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

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]

Ternary content-addressable memory (TCAM) is specialized hardware that can perform in-memory search and pattern matching for data-intensive applications. However, achieving TCAMs with high search capacity, good area efficiency and good energy efficiency remains a challenge. Here, we show that two-transistor–two-resistor (2T2R) transition metal dichalcogenide TCAM (TMD-TCAM) cells can be created by integrating single-layer MoS_2 transistors with metal-oxide resistive random-access memories (RRAMs). The MoS_2 transistors have very low leakage currents and can program the RRAMs with exceptionally robust current control, enabling the parallel search of very large numbers of data bits. These TCAM cells also exhibit remarkably large resistance ratios (*R*-ratios) of up to 8.5×10^5 between match and mismatch states. This *R*-ratio is comparable to that of commercial TCAMs using static random-access memories (SRAMs), with the key advantage that our 2T2R TCAMs use far fewer transistors and have zero standby power due to the non-volatility of RRAMs.

Data-intensive computation typically requires massive parallel processing and demands heavy memory traffic due to frequent off-chip memory access[1]. Off-chip memory access incurs ten times more latency and energy consumption than on-chip memory data manipulation[2]. To eliminate redundant computations and data movements between the computation unit and the off-chip memory, in-memory computing machines directly process the data within the memory chip, thus reducing the energy consumption and latency[3–5]. Ternary content-addressable memory (TCAM) is a hardware realization of in-memory computing for the parallel search of massive datasets, which saves time and energy[6]. TCAM performs the search function by comparing the input data with the stored data in the TCAM in parallel, and returning the data address when a match is detected[7]. Such parallel search allows TCAMs to perform a look-up table function in a single clock cycle. Unlike a binary content-addressable memory (BCAM) cell, which stores bit values of either '0' or '1', a TCAM cell can store an additional 'X' ('don't care') bit, which results in a match state regardless of the input search data, and makes TCAM much more powerful in searching[8]. TCAMs are highly promising for a variety of data-intensive applications that involve searching and matching of large amounts of data with high throughput, such as computing at the edge of the internet, multimedia processing, big data analytics, data mining and information retrieval[3,4,8].

TCAMs based on static random-access memories (SRAMs) are used in commercial applications such as network routers. However, multiple (~16) transistors are required to form a single TCAM cell, which results in large areas and high power consumption[7], limiting its potential for energy-constrained applications. Resistive switching non-volatile memories (NVMs) are promising alternatives for implementing TCAMs that are more area- and energy-efficient[6,8]. TCAMs based on spin-transfer torque (STT) RAMs[9], phase-change memories (PCMs)[8], and resistive random-access memories (RRAMs)[6] have been demonstrated. However, these TCAM cells have had limited (≤100) resistance ratios (*R*-ratios) between the match and mismatch states (R-ratio $= R_{match}/R_{mismatch}$). Such an *R*-ratio is not enough for many practical applications requiring the parallel search of massive datasets[10], because in a TCAM array the leakage currents of the TCAM cells on the same match line add together, and the limited *R*-ratio makes it difficult to distinguish between the all-match state and 1-bit-mismatch state when the array size is large. Large *R*-ratios comparable to those of TCAM cells based on SRAMs (>10^5) are necessary to enable parallel search for large amounts of data while maintaining enough sense margin in a TCAM array.

In this Article, we explore and demonstrate a TCAM architecture formed by integrating monolayer (1 L) molybdenum disulfide (MoS_2) field-effect transistors (FETs) with metal-oxide RRAMs in a two-transistor-two-resistor (2T2R) layout. We show very large *R*-ratios up to 8.5×10^5, three to four orders of magnitude larger than the *R*-ratios of TCAM cells based on other emerging memories, and comparable with those of SRAM-based TCAMs. These results represent a key application of two-dimensional (2D) transistors, taking advantage of their high performance yet low leakage. Through SPICE circuit simulations of a TCAM array with word lengths of up to 2,048 bits and with 1,024 entries (a total of up to 2,048 kb), we show that the measured large *R*-ratios of the TCAM cells lead to a large search capacity. The low processing temperatures (<200 °C) involved in MoS_2 transfer[11] and RRAM fabrication can make the manufacturing of such devices compatible with back-end-of-line (BEOL) metal interconnect wire processes. Therefore, the TCAM array can potentially be integrated into three-dimensional (3D) circuits with dense logic and memory layers, for low-energy and low-latency memory access[12].

Careful selection of the transistor and memory device type is required to achieve TCAM cells with large *R*-ratios. Monolayer MoS_2, a transition metal dichalcogenide (TMD), is an ideal semiconducting material in our application for a multitude of reasons. First, 1 L MoS_2 is a 2D semiconductor with a relatively wide electronic bandgap (~2 eV) and large carrier effective mass[13–15], which

[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

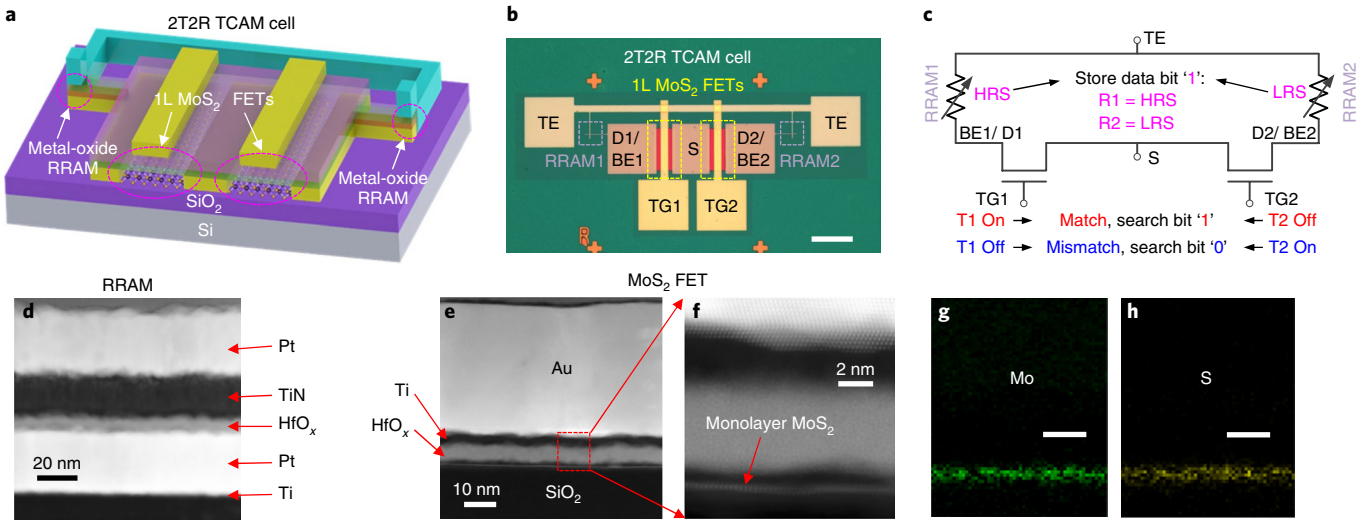

Fig. 1 | The structure of 2T2R TMD-TCAM cells. a, 3D schematic illustration of the 2T2R TCAM cell, with two MoS_2 FETs and two HfO_x-based RRAMs. **b**, Top-view optical image of the 2T2R TCAM cell 1. The MoS_2 FETs have 5 nm HfO_x and 22 nm Al_2O_3 as the top gate dielectric, with MoS_2 channel width $W = 50\,\mu m$ and channel length $L = 0.8\,\mu m$ (the top gate electrodes overlap with part of the source/drain contacts), and the RRAMs have a $1\,\mu m \times 1\,\mu m$ overlap area. Scale bar: 50 μm. **c**, Circuit diagram of the 2T2R TCAM cell showing the definition of match and mismatch states, with the stored data bit '1'. **d**, Cross-sectional TEM image of the HfO_x RRAM, showing 5 nm HfO_x as the RRAM switching layer. **e**, Cross-sectional TEM image of a top-gated monolayer MoS_2 FET with 5 nm HfO_x as the top gate dielectric. **f**, Zoomed-in view of the MoS_2 channel region shown by the red dashed box in **e**. **g**,**h**, EDS maps of Mo and S, respectively, revealing the monolayer MoS_2 as the channel material, with a scale bar of 5 nm.

allow 1 L MoS_2 FETs to have low off-state leakage currents[16,17]. Second, 1 L MoS_2 films can be grown over large areas[18–20], forming FETs with high current drives and large on/off ratios[21–23]. Third, 1 L MoS_2 FETs have superior immunity to short-channel effects, due to the atomic-scale thickness and low dielectric constant[24,25]. Fourth, MoS_2 FETs exhibit good mobility even for the monolayer channels required for gate lengths below 10 nm, compared to bulk semiconductors such as Si, which have much lower mobility at comparable thicknesses[23,26–29]. Finally, MoS_2 is atomically thin and can be transferred and fabricated at low temperatures, which make it suitable for monolithic 3D integration and high-energy-efficiency operation.

Metal-oxide RRAMs are ideal complements to MoS_2 FETs[30] because they are non-volatile[31,32], scalable to high density[33,34], and are as fast as dynamic random-access memory (DRAM)[35]. They can achieve large resistance ratios between high-resistance states (HRS) and low-resistance states (LRS)[36], which are necessary for large *R*-ratios in TCAM cells. In addition, they can readily integrate into TCAMs because they can be fabricated on the same chip as the computing logic transistors at low temperature (typically below 200°C), and are made with materials that are compatible with complementary metal–oxide–semiconductor (CMOS) processes[30,37]. Importantly, the 1 L MoS_2 FETs provide automatic current control to reliably program the RRAMs to very high resistance states, especially during the reset process, which greatly reduces the chance for failure at high reset voltages.

2T2R TMD-TCAM cell structure

The TMD-TCAM cell is formed by a 2T2R structure (Fig. 1a), with each MoS_2 transistor driving a RRAM element, thus using far fewer components than SRAM-based TCAM cells. This scheme builds on the basic one-transistor–one-resistor (1T1R) device configuration, which has been previously demonstrated[38]. Here we further integrate the 1 L MoS_2 FETs and RRAMs into functional TCAM cells with large *R*-ratios, show the understanding of the mechanism for obtaining high RRAM resistances using the 1T1R driving scheme, and analyse, through circuit simulations and using the experimentally obtained device characteristics, the searching capability of a TCAM array with a long (2,048 bits) word length. The 1 L MoS_2 is grown by chemical vapour deposition (CVD) to ensure uniformity and scalability[19,39]. The 2T2R TCAM cell has two top-gated monolayer MoS_2 FETs with a shared source (S), and each FET separately drives a RRAM cell. The bottom electrodes (BEs) of the RRAMs connect to the drain (D) electrodes of the FETs, and the two RRAMs share the top electrode (TE) (Fig. 1b). The top gates (TGs) of the MoS_2 FETs control the two MoS_2 channels separately. The fabrication process is shown in Supplementary Fig. 1.

The circuit diagram with the stored data in the TCAM cell and the searching scheme is shown in Fig. 1c, which we will discuss in detail in later sections. The structure of the fabricated RRAM is shown by the cross-sectional transmission electron microscopy (TEM) image in Fig. 1d, which consists of a HfO_x switching layer between the top and bottom electrodes. The TEM cross section of the top-gated MoS_2 FET is shown in Fig. 1e, with the zoomed-in image showing the 1 L MoS_2 channel (Fig. 1f). Energy dispersive spectroscopy (EDS) maps for Mo and S in Fig. 1g,h confirm the 1 L MoS_2 material.

1T1R programming for data storage in TCAM cells

Data storage in an individual RRAM element is mediated using the MoS_2 access transistors in a 1T1R configuration. When programming one of the RRAMs in a TCAM cell, the MoS_2 FET addressing the other RRAM in the same TCAM cell is turned off to avoid cross-talk. The schematic and the equivalent circuit diagram of the 1T1R structure are illustrated in Fig. 2a,b, respectively. We first characterize the properties of the top-gated monolayer MoS_2 FETs, and the measured drain current I_D versus $V_{TG\text{-}S}$ characteristic in Fig. 2c shows an on/off ratio $\sim 2 \times 10^7$. A higher on/off ratio of 10^9 is obtained in back-gated MoS_2 FETs (Supplementary Fig. 2). We measure the I_D versus V_{DS} curves at different $V_{TG\text{-}S}$, showing a current drive up to $170\,\mu A\,\mu m^{-1}$, with no electrical breakdown up to $V_{DS} = 11$ V (Fig. 2d). These results are important because the MoS_2 FETs not only need to provide enough current drive to the RRAMs, but also need to sustain large enough voltages for programming the RRAMs.

To store the data bits in the TMD-TCAM cells, we then program the RRAMs using the 1T1R scheme. With one FET turned

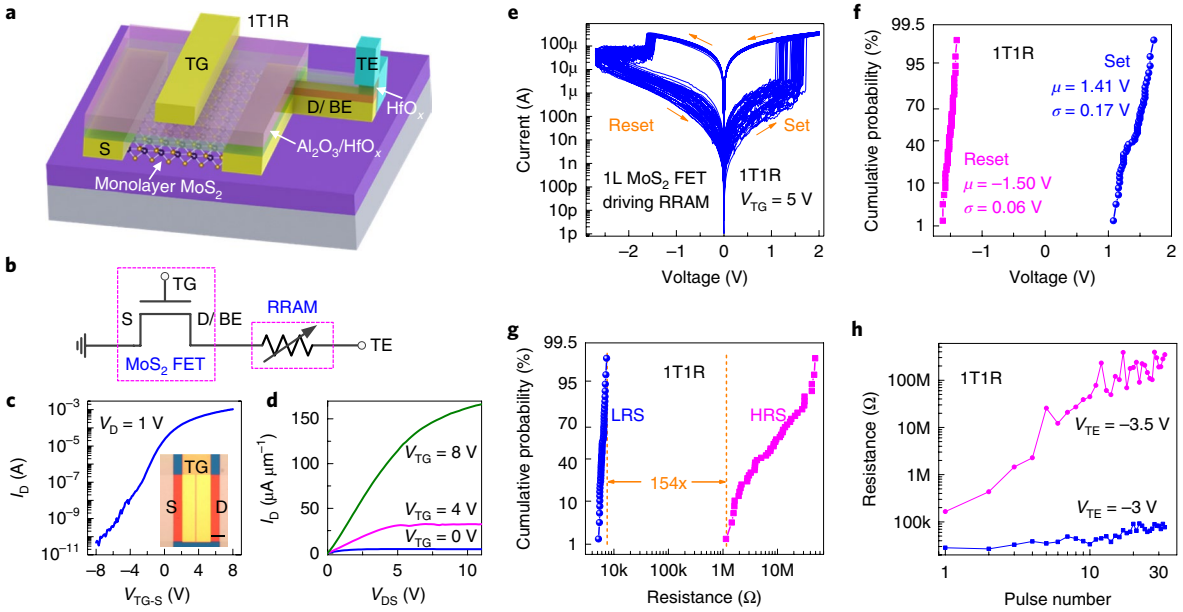

Fig. 2 | Measurements of the 1T1R component with the monolayer MoS_2 FET driving the RRAM. a, 3D schematic of the 1T1R structure, as a component of the 2T2R TCAM cell. **b**, Circuit diagram of the 1T1R structure in **a**. **c**,**d**, Measured representative I_D–$V_{TG\text{-}S}$ (**c**) and I_D–V_{DS} (**d**) characteristics of the top-gated monolayer MoS_2 FETs with 25 nm HfO_x as gate dielectric, with $W = 50$ μm, $L = 0.8$ μm, measured at $V_{BG} = 0$ V and with the source grounded. Inset in **c**: Top-view optical image of the MoS_2 FET, with a scale bar of 10 μm, showing that the top gate overlaps with part of the source/drain so that the MoS_2 channel length is shorter than the gate length. **e**, 45 cycles of 1T1R set and reset measurements, using the top-gated MoS_2 FET with 5 nm HfO_x and 7 nm Al_2O_3 as gate dielectric, with $W = 20$ μm, $L = 0.6$ μm, and RRAM with a 1 μm × 1 μm overlap area. **f**, Distribution of the set and reset voltages in **e**. **g**, Distribution of HRS and LRS resistances during 1T1R measurements in **e**. **h**, Pulsed reset measurements on another 1T1R structure using an increasing number of pulses (with 1 μs pulse width) during RRAM reset, showing multiple resistance states.

on, the programming voltage is applied on the top electrode of the RRAM (Fig. 2b). The HfO_x-based RRAMs typically need a forming process to create the initial conductive filament, which requires a relatively high voltage up to 5 V (Supplementary Fig. 3), and the MoS_2 FET is compatible with such high voltage drive. Then we repeatedly reset/set the RRAM to HRS/LRS by applying negative/positive voltages on TE (V_{TE}) while grounding the source, for 45 cycles using d.c. voltage sweeps (Fig. 2e). We observe that the 1 L MoS_2 FET drives enough current to the RRAM, and that it reliably controls the current compliance during the RRAM set process. The distributions of set and reset voltages during the d.c. sweeps exhibit small variations (Fig. 2f). The distributions of HRS and LRS resistances (including the RRAM resistance and on-state resistance of the MoS_2 FET) are summarized in Fig. 2g, showing that the 'worst-case' ratio between the HRS and the LRS (if we include the tails of the distribution) is 154, and the median value resistance ratio is 1,225. We also control the resistance levels during reset by applying voltage pulses and gradually increasing the pulse number, which provides flexibility to meet certain R-ratio requirements by trading off power and energy consumption (Fig. 2h).

Large R-ratios in TMD-TCAM cells

In the TMD-TCAM cells, the two RRAMs in a TCAM cell are programmed to store one bit of TCAM data. As shown in the circuit diagram in Fig. 1c, the stored TCAM data are defined as follows. Data bit '1': RRAM1 is in HRS, RRAM2 is in LRS. Data bit '0': RRAM1 is in LRS, RRAM2 is in HRS. Data bit 'X' (don't care bit): both RRAMs are in HRS. When we search for the data in the TCAM cell, we send the search signals to the TGs of the transistors, and the data search operation can be defined as follows. Search bit '1': TG1 is high (FET1 is on), TG2 is low (FET2 is off). Search bit '0': TG2 is high (FET2 is on), TG1 is low (FET1 is off). The search signals on the two transistors are always the opposite.

To obtain a large R-ratio in the TCAM cell, it is important to have both a large on/off ratio of the transistor and a large HRS/LRS resistance ratio of the RRAM. In the specific example shown in Fig. 1c, data bit '1' is stored in this TCAM cell. When we search for bit '1', which results in a match state, the resistance of the TCAM cell is high. In this match state, the FET in series with the RRAM in HRS is turned on, while the FET in series with the RRAM in LRS is turned off, so the total resistance in the match state is: $R_{match} = (R_{T,off} + R_{LRS}) || (R_{T,on} + R_{HRS}) \approx R_{T,off} || R_{HRS}$, where $R_{T,off}$, $R_{T,on}$ are the off-state and on-state resistance of the transistor, respectively, and R_{HRS} and R_{LRS} are the HRS and LRS state resistance of the RRAM, respectively. Here we assume that $R_{T,off}$ and R_{HRS} are much larger than $R_{T,on}$ and R_{LRS}, which is true for the properly designed devices in our experiments.

On the other hand, if we search for bit '0' when the stored data is bit '1', this results in a mismatch state, and the resistance of the TCAM cell is low. In the mismatch state, the FET in series with the RRAM in HRS is turned off, while the FET in the series with the RRAM in LRS is turned on, and the cell resistance is: $R_{mismatch} = (R_{T,off} + R_{HRS}) || (R_{T,on} + R_{LRS}) \approx R_{T,on} + R_{LRS}$. For the stored bit 'X', it is also called 'don't care bit' because both RRAMs store HRS, and no matter which transistor is turned on, the TCAM always results in a high resistance, as shown by $R_X = (R_{T,off} + R_{HRS}) || (R_{T,on} + R_{HRS}) \approx (R_{T,off} + R_{HRS}) || R_{HRS}$.

TCAM application is read-dominant, thus to characterize our 2T2R TCAM cells, we program an individual cell to store data bit '1' and repeatedly search for the bit by applying quasi-d.c. voltages to the top gates and reading the resistance at the RRAM top electrode. The measurement results for TMD-TCAM cell 1 storing data bit '1' are summarized in Fig. 3a,b, with the optical image and

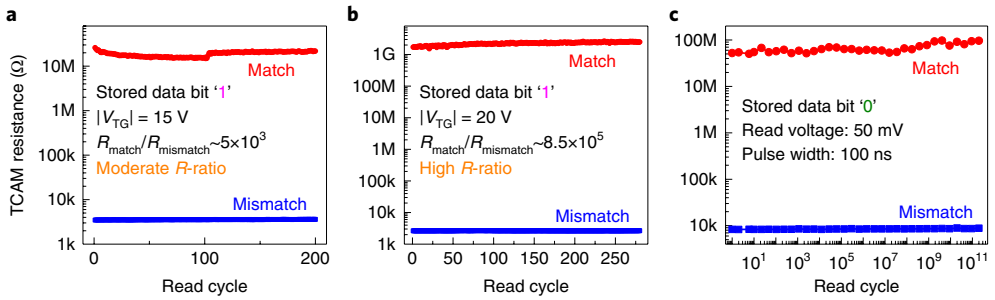

Fig. 3 | 2T2R TMD-TCAM cell characterization. a,b, Repeated quasi-d.c. reading of the resistance of the TMD-TCAM cell 1 from the top electrode of the RRAMs, for both match and mismatch states between the search data and the stored data bit '1', using moderate gate voltages on MoS_2 FETs (**a**), which result in a moderate *R*-ratio, for 200 cycles, and large gate voltages (**b**), which result in a large *R*-ratio, for 280 cycles. **c**, Pulsed read measurements of the match and mismatch states for up to 2×10^{11} cycles after re-programming the TMD-TCAM cell 1 to store data bit '0'. The resistance values are fairly stable, which is consistent with the retention measurement (Supplementary Fig. 4a).

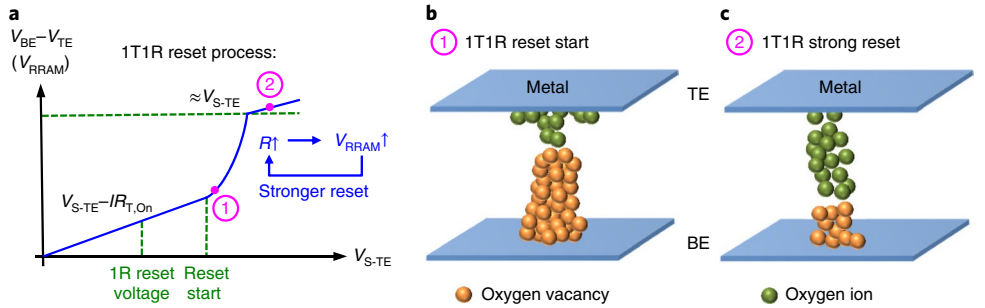

Fig. 4 | Analysis of the large *R*-ratio of the TMD-TCAM cell. a, Schematic of the voltage drop on the RRAM with increasing total applied voltage between S and TE, showing the sudden increase of the voltage when the RRAM reset starts. **b**,**c**, Illustration of the mechanism for obtaining the large resistance ratio using the 1T1R programming scheme based on the MoS_2 FETs and RRAMs, showing the metal-oxide RRAM when the reset just starts (**b**) and when the strong reset happens (**c**) with a large voltage drop on the high-resistance RRAM.

geometry of the TCAM cell shown in Fig. 1b. When $V_{TG1} = 15$ V and $V_{TG2} = -15$ V, a match state is achieved, otherwise a mismatch state is achieved. We get a moderate *R*-ratio ($R\text{-ratio} = \frac{R_{match}}{R_{mismatch}} \approx \frac{R_{T,off} \parallel R_{HRS}}{R_{T,on} + R_{LRS}}$) of ~5,000 for the TCAM cell (Fig. 3a), which is higher than previous TCAM cells based on RRAMs or PCMs, but lower than SRAM-based TCAMs. When we increase $|V_{TG}|$ to 20 V, R_{match} increases and $R_{mismatch}$ decreases, resulting in a high *R*-ratio of up to $\sim8.5\times10^5$ (Fig. 3b), which is comparable to SRAM-based TCAM cells.

TCAM searching does not require constantly re-programming the TCAM. We have measured set/reset of the RRAM using voltage pulses for up to 10^5 cycles (Supplementary Fig. 4), which satisfies the need for most TCAM applications. For the same TCAM cell, we re-program the cell so that it stores data bit '0' (Supplementary Fig. 5). We read the resistance values of the TCAM cell using read voltage pulses for up to 2×10^{11} times, for both match and mismatch states, and observe very stable TCAM resistances, showing that the TCAM is very robust to read disturbances of the programmed state (Fig. 3c). The TCAM cell storing bit '0' also has a high *R*-ratio of ~10^4.

The high *R*-ratios in our 2T2R TCAM cells are due not only to the high on/off ratio of the 1 L MoS_2 FETs, but also to our unique 1T1R programming scheme, which enables higher programmed resistance states for the RRAMs. When programming the RRAM, the $R_{T,on}$ (~1–3 kΩ) of the MoS_2 FET is tuned to be comparable or slightly lower than the R_{LRS} (~2–9 kΩ) of the RRAM, and they form a voltage divider. As shown in Fig. 4a, the voltage on the RRAM is initially $V_{S\text{-}TE} - IR_{T,on}$, and gradually increases with the applied voltage $V_{S\text{-}TE}$. When the reset process starts (with the defect profile in the RRAM illustrated in Fig. 4b), the increasing resistance on the RRAM results in a larger voltage drop that develops across the RRAM, which accelerates the reset process. After the reset finishes, because R_{HRS} is much larger than $R_{T,on}$, the voltage on the RRAM quickly increases to $V_{S\text{-}TE}$. This voltage $V_{S\text{-}TE}$ is typically much higher than the RRAM reset voltage when not using the transistor in series with the RRAM (typically ~1 V), thus resulting in a much stronger reset and higher resistance in HRS (Fig. 4c). As shown in Supplementary Fig. 6, during the RRAM reset process, if the RRAM is programmed in the 1T1R scheme, the chance for reverse set failure is much lower than that without the transistor in series, especially at high reset voltages. Therefore, it is critical that in this 1T1R programming scheme we tune the resistance of the MoS_2 FET to be similar to the R_{LRS} of the RRAM, because it allows us to apply a large voltage across the RRAM during the reset process to achieve a higher resistance in HRS, yet avoids the reset failure or hard breakdown of the RRAM at large voltages by providing robust current control[40]. The gate dielectric thicknesses and top gate voltages of the MoS_2 FETs are chosen to ensure proper programming of the TCAM cells without breakdown of the gate dielectric (Supplementary Note 5 and Supplementary Fig. 7). Measured data from three other representative TCAM cells are presented in Supplementary Figs. 8 and 9 and also show large *R*-ratios.

Simulation of TCAM array with large search capacity

We perform extensive circuit-level analysis of a TCAM array using the measured data from these MoS_2-RRAM 2T2R TCAM cells (with device variations) and using silicon-based 90-nm CMOS technology for peripheral circuits[41,42] (Fig. 5a). The analysis is based on

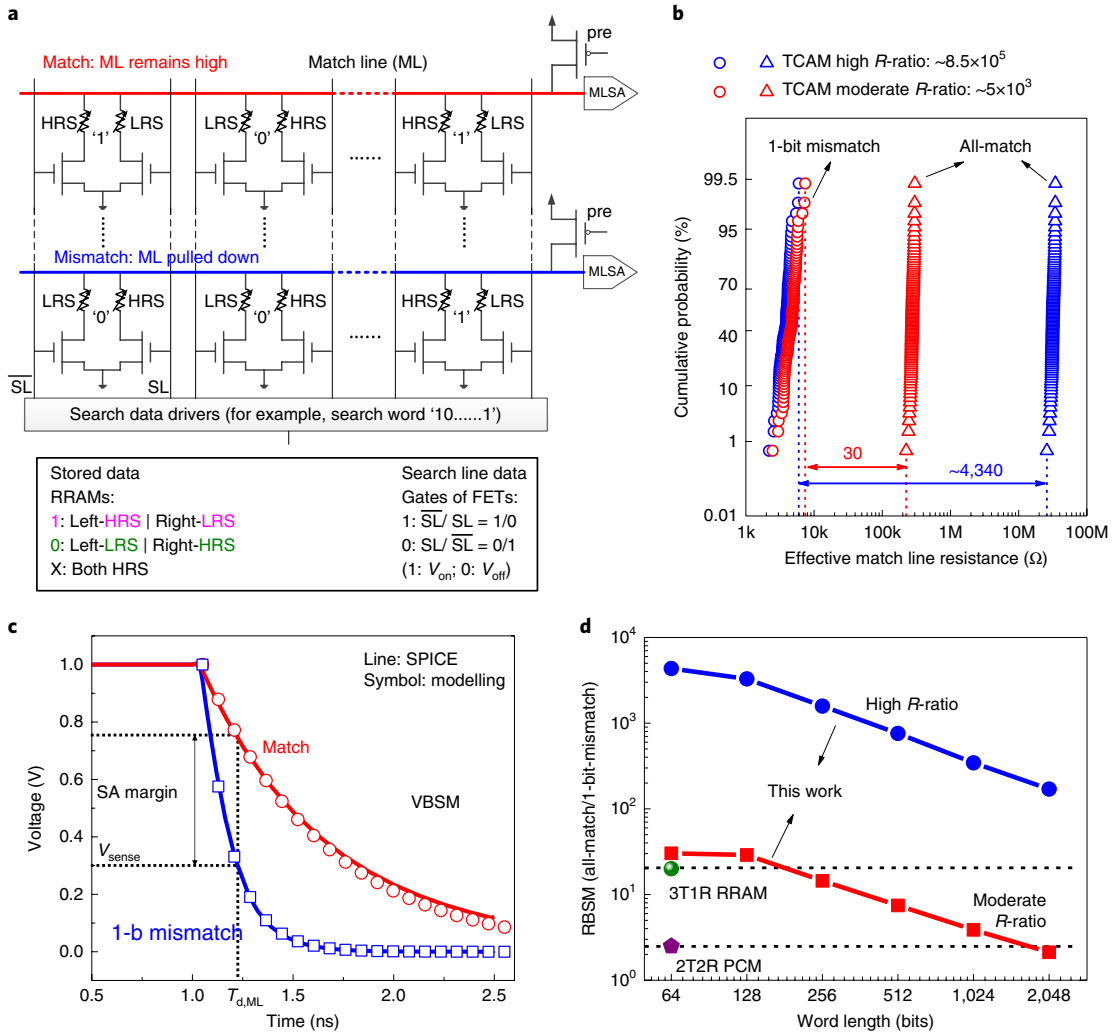

Fig. 5 | Simulation (using SPICE) of the TCAM array using measured 2T2R TMD-TCAM cells. a, Circuit diagram of the simulated TCAM array, using the 2T2R MoS_2-RRAM TCAM cells measured in our experiments, and the standard peripheral circuits based on 90-nm CMOS technology, with the data encoding definitions shown at the bottom. **b**, Simulated distribution of the match line resistance for all-match and 1-bit-mismatch states, for a word length of 64 bit, with 1,024 entries, using the measured *R*-ratios of the TMD-TCAM cells. **c**, Simulated ML discharging behaviours during a search operation (for all-match and 1-bit-mismatch states), showing the VBSM. ML delay ($T_{d,ML}$) is quantified to be 1.25 ns. **d**, Simulation of RBSM at increasing word lengths and comparison with reported results, showing that our TCAM cell with large RBSM can enable very high searching capacity.

SPICE simulations using RRAM array models[43] and 90-nm predictive technology model (PTM)[44], with the detailed parameters for the simulation shown in Supplementary Table 1. For each TCAM cell, the shared source is grounded, the two top gates are on the search lines (SL and $\overline{SL}$), and the two RRAMs share the same TE, which is on the match line (ML). As such, during TCAM operation, the MLs first get charged up to the supply voltage V_{DD}, then the search data are compared with the data stored in the TCAM cells on all MLs together (in one clock cycle). If all the data on a ML match with the search data, then that ML remains high, and the match line sense amplifier (MLSA) reads out the signal. If there is at least one mismatched bit, then that ML will be pulled down to a lower level, which will then be captured by the MLSA.

Sense margin is essential in assessing the parallel search capability of a TCAM. Sense margin can be defined in two methods: the resistance-based sense margin (RBSM), which means the ML resistance ratio between the all-match and 1-bit-mismatch states, and the voltage-based sense margin (VBSM), which means the ML voltage difference between the all-match and 1-bit-mismatch states at certain search latency. The 1-bit-mismatch case is chosen because it is the hardest-to-detect mismatch. The simulated distribution of the resistances on the ML for the all-match and 1-bit-mismatch cases for a 64-bit word is shown in Fig. 5b. The requirements for the transistor and RRAM resistance to achieve certain RBSM is also analysed (Supplementary Fig. 10). The ML development is simulated using VBSM (Fig. 5c), showing search latency of 1.25 ns when the MLSA margin is reached (Supplementary Note 9). The energy consumption is also simulated, showing that a large *R*-ratio not only results in a high search throughput, but also improves the search energy efficiency (effective energy per bit per search) for the 1-bit-mismatch state (Supplementary Fig. 11). Simulation based on our experimentally measured TCAM cells with moderate *R*-ratio and high *R*-ratio provides a RBSM ($R_{\text{all-match}}/R_{\text{1-bit-mismatch}}$) of ~30 and up to ~4,340, respectively (Fig. 5b), where $R_{\text{all-match}}$ is the ML resistance when all the data stored in the TCAM cells on the same ML match with the input data, and $R_{\text{1-bit-mismatch}}$ is the ML resistance when only one bit of data stored in the TCAM cells on the same ML shows a mismatch with the input data. Sense margin decreases with larger word length, which ultimately limits the size of the TCAM array available. The large *R*-ratio of a TCAM cell is

extremely important in obtaining a large enough sense margin for a long word length. Compared with other reported TCAMs based on emerging memories (including PCMs and RRAMs) with the same word length (64 bits)[6,8], the simulated TMD-TCAM arrays based on our measured 2T2R TCAM cells show up to two orders of magnitude larger RBSM (Fig. 5d).

Discussion on using MoS_2 FETs for TCAMs

Although silicon FETs can also be used as the transistors in these TCAM cells[6,8], 2D MoS_2 FETs can outperform silicon FETs in applications requiring low-power, large-scale, and monolithic 3D-integrated TCAMs, due to several advantageous properties. First, 1 L MoS_2 FETs have very low leakage currents and high off-state resistances[16,17]. This contributes to both low-power and large-scale TCAMs: for a large TCAM array, the low-leakage transistors can effectively shut off the leakage paths through them, thus reducing the total power consumption. Furthermore, the large *R*-ratio of each TCAM cell is necessary to maintain a considerable RBSM when the array size is large, and low-leakage transistors can lead to large *R*-ratios. Second, MoS_2 is atomically thin, and MoS_2 FETs have higher mobility than Si FETs at monolayer thicknesses. This can lead to large-scale TCAM and monolithic 3D integration, as 1 L MoS_2 is immune to short channel effects, is scalable to gate lengths below 10 nm (refs. [23,27,28]), and is suitable for large-scale and tight integration with small footprints. Additionally, in monolithic 3D integration, where device layers are laid on top of each other, 1 L MoS_2 results in thinner device layers, which allows denser interlayer vias for interconnects within the limit of the etching aspect ratio. Finally, the MoS_2 transfer and other fabrication steps can be performed below 200°C. This enables the 2T2R TCAM cells to be fabricated in a monolithic 3D scheme, where the MoS_2 FETs can be placed underneath (or above) the RRAM layer without consuming additional chip area. Comparison of the cell sizes of different TCAM cell designs is shown in Supplementary Table 2. Other types of resistive memories can also be considered for vertical scalability in 3D integration[45].

Conclusions

We have reported 2T2R TCAM cells made from CVD monolayer MoS_2 FETs and HfO_x-RRAMs. The TMD-TCAM cells exhibit a high *R*-ratio comparable to SRAM-based TCAMs, due to the low off-state current of 1 L MoS_2 FETs and the robust current control in the 1T1R driving scheme. We have also shown that, when these TCAM cells are integrated into a TCAM array, a very large RBSM can be achieved for a long word length, allowing searches of a large amount of data in parallel. Such a platform is highly promising for data-intensive applications involving high-throughput matching and searching, such as information retrieval and big data analysis[3,4,8], as well as monolithic 3D integration of logic and memory for energy-efficient computing.

Methods

Fabrication process of the 2T2R structure. Fabrication of the 2T2R TCAM starts from CVD growth of continuous, monolayer MoS_2 continuous films on SiO_2 (300 nm) on highly doped Si substrates treated with hexamethyldisilazane (HMDS), using solid sulfur and MoO_3 precursors with the aid of perylene-3,4,9,10 tetracarboxylic acid tetrapotassium salt (PTAS), at a temperature of 850 °C for 15 min, and at a pressure of 760 torr with an Ar environment (Supplementary Fig. 1a). The growth process has been optimized to grow continuous monolayer MoS_2 films with approximately cm^2 size and low variability[39]. 1 L MoS_2 FETs using such CVD materials have shown lower variation in charge carrier density than Si transistors based on 2 nm ultrathin body silicon-on-insulator[19], indicating that the MoS_2 FETs are not expected to increase by much the variability of a large-scale TCAM. Next, electron beam lithography (EBL) is performed to define the MoS_2 channels, followed by etching away the MoS_2 in other regions with a gentle CF_4 plasma (Supplementary Fig. 1b). Another EBL is then performed, followed by physical vapour deposition of pure Au contacts to MoS_2 and lift-off, and the Au contacts ensure low contact resistances[27] and high current drives of these MoS_2 FETs. Then another EBL followed by Ti and Pt evaporation and lift-off define the bottom electrodes of the RRAMs. Next, atomic layer deposition (ALD) of 5 nm HfO_x is performed at 200 °C in vacuum at 0.1 torr base pressure and with N_2 carrier gas, which forms the switching layer of the RRAMs (Supplementary Fig. 1c). After that, EBL, sputter deposition of TiN and Pt, and lift-off define the top electrodes of the RRAMs (Supplementary Fig. 1d). Another optional ALD of HfO_2 at 200 °C or Al_2O_3 at 150 °C following the evaporation of 1.5 nm Al seed adds to the 5 nm HfO_x and forms the top gate dielectric of the MoS_2 FETs together. The combination and thickness of the gate dielectric are chosen to ensure that the MoS_2 FETs effectively program the RRAMs in a 2T2R TCAM cell to obtain certain *R*-ratios without experiencing dielectric breakdown (Supplementary Note 5 and Supplementary Fig. 7). The last EBL, Ti and Au evaporation, and lift-off define the top gates, forming the top-gated MoS_2 FETs (Supplementary Fig. 1e).

Electrical measurement techniques. The static electrical properties including I_D–V_{DS} and I_D–V_{GS} characteristics of the MoS_2 FETs and the 1T1R RRAM programming and 2T2R matching measurements are performed at atmospheric pressure with N_2 flow, using a probe station with up to four probes and four source measurement units (SMUs) connected to a Keithley 4200SCS semiconductor parameter analyser (SPA). During the measurements of the MoS_2 FETs, the voltages are applied to the gates and a drain, and the source is grounded. During the 1T1R RRAM programming, one transistor in the 2T2R cell is turned on for programming the RRAM cell in series with it, while the other transistor in series with the other RRAM is turned off to avoid cross-talk. Then voltage is applied between the top electrode of the RRAM and the source electrode of the FET to program the RRAM through the control of the MoS_2 FET. Si back gate voltages can also be applied to turn on/off the transistor better, together with the top gates. During the pulsed measurements, we use the pulse measurement units (PMUs) connected to the same SPA to provide the voltage pulses.

SPICE simulation of the TCAM array. The simulation of the full TCAM array is performed using SPICE, assuming that the peripheral circuits including the MLSA are based on 90-nm silicon CMOS technology, with detailed parameters in Supplementary Table 1. The TCAM cell characteristics are based on our experimentally measured results for 2T2R structures using MoS_2 FETs and RRAMs. The number of columns, or the word length to search, varies from 64 bits to 2,048 bits, while the number of rows is kept at 1,024 entries. The resistance variation measured from the experimental data has been included in the simulation to estimate the distribution of ML resistance and the sense margin.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Received: 30 August 2018; Accepted: 14 February 2019;
Published online: 15 March 2019

References

1. Wong, H.-S. P. & Salahuddin, S. Memory leads the way to better computing. *Nat. Nanotechnol.* **10**, 191–194 (2015).
2. Theis, T. N. & Wong, H.-S. P. The end of Moore's law: a new beginning for information technology. *Comput. Sci. Eng.* **19**, 41–50 (2017).
3. Ventra, M. D. & Pershin, Y. V. The parallel approach. *Nat. Phys.* **9**, 200–202 (2013).
4. Guo, Q. et al. Resistive ternary content addressable memory systems for data-intensive computing. *IEEE Micro* **35**, 62–71 (2015).
5. Ielmini, D. & Wong, H.-S. P. In-memory computing with resistive switching devices. *Nat. Electron.* **1**, 333–343 (2018).
6. 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 J. Solid-St. Circ.* **52**, 1664–1679 (2017).
7. Pagiamtzis, K. & Sheikholeslami, A. Content-addressable memory (CAM) circuits and architectures: a tutorial and survey. *IEEE J. Solid-St. Circ.* **41**, 712–727 (2006).
8. Li, J., Montoye, R. K., Ishii, M. & Chang, L. 1 Mb 0.41 μm² 2T-2R cell nonvolatile TCAM with two-bit encoding and clocked self-referenced sensing. *IEEE J. Solid-St. Circ.* **49**, 896–907 (2014).
9. Matsunaga, S. et al. A 3.14 μm² 4T-2MTJ-cell fully parallel TCAM based on nonvolatile logic-in-memory architecture. *2012 Symp. VLSI Circuits (VLSIC)* https://doi.org/10.1109/VLSIC.2012.6243781 (IEEE, 2012).
10. Imani, M., Rahimi, A. & Rosing, T. S. Resistive configurable associative memory for approximate computing. In *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)* 1327–1332 (IEEE, 2016).
11. Gurarslan, A. et al. Surface-energy-assisted perfect transfer of centimeter-scale monolayer and few-layer MoS_2 films onto arbitrary substrates. *ACS Nano* **8**, 11522–11528 (2014).

12. Aly, M. M. S. et al. Energy-efficient abundant-data computing: the N3XT 1,000×. *IEEE Computer* **48**, 24–33 (2015).
13. Huang, Y. L. et al. Bandgap tunability at single-layer molybdenum disulphide grain boundaries. *Nat. Commun.* **6**, 6298 (2015).
14. Alam, K. & Lake, R. K. Monolayer MoS_2 transistors beyond the technology road map. *IEEE Trans. Electron Devices* **59**, 3250–3254 (2012).
15. Ryou, J., Kim, Y.-S., Santosh, K. C. & Cho, K. Monolayer MoS_2 bandgap modulation by dielectric environments and tunable bandgap transistors. *Sci. Rep.* **6**, 29184 (2016).
16. 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).
17. Kshirsagar, C. U. et al. Dynamic memory cells using MoS_2 field-effect transistors demonstrating femtoampere leakage currents. *ACS Nano* **10**, 8457–8464 (2016).
18. Wang, H. et al. Large-scale 2D electronics based on single-layer MoS_2 grown by chemical vapor deposition. *2012 IEEE Int. Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM.2012.6478980 (IEEE, 2012).
19. Smithe, K. K. H., Suryavanshi, S. V., Rojo, M. M., Tedjarati, A. D. & Pop, E. Low variability in synthetic monolayer MoS_2 devices. *ACS Nano* **11**, 8456–8463 (2017).
20. Wachter, S., Polyushkin, D. K., Bethge, O. & Mueller, T. A microprocessor based on a two-dimensional semiconductor. *Nat. Commun.* **8**, 14948 (2017).
21. Chhowalla, M., Jena, D. & Zhang, H. Two-dimensional semiconductors for transistors. *Nat. Rev. Mat.* **1**, 16052 (2016).
22. Yang, R., Wang, Z. & Feng, P. X.-L. Electrical breakdown of multilayer MoS_2 field-effect transistors with thickness-dependent mobility. *Nanoscale* **6**, 12383–12390 (2014).
23. English, C. D., Smithe, K. K. H., Xu, R. L. & Pop, E. Approaching ballistic transport in monolayer MoS_2 transistors with self-aligned 10 nm top gates. *2016 IEEE Int. Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM.2016.7838355 (IEEE, 2016).
24. Liu, H., Neal, A. T. & Ye, P. D. Channel length scaling of MoS_2 MOSFETs. *ACS Nano* **6**, 8563–8569 (2012).
25. Frank, D. J., Taur, Y. & Wong, H.-S. P. Generalized scale length for two-dimensional effects in MOSFETs. *IEEE Electron Device Lett.* **19**, 385–387 (1998).
26. Yu, Z. et al. Towards intrinsic charge transport in monolayer molybdenum disulfide by defect and interface engineering. *Nat. Commun.* **5**, 5290 (2014).
27. 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).
28. 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).
29. Cui, X. et al. Multi-terminal transport measurements of MoS_2 using a van der Waals heterostructure device platform. *Nat. Nanotechnol.* **10**, 534–540 (2015).
30. Wong, H.-S. P. et al. Metal-oxide RRAM. *Proc. IEEE* **100**, 1951–1970 (2012).
31. Goux, L. et al. On the gradual unipolar and bipolar resistive switching of TiN/HfO_2/Pt memory systems. *Electrochem. Solid-St. Lett.* **13**, G54–G56 (2010).
32. Wu, H. et al. Stable self-compliance resistive switching in $AlO_\delta/Ta_2O_{5-x}/TaO_y$ triple layer devices. *Nanotechnology* **26**, 035203 (2015).
33. Yu, S., Chen, H.-Y., Gao, B., Kang, J. & Wong, H.-S. P. HfO_x-based vertical resistive switching random access memory suitable for bit-cost-effective three-dimensional cross-point architecture. *ACS Nano* **7**, 2320–2325 (2013).
34. Govoreanu, B. et al. 10×10 nm^2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation. *2011 IEEE Int. Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM.2011.6131652 (IEEE, 2011).
35. Lee, M.-J. et al. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures. *Nat. Mater.* **10**, 625–630 (2011).
36. Dong, W., Liu, D., Xu, S., Chen, B. & Zhao, Y. Demonstrate high R_{OFF}/R_{ON} ratio and forming-free RRAM for rFPGA application based on switching layer engineering. *2017 IEEE 12th Int. Conf. on ASIC (ASICON)* https://doi.org/10.1109/ASICON.2017.8252610 (IEEE, 2017).
37. Nail, C. et al. Understanding RRAM endurance, retention and window margin trade-off using experimental results and simulations. *2016 IEEE Int. Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM.2016.7838346 (IEEE, 2016).
38. Yang, R. et al. 2D molybdenum disulfide (MoS_2) transistors driving RRAMs with 1T1R configuration. *2017 IEEE Int. Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM.2017.8268423 (IEEE, 2017).
39. Smithe, K. K. H., English, C. D., Suryavanshi, S. V. & Pop, E. Intrinsic electrical transport and performance projections of synthetic monolayer MoS_2 devices. *2D Mater.* **4**, 011009 (2017).
40. Sung, C., Song, J., Lee, S. & Hwang, H. Improved endurance of RRAM by optimizing reset bias scheme in 1TIR configuration to suppress reset breakdown. *2016 IEEE Silicon Nanoelectronics Workshop (SNW)* https://doi.org/10.1109/SNW.2016.7577996 (IEEE, 2016).
41. Matsunaga, S. et al. Complementary 5T-4MTJ nonvolatile TCAM cell circuit with phase-selective parallel writing scheme. *IEICE Electronics Express* **11**, 20140297 (2014).
42. Li, H., Wu, T. F., Mitra, S. & Wong, H.-S. P. Resistive RAM-centric computing: design and modeling methodology. *CIEEE Trans. Circuits Syst. I: Reg. Papers* **64**, 2263–2273 (2017).
43. Li, H. et al. Device and circuit interaction analysis of stochastic behaviors in cross-point RRAM arrays. *IEEE Trans. Electron Devices* **64**, 4928–4936 (2017).
44. Cao, Y., Sato, T., Orshansky, M., Sylvester, D. & Hu, C. New paradigm of predictive MOSFET and interconnect modeling for early circuit simulation. *Proc. IEEE 2000 Custom Integrated Circuits Conf. (CICC)* https://doi.org/10.1109/CICC.2000.852648 (IEEE, 2000).
45. Ge, R. et al. Atomristor: nonvolatile resistance switching in atomic sheets of transition metal dichalcogenides. *Nano Lett.* **18**, 434–441 (2018).

Acknowledgements

We acknowledge the discussions with T. Phan, A. Tang, Y. Nishi, K. Saraswat and Y. Chai. We acknowledge the support from the Air Force Office of Scientific Research (AFOSR) Multidisciplinary University Research Initiative (MURI) under award FA9550-16-1-0031, the National Science Foundation (NSF) EFRI 2-DARE grant 1542883, the AFOSR grant FA9550-14-1-0251, the Initiative for Nanoscale Materials and Processes (INMP), the Stanford Non-volatile Memory Technology Research Initiative (NMTRI), and Stanford SystemX Alliance. This work was supported in part by ASCENT, one of six centres in JUMP, a Semiconductor Research Corporation (SRC) programme sponsored by DARPA. R.Y. thanks University of Michigan–Shanghai Jiao Tong University Joint Institute at Shanghai Jiao Tong University for their financial support. K.K.H.S. acknowledges partial support from the Stanford Graduate Fellowship programme and the NSF Graduate Research Fellowship (no. DGE-114747). We are grateful to Stanford Nanofabrication Facility (SNF) and Stanford Nano Shared Facilities (SNSF) for providing the equipment for device fabrication and measurements.

Author contributions

R.Y. and H.-S.P.W. conceived the concept and designed the research. R.Y. performed the fabrication and the electrical measurements of the 2T2R structure. H.L. assisted in the electrical measurement and performed the simulation of the TCAM array. K.K.H.S. performed the CVD growth of MoS_2. T.R.K. and K.O. took the TEM images. E.P., J.A.F. and H.-S.P.W. provided feedback on the experiments. R.Y. and H.-S.P.W. wrote the manuscript, with input from E.P. and J.A.F. All authors have discussed the results and given approval to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41928-019-0220-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.Y.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.