SCIENCE CHINA
Information Sciences

• LETTER •

March 2025, Vol. 68, Iss. 3, 139402:1–139402:2
https://doi.org/10.1007/s11432-024-4253-9

3D self-rectifying memristive ternary content addressable memory for massive and exact in-memory search

Yingjie YU[1†], Shengguang REN[1†], Ling YANG[1], Yi LI[1,2*] & Xiangshui MIAO[1,2]

[1]*School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Hubei Yangtze Memory Laboratories, Wuhan 430205, China*

Received 9 July 2024/Revised 3 September 2024/Accepted 13 December 2024/Published online 12 February 2025

Citation Yu Y J, Ren S G, Yang L, et al. 3D self-rectifying memristive ternary content addressable memory for massive and exact in-memory search. Sci China Inf Sci, 2025, 68(3): 139402, https://doi.org/10.1007/s11432-024-4253-9

Ternary content addressable memory (TCAM) enables high-speed parallel in-situ pattern matching [1], serving as the key processing unit for an efficient in-memory search system, which is widely studied in areas such as routing, similarity computing, and brain-inspired learning. However, traditional TCAM based on static random access memory encounters challenges in terms of low cell density and high static energy consumption due to its complex structure and volatility. To address the above issues, transistorless non-volatile TCAMs have emerged due to their high density and potential for three-dimensional (3D) integration. In this regard, one promising non-volatile device is a self-rectifying memristor (SRM) [2]. Enhancing search parallelism in TCAM arrays is equally vital for applications requiring massive search. Nevertheless, increasing search parallelism results in a reduced sense margin (SM), which is crucial for exact search. This inherent conflict poses a limitation on the search parallelism of TCAM, particularly for low-SM TCAM, such as two-memristor (2R) TCAM [3].

In this study, a 3D SRM-TCAM with cross-layer parallel search capability was proposed, which shows N-fold improvement of density and search parallelism (N represents the number of 3D stacking layers). Low search delay (1.86 ns) and search energy (0.185 fJ/bit/search) are achieved. By virtue of the intrinsic self-rectifying property of SRM, our TCAM alleviates the inherent conflict between SM and search parallelism, overcoming the constraints of conventional 2R TCAM that struggles to perform massive parallel searches. With circuit-level simulations, the 3D SRM-TCAM can support massive and exact searches with word lengths of up to 1024 bits. Our work provides an area-efficient, low-power, and highly parallel solution for future 3D in-memory search systems.

The structure of the 3D SRM-TCAM array is illustrated in Figure 1(a). The proposed SRM-TCAM consists of two SRM devices with their bottom electrodes connected to the match line (ML) (Figure 1(b)). The stored data are defined by the resistance state of the two SRM devices (R_0, R_1). The [high resistance state (HRS), low resistance state (LRS)] represents 1, the (LRS, HRS) represents 0, and the (HRS, HRS) represents X, which is known as "do not care" state. The input voltage pulses (SL, SLB) denote the query data, where (V_S, 0), (0, V_S), and (0, 0) are represented as 1, 0, and X, respectively. The simple TCAM cell structure allows the TCAM array to be implemented in a 3D SRM crossbar array, enabling parallel exact search (Figure 1(c)).

To prevent search errors, it is essential for the voltage difference between 1-bit mismatch and match in ML, known as SM, to be sufficiently large for the sense amplifier (SA) to detect [4]. The ML voltage developing during the search for 2R-TCAM and SRM-TCAM is analyzed. The 2R-TCAM search can be likened to the ML charging circuit as shown in Figure 1(d). During the search operation, where the load capacitance C_L is contributed by the parasitic capacitance of the devices and the input transistor of SA. As two voltage pulses V_S and ground are applied simultaneously to the TCAM array, a charging path from V_S and a discharging path from ground to the ML are established. However, the presence of the discharge path leads to insufficient ML charging, resulting in a low SM. In contrast, for SRM-TCAM, the SRM is functionally equivalent to a one-diode-one-resistor (1D1R) configuration, which possesses a self-rectifying characteristic that prevents the discharging current to ground, as depicted in Figure 1(e). Consequently, compared to 2R-TCAM, the SRM-TCAM allows for a more adequate charging current, leading to a larger SM.

The 3D SRM array of our 3D TCAM exhibits a split-cell structure, which features high integration density and low bit cost. Figure 1(f) shows the top view of this fabricated 3D array and the cross-sectional transmission electron microscopy (TEM) images. Fabrication details can be referred to in [5]. Typical I-V curves of the Pt/TaO_X/Ta SRM in 3D array including self-rectifying behavior are presented in Figure 1(g). The significant self-rectifying effect is attributed to

* Corresponding author (email: liyi@hust.edu.cn)
† These authors contributed equally to this work.

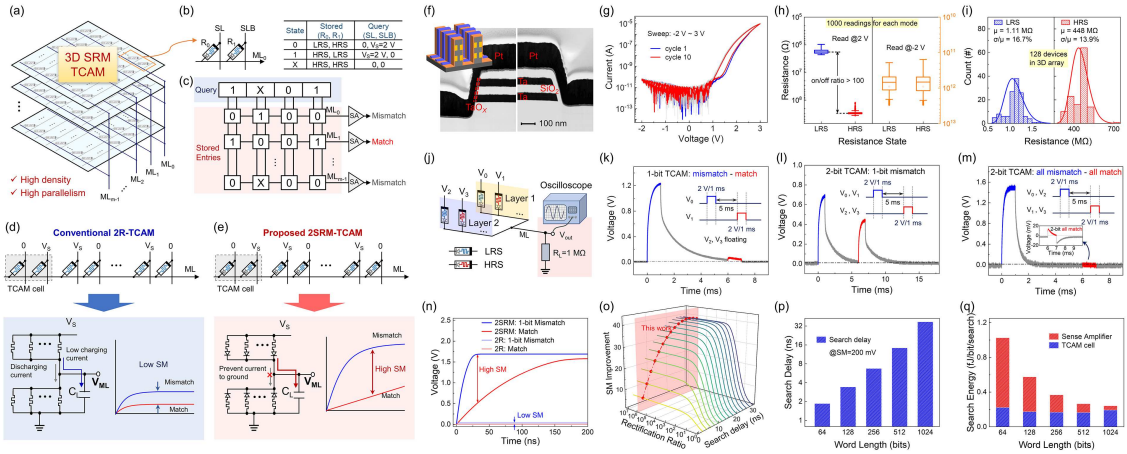

Figure 1 (Color online) (a) Schematic illustration of the 3D SRM-TCAM array; (b) cell structure and state definitions of a 1-bit SRM-TCAM; (c) diagram of SRM-TCAM arrays for performing the exact match; ML charging circuit models for (d) 2R-TCAM and (e) SRM-TCAM; (f) TEM images to distinguish the sidewalls with and without SiO_2, inset: top view of the fabricated 3D SRM array; (g) typical I-V curves of the SRM; (h) 10^3 readings for both LRS and HRS at 2 and −2 V; (i) LRS and HRS distributions for 128 SRMs; (j) measurement setup for our 2-layer 3D SRM-TCAM array; (k) match and mismatch cases of ML response waveform for 1-bit TCAM; (l) 1-bit mismatch, as well as (m) all match and all mismatch cases of ML response waveforms for 2-bit TCAM; (n) transient simulation of 1-bit mismatch and match cases with 2R-TCAM and SRM-TCAM for 64-bit word length search data; (o) SM improvement of SRM-TCAM over 2R-TCAM with different RRs and search delays; (p) search delays when the SM reaches 200 mV for different word lengths; (q) search energy consumption of SRM-TCAM with different word lengths.

the high and asymmetric Schottky barriers between TaO_X and the electrodes on both sides. In Figure 1(h), to further validate the reverse rectification characteristics of our SRM, we performed 10^3 readings with +2 and −2 V read voltages for devices in both LRS and HRS. Stable >100 on/off ratio and TΩ-level reverse reading resistance (rectification ratio, RR $\sim 10^6$) can be observed. Figure 1(i) presents the LRS and HRS distributions of 128 devices in the 3D array. The device-to-device (D2D) variations (σ/μ, σ is the mean square error and μ is the mean value) are calculated to be 16.7% for LRS and 13.9% for HRS.

To verify the 3D cross-layer parallel search capability, two SRMs from each layer of the 2-layer 3D array are selected for 1-bit and 2-bit TCAM tests, as shown in Figure 1(j). The BLs of the 4 SRMs are applied with voltages V_0, V_1, V_2, and V_3, respectively, while the WLs are connected to the ML. The ML voltage is acquired by an oscilloscope. For 1-bit TCAM search, V_2 and V_3 are floating while V_0 and V_1 are applied with (V_S, 0) and (0, V_S) for the mismatch and match cases, respectively (V_S: 2 V/1 ms). The V_0 and V_1 timing diagrams and the corresponding ML voltage responses are shown in Figure 1(k). For the 2-bit TCAM search, the corresponding timing diagram and ML voltage response are shown in Figures 1(l) and (m).

To evaluate the performance of our 3D SRM-TCAM in large-scale search, the SRM is modeled and simulated at the circuit level considering RR and D2D variation based on the experimental data from 128 devices. Figure 1(n) illustrates the ML voltage waveform when searching for 64-bit data using 3D SRM-TCAM and 2R-TCAM, respectively. The SM improvement of SRM-TCAM over 2R-TCAM considering RR and search delay change is presented in Figure 1(o). The current to ground can be prevented when RR is ample ($>10^3$) and the SM improvement of SRM-TCAM over 2R-TCAM peaks at a search delay of 20 ns, exhibiting a 39-fold improvement. The search delay of different word lengths when SM is 200 mV is shown in Figure 1(p). The relationship between search delay and word length appears to be approximately linear. Figure 1(q) presents the energy consumption by the SA and TCAM cells for different word lengths. The simulation of energy consumption considers the SA sensing time and operating frequency. By calculation, the proposed 3D SRM-TCAM excels in search delay (1.86 ns) and search energy (0.185 fJ/bit/search), as illustrated in Table C1 in Appendix C.

In summary, we have demonstrated a compact and low-power 3D SRM-TCAM with $8/N$ F^2 cell area and 0.22 fJ/bit/search of search energy. Experiments demonstrated it can perform a 3D cross-layer parallel search with N-fold improvement in density and search parallelism. Benefiting from the self-rectifying property of SRM, our TCAM alleviates the conflict between SM and search parallelism. By large-scale array and circuit simulation, the proposed TCAM features higher SM compared to 2R-TCAM, supporting it to perform massive parallel (up to 1024 bits) and exact search. Our work provides a promising solution for massive search applications.

Acknowledgements This work was supported by National Key Research and Development Plan of MOST of China (Grant No. 2022YFB4500101), Natural Science Foundation of Hubei Province (Grant Nos. 2024AFA043, 2023AFB335), and Fundamental Research Funds for the Central Universities (Grant No. HUST:5003190012).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Pagiamtzis K, Sheikholeslami A. Content-addressable memory (CAM) circuits and architectures: a tutorial and survey. IEEE J Solid-State Circuits, 2006, 41: 712–727
2 Ren S G, Dong A W, Yang L, et al. Self-rectifying memristors for three-dimensional in-memory computing. Adv Mater, 2024, 36: 2307218
3 Yang L, Zhao R, Li Y, et al. In-memory search with phase change device-based ternary content addressable memory. IEEE Electron Device Lett, 2022, 43: 1053–1056
4 Yang R, Li H, Smithe K K H, et al. Ternary content-addressable memory with MoS_2 transistors for massively parallel data search. Nat Electron, 2019, 2: 108–114
5 Ren S G, Xue Y B, Zhang Y, et al. 3D vertical self-rectifying memristor arrays with split-cell structure, large nonlinearity ($>10^4$) and fJ-level switching energy. IEEE Electron Device Lett, 2023, 44: 2059–2062