

ARTICLE

Open Access

A self-rectifying TaO_y/nanoporous TaO_x memristor synaptic array for learning and energy-efficient neuromorphic systems

Sanghyeon Choi¹, Seonghoon Jang¹, Jung-Hwan Moon², Jong Chan Kim³, Hu Young Jeong³, Peonghwa Jang², Kyung-Jin Lee^{1,2} and Gunuk Wang¹

Abstract

The human brain intrinsically operates with a large number of synapses, more than 10^{15} . Therefore, one of the most critical requirements for constructing artificial neural networks (ANNs) is to achieve extremely dense synaptic array devices, for which the crossbar architecture containing an artificial synaptic node at each cross is indispensable. However, crossbar arrays suffer from the undesired leakage of signals through neighboring cells, which is a major challenge for implementing ANNs. In this work, we show that this challenge can be overcome by using Pt/TaO_y/nanoporous (NP) TaO_x/Ta memristor synapses because of their self-rectifying behavior, which is capable of suppressing unwanted leakage pathways. Moreover, our synaptic device exhibits high non-linearity (up to 10^4), low synapse coupling (S.C, up to 4.00×10^{-5}), acceptable endurance (5000 cycles at 85 °C), sweeping (1000 sweeps), retention stability and acceptable cell uniformity. We also demonstrated essential synaptic functions, such as long-term potentiation (LTP), long-term depression (LTD), and spiking-timing-dependent plasticity (STDP), and simulated the recognition accuracy depending on the S.C for MNIST handwritten digit images. Based on the average S.C (1.60×10^{-4}) in the fabricated crossbar array, we confirmed that our memristive synapse was able to achieve an 89.08% recognition accuracy after only 15 training epochs.

Introduction

Over the past few decades, von Neumann architecture with Si-based complementary metal-oxide-semiconductor (CMOS) technology has served as a mainstay of the modern computer and electronics industries.^{1,2} However, the impending fundamental physical limits of CMOS technology and increased fabrication cost have become the main obstacles to sustainable computing technology envisioned for the future.^{1,3,4} In the big data era, the explosive growth of unstructured data and data complexity have revealed the limits of classical computing

hardware in terms of the von Neumann bottleneck between processor and memory,⁵ the rapid increase of heat flux due to CMOS scaling,⁶ and device packing density restriction due to local energy dissipation.⁷

The neuromorphic electronic system, which imitates the principles of biological synapses in a huge network of neurons, has emerged as a promising approach for implementing exceptionally energy-efficient, time-efficient, and fault-tolerant computing technologies.^{8,9} Many attempts to implement the essential synaptic functions, such as short- and long-term plasticity and spike-timing dependent plasticity (STDP), have resulted in various device architectures, including memristors,^{10–14} phase change memory,^{15,16} and floating-gate transistors.^{17,18} Among these devices, memristors, which consist of a simple storage medium sandwiched between two conductors, are strong candidates for device platforms for

Correspondence: Gunuk Wang (gunukwang@korea.ac.kr)

¹KU-KIST Graduate School of Converging Science and Technology, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

²Department of Materials Science and Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

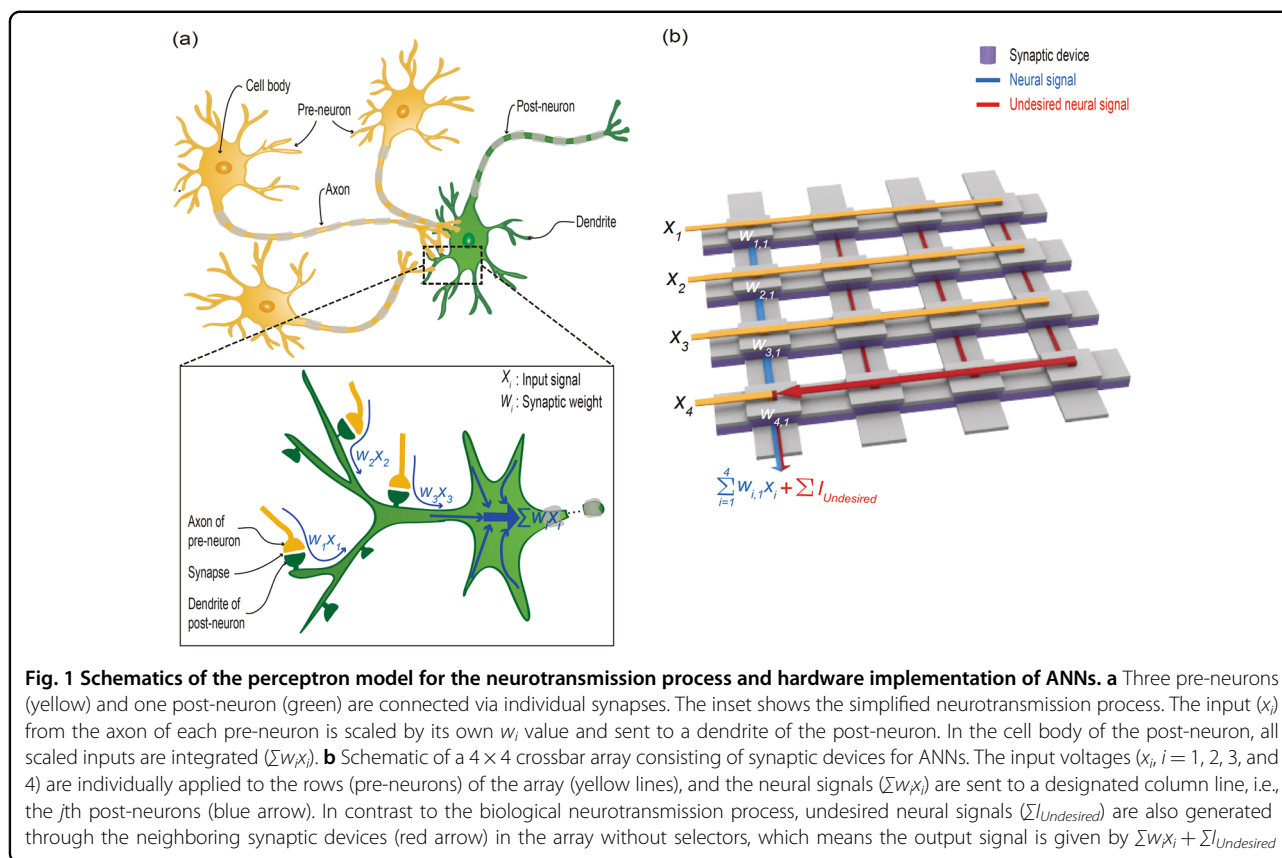
Full list of author information is available at the end of the article.

These authors contributed equally: Sanghyeon Choi, Seonghoon Jang.

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



artificial synapses.^{10–14,19,20} Fundamental device studies have shown that memristors offer desirable properties for artificial synapses, such as a nanoscale footprint,²¹ low-energy²² and low-cost fabrication,²³ and analog switching based on material reconfigurations that can be manipulated by diverse electrical stimuli.²⁴

Because synapses are connectivities between numerous neurons and the analog functionality in the storage of synaptic weight is inherent in the two terminal nodes,^{25,26} neural networks bear some resemblance to crossbar arrays in which memristors are arranged at each crosspoint.^{10,19,20} In real neural networks, each of the hundreds of types of neurons has its own receptor and ion channels that respond to a certain set of inputs through synapse-defined interconnections between the axons of pre-neurons and a dendrite of post-neurons (Fig. 1a), which can perform brain activity.^{25,26} The role of the dendrite is to provide inputs from multiple pre-neurons into the cell body of a single post-neuron. In the perceptron model that simplifies the biological neural network, a post-neuron in the brain receives input signals (x_i) from thousands of other pre-neurons through a biological neurotransmission process (Fig. 1a).²⁷ Each of the input signals is scaled by its own synaptic weight (w_i) and is expected to respond differently depending on the history on its synaptic weight. Then, all inputs are integrated

($\sum w_i x_i$) into the cell body of the post-neuron (Fig. 1a). Some of these outputs ($w_i x_i$) tend to excite the connected neuron whereas others tend to inhibit it.²⁸ If the summed output signal is higher than a threshold value, the signal will be delivered in the form of an action potential firing along the axon of the post-neuron. This signal processing, which is achieved via a complicated interplay among different types of receptors and ion channels, is recognized as the foundation of learning and memory in the human brain.^{25,26}

Considering the signal processing capability of the neural network, the x_i and w_i in the electrical crossbar can be regarded as the applied input voltage and conductance of the i th memristor, respectively. The array refers to an artificial neural network (Fig. 1b).^{10,19,20} In many types of memristors, crosstalk (called cell-to-cell coupling) in which the currents flow through undesired pathways (indicated by the red arrows in Fig. 1b) occurs when the array is comprised of only memristors (i.e., without selectors).^{29,30} This problem may cause the creation of the undesired neural signals (i.e., $\sum I_{Undesired}$) in the ANNs (Fig. 1b),³¹ which could result in misreading of the output current flowing through the designated column lines (i.e., post-neuron), and the learning and memory capability of the array would be severely degraded. More energy would also be consumed because of the

increase in the number of weight updates in the learning process.²⁹ Although diverse selector devices in crossbar arrays have been suggested to suppress these unintentional leakage paths,^{30,32–34} device structures for artificial synapses that only transmit the synaptic signals along the desired pathway without a selector have rarely been reported. In addition, the effect of crosstalk between neighboring synapses on the learning accuracy in ANNs has rarely been investigated. Many studies of memristor synapses have focused on the mimicking of synaptic functions at the single-device level,^{13,14} and have performed learning simulations under the assumption that there is negligible synapse-to-synapse coupling. Therefore, designing and fabricating artificial synapses that perform desirable synaptic functions and suppress unwanted neural signals in the array is a prerequisite for a high-accuracy and low-power learning process.

In this study, we designed and fabricated a self-rectifying memristor synapse and crossbar array (16 × 16) employing a Pt/TaO_y/nanoporous (NP) TaO_x/Ta junction structure that can effectively prevent unwanted neural signals. The essential synaptic and switching characteristics are mimicked and statistically evaluated. Moreover, based on the simulation of MNIST pattern recognition, we verified that the recognition accuracy can be determined by the magnitude of the synapse-coupling (*S.C*) values. Considering the average experimental value of *S.C* (=1.60 × 10⁻⁴), we demonstrated that the trained network consisting only of our memristive synapse can achieve a recognition accuracy of 89.08% after 15 epochs for the MNIST digit images. This value is much higher than the case when a non-rectifying memristor is used (i.e., *S.C* = 1.0). Our proposed TaO_y/NP TaO_x memristor structure could provide an attractive synaptic platform for implementation of ANNs with high-accuracy and energy-efficient learning capability.

Materials and methods

Fabrication of the memristor synapse

Ta metal (200 nm) was deposited on a SiO₂/Si substrate using DC sputtering after the substrate was thoroughly cleaned with acetone, isopropyl alcohol, and DI water for 5 min. Using the designed galvanic cell with an electrolyte consisting of sulfuric acid (150 ml, 98%, DAEJUNG 7683-4100), HF (0.25 ml, 49%, J.T. Baker 9564-06), and H₂O (4.3 ml), the Ta metal was anodized at 50 V for 10 s. The upper Ta metal was simultaneously oxidized and etched, resulting in the formation of NP TaO_x (≈65 nm) on top of the non-anodized Ta metal. The oxygen ions are mostly consumed to form the TaO_x on Ta and other oxidized materials (e.g., SO₃) are rarely produced during the anodization, indicating a low possibility of producing a potential residual of S or SO₃ that could affect the switching.^{35–38} Then, TaO_y (≈10 nm) was deposited using

RF sputtering. To complete the device, patterned Pt metal (100 nm) was sputtered on the TaO_y/NP TaO_x/Ta layer using a shadow mask to act as the top electrode (200-μm junction diameter).

Measurement of device characterization

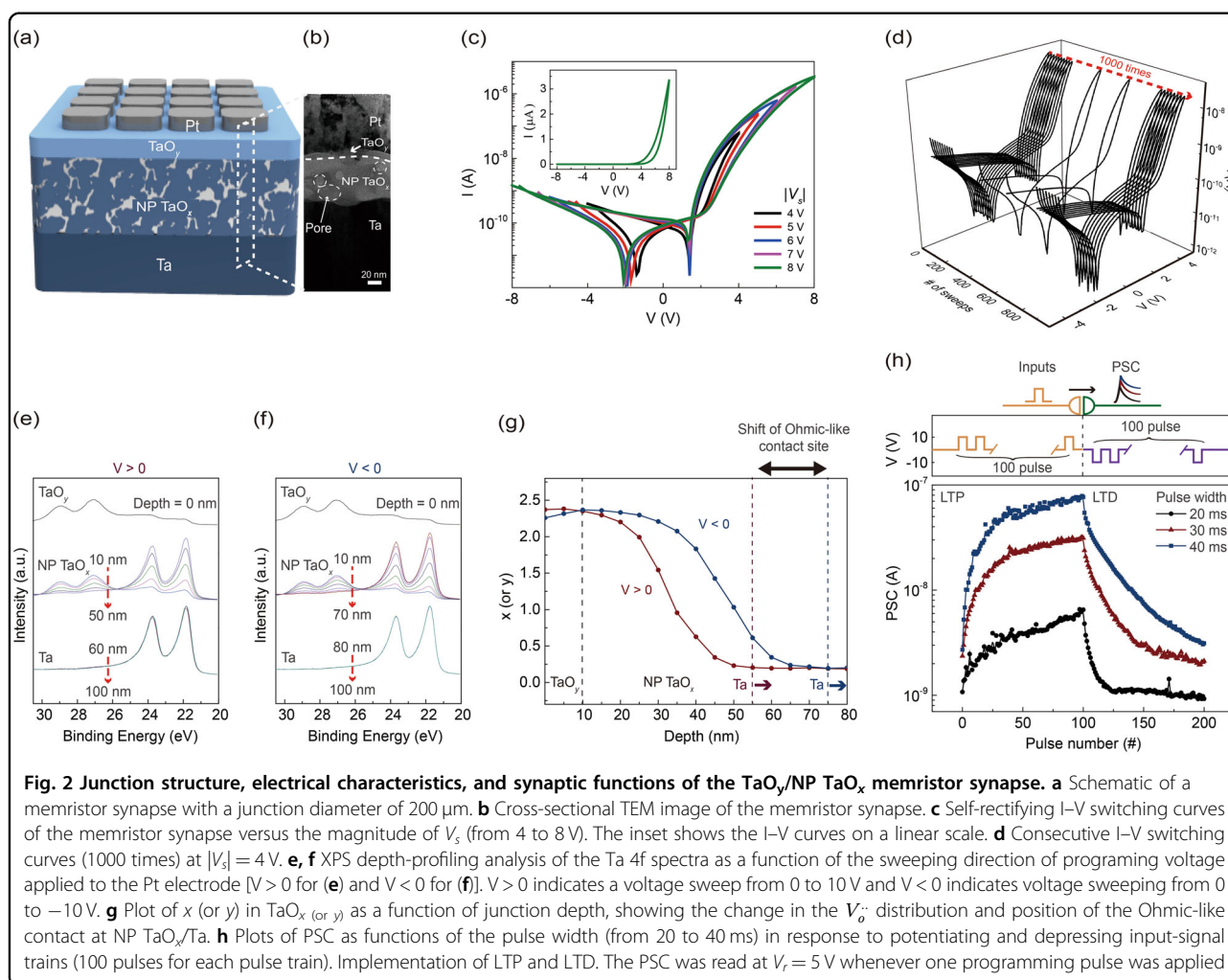
An Agilent 4155C semiconductor parameter analyzer and 41501B pulse generator in the probe station were used to measure the electrical and synaptic characteristics of the TaO_y/NP TaO_x memristor synapse. During the measurements, a voltage was applied to the top Pt electrode and the bottom Ta electrode was grounded. All measurements were performed in air and at room temperature except the endurance tests, which were conducted at 85 °C.

Results and discussion

Single TaO_y/NP TaO_x memristor

To fabricate an artificial synapse that can robustly block undesirable neural signals generated through adjacent synapses in addition to performing essential synaptic functions, we designed a homogeneous bilayer memristor structure consisting of stackable tantalum oxide bilayers (non-porous and nanoporous layers). Figures 2a, b show a schematic diagram of the designed memristor synapse and a cross-sectional transmission electron microscopy image, respectively, in which the junction structure consists of Pt/TaO_y (≈10 nm)/NP TaO_x (≈65 nm)/Ta layers. The NP TaO_x layer was formed from the bottom Ta metal using anodization in a sulfuric acid solution with HF and H₂O, exhibiting an oxygen vacancy (*V_o*) gradient depending on the TaO_x depth. Details of the device fabrication are described in the Methods section. The existence of pores (diameter ≈20–50 nm) inside the TaO_x layer leads to a reduction in the overall switching currents due to its charge/ion trapping ability and the high insulation of the pore itself,^{39,40} decreasing the energy required for switching. The primary role of the top TaO_y layer (*y* ≈2.4) deposited on the NP TaO_x using RF sputtering is to form a robust Schottky barrier between the top Pt metal and the TaO_y layer that is constant under diverse input stimuli. This structure is expected to effectively suppress undesired neural signals in the reverse voltage region at a very low level, regardless of programming voltage variation. In addition, the top TaO_y layer can prevent the potential formation of electrical short circuits based on top Pt metal penetration through the NP oxide layer.

Figure 2c shows the representative current-voltage (*I*-*V*) switching characteristics for the TaO_y/NP TaO_x memristor synapse for different magnitudes of sweeping voltage (*V_S*). All measurements show asymmetric bipolar (self-rectifying) switching behavior controlled by different bias polarities without any formation process, which



resembles the typical switching curve of one-diode–one-resistor (1D–1R) memory devices.^{30,41} In the negative voltage region, the switching current of the device can be suppressed based on the rectifying property of the Schottky barrier at Pt/TaO_y; its level is roughly maintained ($\approx 10^{-10}$ A) regardless of programming voltage variation (from 4 to 8 V). In contrast, in the positive voltage region, the hysteresis window gradually widens as the magnitude of V_S increases. Different voltage points for the minimal current values reflect the different charge-trapping ability according to the applied voltage polarity (Fig. 2c), which may be influenced by the junction asymmetry.^{39,40} The multiple switching conductances of the TaO_y/NP TaO_x memristor controlled by the input voltages can be utilized as variable synaptic weights for the artificial synapse. The maximum non-linearity of the devices was $\approx 10^4$; this value was obtained from the current ratio between the read voltage (defined as $V_r = V_S/2$) and $-V_r$ in the ON state and is comparable to previously reported values for 1D–1R memory devices in a densely

packed crossbar array (Supplementary Information, Fig. S1).^{30,41} Higher non-linearity implies a further reduction in the influence of undesired neural signals flowing through neighboring synapses on a selected synapse. The memristor synapse exhibited excellent operational retention ($\approx 1.2 \times 10^4$ s) and endurance (5.0×10^3 cycles at 85 °C), in addition to repeatable sweeping stability (1000 sweeps for $|V_S| = 4$ V) (Fig. 2d and Supplementary Information, Figs. S2a and S2b).

The programming pulse applied as the pre-synaptic input can determine the consumed energy per the weight update for the network training. The relatively high programming voltages and long pulse width must be further improved for application in energy-efficient and large-scale neuromorphic device arrays. Furthermore, because the pore size reflects the trapping ability, which determines the switching conductance level, and its variation can cause fluctuations in the switching parameters, such as power consumption, non-linearity, and minimal current, the size and uniformity of the pores should be

improved to realize a uniform and nanoscale memristor synapse. We previously suggested that the combination of intentional engineering of the porous structure and operational optimization can enhance the essential switching performance metrics, such as programming power and non-linearity.⁴⁰ There is a potential for our device system to be programmed at a relatively low bias region if we properly engineer the pore structure and optimized the junction structure.

Because the NP TaO_x oxidation state was gradually reduced as its depth was increased, the asymmetric contacts in the memristor were initially established at both interfaces, i.e., the Schottky contact at Pt/TaO_y ($y \approx 2.4$) and the Ohmic-like contact at the NP TaO_x/Ta ($x \approx 0.2$) interface (Figs. 2e, f). Depending on the applied voltage polarity, the distribution of V_o^- in the NP TaO_x layer could be partially altered due to the V_o^- drift in the electric field and its exchange between the TaO_y layer and the NP TaO_x layer, which lead to a change in the position of the Ohmic-like contact in the junction (see the schematic of the switching process in Fig. S3). The positively charged V_o^- values that were additionally produced from TaO_y and the pore edge presumably migrate toward the Ta metal layer when a positive voltage is applied to the Pt top electrode ($V > 0$). In this case, the position of the Ohmic-like contact formed at the NP TaO_x/Ta interface could be located at a shallower depth in the junction, resulting in a reduction of the resistance at V_r (i.e., the ON state). Conversely, V_o^- migrates toward the TaO_y layer and the pore edge when a negative voltage is applied to the Pt top electrode ($V < 0$). The Ohmic-like contact could be located at a deeper depth in the junction, resulting in an increase in the resistance at V_r (i.e., the OFF state). Using ex-situ depth-profiling X-ray photoelectron spectroscopy (XPS) analysis with timed Ar⁺ bombardment, we investigated the differences in the oxygen ratio of TaO_y/NP TaO_x/Ta layer between different voltage polarities. As shown in Fig. 2g, the V_o^- distribution and position of the Ohmic-like contact change based on the polarity of the applied voltages, which supports the suggested switching mechanism. The asymmetric bipolar switching phenomenon of TaO_y/NP TaO_x memristor synapse can be understood based on the shift in the Ohmic-like contact driven by the change in the V_o^- distribution in the NP TaO_x layer.

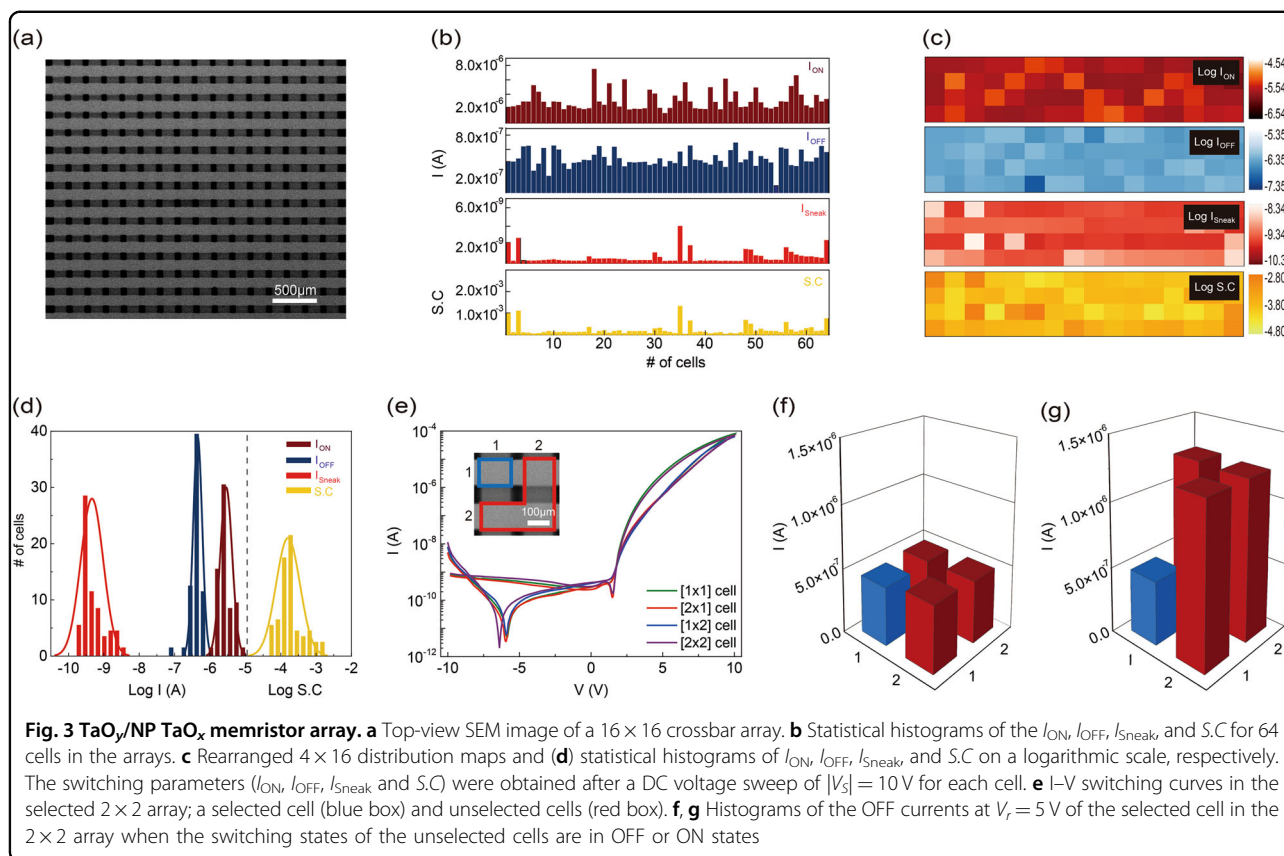
The intrinsic Schottky contact formed at the Pt/TaO_x (≥ 10 nm) interface results in sufficient suppression of the sneak-current in the reverse bias region regardless of the programming voltage (Fig. 2c). However, the sneak-current level was significantly increased when a thinner TaO_y layer (< 8 nm) was inserted between Pt and NP TaO_x (see Supplementary Information, Fig. S4). Therefore, the thickness of the top TaO_y layer determines the rectifying properties and the sneak-current level of the TaO_y/NP

TaO_x memristor synapse. Moreover, because the Schottky barrier formed at the Pt/TaO_y interface is also influenced by the interfacial surface quality, it is important to fabricate a uniform surface with low RMS to suppress undesired pathways and lower the switching variability (Fig. S5).

Long-term plasticity is regarded as essential to biological synaptic functions, and has two forms, long-term potentiation (LTP) and long-term depression (LTD), that generally exhibit gradual potentiation and depression in synaptic weight, respectively, depending on the input stimuli.^{25,26} To mimic LTP and LTD in the Pt/TaO_y/NP TaO_x/Ta memristor synapse, we controlled the degree of change in the synaptic weight by modulating the input stimuli, as shown in Fig. 2h. When a potentiating input-signal train consisting of 100 pulses of 10 V for 20 ms was delivered to the top Pt metal of the memristor synapse (i.e., pre-neuron), the synaptic weight begins to be potentiated and the post-synaptic current (PSC) generated in the bottom Ta metal (post-neuron) gradually increases (mimicking LTP). In contrast, when a depressing input-signal train consisting of 100 pulses at -10 V for 20 ms was delivered to the pre-neuron, the synaptic weight begins to be depressed and the PSC at the post-neuron decreases (mimicking LTD). The change in the PSC for both LTP and LTD is larger when the pulse width is increased from 20 to 40 ms because of the larger change in the V_o^- distribution. This controllability provides input flexibility for controlling the synaptic plasticity. The device also exhibits operational stability for both LTP and LTD under repeatable potentiating and depressing input pulses (see Supplementary Information, Figs. S6, S7). We also imitated the STDP characteristic that is associated with the learning capability of the human brain^{42,43} (see the inset in Fig. S8).

TaO_y/NP TaO_x memristor array

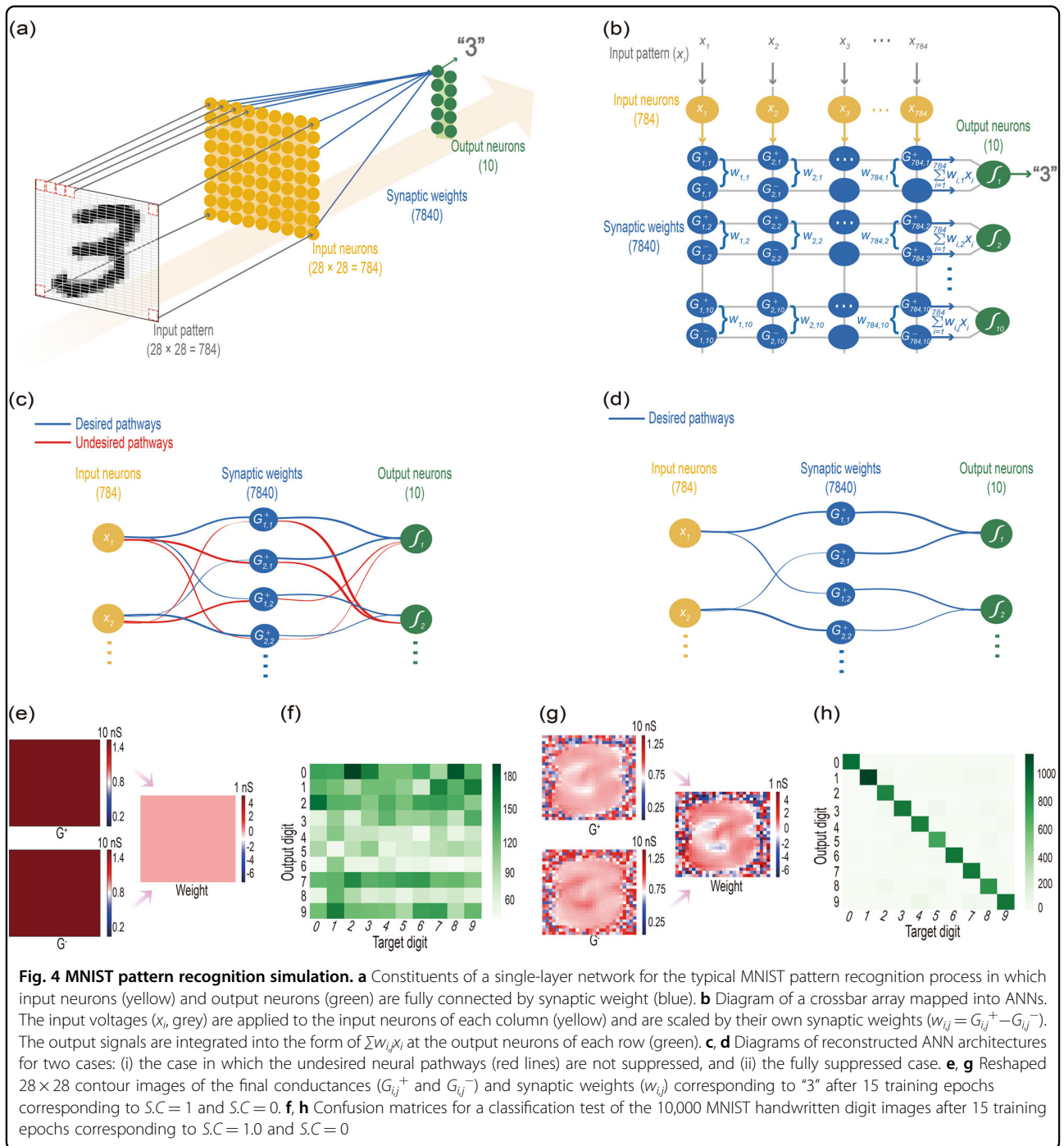
To verify whether the TaO_y/NP TaO_x memristor synapse can suppress undesired neural signals of the crossbar array, a 16×16 crossbar array was fabricated using a shadow mask process and tested as a proof of concept (Fig. 3a). The side length of the cell was 100 μm . An additional selector on every node in the array is not required, significantly reducing the complexity of fabrication. The fabrication of the TaO_y/NP TaO_x crossbar array is described in the Supplementary Information (Fig. S9). As shown in Fig. 3b, the distributions and magnitudes of the ON current (I_{ON}), OFF current (I_{OFF}), sneak-current (I_{Sneak}), and synapse-coupling value (denoted by $S.C$) were statistically investigated (64 cells among 128 cells). I_{Sneak} is the current at $-V_r$ and $S.C$ is defined as $I_{\text{Sneak}}/I_{\text{ON}}$, that is, the inverse of the non-linearity value. Because I_{Sneak} determines the magnitude of the undesired neural signals generated by neighboring



interconnected synapse cells, the $S.C.$ reflects the degree of activation of undesired pathways in the array. Thus, the I_{Sneak} and $S.C.$ in the memristor synapse should be as low as possible. The I_{ON} , I_{OFF} , I_{Sneak} , and $S.C.$ for the TaO_y/NP TaO_x crossbar array were $2.87 \pm 1.96 \times 10^{-6}$ A, $4.49 \pm 3.36 \times 10^{-7}$ A, $4.61 \pm 2.22 \times 10^{-10}$ A, and $1.60 \pm 0.74 \times 10^{-4}$, respectively, for the 64 cells. The distribution maps of each switching parameter show that our memristor array exhibited relatively well-defined switching parameters with acceptable cell uniformity, as shown in Fig. 3c. Figure 3d shows the statistical histograms of I_{ON} , I_{OFF} , I_{Sneak} , and $S.C.$ on a logarithmic scale, which demonstrate that each parameter can be distinguished (based on the distribution). From the selected 2 × 2 matrices within the 16 × 16 array, the I–V switching characteristics for each cell were determined and a crosstalk test was performed, as shown in Figs. 3e–g. Our results confirmed that the OFF states of the selected cell ([1 × 1], shown in the blue box) can be correctly read after programming regardless of the states (ON or OFF) of the neighboring cells ([1 × 2], [2 × 1], and [2 × 2], indicated by red boxes) (Figs. 3f, g). This occurs because the low $S.C.$ of the TaO_y/NP TaO_x memristor synapse can effectively deactivate the unintentional leakage paths; the I_{Sneak} through the [2 × 2] cell in the reverse-bias direction is significantly suppressed.

MNIST pattern recognition simulation

To explore the influence of undesired neural signals in ANNs, we simulated MNIST pattern recognition based on the LTP and LTD fitting results of the TaO_y/NP TaO_x memristor synapse (see Supplementary Information, Fig. S10). Figure 4a shows the constituents of a single-layer network for the typical MNIST pattern recognition process,^{44,45} which simplifies the input/output class of the biological neuron network. Figure 4a shows an example of the recognition process for “3”. The single-layer network consists of 784 input neurons (pre-neurons) and 10 output neurons (post-neurons) fully connected by $784 \times 10 = 7840$ synaptic weights. The input pattern representing the MNIST handwritten digit image (28 × 28 pixels) is connected to each input neuron. The input signal x_i corresponding to each pixel index is individually scaled by $w_{i,j}$; all weighted inputs are delivered into the output neuron in the form of $\sum w_{i,j}x_i$. i and j are integer numbers that range from 1 to 784 for i and from 1 to 10 for j . It is assumed that the signals at the output neurons transmit in the form of a sigmoid activation function $y = f(\sum w_{i,j}x_i)$, where $f = (1 + e^{-x})^{-1}$. This activation function converts the weighted sum into output signal y and acts as a threshold function in a biological neuron. The 60,000 randomly ordered training images corresponding to the MNIST handwritten digits from 0 to 9 are consecutively



fed to the single-layer network (i.e., $\approx 6,000$ images per one-digit number). When a training image is processed, all connected synaptic weights between the input and output neurons are updated based on the delta rule, $\Delta w_{ij} = \eta e x_{ij}$, which is a supervised learning algorithm. η is the learning rate that indicates how fast the synaptic weights change and $e = t - y$ is the error, where t is a target value that is determined by the input pattern. The training process refers to one training epoch. After one epoch is

completed, another 10,000 images that were not used in the above-described training process were tested. This algorithmic approach to the single-layer network process has often been utilized for simple handwritten digit recognition. As shown in Fig. 4b, the single-layer network corresponds to the crossbar array in which input and output neurons intersect at artificial synapses. In the network array, w_{ij} is defined based on the conductance difference between a pair of neighboring memristor

synapses, i.e., $w_{ij} \equiv G_{ij}^+ - G_{ij}^-$.⁴⁸ G_{ij}^+ and G_{ij}^- are the conductances of two neighboring memristors connected at the i th word and j th bit lines, respectively. $w_{ij} > 0$ and $w_{ij} < 0$ represent the synaptic weights for excitatory and inhibitory synapses, respectively. These two-memristor synapses significantly improve the effectiveness of the weight update in the crossbar array. The array implements vector-matrix multiplication and summation, in which all weighted inputs are integrated into the output neuron via Kirchhoff's law and Ohm's law in the form of $\sum G_{ij}^+ x_i$ and $\sum G_{ij}^- x_i$.²⁰ Then, $\sum w_{ij} x_i = \sum G_{ij}^+ x_i - \sum G_{ij}^- x_i$ and the activation function $y = f(\sum w_{ij} x_i)$ are applied and transmitted through circuit-based output neurons.^{49,50} Based on this approach, we simulated two cases: (i) the case that the undesired neural signals are not suppressed and (ii) the fully suppressed case. Figure 4c shows a simple schematic network diagram for case (i), in which neural signals through both desired (blue) and undesired (red) pathways are considered. Each output signal generated from the activation function is determined by all the weighted inputs through the blue and red pathways. Similar to the crosstalk signal in the array, the undesired neural signals through the red pathways could largely be generated at a high $S.C$. In contrast, in case of (ii), the output is determined only by the weighted inputs through the blue pathways (Fig. 4d). All synapses are independent of each other (i.e., $S.C = 0$).

To reflect the effect of undesired neural signals in the ANNs, the sneak current corresponding to the $S.C$ was added to the generated PSC values for simplicity. For example, if $S.C$ ranges from $0 < S.C \leq 1.0$, the red pathways are activated, causing an increased weighted sum ($\sum w_{ij} x_i + \sum I_{Undesired}$) at the connected output neuron. In the case of $S.C = 1.0$, the memristor synapse exhibits a symmetric I-V switching profile regardless of the voltage polarity, i.e., the sneak current is the same as the highest PSC (I_{ON}) of the V_r scheme. In this case, the red pathways are fully activated and the difference in the synaptic weights between synapses almost disappears, as illustrated in Fig. 4c. However, if $S.C = 0$, the red pathways are fully suppressed, which generates only $\sum w_{ij} x_i$ at the output neuron, as illustrated in Fig. 4d. The sneak current becomes zero and each synaptic weight can be independently programmed and updated, which is considered ideal for ANNs.

Figures 4e–h show the simulation results of the pattern recognition for $S.C = 1$ and $S.C = 0$ after 15 training epochs. A detailed flow chart for one epoch is provided in the Supplementary Information (Fig. S11). In both cases, for the “3” input pattern, the final conductance (G_{ij}^+ and G_{ij}^-) and the synaptic weights (w_{ij}) connected to the “3” output neuron can be reshaped to a 28×28 array after 15 epochs (Figs. 4e, g). From the reshaped contour images corresponding to G_{ij}^+ , G_{ij}^- , or w_{ij} , we can estimate

whether “3” is well recognized or not. For $S.C = 1$, the reshaped images in the array do not correspond to the “3” input pattern or other patterns, indicating failure (Fig. 4e). For $S.C = 0$, the reshaped images in the array correspond to only the “3” input pattern, indicating success. Figs. 4f, h show the confusion matrices that correspond to misclassification for $S.C = 1$ and $S.C = 0$ after 15 training epochs, respectively. Based on the similarity of each output signal to a target signal in terms of the individual input digit pattern, the green saturation of an individual tile in the confusion matrices can be determined. The trained network that correctly recognizes all input patterns will render only diagonal tiles in green. For $S.C = 1.0$, the randomly distributed tiles with different saturations of green in the confusion matrix are shown in Fig. 4f, indicating the failure of pattern recognition. For $S.C = 0$, the diagonal tiles are exclusively saturated green, as shown in Fig. 4h, indicating the success of pattern recognition. Using different $S.C$ from 0 to 1 at 0.2 intervals, the single-layer network consisting of TaO_y/NP TaO_x memristor synapses was trained and tested for 15 training epochs. Then, we estimated the accuracy of MNIST pattern recognition for different $S.C$ values and number of training epochs, as shown in Fig. 5. The recognition accuracy significantly improves as the $S.C$ decreases, as indicated by the red arrow in Fig. 5. Interestingly, the accuracy does not further improve despite increasing the number of training epochs, which indicates that the magnitude of undesired neural signals constrains the maximum accuracy of recognition in ANNs. Based on the experimental average $S.C$ (1.60×10^{-4}) for the TaO_y/NP TaO_x memristor synapse, the trained network achieved a

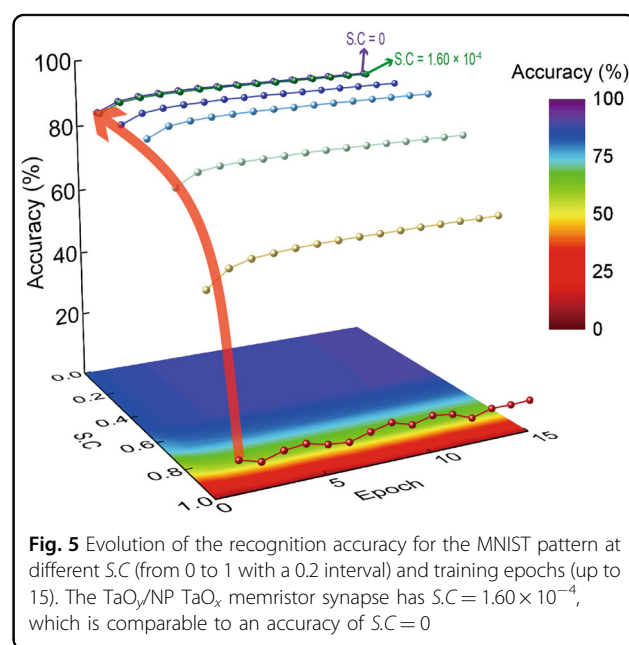


Fig. 5 Evolution of the recognition accuracy for the MNIST pattern at different $S.C$ (from 0 to 1 with a 0.2 interval) and training epochs (up to 15). The TaO_y/NP TaO_x memristor synapse has $S.C = 1.60 \times 10^{-4}$, which is comparable to an accuracy of $S.C = 0$

recognition accuracy of 87.63% after only three epochs and 89.08% after 15 epochs. Notably, other requirements for artificial synaptic devices can significantly affect the learning capability in ANNs, such as the linearity of the weight update (the nonlinearity of the spike-generation process), multilevel states, dynamic range (ON/OFF ratio), and device-to-device and cycle-to-cycle variation. For example, the linearity of the weight update determines the mapping degree of the synaptic weights in the algorithms in the conductance of the device, which affects the learning accuracy.^{3,17} This value ranges from 0 to 1 and should be zero for a perfectly linear weight update. However, in the case of our memristive synapse, the linearity of the weight update is 0.75, which should be improved to achieve higher learning accuracy (Fig. S12). Considering that more conductance states enable the individual pattern to be more clearly distinguished, the multilevel states and dynamic range are related to the resolution capability of information storage for input patterns.^{3,17} In addition, although the device-to-device and cycle-to-cycle variations could cause uncertainty of the weighted sum in the large-scale ANNs during the learning process, a neuro-inspired network features fault-tolerant computing capability.^{3,49} This implies that the network is less vulnerable to device and cycling variation, and could be properly trained to some degree (Fig. S13). Although the linearity of the weight update and the switching window of the device require further improvement for high-recognition accuracy (i.e., closer to that of the human brain), our simulations provide a framework for further research into homogeneous bilayer memristor synapses consisting of non-porous and NP oxide layers with different stoichiometry.

Conclusions

In summary, we fabricated a self-rectifying memristor for an artificial synapse employing a Pt/TaO_y/NP TaO_x/Ta stack on a SiO₂/Si substrate that can effectively suppress the undesired neural signal in ANNs and mimic essential synaptic functions. A potential switching mechanism is suggested and discussed based on the shift in the Ohmic-like contact site driven by the change in the V_o^- distribution in the NP TaO_x layer under an electric field and the intrinsic Schottky contact at the Pt/TaO_y interface. In addition, a 16 × 16 crossbar array consisting of this memristor synapse for ANNs was implemented and statistically evaluated; a crosstalk test was also performed. The effect of undesired neural signals on the accuracy of pattern recognition was simulated for different *S.C.* We demonstrated that a trained network with an experimental $S.C = 1.60 \times 10^{-4}$ can achieve a recognition accuracy of 89.08% after 15 epochs for MNIST digit images. Our results suggest that the rationally designed TaO_y/NP TaO_x memristor synapse offers a novel synaptic

platform that allows for the implementation of ANNs in high-accuracy recognition and low-synaptic-coupling applications.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF-2016R1C1B2007330 and NRF-2018R1A2B6008104), the KU-KIST research fund, Samsung Electronics, a Korea University Future Research Grant, and the KIST Institutional Program (Project No. 2V05750).

Author details

¹KU-KIST Graduate School of Converging Science and Technology, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea. ²Department of Materials Science and Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea. ³UNIST Central Research Facilities and School of Materials Science and Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41427-018-0101-y>.

Received: 16 July 2018 Accepted: 4 November 2018.

Published online: 13 December 2018

References

- International Technology Roadmap for Semiconductors 2.0 (ITRS) 2015 Edition. Beyond CMOS (2015). <http://www.itrs2.net/>. Accessed 16 July 2018.
- Von Neumann, J. The principles of large-scale computing machines. *IEEE Ann. Hist. Comput.* **10**, 243–256 (1988).
- Yu, S. in *Neuro-inspired Computing Using Resistive Synaptic Devices* (ed. Yu, S.) 1, 1–15 (Springer, Gewerbestrasse, Cham, Switzerland, 2017).
- Frank, D. J. et al. Device scaling limits of Si MOSFETs and their application dependencies. *Proc. IEEE* **89**, 259–288 (2001).
- Backus, J. in *ACM Turing Award Lectures* (ed. Ashenurst, R. L.) 614–641 (ACM, Pennsylvania Plaza, New York, 1978).
- Garimella, S. V. et al. Thermal challenges in next-generation electronic systems. *IEEE Trans. Compon. Packag. Technol.* **31**, 801–815 (2008).
- Hoeneisen, B. & Mead, C. A. Fundamental limitations in microelectronics—I. MOS technology. *Solid State Electron.* **15**, 819–829 (1972).
- Kuzum, D., Yu, S. & Wong, H. P. Synaptic electronics: materials, devices and applications. *Nanotechnology* **24**, 382001 (2013).
- Mead, C. Neuromorphic electronic systems. *Proc. IEEE* **78**, 1629–1636 (1990).
- Jo, S. H. et al. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
- Wang, Z. et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* **16**, 101 (2017).
- Ohno, T. et al. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591 (2011).
- Wang, Z. et al. Engineering incremental resistive switching in TaO_x based memristors for brain-inspired computing. *Nanoscale* **8**, 14015–14022 (2016).
- Yoon, C. et al. Synaptic plasticity selectively activated by polarization-dependent energy-efficient ion migration in an ultrathin ferroelectric tunnel junction. *Nano Lett.* **17**, 1949–1955 (2017).
- Kuzum, D., Jeyasingh, R. G., Lee, B. & Wong, H.-S. P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* **12**, 2179–2186 (2011).
- Suri, M. et al. Physical aspects of low power synapses based on phase change memory devices. *J. Appl. Phys.* **112**, 054904 (2012).

17. Kim, S. et al. Pattern recognition using carbon nanotube synaptic transistors with an adjustable weight update protocol. *ACS Nano* **11**, 2814–2822 (2017).
18. Ziegler, M. & Kohlstedt, H. Mimic synaptic behavior with a single floating gate transistor: A MemFlash synapse. *J. Appl. Phys.* **114**, 194506 (2013).
19. Prezioso, M. et al. B. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61 (2015).
20. Zidan, M. A., Strachan, J. P. & Lu, W. D. The future of electronics based on memristive systems. *Nat. Electron.* **1**, 22 (2018).
21. Lee, M.-J. et al. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta₂O_{5-x}/TaO_{2-x} bilayer structures. *Nat. Mater.* **10**, 625 (2011).
22. Zhou, J. et al. Very low-programming-current RRAM with self-rectifying characteristics. *IEEE Electron Device Lett.* **37**, 404–407 (2016).
23. Li, Y. T., et al. A low-cost memristor based on titanium oxide. *Presented at the 2010 10th IEEE International Conference on Solid-state and Integrated Circuit Technology*. 1–4 (IEEE, Shanghai, China, 2010).
24. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453**, 80 (2008).
25. Choi, J.-H. et al. Interregional synaptic maps among engram cells underlie memory formation. *Science* **360**, 430–435 (2018).
26. Whitlock, J. R., Heynen, A. J., Shuler, M. G. & Bear, M. F. Learning induces long-term potentiation in the hippocampus. *Science* **313**, 1093–1097 (2006).
27. Krogh, A. What are artificial neural networks? *Nat. Biotechnol.* **26**, 195 (2008).
28. Micheva, K. D. in *Corsini Encyclopedia of Psychology* (ed. Weiner, I. B. & Craighead, W. E.) 609 (John Wiley & Sons, Hoboken, NJ, 2010).
29. Liang, J. & Wong, H.-S. P. Cross-point memory array without cell selectors—device characteristics and data storage pattern dependencies. *IEEE Trans. Electron Devices* **57**, 2531–2538 (2010).
30. Wang, G. et al. High-performance and low-power rewritable SiO_x 1 kbit one diode–one resistor crossbar memory array. *Adv. Mater.* **25**, 4789–4793 (2013).
31. Gi, S., Yeo, I., Chu, M., Kim, S. & Lee, B. Fundamental issues of implementing hardware neural networks using memristor. *Presented at the 2015 International SoC Design Conference 2–5* (IEEE, Gyeongju, 2015).
32. Huang, J. J., Tseng, Y. M., Luo, W. C., Hsu, C. W. & Hou, T. H. One selector-one resistor (1S1R) crossbar array for high-density flexible memory applications. *Presented at the 2011 IEEE International Electron Devices Meeting 5–7* (IEEE, Washington, DC, 2011).
33. Linn, E., Rosezin, R., Kögeler, C. & Waser, R. Complementary resistive switches for passive nanocrossbar memories. *Nat. Mater.* **9**, 403 (2010).
34. Yoon, J. H. et al. Highly uniform, electroforming-free, and self-rectifying resistive memory in the Pt/Ta₂O₅/HfO_{2-x}/TiN structure. *Adv. Funct. Mater.* **24**, 5086–5095 (2014).
35. El-Sayed, H. & Birss, V. Controlled growth and monitoring of tantalum oxide nanostructures. *Nanoscale* **2**, 793 (2010).
36. El-Sayed, H., Singh, S. & Krusez, S. Formation of dimpled tantalum surfaces from electropolishing. *J. Electrochem Soc.* **154**, 728 (2007).
37. Roy, P., Berger, S. & Schmuki, P. TiO₂ Nanotubes: synthesis and applications. *Angew. Chem. Int. Ed.* **50**, 2904 (2011).
38. Huo, K., Gao, B., Fu, J., Zhao, L. & Chu, P. K. Fabrication, modification, and biomedical applications of anodized TiO₂ nanotube arrays. *RSC Adv.* **4**, 17300 (2014).
39. Wang, G. et al. Three-dimensional networked nanoporous Ta₂O_{5-x} memory system for ultrahigh density storage. *Nano Lett.* **15**, 6009–6014 (2015).
40. Kwon, S. et al. Structurally engineered nanoporous Ta₂O_{5-x} selector-less memristor for high uniformity and low power consumption. *ACS Appl. Mater. Interfaces* **9**, 34015–34023 (2017).
41. Cho, B. et al. Rewritable switching of one diode–one resistor nonvolatile organic memory devices. *Adv. Mater.* **22**, 1228–1232 (2010).
42. Bi, G.-q & Poo, M.-m Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
43. Abbott, L. F. & Nelson, S. B. Synaptic plasticity: taming the beast. *Nat. Neurosci.* **3**, 1178 (2000).
44. Schuman, C. D., et al. A survey of neuromorphic computing and neural networks in hardware. <https://arxiv.org/abs/1705.06963>. 2017. Accessed 16 July 2018.
45. Jain, A. K., Mao, J. & Mohiuddin, K. M. Artificial neural networks: a tutorial. *Computer* **29**, 31–44 (1996).
46. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. in *Parallel Distributed Processing* (ed. Rumelhart, D. E., McClelland, J. L.) 8, 318–363 (MIT Press, Cambridge, MA, 1985).
47. Kotsiantis, S. B. In *Emerging Artificial Intelligence Applications in Computer Engineering* (ed. Maglogiannis, I., Karpouzis, K., Wallace, M. & Soldatos, J.) Part I, 3–24, (IOS Press, Amsterdam, Netherlands, 2007).
48. Zamanidoost, E., Bayat, F. M., Strukov, D. & Kataeva, I. Manhattan rule training for memristive crossbar circuit pattern classifiers. *Presented at the 2015 IEEE 9th International Symposium on Intelligent Signal Processing 15–17* (IEEE, Siena, 2015).
49. Alibart, F., Gao, L., Hoskins, B. D. & Strukov, D. B. High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology* **23**, 075201 (2012).
50. Khodabandehloo, G., Mirhassani, M. & Ahmadi, M. Analog implementation of a novel resistive-type sigmoidal neuron. *IEEE Trans. VLSI Syst.* **20**, 750–754 (2012).