Supporting Information for

Two-terminal MoS2 Memristor and the Homogeneous Integration with MoS2 Transistor for Neural Networks

Shuai Fu[†], Ji-Hoon Park[‡], Hongyan Gao[†], Tianyi Zhang[‡], Xiang Ji[‡], Tianda Fu[†], Lu Sun[†], Jing Kong[‡], Jun Yao[†,§,⊥,*]

[†]Department of Electrical Computer and Engineering, University of Massachusetts, Amherst, MA 01003, USA.

[‡]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[§]Institute for Applied Life Sciences (IALS), University of Massachusetts, Amherst, MA 01003, USA.

[⊥]Department of Biomedical Engineering, University of Massachusetts, Amherst, MA 01003, USA.

* Corresponding author. Emails: juny@umass.edu (J.Y.)

This file includes:

Materials and Methods

Supporting Figures S1 – S9

Supporting references

MATERIALS AND METHODS

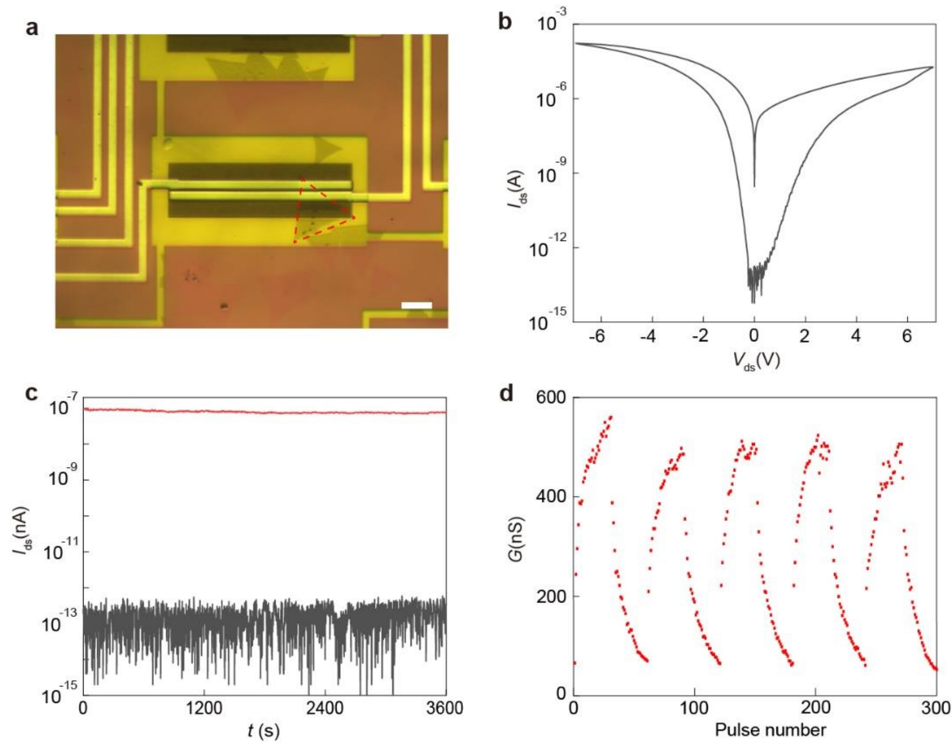
MoS_2 synthesis. Polycrystalline MoS_2 films were grown under low pressure by a metal-organic-chemical-vapor-deposition method.[1] Gas-phase molybdenum hexacarbonyl (98%, Sigma Aldrich) and diethyl sulfide (98%, Sigma Aldrich) serving as precursors of Mo and S, respectively, were supplied into the furnace with a bubbler system using Ar as the carrier gas. A 300 nm-thick SiO_2/Si wafer was used as the growth substrate. The growth took place at the temperature of 350 ^{o}C for 15 hr, with the continuous flow of Ar (100 sccm), molybdenum hexacarbonyl (0.6 sccm), and diethyl sulfide (2.0 sccm). For the synthesis of single-crystalline monolayer MoS_2 flakes, a liquid-phase precursor-assisted CVD method was applied.[2] Powders of MoO_3 (25 mg) and KI (25 mg) were dissolved in ammonia (20 mL), and then spin-coated onto freshly cleaned SiO_2/Si substrates. Subsequently, the precursor-coated substrates were loaded into a tube furnace and sulfurized at 720 ^{o}C for 5 min. Ar was used as the carrier gas throughout the process, with a flow rate of 20 sccm.

Fabrication of MoS_2 devices. For MoS_2 memristor, a global back gate (Ti/Au, 3/45 nm) was first defined by standard electron-beam lithography (EBL), metal deposition, and lift-off processes on a Si substrate covered with 600-nm SiO_2. Then a 30-nm Al_2O_3 was grown by atomic layer deposition (ALD). A floating gate (5-nm Au) was defined by EBL, metal deposition, and lift-off processes. Another 7-nm Al_2O_3 (tunneling layer) was growth by ALD. The as-grown MoS_2 film was transferred onto the Al_2O_3 layer following previous procedure.[1] After transfer, a polymethyl methacrylate (PMMA) sacrificial layer was spin-coated on the MoS_2 layer and patterned by EBL. Oxygen plasma (70 W, 0.8 sccm Ar + 20 sccm O_2, 40 s) was used to etch away unprotected MoS_2 to define device channel, before the PMMA sacrificial layer was removed in acetone. Source and drain contacts (Ti/Au, 3/50 nm) were defined by EBL, metal deposition, and life-off processes. The same procedure was employed for fabricating MoS_2 transistors, only that the floating gate was not included.

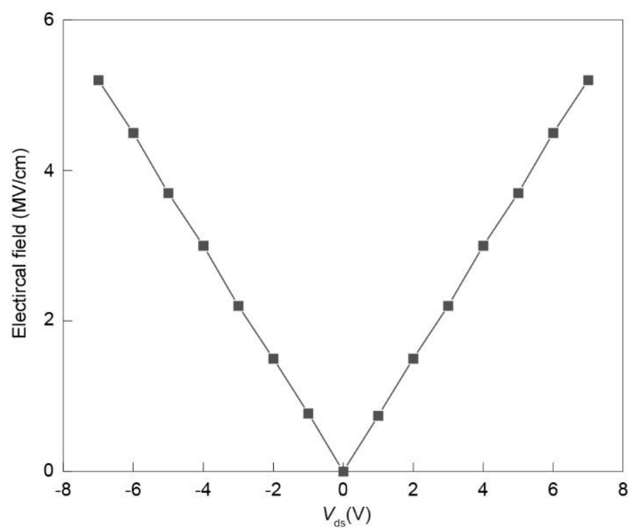
Electrical measurements: The electrical measurements were performed in vacuum (2×10^{-4} Torr) and dark condition using a probe station (Janis ST-500-6TX, Lake Shore Cyrotronics, Inc.). The *I-V* curve measurements and pulse tests were performed by using a semiconductor parameter analyzer (Agilent 4155C).

Simulation of the field distribution. The electric field distribution in the MoS_2 memristor structure was simulated by using the COMSOL Multiphysics 6.0 with the electrostatic (ES) module. The physical parameters in the Au electrodes, Al_2O_3 dielectric layer, and Au floating gate were directly retrieved from the material library of COMSOL, whereas those in the MoS_2 film were obtained from previous study.[3] The geometric dimensions took the actual sizes in the fabricated device.

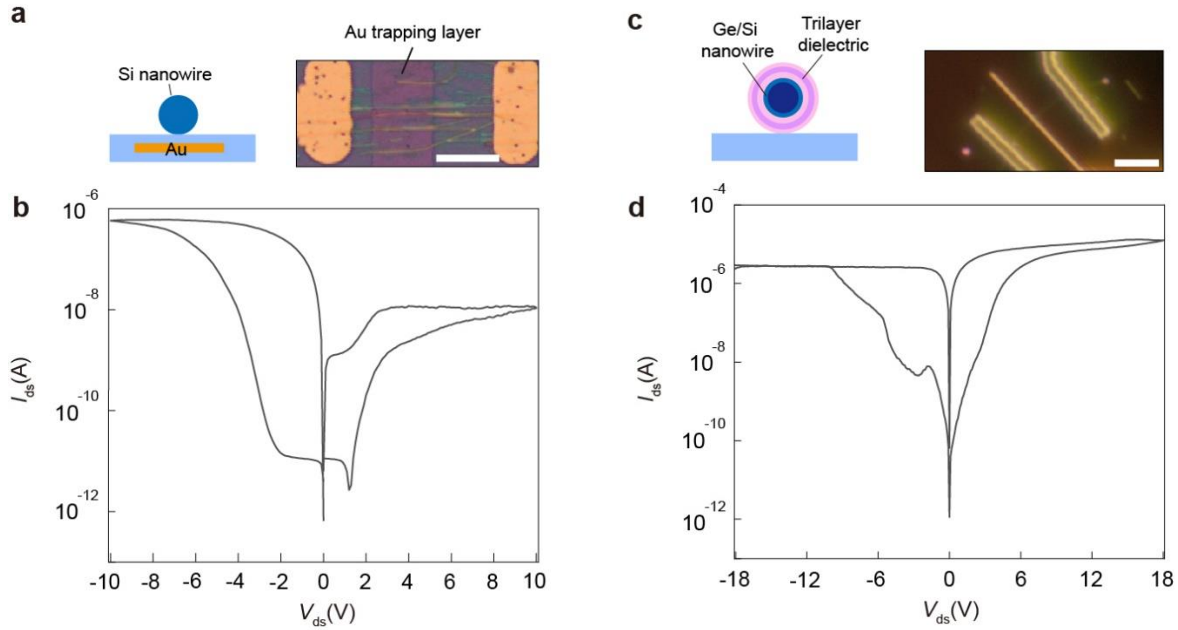
Neural network simulation. The construction and training of the simulated neural network are detailed in Supplementary Information Figure S7.



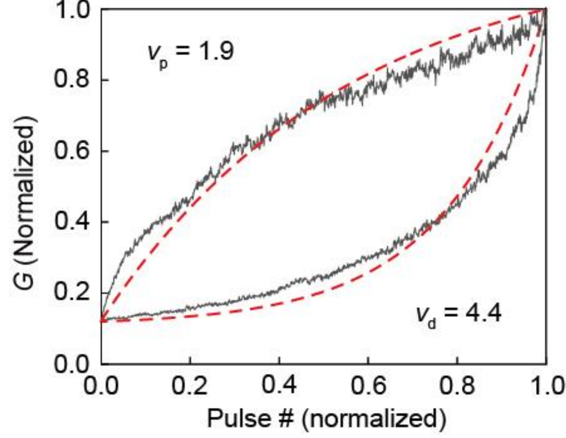
Supplementary Figure S1. Two-terminal memristive effect in single-crystalline MoS_2 device. **a**, Optical image of a memristor device fabricated from single-crystalline MoS_2 (delineated by the dashed triangle). The device shared the same structure as described in Fig. 1a in the main article. Scale bar, 10 µm. **b**, Two-terminal I_{ds}-V_{ds} sweeps, showing hysteresis characteristic of memristive effect. **c**, Retention (with a reading voltage of 0.05 V) in the programmed low-resistance state (LRS) and high-resistance state (HRS) from the MoS_2 memristor. The LRS and HRS were programmed with V_{ds} = +10 V (9 s) and V_{ds} = -10 V (9 s), respectively. **d**, Analog conductance modulation in the MoS_2 memristor. Each cycle involved 30 Set pulses (V_{ds} = +6.8 V, 200 ms) and 30 Reset pulses (V_{ds} = -5.3 V, 200 ms), with the conductance read out by 0.05 V pulses.



Supplementary Figure S2. Strength of the maximal vertical electrical field component at the edge of the drain electrode with respect to applied drain voltage (V_{ds}), revealed by simulation using device parameters same as those in the fabricated MoS_2 device. Details can be found in the *Materials and Methods* section.



Supplementary Figure S3. Two-terminal memristive effect in semiconducting nanowires. **a**, (Left) Cross-section schematic of a Si nanowire device with a 5-nm-thick Au charge-trapping layer defined underneath (*e.g.;* separated by a 7-nm tunneling Al_2O_3 layer). (Right) Bright-field optical image of a Si nanowire device. Scale bar, 10 µm. **b**, Representative two-terminal I_{ds}-V_{ds} sweeps from a Si nanowire device, showing hysteresis characteristic of the memristive effect. **c**, (Left) Cross-section schematic of a Ge/Si core-shell nanowire device. A trilayer dielectric of Al_2O_3–ZrO_2–Al_2O_3 (2–5–5 nm) was coated on the Si/Ge nanowire by atomic layer deposition to serve as the charge trapping layer.[1] (Right) Dark-field optical image of a Ge/Si nanowire device. Scale bar, 5 µm. **d**, Representative two-terminal I_{ds}-V_{ds} sweeps from a Ge/Si nanowire device, showing hysteretic memristive effect. The Si and Ge/Si nanowires were synthesized by a catalyzed chemical vapor deposition method described previously.[4,5] The nanowires were assembled by contact printing method.[6] The devices were fabricated following previous procedures involving lithography, metal evaporation, and lift-off processes.[4,5]



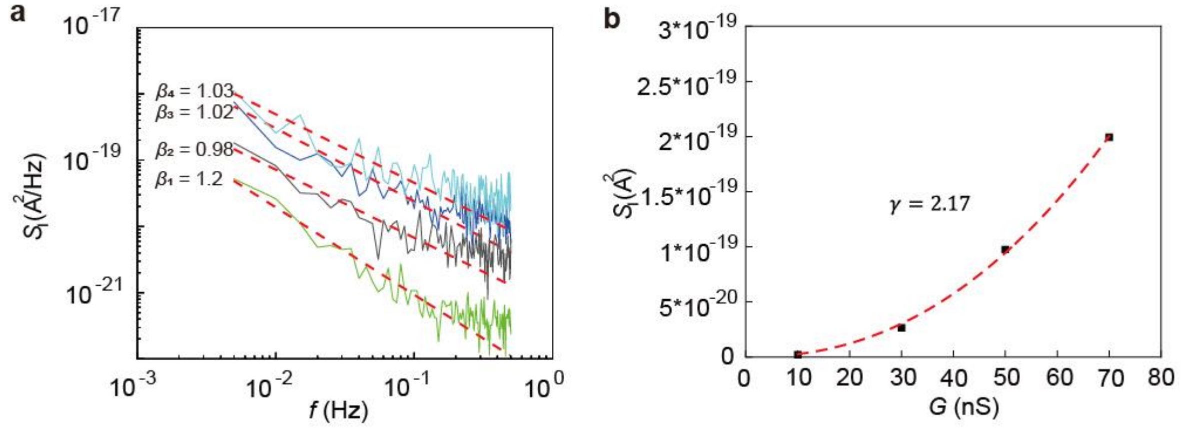
Supplementary Figure S4. Linearity in conductance modulation (based on Fig. 1e in the main paper). The nonlinearity factor (v) is usually introduced to characterize the linearity in multi-state updating. The (normalized) conductance updates during potentiation (G_p) and depression (G_d) are generally described by following relationships[7]:

$$G_p = G_{min} + B * (1 - e^{-v*p}) \qquad (1)$$

$$G_d = G_{max} - B * (1 - e^{v*(p-p_{max})}) \qquad (2)$$

$$B = \frac{G_{max} - G_{min}}{1 - e^{-v*p_{max}}} \qquad (3)$$

Here G_{min} and G_{max} are the minimum and maximum (normalized) conductance, respectively. p and p_{max} are the (normalized) number of applied and maximum pulses, respectively. Based on above equations, the nonlinearity factors extrapolated from fitting (dashed line) during potentiation (v_p) and depression (v_d) are approximately 1.9 and 4.4, respectively. The overall linearity is comparable or better than those estimated value sets of (v_p = 4.5, v_d = 3.8)[8], (v_p = 4, v_d = 6.1)[9], (v_p = 1.2, v_d = 4.5)[10], (v_p = 1.8, v_d = 17.2)[11] from some other MoS_2 memristors based on mechanisms of ion/defect migrations.

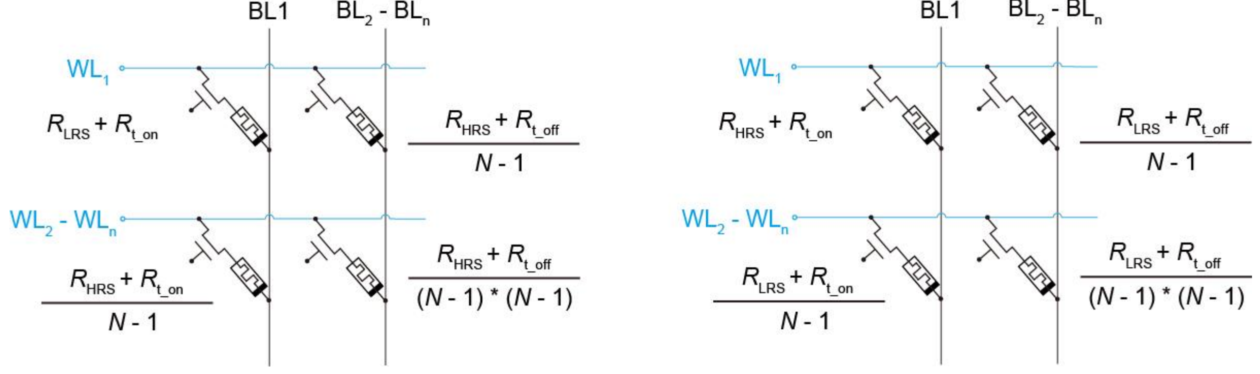


Supplementary Figure S5. Noise spectrum in measured conductance levels. The flicker (or 1/*f*) noise can be expressed by Hooge empirical law[12]:

$$S_I = \frac{AI^\gamma}{f^\beta} \quad ,$$

where S_I, *A,* and *I* represent the noise spectral density, amplitude, and mean current, respectively. The exponents, β and γ have expected values close to 1 and 2, respectively.

a, Fittings (dashed lines) showing that the β values are all close to 1 (1.2, 0.98, 1.02, 1.03) in some representative conductance levels (10, 30, 50, 70 nS) measured from a MoS_2 memristor (Fig. 1g in the main paper). **b**, Considering that A does not change substantially[12], the fitting (dashed line) shows $S_I \propto I^\gamma$ ($\gamma \sim 2.17$, at *f* = 0.2 Hz). These results confirm the 1/*f* origin of the noise observed in the MoS_2 devices.



Supplementary Figure S6. Read margin (RM) analysis in a $N \times N$ array. The current-based RM is defined as[13]:

$$RM = \frac{I_{LRS_min} - I_{HRS_max}}{I_{LRS_min}},$$

where I_{LRS_min} and I_{HRS_max} are the minimum and maximum read-out currents from the selected cell programmed with LRS and HRS states, respectively.

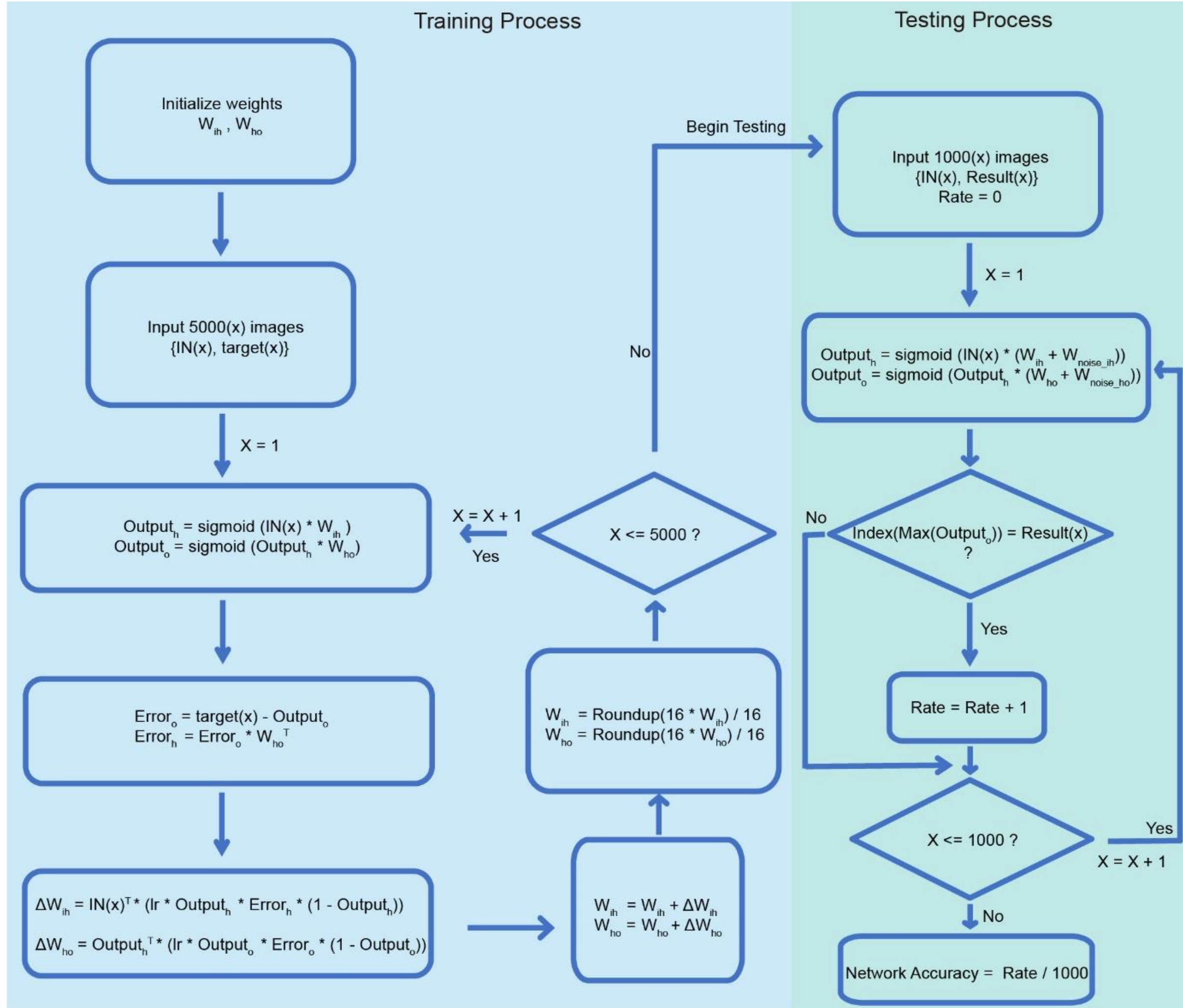
To obtain I_{LRS_min}, the unselected memristors will assume HRS with the resistance R_{HRS}, leading to an equivalent circuit[14] as shown on the left. R_{t_on} and R_{t_off} are the On- and Off-state resistances in the transistor, respectively. The corresponding read-out current can be then expressed as:

$$I_{LRS_min} = \frac{V}{R_{t_on} + R_{LRS}} + \frac{V}{\frac{R_{t_off} + R_{HRS}}{N-1} + \frac{R_{t_off} + R_{HRS}}{(N-1)*(N-1)} + \frac{R_{t_on} + R_{HRS}}{N-1}}.$$

Likewise, to obtain I_{HRS_max}, the unselected memristors will assume LRS with the resistance R_{LRS}, leading to an equivalent circuit[14] as shown on the right. The corresponding read-out current can be then expressed as

$$I_{HRS_max} = \frac{V}{R_{t_on} + R_{HRS}} + \frac{V}{\frac{R_{t_off} + R_{LRS}}{N-1} + \frac{R_{t_off} + R_{LRS}}{(N-1)*(N-1)} + \frac{R_{t_on} + R_{LRS}}{N-1}}.$$

The RM can be estimated by using above equations, with R_{LRS} (~2 MΩ), R_{HRS} (~100 MΩ), R_{t_on} (~0.2 MΩ), and R_{t_off} (~100 GΩ) obtained from experimental data in fabricated MoS_2 devices. A typical RM>10% is considered valid for memory array addressing [15].



Supplementary Figure S7. Training and testing of the emulated neural network.[16,17]

For the three-layer neural network, the input layer and hidden layer is connected by a 784×200 weight matrix and the hidden layer and the output layer by a 200×100 weight matrix. The weight matrices are constructed from the 1T1R MoS_2 cells arranged in a crossbar structure (Fig. 3a). The output vector from each layer is calculated by using the general multiplication rule.[17]

Specifically, the weight matrix between the input neuron *i* and hidden neuron *h* (W_{ih}), the weight matrix between the hidden neuron *h* and the output neuron *O* (W_{ho}), are initialized between -0.1 to 0.1 (*e.g.;* set to -0.05). Each image pixel (with a grayscale of 0-255) is converted to an input value between 0-1. Therefore, an input image x (28×28=784 pixels) is converted to an input row matrix $IN(x)$ of 784 values. The 200-value output in the hidden neuron and 10-value output in the output neuron are represented by one-row matrices of $Output_h(x)$ and $Output_o(x)$, respectively. A 10-value matrix $target(x)$ represents the ideal expectation (*e.g.;* [0,1,0,0,0,0,0,0,0,0] for digit-1 image). Using matrix multiplication and activation function, the output matrix of the hidden neuron $Output_h(x)$ and output neuron $Output_o(x)$ can be expressed as:

$$Output_h(x) = sigmoid\ (IN(x) \times W_{ih}),$$

$$Output_o(x) = sigmoid\ (Output_h \times W_{ho}),$$

where the activation (sigmoid) function is defined as:

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

The error matrix of the output neuron $Error_o(x)$ and hidden neuron $Error_h(x)$ are calculated using matrix multiplication and subtraction as:

$Error_o(x) = target(x) - Output_o(x)$,

$Error_h(x) = Error_o(x) \times W_{ho}{}^T$,

where $W_{ho}{}^T$ is the transposed matrix of W_{ho}. The weight change of each weight matrix (ΔW_{ih} and ΔW_{ho}) is calculated by using:

$\Delta W_{ih} = IN(x)^T \times (lr \cdot Output_h(x) \times Error_h(x) \times (1 - Output_h(x)))$,

$\Delta W_{ho} = Output_h(x)^T \times (lr \cdot Output_o(x) \times Error_o(x) \times (1 - Output_o(x)))$,

where lr is the learning rate (taken as 1 in our case). The weight is updated by:

$W_{ih} = W_{ih} + \Delta W_{ih}$,

$W_{ho} = W_{ho} + \Delta W_{ho}$.

For the discrete weight value in a realistic MoS_2 memristor, the updated weight is rounded up to the closest weight level. For the network taking a 4-bit (16-level) weight precision, the roundup yields:

$W_{ih} = \frac{\text{Roundup}(16 \cdot W_{ih})}{16}$,

$W_{ho} = \frac{\text{Roundup}(16 \cdot W_{ho})}{16}$.

If the calculated weight is >1 or < -1, it is taken remained at 1 or -1.

Above procedure describes a one-time training. For each epoch, a total of 5000 images are used for the training. After each epoch, 1000 images are used for recognition tests.

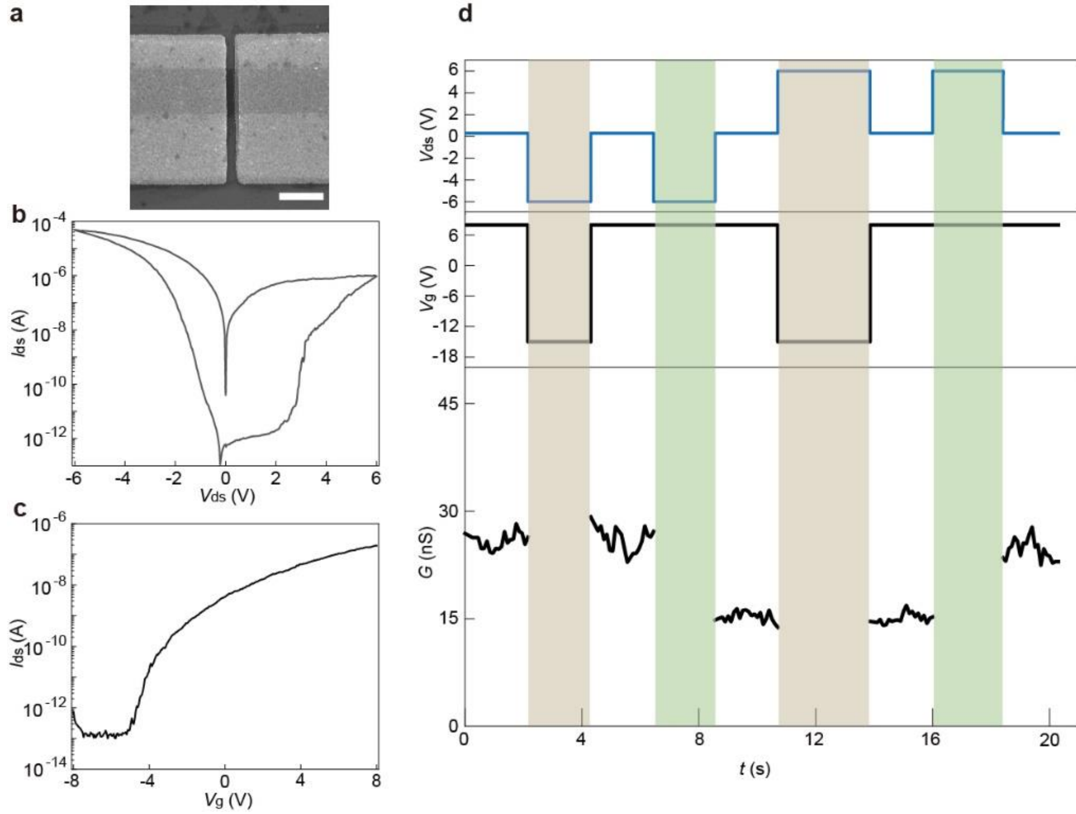
To investigate the influence of noise in the network, noise matrix (W_{noise_ih} and W_{noise_ho}) is generated by using random number generating function (*e.g.;* with Python) and applied them to the calculated weigh matrices:

$Output_h(x) = sigmoid\left(IN(x) \times (W_{ih} + W_{noise_ih})\right)$,

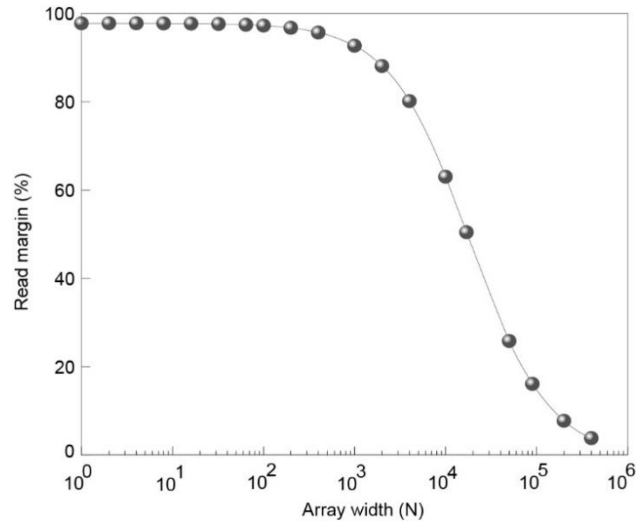
$Output_o(x) = sigmoid\left(Output_h \times (W_{ho} + W_{noise_ho})\right)$.

For realistic emulation, a new set of noise matrices is always generated and added for the calculation of each $Output_o$. The index of the element that has the maximal value in the $Output_o$ matrix is considered to correspond to the recognized digit. And if the result is correct, the countering $Rate$ adds one. After we finish the 1000-image test, the recognition accuracy is defined as:

$Network\ accuracy = \frac{Rate}{1000}$.



Supplementary Figure S8. Device performance at smaller scale. **a**, Scanning electron microscope (SEM) image of a representative device (0.2×1 μm^2) fabricated from single-crystalline monolayer MoS_2 flakes. Scale bar, 1 μm. **b**, Two-terminal *I-V* sweep in a MoS_2 memristor device of the same size, showing drain voltage (V_{ds})-dependent hysteresis. **c**, A representative transport curve (V_{ds} = -0.3 V) in a MoS_2 transistor of the same size, showing maintained On/Off ratio ~10^6. **d**, Demonstration of selective Set and Reset programming in an integrated 1T1R cell (*i.e.;* by connecting the memristor and transistor). The operational detail is similar to that described in the main paper.



Supplementary Figure S9. Read margin analysis in $N \times N$ arrays with device scaling. The performance of scaling MoS_2 device is retrieved from pervious study.[14] For a 40-nm channel length, the On and Off current were revealed to be ~0.1 mA/µm and 1 nA/µm, respectively (V_{ds} = 1 V).[18] As a result, for a scaled MoS_2 transistor (*e.g.*; 40-nm channel length and 50-nm width), the On and Off resistances are expected to be ~ 0.2 MΩ and 20 GΩ, respectively. A modulable conductance range of 2 MΩ ~ 100 MΩ is considered in the MoS_2 memristor. The estimate of the read margin follows the same procedure as described in Supplementary Figure S6.

Supplementary references

1. Park, J. H.; Lu, A. Y.; Shen, P. C.; Shin, B. G.; Wang, H. Z.; Mao, N. N.; Xu, R. J.; Jung, S. J.; Ham, D.; Kern, K.; Han, Y. M.; Kong, J. Synthesis of high-performance monolayer molybdenum disulfide at low temperature. *Small Methods* **2021,** 5, 2000720.
2. Zhang, T. Y.; Fujisawa, K.; Zhang, F.; Liu, M. Z.; Lucking, M. C.; Gontijo, R. N.; Lei, Y.; Liu, H.; Crust, K.; Granzier-Nakajima, T.; Terrones, H.; Elias, A. L.; Terrones, M. Universal in situ substitutional doping of transition metal dichalcogenides by liquid-phase precursor-assisted synthesis. *ACS Nano* **2020**, 14, 4326-4335.
3. Howell, S. L.; Jariwala, D.; Wu, C. C.; Chen, K. S.; Sangwan, V. K.; Kang, J. M.; Marks, T. J.; Hersam, M. C.; Lauhon, L. J. Investigation of band-offsets at monolayer-multilayer MoS_2 junctions by scanning photocurrent microscopy. *Nano Lett*. **2015**, 15, 2278-2284.
4. Yao, J.; Yan, H.; Das, S.; Klemic, J.F.; Ellenbogen, J.C.; Lieber, C.M. Nanowire nanocomputer as a finite-state machine. *Proc Nat Acad Sci* **2014**, *111*, 2431-2435.
5. Gao, H.; Yin, B.; Wu, S.; Liu, X.; Fu, T.; Zhang, C.; Lin, J.; Yao, J. Deterministic assembly of three-dimensional suspended nanowire structures. *Nano Lett.* **2019**, *19*, 5647-5652.
6. Fan, Z.; Ho, J.C.; Jacobson, Z.A.; Yerushalmi, R.; Alley, R.L.; Razavi, H.; Javey, A. Wafer-scale assembly of highly ordered semiconductor nanowire arrays by contact printing. *Nano Lett.* **2008**, *8*, 20-25.
7. Tang, J.; Bishop, D.; Kim, S.; Copel, M.; Gokmen, T.; Todorov, T.; Shin, S.; Lee, K.T.; Solomon, P.; Chan, K.; Haensch, W. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing. *2018 IEEE International Electron Devices Meeting (IEDM)* **2018,** pp. 13.1.1-13.1.4.
8. Li, D.; Wu, B.; Zhu, X.; Wang, J.; Ryu, B.; Lu, W.D.; Lu, W.; Liang, X. MoS2 memristors exhibiting variable switching characteristics toward biorealistic synaptic emulation. *ACS nano* **2018**, *12*, 9240-9252.
9. Feng, X.; Li, S.; Wong, S.L.; Tong, S.; Chen, L.; Zhang, P.; Wang, L.; Fong, X.; Chi, D.; Ang, K.W. Self-selective multi-terminal memtransistor crossbar array for in-memory computing. *ACS nano* **2021**, *15*, 1764-1774.
10. Zhu, X.; Li, D.; Liang, X.; Lu, W.D. Ionic modulation and ionic coupling effects in MoS2 devices for neuromorphic computing. *Nat. Mater* **2019**, *18*, 141-148.
11. Sangwan, V.K.; Lee, H.S.; Bergeron, H.; Balla, I.; Beck, M.E.; Chen, K.S.; Hersam, M.C. Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide. *Nature* **2018**, 554, 500-504.
12. Sangwan, V.K.; Arnold, H.N.; Jariwala, D.; Marks, T.J.; Lauhon, L.J.; Hersam, M.C. Low-frequency electronic noise in single-layer MoS2 transistors. *Nano Lett.* **2013**, *13*, 4351-4355.
13. Sun, W.; Shin, H. Analysis of read margin of crossbar array according to selector and resistor variation. *International Conference on Electronics, Information, and Communication (ICEIC)* **2018**, pp. 1-3.
14. Rao, M.; Song, W.; Kiani, F.; Asapu, S.; Zhuo, Y.; Midya, R.; Upadhyay, N.; Wu, Q.; Barnell, M.; Lin, P.; Li, C. Timing selector: Using transient switching dynamics to solve the sneak path issue of crossbar arrays. *Small Science* **2022**, 2, 2100072.
15. Zhang, L., Cosemans, S., Wouters, D.J., Groeseneken, G., Jurczak, M. and Govoreanu, B. Cell variability impact on the one-selector one-resistor cross-point array performance. *IEEE Trans. Electron Dev.* **2015**, 62, 3490-3497.
16. Wang, Y.; Tang, H.; Xie, Y.; Chen, X.; Ma, S.; Sun, Z.; Sun, Q.; Chen, L.; Zhu, H.; Wan, J.; Xu, Z. An in-memory computing architecture based on two-dimensional semiconductors for multiply-accumulate operations. *Nat. Commun.* **2021**, 12, 3347.
17. Tang, J.; He, C.; Tang, J.; Yue, K.; Zhang, Q.; Liu, Y.; Wang, Q.; Wang, S.; Li, N.; Shen, C.; Zhao, Y. A reliable all-2D materials artificial synapse for high energy-efficient neuromorphic computing. *Adv. Funct. Mater.* **2021**, 31, 2011083.
18. Arutchelvan, G.; Smets, Q.; Verreck, D.; Ahmed, Z.; Gaur, A.; Sutar, S.; Jussot, J.; Groven, B.; Heyns, M.; Lin, D.; Asselberghs, I. Impact of device scaling on the electrical properties of MoS_2 field-effect transistors. *Sci. Rep.* **2021**, 11, 6610.