scientific reports

OPEN

Strategy to improve synaptic behavior of ion-actuated synaptic transistors—the use of ion blocking layer to improve state retention

Seonuk Jeon[1], Nir Tessler[2,3], Nayeon Kim[1], Eunryeong Hong[1], Hyun Wook Kim[1] & Jiyong Woo[1]✉

Synaptic transistors (STs) with a gate/electrolyte/channel stack, where mobile ions are electrically driven across the solid electrolyte, have been considered as analog weight elements for neuromorphic computing. The current (I_D) between the source and drain in the ST is analogously updated by gate voltage (V_G) pulses, enabling high pattern recognition accuracy in neuromorphic systems; however, the governing physical mechanisms of the ST are not fully understood yet. Our previous physics-based simulation study showed that ion movement in the electrolyte, rather than the electrochemical reactions that occur in the channel, plays an important role in switching. In this study, we experimentally explore the properties of the HfO_x electrolyte and show that by tuning the density of oxygen vacancies, it can assume the dual role of electrolyte and channel. We demonstrate analog synaptic behavior using a novel ST with a two-layer stack of CuO_x/HfO_x, where the CuO_x is the gate and Cu ion reservoir, and the HfO_x is the electrolyte and channel. To improve state retention and linearity, we introduce a Cu ion transport barrier in the form of a dense and stoichiometric Al_2O_3 layer. The $CuO_x/Al_2O_3/HfO_x$ exhibits excellent state retention and improved potentiation and depression response. Energy dispersive spectroscopy mapping following potentiation confirms the role of the Al_2O_3 layer in confining the Cu ions in the HfO_x layer. We also show that a two-step programming scheme can further enhance synaptic response and demonstrate high recognition accuracy on the Fashion-MNIST dataset in simulation.

Recently, with the rapid increase in the amount of data, the conventional von Neumann architecture, which processes data through a series of operations between the processing unit and memory, has created a bottleneck effect that slows data processing. To overcome this, neuromorphic computing architecture, based on highly parallel analog computations inspired by data transmission through numerous synapses in the human brain, has been attracting attention[1–3]. To implement this architecture in hardware, a synaptic device that emulates the role of a biological synapse is required[4,5]. Static random-access memory (SRAM) has been utilized as a synaptic device; however, owing to the large size of the SRAM cell (over 100 F^2, where F is the technology node), it is challenging to implement hundreds of millions of synapses in neuromorphic computing systems[6]. For this reason, various two-terminal emerging memory devices such as ferroelectric memory[7], magnetic memory[8], phase change memory[9–11], and resistive memory (RRAM)[12] have been proposed. Among these, RRAM has been mainly explored owing to its low power consumption, sub-10 nm scaling, nonvolatility, and multilevel characteristics[13–15]. However, its filamentary switching mechanism inevitably leads to resistance states, indicating that the synaptic weights are probabilistically tuned, which causes performance degradation in pattern recognition applications[16]. This necessitates a new ion-actuated three-terminal synaptic transistor (ST) with a gate/electrolyte/channel stack for predictable and tunable analog synaptic weights[17]. The physical mechanism of the ST has not yet been elucidated; nevertheless, its plausible working principle has been mainly described by a two-step process: (*i*) ion migration through solid electrolyte and (*ii*) electrochemical reaction at the channel[18]. When a positive gate voltage (V_G) pulse is applied to the gate of the ST, mobile ions originating from the gate or incorporated into the

[1]School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea. [2]Electrical and Computer Engineering, Sara and Moshe Zisapel Nano-Electronic Center, Technion Israel Institute of Technology, 32000003 Haifa, Israel. [3]Department of Earth and Environmental Engineering, Columbia University, New York, NY 10027, USA. ✉email: jiyong.woo@knu.ac.kr

electrolyte are driven toward the channel in the vertical direction. Various mobile ions such as Li^{+}[17,19], O^{2-}[20,21], H^{+}[22,23], and Cu^{+}[24,25] have been explored. The broadly accepted picture is that the switching takes place only when the ions reach the channel and directly dope or convert the valence state of the channel's atoms. Thus, the extent of the ions intercalated into the channel, which is related to the electrochemical potential applied to the gate stack by the V_G bias, analogously increases or decreases the current between the source and drain (I_D), corresponding to potentiation or depression, respectively. We have recently shown that electrolyte charging also induces a charge in the channel material and that, in some cases, this may be the preferred mechanism[26]. Moreover, our physics-based simulation results of CuO_x-gate/HfO_x-electrolyte/WO_x-channel stacks have shown that Cu intercalation into the WO_x channel leads to Cu plating, loss of linearity, and enhanced degradation[18].

In this study, we focus on the HfO_x layer and use a CuO_x/HfO_x ST to demonstrate that by tuning the density of oxygen vacancies, it can also assume the role of the conducting channel making the WO_x layer redundant. We also show that an ultrathin Al_2O_3 film, inserted between the CuO_x and HfO_x layers, acts as an ion barrier that confines the Cu ions to the HfO_x layer and improves the state retention and linearity.

The three-terminal STs composed of all CMOS compatible electrodes and layers were formed on a Si substrate with a thermally grown 100 nm-thick SiO_2 wafer. First, as shown in Fig. 1a, source (S) and drain (D) contacts were patterned and deposited by sputtering with W target at a power of 50 W. The patterning was performed via conventional photolithography, developing, and lift-off processes. Afterwards, a 5 nm-thick HfO_x electrolyte with a length (width) of 50 (150) μm was deposited by sputtering with HfO_2 target at a power of 100 W under Ar and O_2 gas flows at the rate of 25 and 5 sccm, respectively. Next, a 360 nm-thick CuO_x gate electrode was deposited by sputtering with Cu target at 100 W using Ar and O_2 gases at rates of 27 and 3 sccm, respectively. Finally, a W capping layer was deposited to prevent unwanted copper oxidation. The fabricated CuO_x/HfO_x ST was analyzed by transmission electron microscopy (TEM) and X-ray photoelectron spectroscopy (XPS), as shown in Fig. 1b,c. As mentioned in our recent publications[25], to limit the number of mobile Cu ions participating in I_D switching, the CuO_x gate electrode was adopted instead of the previously proposed Cu gate. As shown in Fig. 1d, the intensity of the Cu–O bonding at binding energies of 943 and 948 eV was detected in the measured Cu 2p peak[27]. Further, a non-stoichiometric HfO_x layer comprising both Hf–Hf metal and Hf–O oxide bonds were observed (Fig. 1e)[28]. As discussed below, we consider the oxygen vacancies as facilitators of copper ions transport.

Figure 2 shows the importance of the HfO_x stoichiometry and the role of oxygen vacancies in determining the electrolyte's properties. Figure 2a shows the response for a device using the previously developed HfO_x[25], sputtered using 30 sccm Ar only. Although a low initial I_D of 28 nA was obtained, the potentiation pulses of $V_G = +3$ V and a pulse width of 100 ms flood the entire layer with Cu ions, thus reaching the maximum current within the first pulse. Applying the read voltage of 0.5 V to D and 0 V to S results in a current of about a 1.4 mA. Since the Cu ions transport is expected to be assisted by oxygen vacancies, we introduced oxygen flow and used sccm of 20:10 Ar to O_2, respectively. The response to potentiation pulses is shown in Fig. 2b. Reduced oxygen vacancies render the HfO_x insulating to electron and Cu ion transport. Following a few pulses with low I_D, the

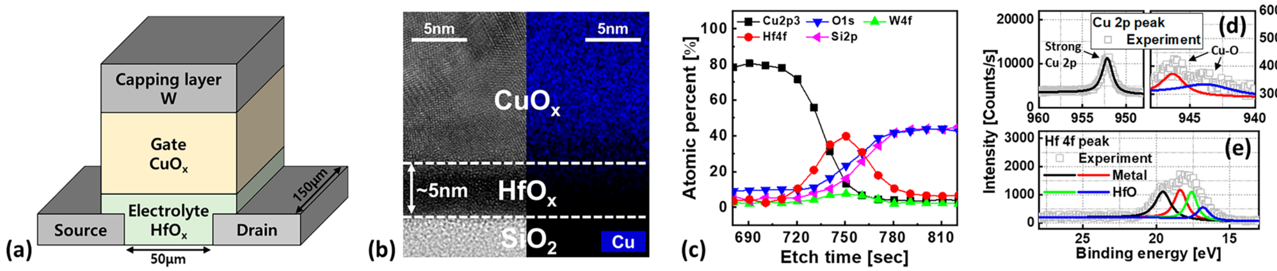

Figure 1. (**a**) Schematic diagram, (**b**) cross-sectional TEM image, and (**c**) XPS depth profiling of the fabricated CuO_x/HfO_x ST. (**d**) Cu 2p and (**e**) Hf 4f peak intensities of CuO_x and HfO_x layers, respectively.

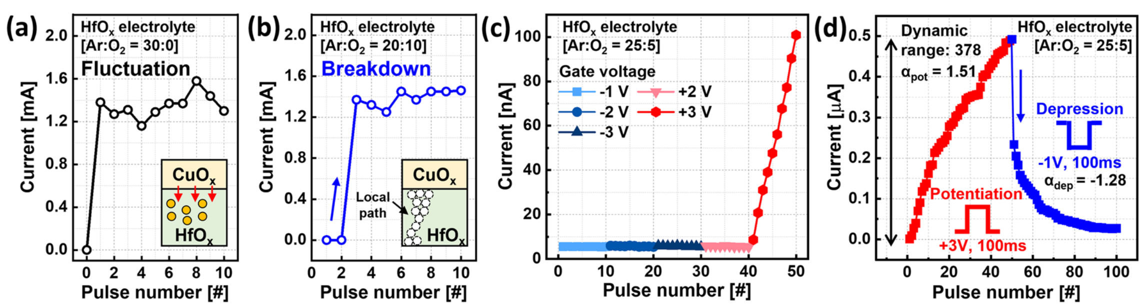

Figure 2. (**a**,**b**) Impact of Ar and O_2 gas flow rates during HfO_x deposition on the I_D response of the CuO_x/HfO_x ST. (**c**) I_D response of the CuO_x/HfO_x ST, with optimized HfO_x stoichiometry, as a function of polarity and amplitude of V_G. (**d**) The update curve of I_D in the ST employing optimized HfO_x electrolyte layer.

current abruptly jumps to its maximum value. It means that the high field concentrated across the insulating HfO_x layer led to permanent oxide breakdown (inset to Fig. 2b), resulting in low gate controllability.

The response of an optimized device using an HfO_x sputtered under 25:5 Ar:O_2 flow is shown in Fig. 2c,d. Figure 2c, shows the response to potentiation by 10 pulses of 100 ms width and V_G values between − 3 V and 3 V. We note that $V_G = 3$ V initiates a linear potentiation response with the Cu ions transport into the 5 nm HfO_x layer being well controlled. An extended potentiation/depression response using 50 pulses is shown in Fig. 2d. While not being an ideal response, it clearly demonstrates the importance of controlling the density of oxygen vacancies to achieve well behaved gate controlled uniform Cu-ion migration throughout the electrolyte.

Based on the synaptic behavior in Fig. 2d, the linearity factor, α, was calculated by the following equation[29]:

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

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 (α_{pot}) of 1.51 was achieved during potentiation in CuO_x/HfO_x ST. A hint for the process causing the sublinear potentiation can be found in the depression response to negative V_G pulses with amplitude of − 1 V and pulse width of 100 ms. The first pulse results in more than 50% reduction of the current with the response to the following pulses saturating quickly. The resulting a non-linear response has a linearity of depression (α_{dep}) of − 1.28. Based on the results of Fig. 2d, we postulated that the nonlinearity is associated with facile Cu ions transport out of the HfO_x layer.

To improve the retention of the Cu ions within the HfO_x layer we introduced a 2 nm Al_2O_3 film between CuO_x and HfO_x. We used atomic layer deposition to ensure a relatively dense and stoichiometric layer that would serve as a partial barrier for Cu ion transport. The use of ultrathin Al_2O_3 film is important to avoid introducing extra ion resistance that may hamper the dynamic range. The deposition was done at a chamber temperature of 200 °C using trimethylaluminum and water sources. The approximately 2 nm-thick Al_2O_3 layer was deposited at a deposition rate of approximately 1.1 Å/cycle. Figure 3a shows an high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image of the $CuO_x/Al_2O_3/HfO_x$ stack where energy dispersive X-ray spectroscopy (EDS) mapping of Hf (Fig. 3b) and of Al (Fig. 3c) confirm the layers' position. Figure 3d shows Cu mapping following 100 potentiation pulses of the $CuO_x/Al_2O_3/HfO_x$ ST. We note a uniform distribution within the HfO_x layer and no Cu signal within the Al_2O_3 one. The uniformity in the HfO_x layer is a testament to the successful morphology and stoichiometry supporting uniform Cu ion injection. Most importantly, the absence of Cu signal in the Al_2O_3 layer supports the notion that it acts as a barrier where the particles can only go through but not reside within.

Figure 4a shows the analog synaptic behavior of the $CuO_x/Al_2O_3/HfO_x$ ST to be compared with that of the CuO_x /HfO_x ST shown in Fig. 2d. The insertion of the Al_2O_3 layer improves the potentiation and somewhat mitigates the initial current drop in the early depression phase. The degree of the I_D change per pulse became relatively constant except for the first pulse during the depression stage, resulting in α_{dep} of − 0.48.

While it should be possible to enhance the device architecture further, we chose to demonstrate the effect of using two gate pulses instead of just one. Essentially, a short pulse of opposite polarity with a width of 50 ms was added to the conventional single V_G pulse (see inset to Fig. 4b). Naturally, adding the opposite polarity pulse

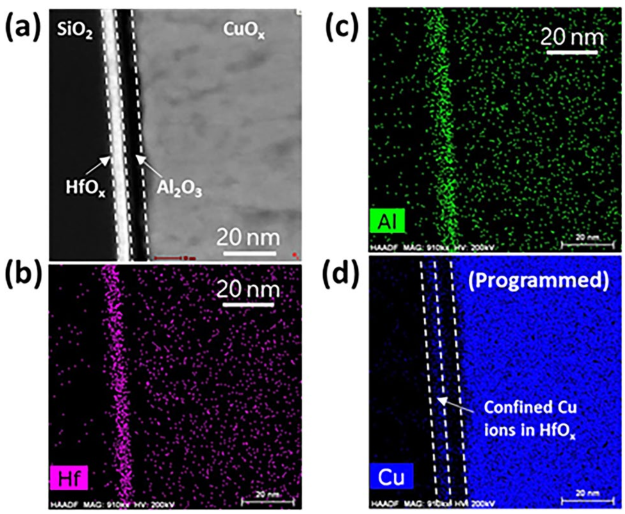

Figure 3. (**a**) HAADF-STEM image of the $CuO_x/Al_2O_3/HfO_x$ stack. (**b**,**c**) EDS mapping image showing the position of the HfO_x layer through (**b**) the Hf signal and of the Al_2O_3 layer through (**c**) the Al signal. (**d**) EDS mapping of $CuO_x/Al_2O_3/HfO_x$ ST following 100 potentiation pulses showing the Al_2O_3 layer being free of Cu ions which are being confined to the HfO_x layer.

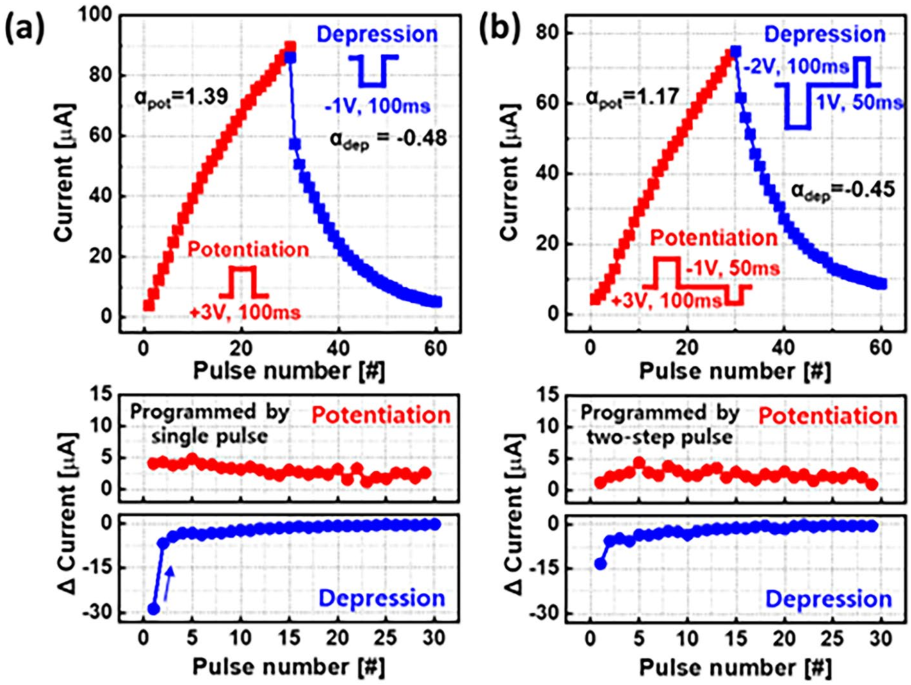

Figure 4. The update curve of I_D in the $CuO_x/Al_2O_3/HfO_x$ ST programmed by (**a**) conventional single V_G pulse method and (**b**) two-step programming pulse scheme. As a result, the degree of the I_D per V_G pulse became uniform even in depression.

reduced the dynamic range in Fig. 4b compared to Fig. 4a. However, the linearity was much improved, and almost symmetric synaptic behaviors were obtained. Most notably, the initial depression drop was mitigated in the ST using the $CuO_x/Al_2O_3/HfO_x$ stack. The calculated linearity parameters α_{pot} and α_{dep} are 1.17 and – 0.45, respectively.

To show that the Al_2O_3 layer confines the Cu ions to the HfO_x layer and acts as partial barrier to Cu ion transport, we tested the state retention during analog switching (Fig. 5). As the top of Fig. 5a shows, the test procedure has a basic block consisting of 10 potentiation pulses followed by a sequence of read pulses for 100 s. This block is then repeated several times. In the context of Fig. 2, we mentioned that the sublinear potentiation is probably associated with poor state retention and fast discharge during potentiation. Figure 5a clearly shows that the CuO_x/HfO_x ST has poor retention. In contrast, the response of the $CuO_x/Al_2O_3/HfO_x$ ST (Fig. 5b) shows a stable state-retention allowing us also to test the stability during depression. Namely, the Al_2O_3 layer is acting as a barrier confining the Cu ions to the HfO_x layer and thus preventing state discharge during read operation.

Finally, we built a multilayer neural network comprising the input, hidden, and output layers based on back-propagation algorithms, as demonstrated in Fig. 5c. The input, first hidden, second hidden, and output layers were composed of 784, 250, 125, and 10 neurons, respectively. The signals were transferred from input neurons to output neurons through synaptic weights, which served as the fabricated STs in this study. The recognition accuracy on the Fashion-MNIST dataset was evaluated using an IBM analog hardware-acceleration simulator kit (AIHWKIT) with a learning rate of 0.01[30]. When highly asymmetric synaptic behavior owing to abrupt I_D drop during depression obtained from a CuO_x/HfO_x ST stack was used, low recognition accuracy of approximately 67% was achieved, which corresponds to an error rate of approximately 33% at 20 iterations, as shown in Fig. 5d. On the other hand, the near-ideal recognition accuracy of approximately 93% with significantly lower error rate of 7% was obtained by exploiting a $CuO_x/Al_2O_3/HfO_x$ ST.

To conclude, we introduced analog switching based on a structure that avoids the use of WO_x channel layer, letting the HfO_x take the role of both the electrolyte and the conducting channel. For this, the stoichiometry of the HfO_x had to be fine-tuned to support stable Cu ion transport as well as electron transport (probably via trap-assisted tunneling)[31]. A sweet spot in terms of oxygen vacancies was found for the process involving 25:5 $Ar:O_2$ sccm flow. However, a 5 nm HfO_x that is directly attached to the Cu ion reservoir (CuO_x) exhibits poor state retention as the Cu ions are easily pulled back into the CuO_x layer. This resulted in a sub-linear potentiation response and, more pronouncedly, a 50% drop during the first depression pulse (Fig. 2). To mitigate the facile pullback of Cu ions, we use an ALD process to introduce a relatively dense and stoichiometric Al_2O_3 layer between the HfO_x and the CuO_x. EDS studies (Fig. 3) showed that the Al_2O_3 layer acts as a barrier confining the Cu ions to the HfO_x layer. Consequently, the $CuO_x/Al_2O_3/HfO_x$ stack shows improved state retention and a better linear response (Figs. 4,5). Lastly, to test the quality of the $CuO_x/Al_2O_3/HfO_x$ ST response, we implemented it to simulate pattern recognition using IBM AIHWKIT, resulting in an error rate as low as 3%.

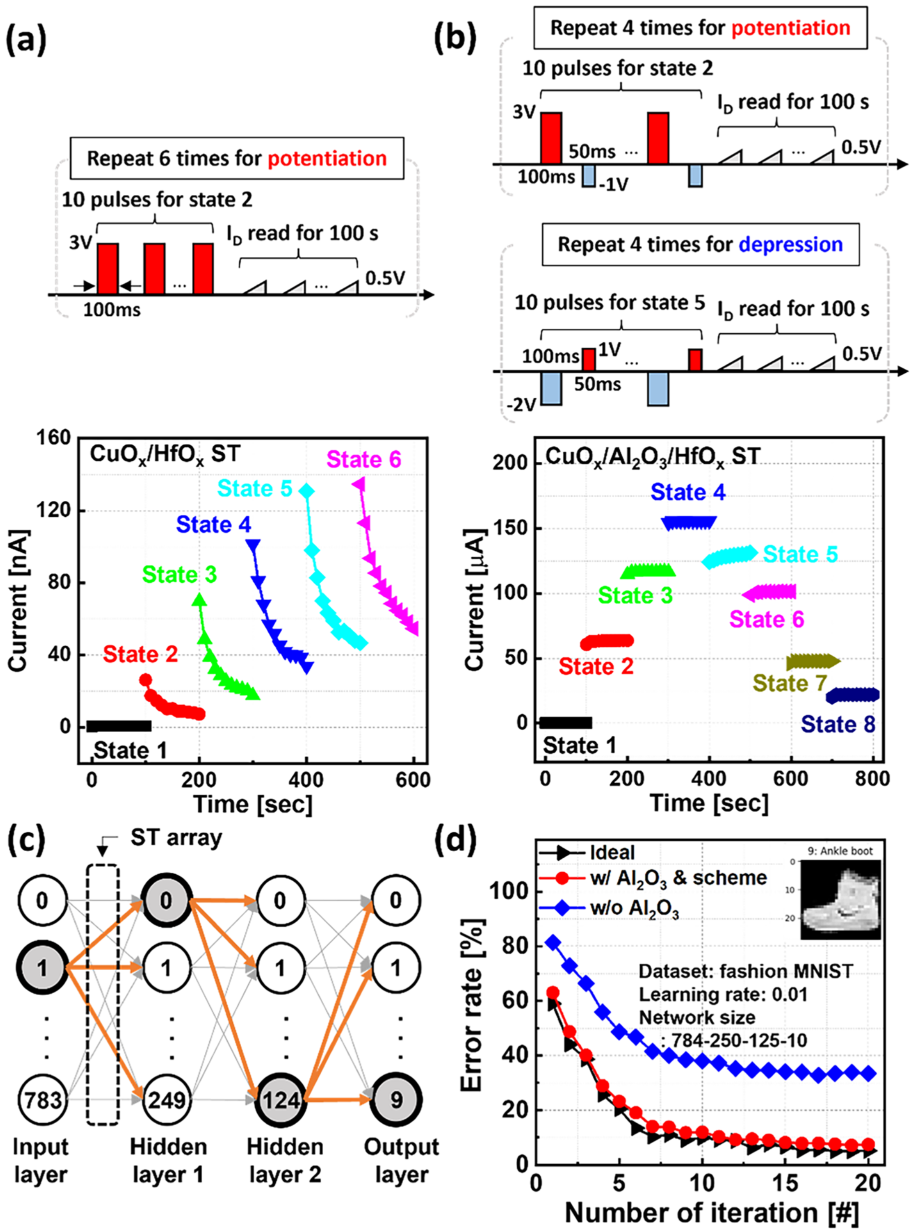

Figure 5. (**a**) Multiple I_D states over time in ST with CuO_x/HfO_x. (**b**) Reliable I_D states for programmed and erased by two-step programming pulse scheme in ST with $CuO_x/Al_2O_3/HfO_x$ stack, measured at room temperature for 100 s. (**c**) Schematic of four-layer-based artificial neural network. (**d**) Recognition accuracy results in the case with and without Al_2O_3 layer and programming scheme. Near-ideal accuracy was evaluated in ST with $CuO_x/Al_2O_3/HfO_x$ programmed under two-step pulse scheme.

Data availability

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

Received: 19 January 2024; Accepted: 26 February 2024
Published online: 29 February 2024

References

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Mead, C. Neuromorphic electronic systems. *Proc. IEEE* **78**(10), 1629–1636 (1990).
3. Bian, J., Cao, Z. & Zhou, P. Neuromorphic computing: Devices, hardware, and system application facilitated by two-dimensional materials. *Appl. Phys. Rev.* **8**, 041313 (2021).
4. Wan, Q., Sharbati, M. T., Erickson, J. R., Du, Y. & Xiong, F. Emerging artificial synaptic devices for neuromorphic computing. *Adv. Mater. Technol.* https://doi.org/10.1002/admt.201900037 (2019).

5. Burr, G. W., Sebastian, A., Vianello, E., Waser, R. & Parkin, S. Emerging materials in neuromorphic computing: Guest editorial. *APL Mater.* **8**, 010401 (2020).
6. Seo, J.-S. *et al. IEEE Custom Integrated Circuits Conference* 1–4 (IEEE, 2011).
7. Kim, M.-K., Kim, I.-J. & Lee, J.-S. Oxide semiconductor-based ferroelectric thin-film transistors for advanced neuromorphic computing. *Appl. Phys. Lett.* **118**, 032902 (2021).
8. Lv, W. *et al.* Stochastic artificial synapses based on nanoscale magnetic tunnel junction for neuromorphic applications. *Appl. Phys. Lett.* **121**, 232406 (2022).
9. Burr, G. W. *et al.*.Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. IEEE International Electron Devices Meeting. 29.5.1-29.5.4 (2014).
10. Eryilmaz, S. B., Kuzum, D., Jeyasingh, R. G. D., Kim, S., BrightSky, M., Lam, C., Wong, H.-S. P., *IEEE International Electron Devices Meeting*, 25.5.1–25.5.4, (2013).
11. Bichler, O. *et al.* Visual pattern extraction using energy-efficient "2-PCM Synapse" neuromorphic architecture. *IEEE Trans. Electron Devices* **59**(8), 2206–2214 (2012).
12. Woo, J. & Yu, S. Resistive memory-based analog synapse: The pursuit for linear and symmetric weight update. *IEEE Nanotechnol. Mag.* **12**(3), 36–44 (2018).
13. Woo, J. *et al.* Improved synaptic behavior under identical pulses using AlO_x/HfO_2 bilayer RRAM array for neuromorphic systems. *IEEE Electron Device Lett.* **37**(8), 994–997 (2016).
14. Cuppers, F. *et al.* Exploiting the switching dynamics of HfO_2-based ReRAM devices for reliable analog memristive behaviour. *APL Mater.* **7**, 091105 (2019).
15. Woo, J. *et al.* Optimized programming scheme enabling linear potentiation in filamentary HfO_2 RRAM synapse for neuromorphic systems. *IEEE Trans. Electron Devices* **63**(12), 5064–5067 (2016).
16. Zhao, M. *et al.* Endurance and retention degradation of intermediate levels in filamentary analog RRAM. *J. Electron Devices Soc.* **7**, 1239–1247 (2019).
17. Tang, J. *et al.* ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing. IEEE International Electron Devices Meeting. 13.1.1-13.1.4 (2018).
18. Tessler, N., Kim, N., Kang, H. & Woo, J. Switching mechanisms of CMOS-compatible ECRAM transistors—Electrolyte charging and ion plating. *J. Appl. Phys.* **134**, 074501 (2023).
19. Fuller, E. J. *et al.* Li-ion synaptic transistor for low power analog computing. Adv. Mater. **29**(4), 1–8 (2017).
20. Kim, S. *et al.* Metal-oxide based, CMOS compatible ECRAM for Deep Learning Accelerator. IEEE International Electron Devices Meeting. 35.7.1-35.7.4 (2019).
21. Lee, C., Choi, W., Kwak, M., Kim, S. & Hwang, H. Impact of electrolyte density on synaptic characteristics of oxygen-based ionic synaptic transistor. *Appl. Phys. Lett.* **119**, 103503 (2021).
22. Yao, X. *et al.* Protonic solid-state electrochemical synapse for physical neural networks. *Nat. Commun.* **11**, 3134 (2020).
23. Cui, J. *et al.* CMOS-compatible electrochemical synaptic transistor arrays for deep learning accelerators. *Nat. Electron.* **6**, 292–300 (2023).
24. 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).
25. 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).
26. 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).
27. Biesinger, M. C., Lau, L. W. M., Gerson, A. R. & Smart, R. S. C. Resolving surface chemical states in XPS analysis of first row transition metals, oxides and hydroxides: Sc, Ti, V, Cu and Zn. *Appl. Surf. Sci.* **257**(3), 887–898 (2010).
28. Luo, X. *et al.* Investigation of HfO_2 thin films on Si by X-ray photoelectron spectroscopy, rutherford backscattering, grazing incidence X-ray diffraction and variable angle spectroscopic ellipsometry. *Crystals* **8**(6), 248 (2018).
29. Jang, J.-W., Park, S., Burr, G. W., Hwang, H. & Jeong, Y.-H. Optimization of conductance change in $Pr_{1-x}Ca_xMnO_3$-based synaptic devices for neuromorphic systems. *IEEE Electron Device Lett.* **36**(5), 457–459 (2015).
30. Rasch, M. J., Moreda, D., Gokmen, T., Gallo, M. L., Carta, F., Goldberg, C., Maghraoui, K. E., Sebastian, A., Narayanan, V. *IEEE International Conference on Artificial Intelligence Circuits and Systems*. 1–4, (2021).
31. Kumar, A., Mondal, S. & Rao, K. S. R. K. Structural, electrical, band alignment and charge trapping analysis of nitrogen-annealed Pt/HfO_2/p-Si (100) MIS devices. *Appl. Phys. A* **122**(12), 1027 (2016).

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 supported by the National R&D Program through the NRF funded by Ministry of Science and ICT (2020M3H2A1078045). This work was supported in part by the Technology Innovation Program (or Industrial Strategic Technology Development Program) (RS-2023-00236568) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea) (1415187361), and in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-RS-2022-00156225) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Author contributions

S.J., performed the experiments and characterized the devices. S.J., N.T., N.K., E.H., H.W.K., and J.W., discussed the results. N.T. and J.W. supervised the study. S.J., N.T., 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.