nature electronics

SUPPLEMENTARY INFORMATION

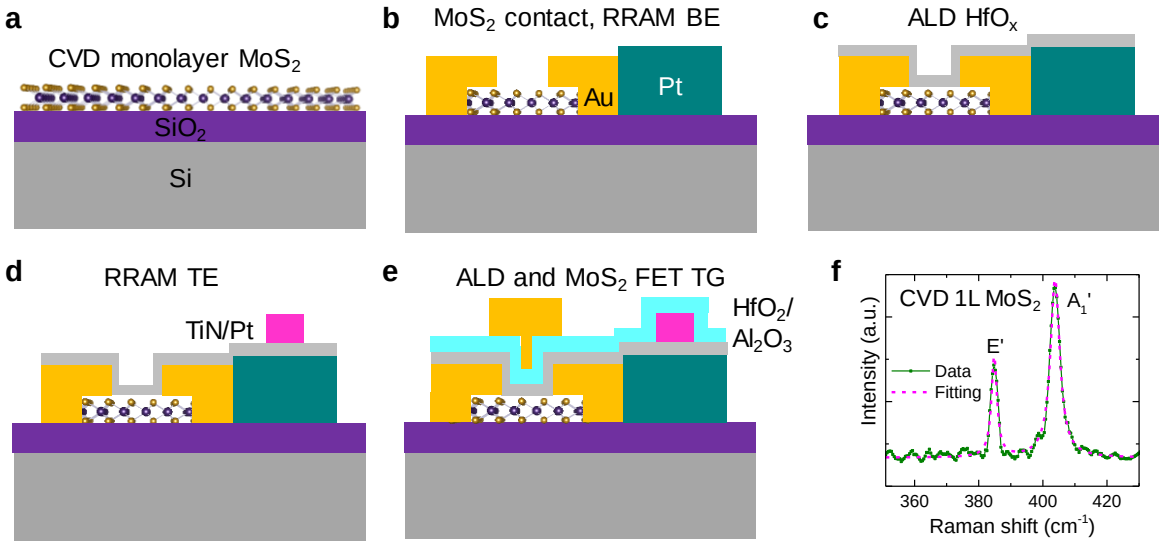
https://doi.org/10.1038/s41928-019-0220-7

In the format provided by the authors and unedited.

Ternary content-addressable memory with MoS_2 transistors for massively parallel data search

Rui Yang[1,2*], Haitong Li[1], Kirby K. H. Smithe[1], Taeho R. Kim[3], Kye Okabe[1], Eric Pop[1,3,4], Jonathan A. Fan[1] and H.-S. Philip Wong[1]

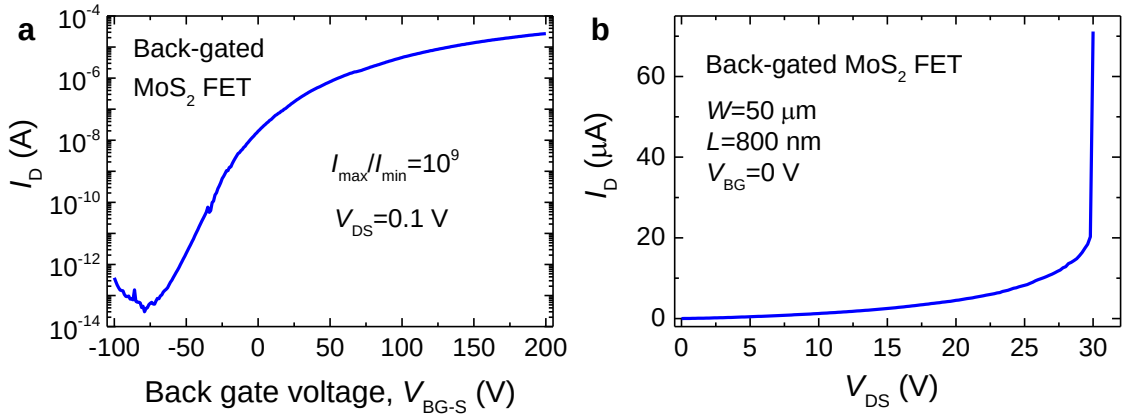
[1]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [2]University of Michigan–Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China. [3]Department of Materials Science and Engineering, Stanford University, Stanford, CA, USA. [4]Precourt Institute for Energy, Stanford University, Stanford, CA, USA. *e-mail: rui.yang@sjtu.edu.cn



Supplementary Fig. 1. Fabrication process of the 2T2R transition metal dichalcogenide (TMD) ternary content-addressable memory (TCAM) cell. a, Chemical vapor deposition (CVD) of monolayer MoS_2 on SiO_2/Si substrate. **b**, Etching the molybdenum disulfide (MoS_2), forming metal contact to the MoS_2, and patterning the bottom electrode (BE) of the resistive random-access memory (RRAM). **c**, Atomic layer deposition (ALD) of HfO_x. **d**, Patterning the top electrode (TE) of the RRAM. **e**, ALD of the top gate dielectric, and patterning the top gate (TG) of the MoS_2 field-effect transistor (FET). **f**, Raman spectrum of the CVD monolayer (1L) MoS_2, showing peak separation of 19.7 cm^{-1} which confirms that the MoS_2 is 1L. Fitting to the Lorentzian is shown by the dashed lines.

Supplementary Note 1. Characterization of the back-gated MoS_2 FETs

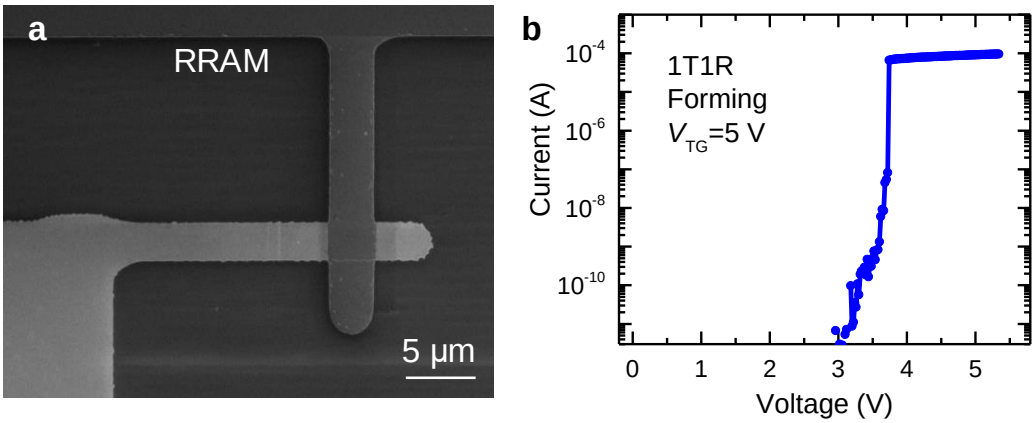
We perform measurements on the back-gated MoS_2 FETs, to examine the performance of the MoS_2 FETs without any effect from the top gates. The back-gated MoS_2 FETs show on/off ratio up to 10^9, which represents the potential of these MoS_2 FETs for low-leakage applications [S1,S2,S3] (Supplementary Fig. 2a). The higher on/off ratio than top-gated MoS_2 FETs indicate n-doping during top gate fabrication, which can be further improved toward top-gated MoS_2 FETs with even lower leakage and TCAM cells with even higher resistance ratio (R-ratio) between match and mismatch states. The I_D – V_{DS} characteristics in Supplementary Fig. 2b shows that the electrical breakdown voltage of these MoS_2 FETs (channel length L = 0.8 μm) can be as high as ~30 V, which shows great potential for further scaling down the MoS_2 channel while still successfully driving the RRAMs.



Supplementary Fig. 2. Characterization of the back-gated MoS_2 FETs. a, I_D – V_{BG-S} curves showing high on/off ratio. **b**, I_D – V_{DS} characteristics at large V_{DS}, until dielectric breakdown occurs.

Supplementary Note 2. 1-transistor-1-resistor (1T1R) forming of the RRAMs

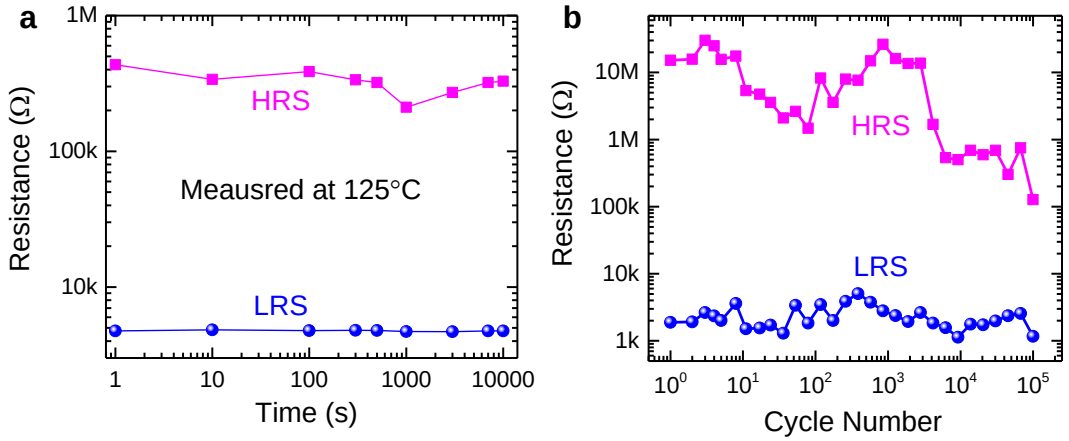
The metal-oxide RRAMs typically require a forming process to form the initial conductive filament, which uses larger voltage than the set and reset processes [S4]. The scanning electron microscopy (SEM) image in Supplementary Fig. 3a shows the cross-bar structure RRAM. The forming process using 1T1R scheme is shown in Supplementary Fig. 3b, where the monolayer MoS_2 FET sets the current compliance.



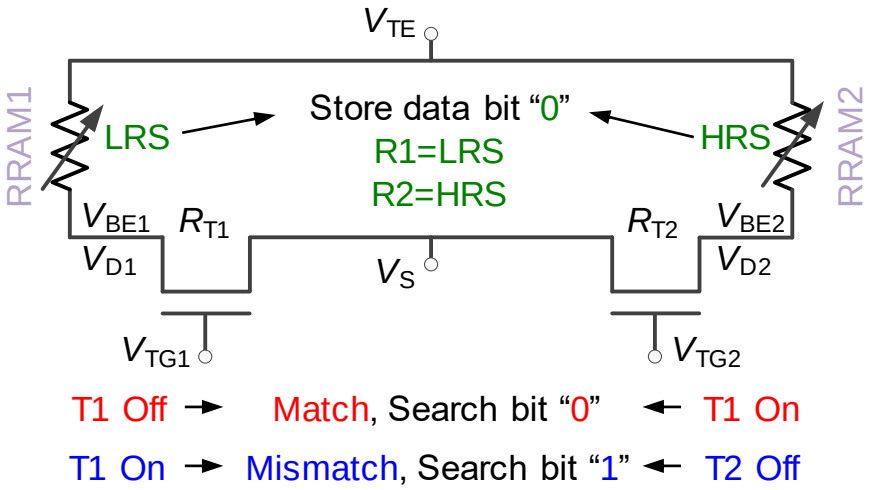
Supplementary Fig. 3. Initial characterization of the RRAMs. a, SEM image of the crossbar structure RRAM. **b**, The forming process of a RRAM using 1T1R configuration, with 5 nm HfO_x and 7 nm Al_2O_3 as top gate dielectric, and at top gate voltage (V_{TG}) of 5 V.

Supplementary Note 3. Retention and cycling of the RRAMs

The retention of the RRAMs measured at 125ºC up to 10,000 s is shown in Supplementary Fig. 4a, which confirms that the resistance states of these RRAMs are very stable over time, and thus are suitable for TCAM applications where the TCAM cells will be repeatedly read for long time. Since the TCAM application is read-dominant, the number of cycles to re-program the RRAMs is not critical [S5]. Nevertheless, we use voltage pulses to set and reset the RRAMs for up to 10^5 cycles (Supplementary Fig. 4b), which meets the requirements for most TCAM applications.



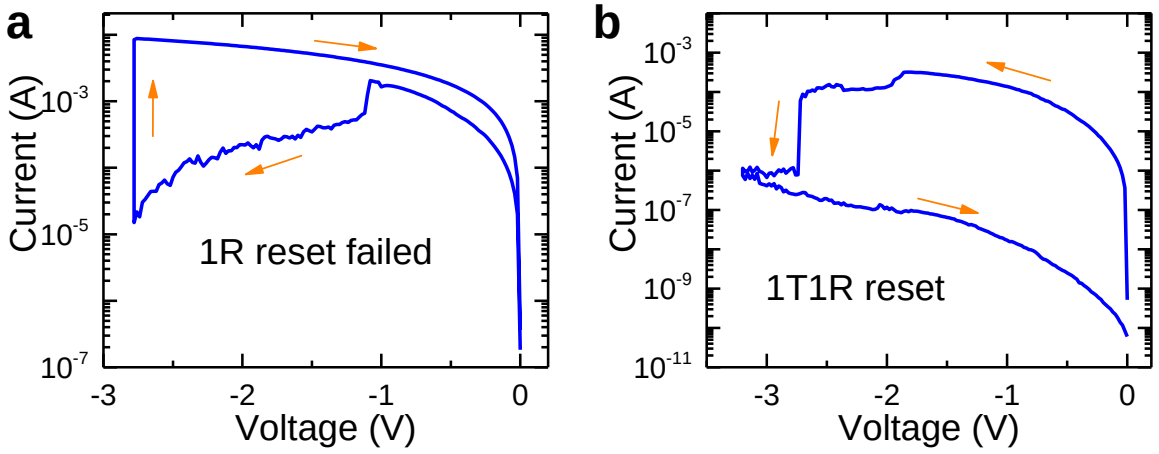
Supplementary Fig. 4. Further characterization of the RRAM properties. a, Retention of the RRAM measured using RRAM only (1R) scheme at 125ºC. **b**, Cycle of RRAM measured using 1R scheme, using programming voltage pulse width of 200 ns.



Supplementary Fig. 5. Circuit diagram of a TCAM cell which stores data bit "0".

Supplementary Note 4. Comparison of reset using 1R and 1T1R

By using 1T1R scheme to perform the reset, the reset voltage on the RRAM is typically higher than that when using 1R scheme, because the RRAM and the on-state MoS_2 FET forms a voltage divider. Applying high voltage during RRAM reset in 1R scheme could result in the device failure due to hard breakdown of the dielectric layer, and become conductive again, as shown in Supplementary Fig. 6a. By using the 1T1R scheme, the MoS_2 FET sets the current compliance of the RRAM, thus limits the current supplied to the RRAM and prevents the electrical breakdown at large voltages [S6]. As shown in Supplementary Fig. 6b, when using 1T1R scheme, although the voltage sweeps to -3.2 V using 1T1R scheme, and the voltage drops almost totally on the RRAM, the RRAM remains in the high resistance state (HRS). While for the 1R scheme in Supplementary Fig. 6a, the RRAM already fails when the voltage sweeps to -2.8 V. Such property of 1T1R driving scheme helps the RRAM to achieve higher HRS resistances, which results in a higher *R*-ratio of the TCAM cell. To obtain higher resistance in HRS and larger *R*-ratio, we program the RRAMs in Fig. 3 using a slightly higher reset voltage than that in Fig. 2e.

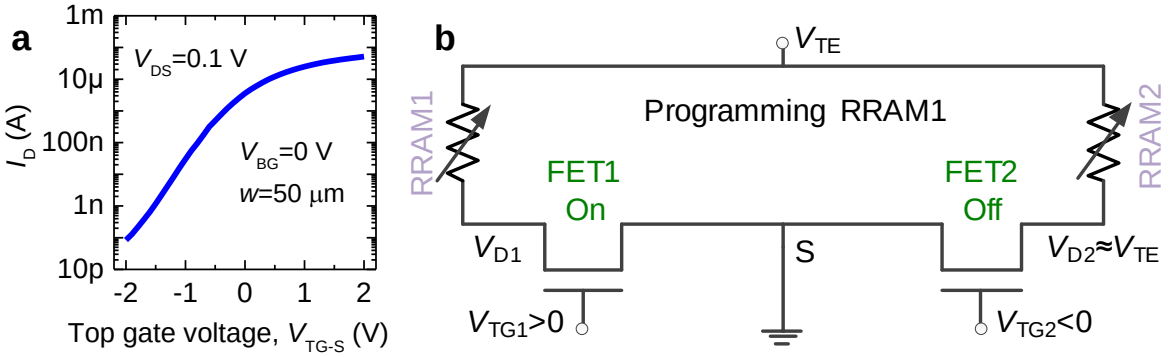


Supplementary Fig. 6. Measurements of RRAMs at high reset voltages using 1R and 1T1R schemes. a, 1R scheme. **b**, 1T1R scheme.

Supplementary Note 5. Optimization of the MoS_2 FETs for TCAM integration

We use a relatively large top gate voltage ($|V_{TG}|$) of 15 to 20 V to operate the TCAM, as shown in Fig. 3. This is because the MoS_2 transistors are integrated with the RRAMs, so to successfully program the RRAMs in the 2T2R TCAM cells without breakdown of the gate dielectric, the MoS_2 cannot be considered alone. Admittedly, if we only consider the MoS_2 transistors, it is possible to decrease V_{TG} by using thinner high-κ dielectric material. We have experimentally demonstrated switching on at 2 V and switching off at -2 V, with on/off ratio of ~ 10^6 (Supplementary Fig. 7a), for a top-gated MoS_2 FET with 5 nm HfO_x as gate dielectric. Similar low voltage operation has also been achieved in other work with thin gate dielectrics [S3,S7]. However, these types of transistors are not suitable for our 2T2R TCAM cell because the programming of the TCAM cell induces high voltage between the gate dielectric, which could result in dielectric breakdown.

Such high voltage during TCAM cell programming is illustrated by the example in Supplementary Fig. 7b. When programming RRAM1 in the 2T2R TCAM cell, we turn on FET1 in series with RRAM1, and turn off FET2 in series with RRAM2. We ground S and apply the programming voltage to the top electrode (TE) of the RRAMs. If RRAM2 is in LRS, then the off-state resistance of FET2 should be much larger than the LRS resistance of RRAM2, so almost all the voltage drop is on FET2, and $V_{D2} \approx V_{TE}$. Then the voltage drop across the top gate dielectric of FET2 is $V_{GD} \approx V_{TG2}$-V_{TE}. Since FET2 is off, we have $V_{TG2} < V_T$, where V_T is the threshold voltage. If the MoS_2 transistor characteristic in Supplementary Fig. 7a is used, and V_{TG2} = -2 V, then we get a large $|V_{GD}|$ on FET2 when we perform the forming process on RRAM1, where V_{TE} is up to 5 V, and $|V_{GD}| \approx 7$ V. This results in a field (14 MV/cm) higher than the breakdown field of HfO_x, which is ~ 7.5 MV/cm from our measured data in Supplementary Fig. 3b.



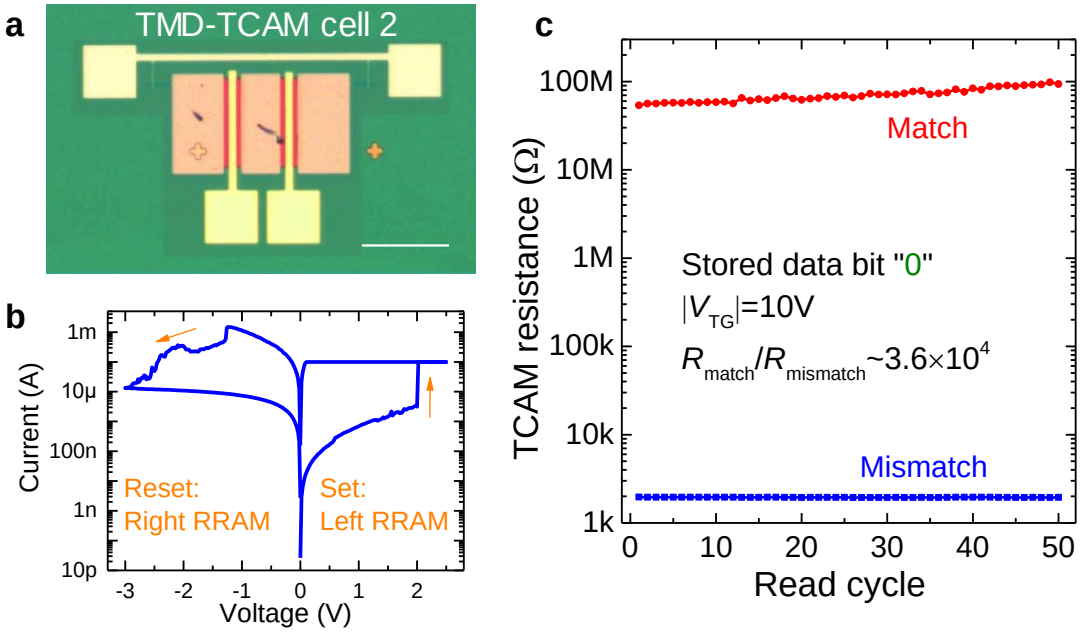
Supplementary Fig. 7. Optimization of the top gate operating voltage. a, Measurement of the I_D – V_{TG-S} curve of a top-gated MoS_2 transistor with 5 nm HfO_x as top gate dielectric, showing switching at ± 2 V, with a grounded source (S). **b**, Equivalent circuit for programming RRAM1 in the 2T2R TCAM cell.

To avoid dielectric breakdown, we use a thicker top gate dielectric in this study (5 nm HfO_x and 22 nm Al_2O_3). The dielectric is thick enough to sustain the highest voltage during the RRAM programming, which results in relatively large top gate voltages required to operate the MoS_2 transistors (15 – 20 V). To ensure good gating of the MoS_2 transistors during the programming or read of the TCAM cells, the Si back gate can also be used to assist in switching the MoS_2 transistors with high on/off ratio. To explore the possibility of scaling down V_{TG}, we have also fabricated FETs with 5 nm HfO_x and 12 nm Al_2O_3 as the top gate dielectric, and $|V_{TG}|$ reduces to 8 – 10 V (Supplementary Fig. 8). With better quality HfO_x, AlO_x, or ZrO_x possessing large dielectric

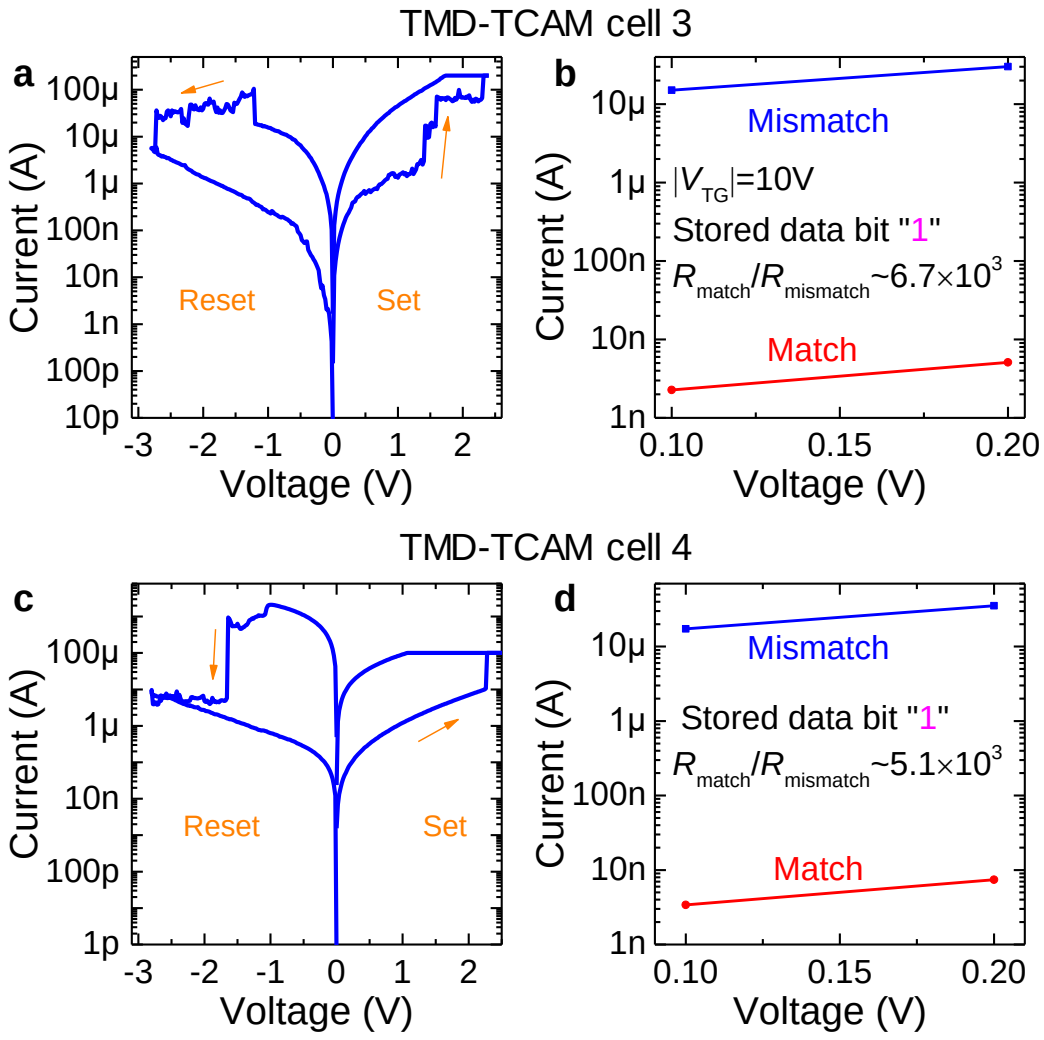
constants and high breakdown fields, thicker dielectrics can be used to maintain a similar equivalent oxide thickness (EOT), and we envision further scaling down of $|V_{TG}|$ to ~ 5 V.

Supplementary Note 6. Measurement results of additional TMD-TCAM cells

We show the measurement results on other representative TMD-TCAM cells in Supplementary Figs. 8 and 9. In Supplementary Fig. 8, we experimentally demonstrate the measurements on the TMD-TCAM cell 2 whose top gate voltage has been decreased to ± 10 V, because we use thinner gate dielectric (5 nm HfO_x and 12 nm Al_2O_3) compared to the TMD-TCAM cell 1 in Fig. 3. This TCAM cell still shows a large *R*-ratio of 3.6×10^4, which is at least 2 orders of magnitude higher than previous TCAMs based on emerging memories (*R*-ratio ~ 100). We have also demonstrated measurement results on two other TMD-TCAM cells (cell 3 and 4) in Supplementary Fig. 9, which show that our technique can be realized in multiple devices.



Supplementary Fig. 8. Measurements on a TCAM cell (TMD-TCAM cell 2), where the MoS_2 FETs have 5 nm HfO_x and 12 nm Al_2O_3 as top gate dielectric, and with channel width of 100 μm. a, Optical image of the device. *Scale bar*: 100 μm. **b**, Programming of the RRAM on the left (set) and the RRAM on the right (reset), resulting in a stored bit "0". **c**, Repeated quasi-DC reading for 50 cycles of the TCAM cell resistance, for both match and mismatch states.



Supplementary Fig. 9. Measurements on other two TCAM cells (TMD-TCAM cell 3 and 4). **a – b**, Experimental data for TMD-TCAM cell 3, showing (**a**) programming of the RRAMs, and (**b**) measurements of the TCAM cell resistances in match and mismatch states. **c – d**, Experimental data for TMD-TCAM cell 4, showing (**c**) programming of the RRAMs, and (**d**) measurements of the TCAM cell resistances in match and mismatch states.

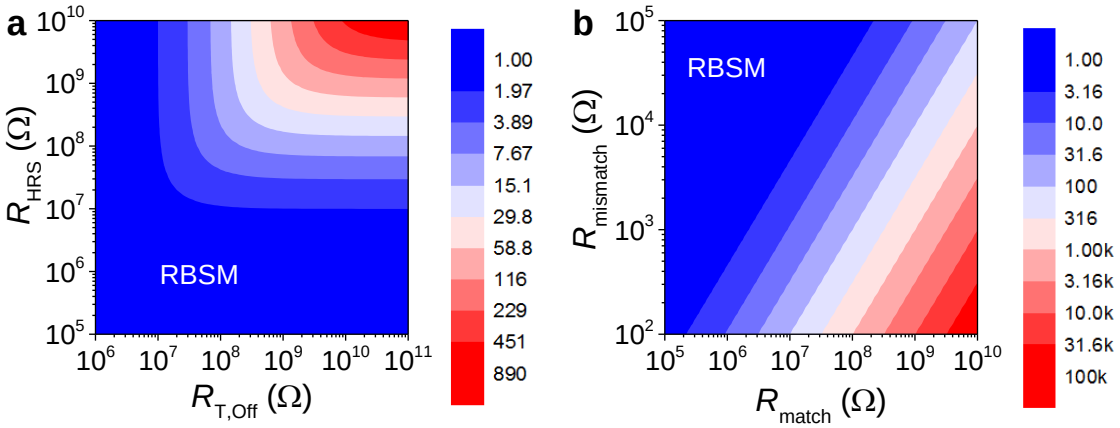
Supplementary Note 7. HSPICE simulation setup

Supplementary Table 1. Key parameters used in the HSPICE simulations

	Value	Unit	Notes
Median LRS	3.5	kΩ	Measured from CAM cells (transistor variations included)
Median HRS	15	MΩ	Measured from CAM cells (transistor variations included)
Technology node	90	nm	
V_{DD}	1.0	V	
Wire resistance	1.0	Ω/unit	
Clock	5	ns	Frequency = 200 MHz
Rise/fall time	0.1	ns	
Duty cycle	1/2	N/A	
NMOS W/L in SenseAmp	1, 2	N/A	5-T sense-and-latch (SAL) design [S8]
PMOS W/L in SenseAmp	4	N/A	5-T sense-and-latch (SAL) design [S8]

Supplementary Note 8. Calculation of the TCAM sense margin

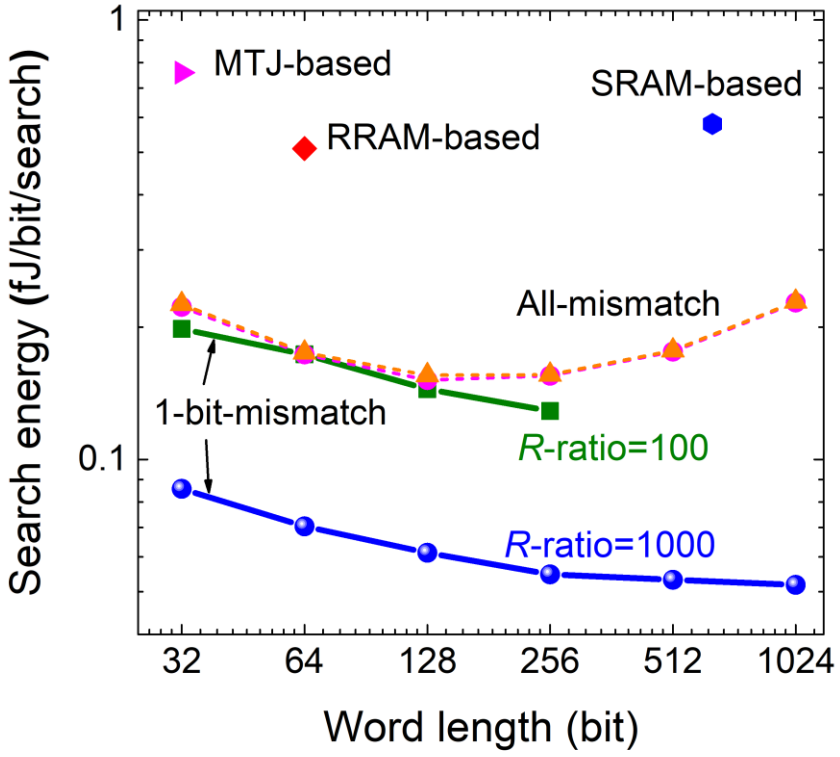
We calculate the resistance-based sense margin (RBSM) of the TCAM based on the *R*-ratio of the individual TCM cells, by using (*R*-ratio)/*n* + (*n*-1)/*n*, where *n* is the word length on each match line (ML), without considering the wire resistance and bit "X". In general, increasing the off-state resistance of the transistor ($R_{T,off}$) and the high-resistance state (HRS) resistance of the RRAM (R_{HRS}) are very effective for increasing the RBSM (Supplementary Fig. 10a). We limit $R_{T,off}$ to be in the range of 10^6 to 10^{11} Ω, much larger than the R_{HRS}, because the MoS_2 FETs have very low leakage. When $R_{T,off}$ is much larger than R_{HRS}, the RBSM is dominated by R_{HRS}. Lower resistance in the mismatch state of the TCAM cells ($R_{mismatch}$) also helps achieve large RBSM, while typically there is limited range to change $R_{mismatch}$ compared with the match state resistance of the TCAM cells (R_{match}), and we assume that $R_{mismatch}$ is in the range of 10^2 to 10^5 Ω (Supplementary Fig. 10b).



Supplementary Fig. 10. Calculated RBSM in color scale. **a**, RBSM with varying $R_{T,off}$ and R_{HRS}, for a TCAM array with 1,024-bit word length, and assuming that $R_{mismatch} = 10$ kΩ. **b**, RBSM with varying R_{match} and $R_{mismatch}$, for a TCAM array with 1,024-bit word length.

Supplementary Note 9. Search latency and energy/bit obtained through simulations

Using voltage-based sense margin (VBSM), we simulate the ML development (*i.e.*, discharging) behaviors during search operations (corresponding to all-match and 1-bit-mismatch states), as shown in Fig. 5c, using the HSPICE parameters listed in Supplementary Table 1. During the discharging, the voltage gap between all-match and 1-bit-mismatch states gradually increases with time, until the sense amplifier (SA) margin and sensing voltage are reached to get the correct digital outputs. We show that a search time of 1.25 ns can be achieved in our TCAM with word length of 128 bits, without intentionally tuning the circuit design to achieve even faster speeds. We do not optimize the latency in this work because: (1) larger word length achieved through high on/off ratio of our TMD-TCAM cells naturally enables higher searching throughput, which is more critical than the latency itself; (2) pure circuit optimization for TCAM is not the focus of this work.



Supplementary Fig. 11. Simulated energy consumption (energy/bit per search operation) for 1-bit-mismatch and all-mismatch states, assuming *R*-ratio of 1000 or 100.

Effective energy/bit per search operation is also simulated and benchmarked against state-of-the-art TCAM designs based on static random-access memories (SRAMs) [S9], RRAMs [S10], and magnetic tunnel junctions (MTJs) [S11]. In our simulation, we estimate the search energy/bit under two situations: 1-bit-mismatch, and all-mismatch states (Supplementary Fig. 11). For the 1-bit-mismatch state, we show that for the TMD-TCAM design, what comes together with higher searching capacity is the search energy efficiency. With increasing word length, the total energy consumption increases, but the effective search energy/bit gradually decreases. This is due to the fact that the increase in the searching capacity (*i.e.*, possible word length per match line) is 'faster' than the increase in the dynamic and static energy consumption (longer wires and more leakage).

Specifically, by using *R*-ratio of ~ 1000, most of the TCAM cells in the match state do not contribute much to the leakage, whereas the mismatch-state cell pulls down the match line voltage during search and dominates the leakage. Thus with an increasing word length, there will be more match-state cells on the ML, but the total leakage will not increase much, and the search energy/bit will decrease. In this case, high throughput and improved energy efficiency can be achieved simultaneously. If we use a hypothetical TCAM cell with *R*-ratio of only ~ 100 (by lowering the high resistance state values of the experimental data) while still using the 1-bit-mismatch state, similar trend is observed, with decreasing search energy/bit for longer word length, but with higher search energy than the case for *R*-ratio ~ 1000. For *R*-ratio ~ 100, we can only simulate to word length of 256 bits, because at larger word length there will not be enough sense margin, which limits the search throughput.

We also perform simulations for the all-mismatch state, representing a sole increase in leakage cells with larger word length. Results in Supplementary Fig. 11 show that at some point the effective search energy/bit will increase for larger TCAM arrays. We perform simulations for both *R*-ratio ~ 1000 an *R*-ratio ~ 100, and the results are almost identical, because the mismatch-state resistances are very similar. While we show comparison with other work, this only applies to specific conditions, and it is hard to accurately compare the energy consumption among different works in general, because of the difference in technology nodes, and the difference between simulation results and experimental results.

Supplementary Note 10. Discussions on the scaling of MoS_2 FETs and TMD-TCAM cells

With further improvements in making low-resistance contacts [S12] to MoS_2 and more reliable gating or doping schemes of MoS_2 FETs, we expect that the TCAM cell size can be further scaled down, and the 2T2R TCAM with a large search capacity can be produced in a large scale. Experimentally, monolayer MoS_2 FETs with 10 nm self-aligned top gates showing onset of ballistic transport [S3], and bilayer MoS_2 FETs with 1 nm carbon nanotube as the gate showing 65 mV/dec subthreshold swing and on/off ratio of 10^6 [S7] have been demonstrated. Although there are still challenges to mass-produce these devices with uniform performance and high yield, these scaled MoS_2 FETs are promising for integration with RRAMs in large scale after further optimization. As reported in the simulation and model in [S13], monolayer MoS_2 transistors with 5 nm gate lengths show performance comparable to the International Technology Roadmap for Semiconductors (ITRS) 2026 requirement for low operating power applications. The large bandgap and large effective mass of the monolayer MoS_2 can help block the direct tunneling current at 5 nm gate length. Other simulations also show the promise for 1L MoS_2 FETs to scale below 10 nm [S14,S15].

A very important characteristic in the integrated TCAM is that the scaling of transistors is linked to the requirement of programming the RRAMs. We design the W/L of the MoS_2 transistors in this manuscript to provide enough drive currents to program the RRAMs, and the W/L can be further scaled down when the programming currents of the RRAMs are reduced. When scaled transistors and scaled RRAMs are integrated together, the TCAM size can also be scaled down.

We further compare the cell sizes of the TMD-TCAMs and TCAMs based on other memory technologies, as shown in Supplementary Table 2. Although the current TMD-TCAM cell is relatively large, it can be significantly scaled down when using more advanced patterning technology, and we show the projected cell size based on our 2T2R design. All the designs are based on 90 nm CMOS technology node, and the contacted gate pitch is assumed to be 260 nm [S16]. The RRAM is assumed to scale down to 4 F^2 [S17], and the transistor has width of 5 F to ensure enough drive current, so the total TMD-TCAM cell size will be around 50 F^2, where F is the feature size of the lithography. The cell size can be further reduced to around 40 F^2 when the RRAMs are directly integrated above the MoS_2 transistors.

Supplementary Table 2. Summary of the TCAM cell sizes based on different designs, at the 90 nm technology node.

Reference	Memory device	TCAM cell design	TCAM cell size (F^2)	TCAM cell size (μm^2)
S18	SRAM	N.A.	560	4.54
S11	MTJ	4T-2MTJ	388	3.14
S19	RRAM	3T1R	194 (logic rule) 107 (SRAM rule)	1.57 (logic rule) 0.87 (SRAM rule)
S20	Phase change memory (PCM)	2T2PCM	50	0.41
This technology	RRAM	2T2R	50 (design)	0.41 (design)

Supplementary References

[S1] Illarionov, Y. Y. et al. Improved hysteresis and reliability of MoS_2 transistors with high-quality CVD growth and Al_2O_3 encapsulation. *IEEE Electron Device Lett.* **38**, 1763-1766 (2017).

[S2] Kshirsagar, C. U. et al. Dynamic memory cells using MoS_2 field-effect transistors demonstrating femtoampere leakage currents. *ACS Nano* **10**, 8457-8464 (2016).

[S3] English, C. D., Smithe, K. K. H., Xu, R. & Pop, E. Approaching ballistic transport in monolayer MoS_2 transistors with self-aligned 10 nm top gates. In *Tech. Digest International Electron Devices Meeting (IEDM 2016)*, San Francisco, CA, December 3-7, 2016, pp. 131-134.

[S4] Wong, H.-S. P. et al. Metal-oxide RRAM. *Proc. IEEE* **100**, 1951–1970 **(**2012).

[S5] Pagiamtzis, K. & Sheikholeslami, A. Content-addressable memory (CAM) circuits and architectures: a tutorial and survey. *IEEE Journal of Solid-State Circuits* **41**, 712–727 (2006).

[S6] Sung, C., Song, J., Lee, S. & Hwang, H. Improved endurance of RRAM by optimizing reset bias scheme in 1TIR configuration to suppress reset breakdown. In *Proc. IEEE Silicon Nanoelectronics Workshop*, Honolulu, HI, June 12-13, 2016, P1-16.

[S7] Desai, S. B. et al. MoS_2 transistors with 1-nanometer gate lengths. *Science* **354**, 99-102 (2016).

[S8] Zheng, L., Shin, S. & Kang, S.-M. S. Memristor-based ternary content addressable memory (mTCAM) for data-intensive computing. *Semicond. Sci. Technol.* **29,** 104010 (2014).

[S9] Arsovski, I., Hebig, T., Dobson, D. & Wistort, R. A 32 nm 0.58-fJ/bit/search 1-GHz ternary content addressable memory compiler using silicon-aware early-predict late-correct sensing with embedded deep-trench capacitor noise mitigation. *IEEE Journal of Solid-State Circuits* **48**, 932-939 (2013).

[S10] Chang, M.-F., et al. A 3T1R nonvolatile TCAM using MLC ReRAM with sub-1ns search time. In *2015 IEEE Int. Solid-State Circuits Conf. (ISSCC 2015)*, San Francisco, CA, Feb. 22-26, 2015, pp. 318-320.

[S11] Matsunaga, S. et al. A 3.14 μm^2 4T-2MTJ-cell fully parallel TCAM based on nonvolatile logic-in-memory architecture. In *Tech. Digest IEEE Symp. VLSI Circuits (VLSIC 2012)*, Honolulu, HI, June 13-15, 2012, pp. 44–45.

[S12] English, C. D., Shine, G., Dorgan, V. E., Saraswat, K. C. & Pop, E. Improved contacts to MoS_2 transistors by ultra-high vacuum metal deposition. *Nano Lett.* **16**, 3824–3830 (2016).

[S13] Alam, K. & Lake, R. K. Monolayer MoS_2 transistors beyond the technology road map. *IEEE Trans. Electron Devices* **59**, 3250–3254 (2012).

[S14] Cao, W., Kang, J., Sarkar, D., Liu, W. & Banerjee, K. 2D semiconductor FETs—projections and design for sub-10 nm VLSI. *IEEE Trans. Electron Devices* **62**, 3459–3469 (2015).

[S15] Liu, L., Lu, Y. & Guo, J. On monolayer MoS_2 field-effect transistors at the scaling limit. *IEEE Trans. Electron Devices* **60**, 4133–4139 (2013).

[S16] https://en.wikichip.org/wiki/90_nm_lithography_process.

[S17] Yu, S. & Chen, P.-Y. Emerging memory technologies. *IEEE Solid-State Circuits Magazine*, 43-56 (2016).

[S18] Arsovski, I. & Nadkarni, R. Low-Noise Embedded CAM with reduced slew-rate Match-lines and asynchronous Search-lines. *IEEE 2005 Custom Integrated Circuits Conference (CICC 2005)*, San Jose, CA. Sept. 21, 2005, pp. 447–450.

[S19] Chang, M. F. et al. A 3T1R nonvolatile TCAM using MLC ReRAM for frequent-off instant-on filters in IoT and big-data processing. *IEEE Journal of Solid-State Circuits* **52**, 1664–1679 (2017).

[S20] Li, J., Montoye, R. K., Ishii, M. & Chang, L. 1 Mb 0.41 μm^2 2T-2R cell nonvolatile TCAM with two-bit encoding and clocked self-referenced sensing. *IEEE Journal of Solid-State Circuits* **49**, 896–907 (2014).