

Design of Spin-Torque Transfer Magnetoresistive RAM and CAM/TCAM with High Sensing and Search Speed

Wei Xu, Tong Zhang, *Senior Member, IEEE*, and Yiran Chen, *Member, IEEE*

Abstract—With a great scalability potential, nonvolatile magnetoresistive memory with spin-torque transfer (STT) programming has become a topic of great current interest. This paper addresses cell structure design for STT magnetoresistive RAM, content addressable memory (CAM) and ternary CAM (TCAM). We propose a new RAM cell structure design that can realize high speed and reliable sensing operations in the presence of relatively poor magnetoresistive ratio, while maintaining low sensing current through magnetic tunneling junctions (MTJs). We further apply the same basic design principle to develop new cell structures for nonvolatile CAM, and TCAM. The effectiveness of the proposed RAM, CAM and TCAM cell structures has been demonstrated by circuit simulation at 0.18 μm CMOS technology.

Index Terms—Content addressable memory (CAM), magnetic tunneling junction (MTJ), magnetoresistive random access memory (MRAM), spin-torque transfer (STT) magnetoresistive memory, ternary CAM (TCAM).

I. INTRODUCTION

DRIVEN by the ever exploding demands for higher capacity nonvolatile solid-state data storage in numerous pervasive computing and communication devices, flash memories have been the fastest growing segment in global semiconductor industry [1]. Nevertheless, besides several drawbacks including low speed, limited endurance, and difficulty of integration in system-on-chip (SoC), flash memories face significant scaling problems at the 32 nm node and beyond [2]–[4]. Hence, it has been a topic of great current interest to search for new nanoscale nonvolatile digital memories that have greater scalability potentials and achieve better performance in terms of speed, endurance, and/or SoC integration. Magnetoresistive random access memory (MRAM) is one of the most promising candidates that have recently attracted a lot of attentions.

The basic building block in MRAM is magnetic tunneling junction (MTJ), and the data storage is realized by configuring the resistance of MTJs into one of two possible states (i.e., high-resistance state and low-resistance state). In conventional MRAM design practice, magnetic fields are explicitly generated and used to switch the state of MTJs [5]–[8]. In stead

of using magnetic fields, a new technique called spin-torque transfer (STT) uses current of spin-aligned electrons (i.e., spin-polarized current) to switch the state of MTJs. Because of its greater scalability potential than conventional MRAM, STT MRAM has received a growing interest [9]–[12]. The resistance state of MTJs can be detected by sensing the current steered through MTJs. Because of the relatively small MTJ resistance values (e.g., a few kilohms) and low magnetoresistive ratio (typically less than 100%), how to achieve high-speed read without incurring large through-MTJ sensing current is a nontrivial task. More importantly, since both writing and read operations in STT MRAM steer current through MTJs, the through-MTJ sensing current must be sufficiently less than the writing current to avoid read disturb. This makes it more challenging to achieve high-speed read in STT MRAM.

Most prior work on MRAM circuit design employs a compact 1 MTJ per cell structure, and the read operation is realized by direct current sensing, i.e., by clamping a voltage to the bit line (BL) through a current source, the through-MTJ sensing current is converted to a sensing voltage that is compared to a reference voltage, as illustrated in Fig. 1(a). As pointed out in [8] and [13], such direct current sensing strategy tends to result in relatively complex and large peripheral circuits. Due to the more severe constraint on through-MTJ current in STT MRAM, it can be more challenging to apply this direct sensing strategy to realize high-speed read in STT MRAM. A 2-MTJ per cell structure [8], [13], as illustrated in Fig. 1(b), has been used to improve the read speed at the cost of cell size. Each cell contains a pair of differential MTJs (i.e., 1 MTJ has high resistance and another has low resistance) that forms a voltage divider, which may greatly simplify the peripheral sensing circuits and improve the sensing speed and reliability. Clearly, the sensing speed and reliability of such 2 MTJ per cell structures are also subject to the magnetoresistive ratio and through-MTJ sensing current constraints, particularly in STT MRAM. To further push the sensing speed limit, sakimura et al. [8] proposed to add a two-transistor common-source amplifier in each cell to amplify the output of the 2-MTJ voltage divider. Nevertheless, in spite of the much higher sensing speed and reliability, the 2-MTJ per cell design strategies presented in [8], [13] result in much larger cell area than a simple 1 MTJ per cell structure.

In this paper, we develop a new STT MRAM structure that applies the same voltage-divide-based sensing strategy in [8] and [13] to improve the sensing speed and reliability, while using only 1 MTJ and one transistor in each cell. In the proposed STT MRAM structure, each BL has one reference cell

Manuscript received April 07, 2008; revised July 31, 2008. First published March 27, 2009; current version published December 23, 2009.

W. Xu and T. Zhang are with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: xuw@rpi.edu; tzhang@ecse.rpi.edu).

Y. Chen is with Alternative Technology Group, Seagate Technology, Bloomington, MN 55435 USA (e-mail: yiran.chen@seagate.com).

Digital Object Identifier 10.1109/TVLSI.2008.2007735

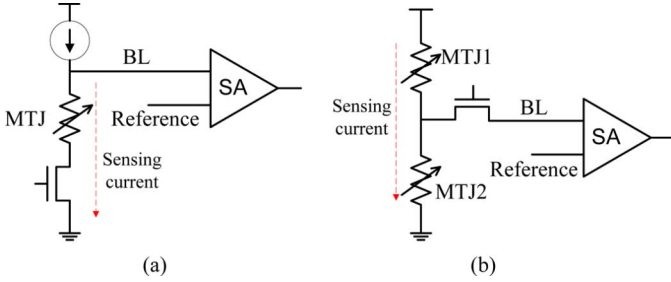


Fig. 1. Conventional sensing strategy when using (a) 1 MTJ per cell and (b) 2 MTJ per cell.

that is used to form a voltage divider with the selected cell during the sensing operation. Both the transistors and MTJs in the reference cell and selected cell participate the voltage divide, and the voltage-modulated ON resistance of transistors can largely amplify the overall resistance ratio. As a result, a significant improvement of sensing voltage margin, and hence, sensing speed/reliability can be achieved. We evaluate this proposed design strategy through simulations at 0.18 μm CMOS technology node. We consider different BL sizes, ranging from 64 up to 256 cells per BL. For 256 cells per BL, we show that the access time of only 7.5 ns is possible while the through-MTJ sensing current is only about 70 μA , assuming that the high and low MTJ resistance fall into the range of 5–6 $\text{K}\Omega$ and 2–3 $\text{K}\Omega$, respectively.

Using the same design principle, we further develop MTJ-based cell structures for nonvolatile content addressable memory (CAM) and ternary CAM (TCAM). Accessed by the content other than address, CAM and TCAM are ideally suited for any systems that require a large amount of data search, e.g., Ethernet address lookup, cache tags, data compression, etc. We develop 2-MTJ per cell CAM structure and 3-MTJ per cell TCAM structure that can achieve high search speed at small through-MTJ search current. As a test vehicle, we designed match-lines of 144 CAM cells and 144 TCAM cells at 0.18 μm CMOS technology. The worst case search time is about 5 ns for CAM at sensing current of 55 μA and 8 ns for TCAM at sensing current of 87 μA , where the search noise margin is over 500 mV.

The paper is organized as follows. Section II provides the fundamental concept of STT MRAM. Section III describes the proposed MRAM cell structure. Section IV describes the MTJ-based CAM cell structures. Finally, a summary of this paper is provided in Section V.

II. STT MRAM BASICS

Like the first generation MRAM, STT MRAM uses MTJs as the storage elements. One MTJ has two ferromagnetic layers and one oxide barrier layer. The resistance of MTJ depends on the relative magnetization directions of the two ferromagnetic layers, i.e., when the magnetization is parallel (or antiparallel), as illustrated in Fig. 2, MTJ is in low (or high) resistance state.

In STT MRAM, parallel and antiparallel magnetization are realized by steering a write current, which must be larger than a threshold, directly through MTJs along opposite directions. In current practice, the write current threshold is as low as few

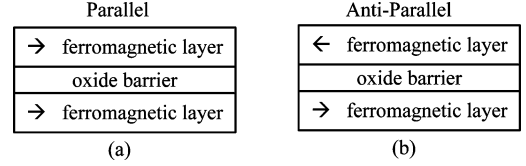


Fig. 2. Two configurations of MTJ. (a) Parallel (low resistance). (b) Anti-parallel (high resistance).

hundred microamperes [9], [10], which directly enables a significant reduction of memory write energy compared with traditional magnetoresistive memories. Let R_h and R_l denote the high and low MTJ resistance, respectively, we define tunneling magnetoresistance ratio (TMR) as $(R_h - R_l)/R_l$. A larger TMR makes it easier to distinguish the two resistance states and hence is highly desirable. As shown in [9], for an MTJ with MgO barrier layer and a dimension of 125 nm \times 220 nm, the high and low resistance are about 5.5 and 2.5 $\text{K}\Omega$, respectively, leading to a typical TMR of 155%. Since both write and read involve through-MTJ current, the sensing current has to be (much) lower than the write current threshold in order to avoid read disturb (e.g., sensing current has to be 100 μA and below).

III. PROPOSED STT MRAM STRUCTURE

In this paper, we propose an STT MRAM structure that realizes high-speed sensing by forming an MTJ-based voltage divider with the same principle, as illustrated in Fig. 1(b). Besides the significant cell size penalty, the straightforward 2-MTJ per cell structure shown in Fig. 1(b) may not be able to realize sufficiently large sensing voltage margin for STT MRAM. This can be illustrated as follows. Let V_h and V_l denote the high and low supply voltages of the 2-MTJ voltage divider. Hence, the output voltage V_o of the voltage divider can be either

$$V_o^{(h)} = V_h + (V_l - V_h)/(2 + \text{TMR})$$

or

$$V_o^{(l)} = V_l + (V_h - V_l)/(2 + \text{TMR})$$

which leads to a voltage margin of

$$V_o^{(h)} - V_o^{(l)} = \text{TMR} \cdot (V_h - V_l)/(2 + \text{TMR}).$$

Because $V_h - V_l$ cannot be too large in order to avoid read disturb and TMR is typically small, the voltage margin may not be sufficiently large. For example, let us consider MTJs whose R_h and R_l fall into the range of 5–6 and 2–3 $\text{K}\Omega$, respectively. If the through-MTJ sensing current is set to 60 μA , the worst case sensing voltage margin $V_o^{(h)} - V_o^{(l)}$ is only about 120 mV, which may result in unsatisfactory reliability and sensing speed.

In this section, we first describe the proposed basic design principle that can amplify the sensing voltage margin, then present a new 1-MTJ per cell memory structure based on this design principle.

A. Basic Design Principle

The basic principle underlying this paper is to insert a pair of transistors into the differential MTJ-based voltage divider to amplify the sensing voltage margin, as illustrated in Fig. 3. Let

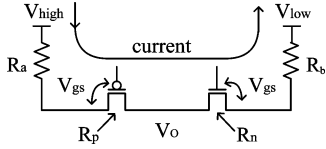


Fig. 3. Basic principle of the proposed sensing strategy.

R_n and R_p represent the ON resistance of the n-MOS (nMOS) and p-MOS (pMOS) transistors, respectively. The sensing voltage margin is determined by $(R_h + R_n)/(R_l + R_p)$.

It is well known that the current I_{ds} of an MOS transistor can be approximated as

$$I_{ds} = K \frac{W}{L} \left[(V_{gs} - V_{th})V_{ds} - \frac{V_{ds}^2}{2} \right] (1 + \lambda V_{ds}) \quad (1)$$

when MOS transistor working at triode region; or

$$I_{ds} = \frac{K}{2} \frac{W}{L} (V_{gs} - V_{th})^2 (1 + \lambda V_{ds}) \quad (2)$$

when MOS transistor working at saturation region. Here, K is the process transconductance parameter, W is the channel width, L is the effective channel length, V_{th} is the threshold voltage, and λ is the channel-length modulation parameter. Clearly, the ON resistance, which is defined as $R_{ON} = V_{ds}/I_{ds}$ is modulated by the gate-source voltage V_{gs} . Such resistance-voltage dependency can improve the sensing voltage margin, which is intuitively explained as follows. Referring to Fig. 3, if $R_a = R_l < R_b = R_h$, due to the relatively large voltage drop across R_b , V_{gs} of the nMOS transistor is relatively small and makes R_n relatively large. Meanwhile, due to the small voltage drop across R_a , V_{gs} of the pMOS transistor is relatively large and makes R_p relatively small. Hence, the ratio $(R_h + R_n)/(R_l + R_p)$ tend to be relatively large, leading to an improved sensing voltage margin. Similarly, if $R_a = R_h > R_b = R_l$, the voltage margin may also improve due to the same resistance-voltage dependency.

Since the ON resistance is modulated by the gate-source voltage V_{gs} , the gate voltage of the two transistors can greatly affect the sensing voltage margin ΔV . Consider MTJs whose R_h and R_l fall into the range of 5–6 and 2–3 $K\Omega$, respectively. Using cadence tool set with 0.18 μm CMOS technology design kit, we implemented the circuit shown in Fig. 3, where V_{high} is 1.8 V and V_{low} is 0 V. We set the gate voltages of the pMOS transistor as 0 V and obtain the dynamic behavior of voltage margin ΔV while varying the gate voltages of the nMOS transistor, as shown in Fig. 4. Table I lists the working region of the nMOS transistor when the voltage divider produces a high and low voltage output, respectively. As shown in Fig. 4, the voltage margin ΔV is maximized within Section III-C, i.e., the nMOS transistor works in the saturation region when the voltage output is high, and in the triode region when the voltage output is low, and ΔV achieves its maximum value when the nMOS gate voltage is around 1.1 V.

Fig. 5 shows the dynamic behavior of voltage margin ΔV while varying the gate voltages of the pMOS transistor and fixing the gate voltage of the nMOS transistor as 1.1 V. Similarly, in order to maximize the margin ΔV , the pMOS transistor should work in the triode and saturation region when the output

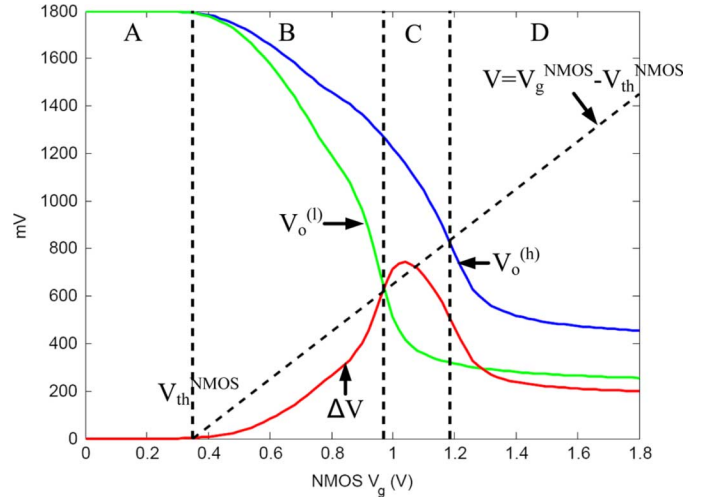


Fig. 4. Sensing margin versus nMOS gate voltages, where $V_{th}^{nMOS} = 0.35$ V. We set V_{high} as 1.8 V, V_{low} as 0 V and the gate voltage of the pMOS transistor as 0 V.

TABLE I
NMOS TRANSISTOR WORKING REGION

	A	B	C	D
High voltage output	cut-off	saturation	saturation	triode
Low voltage output	cut-off	saturation	triode	triode

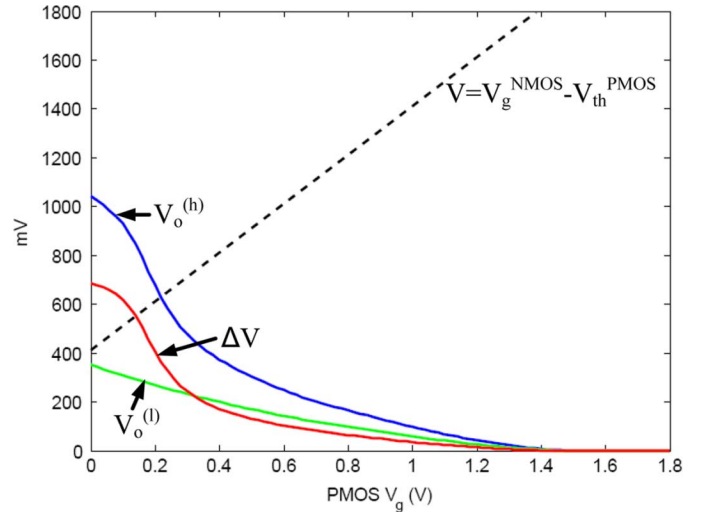


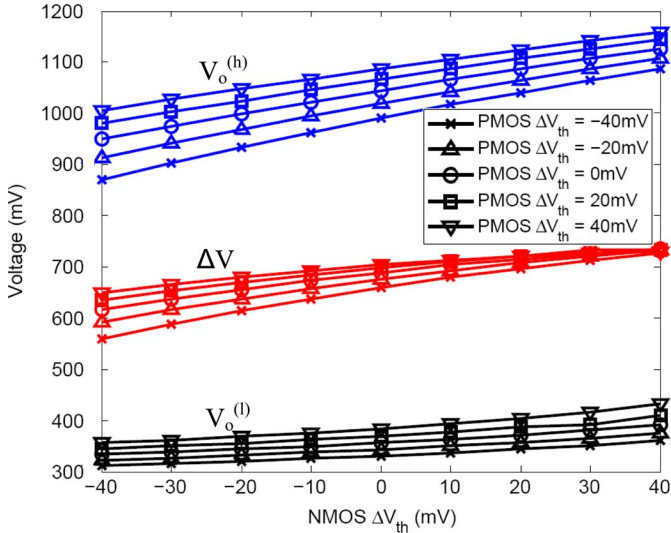
Fig. 5. Sensing margin versus pMOS gate voltages, where $V_{th}^{pMOS} = -0.41$ V. We set V_{high} as 1.8 V, V_{low} as 0 V, and the gate voltage of the nMOS transistor as 1.1 V.

voltage is high and low, respectively. Clearly, we should set the gate voltage of the pMOS transistor as 0 V. With the gate voltages of the nMOS and pMOS transistors as 1.1 and 0 V, Table II shows the worst case output voltage margin assuming that R_h and R_l fall into the range of 5–6 and 2–3 $K\Omega$, respectively. A voltage margin of over 680 mV is achieved with less than 65 μA through-MTJ sensing current. In comparison, under the same through-MTJ current, only about 130 mV voltage margin can be achieved without inserting the two transistors.

According to (1) and (2), the ON resistance of MOS transistors also depends on the threshold voltage. Hence, variation of threshold voltage will cause voltage margin degradation. To

TABLE II
 RESULTS TO DEMONSTRATE THE BASIC DESIGN PRINCIPLE

R_a (K Ω)	R_b (K Ω)	V_o (mV)	Through-MTJ current (μ A)
3	5	1044.0	62.16
5	3	355.5	63.53


 Fig. 6. Sensitivity analysis of the voltage margin ΔV in the presence of threshold voltage variation.

quantitatively evaluate its impact, Fig. 6 shows the simulation results of the output high/low voltages and voltage margin ΔV in the presence of threshold voltage variation of both the nMOS and pMOS transistors. The results show that even with a ± 40 -mV threshold voltage variation of both nMOS and pMOS transistors, the voltage margin ΔV is still over 450 mV. Meanwhile, we note that the low voltage output $V_o^{(l)}$ is less sensitive to the threshold voltage variation than the high voltage output $V_o^{(h)}$.

B. Proposed 1T1MTJ STT MRAM Structure

Straightforwardly, we may have a 2 T2MTJ per cell STT MRAM structure by directly using the circuit shown in Fig. 3 as one cell. During the sensing, the out V_o directly drives the BL, while during write the two transistors steer the programming current through the 2 MTJs independently. However, such a straightforward realization suffers from two main disadvantages. First, the cell size is relatively large. Since the programming current is typically a few hundred microamperes, the two transistors in each cell have to be accordingly sized. Meanwhile, the size of MTJs is usually much smaller than that of MOS transistors [12], [14], [15]. As a result, a 2 T2MTJ cell tends to occupy a large area. Second, since two transistors from each cell attach to BL, it tends to largely increase the BL capacitance and hence degrade the sensing speed.

To more efficiently apply the aforementioned voltage-divide-based sensing principle, we propose a 1 T1MTJ per cell BL structure, as illustrated in Fig. 7. Each cell has 1 MTJ and 1 nMOS transistor, and all the cells on each BL share a reference cell that consists of 1 pMOS transistor and one reference resistor with the resistance of $(R_h + R_l)/2$. The nMOS transistor in each cell should be sized according to the STT write current

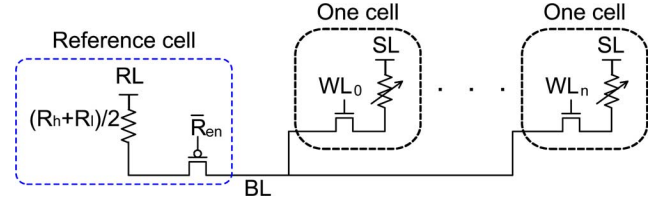


Fig. 7. Proposed 1 T1MTJ per cell BL structure.

threshold. During the read operation, the pMOS transistor in the reference cell is turned on so that the reference cell and the selected cell forms a voltage divider through the BL. According to Fig. 3, we set a high voltage to the read line (RL) connected to the reference cell and set a low voltage to the source line (SL) connected to the 1T1MTJ cells.

To demonstrate the effect of the earlier design principle of replacing 1 MTJ with the constant resistance of $(R_h + R_l)/2$, we implement the similar divider circuit shown in Fig. 3 but with a constant resistance. We assume that R_h and R_l of MTJs fall into the range of 5–6 K Ω and 2–3 K Ω , respectively, and set the reference resistance as 4 K Ω . We set V_{high} as 1.8 V and V_{low} as 0 V, and set the gate voltages of the nMOS and pMOS transistors as 1.05 and 0 V, respectively. A sensing voltage margin of over 550 mV is achieved with less than 70 μ A through-MTJ sensing current.

In order to minimize the read access time, we employ a precharge sensing strategy using the memory array structure, as shown in Fig. 8. Let $V_o^{(h)}$ and $V_o^{(l)}$ denote the BL voltage when the selected cell has a high-resistance MTJ and low-resistance MTJ, respectively. Before turning on the reference cell and selected cell to form a voltage divider, we precharge the BL to $V_{pre} = (V_o^{(h)} + V_o^{(l)})/2$. Hence, we can use the differential sensing scheme, as illustrated in Fig. 8, to improve the speed. Accordingly, the read operation consists of three phases (see Fig. 9), as follows.

- 1) Similar to SRAM and DRAM design practice, we precharge BLs to the median voltage V_{pre} .
- 2) Then, we turn on the reference cell and the selected cell to form a voltage divider along the BL. Depending on the resistance of the MTJ in the selected cell, the voltage divider will start to charge or discharge the BL.
- 3) After a prespecified sensing time, we turn on the differential sense amplifier (SA) to generate the read output by comparing the BL voltage and V_{pre} .

To further demonstrate the earlier proposed STT MRAM structure, we designed BL arrays with the number of cells per BL ranging from 64 up to 256, using Cadence tool set and 0.18 μ m CMOS technology with the same resistance assumption as earlier. The nMOS transistor in each cell is sized to 1 μ m/0.18 μ m in order to allow 400 μ A programming current under 1.8 V of supply voltage. During read operation, we set the gate voltages of the nMOS and pMOS transistors as 1.05 and 0 V, respectively. Correspondingly, the through-MTJ sensing current is below 70 μ A. We designed BLs with the associated number of cells as 64, 128, and 256, respectively. For the purpose of simplicity, we fix the precharge time as 1 ns. The SAs are turned on when $\Delta V \approx 100$ mV and take

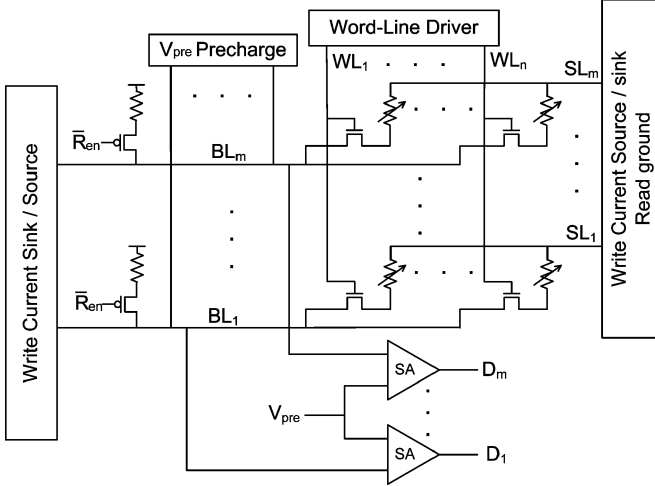


Fig. 8. Proposed 1 T1MTJ MRAM array architecture.

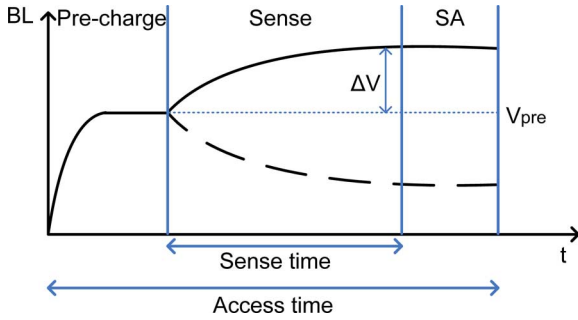


Fig. 9. Illustration of the three-phase read operation.

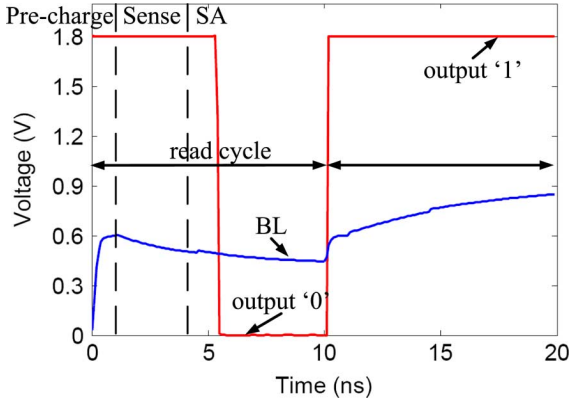


Fig. 10. Simulated read operation waveform.

about 2 ns to deliver the output. Fig. 10 shows the postlayout simulated waveform for the case of 128-cell BL. Clearly, the worst case access delay occurs when the high resistance of MTJ is 5 K Ω or the low resistance of MTJ is 3 K Ω . Table III lists the postlayout simulated worst case access delay for BLs with different number of cells. Table IV compares this paper with prior research, where two types of sensing strategies are used, including voltage-mode sensing and current-mode sensing.

IV. PROPOSED CAM AND TCAM STRUCTURES

In this section, we apply the earlier same basic design principle to develop STT CAM and TCAM cell structures with high

TABLE III
SIMULATION RESULTS

Numbers of cells in BL	64	128	256
Sense time (ns)	2.2	3.0	4.5
Access time (ns)	5.2	6.0	7.5

searching voltage margin and low silicon area. CAM compares input search word against its own content in parallel and returns the address of the stored word that matches (i.e., equals to) the input search word [17]. Each stored word associates with an individual match line (ML) that reports the comparison result (i.e., match or mismatch). As illustrated in Fig. 11(a), CAM typically uses SRAM cell for bit storage. For the purpose of simplification, the figure does not include the access transistors and BLs that are used to read and write the SRAM cell. After the ML has been charged to a high voltage, the search bit is loaded onto the global differential search line $\{SL, \overline{SL}\}$ and compared with the CAM cells in parallel. A mismatch between the search bit and the stored bit will turn on one pulldown path and hence discharge the ML, while a match will turn off both pulldown paths. Fig. 11(b) presents the most straightforward MTJ-based non-volatile CAM cell structure consisting of two differential MTJs. Nevertheless, as discussed earlier, such straightforward realization of CAM cell may not be able to ensure sufficiently large output voltage margin at V_o to turn on/off the pulldown path. TCAM is a special type of CAM, in which each cell can store three values, including logic 0, logic 1, and do not care (denoted as X). During a search operation, a cell that stores X always reports a match regardless to whether the input search bit is logic 0 or 1. Typically, two CAM cells are used to implement one TCAM cell. In this section, we present approaches that apply the aforementioned basic design principle to implement reliable and high-speed CAM and TCAM cells.

A. Proposed CAM Cell Design

Applying the same basic design principle illustrated in Fig. 3, we have two different CAM cell structures, as shown in Fig. 12(a) and (b), where each cell contains two differentially configured MTJs. These two designs largely differ at how the search operation is realized and the search energy consumption. The first CAM cell structure shown in Fig. 12(a) is relatively straightforward. Two pairs of nMOS and pMOS transistors are inserted between the two differential MTJs. The gates of each pair of nMOS and pMOS transistors are driven by differential gate line $\{GL, \overline{GL}\}$. During the search operation, the differential gate line always turns off one pMOS transistor in one pair, and one nMOS transistor in another pair; hence, the cell circuit is always equivalent to the one shown in Fig. 3. The voltage setup of the $\{BL, \overline{BL}\}$ and $\{GL, \overline{GL}\}$ is determined by the corresponding input search bit so that V_o , the output of the cell voltage divider, is relatively high and can turn on the nMOS transistor on the pulldown path if the stored datum does not equal to the input search bit. However, in spite of its simple structure, this CAM cell structure may incur a significant amount of search energy consumption, as intuitively explained as follows. During the search operation, all the $\{BL, \overline{BL}\}$ and $\{GL, \overline{GL}\}$ switch according to input search word and all the CAM cells turn on their internal voltage divider to establish the

TABLE IV
COMPARISON WITH PRIOR WORK

	Voltage-Mode Sensing			Current-Mode Sensing	
	this work	[13]	[8]	[5]	[16]
MTJ TMR	67%	20%	150%	35%	37%
MTJ R_{low}	3K Ω	8K Ω	5K Ω	12K Ω	N/A
Technology	0.18 μm	0.18 μm	0.18 μm	0.18 μm	0.18 μm
Power supply	1.8V	1.8V	1.2V	1.8V	2.0V
Cell structure	1T1MTJ	1T2MTJ	5T2MTJ	1T1MTJ	1T1MTJ
Through-MTJ sensing current	70 μA	36 μA (est.)	96 μA (est.)	N/A	N/A
Sensing margin	550mV	30mV	N/A	-	-
Cell number per bit-line	256	N/A	32	256	N/A
Access time	7.5ns	N/A	0.85ns	5ns	3.5ns

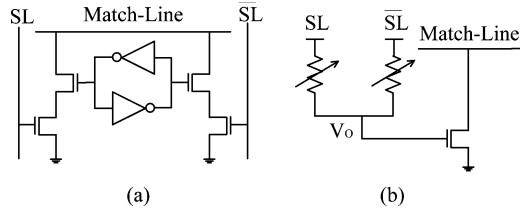


Fig. 11. Structure of (a) a typical SRAM-based CAM cell and (b) a straight-forward MTJ-based CAM cell. For the purpose of simplicity, we omit the transistors and control lines for writing/reading the cell in both (a) and (b).

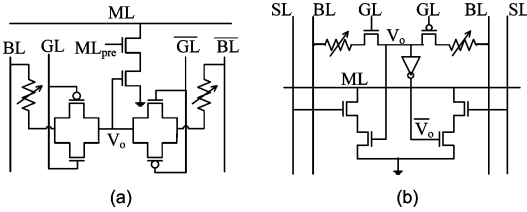


Fig. 12. Two different CAM cell structures.

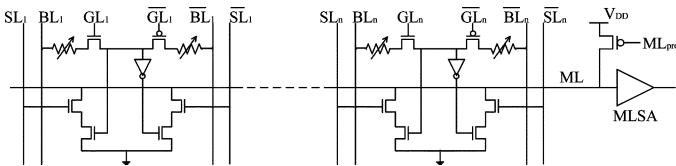


Fig. 13. Match line with n cells. For the purpose of simplicity, we omit the transistors and control lines for writing/reading each CAM cell.

voltage V_o . Due to its current driven nature, each voltage divider consumes static energy during the entire search operation. As we discussed earlier, to establish a large enough voltage margin at V_o , each voltage divider has tens of microamperes static through-MTJ current. This will result in a very significant energy consumption for a reasonably large CAM array.

The CAM cell structure illustrated in Fig. 12(b) can greatly reduce the energy consumption at cost of more transistors per cell. In this design, irrelevant to the input search word, the voltage setup of $\{BL, \overline{BL}\}$ and $\{GL, \overline{GL}\}$ is fixed and set the corresponding voltage divider equivalent to the one shown in Fig. 3. The four nMOS transistors on the two pulldown paths realize XOR operation between the the differential $SL \{\overline{SL}, SL\}$, which is driven by the input search bit, and the voltage divider differential output $\{V_o, \overline{V_o}\}$. If the datum stored in the cell does not equal to the input search bit, one of the two pulldown paths will turn on and discharge the ML.

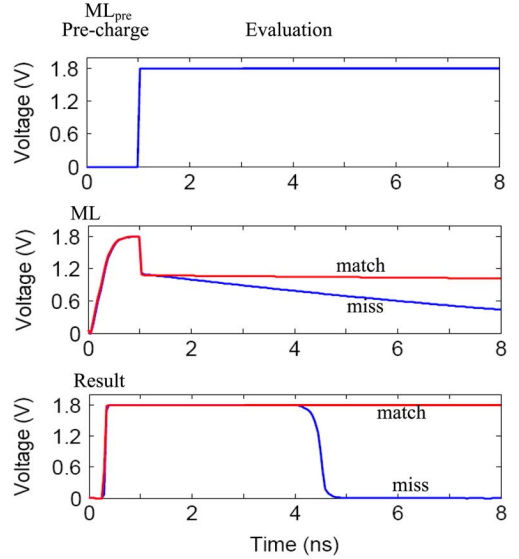


Fig. 14. Simulated worst case CAM search operation waveform.

Using the aforementioned second cell structure shown in Fig. 12(b), we may rely on parasitic capacitor at the voltage divider output to hold V_o and enable the XOR operation with input search bit, while turning off the voltage divider by disabling the differential $GL \{\overline{GL}\}$ to reduce static energy consumption. To compensate the charge leakage, we can periodically turn on the voltage divider within each cell to refresh the output V_o . Clearly, this design may largely reduce the static energy consumed by the voltage dividers within CAM cells, particularly compared with the cell structure shown in Fig. 12(a). Moreover, different from DRAM, the refresh operation in this context is completely independent of and will not interfere with the normal search operation, which can simplify the peripheral control circuitry. Since the high output of the voltage divider $V_o^{(h)}$ may be well below the supply voltage V_{dd} , the inverter that is used to generate $\overline{V_o}$ may have nonnegligible static current if it is also supplied with V_{dd} . To eliminate such static current, we may use a lower voltage $V_{dd}^r < V_{dd}$ as power supply for the inverter.

Fig. 13 shows an ML structure when using the second CAM cell structure presented earlier. The search operation has two phases: first all the MLs are precharged to the supply voltage by setting ML_{pre} to 0 while SL and \overline{SL} are held low. Then, the ML precharge is released by switching ML_{pre} to 1 and each input search bit is loaded onto the differential $\{SL, \overline{SL}\}$ pair.

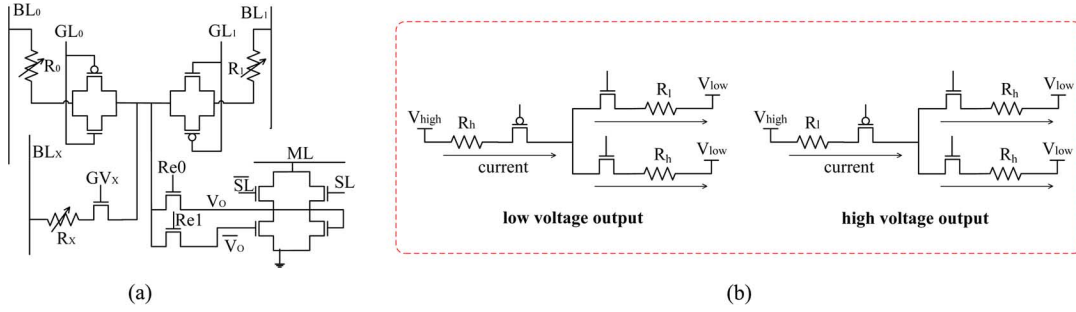


Fig. 15. (a) Proposed TCAM cell structure. (b) Simplified equivalent circuits during refresh operation.

TABLE V
KEY CAM DESIGN METRICS

Technology	0.18 μm
Power supply	1.8V
Cell Area	36 μm^2
Cell Number per Matchline	144
Delay	5ns
Search Energy	7.1fJ/bit/search
Refresh Energy	94.9fJ/bit

TABLE VI
TCAM CELL MTJ CONFIGURATION

Cell value	R_0	R_1	R_x
Logic 1	R_l	R_h	R_h
Logic 0	R_h	R_l	R_h
Don't care X	R_h	R_h	R_l

Only when the entire word matches the input search word, its ML can remain a high voltage, otherwise at least one cell will turn on the pulldown path to discharge the ML. An ML sense amplifier (MLSA) can be used to further speed up the search operation.

To demonstrate the earlier presented low-power CAM structure, we designed an ML with 144-bit CAM cells at 0.18 μm CMOS technology. Similar to the design example presented in Section III, we assume that R_h and R_l of MTJs fall into the range of 5–6 and 2–3 $\text{K}\Omega$. We fix BL as 1.6 V and $\overline{\text{BL}}$ as -0.2 V, and fix GL and $\overline{\text{GL}}$ as 0.99 and 0 V, respectively. The $V_o^{(h)}$ is above 790 mV and the $V_o^{(l)}$ is below 200 mV, leading to a search noise margin of 590 mV. Fig. 14 shows the simulated worst case search operation waveform, i.e., there is only 1-bit mismatch and the differential MTJs in the mismatch cell have the resistance of 3 and 5 $\text{K}\Omega$. Results suggest that the worst case search delay of 5 ns can be achieved, leading to over 200 MHz search frequency. The worst case sensing current through one MTJ is about 55 μA . The energy consumption in worst case is 7.1 fJ/bit per search, assuming that all of 144 bits are missed. The parasitic capacitance at the output of voltage divider is about 3.4 fF, which is sufficient to hold V_o on the order of μs . Hence, given the search operation frequency of 200 MHz, we only need to refresh the CAM cells every 200 search cycles. The energy consumption for one refresh operation is 94.9 fJ/bit, which means only 0.58 fJ/bit for each search cycle. The key metrics are summarized in Table V.

It should be pointed out that as we discussed in Section III-A, $V_o^{(h)}$ and $V_o^{(l)}$ are sensitive to the threshold voltages of the nMOS and pMOS transistor. Therefore, this design tends to make $V_o^{(h)}$ and $V_o^{(l)}$ far away the threshold voltage of the nMOS transistors in the pulldown paths. With a ± 30 -mV threshold voltage variation of both the nMOS and pMOS transistor in the CAM cells, $V_o^{(l)}$ is 100 mV below and $V_o^{(h)}$ is 250 mV above the threshold voltage of the nMOS transistors in the pulldown path.

B. Proposed TCAM Cell Design

It may be intuitive that we can directly use the aforementioned presented CAM cell structure to realize TCAM, where do not care, X that can be represented by configuring the 2 MTJs into the same resistance state. Recall that the operation of the MTJ-based CAM cell depends on the resistance difference between the 2 MTJs. Even though the two MTJs are in the same resistance state, their resistance may still have significant difference due to the process variation, e.g., as assumed in the earlier design examples, the high resistance may vary between 5 and 6 $\text{K}\Omega$ and the low resistance may vary between 2 and 3 $\text{K}\Omega$. Such resistance variation will largely degrade the search noise margin and even result in malfunction; hence, we cannot build a TCAM cell using only one CAM cell.

Instead of using two CAM cells to build a TCAM cell, we develop an area-efficient TCAM cell structure that contains three MTJs, as illustrated in Fig. 15(a). The three MTJs (i.e., R_0 , R_1 , and R_x) are configured based on one-cold coding scheme that configures only 1 MTJ to low resistance state, as shown in Table VI.

Similar to the earlier CAM structure, we rely on parasitic capacitor at the voltage divider output to hold V_o and $\overline{V_o}$ and enable the XOR operation with input search bit, while turning off the voltage divider to reduce static energy consumption. The voltage output V_o and $\overline{V_o}$ are periodically refreshed to compensate the charge leakage. During the refresh operation, the BL BL_x always has a low voltage. To refresh $V_o(\overline{V_o})$, the RL $\text{Re}_0(\text{Re}_1)$ is set to V_{dd} to turn on the nMOS transistor, while the differential bit lines BLs BL_0 and BL_1 are set to $V_{\text{high}}(V_{\text{low}})$ and $V_{\text{low}}(V_{\text{high}})$, respectively. Fig. 15(b) shows the simplified equivalent circuit when the voltage output of the TCAM cell is low or high.

With the same assumption on the range of R_h and R_l as in the earlier CAM design example, we designed a math-line with 144 TCAM cells. We set V_{high} and V_{low} as 1.8 and 0 V, respectively, and set the gate voltages of the nMOS and pMOS transistors as 1.15 and 0 V, respectively. Circuit simulation results show that low voltage output of V_o in each TCAM cell is below 170 mV

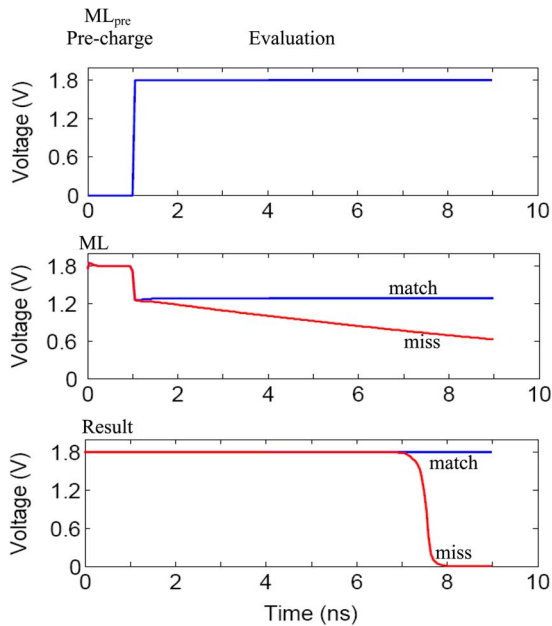


Fig. 16. Simulated worst case TCAM search operation waveform.

TABLE VII
KEY TCAM DESIGN METRICS

Technology	0.18 μm
Power supply	1.8V
Cell Area	42 μm^2
Cell Number per Matchline	144
Delay	8ns
Search Energy	7.4fJ/bit/search
Refresh Energy	152.8fJ/bit

and the high voltage is above 720 mV, leading to a search noise margin of 550 mV. Under a ± 30 -mV threshold voltage variation of both nMOS and pMOS transistor in TCAM cells, the $V_o^{(l)}$ is 80 mV below the threshold voltage of the nMOS transistors and the $V_o^{(h)}$ is 200 mV above the threshold voltage. The worst case sensing current through 1 MTJ is about 87 μA . Fig. 16 shows the simulated worst case TCAM search waveform, which suggests a less than 8 ns worst case search delay. The energy consumption in the worst case is 7.4 fJ/bit per search, assuming that all of 144 bits are missed. The energy consumption for one refresh operation is 152.8 fJ/bit. Given the search operation frequency of 100 MHz and conservative estimation of the parasitic capacitance and charge leakage, we need to refresh the CAM cells every 100 search cycles, leading to 1.5 fJ/bit for one search cycle. Table VII summarizes the key design metrics.

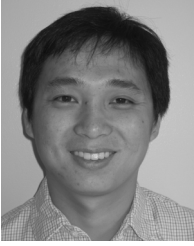
V. CONCLUSION

This paper presents new STT RAM and CAM/TCAM circuit design solutions that can achieve high sensing/search speed and reliability while maintaining low through-MTJ current in the presence of relatively small TMR. The basic design principle is to complement MTJs with transistors to form voltage dividers with enhanced voltage margin. We first apply this design principle to develop a 1T1MTJ per cell STT RAM structure, then further extend it to design energy-efficient and high-speed CAM

and TCAM structures. Circuit design and simulation at 0.18 μm CMOS technology node have been carried out to demonstrate the effectiveness of the proposed STT RAM, CAM, and TCAM structures.

REFERENCES

- [1] C. Hwang, "Nanotechnology enables a new memory growth model," *Proc. IEEE*, vol. 91, no. 11, pp. 1765–1771, Nov. 2003.
- [2] K. Kim and G. Jeong, "Memory technologies for sub-40 nm node," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, Dec. 2007, pp. 27–30.
- [3] H. Pon, "Technology scaling impact on NOR and NAND flash memories and their applications," in *Proc. Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Oct. 2006, pp. 697–700.
- [4] Y. Shin, "Non-volatile memory technologies for beyond 2010," in *Proc. Symp. VLSI Circuits*, Jun. 2005, pp. 156–159.
- [5] J. Debrosse, D. Gogl, A. Bette, H. Hoenigschmid, R. Robertazzi, C. Arndt, D. Braun, D. Casarotto, R. Havreduk, S. Lammers, W. Obermaier, W. R. Reohr, H. Viehmann, W. J. Gallagher, and G. Muller, "A high-speed 128-kb MRAM core for future universal memory applications," *IEEE J. Solid-State Circuits*, vol. 39, no. 4, pp. 678–683, Apr. 2004.
- [6] T. W. Andre, J. J. Nahas, C. K. Subramanian, B. J. Garni, H. S. Lin, A. Omair, and W. L. Martino, "A 4-Mb 0.18- μm 1T1MTJ toggle MRAM with balanced three input sensing scheme and locally mirrored unidirectional write drivers," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 301–309, Jan. 2005.
- [7] T. Sugibayashi, N. Sakimura, T. Honda, K. Nagahara, K. Tsuji, H. Numata, S. Miura, K. Shimura, Y. Kato, S. Saito, Y. Fukumoto, H. Honjo, T. Suzuki, K. Suemitsu, T. Mukai, K. Mori, R. Nebashi, S. Fukami, N. Ohshima, H. Hada, N. Ishiwata, N. Kasai, and S. Tahara, "A 16-Mb toggle MRAM with burst modes," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2378–2385, Nov. 2007.
- [8] N. Sakimura, T. Sugibayashi, T. Honda, H. Honjo, S. Saito, T. Suzuki, N. Ishiwata, and S. Tahara, "MRAM cell technology for over 500-MHz SoC," *IEEE J. Solid-State Circuits*, vol. 42, no. 4, pp. 830–838, Apr. 2007.
- [9] Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen, L. Wang, and Y. Huai, "Spin-transfer torque switching in magnetic tunnel junction and spin-transfer torque random access memory," *J. Phys.: Condens. Matter*, vol. 19, p. 165209, Apr. 2007.
- [10] T. Kawahara, R. Takemura, K. Miura, J. Hayakawa, S. Ikeda, Y. M. Lee, R. Sasaki, Y. Goto, K. Ito, T. Meguro, F. Matsukura, H. Takahashi, H. Matsuoka, and H. Ohno, "2 Mb spin-transfer torque ram (SPRAM) with bit-by-bit bidirectional current write and parallelizing-direction current read," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2007, pp. 480–481.
- [11] X. Wang, Y. Zheng, H. Xi, and D. Dimitrov, "Thermal fluctuation effects on spin torque induced switching: Mean and variations," *J. Appl. Phys.*, vol. 103, p. 034507, 2008.
- [12] Y. Chen, X. Wang, H. Li, H. Liu, and D. Dimitrov, "Design margin exploration of spin-torque transfer RAM (SPRAM)," in *Proc. IEEE Int. Symp. Quality Electron. Design (ISQED)*, 2008, pp. 684–690.
- [13] M. Aoki, H. Iwasa, and Y. Sato, "A novel voltage sensing 1T/2MTJ cell with resistance ratio for highly stable and scalable MRAM," in *Proc. Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2005, pp. 170–171.
- [14] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane, H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, and H. Kano, "A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-RAM," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, 2005, pp. 459–462.
- [15] H. Tanizaki, T. Tsuji, J. Otani, Y. Yamaguchi, Y. Murai, H. Furuta, S. Ueno, T. Oishi, M. Hayashikoshi, and H. Hidaka, "A high-density and high-speed 1T-4MTJ MRAM with voltage offset self-reference sensing scheme," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2006, pp. 3.3–3.6.
- [16] E. K. S. Au, W. Ki, W. H. Mow, S. T. Hung, and C. Y. Wong, "A novel current-mode sensing scheme for magnetic tunnel junction MRAM," *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 483–488, Mar. 2004.
- [17] K. Pagiamtzis, "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey," *IEEE J. Solid-State Circuits*, vol. 41, no. 3, pp. 712–727, Mar. 2006.



Wei Xu received the B.S. and M.S. degrees from Fudan University, Shanghai, China, in 2003 and 2006, respectively. He is currently working toward the Ph.D. degree at the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY.

His current research interests include circuit and system design for data storage systems. He is engaged in magnetoresistive memory system design.



Tong Zhang (M'02–SM'08) received the B.S. and M.S. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2002.

He is currently an Associate Professor with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. His current research interests include algorithm and architecture codesign for communication and data storage systems, variation-tolerant signal processing IC design, fault-tolerant system design for digital memory, and computer architecture.



Yiran Chen (M'05) received the B.S. and M.S. degrees in electronics engineering from Tsinghua University, Beijing, China, and the Ph.D. degree from Electrical and Computer Engineering Department, Purdue University, West Lafayette, IN.

He was with PrimeTime Group of Synopsys, Inc., Sunnyvale, CA, where he developed the award-winning statistical static timing analysis tool "PrimeTimeVX". In 2007, he joined the alternative technology group of Seagate Technology LLC, Bloomington, MN, where he was engaged in the next generation nonvolatile memory. His current research interests include very large scale integration (VLSI) design/computer-aided design (CAD) for nanoscale Silicon and nonsilicon technologies, low-power circuit design and computer architecture, and emerging memory technologies. He has authored or coauthored more than 25 technical papers in refereed journals and conferences. He has more than 50 pending U.S. patents.