Enabling Reliable Two-terminal Memristor Network by Exploiting the Dynamic Reverse Recovery in a Diode Selector

Tianda Fu[1,4], Shuai Fu[1,4], Siqi Wang[1], Jun Yao[1,2,3,5]*

1. Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003, USA.
2. Institute for Applied Life Sciences, University of Massachusetts, Amherst, MA 01003, USA.
3. Department of Biomedical Engineering, University of Massachusetts, Amherst, MA 01003, USA.
4. These authors contributed equally.
5. Lead contact.

* Correspondence: juny@umass.edu

SUMMARY

Transistor is broadly used to address memristor networks, but its three-terminal structure can impose limitations on fully exploiting the potential of efficient integration a two-terminal memristor can offer. While a two-terminal selector is desirable for unlocking this potential, no existing device has attained a similar level of functional maturity. Diode, despite the technological maturity, is still limited by its unipolarity in addressing mainstream bipolar memristors. Here, we demonstrate that a diode can be implemented as a bidirectional selector for constructing two-terminal memristor architecture by exploiting its reverse recovery dynamics. This is demonstrated by the construction of one-diode-one-memristor (1D1R) programmable arrays, which are implemented for *in situ* neural training and classification. Furthermore, a crossbar array made from stacking 1D1R cells is fabricated to demonstrate scalable integration. This dynamic paradigm combines the advantages of functional maturity and structural simplicity of diode selectors to improve the development of memristor integration.

KEYWORDS

neuromorphic, sneak path, crossbar, neural network, reservoir computing, transistor

INTRODUCTION

Neuromorphic computing based on memristors has the potential to improve information processing by circumventing the data shuttle in traditional Von-Neumann architectures [1-8]. Memristors built from emergent new materials (e.g., 2D materials) have further expanded the potential in functions and integration [9-13]. The enabled vector-matrix multiplication (VMM) in the integrated memristor array can parallel previous computationally expensive sequences for efficient throughput [5-8,14,15]. However, the precision in the network programming and subsequent VMM operation heavily depends on the effectiveness in suppressing the sneak-path conduction, which otherwise can yield programming and reading errors to impair computing accuracy[16,17]. One common solution is to pair a transistor with each memristor to control the local current passage. This one-transistor-one-memristor (1T1R) strategy has been successfully implemented in many systems for various functional demonstrations [8,13,14,18-23]. However, the three-terminal transistor adds structural and addressing complexity to the architecture. A two-terminal selector is highly desirable to fully exploit the benefit of simplicity and compactness from a crossbar architecture [16,24,25]. Nevertheless, no existing two-terminal selector has achieved a functional maturity close to that of a transistor for practical implementation [7]. For example, many demonstrated selectors based on filamentary switching devices lack the stability for high endurance and uniformity required in a programmable network [16,26]. Although the diode features a technology maturity similar to the transistor, its unipolarity only allows for addressing unipolar memristors [16,27] but prevents compatibility with the mainstream networks built from bipolar memristors [8,14,18-23]. Bidirectional tunneling devices are limited in current density and introduce large nonlinearity that can prevent the direct use of VMM [16]. During programing, the rest unselected cells are often applied with a voltage to the half or third value of the programming voltage V applied to the selected cell, in order to offset the voltage drop across these cells for avoiding unintended state change; this 1/2V or 1/3V programming theme can substantially increase power consumption in array applications [16].

Despite the limited success, the potential benefits of a two-terminal selector have continued to stimulate the exploration of new concepts for better solutions. One general strategy is to explore the transient dynamics in some devices for timed addressing [24,28]. For example, the delay time in a volatile bipolar memristor, which is inversely related to the applied voltage, is employed for the construction of a timing selector [28]. The selectors along the sneak path, due to the voltage-divider effect, are activated slower than the one in the selected path. This time difference provides a transient window for selective programming. In another strategy, the retention time in a volatile unipolar memristor with the ultralow threshold [29] is exploited to enable bidirectional addressing [24]. On the one hand, the forward selective addressing is readily enabled by the unidirectional conduction in the switch. On the other hand, the retention time also creates a transient window for the reverse addressing. These strategies, however, still suffer from instability and nonuniformity resulted from the inherent stochasticity in the filamentary selectors for practical implementations [30]. Nonetheless, they do offer an insightful perspective: if the transient dynamics of a mature device technology are harnessed, it may lead to the practical solution of a reliable two-terminal selector.

Here, we demonstrate that the transient reverse recovery time (T_{rr}) in a diode[31,32] can be exploited to enable bidirectional addressing for constructing reliable one-diode-one-memristor (1D1R) programmable memristor networks. This strategy can work with all memristors (e.g., both

unipolar and bipolar types), thus greatly expanding the capacity in previous 1D1R architecture that only works with unipolar memristors [16,27]. Meanwhile, the number of addressing lines (2*N*) in a 1D1R architecture (e.g., with a *N*×*N* array size) is reduced from that (3*N*) in the 1T1R architecture. As a result, this strategy has the combined advantages of both functional maturity and structural simplicity for practical applications. We validate the strategy by connecting commercial diodes with the mainstream Ta-HfO_2 nonvolatile memristors [14,19,22,33] to construct an 8×8 array for programmable network. The programmability is analyzed and demonstrated by *in situ* training of the network for digit classifications. The functional validation is followed by the fabrication and integration of 1D1R stacking cells in a crossbar network to demonstrate the potential for integrated neuromorphic computing. This 1D1R integration can also be readily used for constructing two-terminal crossbar memory array.

RESULTS

1. General concept

A 2×2 array that is the constituent unit in larger arrays is used to illustrate the working concept. The 1D1R array shares the same structural configuration as previously used for addressing unipolar memristor networks [16,27,34-38], with the rectifying effect in the diode readily exploited to suppresses the sneak-path conduction in a forward operation (Fig. 1A). Here, the fundamental difference is that the reverse recovery time T_{rr} of the diode is harnessed to enable reverse selective addressing (Fig. 1B). Specifically, T_{rr} describes the finite time needed for the diode to discharge before it switches from the forward-biased (On) state to the reverse-biased (Off) state (Fig. 1B (ii)) [31,32]. This transient window allows the reverse current flow (Fig. 1B (iii)) for the reverse programming of the connected memristor. Note that in the sneak path, this transient window does not exist since the two forward-biased diodes are initially in the Off state (Fig. 1B (i)). Together, this enables the selected reverse programming in the network. The blockage of the sneak-path conduction during both the forward and reverse programming suggests that concurrent programming of multiple cells (*e.g.*, along the same column) is feasible. In addition, this selective programming theme does not need the 1/2V or 1/3V method [16], benefiting the power reduction in array applications.

We first employed commercial diodes for the characterization and demonstration. The measured T_{rr} was positively related to the driving voltage and time (Fig. S1), consistent with the general expectation that the internally stored charge in the junction region increases with the input voltage and charging time. The measured T_{rr} of several μs can provide a sufficient time window for the mainstream (*e.g.*, Ta-HfO_2-based) nonvolatile memristors [39] (Fig. S2 and Experimental Procedures). The fabricated Ta-HfO_2 memristors showed a high yield and relatively uniform performance (Fig. S3) for integrated demonstration.

We then constructed a 1D1R cell by connecting the two devices and tested the programmability (Fig. 1C). A 0.7 V reading pulse (t = 50-150 μs) was applied to the Word Line (WL) with the Bit Line (BL) grounded, showing an initial reading current of ~21 μA (Fig. 1C (i)). For the SET programming, a pulse of 1.2 V, 3 μs was then applied (t ~ 200 μs). Note that although the reverse recovery process yielded a reverse current right after the SET pulse (gray region), the absence of applied reverse voltage did not yield any RESET effect. The subsequent reading (t = 250-350 μs)

showed an increased current of ~80 μA, showing the successful SET process. For the RESET programming (Fig. 1C (ii)), a pulse of 0.7 V, 5 μs (t = 200 μs) was first applied to forwardly activate the diode. The pulse was immediately followed by a RESET pulse of -1.2 V, 3 μs (t = 200-203 μs). This reverse pulse was within the T_{rr} window, so it could apply to the cell. The successful RESET programming was confirmed by that the cell changed from an initial reading current of ~98 μA (t = 50-150 μs) to ~34 μA (t = 250-350 μs)) after the RESET pulse.

2. 8×8 1D1R array

We then connected 64 commercial diodes with 64 Ta-HfO_2 memristors to form an 8×8 1D1R programmable array (Figs. S4-6). Without loss of generality, we used a 2×2 array, which is the constituent unit for an arbitrary-sized array[24], to examine the selective reading (Fig. 2A, Experimental Procedures). We considered the worst case when the memristors in the sneak path (black cells) were set to a Low Resistance State (LRS) and the memristor in the selected path (orange cell) was set to a High Resistance State (HRS). A 0.7 V reading voltage was applied between WL1 and BL1 for current readout. The sneak-path current (blue curve) showed negligible value compared to the reading current of ~8 μA (red curve) from the selected cell, showing successful suppression of the sneak-path current by rectification from the diode. Analysis showed that the read margin could maintain a value >60 % with the array size (N×N) increasing to $N > 10^3$ (Fig. S7). We then used a 3×3 array to examine the selective programmability. Using the addressing strategy in Fig. 1C, we could reversibly adjust the conductance in selected memristors (orange) in a 3×3 array without altering states in unselected (black) memristors (Fig. 2B).

The selective programming and state readout were then applied to the entire 8×8 array (Fig. 2C). A 16×32 pixeled "UM" logo was used as the targeting pattern to program. The background was designed to have a gradient, whereas the "U" and "M" letters had fixed weights. This 16×32 pixeled image can be divided into eight 8×8 sub-images. The same 8×8 array was consecutively programmed to the eight sub-images, showing reliable re-programmability needed for *in situ* neural training.

3. Seven-segment-display (SSD) classification

We further evaluated the enabled selective programmability by implementing the 8×8 array as a double-layer neural network. The synaptic weight in a neural network is often represented by the differential conductance (*e.g.*, G^+-G^-) between a pair of memristors[2]. We therefore used a 1×2 array from the entire array as the pair to test the ability in synaptic weight update (left panel, Fig. 3A). The weight increase was achieved by increasing the conductance G^+ in one memristor using SET pulses, whereas the weight reduction was achieved by increasing G^- in the other memristor using RESET pulses. Both cases employ SET programming that does not involve the use of T_{rr}. Only when the conductance of a memristor reaches a maximum, a RESET programming involving T_{rr} is executed (Fig. S8). This strategy minimizes the time budget added by T_{rr}. In this way, more than 100 weight states could be achieved in the unit pair (Fig. S9). The updates were reversible, maintaining good linearity and symmetry during the repeated cycles (right panel, Fig. 3A), which can be attributed to the reliable and confined Ta conductive channel formed in the HfO_2 matrix[33]. These properties showed reliable programmability in the cells for neural networks.

We first used the array for SSD classification (*e.g.*, "1" to "4"). SSD is used for displaying decimal numerals and intensity variation in the segments represents realistic scenarios that require intelligent recognition, providing a good example for evaluating neural network[40]. To emulate the realistic intensity variation, a 21-level grayscale was used for the segments and different noise levels (*e.g.*, the Gaussian distribution σ = 0.5, 1.5, 2.5, and 3.5) were randomly added (Fig. 3B, Fig. S10).

For the paired representation of synaptic weight, the differential output from a pair of neighboring BLs served as one neuronal output (Fig. 3C). The resultant four outputs thus represented the four classifications of the digits. The converted grayscale intensity from the seven segments (plus one constant bias of 0.7 V) served as the vector input to the 8 WLs. An offset value of 0.3 V was applied to each converted analog input to bypass the nonlinear region of the diode. The 21-level grayscale was thus converted to a voltage between 0.3 to 0.7 V. For simplified training, clear SSD digits (*i.e.*, without added noise) were used (Fig. S11, Experimental Procedures). The trained network showed a dispersion in the synaptic weights (Fig. 4A), which was then used to classify the more realistic SSD digits added with noise. Note that the synaptic weight was converted from the reading current (at V = 0.7 V) by the formula: weight = current/(6×10^{-5} A), where 6×10^{-5} A is the conversion coefficient (Experimental Procedures). For the entire database of 400 samples, the network showed excellent recognition rates of 100 %, 99.5 %, and 95.2 % for noise levels (σ) of 0.5, 1.5, and 2.5, respectively (Fig. 4B). Note that the high recognition rates at low noise levels are reasonable for the small number of classifications with well-defined patterns. Similar results were obtained by simulation using a larger testing dataset with different noise levels (Fig. S12). The recognition rate dropped to ~80 % with the noise level σ = 3.5. This is because the level of noise has blurred the segments beyond easy recognition even by human eyes (bottom panel, Fig. 3B). For example, in one case, the blurring led to ambiguity between the two digits "1" and "3" (Fig. 4C). The close values between the neuronal outputs of "1" and "3" still captured this ambiguity. In another case, even though the ambiguity existed between patterns of "1" and "2", the network could still successfully classify digit "2" since the digit "1" was not displayed in the right place (Fig. 4D).

4. Hand-written digit classification

We then used the 8×8 array for hand-written digit recognition from the Modified National Institute of Standards and Technology (MNIST) dataset, which is a standard benchmark to evaluate learning algorithms [41]. The MINST image is usually flatted to $N \times N$ pixels, yielding an input vector size of N^2 that exceeds the input size (*e.g.*, 8) of the array [22]. Therefore, we used reservoir computing to first compress the N^2-sized input vector to an N-sized one [42]. Specifically, we proposed to use the temporal dynamics innate to a resistor-capacitor (RC) circuit as the reservoir system to compress the data (Fig. 5A). First, each of the images ("1" through "4") from the MNIST database was rescaled to 7 × 7 pixels (Experimental Procedures). The grayscale (out of 256 levels) in each pixel was converted to an analog voltage (between 0.3 to 0.7 V). The sequence of the seven converted values in each row was fed into the RC reservoir for an output. In this way, the 7 × 7 input vector was compressed to a 7×1 vector, which (plus one constant bias of 0.7 V) served as the input to the 8 WLs in the 1D1R array for training and classification.

The time constant (*e.g.*, *RC*) in the reservoir system was chosen to be comparable to the time span of the input sequence (*i.e.*, $N_D \times T$, where N_D and T are the number and width of the converted pixels in an input sequence). Specifically, we used $N_D T = 2.5RC$ with $N_D = 7$, $T = 100$ μs, $R = 2.9$ kΩ, and $C = 100$ nF (Experimental Procedures). For reservoir computing, an activation pulse (400 mV, 200 μs) was first used to raise the output baseline (to ~1/2 V), followed by the seven sequential converted pixel values in a row vector. The instant voltage across the capacitor right after the seventh input was taken as the reservoir output. Fig. 5B shows two special cases when the input streams were "0000000" (green curve) and "VVVVVVV" (V representing the full amplitude of 400 mV, yellow curve) to yield the minimum and maximum reservoir outputs, respectively. For a typical MNIST digit (*e.g.*, "3"), the seven-row vectors (each having seven sequential pixel values) yielded seven differentiated reservoir outputs (Fig. 5C). Inputs converted from other digits also yielded differential reservoir outputs (Fig. S13), suggesting that the RC reservoir system can effectively compress input data.

The compressed input data sets (from a total of 34000 MNIST images) were used to train the 8×8 neural network (Fig. S11 and Experimental Procedures). The 32 synaptic weights in the array showed progressive dispersion after the *in situ* training (Fig. 5D, Fig. S14). After 85 epochs, an average recognition rate of 71 % was achieved from the testing of 400 images (Fig. 5E, green curve), with the recognition rates for individual digits also shown (Fig. 5F). We compared the experimental result to the simulation result (Fig. S15), which showed that the two had very close rates (Fig. 5E, yellow curve). The method can be used to classify images applied with the same degree of rotation to achieve a similar level of recognition rate (Fig. S16), showing the generality in its applicability. Increasing the data variety (e.g., mixing rotated images with unrotated ones) did lead to a reduced classification rate, suggesting that a larger array size is needed for augmenting data complexity. Increasing the array size (*e.g.*, to 64×128) is expected to improve the recognition rate, as revealed by simulation (Figs. S15, S17). Note that with the optimal choice of input voltage range, the effect of possible nonlinearity from the diode can be effectively suppressed (Fig. S18).

5. Diode-memristor integration

Following the above functional validation, we provided a proof-of-concept demonstration of crossbar architecture fabricated from the 1D1R stacking cells for the potential of integrated neuromorphic computing. The diode was fabricated by depositing a layer of n-type ZnO onto a p-type Si layer [43]. ZnO was used for the n-type layer for fabrication convenience. The memristor with a layered structure of Ta/HfO_x/HfO_y/Pt was fabricated on top of the diode to form the stacking 1D1R cell (Fig. 6A, Fig. S19). An 8×8 crossbar array was integrated on a Si substrate (Fig. 6B-D, Fig. S20 and Experimental Procedures, and Supplemental Experimental Procedures).

The 64 diodes in the integrated array showed uniform current-voltage (*I-V*) characteristics featuring the rectifying effect and transient dynamics (Fig. 6E). A reverse current larger than 4 mA (at a reverse bias of -1 V) could pass the diode during the reserve recovery time (T_{rr} ~ 200 ns, inset), which is sufficient for RESET operation in typical memristors[39]. The T_{rr} maintained the stable value after extensive switching (*e.g.*, 10^7 cycles) in the diode (Fig. S21), showing the reliability for scalable integration. The 64 Pt/HfO_x/HfO_y/Ta memristors were fabricated in a reverse layer structure to conform to the driving-current direction in the diode (Fig. S22). The

memristors showed average SET and RESET thresholds of 0.58±0.08 V and -0.47±0.04 V, respectively (Fig. 6F). The On-state curves (set by applying the same compliance current) also showed uniform distribution (Fig. S23.a), which is relevant because the SET operation is predominantly used for weight update in the array (Fig. S8). As a result, the integrated 1D1R cell could be successfully programmed to different conduction states by applying different compliance currents (Fig. 6G). The cell could be reset to an HRS from these programmed states by applying a RESET pulse (*e.g.*, -0.5 V) in the T_{rr} window (Fig. S24). The cell could also be repeatedly programmed to different conduction states (Fig. 6H), showing reliable programmability for synaptic weight update in a network. Note that a one-time RESET was used to closely represent the actual weight update strategy in a cell made from a pair of memristors (Fig. S8). The programmability in the entire array was demonstrated by the consecutive programming of the array into pixeled letters of "UMASS" (Fig. 6I). The dispersity in the programmed pixels in the array was consistent with the conductance dispersity in individual Ta-HfO_2 memristors (Fig. S23), suggesting that the inherent stochasticity in memristors (rather than the programming theme) caused the image dispersion. Improving performance uniformity in the memristors (e.g., by denoising procedure [44]) is expected to improve programming uniformity in the array.

DISCUSSION AND OUTLOOK

In this work, we have demonstrated that a diode can be implemented as a bidirectional selector for constructing two-terminal memristor architecture by exploiting its reverse recovery dynamics. The constructed 1D1R arrays are implemented for *in situ* neural training and classification, and a stacking 1D1R cell structure is exploited to demonstrate scalable integration.

The diode needs to maintain a certain driving current and T_{rr} in order to ensure successful SET and RESET programming in the memristors. We used the mainstream memristor (*e.g.*, Ta-HfO_2 system) to estimate the scaling potential in the diode. An empirical driving current >5 mA (at 2 V) and $T_{rr} \geq 40$ ns were revealed to ensure the successful programming of the memristor (Fig. S25 a-c). Our simulated diode model showed that the diode could maintain a driving current ~8 mA (at 2 V) and T_{rr} ~80 ns with the size reduced to 10×10 μm^2 (Fig. S25 d-f). This size is compatible with the integration density achieved in many current neuromorphic systems [20-22]. Further downsizing the system is possible by engineering high-current/long-retention diode[45] or low-current/fast-switching nonvolatile memristors [46,47].

Besides improving the compactness/simplicity in integration, the two-terminal 1D1R architecture can also reduce the peripheral budget compared to other multiple-terminal systems. One such area is the power consumption involved in digital-to-analog converters (DACs) [14]. The 1D1R architecture was estimated to save 68×(N-1) mW (with an array size of N×N) in DACs compared to other memristor architectures (Note S1).

We acknowledge that the 1T1R architecture can remain as the mainstream strategy for the reliable and efficient integration of memristor networks. Our work points to a potential alternative for constructing two-terminal memristor architectures. The strategy can hold promise for practical implementation due to the technological maturity of diode. The current demonstration is based on

diodes that are not fully optimized for memristor pairing, which means that continuous study in device engineering may progressively improve the function and integration.

EXPERIMENTAL PROCEDURES

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jun Yao (juny@umass.edu).

Materials availability

All processing solvents, such as acetone, isopropyl alcohol, were purchased from commercial sources and used as received. Silicon-on-insulator (SOI) wafer (12 μm *p*-type layer, ρ ~ 0.007-0.015 Ω·cm) was purchased from EI-Cat Inc.

Data and code availability

The data that support the findings of this study are available within the paper and its supplemental information files. Additional data and files are available from the corresponding author upon reasonable request.

Device fabrication

Ta/HfO₂/Pt memristor. The memristor was fabricated following previous procedures [19,33]. Briefly, the bottom electrode (Ti/Pt, 5/25 nm) was defined by standard photolithography, metal deposition, and liftoff processes on a Si substrate capped with 600 nm thermal oxide. A 5 nm-thick HfO_2 dielectric layer was deposited by atom layer deposition (ALD) at 250 ℃. The top electrode (Ta/Pt, 20/15 nm) was defined by similar lithographic processes. Reactive ion etching (RIE) was used to selectively expose the bottom electrode for electrical addressing.

Stacking 8×8 1D1R array. A silicon-on-insulator (SOI) wafer (12 μm *p*-type layer, ρ ~ 0.007-0.015 Ω·cm) was used. The detailed fabrication procedure is described in Fig. S20 and Supplemental Experimental Procedures.

Electrical measurement

All the electrical measurements were performed in the ambient environment. The *I-V*/*I-t*/*V-t* curves and pulse measurements were performed by using a semiconductor analyzer (Keysight B1500). The voltage output from the RC reservoir was collected by an ADC (Digidata 1440A). The diode endurance was measured by the combination of a waveform generator (Keysight 33500B), a current amplifier (DL-1211), and an ADC (Digidata 1440A) as illustrated in Fig. S21b. The programming and processing in the 8×8 1D1R array were performed by a homemade circuit system (Fig. S5). The sneak-path current was measured by a current amplifier (DL-1211;) connected with an ADC (Digidata 1440A).

Dataset

The dataset with noise for SSD classification was generated using the Gaussian random number generator (Fig. S10). The dataset for MNIST classification was cropped from the original 28 × 28 pixels images to 20 × 20 pixels ones. Then, a bicubic interpolation algorithm[22] was used

to further reduce the image size to 7 × 7 pixels (for experiment and simulation) or 8 × 8 pixels (for simulation).

Algorithm

The *in-situ* training involved feedforward interference and feedback weight update[48] (Fig. S11). All memristors were initially set to the HRS. The grayscale (21 levels for SSD images and 256 levels for MNIST images) was linearly converted to a voltage between 0.3 to 0.7 V.

For the SDD training, each epoch used 4 images (e.g., 4-time weight updates per epoch), and 16 epochs were used. For the MNIST training, each epoch used 400 images with a minibatch size of 50 (e.g., 8-time weight updates per epoch), and 85 epochs (a total of 34000 images) were used.

In the forward path, the weight summation was performed in the memristor array by:

$$I_j = \sum_{i=1}^{8} f_{ij}\left(V_x(i)\right) \tag{1}$$

, where I_j is the output current of BL j and $V_x(i)$ the input voltage of WL i for pattern x. $f_{ij}()$ is the *I-V* relationship of each 1D1R cell. Next, an activation function was used to obtain the final neuronal outputs:

$$O(k) = \frac{1}{1 + e^{-(I_{2*k} - I_{2*k-1})/0.00006}} \tag{2}$$

, where k = 1, 2, 3, 4 (note that a pair of BLs forms one neuron) and ($2k$-1, $2k$) is the neighboring pair of BLs. Thus, the current difference $I_{2*k} - I_{2*k-1}$ represents the immediate neuronal output. A coefficient of 0.00006 (A) is used in the activation function.

In the backward path, the weight adjustment Δw_{ik} is first calculated by:

$$\Delta w_{ik} = lr * (T_x(k) - O(k)) * O(k) * (1 - O(k)) * (V_x(i) - 0.3) \tag{3}$$

, where lr is the learning rate ($lr = 2.5$) and $T_x(k)$ the target output from the k^{th} neuron for each specific input pattern x. Then, the accumulated weight adjustment Δw_{ik}^{acc} is updated according to the minibatch size M (M = 1 for SSD and M = 50 for MNIST classification):

$$\Delta w_{ik}^{acc} = \Delta w_{ik}^{acc} + \frac{\Delta w_{ik}}{M} \tag{4}$$

With the accumulated weight, the compliance current used for synaptic weight update for each 1D1R cell is:

$$\begin{cases} CC_{i,2*k} = CC_{i,2*k} + \Delta w_{ik}^{acc} * 0.00012 \\ CC_{i,2*k-1} = CC_{i,2*k-1} \end{cases} if \Delta w_{ik}^{acc} > 0 \tag{5}$$

and

$$\begin{cases} CC_{i,2*k} = CC_{i,2*k} \\ CC_{i,2*k-1} = CC_{i,2*k-1} + \Delta w_{ik}^{acc} * 0.00012 \end{cases} if \Delta w_{ik}^{acc} < 0 \tag{6}$$

, where a coefficient of 0.00012 (A) is used to convert the weight to the compliance current.

RC reservoir design

According to the equation of an RC circuit:

$$C\frac{dV}{dt}+\frac{V}{R}=0 \tag{7}$$

, the output voltage V is:

$$V(t)=V_0e^{-\frac{t}{RC}} \tag{8}$$

We choose the time constant that allows the voltage to reduce to ~10% of the initial value during the input sequence of (0000000) (*i.e.*, the lower-bound output) to ensure a wide output window. This leads to

$$RC\approx\frac{N_D\times T}{2.5} \tag{9}$$

or,

$$N_D\times T=2.5RC \tag{10}$$

, where N_D and T are the number and width of the converted pixels in an input sequence (Fig. 5B).

Acknowledgments: J.Y. acknowledges support from the National Science Foundation (NSF) DMR-2027102, the Army Research Office (W911NF2210027), and Office of Naval Research (N00014-21-1-2593).

Author contributions: T.F. and J.Y. conceived the project. T.F. and S.F. designed the experiments and carried out experimental studies in device fabrication, measurement, and algorithm implementation. T.F. designed the peripheral computing hardware and performed analyses. S.F. did the simulation. S.W. helped with device fabrication. T.F. and J.Y. wrote the manuscript. All authors discussed the results and implications and commented on the manuscript.

Declaration of interests: The authors declare no competing interests.

REFERENCE

1. Kumar, S., Wang, X., Strachan, J.P., Yang, Y., and Lu, W.D. (2022). Dynamical memristors for higher-complexity neuromorphic computing. Nat. Rev. Mater. *7*, 575-591.
https://doi.org/10.1038/s41578-022-00434-z
2. Prezioso, M., Merrikh-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K., and Strukov, D.B. (2015). Training and operation of an integrated neuromorphic network based on metal-oxide memristors. Nature *521*, 61-64.
https://doi.org/10.1038/nature14441

3. Le Gallo, M., Sebastian, A., Mathis, R., Manica, M., Giefers, H., Tuma, T., Bekas, C., Curioni, A., and Eleftheriou, E. (2018). Mixed-precision in-memory computing. Nat. Electron. *1*, 246-253.
https://doi.org/10.1038/s41928-018-0054-8
4. Schuman, C.D., Kulkarni, S.R., Parsa, M., Mitchell, J.P., Date, P., and Kay, B. (2022). Opportunities for neuromorphic computing algorithms and applications. Nat. Comput. Sci. *2*, 10-19.
https://doi.org/10.1038/s43588-021-00184-y
5. Wang, Z., Wu, H., Burr, G.W., Hwang, C.S., Wang, K.L., Xia, Q., and Yang, J.J. (2020). Resistive switching materials for information processing. Nat. Rev. Mater. *5*, 173-195.
https://doi.org/10.1038/s41578-019-0159-3
6. Ielmini, D., and Wong, H.-S.P. (2018). In-memory computing with resistive switching devices. Nat. Electron. *1*, 333-343.
https://doi.org/10.1038/s41928-018-0092-2
7. Xia, Q., and Yang, J.J. (2019). Memristive crossbar arrays for brain-inspired computing. Nat. Mater. *18*, 309-323.
https://doi.org/10.1038/s41563-019-0291-x
8. Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q., Zhang, W., Yang, J.J., and Qian, H. (2020). Fully hardware-implemented memristor convolutional neural network. Nature *577*, 641-646.
https://doi.org/10.1038/s41586-020-1942-4
9. Mullani, N.B., Kumbhar, D.D., Lee, D.H., Kwon, M.J., Cho, S.y., Oh, N., Kim, E.T., Dongale, T.D., Nam, S.Y., and Park, J.H. (2023). Surface Modification of a Titanium Carbide MXene Memristor to Enhance Memory Window and Low‐Power Operation. Adv. Funct. Mater. *33*, 2300343.
https://doi.org/10.1002/adfm.202300343
10. Huh, W., Lee, D., and Lee, C.H. (2020). Memristors based on 2D materials as an artificial synapse for neuromorphic electronics. Adv. Mater. *32*, 2002092.
https://doi.org/10.1002/adma.202002092
11. Nirmal, K.A., Ren, W., Khot, A.C., Kang, D.Y., Dongale, T.D., and Kim, T.G. (2023). Flexible Memristive Organic Solar Cell Using Multilayer 2D Titanium Carbide MXene Electrodes. Adv. Sci. *10*, 2300433.
https://doi.org/10.1002/advs.202300433
12. Duan, H., Cheng, S., Qin, L., Zhang, X., Xie, B., Zhang, Y., and Jie, W. (2022). Low-power memristor based on two-dimensional materials. The Journal of Physical Chemistry Letters *13*, 7130-7138.
https://doi.org/10.1021/acs.jpclett.2c01962
13. Fu, S., Park, J.-H., Gao, H., Zhang, T., Ji, X., Fu, T., Sun, L., Kong, J., and Yao, J. (2023). Two-Terminal MoS2 Memristor and the Homogeneous Integration with a MoS2 Transistor for Neural Networks. Nano. Lett. *23*, 5869-5876.
https://doi.org/10.1021/acs.nanolett.2c05007
14. Kiani, F., Yin, J., Wang, Z., Yang, J.J., and Xia, Q. (2021). A fully hardware-based memristive multilayer neural network. Sci. Adv. *7*, eabj4801.
https://doi.org/10.1126/sciadv.abj4801
15. Wang, W., Danial, L., Li, Y., Herbelin, E., Pikhay, E., Roizin, Y., Hoffer, B., Wang, Z., and Kvatinsky, S. (2022). A memristive deep belief neural network based on silicon synapses. Nat. Electron. *5*, 870-880.
https://doi.org/10.1038/s41928-022-00878-9
16. Shi, L., Zheng, G., Tian, B., Dkhil, B., and Duan, C. (2020). Research progress on solutions to the sneak path issue in memristor crossbar arrays. Nanoscale Adv. *2*, 1811-1827.
https://doi.org/10.1039/D0NA00100G
17. Cassuto, Y., Kvatinsky, S., and Yaakobi, E. (2013). Sneak-path constraints in memristor crossbar arrays. (IEEE), pp. 156-160.
https://doi.org/10.1109/ISIT.2013.6620207

18. Zhong, Y., Tang, J., Li, X., Liang, X., Liu, Z., Li, Y., Xi, Y., Yao, P., Hao, Z., and Gao, B. (2022). A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing. Nat. Electron. *5*, 672-681.
https://doi.org/10.1038/s41928-022-00838-3
19. Li, C., Hu, M., Li, Y., Jiang, H., Ge, N., Montgomery, E., Zhang, J., Song, W., Dávila, N., and Graves, C.E. (2018). Analogue signal and image processing with large memristor crossbars. Nat. Electron. *1*, 52-59.
https://doi.org/10.1038/s41928-017-0002-z
20. Mao, R., Wen, B., Kazemi, A., Zhao, Y., Laguna, A.F., Lin, R., Wong, N., Niemier, M., Hu, X.S., and Sheng, X. (2022). Experimentally validated memristive memory augmented neural network with efficient hashing and similarity search. Nat. Commun. *13*, 6284.
https://doi.org/10.1038/s41467-022-33629-7
21. Wang, R., Shi, T., Zhang, X., Wei, J., Lu, J., Zhu, J., Wu, Z., Liu, Q., and Liu, M. (2022). Implementing in-situ self-organizing maps with memristor crossbar arrays for data mining and optimization. Nat. Commun. *13*, 2289.
https://doi.org/10.1038/s41467-022-29411-4
22. Li, C., Belkin, D., Li, Y., Yan, P., Hu, M., Ge, N., Jiang, H., Montgomery, E., Lin, P., and Wang, Z. (2018). Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. Nat. Commun. *9*, 2385.
https://doi.org/10.1038/s41467-018-04484-2
23. Yao, P., Wu, H., Gao, B., Eryilmaz, S.B., Huang, X., Zhang, W., Zhang, Q., Deng, N., Shi, L., and Wong, H.-S.P. (2017). Face classification using electronic synapses. Nat. Commun. *8*, 15199.
https://doi.org/10.1038/ncomms15199
24. Fu, T., Fu, S., Sun, L., Gao, H., and Yao, J. (2023). An Effective Sneak‐Path Solution Based on a Transient‐Relaxation Device. Adv. Mater. *35*, 2207133.
https://doi.org/10.1002/adma.202207133
25. Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010). Complementary resistive switches for passive nanocrossbar memories. Nat. Mater. *9*, 403-406.
https://doi.org/10.1038/nmat2748
26. Fu, T., Fu, S., and Yao, J. (2023). Recent progress in bio-voltage memristors working with ultralow voltage of biological amplitude. Nanoscale *15*, 4669-4681.
https://doi.org/10.1039/D2NR06773K
27. Wang, G., Lauchner, A.C., Lin, J., Natelson, D., Palem, K.V., and Tour, J.M. (2013). High‐performance and low‐power rewritable SiOx 1 kbit one diode–one resistor crossbar memory array. Adv. Mater. *25*, 4789-4793.
https://doi.org/10.1002/adma.201302047
28. Rao, M., Song, W., Kiani, F., Asapu, S., Zhuo, Y., Midya, R., Upadhyay, N., Wu, Q., Barnell, M., and Lin, P. (2022). Timing selector: Using transient switching dynamics to solve the sneak path issue of crossbar arrays. Small Science *2*, 2100072.
https://doi.org/10.1002/smsc.202100072
29. Fu, T., Liu, X., Gao, H., Ward, J.E., Liu, X., Yin, B., Wang, Z., Zhuo, Y., Walker, D.J., and Joshua Yang, J. (2020). Bioinspired bio-voltage memristors. Nat. Commun. *11*, 1861.
https://doi.org/10.1038/s41467-020-15759-y
30. Lv, Z., Zhou, Y., Han, S.-T., and Roy, V. (2018). From biomaterial-based data storage to bio-inspired artificial synapse. Materials today *21*, 537-552.
https://doi.org/10.1016/j.mattod.2017.12.001

31. Lauritzen, P.O., and Ma, C.L. (1991). A simple diode model with reverse recovery. IEEE Trans Power Electron *6*, 188-191.
https://doi.org/10.1109/63.76804
32. Wang, Y., Zhang, Q., Ying, J., and Sun, C. (2004). Prediction of PIN diode reverse recovery. (IEEE), pp. 2956-2959.
https://doi.org/10.1109/PESC.2004.1355304
33. Jiang, H., Han, L., Lin, P., Wang, Z., Jang, M.H., Wu, Q., Barnell, M., Yang, J.J., Xin, H.L., and Xia, Q. (2016). Sub-10 nm Ta channel responsible for superior performance of a HfO2 memristor. Sci. Rep. *6*, 28525.
https://doi.org/10.1038/srep28525
34. Huang, C.-H., Huang, J.-S., Lin, S.-M., Chang, W.-Y., He, J.-H., and Chueh, Y.-L. (2012). ZnO1–x nanorod arrays/ZnO thin film bilayer structure: from homojunction diode and high-performance memristor to complementary 1D1R application. Acs Nano *6*, 8407-8414.
https://doi.org/10.1021/nn303233r
35. Mun, B.H., You, B.K., Yang, S.R., Yoo, H.G., Kim, J.M., Park, W.I., Yin, Y., Byun, M., Jung, Y.S., and Lee, K.J. (2015). Flexible one diode-one phase change memory array enabled by block copolymer self-assembly. ACS nano *9*, 4120-4128.
https://doi.org/10.1021/acsnano.5b00230
36. Lee, D.K., Kim, G.H., Sohn, H., and Yang, M.K. (2019). Positive effects of a Schottky-type diode on unidirectional resistive switching devices. Applied Physics Letters *115*, 263502
https://doi.org/10.1063/1.5133868
37. Yoon, J., Ji, Y., Lee, S.K., Hyon, J., and Tour, J.M. (2018). Low‐Temperature‐Processed SiOx One Diode–One Resistor Crossbar Array and Its Flexible Memory Application. Adv. Electron. Mater. *4*, 1700665.
https://doi.org/10.1002/aelm.201700665
38. Yoon, K.J., Kim, G.H., Yoo, S., Bae, W., Yoon, J.H., Park, T.H., Kwon, D.E., Kwon, Y.J., Kim, H.J., and Kim, Y.M. (2017). Double‐Layer‐Stacked One Diode‐One Resistive Switching Memory Crossbar Array with an Extremely High Rectification Ratio of 10^9. Adv. Electron. Mater. *3*, 1700152.
https://doi.org/10.1002/aelm.201700152
39. Wang, Z., Li, C., Song, W., Rao, M., Belkin, D., Li, Y., Yan, P., Jiang, H., Lin, P., and Hu, M. (2019). Reinforcement learning with analogue memristor arrays. Nat. Electron. *2*, 115-124.
https://doi.org/10.1038/s41928-019-0221-6
40. Migliato Marega, G., Wang, Z., Paliy, M., Giusi, G., Strangio, S., Castiglione, F., Callegari, C., Tripathi, M., Radenovic, A., and Iannaccone, G. (2022). Low-power artificial neural network perceptron based on monolayer MoS2. ACS nano *16*, 3684-3694.
https://doi.org/10.1021/acsnano.1c07065
41. Deng, L., and Yu, D. (2014). Deep learning: methods and applications. Found Trends Signal Process *7*, 197-387.
http://dx.doi.org/10.1561/2000000039
42. Cao, J., Zhang, X., Cheng, H., Qiu, J., Liu, X., Wang, M., and Liu, Q. (2022). Emerging dynamic memristors for neuromorphic reservoir computing. Nanoscale *14*, 289-298.
https://doi.org/10.1039/D1NR06680C
43. Zhang, L., Zhu, L., Li, X., Xu, Z., Wang, W., and Bai, X. (2017). Resistive switching mechanism in the one diode-one resistor memory based on p+-Si/n-ZnO heterostructure revealed by in-situ TEM. Sci. Rep. *7*, 45143.
https://doi.org/10.1038/srep45143

44. Rao, M., Tang, H., Wu, J., Song, W., Zhang, M., Yin, W., Zhuo, Y., Kiani, F., Chen, B., and Jiang, X. (2023). Thousands of conductance levels in memristors integrated on CMOS. Nature *615*, 823-829. https://doi.org/10.1038/s41586-023-05759-5

45. Pu, T., Younis, U., Chiu, H.-C., Xu, K., Kuo, H.-C., and Liu, X. (2021). Review of recent progress on vertical GaN-based PN diodes. Nanoscale Res. Lett. *16*, 101. https://doi.org/10.1186/s11671-021-03554-7

46. Onen, M., Emond, N., Wang, B., Zhang, D., Ross, F.M., Li, J., Yildiz, B., and Del Alamo, J.A. (2022). Nanosecond protonic programmable resistors for analog deep learning. Science *377*, 539-543. https://doi.org/10.1126/science.abp8064

47. Fuller, E.J., Keene, S.T., Melianas, A., Wang, Z., Agarwal, S., Li, Y., Tuchman, Y., James, C.D., Marinella, M.J., and Yang, J.J. (2019). Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. Science *364*, 570-574. https://doi.org/10.1126/science.aaw5581

48. Vogl, T.P., Mangis, J., Rigler, A., Zink, W., and Alkon, D. (1988). Accelerating the convergence of the back-propagation method. Biol. Cybern. *59*, 257-263. https://doi.org/10.1007/BF00332914

Figure 1. **Addressing concept and basic performance in the 1D1R cell. A**, (i) Schematic of a 2×2 1D1R array during a forward selective addressing, which relies on the (ii) rectification window (yellow) in the diode. (iii) The suppression of reverse current in the rectification window measured from a diode, with the applied voltage following the schematic sequence in (ii). **B**, (i) Schematic of a reverse selective addressing, which relies on the (ii) reverse recovery window (purple) in the diode. (iii) The reverse current flow in the reverse recovery window measured from a diode, with the applied voltage following the schematic sequence in (ii). **C**, (i) A SET operation in a 1D1R cell (left schematic). The SET process used a pulse of 1.2 V, 3 μs. An increase in the reading current was shown by the reading pulses (0.7 V, 100 μs) before and after the SET operation. (ii). The RESET process involved a positive pulse of 0.7 V, 5 μs to first activate the diode, immediately followed by a negative pulse of -1.2 V, 3 μs within the reverse recovery window. A decrease in the reading current was shown by the reading pulses (0.7 V, 100 μs) before and after the RESET operation.

Figure 2. Performance of 1D1R arrays. A, Demonstration of suppression of sneak-path current in a 2×2 array, with the memristors in the sneak path having LRS and memristor in the selected path having HRS. **B**, Selective programming of memristors (colorful) in a 3×3 array. The selected memristors were successfully programmed to different states (color curves), whereas the unselected ones maintained the same state (black curves). **C**, A 16×32 pixeled "UM" logo formed by stitching eight 8×8 sub-images sequentially programmed by the 8×8 array. The programming order was from top to bottom and from left to right, as shown by red dash lines.

Figure 3. SSD *in situ* training. A, Continuous synaptic weight updates (V_{read} = 0.7 V) in a synapse formed by a pair of 1D1R cells (left schematic). Each cycle contains 102 states. **B**, Schematic of noise generation in SSD. The bottom panel shows a series of SSD images added with different noise levels. **C,** Circuit diagram of the 8×8 1D1R array for SSD classification. The converted grayscale in each segment of the SSD was converted to voltage signal and served as the input to the array. The differential current output in each pair of BLs represented the classification of each digit.

Figure 4. Results of SSD classification. A, The weight distributions in the array before and after the training. **B**, Evolution of the recognition rates with different noise levels added in the SSD images. **C**, A case that the visual ambiguity between a"1" and "3" was captured by the close outputs between neurons' 1' and '3'. **D**, A case that the digit "2" was successfully classified.

Figure 5. MNIST classification. A, Schematic of the MNIST classification process, involving the image resizing and pixel vector (e.g., grayscale-to-voltage) conversion (left), reservoir compression (middle), and classification by the 8×8 1D1R array (right). **B**, Schematic of the RC reservoir output. An activation pulse (400 mV, 200 μs) is followed by 7 sequential converted pixel

values in a row. The upper (yellow) and lower (green) output bounds are generated by the sequences of (0000000) and (VVVVVVV), respectively. **C**, An exemplary input series representing the digit "3" displayed in (**A**), with the corresponding output series (right). **D**, Evolution of the 32 synaptic weights during the *in-situ* training. **E**, Comparison of the recognition rates between the experiment and simulation. **F,** the confusion matrix for the experiment result (right).

Figure 6. Integrated 8×8 1D1R stacking array. A, Schematic of the individual stacking cell. **B**, Optical image of a fabricated cell. Scale bar, 150 μm. **C**, Scanning electron microscope (SEM) image of the integrated array. Scale bar, 200 μm. **D**, Optical image of the integrated array on a Si substrate. Scale bar, 5 mm. **E**, *I-V* characteristics and reserve recovery dynamics (inset) measured from the 64 diodes in the array. The average Off current (V = -2 V) was 5.88 ± 0.71 μA (±s.d.), and the average On current (V = 2V) was 7.35 ±0.64 mA (±s.d.). **F**, *I-V* characteristics from the 64 memristors in the array. The green curves show the cumulative probabilities of the threshold voltages. A current compliance of 500 μA was applied for the SET processes. **G**, Representative *I-V* characteristics from a selected cell showing the controlled weight/conductance update by using different compliance currents. **H**, Representative reversible programming from a selected cell. Each cycle covered ten states. **I**, A "UMASS" logo made from five 8×8 pixeled letters that were consecutively programmed by the same array. Each state was read by a reading voltage of 1.2 V. Compliance currents of 0, 200, 400, 700, and 1100 μA and were used for programming the background in the five letters, respectively. Compliance currents of 400, 700, 1100, 0, and 200 μA were used to programming the letters, respectively.