Neuromorphic Engineering for Hardware Computational Acceleration and Biomimetic Perception Motion Integration

Shuiyuan Wang, Xiaozhang Chen, Xiaohe Huang, David Wei Zhang, and Peng Zhou*

In the tide of artificial intelligence evolution, the demand for data computing has exploded, and the von Neumann architecture computer with separate memory and computing units require cumbersome data interaction, which leads to serious degradation in performance and efficiency. Biologically inspired neuromorphic engineering performs digital/analog computations in memory, with massive parallelism and high energy efficiency, making it a promising candidate to get out of the woods. Memristive device-based artificial synapses and neurons are building blocks to form hardware neural networks for computing acceleration. In addition, it enables the implementation of integrated bionic perception and motion systems to mimic the human peripheral nervous system for information sensing and processing. Herein, the biological basis and inspiration are described first, and the memristive synapses and circuit-emulation neurons used for neuromorphic engineering are addressed and evaluated as well as the mechanisms. The computational acceleration and bionic perception motion integration of neuromorphic systems are discussed. Finally, the challenges and opportunities for neuromorphic engineering to accelerate computation and enrich biomimetic perception motion functions are prospected, and it is hoped that light is shed on future advances.

1. Introduction

The computing and memory blocks of traditional von Neumann architecture computers are physically separated.[1] Separate memory and computing modules lead to frequent data access and readout, which brings time and energy consumption of cumbersome data transmission interaction and ultimately encounters energy efficiency and performance limitations.[2] In the past, small data communications had a negligible impact on computational efficiency, and the von Neumann architecture has achieved great success. When the central processing unit (CPU) has to execute simple instructions on massive data, the CPU will be idle during data access and readout, and the data flow becomes the main restraining factor of the overall computing efficiency.[3] In particular, with the evolution of artificial intelligence (AI) and the popularization of big data and Internet of Things scenarios, applications have become more data centric, and performance and energy-efficiency bottlenecks caused by nonessential data interactions have become more serious.[4] The industry has made efforts to use a large number of parallel processors, namely graphics processing units (GPU)[5] or processors customized for dedicated applications,[6] but it is impossible to completely address the issues caused by data transmission. Another alternative idea is to shorten the distance between the memory and the processing unit as much as possible,[7] thereby reducing the time and energy costs required for data movement, the so-called near-memory computing architecture that came into being.[8]

Although the advancement of through silicon via (TSV) die stacking interconnects and memory modules[9,10] booms near-memory computing, computational efficiency is restricted by the TSV interconnect density.[11] The state-of-the-art 3D monolithic integration achieves a higher integration density connection between the memory and the computational unit.[12] Nevertheless, it is worth noting that all the aforementioned technologies aimed at reducing the time and distance of data traffic have not fundamentally eliminated the physical isolation between memory and computing blocks.

The human brain, as a highly advanced system in nature, possesses massive parallelism, robust fault tolerance, self-adaptation, and self-learning ability in information processing, enabling it to efficiently perform various complex tasks, including learning summary, reasoning analogies, generalized memory, etc.[13] A biological neural network composed of approximately $10^{11}$ neurons and $10^{15}$ synapses undertakes the perceptual transmission and processing of environmental information.[14] Synapses, which are the basic units of the nervous system, are distributed and connected in parallel to a variety of neurons and can modulate and remember the activity of related neurons by changing the connection strength with the neurons, that is, the synaptic weight.[13] Neurons receive stimulation from the postsynaptic membrane, and when the membrane potential exceeds a certain threshold, the neurons

© 2020 The Authors. Published by Wiley-VCH GmbH
respond to activation and fire spikes, completing the transmission and processing of physiological signals. The bioplasticity of tunable synaptic weights enables diverse advanced neurobehaviors and functions such as learning and memory. In addition to conducting and processing neural signals, synapses and neurons in the human brain store the current strength of synaptic connections and form memories as well, perfectly interpreting the coexistence of memory and computation. Inspired by the human brain synapse and neuron network architecture, and fundamentally solving the data interaction dilemma of the traditional von Neumann architecture, the construction of high-performance and energy-efficient neuromorphic engineering applied in the era of AI and big data has attracted extensive research interest. Neuromorphic engineering originated from the use of very-large-scale-integrated (VLSI) analog circuits to imitate the structure of biological nervous systems, and now both digital and analog hardware paradigms have been fully developed for neuromorphic engineering. Recently, neuromorphic engineering based on complementary metal-oxide semiconductor (CMOS) technology has been developed for the optimization of special chips or accelerators, which has achieved a significant increase in computing performance. However, CMOS technology usually requires redundant transistors to build artificial synapses and neurons and is far from the biological brain in terms of efficiency, whereas the brain can work at a low frequency of 10 Hz and a low energy consumption of 20 W to realize the interaction and processing of information flow in complex worlds. In addition, the emerging metal-insulator–metal (MIM) phase change, ferroelectric, and van der Waals (vdW) memristive devices perform analog computing, and store information inside the device, without any separation like the brain, which shows great potential in the realization of neuromorphic engineering for energy-efficient computing acceleration. In short, devices that can simulate the functions of biological synapses and neurons are essential for neuromorphic engineering. They imitate the human brain scheme to conduct in situ parallel computation and store information in the hardware unit, enabling to get rid of the von Neumann architecture limitations, improve computing performance, and enhance energy efficiency.

Moreover, neuromorphic engineering simulation expands from the central nervous system where synapses and neurons are interactively connected to the peripheral nervous system, enabling the integration of bionic perception and motion functions. Generally, the artificial sensory neural system is composed of sensing, transmission, processing, and output modules, which can realize the detection and response of the external environment optical, mechanical pressure, acoustics, and other stimuli and correspondingly imitate the perception of vision, touch, and hearing in the human peripheral nervous system. Similar to the acceleration of energy-efficient computing, it is critical for the artificial perceptual neural system to realize artificial synaptic and neural devices with external signal perceiving and information processing functions and require sophisticated integration at the system level. In recent years, remarkable progress has been made in neuromorphic engineering for bionic perception imitation, including the exploration and combination of various inorganic and organic materials to implement the visual, tactile, auditory, and pain perception. In addition, the bionic technology of neuromorphic engineering for artificial afferent nerves with pressure detection, perceptual memory, and motion drive has been demonstrated. The sensorimotor system, which combines photoelectric synapses and neuromuscular components, has been reported to simulate the expansion and contraction of bio-fiber muscle as well. In summary, neuromorphic engineering for simulating the function of the perceptual nervous system has become a promising candidate in the field of biomimetic technology, such as intelligent robotics, neural prosthetics, and brain–machine interfaces and other bioelectronics.

Herein, we review the advanced progress made in neuromorphic engineering and light the way for the development of hardware-implemented energy-efficient neuromorphic computing and neurobionics perception technology. First, we introduce the concept of neuromorphic engineering and discuss the relevant biological basis in a comprehensive manner. Then we review the artificial memristive synapses by functional materials, followed by a detailed discussion of artificial neurons, both of which constitute the basic building blocks of neuromorphic engineering. Subsequently, we focused on the system-level hardware computing strategies and biomimetic technology applications of neuromorphic engineering. Specifically, the neural network acceleration, emerging computing implementation, and bionic perception motion integration subsections are discussed in sequence. Finally, we summarize the challenges and opportunities faced by neuromorphic engineering for computing acceleration and biomimetic perception motion technology and propose promising solutions and perspectives.

2. Biological Basis of Neuromorphic Engineering

Neuromorphic engineering, usually also referred to as neuromorphic computing, was proposed by Carver Mead and initially described the use of VLSI analog circuits to mimic the architecture of biological nervous systems. Nowadays, neuromorphic engineering has been extended to use analog, digital, and digital–analog hybrid VLSI or emerging memristive devices (such as metal oxides, phase change, ferroelectrics, spintronics, and van der Waals [vdW] memristors as well as electrolytic transistors, memtransistors etc.) to implement data-centric artificial neural networks (ANN) and multiple sensory integration and motion control bionic peripheral neural systems through software and hardware strategies. A crucial concept of biologically inspired neuromorphic engineering is to figure out how the biological nervous system expresses, processes, and stores information, develops learning, adapts to changes, and conducts desired functions and behaviors.

Specifically, synapses and neurons are the basic components of physiological signal transmission and processing in the biological nervous system. Various neurons are connected through synapses to form a complex, efficient, and versatile biological neural network. Intelligent behaviors such as memory and forgetting, learning, and decision-making take place in the intricate and advanced neural network. There are 10 neurons and...
$10^{15}$ synapses in the human nervous system, approximately. Neurons are composed of soma, axon, and dendrites as shown in Figure 1a. The soma has the function of receiving and integrating input information and transmitting it. The dendrites are short and branched and are directly expanded by the soma to form a dendritic shape, which is used to receive impulses from other neuron axons and transmit them to the soma. The axon is slender and has almost no branches and receive and transmit external stimuli. Most neurons receive signals through dendrites and somatic cells, then transport signals from axons, and finally transmit them to other neuron cells through synapses at the ends of axons as shown in Figure 1a. In addition, branch-like nerve endings will form at the ends of axons, which are distributed in tissues and organs, including sensory nerve endings that constitute various receptors and motor nerve endings that form different effectors.

The biological synapses consist of the presynaptic membrane, the synaptic cleft, and the postsynaptic membrane, which are the sites where functional connections occur between presynaptic and postsynaptic neurons and are also the key components for information processing, as shown in Figure 1b. Signals propagate from the axons of one neuron to the dendrites of another neuron by synapses, where neurons that emit signals are called presynaptic neurons and neurons that receive signals through synapses are postsynaptic neurons. Synapses enable neurons to transmit electrical or chemical signals to other neurons or related target cells, which correspond to electrical and chemical synapses, respectively. Chemical synapses, which convert electrical signals from presynaptic neurons to chemical signals, that is, neurotransmitters, use neurotransmitters to transfer information to postsynaptic neurons, whereas electrical synapses directly transmit information through electrical signals (i.e., ionic current in the synaptic cleft). It is worth noting that humans or mammals almost transmit information based on chemical synapses. Electrical synapses are mainly found in fish and amphibians. Therefore, the following discussion will focus on chemical synapses. Presynaptic cells contain many synaptic vesicles, which are characteristic structures of the presynapses and contain neurotransmitter chemicals. The nerve impulse is transmitted to the presynaptic membrane along the axon, triggering the opening of calcium channels on the presynaptic membrane; the calcium ions outside the cell enter the presynaptic cells, causing the synaptic vesicles to gradually move to the presynaptic membrane. Finally, the neurotransmitter in the vesicle is released to the synaptic cleft by exocytosis. There are receptors and chemically gated ion channels on the postsynaptic membrane, and as neurotransmitters bind to the corresponding receptors on the postsynaptic membrane, the chemically gated

Figure 1. Biological neurons and synapses and related AP response. a) Neurons are composed of soma, dendrites, and axons. The axons are wrapped with myelin sheaths, and the ends are connected to other neurons through synapses. b) Schematic diagram of a synapse. The neurotransmitters released by synaptic vesicles in presynaptic neurons are received by receptors on the postsynaptic membrane, thus modulating the activity of postsynaptic neurons. c) Ion exchange triggers AP generation, including resting, depolarization, and repolarization phases. d) Neuron response follows the all-or-none law. The neuron fully responds only when the stimulus intensity reaches the threshold; otherwise, no response occurs.
channels coupled to the receptors open, allowing related ions to enter the postsynaptic cells.[15] The distribution of ions on both sides of the postsynaptic membrane changes, showing an excitatory or inhibitory membrane potential, which affects the activity of postsynaptic neurons.[61] The synapse that excites the postsynaptic membrane is an excitatory synapse; otherwise, it is an inhibitory synapse.[58]

Neurotransmitter-induced changes in ion concentration on both sides of the cell membrane lead to the generation of action potentials (AP), a typical excitatory AP is shown in Figure 1c. Usually more potassium ions (K⁺) are accumulated in the cell membrane, and more sodium ions (Na⁺) are distributed outside.[62] The initial nerve cell is at rest, the Na⁺ channel is closed, and the inner and outer cell membrane potentials show a negative difference (that is, the resting potential), corresponding to stage I. With the coupling of neurotransmitters and receptors, some Na⁺ channels on the cell membrane are opened, and a small amount of Na⁺ gradually flows into the cell membrane. The local potential of the cell membrane changes, and the potential gradually increases to the threshold. Subsequently, more Na⁺ channels open, more Na⁺ floods into the cell, a depolarization effect occurs on the cell membrane, the membrane potential increases significantly, the AP is generated and peaked, and the neuron fired, corresponding to stage II. In stage III, the Na⁺ channel is closed, the K⁺ channel is opened, and K⁺ flows out of the cell to refresh the membrane to the original polarization state. After the membrane potential is reset to the resting potential, the K⁺ channel remains open, hyperpolarization of the cell membrane occurs, the membrane potential is lower than the resting potential, and then it recovers within a period of time.[17] In addition, the AP response in the neuron follows the all-or-none law,[63] that is, when the stimulation intensity does not reach the threshold, the AP cannot occur, and when the stimulation intensity reaches the value, the AP is generated and can reach the maximum intensity instantaneously, as shown in Figure 1d. Even if the stimulation intensity continues to increase or decrease, the intensity of the induced AP will no longer change. That is, the response of neurons to stimuli is independent of stimulus intensity. Once the AP is generated, it will be conducted along the axon to the end. During the conduction process, the AP keeps the full response intensity unchanged.[64] The depolarization of the excitatory synaptic membrane generates an excitatory AP, also known as the excitatory postsynaptic potential (EPSP) or current (EPSC).[65]

In contrast, the postsynaptic membrane of the inhibitory synapse strengthens hyperpolarization due to the influx of chloride ions, resulting in inhibitory postsynaptic potential (IPSP) or current (IPSC).[65] Generally, a neuron contains multiple synapses, some of which are excitatory and some are inhibitory. When the sum of excitatory synaptic activity exceeds the inhibitory synapse, and can reach the AP threshold of the neuron, the nerve impulse occurs, the neuron exhibits excitatory activation, and vice versa.[17,62] Thus the synapses complete the transmission and integration of spatio–temporal information to the neurons.[15] The integration of postsynaptic potentials at different synapses on the same neuron and the accumulation of postsynaptic potentials continuously generated at different times in the same synapse reflect the synaptic ability to sum up spatial and temporal information, respectively.[66]

Synapses change the strength of connections with neurons (that is, synaptic weights) or create new connections in response to transmitted AP,[66] resulting in the modulation of neuronal behavior, which is called neuroplasticity or synaptic plasticity.[15] The adjustable synaptic plasticity is considered to be an intrinsic property of biological neurons and the essential basis for biological learning and memory.[67,68] Changes in the number of neurotransmitters and response efficiencies can cause synaptic plasticity modulation. Synaptic plasticity is divided into short-term plasticity and long-term plasticity,[69,70] which correspond to recoverable temporary changes and persistent changes of synaptic weight[71,72] respectively. It is worth noting that the short-term and long-term synaptic plasticity have corresponding excitatory and inhibitory behavior, that is, the short-term or long-term EPSP and IPSP.[16] As shown in Figure 2a, the excitatory and inhibitory short-term synaptic plasticity is shown in sequence. For short-term potentiation (STP), the EPSP amplitude continues to accumulate and the probability of triggering an AP increases, but after the stimulation is complete, the climbing EPSP returns to the baseline level in a short time.[72] For short-term depression (STD), IPSP amplitude continues to accumulate as stimulation increases, and after stimulation ends, IPSP can gradually come to the baseline level as well.[72] The typical quantitative characteristics of STP and inhibition are paired-pulse facilitation (PPF) and paired-pulse depression (PPD), respectively, where the interval time between two consecutive stimulus signals increases, and the induced facilitation (inhibition) of EPSP (IPSP) degrades until disappearance.[73,74] Short-term synaptic plasticity balances potentiation and inhibition in the cerebral cortex and participates in the realization of advanced functions such as attention, sleep rhythm, computation, and short-term memory in the nervous system.[17] The application of repetitive and frequent stimuli will cause changes in the essential morphology of neurons and synapses and lead to a transition from short-term plasticity to long-term plasticity.[72] as shown in Figure 2b. For long-term potentiation (LTP), stimulation number or frequency increases, high-frequency AP is induced, EPSP amplitude increases, and it will not decay to the initial baseline level after stimulation, that is, a permanent strengthening in synaptic weight.[75] Conversely, for long-term depression (LTD), the low-frequency AP leads to a decrease in synapses and neurotransmitter receptors, IPSP amplitude declines progressively, and cannot return to the initial level, which is a persistent reduction in synaptic weights, corresponding to the weakening effect of neural connection strength.[71] Generally, changes in the release of presynaptic neurotransmitters and a decrease in the density of receptors on the postsynaptic membrane are the main causes of LTD.[71] It is generally believed that long-term plasticity is a key neurophysiological activity for organisms to achieve learning and long-term memory behavior.[57,76] The modulation of biological synaptic plasticity follows the Hebbian theory, which indicated that spike stimulation signals with temporal delay in the presynaptic and postsynaptic cells can effectively modulate the synaptic weight potentiation or inhibition.[77] Later, Hebbian learning theory evolved into the well-known spike-timing-dependent plasticity (STDP), which is regarded as the fundamental principle of neural activity and biological learning.[78,79] The timing of spike stimulation in
presynaptic and postsynaptic neurons leads to the potentiation or inhibition of synaptic weight change ($\Delta W$), and as time intervals increase, the changes in synaptic weight decrease.$^{[69]}$ The classic antisymmetric Hebbian STDP was initially found in hippocampal cultures,$^{[80]}$ as schematically shown in the left inset of Figure 2c; when presynaptic spikes arrive before postsynaptic spikes ($\Delta t > 0$), it exhibits excitatory synaptic plasticity, EPSP, or synaptic weight increases ($\Delta W > 0$), and synaptic connections strengthen; However, when the presynaptic spikes arrive after the postsynaptic spikes ($\Delta t < 0$), it shows inhibitory synaptic plasticity, IPSP, or synaptic weight decreases ($\Delta W < 0$), which means that the synaptic connection is weakened.$^{[81–83]}$ The changes in synaptic weight in both potentiation and depression decrease with the increase in the time difference between the presynaptic and postsynaptic spikes. In addition, STDP with different spike temporal orders has been demonstrated, including antisymmetric anti-Hebbian STDP, symmetric Hebbian STDP, and symmetric anti-Hebbian STDP, which exhibit diverse time-dependent plasticity, as shown in Figure 2c.$^{[81,84–87]}$ The changes in synaptic weight based on the timing of nerve spike stimulation. It is worth mentioning that with the development of biology and neuroscience, STDP has been widely implemented in organisms, covering fish, frogs, mice, cats, and humans. STDP, which is ubiquitous in different species, shows that it plays a vital role in synaptic plasticity and neural communication. However, there is no doubt that biologically complex neural activities and computations do not entirely depend on the STDP principle.$^{[81]}$ Spike frequency, activation rate, and even nerve cell location—all have important effects on synaptic plasticity.$^{[79,85]}$ In particular, the synaptic weights are modulated by the frequency of spike stimulation as well; usually, high-frequency stimulation spikes trigger long-term synaptic plasticity (LTP/LTD), whereas low-frequency signals lead to short-term synaptic plasticity (STP/STD), which is called spike-rate-dependent plasticity (SRDP)$^{[83,88–90]}$ Nevertheless, it is worth noting that for different locations and types of biological nerve cells, frequency modulation may be a relatively neutral concept. A certain frequency spike may be a high-frequency stimulus for one neuron, but a low-frequency stimulus for another different location or type of neuron. The relativity of frequency modulation reflects the inconsistency and variability of activation thresholds in various biological neurons, which affects the modulation of short-term and long-term plasticity. Moreover, repeated low-frequency spike stimulation weakens neuronal connections, and the corresponding synapses and neurotransmitter receptors degenerate and disappear, eventually leading to LTD.$^{[71]}$ Typically, repetitive low-frequency spikes of about 1–5 Hz cause long-term inhibition, whereas a high-frequency stimulus of about 20–100 Hz induces an LTP effect.$^{[91]}$ Synaptic plasticity with high fault tolerance, robustness, adaptability, and computational energy efficiency has become the biological basis of neuromorphic engineering, which is widely used in ANN acceleration and bionic perception and motion integration applications. However, the excavation and utilization of advanced neuroplasticity (such as Hebbian STDP and SRDP rules) in neuromorphic engineering is still in its infancy and requires full cooperation and more efforts in the fields of neuroscience and electronic engineering.

Figure 2. Biological short-term and long-term neural plasticity. a) Excitatory and inhibitory short-term plasticity, where the changed postsynaptic potential will recover to baseline in a short time. b) Excitatory and inhibitory long-term plasticity caused by repetitive and frequent stimulation, which forms a persistent change in postsynaptic potential without returning to baseline. c) Diverse STDP types, including antisymmetric Hebbian STDP, antisymmetric anti-Hebbian STDP, symmetric Hebbian STDP, and symmetric Anti-Hebbian STDP.
3. Memristive Synaptic Devices

As the most advanced processor, the human brain has significant advantages of small volume, low power consumption, multifunctional, and fast working speed. Biological researchers show that the human brain is composed of neurons \((\approx 10^{11})\) and connected by synapses \((\approx 10^{15})\). To realize the brain-like function, the imitation of neuron and synapses is the core of neuromorphic engineering researches. In this section, a review of recent progresses in synapses is proposed to indicate unique neuromorphic characteristics.

In the neuromorphic engineering system, the artificial synapses contain a top electrode serving as the presynaptic terminal, interlayer serving as synaptic cleft, and a bottom electrode as postsynaptic membrane, which imitates the structure and transmission process of a biological synapse. The corresponding relationship between biological and artificial synapses is shown in Figure 3. According to the working mechanism of the interlayer, artificial synapses are divided into the following categories.

3.1. Metal-Ion-Migration Synapse

Conductive bridge random access memory (CBRAM), working as a resistive-switching mechanism, has been widely used in artificial synapses. Based on the mechanism of filament formation, CBRAM can be also called programmable metallization cells\(^{[93,94]}\) or electrochemical metallization.\(^{[95]}\) With the structure in Figure 3, the top electrode is commonly made of Ag or Cu serving as the fast diffusing metal layer;\(^{[96]}\) the bottom electrode is made of Pt, Au, or W which serves as a relatively inert layer;\(^{[97]}\) and the dielectric layer is made of amorphous silicon,\(^{[98]}\) oxide,\(^{[99]}\) polymer,\(^{[100]}\) or amorphous carbon.\(^{[101]}\) Because the resistive switching behavior of CBRAM is based on metal-ion migration between two electrodes, the conductive filament is composed of metal oxidation. When a positive voltage is applied to the active electrode, the Ag\(^{+}\) or Cu\(^{2+}\) is ionized and migrates into the dielectric layer, which forms a conducting bridge between two electrodes and a low resistance state. The forming and dissolving processes of the metal filament between two electrodes achieve resistive switching in two-terminal CBRAM. To form the artificial synapse, this physical process can be mapped to synaptic facilitation and connection weight change.

The high scalability and large dynamic range are the two major advantages of CBRAM. Kund\(^ {102} \) et al. fabricated a sub-20 nm CBRAM, which is composed of Ag–Ge–Se. In +240 mV, the metal-ionic filament was formed to switch the resistive state from high \((>10^{11} \, \Omega)\) to low \((<10^5 \, \Omega)\) and the reverse process was achieved in −80 mV. A test chip comprising CBRAM cells was generated by the deposition and etching of a dielectric layer (Si\(_3\)N\(_4\)) on top of a tungsten bottom electrode, filling the respective small via diameter (20 nm) with a tungsten plug, to show the scalability potential, multilevel capability and technology reliability. In sub-20 nm, CBRAM steadily realized resistive state transformation, six orders of magnitudes \(R_{\text{on}}/R_{\text{off}}\), and 10\(^5\) s retention. As the early research of CBRAM, this work showed promising nonvolatile new memory technology. However, it didn’t associate the resistive process to the neural process. Lu and coworkers\(^ {103} \) demonstrated a hybrid silicon-based neuron system with two-terminal CBRAM, which realizes the biological characteristics of STDP. The dielectric layer of this memristor is composed of Ag and Si with a properly designed Ag/Si mixture ratio gradient. With the positive and negative voltage pulses applied on electrodes, the metal-ionic filaments are formed and dissolved from the Ag-rich region to the Ag-poor region, which leads to the resistance switching behavior. The synaptic weight between two neurons is attributed to ionic

![Figure 3. Schematic diagram of biological synapse structure and neuron transmission process and classical two-terminal artificial synapse model.](www.advancedsciencenews.com)
flow in the ionic-migration artificial synapse. The relationship between synaptic weight and spike timing difference is well fitted with exponential decay functions, indicating that STDP characteristics of artificial synapses are similar to biological synapses. Inspired by biological characteristics simulated by CBRAM, the same phenomenon is realized in other material systems, such as Cu–SiO$_2$–W$^{[104]}$, AgNPs–polymethylsiloxane (PDMS)$^{[105]}$, and Ag–TiO$_2$–Pt$^{[106]}$. Except for the ion doping in the interlayer, another mechanism for ion-filament forming is the redox reaction of electrodes and interlayer material due to electrical bias. Kim and coworkers$^{[107]}$ conducted the heteroepitaxial growth of 60 nm SiGe onto the Si substrate, which shows a high threading dislocation density. For limited solid solubility in SiGe and the inability of forming compounds with SiGe, Ag is selected as the active metal layer, which forms a filament along with threading dislocation (Figure 4a). Ag–SiGe–Si CBRAM shows a linear conductance response similar to the biological synapse characteristic. Therefore, an ANN of supervised learning with Mixed National Institute of Standards and Technology database (MNIST) handwritten recognition is simulated based on 28 × 28 preneurons, 300 postneurons, and 10 output neurons, which show 95.1% average recognition accuracy. The CBRAM resistive characteristic and biological performance are shown in Figure 4b–d.$^{[108,109]}

3.2. Oxygen-Vacancy-Migration Synapse

Except for metal-ion-migration mechanism, resistive characteristics can be achieved by oxygen-vacancy migration, especially in transition metal oxides.$^{[110,111]}$ One of the most important

---

**Figure 4.** CBRAM-based artificial synapses. a) A conceptual schematic of Ag–SiGe–Si CBRAM during switching. Reproduced with permission.$^{[107]}$ Copyright 2018, Springer Nature. b) Resistive switching characteristics of Ag/CrPS$_4$/Au CBRAM. The inset is the schematic diagram of the device. c) The experiment data of LRS retention and the fitting curve of the result with exponential decay function. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).$^{[108]}$ Copyright 2018, The Authors, published by Springer Nature. d) The LTP and LTD biological characteristics of Ta/Al$_2$O$_3$/ZrTe CBRAM. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).$^{[109]}$ Copyright 2018, The Authors, published by Springer Nature.
advantages is good compatibility with CMOS-based integrated circuits. Wong and coworkers\textsuperscript{[112]} combined metal oxide resistive switching memory with the electronic synapse, which was composed of TiN/HfO$_x$/AlO$_x$/Pt. In this synapse device, TiN is used as top electrode, bilayer HfO$_x$/AlO$_x$ serves as interlayer, and Pt is the bottom electrode. With oxygen vacancies generation and migration, the device resistance can be set from the high resistive state (HRS) to low resistive state (LRS) and reset from LRS to HRS (like Figure 5a), which shows excellent stability (more than $10^5$ cycle endurance and 7200 s retention). According to the STDP learning rule, the conductance of device STDP learning rule is determined by the delta-T between pre and postspike. If prespike comes before postsing, the resistance of the artificial synapse decreases, called LTP, which shows that the connection between two neurons becomes stronger. In the country, the resistance increases because of prespike delay, called LTD, which shows the weaker connection between two neurons. Except for the broadest PPF bionic behavior shown in Figure 5b, there is much other biological nature that can be imitated in the artificial synapse. Zhou and coworkers\textsuperscript{[115]} proposed hydrogen and sodium titanate nanotubes to emulate the synaptic behavior. The nanotubes are Na$_2$Ti$_3$O$_7$, H$_2$Ti$_3$O$_7$, and intermediate products of hydrothermally synthesized TiO$_2$. Due to the migration of oxygen vacancies driven by electric field and diffusion.

![Figure 5. Oxygen-vacancy-migration synapse-based artificial synapses.](image)

\textsuperscript{a)} I–V characteristics of the Ta/TaO$_x$/TiO$_2$/Ti synapse device measured at the double sweep. b) Measured and simulated PPF artificial curves of the Ta/TaO$_x$/TiO$_2$/Ti device, obtained by applying alternating current (AC) pulses successively. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). Copyright 2015, The Authors, published by Springer Nature. c) 3D schematic diagram of Pt/(Na$_{0.5}$K$_{0.5}$)NbO$_3$/TiN/PI synapses. d) Resistive characteristic generated by Pt/(Na$_{0.5}$K$_{0.5}$)NbO$_3$/TiN/PI synapses, which are excited or suppressed by pulses. e) STDP biological characteristic of Pt/(Na$_{0.5}$K$_{0.5}$)NbO$_3$/TiN/PI synapses. i) The prespike and postspike pulse waveform applied to the synapses. ii) The experimental data and fitted curve of STDP. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). Copyright 2017, The Authors, published by Springer Nature.
spontaneously induced by temperature, hydrogen and sodium titanate nanotube shows a resistive switching characteristic, which leads to the synaptic imitation. These nanotubes are synthesized by the hydrothermal process and uniformly distributed in substrate-deposited Au electrodes. Under the external electrical stimulus, spikes come from preneurons and are transmitted through the synapse (nanotube) to the postneuron. In this device, a series of 50 ms electrical pulses with incremental amplitudes are applied to the left electrode as presynaptic APs and pulse interval is 4 s. Similar to biological excitatory synapses, the peak value of EPSC shows an obvious increase in tendency with the increase in presynaptic pulse, which is called short-term plasticity. In that case, the PPF and posttetanic potentiation (PTP) are separately defined as synaptic gained weight generated by two and ten presynaptic pulses. Under pulse excitation with an interval of 100 ms, PPF and PTP of Au/nanotube/Au are 53% and 436%. The learning and forgetting processes in human memory correspond to potentiation and depression phenomena in artificial synapses. Except for EPSC, the postsynaptic current depression effect is demonstrated, which is generated by a series of negative electrical pulses. After motivating behavior, the postneuron current declines exponentially, defined as \( I = I_0 + A e^{-t/\tau} \), where \( I_0 \) is the saturation current, \( t \) is time, \( A \) is prefactor, and \( \tau \) is time constant depending on the forgetting rate (showing the resistive storage retention capability). So the STP can transform to LTP both in the human brain and in the artificial synapse. With the increase in pulse amounts, \( \tau \) increases, resembling the increase in synaptic strength through frequent stimulation by presynaptic potentials found in biological neural systems. With the same mechanism, a nanogenerator-induced artificial synapse is fabricated and characteristics are shown in Figure 5c–e.\(^{[114]}\) The oxygen-vacancy-migration mechanism can also be discovered in 2D-material-based memristors. In 2018, a full 2D memristor composed of graphene/MoS\(_2\)/O\(_x\)/graphene exhibited an excellent thermal stability (340 °C operating temperature), \( 10^7 \) endurance cycle, and over 1000 bending stability, which illustrate the potential for flexible synapses.\(^{[116]}\)

3.3. Electron-Trapping/Releasing Synapse

Because ionic/oxygen-migration synapses depended on the formation and degradation of filaments to realize resistance modulation, the transformation between different resistance states is discontinued. As the neuroplasticity regulation of biological synapses changes continuously as stimulations accumulate, these memristors with discrete resistance have an advantage in the data storage system and image detector application, but have limitations in neuromorphic engineering. The artificial synapses constructed with electron semiconductors accomplish smooth and gradually resistive variations by the charge trapping and releasing mechanism. Huang and coworkers\(^{[117]}\) constructed an organic CuPc-based memristor, realizing a combination of bioinspired STDP and homeostatic plasticity. The crossbar structure for neuromorphic engineering is indium tin oxide (ITO)/CuPc/Al, showing a typical resistive switching behavior: ten enhancing and suppressing hysteresis loops (Figure 6a). Notably, the CuPc-based memristor displays smooth and continuous \( I-V \) characteristics without saturation and fluctuation during the process of resistive change, implying the uninterrupted charge trapping/releasing process in organic artificial synapses. In the CuPc-based memristor, the CuPc domains serve as charge storage and the interface between CuPc and ITO contains amounts of traps that serve as current barriers. When applying voltage, the current is generated by the carrier hopping from one domain to another, and the resistance arises from intergranule and interfacial impedance. According to carrier wave function and Mott transition mechanism, the increasing concentration of carriers trapped at the domain increases more strong correlations. Therefore, higher conductive percolation channels (conductive path) are formed in an insulating region (between CuPc domains), which leads to a continuous current increase and uninterrupted resistive decrease. This carrier trapping/releasing process is demonstrated by the transport pattern of the carrier shown in Figure 6b. The resistive state in the memristor reflects the efficiency of electrical signal transmission, which is similar to the weight of biological synapses.\(^{[119]}\) On the basis of STDP characteristics, CuPc-based synapses embody a self-adaptive phenomenon, which can be observed from the nonpurely prolonged tendencies under the less-activated excitation (Figure 6c). This biomimetic performance contributes to avoiding excessive excitation or inhibition in the long-term homeostasis of neural activities, which is closer to the real synaptic system. In addition to the self-adaptive response under balanced inputs, the performance under more extensive inputs is shown in Figure 6d. Although CuPc-based synapses originally show unstable resistive fluctuation when unbalanced signals are applied, the overall recitative level gradually reaches a unified steady state after multiperiods of intercoupling processes. Therefore, the stepped-up excitation signal does not result in an excessive response in CuPc-based synapses under both balanced and unbalanced inputs because of self-adjustment. On the basis of previous work, Huang et al.\(^{[118]}\) further studied the neuromorphic engineering performance of CuPc-based synapses in 2017, especially in the auditory application. With the same structure, a LiF buffer layer is inserted between CuPc and ITO to enhance the carrier transport efficiency in organic-based devices and regulate the asymmetric conductive hysteresis in different directions. Compared with previous research, this work analyzes the charge trapping/releasing mechanism of resistive change in the band diagram and the experimental result (Figure 6e) rather than words. According to the rectification characteristics generated by the LiF interlayer, the rectifying device shows uncommon directed signal transmission, which means the electrical signals can only be transmitted significantly in one (positive or negative) direction and the synaptic weight is progressively enhanced under continuous excitatory stimulation. Hence, the rectifying device can be applied to the bidirectional electrical synapse, which can symmetrically transmit the neural-facilitating signals and meanwhile adaptively regulate its synaptic weight in each direction sustainably (Figure 6f). Except for the 3D synaptic mentioned earlier, the electron trapping/releasing mechanism can also be exploited in hybrid one-dimensional-zero-dimensional (1D-0D) memristors. In 2013, a typical hybrid 1D–0D synapse was fabricated with the structure of vertically aligned ZnO nanowires (1D)-attached CeO\(_2\) quantum dots (0D). In these hybrid 1D–0D synapses, \( 10^4 \) on/off ratio, \( 10^5 \) endurance cycle and \( 10^8 \) retention time are
achieved, which shows a promising candidate for the next generation of 1D–0D hybrid synapses.[120] Charge trapping/releasing-based synapses can also be realized by other material systems, such as Al/Nb2O5/Al,[121] BiN2,[122] and Al/HfO2/Al2O3/Si3N4/Si.[123]

3.4. Magnetic-Spin-Torque Synapse

The key component of the spin-torque synapse is a magnetic tunnel junction (MTJ) which is composed of ferromagnetic material/tunnel layer/ferromagnetic material. One ferromagnetic layer is called a reference layer or pinned layer, which has a fixed magnetization direction along with easy axis. The other ferromagnetic layer is called the free layer and has two stable magnetization directions, parallel or reversed to the reference layer, which sets MTJ into LRS or HRS.[124] The tunnel magnetoresistance effect can be explained by the spin-dependent tunneling theory.[124] For metallic–ferromagnetic material, the state distribution of spin-up and spin-down electrons is unbalanced in the Fermi level. When the magnetization direction of the free layer is parallel with the reference layer, most electrons in two ferromagnetic layers have the same magnetization direction, leading to a high tunneling probability and high tunneling current. So the MTJ is in LRS. On the contrary, the MTJ is in HRS. Based on the resistive switching mechanism, MTJ-based synapse has been taken into consideration in neuromorphic engineering application. Thomas and coworkers[125] combined MTJ memristors with synaptic functionality. The device structure is composed of Ta/Cu-N/Ta/Pt37Mn63/Co70Fe30/Ru/Co66Fe22B12/MgO/Co66Fe22B12/Ta/Cu/Ru and MgO serves as the insulating layer. According to the resistive switching behavior, basic

---

Figure 6. Charge trapping/releasing-based artificial synapses. a) $I$–$V$ characteristics of the CuPc-based synapses synapse device measured at the double sweep. b) Log–log plots of the forward-scanning current extracted from (a). c) Balanced resistive responses under coordinate stimulation. d) Unbalanced resistive responses under coordinate stimulation. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).[117] Copyright 2016, The Authors, published by Springer Nature. e) Mechanisms of the rectifying synapse behaviors analyzed by band diagram and experiment result. f) The equivalent circuit model of ITO/LiF/CuPc/Al device. Reproduced with permission.[118] Copyright 2017, Springer Nature.
neuronal behaviors of LTP, LTD, and STDP are achieved. However, the resistive mechanism of this device is voltage-driven oxygen vacancy motion within the MgO layer, rather than magnetic transformation. Querlioz et al.\cite{126} discussed the basic characteristics of spin-transfer torque magnetic memory (STT-MRAM) and the possibility of implementing in learning-capable synapses. According to quantum mechanics, the switching time of STT-MRAM heavily depends on the current and stochastic quantity, which is caused by the physical mechanism of magnetic switching. With STDP rules, a spike neural network system (20 output neurons) is introduced to conduct unsupervised learning. The transmission of neuron signals is similar to other resistive synapses: when asynchronous spikes come from preneurons, the resistance of MTJ acts as the synaptic weight converting voltage to current, which shows the enhancing or suppressing effect to postneurons. For the symmetry in operating voltage between parallel and antiparallel states, the system allows unsupervised learning. Based on Monte Carlo simulations, this small learning network shows a 99% accuracy rate as a vehicle counter and 3.7 μW power consumption, which shows the application potential of MTJs in robust, low-power, cognitive-type systems. The previous work\cite{126} inspired further research of the neuromorphic system consisting of MTJs\cite{127}, which shows widespread stability even in a 25% device variation rate.

Although these two works illustrate the working mechanism of resistive switching and fabricate a small learning network consisting of MTJs, the learning process and neuron characteristics are realized by Monte Carlo simulation rather than the realistic device. Lequeux et al.\cite{128} first showed the experimental achievement of the magnetic-spin-torque synapse, which confirmed the application potential of MTJs in neuromorphic engineering, especially in CMOS technology capability and excellent recyclability. Different from the simplest MTJ structure, this device is composed of four layers to realize controllable multilevel resistive switching: synthetic antiferromagnet layer CoPt/Ru/CoPt/Ta/FeB, tunnel-layer MgO, free-layer FeB/Ta/FeB, and capping-layer MgO/Ta. Figure 7a shows the side view of the schematic device structure and the top view of this device. For the movement of the domain wall prefabricated in the free layer, the resistance of the spin-torque memristor varies from LRS to HRS and generates multilevel resistance shown in Figure 7b. The domain wall can be nucleated by vertically injected current, which means the proportion parallel with the reference layer determines the resistance. With the realization of multilevel resistance, MTJs have the basis of neural simulation. A previous study\cite{132} simulates the number of binary junctions (7–10) needed in fabricating recognizing handwritten digits with a 75% recognition rate, which means the multilevel resistance state of spin-torque synapses shown in Figure 7b can realize classification tasks.

![Figure 7](https://example.com/figure7.png)

**Figure 7.** The different structures and multilevel resistance of spin-torque memristor. a) Side view (left inset): Schematic of two-terminal spin-torque memristor with the magnetic domain wall in the free layer. Top view (right inset): Scanning electron microscope image of two-terminal spin-torque memristor and the black dashed line aims to emphasize the outline of the device. b) Left inset: The function relationship between resistance and vertically injected DC current. Right inset: Micromagnetic simulations of the domain wall propagating in a magnetic track of 100 nm width. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).\cite{128} Copyright 2016, The Authors, published by Springer Nature. c) Schematic diagram of three-terminal spin-torque memristor and its connecting circuit. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).\cite{129} Copyright 2016, The Authors, published by Springer Nature. d) Schematic diagram of four-terminal spin-torque memristor and its connecting circuit. Reproduced with permission.\cite{131} Copyright 2020, Springer Nature.
Except for the advantage of multilevel resistance, spin-torque synapses have two outstanding advantages: one is subnanosecond resistance variations due to the super moving speed of domain walls and the other is excellent CMOS technology compatibility due to the preparation method. However, according to the switching mechanism of spin-torque synapses, the reduction of robustness is important in neuron network construction. In this work, the randomness of domain wall pinning leads to the uncontrollability of resistance between cycles and devices. To solve this problem, the drive signal is changed from direct current (DC) to nanoseconds voltage pulse, which leads to smaller domain wall displacements. In addition to two-terminal spin-torque memristors, synaptic or neuromorphic engineering characteristics have also been realized in three- or multiterminal magnetic structures (Figure 7c,d).129,130,133–137

3.5. Ferroelectric-Polarization Synapse

Due to the noncoincidence of positive and negative charges in the cell, the electric dipole moment is generated spontaneously. With the modulation of the electric field, the orientation of spontaneous polarization turns into consistency and the spontaneous polarization gradually increases to saturation polarization, leading to a ferroelectric hysteresis loop. After removing the electric field, the orientation of spontaneous polarization remains unchanged. Therefore, the ferroelectric tunneling junction (FTJ) composed of the metal-1/ferroelectric-film/metal-2 sandwich structure shows nonvolatile resistive switching, caused by quantum tunneling effect and electrical resistance effect. The different interface potential barrier generated by an incomplete screening of polarization charges leads to the HRS and LRS of the FTJ-based memristor,1138 as shown in Figure 8a. Ferroelectric materials have been applied to neuromorphic engineering since 1993.139 Simple metal-insulator-semiconductor field-effect transistors (MISFETs) with a ferroelectric dielectric film (PbZr1-xTiO3) is designed to fabricate adaptive-learning neuron systems. According to the quantitative analysis, the time dependence between polarization density P and electrical field E is expressed in the following equations and in Figure 8b. With the increase in pulse voltage, the number of learning times shows an exponential decrease in the same pulse width, which means that higher the electrical field, the more complete the polarization ferroelectric material is. The surface charge density of ferroelectric neurons is shown in Figure 8c, which stands for the resistance of the device to some extent. Based on the EPSC, numbers of ferroelectric MISFETs are used to fabricate an adaptive-learning neuron circuit.

\[ t_s = t_{0} \exp \left( \frac{E_s}{E} \right) \]  

\[ P = P_a \left[ 1 - 2 \exp \left( -\left(\frac{t_s}{\tau} \right)^2 \right) \right] \]  

where \( t_s \) is the switching time of ferroelectric MISFETs, \( E_s \) is the activation field, \( E \) is the applied electric field on the device, \( P \) stands for polarization of the ferroelectric material, \( P_a \) is maximum polarization of the material, and \( \tau \) depends on the material.

Although traditional three-terminal ferroelectric field-effect transistors have unique advantages in CMOS-based circuits, such as compatibility with CMOS technology, diversified logical and memory device, and stability in polarization, two-terminal FTJs are more suitable in synaptic applications to realize high integration density. In 2012, Barthelemy et al.140 fabricated a two-terminal FTJ as the synapse which is composed of Co/BaTiO3/La0.67Sr0.33MnO3/Au. The FTJ shows an excellent characteristic of the memristor, including 300 resistive ratios, 10 ns switching time, and quasicontinuous resistive variation, which suggest that two-terminal FTJ has the potential in neuromorphic engineering. With the same width voltage pulse, FTJ-based synapses show a fundamental STDP behavior, as shown in Figure 8d.e. For the lack of the direct synaptic characteristics of FTJ synapses in Agnès Barthélémy’s work, Garcia and coworkers50 reported a Co/BiFeO3/(Ca, Ce)MnO3 FTJ synapse which directly showed STDP characteristic. In biological synapses, the state of postneuron changes when plenty of excitations are released in the postsynaptic membrane, leading to EPSC. Therefore, a voltage thresholds \( V_{th}+V_{th} \) of applied pulse exists to switch LRS to HRS (HRS to LRS) in the artificial synapse. The well-defined and controllable voltage thresholds (affected by the inherent coercivity of ferroelectric materials) guarantee the realization of STDP, reducing the ratio of abnormal resistive switching caused by noise in large integrated circuits. Figure 9a shows the structure of FTJ synapses and neuron spike waveforms emulated by the rectangle voltage pulse and trapezoidal voltage decrease ramp, with a time delay (Δt) between pre and postneuron spikes. According to STDP, if Δt is positive (the spike of preneuron comes before postneuron), the conductance of FTJ artificial synapse will be strengthened, whereas if Δt is negative (the preneuron spikes comes after postneuron), the conductance of the FTJ artificial synapse will be depressed. In Figure 9b, the absolute value of conductance variation (ΔG) shows first an increase and then decrease tendency with the increase of Δt absolute value, in keeping with a biological characteristic that a long time delay between two spikes can’t induce excitation of synapses. To illustrate the physical process and mechanism of conductance transition in FTJ, a dynamic piezoresponse force microscopy (PFM) phase and amplitude change diagram are shown in Figure 9c. The evolution of the phase change diagram reveals that the key conductive switching mechanism of ferroelectric material is material dissociation and domains inhomogeneous nucleation rather than traditional domain expansion mechanism. With the brand new mechanism and experimental data, a mathematic model is designed to simulate resistive behavior, synaptic learning, and spiking network. Based on ferroelectric switching mechanism, synapses can also be implemented in materials such as Ag/PbZr0.52Ti0.48O3 (PZT)/La0.67Sr0.33MnO3 (LSMO),141 Pt/BaTiO3/Nb: SrTiO3,142 Au/Co/BaTiO3/La0.67Sr0.33MnO3/NdGaO3,143 and Au/P(VDF-TrFE)/Nb-doped SrTiO3.144

3.6. Phase-Change Synapse

Phase-change materials used for synapses possess at least two distinct solid-phase structures, usually amorphous (disordered) and crystalline (ordered). The structural distinction between two phases generates a wide difference in optical and electrical characteristics, leading to the switching of resistance. When a long-time and the moderate-amplitude voltage pulse is applied
to phase-change materials, the temperature of phase-change material is raised above the crystallization temperature and below the melting temperature, which implements crystallization. In the crystalline state, the ordered atoms lead the LRS of the phase-change material-based memristor. When a short-time and the high-amplitude voltage pulse are applied to the crystalline material, the temperature reaches above the melting temperature and phase changes from crystalline to amorphous. The disordered atoms lead the HRS of phase-change memory (PCM), as shown in Figure 10a. For the advantage of scalability, reliability, low power consumption, fast operation speed, and multilevel resistance state, metal/phase-change material/metal-based synapse (Figure 10b) has been widely investigated in neuromorphic engineering. Suri et al. produced a phase-change memristor (W/Ge$_2$Sb$_2$Te$_5$ or GeTe/W) connected with two spiking CMOS circuits (acting as pre- and postneurons) to implement synaptic functionality (Figure 10c). A strong reset pulse is applied to the memristor setting phase-change material into an amorphous state (HRS) at the beginning of the resistance test. Similar to the ferroelectric-polarization-based resistive memristor, there exists a threshold voltage ($V_{th}$) for resistive transformation, which is the boundary of resistive switching (Figure 10d). According to resistive switching, two basic synaptic behaviors are achieved, including LTP and LTD. Therefore, a small circuit composed of two-phase-change synapses implements STDP learning rules (Figure 10e). In this circuit, only one synapse’s conductance state is transformed with the modulation of voltage spikes and the operation speed limited.
by nucleation rate or growth velocity. Except for the new circuit structure, a refresh procedure is developed to enable continuous learning in the network without the synapse weight unchanged. Compared with the one-phase-change unit, the two-phase-change synaptic circuit decreases the impact of resistance drift for the synaptic learning system on account of the refresh process. However, the low area efficiency and high power consumption are the fatal weaknesses of this circuit. Because temperature leads to resistance switching of the phase-changed material, the synapses can be modulated not only by the electrical field but also optical field, as long as the material can be heated to the phase-changed temperature. In 2017, Bhaskaran and coworkers\cite{148} delivered a phase-change synapse modulated by the optical pulse to realize STDP characteristics. Different from two-terminal synapses, phase-changed synapses modulated by the waveguide is composed of Ge$_2$Sb$_2$Te$_5$ (GST), with indium tin oxide as the capping layer. To connect the synapse with fiber arrays for optical signal transmission, two diffraction grating couplers are designed in the system, and the structure of the optical synaptic system is shown in Figure 11a. Similar to the electrical field, optical pulses with different widths and densities can realize transformation between the crystalline and amorphous state in phase-change material. The refractive indices of the two phases have a large difference, especially in the near-infrared wavelength regime, which is the most commonly used spectral region for telecommunication applications.\cite{149} The different refractive indexes lead to different intensities of output signals, corresponding to the multilevel of resistance (Figure 11b).\cite{150} The optical signal coming from presynaptic is split into two beams with 50% intensity coupled into GST synapse and interferometer, and a phase modulator is used to adjust the weight synapses (the connection circuit is shown in Figure 9d). Only in the area where two spikes overlap the weight of the synapse increases. Therefore, the intrinsic characteristics, such as fast operation speed, multiple modulation method, and wavelength division multiplexed access and phase-change material, are suitable for neuromorphic engineering. Table 1 shows the representative characteristics of memristive synaptic devices with different mechanisms.
4. Artificial Neuron Circuits

Significant progress has been made in the construction of artificial synapses using memristive devices, and the simulation of diverse synaptic functions has been achieved. In addition to artificial synapses, highly integrated neurons are indispensable basic units for building large-scale neuromorphic engineering, which exhibits completely different design criteria from memristive synapses. Synapses transmit and store signals through weight adjustment, whereas neurons perform complex information processing tasks, that is, integrate the signals fed by synapses into the soma, cause changes in membrane potential, and compare with thresholds to determine whether to activate and fire spikes. The synaptic weight can usually be represented by the conductance of the memristive device, which allows the simulation of synaptic functions in a compact single device. However, it is unrealistic for a separate device to simulate the function of artificial neurons. Instead, hybrid circuits are usually required to construct artificial neurons and implement functions. Traditionally, artificial neurons are constructed using CMOS circuits with integrated transistors, but they face challenges such as poor scalability, low integration, high complexity, and high power consumption. Fortunately, the introduction of memristive devices enables the remarkable simplification of traditional CMOS transistor neuron circuits and has successfully achieved some heuristic progress. It is worth noting that, according to the neuron model used, artificial neurons composed of memristive devices can be generally divided into leaky-integrated-and-fire (LIF) neurons and Hodgkin–Huxley (HH) neurons. LIF neurons focus on
simulating the basic functions of biological neurons, including signal integration, threshold comparison, and spike generation process, which are widely adopted by neuromorphic engineering, thanks to its compact structure and scalable potential. Recently, the integration of nonvolatile PCM memristor devices with amorphous and crystalline phase transitions and auxiliary components such as comparators and pulse generators has been demonstrated to build artificial LIF neuron circuits and achieve integrate-and-fire kinetics simulation. The inherent random reconfiguration of the atomic functional layer in the PCM memristive devices caused by melt quenching exactly mimics the biological neuron stochastic dynamics, which is critical for neural population coding for expressing and transmitting perception and motion signals. In addition, the parallel connection of

---

Table 1. A representative comparison of memristive synapse characteristics with different mechanisms.

| Components                        | Power consumption | Operation speed | Multilevel states | \( R_{on}/R_{off} \) | Retention | Endurance | Reference |
|-----------------------------------|-------------------|-----------------|-------------------|-------------------|-----------|-----------|-----------|
| Metal-ion-migration synapse       | \(-\)              | \(-\)           | 4                 | 10^6              | 10^1 s    | 10^3      | [102]     |
| Ti–Ag–SiGe–Si                     | \(-\)              | 5 \( \mu \)s    | 5                 | \( \approx 10^6 \) | 10^3 s    | 10^6      | [107]     |
| Oxygen-vacancy-migration synapse  | Ti–HfO\(_2\)–AlO\(_2\)–Pt | 6 pJ           | 50 ns             | 7                 | \( \approx 10^3 \) | 7200 s    | 10^3      | [112]     |
| N\(_2\)Ti\(_3\)O\(_7\)–TiO\(_2\)| 212 pJ             | 50 ms           | 6                 | \(-\)             | \( >400 \)s | 5 \times 10^6 | [115]     |
| Electron-trapping/releasing synapse| ZnO NW–CeO\(_2\) QD | \(-\)           | 2                 | 10^2              | 10^4      | 10^2      | [120]     |
| Al–HfO\(_2\)–Al\(_2\)O\(_3\)–Si\(_1\)N\(_2\)–Si | 500 \( \mu \)s | 4               | \( \approx 10^3 \) | 600 s             | 10^4      | [123]     |
| Magnetic-spin-torque synapse      | STT-MRAM           | 3.7 \( \mu \)W  | \( \approx 2 \) ns | 5                 | \( \approx 10^4 \) | \( >10 \) years | 10^10     | [126]     |
| Ferroelectric-polarization synapse | Co–BaTiO\(_3\)–La\(_0.67\)Sr\(_0.33\)MnO\(_3\)–Au | \(-\)           | 10 ns             | 6                 | 300       | \( >10 \) years | 250       | [140]     |
| Phase-change synapse              | Ce\(_2\)Sb\(_2\)Te\(_3\) | \( \approx 480 \)fJ | 500 ps            | 7                 | \( \approx 5 \) | \(-\)      | \( >100 \) | [150]     |

\( ^a \)“–" means this characteristic is not available in the work.
volatile memristive switches (insulator-metal transition devices\textsuperscript{[156,157]} and diffusive memristors\textsuperscript{[154]}) and capacitors has also been used to construct LIF neurons with a compact structure, as shown in Figure 12a. The LIF neuron electrical signal emission can be described as the diffusive memristor simulating the ion channel for firing spikes and the capacitor serving as a cell membrane for integrating charges.\textsuperscript{[154]} Once the voltage decrease across the capacitor exceeds the threshold of the diffusive memristors volatile switch, the neuron activates and fires. It is worth mentioning that for some individual memristive devices, the LIF neuron integration function can be achieved without even introducing a capacitor connection,\textsuperscript{[159,160]} because it works as a capacitor within a specific time, which allows to further simplify the neuron circuit.

Although LIF neurons have successfully simulated the basic LIF function, they sacrificed and ignored the details of ion dynamics in biological neuron cell membranes.\textsuperscript{[151]} As an alternative, HH neurons are used to describe certain complex characteristics of biological neurons, such as ion channel dynamics, all-or-none law and hyperpolarization phenomena, etc.,\textsuperscript{[35]} which

![Figure 12. The implementation of artificial neuron circuits introduced by memristive devices. a) LIF neuron realized by parallel connection of diffusive memristor and capacitor. Reproduced with permission.\textsuperscript{[154]} Copyright 2018, Springer Nature. b) Two Mott memristors are connected in parallel with the capacitors and then connected to the current source to construct the HH neuristor. Reproduced with permission.\textsuperscript{[153]} Copyright 2013, Springer Nature. c) A schematic of the quasi-HH neuron circuit developed by volatile memristive devices combined with a comparator and a Timer and for the first time realized the functional fusion of LIF and HH neurons. Reproduced with permission.\textsuperscript{[158]} Copyright 2018, Wiley-VCH. d) Spatial–temporal integration and biological neuron firing realized by quasi-HH neuron circuit. Reproduced with permission.\textsuperscript{[158]} Copyright 2018, Wiley-VCH.](image-url)
seem to be more biologically authentic and actively adopted in neurocomputing science.\textsuperscript{[161]} However, HH neurons involving sophisticated neuron dynamics usually bring more complex circuit structures, which is different from LIF neurons that can even be implemented in a single memristor. Previously, HH neurons based on two Mott memristors and additional components of capacitors and current sources have been reported (Figure 12b), which realize the functional simulation of Na\textsuperscript{+} and K\textsuperscript{+} channels in neuronal cell membranes.\textsuperscript{[153]} The proposed HH neuron enables the generation of spikes similar to hyperpolarized nerve spikes. Nevertheless, it is worth noting that the previously realized LIF neurons and HH neurons can only imitate part of the biological neuron behaviors, and there is a gap in the combination of LIF and biological neurodynamic functions. Inspiringly, the quasi-HH neuron circuit was developed based on proton migration in volatile memristive devices, which is the first hardware demonstration of the fusion of LIF and HH neuron functions.\textsuperscript{[156]} as shown in Figure 12c. Similarly, threshold-less memristive devices generally have to introduce a comparator to achieve the LIF function. The constructed HH neuron electrical signal transmission can be described as: as long as the local graded potential reaches the threshold, the comparator will generate a falling edge, followed by triggering Timer to generate spikes that contain the timing information of the input signal and are applied to a simple circuit composed of M2 and a resistor R3 connected in series. The voltage drop on R3 acts as an output signal similar to the output spike of the HH neuron, which enables the implementation of spatial–temporal information integration and generates a biologically inspired neuron fire (Figure 12d). In general, HH neurons implement more neuron functions, but additional redundant device components must be introduced, at the expense of circuit compactness, integration, and scalability.

5. Neuromorphic Engineering for Hardware Systems and Biomimetic Techniques

The neuromorphic engineering that integrates the basic components of memristive synapses and neurons has demonstrated excellent capabilities in system-level applications such as ANN computation acceleration and bionic technology. Thanks to the highly parallel memristive array enabling in-memory analog computing, the vector-matrix dot-product operations in the neural network directly mapped by the memristive conductance array are accelerated. In addition, the emerging computing that requires vector-matrix dot-product operation in neuromorphic engineering also benefits from the acceleration of the memristive array system. In addition to computational acceleration, memristive array systems constructed by artificial synapses and neurons have shown great potential in the bionic realization of sensing, perception, and motion integration. In this section, we first review the acceleration effect of memristive systems on hardware neural networks, including single-layer/multilayer perceptron (SLP/MLP), convolutional neural network (CNN), recurrent neural network (RNN), i.e., Hopfield neural network and spiking neural network (SNN). The acceleration of emerging computing paradigms such as sparse coding, reinforcement learning, reservoir computing, and probabilistic computing is also covered. Finally, we discussed the biomimetic technology of the perception and motion integration implemented by the memristive systems, which include visual, tactile, auditory, pain perception, artificial afferent nerves, and sensorimotor systems.

5.1. Hardware Neural Network Acceleration

As the simplest form of the feed-forward neural network, SLP is used for binary linear classifications, where the input is the feature vector of the object and the output is the object category. The perceptron is composed of an input (perception) layer and an output (reaction) layer, which are simple abstractions of biological nerve cells. A single nerve cell can be regarded as a machine with only two states activated, “yes,” whereas not activated as “no.” As mentioned in Section 2, the state of a nerve cell depends on the number of input signals received from other nerve cells and the strength of synapses (potentiation or inhibition). When the sum of the signals exceeds the threshold, the neuron fires and generates an AP, which is transmitted along the axon and through the synapse to other neurons. To simulate the behavior of nerve cells, the basic concepts of perceptron weight, bias, and activation function are proposed, which correspond to synapse, threshold, and soma, respectively. With the first experimental implementation of SLP hardware based on memristor arrays in 2013,\textsuperscript{[162]} the perceptron has been expanded and developed for various classification tasks.\textsuperscript{[163–165]} Recently, Lu’s group implemented an SLP for the classification of 5 × 5 Greek alphabet images through the integrated chip of the 26 × 10 memristor array.\textsuperscript{[166]} Typically, according to the pixel value, the input image is converted into voltage pulses through the peripheral circuit and sent to the memristive array for training, as shown in Figure 13a. The activation of the target neuron is separated from the disturbing neurons, and the output difference gradually accumulates as the training continues, as shown in Figure 13b. As a result, after only five pieces of training, the classification accuracy of training and test set can reach nearly 100%, as shown in Figure 13c, which greatly accelerates the SLP classification process while ensuring accuracy. Despite its simple structure, an SLP can learn and handle complex linear classification problems, but the essential limitation is that it cannot solve nonlinear separable tasks such as simple exclusive OR (XOR). The MLP contains perception, hidden connection, and reaction layers, which can solve the linear inseparable problem as an alternative to the SLP.\textsuperscript{[153]} Both the connection layer and the reaction layer have information processing functions. The first layer implements binary separate classification, and the second layer implements AND operation, which solves the XOR problem. Generally, the weights of the perception layer to the connection layer are fixed, and the weights of the connection layer to the reaction layer are tunable. It is worth mentioning that the large-scale memristive array provides more advanced complexity and functions for the perceptron. Benefiting from the maturity of the transistor process, large-scale one-transistor-one-resistor (1T1R) architecture-based perceptrons have been reported.\textsuperscript{[167,168]} Compared with passive memristive arrays\textsuperscript{[169]} the 1T1R architecture can provide current compliance and enable a more controlled conductance update. In addition, the introduction of transistor selectors enables the prevention of crosstalk effects on unselected devices.
during programming and eliminates the sneak current path in memristive arrays, which result in higher classification accuracy and demonstrate the potential in perceptron applications.

CNN, another type of feed-forward neural network, whose artificial neurons can partially respond to units within the coverage area, shows superiority in applications such as large-scale image document recognition, video analysis, and natural language processing. CNN usually consists of multiple convolutional layers, pooling layers, nonlinear layers, and fully connected layers, as shown in Figure 13d. The convolutional layer extracts the features of the input data, runs the product and sum operation corresponding to the elements between the convolution kernel and the input image, completes the mapping of the information in the receptive field to the elements in the

Figure 13. Hardware-implemented memristive SLP and CNN. a) A chip integrated with a 26 × 10 memristive array implements SLP for image classification. According to the pixel value, the input image is converted into voltage pulse signals and fed into the memristive array for training. The right panel shows the samples in the training data set. b) Signals of output neurons with a given input class as a function of training epochs. c) Evolution of training and test classification errors with epochs. Reproduced with permission.[166] Copyright 2019, Springer Nature. d) Five-layer memristive CNN with convolution, pooling, nonlinear, and fully connected layers for image recognition. e) The error rate under the hybrid training iteration period. The blue and green curves respectively indicate the training and test set trends. f) The error rate of the test set obtained by the parallel memristive convolvers using hybrid training is extremely lower than that of direct weight transfer. Reproduced with permission.[25] Copyright 2020, Springer Nature.
feature map, and finally forms a set of parallel feature maps. It then introduces nonlinear features into the network and applies pooling operations to continuously reduce the data size and the dimension of the feature map, as shown in Figure 13d. Finally, similar to the perceptron, the neurons in the fully connected layer are connected to all activations in the previous layer, which can be directly implemented in the memristive array to achieve advanced inference classification. Traditionally, the convolution operation in CNN is conducted by CPU or GPU, which requires huge energy consumption, and the energy-efficiency dilemma severely restricts the application of CNN such as portable electronic devices. [15] Fortunately, the redundant and complicated convolution operation in CNN is essentially a vector-matrix product operation, which can be efficiently mapped to a memristive array,[4] where the receptive field pixels are used as inputs, and the convolution kernel is stored in the memristor array column. As the convolution operation occupies most of the computations in CNN, the massive multiplication and summation convolution operations directly implemented in the memristive array hardware significantly improve the CNN energy efficiency and realize the computing acceleration.[15] In recent years, the memristor array-based CNN has attracted great interest, showing explosive growth in applications such as pattern classification, image and signal processing, etc.[154,171] In particular, a fully hardware-implemented CNN with eight memristor arrays (each containing 2048 1T1R memristive cells) has been demonstrated to achieve a high parallel efficiency.[25] The introduction of a hybrid training scheme in memristive CNN reduces the error rate of the test set by 1.12% (Figure 13e), effectively compensating for the accuracy degradation caused by the nonideal characteristics of memristive devices and arrays. In addition, the error rate obtained by the hybrid training in the parallel convolution group is far lower than the result of direct measurement after transferring the weights, as shown in Figure 13f, indicating that hybrid training promotes recognition accuracy and boosts the convolution efficiency. It is worth noting that after mapping the weights to the memristive array, all the multiplication and accumulation in the convolution kernel will be completed in one step. In short, the hardware-implemented memristive CNN exhibits accelerated convolution computing, optimized energy area efficiency, and accuracy comparable with digital implementation in data-centric intelligent applications such as image, speech, and video processing.

For temporal sequence input tasks such as image capture, speech translation, and handwriting recognition, RNN similar to the recurrent connections present in the biological nervous systems shows excellent applicability.[172] As the recurrent neurons in RNN are self-connected or connected to each other, the internal memory states can flow into the network, enabling the circulation of information and processing dynamic temporal behavior, as shown in Figure 14a. However, a standard RNN cannot handle the issues of exponential weight explosion or gradient disappearance with recursion, making it difficult to capture long-term time correlations. Although RNN based on long short-term memory (LSTM) has been developed to solve the vanishing or exploding gradient, the LSTM implemented by digital strategies is generally complex in structure and limited in bandwidth and power consumption. Memristive RNN combined with LSTM cells (Figure 14b) has recently been reported to address the energy-efficiency dilemma and achieve network acceleration concurrently.[173] It is worth mentioning that the Hopfield network is usually regarded as a special RNN, whose output is used as the input of the next-loop iteration rather than the next layer.[15] Similarly, the implementation of network connection mapping through memristive arrays is conducive to accelerate Hopfield network computing, which also required redundant vector-matrix multiplication operations. Hopfield networks have become promising candidates for intelligent applications such as content-addressable memory (CAM).[174] associative learning,[175,176] and decision optimization.[177] The Hopfield network using a Y-type flash memristive array with a symmetric differential weight structure has been reported for three-bit CAM.[163] In different initial states, the memristive Hopfield network enables to recall the prestored “110” and “101” (Figure 14c), indicating associative memory function. It is noticeable that the memristive Hopfield network will always evolve to the closest intermediate state until it converges to the prestored state, which is consistent with the minimum energy consumption of retrieval evolution, as shown in Figure 14d. In addition, the introduction of memristive chaotic sources can ensure that the Hopfield network converges to the global minimized state instead of the local minimum.[178] In summary, as vector-matrix multiplications can be physically implemented in memristive arrays and slow- and high-energy weight programming update operations are reduced, which further accelerates memristive RNN or Hopfield network and optimizes computational efficiency.

SNN has also been implemented through memristive arrays, which is closer to the biological nervous system and often referred to as the third-generation neural network. The input/output signal in SNN consists of a series of discrete spikes, and the information is encoded as the amplitude, frequency, or timing of the spikes (Figure 15a), which more realistically simulates the information processing in the brain. Generally, SNN uses an unsupervised biological learning training method, namely Hebbian STDP, for spike feedback, as shown in Section 2. More importantly, unlike the perceptron and CNN, the neurons in SNN will not be fired in every propagation process but will only be activated in the specific event that the neural membrane potential reaches the threshold[161] and then recovers to the resting level, similar to the brain, as shown in Figure 15a. Therefore, compared with the full activation of the perceptron and CNN, the event-driven SNN exhibits superior time and energy efficiency. Although CMOS hardware-based SNN chips (such as Loihi,[23] TrueNorth,[22] etc.) have demonstrated the feasibility of performing image-processing tasks, the complex peripheral circuit components and architectures make them fall into energy-efficiency bottlenecks. In recent years, the use of emerging memristive devices with phase change[20,182] and ferroelectric[50] and other mechanisms[183] to build electrical and even optical SNN[179] has attracted great interest, all of which are committed to providing computing acceleration and optimizing energy efficiency. For example, ferroelectric synapse array-based SNN implements unsupervised learning to recognize images in a predictive manner with low energy consumption.[50] Admittedly, the network recognition accuracy depends on the image presentation number, noise, and spike amplitude, and the highest recognition rate can even reach 100%, as shown in Figure 15b. Moreover, a hybrid hardware
A platform combining ANN and SNN paradigms has recently been explored to stimulate the development of artificial general intelligence. The ANN and SNN hybrid system enables higher-recognition accuracy by introducing the ANN relay to transmit higher-precision intermediate potentials, while tolerating negligible increases in hardware costs such as energy consumption and core area, as shown in Figure 15c. The hybrid chip was used to build a self-driving bicycle system, which completed advanced intelligent tasks including real-time target detection, tracking, voice command, forward avoidance, and balance maintenance (Figure 15d). Fusion of ANN and SNN hybrid strategies can significantly improve system accuracy and become a potential candidate for building universal intelligent platforms with a high energy and area efficiency. In addition, SNN has been applied in more scenarios, including commercial security monitoring, military radar signal processing, drone collision avoidance, etc., which show strong competitive advantages in exploring advanced AI systems.

5.2. Emerging Computing Implementation

In addition to supporting neural network acceleration, the memristive array with highly parallel in-memory analog computing capabilities can also accelerate some emerging computations.
in neuromorphic engineering, which involve a mass of cumbersome vector-matrix multiplication and accumulation operations. Similar to the neural network, the matrix elements in the vector-matrix dot product operation are mapped to the device conductance in the memristive array. Specifically, the voltage vector is the input into the memristive array, and computing is performed following Ohm’s law and Kirchhoff’s law, which enable the vector-matrix dot product operation to be completed in a single-step readout. In this section, we will give some examples of novel computational accelerations implemented by

Figure 15. Memristive array accelerated SNN. a) The SNN communicates through discrete spikes, and only when the integrated power of the postsynaptic spike exceeds the threshold, the neural output spike will be generated. Reproduced with permission.[179] Copyright 2019, Springer Nature. b) The recognition rate is a function of the number of image representations with varying noise levels and the amplitude of postsynaptic spikes, respectively. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/) [50] Copyright 2017, The Authors, published by Springer Nature. c) The hybrid system that introduces ANN relays transmits the intermediate potential with higher accuracy, enabling higher accuracy, while negligible hardware overhead such as power and area. d) The Tianjic chip based on hybrid ANN/SNN integrated with peripheral modules for multifunctional unmanned bicycle demonstration, including tracking, object detection, audio, visual, balance control. Reproduced with permission.[180] Copyright 2019, Springer Nature.
memristive arrays, including sparse coding, reinforcement learning, reservoir computing, and probabilistic computing.

Sparse coding is essentially looking for an overcomplete basis vector and expressing the input vector as a linear combination of the basis vector to represent input data more efficiently, which was inspired by the human visual neural system to encode vast visual information in a sparse way. In other words, sparse coding conducts feature extraction and compression on high-dimensional data to form a sparse expression and data reconstruction processing, which is widely used in machine vision, image recognition, and classification applications. The key to achieving sparse coding is to introduce an inhibition item into the output neurons, which can be interpreted as activated neurons suppressing neurons with similar features and establishing lateral inhibition between neurons to achieve coding sparseness. Sparse coding generally involves two vector-matrix dot-product operations: forward activation of neurons and backward acquisition of reconstructed inputs. The original and the reconstructed input are subtracted to obtain the residual term, which is fed into the network again for repeated iterations until the network reaches stable convergence. The achievable vector-matrix dot-product capability in one read operation enables memristive arrays to effectively perform sparse coding acceleration. In addition, the memristive array is used to perform the forward and backward vector-matrix multiplication and accumulate operations simultaneously, which can conveniently form lateral inhibition necessary for sparse coding, without the need to construct additional physical inhibitory connections between neurons. For example, a $32 \times 32$ WO$_x$-based memristor array has been reported to demonstrate sparse coding, in which horizontal and vertical bars are used to form a dictionary containing 20 basic elements, as shown in Figure 16a. As expected, the memristive array not only accurately reconstructed the input image (taking the 37th pattern as an example) but also selected neurons 8 and 16 as an efficient solution to achieve sparse expression, as shown in Figure 16b. In a word, the sparse coding implemented by the memristive array effectively identifies the data composition features, reduces the data complexity, and accelerates the data processing and information analysis.

Reinforcement learning emphasizes how to take action based on the environment to obtain the maximum expected benefits, that is, how to gradually form expectations for specific stimuli under the stimulus of the rewards or punishments given by the environment and produce habitual behaviors that can maximize the benefits. It is worth noting that reinforcement learning does not require labeled data, nor does it require accurate correction of suboptimization actions, but focuses on online programming, which is to seek a balance between exploration of unknown territory and exploitation with existing knowledge. Reinforcement learning uses samples to optimize behavior and adopts function approximation to describe complex environments, which makes it powerful when dealing with complex scene tasks. A famous example in reality is the AlphaGo, which conducts training from an innovative combination of human Go game supervised learning and self-playing Go game reinforcement learning. The AlphaGo randomly simulates masses of Go games for training and learning and finally defeated human experts for the first time as a machine, but at the cost of large energy consumption. Potentially, memristive array enables direct hardware implementation of the high-energy-efficient vector-matrix dot product, which greatly reduces time complexity and improves the energy efficiency of reinforcement learning. A digital–analog hybrid 128 × 64 ITIR memristive array has been demonstrated to solve representative reinforcement learning tasks, such as cart pole and mountain car games. Here, we take the mountain car game as an example, in which the car with insufficient engine power must climb a steep mountain to reach the target position (marked with a flag), as shown in Figure 16c. The ultimate goal of mountain car game is to reach the target position as soon as possible to end the game, to minimize negative rewards. Reinforcement learning training can be summarized as repeatedly swinging the car between the hills, which uses the left hill to accumulate potential energy and move to the right until climbing up the right hill. After the first epoch, about 800 negative rewards were suffered, the subsequent performance was significantly improved (Figure 16d), which implies the realization of reinforcement learning. The input state and corresponding learnt action mappings once again verify the successful implementation of reinforcement learning, in which a left push will be applied to the car moving left, and vice versa, resulting in rapid accumulation of momentum to reach the right hill target, as shown in Figure 16e. In general, the reinforcement learning implemented by memristive arrays uses the acquired historical experience to optimize subsequent behavioral decisions in strange environments and shows great potential in achieving speed and energy-efficiency improvements.

As an extended framework of RNN, reservoir computing processes the temporal signals by introducing a nonlinear reservoir, which maps the time-series input to a high-dimensional computing space. Generally, the reservoir computing systems include an input layer for signal feeding, a fixed-connected reservoir layer that maps different reservoir states to the high-dimensional space for distinguishing timing inputs, and an output layer that trains and reads the reservoir states to the desired output, as shown in Figure 17a. Reservoir computing can be described as the transfer of temporal data to the reservoir first, followed by simple readout training to read the state in the reservoir, and finally mapped to the expected output. Similar to conventional RNN, neurons in reservoir computing evolve with time-series input, and the states are determined by both previous and current input signals, so the reservoir is required to have short-term memory capacity. It is a remarkable fact that compared with RNN, the primary benefit of reservoir computing is that training only occurs during the reading process, whereas reservoir dynamics do not require training to benefit from fixed connections, which enables to effectively reduce training costs and improve computational efficiency. Benefiting from the inherent ion dynamic mechanism and volatile electrical behavior of memristive devices, it shows great potential in hardware-implemented reservoir computing for solving time-series tasks such as handwriting classification, speech analysis, signal processing, and motion recognition prediction tasks. Recently, a reservoir composed of 88 memristive device arrays has been used to implement reservoir computing, which demonstrates the feasibility of performing handwritten digit classification tasks. After training, the recognition classification result obtained by testing shows that the inference output is consistent with the expected output, and the overall recognition accuracy reaches...
88.1% (Figure 17b), which is comparable with the implementation of large-scale traditional networks and has the potential to be optimized as the reservoir scale increases. It is worth noting that the device variations and sneak path currents in the memristive array are beneficial to the reservoir calculation, which makes the mapped reservoir state more separated,\cite{188} whereas it is quite unfavorable for memory and neural network applications. Furthermore, quantum reservoir computing based on nano-spin torque magnetization\cite{133} and Fermion lattice dynamics\cite{190} has been reported, which are regarded as promising candidates for quantum information interaction and processing. In brief, memristive device arrays allow the convenient realization of reservoir state mapping by utilizing the inherent nonlinear dynamics and volatile memory properties, which greatly simplify the complexity of the reservoir modules, improve the mapping efficiency, and accelerate the reservoir computing.

The universal probabilistic computing aims to transform current computer systems and applications from human-assisted tools to intelligent partners with understanding and decision-making capabilities. It is easy for humans to classify, infer, and predict the possibility of natural information received and fed from the surrounding environment. For example, when...
driving on the street, if a rolling ball suddenly appears in front, humans usually slow down or stop, because they will assume that there may be some children not far away; however, the computer is still unable to perform similar intelligent behaviors. Generally, the input natural information flow shows inherent uncertainty and variability, and probabilistic computing enables statistics and analysis of information possibilities, which introduce the functions of understanding association, predictive judgment, and making decisions for future intelligent systems, so as to make a human-like response. Immaturely, probabilistic computing is usually implemented by a probabilistic neural network, which consists of input, pattern, summation, and output layers, enabling the input pattern data to be mapped into arbitrary classification outputs. The superiority of probabilistic computing is

Figure 17. Realization of reservoir computing and probabilistic computing through memristive devices and arrays. a) Schematic illustration of the reservoir computing system for digital identification, which includes input, memristive reservoir, and readout layers. b) The false-color confusion matrix shows the correspondence between the classification results obtained experimentally from the reservoir computing system and the desired output. The appearance of the predicted output for each test is represented by the colors shown in the matrix. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). Copyright 2017, The Authors, published by Springer Nature. c) A probabilistic computing network is implemented based on Gaussian synapses, which contain input, pattern, summation, and output layers and are used for brainwave pattern recognition. d) The polysomnography records ten test samples, and the corresponding results processed by the probabilistic network are displayed on the color mapping. Reproduced under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). Copyright 2019, The Authors, published by Springer Nature.
mainly reflected in the processing of nonlinear decision boundary problems. That is, it is generally difficult for ANN to complete nonlinear decision boundary pattern classification with satisfactory accuracy. However, a probabilistic computation-based probabilistic neural network uses nonparametric functions to define classification probabilities, which allow accurate classifications of complex patterns with arbitrary shape decision boundaries. Lately, Gaussian synapses composed of 2D heterostructure-based analog memristive devices have been used in hardware to implement probabilistic computing,\textsuperscript{[189]} which enables the sustainable scaling of energy, size, and complexity. It is impressive that the probabilistic computing based on Gaussian synapses demonstrates the detection and classification functions of human brainwave patterns, as shown in Figure 17c. The amplitude and frequency of the preprocessed electroencephalography data by the fast Fourier transform are mapped to the drain and gate voltages of the Gaussian synapse, which is fed from the input layer to the pattern layer. Subsequently, the summation layer integrates the currents from each pattern module and communicates through the winner-take-all circuit, which enables the output layer to implement brainwave patterns detection classification. Probabilistic computing identified and classified brain waves from sleep recordings of healthy subjects, revealing that brain waves are mainly presented by delta and theta waves during sleep, as shown in Figure 17d. In addition to typical tasks such as complex pattern recognition, classification, and speech translation, advances in probabilistic computing hold the potential to accelerate the expansion of AI applications in the real world, including but not limited to home automation, autonomous driving, and even smart cities.

5.3. Biomimetic Perception and Motion Integration

In addition to facilitating the acceleration of diverse neural networks and emerging computing, the memristive array platform composed of artificial synapses and neurons has become a promising candidate for hardware implementing bionic perception and motion integration systems for future applications such as neural prosthetics, human–machine interaction, and even superhumans. In the biological nervous system, the peripheral afferent nerves transmit the sensed information, followed by instructions from the peripheral effect nerves to make the muscles react and produce movement. The current mainstream technical strategy is to innovate and construct synaptic electronics with self-sensing capabilities or integrated sensory motion modules to pursue the perception and reaction functions in the biological peripheral nervous system. In this section, we will review the implemented biomimetic perception and motion-integration technologies based on memristive devices and arrays sequentially from multisensory (visual, tactile, auditory, pain) perception, integrated afferent nerves, and sensorimotor systems.

The eyes, ears, mouth, nose, and skin are distributed with visual, auditory, gustatory, olfactory, and tactile receptors, respectively, which serve as multisensory channels for humans to acquire external information, enabling them to synergistically perceive and react to the complex environments.\textsuperscript{[157]} For visual perception, ambient light enters the human eyes, followed by the photoreceptor cells on the retina that receives optical information input and adaptively generates appropriate electrical spike signals according to the light intensity for conduction output. The optic nerve, which contains multiple neurons and synapses, converts electrical spike stimuli into physiological ions to transmit relevant neurological information and finally feeds into the visual cortex in the brain for recognition and processing. Humans can visually perceive the size, brightness, and color and movement status of external objects to obtain information vital to survival. It is no exaggeration to say that at least 80% of the external information is obtained through visual perception, and vision serves as the most important sense of biology. With the simultaneous realization of optical image sensing memory and efficient neuromorphic preprocessing and recognition functions, visual perception simulation based on memristive devices and arrays has attracted special attention.\textsuperscript{[43,191,192]} In particular, an array of optoelectronic resistive random access memory (ORRAM) with optical nonvolatile memory and adjustable synaptic behavior has been demonstrated to build an efficient neuromorphic visual perception system, as shown in Figure 18a.\textsuperscript{[38]} The neuromorphic visual perception preprocessing function of the ORRAM array allows removing redundant signals, smoothing the background noise of the input image, and highlighting the outline features of the input letters (Figure 18b), followed by feeding the processed image to the neural network for training and recognition, which significantly improves the efficiency and accuracy, as shown in Figure 18c. In addition, it is worth noting that the human visual system enables adaptive adjustment according to the illumination intensity of the environment, resulting in an accurate perception of external diversified and colorful optical information. To this end, photoelectric memristive neuromorphic devices and arrays have been developed to simulate the photopic and scotopic modulation and mixed color recognition capabilities of visual perception.\textsuperscript{[194]}

In addition to visual perception, biological tactile perception exhibits crucial information reception and self-defense functions, which protect humans from mechanical injuries, resist external destructive substances, and are used to diagnose diseases and even discern emotions. Skin tactile receptors receive mechanical stimulation to form tactile perception, the specific process is described as follows: sensitive nerve cells existing deep in the skin sense the mechanical stress caused by the touch, and immediately send out a tiny current signal, which follows the nerve fiber to the brain, so that the brain can perceive the touch and distinguish the location and intensity of the touch. In recent years, various emerging material engineering-based memristive neuromorphic systems that integrate mechanical stress sensing, signal conversion, transmission, and processing modules for tactile perception have been extensively studied, including but not limited to inorganic,\textsuperscript{[195]} organic,\textsuperscript{[196]} self-energizing,\textsuperscript{[44]} and stretchable rubber-like materials.\textsuperscript{[40]} In particular, a proof-of-concept tactile sensing platform with mechanical sensing, neuromorphic coding, learning, and memory capabilities through optical communication was reported,\textsuperscript{[193]} which integrates MXene pressure receptor, stress–light conversion module, and oxide photoelectric memristor to convert mechanical information into optical signals for transmission and processing, as shown in Figure 18d. It is noteworthy that optical communication enables multichannel signal input, which more realistically
simulates multiple synaptic connections and AP integration between axons and dendrites. A 2 × 2 tactile perception array was constructed to realize the detection of motion state (touch intensity and movement direction) through biomimetic rate and timing coding, as shown in Figure 18e. In addition, benefiting from dimensionality reduction feature coding and training learning, the memristive array tactile perception platform successfully demonstrated the recognition and memory of handwritten letters (Figure 18f), in which the weight mapping after 20 training cycles is almost consistent with the extracted feature dictionary. In general, tactile perception built with neuromorphic memristive devices and arrays shows great potential in motion.
tracking, electronic skin, intelligent robots, and human–machine interface applications.

The biological acoustic transmission system transmits detected external sound waves to the inner ear, and the sound receptors in the inner ear subsequently transduce the sound wave energy into nerve impulses, which are transmitted to the auditory cortex of the brain through the auditory nerve to achieve the auditory perception reflecting the complex characteristics of the sound location, distance, strength, etc. The fundamental function of auditory perception is to feel and distinguish sound, which is of great significance to both animals and humans. Animals use hearing to escape enemies or capture food, and human daily speech communication is established on the basis of auditory perception. Recently, memristive neuromorphic devices and arrays have been exploited to simulate the realization of auditory perception in neuromorphic engineering. For instance, the 2D MoS2 transistor with multiple discrete paired gates simulate coincidence detection neurons and further connect discrete gates to the drain of full-gate MoS2 devices in sequence to construct an interaural time-delay neuron (Figure 19a), which successfully implements an audiomorphic perception architecture for sound localization. When the acoustic transducer receives sound signals with the different interaural time difference (ITD), the color coding of the discrete gate potential shows that for a positive ITD, coincidence occurs on the right split gate, and vice versa (Figure 19b), which is essential in processing sound localization applications. Moreover, an unsupervised learning memristive array network demonstrates the auditory recognition of Korean vowels using electroencephalography signals. In general, memristive device and array-based neuromorphic engineering provide a compact, feasible, and promising paradigm for bionic auditory perception.

In addition, pain perception shows a critical effect on body protection. Pain perception is a function of the body that can produce unpleasant sensations when subjected to destructive stimuli and is usually accompanied by emotional fluctuations and defensive reactions, which play a critical protective role for the body. In addition, pain perception is widely used in clinical medicine and has a positive significance for the diagnosis of auxiliary diseases. The process of biological pain perception is as follows: first, the body receives noxious stimuli, causing tissue damage and releasing specific chemicals; then, the pain receptors receive the chemicals to induce excitement and generate a nerve electrical signal. When the signal exceeds the threshold of pain receptors, nerve impulses are transmitted to the pain-sensing area of the cerebral cortex (Figure 19c), and finally pain perception is realized. Pain perception provides a warning message that the body has or will suffer tissue damage, which enables the nervous system to react defensively to avoid injury. The combination of memristive devices or arrays and peripheral components has attracted widespread interest in recent years to demonstrate biomimetic pain perception imitation. For example, Yoon et al. reported a compact and realistic bionic thermal nociceptor system, which integrates a thermoelastic module, a diffusion memristor with ion kinetics, and a series resistances for voltage readout, as shown in Figure 19d. The thermoelastic module is heated to different temperatures with a hot plate at the same time to produce varying voltage stimuli fed into the memristor, and the voltage gradually decreases after the hot plate is removed. Experimental results show that at 40 °C, the partial voltage of the memristor is still below the threshold and is in the off state. However, when the temperature reaches 50 °C and above, a significant voltage signal is the output, which simulates the nerve impulse transmitted to the human nervous system and indicates that the memristor is switched on, as shown in Figure 19e. It is worth mentioning that when the temperature rises, the output signal amplitude increases and the onset time becomes earlier, which means that a higher temperature causes a larger voltage drop on the memristor, resulting in a faster on-state switching (Figure 19e). Moreover, organic and inorganic hybrid transistors have been utilized to imitate the pain perception and sensitization tunable nociceptor as well. In short, pain perception implemented by memristive devices and systems exhibits compact, scalable, and biologically realistic excellent properties and has great potential in applications such as intelligent humanoid robots, human–machine engineering, and neurorehabilitation.

Based on the somatosensory functions discussed earlier, building a complete intelligent system that integrates perception and motion becomes the technical cornerstone for the realization of electronic neuroprosthetics and bionic intelligent robots. Biological afferent nerves are responsible for transmitting nerve impulses from sensory receptors or organs to the central nervous system through afferent pathways, which are indispensable in the biological motor nervous system. Therefore, it is necessary to construct artificial afferent and efferent nerves, which convert the sensed signals into spikes, and then feed them into the effector to make a motion response. Memristive devices and arrays capable of simulating the functions of biological synapses and neurons, integrated with external stimulus sensing, conversion, transmission, and processing response modules, have become potential candidates for the construction of afferent and efferent nerves. For example, a compact artificial spike afferent nerve based on Mott memristor, oscillator, and resistor has been reported to convert continuous analog signals input externally into oscillation spikes suitable for SNN. The constructed universal artificial spike afferent nerve has proved to show great potential in both SNN and bionic robot applications with efficient signal preprocessing and transmission. The bionic electronic system combined with pressure sensors, ring oscillators, and synaptic transistors demonstrates the simulation of biological afferent nerves. The pressure sensor acts as a mechanoreceptor to convert pressure information into a voltage change, followed by a ring oscillator that serves as a nerve fiber to encode a voltage spike, and finally the voltage spikes from multiple nerve fibers are aggregated into the synaptic transistor to generate a postsynaptic current response, as shown in Figure 20a. The postsynaptic current is converted into an amplified voltage, which is connected to the efferent nerve in a discrete cockroach leg through reference and stimulation electrodes. The combination of artificial afferent nerve and biological effector nerve realizes a hybrid motion reflex arc, which is evaluated by measuring the force of the cockroach leg extension (Figure 20b). Figure 20c shows that the input pressure information activates the cockroach leg muscles and produces a clear leg extension response. Typically, the artificial afferent nerves confirmed the feasibility of implementing neural prosthetic technology through neuromorphic systems that integrate sensory, transmission, and responsive capabilities. Furthermore,
neuromorphic engineering that integrates memristive devices and additional auxiliary components demonstrates the imitation of the human sensorimotor nervous system, which can process external stimulus information and make motion responses. For the human sensorimotor system, specific neurons (such as photosensitive neurons) receive external stimuli (corresponding to optical signals) to generate presynaptic spikes, which are transmitted through axons to neuromuscular junctions. Through the release of neurotransmitters, presynaptic stimulation information is transmitted to muscle fibers and induces muscle motor responses (contraction or relaxation), as shown in Figure 20d. Recently, an integrated strategy of the optical synaptic device and metal actuator module successfully simulated the human sensorimotor system. The photoelectric
memristive synapse receives optical stimuli to generate postsynaptic spikes, which are transmitted to the metal actuator device, causing the oscillating motion of the droplets in the electrolyte, and recorded with mechanical or electrical reactions, as shown in Figure 20e. Similarly, the neuromuscular systems based on organic photoelectric synapses and stretchable nanowire synaptic transistors enable the simulation of the human sensorimotor system.\(^{[41]}\) It is worth noting that the light-driven motion-regulated artificial sensorimotor system allows wireless interaction and operation, which can significantly reduce the complexity and energy consumption of next-generation bionic electronics applications. In short, neuromorphic engineering combined with memristive devices and motion modulation units provides a universal paradigm for the development of artificial sensorimotor systems integrating perception and motion functions, which will be widely adopted in future neural prostheses and bionic robotics applications.

6. Conclusion and Perspective

Here we summarize the advanced progress of neuromorphic engineering based on memristive devices and arrays in terms of biological basis, device-level memristive synapses, circuit-level
artificial neurons, system-level hardware neural networks, emerging computation, and bionic perception motion systems. Although memristive devices and array-based neuromorphic engineering have achieved remarkable success, there are indeed challenges and steps that have to be resolved and taken both at the biological and at the electronic levels to advance the application of neuromorphic engineering, including hardware computing acceleration and bionic intelligent system construction.

First of all, on the biological level, the current cognition of information interaction and processing mechanisms in the human brain is superficial, which is the fundamental obstacle to the evolution of brain-inspired neuromorphic engineering applications. Fortunately, in recent years, brain-mapping technology has developed rapidly, and it has become a reality to draw detailed animal-to-human brain neuron-level static connection structures. The dynamic mechanism of brain detection, such as the interpretation of low-level visual cortex information in animals, has also made great progress. It is believed that the experimental evidence of brain science will greatly help make breakthroughs in neuromorphic engineering applications.

In addition, at the electronics level, from memristive devices, circuits to systems are facing certain challenges, severely restricting the application of neuromorphic engineering in computing acceleration and biomimetic implementation. For memristive synaptic devices, different mechanisms can only provide one or two advantages in the brain-inspired learning process, like low power consumption, fast operation speed, high weight precision, high stability, long repeatability, high scalability, linearity/symmetry weight change, large weight modulation range, and compatibility with CMOS technology. Therefore, an ideal artificial synapse that integrates all these advantages is highly desirable for neuromorphic engineering development, which inspires a mass of synaptic researches with the novel mechanism, as shown in the Review. It is worth mentioning that for hardware computing acceleration and biomimetic implementation technology, the ideal synapse properties may be different. For example, an ideal synapse for hardware computing acceleration is expected to have a short decay time, whereas for bionic perception and motion implementation technology, a relatively longer decay time is desired. For artificial neuron circuits, compact circuit modules are required to realize more neuron functions and optimize circuit integration and scalability. Finally, for memristive devices and array-based system-level hardware networks, emerging computations and bionic perception motion integration applications all present specific limitations and challenges that have to be addressed. In terms of hardware networks and emerging computations acceleration that involve massive vector-matrix multiplication, device properties are critical, and nonideal memristive device characteristics such as nonlinear asymmetric weight update will greatly degrade network and computation performance. Typically, PCM conductance exhibits an abrupt change during reset, resulting in severe weight update asymmetry. To this end, an arbitration training scheme with multiple PCM memristive synapses is proposed to compensate for weight potentiation and inhibition mismatch. In addition, the strategy of combining nonvolatile memristive devices with linear and symmetric update devices enables solving the asymmetric and nonlinear dilemma of weight update in hardware network training, which results in a high accuracy comparable with the software implementation. In addition, the sneak current in the memristive array, that is, the current flowing through the unselected cross-point cells, exhibits an adverse effect on the network accuracy. Solutions such as one-selector one-resistor (1S1R) and 1T1R structures have been proposed to compensate for sneak current problems, but at the expense of system compactness, and compatibility concerns still exist. However, it is worth noting that system-level neuromorphic engineering applications are usually immune to device variations. The degree of variation that the system can tolerate depends on the precision required by the specific network and computation application. Moreover, the lack of novel efficient network training algorithms for well-matched memristive device arrays is a core limitation, especially for SNN, which does not possess the mature algorithm such as back-propagation used in deep neural networks. The fundamental reason lies in the fact that the mechanism of energy-efficient brain information processing and learning has not been fully explored, and the inspiration is limited. Hopefully, biologically realistic STDP and SRDP have demonstrated to be promising alternatives to traditional neuromorphic computing learning rules, which still need continuous optimization and development to adapt to diverse applications and further improve computational efficiency. Advances in neuroscience and mathematical algorithm science will spur the development of SNN that realistically mimics the working principle of the brain. Emerging computing based on memristive devices and arrays also encounters challenges. Taking reinforcement learning as an example, the copy operation between networks requires two memristive arrays to achieve conductance replication, whereas the intrinsic random nature of memristive conductance makes it difficult to accurately program each device to the target value. Furthermore, emerging computing requires numerous high-precision registers and computing elements when performing network training for weight update. The modified stochastic gradient algorithm suitable for memristive devices is introduced, which takes into account the convergence guarantee and cost reduction of memristive array training. In addition, system performance including, but not limited to, area efficiency, power consumption, and technical complexity must be evaluated and balanced with uniform standards to promote the design of electronic design automation in neuromorphic engineering. In short, although the current state-of-the-art neuromorphic hardware systems still require collaborative optimization in terms of architecture, algorithms, energy efficiency, and evaluation criteria, neuromorphic computing makes it possible to construct extremely energy-efficient artificial general intelligence hardware systems. Human skin and functional organs usually exhibit superior conformal properties, which indicate that the performance of the memristive devices and auxiliary components that constitute the bionic perception and motion system is required to tolerate stress and deformation. However, despite previous reports of flexible memristive synapses, systematic work on mechanical stress-modulated memristive devices and array performance is lacking. The co-optimized design of structure and materials has become a promising path for memristive devices and peripheral modules to maintain stable and reliable responses under mechanical strain. Essentially, the ultimate goal of building bionic perception and motor systems is to achieve a partial or complete replacement of certain human abilities (such as...
sensory and motor functions). Therefore, the amplitude and frequency matching of the interactive signals between neuromorphic engineering electronics and biology is crucial. Fortunately, the spike frequency (SRDP), duration, and number of memristive components provide the feasibility of flexible and tunable response signals to match the biology in biomimetic electronic systems.

In summary, despite the difficulties, neuromorphic engineering based on memristive devices and arrays shows great potential in hardware computing acceleration and bionic perception motion integrated system construction. With the collaborative optimization of materials, devices, circuits, systems, and neuroscience, neuromorphic engineering for next-generation computing, neural prosthetics, bionic robots, and brain–machine interaction will achieve unprecedented breakthroughs.

Acknowledgements

S. Y. Wang and X. Z. Chen contributed equally to this work. This work was supported by the National Natural Science Foundation of China (61925402, 61851402, and 61734003), Science and Technology Commission of Shanghai Municipality (19JC1416600), National Key Research and Development Program (2017YFB0405600), Shanghai Education Development Foundation, and Shanghai Municipal Education Commission Shuguang Program (18SG01).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

artificial synapses and neurons, biomimetic perception motion systems, hardware computational accelerations, memristive devices and arrays, neuromorphic engineering

Received: June 8, 2020
Revised: July 4, 2020
Published online: August 7, 2020

[1] J. Von Neumann, IEEE Ann. Hist. Comput. 1993, 15, 27.
[2] S. A. McKee, in Proc. of the 1st Conf. on Computing Frontiers, ACM Press, Santa Clara, CA 2004.
[3] M. Horowitz, 2014 IEEE Int. Solid-State Circuits Conf. Digest of Technical Papers (ISSCC), IEEE, Piscataway, NJ 2014.
[4] M. Hu, C. E. Graves, C. Li, Y. Li, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Adv. Mater. 2018, 30, 1705914.
[5] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, D. Glasco, IEEE Micro 2011, 31, 7.
[6] V. Sze, Y.-H. Chen, T.-J. Yang, J. S. Emer, Proc. IEEE 2017, 105, 2295.
[7] G. Singh, L. Chelini, S. Corda, A. J. Awan, S. Stuijk, R. Jordans, H. Corporaal, A.-J. Boonstra, in 2018 21st Euromicro Conf. on Digital System Design (DSD), IEEE, Piscataway, NJ 2018.
[8] S. Khoram, Y. Zha, J. Zhang, J. Li, in Proc. of the 2017 ACM on Int. Symp. on Physical Design, Association for Computing Machinery, New York 2017.
[9] J. T. Pawlowski, in 2011 IEEE Hot Chips 23 Symp. (HCS), IEEE, Piscataway, NJ 2011.
[10] D. U. Lee, K. W. Kim, K. W. Kim, H. Kim, J. Y. Kim, Y. J. Park, J. H. Kim, D. S. Kim, H. B. Park, J. W. Shin, 2014 IEEE Int. Solid-State Circuits Conf. Digest of Technical Papers (ISSCC), IEEE, Piscataway, NJ 2014.
[11] M. Koyanagi, T. Fukushima, T. Tanaka, Proc. IEEE 2009, 97, 49.
[12] M. M. Shulaker, G. Hills, R. S. Park, R. T. Howe, K. Saraswat, H.-S. P. Wong, S. Mitra, Nature 2017, 547, 74.
[13] L. A. Hart, How the Brain Works: A New Understanding of Human Learning, Emotion, and Thinking, Basic Books, New York 1975.
[14] D. Kuzum, S. Yu, H. P. Wong, Nanotechnology 2013, 24, 382001.
[15] J. Tang, F. Yuan, X. Shen, Z. Wang, M. Rao, Y. He, Y. Sun, X. Li, W. Zhang, Y. Li, Adv. Mater. 2019, 31, 1902761.
[16] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. O. Biochemistry, M. B. T. Jessell, S. Siegelbaum, A. Hudspeth, Principles of Neural Science, McGraw-Hill, New York 2000.
[17] M. Mayford, S. A. Siegelbaum, E. R. Kandel, Cold Spring Harb. Perspect. Biol. 2012, 4, a005751.
[18] M. A. Zidan, J. P. Strachan, W. D. Lu, Nat. Electron. 2018, 1, 22.
[19] K. Roy, A. Jaiswal, P. Panda, Nature 2019, 575, 607.
[20] I. Boybat, M. Le Gallo, S. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, E. Eleftheriou, Nat. Commun. 2018, 9, 1.
[21] C. Mead, L. Conway, Introduction to VLSI Systems, Addison-Wesley, Reading, MA 1980.
[22] 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, Science 2014, 345, 668.
[23] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, IEEE Micro 2018, 38, 82.
[24] Z. Wang, S. Joshi, S. E. Savel’ev, H. Jiang, R. Middy, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Nat. Mater. 2017, 16, 101.
[25] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, Nature 2020, 577, 641.
[26] Y. He, S. Nie, R. Liu, S. Jiang, Y. Shi, Q. Wan, Adv. Mater. 2019, 31, 1900903.
[27] Y. van de Burgt, E. Luberberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. A. Talin, A. Salleso, Nat. Mater. 2017, 16, 1414.
[28] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, Nat. Nanotechnol. 2016, 11, 693.
[29] W. Chung, M. Si, D. Y. Peide, in Technical Digest—Int. Electron Devices Meeting (IEDM), IEEE, San Francisco, CA 2018, pp. 15.2.1–15.2.4.
[30] M.-K. Kim, J.-S. Lee, Nano. 2019, 19, 2044.
[31] A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yusa, J. Grollier, D. Querlioz, Nat. Commun. 2018, 9, 1.
[32] J. K. Shen, D. S. Shang, Y. S. Chai, S. G. Wang, B. G. Shen, Y. Sun, Adv. Mater. 2018, 30, 1706717.
[33] S. Wang, C. Chen, Z. Yu, Y. He, X. Chen, Q. Wan, Y. Shi, D. W. Zhang, H. Zhou, X. Wang, Adv. Mater. 2019, 31, 1806227.
[34] V. K. Sangwan, H.-S. Lee, H. Bergeron, I. Balla, M. E. Beck, K.-S. Chen, M. C. Hersam, Nature 2018, 554, 500.
[35] T. Zhang, K. Yang, X. Xu, Y. Cai, Y. Yang, R. Huang, Phys. Status Solidi Rapid Res. Lett. 2019, 13, 1900029.
[36] Y. Lee, T.-W. Lee, Acc. Chem. Res. 2019, 52, 964.
[37] H. L. Park, Y. Lee, N. Kim, D. G. Seo, G. T. Go, T. W. Lee, Adv. Mater. 2019, 32, 1903558.
[38] F. Zhou, Z. Zhou, J. Chen, T. H. Choy, J. Wang, N. Zhang, Z. Lin, S. Yu, J. Kang, H.-S. P. Wong, Nat. Nanotechnol. 2019, 14, 776.
[39] J. H. Yoon, Z. Wang, K. M. Kim, H. Wu, V. Ravichandran, Q. Xia, C. S. Hwang, J. J. Yang, Nat. Commun. 2018, 9, 1.
[40] H. Shim, K. Sim, F. Ershad, P. Yang, A. Thukral, Z. Rao, H.-J. Kim, J. Liu, X. Wang, G. Gu, Sci. Adv. 2019, 5, eaax4961.
[41] Y. Lee, J. Y. Oh, W. Xu, O. Kim, T. R. Kim, J. Kang, Y. Kim, D. Son, J. B.-H. Tok, M. J. Park, Sci. Adv. 2018, 4, eaat7387.
Shuiyuan Wang is a Ph.D. student at the state key laboratory of ASIC and system, School of Microelectronics, Fudan University, China. He received his B.S. degree in 2017 from Lanzhou University, China. His major research interest is in the field of varieties of novel electronic and optoelectronic devices based on nanoscaled 2D materials, especially 2D atomic crystal-based artificial synaptic electronics and neuromorphic systems.

Xiaozhang Chen was born in Heze, Shandong Province, China in 1996. She received the B.S. degrees in IC Design and Integrated System of Xidian University, Xi’an, Shanxi Province, China in 2017. Now, she is currently pursuing M.S degree and Ph.D. degree in microelectronics and solid-state electronics of Fudan University, Shanghai, China. From 2017 to now, her research interests include the electrical properties of low-dimensional materials, computer memory, and fabrication of micro- or nanostructured devices based on layer material. She has published two articles on 2D materials.

Peng Zhou is a full professor at the state key laboratory of ASIC and system, School of Microelectronics, Fudan University, China. He received his B.S. (2000) and Ph.D. (2005) degrees in physics from Fudan University, China. Currently, Professor Zhou is interested in novel high-efficiency and low-power electronic devices based on 2D layered materials, focusing on the application of 2D materials in memory, including 2D flash memory, floating gate memory, semifloating gate quasinonvolatile memory, synaptic electronics, and neuromorphic systems.