physica status solidi rrl
www.pss-rapid.com

Analog-Type Resistive Switching Devices for Neuromorphic Computing

Wenbin Zhang, Bin Gao, Jianshi Tang, Xinyi Li, Wei Wu, He Qian, and Huaqiang Wu**

Brain-inspired neuromorphic computing has attracted considerable attention due to its potential to circumvent the "von Neumann bottleneck" and mimic human brain activity in electronic systems. The key to developing high-performance and energy-efficient neuromorphic computing systems lies in the realization of electronic devices that can closely mimic biological synapses. Resistive random-access memory (RRAM) has shown some important properties for implementing synaptic functions, including analog weight storage and analog switching. Herein, the recent progress in analog-type RRAM is reviewed. The mechanisms underlying the analog switching behavior in RRAM and different types of synaptic plasticity based on the analog switching behavior are discussed. Methods to improve the analog switching behavior and synaptic plasticity are then illustrated. Finally, a summary and a perspective on future research are presented.

1. Introduction

Neuromorphic computing hardware with synaptic devices can supplement conventional hardware in artificial intelligence (AI) applications.[1–9] The conventional hardware based on the von Neumann architecture and on complementary metal–oxide–semiconductor (CMOS) devices faces two main challenges that hinder its performance improvement: 1) low density of on-chip memories and 2) physical separation of the computing units and memory.[10,11] The capacity of the on-chip static random-access memory (SRAM, typically only a few megabytes) is too small to store the millions of weights in state-of-the-art deep neural networks (DNNs) because of its large size (100–200 F^2 per cell, where F is the feature size of the technology node).[12] Alternatively, the use of dynamic random-access memory (DRAM) for off-chip weight storage results in intensive and expensive data movement, inducing severe latency and high energy consumption (even >100 times more than that for on-chip weight storage).[13] Meanwhile, neuromorphic computing has inherited some salient features of the human brain, such as massive parallelism, an in-memory computing architecture, and an adaptive learning ability, making it suitable for high-performance AI applications.[14–16]

The realization of an electronic device that closely emulates a biological synapse is crucial for the development of neuromorphic computing hardware. Similar to that in the human brain, the learning process in neuromorphic computing hardware involves a mass of synapses repeatedly adjusting their weights step by step.[17,18] Therefore, the synaptic device must exhibit analog switching behavior for learning in today's mainstream neural networks, in which synaptic weight is in an analog fashion. Electronic devices enabling analog weight storage and analog switching have been proposed, such as resistive random-access memory (RRAM), FLASH, magnetic random-access memory (MRAM), phase-change memory (PCM), ferroelectric field-effect transistors (FeFETs), and electrochemical random-access memory (ECRAM).[19–23]

Here, we discuss the pros and cons of each type of device for analog synapse application. Each type of device has its advantages, and a single type of device cannot satisfy all application requirements. While FLASH is the most mature technology, its slow programming speed (10 μs–1 ms) and high operation voltage (>10 V) hinder its performance as a synaptic device. Compared with RRAM, MRAM and PCM also have the advantage of technological maturity, while showing little difference in programming speed (~ns) and operation voltage.[24,25] However, despite some breakthroughs,[26,27] MRAM and PCM show relatively poor analog switching behavior due to the low on/off ratio and the uncontrolled RESET, respectively. Multiterminal cells, such as FeFETs, ECRAM, and 3T1C + 2T2R (3-transistor 1-capacitor + 2-transistor 2-resistor), allow for more controllable and linear analog switching as a trade-off for complexity,[28–31] but their reliability requires more evaluation due to the complex cell structure and unclear working mechanisms. Overall, RRAM appears to be more attractive, enabling a compact two-terminal structure (each cell occupies 4 F^2), great scalability (even 2 nm),[32–34] a high programming speed (even <1 ns),[35–38] and low energy consumption (<1 pJ).[39–41] An RRAM array in a crossbar architecture can enable an extremely high density, excellent scalability, and parallel operation.[42,43]

Two approaches can generally be used to develop neuromorphic computing hardware using RRAM. The first approach uses

W. Zhang, Prof. B. Gao, Prof. J. Tang, Dr. X. Li, Dr. W. Wu, Prof. H. Qian, Prof. H. Wu
Institute of Microelectronics
Tsinghua University
Beijing 100084, China
E-mail: gaob1@tsinghua.edu.cn; wuhq@tsinghua.edu.cn

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pssr.201900204.

DOI: 10.1002/pssr.201900204

an RRAM crossbar array to implement DNNs.[1–4] The RRAM crossbar array can directly map DNNs, in which the conductance represents the weights in the DNNs, whereas the word line voltages represent inputs. Based on Ohm's law and Kirchhoff's law, the current through each bit line is the weighted sum of the voltage applied on all the word lines. In this manner, the matrix–vector multiplication, which is the most common and intensive computation in DNNs, is completed with massive parallelism and ultrahigh energy efficiency. The other approach is a biologically realistic one aimed at developing spiking neural networks (SNNs) that resemble synaptic behavior in the human brain.[6–8] During the learning process of SNNs, synapses gradually adjust their weights based on the spike timing of neurons, which is rather similar to the human brain.

This article focuses on analog-type RRAM as the synapse for neuromorphic computing applications. RRAMs with different types of switching behavior and their underlying mechanisms are first discussed, including their pros and cons in the emulation of synapses. We then summarize the different types of synaptic plasticity based on the analog switching behavior, including the plasticity with consecutive pulses and the spiking time-dependent plasticity (STDP). Methods to improve the analog switching behavior of RRAM are subsequently reviewed. Finally, a summary and an outlook on analog-type RRAM are presented.

2. Analog Switching Behavior and the Underlying Mechanisms in RRAM

An RRAM is a two-terminal device with a metal–insulator–metal structure: a switching layer sandwiched between two electrodes.

Bin Gao is an associate professor in the Institute of Microelectronics, Tshinghua University, Beijing, China. He received the B.S. degree in physics from Peking University, Beijing, China, in 2008, and the Ph.D. in microelectronics from Peking University in 2013. He joined Tsinghua University in 2015. His research interests focus on optimization, characterization, and modelling on oxide-based RRAM.

The RRAM operation depends on the difference between resistance states. Switching from a high-resistance state (HRS) to a low-resistance state (LRS) is referred to as SET. Conversely, switching from an LRS to an HRS is referred to as RESET. In a bipolar switching RRAM, the switching direction depends on the polarity of the applied voltage, which is the focus of this review.

Four types of RRAM based on the switching behavior are presented: binary switching RRAM, multilevel switching RRAM, unidirectional analog switching RRAM, and bidirectional analog switching RRAM (referred to as analog-type RRAM herein) (**Figure 1**). First, binary switching RRAM exhibits both abrupt SET and abrupt RESET and has only two stable states.[44–47] Current compliance is usually used in the SET operation, which can be realized by controlling the gate voltage of the transistor in series with the RRAM. Multilevel switching is often achieved by modulating the SET compliance current or the RESET

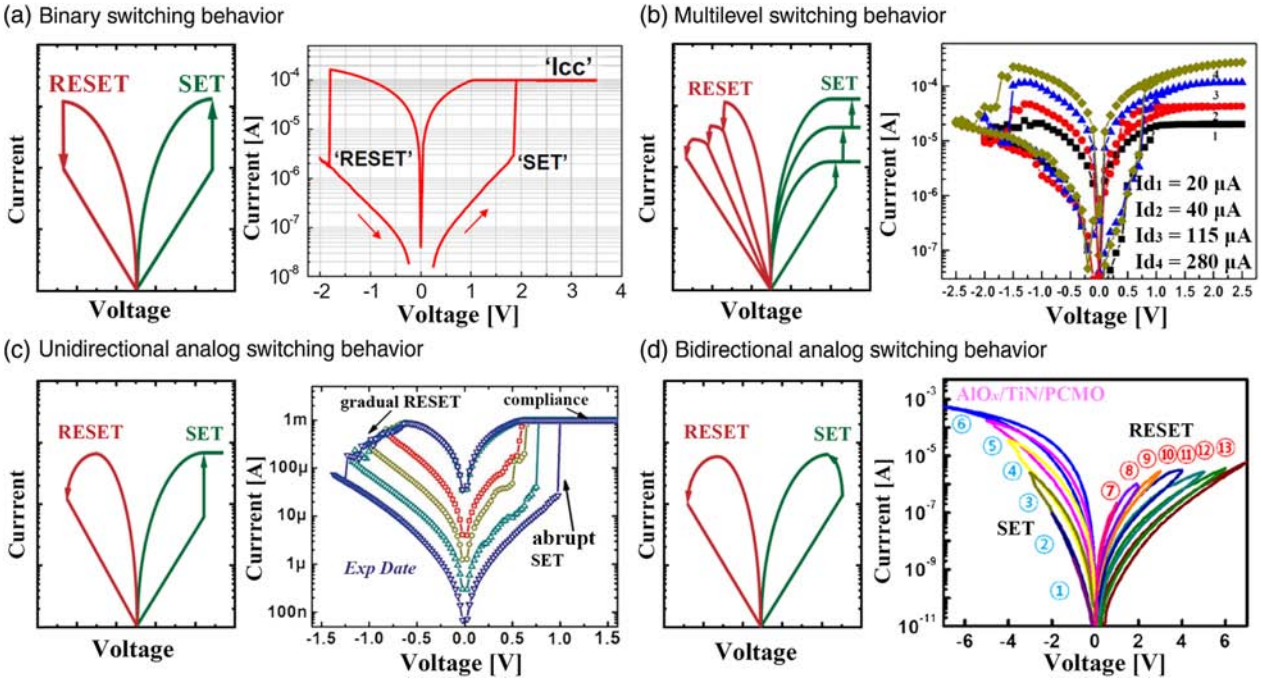


Figure 1. Illustration and experimental data of typical switching behaviors. a) Binary switching, b) multilevel switching, c) unidirectional analog switching, d) bidirectional analog switching. (a) Reproduced with permission.[45] Copyright 2017, IEEE. (b) Reproduced with permission.[49] Copyright 2011, IEEE. (c) Reproduced with permission.[57] Copyright 2013, Wiley-VCH. (d) Reproduced with permission.[63] Copyright 2013, IEEE.

voltage.[37,48–52] Binary switching RRAM and multilevel switching RRAM have mainly been investigated for their data storage applications. Unidirectional analog switching RRAM exhibits either gradual SET or gradual RESET.[53–56] As an alternative, RRAM with only gradual RESET was utilized for adaptive learning.[57] Lastly, bidirectional analog switching RRAM exhibits both gradual SET and RESET.[58–63] The conductance can be gradually modulated by the applied pulses, which is favorable for building neural networks with an adaptive learning capability.

Based on the switching mechanism, RRAMs can be roughly classified into two categories: filamentary RRAM and interfacial RRAM. The switching mechanism of filamentary RRAM is the formation and rupture of the localized conductive filament (CF) consisting of a chain of oxygen vacancies (V_O) or metallic ions (**Figure 2**a). The SET process of filamentary RRAM is typically abrupt and difficult to control.[59,64] Multilevel switching has been demonstrated in filamentary RRAM.[37,48,53] CFs with different thicknesses can be obtained by modulating the SET compliance current, leading to a multilevel LRS, while gaps with different distances can be obtained by modulating the RESET voltage, leading to a multilevel HRS.[65–67] Methods to improve the analog switching behavior of filamentary RRAM will be discussed in detail in Section 4.

In contrast, the switching of interfacial RRAM depends on the modulation of the oxygen ion (O^{2-}) distribution over the entire active interface (e.g., oxide/electrode interface or oxide/oxide interface), which leads to collapse of the Schottky barrier or narrowing of the tunneling barrier.[68,69] Figure 2b shows the energy band diagrams during SET and RESET of a typical interfacial RRAM. During RESET, the motion of O^{2-} toward the metal electrode under a positive bias results in oxidation at the interface, leading to a metal–oxide–semiconductor (MOS) contact with increased device resistance. During SET, O^{2-} moves back to the semiconductor layer under an opposite bias, leading to a metal–semiconductor (M–S) contact with a Schottky diode-like behavior. Interfacial RRAM intrinsically enables analog weight storage and analog switching because the shift of the O^{2-} distribution is gradual.[58,70,71] Generally, interfacial RRAM exhibits better device-to-device and cycle-to-cycle uniformities than filamentary RRAM,[72,73] as the formation and rupture of the CF in filamentary RRAM is a local effect, with only a few O^{2-}/V_O contributing to the switching, while in interfacial RRAM, O^{2-}/V_O throughout the entire area uniformly contributes to the switching. However, interfacial RRAM cannot easily simultaneously provide a fast switching speed, good retention, and a small switching voltage, implying a trade-off when designing an interfacial RRAM device for specific applications.[74]

3. Synaptic Plasticity Based on Analog Switching Behavior

Synaptic plasticity in biology refers to the increase (potentiation) or decrease (depression) in the synaptic weight, which can be emulated in analog-type RRAM via the increase or decrease in conductance induced by SET and RESET operations, respectively.[75] A series of DNNs has been demonstrated on analog-type RRAM crossbar arrays, in which consecutive pulses were used to realize potentiation and depression of analog-type RRAM in the learning process.[1–3] Consecutive identical pulses are preferred because this method simplifies the peripheral circuitry used to program analog-type RRAM.[12,57] The backpropagation learning rule is widely used in the learning process of DNNs, in which linear and symmetric weight updating is crucial in achieving high-accuracy learning.[76–78] **Figure 3**a illustrates the ideal potentiation and depression behaviors with linear and symmetric weight updating, whereas Figure 3b shows the switching behavior of a typical analog-type RRAM, in which obvious nonlinearity can be observed.

The STDP learning rule has also been emulated to build SNNs with analog-type RRAM.[7] In biological neural systems, STDP is a basic learning rule related to Hebbian learning, which states that synaptic plasticity depends on the relative timing between the spikes (activities) of the pre- and postsynaptic neurons.[79–81] Long-term potentiation occurs when the presynaptic spikes precede the postsynaptic spikes. Conversely, long-term depression occurs when the presynaptic spikes follow the postsynaptic spikes. Various programming schemes have been proposed to implement STDP with analog-type RRAM. Most of the schemes were typically realized by the overlap between two spikes on the two terminals,[39,53,82–84] while recently, some nonoverlapping schemes have been developed.[85–87] In the overlapping methods, translating the timing difference into superimposed spikes (pulses) with different amplitudes, widths, or both is crucial.

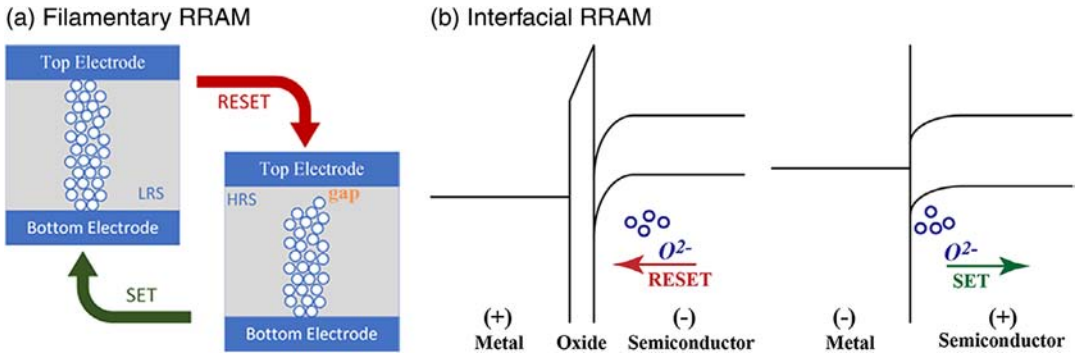


Figure 2. Illustration of the RRAM switching mechanism. a) Switching mechanism of filamentary RRAM. The switching is attributed to CF formation/rupture. b) Simple energy band diagrams of interfacial RRAM. For the RESET operation, O^{2-} moves toward the metal, generating oxide, which results in an MOS contact. For the SET operation, O^{2-} moves toward the semiconductor layer, leading to an M–S contact.

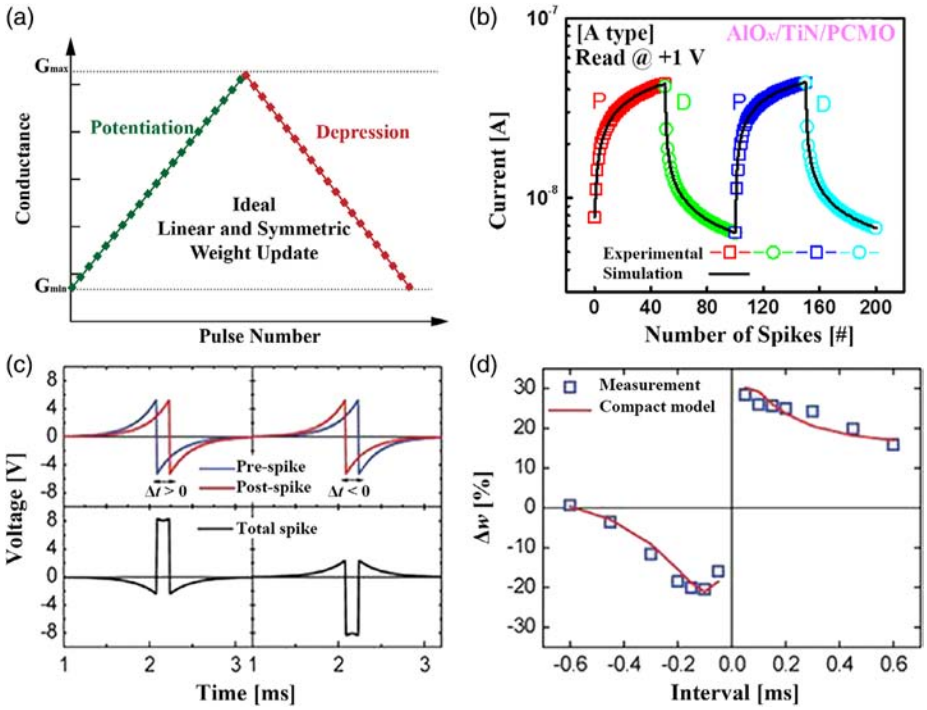


Figure 3. Plasticity in DNNs and SNNs. a) Ideal conductance modulation with the pulse number. The illustrated linear and symmetric weight updating is crucial for DNNs with the backpropagation learning rule. b) Weight updating of a typical analog-type RRAM. The change of the conductance with the pulse number shows obvious nonlinearity. c) Action-potential-like waveforms for the STDP demonstration. The overlap between the pre- and postspikes translates the timing difference into amplitude and width differences of the total spike. d) Synapse weight change with the relative timing. Potentiation occurs for $\Delta t > 0$, whereas depression occurs for $\Delta t < 0$. (b) Reproduced with permission.[63] Copyright 2013, IEEE. (c,d) Reproduced under the terms of the CC-BY Creative Commons Attribution 4.0 International License.[88] Copyright 2015, The Authors, published by Springer Nature.

Consequently, the total spike modulates the conductance as a function of the relative timing. Figure 3c,d shows the demonstration of STDP with a TaO_x/TiO_2-based analog-type RRAM, where the overlap between the two action-potential-like waveforms produces the total spike applied on the RRAM.[88] Other methods to implement STDP with RRAM, including overlapping or nonoverlapping methods, will be presented in Section 5.2.

4. Methods to Improve the Analog Switching Behavior

This section discusses the methods used to improve the analog switching behavior, particularly for filamentary RRAM. To avoid abrupt SET in filamentary RRAM, the origin of this phenomenon should first be clearly understood. Investigations have shown that SET in filamentary RRAM is a positive-feedback process.[59,64] The growth of the CF increases the local electric field or temperature, which in turn promotes V_O generation and migration and CF growth. This process forms a single strong CF, which hinders gradual control of SET. Therefore, methods that avoid the formation of a single strong CF have been proposed, including 1) switching with the formation/rupture of multiple weak CFs; 2) switching with CF size modulation; and 3) switching with channel composition modulation. **Table 1** summarizes the methods used to realize analog-type RRAM, and **Table 2** shows some corresponding examples of analog-type RRAMs and their figures of merit.

4.1. Analog Switching in RRAM with Multiple Weak CFs

Analog switching was achieved in RRAM with multiple weak CFs. Wu et al. reported that a transition from abrupt to gradual SET occurs with increasing temperature in a HfO_x-based RRAM. Therefore, a HfO_x/thermally enhanced layer (TEL) stack was proposed, in which the TEL with a low thermal conductivity confines heat in the HfO_x layer. As an example, a HfO_x/TaO_x RRAM exhibiting gradual SET was demonstrated. The TEL layer promotes a uniform V_O distribution and accelerates the formation of multiple weak CFs. The formation and rupture of each CF contribute only a portion to the total conductance change; hence, a transition from abrupt to gradual switching occurs.[59]

To investigate the mechanism of this transition, the V_O distribution in different types of RRAM was simulated using the kinetic Monte Carlo method (**Figure 4**).[89] Compared with that of a metal/HfO_x RRAM (strong CF type), the V_O distribution of a TEL/HfO_x RRAM is scattered and random. The disordered V_O distribution promotes the existence of multiple weak CFs and enables analog switching, which has been experimentally demonstrated. These simulation results confirmed the relation between the V_O distribution and switching behavior, i.e., a disordered V_O distribution improves the analog switching behavior.

Table 1. Comparison of the analog-type RRAMs for neuromorphic computing.

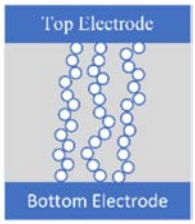
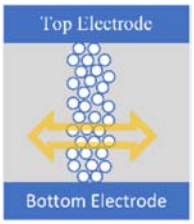
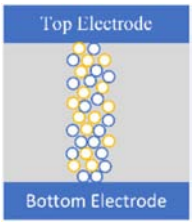
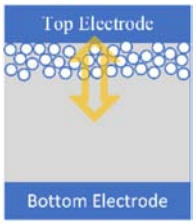
Methods	Formation/rupture of multiple weak CFs	CF size modulation	Channel composition modulation	Interface switching
Schematic				
Advantages	High speed Low current	Good reliability	High endurance High speed	Most gradual switching
Disadvantages	Noise Relaxation	Low on/off ratio High current	High current	Intrinsic trade-off among the switching voltage, speed, and retention
System-level demonstration	Perceptron for face classification on a 1k array[2] Generative adversarial network demonstration on a 1k array[109]	Gray scale image recognition on a 24×24 RRAM array[110]	Multiple-layer network for handwritten digital recognition on an 8k array[3] Fully memristive SNN for letters recognition on an 8×8 array[6]	SNN for handwritten digital recognition on a 1k array[98] Neuromorphic speech systems on a 1k array[63]

Table 2. Examples of analog-type RRAM and their figures of merit.

Device structure[a]	Device size	HRS/LRS	Number of states	Linearity	Operation voltage (SET/RESET)	Operation speed	Retention	Mechanism
ETML/HfO_x	500 nm–2 μm	1/10 μS	128	0.04/−0.63	1.6/1.5 V	50 ns	>10 min (@125 °C)	Formation/rupture of multiple weak CFs
AlO_x/HfO_2	400 nm	20/60 μS	40	−0.01/0.59	0.9/1 V	100 μs	/	CF size modulation
Ta/HfO_2	3 μm	0.4/1.7 mS	26	/	1/1.17 V	100 ns	2.7×10^5 s (@250 °C)	Channel composition modulation
Al/TiN/PCMO	150 nm	0.07/0.5 μS	50	3.68/−6.76	3/3 V	1 ms	$>10^4$ s	Interface switching
TaO_x/TiO_2	20 μm^2	0.1/0.2 μS	105	0.66/−0.69	3/3 V	40/10 ms	/	Interface switching
Desirable	<40 nm	<50/<3 μS	>16	<1	<1.5 V	<100 ns	>1 day (@85 °C)	/

[a]The data are adapted from previous studies[63,94,97,103,111].

The introduction of a dopant into the switching layer also promotes the formation of multiple weak CFs, as demonstrated in a Gd:HfO_x RRAM and a TEL/Al:HfO_x RRAM.[89–91] The V_O tends to be localized around dopants in the switching layer due to the lower formation energy. Therefore, the dopant randomness promotes a scattered V_O distribution, which promotes the formation of multiple weak CFs and improves the analog switching behavior. The TEL/Al:HfO_x RRAM also showed improved cycle-to-cycle and device-to-device uniformities and better retention compared with the TEL/HfO_x RRAM due to the localized V_O.[92] Based on the device enabling reliable bidirectional analog switching, a one-layer perceptron neural network was demonstrated on a 1 k one-transistor one-resistor (1T1R) array, providing face classification using an RRAM array for the first time.[2]

4.2. Analog Switching with CF Size Modulation

CF lateral extension appears to be more controllable than CF formation. Kim et al. fabricated a Ta_2O_5/TaO_x bilayer RRAM and measured its switching behavior under identical pulses.[93] In particular, a comprehensive physical model capturing the switching process was developed. The measured and simulation results showed that the CF evolution in potentiation is a two-step process: 1) initial gap-filling with V_O (CF formation) and 2) subsequent lateral extension of the CF via V_O diffusion. The former step produces an abrupt conductance change, while a gradual conductance change occurs during the latter step, showing the potential to realize analog switching.

Analog switching was also realized with CF size modulation in an AlO_x/HfO_2 bilayer RRAM in which the AlO_x layer acted as a barrier layer to prevent CF rupture.[94] During the device fabrication, HfO_2 was deposited on an Al bottom electrode (BE) by atomic layer deposition, which induced oxidation on the Al surface. The AlO_x layer serves as a barrier layer limiting the motion of V_O through it owing to its lower V_O mobility than that of the HfO_2 layer. Therefore, lateral modulation of the CF preferentially occurs via the movement of V_O through the HfO_2 layer, which enables analog switching.

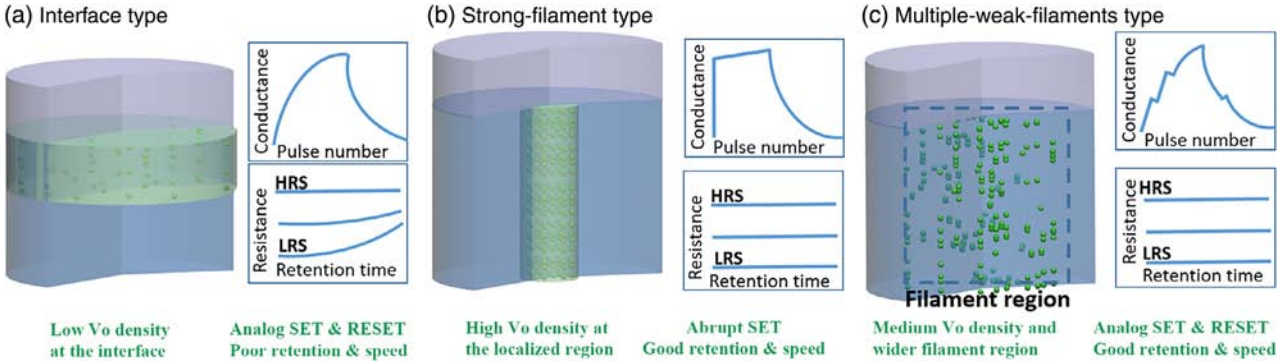


Figure 4. Schematic of the V_O distribution and characteristics of different types of RRAM. These diagrams show the relation between the V_O distribution and the switching mechanism that determines the device characteristics. a) Interfacial RRAM has a low V_O density at the interfacial layer. The device exhibits good analog switching behavior but poor retention and speed. b) RRAM with a single strong CF has a high V_O density in the localized region. The device exhibits abrupt SET. c) RRAM with multiple weak CFs has a medium V_O density over a wider filament region. The device shows a balance between the analog switching behavior and the retention/speed. Reproduced with permission.[89] Copyright 2017, IEEE.

The above link between the CF properties (and evolution) and the switching behavior was further illustrated by a comparative study of a HfO_2 RRAM and a TiO_x/Ta_2O_5 RRAM.[62] In the potentiation of the HfO_2 RRAM, once the CF forms under the first SET pulse, the subsequent SET pulses do not lead to an enlargement of the CF because the electric field is not strong enough for V_O generation (**Figure 5**a–c). In contrast, in the potentiation of the TiO_x/Ta_2O_5 RRAM, V_O is preferentially generated in the vicinity of the CF due to the thermal effect, driving the enlargement of the CF under each SET pulse (Figure 5d–f). Compared with the HfO_2 RRAM, the different dynamics of the TiO_x/Ta_2O_5 RRAM originate from the reduced dependence on the electric field and the higher dependence on the local temperature, as the O^{2-} diffusion in Ta_2O_5 is characterized by a larger activation energy and a smaller field acceleration factor. The simulation results clearly showed that the RESET operation

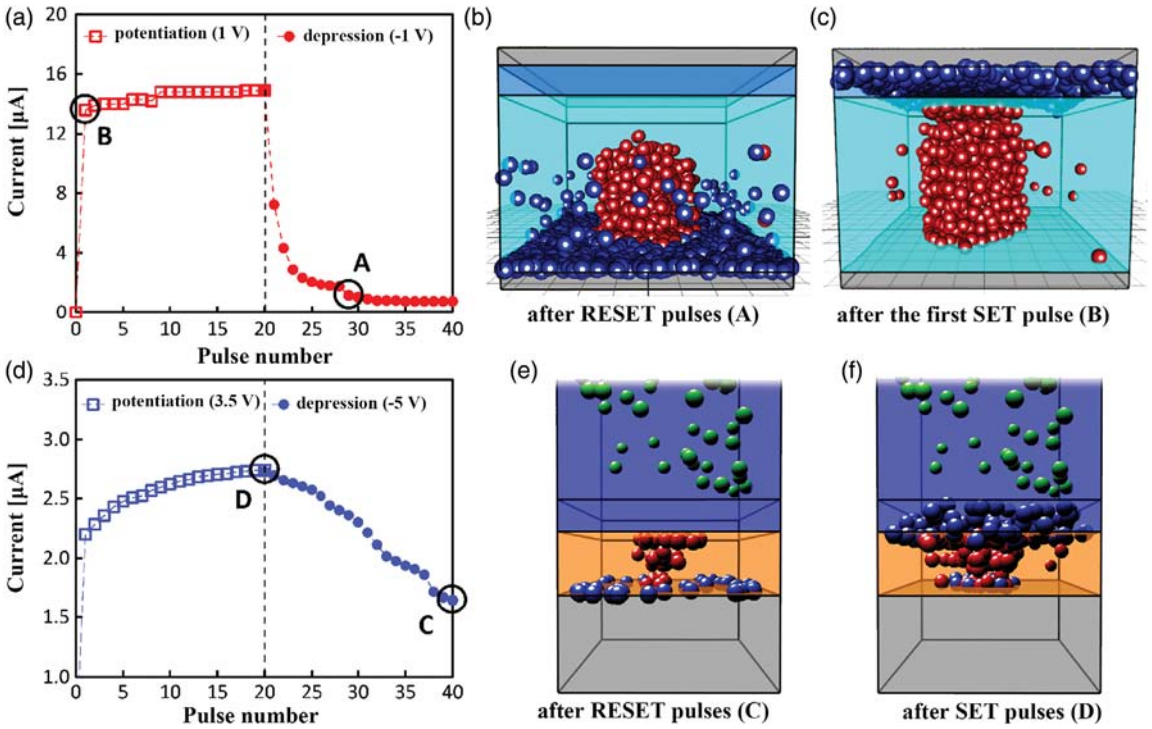


Figure 5. Simulated switching behaviors and switching dynamics of RRAMs with different CF properties. a) Potentiation and depression behavior of the HfO_2 RRAM. Digital switching behavior is observed. b,c) V_O (red spheres) and O^{2-} (blue spheres) distributions corresponding to the resistance states marked A and B in (a). The switching mechanism is the formation and rupture of a single strong CF. d) Potentiation and depression behavior of the TiO_x/Ta_2O_5 RRAM. Analog switching behavior is observed. e,f) V_O and O^{2-} distributions corresponding to the resistance states marked C and D in (d). The CF size modulation leads to the analog switching behavior. Reproduced with permission.[62] Copyright 2017, IEEE.

leads to CF narrowing, rather than to gap widening; hence, analog switching can be obtained through CF size modulation. However, this method enables analog switching at the cost of a reduced on/off ratio, caused by the incomplete dissolution of the CF.

4.3. Analog Switching with Channel Composition Modulation

In addition to the two abovementioned scenarios, analog switching behavior was also observed in RRAM with a dual atom CF, which was realized by modulating the channel composition. Miao et al. developed a Ta/TaO_x-based RRAM, whose conduction channel consisted of an amorphous Ta(O) solid solution.[95,96] Continuous tuning of the oxygen concentration in the channel was demonstrated by controlling the compliance current with the series transistor in SET operation. Herein, the continuous modulation of the channel composition led to analog switching behavior. Very high endurance switching was also realized because the amorphous structure enabled significant mobile species accommodation. Various analysis methods, including cross-sectional transmission electron microscopy, electron-energy-loss spectroscopy, and X-ray photoemission spectroscopy, were used to identify the channel element.

Moreover, analog switching behavior was also demonstrated in a Ta/HfO_2 RRAM via continuous channel composition modulation.[97] The switching is attributed to the growth and reoxidation of Ta-rich and O-deficient conduction channels through the motion of Ta cations (Ta^{x+}) and O^{2-}. Under the SET operation, the migration of Ta^{x+} into and O^{2-} out of the HfO_2 layer leads to a higher Ta but lower O concentration in the conduction channel, with increased conductance of the device; conversely, the RESET operation reverses the process and decreases the device conductance. The simultaneous drift of both ions under the electric field is caused by their comparable mobilities and similar migration barriers within HfO_2. The thermally enhanced lateral diffusion also plays a vital role in the device operation, as the competition between the drift and diffusion is considered to be responsible for the analog switching behavior. Record high endurance (1.2×10^{11} switching cycles and over 2^{20} potentiation/depression epochs) was demonstrated in this device.

4.4. Interface Engineering of Analog-Type RRAM

Theoretically, interfacial RRAM inherently enables analog switching, which was thought to be more suitable for analog synapse applications. However, specific issues still hinder its application. In this section, the development of a $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO)-based analog synapse, which has been intensively investigated, is reviewed.

Park et al. first reported an analog-type RRAM based on an Al/PCMO stack.[98,99] An insulating AlO_x thin layer is spontaneously formed because of the chemical reaction, along with the absorption of O^{2-} from PCMO. A positive bias on the W electrode attracts O^{2-} and forms a thick insulating layer (AlO_x formation), leading to an HRS. In contrast, under a negative bias on the W electrode, O^{2-} moves back from AlO_x to the PCMO bulk layer (AlO_x dissolution), leading to an LRS. Potentiation and depression under consecutive pulses and STDP were demonstrated.

However, the Al/PCMO-based RRAM shows an apparent asymmetric switching behavior. The conductance nearly decreases to the minimum value upon the application of the first RESET pulses during depression (abrupt RESET), whereas the conductance gradually increases during potentiation (gradual SET). This difference emerges because the AlO_x formation is faster than the AlO_x dissolution as a result of the low electronegativity of Al. To overcome this issue, a N-rich TiN layer was inserted between AlO_x and PCMO to balance the speed of the AlO_x formation and dissolution, which led to a more symmetric switching behavior and accompanying gradual conductance modulation during depression.[63]

Another issue with the Al/PCMO RRAM device is that the low electronegativity of Al harms the LRS retention because AlO_x tends to form spontaneously, switching the device from an LRS to an HRS. A Mo electrode was used instead of Al to improve the retention, along with the linearity during depression, at the cost of reducing the on/off ratio.[100] Moon et al. proposed the insertion of a thin Mo buffer layer between Al and PCMO rather than a change of the electrode, which balanced the retention and the on/off ratio.[60]

5. Optimization Methods for DNNs and SNNs

5.1. Optimization of DNNs

Various programming schemes have been proposed to linearly and symmetrically update the conductance with the pulse number in DNNs. Most of them rely on pulses with variable amplitudes or widths. However, such programming schemes may place a burden on the design of the peripheral circuitries, affecting the programming speed and power consumption.

For the RRAM with abrupt SET, gradually increasing its conductance during potentiation is challenging. Accordingly, SET schemes with either SET voltage/pulse width ramping or gate voltage ramping were proposed.[3,53,101] For example, in a HfO_2/Ti-based RRAM, varying the SET conditions (ramping either the SET voltage or pulse width) at a fixed gate voltage results in gradual potentiation. Ramping the gate voltage at a fixed SET voltage, which tends to gradually increase the CF size, leads to a similar result.[101]

The application of consecutive identical pulse pairs was proposed to improve the weight update linearity during potentiation, in which each pulse pair consisted of a SET pulse and a subsequent optimized RESET pulse (**Figure 6**a).[101] During potentiation, the first SET pulse leads to a fully connected CF, and the subsequent RESET pulse partially disconnects the CF, resulting in an intermediate-resistance state (IRS). The next SET pulse leads to a stronger CF because of the higher electric field between the remaining CF and the electrode. The subsequent RESET pulse then drives the same amount of V_O from the thicker CF, resulting in a higher IRS compared with the previous one. Thus, the conductance can be gradually adjusted as a function of the number of pulse pairs. Figure 6c,d shows the potentiation behavior and the physical mechanism.

The use of a current pulse for RESET was also proposed to improve the depression linearity.[102] The pulse power tendencies are different between voltage pulse depression and current pulse

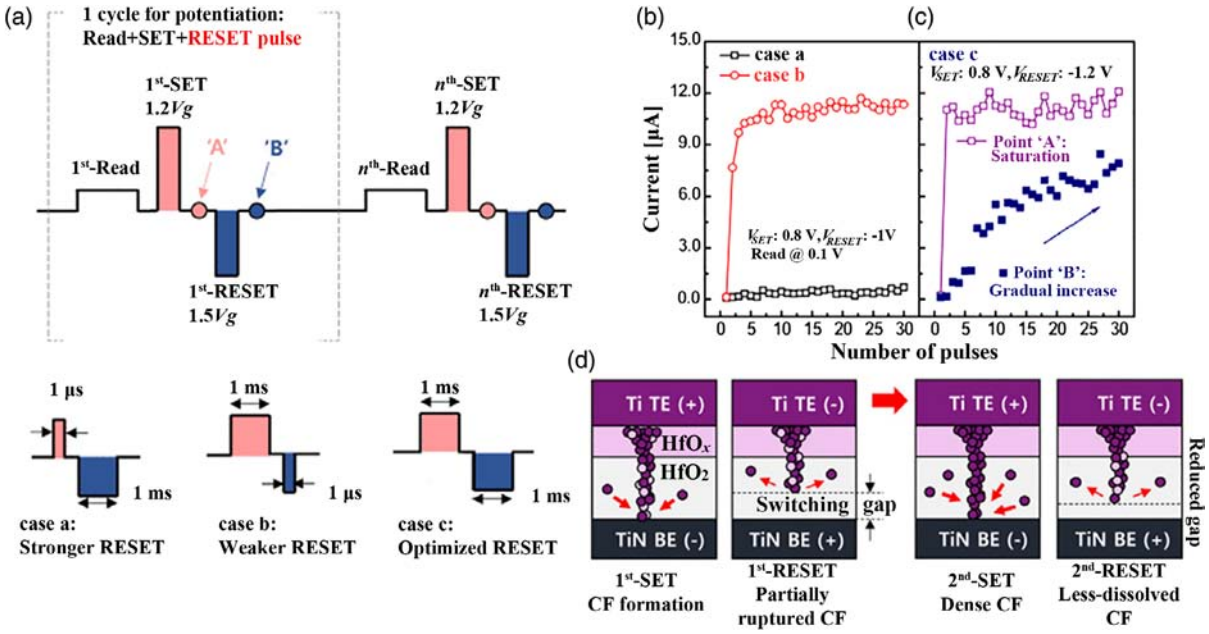


Figure 6. Gradual potentiation with pulse pairs. a) Different pulse pairs for the potentiation operation. A RESET pulse follows a SET pulse in each pair. The corresponding potentiation behaviors are shown in (b) (cases a and b) and (c) (case c). b) Potentiation behaviors of the pulse pairs with a stronger RESET pulse (case a) and a weaker RESET pulse (case b). A strong RESET pulse leads to complete rupture of the CF, whereas a weak RESET pulse is insufficient for disconnecting the CF. c) Potentiation behaviors of the pulse pairs with an optimized RESET pulse (case c). The optimized RESET pulse dissolves the CF to form a proper gap, enabling an IRS. d) Schematic illustration of the potentiation dynamics. Reproduced with permission.[101] Copyright 2016, IEEE.

depression. In the voltage pulse mode, the pulse power ($P = V^2 \times G$) significantly decreases with the pulse number because of the decreasing conductance. Hence, apparent nonlinearity is observed. In contrast, in the current pulse mode, the pulse power ($P = I^2/G$) increases with the pulse number, and the positive power feedback compensates for the change in the switching power (needed to realize linear depression) with the change in the pulse power, thereby significantly improving the linearity.

The application of consecutive nonidentical pulse pairs was proposed for more controllable depression.[3] During depression, a sufficient RESET pulse is first applied to initialize the state, followed by SET operation with a variable gate voltage that decreases with the pulse number (**Figure 7**c). This scheme takes advantage of the transistor to specify a compliance current during depression, which was previously used for potentiation. With gate voltage ramping during potentiation and the abovementioned scheme for depression, Li et al. demonstrated an excellent linear and symmetric weight update behavior with a Ta/HfO_2/Pt-based RRAM (Figure 7b–d). They built a multilayer neural network for MNIST handwritten digit recognition and demonstrated its in situ and self-adaptive learning capability with relatively high accuracy.

In addition to programming schemes, methods based on the elaborate control of the switching dynamics have also been proposed to improve the linearity and symmetry in weight updating.[63,94,103] Taking RRAM with multiple weak CFs as an example, a HfO_x/electrothermal modulation layer (ETML) stack was developed in which the ETML operates not only as a TEL but also as an electric field-modulator.[103] A lower thermal conductivity of the ETML promotes a uniform V_O distribution, which is good for linear potentiation because, in this case, each V_O almost contributes equally to the total conductance. A higher resistivity of the ETML helps decrease the change rate of the electric field in the gap region and improve the linearity during depression.

5.2. Optimization of SNNs

In the above demonstrations of STDP, the spike waveforms must be carefully engineered to ensure that the overlap suitably modulates the conductance, which also induces undesirable energy consumption and complexity of the peripheral circuitries. In this section, we present representative demonstrations of STDP with improved performance.

Lashkare and coworkers proposed an RRAM in series with a bipolar selector as a synapse (1S1R synapse) to reduce the undesirable energy consumption.[82] When implementing STDP with the superposition of pre- and postsynaptic spikes, only a small portion of the total spike changes the RRAM conductance, while the remaining portion induces an inadvertent current, resulting in unnecessary energy consumption. The 1S1R synapse was proposed to avoid this inadvertent energy loss, in which the selector is used to cutoff the current whenever the voltage of the total spike is less than the selector threshold.

The realization of STDP with nonoverlapping spikes has been demonstrated in the synapse using an RRAM in series with a diffusive memristor (**Figure 8**a).[85] The diffusive memristor

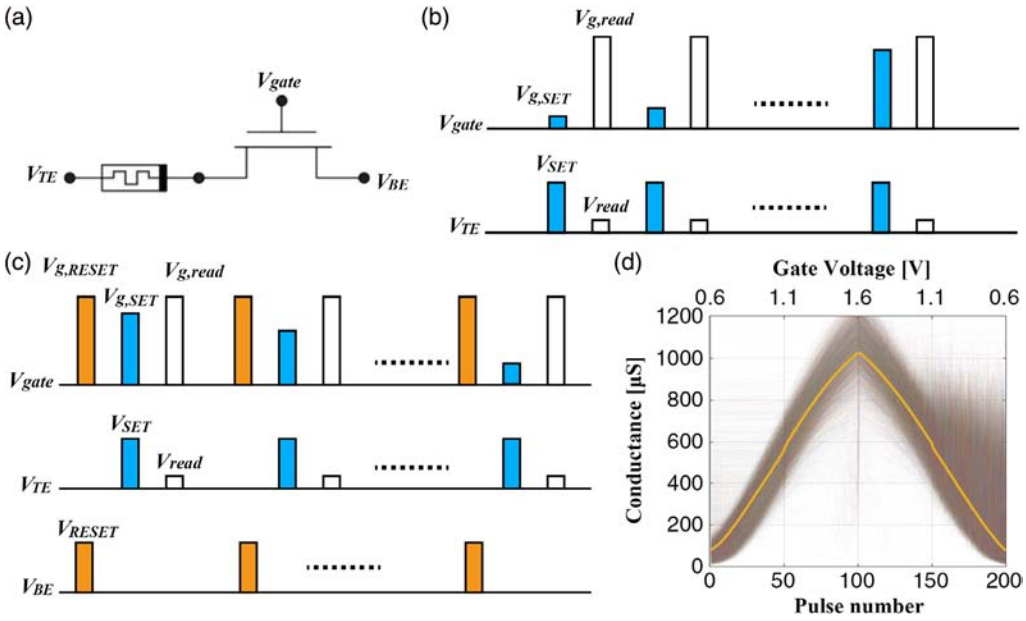


Figure 7. Method to realize linear and symmetric weight updating. a) 1T1R structure. The transistor operates as a current limiter in both potentiation and depression. b) Pulses to increase the conductance. The gate voltage is ramped to obtain linear potentiation. c) Pulses to decrease the conductance. A RESET operation is carried out to initialize the state, followed by a SET operation with a specific gate voltage to increase the conductance to the target value. d) Potentiation and depression behavior in an 8k array over 200 pulse cycles. Excellent linear and symmetric weight updating with minimal cycle-to-cycle and device-to-device variations is realized. Reproduced under the terms of the CC-BY Creative Commons Attribution 4.0 International License.[3] Copyright 2018, The Authors, published by Springer Nature.

exhibits unipolar nonlinear threshold switching behavior with a finite delay time, and it relaxes to its original HRS over a finite time (tens of milliseconds) after the SET pulse ends (Figure 8b). Therefore, the diffusive memristor serves as a "selector" with delay and relaxation characteristics in the synapse emulator. Figure 8c illustrates the spikes for the STDP demonstration. Each spike consists of a short high-voltage pulse followed by a long low-voltage pulse. The diffusive memristor is turned on by the first spike, and its resistance gradually increases over time. The second spike occurs at a time Δt after the end of the first spike and changes the conductance of the RRAM depending on the current state of the diffusive memristor, which is a function of Δt. Therefore, a smaller Δt corresponds to a larger change in the RRAM conductance. Potentiation occurs if the prespike precedes the postspike, whereas depression occurs if the prespike follows the postspike (Figure 8d). With the diffusive memristor providing an intrinsic timing mechanism, STDP realization does not require complex pulse engineering or spike overlap.

STDP with nonoverlapping spikes has also been demonstrated in RRAM with internal dynamics.[86,87] For example, the O^{2-} movement in the WO_x RRAM can be caused by both the electric field during programming and spontaneous diffusion. Moreover, the internal ionic dynamics lead to a conductance decay, which appears to occur on two very different time scales: the conductance exhibits a swift decay immediately after the stimulation, while the decay becomes much slower after a few hundreds of milliseconds. In the STDP implementation, pulses with identical amplitudes and widths are applied on the two terminals at different times. However, their effects do not cancel each other, and the second pulse dominates the net effect due to the spontaneous conductance decay after the first pulse. The conductance changes depending on the relative timing between the nonoverlapping pre- and postsynaptic pulses.[86]

STDP has also been demonstrated in RRAM with multilevel switching behavior with the help of a serial transistor to enforce current compliance.[104] As shown in **Figure 9**a,b, the 1T1R synapse functions with overlapping of the gate voltage and the top electrode (TE) voltage, rather than with the superposition of the voltages at the two terminals of the RRAM. The postsynaptic neuron generally collects current from 1T1R synapses via its virtual ground input node and fires once the integrated current exceeds the internal threshold. Upon firing, in addition to sending a spike pulse to the subsequent layer of neurons, the postsynaptic neuron also delivers a pulse back to the TE of the 1T1R synapse. For $\Delta t > 0$, the positive TE voltage overlaps with the presynaptic spike, leading to a SET operation and thus to potentiation. In contrast, for $\Delta t < 0$, the negative TE voltage overlaps with the presynaptic spike, leading to a RESET operation and thus to depression (Figure 9c).

6. Summary and Outlook

An analog-type RRAM is a promising synaptic device for neuromorphic computing applications due to its excellent scalability and simple two-terminal structure for high-density integration, analog switching behavior for synaptic plasticity emulation, and CMOS process compatibility for easy mass production. Synaptic plasticity under consecutive pulses and STDP were

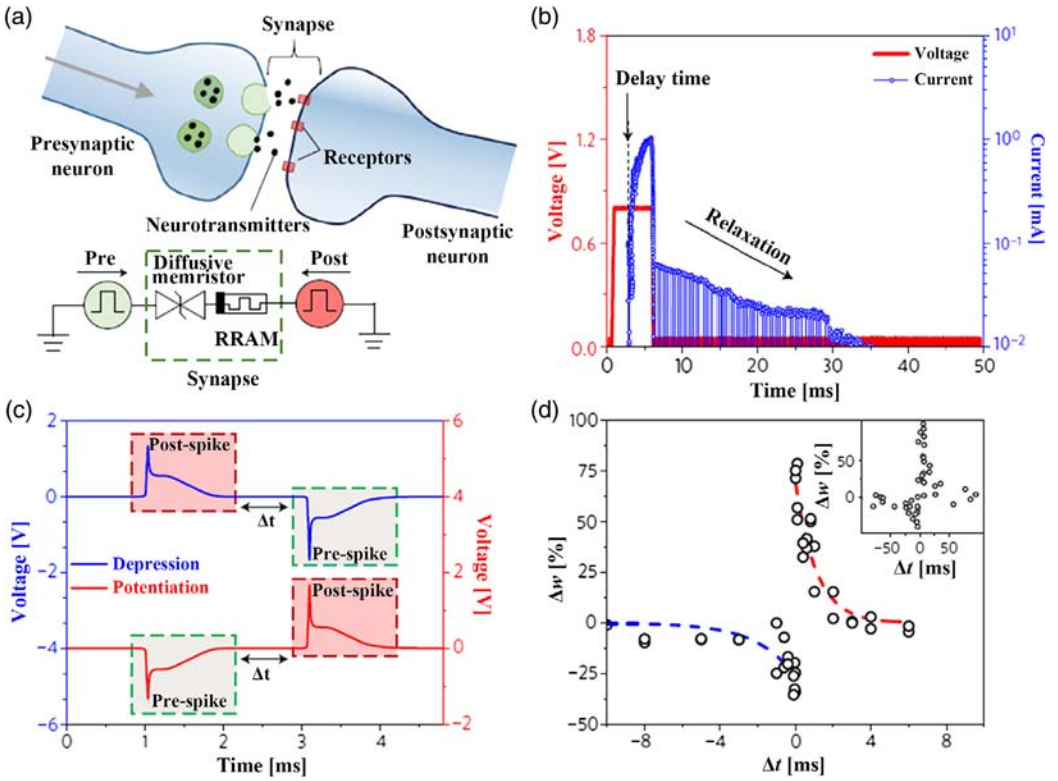


Figure 8. STDP realization with the synapse consisting of an RRAM in series with a diffusive memristor. a) Illustration of the biological synapse and the corresponding electronic emulator. b) Delay and relaxation characteristics of the diffusive memristor. The device requires a finite delay time to turn on and has a finite relaxation time before it transitions to the HRS. c) Schematic of the spikes for the STDP demonstration. The long low-voltage pulse in each spike turns on the diffusive memristor, whereas the short high-voltage pulse switches the RRAM. d) Conductance change of RRAM with Δt. The timing-dependent response of the electronic synapse is similar to that of biological synapses (inset). Reproduced with permission.[85] Copyright 2017, Springer Nature.

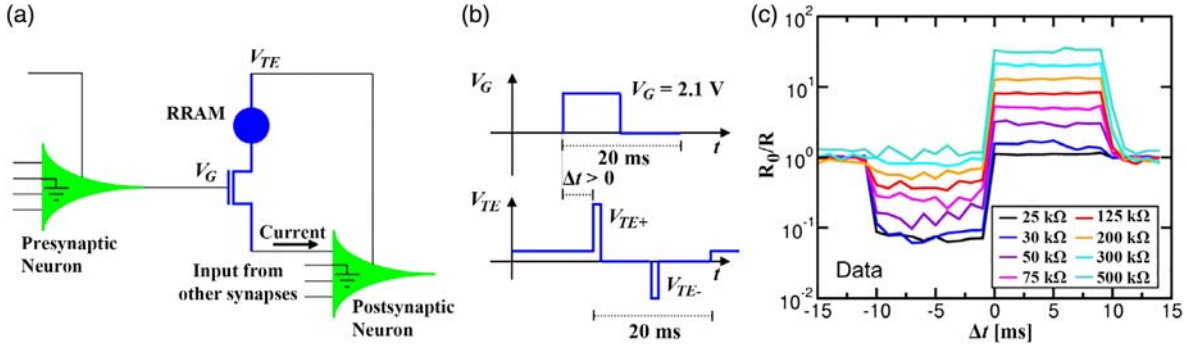


Figure 9. STDP demonstration with the 1T1R synapse. a) Schematic illustration of the 1T1R synapse and the connected neurons. b) Spikes for the STDP implementation. The spike on the gate induces a current, whereas the postsynaptic neuron collects the current and eventually fires. Upon firing, a positive voltage on the TE induces potentiation, whereas a negative voltage induces depression. c) STDP characteristics. The change in conductance is presented as a function of Δt. Reproduced with permission.[104] Copyright 2016, IEEE.

implemented with analog-type RRAM. DNNs and SNNs were demonstrated based on these two types of plasticity.

Analog-type RRAM is classified into filamentary RRAM and interfacial RRAM based on its switching mechanism. Interfacial RRAM was thought to be more suitable for emulating synapses due to its natural bidirectional analog switching behavior. However, this RRAM faces a trade-off among the switching speed, retention, and operation voltage. Filamentary RRAM, which showed abrupt SET in the early stages of research, has been further optimized as an option for the analog synapse. While the formation of a single strong CF leads to abrupt SET, the formation/rupture of multiple weak CFs, CF size

modulation, and channel composition modulation could lead to analog switching behavior. Carefully designed programming schemes have been proposed to enable gradual and linear conductance updating.

The device properties of analog-type RRAM remain to be improved for the development of a large-scale neuromorphic computing system. First, the desired switching behaviors, including analog switching, linearity and symmetry in weight updating, large on/off ratio, and sufficient intermediate resistance levels, should be further pursued. Second, the reliability issues of analog-type RRAM, such as the retention for analog weight storage and the endurance for gradual switching, require more attention.[92,105] Third, the bit-yield, read/write noise, and variations must be further controlled. Although neuromorphic computing to some extent tolerates variations,[27,57,76] variations still harm the learning accuracy and may lead to a time-consuming programming process.[106,107] Compared with the large number of works on system analysis, only a few works have focused on device-level optimization. Some optimization methods to suppress variations have been proposed, including doping/alloying of the switching layer to localize the V_O,[91] utilization of the dislocations of the switching layer to confine the CF,[61] and strengthening of fabrication process control.[108] We hope that this review article will provide some insights into the recent developments of and provide optimization guidelines for analog-type RRAM toward neuromorphic computing applications.

Acknowledgements

This work is supported in part by the NSFC (61874169, 61674087, 61674089, 61674092, 61076115), MOST of China (2016YFA0201801), and Beijing Innovation Center for Future Chips (ICFC).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

analog switching behavior, neuromorphic computing, resistive switching, synaptic devices

Received: April 7, 2019
Revised: June 7, 2019
Published online:

[1] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, *Nature* **2015**, *521*, 61.
[2] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong, H. Qian, *Nat. Commun.* **2017**, *8*, 15199.
[3] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Commun.* **2018**, *9*, 7.
[4] M. Hu, C. E. Graves, C. Li, Y. Li, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Q. Xia, J. P. Strachan, *Adv. Mater.* **2018**, *30*, 1.
[5] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. Di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, G. W. Burr, *Nature* **2018**, *558*, 60.
[6] Z. Wang, S. Joshi, S. Savel'Ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Electron.* **2018**, *1*, 137.
[7] Y. Wang, T. Tang, L. Xia, B. Li, P. Gu, H. Yang, H. Li, Y. Xie, in *Proc. 25th Edition of the Great Lakes Symposium on VLSI–GLSVLSI '15*, ACM Press, New York, USA **2015**, pp. 189–194.
[8] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, *Sci. Rep.* **2017**, *7*, 1.
[9] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J.-P. Han, C. Miller, K. Hosokawa, C. Lam, in *2015 IEEE Int. Electron Devices Meeting*, IEEE **2015**, pp. 17.1.1–17.1.4.
[10] Z. Du, R. Fasthuber, T. Chen, P. Ienne, L. Li, T. Luo, X. Feng, Y. Chen, O. Temam, in *Proc. 42nd Annual Int. Symp. Computer Architecture–ISCA '15*, ACM Press, New York, USA **2015**, pp. 92–104.
[11] H. S. P. Wong, S. Salahuddin, *Nat. Nanotechnol.* **2015**, *10*, 191.
[12] S. Yu, *Proc. IEEE* **2018**, *106*, 260.
[13] S. Han, X. Liu, H. Mao, J. Pu, A. Pedram, M. A. Horowitz, W. J. Dally, in *2016 ACM/IEEE 43rd Annual Int. Symp. Computer Architecture*, IEEE **2016**, pp. 243–254.
[14] C. Mead, *Proc. IEEE* **1990**, *78*, 1629.
[15] D. Monroe, *Commun. ACM* **2014**, *57*, 13.
[16] R. A. Nawrocki, R. M. Voyles, S. E. Shaheen, *IEEE Trans. Electron Devices* **2016**, *63*, 3819.
[17] Y. Lecun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[18] D. A. Drachman, *Neurology* **2005**, *64*, 2004.
[19] D. Kuzum, S. Yu, H.-S. Philip Wong, *Nanotechnology* **2013**, *24*, 382001.
[20] S. Yu, P.-Y. Chen, *IEEE Solid-State Circuits Mag.* **2016**, *8*, 43.
[21] H. S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, K. E. Goodson, *Proc. IEEE* **2010**, *98*, 2201.
[22] H. S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, M. Tsai, *Proc. IEEE* **2012**, *100*, 1951.
[23] A. Chen, *Solid-State Electron.* **2016**, *125*, 25.
[24] Y. J. Song, J. H. Lee, H. C. Shin, K. H. Lee, K. Suh, J. R. Kang, S. S. Pyo, H. T. Jung, S. H. Hwang, G. H. Koh, S. C. Oh, S. O. Park, J. K. Kim, J. C. Park, J. Kim, K. H. Hwang, G. T. Jeong, K. P. Lee, E. S. Jung, in *2016 IEEE Int. Electron Devices Meeting*, IEEE **2016**, pp. 27.2.1–27.2.4.
[25] M. K. Qureshi, V. Srinivasan, J. A. Rivers, *ACM SIGARCH Comput. Archit. News* **2009**, *37*, 24.
[26] S. Lequeux, J. Sampaio, V. Cros, K. Yakushiji, A. Fukushima, R. Matsumoto, H. Kubota, S. Yuasa, J. Grollier, *Sci. Rep.* **2016**, *6*, 1.
[27] G. W. Burr, R. M. Shelby, S. Sidler, C. Di Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, H. Hwang, *IEEE Trans. Electron Devices* **2015**, *62*, 3498.
[28] J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S. Shin, K.-T. Lee, P. Solomon, K. Chan, W. Haensch, J. Rozen, in *2018 IEEE Int. Electron Devices Meeting*, IEEE **2018**, pp. 13.1.1–13.1.4.
[29] M. Jerry, P.-Y. Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 6.2.1–6.2.4.
[30] J. Zhu, Y. Yang, R. Jia, Z. Liang, W. Zhu, Z. U. Rehman, L. Bao, X. Zhang, Y. Cai, L. Song, R. Huang, *Adv. Mater.* **2018**, *30*, 1.
[31] L. P. Romero, S. Ambrogio, M. Giordano, G. Cristiano, M. Bodini, P. Narayanan, H. Tsai, R. M. Shelby, G. W. Burr, *Faraday Discuss.* **2019**, *213*, 371.
[32] S. Pi, C. Li, H. Jiang, W. Xia, H. Xin, J. J. Yang, Q. Xia, *Nat. Nanotechnol.* **2019**, *14*, 35.
[33] B. Govoreanu, G. S. Kar, Y.-Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart,

O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, M. Jurczak, in *2011 Int. Electron Devices Meeting*, IEEE **2011**, pp. 31.6.1–31.6.4.
[34] E. Cha, J. Woo, D. Lee, S. Lee, J. Song, Y. Koo, J. Lee, C. G. Park, M. Y. Yang, K. Kamiya, K. Shiraishi, B. Magyari-Kope, Y. Nishi, H. Hwang, in *2013 IEEE Int. Electron Devices Meeting*, IEEE **2013**, pp. 10.5.1–10.5.4.
[35] C. Wang, H. Wu, B. Gao, W. Wu, L. Dai, X. Li, H. Qian, *Adv. Electron. Mater.* **2017**, *3*, 1700263.
[36] S.-S. Sheu, P.-C. Chiang, W.-P. Lin, H.-Y. Lee, P.-S. Chen, Y.-S. Chen, T.-Y. Wu, F. T. Chen, K.-L. Su, M.-J. Kao, K.-H. Cheng, M.-J. Tsai, in *2009 IEEE Symp. VLSI Technology*, IEEE **2009**, pp. 82–83.
[37] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, M.-J. Tsai, in *2008 IEEE Int. Electron Devices Meeting*, IEEE **2008**, pp. 1–4.
[38] Y. Wu, S. Yu, B. Lee, H. S. P. Wong, *J. Appl. Phys.* **2011**, *110*, 3.
[39] I.-T. Wang, Y.-C. Lin, Y.-F. Wang, C.-W. Hsu, T.-H. Hou, in *2014 IEEE Int. Electron Devices Meeting*, IEEE **2014**, pp. 28.5.1–28.5.4.
[40] L. Goux, A. Fantini, G. Kar, Y.-Y. Chen, N. Jossart, R. Degraeve, S. Clima, B. Govoreanu, G. Lorenzo, G. Pourtois, D. J. Wouters, J. A. Kittl, L. Altimime, M. Jurczak, in *2012 Symp. VLSI Technology*, IEEE **2012**, pp. 159–160.
[41] J. Zhou, F. Cai, Q. Wang, B. Chen, S. Gaba, W. D. Lu, *IEEE Electron Device Lett.* **2016**, *37*, 404.
[42] T. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma, C. Chen, M. Gupta, G. Hilton, A. Kathuria, V. Lai, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, X. Wu, Y. Yin, N. Nagel, Y. Tanaka, M. Higashitani, T. Minvielle, C. Gorla, T. Tsukamoto, T. Yamaguchi, M. Okajima, T. Okamura, S. Takase, H. Inoue, L. Fasoli, *IEEE J. Solid-State Circuits* **2014**, *49*, 140.
[43] S. Zuloaga, R. Liu, P.-Y. Chen, S. Yu, in *2015 IEEE Int. Symp. Circuits System*, IEEE **2015**, pp. 193–196.
[44] Q. Wang, W.-T. Lian, Y. Wang, H.-B. Lü, S. Liu, Y.-T. Li, S.-B. Long, Q. Liu, M. Liu, S. Zhang, *Chin. Phys. B* **2011**, *20*, 017305.
[45] H. Lv, X. Xu, P. Yuan, D. Dong, T. Gong, J. Liu, Z. Yu, P. Huang, K. Zhang, C. Huo, C. Chen, Y. Xie, Q. Luo, S. Long, Q. Liu, J. Kang, D. Yang, S. Yin, S. Chiu, M. Liu, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 2.4.1–2.4.4.
[46] C. Gopalan, Y. Ma, T. Gallo, J. Wang, E. Runnion, J. Saenz, F. Koushan, P. Blanchard, S. Hollmer, *Solid State Electron.* **2011**, *58*, 54.
[47] Y. Zhang, N. Deng, H. Wu, Z. Yu, J. Zhang, H. Qian, *Appl. Phys. Lett.* **2014**, *105*, 063508.
[48] S.-Y. Wang, C.-W. Huang, D.-Y. Lee, T.-Y. Tseng, T.-C. Chang, *J. Appl. Phys.* **2010**, *108*, 114110.
[49] M.-C. Wu, Y.-W. Lin, W.-Y. Jang, C.-H. Lin, T.-Y. Tseng, *IEEE Electron Device Lett.* **2011**, *32*, 1026.
[50] P. Bousoulas, S. Stathopoulos, D. Tsialoukis, D. Tsoukalas, *IEEE Electron Device Lett.* **2016**, *37*, 874.
[51] Q. Wang, W. Wang, J. Yang, Q. Zuo, Q. Liu, M. Liu, S. Zhang, S. Long, M. Zhang, Y. Wang, Y. Li, *Nanotechnology* **2009**, *21*, 045202.
[52] V. Y.-Q. Zhuo, Y. Jiang, R. Zhao, L. P. Shi, Y. Yang, T. C. Chong, J. Robertson, *IEEE Electron Device Lett.* **2013**, *34*, 1130.
[53] Y. Wu, S. Yu, H.-S. P. Wong, Y.-S. Chen, H.-Y. Lee, S.-M. Wang, P.-Y. Gu, F. Chen, M.-J. Tsai, in *2012 4th IEEE Int. Memory Workshop*, IEEE **2012**, pp. 1–4.
[54] B. Gao, Y. Bi, H.-Y. Chen, R. Liu, P. Huang, B. Chen, L. Liu, X. Liu, S. Yu, H.-S. P. Wong, J. Kang, *ACS Nano* **2014**, *8*, 6998.
[55] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, G. Bersuker, *IEEE Trans. Electron Devices* **2015**, *62*, 1998.
[56] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, H. S. P. Wong, *IEEE Trans. Electron Devices* **2011**, *58*, 2729.
[57] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, H. S. P. Wong, *Adv. Mater.* **2013**, *25*, 1774.
[58] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, W. Lu, *Nano Lett.* **2010**, *10*, 1297.
[59] H. Wu, W. Wu, S. Yu, H. Qian, B. Gao, N. Deng, *IEEE Electron Device Lett.* **2017**, *38*, 1019.
[60] K. Moon, A. Fumarola, S. Sidler, J. Jang, P. Narayanan, R. M. Shelby, G. W. Burr, H. Hwang, *IEEE J. Electron Devices Soc.* **2018**, *6*, 146.
[61] S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P. Y. Chen, H. Yeon, S. Yu, J. Kim, *Nat. Mater.* **2018**, *17*, 335.
[62] J. Woo, A. Padovani, K. Moon, M. Kwak, L. Larcher, H. Hwang, *IEEE Electron Device Lett.* **2017**, *38*, 1220.
[63] S. Park, A. Sheri, J. Kim, J. Noh, J. Jang, M. Jeon, B. Lee, B. R. Lee, B. H. Lee, H. Hwang, in *2013 IEEE Int. Electron Devices Meeting*, IEEE **2013**, pp. 25.6.1–25.6.4.
[64] S. Yu, X. Guan, H. S. P. Wong, *IEEE Trans. Electron Devices* **2012**, *59*, 1183.
[65] P. Pavan, A. Padovani, L. Larcher, F. M. Puglisi, G. Bersuker, *IEEE Electron Device Lett.* **2013**, *34*, 387.
[66] P. Huang, X. Y. Liu, B. Chen, H. T. Li, Y. J. Wang, Y. X. Deng, K. L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang, J. Kang, *IEEE Trans. Electron Devices* **2013**, *60*, 4090.
[67] S. R. Lee, S. S.-J. Kim, U.-I. Chung, S. S.-J. Kim, C. J. Kim, Y.-B. Kim, E. Cho, M. Chang, K. M. Kim, I.-K. Yoo, *Sci. Rep.* **2013**, *3*, 1.
[68] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, *Nature* **2008**, *453*, 80.
[69] R. Waser, R. Dittmann, G. Staikov, K. Szot, *Adv. Mater.* **2009**, *21*, 2632.
[70] X. Li, H. Wu, Bin Gao, W. Wu, D. Wu, N. Deng, J. Cai, H. Qian, *Nanotechnology* **2016**, *27*, 305201.
[71] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, H. Hwang, *Nanotechnology* **2011**, *22*, 254023.
[72] T. H. Hou, C.-T. Chou, C.-W. Hsu, W.-L. Lai, Y.-F. Wang, C.-C. Wan, I.-T. Wang, Y.-J. Lee, *Nanotechnology* **2014**, *25*, 165202.
[73] M. Hansen, M. Ziegler, L. Kolberg, R. Soni, S. Dirkmann, T. Mussenbrock, H. Kohlstedt, *Sci. Rep.* **2015**, *5*, 1.
[74] B. Gao, H. Wu, J. Kang, H. Yu, H. Qian, in *2016 IEEE Int. Electron Devices Meeting*, IEEE **2016**, pp. 7.3.1–7.3.4.
[75] A. K. McAllister, L. C. Katz, D. C. Lo, *Annu. Rev. Neurosci.* **1999**, *22*, 295.
[76] H. Wu, P. Yao, B. Gao, W. Wu, Q. Zhang, W. Zhang, N. Deng, D. Wu, H.-S. P. Wong, S. Yu, H. Qian, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 11.5.1–11.5.4.
[77] P.-Y. Chen, S. Yu, *IEEE Des. Test* **2018**, *PP*, 1.
[78] J. Woo, S. Yu, *IEEE Nanotechnol. Mag.* **2018**, *12*, 36.
[79] G. Bi, M. Poo, *J. Neurosci.* **1998**, *18*, 10464.
[80] A. Suvrathan, *Curr. Opin. Neurobiol.* **2019**, *54*, 12.
[81] C. Zamarreño-Ramos, L. A. Camuñas-Mesa, J. A. Pérez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, B. Linares-Barranco, *Front. Neurosci.* **2011**, *5*, 1.
[82] U. Ganguly, B. Das, S. Lashkare, N. Panwar, P. Kumbhare, *IEEE Electron Device Lett.* **2017**, *38*, 1212.
[83] N. Panwar, B. Rajendran, U. Ganguly, *IEEE Electron Device Lett.* **2017**, *38*, 740.
[84] Y. Li, Y. Zhong, L. Xu, J. Zhang, X. Xu, H. Sun, X. Miao, *Sci. Rep.* **2013**, *3*, 1.
[85] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G. L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Mater.* **2017**, *16*, 101.
[86] C. Du, W. Ma, T. Chang, P. Sheridan, W. D. Lu, *Adv. Funct. Mater.* **2015**, *25*, 4290.
[87] S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W. D. Lu, *Nano Lett.* **2015**, *15*, 2203.
[88] Y.-F. Wang, Y.-C. Lin, I.-T. Wang, T.-P. Lin, T. H. Hou, *Sci. Rep.* **2015**, *5*, 10150.

[89] B. Gao, H. Wu, W. Wu, X. Wang, P. Yao, Y. Xi, W. Zhang, N. Deng, P. Huang, X. Liu, J. Kang, H.-Y. Chen, S. Yu, H. Qian, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 4.4.1–4.4.4.
[90] B. Gao, B. Chen, F. Zhang, P. Huang, L. Liu, X. Liu, J. Kang, *Solid State Commun.* **2015**, *205*, 51.
[91] W. Wu, H. Wu, B. Gao, N. Deng, H. Qian, *J. Appl. Phys.* **2018**, *124*, 152108.
[92] M. Zhao, H. Wu, B. Gao, Q. Zhang, W. Wu, S. Wang, Y. Xi, D. Wu, N. Deng, S. Yu, H.-Y. Chen, H. Qian, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 39.4.1–39.4.4.
[93] S. Kim, S. Choi, W. Lu, *ACS Nano* **2014**, *8*, 2369.
[94] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, H. Hwang, *IEEE Electron Device Lett.* **2016**, *37*, 994.
[95] F. Miao, W. Yi, I. Goldfarb, J. J. Yang, M. X. Zhang, M. D. Pickett, J. P. Strachan, G. Medeiros-Ribeiro, R. S. Williams, *ACS Nano* **2012**, *6*, 2312.
[96] F. Miao, J. P. Strachan, J. J. Yang, M. X. Zhang, I. Goldfarb, A. C. Torrezan, P. Eschbach, R. D. Kelley, G. Medeiros-Ribeiro, R. S. Williams, *Adv. Mater.* **2011**, *23*, 5633.
[97] H. Jiang, L. Han, P. Lin, Z. Wang, M. H. Jang, Q. Wu, M. Barnell, J. J. Yang, H. L. Xin, Q. Xia, *Sci. Rep.* **2016**, *6*, 1.
[98] S. Park, H. Kim, M. Choo, J. Noh, A. Sheri, S. Jung, K. Seo, J. Park, S. Kim, W. Lee, J. Shin, D. Lee, G. Choi, J. Woo, E. Cha, J. Jang, C. Park, M. Jeon, B. Lee, B. H. Lee, H. Hwang, in *2012 Int. Electron Devices Meeting*, IEEE **2012**, pp. 10.2.1–10.2.4.
[99] S. Park, J. Noh, M. Choo, A. M. Sheri, M. Chang, Y.-B. Kim, C. J. Kim, M. Jeon, B.-G. Lee, B. H. Lee, H. Hwang, *Nanotechnology* **2013**, *24*, 384009.
[100] K. Moon, E. Cha, J. Park, S. Gi, M. Chu, K. Baek, B. Lee, S. Oh, H. Hwang, in *2015 IEEE Int. Electron Devices Meeting*, IEEE **2015**, pp. 17.6.1–17.6.4.
[101] J. Woo, K. Moon, J. Song, M. Kwak, J. Park, H. Hwang, *IEEE Trans. Electron Devices* **2016**, *63*, 5064.
[102] J. Park, M. Kwak, K. Moon, J. Woo, D. Lee, H. Hwang, *IEEE Electron Device Lett.* **2016**, *37*, 1559.
[103] W. Wu, H. Wu, B. Gao, P. Yao, X. Zhang, X. Peng, S. Yu, H. Qian, in *2018 IEEE Symp. VLSI Technology*, IEEE **2018**, pp. 103–104.
[104] S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z. Q. Wang, A. Calderoni, N. Ramaswamy, D. Ielmini, *IEEE Trans. Electron Devices* **2016**, *63*, 1508.
[105] M. Zhao, H. Wu, B. Gao, X. Sun, Y. Liu, P. Yao, Y. Xi, X. Li, Q. Zhang, K. Wang, S. Yu, H. Qian, in *2018 IEEE Int. Electron Devices Meeting*, IEEE **2018**, pp. 20.2.1–20.2.4.
[106] P.-Y. Chen, X. Peng, S. Yu, in *2017 IEEE Int. Electron Devices Meeting*, IEEE **2017**, pp. 6.1.1–6.1.4.
[107] G. C. Adam, A. Khiat, T. Prodromakis, *Nat. Commun.* **2018**, *9*, 5267.
[108] Y. Fang, Z. Yu, Z. Wang, T. Zhang, Y. Yang, Y. Cai, R. Huang, *IEEE Electron Device Lett.* **2018**, *39*, 819.
[109] Y. Lin, H. Wu, B. Gao, P. Yao, W. Wu, Q. Zhang, X. Zhang, X. Li, F. Li, J. Lu, G. Li, S. Yu, H. Qian, in *2018 IEEE Int. Electron Devices Meeting*, IEEE **2018**, pp. 3.4.1–3.4.4.
[110] Z. Chen, B. Gao, Z. Zhou, P. Huang, H. Li, W. Ma, D. Zhu, L. Liu, X. Liu, J. Kang, H.-Y. Chen, in *2015 IEEE Int. Electron Devices Meeting*, IEEE **2015**, pp. 17.7.1–17.7.4.
[111] L. Gao, I.-T. Wang, P.-Y. Chen, S. Vrudhula, J. Seo, Y. Cao, T.-H. Hou, S. Yu, *Nanotechnology* **2015**, *26*, 455204.