Advances in Memristor-Based Neural Networks

Weilin Xu [1,2,3*†], Jingjuan Wang [1†] and Xiaobing Yan [1,4*]

[1] Key Laboratory of Brain-Like Neuromorphic Devices and Systems of Hebei Province, College of Electron and Information Engineering, Hebei University, Baoding, China, [2] Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin, China, [3] Electrical and Computer Engineering Department, Southern Illinois University Carbondale, Carbondale, IL, United States, [4] Department of Materials Science and Engineering, National University of Singapore, Singapore, Singapore

The rapid development of artificial intelligence (AI), big data analytics, cloud computing, and Internet of Things applications expect the emerging memristor devices and their hardware systems to solve massive data calculation with low power consumption and small chip area. This paper provides an overview of memristor device characteristics, models, synapse circuits, and neural network applications, especially for artificial neural networks and spiking neural networks. It also provides research summaries, comparisons, limitations, challenges, and future work opportunities.

Keywords: memristor, integrated circuit, artificial neural network, spiking neural network, artificial intelligence

INTRODUCTION

Resistance, capacitance and inductance are the three basic circuit components in passive circuit theory. In 1971, Professor Leon O. Chua of the University of California at Berkeley first described a basic circuit that relates flux to charge, called the missing fourth memristor element, and was successfully found by a team led by Stanley Williams at HP Labs in 2008 (Chua, 1971; Strukov et al., 2008). As a non-linear two-terminal passive electrical component, studies have shown that the conductance of a memristor is tunable by adjusting the amplitude, direction, or duration of its terminal voltages. Memristors have shown various outstanding properties, such as good compatibility with CMOS technology, small device area for high-density on-chip integration, non-volatility, fast speed, low power dissipation, and high scalability (Lee et al., 2008; Waser et al., 2009; Akinaga and Shima, 2010; Wong et al., 2012; Yang et al., 2013; Choi et al., 2014; Sun et al., 2020; Wang et al., 2020; Zhang et al., 2020). Thus, although memristors took many years to transform from a purely theoretical derivation into a feasible implementation, these devices have been widely used in applications such as machine learning and neuromorphic computing, as well as non-volatile random-access memory (Alibart et al., 2013; Liu et al., 2013; Sarwar et al., 2013; Fackenthal et al., 2014; Prezioso et al., 2015; Midya et al., 2017; Yan et al., 2017, 2019b,d; Ambrogio et al., 2018; Krestinskaya et al., 2018; Li C. et al., 2018, Li et al., 2019; Wang et al., 2018a, 2019a,b; Upadhyay et al., 2020). Furthermore, thanks to its powerful computing and storage capability, a memristor is a promising device for processing tremendous data and increasing the data processing efficiency in neural networks for artificial intelligence (AI) applications (Jeong and Shi, 2018).

This article intends to analyze the memristor theory, models, circuits, and important applications in neural networks. The contents of this paper are organized as follows. Section Memristor Characteristics and Models introduces the memristor theory and models. Section Memristor-Based Neural Networks presents its applications in the second-generation neural networks, namely artificial neural networks (ANNs) and the third-generation neural networks, namely spiking neural networks (SNNs). Section Summary is the conclusions and future research direction.

MEMRISTOR CHARACTERISTICS AND MODELS

The relationship between the physical quantities (namely charge q, voltage v, flux φ, and current i) and basic circuit elements (namely resistor R, capacitor C, inductor L, and memristor M) is shown in **Figure 1A** (Chua, 1971). Specifically, C defined as a linear relationship between voltage v and electric charge q ($C = dq/dv$), L is defined as a relationship between magnetic flux φ and current i ($L = d\varphi/di$), R is defined as a relationship between voltage v and current i ($R = dv/di$). The missing link between the electric charge and flux is defined as the memristor M and its differential equation is $M = d\varphi/dq$ or $G = dq/d\varphi$. **Figure 1B** shows the current-voltage characteristics of the memristor, where the pinched hysteresis loop is its fundamental identifier (Yan et al., 2018c). As a basic element, the memristor I–V curve cannot be obtained using R, C, and L. According to the shape of the pinched curve, it can be roughly classified into a digital type memristor or an analog type memristor. The resistance of a digital memristor exhibits an abrupt change at higher resistance ratios. The high-resistance and low-resistance states in a digital memristor have a long retention period, making it ideal for memory and logic operations. An analog memristor exhibits a gradual change in resistance. Therefore, it is more suitable for analog circuits and hardware-based multi-state neuromorphic system applications.

Memristor device technology and modeling research are the cornerstones of system applications. As shown in **Figure 2**, top-level system applications (brain-machine interface, face or picture recognition, autonomous driving, IoT edge computing, big data analytics, and cloud computing) are built on the device technology and modeling. Memristor-based analog, digital, and memory circuits play a key role in the link between device materials and system applications. The main usage for bi-stable memristors is binary switches, binary memory, and digital logic circuits, while multi-state memristors are used as multi-bit memories, reconfigurable analog circuits, and neuromorphic circuits.

Since the HP labs verified the nanoscale physical implementation, the physical behavior models of memristors have received a lot of attention. Accuracy, convergence, and
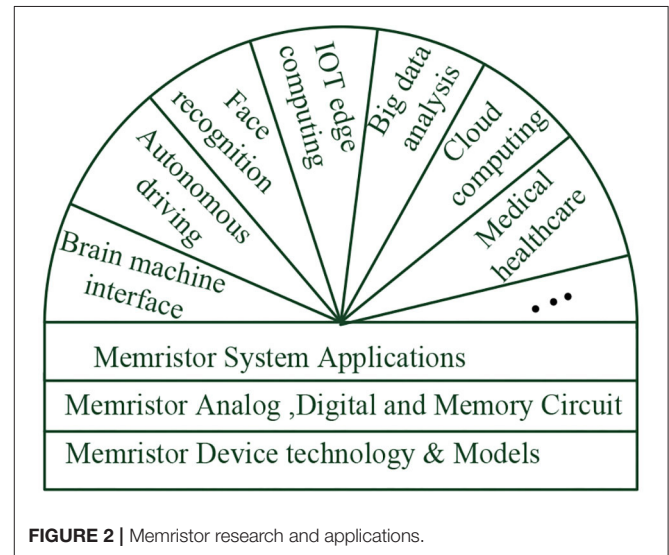
FIGURE 2 | Memristor research and applications.
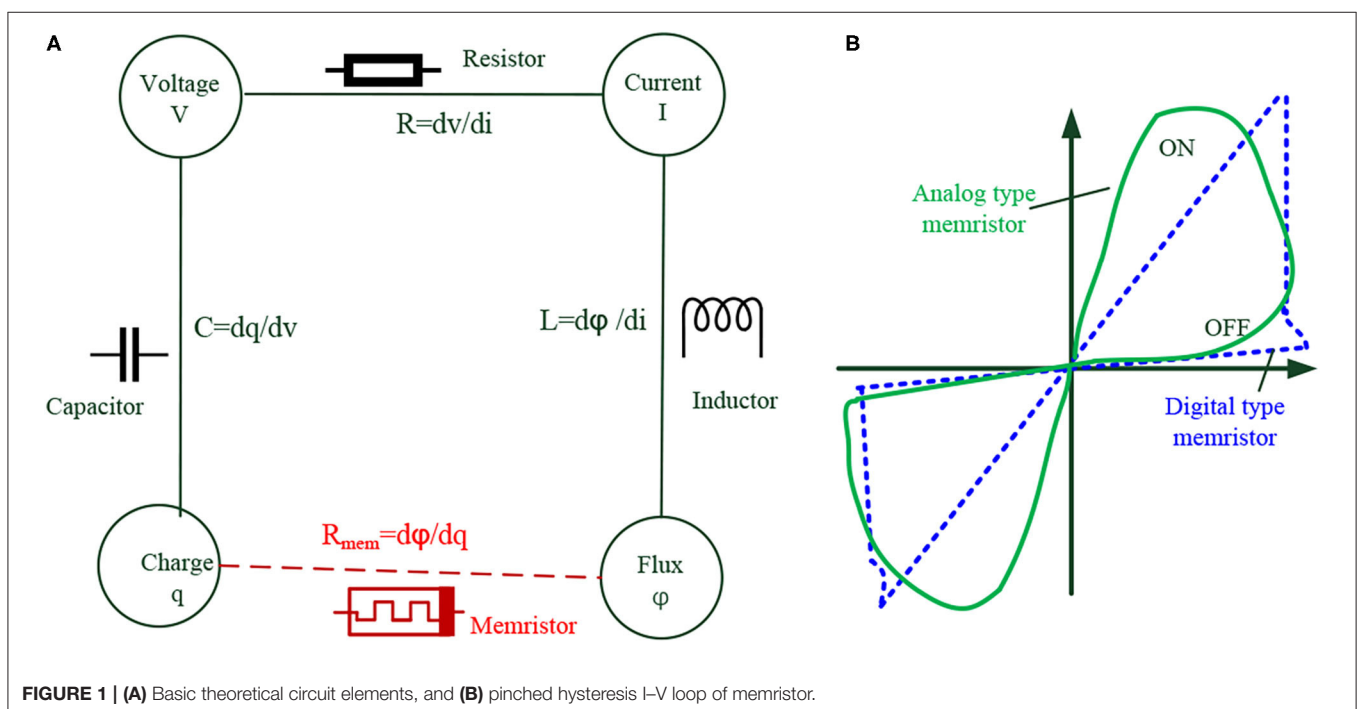
FIGURE 1 | (A) Basic theoretical circuit elements, and **(B)** pinched hysteresis I–V loop of memristor.

TABLE 1 | Classic memristor models.

| Models	Linear ion drift	Non-linear ion drift	Simmons tunnel barrier	TEAM
I–V characteristic	$v(t) = \left(R_{on}\frac{w(t)}{D} + R_{off}\left(1 - \frac{w(t)}{D}\right)\right)i(t)$	$i(t) = w(t)^n\beta\sinh(\alpha v(t)) + \chi[\exp(\gamma v(t)) - 1]$	$v(t) = \left(R_{on} + \frac{R_{off}-R_{on}}{W_{off}-W_{on}}(w - w_{on})\right)i(t)$	$v(t) = R_{on} \cdot \frac{\lambda}{e^{w_{off}-w_{on}}}(w - w_{on})$
State variable $\frac{dw(t)}{dt}$	$\left(u_v\frac{R_{on}}{D}\right)i(t)$	$a \cdot f(w)\,\boldsymbol{v(t)}^m$	$\begin{cases} c_{off}-\sinh\left(\frac{i}{i_{off}}\right)\exp\left[-\exp\left(\frac{w-a_{off}}{w_c} - \frac{|i|}{b}\right) - \frac{w}{w_c}\right], \\ \qquad i > 0 \\ c_{off}-\sinh\left(\frac{i}{i_{off}}\right)\exp\left[-\exp\left(\frac{w-a_{off}}{w_c} - \frac{|i|}{b}\right) - \frac{w}{w_c}\right], \\ \qquad i < 0 \end{cases}$	$\begin{cases} k_{off}\left(\frac{i(t)}{i_{off}} - 1\right)^{a_{off}}, & 0 < i_{off} < i \\ k_{on}\left(\frac{i(t)}{i_{on}} - 1\right)^{a_{on}}, & 0 < i_{off} < i \\ 0, & otherwise \end{cases}$
Interval	$0 \leq w \leq D$	$0 \leq w \leq 1$	$a_{off} \leq w \leq a_{on}$	$a_{on} \leq w \leq a_{off}$
Control mechanism	Current controlled	Voltage-controlled	Current-controlled	Current-controlled
Accuracy	Lowest accuracy	Low accuracy	Highest accuracy	Sufficient accuracy
Threshold exists	No	No	Pracitcally exists	Yes
Linearity	linear	No-linear	No-linear	No-linear

computational efficiency are the most important factors in memristor models. These behavior models are expected to be simple, intuitive, better understood, and closed form. Up to date, various models have been developed, each with its unique advantages and shortcomings. The models listed in **Table 1** are the most popular models, including a linear ion drift memristor model, a non-linear ion drift memristor model, a Simmons tunnel barrier memristor model, a threshold adaptive memristor model (TEAM) (Simmons, 1963; Strukov et al., 2008; Biolek et al., 2009; Pickett et al., 2009; Kvatinsky et al., 2012). In the linear ion drift memristor model, D and u_v represent the full length and device mobility of a memristor film, respectively. ω(t) is a dynamic state variable whose value is limited between 0 and D, taking into account the size of the physical device. The low turn-on resistance R_{on} is the full doped resistance when dynamic variable ω(t) is equal to D. The high turn-off resistance R_{off} is a fully undoped resistance when ω(t) is equal to 0. Besides, a window function multiplied by a state variable is needed to nullify the derivative and provide a non-linear transition for the physical boundary simulation. Several window functions have been presented for modeling memristors such as Biolek, Strukov, Joglekar, and Prodromakis window functions (Strukov et al., 2008; Biolek et al., 2009; Joglekar and Wolf, 2009; Strukov and Williams, 2009; Prodromakis et al., 2011). As the first memristor model, the linear ion drift model shows the features of simple, intuitive, and better understood. However, the state variable ω modulation in nano-scale devices is not a linear process, and the memristor experimental results show non-linear I–V characteristics. The non-linear ion drift model provides a better description of non-linear ionic transport and higher accuracy by experimentally fitting the parameters n, β, α, and χ (Biolek et al., 2009). But more physical reaction kinetics still need to be considered. The Simmons tunnel barrier model consists of a resistor in series with an electron tunnel barrier, which provides a more detailed representation of non-linear and asymmetrical features (Simmons, 1963; Pickett et al., 2009). There are nine fitting parameters in this segmentation model, which makes the mathematical model very complicated and computationally inefficient. The TEAM model can be thought of as a simplified version of the Simmons tunnel barrier model (Kvatinsky et al., 2012). However, all of the above models suffer from smoothing problems or mathematical ill-posedness issues, and they cannot provide robust and predictable simulation results in DC, AC, transient analysis, not to mention complicated circuit analysis such as noise analysis and periodic steady-state analysis (Wang and Roychowdhury, 2016). Therefore, in the face of transistor-level circuit design simulation, circuit designers usually have to replace the actual memristor with an emulator (Yang et al., 2019). The emulator is a complex CMOS circuit used to simulate some performance aspect of a special memristor. An emulator is not a true model, and it is very different from the real memristor model (Yang et al., 2014). Thus, it is urgent to establish a complete memristor model. Correct bias definition and right physical characteristics in SPICE or Verilog-a model are important for complex memristor circuit design. Otherwise, non-physical predictions will confuse circuit engineers in physical chip design.

MEMRISTOR-BASED NEURAL NETWORKS

Neuron Biological Mechanisms and Memristive Synapse

The human brain can solve complex tasks, such as image recognition and data classification, more efficiently than traditional computers. The reason why a brain excels in complicated functions is the large number of neurons and synapses that process information in parallel. As shown in **Figure 3**, when an electrical signal is transmitted between two neurons via axon and synapse, the joint strength or weight is adjusted by the synapse. There are approximately 100 billion neurons in an entire human brain, each with about 10,000 synapses. Pre-synaptic and post-synaptic neurons transfer and receive the signal of excitatory and inhibitory
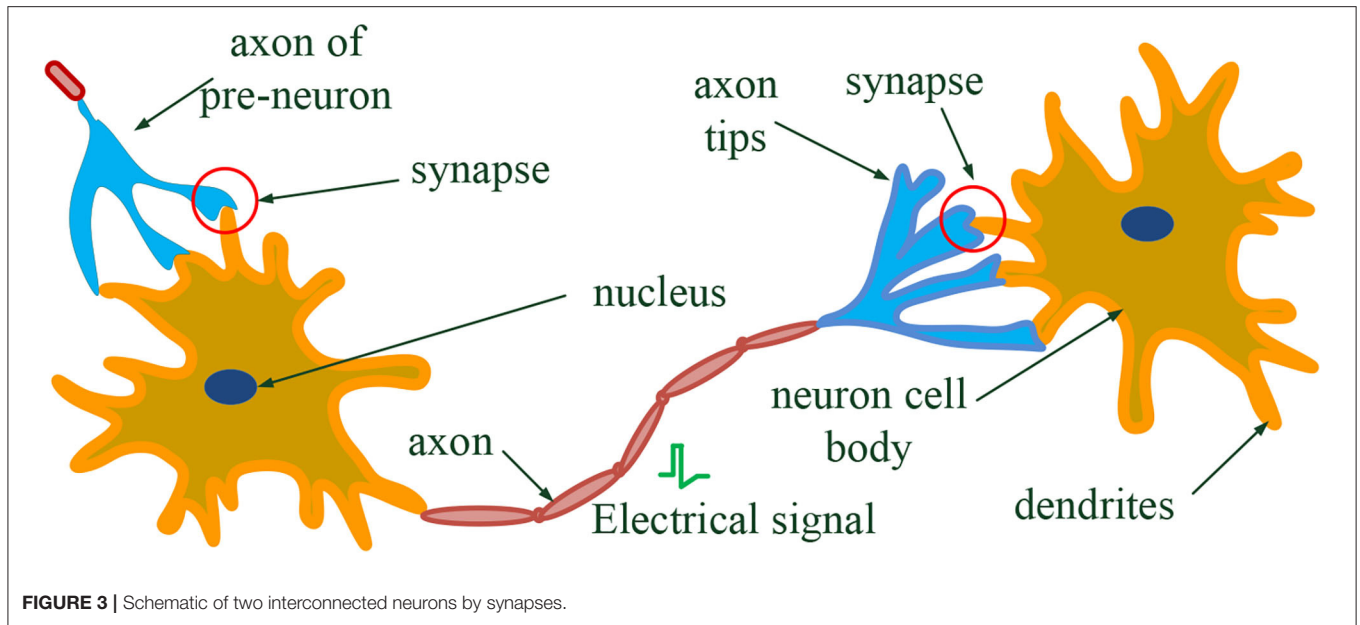
FIGURE 3 | Schematic of two interconnected neurons by synapses.

post-synaptic potentials by updating synaptic weights. Long-term potentiation (LTP) and long-term depression (LTD) are important mechanisms in a biological nervous system, which indicates a deep-rooted transformation in the connection strengths between neurons. According to the interval between pre-synaptic and post-synaptic action potentials or spikes, the phenomenon of synaptic weight modification is known as spike-timing-dependent plasticity (STDP) (Yan et al., 2018a, 2019c). Due to scalability, low power operation, non-volatile features, and small on-chip area, memristors are good candidates for artificial synaptic devices to mimicking the LTP, LTD, and STDP behaviors (Jo et al., 2010; Ohno et al., 2011; Kim et al., 2015; Wang et al., 2017; Yan et al., 2017).

There are some key requirements for memristive devices in neural network applications. For example, a wide range of resistance is required to enable sufficient resistance states; devices are required to have low resistance fluctuations and low device-to-device variability; a higher absolute resistance is required for low power dissipation; and high durability is required for reprogramming and training (Choi et al., 2018; Yan et al., 2018b, 2019a; Xia and Yang, 2019). A concern with device stability is resistance drift, which occurs over time or with the environment. Resistance drift causes undesirable changes in synapse weight and blurs different resistance states, ultimately affecting the accuracy of neural network computation (Xia and Yang, 2019). To deal with this drift challenge, improvements can be made in three aspects: (1) material device engineering, (2) circuit design, and (3) system design (Alibart et al., 2012; Choi et al., 2018; Jiang et al., 2018; Lastras-Montaño and Cheng, 2018; Yan et al., 2018b, 2019a; Zhao et al., 2020). For example, as for the domain of material engineering, threading dislocations can be used to control programming variation and enhance switching uniformity (Choi et al., 2018). In terms of circuit-level design, a module of two series memristors and a transistor with

the smallest size can be used, thus, the resistance ratio of the memristor can be encoded to compensate for the resistance drift (Lastras-Montaño and Cheng, 2018). For the system-design level, device deviation can be reduced by protocols, such as closed loop peripheral circuit with a write-verify function (Alibart et al., 2012). In order to obtain linear and symmetric weight update in LTP and LTD for efficient neural network training, optimized programming pulses can be used to excite memristors with either fixed-amplitude or fixed-width voltage pulses (Jiang et al., 2018; Zhao et al., 2020). Note it is inevitable to increase energy consumption if the memristor resistance value is changed through complex programmable pulses.

The comparison of different memristive synapse circuit structures is shown in **Table 2** (Kim et al., 2011a; Wang et al., 2014; Prezioso et al., 2015; Hong et al., 2019; Krestinskaya et al., 2019). Single memristor synapse (1M) crossbar arrays in neural networks have the lowest complexity and low power dissipation. However, it suffers from sneak path problems and complex peripheral switch circuits. Synapses with two memristors (2M) have a more flexible weight range and better symmetric LTP and LTD, but the corresponding chip area will be doubled. A synapse with one memristor and one transistor (1M-1T) has the advantage of solving the sneak path problem, but it also occupies a large area in the large-scale integration of neural networks. A bridge synapse architecture with four memristors (4M) provides a bidirectional programming mechanism with a voltage input voltage output. Due to the significant on-chip area overhead, the 1M-1T and 4M synapses may not be applicable for large-scale neural networks.

Memristor-Based ANNs

The basic operations of classical hardware ANNs include multiplication, addition, and activation, which are accomplished by CMOS circuits such as GPUs. The weights are typically saved

in SRAM or DRAM. Despite the scalability of CMOS circuits, they are still not enough for ANN applications. Furthermore, the SRAM cell size are too big to be integrated at high density. DRAM needs to be refreshed periodically to prevent data decay. Whether it is SRAM or DRAM, it often needs to interact with

CMOS cores. No matter SRAM or DRAM, the data needs to be fetched by to the cache and register files of the digital processors before processing and returned through the same databus, leading to significant speed limit and large energy consumption, which is the main challenge for deep learning and big data applications (Xia and Yang, 2019). Nowadays, ANNs feature for large number of computational parameters stored in memory compared to classical computation. For example, a two-layer 784-800-10 fully-connected deep neural network in the MNIST dataset has 635,200 interconnections. A state of the art keep neural network like Visual Geometry Group (VGG) has a few millions of parameters. These factors pose a huge challenge to the implementation of ANN hardware. The memristor's non-volatility, lower power consumption, lower parasitic capacitance, and reconfigureable resistance states, high speed, and adaptability lead to a key role in ANN applications (Xia and Yang, 2019). An ANN is an information processing model which are derived from mathematical optimization. A typical ANN architecture and its memristor crossbar are shown in **Figure 4**. The system usually consists of three layers: an input layer, a middle layer or a hidden layer, and an output layer. The connected units or nodes are neurons which are usually series by weighted-sum module and activation function module. Neurons also perform tasks of decoding, control, and signal routing. Due to its powerful signal processing capability, CMOS analog and digital logic circuits are the best candidates for neurons hardware implementation. In **Figure 4**, arrow or connecting lines represent synapses, and their weights represent the connection strengths between two neurons. Assume the weight modulation matrix W_{ij} in a memristor synapse crossbar is a M × N dimensional matrix, where i(i = 1, 2, …, N) and j(i = 1, 2, …, M) are the index numbers of the output and input ports of the memristor crossbar. W_{ij} between pre-neuron input vector X_j and post-neuron output vector Y_i is

TABLE 2 | Comparison of different structure memristive synapse circuit.

Synapses	Structure	Area(F2)	Weight	Weight range	Other features
1M		≈4	G	+	Lower power consumption; least complex; sneak path problem in neural network array
2M		≈8	G^+-G^-	+,0, −	Better symmetric between LTP and LTD; complex post-synaptic neurons
1M-1T		≈24	G	+	Solution for sneak path problem with transistor switch; biggest size; transistor non-ideal effect
4M		≈16	$\dfrac{M_2}{M_1+M_2} - \dfrac{M_4}{M_3+M_4}$	+,0, −	Voltage input voltage output; Bidirectional programming; bigger size
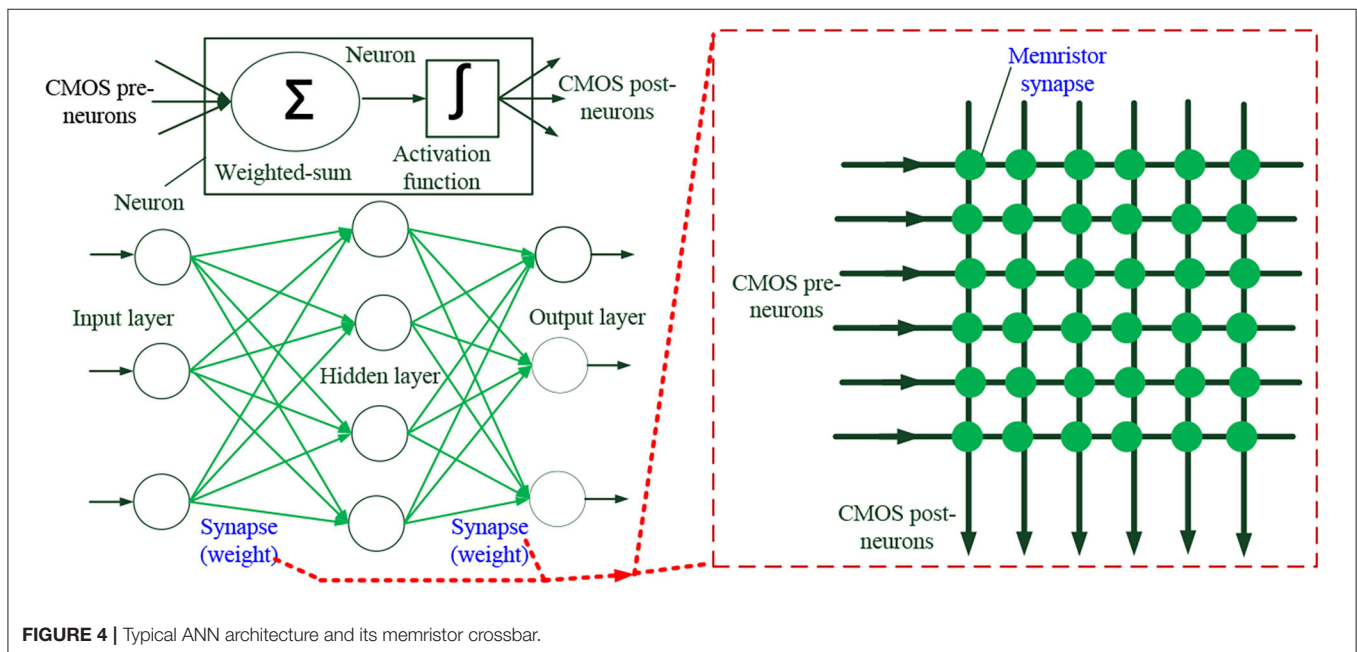
FIGURE 4 | Typical ANN architecture and its memristor crossbar.

TABLE 3 | Typical architectures of Memristive ANNs.

TYPES	Architecture	Layers properties	Applications	Challenges
SLP/MLP	Input layer+hidden layer+output layer	Sigmoid, tanh, etc., activation; Full-connections	Simple pattern classification; Hand-written letter recognition	Power dissipation in deep ANN; Overfitting; non-ideal memristor; Scalability
CNN	Input layer+ Convolution layer+ ReLu layer+Pooling +Fully-connected and output layer	Convolution; Pooling	Image classification; Face recognition; Video analysis	
CeNN	Cell array with templates; 1-D, 2-D, or 3-D	Dissipative non-linear cells; Lyapunov function; Neighborhood communication	Image filtering; Signal processing; moving object detection	Convergence and mulitistability in non-symmetric networks; non-ideal memristor
RNN	Fully recurrent; Elman; Jordan; gated recurrent unit; long short-term memory	Temporal dynamic behavior; directed graph along a temporal sequence; LSTM	Speech recognition; Machine translation; Video processing	Hard to train for long term dependencies; non-ideal memristor

TABLE 4 | ANNs learning accuracy improvement by mitigating memristor non-ideal effects.

Level	Strategies	Tradeoffs
Device materials	Optimizing redox reaction at the metal/oxide interface (Lee et al., 2015), Threading dislocations technology (Tanikawa et al., 2018), Heating element, selectively enhanced filament expansion stage (Jeong et al., 2015)	Manufacturing cost; Power consumption; On-chip area; Peripheral circuit complexity; Algorithm efficiency
Circuits	Hybrid CMOS-memristor Neuromorphic Synapse, 1R+1M1R for better device symmetry (Woo and Yu, 2018), Non-identical pulse excitation (Park et al., 2013; Chang et al., 2017), Bipolar-pulse-training (Li et al., 2016), Spike edge shape design (Li S. J. et al., 2018)	
Architectures	Multiple memristors cell for high redundancy (Chen et al., 2015), Pseudo-crossbar array, peripheral circuit compensation (Chen et al., 2015)	
Algorithms	co-optimization between memristors and ANN algorithms (Li et al., 2016)	

a matrix-vector multiplication operation, expressed as Equation (1) (Jeong and Shi, 2018).

$$Y_i = \Sigma \, W_{ij} \cdot X_j \qquad (1)$$

$$\Delta w_{ij} = r \left[\frac{\partial \left(y - y^*\right)^2}{\partial w_{ij}} \right] \qquad (2)$$

The matrix W can be continuously adjusted until the difference between the output value y and the target value y^* is minimized. The Equation (2) shows the synaptic weight tunning process with the gradient of output error $(y–y^*)^2$ under a training rate (Huang et al., 2018). Therefore, a memristor crossbar is equal to a CMOS adder plus a CMOS multiplier and an SRAM (Jeong and Shi,

2018), because data are computed, stored, and regenerated on the same local device (i.e., a memristor itself). Besides, a crossbar can be vertically integrated into three dimensions (Seok et al., 2014; Lin et al., 2020; Luo et al., 2020). In this way, it saves much chip area and power consumption. Due to the memristor synapse update and save weight data on itself, the memory wall problem with von Neumann bottleneck is solved.

Researchers have developed various topologies and learning algorithms for software-based or hardware-based ANNs. **Table 3** provides a comparison of typical memristive ANNs, including single-layer perceptron (SLP) or multi-layer perceptron (MLP), CNN, cellular neural network (CeNN), and recurrent neural network (RNN). SLP and MLP are classic neural networks with well-known learning rules such as Hebbian learning, backpropagation. Although a lot of ANN studies have been verified by simulations or small-scale implementation, a single-layer neural network with 128×64 1M-1T Ta/HfO$_2$ memristor array has been experimentally demonstrated with an image recognition accuracy of 89.9% for the MNIST dataset (Hu et al., 2018). CNNs (referred to as space-invariant or shift-invariant ANNs) are regularized versions of MLP. Their hidden layers usually contain multiple complex activation functions, and perform convolution or regional maximum value operations. Researchers have demonstrated an over 70% of accuracy in human behavior video recognition with a memristor-based 3D CNN (Liu et al., 2020). It should be emphasized that this verification is only a software simulation result, while the on-chip hardware demonstration is still very challenging, especially for deep CNNs (Wang et al., 2019a; Luo et al., 2020; Yao et al., 2020). CeNN is a massively parallel computing neural network, whose communication features are limited to between adjacent cell neurons. The cells are dissipative non-linear continuous-time or discrete-time processing units. Due to their dynamic processing capability and flexibility, CeNNs are promising candidates for real-time high frame rate processing or multi-target motion detection. For example, a CeNN with 4M memristive bridge circuit synapse has been proposed for image processing (Duan et al., 2014). Unlike classic feed forward ANNs, RNNs have a feedback connection that enables temporal dynamic behavior.

Therefore, it is suitable for speech recognition applications. Long short-term memory (LSTM) is a kind of useful RNN structure for deep learning. Hardware implementation of LSTM networks based on memristors have been reported (Smagulova et al., 2018; Li et al., 2019; Tsai et al., 2019; Wang et al., 2019a).

Due to atomic-level random defects and variability in the conductance modulation process, non-ideal memristor characteristics are the main causes of learning accuracy loss in ANNs. This phenomenon is manifested in the following aspects of memristor: asymmetric non-linear weight change between potentiation and depression, limited ON/OFF weight ratio and device variation. **Table 4** shows the main strategies for how to deal with these issues. One can mitigate the effects of non-ideal memristor characteristics on ANN accuracy from four levels: device materials, circuits, architectures, and algorithms. At device materials level, switching uniformity and analog on/off ratio can be enhanced by optimizing redox reaction at the metal/oxide interface, adopting threading dislocations technology or heating element (Jeong et al., 2015; Lee et al., 2015; Tanikawa et al., 2018). At circuits level, one can use customized excitation pulse or hybrid CMOS-memristor synapses to mitigate memristor non-ideal effects (Park et al., 2013; Li et al., 2016; Chang et al., 2017; Li S. J. et al., 2018; Woo and Yu, 2018). At architectures level, common techniques are multiple memristors cell for high redundancy, pseudo-crossbar array, and peripheral circuit compensation (Chen et al., 2015). Co-optimization between memristors and ANN algorithms is also reported (Li et al., 2016). However, it should be noted that implementation of these strategies inevitably brings side effects, such as high manufacturing cost, large power consumption, large chip area, complex peripheral circuits, or inefficient algorithm. For example, the non-identical pulse excitation or bipolar-pulse-training methods improve the linearity and symmetry of memristor synapses, but it increases the complexity of peripheral circuits, system power consumption, and chip area. Therefore, trade-offs and co-optimization need to be made at each design level to improve the learning accuracy of ANNs (Gi et al., 2018; Fu et al., 2019). **Figure 5** is a collaborative design example from bottom-level memristor devices to top-level training algorithms (Fu et al., 2019). The conductance response (CR) curve of memristors is first measured to obtain its non-linearity factor. Then, the CR curve is divided into piecewise linear segments to obtain their slope, and the pulse width of the excitation pulse is inversely proportional to the slope. These data are stored in memory for comparison and correction by memristor crossbars during the update. Through this method, the ANN recognition accuracy is finally improved.

The memristor-based ANN applications can be software, hardware or hybrid (Kozhevnikov and Krasilich, 2016). Software networks tend to be more accurate than their hardware counterparts because they do not have the analog element non-uniformity issues. However, hardware networks feature better speed and less power consumption due to non-von Neumann architectures (Kozhevnikov and Krasilich, 2016). In **Figure 6**, a deep neuromorphic accelerator ANN chip with 2.4 million Al_2O_3/TiO_2-xmemristors was designed and fabricated (Kataeva et al., 2019). This memristor chip consists of a 24×43 array with a 48×48 memristor crossbar at each intersection, which means its complexity is about 1,000 times higher than previous designs in the literature. This work is a good starting point for the operation of medium-scale memristor ANNs. Similar accelerators have appeared in the last 2 years (Cai et al., 2019; Chen W.-H. et al., 2019; Xue et al., 2020).
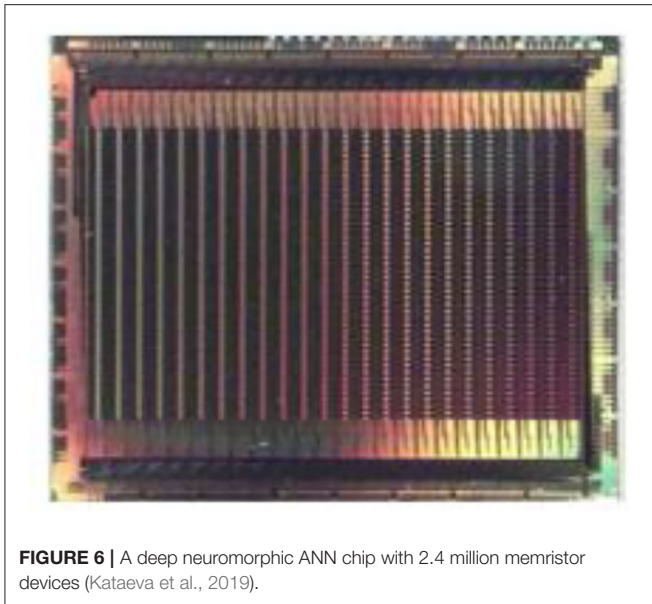
Memristive neural networks can be used to understand human emotion and simulate human operational abilities (Bishop, 1995). The well-known PavlTov associative memory experiment has been implemented in memristive ANNs with a novel weighted-input-feedback learning method (Ma et al., 2018). As more input signals, neurons, and memristor synapses, complex emotional processing will be achieved in further AI chips. Due to the material challenge and the lack of effective models, most of the demonstrations are limited to small-scale simulations for simple tasks. The shortcomings of memristors are mainly the non-linearity, asymmetry, and variability, which seriously affect the accuracy of ANNs. Moreover, the peripheral circuits and interface must provide superior energy efficiency and data throughput.

Memristor-Based SNN

Inspired by cognitive and computational methods of animal brains, the third-generation neural network, SNN, makes desirable properties of compact biological neurons mimic and remarkable cognitive performance. The most prominent feature of SNN is that it incorporates the concept of time into operations
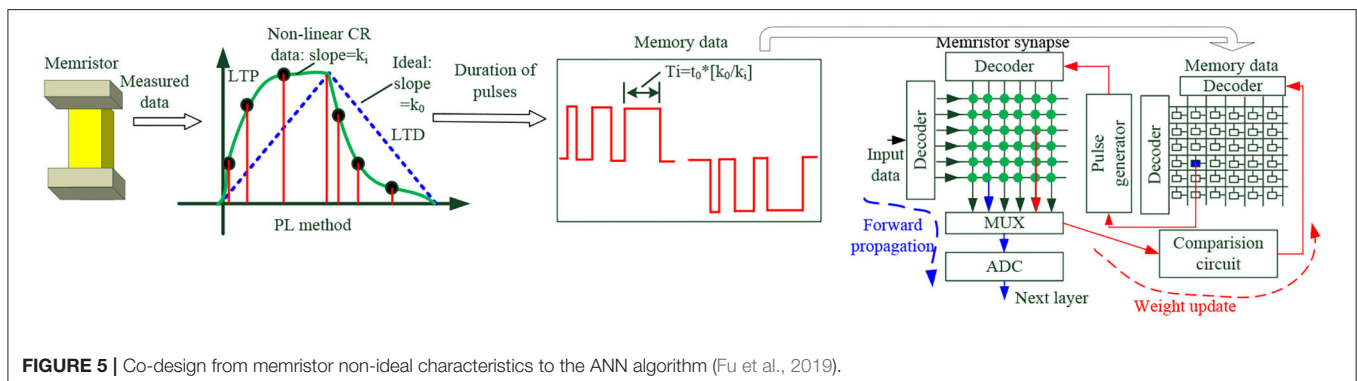
FIGURE 5 | Co-design from memristor non-ideal characteristics to the ANN algorithm (Fu et al., 2019).

FIGURE 6 | A deep neuromorphic ANN chip with 2.4 million memristor devices (Kataeva et al., 2019).

TABLE 5 | Comparison of several memristor-based SNNs.

References	Neuron	Synapse	Learning rules	Size	Applications
Zheng and Mazumder (2018)	LIF	1M1R fixed-polarity memristor	STDP; Supervised learning	784-300-10	Handwritten digits recognition
Chen B. et al. (2019)	LIF	Lithium silicate memristor	STDP, Unsupervised learning, WTA	128-128-12	Motion-style recognition
Shukla and Ganguly (2018)	LIF	HfO$_2$ memristor	STDP; Supervised Hebbian	16-3	Classification problems, Fisher Iris dataset, etc.
Wu and Saxena (2018)	LIF	Stochastic binary memristor	STDP, Dendritic-inspired processing	1-4	Pattern Recognition
Chu et al. (2014)	LIF	Pr$_{0.7}$ Ca$_{0.3}$MnO$_3$-memristor	STDP, Unsupervised learning	30-10	Visual Pattern Recognition
Volos et al. (2015)	H-R, FHN	Flux-controlled memristor	STDP	2	Chaotic oscillators; Neurodynamic behavior
Al-Shedivat et al. (2015)	SRM	Stochastic biolek's memristor model	STDP, WTA	1568-32	Handwritten digits recognition

with discrete values, while the input and output values of the second-generation ANNs are continuous. SNN can better leverage the strength of biological paradigm of information processing, thanks to the hardware emulation of synapses and neurons. ANN is calculated layer by layer, which is relatively simple. However, spike trains in SNN are relatively difficult to understand and efficient coding methods for these spike trains are not easy. These dynamic events driven spikes in SNN enhance the ability to process spatio-temporal or real-world sensory data, with fast adaptation and exponential memorization. The combination of spatio-temporal data allows SNN to process signals naturally and efficiently.

Neuron models, learning rules, and external stimulus coding are key research areas of SNN. The Hodgkin & Huxley (HH) model, leaky Integrate-and-Fire (LIF) model, spike response model (SRM), and Izhikevich model are the most common models of neurons (Hodgkin and Huxley, 1952; Chua, 2013; Ahmed et al., 2014; Pfeiffer and Pfeil, 2018; Wang and Yan, 2019; Zhao et al., 2019; Ojiugwo et al., 2020). The HH model is a continuous-time mathematical model based on conductance. Although this model is based on the study of squid, it is widely used in lower or higher organisms (even humans being). However, since complex non-linear differential equations are set with four variables, this model is difficult to achieve high accuracy. Chua established the memristor model of Hodgkin-Huxley neurons and proved that memristors can be applied to the imitation of complex neurobiology (Chua, 2013). The Izhikevich model integrates the bio-plasticity of HH model with simplicity and higher computational efficiency. The HH and Izhikevich models are calculated by differential equations, while the LIF and SRM models are computed by an integral method. SRM is an extended version of LIF, and the Izhikevich model can be considered as a simplified version of the Hodgkin-Huxley model. These mathematical models are the results of

different degrees of customization, trade-offs and biological neural network optimization. **Table 5** shows a comparison of several memristor-based SNNs. It can be seen that these SNN studies are based on STDP learning rules and LIF neurons. Most of them are still in simple pattern recognition applications, only a few of which have hardware implementations.

The salient features of SNNs are as follows. First, biological neuron models (e.g., HH, LIF) are closer to biological neurons than neurons of ANN. Second, the transmitted information is time or frequency encoded discrete-time spikes, which can contain more information than traditional networks. Third, each neuron can work alone and enter a low power standby mode when there is no input signal. Since SNNs have been proven to be more powerful than ANNs in theory, it is natural to widely use SNNs. Since the spike training cannot be differentiated, the gradient descent method cannot be used to train SNNs without losing accurate temporal information. Another problem is that it takes a lot of computation to simulate SNNs on normal hardware, because it requires analog differential equations (Ojiugwo et al., 2020). Due to the complexity of SNNs, efficient learning rules that meet the characteristics of biological neural networks have not been discovered. This rule is required to model not only synaptic connectivity but also its growth and attenuation. Another challenge is the discontinuous nature of spike sequence, which makes many classic ANN learning rules unsuitable for SNNs, or can only be approximated, because the convergence problem is very serious. Meanwhile, many SNNs
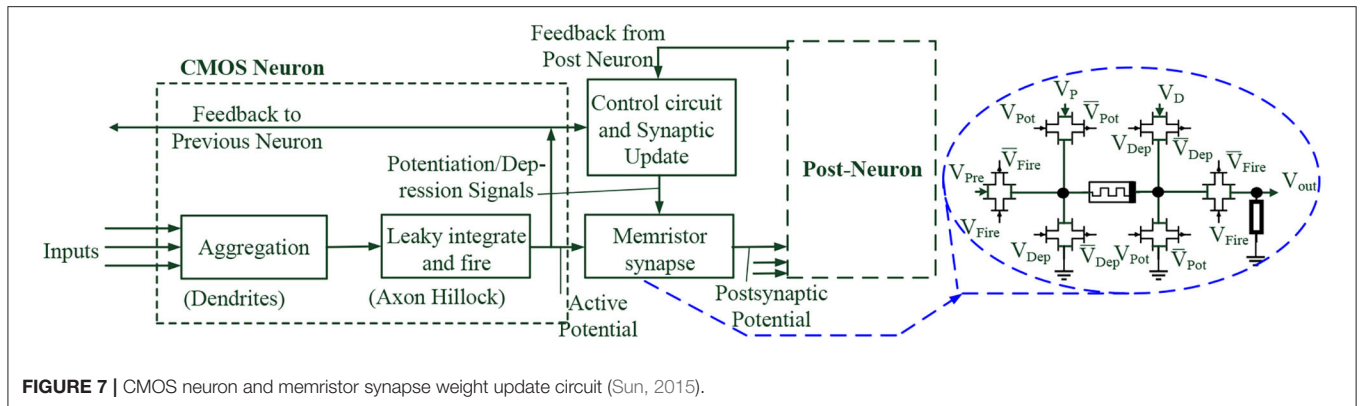
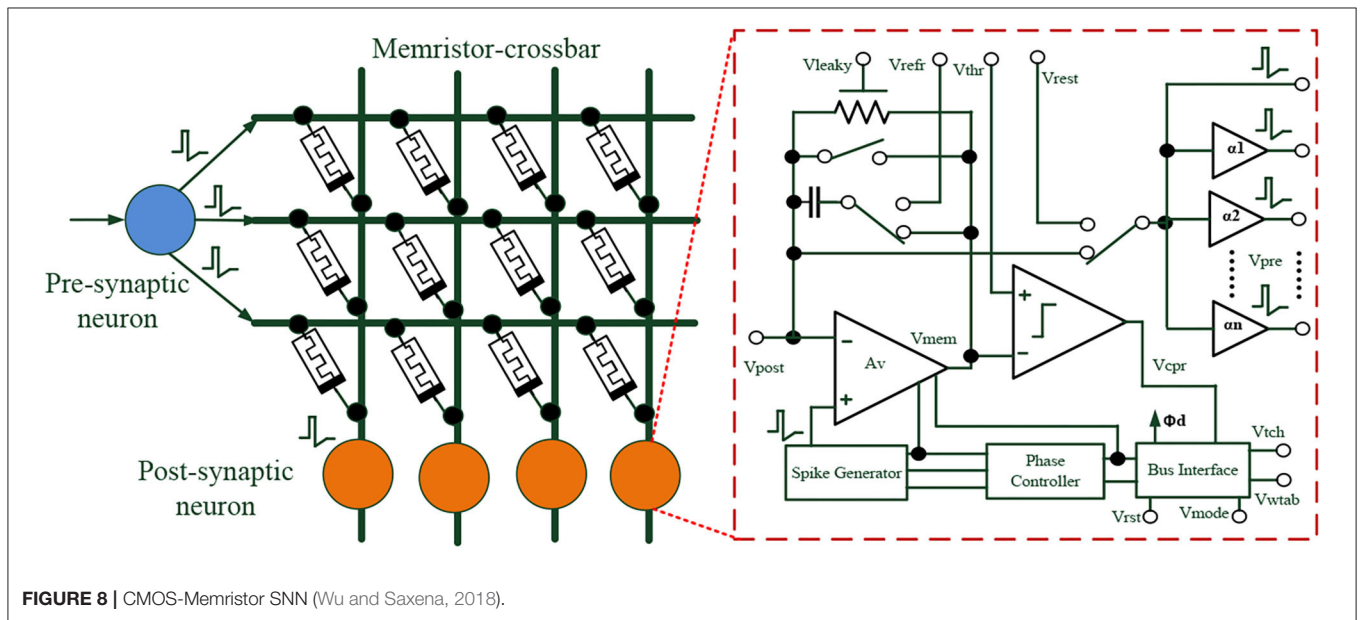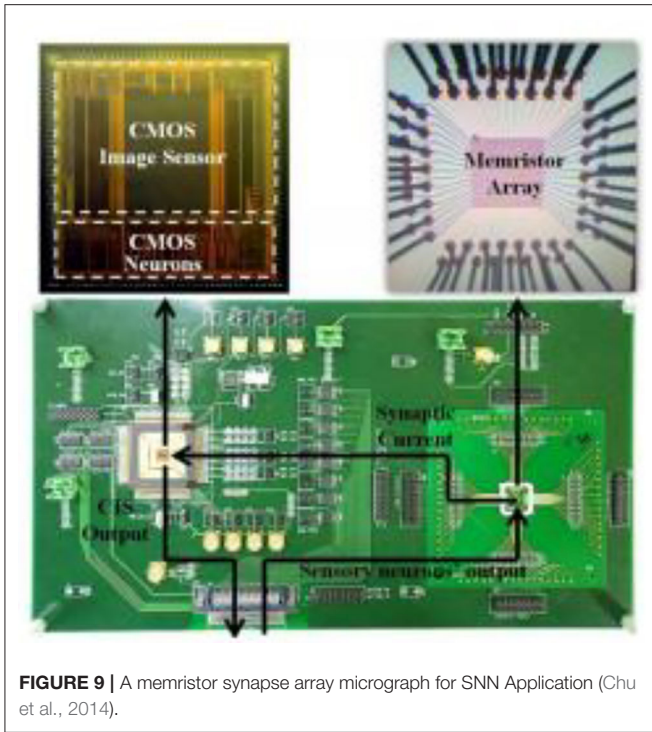**FIGURE 7 |** CMOS neuron and memristor synapse weight update circuit (Sun, 2015).

FIGURE 8 | CMOS-Memristor SNN (Wu and Saxena, 2018).

studies are limited to theoretical analysis and simulation of simple tasks rather than complex and intelligent tasks (e.g., multiple regression analysis, deductive and inductive reasoning, and their chip implementation) (Wang and Yan, 2019). Although the future of SNNs is still unclear, many researchers believe that SNNs will replace deep ANNs. The reason is that AI is essentially a biological brain mimicking process, and SNNs can provide a perfect mechanism for unsupervised learning.

As shown in **Figure 7**, a neural network is implemented with CMOS neurons, CMOS control circuits, and memristor synapses (Sun, 2015). The aggregation module, leaky integrate and fire module are equivalent to the role of dendrites and axon hillocks, respectively. Input neurons signals are temporally and spatially summed through a common-drain aggregation amplifier circuit. A memristor synapse gives the action potential signal a weight and its output signal, that is, a post-synaptic potential signal is transmitted to post-neurons. Using the action potential signal and feedback signals from post-neurons, the control circuit and synaptic update phase provide potentiation or depression signals to memristor synapses. According to the STDP

learning rules, the transistor-level weight adjustment circuit is composed of a memristor device and CMOS transmission gates. The transmission gates are controlled by potentiation or depression signals. The system is very similar to the main features of biological neurons, which is useful for neuromorphic SNN hardware implementation. A more complete description of SNN circuits and system applications is shown in **Figure 8** (Wu and Saxena, 2018). The system consists of event-driven CMOS neurons, a competitive neural coding algorithm [i.e., winner take all (WTA) learning rule], and multi-bit memristor synapse array. A stochastic non-linear STDP learning rule with an exponential shaped window learning function is adopted to update memristor synapse weights *in situ*. The amplitude and additional temporal delay of the half rectangular half-triangular spike waveform can be adjusted for dendritic-inspired processing. This work demonstrates the feasibility and excellence of emerging memristor devices in neuromorphic applications, with low power consumption and compact on-chip area.

Despite the large on-chip area and power dissipation in CMOS implementation of synaptic circuits (Chicca et al., 2003; Seo et al.,

FIGURE 9 | A memristor synapse array micrograph for SNN Application (Chu et al., 2014).

TABLE 6 | Comparison between ANNs and SNNs.

	ANNs	SNNs
Generation	Second-generation NN	Third-generation NN
Biological brain mimicking	General	Better
Signal processing	Continuous multi-level value	Sparse and asynchronous binary time-domain coded spike signals. Event-driven discrete information processing
Energy efficiency	General	Better
Neurons and Synapses	Activation functions;	Hodgkin and Huxley, LIF, etc.
	Digital or analog memristor synapses	Analog memristor synapses
Classical algorithms	Error-backpropagation	SpikeProp, STDP
Chip design	In progress with some achievement.	Preliminary stage
	Near-term application goals	Long-term application goals

2011), Myonglae Chu adopted $Pr_{0.7}Ca_{0.3}MnO_3$-based memristor synaptic array and CMOS leaky IAF neurons in SNN. As shown in **Figure 9**, the SNN chip has been successfully developed for visual pattern recognition with modified STDP learning rules. The SNN hardware system includes 30×10 neurons and 300 memristor synapses. Although this hardware system only recognizes numbers 0–9, it is a good attempt, as most SNN studies have lingered around the software simulation phase (Kim et al., 2011b; Adhikari et al., 2012; Cantley et al., 2012). One can refer to literatures (Wang et al., 2018b; Ishii et al., 2019; Midya et al., 2019b) for more experimental memristor-SNN demos.

Comparison Between ANNs and SNNs

A comparison between ANNs and SNNs is shown in **Table 6** (Nenadic and Ghosh, 2001; Chaturvedi and Khurshid, 2011; Zhang et al., 2020). Traditional ANNs require layer-by-layer computation. Therefore, it is computationally intensive and has a relatively large power consumption. An SNN changes from a standby mode to a working mode, when a large nerve spike is coming with its spike threshold exceeding the membrane voltage. As a result, its system power consumption is relatively low.

SNNs with higher bio-similarity are expected to achieve higher energy efficiency than ANNs. But SNN hardware is harder to implement than ANN hardware. Thus, combining the advantages of ANN and SNN and using ANN-SNN converters to improve SNN performance is a valuable method, which has been experimentally demonstrated (Midya et al., 2019a). The first and second layers of a converter are ordinary ANN structures. The output signals of the second layer are converted to a spike sequence for a 32×1 1M-1T drift memristor synapse array

at the third layer. This ANN-SNN converter may be a good way for SNN hardware implementation. Despite the enormous potential of SNNs, there is currently no fully satisfactory general learning rules and its computational capability has not been demonstrated. Most of these methods lack comparability and generality. Compared to ANNs, the study of dynamic devices and efficient algorithms in SNNs is very challenging. SNNs only need to compute the activated connections, rather than all connections at every time step in ANNs. However, the encoding and decoding of spikes is one of the challenges in SNN research. In fact, it needs further research in neuroscience. ANN is the recent target of memristors, and SNN is the long-term goal in the future.

For neural networks applications, ANN and SNN memristor grids have some common challenges, such as sneak path problems, IR-drop or ohmic drop, grid latency, and grid power dissipation, as shown as **Figure 10** (Zidan et al., 2013; Hu et al., 2014, 2018; Zhang et al., 2017). The large the size of the memristor array, the greater the effect of these parasitic capacitances and resistances. In Figure, the desired weight-update path is the dot-and-dash line, and the sneak path is the dotted line, which is an undesired parallel memristor path due to its relative resistance and non-gated memristor elements. This phenomenon leads to undesired weight changes and a reduction in the accuracy of neural networks. The basic solution for the sneak path is to add a series of connected gate-controlled MOS transistors to memristors as mentioned in **Table 2**. However, this method will lead to large on-chip synapse array and destroy the advantages of high-density integration of memristors. Grounding an unselected memristor array is another solution without the need to add synaptic area. But this approach leads to more power consumption. There are other techniques such as grounding line, floating line, additional bias,
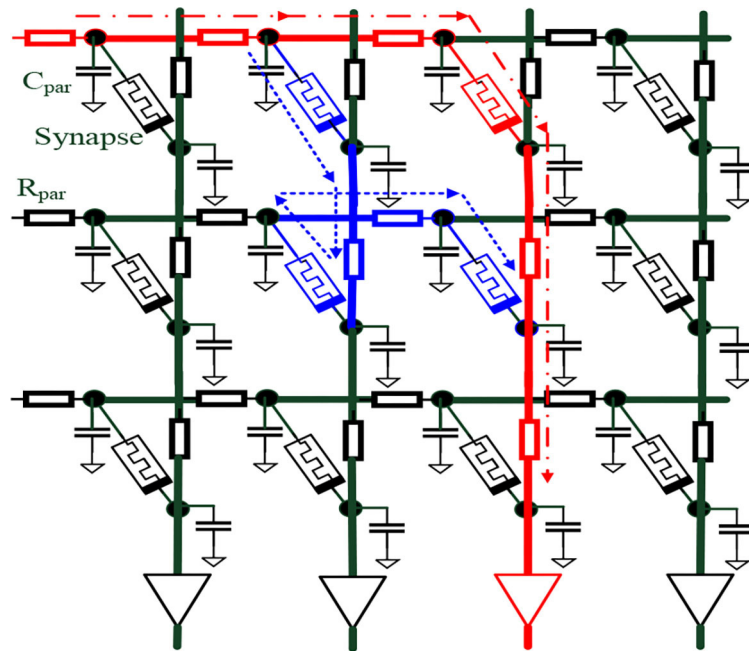
FIGURE 10 | Sneak path, IR-drop, latency, and energy in massive memristor grids of neural networks.

a non-unity aspect ratio of memristor arrays, three-electrode memristor devices. They may be welcome in memristor memory applications, but not necessarily in memristor-based neural network applications (Zidan et al., 2013). In neural network applications, the main concern for memristor arrays is whether the association between input and output signals is correct (Hu et al., 2014). This is one important difference compared to memristor memory applications. IR-drop, memristor grid latency, and power consumption are signal integrity effects caused by grid parasitic resistance Rpar and parasitic capacitance Cpar. These non-ideal factors affect the potential distribution, signal transmission, and ultimately affect the scale of memristor arrays. Similar to CMOS layout and routing techniques, large-scale memristors mesh can be divided into medium-sized modules with high-speed main signal paths for lower parasitic resistance, grid power consumption, and latency. It is worth noting that memristor process variations, gird IR-drop and noise can worsen the sneak path problem.

SUMMARY

The advantage of memristors in neural network applications is their fast processing time and energy efficiency in the computational process. At the device level, memristors have very low power dissipation and high on-chip density. At the architecture level, parallel computing is performed at the same location where data is stored, thereby avoiding frequent data movement and memory wall issues. Due to the quantum effect and non-ideal characteristics in the manufacturing of nanometer memristors, the robust performance of memristor

neural networks still needs to be improved. Meanwhile, the adaptation range of various memristor models is limited and has not been fully explored in chip design. To date, there are no complete unified memristor models for chip designer. Furthermore, wire resistance, sneak path current, and half-select problems are also challenges for high-density integration of memristor crossbar arrays. Memristor neural network research involves engineering, biology, physics, algorithms, architecture, systems, circuits, equipment, and materials. There is still a long way to go for memristive neural networks, as most research remains in single devices or small-scale prototypes. However, with the marketing promotion of the IoT big data and AI, the breakthrough research of memristor-based ANN will be realized by the joint efforts of academia and industry.

AUTHOR CONTRIBUTIONS

WX drafted the manuscript, developed the concept, and conceived the experiments. JW revised the manuscript. XY drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Adhikari, S. P., Yang, C., Kim, H., and Chua, L. O. (2012). Memristor bridge synapse-based neural network and its learning. *IEEE. Trans. Neur. Netw. Learn. Syst.* 23, 1426–1435. doi: 10.1109/TNNLS.2012.2204770

Ahmed, F. Y., Yusob, B., and Hamed, H. N. A. (2014). Computing with spiking neuron networks: a review. *Int. J. Adv. Soft Comput. Appl.* 6. Available online at: https://www.researchgate.net/publication/262374523_Computing_with_Spiking_Neuron_Networks_A_Review

Akinaga, H., and Shima, H. (2010). Resistive random access memory (ReRAM) based on metal oxides. *Proc. IEEE* 98, 2237–2251. doi: 10.1109/JPROC.2010.2070830

Alibart, F., Gao, L., Hoskins, B. D., and Strukov, D. B. (2012). High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology* 23:075201. doi: 10.1088/0957-4484/23/7/075201

Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern classification by memristive crossbar circuits using *ex situ* and *in situ* training. *Nat. Commun.* 4:2072. doi: 10.1038/ncomms3072

Al-Shedivat, M., Naous, R., Cauwenberghs, G., and Salama, K. N. (2015). Memristors empower spiking neurons with stochasticity. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 5, 242–253. doi: 10.1109/JETCAS.2015.2435512

Ambrogio, S., Narayanan, P., Tsai, H., Shelby, R. M., Boybat, I., Di Nolfo, C., et al. (2018). Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* 558, 60–67. doi: 10.1038/s41586-018-0180-5

Biolek, Z., Biolek, D., and Biolkova, V. (2009). SPICE model of memristor with nonlinear dopant drift. *Radioengineering* 18, 210–214. Available online at: https://www.researchgate.net/publication/26625012_SPICE_Model_of_Memristor_with_Nonlinear_Dopant_Drift

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press.

Cai, F., Correll, J. M., Lee, S. H., Lim, Y., Bothra, V., Zhang, Z., et al. (2019). A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations. *Nat. Electron.* 2, 290–299. doi: 10.1038/s41928-019-0270-x

Cantley, K. D., Subramaniam, A., Stiegler, H. J., Chapman, R. A., and Vogel, E. M. (2012). Neural learning circuits utilizing nano-crystalline silicon transistors and memristors. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 565–573. doi: 10.1109/TNNLS.2012.2184801

Chang, C. C., Chen, P. C., Chou, T., Wang, I. T., Hudec, B., Chang, C. C., et al. (2017). Mitigating asymmetric nonlinear weight update effects in hardware neural network based on analog resistive synapse. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 8, 116–124. doi: 10.1109/JETCAS.2017.2771529

Chaturvedi, S., and Khurshid, A. A. (2011). "Review of spiking neural network architecture for feature extraction and dimensionality reduction," in *2011 Fourth International Conference on Emerging Trends in Engineering and Technology* (Port Louis), 317–322. doi: 10.1109/ICETET.2011.57

Chen, B., Yang, H., Zhuge, F., Li, Y., Chang, T. C., He, Y. H., et al. (2019). Optimal tuning of memristor conductance variation in spiking neural networks for online unsupervised learning. *IEEE. Trans. Electron. Dev.* 66, 2844–2849. doi: 10.1109/TED.2019.2907541

Chen, P. Y., Lin, B., Wang, I. T., Hou, T. H., Ye, J., Vrudhula, S., et al. (2015). "Mitigating effects of non-ideal synaptic device characteristics for on-chip learning," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (Austin, TX), 194–199. doi: 10.1109/ICCAD.2015.7372570

Chen, W.-H., Dou, C., Li, K.-X., Lin, W.-Y., Li, P.-Y., Huang, J.-H., et al. (2019). CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors. *Nat. Electron.* 2, 420–428. doi: 10.1038/s41928-019-0288-0

Chicca, E., Indiveri, G., and Douglas, R. (2003). "An adaptive silicon synapse," in *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS'03)* (Bangkok), I–I. doi: 10.1109/ISCAS.2003.1205505

Choi, S., Tan, S. H., Li, Z., Kim, Y., Choi, C., Chen, P. Y., et al. (2018). SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations. *Nat. Mater.* 17, 335–340. doi: 10.1038/s41563-017-0001-5

Choi, S., Yang, Y., and Lu, W. (2014). Random telegraph noise and resistance switching analysis of oxide based resistive memory. *Nanoscale* 6, 400–404. doi: 10.1039/C3NR05016E

Chu, M., Kim, B., Park, S., Hwang, H., Jeon, M., Lee, B. H., et al. (2014). Neuromorphic hardware system for visual pattern recognition with memristor array and CMOS neuron. *IEEE. Trans. Ind. Electron.* 62, 2410–2419. doi: 10.1109/TIE.2014.2356439

Chua, L. (1971). Memristor-the missing circuit element. *IEEE Trans. Circ. Theor.* 18, 507–519. doi: 10.1109/TCT.1971.1083337

Chua, L. (2013). Memristor, Hodgkin–Huxley, and edge of Chaos. *Nanotechnology* 24:383001. doi: 10.1088/0957-4484/24/38/383001

Duan, S., Hu, X., Dong, Z., Wang, L., and Mazumder, P. (2014). Memristor-based cellular nonlinear/neural network: design, analysis, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1202–1213. doi: 10.1109/TNNLS.2014.2334701

Fackenthal, R., Kitagawa, M., Otsuka, W., Prall, K., Mills, D., Tsutsui, K., et al. (2014). "0.19.7 A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* (San Francisco, CA), 338–339. doi: 10.1109/ISSCC.2014.6757460

Fu, J., Liao, Z., Gong, N., and Wang, J. (2019). Mitigating nonlinear effect of memristive synaptic device for neuromorphic computing. *IEEE. J. Emerg. Select. Top. Circ. Syst.* 9, 377–387. doi: 10.1109/JETCAS.2019.2910749

Gi, S. G., Yeo, I., Chu, M., Moon, K., Hwang, H., and Lee, B. G. (2018). Modeling and system-level simulation for nonideal conductance response of synaptic devices. *IEEE. Trans. Electron. Dev.* 65, 3996–4003. doi: 10.1109/TED.2018.2858762

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Hong, Q., Zhao, L., and Wang, X. (2019). Novel circuit designs of memristor synapse and neuron. *Neurocomputing* 330, 11–16. doi: 10.1016/j.neucom.2018.11.043

Hu, M., Graves, C. E., Li, C., Li, Y., Ge, N., Montgomery, E., et al. (2018). Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mater.* 30:1705914. doi: 10.1002/adma.201705914

Hu, M., Li, H., Chen, Y., Wu, Q., Rose, G. S., and Linderman, R. W. (2014). Memristor crossbar-based neuromorphic computing system: a case study. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 1864–1878. doi: 10.1109/TNNLS.2013.2296777

Huang, A., Zhang, X., Li, R., and Chi, Y. (2018). *Memristor Neural Network Design*. London: IntechOpen.

Ishii, M., Kim, S., Lewis, S., Okazaki, A., Okazawa, J., Ito, M., et al. (2019). "On-Chip Trainable 1.4M 6T2R PCM Synaptic Array with 1.6K Stochastic LIF Neurons for Spiking RBM," in *IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA), 7–11.

Jeong, H., and Shi, L. (2018). Memristor devices for neural networks. *J. Phys. D Appl. Phys.* 52:023003. doi: 10.1088/1361-6463/aae223

Jeong, Y., Kim, S., and Lu, W. D. (2015). Utilizing multiple state variables to improve the dynamic range of analog switching in a memristor. *Appl. Phys. Lett.* 107:173105. doi: 10.1063/1.4934818

Jiang, H., Yamada, K., Ren, Z., Kwok, T., Luo, F., Yang, Q., et al. (2018). "Pulse-width modulation based dot-product engine for neuromorphic computing system using memristor crossbar array," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Florence), 1–4. doi: 10.1109/ISCAS.2018.8351276

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., and Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* 10, 1297–1301. doi: 10.1021/nl904092h

Joglekar, Y. N., and Wolf, S. J. (2009). The elusive memristor: properties of basic electrical circuits. *Eur. J. Phys.* 30:661. doi: 10.1088/0143-0807/30/4/001

Kataeva, I., Ohtsuka, S., Nili, H., Kim, H., Isobe, Y., Yako, K., et al. (2019). "Towards the development of analog neuromorphic chip prototype with 2.4 M integrated memristors," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (Sapporo), 1–5. doi: 10.1109/ISCAS.2019.8702125

Kim, H., Sah, M. P., Yang, C., Roska, T., and Chua, L. O. (2011a). Memristor bridge synapses. *Proc. IEEE* 100, 2061–2070. doi: 10.1109/JPROC.2011.2166749

Kim, H., Sah, M. P., Yang, C., Roska, T., and Chua, L. O. (2011b). Neural synaptic weighting with a pulse-based memristor circuit. *IEEE Trans. Circ. Syst. I Reg. Pap.* 59, 148–158. doi: 10.1109/TCSI.2011.2161360

Kim, S., Du, C., Sheridan, P., Ma, W., Choi, S., and Lu, W. D. (2015). Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* 15, 2203–2211. doi: 10.1021/acs.nanolett.5b00697

Kozhevnikov, D. D., and Krasilich, N. V. (2016). Memristor-based hardware neural networks modelling review and framework concept. *Proc. Inst. Syst. Prog. RAS* 28, 243–258. doi: 10.15514/ISPRAS-2016-28(2)-16

Krestinskaya, O., James, A. P., and Chua, L. O. (2019). Neuromemristive circuits for edge computing: a review. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4–23. doi: 10.1109/TNNLS.2019.2899262

Krestinskaya, O., Salama, K. N., and James, A. P. (2018). Learning in memristive neural network architectures using analog backpropagation circuits. *IEEE Trans. Circ. I* 66, 719–732. doi: 10.1109/TCSI.2018.2866510

Kvatinsky, S., Friedman, E. G., Kolodny, A., and Weiser, U. C. (2012). TEAM: threshold adaptive memristor model. *IEEE Trans. Circ. Syst. I Reg. Pap.* 60, 211–221. doi: 10.1109/TCSI.2012.2215714

Lastras-Montaño, M. A., and Cheng, K. T. (2018). Resistive random-access memory based on ratioed memristors. *Nat. Electron.* 1, 466–472. doi: 10.1038/s41928-018-0115-z

Lee, D., Park, J., Moon, K., Jang, J., Park, S., Chu, M., et al. (2015). "Oxide based nanoscale analog synapse device for neural signal recognition system," in *IEEE International Electron Devices Meeting (IEDM)* (Washington, DC), 4–7. doi: 10.1109/IEDM.2015.7409628

Lee, K. J., Cho, B. H., Cho, W. Y., Kang, S., Choi, B. G., Oh, H. R., et al. (2008). A 90 nm 1.8 V 512 Mb diode-switch PRAM with 266 MB/s read throughput. *IEEE. J. Solid State Circ.* 43, 150–162. doi: 10.1109/JSSC.2007.908001

Li, C., Belkin, D., Li, Y., Yan, P., Hu, M., Ge, N., et al. (2018). Efficient and self-adaptive *in-situ* learning in multilayer memristor neural networks. *Nat. Commun.* 9:2385. doi: 10.1038/s41467-018-04484-z

Li, C., Wang, Z., Rao, M., Belkin, D., Song, W., Jiang, H., et al. (2019). Long short-term memory networks in memristor crossbar arrays. *Nat. Mach. Intell.* 1, 49–57. doi: 10.1038/s42256-018-0001-4

Li, S., Wen, J., Chen, T., Xiong, L., Wang, J., and Fang, G. (2016). *In situ* synthesis of 3D CoS nanoflake/Ni (OH) 2 nanosheet nanocomposite structure as a candidate supercapacitor electrode. *Nanotechnology* 27:145401. doi: 10.1088/0957-4484/27/14/145401

Li, S. J., Dong, B. Y., Wang, B., Li, Y., Sun, H. J., He, Y. H., et al. (2018). Alleviating conductance nonlinearity via pulse shape designs in TaO x memristive synapses. *IEEE. Trans. Electron. Dev.* 66, 810–813. doi: 10.1109/TED.2018.2876065

Lin, P., Li, C., Wang, Z., Li, Y., Jiang, H., Song, W., et al. (2020). Three-dimensional memristor circuits as complex neural networks. *Nat. Electron.* 3, 225–232. doi: 10.1038/s41928-020-0397-9

Liu, J., Li, Z., Tang, Y., Hu, W., and Wu, J. (2020). 3D Convolutional Neural Network based on memristor for video recognition. *Pattern. Recogn. Lett.* 130, 116–124. doi: 10.1016/j.patrec.2018.12.005

Liu, T. Y., Yan, T. H., Scheuerlein, R., Chen, Y., Lee, J. K., Balakrishnan, G., et al. (2013). "A 130.7mm^2 2-layer 32Gb ReRAM memory device in 24nm technology," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers* (San Francisco, CA), 210–211. doi: 10.1109/JSSC.2013.2280296

Luo, Q., Xu, X., Gong, T., Lv, H., Dong, D., Ma, H., et al. (2020). "8-Layers 3D vertical RRAM with excellent scalability towards storage class memory applications," in *IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA), 2.7.1–2.7. 4.

Ma, D., Wang, G., Han, C., Shen, Y., and Liang, Y. (2018). A memristive neural network model with associative memory for modeling affections. *IEEE Access.* 6, 61614–61622. doi: 10.1109/ACCESS.2018.2875433

Midya, R., Wang, Z., Asapu, S., Joshi, S., Li, Y., Zhuo, Y., et al. (2019a). Artificial neural network (ANN) to spiking neural network (SNN) converters based on diffusive memristors. *Adv. Electron. Mater.* 5:1900060. doi: 10.1002/aelm.201900060

Midya, R., Wang, Z., Asapu, S., Zhang, X., Rao, M., Song, W., et al. (2019b). Reservoir computing using diffusive memristors. *Adv. Intell. Syst.* 1:1900084. doi: 10.1002/aisy.201900084

Midya, R., Wang, Z., Zhang, J., Savel'ev, S. E., Li, C., Rao, M., et al. (2017). Anatomy of Ag/Hafnia-based selectors with 10^{10} nonlinearity. *Adv. Mater.* 29:1604457. doi: 10.1002/adma.201604457

Nenadic, Z., and Ghosh, B. K. (2001). "Computation with biological neurons," in *Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148)* (Arlington, VA), 257–262. doi: 10.1109/ACC.2001.945552

Ohno, T., Hasegawa, T., Tsuruoka, T., Terabe, K., Gimzewski, J. K., and Aono, M. (2011). Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* 10, 591–595. doi: 10.1038/nmat3054

Ojiugwo, C. N., Abdallah, A. B., and Thron, C. (2020). "Simulation of biological learning with spiking neural networks," in *Implementations and Applications of Machine Learning,* Vol. 782, eds S. A. Subair and C. Thron (Cham: Springer),207–227. doi: 10.1007/978-3-030-37830-1_9

Park, S., Sheri, A., Kim, J., Noh, J., Jang, J., Jeon, M., et al. (2013). "Neuromorphic speech systems using advanced ReRAM-based synapse," in *IEEE International Electron Devices Meeting* (Washington, DC), 25.6.1–25.6.4. doi: 10.1109/IEDM.2013.6724692

Pfeiffer, M., and Pfeil, T. (2018). Deep learning with spiking neurons: opportunities and challenges. *Front. Neurosci.* 12:774. doi: 10.3389/fnins.2018.00774

Pickett, M. D., Strukov, D. B., Borghetti, J. L., Yang, J. J., Snider, G. S., Stewart, D. R., et al. (2009). Switching dynamics in titanium dioxide memristive devices. *J. Appl. Phys.* 106:074508. doi: 10.1063/1.3236506

Prezioso, M., Merrikh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K., and Strukov, D. B. (2015). Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61–64. doi: 10.1038/nature14441

Prodromakis, T., Peh, B. P., Papavassiliou, C., and Toumazou, C. (2011). A versatile memristor model with nonlinear dopant kinetics. *IEEE. Trans. Electron Dev.* 58, 3099–3105. doi: 10.1109/TED.2011.2158004

Sarwar, S. S., Saqueb, S. A. N., Quaiyum, F., and Rashid, A. H. U. (2013). Memristor-based nonvolatile random access memory: hybrid architecture for low power compact memory design. *IEEE Access* 1, 29–34. doi: 10.1109/ACCESS.2013.2259891

Seo, J. S., Brezzo, B., Liu, Y., Parker, B. D., Esser, S. K., Montoye, R. K., et al. (2011). "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," in *IEEE Custom Integrated Circuits Conference (CICC)* (San Jose, CA), 1–4. doi: 10.1109/CICC.2011.6055293

Seok, J. Y., Song, S. J., Yoon, J. H., Yoon, K. J., Park, T. H., Kwon, D. E., et al. (2014). A review of three-dimensional resistive switching cross-bar array memories from the integration and materials property points of view. *Adv. Funct. Mater.* 24, 5316–5339. doi: 10.1002/adfm.201303520

Shukla, A., and Ganguly, U. (2018). An on-chip trainable and the clock-less spiking neural network with 1R memristive synapses. *IEEE. Trans. Biomed. Circ. Syst.* 12, 884–893. doi: 10.1109/TBCAS.2018.2831618

Simmons, J. G. (1963). Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* 34, 1793–1803. doi: 10.1063/1.1702682

Smagulova, K., Krestinskaya, O., and James, A. P. (2018). A memristor-based long short term memory circuit. *Analog. Integr. Circ. Signal Process.* 95, 467–472. doi: 10.1007/s10470-018-1180-y

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature* 453, 80–83. doi: 10.1038/nature06932

Strukov, D. B., and Williams, R. S. (2009). Exponential ionic drift: fast switching and low volatility of thin-film memristors. *Appl. Phys. A* 94, 515–519. doi: 10.1007/s00339-008-4975-3

Sun, J. (2015). *CMOS and Memristor Technologies for Neuromorphic Computing Applications.* Technical Report No. UCB/EECS-2015-S-2218, Electrical Engineering and Computer Sciences University of California at Berkeley.

Sun, K., Chen, J., and Yan, X. (2020). The future of memristors: materials engineering and neural networks. *Adv. Funct. Mater.* 31:2006773. doi: 10.1002/adfm.202006773

Tanikawa, T., Ohnishi, K., Kanoh, M., Mukai, T., and Matsuoka, T. (2018). Three-dimensional imaging of threading dislocations in GaN crystals using two-photon excitation photoluminescence. *Appl. Phys. Express.* 11:031004. doi: 10.7567/APEX.11.031004

Tsai, H., Ambrogio, S., Mackin, C., Narayanan, P., Shelby, R. M., Rocki, K., et al. (2019). "Inference of long-short term memory networks at software-equivalent accuracy using 2.5M analog phase change memory devices," in *Symposium on VLSI Technology* (Kyoto), 82–83. doi: 10.23919/VLSIT.2019.8776519

Upadhyay, N. K., Sun, W., Lin, P., Joshi, S., Midya, R., Zhang, X., et al. (2020). A memristor with low switching current and voltage for 1s1r integration and array operation. *Adv. Electron. Mater.* 6:1901411. doi: 10.1002/aelm.201901411

Volos, C. K., Kyprianidis, I. M., Stouboulos, I. N., Tlelo-Cuautle, E., and Vaidyanathan, S. (2015). Memristor: a new concept in synchronization of coupled neuromorphic circuits. *J. Eng. Sci. Technol. Rev.* 8, 157–173. doi: 10.25103/jestr.082.21

Wang, H., and Yan, X. (2019). Overview of resistive random access memory (RRAM): materials, filament mechanisms, performance optimization, and prospects. *Phys. Status. Solidi R.* 13:1900073. doi: 10.1002/pssr.201900073

Wang, T., and Roychowdhury, J. (2016). Well-posed models of memristive devices. *arXiv* preprint arXiv:1605.04897.

Wang, Z., Joshi, S., Savel'ev, S., Song, W., Midya, R., Li, Y., et al. (2018a). Fully memristive neural networks for pattern classification with unsupervised learning. *Nat Electron.* 1, 137–145. doi: 10.1038/s41928-018-0023-2

Wang, Z., Joshi, S., Saveliev, S. E., Jiang, H., Midya, R., Lin, P., et al. (2017). Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* 16, 101–108. doi: 10.1038/nmat4756

Wang, Z., Li, C., Lin, P., Rao, M., Nie, Y., Song, W., et al. (2019a). *In situ* training of feed-forward and recurrent convolutional memristor networks. *Nat. Mach. Intell.* 1, 434–442. doi: 10.1038/s42256-019-0089-1

Wang, Z., Li, C., Song, W., Rao, M., Belkin, D., Li, Y., et al. (2019b). Reinforcement learning with analogue memristor arrays. *Nat. Electron.* 2, 115–124. doi: 10.1038/s41928-019-0221-6

Wang, Z., Rao, M., Han, J. W., Zhang, J., Lin, P., Li, Y., et al. (2018b). Capacitive neural network with neuro-transistors. *Nat. Commun.* 9:3208. doi: 10.1038/s41467-018-05677-5

Wang, Z., Wu, H., Burr, G. W., Hwang, C. S., Wang, K. L., Xia, Q., et al. (2020). Resistive switching materials for information processing. *Nat. Rev. Mater.* 5, 173–195. doi: 10.1038/s41578-019-0159-3

Wang, Z., Zhao, W., Kang, W., Zhang, Y., Klein, J. O., and Chappert, C. (2014). "Ferroelectric tunnel memristor-based neuromorphic network with 1T1R crossbar architecture," *International Joint Conference on Neural Networks (IJCNN)* (Beijing), 29–34. doi: 10.1109/IJCNN.2014.6889951

Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Adv. Mater.* 21, 2632–2663. doi: 10.1002/adma.200900375

Wong, H. S. P., Lee, H. Y., Yu, S., Chen, Y. S., Wu, Y., Chen, P. S., et al. (2012). Metal–oxide RRAM. *Proc. IEEE* 100, 1951–1970. doi: 10.1109/JPROC.2012.2190369

Woo, J., and Yu, S. (2018). Resistive memory-based analog synapse: the pursuit for linear and symmetric weight update. *IEEE Nanotechnol. Mag.* 12, 36–44. doi: 10.1109/MNANO.2018.2844902

Wu, X., and Saxena, V. (2018). Dendritic-inspired processing enables bio-plausible STDP in compound binary synapses. *IEEE. Trans. Nanotechnol.* 18, 149–159. doi: 10.1109/TNANO.2018.2871680

Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* 18, 309–323. doi: 10.1038/s41563-019-0291-x

Xue, C.-X., Chiu, Y.-C., Liu, T.-W., Huang, T.-Y., Liu, J.-S., Chang, T.-W., et al. (2020). A CMOS-integrated compute-in-memory macro based on resistive random-access memory for AI edge devices. *Nat. Electron.* 4, 81–90. doi: 10.1038/s41928-020-00505-5

Yan, X., Li, X., Zhou, Z., Zhao, J., Wang, H., Wang, J., et al. (2019a). Flexible transparent organic artificial synapse based on the tungsten/egg albumen/indium tin oxide/polyethylene terephthalate memristor. *ACS. Appl. Mater. Inter.* 11, 18654–18661. doi: 10.1021/acsami.9b04443

Yan, X., Pei, Y., Chen, H., Zhao, J., Zhou, Z., Wang, H., et al. (2019b). Self-assembled networked PbS distribution quantum dots for resistive switching and artificial synapse performance boost of memristors. *Adv. Mater.* 31:1805284. doi: 10.1002/adma.201805284

Yan, X., Wang, K., Zhao, J., Zhou, Z., Wang, H., Wang, J., et al. (2019c). A new memristor with 2D Ti3C2Tx MXene flakes as an artificial bio-synapse. *Small* 15, 1900107. doi: 10.1002/smll.201900107

Yan, X., Zhang, L., Chen, H., Li, X., Wang, J., Liu, Q., et al. (2018a). Graphene oxide quantum dots based memristors with progressive conduction tuning for artificial synaptic learning. *Adv. Funct. Mater.* 28:1803728. doi: 10.1002/adfm.201803728

Yan, X., Zhang, L., Yang, Y., Zhou, Z., Zhao, J., Zhang, Y., et al. (2017). Highly improved performance in Zr 0.5 Hf 0.5 O_2 films inserted with graphene oxide quantum dots layer for resistive switching non-volatile memory. *J. Mater. Chem. C.* 5, 11046–11052. doi: 10.1039/C7TC03037A

Yan, X., Zhao, J., Liu, S., Zhou, Z., Liu, Q., Chen, J., et al. (2018b). Memristor with Ag-cluster-doped TiO_2 films as artificial synapse for Neuroinspired computing. *Adv. Funct. Mater.* 28:1705320. doi: 10.1002/adfm.201705320

Yan, X., Zhao, Q., Chen, A. P., Zhao, J., Zhou, Z., Wang, J., et al. (2019d). Vacancy-induced synaptic behavior in 2D WS2 nanosheet–based memristor for low-power neuromorphic computing. *Small* 15:1901423. doi: 10.1002/smll.201901423

Yan, X., Zhou, Z., Zhao, J., Liu, Q., Wang, H., Yuan, G., et al. (2018c). Flexible memristors as electronic synapses for neuro-inspired computation based on scotch tape-exfoliated mica substrates. *Nano. Res.* 11, 1183–1192. doi: 10.1007/s12274-017-1781-2

Yang, C., Adhikari, S. P., and Kim, H. (2019). On learning with nonlinear memristor-based neural network and its replication. *IEEE. Trans. Circ. I.* 66, 3906–3916. doi: 10.1109/TCSI.2019.2914125

Yang, C., Choi, H., Park, S., Sah, M. P., Kim, H., and Chua, L. O. (2014). A memristor emulator as a replacement of a real memristor. *Semicond. Sci. Technol.* 30, 015007. doi: 10.1088/0268-1242/30/1/015007

Yang, J. J., Strukov, D. B., and Stewart, D. R. (2013). Memristive devices for computing. *Nat. Nanotechnol.* 8, 13–24. doi: 10.1038/nnano.2012.240

Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q., Zhang, W., et al. (2020). Fully hardware-implemented memristor convolutional neural network. *Nature* 577, 641–646. doi: 10.1038/s41586-020-1942-4

Zhang, Y., Wang, X., and Friedman, E. G. (2017). Memristor-based circuit design for multilayer neural networks. *IEEE. Trans. Circ. I.* 65, 677–686. doi: 10.1109/TCSI.2017.2729787

Zhang, Y., Wang, Z., Zhu, J., Yang, Y., Rao, M., Song, W., et al. (2020). Brain-inspired computing with memristors: challenges in devices, circuits, and systems. *Appl. Phys. Rev.* 7:011308. doi: 10.1063/1.5124027

Zhao, J., Zhou, Z., Zhang, Y., Wang, J., Zhang, L., Li, X., et al. (2019). An electronic synapse memristor device with conductance linearity using quantized conduction for neuroinspired computing. *J. Mater. Chem. C* 7, 1298–1306. doi: 10.1039/C8TC04395G

Zhao, Q., Xie, Z., Peng, Y. P., Wang, K., Wang, H., Li, X., et al. (2020). Current status and prospects of memristors based on novel 2D materials. *Mater. Horiz.* 7, 1495–1518. doi: 10.1039/C9MH02033K

Zheng, N., and Mazumder, P. (2018). Learning in memristor crossbar-based spiking neural networks through modulation of weight-dependent spike-timing-dependent plasticity. *IEEE. Trans. Nanotechnol.* 17, 520–532. doi: 10.1109/TNANO.2018.2821131

Zidan, M. A., Fahmy, H. A. H., Hussain, M. M., and Salama, K. N. (2013). Memristor-based memory: the sneak paths problem and solutions. *Microelectron. J.* 44, 176–183. doi: 10.1016/j.mejo.2012.10.001