PAPER • OPEN ACCESS

Ultrathin HfO$_2$/Al$_2$O$_3$ bilayer based reliable 1T1R RRAM electronic synapses with low power consumption for neuromorphic computing

View the article online for updates and enhancements.

You may also like

NEUROMORPHIC
Computing and Engineering

PAPER

Ultrathin HfO$_2$/Al$_2$O$_3$ bilayer based reliable 1T1R RRAM electronic synapses with low power consumption for neuromorphic computing

Qiang Wang[1,7] ⓘ, Yankun Wang[1,7] ⓘ, Ren Luo[1], Jianjian Wang[2,3], Lanlong Ji[2,3], Zhuangde Jiang[4], Christian Wenger[5], Zhitang Song[6], Sannian Song[6], Wei Ren[1,*], Jinshun Bi[2,3,*] and Gang Niu[1,*] ⓘ

1. Electronic Materials Research Laboratory, Key Laboratory of the Ministry of Education & International Center for Dielectric Research, School of Electronic Science and Engineering & The International Joint Laboratory for Micro/Nano Manufacturing and Measurement Technology, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China
2. Key Laboratory of Microelectronics Device and Integrated Technology, The Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, People's Republic of China
3. University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
4. The State Key Laboratory for Manufacturing Systems Engineering & The International Joint Laboratory for Micro/Nano Manufacturing and Measurement Technology, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China
5. IHP-Leibniz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt, Germany
6. State Key Laboratory of Functional Materials for Informatics, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, 865 Changning Road, Shanghai 200050, People's Republic of China
7. These authors contributed equally for this work.
* Authors to whom any correspondence should be addressed.

E-mail: wren@mail.xjtu.edu.cn, bijinshun@ime.ac.cn and gangniu@xjtu.edu.cn

Abstract

Neuromorphic computing requires highly reliable and low power consumption electronic synapses. Complementary-metal-oxide-semiconductor (CMOS) compatible HfO$_2$ based memristors are a strong candidate despite of challenges like non-optimized material engineering and device structures. We report here CMOS integrated 1-transistor-1-resistor (1T1R) electronic synapses with ultrathin HfO$_2$/Al$_2$O$_3$ bilayer stacks (<5.5 nm) with high-performances. The layer thicknesses were optimized using statistically extensive electrical studies and the optimized HfO$_2$(3 nm)/ Al$_2$O$_3$(1.5 nm) sample shows the high reliability of 600 DC cycles, the low Set voltage of ∼0.15 V and the low operation current of ∼6 μA. Electron transport mechanisms under cycling operation of single-layer HfO$_2$ and bilayer HfO$_2$/Al$_2$O$_3$ samples were compared, and it turned out that the inserted thin Al$_2$O$_3$ layer results in stable ionic conduction. Compared to the single layer HfO$_2$ stack with almost the same thickness, the superiorities of HfO$_2$/Al$_2$O$_3$ 1T1R resistive random access memory (RRAM) devices in electronic synapse were thoroughly clarified, such as better DC analog switching and continuous conductance distribution in a larger regulated range (0–700 μS). Using the proposed bilayer HfO$_2$/Al$_2$O$_3$ devices, a recognition accuracy of 95.6% of MNIST dataset was achieved. These results highlight the promising role of the ultrathin HfO$_2$/Al$_2$O$_3$ bilayer RRAM devices in the application of high-performance neuromorphic computing.

1. Introduction

Neuromorphic computing aims to construct a hardware platform to fully achieve the function of biological neural networks. Human brains are composed by ∼10^{11} neurons ∼10^{14} synapses, which interact with each other to realize the cognition function like learning and memory. Therefore, the mimic of neurons and synapses using electronic devices is of significance for the neuro-inspired computing. Memristors are considered as the natural artificial synapse devices due to the similarity between the conductance modulation and the weight tuning via the regulation of ions conduction. Single artificial synapses based on memristor

devices have been intensively studied and both long-term and short-term synaptic plasticity have been achieved, which strongly contributes to the cognition function. Complementary-metal-oxide-semiconductor (CMOS) compatible large scale integrated artificial synapses are still under investigation and require suitable technologies. HfO$_2$ based resistive random access memories (RRAM) become strong candidates for such applications thanks their virtues like high CMOS compatibility, low energy consumption, analog switching with multistate regulation and low cost etc [1, 2]. Different from universal memories using two resistance states, RRAM for artificial synapses require resistive switching with continuous, symmetric, reliable and uniform multilevel resistance regulation, which thus need the large On/Off ratio, analog RS, high endurance and retention of the RRAM device. Current dominating RRAMs using the single HfO$_2$ layers can hardly satisfy these requests, which leads to extensive efforts to engineer both the materials and device structure. The formation and the rupture of conductive filaments (CFs) consist with oxygen vacancies (V$_O^{..}$) is the widely accepted mechanism for the resistive switching (RS) of HfO$_2$ based RRAM. One approach to optimize to improve analog RS is doping. Ti [3], Al [4, 5], Mg [6], Ni [7], Au [8], S [9] have been introduced as dopants into HfO$_2$ layers to tailor the formation of CFs, resulting in improved RS performance and gradual switching. On the other hand, the metal/oxide interface is significant for the formation and dynamics of V$_O$ and thus many efforts have been carried out to engineer the interfaces to optimize the multilevel RS and the reliability. For example, Ti [10], TiO$_x$ buffer layers [11], plasma-enhanced [12–14] and electrode treatments [14, 15] have been demonstrated to be able to engineer interfacial defects to improve the device reliability and analog switching. Moreover, recently it is reported that bilayer or multilayer HfO$_2$ stacks using AlO$_x$ [16–19], TiO$_x$ [20, 21], ZrO$_x$ [22], ZnO [15] insert layers to form hetero metal-oxide interfaces show excellent performances for multilevel RS, thanks to the possible control and confinement of CFs at interface in such heterostructures. Among them, HfO$_2$/Al$_2$O$_3$ bilayer turns out to be the most interesting candidate for the application of neuromorphic hardware due to the compatibility with industrial mass fabrication. Azzaz *et al* [23] reported 16k bit HfO$_2$(5 nm)/Al$_2$O$_3$(1 nm) based memory in 2015. Woo *et al* [24], verified HfO$_2$/AlOx bilayer RRAM arrays showing the achievement of multiple conductance states and resulting in high accuracy in the neuromorphic task of pattern recognition. In these demonstrations, the 1-transistor-1-resistor (1T1R) scheme based on CMOS platform was widely adopted for the fabrication of arrays, because it enables resistors to address the crosstalk issue of arrays and independently perform a linear current–voltage relationship with the transistor gate control in such arrays, realizing the simulation of analogue multiplication for neuromorphic applications [25, 26]. Despite the reported promising results based on HfO$_2$/Al$_2$O$_3$ bilayer RRAM arrays, the study of such electronic synapses is still at its early stage. The understanding of the impact of the film stacking structure needs to be improved and the film thickness requires further optimized to achieve better performances of the electronic synapses. Moreover, the request of the reduction of operating energy consumption of devices possibly asks for ultrathin oxide layers, which also strongly impact the multilevel RS and reliability of RRAM synapses.

In this work, we report ultrathin HfO$_2$/Al$_2$O$_3$ bilayer (<5.5 nm) based reliable integrated 1-transistor-1-resistor (1T1R) RRAM electronic synapses with high performance. The 1T1R RRAM array was fabricated using 0.18 μm back-end-of-line CMOS process. The thickness of the Al$_2$O$_3$ layer was set to be 1.5 nm and the thickness of HfO$_2$ layer varies from 2 nm to 4 nm. Multilevel RS properties of the bilayer 1T1R RRAM array with different film thicknesses have been in-depth studied in a statistical manner. The conduction mechanism of the devices has been investigated in detail by analyzing the current–voltage (I–V) curves and the advantages of bilayer structure have been clarified, compared to the single-layer devices. Using the optimized devices and a three-layer full-connected neural network, we demonstrate a recognition rate of MNIST dataset of 95.6%, which is much higher than that achieved using conventional single-layer HfO$_2$ devices. Our results highlight the significance of materials engineering for the improvement of electronic synapses and pave the way to eventually achieve high-performance neuromorphic computing using 1T1R HfO$_2$/Al$_2$O$_3$ RRAM arrays.

2. Experimental section

All integrated 1T1R RRAM arrays were fabricated in a 0.18 μm CMOS processing line (8 inch) at Institute of Microelectronics, Chinese Academy of Sciences (IMECAS). Detailed preparation process was illustrated in figure S1(a) in the supporting materials. As resistors, metal–insulator-metal (MIM) stacks, with the same thickness of ∼40 nm, were fabricated on Drain port of transistor. As a reference device, TiN/Ti/HfO$_2$(5 nm)/W was firstly fabricated. In order to optimize the HfO$_2$ thickness and investigate its impact on the RS property, the thickness of the Al$_2$O$_3$ layer was fixed to be 1.5 nm, a series of TiN/Ti/HfO$_2$/Al$_2$O$_3$/W bilayer RRAM arrays with the HfO$_2$ thickness of 2 nm, 3 nm and 4 nm were fabricated. Detailed information of all fabricated samples was listed in table 1. Both Al$_2$O$_3$ and HfO$_2$ layer were deposited by atomic layer deposition. TiN/Ti as the top electrode was deposited by physical vapor

Table 1. Structures of all fabricated samples.

Samples	Stacks
H5	TiN/Ti/HfO$_2$(5 nm)/W
H2A1.5	TiN/Ti/HfO$_2$(2 nm)/Al$_2$O$_3$(1.5 nm)/W
H3A1.5	TiN/Ti/HfO$_2$(3 nm)/Al$_2$O$_3$(1.5 nm)/W
H4A1.5	TiN/Ti/HfO$_2$(4 nm)/Al$_2$O$_3$(1.5 nm)/W

deposition. The photo of the 8 inch wafer scale RRAM arrays was shown in the figure S1(b) in the supporting materials (SM).
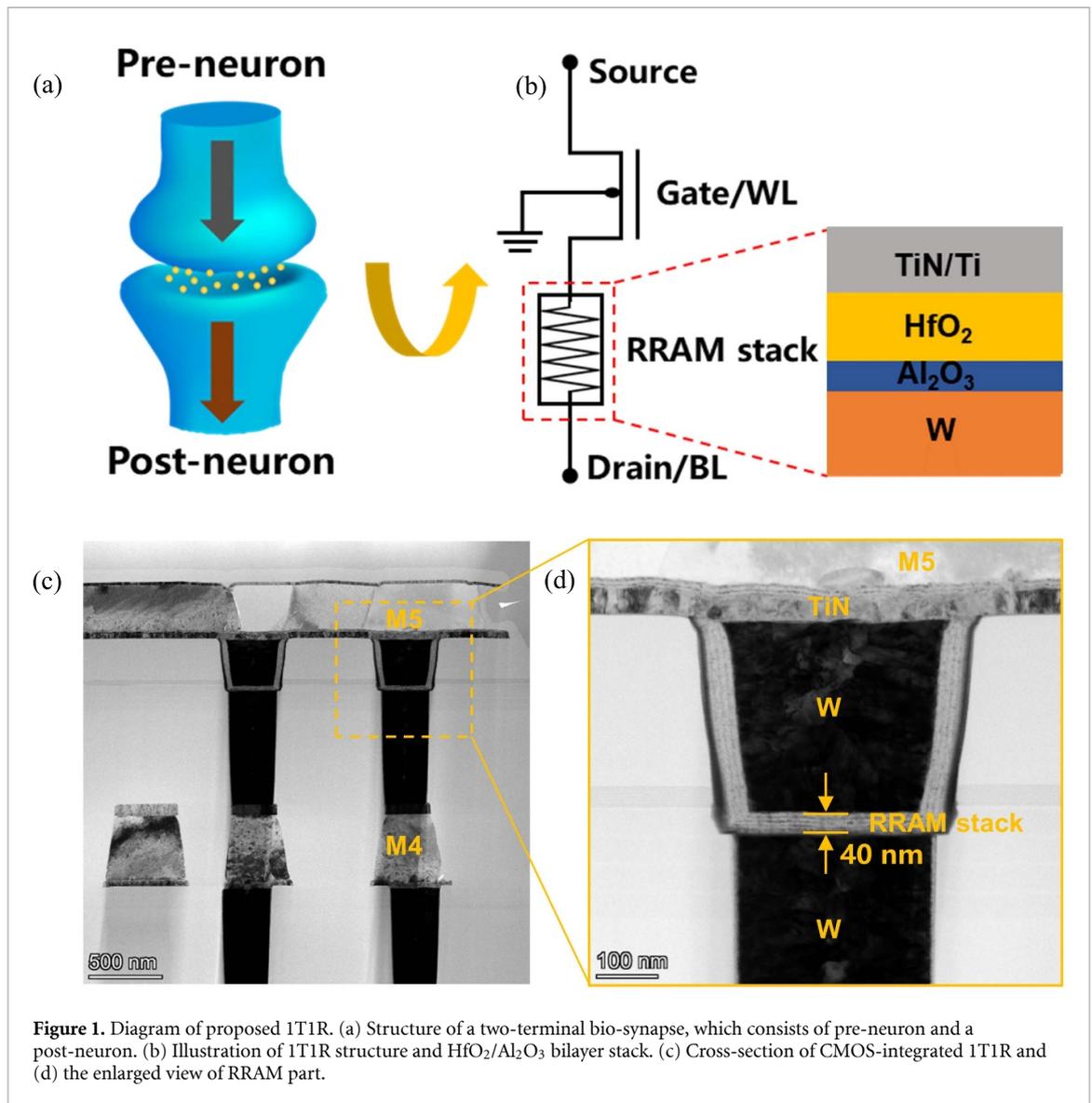
The transmission electron microscopy (TEM) and the energy dispersive x-ray spectroscopy (EDX) characterization were performed with the acceleration voltage of 300 kV. All electrical measurements under the direct current (DC) mode and the pulse mode were conducted in a high shielding chamber, using a Keithley 4200 A semiconductor parameter analyzer that containing three source measure unit and one ultra-fast pulse measure unit to generate arbitrary pulses. The *I–V* characterizations of each sample were performed and summarized from at least five devices. Simulation of MNIST recognition task was performed on graphics processing unit by python.

3. Results and discussion

The device structure of the 1T1R RRAM electronic synapses was shown in figure 1. Figure 1(a) illustrates a two-terminal bio-synapse structure, which is composed by a pre-neuron and a post-neuron. Such structure can be well mimicked using the 1T1R RRAM structure between the Gate and Drain, as shown in figure 1(b). Meanwhile, the behavior of synaptic weight update can be realized by regulating the conductance of 1T1R RRAM devices. As shown in figure 1(b), the TiN/Ti/HfO$_2$/Al$_2$O$_3$/W bilayer stack as a resistor was fabricated in the Drain port of 1T1R structure. The cross-sectional TEM image of the bilayer stack was shown in figure 1(c) and its enlarging image was displayed in figure 1(d). The whole structure was fabricated in the Via between metal 4 (M4, aluminum) and metal 5 (M5, aluminum) using the 0.18 μm CMOS processing line. The thickness of the RRAM stack is ~40 nm. The dark part is tungsten (W) plug. In addition, the TEM-EDX results of TiN/Ti/HfO$_2$/W stack were presented in figure S2 in the SM, suggesting the as-expected well-defined element distribution.

The RS properties of all four samples were statistically studied and typical *I–V* curves of H5, H2A1.5, H3A1.5 and H4A1.5 were selected after 50 cycles and shown in figure 2. It is noted here that the output current character (I_{DS}) of the transistor, as the current compliance function, was used to decide Gate voltage (V_G) and Drain/Source voltage (V_{DS}), as shown in figure S3 in SM. Figure S3 shows that I_{DS} can be reliably regulated within 1 mA when V_G increases from 1.0 V to 3.0 V. All subsequent electrical characteristics were measured using the same parameters, i.e. $V_G = 1.0$ V, $V_S = 0$ V and V_D being positive bias for the Forming operation; $V_G = 1.5$ V, $V_S = 0$ V and V_D being positive bias for Set operation; $V_G = 2.5$ V, $V_D = 0$ V and V_S being positive bias for Reset operation. Figure 2(a) indicates that H5 has an abrupt switching during Set process while all bilayer samples exhibit more gradual characteristic. Moreover, such gradual switching is gradually unobvious as the thickness of HfO$_2$ layer in bilayer stacks increases from 2 nm to 4 nm. The H2A1.5 sample shows the quite gradual (as shown in figure 2(c)) and the H4A1.5 sample (as shown in figure 2(g)) reproduces the slight abrupt Set process. Figures 2(d), (f) and (h) indicate that bilayer samples have a higher I_{DS} during Reset process as compared to H5 sample (as shown in figure 2(b)).

In order to understand the RS properties of different samples, the Forming, Set and Reset voltages of samples shown in figure 2 were analyzed and summarized in figure 3. The forming process was carried out on each sample and detailed *I–V* characteristics can be found in figure S4 in SM. The variation of the Forming voltage (V_F) as a function of different samples was plotted in figure 3(a), the data of which were collected from ten devices for each sample. It can be evidently observed that the V_F firstly decreases from ~2.3 V of H5 single layer sample to ~1.8 V of H2A1.5 and then gradually increases to ~2.2 V of H3A1.5 and ~2.9 V of H4A1.5. This suggests that the V_F is almost proportional to the whole dielectric layer thickness. Then, each device was characterized with 50 DC cycles and the original data can be found in the figures S5 and S6 in SM. The Set and Reset voltages of samples were summarized in figures 3(b) and (c), respectively. It can be seen in figure 3(b) that all bilayer samples have much lower Set voltage of <0.15 V and higher uniformity (smaller box height), compared to those of the single-layer H5 sample. Corresponding to the Set voltage, statistical operation currents were presented in figure S7 in the SM, suggesting the ultralow average operation current <7 μA of all samples. Particularly, H3A1.5 sample has a lowest average operation current of <6 μA, which is a very low value in comparison with previously reported 1T1R RRAM devices, as shown

Figure 1. Diagram of proposed 1T1R. (a) Structure of a two-terminal bio-synapse, which consists of pre-neuron and a post-neuron. (b) Illustration of 1T1R structure and HfO_2/Al_2O_3 bilayer stack. (c) Cross-section of CMOS-integrated 1T1R and (d) the enlarged view of RRAM part.
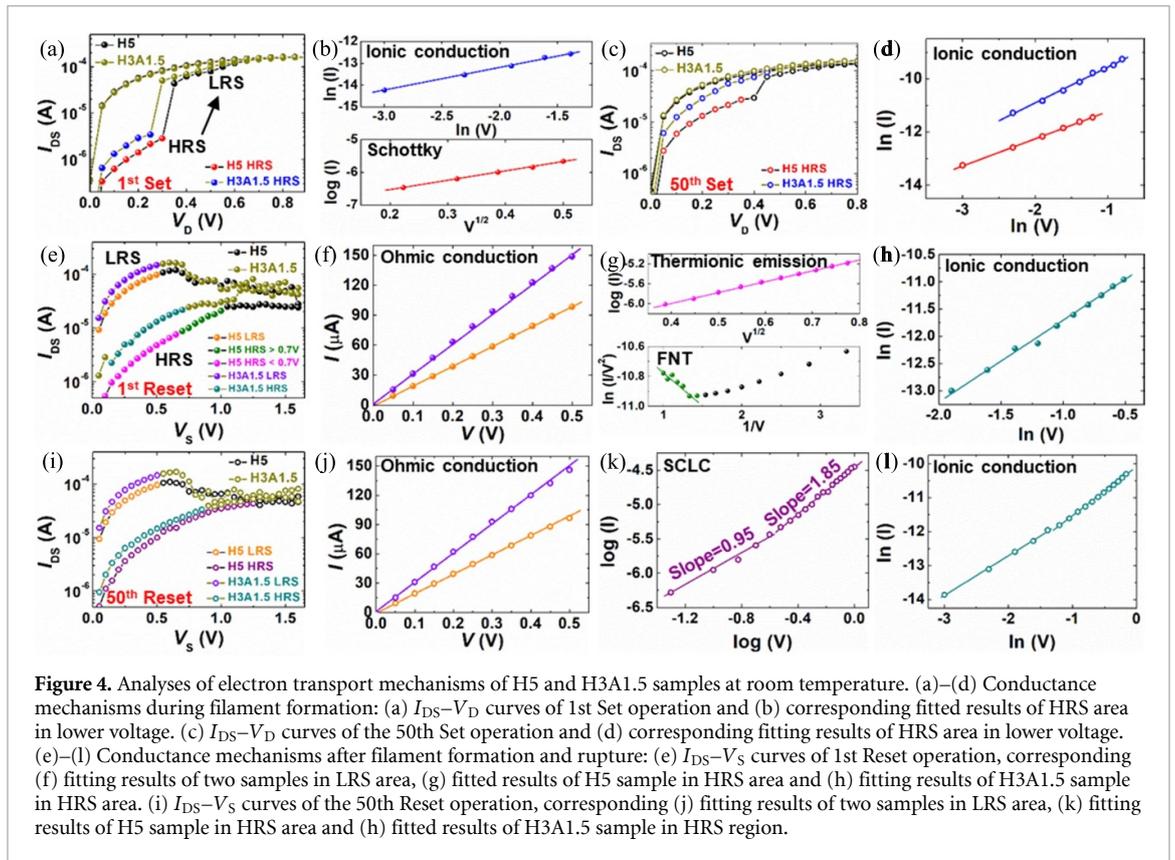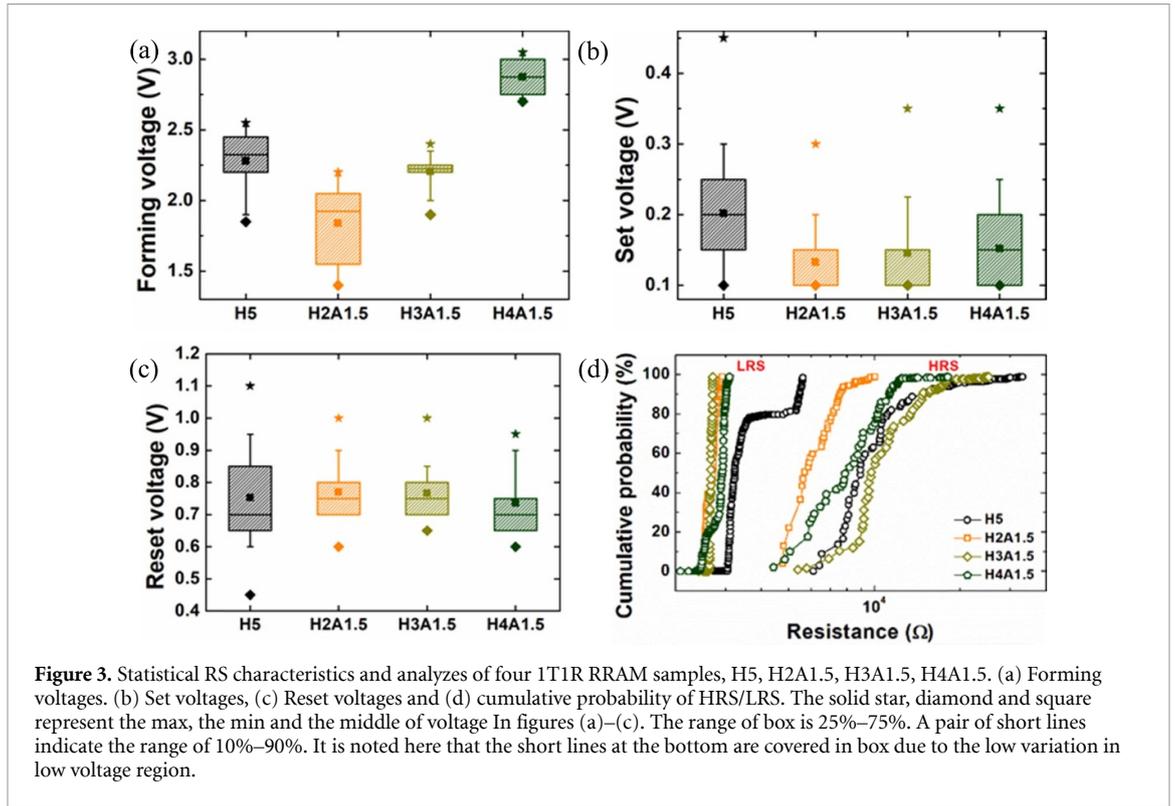
in the table S1 in SM. Figure 3(c) shows the variation of the Reset voltage for different samples. It reveals that the Reset voltage remains 0.7–0.8 V, which is independent from the bilayer stack structure. Nevertheless, the bilayer devices show smaller voltage boxes indicating the better uniformity compared to the single layer H5 sample. Figure 3(d) presents the cumulative probability of high resistance state (HRS) and low resistance state (LRS) of all samples, which indicates that all bilayer samples have lower LRS values compared to the single layer H5 sample whereas the HRS values have a thickness dependent characteristic. In particular, the H3A1.5 sample has the highest HRS value resulting in the largest On/Off ratio, i.e. >11 up to 600 cycles. Figure S8 shows more details of typical characteristics of H5 and H3A1.5 samples, which suggest that H3A1.5 bilayer samples have enhanced RS properties including On/Off ratio and reliability, which probably originates from the Al_2O_3 layer between bottom electrode and HfO_2, serving as a O-scavenging barrier thus stabilizing the CFs [16, 18].

To get insight to the possible conductance mechanism of the enhanced RS performance of the H3A1.5 bilayer sample, the I–V curves (the first and the 50th cycle) of both H5 and H3A1.5 samples were fitted and the electron transport mechanism was analyzed. As shown in figure 4(a) of the I_{DS}–V_D curve, the devices firstly undergo the states from HRS to LRS. Figure 4(b) shows the corresponding fitting results of the HRS of H5 (red, bottom) and H3A1.5 (blue, top) samples, respectively. It can be observed that the H5 sample and the H3A1.5 samples have linear characteristic for the $\log(I) - V^{1/2}$ fitting and the $\ln(I) - \ln(V)$ fitting, suggesting the Schottky emission and the ionic conduction mechanism, respectively. After 50 RS cycles, the Set I_{DS}–V_D curves of two samples were presented in figure 4(c), which shows that the On/Off ratios decrease for both samples, mainly due to the augmentation of the HRS values. Interestingly, the I–V curve of the H3A1.5 sample (yellow) displays a gradual Set characteristic, which is possibly because of the difference

Figure 2. Typical *I–V* curves of (a), (b) H5, (c), (d) H2A1.5, (e), (f) H3A1.5 and (g), (h) H4A1.5 samples.
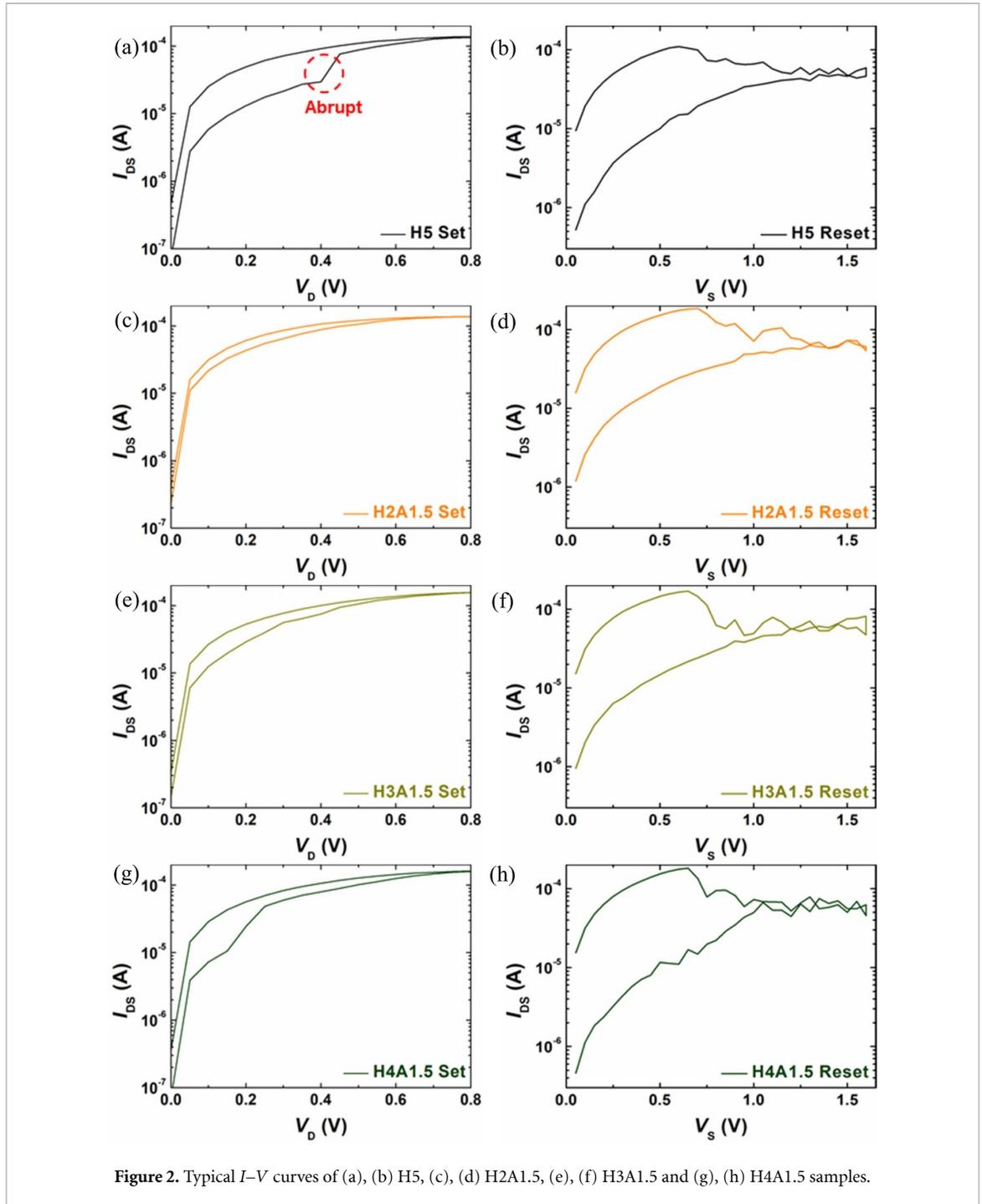
mobility rate of $V_O^{\cdot\cdot}$ in HfO$_2$ and Al$_2$O$_3$ and the HfO$_2$/Al$_2$O$_3$ interface [24]. The oxide interface and the different mobility of $V_O^{\cdot\cdot}$ in Al$_2$O$_3$ layer possibly change the formation and/or the rupture of the CFs and thus modify the Set process from an abrupt one of H5 sample to a gradual one of the H3A1.5 sample [18, 24]. Moreover, fitting results of HRS state of 50th Set I_{DS}–V_D curves in figure 4(d) show that both samples show a mechanism of ionic conduction.

The 1st cycle of the Reset processes for both H5 and H3A1.5 samples shown in figures 4(e) and (f)–(h) show the corresponding fitting results at different regions. Figure 4(f) presents the linear *I–V* fitting for LRS area, indicating the ohmic conduction for both samples, which is in consistent with the CF related HfO$_2$-based RRAM. Figure 4(g) shows the fitting of 1st Reset of H5 sample in HRS. The linear $\log(I) - V^{1/2}$ (top panel) in the area below 0.7 V might suggest thermionic emission while the linear $\ln(I/V^2) - 1/V$ (bottom panel) with the negative slope in the area >0.7 V represents a Fowler-Nordheim tunneling (FNT) mechanism. Differently, the bilayer H3A1.5 sample displays a linear $\ln(I) - \ln(V)$ fitting, as shown in figure 4(h), indicating an ionic conduction mechanism. After 50 cycles, the 50th Reset processes of both

Figure 3. Statistical RS characteristics and analyzes of four 1T1R RRAM samples, H5, H2A1.5, H3A1.5, H4A1.5. (a) Forming voltages. (b) Set voltages, (c) Reset voltages and (d) cumulative probability of HRS/LRS. The solid star, diamond and square represent the max, the min and the middle of voltage In figures (a)–(c). The range of box is 25%–75%. A pair of short lines indicate the range of 10%–90%. It is noted here that the short lines at the bottom are covered in box due to the low variation in low voltage region.

Figure 4. Analyses of electron transport mechanisms of H5 and H3A1.5 samples at room temperature. (a)–(d) Conductance mechanisms during filament formation: (a) I_{DS}–V_D curves of 1st Set operation and (b) corresponding fitted results of HRS area in lower voltage. (c) I_{DS}–V_D curves of the 50th Set operation and (d) corresponding fitting results of HRS area in lower voltage. (e)–(l) Conductance mechanisms after filament formation and rupture: (e) I_{DS}–V_S curves of 1st Reset operation, corresponding (f) fitting results of two samples in LRS area, (g) fitted results of H5 sample in HRS area and (h) fitting results of H3A1.5 sample in HRS area. (i) I_{DS}–V_S curves of the 50th Reset operation, corresponding (j) fitting results of two samples in LRS area, (k) fitting results of H5 sample in HRS area and (h) fitted results of H3A1.5 sample in HRS region.

samples were shown in figure 4(i). Figure 4(j) shows that the LRS regions still have the same ohmic conduction for both samples. However, the HRS of H5 sample shows a linear relationship in $\log(I)$–$\log(V)$ curve with the slopes of 0.95 and 1.85 (figure 3(k)), corresponding to a space-charge-limited-conduction (SCLC) mechanism [27, 28], which is different from the 1st Reset cycle. The HRS of H3A1.5 sample remains the ionic conduction mechanism, as shown in figure 4(l). All formulas of conduction mechanisms were
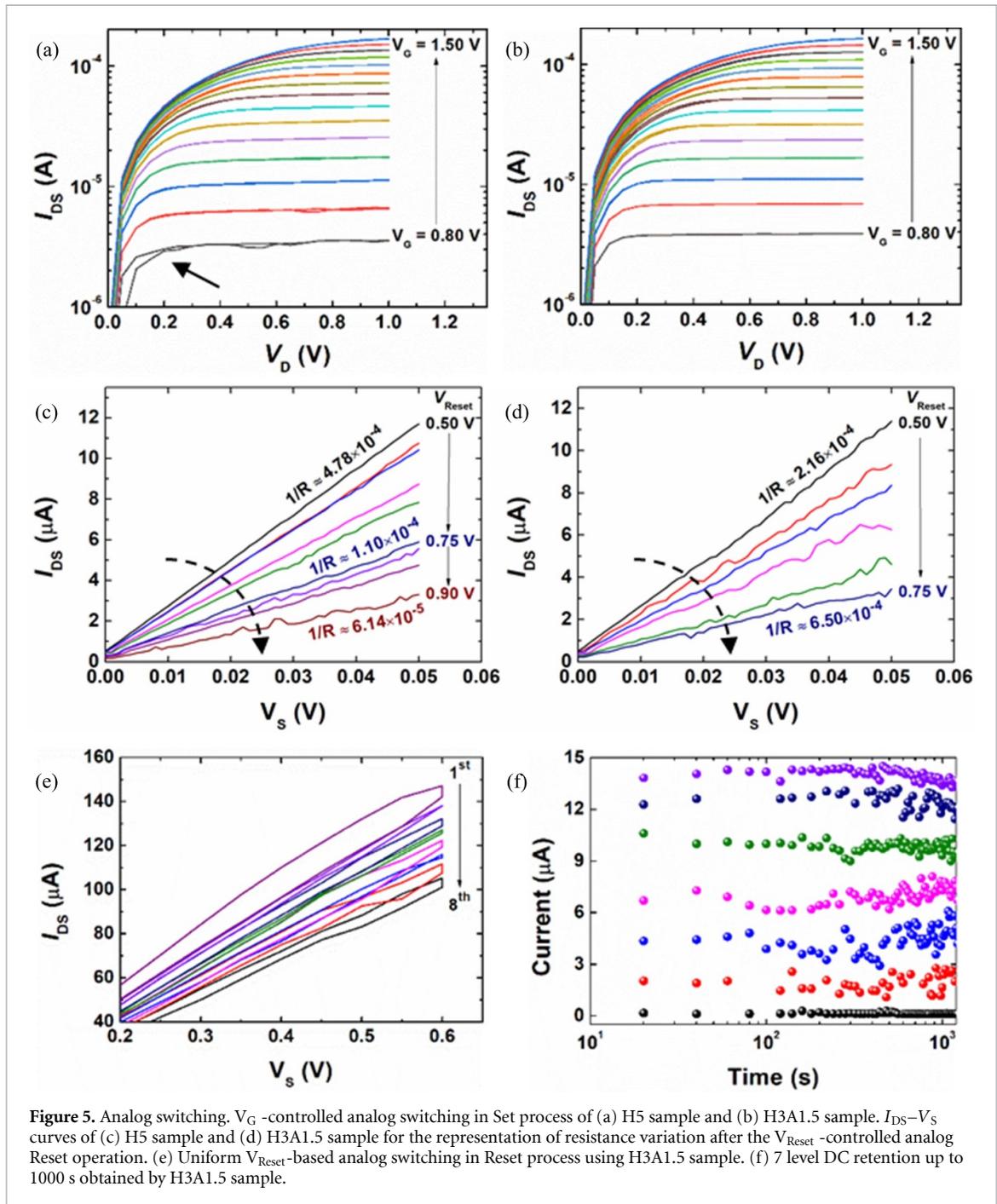
Table 2. Expressions of typical conduction mechanism in proposed 1T1R structures.

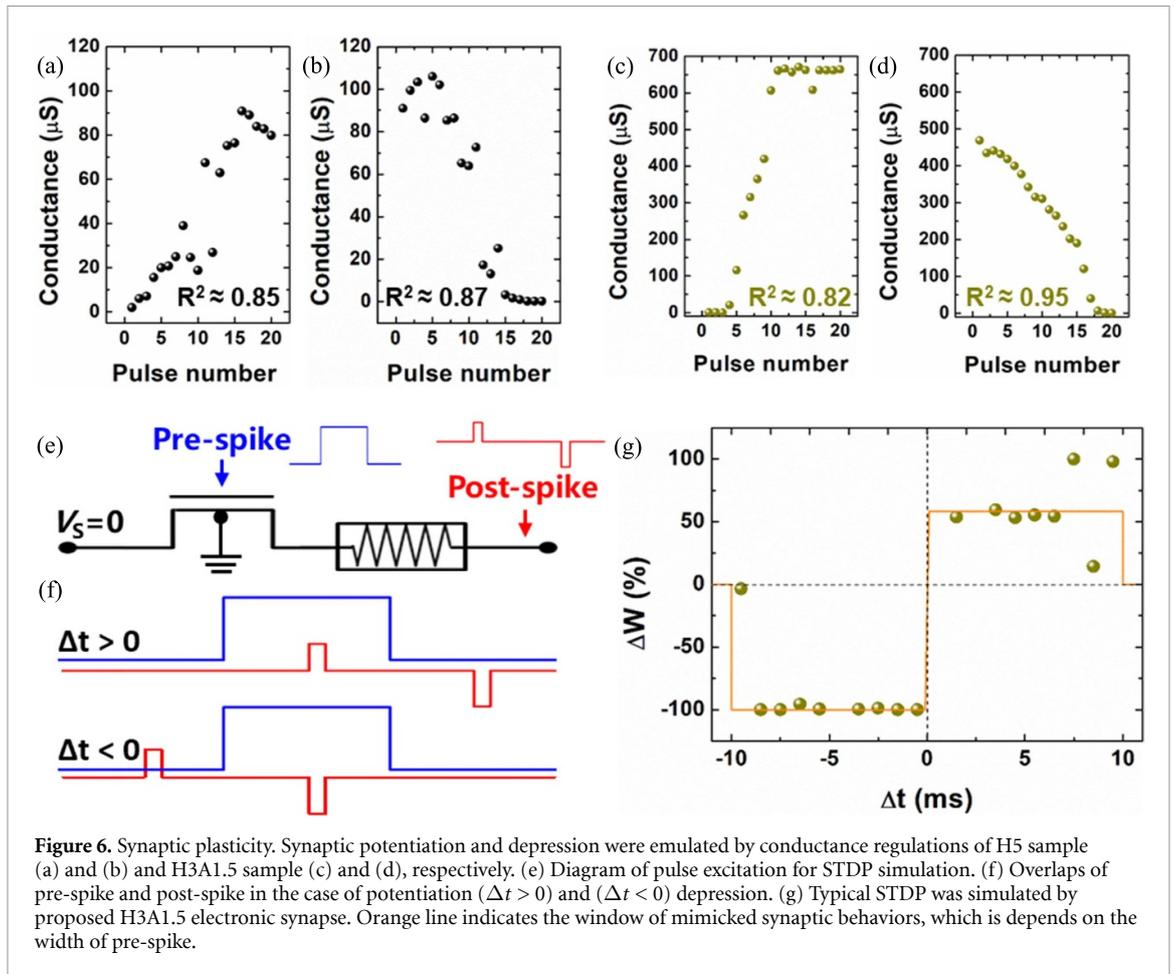| Conduction mechanism	Expressions	Relation between J and E
Schottky/Thermionic Emission	$J_{\text{Sch or TE}}(E) =$ $A^* T^2 \exp\left[-\frac{q}{kT}\left(\varphi_b - \sqrt{\frac{qE}{4\pi\varepsilon_0\varepsilon_r}}\right)\right]$	$\text{Log}\,(J_{\text{Sch or TE}}) \propto E^{1/2}$
Ionic Conduction	$J_{\text{Ionic}}(E) = J_0 \exp\left[-\left(\frac{q\varphi_b}{kT} - \frac{qEd}{2kT}\right)\right]$	$Ln\,(J_{\text{Ionic}}) \propto \ln(E)$
Ohmic Conduction	$J_{\text{Ohm}}(E) = nq\mu E,$ $n = N_c \exp\left[\frac{-(E_C - E_F)}{kT}\right]$	$J_{\text{Ohm}} \propto E$
Fowler-Nordheim Tunneling (FNT)	$J_{\text{FNT}}(E) =$ $\frac{q^3 E^2}{8\pi h q \varphi_b} \exp\left[\frac{-8\pi\sqrt{2qm^*}}{3hE}\varphi_b^{3/2}\right]$	$Ln\,(J_{\text{FNT}}/E^2) \propto 1/E$
Space-charge-limited Conduction (SCLC)	$J_{\text{ohm}}(E) = n_0 q\mu E,\ J_{\text{Child}}(E) = \frac{9}{8}\mu\varepsilon\frac{V^2}{d^3}$	(1) $J_{\text{Ohm}} \propto E$, (2) $J_{\text{Child}} \propto E^2$

shown in table 2, where J is the current density, (For ionic conduction, J is linearly related to electric field E, in the case of lowelectric fields), A^* is the effective Richardson constant, T is the absolute temperature, q is the electronic charge, $q\phi_b$ is the Schottky barrier height (i.e. conduction band offset), E is the electric field across the dielectric, K is the Boltzmann's constant, ε_0 is the permittivity in vacuum, ε_r is the optical dielectric constant, h is the Planck's constant, $m*$ is the effective electron mass in dielectric, n_O is the concentration of the free charge carriers in thermal equilibrium, V is the applied voltage, d is the thickness of thin films, ε is the static dielectric constant [27, 28].

It is noted here that, according to the results shown in figure 4, single layer HfO$_2$ based RRAM devices experiences the complex transformation of electron transport during the RS cycles. The Schottky mechanism changes to ionic conduction mechanism for HRS during the CF formation, and FNT/Thermionic emission changes to SCLC for HRS after CF rupture. These facts may lead to the instability of the devices during cycling. On the contrary, the HfO$_2$/Al$_2$O$_3$ bilayer RRAM devices present the ionic conduction mechanism for HRS during both CF formation and rupture, which could be an important reason to explain better device stability. In addition, the HfO$_2$/Al$_2$O$_3$ bilayer stack presents a gradual switching characteristic which is favorable for the realization of the analog switching for the implementation of electronic synapse.

In order to investigate the capability of both H5 and H3A1.5 samples to achieve analog RS, the multilevel RS characteristics of both samples were examined and the results were shown in figure 5. For bipolar memristors, analog switching can be achieved using both Set process and Reset process. After a complete Reset operation, analog Set process was obtained by finely controlling V_G from 0.8 V to 1.5 V, with a step of 0.05 V for both H5 (figure 5(a)) and H3 A1.5 samples (figure 5(b)). Two samples show quite similar analog I_{DS}–V_D curves but there is an obvious abrupt changing of the I_{DS} at $V_G = 0.8$ V (black curve) of the H5 sample (figure 5(a)). It is induced by the abrupt resistance state changing during the Set process of the H5 sample. On the contrary, after a complete Set operation, analog Reset process was achieved by increasing the Reset voltage (V_{Reset}) from the initial value of 0.5 V, with a step of 0.05 V and the V_G being fixed to 2.5 V. More details of the original analog Reset curves of H5 and H3A1.5 samples can be found in in figures S9(a) and (b), respectively in SM. Every Reset operation was followed with a Read operation. Corresponding I_{DS}-V_S curves were then obtained, as displayed in figures 5(c) and (d) for H5 and H3A1.5 samples, respectively. The slope of I_{DS}–V_S curve represents $1/R$. For H5 as shown in figure 5(c), the value of $1/R$ decreased from 4.78×10^{-4} (i.e. $R = 2090\ \Omega$) to 6.14×10^{-4} (i.e. $R = 16\,300\ \Omega$) as the V_{Reset} increases from 0.5 V to 0.9 V with a step of 0.05 V. However, these I_{DS}-V_S curves were not continuous and even overlapped. This is due to the abrupt formation of CF in single layer HfO$_2$ RRAM devices. On the contrary, I_{DS}-V_S curves of H3A1.5 (figure 5(d)) shows obvious well-proportioned distribution with the V_{Reset} increasing from 0.5 V to 0.75 V with a step of 0.05 V. The corresponding $1/R$ values are from 2.16×10^{-4} ($R = 4630\ \Omega$) to 6.50×10^{-4} ($R = 15\,400\ \Omega$), being very closed to the result obtained from H5 with the V_{Reset} ranging from 0.5 V to 0.9 V. Furthermore, within the limited range of 0.5–0.75 V, H5 sample only achieves a $1/R$ range from 4.78×10^{-4} to 1.10×10^{-4} which is obviously smaller than that of H3A1.5 sample. Moreover, as presented in figure 5(e), the 8-level analog switching was achieved using the H3A1.5 sample by continuously applying eight times identical 0.6 V DC sweep. Finally, the H3A1.5 sample was able to retain 7 level current retention up to 1000 s, as shown in figure 5(f), exhibiting a promising stability in multilevel RS. Such excellent analog switching characteristics of H3A1.5 sample compared to the single layer H5 sample can be attributed to the 1.5 nm thick Al$_2$O$_3$ layer. It is probable that the Al$_2$O$_3$ layer and the HfO$_2$/Al$_2$O$_3$ interface facilitates the constrain of CFs, which subsequently helps to achieve the analog switching and continuous conductance distribution [19, 24].

Figure 5. Analog switching. V_G-controlled analog switching in Set process of (a) H5 sample and (b) H3A1.5 sample. I_{DS}–V_S curves of (c) H5 sample and (d) H3A1.5 sample for the representation of resistance variation after the V_{Reset}-controlled analog Reset operation. (e) Uniform V_{Reset}-based analog switching in Reset process using H3A1.5 sample. (f) 7 level DC retention up to 1000 s obtained by H3A1.5 sample.
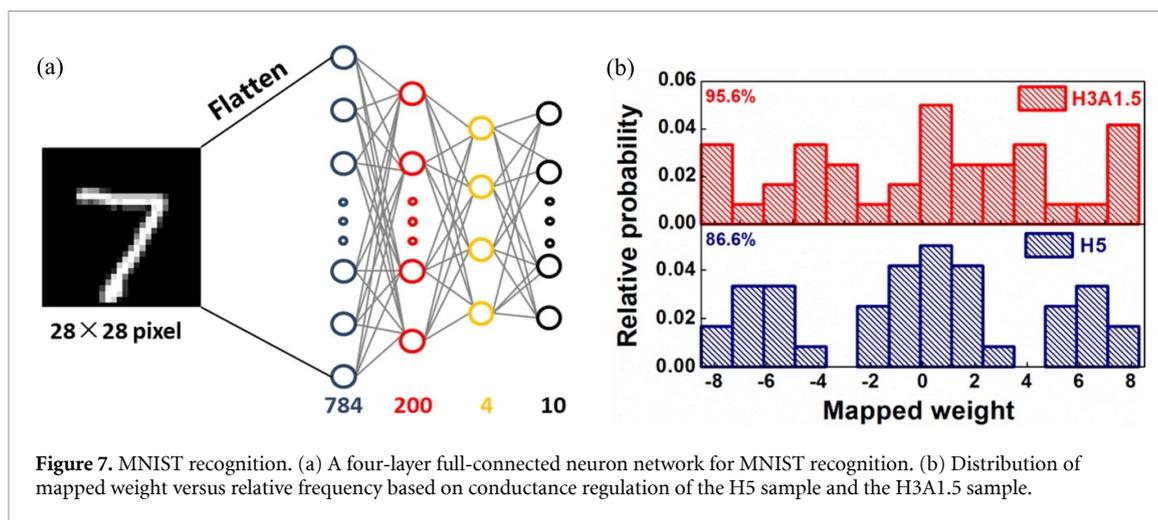
In order to examine the potential of single layer H5 and bilayer H3A1.5 samples for the application for electronic synapses, synaptic potentiation and synaptic depression, as the basic biological plasticity, were emulated by conductance regulation using H5 and H3A1.5 based 1T1R RRAM devices. Figure 6(a) shows the increase of conductance in the range of 0– 90 μS was obtained for synaptic potentiation, using identical 20 pulses with the amplitude of 0.6 V being applied to the Drain of H5 sample, with a fixed V_G of 1.5 V. The correlation coefficient (R^2) representing the degree of the linearity was calculated as 0.85 by the linear fitting. Figure 6(b) shows the decrease in the range of 110–0 μS for synaptic depression (calculated $R^2 \approx 0.87$) using identical 20 pulses with the amplitude of −0.8 V being applied to the Drain of H5 sample, with the fixed V_G of 2.5 V. Obviously, the sudden drop of conductance can be observed in range of 30–60 μS in both potentiation and depression. This is related to the inherent abrupt switching characteristics of the single layer HfO$_2$ sample. Using the same measurement configuration for the H3A1.5 sample, the conductance augmentation (calculated $R^2 \approx 0.82$) with a large range of 0–700 μS was achieved using the 0.2 V pulse (as shown in figure 6(c)). The conductance reduction with a large range of 500–0 μS and the higher R^2 of 0.95 was also achieved using the −0.8 V pulse (as shown in figure 6(d)). Compared to the H5 sample, the

Figure 6. Synaptic plasticity. Synaptic potentiation and depression were emulated by conductance regulations of H5 sample (a) and (b) and H3A1.5 sample (c) and (d), respectively. (e) Diagram of pulse excitation for STDP simulation. (f) Overlaps of pre-spike and post-spike in the case of potentiation ($\Delta t > 0$) and ($\Delta t < 0$) depression. (g) Typical STDP was simulated by proposed H3A1.5 electronic synapse. Orange line indicates the window of mimicked synaptic behaviors, which is depends on the width of pre-spike.

conductance regulation of H3A1.5 sample demonstrates advantages like the larger range of regulation, the better linearity and the better continuity, which are significant for high-performance simulations of synaptic plasticity and applications of neuromorphic computing. It is noted here that, during the augmentation, the conductance increases from 0 to $\sim$700 μS only experiencing serval pulses. This is because the applied bias of 0.2 V is slightly higher than the operation voltage of the H3A1.5 sample ($\sim$0.15 V).

We further study the realization of the STDP using the H3A1.5 sample and the results were shown in figure 6(e). It is known that STDP is a biological plasticity to describe the correlation between the synaptic weight and the time difference. As revealed in figure 6(e), the Gate port and the Drain port served as the pre-synapse and the post-synapse, being excited by the pre-spike and the post-spike, respectively. Two spikes have the same length and do not overlap at the beginning. The pre-spike is a long square pulse, with a voltage of 2 V and a width of 10 ms, triggering the synaptic event. The post-spike consists of two triggered pulses: a 0.6 V/100 μs pulse and a -0.8 V/100 μs pulse. Time difference between two spikes was noted as Δt. Change of synaptic weight (ΔW) was described by the percentage of incremental conductance, using the following formula: $100\% \times (G_i - G_0)/G_0$, $i = 1,2,3\ldots$. No behavior will be triggered only if the triggered pulse of the post-spike encounters with the pre-spike pulse. As shown in figure 6(f), when the pre-spike precedes the post-spike, $\Delta t > 0$, it is overlapped with the positive pulse of post-spike, resulting in the synaptic potentiation. Otherwise, $\Delta t < 0$, the pre-spike encounters with the negative pulse of post-spike, resulting in the synaptic depression. Eventually, the shape of mimicked STDP, ΔW versus Δt, was obtained by changing Δt and plotted in figure 6(g).

It is noted here that the STDP result is different from that of the biological synapse, which expresses the controllable change of weight as a function of the spike time difference, which normally can be mimicked using two-terminal 1 R RRAM devices [5, 14]. In this work, the memristor part of the three terminal 1T1R RRAM can only be regulated when the transistor is turned on. Thus, the pre-spike must be applied to the Gate of transistor when the proposed 1T1R RRAM was used to mimic two-terminal biological synapse. Under this working mechanism, only the post-spike that consists of two square pulse decides the emulation of synaptic weight. The amplitudes of two square pulses of the post-spike are enough to perform the full switching during Set and Reset, resulting in two distinct states. Meanwhile, the conductance of 1T1R RRAM

Figure 7. MNIST recognition. (a) A four-layer full-connected neuron network for MNIST recognition. (b) Distribution of mapped weight versus relative frequency based on conductance regulation of the H5 sample and the H3A1.5 sample.

is hardly progressive due to the long interval of time between adjacent STDP test. Thus, the shape of STDP with binary state shown in figure 6(g) was obtained.

The HfO_2/Al_2O_3 bilayer based electronic synapses were then used to fulfill the recognition task of MNIST database and the results were show in figure 7. As shown in figure 7(a), (a) four-layer full-connected neural network, with 784 inputs and 10 outputs, was constructed by Python. The 28×28 pixel images for the handwriting numbers from 0 to 9 was flatten in turn and input into the neural network. Conductance values of the potentiation and depression (in figures 6(a)–(d)) were mapped onto the optimized weight matrix of neural network between the third layer and the output layer. Then, 95.6% and 86.6% as recognition accuracy were computed after once interference based on H3A1.5 and H5 samples, respectively. Detailed simulation in software was can be found in the figure S10 in SM. Figure 7(b) shows the relative probability of weight distribution of the H3A1.5 sample (top panel) and the H5 sample (bottom panel), indicating the obvious uncontinuity in H5 sample as compared to the that of H3A1.5 sample.

4. Conclusion

In conclusion, ultrathin HfO_2/Al_2O_3 bilayer (<5.5 nm) based reliable 1T1R RRAM electronic synapses with low power consumption were successfully realized using 0.18 μm CMOS technology. The layer thickness of the bilayer stack was optimized to be HfO_2 (3 nm)/Al_2O_3 (1.5 nm) using a 5 nm single layer HfO_2 sample as a reference. The optimized H3A1.5 sample shows the high reliability of 600 DC cycles with On/Off ratio >11, the low Set voltage of ~0.15 V, the low operation current of ~6 μA and the excellent analog RS with 7-level states retention up to 1000 s. Detailed analyses of the electron transport mechanisms indicated that, differing from complex conduction mechanisms of single layer H5 sample, 1.5 nm Al_2O_3 inserted layer led to the stable ionic conduction mechanism in both HRS and LRS of the bilayer H3A1.5 sample, which possibly contributes to the improvement of device reliability. Moreover, the synaptic potentiation, the synaptic depression and STDP were achieved, suggesting the superiorities of bilayer HfO_2/Al_2O_3 1T1R RRAM devices for the realization of electronic synapses, particularly considering the continuous conduction regulation in 0–700 μS the as compared to the single layer H5 sample. High recognition accuracy of 95.6% for MNIST database was eventually obtained using the bilayer HfO_2/Al_2O_3 1T1R RRAM device. Our results provide significant guidance to achieve high-performance electronic synapses using ultrathin bilayer HfO_2/Al_2O_3 1T1R RRAM devices for the neuromorphic computing hardware.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

Conflict of interest

The authors declare no competing financial interest.

ORCID iDs

Qiang Wang https://orcid.org/0000-0002-6147-3747
Yankun Wang https://orcid.org/0000-0003-4090-1850
Gang Niu https://orcid.org/0000-0002-8813-8885

References

[1] Sokolov A S, Abbas H, Abbas Y and Choi C 2021 Towards engineering in memristors for emerging memory and neuromorphic computing: a review *J. Semiconduct.* **42** 013101

[2] Ielmini D 2018 Brain-inspired computing with resistive switching memory (RRAM): devices, synapses and neural networks *Microelectron. Eng.* **190** 44–53

[3] Athena F F, West M P, Hah J, Hanus R, Graham S and Vogel E M 2022 Towards a better understanding of the forming and resistive switching behavior of Ti-doped HfO$_x$ RRAM *J. Mater. Chem.* C **10** 5896–904

[4] Chandrasekaran S, Simanjuntak F M, Saminathan R, Panda D and Tseng T Y 2019 Improving linearity by introducing Al in HfO$_2$ as a memristor synapse device *Nanotechnology* **30** 445205

[5] Roy S *et al* 2020 Toward a reliable synaptic simulation using Al-doped HfO$_2$ RRAM *ACS Appl. Mater. Interfaces* **12** 10648–56

[6] Li L H, Xue K H, Zou L Q, Yuan J H, Sun H J and Miao X S 2021 Multilevel switching in Mg-doped HfO$_x$ memristor through the mutual-ion effect *Appl. Phys. Lett.* **119** 153505

[7] Tan T T, Du Y H, Cao A, Sun Y L, Zha G Q, Lei H and Zheng X S 2018 The resistive switching characteristics of Ni-doped HfO$_x$ film and its application as a synapse *J. Alloys Compd.* **766** 918–24

[8] Tan T T, Guo T T, Chen X, Li X J and Liu Z T 2014 Impacts of Au-doping on the performance of Cu/HfO$_2$/Pt RRAM devices *Appl. Surf. Sci.* **317** 982–5

[9] Zhang Z Z, Wang F, Hu K, She Y, Song S N, Song Z T and Zhang K L 2021 Improvement of resistive switching performance in sulfur-doped HfO$_x$-based RRAM *Materials* **14** 3330

[10] Hah J, West M P, Athena F F, Hanus R, Vogel E M and Graham S 2022 Impact of oxygen concentration at the HfO$_x$/Ti interface on the behavior of HfO$_x$ filamentary memristors *J. Mater. Sci.* **57** 9299–311

[11] Calka P *et al* 2014 Engineering of the chemical reactivity of the Ti/HfO$_2$ Interface for RRAM: experiment and theory *ACS Appl. Mater. Interfaces* **6** 5056–60

[12] Kim G S, Song H, Lee Y K, Kim J H, Kim W, Park T H, Kim H J, Min Kim K and Hwang C S 2019 Defect-engineered electroforming-free analog HfO$_x$ memristor and its application to the neural network *ACS Appl. Mater. Interfaces* **11** 47063–72

[13] Ku B, Abbas Y, Kim S, Sokolov A S, Jeon Y-R and Choi C 2019 Improved resistive switching and synaptic characteristics using Ar plasma irradiation on the Ti/HfO$_2$ interface *J. Alloys Compd.* **797** 277–83

[14] Wang Q *et al* 2019 Interface-engineered reliable HfO$_2$-based RRAM for synaptic simulation *J. Mater. Chem.* C **7** 12682–7

[15] Zhang W, Lei J Z, Dai Y X, Zhang X H, Kang L M, Peng B W and Hu F R 2022 Switching-behavior improvement in HfO$_2$/ZnO bilayer memory devices by tailoring of interfacial and microstructural characteristics *Nanotechnology* **33** 255703

[16] Chand U, Huang K-C, Huang C-Y, Ho C-H, Lin C-H and Tseng T-Y 2015 Investigation of thermal stability and reliability of HfO$_2$ based resistive random access memory devices with cross-bar structure *J. Appl. Phys.* **117** 184105

[17] Akbari M, Kim M-K, Kim D and Lee J-S 2017 Reproducible and reliable resistive switching behaviors of AlO$_x$/HfO$_x$ bilayer structures with Al electrode by atomic layer deposition *RSC. Adv.* **7** 16704–8

[18] Goux L, Fantini A, Govoreanu B, Kar G, Clima S, Chen Y Y, Degraeve R, Wouters D J, Pourtois G and Jurczak M 2012 Asymmetry and switching phenomenology in TiN\(Al$_2$O$_3$)\HfO$_2$\Hf systems *ECS Solid State Lett.* **1** P63–P5

[19] Goux L *et al* 2013 Understanding of the intrinsic characteristics and memory trade-offs of sub-μA filamentary RRAM operation *2013 Symp. on VLSI Technology Digest of Technical* (*Honolulu, HI*) pp 1–12

[20] Ding X, Feng Y, Huang P, Liu L and Kang J 2019 Low-power resistive switching characteristic in HfO$_2$/TiO$_x$ Bi-layer resistive random-access memory *Nanoscale. Res. Lett.* **14** 157

[21] Liu J, Yang H, Ji Y, Ma Z, Chen K, Zhang X, Zhang H, Sun Y, Huang X and Oda S 2018 An electronic synaptic device based on HfO$_2$TiO$_x$ bilayer structure memristor with self-compliance and deep-RESET characteristics *Nanotechnology* **29** 415205

[22] Huang C-Y, Huang C-Y, Tsai T-L, Lin C-A and Tseng T-Y 2014 Switching mechanism of double forming process phenomenon in ZrO$_x$/HfO$_y$ bilayer resistive switching memory structure with large endurance *Appl. Phys. Lett.* **104** 062901

[23] Azzaz M *et al* 2016 Improvement of performances HfO$_2$-based RRAM from elementary cell to 16 kb demonstrator by introduction of thin layer of Al$_2$O$_3$ *Solid State Electron.* **125** 182–8

[24] Woo J, Moon K, Song J, Lee S, Kwak M, Park J and Hwang H 2016 Improved synaptic behavior under identical pulses using AlO$_x$/HfO$_2$ bilayer RRAM array for neuromorphic systems *IEEE Electron Device Lett.* **37** 994–7

[25] Wan H J, Zhou P, Ye L, Lin Y Y, Tang T A, Wu H M and Chi M H 2010 *In situ* observation of compliance-current overshoot and its effect on resistive switching *IEEE Electron Device Lett.* **31** 246–8

[26] Li C *et al* 2018 Analogue signal and image processing with large memristor crossbars *Nature Electron.* **1** 52–59

[27] Chiu F-C 2014 A review on conduction mechanisms in dielectric films *Adv. Mater. Sci. Eng.* **2014** 578168

[28] Lim E W and Ismail R 2015 Conduction mechanism of valence change resistive switching memory: a survey *Electronics* **4** 586–613