power Concerns

The write power concern of STT-MRAM mainly comes from three factors. The first one is that the switching of the MTJ by STT effect requires a relatively high current density. The critical current densities for in-plane and out-of-plane MTJs are expressed as [11],

$$J_{in-plane} = \frac{2e}{\hbar}\frac{\alpha}{\eta}(M_s t_F)(4\pi M_s + H_K). \quad (1)$$

$$J_{out-of-plane} = \frac{2e}{\hbar}\frac{\alpha}{\eta}(M_s t_F)H_K. \quad (2)$$

$$\Delta = E_b/k_B T = H_K M_s V/2k_B T. \quad (3)$$

where e is the charge of electrons, $\hbar$ the reduced planck's constant, η the spin transfer efficiency, α the Gilbert damping constant, M_s the saturation magnetization, t_F the thickness of the FL, V the volume of the FL and H_K the effective anisotropy energy. As can be compared from Eq. (1) and (2), the critical current density for switching the out-of-plane MTJ is small than that of the in-plane

MTJ while maintaining the same thermal stability. In addition to the critical switching current density, the out-of-plane MTJ also has better scalability than that of the in-plane MTJ. Thereby we choose the out-of-plane MTJ in this work. The typical value of the critical current density amplitude and duration for the out-of-plane MTJ are about several MA/cm^2 and several nanoseconds respectively [11].

The second concern comes from the asymmetry, including the STT switching asymmetry and the 1T1MTJ bit-cell asymmetry. The STT switching asymmetry is intrinsically physical, as the spin transfer efficiency η is determined by the relative magnetization orientation of the two FMs, as [20],

$$\eta = (P/2)/(1 + P^2 cos\theta). \quad (4)$$
$$J_{P-AP}/J_{AP-P} = (1 + P^2)/(1 - P^2). \quad (5)$$

where P is the tunneling spin polarization, θ the angle between the magnetization orientation of the two FMs, J_{P-AP} and J_{AP-P} are the critical current density for the MTJ switching operation of P → AP and AP → P, respectively. We found that J_{P-AP} is much higher than J_{AP-P}. The 1T1MTJ bit-cell asymmetry is due to the source degradation of the access NMOS transistor, which affects the transistor driving capability because different bias conditions are utilized for writing data bits "0" and "1" [20]. In practice, we should consider the worst-case corner for reliable write operations, leading to write power wastage.

The third concern comes from the intrinsic stochastic STT-driven MTJ switching property, random thermal fluctuation and the process variations. The stochastic magnetization dynamics of the FL of the MTJ can be characterized by modifying the Landau-Lifshitz-Gilbert (LLG) equation, taking into consideration the random thermal effect [18, 19].

$$dm/dt = \gamma m_f \times (H_{eff} + H_{fluc}) - \alpha m_f \times (m_f \times (H_{eff} + H_{fluc})) + J(\theta)(m_f \times m_f \times m_p) \quad (6)$$

where γ is the gyro-magnetic constant, m_f and m_p are the unit magnetization vectors of the FL and PL respectively, H_{eff} the effective magnetic field, H_{fluc} the thermal induced random field, $J(\theta)$ is the coefficient of the STT term, depending on the initial angle between the magnetization orientation of the two FMs. The stochastic MTJ magnetization switching behaviors introduce time-to-time variations. In addition, the PVT variations of the MTJs and CMOS transistors further add cell-to-cell stochasticity across the whole memory chip. Again, much power consumption is wasted to cover the worst case of the chip. Taking all the above-mentioned concerns into consideration, Fig. 2 shows our simulation results of the write power distributions for the 1T1MTJ cell at 40 nm node.
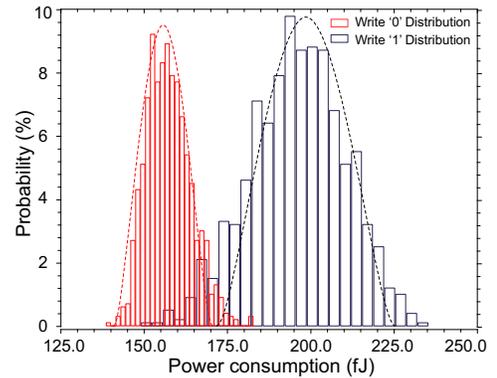
Fig. 2. 1T1MTJ cell write power distributions at 40 nm technology.

2.3 Read Reliability Concerns

In the TS scheme, the data bit stored in an MTJ is readout by firstly applying an external current I_{read} (or voltage) on the data and reference cells, then comparing the sensed voltage V_{data} (or current) with that (V_{ref}) of the reference cell, finally amplifying the voltage (or current) difference between the data and reference cells to a digital value [21,22]. As the TMR ratio of MTJ is limited by the device material and structure, the RM is also limited. If the RM cannot overcome the input-offset of the sensing amplifier, decision error occurs. Increasing the read current can improve the RM to reduce the decision error (see Eq. (7)) [23], however, this may in turn increase the read disturbance probability Pr_{dis} (see Eq. (8)).

$$\text{RM} = \min\{|V_{ref} - V_{data0}|, |V_{ref} - V_{data1}|\}. \tag{7}$$

$$Pr_{dis} = 1 - exp\left\{-\frac{t_{read}}{\tau_0}\exp\left[-\Delta\left(1-\frac{I_{read}}{I_{C0}}\right)\right]\right\}. \tag{8}$$

where I_{read} and t_{read} are the sensing current pulse amplitude and duration, τ_0 the attempt period, I_{C0} the critical current amplitude, V_{data0} and V_{data1} are the sensed voltages of the data memory cells in low and high resistance states respectively.

As technology scales, on one hand, the critical current of the MTJ reduces, that means the sensing current should also be reduced to avoid increasing the read disturbance probability (see Eq.(8)). On the other hand, the PVT variations increase, resulting in higher mismatch or input-offset of the sensing circuit. Technology scaling further adds design challenges on the read reliability of future STT-MRAM. Fig. 3 shows the simulation results of the decision error and read disturbance probability in different technology nodes.
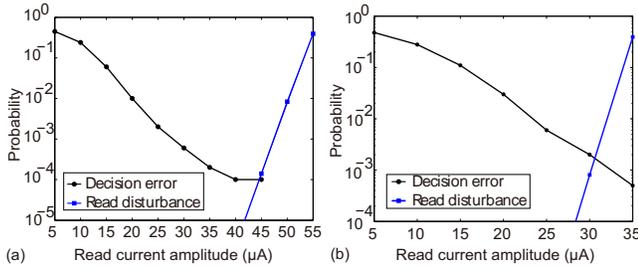
Fig. 3. The decision error and read disturbance with respect to the read current (pulse width of 0.5 ns) in (a) 40 nm; (b) 28 nm.

2.4 Motivation

For the write power concerns, we know that the critical current density and the stochasticity of the STT-driven MTJ switching are determined by the device characteristics or manufacturing process based on the above analyses. However we can reduce the write power through circuit or architecture design techniques taking into consideration the asymmetry features. As writing data bit "1" (i.e., P → AP switching of the MTJ) consumes much more power than writing bit "0" (i.e., AP → P switching of the MTJ), we should design to reduce writing "1" operations as much as possible. For the read reliability, as the process variations significantly increase while the critical current reduces, the conflict between decision error and read disturbance becomes more serious as technology scales. DS is an effective solution to keep the RM while reduce the read disturbance since it is able to double the RM compared to the TS scheme. The problems of DS scheme are its area and power overheads, as it requires two MTJs (named 2T2MTJ cell) to store only one single bit of data information. In this work, by optimizing the conventional DS concept and utilizing the asymmetry features, we propose a synergistic PDS scheme to address the read reliability of TS, while reduce the area and write power concerns of DS.

3. PSEUDO-DIFFERENTIAL SENSING

The schematic of the proposed PDS scheme is shown in Fig. 4, which mainly includes three design techniques, including cell cluster structure, asymmetric sensing amplifier (ASA) as well as self-error-detection-correction (SEDC) module. The concept of the PDS scheme is to take advantages of the DS scheme to improve RM, meanwhile to reduce the area and power overheads. In order to achieve this goal, we propose to compare the resistance states of the MTJs with each other in a local cell cluster. Within the proposed PDS architecture, the data representation, memory cell structure and the peripheral sensing amplifier should all be re-designed.
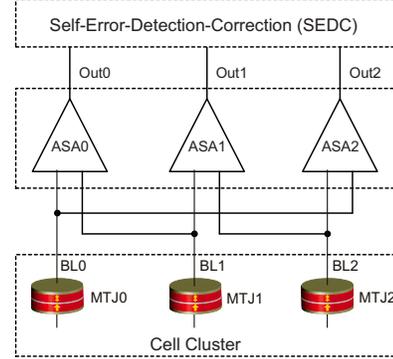
Fig. 4. The overall schematic of the proposed PDS scheme.

3.1 Cell Cluster Structure

In the proposed cell cluster structure, several (e.g., three) MTJs are formed together to represent one data symbol of two bits (see Table 1) and the data are readout by comparing the resistance states of the MTJs in the cell cluster. We utilize three MTJs (with eight states) to represent only two data bits (four states), two different states of the cell cluster are used to stand for the same data symbol. This is named state-restrict mapping (SR-mapping). Specifically, the states {000} and {111} stand for data symbol (00), {001} and {011} stand for (01), {010} and {110} stand for (10) and {100} and {111} stand for (11). Based on the analyses in Section 2.2, we have known that writing data "1" consumes much more power than writing data "0". In this case, we can choose the cell cluster states with less "1" to represent the data symbol i.e., {000} ↔ (00), {001} ↔ (01), {010} ↔ (10) and {100} ↔ (11). As a result, any write operation of the cell cluster involves at most one 0 → 1 and 1 → 0 transitions (see Fig. 5(a)). Another strategy is to first reset the cell cluster to the {000} state, then to the target state based on the new data symbol (see Fig. 5(b)). In addition, as two cell cluster states represent only one data symbol, which adds redundancy for improving data robustness (which will be shown in Section 3.3).

Table 1. Data representation in the proposed PDS scheme.

Data symbols	MTJ states	Outputs of ASAs
(2 bits)	MTJ0, MTJ1, MTJ2	ASA0, ASA1, ASA2
00	**0 0 0**	0 0 0
	1 1 1	
01	**0 0 1**	0 0 1
	0 1 1	
10	**0 1 0**	0 1 0
	1 1 0	
11	**1 0 0**	1 0 0
	1 0 1	
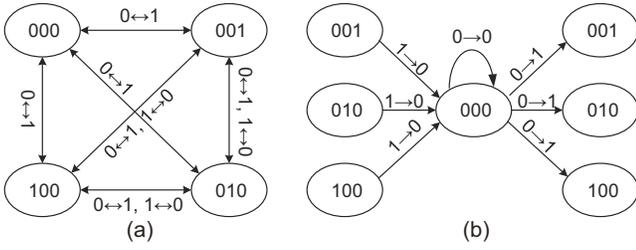
Fig. 5. The state transition diagram of the proposed cell cluster; (a) Direct write strategy; (b) Two-step write strategy.
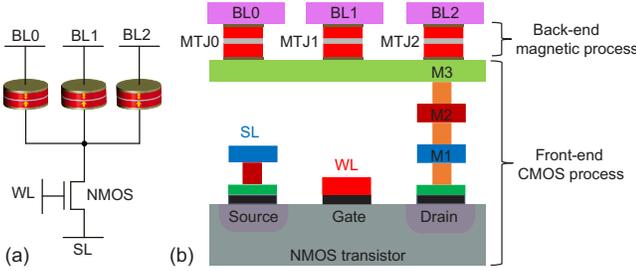
Fig. 6. The proposed cell cluster structure; (a) symbol; (b) layout.

Furthermore, the direct combination of three 1T1MTJ cells to form a cell cluster (named 3T3MTJ cluster) results in much area penalty. As we know that the MTJ is fabricated on top of the access transistor using the back-end-of-line (BEOL) process and the bit-cell area mainly depends on the access transistor size [24]. Thereby we further propose to cluster several (e.g., three) MTJs on top of one transistor, named 1T3MTJ cell cluster, as shown in Fig. 6. With this proposed cell cluster, the memory array structure (which will be shown in Section 3.4) should be reorganized, where the BLs within a local cell cluster form a BL cluster and are accessed simultaneously through a global BL (G_BL). This cell cluster, however, may fail to provide enough current drivability for writing the three MTJs at the same time if the transistor size is limited. In this case, we utilize a read-before-write (RBW) method and employ the two-step write strategy (see Fig. 5(b)) to solve this problem by sacrificing the write performance. As any data symbol has only one MTJ with states "1" (see Table 1), we only need to switch the MTJ with state "1" to be "0" and then program another MTJ to be "1" based on the new data symbol. Thereby in the RBW method, any write cycle only needs to switch one of the three MTJs, adding no area overhead on the transistor compared to the 1T1MTJ cell.

3.2 Asymmetric Sensing Amplifier

In the DS scheme, the two MTJs (or the two inputs) of the SA are always in complementary states. In the proposed PDS scheme, however, as can be seen from Fig. 4 and Table 1, the inputs of the SA may be the same. For example, given the cell cluster state of {000}, the MTJ states (inputs of the three SAs) are all in the low resistance states. In this case, the output of the SA will be instable mainly depending on the PVT variations. To solve this problem, we propose an asymmetric sensing amplifier (ASA) particularly for the PDS scheme, as illustrated in Fig. 7(a), in which we intentionally add a pre-known offset (ΔV) between the two inputs of the SA. This input-offset can be introduced, for example, by changing the load resistances between the two branches or by pre-charging one of the inputs of the conventional SA. For simplicity, we assume that: (a) the right branch of the ASA always has a positive ΔV compared to that of the left branch; and (b) the ASA outputs a digital "0" if the potential of the right input is higher than that of the left input,
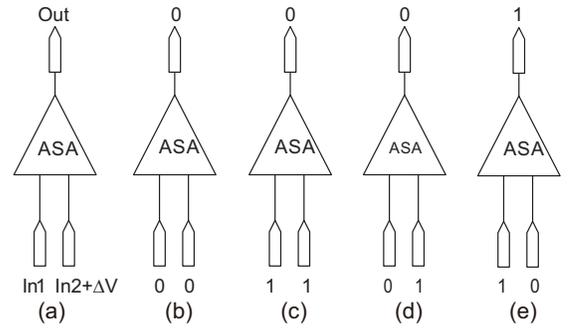
Fig. 7. (a) The concept of the ASA; (b)-(e) Four input cases.
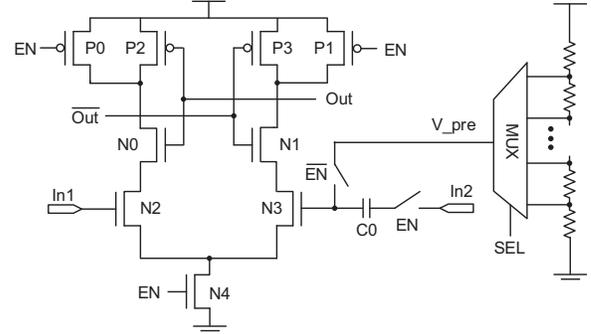
Fig. 8. The schematic of the proposed ASA design (an example).

otherwise, it output a digital "0". As a result, there are three cases (ideally) for the proposed ASA (see Fig. 7(b)-(e)): (a) The MTJ states for the two inputs of the ASA are the same, i.e., {00} or {11}, then the right input of the ASA has a larger potential and the ASA outputs a digital bit "0". The RM depends on the amplitude of ΔV; (b) The MTJ states for the two inputs of the ASA are {01}, then the ASA outputs a data bit "0" and the RM is $(|V_{data1} - V_{data0}| + \Delta V)$, which is even larger than that of the DS scheme by ΔV ; (c) The MTJ states for the two inputs of the ASA are {10}, then the ASA outputs a digital bit "1" and the RM is $(|V_{data1} - V_{data0}| - \Delta V)$, which is less than that of the DS scheme by ΔV. In summary, we will have the digital outputs of the three ASAs for all the cell cluster states as following: $[000] \leftrightarrow \{000, 111\}$, $[001] \leftrightarrow \{001, 011\}$, $[010] \leftrightarrow \{010, 110\}$, $[100] \leftrightarrow \{100, 101\}$, as shown in Table 1. Then the challenge turns to how to set the value of ΔV to achieve the optimal RM. An intuitive value is $\Delta V = (V_{data0} + V_{data1})/2$ to trade off among all the three cases, then the average RM of the proposed ASA is exactly the same as that of the typical SA when considering no PVT variations. This is the origin of the name "pseudo-DS". Even in this case, the PDS (or ASA) outperforms the typical SA when taking into consideration the PVT variations for the following three reasons: (a) In the proposed PDS scheme, there is no need of reference cell. All the data cells at the same time act as reference cells for each other in a local cell cluster. Therefore there is no regularity problem, which denotes the process parameter difference between the data and reference cells; (b) The proposed ASA only involves MTJs locally in a cell cluster, thereby it has the advantage of little parasitic mismatch; (c) As three outputs can be obtained to represent only two data bits, adding redundancy for error detection and correction (which will be shown in the next sub-section 3.3). In practical applications, we can first test the memory chip after fabrication and then set the value of ΔV adaptively according to the real PVT variations. Fig. 8 shows an example of ASA design with the pre-charging method, in which we can change ΔV dynamically by modulating the pre-charge voltage (V_pre).

3.3 Self-Error-Detection-Correction

By combining the cell cluster structure and the ASA design, the PDS scheme provides also self-error-detection-correction (SEDC) capability. First we discuss the self-error-correction (SEC). As two cell cluster states represent the same data symbol (see Table 1), the correct data information can be readout if the error result in cell cluster state transition between these two states. For example, assume that the data symbol is (01) and then the initial cell cluster state is {001}, if the state of the middle MTJ flips due to radiation particle or thermal fluctuation, then the cell cluster state turns to be {011}, which is correctly readout to be {001}, thanks to the ASA design. This SEC process is automatic and transparent, however the error correction capability is limited by the specific error pattern. Now we discuss the self-error-detection (SED) capability. As can be seen from Table 1, for all the cell cluster states, there are four combinations of outputs for the three ASAs, including [000], [001], [010] and [100]. We can find that at most one output bit "1" is valid for a correct data symbol. With this finding, we can introduce SED capability into the PDS scheme by only adding a majority voter (see Fig. 9). If two or more SASs output digital bits "1" due to radiation effects or voltage-current fluctuations, the SEDC module is able to detect the error and output an acknowledge signal.
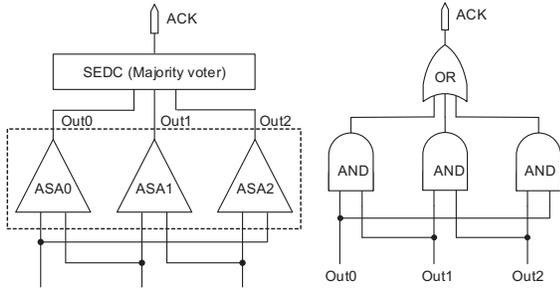
Fig. 9. (a) The schematic of the SEDC module combined with the ASA design; (b) Majority implementation.

3.4 Overall Architecture and Discussion

Integrating the above three design techniques, we present the overall PDS architecture for STT-MRAM, as shown in Fig. 10. The main differences from the typical STT-MRAM architecture are the memory cell array, read circuit, data representation and the memory controller. The local BLs in the cell cluster form a G_BL and are accessed simultaneously. During write operations, input data bit sequence is firstly truncated into (2-bit) data symbols, and then the data symbols are stored in the cell clusters through the write driver. During read operations, data symbols stored in the cell clusters are readout by the ASAs, then they are mapped back to the original data bit sequence before entering the input/output (I/O) module. If an error is detected by the SEDC module, an error acknowledgement signal is generated and transferred to the memory controller.
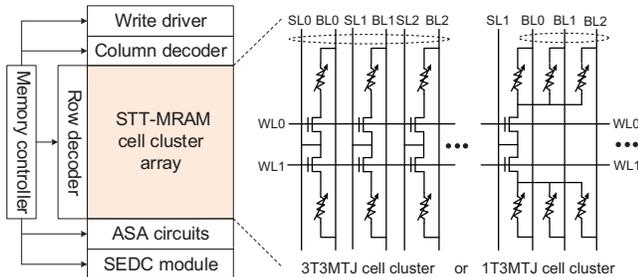
Fig. 10. The overall PDS architecture for the STT-MRAM.

One intuitive question of the PDS scheme maybe proposed, that is, how about forming the cell cluster with more (> three) MTJs. This extension is theoretically reasonable but it adds design issues and challenges in practice. First of all, from device point of view, the size of the transistor is limited for putting more than three MTJs on top of it. Second, much more ASAs are required to readout the data of the cell cluster. For example, if the cell cluster has four MTJs, then it needs six ASAs. The number of the ASAs is more than that of the local BLs in the cell cluster. In addition, the data representation and array structure will be much more complex.

4. EVALUATION

This section presents our evaluation results of the PDS scheme, which include two parts: circuit level and architecture level. The circuit level evaluations were performed on Cadence platform with a physics-based STT-MTJ model [25] under the 40 nm technology node. The architecture level evaluations were conducted on NVSim [26] with word width of 64-bytes, 8-way cache configuration.

4.1 Circuit Level Evaluation

Fig. 11 shows the transient waveforms of the ASAs. Here "R" and "W" denote read and write cycles respectively. As can be seen, the output results of the three ASAs are consistent with those listed Table 1. Fig. 12 shows the RM of the ASA compared to that of the typical SA. The case with input states {01} achieves the biggest RM, while the cases with input states {00} and {11} result in the smallest RM. Anyway, the average RM in the worst-case of the proposed ASA is much better (~35.6%) than that of the typical SA.

4.2 Architecture Level Evaluation

We compare the performance in terms of area, read latency, read power, write latency and write power among SRAM, 1T1MTJ (TS) cell, 2T2MTJ cell (DS), 3T3MTJ cell cluster (PDS) as well as 1T3MTJ cell cluster (PDS), as shown in Fig. 13. We find that the performance of the PDS scheme with the 3T3MTJ cell cluster is between the TS (1T1MTJ) and DS (2T2MTJ) schemes, which is consistent with our intuition. Nevertheless, the PDS scheme with the 1T3MTJ cell cluster outperforms both the TS and DS schemes in terms of area (~40.0%, ~66.1%), read latency (~16.6%, ~32.1%), read power (~16.7%, ~37.1%), write latency (~5.4%, 16.3%) and write power (~18.6%, ~43.4%) with the 16 Mbit cache for example.
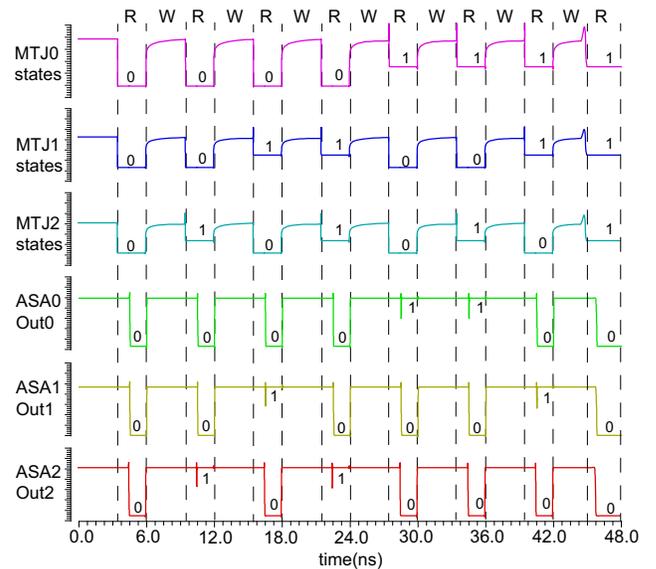
Fig. 11. Transient waveforms of the proposed ASA.
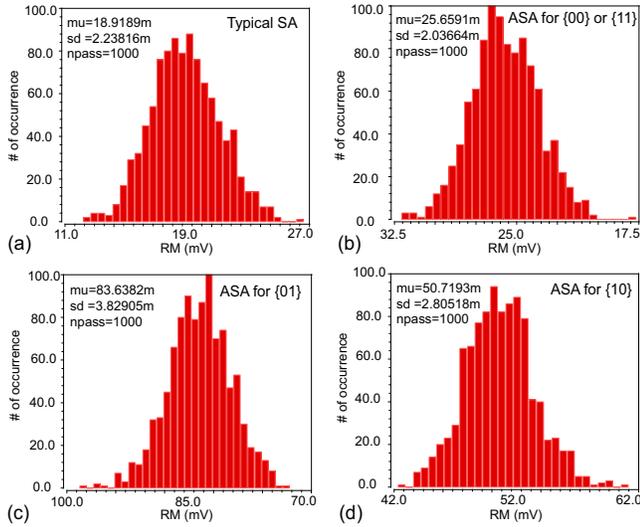
Fig. 12. RM statics of the (a) typical SA, (b) ASA with inputs {00} or {11}, (c) ASA with inputs {01} and (d) ASA with inputs {10}.
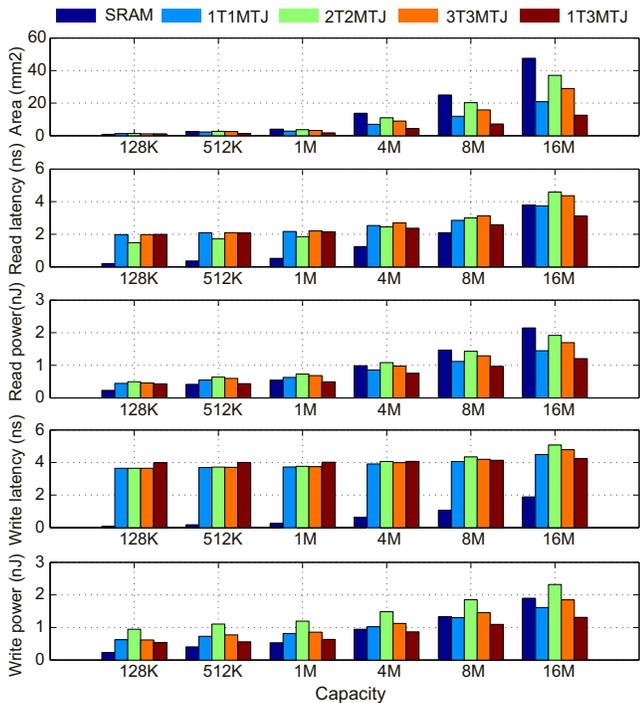
Fig. 13. Architecture level evaluations; (a) area, (b) read latency; (c) read power; (d) write latency; (e) write power.

5. CONCLUSION

Write power and read reliability are two big challenges for the real application of STT-MRAM. In this work, we proposed PDS, a synergistic solution integrating three design techniques, i.e., cell cluster, ASA and SEDC, to jointly address these two concerns. In addition, a new data representation, SR-mapping, was designed to coordinate with the PDS scheme. We performed circuit level and architecture level simulations to evaluate the performance of the proposed PDS scheme. Our experimental results show that the PDS scheme with the 1T3MTJ cell cluster improves the read reliability of the TS scheme while reduces the area, read latency, read power, write latency and write power of both the TS and DS schemes.

6. REFERENCES

[1] "International Technology Roadmap for Semiconductors". Online http://www.itrs.net/, 2013.

[2] C. J. Xue et al., "Emerging non-volatile memories: opportunities and challenges", in *IEEE/ACM/IFIP Int. Conf. Hardware/ Software codesign and system synthesis*, pp. 325-334, 2011.

[3] H.-S. Philip Wong and S. Salahuddin, "Memory leads the way to better computing", *Nature Nanotechnol.*, vol. 10, pp. 191-194, 2015.

[4] G. W. Burr, et al., "Overview of candidate device technologies for storage-class memory", *IBM J. Res. Dev.*, vol. 52, no. 4.5, pp. 449-464, 2008.

[5] W. Kang, et al., "Spintronics: Emerging ultra-low power circuits and systems beyond MOS technology", *ACM J. Emerg. Technol. Comput. Syst.*, vol. 12, no. 2, Article 16, 2015.

[6] A. Jog, et al., Cache revive: architecting volatile STT-RAM caches for enhanced performance in CMPs, in *ACM/EDAC/IEEE DAC*, pp. 1187-1192, 2012.

[7] D. Lee and K. Roy, "Energy-delay optimization of the STT-MRAM write operation under process variations. *IEEE Trans. Nanotechnol.,*, vol. 13, no. 4, pp. 714-723, 2014.

[8] D. Suzuki, et al. "Cost-efficient self-terminated write driver for spin-transfer-torque RAM and logic", *IEEE Trans. Magn.*, vol. 50, no. 11, pp. 3402104, 2014.

[9] Y. Zhang et al., "Read Performance: The Newest Barrier in Scaled STT-RAM", *IEEE Trans. VLSI.*, vol. 23, no. 6, pp. 1170-1174, 2015.

[10] W. Kang, et al., "Variation-tolerant and disturbance-free sensing circuit for deep nanometer STT-MRAM". *IEEE Trans. Nanotechnol.*, vol. 13, no. 6, pp. 1088-1092, 2014.

[11] R. Sbiaa, et al., "Reduction of switching current by spin transfer torque effect in perpendicular anisotropy magneto-resistive devices", *J. Appl. Phys.*, vol. 109, no. 7, pp. 07C707, 2011.

[12] E. Eken, et al. "A new field-assisted access scheme of STT-RAM with self-reference capability", in *ACM/EDAC/IEEE DAC*, pp. 1-6, 2014.

[13] P. Zhou, et al. "Energy reduction for STT-RAM using early write ermination", in *IEEE/ACM ICCAD*, pp. 264-268, 2009.

[14] R. Bishnoi, et al., "Avoiding unnecessary write operations in STT-MRAM for low power implementation", in *ISQED*, pp. 548-553, 2014.

[15] H. Tanizaki, et al., "A high-density and high-speed 1T-4MTJ MRAM with voltage offset self-reference sensing scheme", in *IEEE ASSCC*, pp. 303-306, 2006.

[16] Y. Chen, et al., "A nondestructive self-reference scheme for spin-transfer torque random access memory (STT-RAM)," in *IEEE DATE*, pp. 148-153, 2010.

[17] H. Noguchi, et al., "Variable nonvolatile memory arrays for adaptive computing systems", in *IEEE IEDM.*, pp. 25.4.1-25.4.4, 2013.

[18] W. Kang, et al, Reconfigurable codesign of STT-MRAM under process variations in deeply scaled technology. *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1769-1777, 2015.

[19] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *J. Magn. Magn. Mater.*, vol. 159, nos-. 1-2, pp. L1–L7, 1996.

[20] Y. Zhang, et al., "Asymmetry of MTJ Switching and Its Implication to the STT-RAM Designs", in *IEEE DATE*, pp. 1313-1318, 2012.

[21] J. Kim, et al., "A novel sensing circuit for deep submicron spin transfer torque MRAM (STT-MRAM)". *IEEE Trans. VLSI. Syst.*, vol. 20, no. 1, pp. 181-186, 2012.

[22] W. Kang, et al, High reliability sensing circuit for deep submicron spin transfer torque magnetic random access memory. *Electron. Lett.*, vol. 49, no. 20, pp. 1283-1285, 2013.

[23] W. S. Zhao, et al., "Failure and reliability analysis of STT-MRAM," *Microelectron. Releliability.*, vol. 52, no. 9, pp. 1848-1852, 2012.

[24] M. C. Gaidis, et al., "Two-level BEOL processing for rapid iteration in MRAM development", *IBM J. Res. Dev.*, vol. 50, no. 1, pp. 41-54, 2006.

[25] Y. Zhang, et al., "Compact model of subvolume MTJ and its design application at nanoscale technology nodes", *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 2048-2055, 2015.

[26] X. Dong et al., "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory", *IEEE Trans. Comput. Aided Des. Integr. Circ. Syst.*, vol. 31, no. 7, pp. 994-1007, 2012.