A 40nm 48.7F^2/bit 1T2R Resistive Memory Bitcell Array For Adaptive Vector-Symbolic In-Memory Computing

Wenshuo Yue[†‡], Zhaokun Jing[†], Lintao Ye[‖], Tianyao Xiao[‖], Pek Jun Tiw[†], Yihan Fu[††], Yuxiang Yang[†], Mo Guang[††], Kaiwen Long[††], Daijing Shi[†‡], Hongxiao Zhao[†‡], Teng Zhang[†], Bonan Yan[†‡*], and Yuchao Yang[†‡§¶*]
[†]Beijing Advanced Innovation Center for Integrated Circuits, School of Integrated Circuits, Peking University, Beijing, China.
[‡]Institute for Artificial Intelligence, Peking University, Beijing, China. [††] Systems and Computing Group, Li Auto.
[‖] School of Electronics Engineering and Computer Science, Peking University, Beijing, China.
[§]Guangdong Provincial Key Laboratory of In-Memory Computing Chips, School of Electronic and Computer Engineering, Peking University, Shenzhen, China. [¶]Chinese Institute for Brain Research (CIBR), Beijing, China.
[*]Email: bonanyan@pku.edu.cn, yuchaoyang@pku.edu.cn

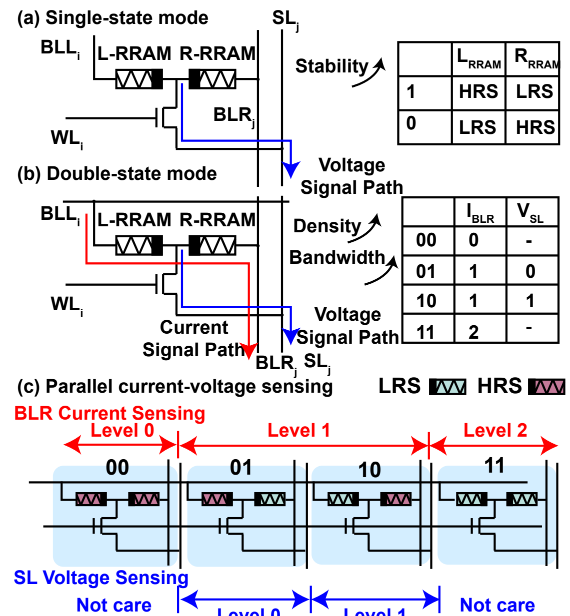
***Abstract*—This work introduces a 1T2R (1-Transistor-2-RRAM) resistive memory with high memory density, stability, and adaptive in-memory computing capabilities. Enabled by a parallel current-voltage sensing scheme, the fabricated memory attains a high cell density with the size of 48.7F^2/bit. The scheme enhances the high/low ratio by 1.29× and ensures stability across varying temperatures. The 1T2R array enables adaptable in-memory computing for different operations in vector-symbolic architectures, supporting both encoding and prediction phases while achieving 67.5%-71.0% energy savings and 22.0%-49.7% memory reduction compared to conventional approaches.**

I. Introduction

Neuro-symbolic artificial intelligence models, especially vector-symbolic architecture (VSA), operate on hyperdimensional vectors and demonstrate effective cognitive and reasoning abilities [1]. However, their high demand for memory capacity and bandwidth, as well as the support for various operations, including similarity calculation, vector-matrix multiplication (VMM), vector bundling, and vector binding, leads to excessive data access and substantial memory requirements. The growing need to store larger, more complex models on edge devices while maintaining robust computation poses significant challenges in developing high-density, adaptive on-chip memory and in-memory computing solutions.

Prior research highlights that Resistive Random-Access Memory (RRAM) achieves high memory density via its 1-Transistor-1-RRAM (1T1R) structure, yet conventional 1T1R single-level cell (SLC) and multi-level cell (MLC) designs are limited by transistor size requirements for read/write operations. While the 1TnR architecture enhances density, it exacerbates environmental impacts like temperature drift. Additionally, conventional RRAM arrays must improve operational adaptability with novel structures to minimize frequent data loading across functional cores in VSA implementations.

Some recent works propose the 1-Transistor-2-RRAM (1T2R) cell structure for memory and in-memory computing [2]. It utilizes the orthogonal BLR and BLL to attach the L-RRAM and R-RRAM separately. It shows a large improvement in the read margin, compared to the conventional 1T1R scheme. Moreover, it reduces state overlap by applying the voltage-sensing scheme. However, the former work only presents this scheme at a theoretical level. Also, it only supports memory mode and VMM operation.



	L_{RRAM}	R_{RRAM}
1	HRS	LRS
0	LRS	HRS

	I_{BLR}	V_{SL}
00	0	-
01	1	0
10	1	1
11	2	-

Fig. 1. Schematic of the proposed 1T2R resistive memory. (a) The single-bit mode uses an LRS-HRS pair to store 1 bit, with voltage sensing on SL for readout. (b) Double-state mode uses 4 states in a 1T2R cell to store 2 bits, employing voltage sensing on SL and current sensing on BLR for readout. (c) The 4 states (00, 01, 10, 11) represent HRS as 0 and LRS as 1.

This work proposes and fabricates a 1T2R resistive memory for high-density processing. It applies a parallel current-voltage sensing scheme to provide the double-state mode of 1T2R memory. It achieves 2bits/cell and 48.7F^2/bit. The voltage-sensing achieves 1.29× improvement in memory win-

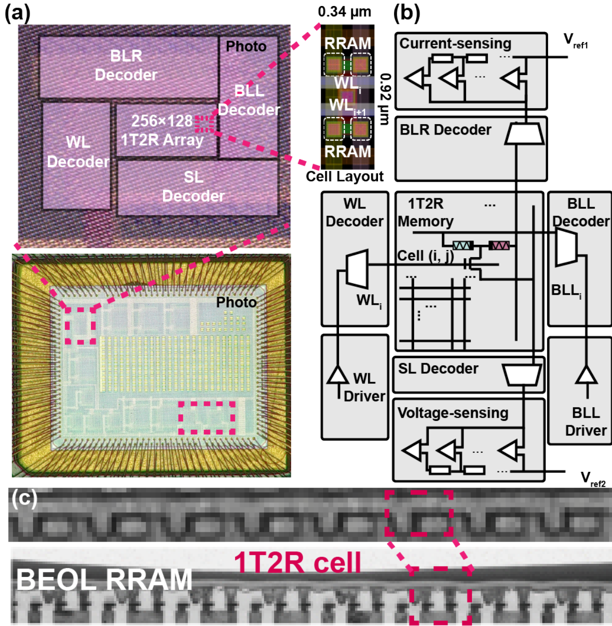


Fig. 2. Microscopic photography of the 1T2R RRAM array with peripheral circuits. (a) The 64Kb array with BLR, WL, SL, and BLL decoders. (b) Schematic of the proposed scheme. (c) Microscopic photography of the cells.

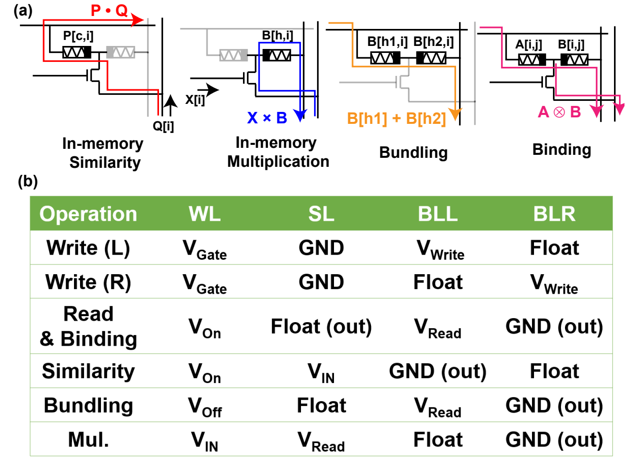
dow and enhances memory stability against temperature variation. Programming RRAM with MLC enabled the double-state mode to achieve a $2 \sim 2.5\times$ density improvement compared with 1T1R MLC. The support for locally computing various VSA operations, including similarity calculation, in-memory multiplication, bundling, and binding, with high memory density and bandwidth, provides an advantage for VSA.

II. 1T2R Device, Array, & Parallel I-V Sensing

A. Array design and characterization

As shown in Fig. 1a, a 1T2R cell consists of a single transistor and two RRAM devices (L-RRAM and R-RRAM) connected to its drain, with BLL and BLR lines arranged orthogonally. In conventional voltage-sensing, BLL applies the read voltage, and SL detects the voltage difference. This work introduces a parallel current-voltage (I-V) sensing scheme (Fig. 1b), encoding 2 bits per cell. Besides voltage-sensing via SL, the double-state mode adds current-sensing through BLR, enabling storage of HRS-HRS and LRS-LRS states. The scheme accurately identifies all four states (Fig. 1c): current sensing via BLR detects the states “00”, “01/10”, and “11”, while voltage sensing via SL distinguishes the states “01” and “10”. The voltage- and current-sensing can be executed in parallel to expedite 2-bit sensing latency.

We fabricated a chip-level 1T2R RRAM array with peripheral circuits, as illustrated in Fig. 2a, comprising the 1T2R array alongside BLR, WL, SL, and BLL decoders. The chip features multiple 1T2R RRAM arrays with sizes of 256×128 and so on. The double-mode memory integrates a voltage-sensing module to measure SL voltages and a current-sensing module on the BLR side to monitor currents. The BLL and WL drivers provide the necessary input voltages for read/write operations. Operating concurrently, these sensing modules double



Operation	WL	SL	BLL	BLR
Write (L)	V_{Gate}	GND	V_{Write}	Float
Write (R)	V_{Gate}	GND	Float	V_{Write}
Read & Binding	V_{On}	Float (out)	V_{Read}	GND (out)
Similarity	V_{On}	V_{IN}	GND (out)	Float
Bundling	V_{Off}	Float	V_{Read}	GND (out)
Mul.	V_{IN}	V_{Read}	Float	GND (out)

Fig. 3. Adaptive storage and operations in 1T2R structure. (a) The schematic of realizing in-memory similarity computation, in-memory multiplication, bundling, and binding with 1T2R cell. (b) The detailed computing diagram.

both memory density and word length (memory bandwidth). Prior work has demonstrated that the 1T2R array design effectively suppresses the sneak path [2].

The RRAM devices are fabricated with a back-end-of-line (BEOL) process on a standard 40nm CMOS platform. The whole 1T2R cell area is 97.8F^2, as the layout proved in the former work [2]. With the help of the proposed double-state memory scheme, this work reaches the compact 48.7F^2/bit, indicating high memory density. The top-view and the side-view microscopic photos prove the fabricated structure (Fig. 2c). The 1T2R structure, featuring orthogonal BLL and BLR, enables diverse operations. Current summation on BLL and BLR facilitates vector similarity and VMM, respectively (Fig. 3a). The cascade of two RRAM devices implements bundling, with two bundling matrices stored locally within the same array, and the double-state read mode inherently supports vector binding.

The 1T2R array programming and resistance measurement follow the 1T1R methodology. Writing/reading L-RRAM uses the BLL-SL path (1T1R behavior), while R-RRAM uses BLR-SL path. Measured 1T1R resistance distributions show a favorable H/L ratio (Fig. 4a). The voltage-sensing mode on 1T2R cells significantly enhances the memory window (Fig. 4b). Furthermore, the resistance states in the 1T2R cells fully support the double-state memory scheme. As illustrated in Fig.4c, under a fixed read voltage, the current distribution of the signal path distinctly separates the HH, LH/HL, and LL states. Combined with the voltage sensing distribution, the results confirm that the double-state memory scheme effectively doubles both memory density and bandwidth.

B. High/Low ratio improvement and stability

The High/Low (H/L) ratio of the conventional 1T1R scheme is defined as $\frac{R_T+R_{HRS}}{R_T+R_{LRS}}$, where R_T represents the effective resistance of the transistor, and R_{HRS} and R_{LRS} denote the resistances of the high-resistance state and low-resistance state, respectively. The transistor resistance diminishes the H/L ratio. In contrast, the H/L ratio of 1T2R voltage-sensing mode is defined as $1 + \frac{|V-V_{mid}|}{V_{mid}}$. Here, V_{mid} is $V_{Read}/2$

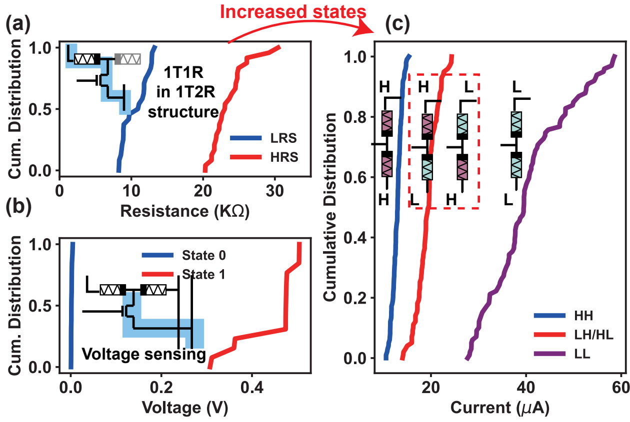


Fig. 4. Cumulative distribution function (CDF) of experimental measurements on 1T2R cells. (a) Resistance distributions of each 1T1R cell in 1T2R, measured via current sensing, including RRAM (HRS/LRS) and transistor resistances. (b) Readout voltage distributions of 1T2R cells, demonstrating an enlarged memory window and sensing range. (c) Readout current states of the 1T2R double-state mode with assumed $\sim 4k\Omega$ transistor resistance.

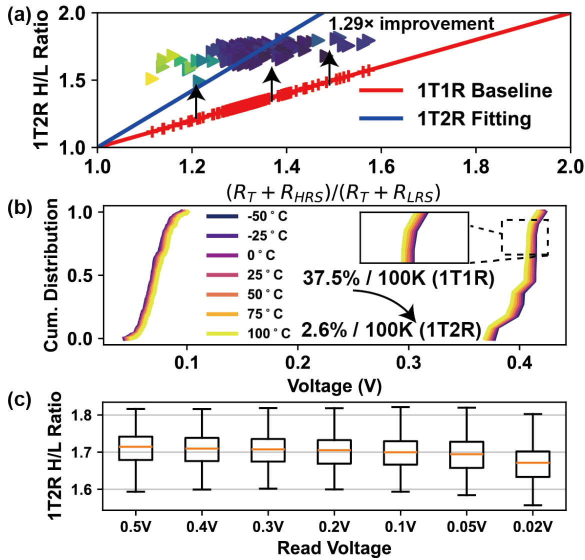


Fig. 5. The experimentally measured H/L ratio and its stability. (a) An average 1.29× improvement with using 1T2R cells, compared with conventional 1T1R. (b) The measured voltage-sensing results under various temperature conditions. (c) The measured H/L ratio with various read voltages.

and V is $V_{Read} \times \frac{R_{HRS}}{R_{HRS}+R_{LRS}}$ or $V_{Read} \times \frac{R_{LRS}}{R_{HRS}+R_{LRS}}$. This formulation ensures a H/L ratio of 1 when $R_{HRS} = R_{LRS}$. This memory scheme significantly mitigates the impact of transistor resistance, which otherwise reduces the H/L ratio. To assess the H/L ratio enhancement of 1T2R memory, RRAM devices were deliberately programmed to low initial H/L ratios (1.2-1.6). As demonstrated in Fig. 5, the 1T2R array's voltage sensing scheme elevates this ratio to 1.5-1.8, yielding a 1.29× average improvement over 1T1R configurations.

The resistance states of 1T1R RRAM exhibit temperature-induced drift [3], introducing errors in RRAM-based edge devices operating in diverse environments. The 1T2R structure mitigates this effect by leveraging its dual devices for compensation. Across a temperature range of -50°C to 100°C, the H/L ratio of the 1T2R cell remains stable (Fig. 5b, c), reducing temperature variation from 37.5%/100K in 1T1R [4] to 2.6%/100K. These results demonstrate the enhanced robustness of the 1T2R structure under varying thermal conditions.

III. VSA Computation with 1T2R Array

VSA classifies data by assessing the similarity between queried and stored features. During inference, it encodes input data into a query hyperdimensional vector (HV), compares it with stored HVs of various classes, and outputs the class with the highest similarity. The 1T2R supports VSA through adaptive in-memory computing and double-state mode for local memory storage. The encoding phase involves random basis vector generation, binding, and bundling/in-memory multiplication, with basis vectors stored in the item memory (IM). The comparison phase focuses on in-memory similarity calculation, with HVs stored in the associated memory (AM).

The array generates random basis vectors leveraging inherent device-to-device variation. During training, both IM and AM are stored. Input data are processed locally through binding and bundling, as illustrated in Fig. 3a. In the comparison phase, the query is sent to AM for in-memory similarity calculation. The 1T2R RRAM array provides a compact VSA solution by storing and processing both IM and AM or other processing data in the same array, reducing memory overhead (Fig. 6a). Its double-state mode stores IM in the right RRAM and AM in the left RRAM within a single cell. This eliminates inter-array communication, significantly lowering data loading costs for long vectors (Fig. 6b) and reducing energy consumption for read, bundling, and binding by 67.5% $\sim$ 71.0%. In actual applications, VSA may also require the computation of neural networks, HV permutation, etc. Considering these facts, graph classification (MUTAG, PROTEIN) and evaluations on MNIST demonstrate its low bit-cell area and overall memory cost savings (Fig. 6c, d).

IV. Discussion on MLC-1T2R Configuration

Previous studies have employed MLC to increase memory density in 1T1R RRAM [5]. Integrating the double-state mode

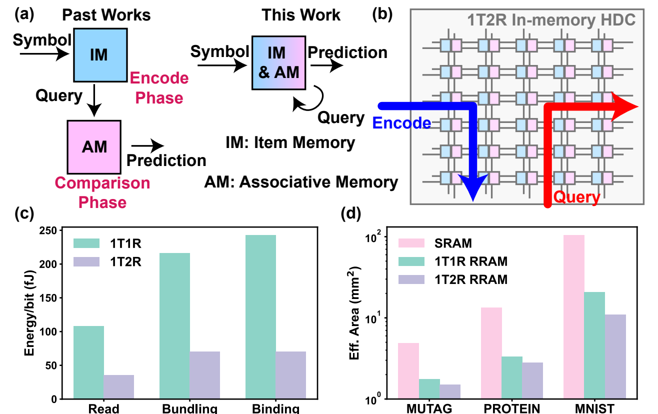


Fig. 6. Evaluations on vector symbolic architecture. (a) Using 1T2R structure, this work integrates item memory (IM) and associative memory (AM) into a single RRAM array. (b) The encode and query happen in the same array. (c) The energy consumption per bit of read, bundling, and binding with 1T1R and 1T2R structure (40ns). (d) The effective area cost of implementing different vector-symbolic architecture tasks with SRAM, 1T1R, and 1T2R RRAM.

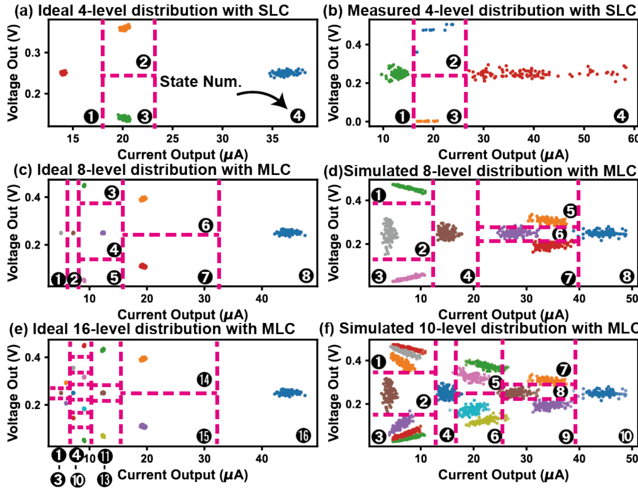


Fig. 7. Discussion on 1T2R memory with SLC and MLC RRAM devices and parallel current-voltage sensing. (a) Ideal I-V distribution of 4 states (labeled) in a 1T2R SLC cell. (b) Measured 4 states in a 1T2R SLC cell, with measured voltages, and assumed $\sim 4k\Omega$ transistor resistance for currents. (c) Ideal I-V distribution of 8 states in a 1T2R MLC cell. (d) Simulated 8-state distribution with RRAM resistance; dotted lines show state separation thresholds. (e) Ideal I-V distribution of 16 states with two 2-bit RRAM devices. (f) 16 combinations were simulated, supporting up to 10 separable states.

with 2-bit MLC RRAM could further enhance density and enable more compact VSA computing. Fig. 7a, b compares the current-voltage distribution of 1T2R cells with binary-resistance devices, contrasting ideal linear states (negligible variation) with measured nonlinear states (realistic variation). This work primarily programs devices to SLC, though the resistance design space demonstrates 1T2R's potential for MLC when simulated with realistic non-ideal 1T1R cells.

The double-state mode, as illustrated in Fig. 7c,d, enables at least 8 distinguishable states through specific L-RRAM and R-RRAM state combinations, effectively doubling memory density. Each state is uniquely identifiable via parallel I-V sensing of distinct current-voltage pairs. Furthermore, a 1T2R cell incorporating 2-bit MLC RRAM devices can achieve more than double the memory density. The total 16 states, accounting for variation, support up to 10 separable states per cell (Fig. 7e, f), which is $2.5\times$ states than SLC 1T2R. Enabled by the parallel sensing scheme, these 10 states can be accessed simultaneously in a single readout cycle. In an ideal case, all 16 states within a single 1T2R cell could be distinguished, achieving a $4\times$ enhancement in memory density.

V. Conclusion

The 1T2R resistive memory achieves a memory density of 48.7F^2/bit, with its double-state mode, enabled by parallel current-voltage sensing, doubling both density and bandwidth for binary and multi-level devices (Table I). Voltage sensing enhances the High/Low ratio by $1.29\times$ and ensures thermal stability. The 1T2R array is particularly suited for VSA, supporting diverse operations, increasing density and bandwidth, and reducing computational energy and memory costs.

TABLE I
Comparison with other RRAM memory at aligned access time.

	[6]	[7]	[8], [9]	[10]	This work
Structure	1T1R	1T1R	1T2R	1T2R	1T2R
Cell Area (F^2/bit)	60.3	172.9	67.0	64.7	48.7
MLC Area (F^2/bit)	30.2	86.5	33.5	-	29.4
Port	Single port	Single port	Single port	Single port	Dual port
Technology	40nm	12nm	28nm	28nm	40nm
Read Cell Energy (fJ/bit)	-	-	23	-	15.5
Sensing Scheme	Current	Current	Current	Voltage	Parallel I-V
Adaptive Store	No	No	No	No	Yes

Acknowledgment

This work was supported by National Key R&D Program of China (2023YFB4502200), National Natural Science Foundation of China (61925401, 92064004, 61927901, 92164302, 92364102, T2350006), Guangdong Provincial Key Laboratory of In-Memory Computing Chips (2024B1212020002), Shenzhen Science and Technology Program (JCYJ20241202125907011), Beijing Natural Science Foundation (L234026) and 111 Project (B18001). This work is sponsored by Beijing Nova Program.

References

[1] M. Hersche *et al.*, "A neuro-vector-symbolic architecture for solving Raven's progressive matrices," *Nature Machine Intelligence*, vol. 5, no. 4, pp. 363–375, 2023.

[2] Z. Jing *et al.*, "VSDCA: a voltage sensing differential column architecture based on 1T2R RRAM array for computing-in-memory accelerators," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 10, pp. 4028–4041, 2022.

[3] Y. Ling *et al.*, "Temperature-dependent accuracy analysis and resistance temperature correction in RRAM-based in-memory computing," *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 294–300, 2023.

[4] C. Walczyk *et al.*, "Impact of temperature on the resistive switching behavior of embedded HfO_2-based RRAM Devices," *IEEE Transactions on Electron Devices*, vol. 58, no. 9, pp. 3124–3131, 2011.

[5] W. Li *et al.*, "A 40-nm MLC-RRAM compute-in-memory macro with sparsity control, on-chip write-verify, and temperature-independent ADC references," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 9, pp. 2868–2877, 2022.

[6] S. D. Spetalnick *et al.*, "A 2.38 MCells/mm^2 9.81-350 TOPS/W RRAM compute-in-memory macro in 40nm CMOS with hybrid offset/I_{OFF} cancellation and $I_{CELL}R_{BLSL}$ drop mitigation," in *2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*. IEEE, 2023, pp. 1–2.

[7] Y.-C. Huang *et al.*, "A 32Mb RRAM in a 12nm FinFet Technology with a 0.0249 μm^2 bit-cell, a 3.2 GB/S Read Throughput, a 10KCycle Write Endurance and a 10-Year Retention at 105°C," in *2024 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 67. IEEE, 2024, pp. 288–290.

[8] J. Yang *et al.*, "A 28nm 1.5Mb Embedded 1T2R RRAM with 14.8 Mb/mm^2 using Sneaking Current Suppression and Compensation Techniques," in *2020 IEEE Symposium on VLSI Circuits*, 2020, pp. 1–2.

[9] Y. Cao *et al.*, "A 67F^2 Reconfigurable PUF Using 1T2R RRAM Switching Competition in 28nm CMOS with 5e-9 Bit Error Rate," in *2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*. IEEE, 2024, pp. 1–2.

[10] K. Zhou *et al.*, "An energy efficient computing-in-memory accelerator with 1T2R cell and fully analog processing for edge ai applications," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 8, pp. 2932–2936, 2021.