



Bridging Biological and Artificial Neural Networks with Emerging Neuromorphic Devices: Fundamentals, Progress, and Challenges

Jianshi Tang, Fang Yuan, Xinke Shen, Zhongrui Wang, Mingyi Rao, Yuanyuan He, Yuhao Sun, Xinyi Li, Wenbin Zhang, Yijun Li, Bin Gao, He Qian, Guoqiang Bi, Sen Song, J. Joshua Yang,* and Huaqiang Wu*

As the research on artificial intelligence booms, there is broad interest in brain-inspired computing using novel neuromorphic devices. The potential of various emerging materials and devices for neuromorphic computing has attracted extensive research efforts, leading to a large number of publications. Going forward, in order to better emulate the brain's functions, its relevant fundamentals, working mechanisms, and resultant behaviors need to be re-visited, better understood, and connected to electronics. A systematic overview of biological and artificial neural systems is given, along with their related critical mechanisms. Recent progress in neuromorphic devices is reviewed and, more importantly, the existing challenges are highlighted to hopefully shed light on future research directions.

Prof. J. Tang, Dr. F. Yuan, Dr. X. Li, W. Zhang, Y. Li, Prof. B. Gao,
 Prof. H. Qian, Prof. H. Wu

Institute of Microelectronics
 Beijing Innovation Center for Future Chips (ICFC)
 Tsinghua University
 Beijing 100084, China
 E-mail: wuhq@tsinghua.edu.cn

Prof. J. Tang, Prof. B. Gao, Prof. H. Qian, Prof. H. Wu
 Beijing National Research Center for Information Science
 and Technology (BNRist)
 Tsinghua University
 Beijing 100084, China

X. Shen, Y. He, Y. Sun, Prof. S. Song
 Tsinghua Laboratory of Brain and Intelligence and Department
 of Biomedical Engineering
 School of Medicine
 Tsinghua University
 Beijing 100084, China

Dr. Z. Wang, M. Rao, Prof. J. J. Yang
 Department of Electrical and Computer Engineering
 University of Massachusetts
 Amherst, MA 01003, USA
 E-mail: jjyang@umass.edu

Prof. G. Bi
 School of Life Sciences
 University of Science and Technology of China
 Hefei 230027, China

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.201902761>.

DOI: [10.1002/adma.201902761](https://doi.org/10.1002/adma.201902761)

1. Introduction

The history of artificial intelligence (AI) may be traced back to the 1940s, when the British mathematician, Turing, suggested that AI would be best researched by programming computers.^[1] Encouragingly, AI has indeed taken off during the past decade or so, and has already experienced explosive growth in applications such as image and speech recognition, social networks, finance, medical science, cyber security, gaming, etc. The boom of AI has been largely driven by the availability of large datasets and the use of

graphics processing unit (GPU) to implement the training of artificial neural networks (ANNs) with backpropagation algorithm. While different algorithms and architectures are still being pursued to improve the training efficiency, the speed and energy efficiency of silicon CMOS-based computing hardware is quickly approaching its theoretical limit.^[2] The main obstacles that hinder further improvement in computing efficiency include: 1) the Moore's law-based device scaling and accompanying technology development have been significantly slowing down; 2) the physical separation of memory and data processing units increases the cost for "big data" movements, known as the "von Neumann bottleneck." This issue becomes more and more alarming as the performance gap between memory and processor keeps expanding (known as the "memory wall" problem). Therefore, to address those challenges, new computing hardware and architectures are in urgent need to meet the requirements for data-intensive computing like AI.

Brain-inspired computing has emerged as a new computing paradigm to enable massively parallel analog computing for deep learning. For example, it is projected that a new architecture of resistive processing unit (RPU) could provide 30 000x acceleration compared to state-of-the-art CPU/GPU in training deep neural networks (DNNs).^[3] Experimentally, the first attempt to build an artificial neural machine is SNARC, short for "Stochastic Neural-Analog Reinforcement Computer," using vacuum tubes to simulate a network of 40 neurons, dated back to the 1950s.^[4] Since then, extensive

research has been done to mimic the complex dynamics of neurons and synapses using CMOS integrated circuits, and this subfield of brain-inspired computing is often called neuromorphic computing because of the more faithful emulation of biological neurons and synapses. The term “neuromorphic” was coined by Carver Mead in the late 1980s,^[5–7] who has carried out considerable pioneer work on neuromorphic silicon neuron circuits (see a recent review article by Indiveri et al.^[8]). A more recent example along this direction is IBM’s TrueNorth chip, which has 4096 cores built with 5.4 billion transistors that integrate 1 million programmable spiking neurons and 256 million configurable synapses, taking full advantage of the mature silicon technology.^[9] The TrueNorth chip is shown to be well suited for multi-object detection and classification. Another example is Intel’s Loihi chip announced in 2017, which has 128 neuromorphic cores, each containing 1024 primitive spiking neural units grouped into tree-like structures, for a total of around 131 thousand simulated neurons and nearly 130 million synapses. Different from TrueNorth chip, the Loihi chip can implement not only inference but also self-learning based on spike neural networks (SNNs).

However, such CMOS-based approaches may not be the ultimate solutions for future AI applications, because they usually have complex circuit architectures that are power-hungry and occupy massive chip area, and also have limits in further scalability amid the slowdown in CMOS scaling. In parallel, emerging materials and devices, especially non-volatile memories (NVMs) such as resistive random-access memory (RRAM), conducting bridge random-access memory (CBRAM), phase-change memory (PCM), and magnetic random-access memory (MRAM), have been extensively studied as synaptic elements to build prototype ANNs for energy-efficient neuromorphic computing. Volatile memory devices, on the other hand, are intrinsically equipped with desirable dynamics, which can be explored to implement the critical dynamics in biological neural networks (BioNNs).^[10] Both those nonvolatile and volatile devices fit in the general definition of memristive devices and are also called memristors.^[11,12] The beauty of those NVMs is that they can be readily built into cross-point arrays that directly map ANNs, where the conductances of NVMs are used to represent the weights in ANNs, the voltage on each bit line represents the input, and the current collected from each word line already completes the “vector-matrix” multiplication based on Ohm’s law and Kirchhoff’s law. Having weights (data) stored locally for in-memory computing, such architecture can provide enormous parallel computation and hardware acceleration for neural network inferencing and training. Therefore, emerging NVMs have tremendous potential for highly efficient neuromorphic computing, and hence attracted broad research interest from material scientists, electrical engineers, physicists, and computer scientists.

Nevertheless, it should be noted that there are still many technical challenges to implement ANNs with existing NVMs for neuromorphic computing, many of which arise from their imperfect device performance. For example, RRAM and CBRAM usually exhibit large stochasticity and limited endurance, while PCM typically shows asymmetric switching (i.e.,



Jianshi Tang received his B.S. degree from Tsinghua University in 2008 and Ph.D. degree from University of California, Los Angeles in 2014. From 2015 to 2019, he worked at IBM T. J. Watson Research Center. He is currently an assistant professor at the Institute of Microelectronics, Tsinghua University. His research interest includes neuromorphic computing, carbon electronics, spintronics, etc.



J. Joshua Yang received his B.A. degree in mechanical engineering from Southeast University in China and Ph.D. from the University of Wisconsin-Madison in Material Science Program in 2007. He is currently a professor in the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst. Before Joining UMass in 2015, he spent over 8 years at HP Labs, leading the materials and devices team. His current research interests are nanoelectronics and nanoionics, especially for energy and computing applications.



Huaqiang Wu received his double B.S. degrees in material science & engineering and enterprise management from Tsinghua University in 2000 and Ph.D. degree in electrical engineering from Cornell University in 2005. From 2006 to 2008, he was a Senior Engineer at Spansion LLC. He is currently a full professor and the Deputy Director at the Institute of Microelectronics, Tsinghua University, and also serves as the Deputy Director of the Beijing Innovation Center for Future Chip. His research interests include emerging memory and neuromorphic computing.

can be explored to mimic certain functions of biological neural systems, such as forgetting and stochasticity. The number of biological structures and functions that have been mimicked electrically using emerging devices is still very limited. To better emulate the brain's functions, it is believed that the fundamental basis of the brain, including its structure and components, as well as the behaviors and working mechanisms of neural circuits, needs to be re-visited and better understood in the context of neuromorphic computing. There have been many valuable review articles published in this field. For example, a classic tutorial by Basheer and Hajmeer provides a practical guide to design ANNs as well as a brief history of the evolution of neuromorphic computing.^[13] In addition, Rajendran and Alibart,^[14] along with Jeong and Hwang,^[15] have surveyed the mechanisms of resistive switching, electrical properties of resistive switches, and memristive ANNs. A similar review by Kim et al.^[16] focuses on the resistive switching physics and ANN learning rules in memristive neural networks. In this review, we intend to bridge the gap between BioNNs and ANNs, including both DNNs and SNNs, by giving a comprehensive overview of both systems, including their key computing elements and related functions. As a comparison, state-of-the-art ANNs using emerging neuromorphic devices will be reviewed. In parallel, the recent progress of neuroscience in the related field will also be covered. The connections and gaps between these two systems as well as the existing challenges for building ANNs to more faithfully mimic BioNNs will be discussed. By doing so, we hope to provide a broad perspective on neuromorphic computing and inspire future research on neuromorphic materials and devices.

This review paper is structured as follows: Section 2 gives a high-level overview of BioNNs and ANNs with a side-by-side hierarchical comparison between human's nervous system and artificial neural system. Section 3 then gives a detailed comparison between the critical functional components of BioNNs and ANNs, including neurons and synapses along with the associated gated ions channels. The recent demonstrations of artificial neurons and synapses using emerging devices, in particular with resistive switching characteristics, are also reviewed in this section. Section 4 comprehensively discusses the key concept of plasticity, including synaptic plasticity (long-term and short-term as well as wiring plasticity) and nonsynaptic plasticity (intrinsic plasticity). Metaplasticity and synaptic consolidation are also introduced. The electronic implementations (if any) of each type of plasticity are also briefly reviewed in this section. Section 5 further reviews the classical theories of learning and memory and their principles in biological neural systems, including Hebbian learning and homeostatic plasticity as well as memory engram and consolidation. The implementations of those learning rules using emerging devices are also reviewed in this section, including unsupervised and supervised learning, reinforcement learning, reservoir computing, and one-shot learning. Finally, Section 6 provides an outlook for future research on biological and artificial neural systems by discussing the considerable gap between them, in light of the great diversity and complexity of plasticity and learning rules in biological systems.

2. An Overview of Biological and Artificial Neural Systems

The nervous system in vertebrates (e.g., human beings) is where intelligence resides. It is composed of the brain, the spinal cord, and peripheral nerves all over the body.^[17,18] In this system, computations that support intelligent functions such as memorizing and forgetting, learning, and decision making are carried out in various neural circuitries in the brain. As a side-by-side comparison, **Figure 1** provides a high-level overview of the human nervous system and artificial neural system built with emerging neuromorphic devices. Both systems span multiple levels of organization. Three basic levels can be distinguished. On the top level, the human nervous system is a vast mesh of different types of neural networks organized hierarchically to support different computational functions such as vision, audition, emotion, etc. On the medium level, the basic unit is a neuron, which is composed of a soma, with many dendrites to receive inputs, and a single axon (usually with many branches) to send out outputs. These neurons connect to one another via synapses. On the bottom level, different types of ion channels form the molecular basis for electrical activities in neurons and support the transmission and processing of information. The number and properties of ion channels in a neuron are regulated by other cellular signaling machinery. Similarly, a typical AI chip is composed of different types of ANNs that can be mapped onto cross-point arrays of resistive switching elements, where the artificial synapses and neurons can be operated through conduction channels (filaments) that are associated with ion movements driven by electric field, Joule heating or electrochemical potential. Before going into details, let's first compare the two systems on the network level, that is, neural circuitries and ANNs.

2.1. Neural Circuitries

In biological neural systems, numerous neurons interconnect to form circuitries of different scales, from microcircuits to large scale networks that carry out various functions.^[17,18] Among them, the architecture of the visual system is the most thoroughly studied and gives direct inspiration to convolutional neural networks (CNNs) in AI.^[19] Here we take the mammalian visual system as an example to explain neural circuitries, as illustrated in **Figure 2**. The nervous systems use two very important principles to parallel streams and hierarchical processing to organize computations.

In the first part of mammalian visual system, retina is an extraordinary sensor and processor of visual information. Light energy is transformed into electrical signals in two types of photoreceptors, rods and cones, in the retina. Rods are very sensitive to light, while three types of cones specialize in color vision. Therefore, the brain starts to use the principle of parallel streams from the earliest stages. The signals are passed from rods and cones to bipolar cells, and ultimately ganglion cells in a vertical direction (Figure 1). Each ganglion cell only responds to luminance contrast (i.e., bright in the center and dark on the periphery, or vice versa) in a specific position, which is called its receptive field. This is another example of

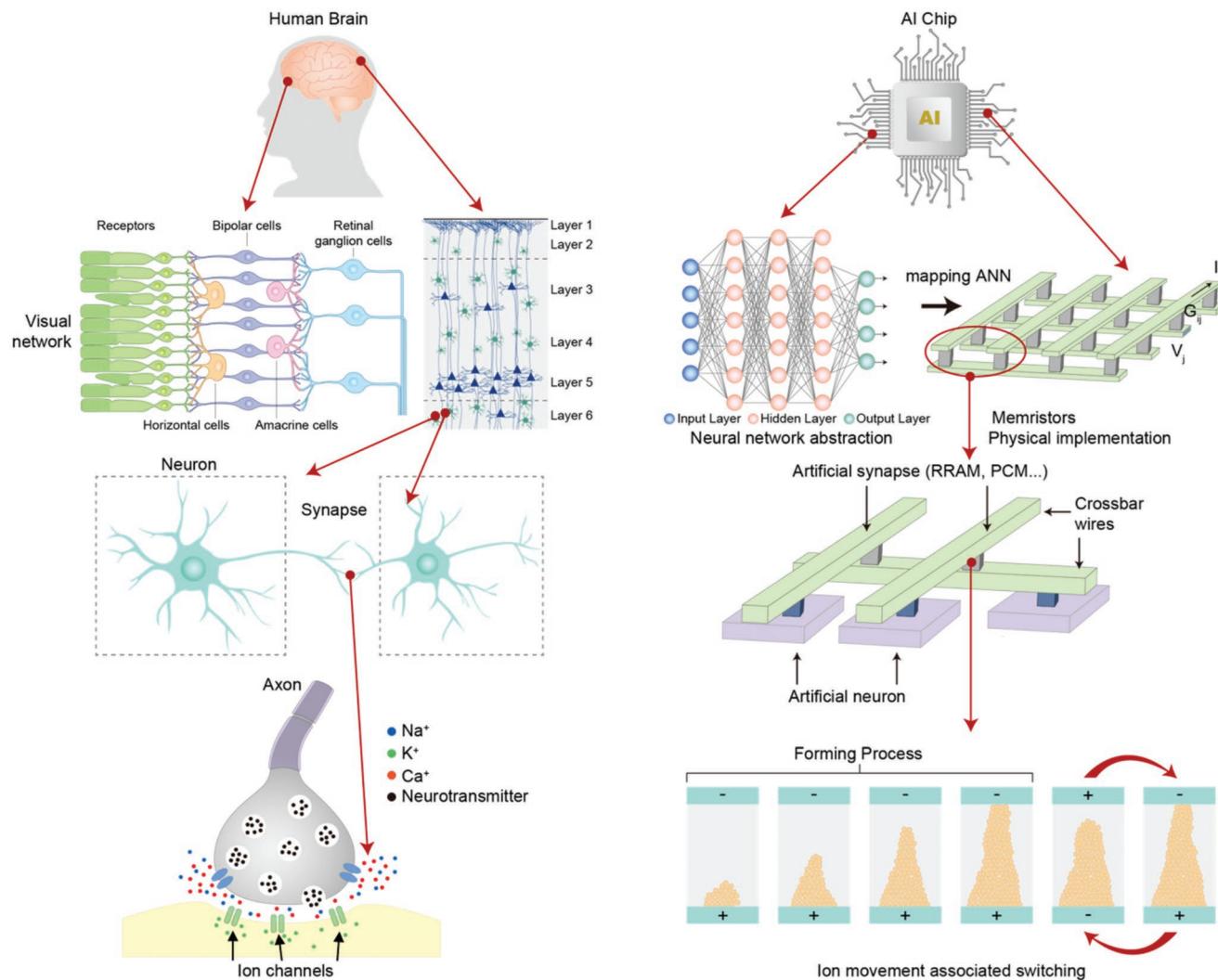


Figure 1. A high-level comparison of the human nervous system and an artificial neural system built with emerging neuromorphic devices. The human nervous system (left panel) has different types of neural networks whose basic functional elements are neurons and synapses, in which different types of ion channels underlie electrical neuronal activity. Similarly, an AI chip (right panel) is composed of different types of ANNs that can be mapped with cross-point arrays of artificial synapses and neurons, whose operation mechanism could be conductive filaments associated with electrically induced ion movements as in the case of RRAM.

parallel processing. The most important computational function of the center-surround receptive field is to form an edge map. In order to carry out this computation, there are so-called interneurons, specifically horizontal cells, and amacrine cells gathering information along the horizontal dimension. Ganglion cells are projection neurons that send axons which go out of the retina. The specialization of neurons into primarily information transmission and local information processing roles is another important principle.

Visual signals are then sent from the retina to primary visual cortex, or V1 area, relayed by lateral geniculate nucleus (LGN). Unlike ganglion cells, some neurons in V1 respond to light-dark bars or edges with particular orientation, rather than a light spot in the center. These neurons are called simple cells. The information encoded by those neurons is mainly determined by the connection pattern between the upstream LGN neurons and themselves. Another type of neurons, complex

cells, have similar receptive field with simple cells, but they are more positionally invariant, which means light-dark edges in a large range of its receptive field can activate the neuron. The main challenge of objective recognition is to achieve high selectivity to object identity, while remaining invariant to the appearance of the object. The two types of neurons embody the neuronal basis of the selection and invariance operations in the brain, which are imitated by the currently very successful convolutional ANNs.

Beyond V1, visual information is divided into two separate pathways; one is specialized for object recognition, and the other for spatial relationship processing and movement detection. Both pathways are organized with hierarchical architecture, with higher-level areas responding to more specialized and abstract information, like faces or houses.^[17,18] Therefore, on a more macrolevel, we also see the principles of parallel and hierarchical processing at work.

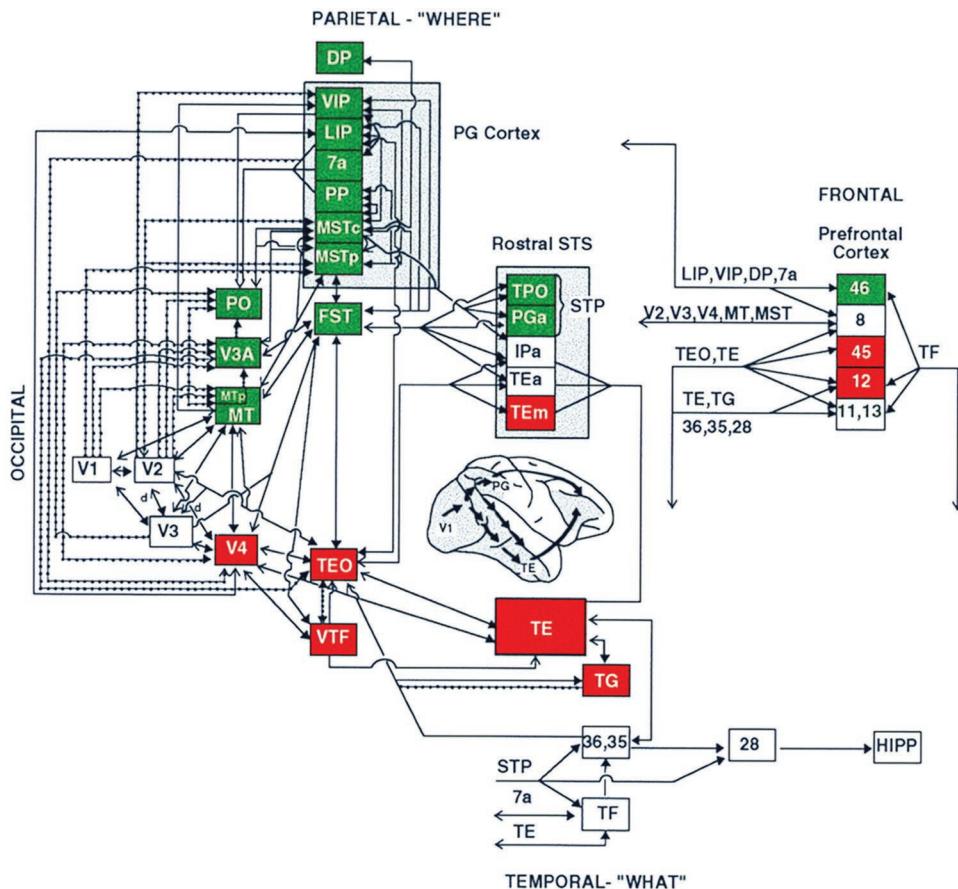


Figure 2. Hierarchical and parallel processing in primate visual system. There are two parallel pathways beyond V1: “What” pathway (shown in red) for object detection and “Where” pathway (shown in green) for spatial information processing. The areas in the diagram are organized from lower (left) to higher levels of hierarchy (right). Lines between areas represent known anatomical connections. Arrows with heavy heads indicate connections from lower to higher levels of hierarchy. Open Arrows indicate connections from higher to lower levels of hierarchy. The anatomical layout of areas illustrated in the graph is shown in the center brain figure. Adapted with permission.^[20] Copyright 1995, American Association for the Advancement of Science.

2.2. ANNs

Inspired by biological neural systems, ANNs have been greatly developed for decades since McCulloch and Pitts first proposed a computational model for neural networks based on mathematics and algorithms in 1943.^[21] ANNs are now widely used for many AI applications including image analysis, speech recognition, and robotics. The neuron and synapse models used in current ANNs typically only consist of simple multiplication, addition, and activation operations, and they are greatly simplified compared to actual BioNNs. There are a large variety of network architectures, differing mostly in the connectivity structure, among which the most popular ANNs include CNNs (sometimes interchangeable with DNNs), recurrent neural networks (RNNs), and SNNs. Figure 3 illustrates different types of ANNs.

2.2.1. CNNs and DNNs

As shown in Figure 3a, CNN is a class of feedforward ANNs most commonly applied to feature extraction in vision analysis.^[19] CNNs apply a sequence of convolution operations and

pooling operations to the input. The convolution emulates the response of an individual neuron to visual stimuli in the brain. Conventionally, convolution operations are performed by power-hungry CPUs and/or GPUs, which severely hinders their integration into consumer electronics. However, with the help of emerging devices, such as RRAM, and PCM, massive computations of multiplication can be implemented in cross-point arrays, so that image recognitions with high efficiency can be realized.^[22–24]

CNNs structured with other computational layers that have a depth of more than three layers are often called DNNs.^[25] When such multi-layer systems are trained using the backpropagation algorithm, it is often referred to deep learning.^[26] Today, the typical numbers of network layers used in deep learning range from five to more than one thousand. These many layers usually include fully connected (FC) layers, convolutional (CONV) layers, nonlinearity layers (rectified linear unit, ReLU), pooling (POOL). In an FC layer, all outputs are connected to all inputs, thus the weighted sum of each output is computed from all the input information. In a CONV layer, however, the weighted sum for each output is computed using only a small patch of inputs and moreover, the same set of weights are shared for every

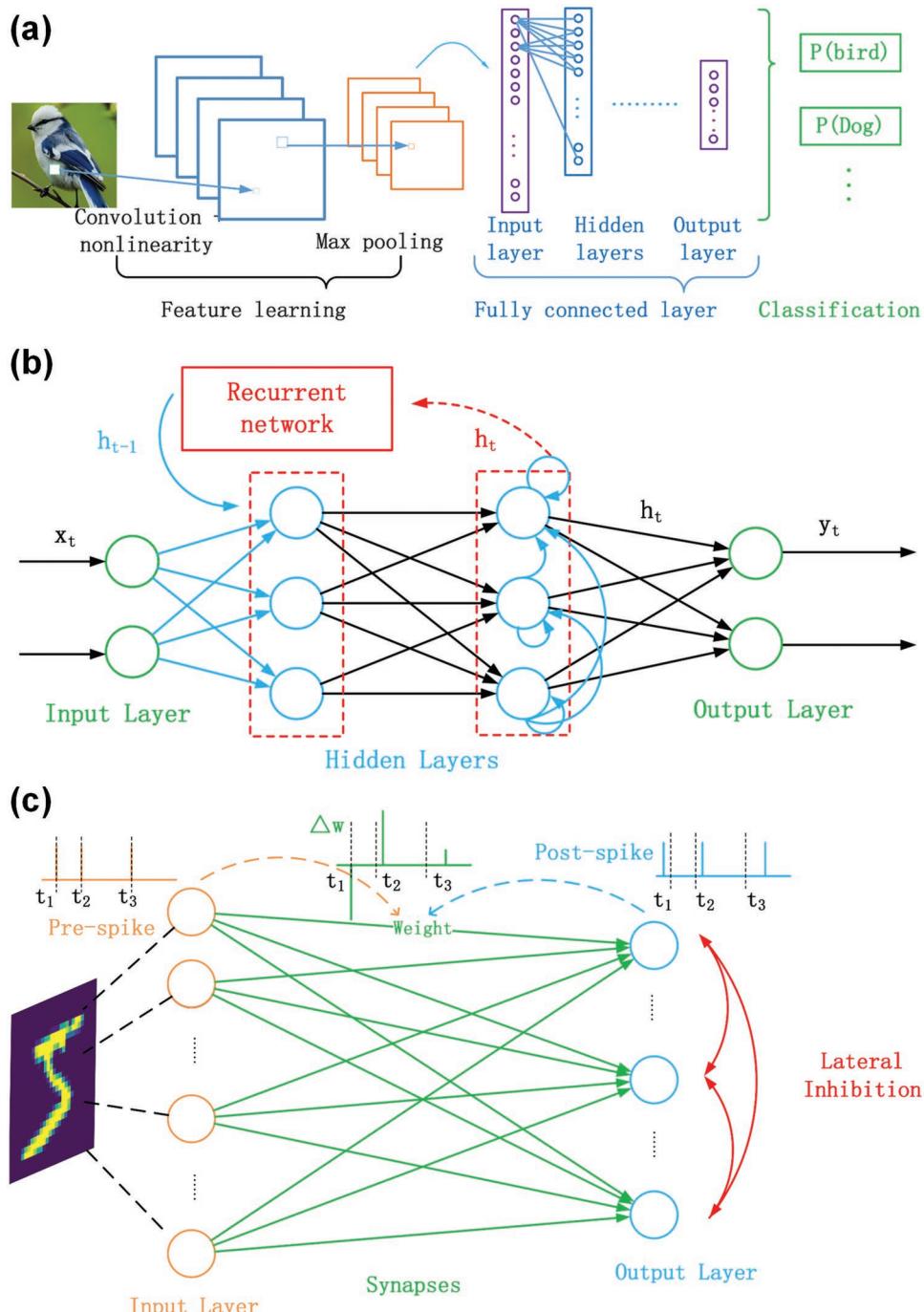


Figure 3. Illustration of different types of ANNs: a) CNN and DNN are the basic class of feedforward ANNs with multiple hidden layers between the input and output layers, which can be trained with the standard backpropagation algorithm. The nodes on each layer and the connections among them are conceptualized as artificial neurons and synaptic weights, respectively. b) RNN is a class of ANNs where the output of some neurons are fed back to themselves or to neurons in preceding layers, which enables the information to flow in both forward and backward directions and provides RNN with a dynamic memory. c) SNN is a more bio-plausible type of ANNs in which neurons communicate with spikes and information is encoded into the timing and frequency of spikes. The information processing in SNN is event based, and hence could be very energy efficient.

output, leading to a successively higher-level abstraction of the input data. A ReLU layer is typically applied after each CONV or FC layer to introduce nonlinearity, while a POOL layer is applied to each channel separately to reduce the dimensionality of a feature map. DNNs are capable of delivering the highest

accuracy of many AI tasks like pattern recognitions, some of which can even exceed human brain. However, the drawback of DNNs is that the computational cost correspondingly increases with the network complexity. The most common and intensive computation in DNN is vector-matrix multiplication, which can

be performed with emerging devices with orders of magnitude improvements in efficiency over traditional computing systems because of their simple cross-point array structure and direct weight update scheme.^[3,27–30]

In the hardware implementation of FC networks, e.g., multi-layer perceptrons (MLPs), the weight matrices could be directly mapped to the conductance matrices of memristive crossbar arrays. On the other hand, for CNNs, the weights of 2D CONV layers are usually 4D. Thus, 4D-to-2D mapping is needed to encode weights on memristive crossbars, which differs from that of MLPs. Depending on the kernel size, such mapping is not unique, bearing a tradeoff between performance boost and chip area expenditure.^[31,32] In addition, it is found that the programming noise and signal bounds show a more significant effect on the training accuracy of the memristive CNNs compared to that of MLPs, particularly the noise in the backward cycle and the signal bounds on the output layer.^[31] Techniques such as input vector rescaling and signal bound detection have been proposed to help the training of memristive CNNs.^[31] Compared to digital implementations, one advantage of memristive CNNs is the clear boost in terms of energy-area efficiency. The computational complexity for a convolution stride is usually $O(k_{\text{dim1}} \times k_{\text{dim2}} \times k_{\text{dim3}} \times k_{\text{num}})$ where k_{dim1} , k_{dim2} , k_{dim3} , and k_{num} are the height, width, depth, and population of the kernels, respectively. On the other hand, the complexity becomes $O(1)$ once the weights are physically mapped onto the memristor crossbar array because all multiplications and accumulations within the same kernel and across all kernels of the entire layer are performed in a single step.

2.2.2. RNNs

As shown in Figure 3b, RNN is a class of ANNs where neurons on the same level (or the same layer) have connections to each other or themselves; therefore information can flow in a loop. This can endow such networks with temporal depth, where connections between nodes follow a certain sequence.^[33] Neural networks in the biological brain almost always have recurrent connections. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process the sequences of inputs. RNNs based on long short-term memory (LSTM) and bidirectional (BRNN) architectures have demonstrated groundbreaking performance on various AI tasks from image captioning, language translation, to handwriting recognition.^[34,35] Take the LSTM unit as an example. Typically, an LSTM unit consists of an input node, an input gate, a cell, a forget gate, and an output gate. The output gate processes the current input information together with the stored data in the cell as well as the forgetting information to pass the result to the next layer. The recurrent processing occurs in the hidden layers on the basis of the backpropagation algorithm. Using RRAM as the building blocks for LSTM units in an RNN architecture, it has been shown that RRAM cross-point array can potentially provide a significant acceleration for RNNs as well as CNNs.^[24,36–38]

In the hardware implementation of RNNs using memristors, depending on the types of the recurrent layers (e.g., simple RNN, LSTM, or convolutional LSTM), the network

topologies need to be mapped to 2D conductance matrices of memristive crossbars, and the recurrent connections are usually implemented in digital circuits.^[37–39] Compared to memristive MLPs, it has been shown that device-to-device variation in the switching (weight update) asymmetry could yield more significant performance degradation for memristive RNNs.^[37] Also, the convergence of memristive LSTM RNNs is much more sensitive than memristive MLPs to the precision of input signals.^[37] Thus a higher input signal resolution or techniques like stochastic rounding scheme are favored.^[37] Compared to digital implementations, one advantage of memristive RNNs is the increased ratio between the number of operations on hardware forward/backward passes (e.g., memristive vector-matrix multiplications) and that on weight updates (e.g., memristive crossbar programming), which could further benefit the overall energy-area efficiency since the programming is relatively slow and power hungry.

2.2.3. SNNs

Beyond the ANNs described above, another class of ANNs, the SNNs have recently attracted more and more attention, as the next-generation network architecture. As shown in Figure 3c, neurons in SNNs communicate with spikes, which are unitary events that mark time, while CNNs or RNNs that transmit information with floating point numbers. In SNNs, rather akin to the brain, information is encoded into the timing and frequency of spikes. Neurons in the brain and in SNN are typically silent and transmit spikes only occasionally, therefore saving enormous amount of energy. This computational paradigm is often called event-based processing. A widely used computationally simple and yet powerful model for a spiking neuron is the leaky integrate-and-fire (LIF) model.^[40] The LIF model describes the potential of a neuron as the voltage across a capacitor that is connected in parallel with a leaky conductance path and charged by incoming input currents. It fires a spike (action potential, AP) when a threshold is reached and the resting potential is subsequently regained. Therefore, SNNs can more faithfully mimic BioNNs thus to inspire new computational paradigms. Several hardware demonstrations using RRAM and PCM type devices have shown the feasibility of using SNNs to complete classification tasks.^[41–47] However, to build an energy-efficient hardware for SNNs, the lack of efficient training algorithms (like backpropagation for DNNs) and well-established datasets is a critical impediment that limits current applications of SNNs.^[48,49]

3. Critical Functional Components of BioNNs and ANNs

From the top level of BioNNs down to the cell level, the basic units in the nervous system where signal processing happens are mainly neurons and synapses. They are also the mostly mimicked biological components using emerging neuromorphic devices so far. It is thus worthwhile to carefully review and compare the structures and mechanisms of biological and artificial neurons and synapses.

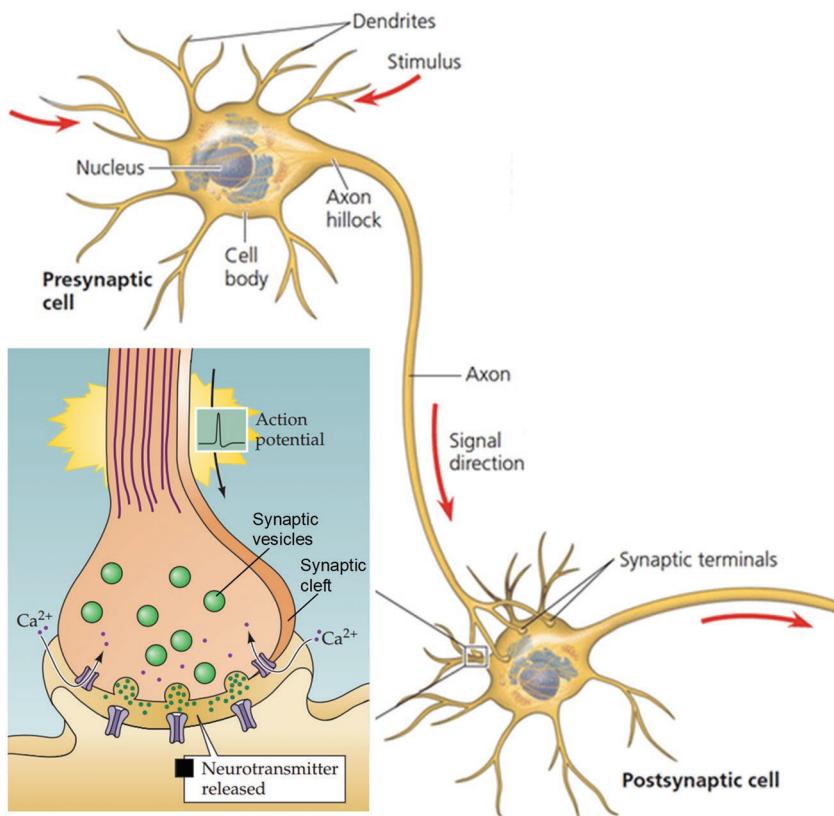


Figure 4. Schematics of a typical neuron and chemical synapse. A typical neuron is composed of a soma, an axon, and dendrites. Synapse connects the axon terminal of presynaptic neuron and the dendrite of postsynaptic neuron. The space at the chemical synapses is called synaptic cleft, and the critical feature is the presence of small, spherical, membrane-bonded organelles called synaptic vesicles within the presynaptic neuron. Calcium ion influx through ion channels causes the release of neurotransmitters from these vesicles, and when they reach the synaptic cleft, they bind with the receptors in the postsynaptic neuron, leading to the ion channels open or close (will be discussed in the next Section 3.2), causing the potential change in the postsynaptic neuron. Adapted with permission.^[17] Copyright 2017, Oxford University Press.

3.1. Fundamental Biological Structures of BioNNs: Neurons and Synapses

In biological nervous systems, a typical neuron, as illustrated in Figure 4, consists of several structural and functional parts: a soma, an axon, and dendrites. The junction that connects an axon terminal of one (presynaptic) neuron and a dendrite of another (postsynaptic) neuron is the synapse. Each of these parts plays distinct roles in the generation and communication of signals between cells.^[17,18] Since biological neurons are delimited by lipid bilayer membranes which act as capacitors and are also punctuated by small holes on the membrane (ion channels) acting as reconfigurable resistors. A membrane potential gradient is also set up by the difference in ion concentrations across the membrane (which is maintained by Na^+/K^+ pumps that consume energy in the form of adenosine triphosphate (ATP), as shown in Figure 5b) and acts as a battery. Typically, the membrane resting potential is around -70 mV with the inside of the cell being at a lower potential. The most basic electric operation of a biological neuron can be understood as a RC circuit with a battery. The membrane potential (voltage) can be changed when currents are injected mainly through the ion channels (Section 3.2). The capacitance and electric potential are normally impossible to change over a short time scale, but the resistance can be changed by opening and closing of ion channels, which give rise to current flows and underlie diverse electric

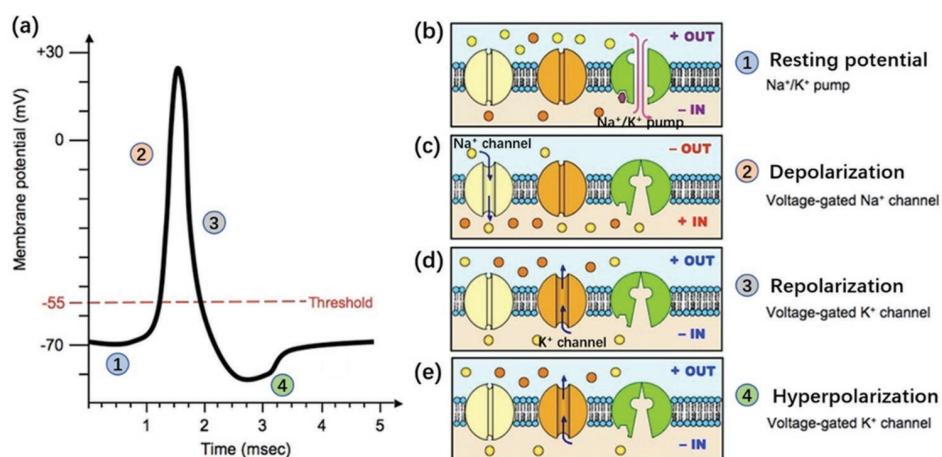


Figure 5. Schematics of an action potential (AP). a) Illustration of the change of the membrane potential during an AP. Diagrams b–e) illustrate the ion channels that act in each phase: b) In the resting state, Na^+/K^+ pumps transport Na^+ to outside of the membrane and K^+ to inside of the membrane. The ion concentration gradient drives K^+ to flow outside of the membrane, which results in approximately -70 mV membrane potential (the potential inside minus the potential outside of the membrane). c) When the membrane potential increases up to a threshold, the voltage-gated Na^+ channels are activated and Na^+ flows from outside of the membrane to inside, thus the membrane potential goes up rapidly. This process is called depolarization. d) With the increase of membrane potential, voltage-gated K^+ channels are activated and the voltage goes down again, which is called repolarization. e) The voltage usually “undershoot” below resting potentials due to the outflow of K^+ . This is called hyperpolarization. Adapted with permission.^[55] Copyright 2019, BioNinja.

operations in the neurons. We review the main structural and functions of each part of the neuron below.

3.1.1. Soma

The soma, or the cell body of a neuron, contains the cell nucleus and cellular organelles. From the soma grows two types of cytoplasmic protrusions, namely the axon and the dendrites. A specialized region of the neuron, the axon hillock, is where the axon originates. Within the axon hillock, an AP (Figure 5a) is generated by integrating the synaptic inputs on dendrites that cause sufficient excitation over a certain threshold, and then pass to the axon (output) as a traveling impulse. The generation of APs is generally termed as neuronal “firing.”

3.1.2. Axon

The axon is considered as the output element of the neuron according to the description of “dynamic polarization” by Cajal.^[50] APs generated at the axon hillock propagate along the axon to the nerve terminals, where synapses are formed on the postsynaptic dendrites (and sometimes soma). The most important function of AP is probably to ensure reliable transmission of signals across a long distance while it might also have important computational functions. Some recent research has reported that despite the faithful conduction in the axons, the complex time and voltage dependences of APs can lead to activity-dependent changes in spike shapes and the resting potential, affecting the temporal fidelity of spike conduction.^[51] With the exception of SNNs, “axons” in most ANNs typically transmit a real number firing rate rather than all-or-none spike events.

3.1.3. Dendrites

Dendrites are the input elements of a neuron that branch off from the soma. Dendrites play a critical role in filtering and integrating the synaptic inputs to determine whether to fire a signal or not. Traditionally the dendrite is considered to be a passive element and does not amplify the input signals. Different from the single axon derived from the soma, there are many dendrites and thus, the morphology such as branch density and grouping patterns are highly relevant to the function of the neuron. Recent studies however have emphasized the generation of active electric events in the dendrites akin to those happening at the axon hillock.^[52] Therefore, dendrites themselves may be capable of complicated processing of incoming stimuli. Current ANN model only describes the passive integration part of the dendrite function in a very simplified manner, such as summations.

3.1.4. Synapses

Synapses are the connecting structure where the axon terminal of the presynaptic neuron transmits signals to the postsynaptic neuron. The branches of a single axon may form synapses with

as many as thousands of other neurons. There are two fundamentally different types of synapses: electrical and chemical synapses. Electrical synapses pass ionic current to induce the voltage changes in the postsynaptic cell directly. While for the chemical synapse, which is the majority of synapses in the human brain, electric activity in the presynaptic cell is converted to chemicals (neurotransmitters) and chemical signals are converted back to electrical activity in the postsynaptic cell. This arrangement isolates the two cells electrically and can give rise to many complex signal transformations at the synapse in both amplitude and time domain. Depending on the direction (inward or outward) of the synaptic current, the membrane potential is increased (depolarized) or decreased (hyperpolarized), the current and potential are termed excitatory/inhibitory postsynaptic current and potential (EPSC/IPSC and EPSP/ IPSP), respectively. In this way, the connection between the two communicating neurons is modulated, i.e., the weight of the synapse is changed via synaptic plasticity (will be discussed in Section 4). The 3D structures of two common types of chemical synapses, excitatory and inhibitory synapses, have been revealed by the latest cryo electron microscopy technique.^[53] In biological systems, it is generally believed that synapses are either excitatory or inhibitory, and furthermore each neuron only sends out either excitatory or inhibitory synapses but can receive neurotransmitters from both types of synapses. This is termed as Dale’s law. In ANNs, the synapse is normally simply modeled as a number (synaptic weight) to be multiplied with and can be of both positive and negative signs. Change in synaptic properties, most importantly synaptic weight is the main manner that learning is realized in both ANNs and BioNNs.

3.2. Fundamental Mechanism of Activity in Biological Neurons and Synapses: Gated Ion Channels

In the biological neural system, ion channels are the molecular basis for all of electrical activities in initiation, processing, and transmission of information. There are two main classes of ion channels in the neurons: voltage-gated ion channels (VGICs) and ligand-gated ion channels (LGICs). Neurons receive thousands of synaptic inputs through the inhibitory or excitatory ligand-gated channels located on postsynaptic membranes, transform chemical neurotransmitters to electivity signals, and change the synaptic potentials. Then the changed membrane potentials activate the VGICs to generate APs (Figure 5) at the axon initial segment (AIS, adjacent to the axon hillock). Next, the APs are conveyed along axons and trigger neurotransmitter to release at the axonal terminals to activate other neurons. Therefore, it is important to understand the crucial role of the gating mechanisms in VGICs and LGICs, so as to further comprehend its function in communication with electrical signals.^[54]

3.2.1. VGICs

VGICs are voltage-sensing ion channels where the conduction of ions is controlled by the changing of membrane potential (Figure 6a). Therefore, VGICs are critical for the transmission of electrical signals in exciting cells. VGICs are mainly involved

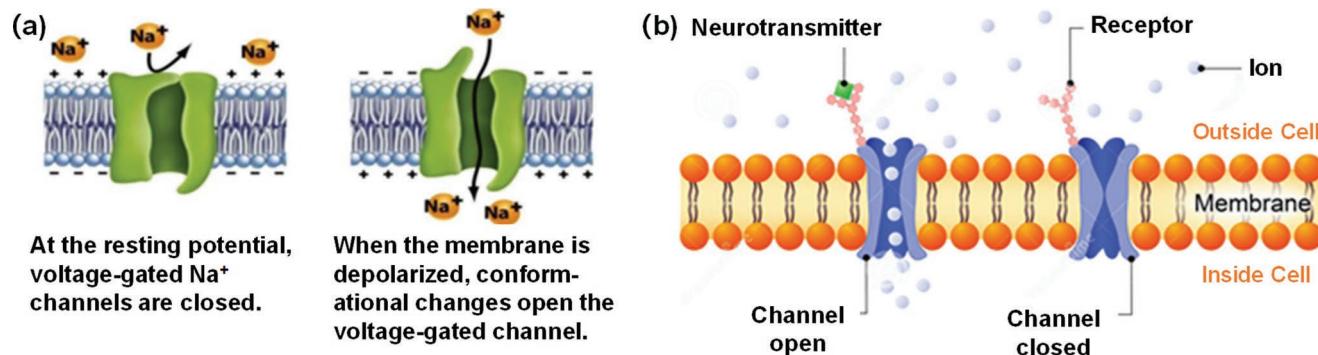


Figure 6. Schematics of typical gated ion channels. a) Voltage-gated ion channels (VGICs) Adapted with permission.^[70] Copyright 2005, Pearson Prentice Hall. b) Ligand-gated ion channel (LGICs). Adapted with permission.^[71] Copyright 2019, Designua.

in the initiation and regulation of AP, the transmembrane transport of excitation and inhibition signals. They are the largest and most extensively studied superfamily of ion channels, which include the sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), chloride (Cl^-), and proton channel family members.^[56,57]

Voltage-gated sodium channels (Navs) are transmembrane proteins which possess voltage sensitivity and transduction sodium ions. The potential gradient set up by the ion concentration gradient (Section 3.1) is such that opening of Nav leads to depolarization (increased potential inside the membrane due to the inflow of Na^+) of neuronal membrane potentials. In neurons, the activation of Navs accumulated at high density in the AIS. They open when the membrane potential is sufficiently positive, thus mediating a positive-feedback loop leading to rapid depolarization and the all-or-none event called the AP (Figure 5c).^[58] The membrane potential at which they start to open to a sufficient degree determines the threshold of AP generation. Voltage-gated potassium channels (KVs) mediate outward K^+ currents and lead to the repolarization and hyperpolarization of membrane potentials (decreased potential inside the membrane due to the outflow of K^+ , Figures. 5d–e). They are activated by depolarization of the membrane potential completing a negative feedback loop and regulate the intrinsic excitability of neurons.^[59] Together, these two types of channels, along with ATP-driven Na^+/K^+ pump, regulate the basic electric integration and AP generation processes.

In addition, CaV channel is a channel especially important in synaptic functions. Ca^{2+} entry to a presynaptic terminal is mainly through voltage-gated Ca^{2+} channels (CaV) and can trigger neurotransmitter release from the active zone.^[60–62] Finally, chloride (Cl^-) channels regulate the membrane excitability, repolarize APs, stabilize the resting voltage, and regulate the IPSC.^[57]

3.2.2. LGICs

LGICs are mainly distributed in postsynaptic membrane, and they sense the presence of a particular type of messenger (Figure 6b). They open when they are activated by the presence of excitatory or inhibitory neurotransmitters in neuronal networks, such as glutamate or GABA. The AMPA- and NMDA-type ionotropic glutamate receptors (iGluRs) are the most abundant excitatory neurotransmitter receptors, responsible

for sensing glutamate to mediate the vast majority of excitatory transmission in the brain. GABA receptor receives GABA to mediate inhibitory neurotransmission in the brain and conduct Cl^- ions.^[57,63–65] AMPA receptor is important for normal transmission of positive (excitatory) signals. The number and electric properties of AMPA receptors can be modulated in many different ways to change the synaptic properties especially weights of that particular synapse.^[66,67] N-methyl-D-aspartate receptors (NMDARs) are special in that they are both ligand-gated and voltage-gated and they are especially important in synaptic plasticity and learning. They only open when the presynaptic neuron is active and release glutamate and the postsynaptic neuron is sufficiently active and the membrane potential is sufficiently depolarized. NMDARs therefore serve as prototypical coincidence detectors in the central nervous system and provide a molecular mechanism for the Hebbian rule of synaptic learning (Section 5.1).^[68,69]

3.3. Demonstrations of Artificial Neurons and Synapses Using Emerging Devices

In the past decades, tremendous efforts have been made to implement artificial neurons and artificial synapses using a variety of emerging materials and devices, as many excellent review articles have previously summarized.^[72–74] Here we pick some representative implementations as listed in Table 1.

3.3.1. Artificial Neurons

To emulate the integration and firing dynamics of the biological neuron (Figure 5a), there is a growing research interest to exploit the physical natures of emerging devices. PCM, RRAM, and CBRAM are among the most popular options. For example, a NbO_2 Mott memristor was first used to demonstrate the biomimetic properties, including all-or-none spiking of an AP, a bifurcation threshold to a continuous spiking regime, signal gain and a refractory period.^[82] Similarly, a PCM-based stochastic neuron has been proposed in which the neuron membrane potential is presented by the phase configuration of the chalcogenide film (Figure 7a).^[75] The temporal integration of postsynaptic potentials can be achieved on a nanosecond

Table 1. Comparison of artificial neurons and synapses using emerging devices.

Devices	Schematic ^{a)}	Mechanism	Materials	Pros	Cons	Research status	Artificial neuron	Artificial synapse	Switching energy ^{b)}	On/off ratio ^{b)}	# of States ^{b)}
PCM		Amorphous-crystalline phase change; Ovonic threshold switching	Ge ₂ Sb ₂ Te ₅ (GST)	CMOS compatible	Asymmetric switching Conductance drift	165k cells network to realize MNIST classification	[75–77]	[78–81]	<10 ³	≈20	
Mott Memristor		Metal-to-insulator transition	Nb ₂ O ₅ VO ₂	CMOS compatible	Low resistance in the OFF state	Single (few) device demo	[82]	[83]	>1 pJ	>500	≈2
RRAM		Oxygen filament growth and rupture (filamentary) Defects/ions migration (interface)	Metal oxides 2D materials Perovskite	CMOS compatible High density Low power Gradual G change	Variation Stochasticity	8k array to realize pattern recognition	[84,85]	[30,45,86–95]	>20 fJ	10–100	64–500
CBRAM		Metal filament growth and rupture	Cu/Ag in oxide or chalcogenide	Scalable Low power	Variation Stochasticity	Single device and < 10 × 10 array lab demo	[41,96,97]	[98–101]	>100 fJ	10–240	≈100
FeFET/FTJ		Polarization switching	Doped HfO _x Perovskite	CMOS compatible	Hard to realize gradual G change and multilevel (relying on number of domains).	Single device demo	[102]	[94,103]	>100 fJ	45–300	32–100
MRAM/magnetic domain wall device		Magnetization switching	CoFeB MgO	High endurance Fast speed	Limited number of states Small on/off ratio	Single device demo	[104–107]	[108–111]	>10 fJ	2–3	≈2
EcRAM		Electrochemically driven ion intercalation	Metal oxides PEDOT:PSS Graphene	Good symmetry and linearity Gradual G change Multi-states	Write speed Complex unit cell CMOS compatibility	Single device and 3 × 3 array demo	N/A	[29,112–119]	>10 fJ	2–40	50–1000
FETs		Floating gate (FG)/charge trapping based threshold change	Silicon CNT Polymers 2D materials	Technologically more mature	Asymmetric switching Usually high voltage (FG) Speed/endurance	>100k cells for Flash-based demo	N/A	[120–131]	>1 fJ	10–100	20–100

^{a)}The first six inset figures are adapted with permission.^[132] Copyright 2018, Springer Nature; ^{b)}The numbers represent the typical values reported in literature for neuromorphic computing.

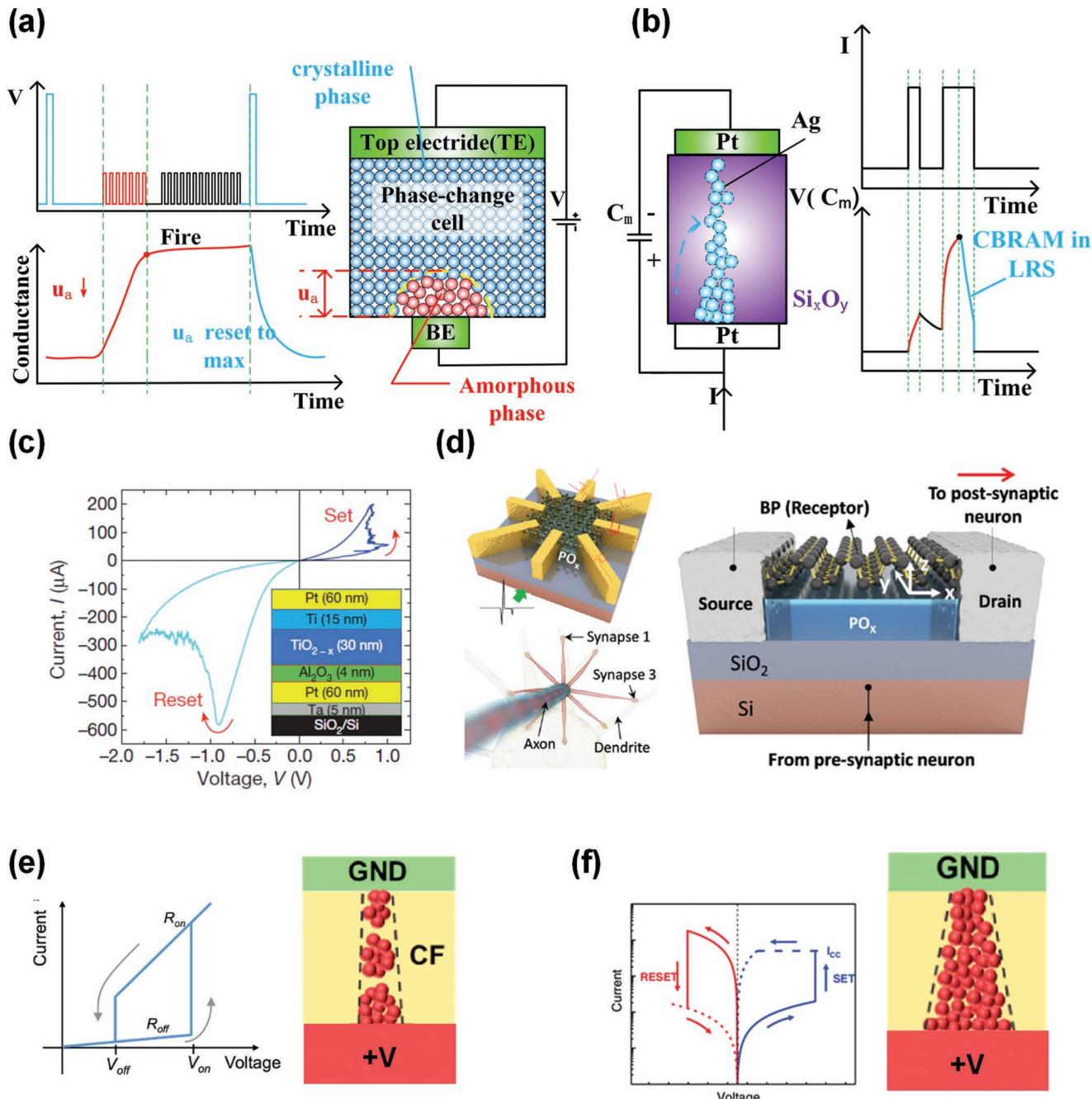


Figure 7. Selected demonstrations and mechanisms of artificial neurons and synapses. a) PCM as artificial neuron. Adapted with permission.^[75] Copyright 2016, Springer Nature. b) CBRAM as artificial neuron. Adapted with permission.^[41] Copyright 2018, Springer Nature. c) RRAM as artificial synapse. Adapted with permission.^[30] Copyright 2015, Springer Nature. d) Black phosphorus (BP) FET as artificial synapse. Adapted with permission.^[123] Copyright 2016, Wiley. e) Threshold switching in artificial neuron. f) Memory switching in artificial synapse.

timescale via the reversible amorphous-to-crystal phase transitions. Moreover, the inherent stochasticity of PCM leads to stochastic firing response of individual neurons in a population as well as the interneuron variability, which are key for population-based neuronal computation.

In the meanwhile, oxygen filament-based RRAM memristors have also been developed to mimic the neuron functions.^[84,85] A SiO_x RRAM device can generate controlled voltage transients,

which resemble spike-like responses.^[85] Thus, it is capable of integrating input current pulses over time to produce threshold voltage transients. Metal filament-based CBRAM-like devices (diffusive memristors) have also been proved feasible to achieve the integrate-and-fire function.^[41,96,97] Importantly, these devices relax back to their highly resistive off state spontaneously (without the need of RESET operation) after firing, closely resembling the repolarization process of biological

neurons. Wang et al. integrated the Ag diffusive memristors onto the gate of a transistor, which then becomes the active front end of an “axon” for a stochastic LIF neuron emulator (Figure 7b).^[41] Such neurotransistor is capable of propagating signals as biological neurons, which is important for efficiently implementing multiple layer neural networks. Input signals are integrated by the memristor neuron that mimics dendritic spatial and temporal summation.^[41,96] In addition, both ferroelectric field-effect transistors (FeFETs) relying on ferroelectric polarization switching and MRAM devices relying on magnetization switching have also been employed to emulate biological neuron behaviors.^[102,104–107]

As we can see, in most of these researches, the functions of soma, axon, and dendrites are combined into a simple integrate-and-fire neuron and realized at the device level. However, each individual part of the neuron is worthy of studying and mimicking, so as to achieve more complex and modulative neuronal characteristics.

3.3.2. Artificial Synapses

In the meanwhile, artificial synapses have also been even more extensively studied using emerging devices. From electronics point of view, the key feature of biological synapses is tunable weight, which can be conveniently translated to the conductance (or resistance) change in electronic devices. Desirable properties and switching dynamics of a variety of neuromorphic devices for neuromorphic applications have been discussed by several papers.^[74,133,134] Depending on the applications and algorithms, different switching characteristics may be emphasized, such as switching symmetry, linearity, number of conductance states.

In PCM synapses, only the SET process (amorphous to crystalline transition) can be made incremental, with repetitive pulses slowly crystallizing a high-resistance amorphous plug within the device, while the quenched RESET process (crystalline to amorphous transition) is usually abrupt.^[78,79] In order to achieve gradual conductance switching, several approaches have been proposed, e.g., using two PCM per synapse to realize potentiation and depression respectively, one PCM per synapse with carefully tuned programming pulses.^[79,80] Also, gradual RESET in PCM has been recently demonstrated by optimizing the device structure (using a narrow heater as the bottom electrode) and pulse scheme.^[81] Another issue with PCM is the conductance drift, or relaxation, which is frequently observed especially for the amorphous state (i.e., off-state).^[135] This phenomenon leads to a slow but steady increase in the device resistance over time, which follows the equation $R(t) = R_0(t/t_0)^\nu$, where R_0 represents the initial programmed resistance and t is the time elapsed since programming.^[136] The drift coefficient ν and time constant t_0 are material system dependent. A typical ν is ≈ 0.1 , while t_0 shows a large variation.^[136] In PCM fabrication, the drift coefficient can be largely reduced using a metallic surfactant liner as a resistance drift stabilizer.^[137] In comparison, no universal trend has been observed on the off-state retention of RRAM,^[138,139] while the ON-state retention usually follows the Arrhenius equation with an activation energy (e.g., barrier of ionic hopping) determined by the corresponding material systems.

Compared with PCM, oxide RRAM devices are more promising to achieve incremental switching characteristics.^[86–89] In addition, RRAM has a simple device structure that uses mature metal oxides such as HfO_x , TiO_x , WO_x , TaO_x and mixtures or stacks of those films (Figure 7c), almost all of which are CMOS compatible.^[30,45,87–93] Similarly, in CBRAM synapses, Cu or Ag (or other electrochemically active metal) filaments or nanoclusters are formed or dissolved to tune the conductance so as to imitate the synaptic weight change.^[98–100] For both RRAM and CBRAM, the filament formation and rupture usually cannot be very precisely controlled, leading to a relatively large variation and stochasticity.

More recently, novel 2D layered materials with unique merits like flexibility, transparency, anisotropy have been studied to explore synaptic behaviors as well (Figure 7d).^[123,140–142] Also, FETs based on silicon, organic materials, perovskite or carbon nanotube (CNT) have been reported as synaptic elements based on either charge trapping/detrapping mechanism or floating-gate (FG) memory structure, some of which have a structure of multi-gate FET.^[120–128] Three-terminal FeFET and two-terminal ferroelectric tunnel junction (FTJ) relying on ferroelectric polarization switching have been explored as artificial synapses, where the pulsing scheme needs to be carefully designed to improve the switching symmetry and linearity.^[94,103] Similarly, two-terminal spin-transfer torque MRAM (STT-MRAM) based on magnetization switching and multi-terminal magnetic devices based on domain wall motion have also been studied to implement artificial synapses.^[108–111] In addition, electrochemical random-access memory (ECRAM) based on ion intercalation has recently been reported as a promising synaptic cell, showing multi-states and incremental switching with near-ideal switching symmetry and linearity.^[29,112–119] The electrochemically driven ion intercalation process is more controllable than filament-related ion movements in RRAM; therefore, ECRAM also exhibits a much smaller stochasticity. In addition, by borrowing the battery concept, those devices successfully decouple the read and write operations and thus realize low programming energy and long retention time simultaneously.^[143] However, as a three-terminal device, ECRAM has a more complex unit cell design for array implementation, and it also faces other challenges in CMOS compatibility and speed (limited by write-induced transients associated with the ion movements across the electrolyte/channel interface),^[144] etc.

Overall, the most critical characteristic of artificial synapses is to exhibit incremental and reliable conductance changes. For practical conductance update, mainly three approaches have been developed in literature. One is to use two precisely designed pulses applied on the two terminals of the device so that the conductance is modulated only when they overlap in time (which can enable parallel update in an array).^[3,30] The second approach is to use pulse trains continuously program the target device until the conductance change reaches the desired value before programming the next one (sequential update).^[145] The third approach is to build relaxation dynamic into the artificial synapse so that it can sense the timing difference between the pre- and post-spikes and modulate its synaptic weight accordingly, as what a biological synapse does.^[99]

3.4. Fundamental Mechanism of Artificial Neurons and Synapses: Resistive Switching

Different types of physical mechanisms have been used in previous neuromorphic device demonstrations, among which resistive switching is the most widely used. The neuron-like integrate-and-fire behaviors are usually realized by the threshold switching characteristics of memristors (Figure 7e), while the synaptic weight changes in artificial synapses are modulated by means of analog (incremental) nonvolatile switching onto different levels of resistance (or conductance) (Figure 7f). We next review both mechanisms in neuromorphic devices below.

3.4.1. Threshold Switching in Artificial Neurons

When a memristor is employed as a neuron, the key behavior is to generate a current or voltage spike (in some cases may be a negative spike) when the threshold is reached after integrating a certain pulse train. In PCM neurons, the as-made device consists of a nanoscale volume of phase-change material initially in the crystalline phase.^[75] Voltage pulses with a sufficiently high amplitude applied to the devices (also referred as RESET) flow a Joule heating current through the switching layer and melt a substantial portion of the phase-change material. If the pulse is cut off abruptly, the molten part will rapidly quench into the amorphous phase following a glass transition. By changing the voltage amplitude, the effective thickness of this amorphous region can be tuned to change the device conductance. Once the conductance exceeds the preset threshold value, the neuron fires and the PCM is then reset to its initial state, which effectively realizes the nonlinear integrate-and-fire dynamics. It should be noted that such RESET process is controlled by peripheral instrumentation rather than an intrinsic property of PCM itself.^[75] In addition to phase transition, chalcogenide-based Ovonic threshold switching has also used to implement the integrate-and-fire function by pairing them with capacitors (resistors) that simulate neural membranes (axial resistance).^[76,77]

Threshold switching also occurs in RRAM devices, especially in the low operation voltage or current regime, and hence it can be exploited as volatile “turn-on” behavior, resembling the function of neurons.^[84,85] Oxygen vacancies are generated within the oxide switching layer under an external electric field (also referred as SET) and drift towards counter electrode to form a conductive filament. Low voltage or current compliance allows the formed filament to remain thin and unstable so that it ruptures spontaneously upon removal of the electrical bias. This transient conductance spiking is similar to that often seen in biological neurons.

When it comes to the CBRAM-like neurons, diffusive dynamics of metal filaments or nanoclusters are reported to contribute to the threshold switching, leading to threshold firing.^[99,146] Under electrical bias, Ag or Cu atoms are oxidized to ions, which migrate along the electric field and then are reduced to neutral atoms in the electrolyte, forming a metallic filament(s) and suddenly increase the conductance of the device. But different from those in RRAMs, these metallic filaments tend to spontaneously rupture and form disconnected

islands to minimize their interfacial energy as the external bias is removed. Therefore, the conductance falls at a rather rapid speed, realizing a spiking firing like a neuron.

3.4.2. Memory Switching in Artificial Synapses

When a memristor is used as a synapse, it uses the continuous conductance change to implement the weight tuning process. There are several typical nonvolatile resistive switching mechanisms: phase change in the switching layers (PCM), filament modulation (filamentary RRAM and CBRAM), and interface valence change (nonfilamentary RRAM), etc.

In PCM devices, as mentioned before, they exhibit a unique switching behavior between amorphous (low conductivity) and crystalline (high conductivity) states with the application of electrical pulses that are large enough to generate the heat required for phase transition. Different levels of stable conductance can be obtained through carefully adjusting the pulse amplitude or width to realize gradual SET programing.^[78,147]

Among these candidates for artificial synapses, RRAM devices appear to be the most attractive option because of the low energy consumption down to sub-pJ per synaptic event, the extreme scalability as crossbar RRAM can have an area of down to $2 \times 2 \text{ nm}^2$, the potential for high-density integration with selector and 3D stacking,^[148,149] as well as excellent CMOS fabrication compatibility.^[150–152] These merits make it feasible to build large-scale RRAM synaptic arrays. It should be pointed out that, for future manufacturing using advanced technology nodes (e.g., 22 nm node and beyond),^[153] the operation voltage of RRAM, which varies significantly with different material systems, need to be further reduced (e.g., sub-1 V) while retaining the desired analog switching properties for neuromorphic computing.^[154,155] In comparison, STT-MRAM has demonstrated both low operation voltage and high endurance as well as readiness for high-volume production in major foundries (e.g., Samsung, Intel and TSMC);^[156,157] however, it also suffers from small on/off ratio (typically 2–3) and limited number of resistance states (typically ≈ 2). Therefore, in practice the choice of NVMs for implementing neuromorphic computing depends on the specific applications.

For filamentary RRAM and CBRAM synapses, the control of conducting filament size (or its number) is likely the main approach to obtain analog conductance modulation.^[158] The amplitude, duration, and numbers of electrical pulses all contribute to the change of filament size and the conductance of the device. Besides, there are also nonfilamentary RRAM devices that exhibit synaptic behaviors.^[159,160] In those devices, defect or ion migration takes place over the entire device area uniformly, typically at an interface between two materials such as an oxide and a metallic electrode, modulating the energy barrier at the interface, leading to observable conductance change. Moreover, ionic motion can change the conductivity of the local oxide, narrowing the tunneling gap and enabling gradual switching that is highly suitable for implementing analog synapses.

Tremendous progress has been made in neuromorphic devices; however, it should be pointed out that it still remains extremely challenging to emulate the entire neuron or synapse faithfully using current electronic devices. This is because

even a single biological neuron or synapse contains a number of complicated sub-components, such as different types of ion channels, and all of them collaboratively contribute to the neuromorphic functions. In an electronic device, usually only a single type of switching behavior dominates due to the limited complexity of the artificial system, which unfortunately has already been very challenging to be completely understood and modeled so far. Although there have been some attempts along this direction, the complexity and cooperativity of the natural intelligent systems are still far out of reach.^[96,161,162] More fundamental studies at the materials and devices levels are needed to achieve more capable artificial intelligent components and systems.

4. Plasticity

In this part we will discuss the key concept in neuroscience and brain-inspired computing: plasticity. Neural plasticity in the biological system refers to the ability of neurons or synapses to change their properties, often dependent on neural activity, and is believed to underpin the learning and memory functions of the brain.^[163,164] In artificial computing, plasticity is usually presented by changes in device conductance. The biological behavior, origins and functions, as well as the electronic implementations (if any) of each type of plasticity will be reviewed as follows.^[165–169]

4.1. Synaptic and Nonsynaptic Plasticity

Synaptic plasticity is the ability of synaptic properties to change, most commonly the strengthening or weakening of synaptic strength. It is the most well studied form of neural plasticity, and is considered the most prominent form of neural plasticity underlying biological learning and memory.^[170,171] Synaptic plasticity can be categorized into short-term and long-term according to the duration of its effect on the synaptic strength. Besides the change of synaptic weights, plasticity of wiring patterns is also possible in the brain. Nonsynaptic plasticity also exists, and the most common form is intrinsic plasticity, which regulates the voltage-dependent ion-channels and modifies the neuronal excitability.

4.1.1. Short-Term Synaptic Plasticity (STSP)

STSP, also referred to as short-term synaptic dynamics, lasts for several minutes or shorter.^[17,18] The effects of STSP on synaptic efficacy (the capacity of a presynaptic input to influence postsynaptic output) can be either enhancement or depression. Synaptic enhancement can be further differentiated as paired pulse facilitation (PPF), and post-tetanic potentiation (PTP) according to the time scale.^[172] PPF occurs when two APs arrive at the synapse in a short interval within tens of milliseconds. It is most likely due to the increased level of Ca^{2+} in presynaptic cell after an AP. High Ca^{2+} concentration enables more neurotransmitters to be released and results in a larger excitatory postsynaptic potential. PTP acts on a longer time scale of tens of seconds to minutes. It occurs after a high-frequency

burst of presynaptic APs. It is associated with Ca^{2+} dependent activation of presynaptic protein kinases.^[173] Short-term synaptic depression occurs in some synapses after rapidly repeated presynaptic APs. It is mostly due to the depletion of releasable pools of neurotransmitter containing vesicles. These forms of STSP can occur in a time series (Figure 8a): repeated presynaptic APs induce paired pulse facilitation in a short period, then the depletion of vesicles allows synaptic depression to dominate; After a period of recovery, new arrival of presynaptic AP may cause enhanced neurotransmission due to PTP.^[17,18] Overall, STSP enables different synapses to act as low-pass, high-pass, or band-pass filters of information transmission.^[174] It is also hypothesized to play an important role in sound-source localization and working memory.^[166,168,175]

In an electronic synapse, when a proper electrical stimulus applies to one electrode of the device, an abrupt increase of current through the device can be sensed, which emulates the excitatory spike onto the presynaptic membrane that induces an increase in the postsynaptic current level, i.e., EPSC behavior. After that, the current would decay rapidly with time. This transient current increase and recovery are likely dominated by the internal ion movements (RRAM, CBRAM), which are typically volatile, and to some extents resemble the neurotransmitter release in the biological synapse. Further, when two excitatory presynaptic spikes are applied successively, the second pulse generates a larger EPSC than the first one, it is termed PPF, that is one type of biological STSP.^[92,99,118,176–179] By varying the time interval between the paired pulses, the amplitude of EPSC caused by the second pulse can be tuned; generally, a longer interval results in a reduced EPSC amplitude enhancement. On the contrary is IPSC which is the current decrease when facing a presynaptic spike to the electronic device.^[126,142] The current response in the device is actually a rise in the amplitude towards an opposite direction, e. g., a sudden negative current triggered by a positive presynaptic spike (gate bias).^[126,142]

4.1.2. Long-Term Synaptic Plasticity (LTSP)

Different from STSP, LTSP can last for several hours or longer.^[180] Persistent increase in synaptic strength is called long-term potentiation (LTP). On the contrary, persistent decrease in synaptic strength is known as long-term depression (LTD). The cellular and molecular mechanisms of LTP or LTD vary at different synapses. Here we only introduce the canonical phenomena and signal pathways underlying LTP and LTD.

LTP and LTD were first discovered in hippocampus. After a high-frequency electrical stimulation for seconds, synaptic strength can be increased for days to weeks (LTP).^[181] Meanwhile, after a low-frequency (≈ 1 Hz) stimulation for long periods (10–15 min), synaptic strength undergoes long-lasting decrease (i.e., LTD) (Figure 8b).^[180] With physiological stimulation paradigms involving correlated activities of pre- and postsynaptic neurons, it was found that the extent and the direction of synaptic modification depend on the relative timing of pre- and postsynaptic activation (spike-timing-dependent-plasticity, or STDP):^[182,183] timing-dependent LTP (t-LTP) was induced when the postsynaptic spike follows presynaptic activation within about 20 ms, whereas t-LTD was induced when

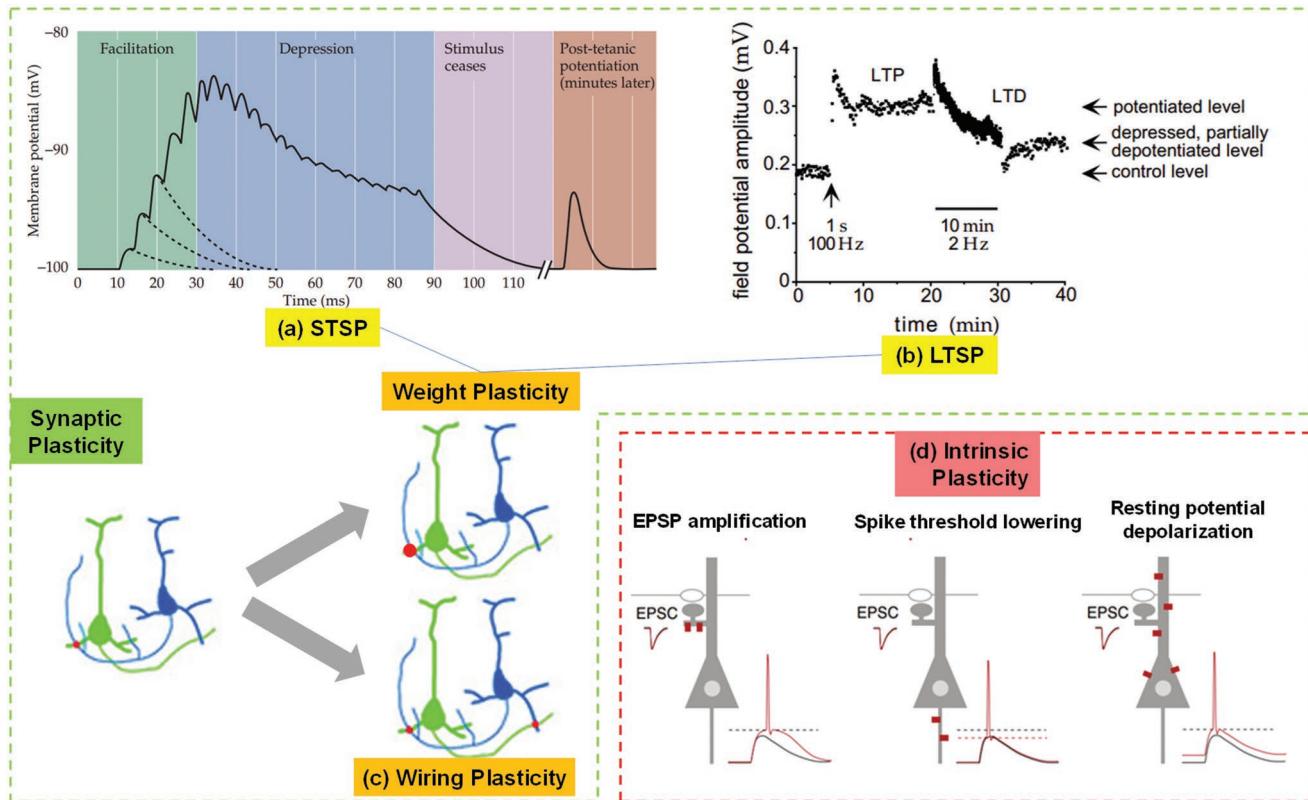


Figure 8. Illustrations of different types of neuroplasticity. a) Time course of STSP. Adapted with permission.^[17] Copyright 2017, Sinauer Associates. b) Schematic of LTP and LTD. High-frequency stimulation for a short period induces LTP, while low-frequency stimulation for a long period induces LTD. Adapted with permission.^[20] Copyright 2005, The MIT Press. c) Schematic of wiring plasticity. Adapted with permission.^[19] Copyright 2004, Springer Nature. d) Schematic of intrinsic plasticity. Three mechanisms are illustrated here: i) EPSP amplification: Excitatory postsynaptic potential can be amplified by the regulation of dendritic ion channels. It is similar to increased synaptic weights. ii) Spike threshold lowering: the voltage threshold at which an action potential is triggered can be lowered by the regulation of voltage-gated channels on the initial segment of axons, thus it is easier for the neuron to fire. iii) Resting potential depolarization: the membrane resting potential can be raised by the regulation of voltage-gated channels, which also makes the spike threshold easier to reach. Adapted with permission.^[19] Copyright 2019, Elsevier.

the temporal order was reversed.^[182] Besides the hippocampus, LTP, LTD and STDP were observed in a variety of brain regions, including cortex, cerebellum, etc.^[163,164,184]

A special ligand-gated ion channel, the NMDAR on post-synaptic membrane, plays an important role in hippocampal LTP and LTD. These receptors can be activated by the neurotransmitter glutamate from the presynaptic cell only when the post-synaptic membrane is depolarized at the same time.^[163] The activation of NMDAR enables the influx of Ca^{2+} to post-synaptic cell, which triggers the subsequent signal pathway. Whether LTP or LTD occurs seems to depend on the amount and kinetics of Ca^{2+} influx.^[185] Generally speaking, rapid, large amount of Ca^{2+} influx will cause the AMPA receptors to be inserted to the post-synaptic membrane, and the synaptic transmission is strengthened (LTP), while more sustained low level Ca^{2+} influx triggers the signal pathway that mediates internalization of AMPA receptors and weakens synaptic transmission (LTD). STDP may involve more complex readout of Ca^{2+} influx by intracellular signaling machinery.^[185,186] Recent evidence suggests that some forms of LTD may even involve presynaptic NMDARs.^[187] At a longer timescale, late-phase LTP involves changes in gene expression.^[188]

To mimic LTSP, many schemes have been developed for different kinds of electronic devices, basically most of which are

based on nonvolatile switching characteristics of the devices. Once upon the presynaptic spikes, EPSC or IPSC is generated in the electronic synapse, then the current decays with a relatively long scale of time, from seconds to even days. The slow decay is ascribed to the device retention behavior. PCMs use the reversible crystalline-amorphous transition while RRAMs and CBRAMs rely on the control of ion movements to achieve stable conductance states in response to incoming presynaptic pulses. Therefore, the devices show an LTP. In some cases, CBRAM is difficult to realize LTD due to the difficulty in controlling the dissolution of metal filament.^[189] However, the conductance relaxation dynamics of such CBRAM can be utilized to gauge the relative timing of the pre- and post-synaptic spikes, which can be combined with a nonvolatile RRAM device to form a more faithful synapse that naturally exhibits short-term plasticity and long-term plasticity.^[190]

4.1.3. Wiring Plasticity

The two types of plasticity discussed above are both based on the weight change of connections already existed between neural units. (A neural unit can be an individual neuron, or part of a

neuron, or neurons with the same functional role.) These forms of plasticity are termed as weight plasticity. In addition, connections can be formed or eliminated entirely between neural units. This is termed as wiring plasticity (Figure 8c).^[190] Structural learning in computational models is an analogy to wiring plasticity in brain. Wiring plasticity is a less thoroughly studied area than weight plasticity, although it can be considered an extension of synaptic weight plasticity (to the extreme). Optical microscopy shows that dendritic spines (where synapses form) are formed and eliminated in healthy adult brain dynamically. Dendritic growth and axon remodeling are also hinted in adult brain.^[190,191] Molecular mechanisms behind the phenomena are still largely unexplored. Evidence indicates that wiring and weight plasticity share some common molecular pathways like the calcium/calmodulin-dependent protein kinase (CaMKII) pathway.^[192–194] Wiring plasticity extends the capacity of brain to encode information and is indicated to be relevant to a variety forms of learning, including perceptual learning, motor learning, spatial learning, etc.^[190,195–197] So far, weight plasticity has been widely studied and realized with electronic devices, while wiring plasticity is mainly adopted in computing algorithms only.

4.1.4. Intrinsic Plasticity

Although synaptic modification is generally considered as the dominant form of neuroplasticity relevant for learning and memory, parallel studies have shown that nonsynaptic changes may also play a role. For example, intrinsic plasticity involves the modification of neuronal excitability through regulation of voltage-dependent ion-channels and can be viewed as a change in the input–output function of a neuron. There are at least three mechanisms underlying the modulation of neuronal excitability (Figure 8d). Firstly, the regulation of dendritic channels amplifies EPSP. This effect is local because it only affects specific dendritic branches. Secondly, the voltage-dependent channels on the initial segment of axons are regulated to lower the spike threshold, thus it is easier to trigger an AP. Thirdly, alteration of the voltage-gated channels depolarizes the membrane resting potentials. The last two effects are global as they have influence on all synaptic inputs. Similar to synaptic plasticity, intrinsic plasticity is bidirectional, i.e., neuronal excitability can be either enhanced or decreased. Intrinsic plasticity is found to act synergistically with LTP or LTD, and show homeostatic properties. It is involved in many different forms of learning, such as classical conditioning, spatial learning, fear conditioning and odor conditioning.^[198] From the behavior point of view, modulating the threshold of electronic devices seems to be a possible way to mimic the regulation of the neuronal excitability. Recently, there have been attempts to demonstrate intrinsic plasticity using diffusive memristor-based nociceptor, which is a special type of sensory neurons that generate APs upon noxious stimulus to emulate the nociceptive behaviors including hyperalgesia and allodynia.^[199,200]

4.2. Metaplasticity and Synaptic Consolidation

Synaptic plasticity is determined not only by current stimulation, but also by the history of the synaptic states. Therefore, the

same stimuli may cause different effects in synaptic changes, and such phenomenon is described as metaplasticity.^[202] Besides, synaptic plasticity is closely related to memory formation and maintenance. Computational models indicate that the dilemma of maintaining old memory and encoding new memory can be solved by complex interactions of fast and slow cascade hidden variables.^[203–205] This is accordant with experimental evidence of synaptic consolidation.

4.2.1. Metaplasticity

Metaplasticity refers to the phenomenon that the ability of a neuron to generate synaptic plasticity is modulated by its history of activity. It is the “plasticity of synaptic plasticity.”^[202] For example, enriched-environment exposure of animals can facilitate later induction of LTP, while stressful stimuli with high intensity causes long-lasting (up to 24 h) inhibition of hippocampal LTP and facilitation of LTD.^[202] The mechanism of metaplasticity can be categorized as synaptic-specific and heterosynaptic regulation. In synaptic-specific metaplasticity, history activities of a synapse affect its own plasticity. NMDAR-mediated and mGluR receptor-mediated signaling pathways are the two major forms of synaptic-specific metaplasticity. On the other hand, plasticity of a synapse can also be affected by the history activities of its neighboring synapses.^[202] This can be explained by synaptic tagging and capture hypothesis (Section 4.2.2).

Some novel devices have been reported to implement the metaplasticity behavior.^[93,206–210] Kim et al. proposed a second-order memristor device, in which the internal thermal dissipation dynamics allows the oxide-based memristor to exhibit Ca^{2+} -like dynamics that encodes timing information and regulate synaptic weights.^[93] Ag diffusion dynamics in diffusive memristors has also been used to emulate Ca^{2+} dynamics with high fidelity.^[99] Furthermore, the history-dependent response can be embraced to present a modulation of plasticity, i.e., the polarity and the rate of conductance change, which is the metaplasticity.^[210] In other demonstrations, parameters like current compliance and stimuli (electrical pulse) interval are programmed to modulate the device plasticity change, which might be viewed as the metaplasticity of the electronic synapse.^[206–209]

4.2.2. Synaptic Consolidation

Synaptic consolidation is also referred to as late-phase LTSP (L-LTSP), during which the change of synaptic strengths is stabilized over minutes to hours after the initial induction event and is the first step of memory consolidation. It concerns how a STSP and early-phase LTSP (E-LTSP) are either maintained for transferring to LTSP or forgotten either through passive decay, interference by a similar stimulus, or active inhibitory processes.^[165] Repeated presentation (often termed rehearsal) can greatly enhance synaptic consolidation (Figure 9a), especially when they are spread out in time over the course of 24 h, i.e., distributed learning.^[211] The standard model of synaptic consolidation suggests that alterations of synaptic protein synthesis and changes in membrane potential are achieved through activating intracellular transduction cascades which lead to gene

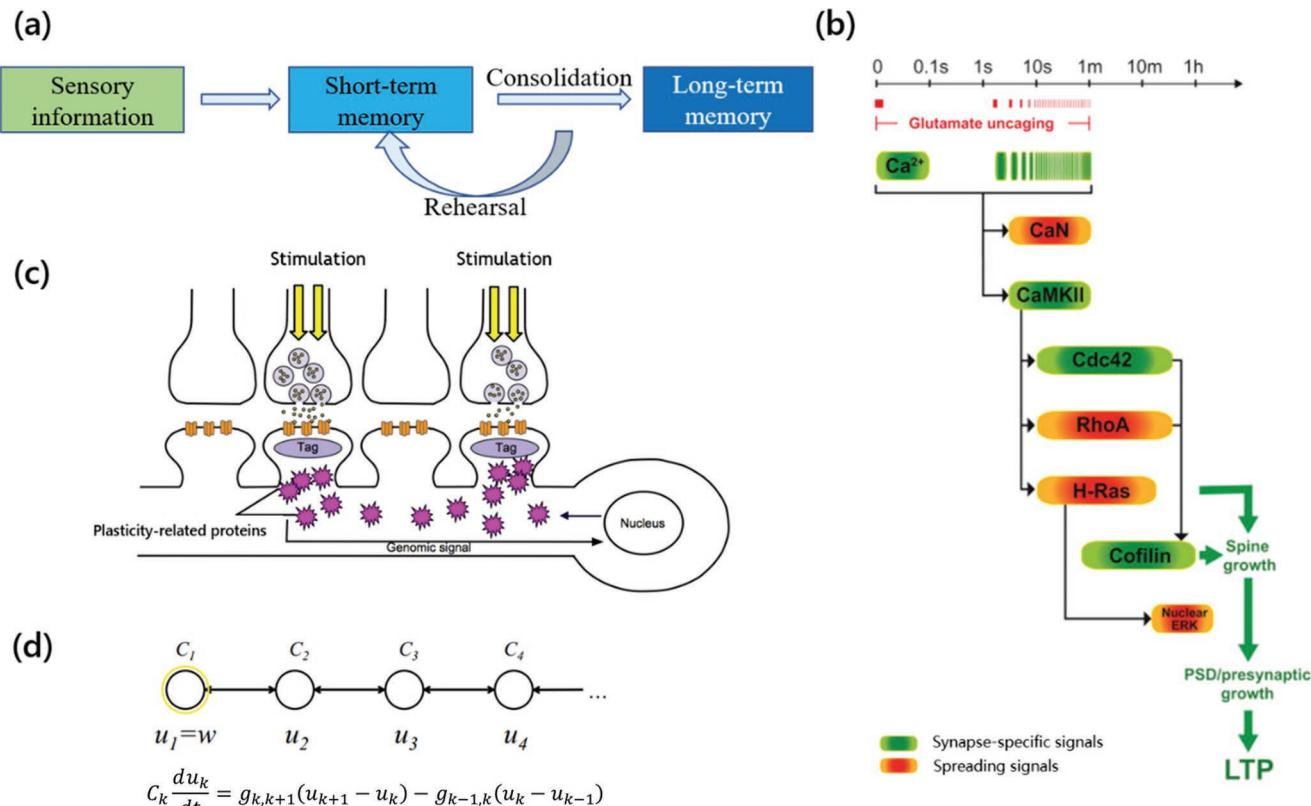


Figure 9. Illustration of synaptic consolidation and cascade models. a) Schematic of synaptic consolidation. b) Intracellular transduction cascades that alter synaptic protein synthesis and change membrane potential. Adapted with permission.^[212] Copyright 2014, Elsevier. c) Illustration of synaptic tagging and capture hypothesis. Adapted with permission.^[215] Copyright 2013, Oxford University Press. d) Illustration of the cascade model. u_1, u_2, \dots are bidirectional linked variables of different time scales. The formula describes the dynamics of each variable. Adapted with permission.^[203] Copyright 2017, Stefano Fusi.

transcription and protein synthesis (Figure 9b).^[212] Synaptic consolidation is also modulated by molecules such as dopamine which is associated with reward. This can potentially explain why behaviorally salient and emotionally charged memories are particularly persistent.^[213] One prominent theory to explain the process of synaptic consolidation is the synaptic tagging and capture hypothesis (Figure 9c): during E-LTSP, AMPA receptors are inserted to postsynaptic membrane and synapses grow in size.^[214] The initial induction event also activates molecular cascades and changes the state of the synapse to be as tagged. Then in L-LTSP, synthesized plasticity-related proteins (PRPs) move to tagged synapses and stabilize the AMPA receptors on membrane, thus maintain the change of synaptic strengths. As some synaptic tagging molecules can diffuse out of the synapse, PRPs can be transported to neighboring synapses. This enables different plasticity inducing events encoded by different synapses but close in time to therefore either constructively or destructively interfere.

Owning to the dual switching behaviors of some memristors, both volatile and nonvolatile conductance states can be obtained. The transition between two switching behaviors is quite similar to synaptic consolidation here. For example, Chang et al. showed experimentally that the retention loss in a RRAM memristor device bears striking resemblance to memory loss in biological systems.^[177] By stimulating the memristor with

repeated voltage pulses, an effect analogous to synaptic consolidation with a much improved retention time was observed in the memristor.^[176] Similarly, increasing the current compliance or the number of pulses applied to a CBRAM synapse, the conducting filaments will grow wider, thus the conductance state can be maintained for a longer period, which is a modulation from STSP to LTSP.^[98,216–218]

4.2.3. Cascade Model, Catastrophic Forgetting

As described above, synaptic plasticity acts at multiple time-scales and is affected by the history states. Fusi and colleagues developed a model with cascade variables of different time scales and bidirectional connections to represent the processes (Figure 9d).^[203–205] The fastest variable represents synaptic weights, and slower variables represent the concentration of signaling and regulatory molecules in the synapse. These slower variables have bi-stable (for example, phosphorylated or dephosphorylated) or multi-stable states. Some molecules are relatively unstable and may change their states in hours, like CaMKII involved in E-LTSP. They have a smaller time constant in the model, while molecules involved in L-LTSP have a larger time constant.^[203] This model enables synapses to have a great capacity of encoding memory due to the complex

cascade variables. It can explain metaplasticity and memory consolidation in a unified framework. Researchers have applied this model to a common problem in ANNs called catastrophic forgetting,^[219] which refers to the rapid forgetting of old memories due to a new learning that leads to catastrophic effects in a continual learning setting. An elastic synaptic plasticity algorithm inspired by this work was recently proposed for artificial neural network training.^[219]

5. Theories and Principles of Learning and Memory

In this section, the classical theories of learning and memory in biological neural systems will be reviewed. The mechanisms, phenomena, and the related electrical demonstrations (if any) will be introduced. We will also discuss the gap between biological learning and artificial emulations, and try to point out whether there are some important memory and learning functions in the neural system but have not been demonstrated with electronic devices yet.

5.1. Learning and Memory Theories for Biological Neural Systems

Learning is the most important brain function of intelligent animals in the world. Biologists and neuroscientists have made great endeavors to decipher the rules and theories of learning. Learning theories and rules could be revealed by experiments, based on which complicated models are formulated to explain biological behaviors. A major difference between biological learning and deep learning rules is the locality and globality. Efforts of theoretical studies have been aiming at bridging this gap, but few results have been accomplished. There are several distinct forms of memory in the field of neurobiology. For example, episodic memory is the memory of autobiographical events that can be explicitly expressed;^[220] working memory is a memory for temporary processing, like a brain cache; and body memory is a hypothesis of the memory stored in the body itself.^[221]

5.1.1. Hebbian Learning

Hebbian learning was proposed by Hebb over 70 years ago, and shaped much of the thinking about biological basis of learning ever since.^[201,222] A simple description of the Hebbian learning rule, is that an increase in synaptic efficacy arises from a presynaptic cell's repeated and persistent activation of a postsynaptic cell (**Figure 10a**). The principle is often summarized and referred to as "neurons that fire together wire together" which somewhat overlooked the requirement of temporal order of Hebb's original postulate.^[164] The brilliant insights of Hebb were later confirmed by the discovery of biological phenomenon of synaptic plasticity described above.

Spike Rate-Dependent Plasticity (SRDP): The mathematical model the Hebb's rule is originally formulated in terms of firing rates which refer to the average number of APs per unit time. Hebb's rule can be formalized as follows

$$\Delta w_{i,j} = \beta \cdot f_i(a_i) \cdot f_j(a_j) \quad (1)$$

where $\Delta w_{i,j}$ is the change in the strength of the connection between neurons i and j, β is the learning rate parameter and determines the magnitude of the change, while $f_i(a_i)$ and $f_j(a_j)$ are functions that depend on presynaptic activity (a_i) and postsynaptic activity (a_j), respectively.

The original form of Hebbian learning did not include LTD. The Bienenstock, Cooper, and Munro (BCM) learning rule was later proposed, in which a sliding threshold of the induction of LTP or LTD is determined by overall postsynaptic activity: a low postsynaptic activity favors LTD while a high postsynaptic activity favors LTP.^[223]

SRDP is a terminology used in the field of electronics rather than neuroscience to describe these rules. It describes the conductance change in memristors according to a simulation of a spike train of certain spike rate (Figure 10b). CMOS devices,^[162] CBRAM,^[99] RRAM,^[179] hybrid CMOS/RRAM,^[224] perovskite devices,^[225] and many others have been employed to show SRDP in response to incoming pulses at different frequencies: a high-frequency (20–100 Hz) train of presynaptic pulses results in LTP of the synaptic strength, whereas a low-frequency (1–5 Hz) train results in LTD. These protocols are consistent with aspects of the BCM rule.

STDP: STDP provides a spiking-based learning framework under the similar postulate of Hebbian learning.^[182,183,226,227] The STDP rule states that connection of synapse is strengthened when presynaptic activity proceeds postsynaptic activity for a few tens of milliseconds, and similarly, is weakened when presynaptic and postsynaptic activity in converse temporal order (Figure 10c). This beautiful and precise timing learning rule has been supported by many studies on different regions of brain and species. One way to quantify STDP is

$$\Delta \omega = \begin{cases} \omega_0 \exp(-\Delta t/\tau) & \Delta t > 0 \\ -\omega_0 \exp(\Delta t/\tau) & \Delta t < 0 \end{cases} \quad (2)$$

where Δt is the time difference between post- and presynapses, ω is the connectivity, and ω_0 and τ are two constants, controlling the learning rate and timing sensitivity, respectively.

Although the simple form of STDP is a well-tested model, limitations exist for this learning theory.^[226] First, a gap between macro and micro temporal scale remains, as a typical learning behavior is often on a timescale of minutes while STDP operates on tens of milliseconds. This gap may be filled by persistent neuronal or network activity,^[228] or by new learning rules that operate on a longer time-scale involving delayed reward signal^[229] or long-lasting plateau potential.^[230] Second, STDP is a pair-based spike plasticity model, not built for multiple pairs situation, and attempts have been made to expand STDP to the situation of multiple spike pairs.^[231] Third, STDP rule itself is subject to modulation and control by additional factors.^[232,233] This diversity of phenomena have not converged onto a canonical model yet, partly because STDP is still a phenomenological model, while building a mechanistic model of the synapse remains an important objective for the future.^[234]

Experimentally, the implementation of STDP using memristor typically relies on the overlap between pre- and postspikes applied on its two terminals.^[178,235–237] In this way, it is crucial to

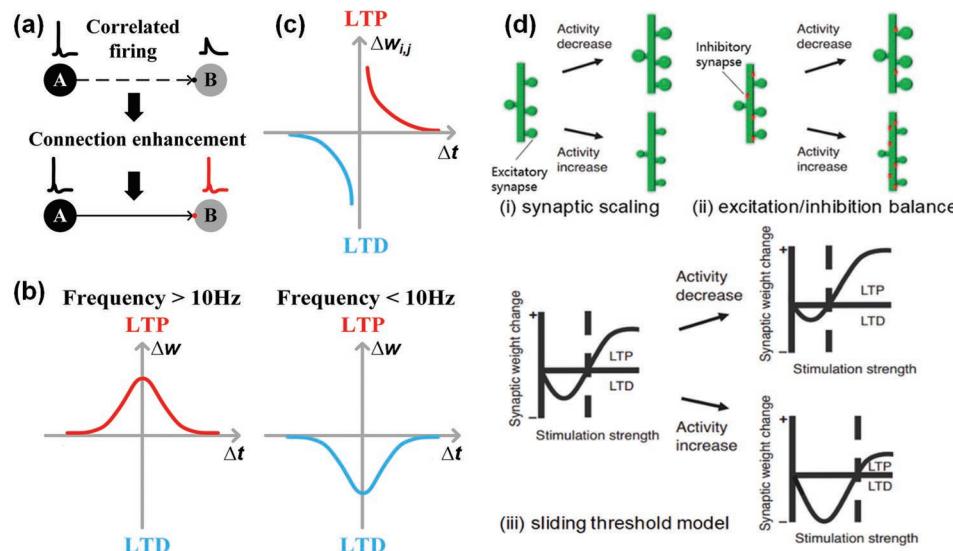


Figure 10. Illustration of different learning theories. a) Hebbian learning. Correlated firing makes the synapse between two neurons stronger. b) Schematic illustration of SRDP. c) Schematic illustration of STDP. d) Schematic of three different mechanisms for homeostatic plasticity: i) synaptic scaling; ii) shift of excitation and inhibition ratio; iii) sliding threshold. Adapted with permission.^[239] Copyright 2017, Elsevier.

translate the timing difference into the difference in amplitude or width of the superimposed spikes, which then modulate the synaptic weight as a function of the relative timing of the two spikes. In addition, certain memristor dynamics could provide a timing mechanism, from which STDP with non-overlapping spikes has also been realized. For example, a diffusive memristor in series of RRAM has been employed as a synaptic emulator, in which the relaxation characteristic of the diffusive memristor offers the timing mechanism.^[99] In addition, a second-order memristor with internal decay dynamics could also offer an intrinsic timing mechanism for realizing STDP.^[92,238]

5.1.2. Homeostatic Plasticity and Stability

Homeostatic plasticity is a form of plasticity that stabilizes the activity of a neuron or a neural circuit around a specific balance point.^[240] The counterpart of homeostatic plasticity is Hebbian plasticity. Although LTP and LTD act on opposite directions, Hebbian plasticity alone cannot maintain the homeostasis of neural circuits, because synapses undergoing LTP are more likely to activate the postsynaptic neurons, thus undergoing further LTP rather than LTD.^[240] Therefore, homeostatic plasticity is needed as a negative feedback to keep the normal activation level of neural circuits.^[241] Homeostatic plasticity often acts on the timescale of days, but more rapid homeostatic changes are also observed.^[242] Several mechanisms have been proposed underlying homeostatic plasticity (Figure 10d) including but not restricted to synaptic scaling, shift of excitation and inhibition ratio, sliding threshold for LTP and LTD, and changes in neuronal excitability.^[239,243] Experimentally, there has been attempt to demonstrate homeostatic plasticity using neuromorphic devices with global connectivity through electrolyte gating.^[244]

Synaptic scaling refers to the fact that synaptic strengths can make compensatory changes to restore average AP firing

rate in response to prolonged decreased or increased activity. In synaptic scaling, virtually all synaptic weights of a neuron are scaled down or up multiplicatively, which preserves the relative strength of the synapses.^[240] The balance between excitation and inhibition synapses (*E/I* balance) on a neuron can be shifted in response to decreased or increased activity for an extended period. Evidence shows that total inhibitory inputs to excitatory neurons decrease within hours to days after activity deprivation.^[239] The sliding threshold theory in the aforementioned BCM learning rule (Section 5.1.1.1) indicates that the altered activity level on a longer time can shift the thresholds for LTP and LTD, which is crucial for the stability of the rule and also regarded as a metaplasticity mechanism. Recently, these phenomena have been shown to share common molecular mechanisms among themselves and with Hebbian plasticity and a common metaplasticity framework has been proposed to explain them.^[239,241] At last, changes in neuronal excitability through a variety of intrinsic plasticity mechanisms are also possible to maintain homeostasis.^[193]

5.1.3. Memory Engram and Hebbian Assembly

Engrams are the specific changes in the brain that are formed by experience and stored in a quiescent state, which can be reactivated and become functional under appropriate presentation of external stimuli and conditions which lead to the retrieval of the stored memory.^[245,246] Hebb proposed that memories are stored in the brain with population of neurons which is now termed Hebbian assemblies. He proposed that Hebbian assemblies are formed by strengthening connections between neurons activated by a particular stimulus using the Hebbian rule for synaptic plasticity. This proposal was later formalized mathematically in the Hopfield model, which memorizes the patterns using the Hebbian rule and ensures the retrieval of a

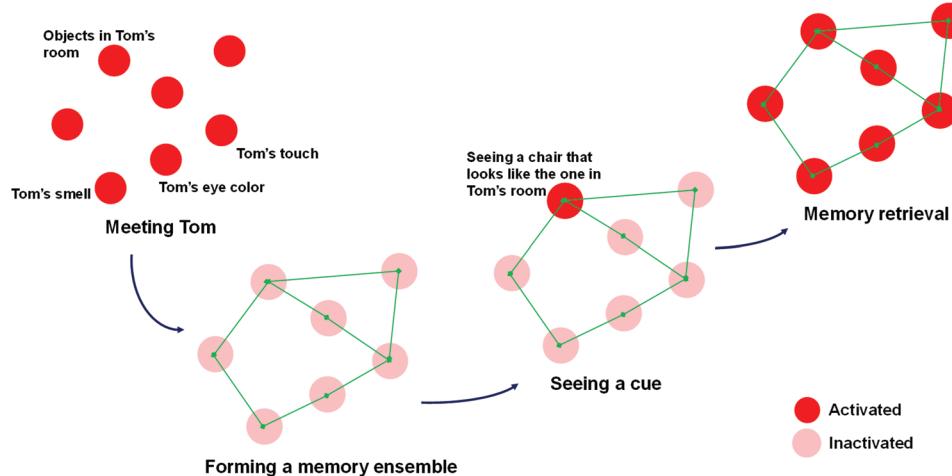


Figure 11. Illustration of memory engram. Here we take an example that an event, “Meeting Tom” can be retrieved by seeing a cue, “a chair that looks like the one in Tom’s room.” Adapted with permission.^[256] Copyright 2016, Lisa.

previously stored one upon a stimulus sufficiently similar to the original pattern.^[247] The existence of engrams is posited to explain the persistence of memory (Figure 11).^[248]

Recently, engram cells for particular memories have been demonstrated in the brain.^[249] They are reactivated by specific cues associated with the training experience and has been shown to be able to mediate memory retrieval when activated artificially with optogenetic techniques. Interestingly, it has been discovered that recruitment of a particular neuron into the memory ensemble is not only determined by the synaptic properties but also by intrinsic plasticity mechanisms.^[250,251] Furthermore, such changes in memory excitability through intrinsic plasticity mechanisms last for a period of time, therefore, memories close in time are often allocated to similar sets of neurons, while distant memories are stored in separate sets of neurons, leading to desired properties of memory generalization and separation.^[252]

5.1.4. System Consolidation of Memory

On a longer time scale, memories can also be shifted from one part of the nervous system to another part of nervous system. In a well-studied example, memories from the hippocampal region, where memories are first encoded, can be moved to the neocortex in a more permanent form of storage. Systems consolidation is a slow dynamic process over days and months that often requires sleep, and can take as long as one to two decades to be fully completed in humans, unlike synaptic consolidation that only takes minutes to hours for new information to stabilize into memories. System consolidation is a spatial transfer of memory rather than temporal transfer as in synaptic consolidation. This mechanism has been borrowed by DeepMind and inspired their work in building artificial intelligence systems.^[253–255]

5.2. Principles of Memory and Learning in Biological Neural System

In the previous section, we have outlined major theories that explain learning and memory from subcellular to system levels

and span from milliseconds to years. In this section, we would like to outline some important principles of learning and memory, also supported by experiments, which are features of the basic biological mechanisms that are essential for the proper functioning of the system.

5.2.1. Synaptic Competition and Normalization

Synaptic competition happens when synapses compete for limited resources in the brain, such as space and synaptic building blocks, or more abstract quantities such as a limited time window to control the firing of the postsynaptic neuron.^[227,257,258] These competitive interactions have been linked to Hebbian learning in an elegant theory by Miller and MacKay.^[259] Two types of competition or normalization are possible. The multiplicative form of normalization scales all synapses multiplicatively to maintain homeostasis and can be realized by synaptic scaling (Section 5.1.2).^[260] The subtractive form of normalization subtracts a constant number from the strengths of all synapses leading to the complete elimination of weak synapses. This can help to make synaptic connectivities sparse, which might improve the coding capacities of the neural network and encourage neurons to specialize and allow feature selection. The STDP rule (Section 5.1.1.2) has been shown to exhibit properties similar to subtractive normalization.^[227] In the artificial neural network literature, schemes such as weight decay, weight normalization has been proven essential to accelerate network learning.^[261] Multiplicative and subtractive normalizations are similar to L2 and L1 weight normalizations respectively.^[262]

5.2.2. Stochastic Synaptic Release and Synaptic Sampling

Another prominent property of the central nervous system, i.e., the brain, is that it appears to be quite noisy, e.g. the firing time of spikes can be modeled as a Poisson process and release of neurotransmitters from synapses are stochastic. A recent theory has linked this feature to a technique called sampling in Bayesian learning.^[263–265] Inference is a terminology in Bayes

probability, which means measuring the probability distribution of unknown variables. Similarly, brain is modeling outside world with its received stimulus, like visual, auditory, and somatosensory inputs and is often uncertain. In mathematics, there are two ways of doing inference, sampling-based inference and optimization-based inference. One of the most famous sampling-based inferences is Markov Chain Monte Carlo. Traditionally, learning in the brain is considered as optimization rather than sampling. With inherited stochastic processes inside neurons, like stochastic synaptic release and noise, learning can also be viewed as a sampling process. Synaptic sampling can help the network to reach global minimum and converge faster. Theory of synaptic sampling could be the source of better learning ability of humans than computers.

5.2.3. Control of Plasticity

Hebbian learning dictates that synaptic plasticity is determined by the firing rate of pre- and postsynaptic neurons and this rule is therefore considered local. However, synaptic strength between two individual neurons in a neuronal system could be affected and controlled by other parameters in the system. Three main routes for the control of plasticity have been relatively well studied.^[266] First, neuromodulator is a class of chemical molecule, which binds receptors on dendrites and regulate activity of a neuron. Neuromodulator could be released diffusely and affect and control many synapses within a volume. Neuromodulators are normally produced by neuromodulator neurons, which reside in specialized parts of the brain and represent globally important information such as reward, prediction error or salience.^[267] In adult animals, neuromodulators are usually necessary for the induction of plasticity and provide a temporal gating mechanism for the brain to update the targeted part of the internal representation with relevant information.^[268] Izhikevich has built a model based on STDP, bridging the temporal assignment gap of STDP for reinforcement learning (Section 5.3.3) by introducing dopamine, a type of neuromodulator associated with reward.^[269]

Second, plasticity in a local circuit can be controlled by inhibitory interneurons. In contrast to the excitatory neuron which can communicate information over long distances, the inhibitory neurons often only send out axons to a local volume and are often considered to serve control functions or carry out local computations. Such control neurons often send axons to many synapses to close their gates for plasticity. Another type of interneuron inhibits those neurons and are considerably disinhibitory neurons to allow local learning to occur, and often receive their inputs from distant sites.^[270–272] This mechanism could afford a more refined level of control than the diffuse release of neuromodulators.

Third, control of plasticity can also be realized through dendritic computation. The distant parts of the dendrites of excitatory neurons in the brain are also often considered modulatory as inputs at these locations often cannot cause the neurons to fire but can control the efficacy and plasticity of synaptic inputs closer to the soma. In the cerebral cortex, feedforward inputs from the sensory periphery that carry outside world information often form synapses close to the soma and are considered driving, while the modulatory inputs often are feedback

carrying cognitive variables such as task demand, reward and attention.^[273] This mechanism potentially affords the finest level of control of plasticity. Recently, several researchers have proposed alternatives to the well-adopted backpropagation algorithm in ANN training based on this mechanism.^[266,274]

Using multi-terminal Ag-based CBRAM memristor, the extra terminal with electrical bias is shown to mimic the use of external variables to modulate its plasticity and can be wired into more complicated circuits to potentially mimic such control mechanisms in the brain in the future.^[275]

5.3. Implementations of Learning with Emerging Devices

Based on the theories and principles discovered in neuroscience, many models and learning rules have been established to describe the function mathematically. Furthermore, hardware implementations of these learning rules with various emerging devices have been demonstrated, as to be reviewed next.

5.3.1. Unsupervised Learning

Unsupervised learning is to deal with unlabeled data, learn the relationship between elements in the dataset, and find the hidden structures, patterns, or features to analyze the new incoming data. The most classical example of unsupervised learning is Hebbian learning. In Hebbian learning, the connection is reinforced irrespective of an error, but is exclusively a function of the coincidence between APs between the two neurons. Similar to Hebbian learning, STDP is another version of unsupervised learning that takes into account the spike timing between the APs of pre- and postneurons hence to modulate the synaptic weights. Hebbian Learning has been hypothesized to underlie a wide range of neural and cognitive functions, such as neural development of receptive fields, pattern recognition, and experiential learning. Many other learnings are derivations from Hebbian learning plus some restrictions, such as competitive learning, in which certain nodes in the networks are strengthened to compete for the right to respond to a subset of the input data, and Sanger's learning, which is a linear feed-forward neural network model.^[276–278]

To demonstrate unsupervised learning in artificial neural system, some approaches have been developed using emerging devices. One is to employ memristors (e.g., CBRAM or RRAM) as synapses along with CMOS circuits as neurons (**Figure 12a**).^[45,189,279] The gradual and probabilistic switching behavior of the memristor synapses is exploited to realize the STDP weight update in learning. More recently, another approach is proposed, that is to use diffusive memristor as a neuron that can utilize the integrate-and-fire function to enable unsupervised synaptic weight updating and pattern classification on a fully memristor-composed neural networks.^[41]

5.3.2. Supervised Learning

Supervised learning is to infer a function through a labeled dataset and map the input (typically a vector) to the correct

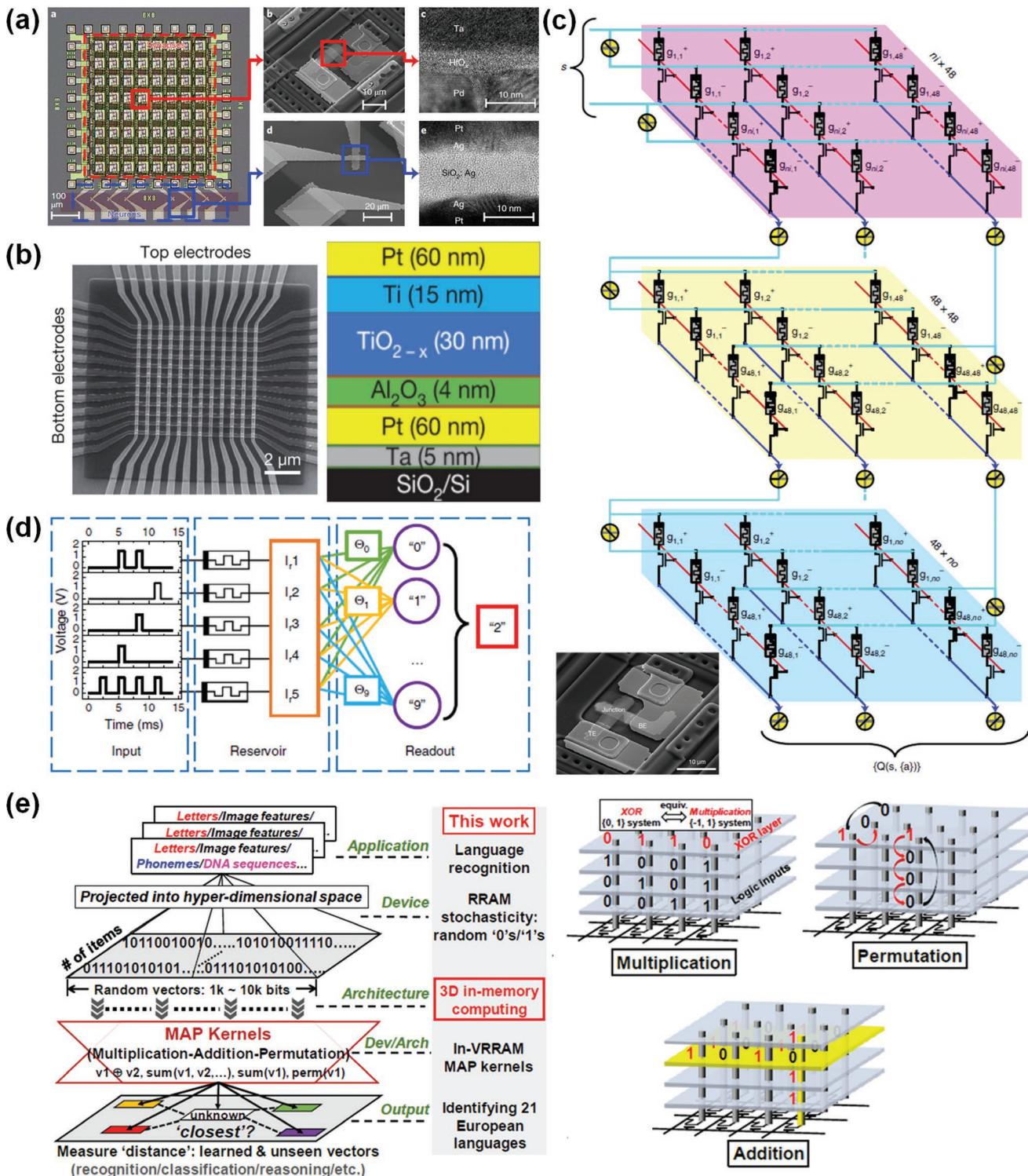


Figure 12. Demonstrations of emerging devices for learning and memory. a) An 8 × 8 crossbar array of Pt/HfO₂/Ta memristor synapses connected with Pt/Ag/SiO₂/Ag/Pt diffusive memristor neurons for unsupervised learning. Adapted with permission.^[41] Copyright 2018, Springer Nature. b) A 12 × 12 crossbar array of Pt/Al₂O₃/TiO_{2-x}/Ti/Pt memristors for supervised learning. Adapted with permission.^[30] Copyright 2015, Springer Nature. c) A three-layer fully connected Q-network for reinforcement learning using memristors. The inset is the scanning electron micrograph of a Pd/HfO₂/Ta memristive synapse. Adapted with permission.^[28] Copyright 2019, Springer Nature. d) Reservoir computing using W/WO_x/Pd memristors. Adapted with permission.^[36] Copyright 2017, Springer Nature. e) Proposed hyper-dimensional computing framework based on 3D vertical RRAM for one-shot learning. Adapted with permission.^[29] Copyright 2016, IEEE.

output value. By analyzing the training data, supervised learning can produce an inferred function and then determine the class labels of previously unseen examples. In supervised learning, backpropagation is widely used and well developed in training multi-layer feedforward and recurrent networks. Errors are backpropagated between the layers, adjusting each weight to minimize the loss function by gradient descent and reducing the classification error between the computed and the expected output vectors.

Generally, massive parallelism of multiplication and accumulation operations is required during synaptic weight update in supervised learning, and emerging devices in the crossbar structure meet exactly such requirement. RRAM, PCM, and organic devices have been demonstrated to be suitable and efficient in image recognition (e.g., the MNIST database of handwritten digits) with high accuracy and low energy consumption (Figure 12b).^[29,30,79,145]

For memristive MLPs and CNNs, depending on the available training sets, both supervised learning (e.g., MLP^[145] and CNN^[23] for pattern classification) and unsupervised learning (e.g., MLP^[280] for principal component analysis, CNN^[41] for pattern clustering) have been experimentally demonstrated. In comparison, memristive SNNs are mainly used for unsupervised learning so far, which relies on bio-plausible learning rules, such as STDP where the weight adjustment is both local to the synapse and local in time.^[41,280] The hardware-implemented plasticity makes memristors inherently attractive for unsupervised learning in SNNs. On the other hand, supervised learning rules using gradient-based optimization is difficult for SNNs because of the nondifferentiability of spike trains and weight transport problem, although modified rules such as SpikeProp^[281] (gradients of postsynaptic potential) and Chronotron^[282] (gradients of Victor–Purpura distance) have been reported. But such modified rules are more complicated than ANN supervised learning while not showing better performance yet. Other nongradient optimization methods includes ReSuMe rule,^[283] where the weight change is determined by the STDP with supervised signal and anti-STDP with neural output signal. A derivative of such learning has been implemented on a 16×1 RRAM array for spatiotemporal correlation detection.^[42] But the complicated supervised spiking neurons were built with digital circuits.^[42] A similar SPAN rule demands analog kernels for spike train conversion,^[284] which introduces extra complexity compared to that of unsupervised SNNs. In short, memristive SNNs may demand complicated digital circuitry for supervised learning rules while not showing much improved performance yet, which makes it less favorable so far.

5.3.3. Reinforcement Learning

Reinforcement learning refers to goal directed algorithms. Different from unsupervised and supervised learning, reinforcement learning usually involves a sequence of actions. A system, called “agent,” has the goal to maximize the cumulative reward when certain goals are achieved. Therefore, reinforcement learning refers to acting within an environment to maximize the reward through trial-and-error experience, e. g., game theory, optimal control. There are two main strategies

for solving reinforcement learning problems. The first strategy, called direct policy search, is to search in the space of behaviors in order to find one that performs well in the environment. The second one, called value function estimation, is to use statistical techniques and dynamic programming methods to estimate the utility of taking actions in states of the world.

An exemplary implementation of reinforcement learning is AlphaGo, in which DNNs are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play.^[285,286] Without any look-ahead search, the neural networks play “Go” at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. It is the first time that a computer program defeats a human professional player in the full-sized game of Go, but with enormous power consumption. An updated version of AlphaGo, i.e., AlphaGo Zero, was trained solely by reinforcement learning with zero inputs from human expert games, defeated AlphaGo 100 times in 100 games.^[285] In recent years, emerging materials and devices have shown great potentials for energy-efficient hardware implementation of reinforcement learning (Figure 12c).^[287] The memristor device with its cross-point array structure can realize the vector-matrix multiplication with ultrahigh energy efficiency, so that transfer time complexity from $O(n^2)$ to $O(1)$, boosting the energy efficiency of reinforcement learning.^[288] However, there still exist some challenges to realize reinforcement learning on memristors. One major obstacle is that the copy operation between neural networks needs two crossbar arrays to realize conductance copy. But due to the stochastic characteristic of memristors, it is rather difficult to precisely program each cell to a target conductance. Another challenge is in the training phase of neural network when calculating the weights variation, which needs a large amount of high-accuracy registers and computing units, such as multipliers. The stochastic gradient algorithm is modified to adapt to memristors without loss of convergence so as to reduce the cost of training phase in memristors.

In comparison, supervised, unsupervised, and reinforcement learning algorithms have the common goal of estimating the class conditional distribution from the training data. The essential difference among these different learning algorithms is the availability of feedback information from the environment indicating which pattern class generated the feature vector. In supervised learning, each feature vector in the training data set contains explicit information regarding which pattern class generated that feature vector. In unsupervised learning, information regarding which pattern class generated a particular feature vector is not available in either the training data set or the test data set. In reinforcement learning, “hints” about which pattern class generated a particular feature vector are provided either periodically or aperiodically throughout the learning process.^[289]

5.3.4. Reservoir Computing

Reservoir computing is an often used method to learn time-series. It employs dynamic reservoirs having short-term memory to project features from the temporal inputs into a

high-dimensional feature space. Typically, input signals are transmitted into dynamical system with fixed connections called a reservoir and the dynamics of the reservoir map the inputs to a higher dimension. Then a readout function is trained to read the inner state of the reservoir and map it to the desired output. The main benefit is that training is performed only at the readout stage and the reservoir is fixed, so the training cost can be greatly reduced compared to traditional RNNs. During the computing, the network nodes (neurons) in the reservoir evolves with the temporal input, so the inner states are determined by both the current inputs as well as the inputs within a certain period in the past. Therefore, the reservoir itself must have short-term memory.

There are some researches that have demonstrated the implementation of reservoir computing, via field programmable gate arrays (FPGAs), silicon photonics chips, and other hardware systems, to solve temporal classification, regression or prediction tasks such as speech recognition.^[290,291] Memristors, due to the intrinsic nonlinearity and short-term memory effects, has also been proved as a perfect fit for the reservoir computing system.^[36,292] Du et al. used dynamic memristor devices to map temporal input patterns into different reservoir states, using the collective memristor resistance states, and then further process the states through a simple readout function (Figure 12d).^[36] The memristor-based system is able to perform hand digit recognition tasks and solve a second-order nonlinear task.

5.3.5. One-Shot and Few-Shot Learning and Transfer Learning

One-shot and few-shot learning is one of the current frontiers of deep learning. Traditionally supervised deep learning has achieved great success when a large labeled dataset is available. However, humans can learn an object category even when only one or few instances are shown. Even at a young age of six, one has learned almost all of the 10 to 30 thousand object categories in the world. This is due not only to the computational power of human brains, but also to the ability to synthesize and learn new object classes from limited information about different, previously learned classes. One-shot and few-shot learning aim to achieve a similar goal. In machine learning, one-shot learning can be performed via two methods. One is the directly supervised learning-based method, which directly learns the limited dataset without any auxiliary data. This often requires an explicit a priori model to be available. The other method is to transfer learned knowledge from auxiliary data to do the classifications and this paradigm is often called transfer learning.^[293]

Although it seems to present a particularly challenging problem for AI, several hardware architectures have been built via emerging devices to implement one-shot learning. For example, in the hyper-dimensional computing framework, information (letters, phonemes, DNA sequences, etc.) is represented and distributed in binary vectors with thousands of random “0” s and “1” s. A four-layered vertical RRAM array with FinFET selectors is fabricated and used as nonvolatile kernels in computing.^[294] Thanks to the array structure, multiplication, addition, and permutation are mapped onto 3D vertical RRAM with

low-resistance state (LRS) representing “1” and high-resistance state (HRS) representing “0,” hence the computing can be completed with high efficiency (Figure 12e). Besides, FinFET, and carbon nanotube FET with RRAM have also been utilized as the computing elements for one-shot learning.^[295,296]

6. Outlook

Although ANNs have been inspired by the biological nervous system, it should be pointed out that the gap between them is still very large. **Table 2** briefly summarizes the research status of biological and artificial neural systems to illustrate the details. A few examples are briefly discussed below.

6.1. Diversity

In biological system, there is a large variety of neurons that collaboratively function. The most important division is the division into excitatory and inhibitory neurons, while in current ANNs neurons send out mixed excitatory and inhibitory types of synapses. This division allows for further functional specialization of those neurons. Currently, a great concerted international effort is underway to create a catalog of all cell types in the nervous system.^[297] We estimate there are hundreds of different types of neurons. For the retina and cerebral cortex, a complete cell census appears within reach. Some of those types are present at different parts of brain, but even at the level of a local circuit in the cerebral cortex, there are at least four types of inhibitory neurons which are well studied and the final number will probably be in the tens. For each type of neuron, there is a set of morphological, physiological, molecular and connectional properties associated, which might dictate different computational and plasticity rules. Elucidating how this diversity contributes to computational function and delineating the detailed role of each type of neuron will occupy much of the effort of neuroscientist in the coming years. Examples of different types of neurons are illustrated in **Figure 13**. In a hardware composed of electronic devices, however, there are usually limited types of devices with unique function(s) for each type. How to model this diversity on the neuronal level in electronic devices will be one of the big challenges in the future.^[297]

Some of the diversities that we think are of particular importance will be discussed in more details below.

6.1.1. Excitatory and Inhibitory Neurons

Divided by the type of responses it produces in the downstream neuron upon firing, there are two types of neurons in the nervous system, excitatory and inhibitory neurons. The excitatory neurons are considered to be information coding and carry information across different parts of the brain. The inhibitory neurons are usually local interneurons, although they can also be long range projecting neurons in other parts of the brain. This makes it possible for the inhibitory neurons to perform specialized computations. As explained under the previous section, the electrical properties and plasticity rules associated with

Table 2. Comparison of research status on biological and artificial neural systems.

Component/function	BioNNs				ANNs		
	Importance	Experimentally observed	Model established	Mechanism clarified	Algorithm realized	Material/device demonstrated	System implemented
Ion channels	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	- ^{d)}	+ ^{c)}	- ^{d)}
AP generation	++ ^{b)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	++ ^{b)}	- ^{d)}
EPSC/EPSP, IPSC/IPSP	++ ^{b)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	++ ^{b)}	- ^{d)}
Short-term plasticity (PPF, PPD)	++ ^{b)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}
Long-term plasticity	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	++ ^{b)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}
Wiring plasticity	+ ^{c)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}	+++ ^{a)}	- ^{d)}	- ^{d)}
Intrinsic plasticity	++ ^{b)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}	- ^{d)}	+ ^{c)}	- ^{d)}
Homeostatic plasticity	+++ ^{a)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}	- ^{d)}	+ ^{c)}	- ^{d)}
Metaplasticity	+++ ^{a)}	+++ ^{a)}	+ ^{c)}	+ ^{c)}	- ^{d)}	+ ^{c)}	- ^{d)}
Synaptic consolidation	++ ^{b)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}	- ^{d)}	+ ^{c)}	- ^{d)}
Hebbian rule	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	++ ^{b)}
SRDP (BCM rule)	++ ^{b)}	++ ^{b)}	+++ ^{a)}	+ ^{c)}	+++ ^{a)}	+++ ^{a)}	++ ^{b)}
STDP	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}	+++ ^{a)}
Memory engram	+++ ^{a)}	++ ^{b)}	++ ^{b)}	++ ^{b)}	- ^{d)}	- ^{d)}	- ^{d)}
Memory system consolidation	++ ^{b)}	++ ^{b)}	++ ^{b)}	++ ^{b)}	- ^{d)}	- ^{d)}	- ^{d)}
System competition	++ ^{b)}	++ ^{b)}	++ ^{b)}	++ ^{b)}	- ^{d)}	- ^{d)}	- ^{d)}
Synaptic sampling	++ ^{b)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}	- ^{d)}	- ^{d)}	- ^{d)}
Control of plasticity	+++ ^{a)}	++ ^{b)}	++ ^{b)}	++ ^{b)}	- ^{d)}	+ ^{c)}	- ^{d)}
Supervised learning	+ ^{c)}	N/A ^{e)}	N/A ^{e)}	N/A ^{e)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}
Unsupervised learning	++ ^{b)}	N/A ^{e)}	N/A ^{e)}	N/A ^{e)}	+++ ^{a)}	++ ^{b)}	++ ^{b)}
Reinforcement learning	+++ ^{a)}	N/A ^{e)}	N/A ^{e)}	N/A ^{e)}	+++ ^{a)}	+ ^{c)}	+ ^{c)}
Reservoir computing	++ ^{b)}	N/A ^{e)}	N/A ^{e)}	N/A ^{e)}	+++ ^{a)}	+ ^{c)}	+ ^{c)}
Low-shot learning	+++ ^{a)}	N/A ^{e)}	N/A ^{e)}	N/A ^{e)}	+++ ^{a)}	+ ^{c)}	+ ^{c)}
Dendritic computation	++ ^{b)}	++ ^{b)}	++ ^{b)}	+ ^{c)}	+ ^{c)}	+ ^{c)}	- ^{d)}

^{a)}Extensively studied; ^{b)}Studied; ^{c)}Just started; ^{d)}Not studied yet; ^{e)}Not applicable.

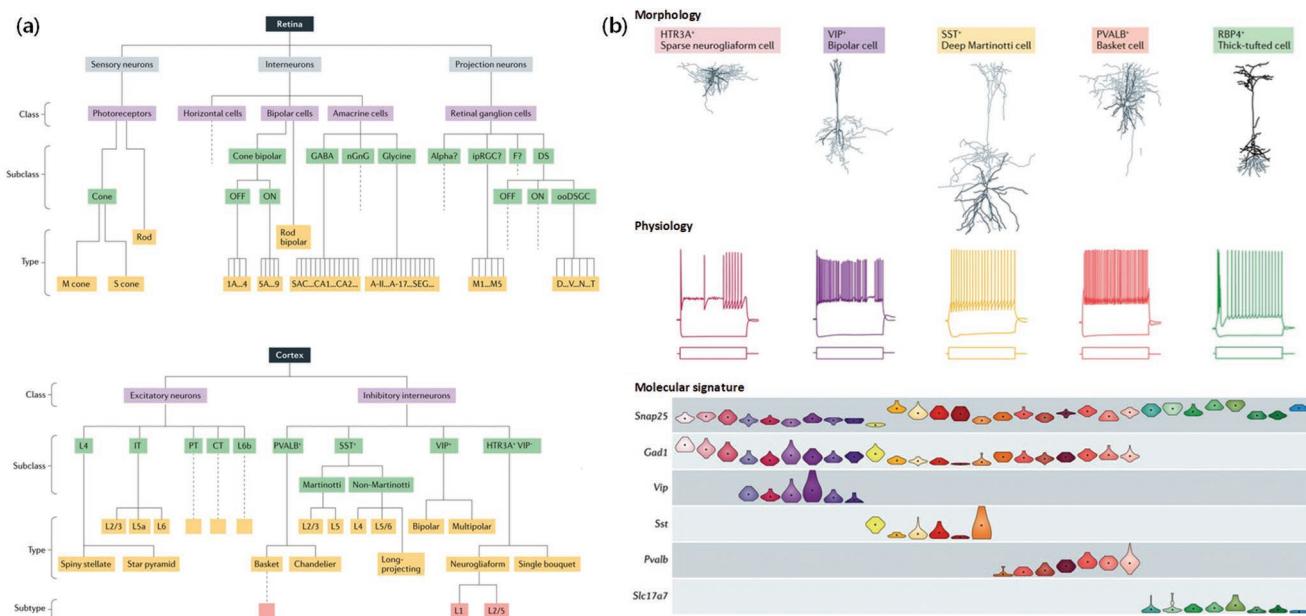


Figure 13. Examples of different types of neurons. a) Neuronal cell-type classification of the retina and cerebral cortex. b) Criteria by which neurons are classified: morphology, physiology, and molecular signature. Adapted with permission.^[297] Copyright 2017, Springer Nature.

the inhibitory neurons are often quite different from the excitatory neurons, which would not be possible if a single type of neurons send out mixed excitatory and inhibitory synapses as in current ANNs.

Some of the currently known functions performed by inhibitory neurons include the following:^[298] First, the inhibitory neurons can mediate lateral competition and winner-take-all type of operations, which can be useful in generating sparse codes and solve constraint-satisfaction problems.^[299,300] Second, brain uses sophisticated mechanisms to maintain the total amount of excitation and inhibition in the circuit to be roughly the same. Besides ensuring stability and giving robustness to the system, circuits maintained in this approximate balance can give fast responses to incoming stimuli. Third, inhibitory neurons act as gates to control aspects of local processing and ensure spatiotemporal coordination, often mediated by different kinds of interneurons, such as output gates (chandelier neurons), input gates (somatostain neurons), and plasticity gates (VIP neurons). This divides the local circuit into different computing subroutines that can be called upon at different times. Fourth, interneuron system can act as a distributed clock. The brain has numerous rhythms spanning the time epochs from tens of seconds to milliseconds, which can be set up by negative feedback loops using inhibitory neurons. These local clocks can also synchronize with each other giving rise to sophisticated communication schemes. How to model those aspects of the neuronal circuits would be an interesting challenge for the neuromorphic engineering community. It is worth mentioning that there are recently some interesting studies on neuromorphic computing based on device synchronization.^[301,302]

6.1.2. Dendritic Computation

Recently, there is growing research interest driven by recent technological advances to the capability of dendrites in single neuron in computing.^[303] Recent work show that neuronal dendrites exhibit a range of nonlinear mechanisms that allow them to implement complex computations beyond that of simple summation in perceptron neurons (Figure 14a).^[304,305] For example, a single neuron can carry out computations that are similar to a network of two-layer perceptron (palimpsest neuron, Figure 14c) or even more sophisticated operations such multiplication of two sets of inputs (complex neurons, Figure 14d). Emerging neuromorphic devices, e.g., RRAM, have been used to emulate the dendritic computation,^[306] which could be a potential approach to improve computation ability for future high-volume computing tasks.

6.2. Temporal Coordination and Control of Information Processing

The brain uses spikes to communicate. This gives the opportunity to use complex spatiotemporal codes and coordination schemes for computation, and there are recently several implementations based on RRAM.^[42,307] As mentioned above, the synchronization of the firing times can be used as distributed clocks to temporarily bind several items together for processing. Besides, the brain activities rely greatly on historical events while ANNs emphasize more on the current input. The brain is constantly performing predictions into the future which might underlie some of its advanced intelligence.

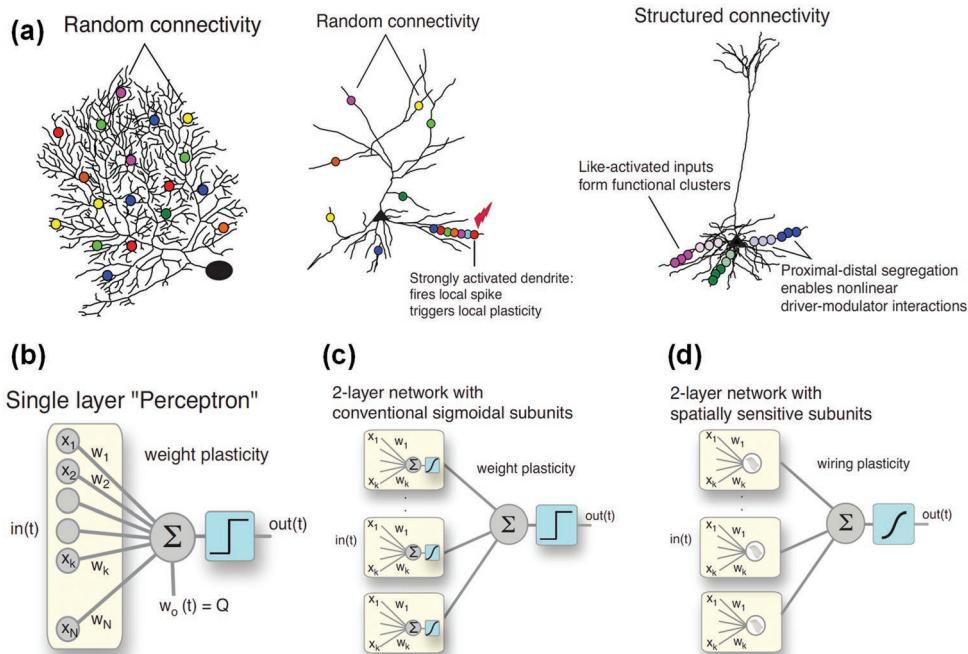


Figure 14. The diversity of dendritic processing in BioNNs. a) Schematic of connectivity on the dendritic tree in biological neural system. b-d) Schematic of computational models that mimic each situation: b) perceptron neuron, c) palimpsest neuron, and d) complex neuron. Adapted with permission.^[303] Copyright 2017, Elsevier.

One of the largest gaps is that the brain uses a sophisticated control to regulate the flow of information and plasticity. There are also dedicated areas such as thalamus that serve as gate and hub for information flow. Such modulation can be realized through neuromodulators released more diffusely or specialized circuits using inhibitory neurons or specialized compartments on the dendrites (Section 5.2.3). Together with the distributed clocking system mentioned above, the brain can be set up to have a different network configuration for different tasks whenever needed. This is somewhat similar to the mixture-of-experts network in current artificial neural network literature but in a much more sophisticated manner.^[268,308]

6.3. Advanced Learning Capabilities

Although deep learning has made big strides in recent years, the automatic learning ability of biological systems is still unrivaled in robustness and flexibility. Several researchers have voiced criticisms of deep learning in recent years on those grounds.^[309]

6.3.1. Memory Engram and Innate Structures in Neural Circuit

Recent advances in neuroscience techniques have allowed us to observe and manipulate a population of neurons encoding a memory engram, the engram cells.^[248] This brings the investigation of learning and memory to the cell population and circuitry level. Such memory might not be formed arbitrarily from a tabula rasa as typically done in computer memories, but might be highly regulated and organized according to innate principles.^[310] How such rules relate to the diversity of cell types also remain to be investigated. Progress is currently rapid in this field. This aspect has rarely been emulated using neuromorphic devices yet. There seems a possibility that by using nonvolatile electronic devices, the storage and retrieve function of memory can be realized.

6.3.2. One-Shot Learning and Continual Learning

Current deep learning algorithms achieve success by learning from big volume of labeled data. But the biological nervous system is able to learn from a few examples and continue to learn without explicit instructions automatically while performing functions throughout the life span, not just at specialized times (Sections 4.2.3 and 5.3.5). How to close this gap is a currently very active area in artificial intelligence. Researchers in the neuromorphic computing field are also starting to tackle these problems and much progress is expected in this area.

In summary, there are still many inevitable challenges in materials and devices to implement bio-realistic neural systems. A great diversity of neuronal and synaptic types and subtypes exists in biological system, as well as the complexity of plasticity or learning rules (including modulation). Such diversity originates from different ways to assemble many parts (e.g., protein molecules and complexes) selected in evolution, and may be critical for the robustness and other properties of biological systems. Implementing similar diversity in electronic systems with current technology will be a daunting task. Yet, it

could be worthwhile to implement certain aspects of it, and to explore how critical these aspects are for the proper functionalities of biological and artificial systems.

Acknowledgements

J.T., F.Y., and X.S. contributed equally to this work. This work was supported in part by the National Key R&D Program of China (2018YFB0407604), National Natural Science Foundation of China (61851404, 61874169, 61974081), National Major Research Program (2017ZX02315001-005), Beijing Municipal Science and Technology Project (Z181100003218001, Z191100007519008), and National Young Thousand Talents Plan.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

artificial neural systems, biological neural systems, brain-inspired computing, memristors, neural networks, neuromorphic devices

Received: April 30, 2019
Revised: August 16, 2019
Published online:

- [1] A. M. Turing, *Mind* **1950**, LIX, 433.
- [2] M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **2018**, 1, 22.
- [3] T. Gokmen, Y. Vlasov, *Front. Neurosci.* **2016**, 10, 333.
- [4] M. L. Minsky, *Ph.D. Dissertation*, Princeton University **1954**.
- [5] C. Mead, *Proc. IEEE* **1990**, 78, 1629.
- [6] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA **1989**.
- [7] R. Douglas, M. Mahowald, C. Mead, *Annu. Rev. Neurosci.* **1995**, 18, 255.
- [8] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbrück, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, K. Boahen, *Front. Neurosci.* **2011**, 5, 73.
- [9] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, *Science* **2014**, 345, 668.
- [10] Q. Xia, J. J. Yang, *Nat. Mater.* **2019**, 18, 309.
- [11] L. Chua, *IEEE Trans. Circuit Theory* **1971**, 18, 507.
- [12] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, *Nature* **2008**, 453, 80.
- [13] I. Basheer, M. Hajmeer, *J. Microbiol. Methods* **2000**, 43, 3.
- [14] B. Rajendran, F. Alibart, *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2016**, 6, 198.
- [15] D. S. Jeong, C. S. Hwang, *Adv. Mater.* **2018**, 30, 1704729.
- [16] C.-H. Kim, S. Lim, S. Y. Woo, W.-M. Kang, Y.-T. Seo, S.-T. Lee, S. Lee, D. Kwon, S. Oh, Y. Noh, H. Kim, J. Kim, J.-H. Bae, J.-H. Lee, *Nanotechnology* **2019**, 30, 032001.
- [17] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, R. D. Mooney, M. L. Platt, L. E. White, *Neuroscience*, Oxford University Press, Oxford **2017**.

- [18] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, *Principles of Neural Science*, McGraw-Hill Education, New York **2000**.
- [19] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, *Proc. IEEE* **1998**, *86*, 2278.
- [20] L. G. Ungerleider, *Science* **1995**, *270*, 769.
- [21] W. S. McCulloch, W. Pitts, *Bull. Math. Biophys.* **1943**, *5*, 115.
- [22] D. Garbin, O. Bichler, E. Vianello, Q. Rafhay, C. Gamrat, L. Perniola, G. Ghibaudo, B. DeSalvo, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2014**, pp. 28.4.1–28.4.4.
- [23] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, G. W. Burr, *Nature* **2018**, *558*, 60.
- [24] Z. Wang, C. Li, P. Lin, M. Rao, Y. Nie, W. Song, Q. Qiu, Y. Li, P. Yan, J. P. Strachan, N. Ge, N. McDonald, Q. Wu, M. Hu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Mach. Intel.* **2019**, *1*, 434.
- [25] V. Sze, Y.-H. Chen, T.-J. Yang, J. S. Emer, *Proc. IEEE* **2017**, *105*, 2295.
- [26] Y. Lecun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
- [27] F. M. Bayat, M. Prezioso, B. Chakrabarti, H. Nili, I. Kataeva, D. Strukov, *Nat. Commun.* **2018**, *9*, 2331.
- [28] I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, E. Eleftheriou, *Nat. Commun.* **2018**, *9*, 2514.
- [29] Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. Alec Talin, A. Salleo, *Nat. Mater.* **2017**, *16*, 414.
- [30] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, *Nature* **2015**, *521*, 61.
- [31] T. Gokmen, M. Onen, W. Haensch, *Front. Neurosci.* **2017**, *11*, 538.
- [32] X. Peng, R. Liu, S. Yu, *IEEE Int. Symp. Circuits Syst.*, IEEE, Piscataway, NJ **2019**, pp. 1–5.
- [33] Z. C. Lipton, J. Berkowitz, C. Elkan, *arXiv preprint arXiv:1506.00019*, **2015**.
- [34] S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9*, 1735.
- [35] M. Schuster, K. K. Paliwal, *IEEE Trans. Signal Process.* **1997**, *45*, 2673.
- [36] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, W. D. Lu, *Nat. Commun.* **2017**, *8*, 2204.
- [37] T. Gokmen, M. J. Rasch, W. Haensch, *Front. Neurosci.* **2018**, *12*, 1.
- [38] C. Li, Z. Wang, M. Rao, D. Belkin, W. Song, H. Jiang, P. Yan, Y. Li, P. Lin, M. Hu, N. Ge, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Mach. Intell.* **2019**, *1*, 49.
- [39] H. Tsai, S. Ambrogio, C. Mackin, P. Narayanan, R. M. Shelby, K. Rocki, A. Chen, G. W. Burr, *Symp. VLSI Technol.*, IEEE, Piscataway, NJ **2019**, pp. T82–T83.
- [40] L. Abbott, *Brain Res. Bull.* **1999**, *50*, 303.
- [41] Z. Wang, S. Joshi, S. Savel'Ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Electron.* **2018**, *1*, 137.
- [42] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, *Sci. Adv.* **2018**, *4*, eaat4752.
- [43] T. Werner, E. Vianello, O. Bichler, D. Garbin, D. Cattaert, B. Yvert, B. De Salvo, L. Perniola, *Front. Neurosci.* **2016**, *10*, 474.
- [44] S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z.-Q. Wang, A. Calderoni, N. Ramaswamy, D. Ielmini, *IEEE Trans. Electron Devices* **2016**, *63*, 1508.
- [45] A. Serb, J. Bill, A. Khiat, R. Berdan, R. Legenstein, T. Prodromakis, *Nat. Commun.* **2016**, *7*, 12611.
- [46] S. Ambrogio, N. Ciocchini, M. Laudato, V. Milo, A. Pirovano, P. Fantini, D. Ielmini, *Front. Neurosci.* **2016**, *10*, 56.
- [47] S. R. Nandakumar, I. Boybat, M. Le Gallo, E. Eleftheriou, A. Sebastian, B. Rajendran, *arXiv preprint arXiv:1905.11929*, **2019**.
- [48] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, A. Maida, *Neural Networks* **2019**, *111*, 47.
- [49] A. Chattopadhyay, C. H. Chang, H. Yu, *Emerging Technology and Architecture for Big-Data Analytics*, Springer International Publishing, Cham **2017**.
- [50] S. R. y Cajal, *Nobel Lect. Physiol. or Med.* **1906**, 220.
- [51] D. Bucher, J.-M. Goaillard, *Prog. Neurobiol.* **2011**, *94*, 307.
- [52] C. Grienberger, X. Chen, A. Konnerth, *Trends Neurosci.* **2015**, *38*, 45.
- [53] C.-L. Tao, Y.-T. Liu, R. Sun, B. Zhang, L. Qi, S. Shivakoti, C.-L. Tian, P. Zhang, P.-M. Lau, Z. H. Zhou, G.-Q. Bi, *J. Neurosci.* **2018**, *38*, 1493.
- [54] A. H. Zhang, G. Sharma, E. A. B. Undheim, X. Jia, M. Mobli, *Neurosci. Lett.* **2018**, *679*, 35.
- [55] BioNinja, Nervous System, <http://www.vce.bioninja.com.au/aos-2-detecting-and-respond/coordination-regulation/nervous-system.html> (accessed: August 2019).
- [56] Y.-H. Tu, A. J. Cooper, B. Teng, R. B. Chang, D. J. Artiga, H. N. Turner, E. M. Mulhall, W. Ye, A. D. Smith, E. R. Liman, *Science* **2018**, *359*, 1047.
- [57] T. J. Jentsch, W. Günther, M. Pusch, B. Schwappach, *J. Physiol.* **1995**, *482*, 19.
- [58] A. D. Nelson, P. M. Jenkins, *Front. Cell. Neurosci.* **2017**, *11*, 136.
- [59] C. Gu, J. Barry, *Prog. Neurobiol.* **2011**, *94*, 115.
- [60] R. Schneggenburger, Y. Han, O. Kochubey, *Cell Calcium* **2012**, *52*, 199.
- [61] S. Mochida, *Neurosci. Res.* **2018**, *127*, 33.
- [62] M. A. Böhme, A. T. Grasskamp, A. M. Walter, *FEBS Lett.* **2018**, *592*, 3516.
- [63] C. Guevara, G. Farias, C. Silva-Rosas, P. Alarcon, G. Abudinen, J. Espinoza, A. Caro, H. Angus-Leppan, J. de Grazia, *Front. Immunol.* **2018**, *9*, 2568.
- [64] E. Szodorai, K. Bampali, R. A. Romanov, S. Kasper, T. Hökfelt, M. Ernst, G. Lubec, T. Harkany, *Cell. Signal.* **2018**, *50*, 142.
- [65] R. L. Macdonald, R. W. Olsen, *Annu. Rev. Neurosci.* **1994**, *17*, 569.
- [66] I. H. Greger, J. F. Watson, S. G. Cull-Candy, *Neuron* **2017**, *94*, 713.
- [67] N. Scheehals, H. D. MacGillavry, *Mol. Cell. Neurosci.* **2018**, *91*, 82.
- [68] J. D. Clements, G. L. Westbrook, *Neuron* **1991**, *7*, 605.
- [69] G. Riedel, *Behav. Brain Res.* **2003**, *140*, 1.
- [70] S. Freeman, *Biological Science*, Pearson, Prentice Hall, NJ **2005**.
- [71] Designua, Ligand-gated ion channel, <https://www.dreamstime.com/stock-illustration-ligand-gated-ion-channel-proteins-which-open-to-ions-na-k-ca-cl-image76834578> (accessed: August 2019).
- [72] J. J. Yang, D. B. Strukov, D. R. Stewart, *Nat. Nanotechnol.* **2012**, *8*, 13.
- [73] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. Le Gallo, K. Moon, J. Woo, H. Hwang, Y. Leblebici, *Adv. Phys.: X* **2017**, *2*, 89.
- [74] D. Kuzum, S. Yu, H.-S. Philip Wong, *Nanotechnology* **2013**, *24*, 382001.
- [75] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, *Nat. Nanotechnol.* **2016**, *11*, 693.
- [76] H. Lim, H.-W. Ahn, V. Kornijcuk, G. Kim, J. Y. Seok, I. Kim, C. S. Hwang, D. S. Jeong, *Nanoscale* **2016**, *8*, 9629.
- [77] D. Lee, M. Kwak, K. Moon, W. Choi, J. Park, J. Yoo, J. Song, S. Lim, C. Sung, W. Banerjee, H. Hwang, *Adv. Electron. Mater.* **2019**, *78*, 1800866.
- [78] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, B. DeSalvo, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2011**, pp. 4.4.1–4.4.4.
- [79] G. W. Burr, R. M. Shelby, S. Sidler, C. di Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, H. Hwang, *IEEE Trans. Electron Devices* **2015**, *62*, 3498.

- [80] D. Kuzum, R. G. D. Jeyasingh, S. Yu, H.-S. P. Wong, *IEEE Trans. Electron Devices* **2012**, *59*, 3489.
- [81] S. La Barbera, D. R. B. Ly, G. Navarro, N. Castellani, O. Cueto, G. Bourgeois, B. De Salvo, E. Nowak, D. Querlioz, E. Vianello, *Adv. Electron. Mater.* **2018**, *4*, 1800223.
- [82] M. D. Pickett, G. Medeiros-Ribeiro, R. S. Williams, *Nat. Mater.* **2013**, *12*, 114.
- [83] T. Driscoll, H.-T. Kim, B.-G. Chae, M. Di Ventra, D. N. Basov, *Appl. Phys. Lett.* **2009**, *95*, 043503.
- [84] J. Woo, D. Lee, Y. Koo, H. Hwang, *Microelectron. Eng.* **2017**, *182*, 42.
- [85] A. Mehonic, A. J. Kenyon, *Front. Neurosci.* **2016**, *10*, 57.
- [86] S. Park, J. Noh, M. Choo, A. M. Sheri, M. Chang, Y.-B. Kim, C. J. Kim, M. Jeon, B.-G. Lee, B. H. Lee, H. Hwang, *Nanotechnology* **2013**, *24*, 384009.
- [87] Y.-F. Wang, Y.-C. Lin, I.-T. Wang, T.-P. Lin, T. H. Hou, *Sci. Rep.* **2015**, *5*, 10150.
- [88] H. Wu, P. Yao, B. Gao, W. Wu, Q. Zhang, W. Zhang, N. Deng, D. Wu, H. S. P. Wong, S. Yu, H. Qian, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2018**, pp. 11.5.1–11.5.4.
- [89] W. Wu, H. Wu, B. Gao, P. Yao, X. Zhang, X. Peng, S. Yu, H. Qian, *IEEE Symp. VLSI Technol.*, IEEE, Piscataway, NJ **2018**, pp. 103–104.
- [90] B. Gao, Y. Bi, H.-Y. Chen, R. Liu, P. Huang, B. Chen, L. Liu, X. Liu, S. Yu, H.-S. P. Wong, J. Kang, *ACS Nano* **2014**, *8*, 6998.
- [91] M. Hu, C. E. Graves, C. Li, Y. Li, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Q. Xia, J. P. Strachan, *Adv. Mater.* **2018**, *30*, 1705914.
- [92] C. Du, W. Ma, T. Chang, P. Sheridan, W. D. Lu, *Adv. Funct. Mater.* **2015**, *25*, 4290.
- [93] S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W. D. Lu, *Nano Lett.* **2015**, *15*, 2203.
- [94] A. Chanthbouala, V. Garcia, R. O. Cherifi, K. Bouzehouane, S. Fusil, X. Moya, S. Xavier, H. Yamada, C. Deranlot, N. D. Mathur, M. Bibes, A. Barthélémy, J. Grollier, *Nat. Mater.* **2012**, *11*, 860.
- [95] C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Electron.* **2018**, *1*, 52.
- [96] Z. Wang, M. Rao, J.-W. Han, J. Zhang, P. Lin, Y. Li, C. Li, W. Song, S. Asapu, R. Midya, Y. Zhuo, H. Jiang, J. H. Yoon, N. K. Upadhyay, S. Joshi, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, Q. Qiu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Commun.* **2018**, *9*, 3208.
- [97] X. Zhang, W. Wang, Q. Liu, X. Zhao, J. Wei, R. Cao, Z. Yao, X. Zhu, F. Zhang, H. Lv, S. Long, M. Liu, *IEEE Electron Device Lett.* **2018**, *39*, 308.
- [98] S. La Barbera, D. Vuillaume, F. Alibart, *ACS Nano* **2015**, *9*, 941.
- [99] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Mater.* **2017**, *16*, 101.
- [100] M. Ziegler, R. Soni, T. Patelczyk, M. Ignatov, T. Bartsch, P. Meuffels, H. Kohlstedt, *Adv. Funct. Mater.* **2012**, *22*, 2744.
- [101] S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P.-Y. Chen, H. Yeon, S. Yu, J. Kim, *Nat. Mater.* **2018**, *17*, 335.
- [102] H. Mulaosmanovic, E. Chicca, M. Bertele, T. Mikolajick, S. Slesazeck, *Nanoscale* **2018**, *10*, 21755.
- [103] M. Jerry, P. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2017**, pp. 6.2.1–6.2.4.
- [104] A. Sengupta, K. Roy, *Int. Jt. Conf. Neural Networks*, IEEE, Piscataway, NJ **2015**, pp. 1–7.
- [105] M. Sharad, D. Fan, K. Roy, *Proc. 50th Annu. Des. Autom. Conf. – DAC'13*, ACM Press, New York, **2013**, p. 1.
- [106] A. Sengupta, S. H. Choday, Y. Kim, K. Roy, *Appl. Phys. Lett.* **2015**, *106*, 143701.
- [107] D. Fan, Y. Shim, A. Raghunathan, K. Roy, *IEEE Trans. Nanotechnol.* **2015**, *14*, 1013.
- [108] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, J. Grollier, *Nature* **2017**, *547*, 428.
- [109] A. F. Vincent, J. Larroque, W. S. Zhao, N. Ben Romdhane, O. Bichler, C. Gamrat, J.-O. Klein, S. Galdin-Retailleau, D. Querlioz, *IEEE Int. Symp. Circuits Syst.*, IEEE, Piscataway, NJ **2014**, pp. 1074–1077.
- [110] A. F. Vincent, J. Larroque, N. Locatelli, N. Ben Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J.-O. Klein, S. Galdin-Retailleau, D. Querlioz, *IEEE Trans. Biomed. Circuits Syst.* **2015**, *9*, 166.
- [111] A. Sengupta, Y. Shim, K. Roy, *IEEE Trans. Biomed. Circuits Syst.* **2016**, *10*, 1152.
- [112] J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S. Shin, K.-T. Lee, P. Solomon, K. Chan, W. Haensch, J. Rozen, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2018**, pp. 13.1.1–13.1.4.
- [113] E. J. Fuller, F. El Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, A. A. Talin, *Adv. Mater.* **2017**, *29*, 1604310.
- [114] M. T. Sharbati, Y. Du, J. Torres, N. D. Ardolino, M. Yun, F. Xiong, *Adv. Mater.* **2018**, *30*, 1802353.
- [115] J.-T. Yang, C. Ge, J.-Y. Du, H.-Y. Huang, M. He, C. Wang, H.-B. Lu, G.-Z. Yang, K.-J. Jin, *Adv. Mater.* **2018**, *30*, 1801548.
- [116] C.-S. Yang, D.-S. Shang, N. Liu, E. J. Fuller, S. Agrawal, A. A. Talin, Y.-Q. Li, B.-G. Shen, Y. Sun, *Adv. Funct. Mater.* **2018**, *28*, 1804170.
- [117] E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, A. A. Talin, *Science* **2019**, *364*, 570.
- [118] C. J. Wan, L. Q. Zhu, Y. H. Liu, P. Feng, Z. P. Liu, H. L. Cao, P. Xiao, Y. Shi, Q. Wan, *Adv. Mater.* **2016**, *28*, 3557.
- [119] C. Qian, L. Kong, J. Yang, Y. Gao, J. Sun, *Appl. Phys. Lett.* **2017**, *110*, 083302.
- [120] L. Wang, W. Liao, S. L. Wong, Z. G. Yu, S. Li, Y. Lim, X. Feng, W. C. Tan, X. Huang, L. Chen, L. Liu, J. Chen, X. Gong, C. Zhu, X. Liu, Y. Zhang, D. Chi, K. Ang, *Adv. Funct. Mater.* **2019**, *29*, 1901106.
- [121] Q. A. Vu, H. Kim, V. L. Nguyen, U. Y. Won, S. Adhikari, K. Kim, Y. H. Lee, W. J. Yu, *Adv. Mater.* **2017**, *29*, 1703363.
- [122] I. Sanchez Esqueda, X. Yan, C. Rutherglen, A. Kane, T. Cain, P. Marsh, Q. Liu, K. Galatsis, H. Wang, C. Zhou, *ACS Nano* **2018**, *12*, 7352.
- [123] H. Tian, Q. Guo, Y. Xie, H. Zhao, C. Li, J. J. Cha, F. Xia, H. Wang, *Adv. Mater.* **2016**, *28*, 4991.
- [124] C. Diorio, P. Hasler, A. Minch, C. A. Mead, *IEEE Trans. Electron Devices* **1996**, *43*, 1972.
- [125] F. Alibart, S. Pleutin, D. Guérin, C. Novembre, S. Lenfant, K. Lmimouni, C. Gamrat, D. Vuillaume, *Adv. Funct. Mater.* **2010**, *20*, 330.
- [126] W. Xu, S.-Y. Min, H. Hwang, T.-W. Lee, *Sci. Adv.* **2016**, *2*, e1501326.
- [127] Y. H. Liu, L. Q. Zhu, P. Feng, Y. Shi, Q. Wan, *Adv. Mater.* **2015**, *27*, 5599.
- [128] Y. Wang, Z. Lv, J. Chen, Z. Wang, Y. Zhou, L. Zhou, X. Chen, S.-T. Han, *Adv. Mater.* **2018**, *30*, 1802883.
- [129] X. Guo, F. M. Bayat, M. Bavandpour, M. Klachko, M. R. Mahmoodi, M. Prezioso, K. K. Likharev, D. B. Strukov, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2017**, pp. 6.5.1–6.5.4.
- [130] X. Guo, F. M. Bayat, M. Prezioso, Y. Chen, B. Nguyen, N. Do, D. B. Strukov, *IEEE Cust. Integr. Circuits Conf.*, IEEE, Piscataway, NJ **2017**, pp. 1–4.
- [131] M. R. Mahmoodi, D. Strukov, *Proc. 55th Annu. Des. Autom. Conf. – DAC'18*, ACM Press, New York, NY **2018**, pp. 1–6.
- [132] D. Ielmini, H.-S. P. Wong, *Nat. Electron.* **2018**, *1*, 333.
- [133] R. Islam, H. Li, P.-Y. Chen, W. Wan, H.-Y. Chen, B. Gao, H. Wu, S. Yu, K. Saraswat, H.-S. Philip Wong, *J. Phys. D: Appl. Phys.* **2019**, *52*, 113001.

- [134] S. Saighi, C. G. Mayr, T. Serrano-Gotarredona, H. Schmidt, G. Lecerf, J. Tomas, J. Grollier, S. Boyn, A. F. Vincent, D. Querlioz, S. La Barbera, F. Alibart, D. Vuillaume, O. Bichler, C. Gamrat, B. Linares-Barranco, *Front. Neurosci.* **2015**, 9, 1.
- [135] M. Le Gallo, D. Krebs, F. Zipoli, M. Salinga, A. Sebastian, *Adv. Electron. Mater.* **2018**, 4, 1700627.
- [136] G. W. Burr, M. J. BrightSky, A. Sebastian, H.-Y. Cheng, J.-Y. Wu, S. Kim, N. E. Sosa, N. Papandreou, H.-L. Lung, H. Pozidis, E. Eleftheriou, C. H. Lam, *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2016**, 6, 146.
- [137] S. Kim, N. Sosa, M. BrightSky, D. Mori, W. Kim, Y. Zhu, K. Suu, C. Lam, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2013**, pp. 30.7.1–30.7.4.
- [138] B. Gao, H. Zhang, B. Chen, L. Liu, X. Liu, R. Han, J. Kang, Z. Fang, H. Yu, B. Yu, D.-L. Kwong, *IEEE Electron Device Lett.* **2011**, 32, 276.
- [139] A. Prakash, S. Maikap, H.-C. Chiu, T.-C. Tien, C.-S. Lai, *Nanoscale Res. Lett.* **2014**, 9, 125.
- [140] A. A. Bessonov, M. N. Kirikova, D. I. Petukhov, M. Allen, T. Ryhänen, M. J. A. Bailey, *Nat. Mater.* **2015**, 14, 199.
- [141] Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H.-S. P. Wong, M. Lanza, *Nat. Electron.* **2018**, 1, 458.
- [142] A. J. Arnold, A. Razavieh, J. R. Nasr, D. S. Schulman, C. M. Eichfeld, S. Das, *ACS Nano* **2017**, 11, 3110.
- [143] J. J. Yang, Q. Xia, *Nat. Mater.* **2017**, 16, 396.
- [144] D. Bishop, P. M. Solomon, S. Kim, J. Tang, J. Tersoff, T. Todorov, M. Copel, J. Collins, K.-T. Lee, S. Shin, W. Haensch, J. Rozen, *Int. Conf. Solid State Device Mater.* **2018**, pp. 23–24.
- [145] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong, H. Qian, *Nat. Commun.* **2017**, 8, 15199.
- [146] F. Yuan, Z. Zhang, C. Liu, F. Zhou, H. M. Yau, W. Lu, X. Qiu, H.-S. P. Wong, J. Dai, Y. Chai, *ACS Nano* **2017**, 11, 4097.
- [147] C. D. Wright, Y. Liu, K. I. Kohary, M. M. Aziz, R. J. Hicken, *Adv. Mater.* **2011**, 23, 3408.
- [148] Q. Luo, X. Xu, T. Gong, H. Lv, D. Dong, H. Ma, P. Yuan, J. Gao, J. Liu, Z. Yu, J. Li, S. Long, Q. Liu, M. Liu, *IEEE Int. Electron Devices Meet.*, Piscataway, NJ **2017**, pp. 2.7.1–2.7.4.
- [149] K. J. Yoon, Y. Kim, C. S. Hwang, *Adv. Electron. Mater.* **2019**, 5, 1800914.
- [150] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, H.-S. P. Wong, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2012**, pp. 10.4.1–10.4.4.
- [151] B. Govoreanu, G. S. Kar, Y.-Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, M. Jurczak, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2011**, pp. 31.6.1–31.6.4.
- [152] S. Pi, C. Li, H. Jiang, W. Xia, H. Xin, J. J. Yang, Q. Xia, *Nat. Nanotechnol.* **2019**, 14, 35.
- [153] O. Golonzka, U. Arslan, P. Bai, M. Bohr, O. Baykan, Y. Chang, A. Chaudhari, A. Chen, N. Das, C. English, P. Jain, H. Kothari, B. Lin, J. Clarke, C. Connor, T. Ghani, F. Hamzaoglu, P. Hentges, C. Jezewski, I. Karpov, R. Kotlyar, M. Metz, J. Odonnell, D. Ouellette, J. Park, A. Pirkle, P. Quintero, D. Seghete, M. Sekhar, A. Sen Gupta, M. Seth, N. Strutt, C. Wiegand, H. J. Yoo, K. Fischer, *Symp. VLSI Technol.*, IEEE, Piscataway, NJ **2019**, pp. T230–T231.
- [154] M. N. Kozicki, M. Park, M. Mitkova, *IEEE Trans. Nanotechnol.* **2005**, 4, 331.
- [155] P. Jain, U. Arslan, M. Sekhar, B. C. Lin, L. Wei, T. Sahu, J. Alzate-vinasco, A. Vangapaty, M. Meterelliyozi, N. Strutt, A. B. Chen, P. Hentges, P. A. Quintero, C. Connor, O. Golonzka, K. Fischer, F. Hamzaoglu, *IEEE Int. Solid-State Circuits Conf.*, IEEE, Piscataway, NJ **2019**, pp. 212–214.
- [156] A. D. Kent, D. C. Worledge, *Nat. Nanotechnol.* **2015**, 10, 187.
- [157] O. Golonzka, J. G. Alzate, U. Arslan, M. Bohr, P. Bai, J. Brockman, B. Buford, C. Connor, N. Das, B. Doyle, T. Ghani, F. Hamzaoglu, P. Heil, P. Hentges, R. Jahan, D. Kencke, B. Lin, M. Lu, M. Mainuddin, M. Meterelliyozi, P. Nguyen, D. Nikonor, K. O'brien, J. Donnell, K. Oguz, D. Ouellette, J. Park, J. Pellegren, C. Puls, P. Quintero, T. Rahman, A. Romang, M. Sekhar, A. Selarka, M. Seth, A. J. Smith, A. K. Smith, L. Wei, C. Wiegand, Z. Zhang, K. Fischer, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2018**, pp. 18.1.1–18.1.4.
- [158] W. Zhang, B. Gao, J. Tang, X. Li, W. Wu, H. Qian, H. Wu, *Phys. Status Solidi RRL* **2019**, 1900204.
- [159] I. T. Wang, Y. C. Lin, Y. F. Wang, C. W. Hsu, T. H. Hou, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2015**, pp. 28.5.1–28.5.4.
- [160] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, H. Hwang, *Nanotechnology* **2011**, 22, 254023.
- [161] J. S. Najem, G. J. Taylor, R. J. Weiss, M. S. Hasan, G. Rose, C. D. Schuman, A. Belianinov, C. P. Collier, S. A. Sarles, *ACS Nano* **2018**, 12, 4702.
- [162] G. Rachmuth, H. Z. Shouval, M. F. Bear, C.-S. Poon, *Proc. Natl. Acad. Sci. USA* **2011**, 108, E1266.
- [163] T. V. P. Bliss, G. L. Collingridge, *Nature* **1993**, 367, 31.
- [164] G. Bi, M. Poo, *Annu. Rev. Neurosci.* **2001**, 24, 139.
- [165] R. L. Davis, Y. Zhong, *Neuron* **2017**, 95, 490.
- [166] G. Mongillo, O. Barak, M. Tsodyks, *Science* **2008**, 319, 1543.
- [167] O. Barak, M. Tsodyks, *Curr. Opin. Neurobiol.* **2014**, 25, 20.
- [168] Y. Mi, M. Katkov, M. Tsodyks, *Neuron* **2017**, 93, 323.
- [169] A. Citri, R. C. Malenka, *Neuropsychopharmacology* **2008**, 33, 18.
- [170] E. R. Kandel, B. Milner, L. R. Squire, *Neuron* **1998**, 20, 445.
- [171] R. Lamprecht, J. LeDoux, *Nat. Rev. Neurosci.* **2004**, 5, 45.
- [172] A. M. Thomson, *Trends Neurosci.* **2000**, 23, 305.
- [173] R. S. Zucker, W. G. Regehr, *Annu. Rev. Physiol.* **2002**, 64, 355.
- [174] L. F. Abbott, W. G. Regehr, *Nature* **2004**, 431, 796.
- [175] K. M. MacLeod, *Hear. Res.* **2011**, 279, 13.
- [176] W. Xu, H. Cho, Y.-H. Kim, Y.-T. Kim, C. Wolf, C.-G. Park, T.-W. Lee, *Adv. Mater.* **2016**, 28, 5916.
- [177] T. Chang, S.-H. Jo, W. Lu, *ACS Nano* **2011**, 5, 7669.
- [178] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, W. Lu, *Nano Lett.* **2010**, 10, 1297.
- [179] J. Yin, F. Zeng, Q. Wan, F. Li, Y. Sun, Y. Hu, J. Liu, G. Li, F. Pan, *Adv. Funct. Mater.* **2018**, 28, 1706927.
- [180] R. C. Malenka, M. F. Bear, *Neuron* **2004**, 44, 5.
- [181] T. V. P. Bliss, T. Lømo, *J. Physiol.* **1973**, 232, 357.
- [182] G. Bi, M. Poo, *J. Neurosci.* **1998**, 18, 10464.
- [183] H. Markram, J. Lübke, M. Frotscher, B. Sakmann, *Science* **1997**, 275, 213.
- [184] R. C. Malenka, *Science* **1999**, 285, 1870.
- [185] G.-Q. Bi, J. Rubin, *Trends Neurosci.* **2005**, 28, 222.
- [186] R. C. Gerkin, P.-M. Lau, D. W. Nauen, Y. T. Wang, G.-Q. Bi, *J. Neurophysiol.* **2007**, 97, 2851.
- [187] A. Banerjee, R. S. Larsen, B. D. Philpot, O. Paulsen, *Trends Neurosci.* **2016**, 39, 26.
- [188] M. A. Lynch, *Physiol. Rev.* **2004**, 84, 87.
- [189] M. Suri, O. Bichler, D. Querlioz, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, B. DeSalvo, *Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2012**, pp. 10.3.1–10.3.4.
- [190] D. B. Chklovskii, B. W. Mel, K. Svoboda, *Nature* **2004**, 431, 782.
- [191] S. H. Bennett, A. J. Kirby, G. T. Finnerty, *Neurosci. Biobehav. Rev.* **2018**, 88, 51.
- [192] K. G. Pratt, A. J. Watt, L. C. Griffith, S. B. Nelson, G. G. Turrigiano, *Neuron* **2003**, 39, 269.
- [193] P. Jourdain, K. Fukunaga, D. Muller, *J. Neurosci.* **2003**, 23, 10645.
- [194] G.-Y. Wu, H. T. Cline, *Science* **1998**, 279, 222.

- [195] T. Xu, X. Yu, A. J. Perlik, W. F. Tobin, J. A. Zweig, K. Tenant, T. Jones, Y. Zuo, *Nature* **2009**, *462*, 915.
- [196] M. B. Moser, M. Trommald, P. Andersen, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12673.
- [197] S. B. Hofer, T. D. Mrsic-Flogel, T. Bonhoeffer, M. Hübener, *Nature* **2009**, *457*, 313.
- [198] D. Debanne, Y. Ingelbert, M. Russier, *Curr. Opin. Neurobiol.* **2019**, *54*, 73.
- [199] J. H. Yoon, Z. Wang, K. M. Kim, H. Wu, V. Ravichandran, Q. Xia, C. S. Hwang, J. J. Yang, *Nat. Commun.* **2018**, *9*, 417.
- [200] Y. Kim, Y. J. Kwon, D. E. Kwon, K. J. Yoon, J. H. Yoon, S. Yoo, H. J. Kim, T. H. Park, J.-W. Han, K. M. Kim, C. S. Hwang, *Adv. Mater.* **2018**, *30*, 1704320.
- [201] P. Dayan, L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, Cambridge, MA **2005**.
- [202] W. C. Abraham, *Nat. Rev. Neurosci.* **2008**, *9*, 387.
- [203] S. Fusi, *arXiv preprint arXiv:1706.04946*, **2017**, *1*.
- [204] M. K. Benna, S. Fusi, *Nat. Neurosci.* **2016**, *19*, 1697.
- [205] S. Fusi, P. J. Drew, L. F. Abbott, *Neuron* **2005**, *45*, 599.
- [206] B. Liu, Z. Liu, I.-S. Chiu, M. Di, Y. Wu, J.-C. Wang, T.-H. Hou, C.-S. Lai, *ACS Appl. Mater. Interfaces* **2018**, *10*, 20237.
- [207] Z.-H. Tan, R. Yang, K. Terabe, X.-B. Yin, X.-D. Zhang, X. Guo, *Adv. Mater.* **2016**, *28*, 377.
- [208] B.-Y. Kim, H.-G. Hwang, J.-U. Woo, W.-H. Lee, T.-H. Lee, C.-Y. Kang, S. Nahm, *NPG Asia Mater.* **2017**, *9*, e381.
- [209] Q. Wu, H. Wang, Q. Luo, W. Banerjee, J. Cao, X. Zhang, F. Wu, Q. Liu, L. Li, M. Liu, *Nanoscale* **2018**, *10*, 5875.
- [210] X. Zhu, C. Du, Y. Jeong, W. D. Lu, *Nanoscale* **2017**, *9*, 45.
- [211] L. Litman, L. Davachi, *Learn. Mem.* **2008**, *15*, 711.
- [212] J. Nishiyama, R. Yasuda, *Neuron* **2015**, *87*, 63.
- [213] L. A. Atherton, D. Dupret, J. R. Mellor, *Trends Neurosci.* **2015**, *38*, 560.
- [214] R. L. Redondo, R. G. M. Morris, *Nat. Rev. Neurosci.* **2011**, *12*, 17.
- [215] Jerry w. Rudy, *The Neurobiology of Learning and Memory*, Oxford University Press, Oxford **2013**.
- [216] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, *Nat. Mater.* **2011**, *10*, 591.
- [217] W. Wang, M. Wang, E. Ambrosi, A. Bricalli, M. Laudato, Z. Sun, X. Chen, D. Ielmini, *Nat. Commun.* **2019**, *10*, 81.
- [218] W. Wang, A. Bricalli, M. Laudato, E. Ambrosi, E. Covi, D. Ielmini, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2018**, pp. 40.3.1–40.3.4.
- [219] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, R. Hadsell, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3521.
- [220] E. J. Shoben, *Semantic and Episodic Memory*, Lawrence Erlbaum, Hillsdale, NJ **1984**.
- [221] P. Miyake, A. Shah, *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge University Press, Cambridge **1999**.
- [222] F. Attnave, M. B., D. O. Hebb, *Am. J. Psychol.* **1950**, *63*, 633.
- [223] C. C. Law, L. N. Cooper, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7797.
- [224] V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, D. Ielmini, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2016**, pp. 16.8.1–16.8.4.
- [225] Z. Xiao, J. Huang, *Adv. Electron. Mater.* **2016**, *2*, 1600100.
- [226] H. Markram, W. Gerstner, P. J. Sjöström, *Front. Synaptic Neurosci.* **2012**, *4*, 2.
- [227] S. Song, K. D. Miller, L. F. Abbott, *Nat. Neurosci.* **2000**, *3*, 919.
- [228] P. J. Drew, L. F. Abbott, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8876.
- [229] Z. Brzozko, W. Schultz, O. Paulsen, *Elife* **2015**, *4*, e09685.
- [230] A. Suvrathan, *Curr. Opin. Neurobiol.* **2019**, *54*, 12.
- [231] R. C. Froemke, D. Debanne, G.-Q. Bi, *Front. Synaptic Neurosci.* **2010**, *2*, 19.
- [232] V. Pawlak, J. R. Wickens, A. Kirkwood, J. N. D. Kerr, *Front. Synaptic Neurosci.* **2010**, *2*, 146.
- [233] N. Frémaux, W. Gerstner, *Front. Neural Circuits* **2016**, *9*, 85.
- [234] H. Shouval, S. S.-H. Wang, G. M. Wittenberg, *Front. Comput. Neurosci.* **2010**, *4*, 19.
- [235] Y. Wu, S. Yu, H.-S. P. Wong, Y.-S. Chen, H.-Y. Lee, S.-M. Wang, P.-Y. Gu, F. Chen, M.-J. Tsai, *IEEE Int. Mem. Work.*, IEEE, Piscataway, NJ **2012**, pp. 1–4.
- [236] U. Ganguly, B. Das, S. Lashkare, N. Panwar, P. Kumbhare, *IEEE Electron Device Lett.* **2017**, *38*, 1212.
- [237] N. Panwar, B. Rajendran, U. Ganguly, *IEEE Electron Device Lett.* **2017**, *38*, 740.
- [238] R. Yang, H.-M. Huang, Q.-H. Hong, X.-B. Yin, Z.-H. Tan, T. Shi, Y.-X. Zhou, X.-S. Miao, X.-P. Wang, S.-B. Mi, C.-L. Jia, X. Guo, *Adv. Funct. Mater.* **2018**, *28*, 1704455.
- [239] T. Keck, M. Hübener, T. Bonhoeffer, *Curr. Opin. Neurobiol.* **2017**, *43*, 87.
- [240] G. Turrigiano, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a005736.
- [241] J. Li, E. Park, L. R. Zhong, L. Chen, *Curr. Opin. Neurobiol.* **2019**, *54*, 44.
- [242] I. Delvendahl, M. Müller, *Curr. Opin. Neurobiol.* **2019**, *54*, 155.
- [243] G. G. Turrigiano, S. B. Nelson, *Nat. Rev. Neurosci.* **2004**, *5*, 97.
- [244] P. Gkoupidenis, D. A. Koutsouras, G. G. Malliaras, *Nat. Commun.* **2017**, *8*, 15448.
- [245] R. W. Semon, *The Mneme*, George Allen & Unwin., London **1921**.
- [246] E. Tulving, *Elements of Episodic Memory*, Oxford University Press, New York **1983**.
- [247] J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554.
- [248] R. L. Ressler, S. Maren, *Curr. Opin. Neurobiol.* **2019**, *54*, 54.
- [249] S. Tonegawa, M. Pignatelli, D. S. Roy, T. J. Ryan, *Curr. Opin. Neurobiol.* **2015**, *35*, 101.
- [250] S. A. Josselyn, P. W. Frankland, *Annu. Rev. Neurosci.* **2018**, *41*, 389.
- [251] J. Lisman, K. Cooper, M. Sehgal, A. J. Silva, *Nat. Neurosci.* **2018**, *21*, 309.
- [252] D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, S. E. Flores, I. Kim, Y. Sano, M. Zhou, K. Baumgaertel, A. Lavi, M. Kamata, M. Tuszyński, M. Mayford, P. Golshani, A. J. Silva, *Nature* **2016**, *534*, 115.
- [253] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, *Nature* **2015**, *518*, 529.
- [254] D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick, *Neuron* **2017**, *95*, 245.
- [255] D. Kumaran, D. Hassabis, J. L. McClelland, *Trends Cognit. Sci.* **2016**, *20*, 512.
- [256] Lisa, Memory and the manipulations thereof, <http://overthebrainbow.com/blog/2016/5/25/memory-and-how-it-can-be-manipulated> (accessed: August 2019).
- [257] Y. Ramiro-Cortes, A. F. Hobbiss, I. Israely, *Philos. Trans. R. Soc., B* **2013**, *369*, 20130157.
- [258] B. Babadi, L. F. Abbott, *PLoS Comput. Biol.* **2016**, *12*, e1004750.
- [259] K. D. Miller, D. J. C. MacKay, *Neural Comput.* **1994**, *6*, 100.
- [260] M. C. W. van Rossum, G. Q. Bi, G. G. Turrigiano, *J. Neurosci.* **2000**, *20*, 8812.
- [261] T. Salimans, D. P. Kingma, *arXiv preprint arXiv:1602.07868*, **2016**.
- [262] A. Nagpal, L1 and L2 Regularization Methods – Towards Data Science, <https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c> (accessed: August 2019).
- [263] M. Llera-Montero, J. Sacramento, R. P. Costa, *Curr. Opin. Neurobiol.* **2019**, *54*, 90.
- [264] Z. Yu, D. Kappel, R. Legenstein, S. Song, F. Chen, W. Maass, *arXiv preprint arXiv:1606.00157v1*, **2016**, *1*.

- [265] D. Kappel, S. Habenschuss, R. Legenstein, W. Maass, presented at *Advances in Neural Information Processing Systems*, Montreal, QC, Canada, December 2014.
- [266] P. R. Roelfsema, A. Holtmaat, *Nat. Rev. Neurosci.* **2018**, *19*, 166.
- [267] F. Nadim, D. Bucher, *Curr. Opin. Neurobiol.* **2014**, *29*, 48.
- [268] J. Palacios-Filardo, J. R. Mellor, *Curr. Opin. Neurobiol.* **2019**, *54*, 37.
- [269] E. M. Izhikevich, *Cereb. Cortex* **2007**, *17*, 2443.
- [270] L. Topolnik, O. Camiré, *Curr. Opin. Neurobiol.* **2019**, *54*, 98.
- [271] K. Lamsa, P. Lau, *Curr. Opin. Neurobiol.* **2019**, *54*, 20.
- [272] S. F. Owen, J. D. Berke, A. C. Kreitzer, *Cell* **2018**, *172*, 683.
- [273] M. Larkum, *Trends Neurosci.* **2013**, *36*, 141.
- [274] J. Guerguiev, T. P. Lillicrap, B. A. Richards, *Elife* **2017**, *6*, e22901.
- [275] Y. Yang, B. Chen, W. D. Lu, *Adv. Mater.* **2015**, *27*, 7720.
- [276] D. Rumelhart, D. Zipser, *Cogn. Sci.* **1985**, *9*, 75.
- [277] T. D. Sanger, *Neural Networks* **1989**, *2*, 459.
- [278] S. Choi, P. Sheridan, W. D. Lu, *Sci. Rep.* **2015**, *5*, 10492.
- [279] H. Li, K.-S. Li, C.-H. Lin, J.-L. Hsu, W.-C. Chiu, M.-C. Chen, T.-T. Wu, J. Sohn, S. B. Eryilmaz, J.-M. Shieh, W.-K. Yeh, H.-S. P. Wong, *IEEE Symp. VLSI Technol.*, IEEE, Piscataway, NJ **2016**, pp. 1–2.
- [280] S. Choi, J. H. Shin, J. Lee, P. Sheridan, W. D. Lu, *Nano Lett.* **2017**, *17*, 3113.
- [281] S. M. Bohte, J. N. Kok, H. La Poutré, *Neurocomputing* **2002**, *48*, 17.
- [282] R. V. Florian, *PLoS One* **2012**, *7*, e40233.
- [283] A. Kasiński, F. Ponulak, *Int. J. Appl. Math. Comput. Sci.* **2006**, *16*, 101.
- [284] A. Mohemmed, S. Schliebs, S. Matsuda, N. Kasabov, *Int. J. Neural Syst.* **2012**, *22*, 1250012.
- [285] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, *Nature* **2017**, *550*, 354.
- [286] M. Lanctot, D. Hassabis, T. Graepel, V. Panneershelvam, T. Lillicrap, J. Nham, I. Antonoglou, D. Silver, C. J. Maddison, A. Guez, I. Sutskever, A. Huang, J. Schrittwieser, N. Kalchbrenner, D. Grewe, G. van den Driessche, M. Leach, L. Sifre, K. Kavukcuoglu, S. Dieleman, *Nature* **2016**, *529*, 484.
- [287] Z. Wang, C. Li, W. Song, M. Rao, D. Belkin, Y. Li, P. Yan, H. Jiang, P. Lin, M. Hu, J. P. Strachan, N. Ge, M. Barnell, Q. Wu, A. G. Barto, Q. Qiu, R. S. Williams, Q. Xia, J. J. Yang, *Nat. Electron.* **2019**, *2*, 115.
- [288] M. Cheng, L. Xia, Z. Zhu, Y. Cai, Y. Xie, Y. Wang, H. Yang, *Proc. 54th Annu. Des. Autom. Conf.* 2017, ACM Press, New York **2017**, pp. 1–6.
- [289] R. S. Sutton, A. G. Barto, *IEEE Trans. Neural Networks* **1998**, *9*, 1054.
- [290] M. C. Soriano, S. Ortín, L. Keuninckx, L. Appeltant, J. Danckaert, L. Pesquera, G. van der Sande, *IEEE Trans. Neural Networks Learn. Syst.* **2015**, *26*, 388.
- [291] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, P. Bienstman, *Nat. Commun.* **2014**, *5*, 3541.
- [292] R. Midya, Z. Wang, S. Asapu, X. Zhang, M. Rao, W. Song, Y. Zhuo, N. Upadhyay, Q. Xia, J. J. Yang, *Adv. Intell. Syst.* **2019**, <https://doi.org/10.1002/aisy.201900084>.
- [293] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, T. Lillicrap, *arXiv preprint arXiv:1605.06065*, **2016**.
- [294] H. Li, T. F. Wu, A. Rahimi, K.-S. Li, M. Rusch, C.-H. Lin, J.-L. Hsu, M. M. Sabry, S. B. Eryilmaz, J. Sohn, W.-C. Chiu, M.-C. Chen, T.-T. Wu, J.-M. Shieh, W.-K. Yeh, J. M. Rabaey, S. Mitra, H.-S. P. Wong, *IEEE Int. Electron Devices Meet.*, IEEE, Piscataway, NJ **2016**, pp. 16.1.1–16.1.4.
- [295] J. L. Kuo, H. W. Chen, E. R. Hsieh, S. S. Chung, T. P. Chen, S. A. Huang, J. Chen, O. Cheng, *IEEE Symp. VLSI Technol.*, IEEE, Piscataway, NJ **2018**, pp. 29–30.
- [296] T. F. Wu, H. Li, P.-C. Huang, A. Rahimi, J. M. Rabaey, H.-S. P. Wong, M. M. Shulaker, S. Mitra, *Int. Solid – State Circuits Conf.*, IEEE, Piscataway, NJ **2018**, pp. 492–494.
- [297] H. Zeng, J. R. Sanes, *Nat. Rev. Neurosci.* **2017**, *18*, 530.
- [298] J. S. Isaacson, M. Scanziani, *Neuron* **2011**, *72*, 231.
- [299] Z. Jonke, S. Habenschuss, W. Maass, *Front. Neurosci.* **2016**, *10*, 118.
- [300] Z. Jonke, R. Legenstein, S. Habenschuss, W. Maass, *J. Neurosci.* **2017**, *37*, 8511.
- [301] M. Ignatov, M. Ziegler, M. Hansen, H. Kohlstedt, *Sci. Adv.* **2017**, *3*, e1700849.
- [302] M. Romera, P. Talatchian, S. Tsunagi, F. Abreu Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, J. Grollier, *Nature* **2018**, *563*, 230.
- [303] B. W. Mel, J. Schiller, P. Poirazi, *Curr. Opin. Neurobiol.* **2017**, *43*, 177.
- [304] B. A. Richards, T. P. Lillicrap, *Curr. Opin. Neurobiol.* **2019**, *54*, 28.
- [305] M. E. Sheffield, D. A. Dombeck, *Curr. Opin. Neurobiol.* **2019**, *54*, 1.
- [306] X. Li, H. Wu, Q. Zhang, B. Gao, W. Wu, W. Zhang, N. Deng, J. Tang, J. J. Yang, S. Song, L. Deng, Y. Xie, H. Qian, under review **2019**.
- [307] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, *Faraday Discuss.* **2019**, *213*, 453.
- [308] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, J. Dean, *arXiv preprint arXiv:1701.06538*, **2017**.
- [309] G. Marcus, *arXiv preprint arXiv:1801.00631*, **2018**.
- [310] J. Z. Tsien, *Trends Neurosci.* **2015**, *38*, 669.