



OPEN

WO_x channel engineering of Cu-ion-driven synaptic transistor array for low-power neuromorphic computing

Seonuk Jeon¹, Heebum Kang¹, Hyunjeong Kwak², Kyungmi Noh², Seungkun Kim², Nayeon Kim¹, Hyun Wook Kim¹, Eunryeong Hong¹, Seyoung Kim² & Jiyong Woo¹✉

The multilevel current states of synaptic devices in artificial neural networks enable next-generation computing to perform cognitive functions in an energy-efficient manner. Moreover, considering large-scale synaptic arrays, multiple states programmed in a low-current regime may be required to achieve low energy consumption, as demonstrated by simple numerical calculations. Thus, we propose a three-terminal Cu-ion-actuated CuO_x/HfO_x/WO₃ synaptic transistor array that exhibits analogously modulated channel current states in the range of tens of nanoamperes, enabled by WO₃ channel engineering. The introduction of an amorphous stoichiometric WO₃ channel formed by reactive sputtering with O gas significantly lowered the channel current but left it almost unchanged with respect to consecutive gate voltage pulses. An additional annealing process at 450 °C crystallized the WO₃, allowing analog switching in the range of tens of nanoamperes. The incorporation of N gas during annealing induced a highly conductive channel, making the channel current modulation negligible as a function of the gate pulse. Using this optimized gate stack, Poole–Frenkel conduction was identified as a major transport characteristic in a temperature-dependent study. In addition, we found that the channel current modulation is a function of the gate current response, which is related to the degree of progressive movement of the Cu ions. Finally, the synaptic characteristics were updated using fully parallel programming and demonstrated in a 7 × 7 array. Using the CuO_x/HfO_x/WO₃ synaptic transistors as weight elements in multilayer neural networks, we achieved a 90% recognition accuracy on the Fashion-MNIST dataset.

In this era of big data, the amount of data to be processed has increased significantly. Data processing through the conventional von Neumann computing architecture, where the processor and memory units are physically separated, has caused significant latency bottlenecks and power consumption¹. To address this challenge, brain-inspired neuromorphic systems using parallel-connected biological synapses have been proposed as new computing architectures to enhance computation efficiency^{2–5}. To implement this architecture in hardware systems, it is important to emulate the functions of biological synapses using analog memory devices. Among the various candidates for synaptic devices, resistive switching random-access memory (RRAM) has mainly been explored because of its scalability^{6–8}. The multiple resistance states of RRAM can be demonstrated by the precise control of the formation and rupture of local conductive filaments through materials engineering and novel programming schemes⁹. However, stochastic ion motion in the switching layer interferes with the filament, causing a nonuniform synaptic response with respect to the programming pulses^{10–14}.

For uniform analog synaptic devices, ion-actuated transistors, which have also been referred to as electrochemical random-access memory or redox transistors and were inspired by the field-driven Li ion motion of rechargeable batteries, have recently been suggested¹⁵. Instead of the gate dielectric of a conventional transistor, an electrolyte that not only contains mobile ion sources but also allows ionic motion driven by the field can be utilized. As the number of ions attracted (or repulsed) to the channel layer increases (or decreases), the channel current (I_D) between the source (S) and drain (D) is gradually tuned towards higher (or lower) levels as a function of the positive (or negative) programming pulses applied to the gate (G). Thus far, this analog switching behavior

¹School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea. ²Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea. ✉email: jiyong.woo@knu.ac.kr

has been demonstrated in synaptic transistor unit cells for various combinations of electrolyte/channel material stacks and mobile ion species such as Li, H, O, and Cu ions^{15–18}. Recent advances in synaptic transistors have been extended to array-level analyses^{19–21}. Analog switching of some or all cells in arrays that can be updated simultaneously has been experimentally achieved using a half-bias scheme, where the drain voltage (V_D) and gate voltage (V_G) are used during programming. It is worth noting that the I_D of the synaptic transistor should be as low as possible considering the energy consumption at the array level²¹. This is because, unlike usual transistors serving as switch elements, analog synaptic transistors are always on-channel without a threshold voltage required to allow current to flow in the channel. The magnitude of the I_D of the synaptic transistor is tuned only by the V_G . Assuming that only the farthest synaptic transistor in the array is programmed, unselected synaptic transistors located on a horizontal (or vertical) line applied to V_G (or V_D) generate unwanted gate (or drain) leakage currents, as shown in Fig. 1a. In particular, although extremely low gate currents (I_G s) of several tens of picoamperes have been reported to date, the I_D has mainly been several microamperes^{15–18,22}. This implies that the magnitude of the I_D is primarily related to the energy consumption in the synaptic array and its impact becomes significant as the array size increases, as shown in Fig. 1b.

Therefore, we fabricated a $\text{CuO}_x/\text{HfO}_x/\text{WO}_x$ synaptic transistor array that exhibited analog switching in the range of tens of microamperes. Next, the effect of the WO_x channel layer was investigated using the sputtering method and annealing conditions related to the stoichiometry and crystallinity of the material.

The Cu-ion-actuated synaptic transistor structure developed in our previous study was used in this study¹⁸. For the synaptic transistor array, S and D lines consisting of W were patterned on an Si/SiO₂ wafer, as shown in Fig. 2a. Via-holes for the S/D contacts were defined after Si₃N₄ interlayer deposition. A WO_x channel with a length (or width) of 20 (or 5) μm was deposited by sputtering with a WO_3 single target. For comparison, a different channel layer was formed by reactive sputtering with a W metal target under ambient Ar and O plasma conditions. Additional annealing was performed at 450 °C using a rapid thermal annealing system to lower I_D in the range of nanoamperes. Subsequently, an HfO_x electrolyte and CuO_x G with a length of 10 μm were deposited by sputtering using HfO_x and Cu targets, respectively. Finally, a W capping layer was formed on top of the G and S/D contacts by sputtering. The thickness of each layer was evaluated by alpha step at the 3-D convergence center of Inha University. The fabricated $\text{CuO}_x/\text{HfO}_x/\text{WO}_x$ synaptic transistor was analyzed using transmission electron microscopy (TEM) and X-ray photoelectron spectroscopy (XPS) depth profiling, and each layer was clearly distinguished, as shown in Fig. 2b and c.

The electrical characteristics of the $\text{CuO}_x/\text{HfO}_x/\text{WO}_x$ unit synaptic transistor were evaluated by applying a V_G of +6 V (or -5 V) with a pulse width of 100 ms to increase (or decrease) the I_D , as shown in Fig. 3a. The tuned I_D was sensed by applying a V_D of 0.5 V. The observed switching behavior was attributed to the reversible driving of Cu ions back and forth through the HfO_x layer. The effective thickness of HfO_x was thinned or thickened, and the electric field-adjusted channel conductivity was enhanced or weakened²³. As reported in a previous study, a limited number of Cu ions was provided by the CuO_x G electrode, which could be effectively and precisely controlled by successive V_G pulses to allow linear and symmetric synaptic behavior in the range of tens of microamperes¹⁸. Under the given programming conditions, we first analyzed the effect of the stoichiometry of the channel layer. A stoichiometric WO_3 layer was formed by reactive sputtering with O gas injection, as shown in Fig. 3b. Although an extremely low I_D at a level below tens of picoamperes was measured, no change in I_D was observed as a function of V_G pulses. When annealing was conducted at 450 °C and atmospheric pressure for 30 min, I_D became higher and began to be analogously controlled by the V_G pulses, as shown in Fig. 3c. The WO_3 layer crystallizes at temperatures above 400 °C²⁴. The polycrystalline WO_3 layer, which was verified by the fast Fourier transform in the TEM analysis (inset of Fig. 3c), boosted the conductivity. Thus, most of the applied V_G was transferred to the HfO_x electrolyte to drive the Cu ions, leading to a progressive increase and decrease in I_D for 200 consecutive V_G pulses. However, when ambient N gas was applied during annealing, I_D was increased to hundreds of microamperes, making the I_D modulation negligible, as shown in Fig. 3d. To identify the impact of

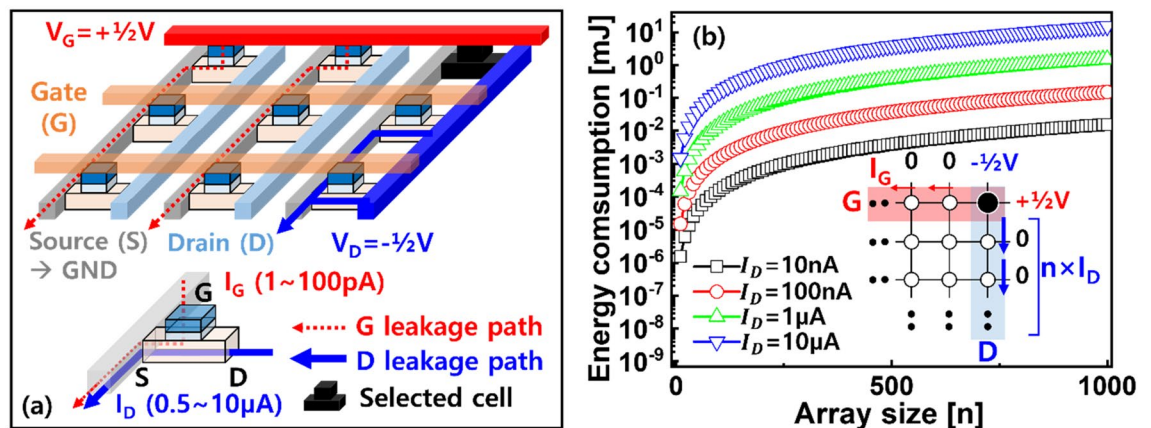


Figure 1. (a) Schematic illustration of leakage current sources in synaptic arrays. The half-bias scheme was used to program the farthest cell for simplicity. (b) Numerical calculations show the impact of leakage currents on energy consumption with respect to array size. Energy was calculated considering the leakage currents originated from unselected synaptic transistors, where either V_G or V_D was applied.

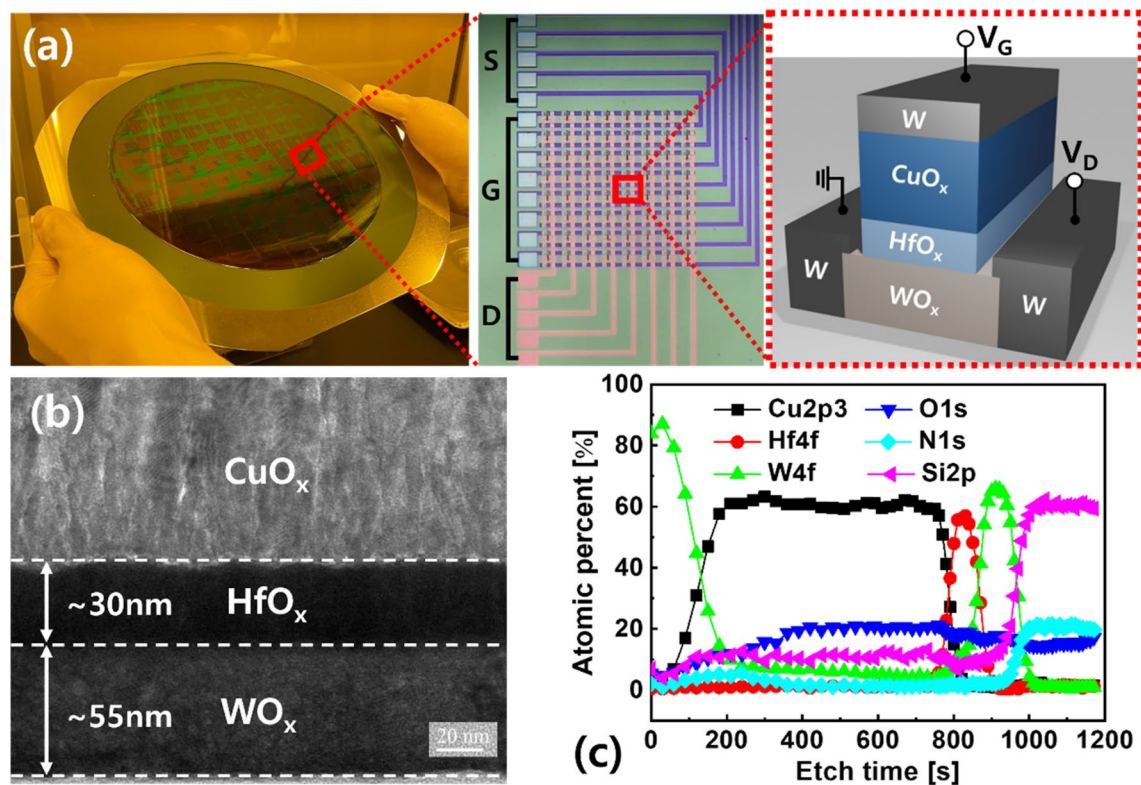


Figure 2. (a) Fabricated $\text{CuO}_x/\text{HfO}_x/\text{WO}_x$ synaptic transistor arrays. I_D of the synaptic transistor was updated by applying V_G and grounding S. Updated I_D was read by addressing voltage to D. (b) Cross-sectional TEM image and (c) XPS depth profiling of the synaptic transistor.

the annealing gas on analog switching, the W 4f peak at the WO_3 channel interface of the two annealed devices, obtained through XPS analysis, was investigated, as shown in Fig. 3e. Two peaks at binding energies of approximately 31 and 34 eV, indicating metallic states related to nonstoichiometric WO_x were clearly observed for both samples. However, two peaks corresponding to WO_3 observed at 36.5 and 38 eV were noticeable for the annealed sample without N gas. This implies that N preferentially bonds to W. W–N bonding is usually represented by metallic peaks (approximately 31 and 34 eV); therefore, the enhanced channel conductivity can be attributed to the conductive WN formed.

Next, we studied the temperature dependence of the I_G to identify the mechanisms of analog switching in the optimized gate stack. I_G , which was obtained by applying voltage to G and grounding S, was measured by raising the temperature to 140 °C. As a result, the transport characteristics were suitably fitted using the Poole–Frenkel emission, as shown in Fig. 4a. This indicates that bulk switching across the entire electrolyte area was involved, rather than metallic conduction, owing to the locally clustered Cu ions. Specifically, by comparing simultaneously obtained I_G and I_D , we found that the I_G response to V_G pulses was related to I_D modulation, as shown in Fig. 4b. That is, the gradual transition of I_G by a single V_G pulse during initial programming was typically projected onto the lateral I_D , resulting in a linear I_D response in working cell. However, at the given programming condition, nonlinear I_D behavior was often observed in non-optimized stacks. When we examined the I_G response of these failed cells, the abrupt increase and saturation of I_G was observed. These results imply that the extent to which Cu ions move steadily through the electrolyte per V_G pulse is related to the linearity and symmetry of I_D .

Finally, we evaluated synaptic arrays employing the engineered WO_3 channel. As the channel resistance was reduced, a larger electric field across the HfO_x layer was used to drive the Cu ions; therefore, analog switching could be achieved by lowering V_G to +3 V. To facilitate fully parallel programming, all the G, S, and D pads of the 7×7 array were contacted by a probe card, as shown in Fig. 5a. In the half-bias scheme, the +3 V (or –3 V) required to program the unit cell was halved, and +1.5 V (or –1.5 V) was simultaneously applied to V_G and V_D . Consequently, a full voltage of approximately ± 3 V was applied to the selected cells. Reliable analog switching behavior was obtained for all synaptic transistors. The cell-to-cell update curves expressed as the median repeatedly increase and decrease, as the polarity of the pulse changes every 50 times for a total of 200 pulses, as shown in Fig. 5b. Specifically, the I_D states updated by the first 50 potentiation pulses reached its maximum value and returned to the initial state by the next 50 depression pulses. As a result, the I_D states applied at the first, 100th, and 200th pulses (or 50th and 150th) showed uniform cycle-to-cycle distribution, as shown in Fig. 5c. Note that although the reversible analog I_D modulation from all cells was achieved, cell-to-cell uniformity in the array needed to be improved. We recently revealed that the effect of Cu ion distribution (or concentration) in the stack on the synaptic behavior through physics-based simulation²⁵. Based on the clue, further study is currently underway on process development and strategies to ensure reliability as well as uniformity by controlling the adequate amount of Cu ions in the device.

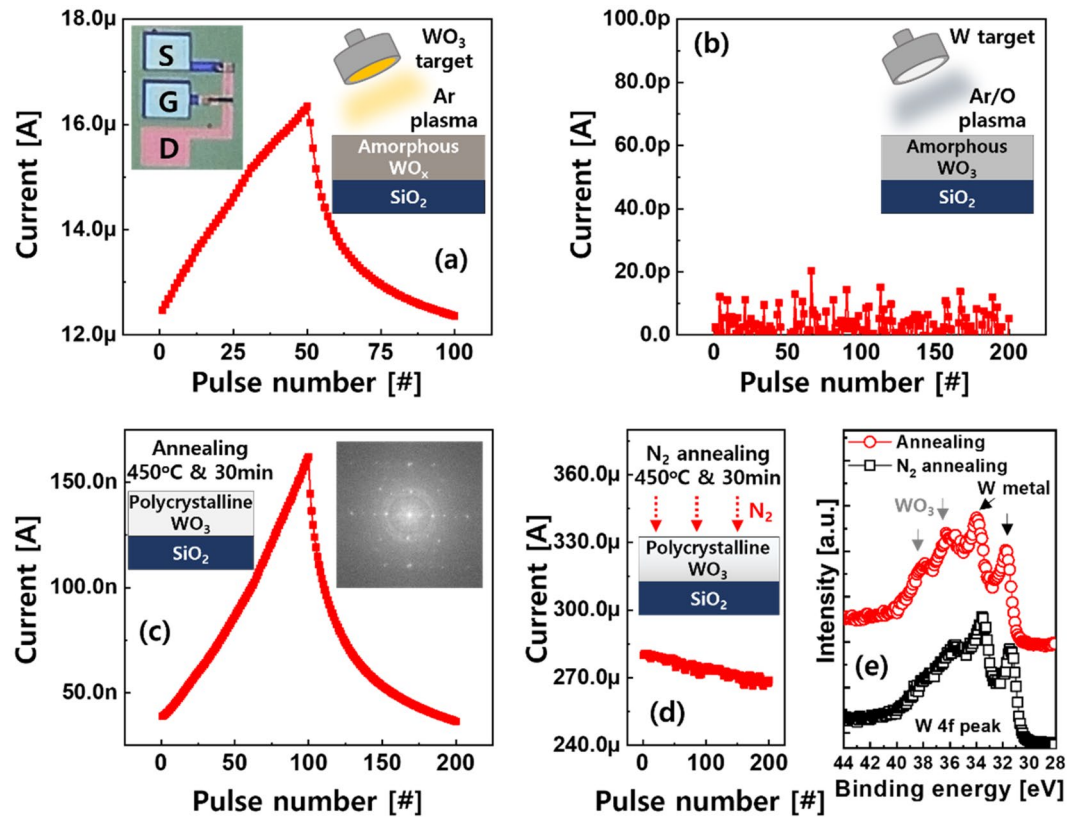


Figure 3. V_G of +6 V (or -5 V) with a pulse width of 100 ms was used for all devices to evaluate device characteristics. (a) Analog synaptic behavior observed in the synaptic transistor employing amorphous WO_x channel. (b) No switching was observed when relatively stoichiometric WO₃ channel was used. (c) Crystallized WO₃ channel, which was verified by fast Fourier transformation image, allowed the I_D to be adjusted by pulse number. (d) Increased I_D level due to N annealing hid the switching behavior. (e) W 4f peak obtained through XPS analysis.

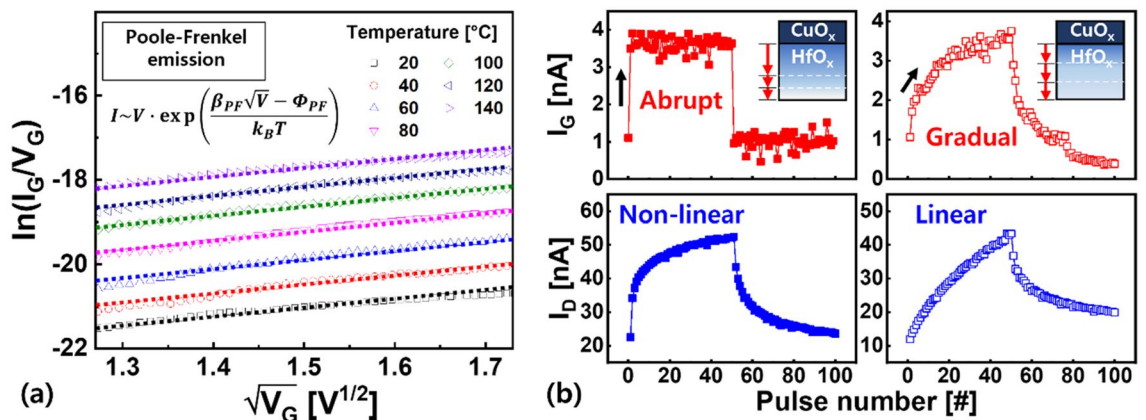


Figure 4. (a) Poole-Frenkel conduction of the gate stack. (b) Link between the I_G response, which can be described by Cu ion motion over the electrolyte, and I_D modulation.

Based on these results, we built multilayer neural networks with size 784-256-128-10 (from input to output neurons) designed for image recognition, as shown in Fig. 5d. The input signals were transferred from input neurons to output neurons through synaptic weights, which served as the optimized synaptic transistor in this study. The degree of linearity of synaptic characteristics is expressed by a linearity factor, α , which is calculated by the following equation:²⁶

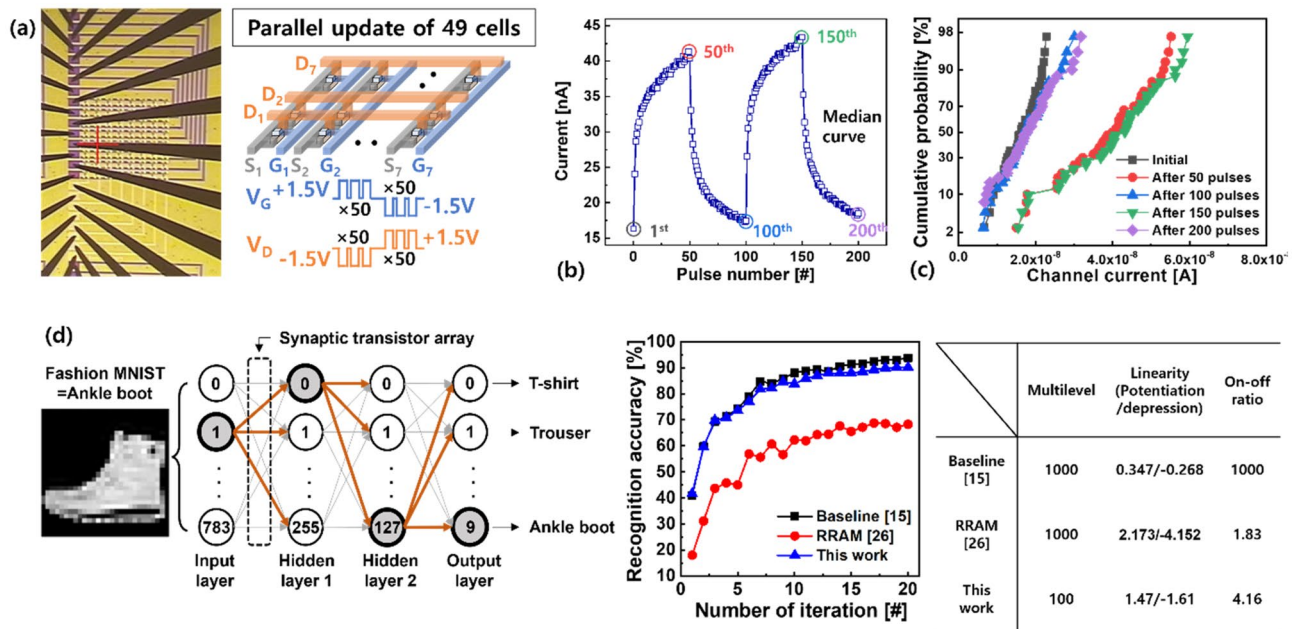


Figure 5. (a) Optical microscope image of fabricated synaptic transistor array contacted by probe card. To update all synaptic transistor cells simultaneously, 50 V_G (or V_D) pulses of +1.5 V (or -1.5 V) were consecutively addressed to all G (or D) lines to increase I_D . By changing the polarity of each pulse, I_D can be steadily lowered. (b) Reliable synaptic behaviors expressed as median were obtained from all cells. (c) Distributions of I_D s when pulses were applied every 50 cycles from the initial state. (d) Pattern recognition accuracy of multilayer perceptron evaluated using IBM analog hardware acceleration kit simulator. The parameters used in this simulation were listed in the table.

$$G = \begin{cases} ((G_{MAX}^\alpha - G_{MIN}^\alpha) \times \omega + G_{MIN}^\alpha)^{\frac{1}{\alpha}} & (if \alpha \neq 0) \\ G_{MIN} \times (G_{MAX}/G_{MIN})^\omega & (if \alpha = 0) \end{cases}$$

where, G_{MAX} and G_{MIN} are conductance at the maximum and minimum I_D state, respectively, and ω is an internal variable which ranges from 0 to 1. Moreover, α is equal to 1 in the case of the ideal synaptic behavior. Based on these equations, linearity of potentiation (or depression) of 1.47 (or -1.61) was achieved for the developed synaptic transistor. The difference between the maximum and minimum I_D states of ~4 was also used in the simulations. Subsequently, the recognition accuracy on the Fashion-MNIST dataset was evaluated with a learning rate of 0.01 using the IBM analog hardware acceleration kit simulator by assuming that the achieved synaptic transistors served as weight elements²⁶. Consequently, our device exhibited a recognition accuracy of approximately 90%, which is comparable to the baseline, when the multilevel states were almost linearly adjusted^{15,27}.

Finally, through simple numerical analysis, we showed that update energy consumption is reduced by lowering I_D of the synaptic transistors, which are assumed to be configured in 1 K array, as shown in Fig. 6. Compared to reported synaptic transistors that utilize the Li or O ions, the V_G to drive the Cu ions seemed to be larger. Current has a greater impact on power than voltage because its magnitude can be changed significantly. To date, the energy consumed to change the I_D state by a single V_G pulse has been mainly discussed considering unit cell operation²⁸. However, to sequentially update all cells in the array, the V_G and V_D signals were applied to the array via all interconnect lines. Thus, synaptic transistors with tunable I_D in the low current regime enable low-power operation because the I_D is typically higher than the I_G . In addition to energy consumption, recent study showed that recognition accuracy is degraded due to interference from high I_D s of neighboring cells²⁹. These results mean that the range of I_D , which has often been ignored, should be considered when designing a three-terminal synaptic transistor.

In this work, we achieved analog switching in a low I_D regime (< 100 nA) of a 7×7 $CuO_x/HfO_x/WO_3$ synaptic transistor array using WO_3 channel engineering. Specifically, we revealed that the I_D range could be lowered or increased by utilizing a stoichiometric WO_3 layer or by strengthening the crystallinity. In addition, the steady and gradual motion of the Cu ions across the gate stack enabled a linearly tuned I_D response. The obtained synaptic behavior not only accurately inferred the Fashion-MNIST dataset but also performed recognition with low energy consumption.

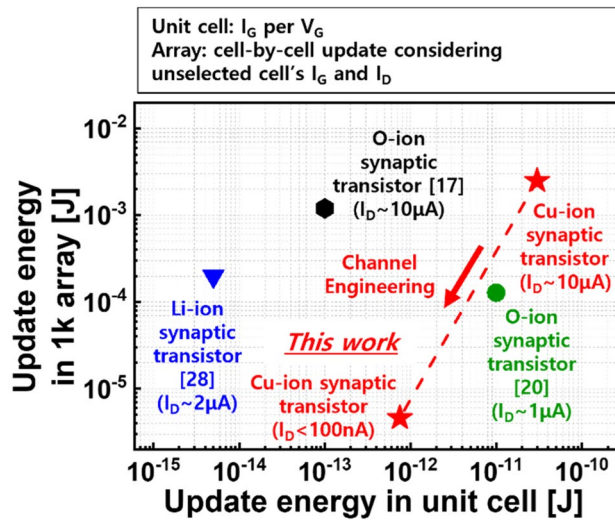


Figure 6. From a unit cell perspective, synaptic transistors that exhibit low I_G and are driven by low V_G seem to be preferred for low-power application. However, considering update operation in the array, the I_D range of the synaptic transistors contributes more to energy consumption.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 27 September 2023; Accepted: 6 December 2023

Published online: 13 December 2023

References

- Xia, Q. & Yang, J. J. Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* **18**, 309–323 (2019).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Mead, C. Neuromorphic electronic systems. *Proc. IEEE* **78**(10), 1629–1636 (1990).
- Esser, S. K. *et al.* Convolutional networks for fast, energy-efficient neuromorphic computing. *PNAS* **113**(41), 11441–11446 (2016).
- Tang, J. *et al.* Bridging biological and artificial neural networks with emerging neuromorphic devices: Fundamentals, progress, and challenges. *Adv. Mater.* **31**, 1902761 (2019).
- Burr, G. W. *et al.* Neuromorphic computing using non-volatile memory. *Adv. Phys.: X* **2**(1), 89–124 (2017).
- Boybat, I. *et al.* Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **9**, 2514 (2018).
- Ielmini, D. & Ambrogio, S. Emerging neuromorphic devices. *Nanotechnology* **31**, 092001 (2019).
- Woo, J. *et al.* Improved synaptic behavior under identical pulses using $\text{AlO}_x/\text{HfO}_2$ bilayer RRAM array for neuromorphic systems. *IEEE Trans. Electron Dev.* **63**(12), 5064–5067 (2016).
- Wu, W. *et al.* Improving analog switching in HfO_x -based resistive memory with a thermal enhanced layer. *IEEE Electron Dev. Lett.* **38**(8), 1019–1022 (2017).
- Choi, S. *et al.* SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations. *Nat. Mater.* **17**, 335–340 (2018).
- Woo, J. *et al.* Optimized programming scheme enabling linear potentiation in filamentary HfO_2 RRAM synapse for neuromorphic systems. *IEEE Electron Dev. Lett.* **37**(8), 994–997 (2016).
- Jeong, Y., Kim, S. & Lu, W. D. Utilizing multiple state variables to improve the dynamic range of analog switching in a memristor. *Appl. Phys. Lett.* **107**, 173105 (2015).
- Zhao, M., Gao, B., Tang, J., Qian, H. & Wu, H. Reliability of analog resistive switching memory for neuromorphic computing. *Appl. Phys. Rev.* **7**, 011301 (2020).
- Tang, J. *et al.* ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing. In *2018 IEEE International Electron Devices Meeting (IEDM)* 13.1.1–13.1.4 (2018).
- Liu, L. *et al.* Proton-assisted redox-based three-terminal memristor for synaptic device applications. *Appl. Mater.* **15**(24), 29287–29296 (2023).
- Kim, S. *et al.* Metal-oxide based, CMOS-compatible ECRAM for deep learning accelerator. In *2019 IEEE International Electron Devices Meeting (IEDM)* 35.7.1–35.7.4 (2019).
- Kang, H., Kim, H. W., Hong, E. R. & Woo, J. Analog synaptic behavior of mobile ion source-limited electrochemical RAM using CuO_x oxide electrode for deep learning accelerator. *Appl. Phys. Lett.* **120**, 122101 (2022).
- Li, Y. *et al.* In situ parallel training of analog neural network using electrochemical random-access memory. *Front. Neurosci.* **15**, 636127 (2021).
- Lee, C., Kwak, M., Choi, W., Kim, S. & Hwang, H. Improved on-chip training efficiency at elevated temperature and excellent inference accuracy with retention (> 108 s) of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_{3-x}$ ECRAM synapse device for hardware neural network. In *2021 IEEE International Electron Devices Meeting (IEDM)* 12.3.1–12.3.4 (2021).
- Kang, H. *et al.* Analysis of electro-chemical RAM synaptic array for energy-efficient weight update. *Front. Nanotechnol.* **4**, 1034357 (2022).
- Kang, H. & Woo, J. Cu-ion-actuated three-terminal neuromorphic synaptic devices based on binary metal-oxide electrolyte and channel. *Appl. Phys. Lett.* **119**, 072103 (2021).
- Kim, N., Kang, H., Kim, H. W., Hong, E. & Woo, J. Understanding synaptic characteristics of nonvolatile analog redox transistor based on mobile ion-modulated-electrolyte thickness model for neuromorphic applications. *Appl. Phys. Lett.* **121**, 072105 (2022).

24. Woo, J. *et al.* Introduction of WO₃ layer in a Cu-based Al₂O₃ conductive bridge RAM system for robust cycling and large memory window. *IEEE J. Electron Dev. Soc.* **4**(3), 163–166 (2016).
25. Tessler, N., Kim, N., Kang, H. & Woo, J. Switching mechanisms of CMOS-compatible ECRAM transistors-electrolyte charging and ion plating. *J. Appl. Phys.* **134**, 7 (2023).
26. Rasch, M. J. *et al.* A flexible and fast PyTorch toolkit for simulating training and inference on analog crossbar arrays. In *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)* 1–4 (2021).
27. Gong, N. *et al.* Signal and noise extraction from analog memory elements for neuromorphic computing. *Nat. Commun.* **9**, 2102 (2018).
28. Lee, J., Nikam, R. D., Kim, D. & Hwang, H. Highly scalable (30 nm) and ultra-low energy (~5fJ/pulse) vertical sensing ECRAM with ideal synaptic characteristics using ion-permeable graphene electrodes. In *2022 IEEE International Electron Devices Meeting (IEDM)* (2022).
29. Kim, S., Son, J., Kwak, H. & Kim, S. Accurate weight update in an electrochemical random-access memory based cross-point array using channel-high half-bias scheme for deep learning accelerator. *Adv. Electron. Mater.* **2023**, 2300476 (2023).

Acknowledgements

This research was supported by the National R&D Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (RS-2023-00258227), and also supported by the National R&D Program through the NRF funded by Ministry of Science and ICT (2020M3H2A1078045). This research was supported by the MOTIE (Ministry of Trade, Industry & Energy (RS-2023-00236568 & RS-2023-00231956) and KSRC (Korea Semiconductor Research Consortium) (1415187475 & 1415187361) support program for the development of the future semiconductor device.

Author contributions

S. J., H. K., and H. K. conducted the experiments and characterized the devices. K. N. and S. K. characterized the arrays. S. J., H. K., H. K., K. N., S. K., N. K., H. W. K., E. H., S. K., and J. W. discussed the results. S. K. and J. W. supervised the study. S. J., H. K., S. K. and J. W. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.W.

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

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



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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023