4-Transistor Ternary Content Addressable Memory Cell Design using Stacked Hybrid IGZO/Si Transistors

Munhyeon Kim
Seoul National University
munhyeon.kim@snu.ac.kr

Jae-Joon Kim
Seoul National University
kimjaejoon@snu.ac.kr

ABSTRACT

In this paper, we propose a 4T-based paired orthogonally stacked transistors for random access memory (POST-RAM) cell structure and also suggest ternary content addressable memory (TCAM) applications. POST-RAM cells feature vertically stacked read and write transistors, maximizing area efficiency by utilizing only two transistors' space. POST-RAM employs InGaZnO (IGZO) channels for write transistors and single crystal silicon channels for read transistors, which results in both extremely long memory retention and fast reading performance. A comprehensive 3D-TCAD simulation is conducted to validate the procedural design of the proposed device structure. Furthermore, we introduced a self-clamped searching scheme (SC2S) designed to enhance the efficiency of TCAM operations. The results conclusively demonstrate that operating a TCAM based on the proposed POST-RAM architecture can lead to a 20% improvement in energy-delay product (EDP). Notably, the delay performance can be enhanced by up to 40% when compared to a 16T SRAM-based TCAM. Additionally, the proposed scheme enables a more than sixfold reduction in cell area, demonstrating an efficient use of space.

CCS CONCEPTS

• **Hardware** → **Dynamic memory**; *Static memory*.

KEYWORDS

2T0C eDRAM, IGZO Transistor, monolithic 3D stacked Transistors, Ternary content-addressable memory (TCAM)

ACM Reference Format:
Munhyeon Kim and Jae-Joon Kim. 2024. 4-Transistor Ternary Content Addressable Memory Cell Design using Stacked Hybrid IGZO/Si Transistors. In *61st ACM/IEEE Design Automation Conference (DAC '24), June 23–27, 2024, San Francisco, CA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3649329.3655965

1 INTRODUCTION

Content addressable memory (CAM) returns addresses by comparing the stored data with the search data [1]. Specifically, CAM conducts a query on the content being searched, leading the match line to point the address of the result. Additionally, ternary content addressable memory (TCAM) distinguishes itself from binary content addressable memory (BCAM), which deals exclusively with "1" and "0", by introducing a third state known as "don't care" [2]. This inclusion significantly enhances lookup speed of TCAM compared to BCAM. Traditionally, CAM has been applied to tasks such as internet protocol packet classification and packet forwarding in high-performance network routers [3, 4]. Recently, there is potential for applications in artificial intelligence (AI), such as memory augmented neural networks (MANNs) [5].

However, TCAM traditionally relies on static random-access memory (SRAM)-based 16T NOR-type cells, leading to challenges of high cost and poor area efficiency [6–8]. The challenges of recent CMOS scaling have compounded, making energy-efficient operations even more difficult [9]. As a result, there is ongoing research into TCAM cells based on emerging non-volatile memory (eNVM), encompassing technologies like resistive random access memory (RRAM) [10], magnetic random access memory (MRAM) [11, 12], phase-change random access memory (PRAM) [13] and ferroelectric field-effect transistor (FeFET) [5]. These technologies demonstrate promising outcomes from both energy efficiency and compact cell perspectives. Nonetheless, to ultimately enhance search speed while increasing density, further research is necessary. And reducing the cell area with eNVM-based cells and achieving speeds beyond CMOS pose challenges. Furthermore, there is active research in the field of increasing the ratio between high-resistance state (HRS) and low-resistance state (LRS) [14]. Overcoming the variations in eNVM has become a significant challenge, and dedicated error correction code (ECC) schemes are consistently under discussion [15]. Due to the widespread adoption of CMOS-based cells, there is an urgent need for the proposition of technologies that are not only efficient but also capable of minimizing the extent of change.

In this study, we propose a 4T-based TCAM cell with a hybrid stack of single crystalline silicon and InGaZnO (IGZO) channel MOSFETs. The proposed paired orthogonally stacked transistors for random access memory (POST-RAM) cell has a great potential as a next-generation memory candidate capable of simultaneously achieving significantly high speed and density. In particular, the compatibility of the read transistor with CMOS technology makes it highly convenient for leveraging existing technology and circuit techniques.

Our contributions can be succinctly summarized as follows:

- We introduce POST-RAM, a 4T-based TCAM cell with a monolithic 3D structure in which transistors are vertically stacked with single crystal channels and IGZO channels. The write transistor is rotated and stacked to form a 90° angle with the read transistor.

DAC '24, June 23–27, 2024, San Francisco, CA, USA

ACM ISBN 979-8-4007-0601-1/24/06.
https://doi.org/10.1145/3649329.3655965

- We propose the storage node patterning technology (SNPT) as a crucial process integration scheme to form the structure of POST-RAM. Through rigorous validation using 3D-TCAD, we have verified its precision from a single cell to an array.
- We propose the self-clamped searching scheme (SC2S) for TCAM operations in POST-RAM. This scheme, while maintaining a compact cell area, operates effectively and has been confirmed to exhibit robust performance against process, voltage, and temperature (PVT) variations. Additionally, we demonstrate its superiority in power, performance, and area (PPA) compared to 16T-TCAM cells based on SRAM.

2 BACKGROUND AND RELATED WORK

2.1 Expansion of TCAM cell research into diverse memories

Traditionally, as depicted in Fig. 1(a) and (b), TCAM has conventionally employed an SRAM-based 16T NOR-type cell, leading to high costs and suboptimal area efficiency. Furthermore, the advantages that could potentially be obtained by scaling are drastically declining due to the slowdown in CMOS scaling. However, leveraging eNVM technologies allows for downsizing compared to the conventional 16T NOR-type SRAM-based TCAM cell, thereby improving area efficiency. Cells based on RRAM, PRAM, and MRAM exhibit compatibility with the back-end-of-line (BEOL) through the resistor, resulting in an occupied area that closely aligns with the area of the access transistor. Furthermore, it has been reported that FeFET, utilizing only two transistors, enables eNVM-based TCAM to achieve a compact design corresponding to the area of two transistors [5]. However, eNVM-based TCAM cells face challenges that eNVM components cannot entirely overcome. Limitations in endurance must be addressed for their positioning as on-chip memory to be immediately feasible in the future [16]. Additionally, to replace traditional TCAM, it is necessary to efficiently configure cells with speeds comparable to CMOS.

2.2 2T0C eDRAM cells adopting IGZO channel

Traditional dynamic random-access memory (DRAM) cell is composed of an access transistor and a capacitor (1T1C). Scaling cell areas while maintaining an adequate capacitor capacity has proven to be a challenging endeavor. Recently, significant attention has been directed towards a novel alternative, the 2T0C cell, which employs two transistors utilizing an IGZO channel without a capacitor [17, 18]. The 2T0C cell, known for achieving under 10^{-17} μA/μm off current (I_{OFF}) and demonstrating retention exceeding 10^2 s after a single write operation, has garnered attention as a potential advancement in memory technology. Moreover, the choice to control charge flow through transistors rather than destructively manipulating the switching layer as in eNVMs contributes to the high endurance of IGZO-based 2T0C cells [17]. The achieved endurance is noteworthy, considering it aligns with the endurance levels observed in conventional DRAM systems. This finding underscores the potential of IGZO-based 2T0C cells as a viable and resilient alternative in the landscape of memory technologies. However, despite these promising attributes, IGZO-based 2T0C cells face several challenges. Firstly, their mobility is significantly lower compared

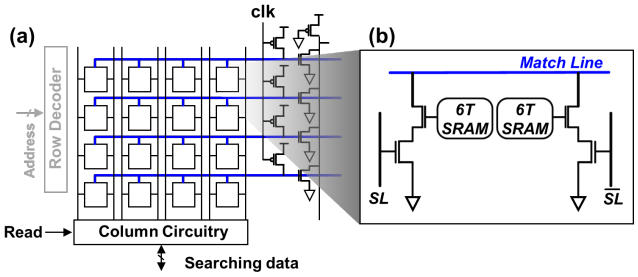

Figure 1: (a) Conceptual view of conventional content-addressable memory (CAM), (b) 16T-SRAM based nor-type ternary content-addressable memory (TCAM) cell [6].

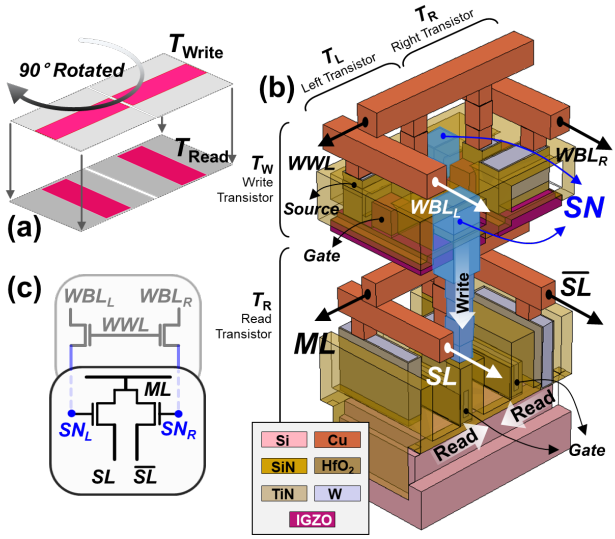

Figure 2: (a) Concept of paired orthogonally stacked random-access memory (POST-RAM) cell, (b) 3D view of POST-RAM cell structure, (c) circuit diagram of the POST-RAM cell.

to single crystalline silicon, necessitating further research [19]. Secondly, achieving reliability in transistors when vertically stacking two or more layers is challenging. Although engineering efforts have enabled processes below 400°C and successful reliability for vertically stacked cells, additional research is required when stacking more than two cells to ensure continued reliability [20].

3 PROPOSED TCAM CELL: POST-RAM

3.1 Concept of POST-RAM

The proposed POST-RAM has two key features in terms of the device structure. Firstly, as shown in Fig.2(a), the upper and lower transistors are stacked in a vertical orientation through a 90° rotation. In the 2T0C cell based on the gain cell, the drain of the write transistor is directly connected to the gate of the read transistor, utilizing this node as a storage node. Therefore, by employing a 90° rotation of the write transistor compared to the read transistor, as in POST-RAM, the drain of the write transistor precisely aligns with the gate of the read transistor. Secondly, as shown in Fig.2(b), the lower read transistor, which utilizes a single crystal silicon channel,

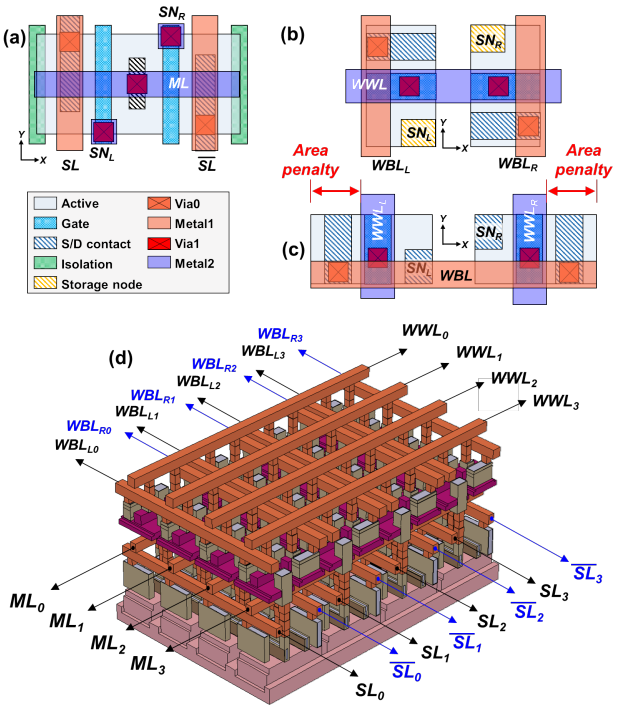

Figure 3: (a) Read transistor layout and (b) write transistor layout of POST-RAM, (c) Write transistor layout of vertically stacked 4T0C (VS-4T0C) cell. VS-4T0C has read and write transistors stacked in parallel. (d) Bird's eye view of 4×4 array of POST-RAM.

and the upper write transistor, which employs an IGZO channel, are vertically stacked. The use of an IGZO channel in the upper write transistor allows for excellent retention characteristics, while the lower read transistor, utilizing a single crystal silicon channel, enables a CMOS-compatible process scheme and performance comparable to SRAM. The circuit diagram of the POST-RAM cell is shown in Fig.2(c). The layout of the read transistor and write transistor in POST-RAM is presented in Fig.3(a) and (b) respectively. It can be observed that the left storage node (SN_L) and right storage node (SN_R) are aligned. In contrast, as depicted in Fig.3(c), in the case of the 4T0C (VS-4T0C) cell where the read transistor and write transistor are vertically stacked without rotation, SN_L and SN_R are separated and need to be controlled separately. Therefore, an additional area equivalent to the contacted gate pitch (CGP) of each side is required. Additionally, the bird's eye view in Fig. 3(d) allows verification of the process integration scheme of POST-RAM, ensuring that cells in the array are connected in an orderly manner. Specifically, it enables checking whether there is sufficient margin between adjacent cells for SN_L and SN_R, and whether isolation is effectively achieved.

3.2 Fundamental PPA gains of POST-RAM

POST-RAM exhibits fundamental gains in PPA due to its structure where the read and write transistors are vertically stacked in an orthogonal direction. As shown in Fig. 4(a), the ultimate POST-RAM structure, when evaluated in a 28-nm technology, occupies

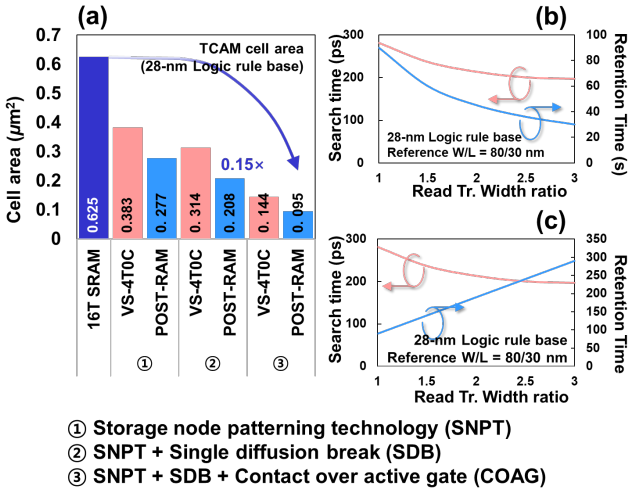

Figure 4: (a) Cell Area Evaluation of POST-RAM, Trends in read transistor width for (b) VS-4T0C and (c) POST-RAM in terms of search performance and retention time.

only 15% of the area compared to a 16T SRAM-based TCAM cell. Achieving the ultimate POST-RAM engineering requires several process integration technologies as prerequisites. Notably, SNPT is a key element, followed by single diffusion break (SDB) [21], and finally, contact over active gate (COAG) [22]. SDB is an essential process that must be adopted when manufacturing POST-RAM for processes below the 7-nm technology node, due to its significantly more efficient isolation method compared to the paired cell structure of double diffusion break (DDB). Furthermore, adopting COAG in POST-RAM offers substantial benefits, as it yields much greater advantages in unipolar structures using logic technology, as opposed to structures like logic standard cells led by inverters. With the application of all these technologies, POST-RAM can reduce the cell area by 35%, compared to VS-4T0C. Furthermore, Fig.4(b) and (c) present the results of simulating and monitoring the behavior of search speed and retention time by varying the width of the read transistor for VS-4T0C and POST-RAM, respectively. The width of the read transistor is directly related to and proportional to the discharging speed of the match line (ML).

However, due to the use of vertically stacked cells, allocating area in the width direction of the read transistor also requires additional area for the write transistor. The retention characteristics vary depending on whether the additional area allocated to the write transistor is in the form of gate length or channel width. In the case of VS-4T0C shown in Fig.4(b), as the width of the read transistor increases, the search time decreases. However, since the width of the write transistor also increases, the retention time decreases. Considering that excluding the option of not increasing the width of the write transistor is not feasible, the best-case scenario is maintaining retention. On the other hand, examining the case of POST-RAM in Fig.4(c), increasing the width of the read transistor results in an increase in the gate length of the write transistor. Consequently, both search speed and retention time increase. Specifically, doubling the width of the read transistor reduces the search time by 68 ps, and the retention time approximately doubles.

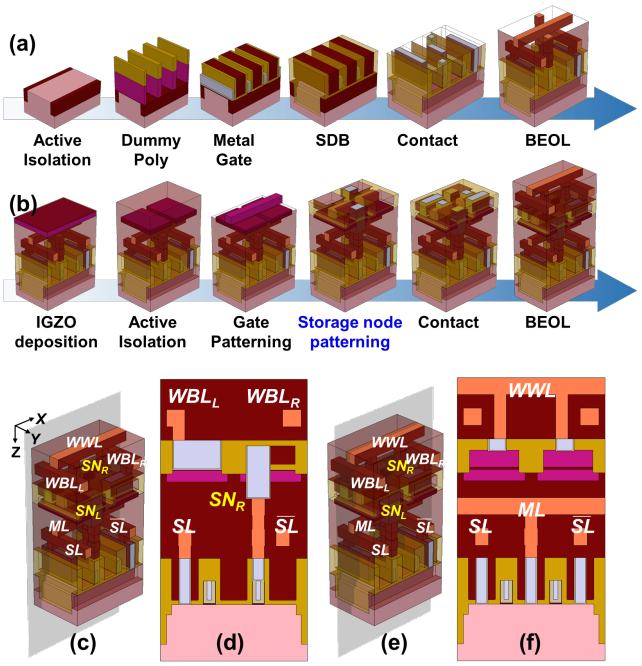

Figure 5: The process integration scheme of POST-RAM is verified by the results of a 3D process simulation. (a) Read transistor and (b) write transistor. (c)-(d) storage node (SN)-centered cross-section view, (e)-(f) match line (ML)-centered cross-section view.

The relationship between search speed and retention, stemming from the structure of POST-RAM, is expected to increase the timing margin during cell design. This is attributed to its role as a device that helps avoid critical trade-offs.

3.3 Process integration scheme

The key process scheme for POST-RAM is SNPT, SDB and COAG. Among them, SDB and COAG have already been validated in advanced logic technology [21, 22]. To validate SNPT, 3D process simulations are conducted. Fig.5(a) shows the simulation for the read transistor process in POST-RAM, incorporating the SDB process. It is noteworthy that the read transistor can be sufficiently implemented with existing logic technology, and employing advanced transistor structures such as FinFET [23], gate-all-around (GAA)-FET, and nanosheet (NS) FET [24] could further enhance its PPA. Furthermore, as demonstrated in Fig.5(b), the use of the SNPT allows verification of error-free fabrication of the POST-RAM structure. Furthermore, to concretely verify the results of the process, cross-sectional views are examined. The cross-sectional view centered on SN_R is shown in Fig. 5(c) and (d), clearly revealing that the drain of the right write transistor is connected to the gate of the right read transistor through SN_R. Additionally, the cross-sectional view centered on ML is observable in Fig. 5(e) and (f), where, from the perspective of the read transistor, the shared drain is connected to ML, maximizing area efficiency. Moreover, its parallel formation with the write word-line (WWL) demonstrates the rotation of the write transistor in POST-RAM.

Input			Storage Node			ML	
Val.	SL	$\overline{SL}$	Val.	SN_L	SN_R	Val.	ML_{SO}
-1	V_{SS}	V_{DD}	-1	V_{SS}	V_{DD}	1	V_{DD}
-1	V_{SS}	V_{DD}	1	V_{DD}	V_{SS}	-1	V_{SS}*
1	V_{DD}	V_{SS}	-1	V_{SS}	V_{DD}	-1	V_{SS}*
1	V_{DD}	V_{SS}	1	V_{DD}	V_{SS}	1	V_{DD}

* The voltage of the match line is determined by the quantity of miss cells.

Table 1: TCAM input and output voltage conditions using POST-RAM

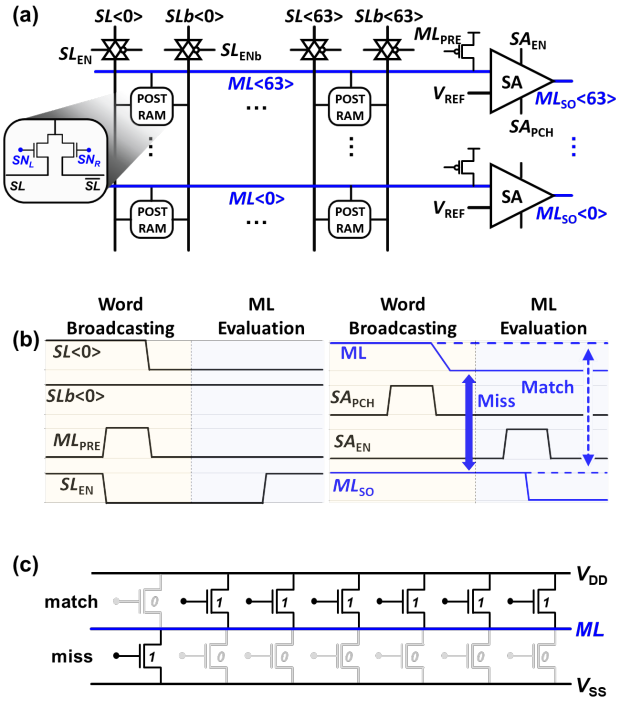

Figure 6: (a) ML sensing circuit consisting of POST-RAM TCAM cell array. (b) Waveform of a search operation with word broadcasting and ML evaluation. (c) concept of self-clamped search scheme (SC2S).

4 TCAM OPERATION BASED ON POST-RAM

4.1 Self-clamped search scheme

In this section, we introduce a SC2S that effectively enables TCAM operations using POST-RAM. In Table 1, the inputs and outputs for SC2S can be observed. However, it should be noted that in cases of match line sensing out (ML_{SO}) = V_{SS}, the voltage discharge levels differ depending on the ratio of miss cells to match cells due to the self-clamping mechanism. Additionally, it can be deduced that there is a logical relationship of XNOR between the input and output. Specifically as shown in Fig.6(a), the circuit for SC2S includes transmission gates controlled by the search line enable (SL_{EN}) signal positioned at each search line (SL). Additionally, it incorporates inputs for ML pre-charge (ML_{PRE}), sense amplifier enable (SA_{EN}), and outputs fo ML_{SO}. TCAM operations, as depicted in Fig.6(b), are

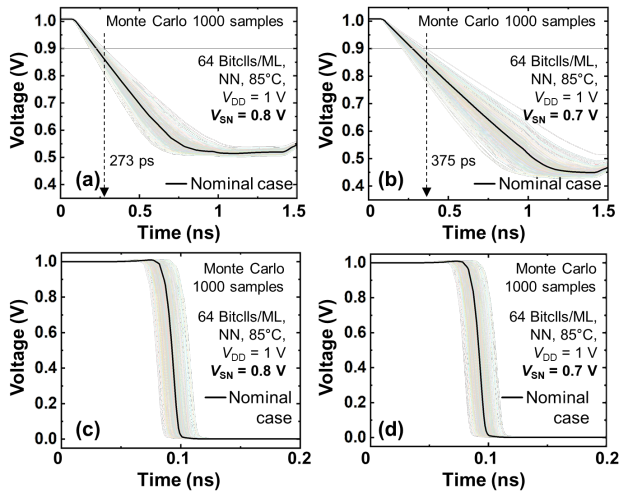

Figure 7: Monte Carlo simulation using 1000 samples for the time-ML voltage curve of TCAM composed of POST-RAM (a) 0 s after write (V_{SN} = 0.8 V), (b) 100 s after write(V_{SN} = 0.7 V). time - match line sensing out (ML_{SO}) voltage for (c) 0 s after write (V_{SN} = 0.8 V), (d) 100 s after write(V_{SN} = 0.7 V).

divided into two processes: word broadcasting and ML evaluation. In cases where the stored bits in each entry match the broadcasted search bits, the ML does not discharge, thus maintaining the pre-charge. Otherwise, when there is a mix of miss and match cells, both charging and discharging paths exist. However, since the read transistors of POST-RAM are composed of NMOS only, the charging path charges only up to V_{DD}- (NMOS V_{TH} (V_{TH-N})), resulting in self-clamping to V_{TH-N}. The concept of SC2S, exemplified in Fig. 6(c), involves a scenario in which out of 7 cells, only one cell is in a mismatched state. Matching SL cells charge ML, while non-matching ML cells discharge ML. The structure exhibits varying parallel resistance values based on the number of corresponding cells, consequently causing voltage distribution results depending on the final resistance values. ML voltage (V_{ML} equation is as follows, where nMatch is the number of match cells per unit word and nMiss is the number of miss cells:

$$V_{ML} = (V_{DD} - V_{TH-N}) \times \frac{1}{1 + \frac{nMatch}{nMiss}} \quad (1)$$

Consequently, by adopting SC2S without the need for a separate clamping circuit, a reliable sensing margin can be ensured. Furthermore, as the clamping voltage corresponds to the V_{TH} of the NMOS, setting the reference voltage (V_{REF}) of the sense amplifier to half of the V_{TH-N} enables reliable detection. In the experiments conducted in this paper, V_{REF} is set to 900 mV.

4.2 PVT verifications

To validate the TCAM operation of POST-RAM, a 64×64 array is designed, employing a voltage sense amplifier (VSA). From the perspective of the SC2S scheme, the worst-case scenario involves a single miss cell in an entry composed of 64 cells. Validating this case covers all scenarios, as it maximizes the time taken for the pre-charged ML to discharge. Verification conditions include V_{DD} = 1 V, and sensitivity to a temperature of 85°C is also considered. Through the experimental conditions, the robustness of POST-RAM to PVT is evaluated. Fig.7(a) and (b) represent cases where the storage node voltage (V_{SN}) is 0.8 and 0.7 V, respectively. These experiments validate the operation after writing 0 and 100 s [17], with the moment when ML starts to fall from V_{DD} being when SL_{EN} is applied. In conclusion, in Fig.7(b), SL_{EN} is applied, and within 400 ps, all samples fall below V_{REF} = 0.9 V, confirming that there are no operational issues. Furthermore, under the same conditions, the ML_{SO} voltage is monitored from the time SA_{EN} signal is enabled, as shown in Fig.7(c) and (d). In the worst-case scenario depicted in Fig.7(d), all output voltages fall to V_{SS} within approximately 110 ps after the input of the SA_{EN} signal. In conclusion, the operation of POST-RAM has been successfully verified even under worst-case scenarios within the retention time. However, considering that V_{SN} decreases over time, assuming a retention of 100 s, an additional timing margin of around 100 ps needs to be allocated. To apply POST-RAM to TCAM, it is essential to consider a design that balances retention time and timing margin.

4.3 Comparison with SRAM-based TCAM cell

POST-RAM distinguishes itself from other eNVMs in that only read transistors are present in the read path, and it is 100% logic compatible on the read-transistor side. Consequently, the only valid benchmark for comparison is with 16T-SRAM-based SRAM, representing a memory cell structure that can seamlessly replace SRAM-based TCAM. As shown in Table 2, PPA analysis conducted on the same 28-nm technology node highlights notable improvement in area efficiency. The area is reduced by more than 6 times compared to a 16T-based SRAM-based CAM cell, a result of the vertical stacking that achieves 2T footprint. In addition, a noticeable reduction in delay is also observed, with an approximately 40% decrease. This delay reduction is attributed to the reduced number of unit transistors present in the ML discharging path during search operations. However, POST-RAM adopts SC2S, requiring the forcing of search bit bias in SL, leading to the disadvantage of consuming more energy than SRAM-based cells for discharging the floating node. The energy difference is 15%, forming a trade-off relationship with area gains. By adding a pair of transistors and adopting the same ML technique as the 16T SRAM-based cell, it is possible to sufficiently overcome the energy difference. In conclusion, from the perspective of the energy-delay product (EDP), POST-RAM exhibits a 20% higher performance due to the greater delay gain. The EDP results indicate that there is a much broader range of process and electrical margins for future optimizations in POST-RAM.

5 FUTURE WORKS

The future research directions can be categorized into two main branches. Firstly, enhancing the performance of POST-RAM itself is a priority. POST-RAM, as a new memory cell structure, has many optimization points rooted in the process scheme. Additionally, conducting separate analysis of component structures related to write and read schemes is also a crucial aspect of the ongoing work. Secondly, there is a specific focus on applications in the field of AI. Considering the momentum of MANN, there is a significant demand for content-based deep learning acceleration. We anticipate that

Parameters		Unit	16T SRAM	POST-RAM
Technology node		nm	28	28
Search voltage		V	1	1
Cell area		μm^2	0.625 (6.5×)	0.095 (1.0×)
ML discharge	Energy	fJ/bit/search	1.95 (0.85×)	2.29 (1.0×)
	Delay	ps	396 (1.4×)	281 (1.0×)
	EDP	-	1.2×	1.0×
ML technique			Nor-type read	Self-clamping

Table 2: Comparison of power, performance, and area (PPA) parameters of SRAM-based cell and POST-RAM

AI applications based on POST-RAM could yield highly efficient computational results in this context.

6 CONCLUSION

In this paper, we propose the POST-RAM, and the implementation of the proposed device structure is rigorously validated through comprehensive 3D-TCAD simulations for process integration. We demonstrate the efficacy of SNPT by rotating transistors using IGZO and single-crystalline silicon as channels by 90°, enabling not only area savings but also securing a margin for retention. Furthermore, we implement TCAM operations using POST-RAM and introduce the SC2S, demonstrating superior TCAM performance. Monte Carlo simulations confirm its robustness against PVT variations, establishing it as an operation scheme with a rational timing margin. In conclusion, POST-RAM achieves a reduction of over 6 times in area compared to 16T SRAM-based nor-type TCAM, with a 20% enhancement in EDP. This underscores the ample potential of POST-RAM as the next-generation memory for TCAM applications.

ACKNOWLEDGEMENT

This work was supported in part by by the New Faculty Startup Fund from Seoul National University, Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00266, Development of Ultra-Low Power Low-Bit Precision Mixed-mode SRAM PIM, IITP-2023-RS-2023-00256081: artificial intelligence semiconductor support program to nurture the best talents), BK21 FOUR program at Seoul National University, and Inter-university Semiconductor Research Center at Seoul National University. The EDA tool was supported by the IC Design Education Center(IDEC). (Corresponding Author: Jae-Joon Kim).

REFERENCES

[1] S. Stas, "Associative processing with cams," in *Proceedings of NORTHCON'93 Electrical and Electronics Convention*, pp. 161–167, IEEE, 1993.

[2] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (cam) circuits and architectures: A tutorial and survey," *IEEE journal of solid-state circuits*, vol. 41, no. 3, pp. 712–727, 2006.

[3] G. Qin, S. Ata, I. Oka, and C. Fujiwara, "Effective bit selection methods for improving performance of packet classifications on ip routers," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 3, pp. 2350–2354, IEEE, 2002.

[4] H. J. Chao, "Next generation routers," *Proceedings of the IEEE*, vol. 90, no. 9, pp. 1518–1558, 2002.

[5] K. Ni, X. Yin, A. F. Laguna, S. Joshi, S. Dünkel, M. Trentzsch, J. Müller, S. Beyer, M. Niemier, X. S. Hu, *et al.*, "Ferroelectric ternary content-addressable memory for one-shot learning," *Nature Electronics*, vol. 2, no. 11, pp. 521–529, 2019.

[6] K. Nii, T. Amano, N. Watanabe, M. Yamawaki, K. Yoshinaga, M. Wada, and I. Hayashi, "13.6 a 28nm 400mhz 4-parallel 1.6 gsearch/s 80mb ternary cam," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 240–241, IEEE, 2014.

[7] O. Tyshchenko and A. Sheikholeslami, "Match sensing using match-line stability in content-addressable memories (cam)," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 1972–1981, 2008.

[8] I. Hayashi, T. Amano, N. Watanabe, Y. Yano, Y. Kuroda, M. Shirata, K. Dosaka, K. Nii, H. Noda, and H. Kawai, "A 250-mhz 18-mb full ternary cam with low-voltage matchline sensing scheme in 65-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 11, pp. 2671–2680, 2013.

[9] A. Razavieh, P. Zeitzoff, and E. J. Nowak, "Challenges and limitations of cmos scaling for finfet and beyond architectures," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 999–1004, 2019.

[10] X. Wang, L. Wang, Y. Wang, J. An, C. Dou, Z. Wu, X. Zhang, J. Liu, C. Zhang, Z. Yao, *et al.*, "A 4t2r rram bit cell for highly parallel ternary content addressable memory," *IEEE Transactions on Electron Devices*, vol. 68, no. 10, pp. 4933–4937, 2021.

[11] B. Song, T. Na, J. P. Kim, S. H. Kang, and S.-O. Jung, "A 10t-4mtj nonvolatile ternary cam cell for reliable search operation and a compact area," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 6, pp. 700–704, 2016.

[12] E. Garzón, M. Lanuzza, A. Teman, and L. Yavits, "Am 4: Mram crossbar based cam/tcam/acam/ap for in-memory computing," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 13, no. 1, pp. 408–421, 2023.

[13] J. Li, R. Montoye, M. Ishii, K. Stawiasz, T. Nishida, K. Maloney, G. Ditlow, S. Lewis, T. Maffitt, R. Jordan, *et al.*, "1mb 0.41 μm 2 2t-2r cell nonvolatile tcam with two-bit encoding and clocked self-referenced sensing," in *2013 Symposium on VLSI Technology*, pp. C104–C105, IEEE, 2013.

[14] L. Wang, W. Ye, J. Lai, J. Liu, J. Yang, X. Si, C. Huo, C. Dou, X. Xu, Q. Liu, *et al.*, "A 14nm 100kb 2t1r transpose rram with> 150x resistance ratio enhancement and 27.95% reduction on energy-latency product using low-power near threshold read operation and fast data-line current stabling scheme," in *2021 Symposium on VLSI Technology*, pp. 1–2, IEEE, 2021.

[15] C.-F. Yang, C.-Y. Wu, M.-C. Shih, M.-T. Yang, M.-H. Yang, Y.-T. Wu, T.-C. Chien, C.-W. Lai, S.-C. Tsai, W.-T. Chu, *et al.*, "Demonstration of high endurance capability on mega-bit rram macro and model of ppm level failures," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, pp. 318–319, IEEE, 2022.

[16] Y.-S. Chen, H.-Y. Lee, P.-S. Chen, P.-Y. Gu, W.-H. Liu, W.-S. Chen, Y.-Y. Hsu, C.-H. Tsai, F. Chen, M.-J. Tsai, and C. Lien, "Good endurance and memory window for Ti/HfO_x pillar rram at 50-nm scale by optimal encapsulation layer," *IEEE Electron Device Letters*, vol. 32, no. 3, pp. 390–392, 2011.

[17] A. Belmonte, H. Oh, S. Subhechha, N. Rassoul, H. Hody, H. Dekkers, R. Delhougne, L. Ricotti, K. Banerjee, A. Chasin, *et al.*, "Tailoring igzo-tft architecture for capacitorless dram, demonstrating> 10 3 s retention,> 10 11 cycles endurance and l g scalability down to 14nm," in *2021 IEEE International Electron Devices Meeting (IEDM)*, pp. 10–6, IEEE, 2021.

[18] M. Oota, Y. Ando, K. Tsuda, T. Koshida, S. Oshita, A. Suzuki, K. Fukushima, S. Nagatsuka, T. Onuki, R. Hodo, *et al.*, "3d-stacked caac-in-ga-zn oxide fets with gate length of 72nm," in *2019 IEEE International Electron Devices Meeting (IEDM)*, pp. 3–2, IEEE, 2019.

[19] K. Huang, X. Duan, J. Feng, Y. Sun, C. Lu, C. Chen, G. Jiao, X. Lin, J. Shao, S. Yin, J. Sheng, Z. Wang, W. Zhang, X. Chuai, J. Niu, W. Wang, Y. Wu, W. Jing, Z. Wang, J. Xu, G. Yang, D. Geng, L. Li, and M. Liu, "Vertical channel-all-around (caa) igzo fet under 50 nm cd with high read current of 32.8 μa/μm (vth + 1 v), well-performed thermal stability up to 120 ℃ for low latency, high-density 2t0c 3d dram application," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, pp. 296–297, 2022.

[20] M. Oota, Y. Ando, K. Tsuda, T. Koshida, S. Oshita, A. Suzuki, K. Fukushima, S. Nagatsuka, T. Onuki, R. Hodo, *et al.*, "3d-stacked caac-in-ga-zn oxide fets with gate length of 72nm," in *2019 IEEE International Electron Devices Meeting (IEDM)*, pp. 3–2, IEEE, 2019.

[21] W. Jeong, S. Maeda, H. Lee, K. Lee, T. Lee, D. Park, B. Kim, J. Do, T. Fukai, D. Kwon, *et al.*, "True 7nm platform technology featuring smallest finfet and smallest sram cell by euv, special constructs and 3 rd generation single diffusion break," in *2018 IEEE Symposium on VLSI Technology*, pp. 59–60, IEEE, 2018.

[22] K. Cheng, C. Park, H. Wu, J. Li, S. Nguyen, J. Zhang, M. Wang, S. Mehta, Z. Liu, R. Conti, *et al.*, "Improved air spacer co-integrated with self-aligned contact (sac) and contact over active gate (coag) for highly scaled cmos technology," in *2020 IEEE Symposium on VLSI Technology*, pp. 1–2, IEEE, 2020.

[23] S.-Y. Wu, C. Lin, M. Chiang, J. Liaw, J. Cheng, S. Yang, C. Tsai, P. Chen, T. Miyashita, C. Chang, V. Chang, K. Pan, J. Chen, Y. Mor, K. Lai, C. Liang, H. Chen, S. Chang, C. Lin, C. Hsieh, R. Tsui, C. Yao, C. Chen, R. Chen, C. Lee, H. Lin, C. Chang, K. Chen, M. Tsai, K. Chen, Y. Ku, and S. M. Jang, "A 7nm cmos platform technology featuring 4th generation finfet transistors with a 0.027um2 high density 6-t sram cell for mobile soc applications," in *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 2.6.1–2.6.4, 2016.

[24] N. Loubet, T. Hook, P. Montanini, C.-W. Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, *et al.*, "Stacked nanosheet gate-all-around transistor to enable scaling beyond finfet," in *2017 Symposium on VLSI Technology*, pp. T230–T231, IEEE, 2017.