

A High-Speed Robust NVM-TCAM Design Using Body Bias Feedback

Bonan Yan[†], Zheng Li[†], Yaojun Zhang[†], Jianlei Yang[†], Weisheng Zhao^{*},
Pierre Chor-Fung Chia[‡], Hai Li[†]

[†]University of Pittsburgh, Pittsburgh, Pennsylvania USA

^{*}Beihang University, Beijing, China

[‡]Cisco Systems, Inc., San Jose, California

{boy12, zhl85, yaz24, jiy64, hal66}@pitt.edu, weisheng.zhao@u-psud.fr,
chchia@cisco.com

ABSTRACT

As manufacture process scales down rapidly, the design of *ternary content-addressable memory* (TCAM) requiring high storage density, fast access speed and low power consumption becomes very challenging. In recent years, many novel TCAM designs have been inspired by the research on emerging nonvolatile memory technologies, such as *magnetic tunneling junction* (MTJ), *phase change memory* (PCM), and memristor. These designs store a data as the resistive variable of a nonvolatile device, which usually results in limited sensing margin and therefore constrains the searching speed of TCAM architecture severely. To further enhance the performance and robustness of TCAMs, we proposed two novel cell designs that utilize MTJs as data storage units—the symmetrical dual-N structure and the asymmetrical P-N scheme. In both designs, a body bias feedback circuit is integrated to enlarge the sensing margins. Compared with an existing MTJ-based TCAM structure, the tolerance in gate voltage variation of the symmetrical dual-N (asymmetrical P-N) scheme can significantly improve 59.5% (21.2%). The latency and the dynamic energy consumption in one searching operation at the word length of 256 bits are merely 590.35ps (97.89ps) and 65.05fJ/bit (36.85fJ/bit), not even mentioning that the use of nonvolatile MTJ devices avoids unnecessary leakage power consumption.

Categories and Subject Descriptors

B.7.1 [Integrated Circuits]: Types and Design Styles—
Memory technologies

General Terms

Design, Performance, Reliability

Keywords

Ternary content-addressable memory (TCAM), nonvolatile memory (NVM), body bias feedback

1. INTRODUCTION

Ternary content-addressable memory (TCAM) compares an input searching data against a table of stored data and returns the address of the matching one(s) [1]. As process technology scales down, the design of conventional SRAM-based TCAMs faces severe difficulties due to the relatively low cell density, large leakage power consumption and deteriorated reliability. Although data is rarely updated in TCAMs, maintaining them in SRAM cells results in huge leakage power consumption which starts dominating the overall energy of TCAM architecture. Moreover, the scaling of SRAM designs has been much slower than that of fabrication processes due to the deteriorating device reliability. Furthermore, the searching speed of SRAM-based TCAMs approaches the design limitation [2].

To increase data storage density, reduce power consumption and improve searching speed, TCAM designs based on emerging *nonvolatile memories* (NVMs) have been extensively studied. The nonvolatile storage of these technologies makes significant reduction in static power consumption [3][4]. The fast access speed and good scalability potentially help improve the TCAM density and enhance the searching speed as well. Among various NVMs, *magnetic tunneling junction* (MTJ) could be one of the best candidates for TCAM design considering its technology readiness and commercialization status [5].

Some examples of the latest MTJ-based TCAM designs include 9-transistors-6-MTJs (9T-6MTJ) [6], 3T-2MTJ [7] and 20T-4MTJ [8] structures. In these designs, a logic bit is represented by the *high or low resistance states* (HRS or LRS) of one or a few MTJ devices and the searching operation is realized through detecting the resistance value. Compared to SRAM-based TCAM with 12 transistors (12T-SRAM) [9], these designs significantly decrease the number of transistors and therefore reduce the cell area [6][7][10]. Particularly, Xu *et al.* proposed a structure which utilizes the voltage-dividing, instead of detecting the exact MTJ resistance value in reading and searching operations [7]. Such a structure effectively reduces the transistor numbers of a single TCAM cell to 3.

However, all these MTJ-based TCAM designs face a severe design challenge—the small difference of MTJ's high and low resistance values (usually only a few K Ω) produces a very limited sensing margin, which significantly degrades the reliability of TCAM design, prolongs the searching latency and induces extra power overhead. The benefits from the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
GLSVLSI'15, May 20–22, 2015, Pittsburgh, PA, USA.
Copyright © 2015 ACM 978-1-4503-3474-7/15/05 ...\$15.00.
<http://dx.doi.org/10.1145/2742060.2742077>.

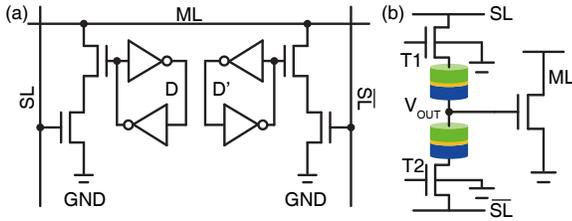


Figure 1: (a) SRAM-based TCAM design; (b) A MTJ-based TCAM cell structure [7].

nonvolatile device utilization will be also amortized. Very recently, Onizawa *et al.* proposed a 20T-4MTJ TCAM structure with very fast searching performance [8]. However, its cell area and leakage power increase dramatically, even compared with conventional 12T-SRAM version [9].

To conquer the small sensing margin issue and improve the sensing speed and design robustness of MTJ-based TCAMs, we propose two new cell structures, namely, *symmetrical dual-N* and *asymmetrical P-N* designs. Both schemes utilize the voltage dividing scheme [7] for data reading and searching. Besides, a body biasing feedback circuit is embedded. It forms a positive feed mechanism to adaptively adjust the effective resistances of select transistors and therefore improve the sensing margin. We evaluated our designs at 45nm CMOS technology [11] and the MTJ model based on the implementation of 40nm perpendicular anisotropy structure [12]. The simulations demonstrated significant improvement in sensing speed and energy consumption. For the word length of 256 bits, the symmetrical dual-N scheme can obtain a searching latency of 590.35ps and the dynamic energy consumption of 65.05fJ/bit, under the worst-case scenario. The asymmetrical P-N design has even larger sensing margin, further reducing the searching latency to 97.89ps and the dynamic energy consumption to 65.05fJ/bit.

2. PRELIMINARY

2.1 TCAM

A TCAM cell allows three types of data values: logic 1, logic 0, and X representing a *don't care* value—the state that always output *match* no matter what the searched bit is. In conventional SRAM-based TCAM design, the three different data values are represented by two SRAM cells, as shown in Figure 1(a). The *match line* (ML) is connected to a sense amplifier in reading and searching operations, while a pair of *source lines* (SL and \overline{SL}) are used to supply the programming or searching data. Such an SRAM-based TCAM cell requires 12 transistors to achieve a satisfactory reliability and performance, resulting in high leakage power as well as a large cell area.

Figure 1(b) illustrates the fundamental storage principle of an NVM-based TCAM cell [7]. Two *magnetic tunnel junction* (MTJ) devices are used to represent the three possible logic values. The matching function is realized by detecting the voltage at the internal node (V_{OUT}) which is determined by the resistances of the two MTJs as well as transistors T1 and T2. Here, the voltage margin ΔV is defined as

$$\Delta V = |V_{OUT}^H - V_{OUT}^L|, \quad (1)$$

where V_{OUT}^H and V_{OUT}^L respectively denote the lowest bound of *match* condition and the highest possible value in *miss* condition.

The two select transistors (T1 and T2) are necessary to control data access. However, their effective drain-to-source

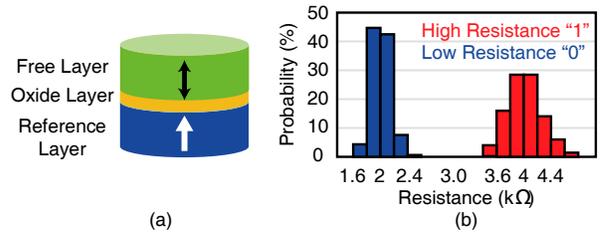


Figure 2: (a) An illustration of MTJ; (b) the MTJ resistance distribution.

resistances squeeze the voltage margin ΔV to a limited range, which severely constrains the searching speed and therefore the applications of the design. To address the issue, we propose to utilize body-bias feedback circuit which can enlarge the sensing margin by tuning the effective resistances of the two select transistors properly.

2.2 Basics of MTJ

The structure of a *magnetic tunnel junction* (MTJ) is shown in Figure 2(a). It composes of two ferromagnetic layers (namely the *reference layer* and the *free layer*) and a tunneling oxide layer (e.g., MgO) between the two ferromagnetic layers.

A MTJ can be programmed into two resistance states. When injecting a current from the free layer to the reference layer, the magnetization orientation of the free layer can be switched to be parallel to that of the reference layer. In this case, the MTJ demonstrates a lower resistance value (R_{low}), representing logic “0”. Otherwise, the anti-parallel orientation of the two ferromagnetic layers result in a higher resistance (R_{high}), denoting logic “1”. The difference between two resistance states is denoted by *tunnel magnetoresistance ratio* (TMR) as $TMR = (R_{high} - R_{low})/R_{low}$.

The logic state stored in a MTJ device can be detected by sensing out the its resistance value R_{MTJ} , which is related to the thickness of the tunneling oxide and the surface area of the device. Thus, process variations can significant affect the MTJ resistance. For example, Figure 2(b) shows the distributions of the high and low resistance values of MTJs in a 64×64 array [13]. The means of the high and low resistance states are $R_{high} = 4K\Omega$ and $R_{low} = 2K\Omega$, respectively. As shown in the figure, the gap between the two resistance states reduces to merely 1KΩ after including the impact of the process variations, making fast read/searching operations very difficult [12].

3. THE BODY BIAS FEEDBACK SCHEME

3.1 Design Principle and Mechanism

To address the small sensing margin and slow searching speed of MTJ-based TCAMs, we propose two new cell structures based on the same basic circuit scheme depicted in Figure 3(a). Same as the 3T-2MTJ TCAM in [7], our designs also use a pair of MTJs (b1 and b2) to represent the three possible logic values. Each MTJ is associated with a select transistor for access control. Instead of directly detecting the exact resistances of MTJs, the proposed design adopts the voltage-dividing property of two serially-connected MTJs as the searching principle.

More importantly, we propose to integrate a body-biasing feedback circuit to enhance the sensing margin. In the work, we denote V_{OUT} as the gate control of the discharging tran-

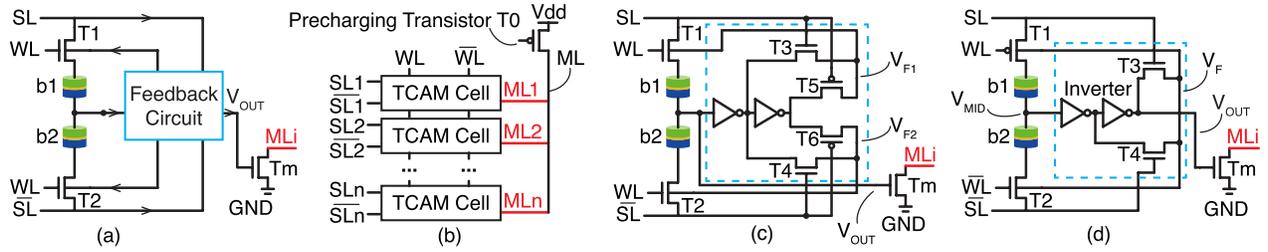


Figure 3: The proposed TCAM designs: (a) The fundamental design, where the arrows indicates signal flow. V_{out} is the gate voltage of T_m . (b) The NVM-based TCAM architecture, in which each word consists of n bits. The match line (ML) is precharged to V_{dd} through T_0 . (c) The symmetrical Dual-N TCAM cell. V_{F1} and V_{F2} are the feedback signals that respectively control the body bias of T_1 and T_2 . (d) The asymmetrical P-N TCAM cell. A single feedback signal V_F is used to control the body bias of the two select transistors.

sistor T_m . As aforementioned in Section 2.1, the effective drain-to-source resistances of T_1 and T_2 involve into the voltage dividing function, greatly constraining the ranges of V_{OUT}^H and V_{OUT}^L . Determined by the voltage level of V_{OUT} , the feedback circuit in our designs can adaptively adjust the body controls and hence tune the resistances of T_1 and T_2 , further pulling V_{OUT} toward the expected condition. As such, the voltage margin ΔV increases.

Table 1 summarizes the truth table of the proposed TCAM designs. Two MTJs b_1 and b_2 together represent the stored data. For example, to save logic “1”, b_1 and b_2 shall be programmed to the high and the low resistance states, respectively, denoting as $(b_1, b_2) = (1, 0)$. And the searching output V_{OUT} is determined by the combination of b_1 , b_2 , SL and \overline{SL} . To continue the example, if SL and \overline{SL} are respectively set to V_{dd} (“1”) and gnd (“0”), V_{OUT} will end at a relatively low voltage level which is not sufficient to turn on the discharging transistor T_m , implying a *match*. On the contrary, a relatively high V_{OUT} can be generated when SL and \overline{SL} are set as “0” and “1”, respectively. T_m is turned on to discharge ML , resulting in a *miss*. The *don't care* value in a TCAM cell can be obtained by programming both b_1 and b_2 to high resistance state. In this case, no matter what combination of SL and \overline{SL} , V_{OUT} remains at a relatively low level, corresponding to the *match* condition.

The key of the design is to keep the low level of V_{OUT} below the threshold voltage V_{th} of T_m while making sure that its high level beyond V_{th} [14]. However, V_{OUT} is determined by the voltage dividing of the upper and lower branches so it might not be able to reach the ideal V_{dd} (gnd) as its high (low) level. For convenience and accuracy, we will use “semi-high” and “semi-low” to indicate the level of V_{OUT} from henceforth.

Figure 3(b) illustrates the TCAM architecture of the proposed design. Prior a searching operation, the *match line* (ML) shall be precharged to V_{dd} . During the following searching phase, the *wordlines* (WLs) of the selected cells are raised to high to turn on the select transistors. For any bit i , the searching data and its complementary are supplied to

SL_i and \overline{SL}_i , respectively. In the case that the searching bit does not match to the stored bit (*i.e.*, *miss*), $V_{OUT,i}$ switches to a high level to turn on T_{mi} and discharge ML . Even there is only one *miss* condition among the n -bits of a word, ML that is connected to all the n TCAM cells cannot maintain at the high voltage level. In other words, a positive matching signal can be generated only when all the bits produce *match* results and the discharging transistors in all the cells are disabled.

3.2 Symmetrical Dual-N Scheme

We first present the symmetrical dual-N scheme which is shown in Figure 3(c). Here, V_{OUT} is connected to the input of the feedback circuit.

Let's use the case of $(b_1, b_2) = (0, 1)$ to explain how the feedback circuit works in the scheme. If the searching inputs $(SL, \overline{SL}) = (1, 0)$, V_{OUT} is semi-high. Transistors T_3 and T_6 are turned on while T_4 and T_5 remain off. In the situation, V_{F2} as the body bias of T_2 becomes much larger than V_{F1} , the body bias of T_1 . As such, the threshold voltage and hence the resistance of T_2 grows faster than those of T_1 . This scenario further enlarges the resistance difference of $T_1 + b_1$ and $T_2 + b_2$, boosting up V_{OUT} to high. The opposite searching inputs, on the contrary, will turn on T_4 and T_5 and greatly enlarge the resistance of $T_1 + b_1$. As a result, V_{OUT} is further pulled down. In summary, a positive feedback loop is formed, which can adaptively tune the resistances of T_1 and T_2 according to the logic of (b_1, b_2) and input searching data (SL, \overline{SL}) . The design aggravates the voltage difference between match and miss conditions, that is, the sensing margin ΔV .

There are totally 11 transistors required in a symmetrical dual-N TCAM cell. Though the feedback circuit can help enhance the sensing margin and improve the robustness of the TCAM design, the access control by NMOS type transistors T_1 and T_2 always cause threshold voltage drop at the circumstance of passing V_{dd} from either SL or \overline{SL} . The corresponding simulation results can be found in Section 4.

3.3 Asymmetrical P-N Scheme

We propose the asymmetrical P-N TCAM design to moderate the voltage drop problem in the symmetrical dual-N

Table 1: Truth Table of NVM-based TCAM Design

Stored Data		SL	\overline{SL}	V_{OUT}	Condition
Logic	(b_1, b_2)				
0	(0, 1)	0	1	L	<i>match</i>
		1	0	H	<i>miss</i>
1	(1, 0)	0	1	H	<i>miss</i>
		1	0	L	<i>match</i>
<i>Don't care</i>	(1, 1)	X	\overline{X}	L	<i>match</i>

Table 2: The Imbalanced V_{OUT} of The Asymmetrical P-N Design

SL	V_{OUT}^H when <i>miss</i>	V_{OUT}^L when <i>match</i>	ΔV
1.0V	997.85mV	194.88mV	802.97mV
0.0V	998.06mV	370.69mV	627.37mV

Table 3: The Comparison of Various TCAM Designs

Design		Dual-N (this work)	P-N (this work)	3T-2MTJ [7]	12T-SRAM [9]
MTJ	TMR	1	1	1	N/A
	R_{low}	$2K\Omega$	$2K\Omega$	$3K\Omega$	N/A
Circuit	Technology Node	$45nm$	$45nm$	$45nm$	$45nm$
	V_{dd}	$1.0V$	$1.0V$	$1.0V$	$1.0V$
	Cell Structure	10T-2MTJ	8T-2MTJ	3T-2MTJ	12T
	Sensing Margin	$189.7mV$	$802.4mV$	$56.45mV$	$792.9mV$
	Sensing Latency [†]	$590.34ps$	$97.89ps$	$2.3ns$	$132ps$
	Sensing Energy per Search	$65.05fJ/bit$	$36.85fJ/bit$	$80.56fJ/bit$	$90.75fJ/bit$

[†] The sensing latency was obtained for the word length of 256 bits under the condition that 255 bits match to the input searching data while 1 bit is missed.

scheme. The circuit structure is illustrated in Figure 3(d), where PMOS transistor T1 and NMOS transistor T2 are adopted for access control. Here, V_{OUT} is supplied by the post-amplified signal.

The asymmetrical P-N scheme follows the same fundamental searching principle and the same feedback mechanism as the symmetrical dual-N cell design. Note that the effective resistances of NMOS and PMOS transistors have the opposite dependency on the body bias voltage. More specific, an identical body bias can result in the increase of a NMOS and decrease of a PMOS simultaneously, or vice versa. Thus, we only need two transistors T3 and T4 to supply the body bias in the asymmetrical P-N scheme. The total transistor number reduces to nine.

The most significant difference of the asymmetrical structure from the previous symmetrical version is the use of select transistors. As aforementioned, the threshold voltage drop of NMOS transistor occupies a large portion of the voltage margin, inspiring the symmetrical P-N design. However, the issue cannot be completely solved under all the possible combinations of (b1, b2) and (SL, \overline{SL}).

The simulation results in Table 2 show that the asymmetrical P-N scheme helps improve ΔV in most of scenarios except when (b1, b2) = (1, 0) and (SL, \overline{SL}) = (0, 1), which in fact exacerbates swiftly the voltage margin due to the unbalanced gate-source voltage condition. This is a typical design trade-off between the cell area and the sensing margin. Even though, the simulation and analysis in Section 4 still show the asymmetrical P-N design can obtain the best sense margin and therefore the fastest sensing performance.

4. SIMULATIONS AND EVALUATION

We implemented the proposed TCAM designs and evaluated the performance through circuit simulations in Cadence Virtuoso environment. The $45nm$ CMOS technology with the power supply of $V_{dd} = 1V$ was adopted [11]. The MTJ device model was based on the implementation of $40nm$ perpendicular anisotropy structure [15].

We first examine and compare the sensing margins of the proposed schemes with existing MTJ-based TCAM designs [7] as well as conventional 12T-SRAM based on the design in [9]. Table 3 summarizes the simulation results, including the sensing margins, the sensing latency, as well as the energy consumption.

In this section, the major design factors that affects the performance and robustness of MTJ-based TCAMs are discussed and analyzed based on simulation details. Afterwards, the searching speed and power consumptions of the proposed designs are presented and discussed.

4.1 Impact Factors of Sensing Margin

For the proposed TCAM design structures that leverage the voltage dividing for data detection and searching, the sensing performance is greatly affected by the balance of the two access transistors (T1 and T2) as well as the resistance variations of MTJs (b1 and b2). Accordingly, there are three major factors that potentially degrade the sensing margin: the unbalance in gate voltage supplies to T1 and T2, the CMOS process variations, and the MTJ resistance variations. In this subsection, the impacts of these three design factors on the TCAM robustness will be investigated.

4.1.1 The Unbalanced Transistor Gate Voltage

The gate voltage variations of T1 and T2 are firstly investigated. The impacts on V_{OUT} and the sensing margin ΔV are shown in Figure 4. In the simulations, we assume that T1 is turned on with an ideal gate voltage, *i.e.*, $1.0V$ in the symmetric dual-N scheme and $0V$ for an asymmetric P-N cell. The gate voltage of T2 is then swept from $0V$ to $V_{dd} = 1.0V$.

For a design x , $V_{OUT}^H(x)$ and $V_{OUT}^L(x)$ are used to denote the lowest voltage level of V_{OUT} under the *miss* condition and its highest possible value when the searched and stored bits *matches*, respectively. The simulation results for three MTJ-based TCAM designs are presented, including the two schemes proposed in this work (*Dual-N* and *P-N*) and the 3T-2MTJ design [7] used as the baseline in the work.

Figure 4(a) and (b) demonstrate the exact V_{OUT} and the corresponding sensing margin change of the three designs, which is obtained by

$$\Delta V(x) = |V_{OUT}^H(x) - V_{OUT}^L(x)|. \quad (2)$$

As shown in the figure, V_{OUT} that is dependent on the *match* and *miss* conditions changes dramatically with the gate volt-

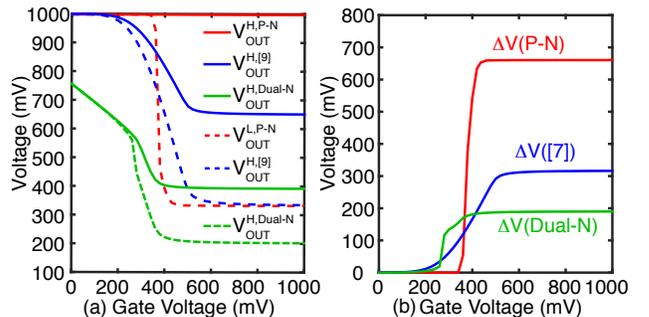


Figure 4: The impact of the gate voltage of access transistors T1 or T2 on (a) V_{OUT} and (b) the voltage margin.

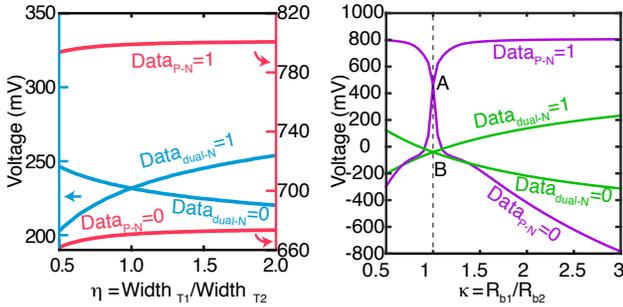


Figure 5: (a) The disequilibrium represented the relation of the sensing margin and the select transistor size ratio η . The y -axis on the left is for the symmetrical dual-N scheme and the right one is for the asymmetrical P-N design. (b) The impact of MTJ resistance variation represented by the change of the voltage margin with the MTJ resistance ratio κ .

age of T2. $\Delta V(x)$ is a key criteria to evaluate the robustness of a given design x . We observe that the asymmetric P-N scheme obtains the highest sensing margin of $627mV$, even when the gate voltage of T2 reduces to about $450mV$. The results demonstrates that the P-N scheme has the strongest resilience on the variations of the access transistor effective resistance .

Comparably, the symmetrical dual-N design has the smallest margins as T2’s gate voltage exceeds $600mV$. However, it can steadily maintain the voltage margin in the largest input range – even when the gate voltage of T2 approaches to $300mV$. Thus, under the extremely worst-case scenario when considering all the forms of reliability hazards, the robustness of the symmetrical dual-N design could be the best among these three designs.

4.1.2 The CMOS Process Variations

The process variations can also break the balance of T1 and T2 to cause disequilibrium. So we investigate the impact of variations resulted by CMOS technology.

Figure 5(a) shows the manifestation of disequilibrium represented by the sensing margin versus parameter η which the ratio of transistor dimension, that is, $\eta = \text{Width}_{T1}/\text{Width}_{T2}$. Here, Width_{T1} and Width_{T2} denote the width of transistors T1 and T2, respectively. In the simulation, we keep Width_{T2} unchanged meanwhile changing Width_{T1} to obtain different η . The simulated curves indicate that the variation in η results in distinguished voltage margins.

For the symmetrical dual-N scheme, the largest sensing margin is obtained at $\eta = 1$ when T1 and T2 are identical. Increasing or decreasing η can result in the sensing margin reduction when “0” or “1” is stored in the cell, receptively. Thus the overall system performance degrades. Note that the two curves are not symmetric. This is because we kept Width_{T2} unchanged so that a bigger η corresponds to larger transistors in use. The simulation results also imply that the system performance degradation has a slower rate when increasing the size of T1 and T2, thought it will results in large design area. As for the disequilibrium of P-N design, the voltage margin stays a high level because the V_{OUT} remains amplified after the two-inverter buffer.

4.1.3 The MTJ Resistance Variation

An unavoidable variation factor comes from the NVM storage unit itself. For a MTJ device, the linear variation of

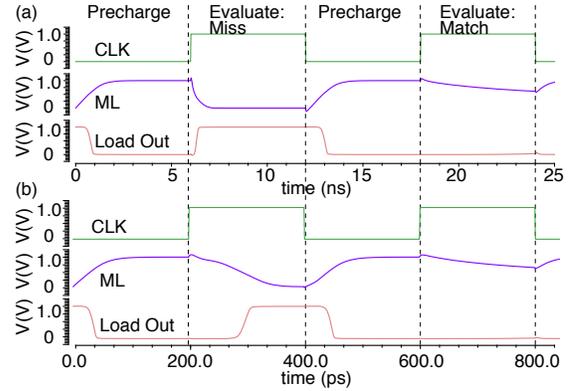


Figure 6: The transient response of key signals: (a) symmetrical dual-N design; (b) asymmetrical P-N design.

the oxide layer thickness can cause the exponential change of the effective MTJ resistance according to the quantum tunnel effect. Thus the impact of MTJ resistance variations should be addressed.

Figure 5(b) shows the change of voltage margins according to the ratio of MTJ resistance $\kappa = R_{b1}/R_{b2}$. Similarly, we fixed R_{b2} in the simulation and changed the value of R_{b1} . Thus, the change of κ indeed reflects the variation of MTJ’s TMR ratio, such as

$$\kappa = \begin{cases} (\text{TMR} + 1)^{-1} & \text{if } \kappa \leq 1 \\ \text{TMR} + 1 & \text{if } \kappa > 1 \end{cases} \quad (3)$$

As expected, the MTJ variation plays a crucial role in the proposed designs. Naturally, the larger κ induces a bigger gap between the sensing voltages of match and miss conditions. Consequently, the voltage margin is larger.

We also want to point out that at $\kappa = 1$, the cross point of the two curves of the same design refers to the sensing margin induced solely by the feedback circuit. More specific, point A is for the asymmetrical P-N scheme and point B is the one of the symmetric dual-N design. The observation of $\Delta V(A) > \Delta V(B)$ is consistent to the result in Figure 4(b) – the asymmetrical P-N design has a larger voltage margin than that of the symmetrical dual-N design.

4.2 The Searching Speed

The searching speed is the most crucial performance factor of TCAMs, which is directly related to the sensing margin. We compare the sensing speeds of the proposed schemes with the baseline MTJ-based TCAM design [7] and conventional 12T-SRAM [9]. The results are shown in Table 3. Here, the sensing latency was obtain for the word length of 256 bits under the worst-case operation condition: only one bit is different from the input searching bit so that the discharging current at ML is minimal. Compared to the baseline design [7], the searching speed of the asymmetrical dual-N scheme improves $12.7\times$ while the enhancement of the asymmetrical P-N design is more than $76.5\times$.

Figure 6(a,b) presents the transient process during the searching operations for the symmetrical dual-N design and the asymmetrical P-N scheme, respectively. The simulations show a slight degradation on ML voltage level when the evaluation results in a matching. This is because of the electrical charge leakage through the discharging transistors in each TCAM cell, since no keeper is used in the proposed designs. The signal of ML is regulated via an inverter, generating the output signal Load-Out in the figure.

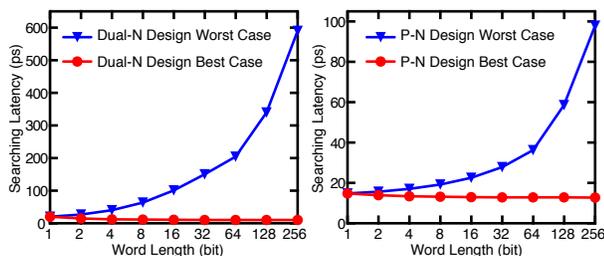


Figure 7: The dependency of the searching latency on the word length. The y -axis on the left is for the asymmetrical P-N design and the right one is for the symmetrical dual-N scheme.

The length of word and the operation condition can also greatly affect the speed of sensing operations as our simulations shown in Figure 7. Same as the above, the worst-case condition occurs when only one bit in the stored data is different from the input searching one. The best-case scenario, in contrast, assume the discrepancy happens at all the bits so that the charge on ML is sunk through the discharging transistors for all the TCAM cells. The simulation shows that searching speed under the best-case situation doesn't change much as the word length increases. However, the worst-case latency increase quickly as bit number grows and therefore the overall capacitance on ML increases. Comparably, the symmetrical dual-N design with smaller sensing margin is much slower than the asymmetrical P-N design. ical dual-N design are lower than those of the asymmetrical P-N design.

4.3 The Searching Energy Consumption

The energy saving is a key incentive in developing the NVM-based TCAM designs. Compared with conventional SRAM-based TCAMs, the non-volatility of these emerging devices allows zero standby power consumption. Besides, we evaluated the dynamic energy consumption in searching operation and summarized the results in Table 3. Compared with 12T-SRAM design, the proposed symmetrical dual-N and asymmetrical P-N designs can obtain about 35.6% and 63.5% energy savings, respectively.

5. CONCLUSIONS

At advanced technologies, NVM-based TCAM designs possess exceptional potential in density improvement and power saving. However, the emerging nonvolatile storage units encounter insufficient sensing margins that results in low design reliability and poor access speed. We proposed to utilize an adaptive body bias feedback scheme to enhance the sensing margin of MTJ-based TCAMs. Determined by the select transistor types in use, two cell structures were presented: the symmetrical dual-N design and the asymmetrical P-N scheme. We thoroughly analyzed the key design factors that affect the performance and reliability of the proposed TCAMs. Both designs demonstrated distinguished enhancement with the enlarged sense margins, fast searching speed and dynamic energy reduction. These characteristics are remarkably beneficial to commercial applications requiring high-performance TCAMs.

Acknowledgments

This work was supported in part by NSF CNS-1311706, NSF CNS-1342566, and Cisco Systems, Inc. Any opinions, findings and conclusions or recommendations expressed in this

material are those of the authors and do not necessarily reflect the views of NSF, Cisco Systems, or their contractors.

6. REFERENCES

- [1] K. Pagiamtzis and A. Shekholeslami, "Content-addressable memory (cam) circuits and architectures: A tutorial and survey," *IEEE Jour. Solid-State Circuits*, vol. 41, no. 3, pp. 712–727, 2006.
- [2] Z. Ullah, M. K. Jaiswal, and R. C. Cheung, "E-tcam: An efficient sram-based architecture for tcam," *Circuits, Systems, and Signal Processing*, vol. 33, no. 10, pp. 3123–3144, 2014.
- [3] H. Li, X. Wang, Z.-L. Ong, W.-F. Wong, Y. Zhang, P. Wang, and Y. Chen, "Performance, power, and reliability tradeoffs of stt-ram cell subject to architecture-level requirement," *IEEE Trans. Magnetics*, vol. 47, no. 10, pp. 2356–2359, 2011.
- [4] Y. Chen, Y. Zhang, and P. Wang, "Probabilistic design in spintronic memory and logic circuit," in *17th Asia and South Pacific Design Automation Conference*, pp. 323–328, IEEE, 2012.
- [5] J. Janesky, N. Rizzo, D. Houssameddine, R. Whig, F. Mancoff, M. DeHerrera, J. Sun, M. Schneider, H. Chia, et al., "Device performance in a fully functional 800mhz ddr3 spin torque magnetic random access memory," in *5th IEEE International Memory Workshop*, pp. 17–20, IEEE, 2013.
- [6] N. Onizawa, S. Matsunaga, and T. Hanyu, "Design of a soft-error tolerant 9-transistor/6-magnetic-tunnel-junction hybrid cell based nonvolatile tcam," in *IEEE 12th International New Circuits and Systems Conference*, pp. 193–196, IEEE, 2014.
- [7] W. Xu, T. Zhang, and Y. Chen, "Design of spin-torque transfer magnetoresistive nan and cam/tcam with high sensing and search speed," *IEEE Trans. Very Large Scale Integration Systems*, vol. 18, no. 1, pp. 66–74, 2010.
- [8] N. Onizawa, S. Matsunaga, and T. Hanyu, "A compact soft-error tolerant asynchronous tcam based on a transistor/magnetic-tunnel-junction hybrid dual-rail word structure," in *20th IEEE International Symposium on Asynchronous Circuits and Systems*, pp. 1–8, IEEE, 2014.
- [9] Y. Nishi, "Advances in non-volatile memory and storage technology," 2014.
- [10] S. Matsunaga, A. Katsumata, M. Natsui, S. Fukami, T. Endoh, H. Ohno, and T. Hanyu, "Fully parallel 6t-2mtj nonvolatile tcam with single-transistor-based self match-line discharge control," in *Symposium on VLSI Circuits*, pp. 298–299, IEEE, 2011.
- [11] W. Zhao and Y. Cao, "Predictive technology model for nano-cmos design exploration," *ACM Jour. Emerging Technologies in Computing Systems*, vol. 3, no. 1, p. 1, 2007.
- [12] Y. Zhang, X. Wang, and Y. Chen, "Stt-ram cell design optimization for persistent and non-persistent error rate reduction: a statistical design view," in *Proceedings of the International Conference on Computer-Aided Design*, pp. 471–477, IEEE, 2011.
- [13] Y. Iba, A. Takahashi, A. Hatada, M. Nakabayashi, C. Yoshida, Y. Yamazaki, et al., "A highly scalable stt-mram fabricated by a novel technique for shrinking a magnetic tunnel junction with reducing processing damage," in *Symposium on VLSI Technology: Digest of Technical Papers*, pp. 1–2, IEEE, 2014.
- [14] J. W. Tschanz et al., "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE Jour. Solid-State Circuits*, vol. 37, no. 11, p. 1396, 2002.
- [15] Y. Zhang, W. Zhao, Y. Lakys, J.-O. Klein, J.-V. Kim, D. Ravelosona, and C. Chappert, "Compact modeling of perpendicular-anisotropy cfeb/mgo magnetic tunnel junctions," *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 819–826, 2012.