Brain-inspired computing via memory device physics

D. Ielmini[1], Z. Wang[2], Y. Liu[2]
[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano and IU.NET, Piazza L. da Vinci 32, 20133 Milano, Italy.
[2]Key Laboratory for UV Light-Emitting Materials and Technology (Northeast Normal University), Ministry of Education, Renmin Street, 5268 Changchun, China.
Email: daniele.ielmini@polimi.it; wangzq752@nenu.edu.cn; ycliu@nenu.edu.cn

Abstract–In our brain, information is exchanged among neurons in the form of spikes where both the space (which neuron fires) and time (when the neuron fires) contain relevant information. Every neuron is connected to other neurons by synapses, which are continuously created, updated and stimulated to enable information processing and learning. Realizing the brain-like neuron/synapse network in silicon would enable artificial autonomous agents capable of learning, adaptation and interaction with the environment. Toward this aim the conventional microelectronic technology, which is based on complementary metal-oxide semiconductor (CMOS) transistors and the von Neumann computing architecture, does not provide the desired energy efficiency and scaling potential. A generation of emerging memory devices, including resistive switching random access memory (RRAM) also known as the memristor, can offer a wealth of physics-enabled processing capability, including multiplication, integration, potentiation, depression and time-decaying stimulation, which are suitable to recreate some of the fundamental phenomena of the human brain in silico. This work provides an overview about the status and the most recent updates on brain-inspired neuromorphic computing devices. After introducing the RRAM device technologies, we discuss the main computing functionalities of the human brain, including neuron integration&fire, dendritic filtering, short- and long-term synaptic plasticity. For each of these processing function we discuss their proposed implementation in terms of materials, device structure and brain-like characteristics. The rich device physics, the nano-scale integration, the tolerance to stochastic variations and the ability to process information in-situ make the emerging memory devices a promising technology for future brain-like hardware intelligence.

1. Introduction

The human brain is one of the most complex objects in the universe. It is capable of executing high-level cognitive tasks, such as abstraction, generalization, prediction, decision making, recognition and navigation in a continuously changing environment. Such high cognitive capability of the brain comes at the expense of an extremely low power consumption of only 20 W. There are mainly two reasons for the high energy efficiency of the brain: first, information exchange and processing are event driven, therefore spiking energy is consumed only when and where it is needed. Second, neurons and synapses are co-located within the same, highly interconnected neural network, where each neuron is connected to other 10^4 neurons, on the average. Neuron/synapse co-location means that data processing, consisting of synaptic excitation and neuron firing, and memory, consisting of the synaptic weight and the neuron threshold, share the same location within the brain.[1]

Many research efforts aim at mimicking the type of computation of the human brain, to achieve its outstanding energy efficiency. This is the objective of neuromorphic engineering, where spiking neural networks (SNNs) are developed with artificial neurons and synapses. SNNs generally adopt the same fully connected (FC) architecture of the conventional perceptron networks pioneered by Rosenblatt and Minsky.[2,3] In a SNN, however, neurons and synapses usually display a time-dependent response to the applied spikes, such as integration and fire in a neuron and excitatory

post-synaptic current (EPSC) across a synapse. This is different from the conventional artificial neural networks (ANNs) used in artificial intelligence (AI) accelerators for computer vision and speech recognition, where the information is synchronous and based on the amplitude of the signal, instead of its time.[4]

Most SNNs generally relies on the complementary metal-oxide-semiconductor (CMOS) technology, with two main significant advantages: First, the CMOS technology is widely available in the semiconductor industry ecosystem, including design, fabrication, and qualification, therefore creating the conditions to make CMOS-based neuromorphic engineering a mature topic. Second, the CMOS transistor can scale down according to the Moore's law, where a reduction of the lithography feature size allows for a larger density and a better performance of the circuit. On the other hand, there are significant limitations in CMOS technology. For instance, time-dependent functions such as spike integration in an artificial neuron generally requires large capacitors in CMOS technology, therefore limiting the cost effectiveness of neuromorphic circuits.[5] Synaptic weights are generally stored in static random access memory (SRAM), which are volatile, i.e., all synaptic values are lost when the circuit is switched off.[6] In addition, SRAM devices are large and binary, i.e., they can only store 0 and 1, thus they are not suitable for gradual potentiation and depression which are typical of synaptic plasticity phenomena.[7-9]

To overcome these limitations, neuromorphic materials and devices are intensively explored to complement CMOS technology. The aim of this new wave of research is to reproduce bio-neurological phenomena typical of the human brain with device physics. For instance, phase change materials have been shown to accumulate applied voltage spikes and consequently change their resistivity, which can be used as the physical mechanism for integrate-and-fire (I&F) neurons without capacitors.[10-12] The fire process of the typical I&F neuron can be reproduced in a nanoelectronics device by abrupt current switching at the onset of the negative differential resistance (NDR) region, such as the electronic threshold switching in ovonic threshold switch (OTS) elements[13] or ferroelectric transition in HfO_2.[14] Similarly, all other key mechanisms in the biological neural network can be emulated by specifically-engineered devices through their physics. The objective is the recreation of a brain-like circuit system with extremely low power consumption and compact, scalable architecture.

This work provides an overview about the status on the development of neuromorphic devices that emulate biological neural processes by device physics. The work will focus on the resistive switching random access memory (RRAM) as the device technology for the implementation of various neuromorphic functions, including artificial synapse, neuron and dendrite. Circuits demonstrating the full neuromorphic function, such as unsupervised learning and pattern recognition, will also be presented. The rest of the paper is organized as follows: Sec. 2 will illustrate the major categories of RRAM devices in terms of switching mechanism and device structure. Sec. 3 will provide an overview of the neuromorphic processes and their implementation in RRAM devices. Sec. 4 will deal with artificial neurons with integration, fire, oscillations and dendritic filtering capability. Finally, Sec. 5 will focus on artificial synapses including learning functions via plasticity and sensing/computation via short-term memory.

2. Resistive switching devices

Urged by the scaling limitation of CMOS-based memories, various types of emerging memory devices have been proposed in the last 20 years. These include phase change memory (PCM),[15] magnetic random-access memories,[16, 17] ferroelectric random-access memory (FERAM)[18, 19] and RRAM.[20-22] These memories have the ability to change their resistance state by a permanent

modification of the active material, thus serving as scalable nonvolatile memories for standalone and embedded memories.[23-25]

Among the novel emerging memory technologies, RRAM has attracted strong research interest partly due to the simple structure that allows for a relatively straightforward fabrication in academic labs and integration within industrial CMOS process. RRAM has been recognized as a potential technology for synaptic connections in ANNs and SNNs, thanks to the small size, easy integration and scalability that allows for high connectivity within the neural network.[26] The high synaptic density is further supported by the ability of 3D integration by array stacking[27] of vertical structures.[28-32] The programming energy of RRAM is generally low thus enabling energy-efficient computation and reconfiguration of the neural network.[31]

2.1 2-terminal devices

Fig. 1 shows a schematic illustration of 2-terminal RRAM device, including a filamentary switching RRAM (a) and a uniform switching RRAM (b). While both devices are based on a metal-insulator-metal (MIM) structure with a top electrode (TE), a bottom electrode (BE) and at least one dielectric layer, the switching mechanisms is fundamentally different. In the filamentary structure, a forming process is first applied, by applying a relatively large voltage that leads to soft breakdown of the MIM.[22] The breakdown spot, consisting of a filamentary path with low resistivity, is then subjected to set/reset processes by the application of voltage pulses. Typically, the RRAM device shows a bipolar switching characteristic, where the applied electric field across the conductive filament causes ionic migration and a consequent change of resistance.[33] For instance, a negative voltage applied to the TE leads to the migration of positively-ionized defects toward the TE, thus resulting in depletion of defects at the BE side with an increase of resistance, or reset transition.[34] A positive voltage applied to the TE results in migration of the defects toward the BE, thus refilling the depleted gap and causing a decrease of resistance, or set transition.[34] Filamentary set transition is generally abrupt due to the positive feedback in the gap filling process: As the defects start migrate toward the gap, the electric field increases, thus causing an acceleration of the ionic migration. To avoid uncontrolled filament growth during the abrupt set transition, usually a transistor is added in series with the RRAM device to enable current limitation below a certain compliance current I_C.[35] Fig. 2a shows typical I-V characteristics for a HfO_2 RRAM device with 1T1R structure.[36] As I_C increases, the device conductance in the low resistance state (LRS) increases thus indicating a larger size of the conductive filament. The reset current correspondingly increases, as a result of the larger filament size. The adoption of the one-transistor/one-resistor (1T1R) structure of the RRAM device thus allows for low current operation and tight control of the device conductance, which is beneficial for analogue in-memory computing.[36]

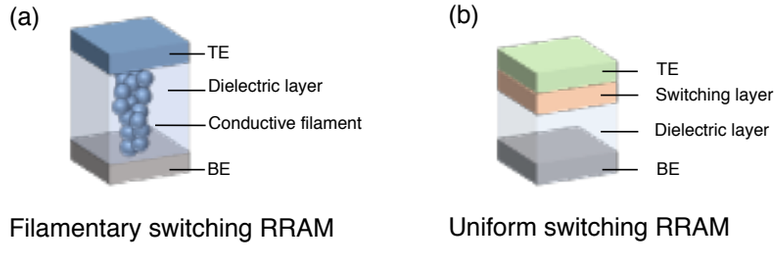

Fig. 1 Illustration of 2-terminal resistive switching memory (RRAM) devices for neuromorphic computing. (a) Filamentary RRAM, where the device resistance is changes due to the formation and modulation of a conductive filament across a high resistance dielectric layer. The filament can connect the top electrode (TE) and the bottom electrode (BE) in the low resistance state (LRS), whereas the filament is disconnected between TE and BE in the high

resistance state (HRS). (b) Uniform switching RRAM, where the device resistance is controlled by a switching layer, usually a metal oxide, which shows a high resistance, due to a low concentration of defects, in the HRS, or a low resistance, due to high concentration of defects, in the LRS.

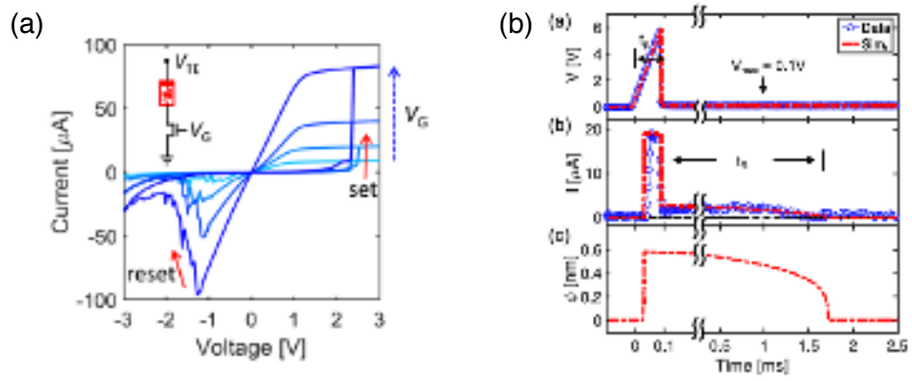

Fig. 2 Filamentary RRAM characteristics. (a) Measured I-V characteristics of a filamentary RRAM device with 1T1R structure and HfO_2 switching layer. As the compliance current IC increases, the filament conductance increases due to the increased size. (b) Pulsed characteristic of a volatile RRAM device, including applied voltage (top), response current (center) and calculated filament diameter (bottom). After the set transition at time 0, the filament spontaneously retracts to the electrodes, thus resulting in a fast drop of conductance within a retention time t_R of about 1.5 ms. Reprinted with permission from Sun et al. IEEE Trans. Electron Devices 67, 1466 (2020) [36] and Wang et al. IEEE Trans. Electron Devices 66, 3802 (2019). [44] Copyright IEEE (2019, 2020).

Depending on the electrode materials, the conductive filament can be stable for long time even at high temperature,[37] or be metastable due to defect diffusion after the set transition.[38-44] In particular, RRAM with Ag TE generally tend to display this type of volatile behavior due to the spontaneous diffusion of Ag from the filament location. This was attributed to surface diffusion of Ag to minimize the total energy of the filament by minimizing the surface to volume ratio.[45] Fig. 2b shows the typical pulsed programming characteristics for Ag/SiO_2 RRAM device.[44] Under a triangular pulsed of applied voltage, the device shows a set transition, marked by the abrupt rise of current to the I_C level. After the pulse, the read current remains active for a finite retention time t_R of about 1.5 ms, thus revealing the spontaneous decay of the conductive filament diameter ϕ (see simulation results at the bottom of Fig. 2b).[44] Such a volatile behavior has been proposed for selector elements in crosspoint device,[46] thanks to the steep switching slope and extremely large on/off ratio exceeding 10 orders of magnitude.[42] However, due to the relatively long retention time in the range between 1 μs and several ms, the device is most suitable as a physics-based neuromorphic device to implement transient biological phenomena, such as short-term memory [47] and spike-timing dependent plasticity.[41]

The filamentary set/reset process causes intrinsic variability issues due to individual defect diffusion and instability.[48-51] Variations include cycle-to-cycle changes of conductance, due to the variability in filament shape and volume,[48] and device-to-device variations due to the difference in structure and geometry among various RRAM devices.[49] Generally, device-to-device variation plays the key role in technology reliability, due to the sensitivity to local defect concentration, dielectric film microstructure, interface roughness and filament shape originating from the breakdown event at forming.[50] In addition to programming variations, read variation causes the device resistance to vary even after the device has undergone the set/reset process. The resistance can in fact display time-dependent fluctuations such as random telegraph noise (RTN) and random walk due to defect instability.[51] The conductance variations can cause a degradation

of neural network accuracy,[52, 53] although some stochastic computing algorithms may take advantage from noise.[54-59]

Note that variations are not intrinsic to filamentary switching, rather they arise generally in most types of memory technologies. For instance, PCM displays programming variations due to the stochastic nature of nucleation and growth in the crystallization process.[60] Similarly, FERAM shows variations in the multilevel conductance due to the stochastic switching of individual ferroelectric domains.[61] However, filamentary RRAM is more critically affected by post-programming fluctuations of the resistance, as a result of the localized conduction at atomically-thin channels, where trapping, detrapping and atomic relaxation can induce a strong variation in the device resistance.[62] To mitigate the cycle-to-cycle variations, a new concept of filamentary switching RRAM was developed, where the conductive path originates from threading dislocations within epitaxially-grown SiGe layers on Si substrates.[63] The materials-based approach to induce switching at predetermined channels is extremely promising for reducing the programming variations, although post-programming variations at the dislocation filament may still be a concern for reliability.

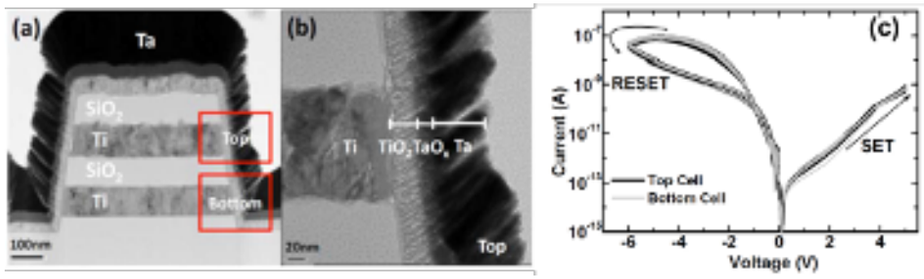

Fig. 3 Uniform switching RRAM structure and characteristics. (a,b) Device structure, including vertical Ta TE, TaO_x/TiO_2 stack as switching/dielectric layers and horizontal Ti BE. The TE and oxide bi-layer are deposited on the side wall of a stack of multiple BE/SiO_2 layers for a cost-effective vertical RRAM structure.[30] (c) I-V curves of uniform switching for top and bottom cells.[30] Reprinted with permission from Hsu et al., IEEE Int. Electron Devices Meeting (IEDM) (2013), pp. 10.4.1-10.4.4.[30] Copyright IEEE (2013).

The conductance variations and their impact on the neural network accuracy can be mitigated by the uniform switching RRAM in Fig. 1b. The conductance in this device changes as a result of oxygen vacancy exchange at the interface between two oxide layers, the dielectric layer and the switching layer.[64] For instance, the switching layer can consist of an interfacial oxide layer between an active electrode, e.g., Sm, and a relatively-high conductive oxide layer, such as $La_{0.7}Ca_{0.3}MnO_3$ (LCMO).[65] Fig. 3 shows a possible implementation of a uniform switching device with vertical structure.[30] The device stack includes a Ta TE, TaO_x, TiO_2 and Ti BE. The bipolar switching takes place by the oxygen exchange between the TaO_x and TiO_2 layers. Fig. 3c shows the I-V curves of the uniform switching device, indicating a smooth and gradual change of resistance.[30] Thanks to the gradual set/reset dynamics, the uniform switching is suitable to perform pulsed potentiation/depression for analogue artificial synapses.[30,66] Also, the low conductance around 100 nS in uniform switching allows for an extremely low energy per spike below 10 fA.[31]

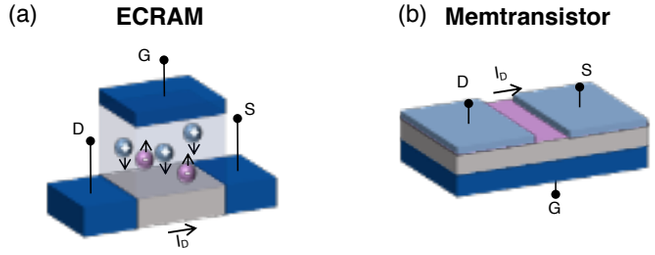

Fig. 4 Illustration of 3-terminal devices. (a) ECRAM, where the channel conductance changes by the field-induced migration of ionized defects, such as Li^+ or oxygen vacancies. (f) Mem-transistor, where the conductance is controlled by the migration of defects across a 2D semiconductor channel. Reprinted with permission from.[67] Copyright IEEE (2013), IOP (2016).

2.2 3-terminal devices

The need for analogue conductance, low variation and low energy in neuromorphic circuits has stimulated the study of advanced 3-terminal devices based on ionic migration. Fig. 4a shows a 3-terminal device called electro-chemical random access memory (ECRAM).[67] The ECRAM displays a transistor structure with gate, source and drain contacts, where the read path is from source to drain, while the programming takes place by gate pulses. Application of positive/negative gate pulses results in the migration of ionized defects from a reservoir, close to the gate terminal, to the channel between source and drain. Defects can be either Li^+ impurities,[68-70] H^+ [71, 72] or oxygen ions/vacancies.[73, 74] Li^+ intercalation and oxygen exchange within the channel can change its conductivity, thus resulting in weight potentiation or depression. ECRAM devices, also referred to as redox transistors or ionic transistors, are characterized by extremely low conductance in the range from few nS [69] to few μS, [73] thanks to the low mobility and low carrier concentration in the channel material, e.g., WO_3. Such a low conductance is essential to minimize the signal current within the synaptic array, thus enabling low parasitic IR drop[67] and small size of the circuit periphery to handle the output current, including select transistors and integrating capacitances. Most importantly, the potentiation and depression characteristics are extremely gradual and linear, thanks to the bulk conduction mechanism in the device [74] and for accurate integration in I&F neurons. On the other hand, the ECRAM technology usually requires selector devices to properly execute program and read operations.[72]

Fig. 4b shows the mem-transistor device, where the channel consists of a 2D semiconductor region with atomic thickness.[75, 76] The drain current of the device can be modulated by applying a suitable gate voltage, thanks to the semiconductor properties of transition metal dichalcogenides (TMDs) such as MoS_2.[77] In addition, the application of a large drain bias can lead to a persistent modification of the channel conductivity due to migration of defects such as grain boundaries[75, 76] or Li^+ impurities.[78] The mem-transistor thus allows in principle both transistor effect (by gate stimulation) and memory effect (by drain stimulation), which can support various neuromorphic functions, such as synaptic potentiation/depression and spike dependent plasticity.[75]

3. Neuromorphic processes by device physics

Memory devices allow to embody neurobiological processes within a single device with extremely compact size and highly bio-realistic properties. This is made possible by the rich physics of the emerging memory devices, where electric/magnetic polarization, phase structure and local chemical composition contribute to the electrical conductance, which is in turn affected by

atomic/ionic drift/diffusion, electro-chemical redox reactions, phase transitions, dielectric breakdown phenomena, and ferroelectric/ferromagnetic transitions.

Fig. 5 shows a summary of neurobiological features and functions and their respective implementation in resistive memory devices. Generally, neurons in the biological neural network of the human brain consists of soma, dendrites and an axon. The temporal spikes containing the incoming information are collected by dendrites and processed by the soma. Depending on the incoming stimulation and the type of information processing, e.g., I&F with a characteristic threshold, the neuron can fire, i.e., send an output spike through the axon toward the receiving neurons. The spike transmission from a neuron axon to other neuron dendrites takes place via a synapse, called axo-dendritic synapse, each having a specific weight and a corresponding weight update behavior. The synaptic weight describes the efficacy of an input spike to stimulate the receiving neuron. Synapses display synaptic plasticity, namely the ability to change their weight in response to the stimulation. Although the synaptic plasticity mechanism is not yet fully understood, several plasticity rules have been proposed, including spike-timing dependent plasticity (STDP)[7, 9, 79-81] and triplet-based plasticity,[8, 82, 83] where the timing of spikes, e.g., their respective delay or relative frequency, dictates the potentiation or depression of the synapse. Synaptic plasticity controls learning within the human brain, thus it is of utmost importance in all neuromorphic circuits.

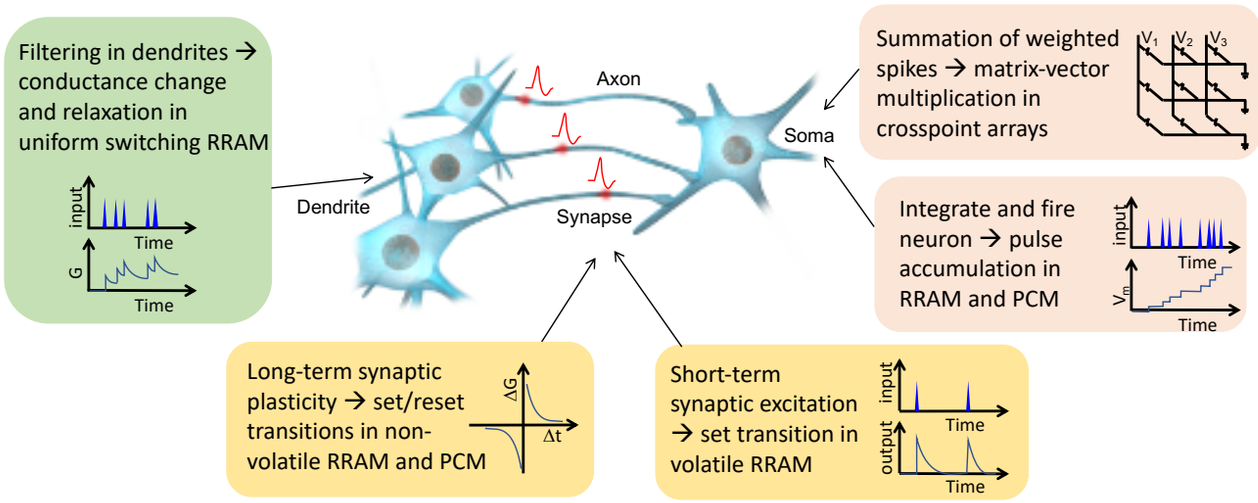

Fig. 5 Illustration of various possible circuit/device implementations of neuro-biological processes. The neuron soma weighted summation can be reproduced by the matrix vector multiplication (MVM) in crosspoint array circuits, while integration of I&F neurons is mimicked by pulse accumulation mechanisms in PCMs and RRAMs. The filtering function of dendrites is described by the conductance change and relaxation of uniform switching RRAMs. Short- and long-term plasticity of biological synapses can be implemented by set/reset dynamics of volatile/nonvolatile RRAMs or PCMs.

To implement the individual elements of Fig. 5, several devices, circuits and their respective physics can be adopted. The summation and integration functions of the soma can be implemented in hardware by matrix vector multiplication (MVM) in crosspoint arrays and integration in nanoscale memory devices. Time-dependent dendrite filtering and synaptic plasticity effects can be described by the switching properties of RRAM devices. The rich physics of memory devices and their combination can thus be used to reproduce neuro-biological phenomena at the nanoscale, which benefits the massive connectivity, high scalability and low-cost of neuromorphic circuits.

4. Hardware neurons

The neuron soma can be described by the popular McCulloch-Pitts model, [84] where the neuron input is given by the weighted summation of the incoming spike, while the output signal is given by a suitable nonlinear activation function. This can be expressed by the formula:

$$y_i = f\left(\sum_j w_{ij} x_j\right), \quad (1)$$

where y_i is the output of post-synaptic neuron *i*, *f* is the activation function, w_{ij} are the weights of the synapses connecting presynaptic neurons *j* with postsynaptic neuron *i*, and x_j is the signal of pre-synaptic neuron *j*. While other models, such as the Hodgkin-Huxley (HH) model, [85] are more accurate in the description of the temporal shape of the spike and the bio-chemical details of Ca and K ion transport, the McCulloch-Pitts model provides a simple mathematical description to elaborate the interaction between presynaptic and postsynaptic neurons.

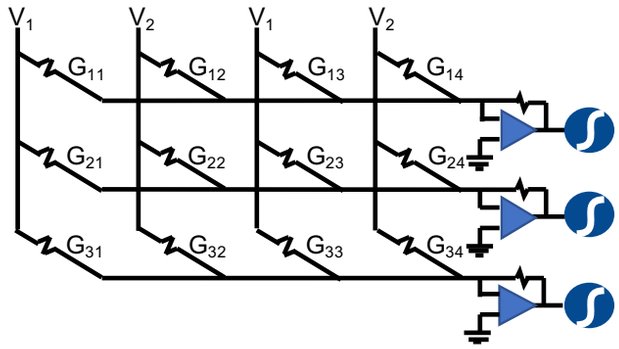

Fig. 6 Illustration of a crosspoint memory array to execute the MVM. The resistive memory devices playing the role of synapses are preliminarily programmed to have conductance G_{ij}. A voltage vector V_j is applied to the array columns, thus resulting in output currents I_i given by Eq. (2).

The weighted summation of the McCulloch-Pitts model can be well described in hardware by the matrix-vector multiplication (MVM) in a crosspoint memory array, which is depicted in Fig. 6. In the crosspoint array, each resistive memory device is preprogrammed with conductance G_{ij}. The application of a voltage vector V_j at the array columns thus results in the generation of currents $G_{ij}V_j$ at the memory element with coordinates (i,j) via Ohm's law. All these currents are then collected at the array rows by Kirchhoff's law, thus yielding a total row current I_i given by:

$$I_i = \sum_j G_{ij} V_j, \quad (2)$$

which is line with the argument of the activation function in Eq. (1). The output current is then typically converted into voltage by transimpedance amplifiers and passed through an activation function to fully emulate the neuronal information processing. The significant advantage of the crosspoint array circuit is that it allows to accelerate MVM by simultaneous multiplication and summation by physical laws, in contrast with the iterative multiply-accumulate (MAC) algorithm for MVM execution in digital processing units.[86-88] Another strong advantage is the ability to process information within the memory, thus eliminating any data transfer between the memory and the separate processing unit which would be affected by the memory bottleneck of von Neumann architectures.[89, 90] On the other hand, in-memory MVM is executed in the analogue domain, which raises a number of concerns such as electronic noise, limited precision of the conductance values G_{ij}, non-linear memory characteristics and parasitic IR drop along the

row/column lines in the array circuit.[67] Nonetheless, MVM has been demonstrated in several applications, such as neural network acceleration,[49, 91-94] sparse coding,[95] mixed-precision computing,[96] compressed sensing,[97] solution of differential equations [98] and the solution of linear matrix problems such as matrix inversion [99] and linear regression.[100]

4.1 Neuron integration

Biological neurons are also known to have a memory effect, where input spikes are integrated, instead of being summed simultaneously. The incoming signals from synapses cause the increase of a local graded potential (LGP) in the dendritic membrane. The neuron then generates an action potential if the LGP reaches a threshold, otherwise relaxes to its resting state if the LGP is below the threshold. The neuron can thus conduct the signal-processing functions by information integration and the threshold firing.[101] This functionality of the biological neuron is expressed by the concept of I&F neuron, where spike integration causes the increase of an internal state variable, generally named membrane potential V_m. As the membrane potential reaches a given threshold V_{th}, then the neuron responds with a fire, i.e., by sending an output spike.[102, 103] In addition to this simple I&F concept, many other bio-plausible models have been proposed to implement artificial neurons, such as the leaky I&F model [104] and the biophysical HH model.[105, 106] The I&F neuron is usually implemented by relatively large CMOS circuits containing tens of transistors [6, 107] and large integrating capacitors.[108] For instance, a memory capacitor C_{mem} with capacitance of 432 fF was reported to have a layout area of 244 μm^2 in 0.35 μm CMOS technology for injection currents of the order of tens of picoamperes.[108] A larger capacitance may be needed in the case of larger synaptic currents, which might be the case for memory-based neural networks.[109-111] To reduce the circuit area for I&F neurons, one can take advantage of device physics of memory devices, typically in hybrid combination with CMOS transistors, to fully realize integration, firing and bursting modes of biological neurons.

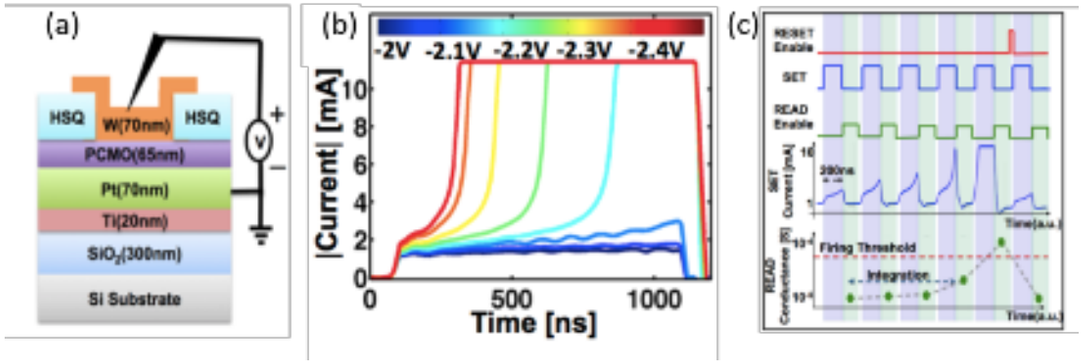

Fig. 7 I&F neuron based on PCMO RRAM. (a) Sketch of the RRAM device structure. (b) Time dependent current during an applied pulse at increasing voltage amplitude. As the voltage increases, the integration phase increases its slope and the time-to-fire decreases. (c) Results of the application of voltage spikes with constant voltage. The gradual increase of current under the first four pulses indicates the integration function, while the abrupt conductance increase induced by the fifth pulse represents the fire function. Reprinted with permission from Lashkare et al. IEEE Electron Device Lett. 39, 484 (2018).[112] Copyright IEEE (2018).

To reduce the area of the neuron integration circuit, it is possible to take advantage of pulse accumulation processes in nanoscale memory elements. For instance, the application of voltage pulses across a PCM can lead to incremental crystallization due to local Joule heating and a consequent increase of conductance, which can be used as an equivalent membrane potential.[10-12]

Similar pulse accumulation processes in FERAM [14] and RRAM [112-115] can be used for compact spike integration, thus allowing to minimize the neuron area.

Fig. 7 shows the implementation of I&F artificial neuron by using a $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO) RRAM, where the integration function is performed due to the gradual conductance increase during set process.[112] Fig. 7a shows the structure of the PCMO RRAM device, where 70 nm-thick PCMO layer is inserted between a Ti BE and a W TE. Fig. 7b shows the measured current in response to applied pulses of fixed width and increasing amplitude. In general, the current shows an initial gradual increase, which can be understood as the integration phase, followed by a steep rise, representing the fire response. The non-linear current response is the result of the ion migration dynamics in PCMO, where the field-driven defect migration lead to an increase of conductance. Fig. 7c shows the measured current in response to the application of a sequence of 5 voltage spikes. The conductance first gradually increases under the stimulations of repeated set pulses (i.e., integration function), followed by an abrupt increase of spike current once reaching a threshold (i.e., fire function). Subsequently, a reset pulse is used to reset the RRAM device to the initial conductance. The experimental results of current transient support the feasibility of I&F neuron based on PCMO RRAM.[112] Similarly, the neuron integration function can be performed in a $SrTiO_3$ based memristor device with uniform switching.[115]

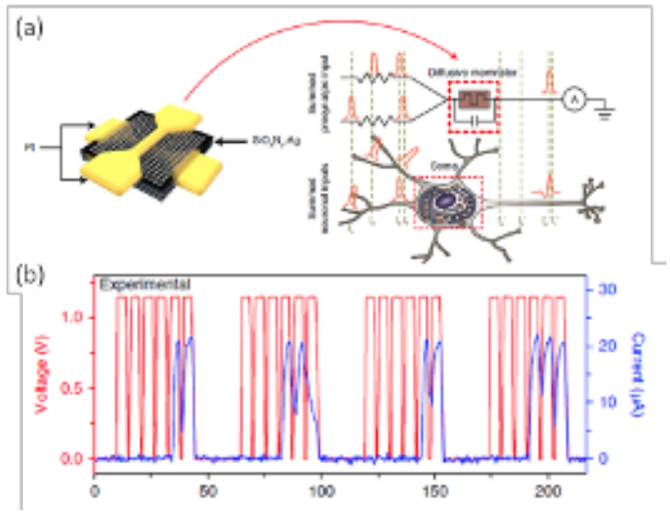

Fig. 8 The I&F artificial neuron comprising a SiO_xN_y:Ag diffusive memristor and a parallel capacitor. (a) Schematic illustration of the I&F neuron. (b) the experimental response of the artificial neuron. Reprinted with permission from Wang et al., Nat. Electron. 1, 137 (2018).[120] Copyright Springer Nature (2018).

4.2 Neuron fire

In general, devices exhibiting intrinsic threshold switching allow to perform the fire function in a simple way, i.e., within a nanosized device instead of using bulky comparators and pulse generators. In fact, firing, bursting and oscillating functions of the neurons have been reported by using threshold-switching devices based on Mott transition [116-118] and RRAM.[119-123] Fig. 8a illustrates a typical implementation of an artificial I&F neuron, consisting of a volatile RRAM device or diffusive memristor based on SiO_xN_y:Ag and a parallel capacitor.[120] In the figure, the diffusive memristor with volatile behavior executes the fire function by threshold switching, while the capacitor conducts the integration function through the charging process. Additionally, a resistor in series with the artificial neuron is adopted as artificial synapse and to monitor the output current versus time.

Fig. 8b shows the experimental response of the artificial neuron to a sequence of sub-threshold stimulations. [120] By applying the super-threshold pulse train on the I&F neuron, the capacitor is charged with a typical time constant, resulting in the increase of voltage across the diffusive memristor, thus serving as the LGP state variable. This integration process results in a negligible current during the first several pulses in the experimental data. Once the LGP reaches the threshold after a certain number of pulses, the volatile RRAM device switches to the high-conductance state, thus resulting in a fire output signal with high current. The delay time between the arrival of input spikes and the fire operation depends on the RC time constant and the internal Ag dynamics of the memristor. After fire, the device spontaneously relaxes to a low conductance state, corresponding to its resting state, as a result of the discharge of the capacitor and the volatile behavior of the RRAM device. The results in Fig. 8b supports the feasibility of the artificial I&F neuron enabled by the volatile RRAM physics. Similar I&F neuron implementations were reported by using a vertical MoS_2/graphene threshold switching memristor.[124] In some cases, RRAM devices with capacitive effect, referred to as memcapacitors, are also used to replace the common capacitor to implement the I&F neurons.[125-127] For instance, I&F neurons with various neuron functions were reported using single RRAM devices based on $GaTa_4Se_8$ [128] and a stack of $Ag/FeO_x/Pt$.[129] The combination of I&F functions in nanoscale memory device is the most promising to improve the scalability of artificial spiking neurons.

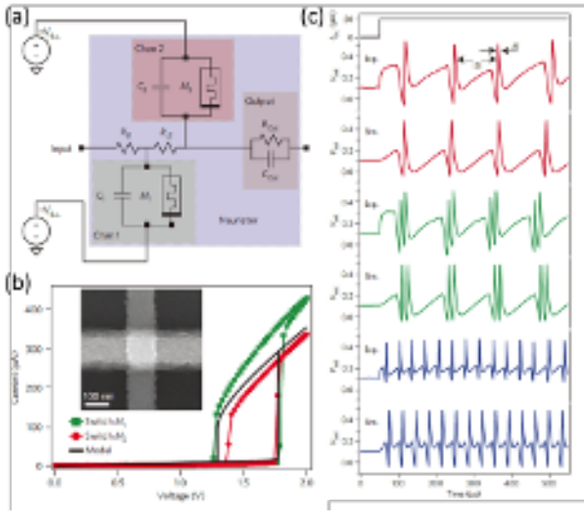

Fig. 9 Implementation of an oscillatory neuron using a HH model. (a) Schematic of the HH neuron circuit comprising two RRAM devices based on Mott insulator NbO_2 (M1 and M2) and two parallel capacitors (C_1 and C_2). (b) the I-V curve and the typical SEM image of the NbO_2 memristor. (c) Spike burst and trains of the oscillatory neuron. Reprinted with permission from Pickett et al., Nat. Mater. 12, 114 (2013).[116] Copyright Springer Nature (2013).

4.3 Oscillating neurons

Threshold switching in volatile RRAM devices provides the basis for generating self-sustained oscillations, thus enabling bio-plausible artificial neurons. Fig. 9 illustrates an oscillating neuron based on the Mott insulator NbO_2. [116] As shown in Fig. 9a, the application of a voltage close to a characteristic threshold voltage V_T causes the NbO_2 layer to switch from a high resistance (off) state to a low resistance (on) state, followed by a fast recovery of the initial off state. This threshold switching effect was explained as due to internal Joule heating triggering a higher conductance due to Poole Frenkel transport [130-133] or insulator-metal transition typical of Mott

insulators [134] or by a coexistence of these phenomena.[135] To describe the complex dynamics of Na^+ and K^+ ion channels in the HH neuron model, two elements are used in the HH neuron circuit of Fig. 9b each including a parallel combination of a NbO_2 RRAM device and a capacitor. These two ionic channels are stimulated by pulses with opposite polarity bias and coupled to each other by the load resistor R_{L2}, while the load resistor R_{L1} serves as input resistance. The parallel combination of the threshold switching device and a capacitor is able to induce oscillatory spike trains with various shapes. Assuming a constant input current, a time-oscillating response can be obtained by the HH circuit. Fig. 9c shows experimental results of the output of the HH neuron circuit, compared to circuit simulations for a constant input current of 20 μA. The inter-spike time interval can be controlled by the value of capacitances C_1 and C_2. Similar oscillatory HH neurons have been developed based on other types of devices exhibiting threshold switching, such as other Mott insulators VO_2 [136] and TaO_x,[137] and chalcogenide glass GeSe.[13] Oscillatory neurons have also been demonstrated by using SiO_xN_y:Ag volatile RRAM, which is capable to controlling the oscillation frequency by the conductance value.[138] Threshold switching in a HfO_2 layer with Pt/Ag nanodot top electrode and Pt bottom electrode was reported to display low operation voltage (<0.6 V) and ultralow power consumption (<1.8 μW), thus enabling low voltage/low power oscillatory neurons.[123]

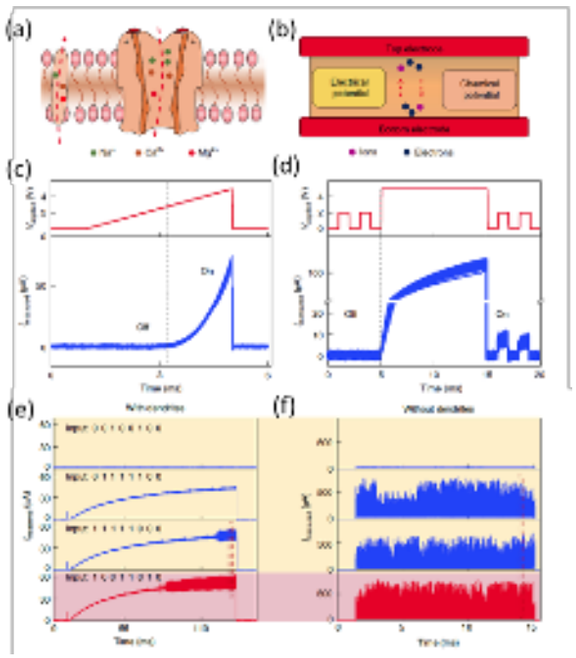

Fig. 10 Illustration of artificial dendrites by RRAM devices. (a) Schematic of the membrane of a biological dendrite (b) a metal-oxide-based dynamic memriastor as an artificial dendrite. (c) Measured nonlinear current response to the applied voltage on the fabricated artificial dendrite. The applied voltage ramped linearly from 0 V to 5 V. (d) Measured current response of the artificial dendrite device in the off and on states, exhibiting a nonlinear filtering and integration property. (e) Measured the output current of neural network with artificial dendrites for different input patterns. (f) Measured the output current of neural network without artificial dendrites for different input patterns. Reprinted with permission from Li et al., Nat. Nanotechnol. 15 (9), 776 (2020).[150] Copyright Springer Nature (2020).

4.4 Dendritic filtering

In the biological nervous system, dendrites are important components of neuronal units that extend from the cell body of neurons and play a critical role in information processing.[139-141] Dendrites are generally considered to be passive elements that merely transmit synaptic currents

to the soma. They can integrate synaptic inputs and output signals nonlinearly and filter out insignificant background information.[142-144] Recently, several CMOS-based circuits have been reported to emulate the dendrite functions .[145, 146] Implementing the dendrite function in nanoscaled devices is thus highly desirable for neuromorphic engineering. Dendritic integration was shown by using starch-based electrolyte-gated oxide transistors.[147] Spatiotemporal dendritic integration and linear/superlinear dendritic algorithms were demonstrated within transistor structures.[148, 149]

Fig. 10 illustrates the analogy between the ionic channel in a biological synapse (a) and the two-terminal RRAM device (b), which provides the foundation to implement the key dendritic functions.[150] To implement dendritic nonlinear integration and filtering functions, a volatile RRAM with Pt/TaO_x/AlO_δ/Al stack was proposed, where a positive voltage stimulus leads to conductance increase followed by a gradual relaxation to the initial high-resistance state as the voltage bias is removed. Fig. 10c shows the measured electric characteristics of the artificial dendritic device, indicating a nonlinear current response to a linearly increasing voltage from 0 V to 5 V, which is similar to that of N-methyl-d-aspartate (NMDA) channels in the biological dendrite. Fig. 10d shows that the RRAM device can filter out sub-threshold input signals smaller than the threshold of 3 V and performs nonlinear integration of input signals larger than the threshold voltage, resulting in a continuously increasing current response over time. The filtering effect can be explained by the energy barrier for oxygen ion migration. Only the input signals with amplitude larger than the threshold voltage can induce the oxygen ion migration toward the Al electrode, resulting in a decrease of the barrier height and an increasing current response to the applied voltages. On the other hand, the sub-threshold input signals are filtered out. Fig. 10e and f shows the measured current during the inference process of a neural network for various input patterns with and without artificial dendrites, respectively. The pattern recognition accuracy and power consumption are significantly improved by including the dendritic devices into the neural network, thanks to the filtering effect.

5. Hardware synapses

Synapses in the biological neural system are responsible for the weighted transmission of spikes from a pre-synaptic neuron to a post-synaptic neuron, as depicted in Fig. 11a.[151] Most importantly, the synaptic weight should be able to adjust depending on the history of spiking stimulation, a phenomenon known as synaptic plasticity which is regarded as the basis for learning and memory functions. Synaptic plasticity can be realized in hardware via the conductance change in memory device, such as the set and reset processes in RRAM devices which have been widely developed to mimicking biological synapse.[151-155] The close emulation of synaptic functions is a critical step to achieve a neuromorphic system with the ability to learn and adapt in response to environmental changes. Generally, the synaptic plasticity can be categorized into long-term plasticity and short-term plasticity depending on the retention time, representing the permanent and temporary synaptic modification, respectively [47,155,156] Various long- and short-term synaptic functions have been demonstrated by utilizing memory devices, such as STDP, spike-rate-dependent plasticity (SRDP), paired-pulse facilitation (PPF) and paired-pulse depression (PPD).[47,151-159]

5.1 Long-term potentiation and depression

Long-term potentiation (LTP) and long-term depression (LTD) consist of the permanent increase or decrease, respectively, of the synaptic weight as a result of the spiking stimulation. LTP and LTD is possible in nonvolatile memory devices by the pulse-induced change of the conductance according to the input pulse shape and number.[155] Both digital (binary) and analogue (multilevel) conductance change are reported.[160-162] Binary states are more suitable for memory storage due

to the clear difference between high-resistance state (HRS) and low-resistance state (LRS).[163] On the other hand, analogue states are ideal for synaptic devices with incremental weight update.[151-155] In particular, analogue-type conductance states with linear and symmetric LTP/LTD are essential in hardware accelerators of inference and training.[49, 93, 94, 164-166] Non-linear and asymmetric LTP/LTD are commonly observed in most synaptic devices.[67, 94, 167]Algorithmic and engineering methods should be identified to compensate the intrinsic linearity of synaptic weight update.

The linearity of the update characteristics can be improved by optimization of the programming pulse, [168] utilization of defects engineering[169] and adoption of three terminal devices such as the ECRAM.[68, 74] Fig. 11b shows the LTP and LTD behaviors of a $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO) based memristor under the programming spikes with different pulse scheme.[168] The A-type behavior is a typical update characteristic with nonlinear LTP and abrupt LTD, which was obtained by using spikes with constant voltage amplitude. The update linearity of LTP/LTD can be clearly improved by adopting spikes with incremental amplitude (type B) and pulse width (type C). These results indicate that nonidentical pulses are most effective in controlling and improving the synaptic update linearity. This is because the increasing amplitude/pulse-width compensate the typical saturating behavior of the conductance for constant pulses.[94] However, note that the increasing amplitude and increasing width methodologies are not compatible with the outer product scheme of weight update, where the whole crosspoint array is updated simultaneously by applying voltage vectors at the rows and columns with variable pulse widths.[169] The increasing amplitude and width thus results in more complicated updated schemes requiring longer update time and larger energy consumption.

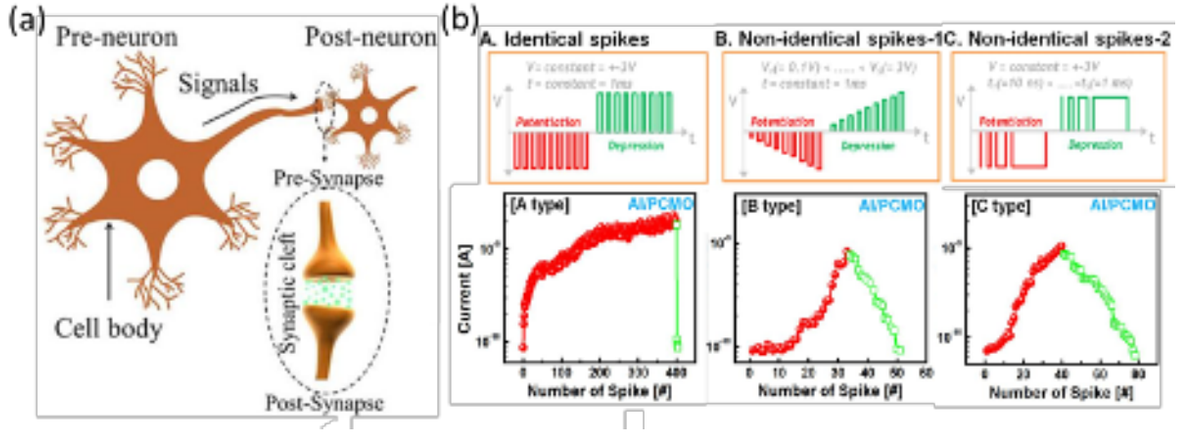

Fig. 11 (a) Schematic illustration of a biological synapse.[151](b) LTP/LTD processes operated using different pulse spikes, namely identical spikes and non-identical spikes with incremental amplitude and pulse width.[168] Reprinted with permission from Lin et al., NPG Asia Mater. 12, 64 (2020)[151] and Park et al., IEEE Int. Electron Devices Meeting (IEDM) (2013), pp. 25.6.1-25.6.4.[168] Copyright Springer Nature (2020), IEEE (2013).

The LTP/LTD linearity can be also enhanced in ECRAM devices thanks to the bulk-type of switching and to the separation between the programming path (between gate and channel) and the read path (across the channel between source and drain). This allows for better controllability of the device conductance by field-induced migration of impurities, such as Li ions in inorganic ECRAM[68-70] or protons in organic ECRAM.[71]

5.2 Spike-timing-dependent plasticity (STDP)

STDP, namely the weight modification relying on the temporal order of pre- and post-synaptic spikes, is regarded as one of the essential learning rules for unsupervised learning.[7,9,154,155] Thus, implementing STDP rules in hardware SNNs is a critical step toward achieving neuromorphic systems capable of learning and adaptation. For the typical STDP rule, the synaptic weight undergoes LTP if a pre-synaptic spike occurs earlier than a post-synaptic spike, i.e. it the spike delay $\Delta t = t_{post} — t_{pre}$ between the post-synaptic spike time t_{post} and the pre-synaptic spike time t_{pre} is positive. Conversely, LTD takes place for the case $\Delta t < 0$.[154,155] To achieve the above STDP function, the synaptic device usually needs to satisfy the requirement of gradual conductance change and fast response to individual spikes.

Various STDP methods have been reported for both digital- and analogue-type memory devices.[154,155,170-173] A typical approach for STDP is the overlap method where the neuron spike is designed such that the Δt-dependent overlap between pre- and post-synaptic spikes leads to the desired LTP or LTD.[154,172,173] Fig. 12 shows typical examples of overlap-type implementations of STDP for HfO_2-based RRAMs.[173,174] The pre- and post-spikes can be designed as series of 6 pulses, where the first negative pulse is followed by 5 positive pulses with decreasing amplitude, as shown in Fig. 12a. An important design principle is that each individual spike is unable to induce a conductance change ΔG, i.e., all pulses should be below the threshold for set/reset processes. However, the overlap between pre-spike applied at one electrode and the post-spike applied to the other electrode causes a voltage drop across the memory device that is large enough to change the conductance. As illustrated in Fig. 12a, when the pre-spike is applied earlier than the post-spike ($\Delta t > 0$), the overlapping spikes result in a positive pulse with a relatively large amplitude, hence LTP.[173] On the other hand, for $\Delta t < 0$, the overlapping spikes cause a negative pulse with large amplitude, hence LTD. Most importantly, Δt controls the amplitude of the resulting pulse amplitude, hence the degree of conductance change ΔG. Fig. 12b shows the measured ΔG as a function of the spike timing, indicating that the amplitude of positive and negative ΔG decreases for increasing delay $|\Delta t|$, in agreement with the STDP rule.[173] Previous work suggests that the correlation between ΔG and Δt can be tuned by adjusting the pulse shape and the programming scheme.[175] Various types of STDP curves where obtained by the overlap approach in various synaptic devices, including RRAM,[176, 177] PCM,[178-180] STT-MRAM[181] and FERAM.[172]

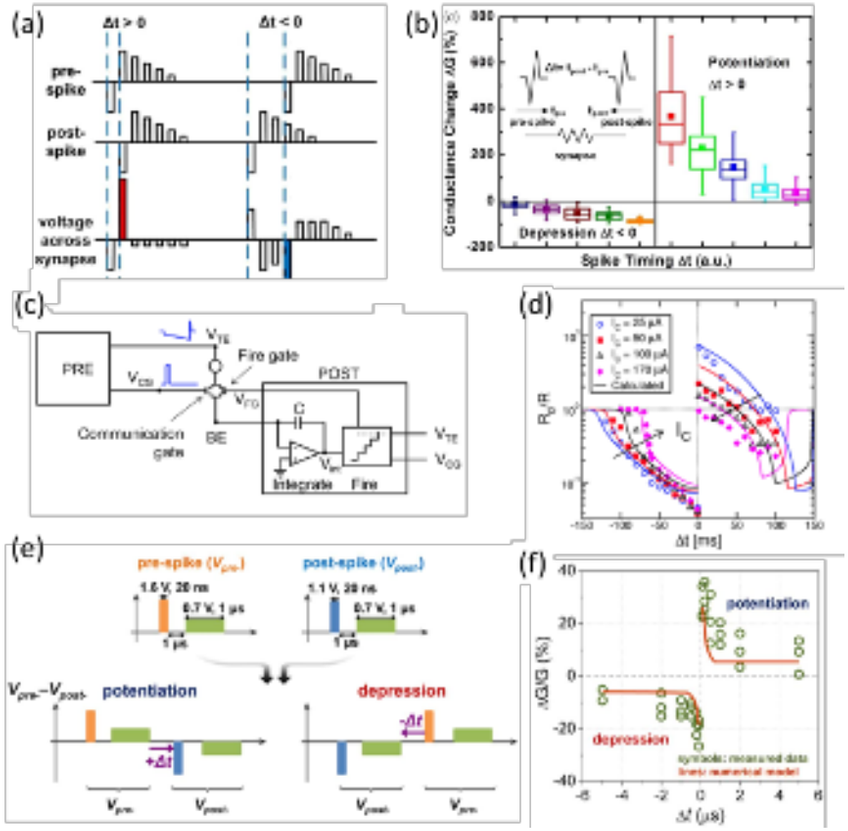

Fig. 12 Schemes for the implementation of STDP rule using the overlap approach and the 2T1R structure and the nonoverlap method in RRAM devices. (a-b) the pulse design and measured STDP of the HfO_2 based RRAM through overlapping the pre- and post-spikes.[173] (c) the illustrative scheme for 2T1R synapse containing a memristor and 2 transistors. (d) the tunable STDP curves with different initial conductance states obtained by set the device under increasing I_C from 25 μA to 170 μA.[174] (e-f) the operation design and the realization of STDP function using the nonoverlap spikes in a second-order Ta_2O_{5-x}/TaO_y memristor.[188]Reprinted with permission from Yu et al., IEEE Trans. Electron Devices 58, 2729 (2011),[173] Wang et al., Front. Neurosci. 8, 438 (2015)[174] and Kim et al., Nano Lett. 15, 2203 (2015). [188] Copyright IEEE (2011), Frontiers (2015), ACS (2015).

The overlap STDP scheme may suffer from a relatively large variation of ΔG, since there is no compliance current to control the growth of conducting filament during the LTP process. To overcome this issue, the 2-transistor/1-resistor (2T1R) synaptic circuit structure was proposed to implement STDP function in RRAM[174] and PCM.[183] As shown in Fig. 2, a series MOS transistor can limit the current during the set transition for better controlling the resistance in RRAM.[35,184] The additional transistor also provides a multiple-input control for handling the various synaptic functions, i.e., spike transmission, LTP and LTD. Fig. 12c illustrates the 2T1R synapse, where the PRE spike is applied to the RRAM TE, while the POST spike is applied to the fire gate (FG). Additionally, a short positive pulse is given to the communication gate (CG) of the second transistor for synaptic transmission. The coincidence of the PRE spike with amplitude V_{TE} and the POST spike with amplitude of V_{FG} can induce the set and reset transition of resistive switching memory, thus leading to LTP and LTD, respectively. Importantly, the filament growth is controlled by the V_{FG}, which in turn depends on the spike timing thus enabling time-dependent potentiation according to the STDP function. Fig. 12d shows the resulting STDP characteristics, namely the relative change of conductance R_0/R, where R_0 is the resistance before the spike application and R is the final resistance, for various initial states R_0 obtained for various I_C. the results indicate LTD for $\Delta t < 0$ and LTP for $\Delta t > 0$, where the change of conductance tends to vanish at increasing $|\Delta t|$,

which is in line with the observed biological STDP.[7] A simplified STDP synaptic circuit was reported by adopting a 1-transistor/1-resistor (1T1R) structure with RRAM synapse.[111,185] This was later extended to a four-transistor/one-resistor (4T1R) structure to demonstrate the SRDP.[186, 187]

Although the overlap method allows for efficient STDP function with local activity, it does not fully account for the observed biological STDP, where overlapping spikes are generally not necessary for weight update. To overcome this limitation, the second-order memristor was proposed to execute LTP/LTD according to the STDP rule without any overlap between pre- and post-spikes.[188] The second-order memristor consists of a RRAM device where the conductance change is not only determined by the first variable, e.g., the filament size or interface barrier, but also by a second variable, e.g., the local temperature or oxygen mobility, which impacts the dynamics of the first variable.[41,155,188,189] The second variable usually displays a transient dynamics, such as a spontaneous decay after stimulation, which is similar to the Ca2+ dynamics in the biological synapse. As a result, the second-order memristor can display non-overlap, biorealistic emulation of STDP rule and other synaptic learning functions. [41,155,188,189] Fig. 12e illustrates the pre-/post-spikes, including a programming pulse with high amplitude and a heating pulse with long pulse width. By applying the pre- and post-spikes at the TE and BE, the interaction between the applied electric field and the local temperature can lead to a Δt-dependent conductance change, as indicated by the STDP characteristic for a Ta_2O_{5-x}/TaO_y second-order memristor in Fig. 12g. Similarly, a second-order memristor consisting of a $Pt/WO_{3-x}/W$ stack was reported, where the two variables are the Schottky barrier and the oxygen ion mobility.[177] Second-order memristor were experimentally demonstrated for various material systems, such as InGaZnO,[155] Ta_2O_{5-x}/TaO_y,[188] WO_{3-x},[177,189] $SrTiO_3$,[190,] SiO_xN_y:Ag.[41] and TiO_2:Ag.[191]

5.3 Spike-rate-dependent plasticity (SRDP)

In the human brain, there are two main types of information coding, namely time coding and rate coding. While STDP is most suitable for learning in the presence of time coding, SRDP can serve as learning rule for rate coding.[192] Frequency dependent LTP/LTD have been extensively reported in memory devices with dynamic effects, e.g., oxygen diffusion.[193,194] SRDP generally relies on the Bienenstock-Cooper-Munro (BCM) learning rule as a high-order function of SRDP.[195-197]According to the BCM rule, spike trains with a frequency larger than a certain threshold induce LTP, while spike trains with a lower frequency lead to LTD. A threshold slide effect has been reported, where the threshold frequency changes depending on the learning experience, thus enabling a history-dependent synaptic adaptation.[83,198] Many efforts have been made to realize the BCM rule by using the rate-based pre-spikes in the second-order memristors.[189,199] In these schemes, the forgetting effect of the learning experiences and the potentiation effect induced by the rate-based pre-spikes were compared, thus achieving the BCM learning rule with monotonic trend. The effect of tunable forgetting rate on the BCM curve was studied for $SrTiO_3$-based RRAM devices.[200] However, monotonic SRDP is not consistent with the "tick" shape of BCM rule in biological systems. Also, the BCM rule should represent the long-term characteristics rather than the short-term modification implemented in some studies.[189,200]

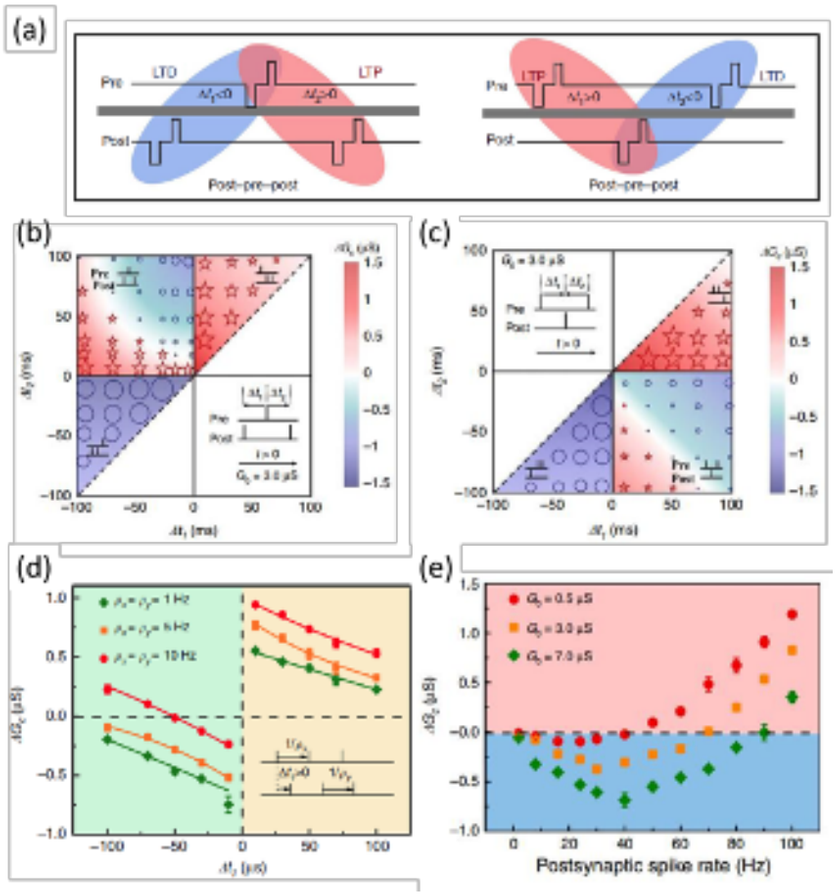

Fig. 13 Bio-realistic demonstration of BCM learning rule using a triplet-STDP scheme.[199] (a) Scheme for the typical spike triplets of 'post-pre-post' and 'pre-post-pre'. (b-c) the experimental results of triplet-STDP measured in WO_{3-x} memristor. It summarized the LTP and LTD using the sequences of 'post-pre-post' and 'pre-post-pre' with various timing intervals. The degree of ΔG is indicated by both the symbol size and background color. (d) The dependence of ΔG on both the pre-spike rate ρ_x and post-spike rate ρ_y. (e) The triplet-STDP based BCM rules with various learning experiences (i.e., different initial conductance G_0). Reprinted with permission from Wang et al. Nat. Commun. 11, 1510 (2020).[199] Copyright Springer Nature (2020).

Compared to the standard STDP with paired spikes, a third spike is introduced in the triplet-STDP, thus resulting in a triplet of interacting spikes. The interaction of paired spikes with the third spike leads to the multiplicative term to enable the BCM rule. In biological systems, there are two types of triplet STDP, namely, the first-spike-dominating rule and last-spike-dominating rule. The former was demonstrated in a $Pt/SrTiO_3/Nb$-STO stack RRAM exhibiting synaptic suppression triplet-STDP.[115] Last-spike-dominating triplet-STDP was reported for a $Pt/WO_{3-x}/W$ second-order memristor.[199,201] This is shown in Fig. 13a reporting the typical triplets of 'post-pre-post' and 'pre-post-pre' for stimulating the WO_{3-x} synaptic RRAM device. Fig. 13b shows the conductance change as a function of the first and the second spike delay in the post-pre-post triplet, Δt_1 and Δt_2, respectively, while Fig. 13c shows the same for the pre-post-pre triplet. Fig. 13d shows the measured ΔG as a function of Δt_1 for increasing spiking rate, indicating that plasticity depends on both the pre-spike rate ρ_x and post-spike rate ρ_y. Based on these results, the BCM learning rule can be implemented by designing a proper triplet-STDP scheme. Fig. 13e illustrates the triplet-based BCM learning rule by extracting the data from the diagonal line of quadrant II in Fig. 13b and defining the post-spike rate as given by $\rho_y = 1/(|\Delta t_1| + |\Delta t_2|)$. The experience-dependent sliding threshold characteristic is also demonstrated by tuning the initial conductance G_0 in such BCM implementation, resulting in a close emulation of the biological BCM curve.[199]

5.4 Short-term synaptic plasticity and memory

While long-term plasticity can last for the entire lifetime, short-term plasticity or short-term memory (STM) in the human brain can be as short as milliseconds to minutes.[148-150] Several typical types of STM have been realized in hardware memory devices, including the excitatory postsynaptic current (EPSC), PPF/PPD and SRDP.[47,155-157,193,194] Usually STM is implemented by directly taking advantage of the inherent transient behavior of volatile memory devices. For instance, Fig. 14 shows the analogy between the transient dynamics of the EPCS [202] and the volatile nature of an Ag filament in a HfO2-based RRAM device.[203] In a biological synapse of Fig. 14a, a pre-synaptic spiking stimulation induces the release of neurotransmitter from synaptic vesicles into the synaptic cleft. The neurotransmitter, e.g., L-glutamate, then binds to the receptor to activate an ion channel, thus triggering the ionic inflow of Na^+ and Ca^{2+} into the post-synaptic neuron, which is responsible for the EPSC.[202] The opening of the ion channels has limited duration in time, which accounts for the transient nature of the EPSC. In a volatile RRAM, the electrical pulse results in the formation of an Ag filament, which then serves as a conductive bridge for electrons across the RRAM. Both the EPSC and the conductive filament remain active for a short time, typically in the range from few ms to several minutes. The physics of the volatile RRAM can thus serve as a basis for replicating STM in hardware via a small-scale device, i.e., without the need for large capacitors to emulate relatively-long time constants.

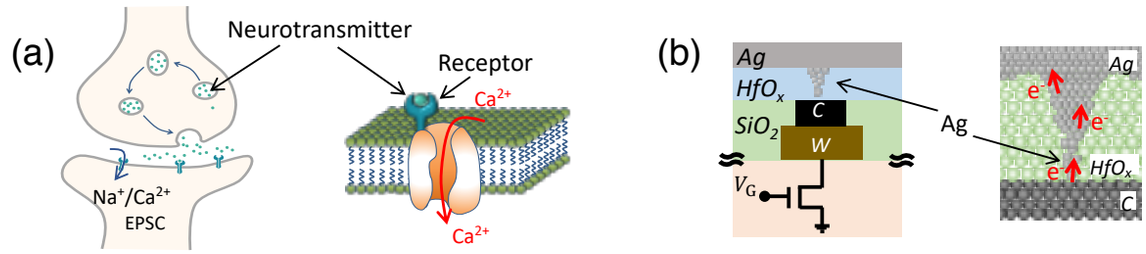

Fig. 14 Analogy between EPSC in biological synapses and the diffusive Ag filament in volatile RRAM devices. (a) The pre-synaptic stimulation causes the release of a neurotransmitter, which activates Ca^{2+} transport across the ionic channels at the basis of EPSC. (b) An Ag filament is formed by a voltage pulse via Ag ion migration. The Ag filament then serves as a bridge for electron conduction across the Ag filament. Reprinted with permission from Lester et al., J. Neurosci. 12, 635 (1992) [202] and Wang et al., Adv. Intell. Syst. 2000224 (2020).[203] Copyright the Society for Neuroscience (1992), Wiley (2020).

Volatile memory effects have been used to naturally emulate the EPSC in several two- and 3-terminal memory devices.[41,155,204-207] Similarly, volatile memory devices can also mimic the PPF induced by paired spikes.[204-207] In a biological PPF, the second spike can generate much larger change of synaptic weight than the first spike, thus resulting in a strong spike interaction and correlation of spikes in the Ca^{2+} dynamics. On the other hand, paired spikes may also cause synaptic depression, hence PPD, which has been also mimicked in several memory devices.[157,208,209]

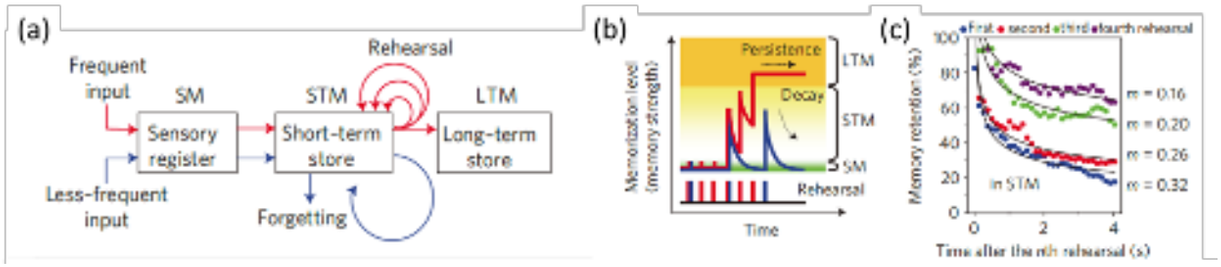

Fig. 15 Illustration of the transition from STM to LTM in volatile RRAM devices.[47] (a) The psychological model of human memory and (b) the simplified memory model for RRAM synapse. (c) the STM-to-LTM transition with repeated learning in the Ag_2S based atomic switching memristor. Reprinted with permission from Ohno et al., Nat. Mater. 10, 591 (2011).[47] Copyright Springer Nature (2011).

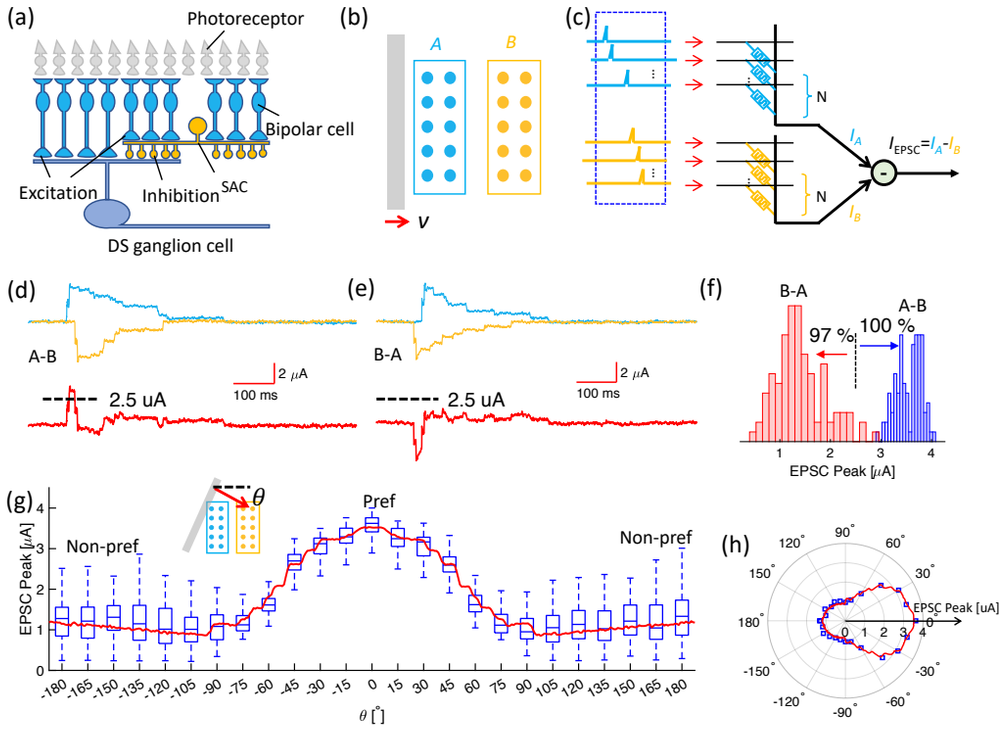

Fig. 16 Illustration of movement recognition by volatile RRAM devices.[203] (a) Structure of the direction selective (DS) ganglion cell, comparing excitatory and inhibitory EPSCs induced by light-stimulated photoreceptors. (b,c) The movement of a light bar across the receptive field causes the activation of excitatory current spikes, followed by inhibitory current spikes, which result in an EPSC with large positive current. (d,e) Comparison of excitatory and inhibitory currents from volatile RRAM devices for the case of an image bar moving from left to right (preferred direction) and from right to left (non-preferred direction). Only the EPSC of the preferred direction can exceed the threshold, thus being recognized. (f) Histogram of EPSC for preferred and non-preferred directions. (g,h) EPSC peak as a function of the movement direction over the whole range of angles. Reprinted with permission from Wang et al., Adv. Intell. Syst. 2000224 (2020).[203] Copyright Wiley (2020).

According to our daily experience, it is known that STM is capable to transition to long-term memory (LTM) by repeated training, as illustrated in Fig. 15a.[47] Ag_2S-based volatile RRAM, also called atomic switches, can replicate a similar function. Fig. 15b depicts a simplified memory model to implement the transition from STM to LTM transition in a volatile RRAM synapse, where the memorization level can increase from the sensory memory (SM) to STM and LTM by increasing the number of stimulations, similar to repeated rehearsals in the human experience. Fig. 15c shows that data retention is clearly enhanced by repeated stimulations, supporting the transition

from STM to LTM in the Ag_2S RRAM device. Similar to the Ag_2S RRAM device, the transition from STM to LTM has been extensively reported for various memory devices.[155,156,210,211]

5.5 Cognitive computing functions enabled by STM

STM is an essential function in the human brain that is functional for several sensing and recognition functions, such as the recognition of speech, movement and other types of dynamic information. Fig. 16 shows an example of the use of volatile RRAM for the movement recognition and direction selectivity similar to the human retina. The biological visual system is capable of fast motion detection by direction-selective (DS) ganglion cells.[212] As shown in Fig. 16a, the retina includes bipolar cells and starburst amacrine cells (SACs) with receptive fields capable of stimulating the ganglion cells with excitatory and inhibitory inputs, respectively.[203] The combination of excitatory and inhibitory signals causes an EPSC into the ganglion cells, which enables the recognition of various moving directions under sight. Fig. 16b shows receptive field stimulated by a moving light bar, which first induces excitatory current spikes, followed by inhibitory current spikes. The comparison between the transient excitatory and inhibitory currents result in an EPSC with large positive current, which exceeds a threshold thus triggering the detection of the preferred direction. The transient excitatory and inhibitory currents was replicated in hardware by volatile RRAM with Ag TE and HfO_2 as switching material.[203] Fig. 16c schematically shows the circuit with several volatile RRAM to enable the averaging of stochastic excitatory and inhibitory currents. The overall EPSC, obtained as the subtraction of excitatory and inhibitory currents, shows a positive peak for the preferred direction (left to right, Fig. 16d) and negative peak for the non-preferred direction (right to left, Fig. 16e). Fig. 16f shows the distribution of preferred and non-preferred EPSCs, indicating that the two directions can be efficiently discriminated by comparing the EPSC to a threshold. The same concept can be extended to the full range of movement directions (Fig. 16g and h), thus enabling fast direction sensitivity by direct current sensing in the analogue domain.[203]

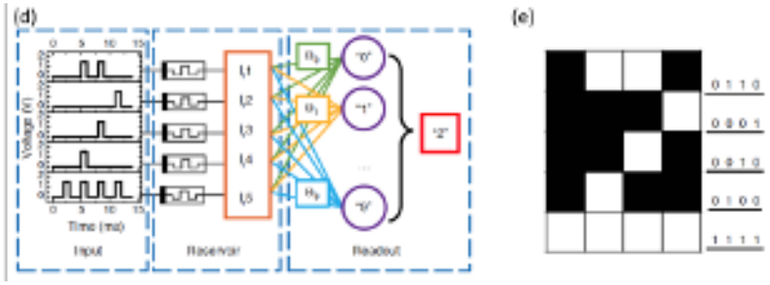

Fig. 17 Illustration of reservoir computing enabled by STM. (a) Network for digit recognition based on memristive reservoir computing system. (b) Image pattern for digit '2', based on 5 memristors each receiving a sequence of 4 spikes. Reprinted with permission from Du et al., Nat. Commun. 8, 2204 (2017).[213] Copyright Springer Nature (2017).

STM is also at the basis of reservoir computing (RC) systems, [214,215] which are widely utilized to implement temporal and sequential data processing. Generally, a RC system consists of a reservoir network for mapping the input stimuli into a high-dimensional feature space and a readout network for the analysis of the response from the reservoir states and final inference. Volatile memory devices with intrinsic STM behavior offer an ideal platform for brain-inspired implementing RC systems. For instance, volatile WO_x-based RRAM with STM dynamic effect were used to implement a RC network for image recognition.[213] Fig. 17 shows the network architecture of the RC system for digit recognition using 5 volatile RRAM devices. Each digit is mapped into 20

pixels, as shown in Fig. 17b for the case of digit 2. The 20 pixels are divided into 5 rows, each row stimulated with a sequence of 4 consecutive spikes applied to one of the 5 RRAM synapses. As a result, each RRAM device is stimulated by a 4-spike timeframe input stream. The image is thus represented by a spatiotemporal coding, i.e., not only using the spatial location in the rows but also the temporal sequence of the stream. For the readout function, a fully-connected network with 5 input neurons and 10 output neurons is employed to measure the conductance states of the 5 memristors in the reservoir network and recognize the digit. The recognition of the ten digits is executed only using 5 memristors, which is far less than the 200 weights in a conventional neural network. Similar spatiotemporal RC networks based on RRAM have been shown for handwritten digit recognition,[213] solution of second-order nonlinear tasks,[213] spoken-digit recognition[214] and autonomous chaotic time-series forecasting.[214] Besides top-down memory devices fabricated with conventional microelectronic technology, bottom-up approaches have been proposed. For instance, volatile switching was demonstrated in a network of switching nanowires capable of learning via homo-synaptic and hetero-synaptic plasticity.[215] This concept might pave the way for hardware implementation of unconventional computing paradigms in self-organizing stochastic networks of nanowires.

6. Technological challenges and potential solutions

While RRAM devices offer a wealth of physical properties that are attractive for neuromorphic computing primitives, there are several technological challenges that currently prevent the widespread adoption of RRAM for memory and computing.

A major technological limitation is given by the programming and read variations that prevent repeatable, reliable storage of data. This is a strong issue especially for MLC storage where the drift and fluctuation of the conductance cause time-dependent retention failure.[51] For instance, DNNs adopting MLC weights for MVM are heavily affected by fluctuations and drift that can cause a significant drop of accuracy during time. [49,217]Programming variations can be improved by accurate program-verify algorithms, based on voltage- or current ramping during the set or reset operation. In particular, current-based approaches appear most promising thanks to a relatively shallow programming characteristics, compared to voltage-based techniques. [218] Relaxation effects might be mitigated by redundancy techniques, where averaging among various devices allow for a better robustness toward individual fluctuations and noise. [219] It has been observed that low-current LRS are more affected by drift and variations after programming, as a result of the smaller size of the conductive filament.[220] Therefore, one may configure the network algorithm in such a way that the number of intermediate weights is minimized, whereas the presence of HRS and full LRS with high stability is maximized. [221] RRAM technologies with higher stability, such epitaxial[63] and uniform-switching RRAM[64-66] might also improve the immunity to resistance fluctuations for high precision in-memory computing. On the other hand, neuromorphic computing appears less affected by variations and stochasticity, thanks to the self-adaptation and continuous learning, where a change in the device parameters can be compensated in real time. However, note that variability in neuromorphic circuits does not affect only the synaptic weight but also all other brain-inspired properties such as neuron integration and retention time of short-term synapses.

Another key issue for RRAM is the excessive read current, which is due to the filamentary conduction across a metallic path across a nanometric length in the active oxide. Typically, the LRS shows a conductance in the range between 10 kΩ and 100 kΩ, which corresponds to a synaptic current between 1 μA and 10 μA for a typical read voltage of 0.1 V. While this current is reasonable for typical memory applications, aimed at fast read in presence of large parasitic

capacitances within large arrays, it represents a critical limitation for MVM implementations in DNNs and other neuromorphic applications. For instance, assuming a crosspoint array of size 128x128 in Fig. 6 with an average synaptic current of I_{read} = 10 μA, the current in an individual row/column would reach 640 μA, which requires a decoder transistor with the proper channel size for sensing and amplification. In addition, the large read current may lead to a significant voltage drop, also known as IR drop, along the row/columns of the crosspoint array. For instance, assuming a cell-to-cell wire resistance of r = 1 Ω in the array columns/rows,[222] the total voltage drop would be approximately given by $\Delta V \approx rI_{read}N^2/2$, which gives ΔV ≈82 mV, which contributes an error around 82% with respect to the applied voltage. Reducing the operating current in the device typically requires LRS at relatively small filament size, which are in turn less stable with respect to drift and fluctuations. At architecture level, the IR drop issue is addressed by adopting crosspoint arrays, also referred to as tiles, with relatively small size, e.g., below the 32x32 range.[223] Sparsity, which is typical of the human brain, hence of many hardware neuromorphic circuits, can alleviate the IR-drop problem, as it reduces the number of active synapses within the array. Alternative device concepts, such as uniform-switching RRAM[64-66] or ECRAM[67-74] characterized by bulk-type conduction, appear more promising in reducing the read current, thus enabling a larger size of the neuromorphic array.

More on the technological side, provided that synaptic currents can be substantially reduced, a significant issue is the development of high density crosspoint array, possibly with 3D integration. The brain is in fact characterized by a high connectivity, where each neuron is connected, on average, to 10,000 neurons. [224] Achieving such a large connectivity thus requires arrays with extremely large numbers of rows and columns, which makes 3D integration mandatory to fit the neuromorphic circuit within a single chip. Recently, 3D crosspoint arrays with 8 layers of RRAM devices with vertically-aligned electrode have been demonstrated for DNN implementation,[27] although the extension of this technology to brain-inspired cognitive circuits has not been reported yet. In this regard, a significant challenge is the RRAM selector, since the 3D integration of CMOS transistors is not straightforward. Several non-linear selectors with the capability of 3D integration have been reported, including Mott insulator, [225-227] chalcogenide glasses, [228,229] mixed ion-electron conduction (MIEC) devices, [230] multilayer tunnel junctions, [231] and threshold vacuum switches.[232] The resulting one-selector/one-resistor (1S1R) structure is extremely compact and suitable for 3D integration, thus being very attractive for both memory[233] and computing applications.[67] 3D-integrated, monolithic circuits capable of hetero-integration of various RRAM technologies, each serving a different function for sensing, neurons and synapses, would provide the ideal technology platform for neuromorphic system scapable of paralleling the brain computing functionality via device physics.

7. Conclusions

Neuromorphic computing requires a set of ad-hoc hardware capable of harnessing device physics to recreate the neuron and synapse functions in the human brain. RRAM offers a range of physical phenomena, arising from electrical transport, switching and ion migration, that can be used to approximate neuromorphic functions, such as neuronal integration, fire, oscillations, dendritic filtering and synaptic plasticity according to various spike-time, spike-rate learning rules experimentally observed in the brain. Ionic diffusion allows for short-term plasticity and STM, which form the basis of direction selectivity, RC and other emerging cognitive computing concepts. While many of these phenomena have individually been demonstrated by proof of concept, their combination into full neural networks and their extension to alternative architectures, such as multiterminal devices and bottom-up nanostructures, may further develop

this field of neuromorphic devices into a mature technology for manufacturable cognitive computing hardware.

8. Acknowledgments

This work was supported by the Italian Ministry of Foreign Affairs and International Cooperation (grant number PGR01011) and by Chinese MOST (grant number 2018YFE0118300).

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. G. Indiveri and S.-C. Liu, Proc. IEEE **103**, 1379 (2015).
2. F. Rosenblatt, Cornell Aeronautical Laboratory Report 85-460-1, (1957).
3. M. L. Minsky and S. A. Papert, MIT Press (2017).
4. Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
5. E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri, Proc. IEEE **102**, 1367 (2014).
6. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, Science **345**, 668 (2014).
7. G. Q. Bi and M. M. Poo, J. Neurosci. **18**, 10464 (1998).
8. J. P. Pfister and W. Gerstner, J. Neurosci. **26**, 9673 (2006).
9. T. Masquelier, R. Guyonneau, and S. J. Thorpe, PLoS ONE **3**, e1377 (2008).
10. S. R. Ovshinsky, Jap. J. Appl. Phys. **43**, 4695 (2004).
11. C. D. Wright, P. Hosseini, and J. A. V. Diosdado, Adv. Funct. Mater. **23**, 2248 (2013).
12. T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, Nat. Nanotechnol. **11**, 693 (2016).
13. M. Lee, S. W. Cho, S. J. Kim, J. Y. Kwak, H. Ju, Y. Yi, B.-k. Cheong, and S. Lee, Phys. Rev. Appl. **13**, 064056 (2020).
14. H. Mulaosmanovic, E. Chicca, M. Bertele, T. Mikolajick, and S. Slesazeck, Nanoscale **10**, 21755 (2018).
15. S. Raoux, W. Welnic, and D. Ielmini, Chem. Rev. **110**, 240 (2010).
16. C. Chappert, A. Fert, and F. N. Van Dau, Nat. Mater. **6**, 813 (2007).
17. M. Cubukcu, O. Boulle, M. Drouard, K. Garello, C. Onur Avci, I. Mihai Miron, J. Langer, B. Ocker, P. Gambardella, and G. Gaudin, Appl. Phys. Lett. **104**, 042406 (2014).
18. T. Mikolajick, C. Dehm, W. Hartner, I. Kasko, M. J. Kastner, N. Nagel, M. Moert, and C. Mazure, Microelectron. Reliab. **41**, 947 (2001).
19. H. Mulaosmanovic, J. Ocker, S. Müller, M. Noack, J. Müller, P. Polakowski, T. Mikolajick, and S. Slesazeck, In Symposium on VLSI Technology, IEEE (2017), PP. T176-T177.
20. R. Waser and M. Aono, Nat. Mater. **6**, 833 (2007).
21. H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, Proc. IEEE. **100**, 1951 (2012).
22. D. Ielmini, Semicond. Sci. Technol. **31**, 063002 (2016).
23. C.-C. Chou, Z.-J. Lin, P.-L. Tseng, C.-F. Li, C.-Y. Chang, W.-C. Chen, Y.-D. Chih, T.-Y. J. Chang, In IEEE Int. Solid-State Circuits Conference (ISSCC) (2018), pp. 478-480.
24. F. Arnaud, P. Zuliani, J.P. Reynard, A. Gandolfo, F. Disegni, P. Mattavelli, E. Gomiero, G. Samanni, C. Jahan, R. Berthelon, O. Weber, E. Richard, V. Barral, A. Villaret, S. Kohler, J.C. Grenier, R. Ranica, C. Gallon, A. Souhaite, D. Ristoiu, L. Favennec, V. Caubet, S. Delmedico, N. Cherault, R. Beneyton, S. Chouteau, P.O. Sassoulas, A. Vernhet, Y. Le Friec, F. Domengie, L. Scotti, D. Pacelli, J.L. Ogier, F. Boucard, S. Lagrasta, D. Benoit, L. Clement, P. Boivin, P. Ferreira, R. Annunziata, and P. Cappelletti, In IEEE Int. Electron Devices Meeting (IEDM) (2018), pp.18.4.1-18.4.4.
25. D. Shum, D. Houssameddine, S. T. Woo, Y. S. You, J. Wong, K. W. Wong, C. C. Wang, K. H. Lee, K. Yamane, V. B. Naik, C. S. Seet, T. Tahmasebi, C. Hai, H. W. Yang, N. Thiyagarajah, R. Chao, J. W. Ting, N. L. Chung, T. Ling, T. H. Chan, S. Y. Siah, R. Nair, S. Deshpande, R. Whig, K. Nagel, S. Aggarwal, M. DeHerrera, J. Janesky, M. Lin, H.-J. Chia, M. Hossain, H. Lu, S. Ikegawa, F. B. Mancoff, G. Shimon, J. M. Slaughter, J. J. Sun, M. Tran, S. M. Alam, and T. Andre, In Symposium on VLSI Technology, IEEE (2017), pp. T208-T209.
26. D. Ielmini, Microelectron. Eng. **190**, 44 (2018).

27. P. Lin, C. Li, Z. Wang, Y. Li, H. Jiang, W. Song, M. Rao, Y. Zhuo, N. K. Upadhyay, M. Barnell, Q. Wu, J. J. Yang, and Q. F. Xia, Nat. Electron. **3**, 225 (2020).
28. S.-G. Park, M. K. Yang, H. Ju, D.-J. Seong, J. M. Lee, E. Kim, S. Jung, L. J. Zhang, Y. C. Shin, I.-G. Baek, J. Choi, H.-K. Kang, and C. Chung, In IEEE Int. Electron Devices Meeting (IEDM) (2012), pp. 20.8.1-20.8.4.
29. S. Yu, H.-Y. Chen, B. Gao, J. Kang, and H.-S. P. Wong, ACS Nano **7**, 2320 (2013).
30. C.-W. Hsu, C.-C. Wan, I.-T. Wang, M.-C. Chen, C.-L. Lo, Y.-J. Lee, W.-Y. Jang, C.-H. Lin, and T.-H. Hou, In IEEE Int. Electron Devices Meeting (IEDM) (2013), pp. 10.4.1-10.4.4.
31. I.-T. Wang, Y.-C. Lin, Y.-F. Wang, C.-W. Hsu, and T.-H. Hou, In IEEE Int. Electron Devices Meeting (IEDM) (2014), pp. 28.5.1-28.5.4.
32. M. Yu, Y. Cai, Z. Wang, Y. Fang, Y. Liu, Z. Yu, Y. Pan, Z. Zhang, J. Tan, X. Yang, M. Li, and R. Huang, Sci. Rep. **6**, 21020 (2016).
33. F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer, and D. Ielmini, IEEE Trans. Electron Devices **59**, 2461 (2012).
34. S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, and D. Ielmini, IEEE Trans. Electron Devices **59**, 2468 (2012).
35. D. Ielmini, IEEE Trans. Electron Devices **58**, 4309 (2011).
36. Z. Sun, E. Ambrosi, G. Pedretti, A. Bricalli, and D. Ielmini, IEEE Trans. Electron Devices **67**, 1466 (2020).
37. A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez, and D. Ielmini, IEEE Trans. Electron Devices **65**, 115 (2018).
38. A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez, and D. Ielmini, IEEE Trans. Electron Devices **65**, 122 (2018).
39. W. Chen, H. J. Barnaby, and M. N. Kozicki, IEEE Electron Device Lett. **37**, 580 (2016).
40. J. Song, J. Woo, A. Prakash, D. Lee, and H. Hwang, IEEE Electron Device Lett. **36**, 681 (2015).
41. 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, and J. J. Yang, Nat. Mater. **16**, 101 (2017).
42. R. Midya, Z. Wang, J. Zhang, S. E. Savel'ev, C. Li, M. Rao, M. H. Jang, S. Joshi, H. Jiang, P. Lin, K. Norris, N. Ge, Q. Wu, M. Barnell, Z. Li, H. L. Xin, R. S. Williams, Q. Xia, and J. J. Yang, Adv. Mater. **29**, 1604457 (2017).
43. W. Wang, M. Laudato, E. Ambrosi, A. Bricalli, E. Covi, Y.-H. Lin, and D. Ielmini, IEEE Trans. Electron Devices **66**, 3795 (2019).
44. W. Wang, M. Laudato, E. Ambrosi, A. Bricalli, E. Covi, Y.-H. Lin, and D. Ielmini, IEEE Trans. Electron Devices **66**, 3802 (2019).
45. W. Wang, M. Wang, E. Ambrosi, A. Bricalli, M. Laudato, Z. Sun, X. Chen and D. Ielmini, Nat. Commun. **10**, 1 (2019).
46. M. Wang, W. Wang, W. R. Leow, C. Wan, G. Chen, Y. Zeng, J. Yu, Y. Liu, P. Cai, H. Wang, D. Ielmini, and X. Chen, Adv. Mater. **30**, 1802516 (2018).
47. T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, and M. Aono, Nat. Mater. **10**, 591 (2011).
48. S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, IEEE Trans. Electron Devices **61**, 2912 (2014).
49. V. Milo, C. Zambelli, P. Olivo, E. Pérez, M. K. Mahadevaiah, O. G. Ossorio, Ch. Wenger, and D. Ielmini, APL Mater. **7**, 081120 (2019).
50. A. Calderoni, S. Sills, C. Cardon, E. Faraoni, and N. Ramaswamy, Microelectron. Eng. **147**, 145 (2015).
51. S. Ambrogio, S. Balatti, V. McCaffrey, D. C. Wang, and D. Ielmini, IEEE Trans. Electron Devices **62**, 3812 (2015).

52. M. Rao, Z. Wang, C. Li, H. Jiang, R. Midya, P. Lin, D. Belkin, W. Song, S. Asapu, Q. Xia, and J. J. Yang, In IEEE Int. Electron Devices Meeting (IEDM) (2019), pp.35.4.1-35.4.4.
53. T.-J. Yang and V. Sze, In IEEE Int. Electron Devices Meeting (IEDM) (2019), pp. 22.1.1-22.1.4.
54. W. Maass, Proc. IEEE **102**, 860 (2014).
55. M. R. Mahmoodi, M. Prezioso, and D. B. Strukov, Nat. Commun. **10**, 5113 (2019).
56. F. Cai, S. Kumar, T. V. Vaerenbergh, X. Sheng, R. Liu, C. Li, Z. Liu, M. Foltin, S. Yu, Q. Xia, J. J. Yang, R. Beausoleil, W. D. Lu, and J. P. Strachan, Nat. Electron. **3**, 409 (2020).
57. K. Yang, Q. Duan, Y. Wang, T. Zhang, Y. Yang, and R. Huang, Sci. Adv. **6**, eaba9901 (2020).
58. G. Pedretti, P. Mannocci, S. Hashemkhani, V. Milo, O. Melnic, E. Chicca, and D. Ielmini, IEEE J. Explor. Solid-State Computat. Devices Circuits **6**, 89 (2019).
59. T. Dalgaty, N. Castellani, C. Turck, K.-E. Harabi, D. Querlioz, and E. Vianello, Nat. Electron. **4**, 151 (2021).
60. D. Mantegazza, D. Ielmini, A. Pirovano, A. L. Lacaita, E. Varesi, F. Pellizzer, and R. Bez, Solid-State Electron. **52**, 584 (2008).
61. H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, and S. Slesazeck, ACS Appl. Mater. Interfaces **9**, 3792 (2017).
62. S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, IEEE Trans. Electron Devices **61**, 2920 (2014).
63. S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P.-Y. Chen, H. Yeon, S. Yu and J. Kim, Nat. Mater. **17**, 335 (2018).
64. A. Sawa, Mater. Today **11**, 28 (2008).
65. M. Hasan, R. Dong, H. J. Choi, D. S. Lee, D.-J. Seong, M. B. Pyun, and H. Hwang, Appl. Phys. Lett. **92**, 202102 (2008).
66. I.-T. Wang, C.-C. Chang, L.-W. Chiu, T. Chou, and T.-H. Hou, Nanotechnology **27**, 365204 (2016).
67. D. Ielmini and G. Pedretti, Adv. Intell. Syst. 2000040 (2020).
68. E. J. Fuller, F. E. Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, and A. A. Talin, Adv. Mater. **29**, 1604310 (2017).
69. J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S. Shin, K.-T. Lee, P. Solomon, K. Chan, W. Haensch, and J. Rozen, In IEEE Int. Electron Devices Meeting (IEDM) (2018), pp. 13.1.1-13.1.4.
70. Y. Li, E. J. Fuller, S. Asapu, S. Agarwal, T. Kurita, J. J. Yang, and A. A. Talin, ACS Appl. Mater. Interfaces **11**, 38982 (2019).
71. Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. A. Talin, and A. Salleo, Nat. Mater. **16**, 414 (2017).
72. E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, and A. A. Talin, Science **364**, 570 (2019).
73. S. Kim, J. A. Ott, T. Ando, H. Miyazoe, V. Narayanan, J. Rozen, T. Todorov, M. Onen, T. Gokmen, D. Bishop, P. Solomon, K.-T. Lee, M. Copel, and D. B. Farmer, In IEEE Int. Electron Devices Meeting (IEDM) (2019), pp. 35.7.1-35.7.4.
74. Y. Li, E. J. Fuller, S. Yoo, D. S. Ashby, C. H. Bennett, R. D. Horton, M. S. Bartsch, M. J. Marinella, W. D. Lu, and A. A. Talin, Adv. Mater. **32**, 2003984 (2020).
75. V. K. Sangwan, H.-S. Lee, H. Bergeron, I. Balla, M. E. Beck, K.-S. Chen, and M. C. Hersam, Nature **554**, 500 (2018).
76. V. K. Sangwan, D. Jariwala, I. S. Kim, K.-S. Chen, T. J. Marks, L. J. Lauhon, and M. C. Hersam, Nat. Nanotechnol. **10**, 403 (2015).
77. G. Fiori, F. Bonaccorso, G. Iannaccone, T. Palacios, D.Neumaier, A. Seabaugh, S. K. Banerjee, and L. Colombo, Nat. Nanotechnol. **9**, 768 (2014).
78. X. Zhu, D. Li, X. Liang, and W. D. Lu, Nat. Mater. **18**, 141 (2019).

79. L. F. Abbott and S. B. Nelson, Nat. Neurosci. **3**, 1178 (2000).
80. M. R. Azghadi, N. Iannella, S. F. Al-Sarawi, G. Indiveri, and D. Abbott, Proc. IEEE **102**, 717 (2014).
81. S. Mitra, S. Fusi, and G. Indiveri, IEEE Trans. Biomed. Circuits Syst. **3**, 32 (2009).
82. J.-P. Pfister and W. Gerstner, Adv. Neural Inf. Process. Syst. 1081 (2006).
83. J. Gjorgjieva, C. Clopath, J. Audet, and J.-P. Pfister, PNAS **108**, 19383 (2011).
84. W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).
85. A. L. Hodgkin and A. F. Huxley, J. Physiol. **116**, 473 (1952).
86. S. N. Truong and K.-S. Min, JSTS: J. Semicond. Tech. Sci. **14**, 356 (2014).
87. C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, Nat. Electron. **1**, 52 (2018).
88. 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, and J. P. Strachan, Adv. Mater. **30**, 1705914 (2018).
89. M. A. Zidan, J. P. Strachan, and W. D. Lu, Nat. Electron. **1**, 22 (2018).
90. D. Ielmini and H.-S. Philip Wong, Nat. Electron. **1**, 333 (2018).
91. M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, Nature **521**, 61 (2015).
92. 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, and H. Hwang, IEEE Trans. Electron Devices **62**, 3498 (2015).
93. T. Gokmen and Y. Vlasov, Front. Neurosci. **10**, 333 (2016).
94. S. Yu, Proc. IEEE **106**, 260 (2018).
95. P. M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang, and W. D. Lu, Nat. Nanotechnol. **12**, 784 (2017).
96. M. Le Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, and E. Eleftheriou, Nat. Electron. **1**, 246 (2018).
97. M. Le Gallo, A. Sebastian, G. Cherubini, H. Giefers, and E. Eleftheriou, IEEE Trans. Electron Devices **65**, 4304 (2018).
98. M. A. Zidan, Y. Jeong, J. Lee, B. Chen, S. Huang, M. J. Kushner, and W. D. Lu, Nat. Electron. **1**, 411 (2018).
99. Z. Sun, G. Pedretti, E. Ambrosi, A. Bricalli, W. Wang, and D. Ielmini, PNAS **116**, 4123 (2019).
100. Z. Sun, G. Pedretti, A. Bricalli, and D. Ielmini, Sci. Adv. **6**, eaay2378 (2020).
101. M. F. Bear, B. W. Connors, and M. A. Paradiso, Philadelphia, USA (2007).
102. A. N. Burkitt, Biol. Cybern. **95**, 1 (2006).
103. E. M. Izhikevich, IEEE Trans. Neural Networks **14**, 1569 (2003).
104. W. Gerstner and W. M. Kistler, Cambridge university press (2002).
105. E. N. Miranda, Acta Biotheor. **45**, 171 (1997).
106. W. Gerstner and R. Naud, Science **326**, 379 (2009).
107. G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K.Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen, Front. Neurosci. **5**, 73 (2011).
108. G. Indiveri, E. Chicca, and R. Douglas, IEEE Trans. on Neural Networks **17**, 211 (2006).
109. G. Indiveri, B. Linares-Barranco, R. Legenstein, G. Deligeorgis, and T. Prodromakis, Nanotechnology **24**, 384010 (2013).
110. C. Zamarreño-Ramos, L. A. Camuñas-Mesa, J. A. Pérez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, and B. Linares-Barranco, Front. Neurosci. **5**, 26 (2011).
111. G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, Sci. Rep. **7**, 5288 (2017).

112. S. Lashkare, S. Chouhan, T. Chavan, A. Bhat, P. Kumbhare, and U. Ganguly, IEEE Electron Device Lett. **39**, 484 (2018).
113. I. Gupta, A. Serb, A. Khiat, R. Zeitler, S. Vassanelli, and T. Prodromakis, Nat. Commun. **7**, 12805 (2016).
114. D. Lee, M. Kwak, K. Moon, W. Choi, J .Park, J. Yoo, J. Song, S. Lim, C. Sung, W. Banerjee, and H. Hwang, Adv. Electron. Mater. **5**, 1800866 (2019).
115. R. Yang, H.-M. Huang, Q.-H. Hong, X.-B. Yin, Z.-H. Tan, T. Shi, Y.-X. Zhou, X.-S. Miao, X.-P. Wang, S.-B. Mi, C.-L. Jia, and X. Guo, Adv. Funct. Mater. **28**, 1704455 (2018).
116. M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, Nat. Mater. **12**, 114 (2013).
117. H. Lim, V. Kornijcuk, J. Y. Seok, S. K. Kim, I. Kim, C. S. Hwang, and D. S. Jeong, Sci. Rep. **5**, 9776 (2015).
118. C. Adda, B. Corraze, P. Stoliar, P. Diener, J. Tranchant, A. Filatre-Furcate, M. Fourmigué, D. Lorcy, M.-P. Besland, E. Janod, and L. Cario, J. Appl. Phys. **124**, 152124 (2018).
119. A. Mehonic and A. J. Kenyon, Front. Neurosci. **10**, 57 (2016).
120. 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, and J. J. Yang, Nat. Electron. **1**, 137 (2018).
121. H.-M. Huang, R. Yang, Z.-H. Tan, H.-K. He, W. Zhou, J. Xiong, and X. Guo, Adv. Mater. **31**, 1803849 (2019).
122. X. Zhang, W. Wang, Q. Liu, X. Zhao, J. Wei, R. Cao, Z. Yao, X. Zhu, F. Zhang, H. Lv, S. Long, and M. Liu, IEEE Electron Device Lett. **39**, 308 (2018).
123. Q. Hua, H. Wu, B. Gao, Q. Zhang, W. Wu, Y. Li, X. Wang, W. Hu, and H. Qian, Glob. Chall. **3**, 1900015 (2019).
124. H. Kalita, A. Krishnaprasad, N. Choudhary, S. Das, D. Dev, Y. Ding, L. Tetard, H.-S.Chung, Y. Jung, and T. Roy, Sci. Rep. **9**, 53 (2019).
125. M. Ignatov, M. Ziegler, M. Hansen, A. Petraru, and H. Kohlstedt, Front. Neurosci. **9**, 376 (2015).
126. M. S. Feali, A. Ahmadi, and M. Hayati, Neurocomputing **309**, 157 (2018).
127. Z. Wang, M. Rao, J.-W. Han, J. Zhang, P. Lin, Y. Li, C. Li, W. Song, S. Asapu, R. Midya, Y. Zhuo, H. Jiang, J. H. Yoon, N. K. Upadhyay, S. Joshi, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, Q. Qiu, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Commun. **9**, 3208 (2018).
128. P. Stoliar, J. Tranchant, B. Corraze, E. Janod, M.-P. Besland, F. Tesler, M. Rozenberg, and L. Cario, Adv. Funct. Mater. **27**, 1604740 (2017).
129. Y. Zhang, W. He, Y. Wu, K. Huang, Y. Shen, J. Su, Y. Wang, Z. Zhang, X. Ji, G. Li, H. Zhang, S. Song, H. Li, L. Sun, R. Zhao, and L. Shi, Small **14**, 1802188 (2018).
130. S. Slesazeck, H. Mähne, H. Wylezich, A. Wachowiak, J. Radhakrishnan, A. Ascoli, R. Tetzlaff, and T. Mikolajick, RSC Adv. **5**, 102318 (2015).
131. G. A. Gibson, S. Musunuru, J. Zhang, K. Vandenberghe, J. Lee, C.-C. Hsieh, W. Jackson, Y. Jeon, D. Henze, Z. Li, and R. S. Williams, Appl. Phys. Lett. **108**, 023505 (2016).
132. C. Funck, S. Menzel, N. Aslam, H. Zhang, A. Hardtdegen, R. Waser, and S. Hoffmann-Eifert, Adv. Electron. Mater. **2**, 1600169 (2016).
133. Z. Wang, S. Kumar, Y. Nishi, and H.-S. P. Wong, Appl. Phys. Lett. **112**, 193503 (2018).
134. X. Liu, S. Li, S. K. Nandi, D. K. Venkatachalam, and R. G. Elliman, J. Appl. Phys. **120**, 124102 (2016).
135. S. Kumar, Z. Wang, N. Davila, N. Kumari, K. J. Norris, X. Huang, J. P. Strachan, D. Vine, A. L. D. Kilcoyne, Y. Nishi, and R. S. Williams, Nat. Commun. **8**, 658 (2017).
136. W. Yi, K. K. Tsang, S. K. Lam, X. Bai, J. A. Crowell, and E. A. Flores, Nat. Commun. **9**, 4661 (2018).

137. T. C. Jackson, A. A. Sharma, J. A. Bain, J. A. Weldon, and L. Pileggi, IEEE. J. Emerg. Sel. Topics Circuits. Syst. **5**, 230 (2015).
138. R. Midya, Z. Wang, S. Asapu, S. Joshi, Y. Li, Y. Zhuo, W. Song, H. Jiang, N. Upadhay, M. Rao, P. Lin, C. Li, Q. Xia, and J. J. Yang, Adv. Electron. Mater. **5**, 1900060 (2019).
139. J. J. Moore, P. M. Ravassard, D. Ho, L. Acharya, A. L. Kees, C. Vuong, and M. R. Mehta, Science **355**, eaaj1497 (2017).
140. T. Branco and M. Häusser, Curr. Opin. Neurobiol. **20,** 494 (2010).
141. A. Losonczy, J. K. Makara, and J. C. Magee, Nature **452**, 436 (2008).
142. S. Trenholm, A. J. McLaughlin, D. J. Schwab, M. H. Turner, R. G. Smith, F. Rieke, and G. B. Awatramani, Nat. Neurosci. **17,** 1759 (2014).
143. S. D. Antic, W.-L. Zhou, A. R. Moore, S. M. Short, and K. D. Ikonomu, J. Neurosci. Res. **88,** 2991 (2010).
144. M. Lavzin, S. Rapoport, A. Polsky, L. Garion, and J. Schiller, Nature **490,** 397 (2012).
145. A. Bhaduri, A. Banerjee, S. Roy, S. Kar, and A. Basu, Neural. Comput. **30,** 723 (2018).
146. J. Schemmel, L. Kriener, P. Müller, K. Meier, In International Joint Conference on Neural Networks (IJCNN) (2017), pp. 2217-2226.
147. W. T. Gao, L. Q. Zhu, J. Tao, D. Y. Wan, H. Xiao, and F. Yu, ACS Appl. Mater. Interfaces **10,** 40008 (2018).
148. C. J. Wan, L. Q. Zhu, J. M. Zhou, Y. Shi, and Q. Wan, Nanoscale **6,** 4491 (2014).
149. L. Q. Zhu, C. J. Wan, P. Q. Gao, Y. H. Liu, H. Xiao, J. C. Ye, and Q. Wan, ACS Appl. Mater. Interfaces **8,** 21770 (2016).
150. X. Li, J. Tang, Q. Zhang, B. Gao, J. J. Yang, S. Song, W. Wu, W. Zhang, P. Yao, N. Deng, L. Deng, Y. Xie, H. Qian, and H. Wu, Nat. Nanotech. **15,** 776 (2020).
151. Y. Lin, Z. Wang, X. Zhang, T. Zeng, L. Bai, Z. Kang, C. Wang, X. Zhao, H. Xu, and Y. Liu, NPG Asia Mater. **12**, 64 (2020).
152. X. Shan, Z. Wang, Y. Lin, T. Zeng, X. Zhao, H. Xu, and Y. Liu, Adv. Electron. Mater. **6**, 2070039 (2020).
153. Y. Park and J.-S. Lee, ACS Nano **11**, 8962 (2017).
154. S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, Nano Lett. **10**, 1297 (2010).
155. Z. Q. Wang, H. Y. Xu, X. H. Li, H. Yu, Y. C. Liu, and X. J. Zhu, Adv. Funct. Mater. **22**, 2759 (2012).
156. T. Chang, S.-H. Jo, and W. Lu, ACS Nano **5**, 7669 (2011).

157. L. Sun, Y. Zhang, G. Hwang, J. Jiang, D. Kim, Y. A. Eshete, R. Zhao, and H. Yang, Nano Lett. **18**, 3229 (2018).
158. P. K. R. Boppidi, B. Suresh, A. Zhussupbekova, P. Biswas, D. Mullarkey, P. M. P. Raj, I. V. Shvets, and S. Kundu, IEEE Trans. Electron Devices **67**, 3451 (2020).
159. Z. Y. Ren, L. Q. Zhu, Y. B. Guo, T. Y. Long, F. Yu, H. Xiao, and H. L. Lu, ACS Appl. Mater. Interfaces **12**, 7833 (2020).
160. Z. Wang, S. Ambrogio, S. Balatti, S. Sills, A. Calderoni, N. Ramaswamy, and D. Ielmini, IEEE Trans. Electron Devices **63**, 4279 (2016).
161. X. Zhao, H. Xu, Z. Wang, L. Zhang, J. Ma, and Y. Liu, Carbon **91**, 38 (2015).
162. X. Li, H. Wu, B. Gao, W. Wu, D. Wu, N. Deng, J. Cai, and H. Qian, Nanotechnology **27**, 305201 (2016).
163. X. Zhao, Z. Wang, W. Li, S. Sun, H. Xu, P. Zhou, J. Xu, Y. Lin, and Y. Liu, Adv. Funct. Mater. **30**, 1910151 (2020).
164. P.-Y. Chen, B. Lin, I.-T. Wang, T.-H. Hou, J. Ye, S. Vrudhula, J.-S. Seo, Y. Cao, and S. Yu, In IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (2015), pp. 194-199.

165. Z. Chen, B. Gao, Z. Zhou, P. Huang, H. Li, W. Ma, D. Zhu, L. Liu, X. Liu, J. Kang, and H.-Y.Chen, In IEEE Int. Electron Devices Meeting (IEDM) (2015), pp. 17.7.1-17.7.4.
166. S. Park, M. Chu, J. Kim, J. Noh, M. Jeon, B. H. Lee, H. Hwang, B. Lee, and B. G. Lee, Sci. Rep. **5**, 10123 (2015).
167. P.-Y. Chen, L. Gao, and S. Yu, IEEE Trans Multi-Scale Comput Syst. **2**, 257 (2016).
168. S. Park, A. Sheri, J. Kim, J. Noh, J. Jang, M. Jeon, B. Lee, B. R. Lee, B. H. Lee, and H. Hwang, In IEEE Int. Electron Devices Meeting (IEDM) (2013), pp. 25.6.1-25.6.4.
169. S. Agarwal, S. J. Plimpton, D. R. Hughart, A. H. Hsia, I. Richter, J. A Cox, C. D. James, and M. J. Marinella, In International Joint Conference on Neural Networks (IJCNN) (2016), pp.929-938.
170. Y. Lin, C. Wang, Y. Ren, Z. Wang, H. Xu, X. Zhao, J. Ma, and Y. Liu, Small Methods, **3**, 1900160, (2019).
171. T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco, Front. Neurosci. **7**, 2(2013).
172. G. S. Snider, In IEEE International Symposium on Nanoscale Architectures (NANOARCH) (2008), pp.85-92.
173. S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S. Wong, IEEE Trans. Electron Devices **58**, 2729 (2011).
174. Z. Wang, S. Ambrogio, S. Balatti, and D. Ielmini, Front. Neurosci. **8**, 438 (2015).
175. Z.-H. Tan, R. Yang, K. Terabe, X. B. Yin, X.-D. Zhang, and X. Guo, Adv. Mater. **28**, 377 (2016).
176. M. Prezioso, F. M. Bayat, B. Hoskins, K. Likharev, and D. Strukov, Sci. Rep. **6**, 21331 (2016).
177. Y. Lin, T. Zeng, H. Xu, Z. Wang, X. Zhao, W. Liu, J. Ma, and Y. Liu, Adv. Electron. Mater. **4**, 1800373 (2018).
178. Y. Li, Y. Zhong, J. Zhang, L. Xu, Q. Wang, H. Sun, H. Tong, X. Cheng, and X. Miao, Sci. Rep. **4**, 4906 (2014).
179. D. Kuzum, R. G. D. Jeyasingh, B. Lee, and H.-S. Philip Wong, Nano Lett. **12**, 2179 (2012).
180. T. Tuma, M. L. Gallo, A. Sebastian, and E. Eleftheriou, IEEE Electron Device Lett. **37**, 1238 (2016).
181. G. Srinivasan, A. Sengupta, and K. Roy, Sci. Rep. **6**, 29545 (2016).
182. S. Boyn, J. Grollier, G. Lecerf, B. Xu, N. Locatelli, S. Fusil, S. Girod, C. Carrétéro, K. Garcia, S. Xavier, J. Tomas, L. Bellaiche, M. Bibes, A. Barthélémy, S. Saïghi, and V. Garcia, Nat. Commun. **8**, 14736 (2017).
183. 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, and C. Lam, In IEEE Int. Electron Devices Meeting (IEDM) (2015), pp. 17.1.1-17.1.4.
184. J. Xu, X. Zhao, Z. Wang, H. Xu, J. Hu, J. Ma, and Y. Liu, Small **15**, 1803970 (2019).
185. S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z.-Q. Wang, A. Calderoni, N. Ramaswamy, and D. Ielmini, IEEE Trans. Electron Devices **63**, 1508 (2016).
186. V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, and D. Ielmini, In IEEE Int. Electron Devices Meeting (IEDM) (2016), pp. 16.8.1-16.8.4.
187. V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, and D. Ielmini, IEEE Transactions on Very Large Scale Integration (VLSI) Systems **26**, 2806 (2018).
188. S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, and W. D. Lu, Nano Lett. **15**, 2203 (2015).
189. C. Du, W. Ma, T. Chang, P. Sheridan, and W. D. Lu, Adv. Funct. Mater. **25**, 4290 (2015).
190. Z.-H. Tan, X.-B. Yin, R. Yang, S.-B. Mi, C.-L. Jia, and X. Guo, Sci. Rep. **7**, 713 (2017).
191. X. Yan, J. Zhao, S. Liu, Z. Zhou, Q. Liu, J. Chen, and X. Y. Liu, Adv. Funct. Mater. **28**, 1705320 (2018).
192. J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A.Reyes, Nature **448**, 802 (2007).

193. G. Liu, C. Wang, W. Zhang, L. Pan, C. Zhang, X. Yang, F. Fan, Y. Chen, and R.-W. Li, Adv. Electron. Mater. **2**, 1500289 (2016).
194. Y. Wang, Z. Zhang, M. Xu, Y. Yang, M. Ma, H. Li, J. Pei, and L. Shi, ACS Appl. Mater. Interfaces **11**, 24230 (2019).
195. A. Kirkwood, M. G. Rioult, and M. F. Bear, Nature **381**, 526 (1996).
196. M. F. Bear, L. N. Cooper, and F. F. Ebner, Science **237**, 42 (1987).
197. L. N. Cooper and M. F. Bear, Nat. Rev. Neurosci. **13**, 798 (2012).
198. H. Z. Shouval, PNAS **108**, 19103 (2011).
199. Z. Wang, T. Zeng, Y. Ren, Y. Lin, H. Xu, X. Zhao, Y. Liu, and D. Ielmini, Nat. Commun. **11**, 1510 (2020).
200. J. Xiong, R. Yang, J. Shaibo, H.-M. Huang, H.-K. He, W. Zhou, and X. Guo, Adv. Funct. Mater. **29**, 1807316 (2019).
201. H.-X. Wang, R. C. Gerkin, D. W. Nauen, and G.-Q. Bi, Nat. Neurosci. **8**, 187 (2005).
202. R. A. J. Lester and C. E. Jahr, J. Neurosci. **12**, 635 (1992).
203. W. Wang, E. Covi, A. Milozzi, M. Farronato, S. Ricci, C. Sbandati, G. Pedretti, and D. Ielmini, Adv. Intell. Syst. 2000224 (2020).
204. M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, and F. Xiong, Adv. Mater. **30**, 1802353 (2018).
205. J. Zhu, Y. Yang, R. Jia, Z. Liang, W. Zhu, Z. U. Rehman, L. Bao, X. Zhang, Y. Cai, L. Song, and R. Huang, Adv. Mater. **30**, 1800195 (2018).
206. Y. H. Liu, L. Q. Zhu, P. Feng, Y. Shi, and Q. Wan, Adv. Mater. **27**, 5599 (2015).
207. X. Yao, K. Klyukin, W. Lu, M. Onen, S. Ryu, D. Kim, N. Emond, I. Waluyo, A. Hunt, J. A. del Alamo, J. Li, and B. Yildiz, Nat. Commun. **11**, 3134 (2020).
208. L. Zhou, S. Yang, G. Ding, J.-Q. Yang, Y. Ren, S.-R. Zhang, J.-Y. Mao, Y. Yang, Y. Zhou, and S.-T. Han, Nano Energy **58**, 293 (2019).
209. X. Yan, Z. Zhou, J. Zhao, Q. Liu, H. Wang, G. Yuan, and J. Chen, Nano Res. **11**, 1183 (2018).
210. M.-K. Kim and J.-S. Lee, ACS Nano **12**, 1680 (2018).
211. H. Chen, C. Liu, Z. Wu, Y. He, Z. Wang, H. Zhang, Q. Wan, W. Hu, D. W. Zhang, M. Liu, Q. Liu, and P. Zhou, Adv. Sci. **6**, 1901072 (2019).
212. H. B. Barlow, R. M. Hill, and W. R. Levick, J. Physiol. **173**, 377 (1964).
213. C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Nat. Commun. **8**, 2204 (2017).
214. J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, Nat. Electron. **2,** 480 (2019).
215. R. Midya, Z. Wang, S. Asapu, X. Zhang, M. Rao, W. Song, Y. Zhuo, N. Upadhyay, Q. Xia, and J. J. Yang, Adv. Intell. Syst. **1**, 1900084 (2019).
216. G. Milano, G. Pedretti, M. Fretto, L. Boarino, F. Benfenati, D. Ielmini, I. Valov, and C. Ricciardi, Adv. Intell. Syst. **2**, 2000096 (2020).
217. Y. H. Lin, C. H. Wang, M. H. Lee, D. Y. Lee, Y. Y. Lin, F. M. Lee, H. L. Lung, K. C. Wang, T. Y. Tseng, and C. Y. Lu, IEEE Trans. Electron Devices **66**, 1289 (2019).
218. V. Milo, F. Anzalone, C. Zambelli, E. Pérez, M. Mahadevaiah, O. Ossorio, P. Olivo, C. Wenger, and D. Ielmini, IRPS 2C.4 (2021).
219. I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, and E. Eleftheriou, Nat. Commun. **9**,2514 (2018).
220. D. Ielmini, F. Nardi, C. Cagli, and A. L. Lacaita, IEEE Electron Device Lett. **31**, 353 (2010).
221. W. He, W. Shim, S. Yin, X. Sun, D. Fan, S. Yu, and J.-S. Seo, IRPS 2C.3 (2021).
222. B. Hoefflinger, "ITRS: The international technology roadmap for semiconductors," in Chips 2020, 161 (Springer, 2011).
223. Q. Wang, X. Wang, S. H. Lee, F.-H. Meng, and W. D. Lu, in IEEE Int. Electron Devices Meeting (IEDM)(2019), pp. 14.4.1-14.4.4.

224. W. Zhang, B. Gao, J. Tang, P. Yao, S. Yu, M.-F. Chang, H.-J. Yoo, H. Qian, and H. Wu, Nat. Electron. **3**, 371 (2020).
225. 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-Köpe, Y. Nishi, and H. Hwang, In IEEE Int. Electron Devices Meeting (IEDM) (2013), pp. 268-271.
226. M. Son, J. Lee, J. Park, J. Shin, G. Choi, S. Jung, W. Lee, S. Kim, S. Park, and H. Hwang, IEEE Electron Device Lett. **32**, 1579 (2011).
227. X. Liu, S. M. Sadaf, M. Son, J. Park, J. Shin, W. Lee, K. Seo, D. Lee, and H. Hwang, IEEE Electron Device Lett. **33**, 236 (2012).
228. D. Kau, S. Tang, I. V. Karpov, R. Dodge, B. Klehn, J. A. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K.-W. Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro, and G. Spadini, In IEEE Int. Electron Devices Meeting (IEDM) (2009), pp. 617-620.
229. M.-J. Lee, D. Lee, S.-H. Cho, J.-H. Hur, S.-M. Lee, D. H. Seo, D.-S. Kim, M.-S. Yang, S. Lee, E. Hwang, M. R. Uddin, H. Kim, U.-I. Chung, Y. Park, and I.-K. Yoo, Nat. Commun. **4**, 2629 (2013).
230. K. Virwani, G. W. Burr, R. S. Shenoy, C. T. Rettner, A. Padilla, T. Topuria, P. M. Rice, G. Ho, R. S. King, K. Nguyen, A. N. Bowers, M. Jurich, M. BrightSky, E. A. Joseph, A. J. Kellock, N. Arellano, B. N. Kurdi, and K. Gopalakrishnan, In IEEE Int. Electron Devices Meeting (IEDM) (2012), pp. 2.7.1-2.7.4.
231. W. Lee, J. Park, S. Kim, J. Woo, J. Shin, G. Choi, S. Park, D. Lee, E. Cha, B. H. Lee, and H. Hwang, ACS Nano **6**, 8166 (2012).
232. C. H. Ho, H.-H. Huang, M.-T. Lee, C.-L.Hsu, T.-Y. Lai, W.-C. Chiu, M. Y. Lee, T.-H. Chou, I. Yang, M.-C. Chen, C.-S. Wu, K.-H. Chiang, Y.-D. Yao, C. Hu, and F.-L. Yang, In IEEE Int. Electron Devices Meeting (IEDM) (2012), p. 40.
233 D. Alfaro Robayo, G. Sassine, L. Grenouillet, C. Carabasse, T. Martin, N. Castellani, A. Verdy, G. Navarro, L. Ciampolini, B. Giraud, M. Bernard, T. Magis, V. Beugin, E. Vianello, G. Ghibaudo, G. Molas, E. Nowak, 2019 IEEE 11th International Memory Workshop (IMW) (2019).