A Novel Multi-Context Non-Volatile Content-Addressable Memory Cell and Multi-Level Architecture for High Reliability and Density

Xian Wang†, Deming Zhang†, Kaili Zhang†, Erya Deng†, You Wang∗ and Weisheng Zhao†,∗

†School of Electronic and Information Engineering, Beihang University, Beijing 100191, China
‡School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100191, China
∗Hefei Innovation Research Institute, Beihang University, Hefei 230013, China

Abstract—Currently, non-volatile content-addressable memory (NV-CAM) based on magnetic tunnel junction (MTJ) has huge potential in search applications, owing to its non-volatility, zero standby power and high speed. However, it still suffers from severe reliability and energy dissipation issues especially when the searched data information is large. To address these issues, we propose a multi-context cell (MCC) circuit by employing an output selector (OS) instead of a logic tree (LT) and a multi-level architecture (MLA) by employing SEN generators. In this proposed M-level NV-CAM (e.g., M=2), every M 1T/2MTJ memory cells share one search circuit composed of a M-level selector, a pre-charge sense amplifier (PCSA), a OS and a math-line (ML) switch to improve area efficiency, and the search circuit together with M memory cells combine into a MCC. Moreover, the search-enable signal SEN influenced by ML can bring inessential search operations inactivity. Hybrid 40nm CMOS/MTJ simulation results show that the proposed MCC circuit can reach a lower search-error-rate (SER) of 0.5 % and a lower search delay of 39.45 ps compared with the previous cell circuit with LT. On the other hand, the SER of searching a 144-bit data information in the proposed 2-level-architecture NV-CAM can be only 4.5 %, about 2.93 times lower than that in the traditional-architecture NV-CAM.

Index Terms—non-volatile content-addressable memory (NV-CAM), magnetic tunnel junction (MTJ), multi-context cell (MCC), output selector (OS), multi-level architecture (MLA), search-error-rate (SER)

I. INTRODUCTION

Content-addressable memory (CAM) is widely used in various lookup applications for its high search speed [1]. In recent years, non-volatile CAM (NV-CAM) has been researched to reduce the standby power consumption by employing magnetic tunnel junction (MTJ) to store data information [2], [3]. As the memory element, MTJ is mainly composed of one ultra-thin oxide barrier layer sandwiched by two ferromagnetic (FM) layers, reference layer and free layer [4]. Generally, the magnetization direction of reference layer is fixed, whereas that of free layer can be switched by a large enough bidirectional current. When the relative magnetization direction of these two FM layers is parallel (anti-parallel), the MTJ will exhibit a low (high) resistance state, i.e., $R_P$ ($R_{AP}$). And the resistance difference can be characterized by the tunnel magneto-resistance ratio (TMR = $(R_{AP} - R_P)/R_P$). To address the severe reliability issue caused by small TMR, the 1T/2MTJ memory cell has been proposed at the expense of density, where two MTJs in complementary resistance states are used to represent 1-bit stored data information [5].

However, the speed and energy consumption of previous NV-CAM cell circuits come at the cost of reliability and area [6]–[10]. As larger capacity of NV-CAM is being required, ensuring high reliability and low energy consumption without
sacrificing area is the main thread of recent research. Thus, the NV-CAM in multi-context architecture [10], [11] has been presented to reduce the area overhead by memory cells in the same location of words sharing the same comparison and writing circuit, but its total search delay rises greatly with the amount of words increasing. In addition, the NV-CAM in segmented match-line (ML) architecture [12] has been presented to minimize the active energy dissipation, but it suffers from lower density and reliability.

In this paper, we propose a novel NV-CAM to meet the development of big data in both cell circuit and architecture sides. The contributions of this paper can be summarized as follows:

1) We firstly propose a multi-context cell (MCC) circuit for NV-CAM by employing an output selector (OS) instead of a logic tree (LT) to reduce search-error-rate (SER) and search delay, which also combines multi 1T/2MTJ memory cells by employing the M-level selector to improve area efficiency.

2) We firstly propose a SEN generator to realize the multi-level architecture (MLA) for NV-CAM, which brings inessential search operations to idle state and results in high reliability and low energy consumption. Although its search delay is a little higher than that of traditional one, they are getting closer as the length of data information rises. Hybrid 40 nm CMOS/MTJ simulations are also performed to analyze the influence of level amounts on its performances.

The rest of paper is organized as follows. Section II introduces the proposed MCC circuit with OS and its performance. Section III describes and evaluates the proposed MLA. Finally, Section IV makes a conclusion.

II. PROPOSED MCC FOR HIGH RELIABILITY AND SPEED

A. Proposed MCC

Fig. 1(a) presents the proposed MCC, which is mainly composed of five parts: a pre-charge sense amplifier (PCSA), a M-level selector (M denotes the amount of levels in the MLA), a configurable MTJ array, an output selector (OS) and a match-line (ML) switch. The PCSA contains two pre-charge transistors (P1,2) controlled by the search-enable signal SENi (i and j denote the number of bit and word in NV-CAM, respectively) and two cross-coupled inverters (P3,4 and N1,2) to improve sense reliability. The M-level selector contains 2M switch transistors (e.g., N1,6 when M=2) controlled by the bit-search signals SBi (e.g., SB0 and SBN when M=2) to select a memory cell with two complementary MTJs (e.g., d1-2 or d3-4) in one level of search operation (e.g., level I or level II). The configurable MTJ array contains M 1T/2MTJ memory cells, in which the switch transistors (e.g., N7) are controlled by the bit-write signals WBi (e.g., WB0) and the bit-line (BL) together with bit-line-bar (BLB) can come into a write current to configure every pair of MTJs in complementary resistance states. The OS contains an inverter and two switch transistors (P9,10) controlled by the search-line (SL) and search-line-bar (SLB) which have different inputs in different levels to compare the searched data information with the stored data information. Especially, replacing the LT [10] with OS causes area decrease of proposed MCC for the sake of transistors. The ML switch contains a pass transistor (N0) as a critical path between ML and the ground and the dis-charge delay of ML is limited due to the full-swing cell voltage (Vcell).

Fig. 1(b) presents the SEN generator required per word to control the search operation according to the voltage of ML, which is composed of two inverters and a OR gate.

B. Write Operation

The write operation of the proposed MCC includes M levels: e.g., level I for writing the d1-2 and level II for writing the d3-4 when M is 2.

In level I, the signals SB0 and SBN are driven to GND and the transistor N7 is activated by the signal WB0. Then, the free layers of d1 and d2 are connected together and the write current is generated by the voltage difference between the BL and complementary BLB as shown in Fig. 2(a). For writing ‘1’, the BL and BLB are driven to Vwrite and GND, respectively. Subsequently, the d1 and d2 are written into RAP and RP, respectively. For writing ‘0’, it is reversed that the BL and BLB are driven to GND and Vwrite, respectively, which results in the RP and RAP states of d1 and d2, respectively. Similarly, the operation of writing d3-4 in level II is described in Fig. 2(b).

C. Search Operation

The search operation of proposed MCC includes M levels: e.g., level I for searching the d1-2 and level II for searching...
the d3-4 when M is 2. And each level contains two phases: pre-charge phase and sense phase. Note that the ML will be charged to VDD by the signal PRE_ML before the search operation. Thereby, the signal SEN0 in level I only depends on the signal PRE_PCSA according to the circuit in Fig. 1(b).

In the pre-charge phase of level I as shown in Fig. 3(a), the signals SB0 and SBN are driven to GND and the transistors P1_2 are activated by the signal SEN0. Subsequently, both the node A and B are charged to VDD. At the same time, the SL and SLB are driven to SL0 and SLB0, respectively, which results in a high Vcell for there must be a high voltage between SL0 and SLB0. Therefore, the ML will maintain VDD in the whole pre-charge phase.

After that, the signals SEN0 and SB0 are driven to VDD, and the signals WB0, WBN, SBN, BL and BLB are driven to GND, i.e., it enters into the sense phase of level I. Meanwhile, the SL and SLB are driven to SL0 and SLB0, respectively. In this way, the transistors P1_2 are turned off, the transistors N1_4 are turned on and the d1 and d2 are connected to the node A and B, respectively. As summarized in Table I, there exits four kinds of search cases corresponding to the stored data information in memory cell and the searched data information inputted by SL0 and SLB0.

- For the stored data information is ‘0’, the (d1, d2) are in (RP, RA) states. Then, the node A discharges at a quicker speed than the node B at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node A firstly becomes less than the threshold of the inverter (P4 and N2). Eventually, the node A and B can be amplified to GND and VDD, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned on and off, respectively, which results in a low Vcell. In this case, the ML keeps the initial voltage of VDD and represents the search result of mismatch.

- For the stored data information is ‘1’, the (d1, d2) are in (RA, RP) states. Then, the node B discharges at a quicker speed than the node A at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node B firstly becomes less than the threshold of the inverter (P3 and N1). Eventually, the node A and B can be amplified to VDD and GND, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned off and on, respectively, which results in a high Vcell. Then, the pass transistor N0 is fully turned on. In this case, the ML will be discharged to GND and represents the search result of mismatch.

- For the stored data information is ‘1’, the (d1, d2) are in (RP, RA) states. Then, the node B discharges at a quicker speed than the node A at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node B firstly becomes less than the threshold of the inverter (P3 and N1). Eventually, the node A and B can be amplified to VDD and GND, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned on and off, respectively, which results in a low Vcell. Then, the pass transistor N0 is fully turned on. In this case, the ML will be discharged to GND and represents the search result of mismatch.

- For the stored data information is ‘1’, the (d1, d2) are in (RA, RP) states. Then, the node A discharges at a quicker speed than the node B at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node A firstly becomes less than the threshold of the inverter (P4 and N2). Eventually, the node A and B can be amplified to GND and VDD, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned off and on, respectively, which results in a high Vcell. Then, the pass transistor N0 is fully turned on. In this case, the ML will be discharged to GND and represents the search result of mismatch.

---

**Table I**

<table>
<thead>
<tr>
<th>Stored Data</th>
<th>Searched Data</th>
<th>Pre-charge Phase</th>
<th>Sense Phase</th>
<th>Search Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>0'</td>
<td>0'</td>
<td>VDD</td>
<td>VDD</td>
<td>Match (ML='1')</td>
</tr>
<tr>
<td>(RP, RA)</td>
<td>'1'</td>
<td>VDD</td>
<td>VDD</td>
<td>Mismatch (ML='0')</td>
</tr>
<tr>
<td>'1'</td>
<td>0'</td>
<td>VDD</td>
<td>VDD</td>
<td>Mismatch (ML='0')</td>
</tr>
<tr>
<td>(RA, RP)</td>
<td>'1'</td>
<td>VDD</td>
<td>VDD</td>
<td>Match (ML='1')</td>
</tr>
</tbody>
</table>

- For the stored data information is ‘0’, the (d1, d2) are in (RP, RA) states. Then, the node A discharges at a quicker speed than the node B at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node A firstly becomes less than the threshold of the inverter (P4 and N2). Eventually, the node A and B can be amplified to GND and VDD, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned on and off, respectively, which results in a low Vcell. In this case, the ML keeps the initial voltage of VDD and represents the search result of mismatch.

- For the stored data information is ‘1’, the (d1, d2) are in (RA, RP) states. Then, the node B discharges at a quicker speed than the node A at the beginning of sense phase owing to the larger discharge current. As a result, the voltage of node B firstly becomes less than the threshold of the inverter (P3 and N1). Eventually, the node A and B can be amplified to VDD and GND, respectively, owing to the positive feedback of cross-coupled inverters. If the searched 1-bit data information is ‘0’, the transistors N9 and N10 will be turned off and on, respectively, which results in a high Vcell. Then, the pass transistor N0 is fully turned on. In this case, the ML will be discharged to GND and represents the search result of mismatch.
Note that the voltages of nodes A and B only depend on the stored data information and the voltage of \( V_{\text{cell}} \) depends on the searched data information and the voltages of nodes A and B.

Especially, the operation of searching d3-4 in level II will be similar as described above and the SL and SLB will be driven to SL\(_A\) and SLB\(_N\), respectively, if the search result in level I is match. Otherwise, the voltage of signal \( \text{SEN}_0 \) will stay high according to the circuit in Fig. 1(b), which will result in an idle state in level I, i.e., the MCC will stop searching d3-4 if the search result of d1-2 is mismatch, which successfully reduces the inessential operations.

Moreover, only when the ML still keeps the initial voltage of \( V_{\text{DD}} \) unchanged after operations in all levels, the final M-bit search result (e.g., \( M=2 \)) of MCC is match.

**D. Functional Simulation**

By using a physics-based STT-MTJ compact model [13] and a commercial CMOS 40 nm design kit, hybrid simulations have been performed to demonstrate the functionality of the proposed MCC (e.g., \( M=2 \)). Table II shows the critical parameters and their default values of both the STT-MTJ and the transistors in the simulations.

Fig. 4(a) shows the transient simulation results of the proposed 2-bit MCC for writing four kinds of data, i.e., ‘00’, ‘01’, ‘10’ and ‘11’. As seen, each write operation includes two levels: level I where the signal \( WB_0 \) is set to be \( V_{\text{DD}} \) to connect d1 and d2 and level II where the signal \( WB_N \) is set to be \( V_{\text{DD}} \) to connect d3 and d4. For writing ‘0’, the BL and BLB are driven to GND and \( V_{\text{Vwrite}} \), respectively, thereby leading to the (d1, d2) written into \((R_P, R_{AP})\). For writing ‘1’, the BL and BLB are driven to \( V_{\text{Vwrite}} \) and GND, respectively, thereby leading to the (d1, d2) written into \((R_{AP}, R_P)\).

Fig. 4(b) shows the transient simulation results of the proposed 2-bit MCC for searching four kinds of data, i.e., ‘00’, ‘01’, ‘10’ and ‘11’ in the case where the stored data information is ‘01’ as example. As seen in the pre-charge phase, the signals \( \text{SEN}_0, \text{SB}_0 \) and \( \text{SB}_N \) are all set to be GND and the \( V_{\text{cell}} \) is GND, which keeps the pass transistor \( N_0 \) in off state. Then in the sense phase, the SL and SLB are set to be GND and \( V_{\text{DD}} \) for searching ‘0’ or \( V_{\text{DD}} \) and GND for searching ‘1’. Moreover, in the sense phase of level I the signal \( SB_0 \) is set to be \( V_{\text{DD}} \) to select d1 and d2 and in the sense phase of level II the signal \( SB_N \) is set to be \( V_{\text{DD}} \) to select d3 and d4. Finally, in the match case, the \( V_{\text{cell}} \) is still GND and then the ML maintains the voltage of \( V_{\text{DD}} \). But in the mismatch case, the \( V_{\text{cell}} \) is \( V_{\text{DD}} \) and then the ML is eventually discharged to GND. Especially, the operation in level II will be in idle state and the voltages of nodes A and B will be unchanged if the search result in level I is mismatch.

**E. Performance Evaluation**

Monte-Carlo simulations of 1000 runs per case are performed with the consideration of 1 \( \sigma \) CMOS probability distribution and 3 % MTJ process variation to evaluate the level-
I search reliability of proposed MCC (e.g., M=2). Fig. 5(a) shows the influence of the TMR on the SER of the MCC circuits. As seen, the SER of the proposed MCC with OS is much lower than that of the previous MCC with LT [10] due to the decrease of transistor amounts in the discharge branches. In addition, its SER decreases greatly from 13.7 % to 0.2 % when the TMR increases from 100 % to 300 %, which means that TMR improves the reliability of search operation. Fig. 5(b) shows the influence of transistor width on the SER of the proposed MCC, where the TMR is 150 %. T_{PCSA} and T_{MLS} denote the widths of transistors N_{1-2} in the PCSA and of transistors N_{3-6} in the M-level selector, respectively. As seen, its SER decreases violently with the increase of T_{PCSA} or T_{MLS} owing to the increase of resistance difference between two discharge branches, which means that larger width is required for high-reliable search operation. It should be noted that only 0.5 % SER can be observed in the proposed 2-bit MCC when the TMR, T_{PCSA}, T_{MLS} and V_{DD} are set to be 150 %, 240 nm, 240 nm and 1 V, respectively, which will be employed in the following simulations.

Additionally, the level-I search delay and energy of proposed MCC (e.g., M=2) are evaluated in the mismatch case. Here, the 1-bit search delay is measured from the time when the signal SEN_0 rises to the half of V_{DD} to the time when the ML is discharged to the half of V_{DD}. As illustrated in Fig. 5(c), the search delay of proposed MCC with OS is a little lower than the previous MCC with LT and decreases greatly from 113.82 ps to 25.97 ps when the V_{DD} increases from 0.7 V to 1.3 V. As illustrated in Fig. 5(d), the difference between the level-I search energy of two MCC circuits can almost be negligible and the search energy of proposed MCC rises from 0.820 fJ to 2.865 fJ with the increase of V_{DD}.

III. PROPOSED MLA FOR HIGH RELIABILITY AND DENSITY

A. Proposed MLA

Furthermore, a MLA for NV-CAM is proposed to improve the reliability and density and reduce energy consumption notably. Fig. 6 describes a 4×2N (i.e., four 2N-bit words, M=2 and N denotes the amount of bits per level) NV-CAM as example, which consists of 4N proposed 2-bit MCCs with OS, 4 MLs, 2N SLs, 2N SLBs, 4 pre-charged transistors and 4 SEN generators. As seen, it is equally divided into two levels and every two 1-bit memory cell (e.g., B_0 and B_N) share one same search circuit, which improves the area efficiency. In the search operation of a 2N-bit data information, level I is firstly performed to search the former N bits. If its search result is matched, i.e., the voltage of ML_j stays high, level II will be performed to search the latter N bits continually. Otherwise, level II will be in the idle state and the inessential operations will be saved. In this way, the search energy dissipation can be reduced and higher search reliability can be achieved due to the shortened data length per search. Moreover, most words in a NV-CAM can be excluded without being searched all bits owing to the high probability of mismatch.

B. Performance Evaluation

Fig. 7(a) illustrates that the SER of a 1×2N NV-CAM in the match case rises with the increase of word length (2N) from 2 to 144. As seen, the SER of NV-CAM in the MLA of two levels is much lower than that in the traditional architecture without levels and it is only 4.5 % even when the word length reaches 144 bits, which proves the high reliability improved by our proposed MLA.

Fig. 7(b) shows the influence of word length on search delay of a 1×2N NV-CAM. Evaluated in the worst case where only
1-bit mismatch exits, the 1-word search delay is M times as much as the value measured from the time when the signal SEN\textsubscript{0} rises to 1/2\textsubscript{VDD} to the time when the ML is discharged to 1/2\textsubscript{VDD}. As seen, the search delay of NV-CAM in the MLA of two levels increases linearly with the rise of 2N and is always about 45 ps larger than that in the traditional architecture. Nevertheless, the delay ratio defined as the quotient of them gets closer to ‘1’ and the search delay of NV-CAM in the MLA is only 0.167 ns even when the word length reaches 144 bits, which means that the cost of speed is little.

Fig. 7(c) shows the search delay and energy of a 1×144 NV-CAM with respect to different \textsubscript{VDD}. The 1-word search energy is measured in three different cases: case I where 1-bit mismatch exits in level II, case II where 1-bit mismatch exits in level I and case III where both two levels match. As seen, the search delay decreases and the search energy increases with the rise of \textsubscript{VDD}. Especially, even in case I the search energy of NV-CAM in the MLA of two levels (e.g., 1.818 fJ/bit) is lower than that in the traditional architecture (e.g., 2.192 fJ/bit), let alone in case II (e.g., 0.872 fJ/bit), which proves the low energy consumption improved by proposed MLA.

Fig. 7(d) shows the performance of a 1×144 (i.e., M×N=144) NV-CAM with respect to different M. As seen, the SER and search energy in case II reduce greatly, while the search delay increases linearly with respect to M from 1 to 4. Even though the search energy in case III rises when M increases, the total energy consumption of a multi-word NV-CAM must be lower due to the few amount of words required to be searched in all levels. Moreover, the density increases greatly for the amount of transistors per bit is only (11M+3). Thus, it can be predicted that more levels are called for big data.

IV. CONCLUSION

In this paper, we firstly realize a multi-context cell with output selector and a multi-level architecture with SEN generator, which improves the reliability, energy consumption and area efficiency by reducing inessential operations. Hybrid simulations demonstrate its low SER of 4.5 % /144-bit, low energy in case I of 261.8 fJ /144-bit and low delay of 167.08 ps /144-bit when the M, TMR, T\textsubscript{PCSA}, T\textsubscript{MLS} and V\textsubscript{DD} are set to be 2, 150 %, 240 nm, 240 nm and 1 V, respectively. In addition, the multi-level architecture of more levels are called for big data with the rapid development of NV-CAM.

REFERENCES