Applications of Memristors in Neural Networks and Neuromorphic Computing: A Review

Ye-Guo Wang

Abstract—Memristor is one of the best choices for neuromorphic computing because of its synapse-like structure and function. The single memristor with ion dynamics enables emulations of diverse synaptic plasticity significant for learning and memory. Moreover, several memristive crossbar arrays show low power consumption, high precision and high efficiency on physically achieving algorithmic functions. Although a large number of experiments have demonstrated great potential of memristive devices in the field of computer architecture design and integrated circuits, there is still a long way to go for their practical industrialization. This review concentrates on the application and function of memristors, as well as some critical challenges and perspectives on their future development.

Index Terms—Memristor, neuromorphic computing, neural network, oxide, synaptic plasticity.

I. INTRODUCTION

Memristor [1], [2] is an intriguing electrical unit with an intrinsic hysteresis conductance. Physically, memristor is a two-terminal [3] device with a sandwich structure of metal/insulator/metal (MIM). The evolution of metal/defect conductive filaments [4] based on ion migration [5], [6] inside the insulator layer is the reason of that hysteresis conductance. When one electrode is under positive potential while another one is connected with negative potential, the metal/defect ions start to move under the external electrical field from one side to another, its main function is to regulate the high/low resistance of the memristor. It is widely acknowledged that synapses exist in cerebral cortex, [7] which are basically gray matters covering the surface of the cerebral hemisphere that enables the neuron communications between the left- and right-brain, and the structures of synapses are partly related to communicated function of human brain. Single synapse is composed of a pre-synaptic membrane, a post-synaptic membrane and a synaptic cleft between the two membranes, synaptic memory function is related to the inflow and outflow of calcium ions among those three different structures. As aforementioned, a single memristor unit is a sandwich structure, e.g. Ti/TiO₂/Al₂O₃/Pt, [8] and it also includes ion dynamic process. The memristor is therefore an ideal candidate for imitating the biological synapses morphologically and functionally.

Traditionally, since the data in dynamic random access memory (DRAM) [9] will vanish without external power, in other words, DRAM is short retention, it must be refreshed every 16 milliseconds to ensure the proper operation. However, memristor does not need to be updated in a few days or weeks even if there is no external power because of its non-volatile [10] characteristic, e.g. resistant random access memory (RRAM) [10]-[13]. Additionally, RRAM device is proved to be the next generation non-volatile memory due to its low energy consumption (<0.1 pJ), [13]-[15] high speed (<1 ns), [16] excellent endurance (>10¹² switching cycles), [13] long retention (10 years) [17] and the smallest size (4F²) [18]. Hence, it can gracefully solve the size limitation of conventional transistors. Moreover, since memristors can be fabricated as simple cross-point devices, i.e. crossbar structures, [19], [20] they are able to be readily stacking as 3D integration array [12] with high density, which is feasible for breaking the memory-wall based on Von Neumann architecture and extending Moore’s law.

Recently, memristive arrays are highlighted for in-memory computing because they are based on physical diagrams of neural networks by using analog memristors with cross-point connections, enabling human brain bionics of integrated circuits, faster computing and high accuracy. The researches have tried to construct several feasible and efficient memristive arrays, including one diode and one memristor (1D1R) [13] structure, one transistor and one memristor (1T1R) [21], [22] structure and one selector and one memristor (1S1R) [20], [23] structure. By integrating diodes, transistors or selectors, the more precise and repetitive programming of memristor conductance as well as higher programming speed can be simply accomplished than those of coding and encoding based on conventional computing architectures.

In this review, we demonstrated the characteristics and distinctive functions of memristor in both single devices and three forms of arrays, memristors have great potential compared to existing hardware devices. In addition, artificial neural networks with memristive devices [24] possess an outstanding advantage in the fields of face recognition and image processing [25]. We expect to presents phased achievements and the future challenges of memristors, and to provide guidance of relevant researches.

II. APPLICATIONS AND FUNCTIONS OF MEMRISTORS

A. Single Memristor Device

Multiple researches have shown that the source of memory is synapse, so it is intended to develop a chip to simulate brain computing, the core of this chip is the function of simulating synapses, which is the technical advantage of
memristive device [26], [27]. Apparently, the top electrode acts like a presynaptic membrane, the biosynaptic ions (sodium, potassium) are like conductive ions, [6] and the low electrode is like a postsynaptic membrane. Functionally, it is widely accepted that people remember in various and intricate ways, such as short-term plasticity (STP) and long-term plasticity (LTP), [28], [29] and it can be learned that both STP and LTP simulated to work in organic synapses nanoscale memristors in an experimental research, which possess the capacity to serve as synapse in neuromorphic systems.

![Isolated memristive device.](image)

Fig. 1. Isolated memristive device. (a) millimeter scale configuration of Ag/Ag₂S/Pt memristive device and nanometer scale configuration. Reproduced from[4], American Chemistry Society. (b) stack structure of Pd/WOₓ/W memristive device. Reference [31], American Chemistry Society.

![Simulation of method of memory](image)

Fig. 2. Simulation of method of memory (a) STM-to-LTM transition. (b) I–V characteristics of the Ag₂S atomic switch under the ON/OFF switching operation. (c) (insertion) Anti-STDP synaptic feature map / STDP synaptic feature map (d) the simulated conductance of pulse programming in a PCMO device (Inset (a). The current compliance level can control the set resistance. Inset (b). Reset voltage control the reset resistance completely). (e)The STDP waveform is decomposed into two part as pre (red) and post (green) neuronal waveforms. (f) The upper panel shows that Neuron waveform shift to different ranges in time (Δ t) i.e. (i) negative (Δ t < 0), (ii) Δ t = 0, (iii)positive (Δ t > 0). The bottom panel shows that resultant superposition generates peak voltage Vₚₑᵃᵏ(Δ t) which is similar to STDP waveform. Reproduced from [31], American Chemistry Society. (a);Reference [33], Springer Nature Ltd. (b); Reference [36], Institute of Physics. (c); Reference [37], IEEE. (d-f).

S. L. Barbera et al. fabricated the Ag/Ag₂S/Pt device in millimeter size (Fig. 1a), [4] STP and LTP operating modes can be controlled by simply adjusting the volatility of the memristive device. There were two parameters, Iₚ (pulse mode) and the number of pulse which controlling Gₘₐₓ, altering the electrical conductance alongside filament volatility. Also, A memristor device (Fig. 1b) consisting of a W bottom electrode, a Pd top electrode, and a WOₓ film, was developed by T. Chang et al. in order to mimic the STM(Short Term Memory)-to-LTM(Long Term Memory) [30] conversion process as shown in Fig. 2a: when different amounts of uniform irritation (dots) and fitted curves which employing the SEF (solid lines) have been applied, the memory retention data can be recorded, and the data are scaled by a prefactor I₀. The memristor received a repeated stimulation pulse (duration for 0.4ms, amplitude for +1.3V, pulse interval for 60ms). After that, through reading current with +0.5V, 8ms loading pulses when the last incentives in the series [31] have finished, retention curves were successfully acquired. In addition, the Ag₂S inorganic synapse was discovered by T. Ohno et al., [32] by means of inputting pulse repetition time, the synaptic functions of STP and LTP peculiarities can be imitated. In fact, the structure named atomic switch operating at threshold voltage (Fig. 2b) retains information as STP and spontaneously decays in response to the conductance level of the intermittent input stimulus, and frequent stimulation leads to the conversion to LTP. A valuable feature of Ag₂S inorganic synapse is similar to a single biological synapse, enabling dynamic memory in every device without external pre-programming [33].
Besides, numerous scientists focus on spike-timing-dependent-plasticity (STDP) [34], [35]. K. Seo. et al. used nanoscale titanium dioxide double-layer resistive switching devices to demonstrate synaptic plasticity alongside spike-timing-dependent-plasticity and simulated memory with a simple manufacturing process and good yield uniformity [36], as shown in Fig. 2c, t in horizontal axis indicates that the presynaptic membrane and the protruding posterior membrane are simultaneously stimulated or not stimulated at the same time, whereas the vertical axis represents the synaptic weight, the strength of the synaptic intrinsic connection has confirmed the separation state of the conductivity change accuracy and multi-level conductance along with simulation of memory features. Subsequently, it also proves the potential of titanium dioxide double-layer resistance converters as synapses in achieving the function of neuromorphic devices by analyzing STDP and biological triple models deeply. And N. Panwar. et al. share completely new method of a series of spike-related time scales for STDP with W/Pt:Ca0.5MnO2/Pt based memristor. To generate any arbitrary STDP, the approach that for programing the required pre- and post-neuron waveforms was used. In Fig. 2d, since G is proportional to V peak, the waveform of STDP (ΔG vs. Δt) is substituted by a voltage waveform (i.e. V vs. t) in Fig. 2e, and the V waveform is divided into two forms, the green one (V>0) is to indicate LTP, whereas the red one (V < 0) indicates LTD. And a set or reset appears when VTa (resistance switching threshold) was surpassed by |V| (Fig. 2f). By reason of the approx. linear G vs. V peak (Fig. 2d), the STDP is produced [37]-[39].

In theory, the function of the synapse in the brain can be simulated according to the characteristics of the memristor. However, the structure of the brain is very complicated. The synapse exists as a part of the structure in the brain, realizing the function of the brain. Therefore, if scientists want to realize the function of the brain. For example, memory learning, it requires not only the participation of synapses. In other words, to simulate the computational functions of the brain, [40] not only memristor devices are required but also other unknown fields need to be developed. The truth is simulating synapses through memristors alone cannot fully realize brain function. For instance, our brain consists of two hemispheres is immersed in the cerebrospinal fluid, a liquid environment, with two lateral hemispheres connected by nerve fibers, there are liquid between each structure. When the synapse realizes its function, the presynaptic cells transfer information to the postsynaptic cells by means of chemical signals called neurotransmitters, and this kind of synapse called chemical synapse, which is different from the way the memristor works on the integrated circuit. To completely simulate the brain and realize the brain-like work, not only microelectronic scientists but also brain neuroscientists are needed, in order to better simulate the device, we have to understand the brain more deeply.

**B. Memristor Arrays**

There are many types of memristor arrays. One is single memristor arrays. The structure of a single-transistor single-resistor (1T1R) array requires an access transistor at each intersection to independently gate each cell. E. J. Merced-Grafals et al. study the suitability of the 1T1R array as a synaptic component of the neuromorphic system [21]. It was made up of the integration of CMOS—transistors and tantalum oxide TaOx, memristor devices with its switching material was put between Ta/Pt electrodes. By utilizing an simple pulse algorithm that makes use of the transistor gate voltage which controlling the SET switch operation and augment the programming speed of the 1T1R cell, testing and programming are implemented. In the Fig. 3a, the first action is to start a SET operation, TE (top electrode) received a positive voltage, while for RESET, BE (bottom electrode) received a positive voltage. When GE (gate electrode) is applied for selecting and programming the cell, a voltage is applied to either TE or BE [41]. By biasing the the transistors’ gate to the peak voltage and applying a 0.1 V bias at BE (see Fig. 3b), the memristor’s conductance is acquired [42], [43]. In order to accurately adjust the memristor’s conductance during the test, in Fig. 3c, After single SET and RESET square pulses with duration for 100 ns and amplitude of 1.1 V and 1.4 V separately, low bias conductance was be adapted as a function of V SET. Besides, For the purpose of speeding up the programming process, an accommodated algorithm is used to incorporate the aforesaid model, also, a second order polynomial using least squares is fitted with SET data. In addition, 1k-bit PCMO-based resistive memory arrays was fabricated by J. W. Jang et al. for evaluation as synaptic devices by means of conventional lithography. In the PCMO crossbar array, which can implement the signed weights required for learning algorithms by using G and G* to simulate positive conductance and negative conductance for single synapse. It is also applied to conductance difference, which is w (w = G*−G), serves as the synaptic weight (see Fig. 3d). The experimental simulation process is as follows [44]:

\[
G = \frac{1}{G_{\text{LRS}} - G_{\text{HRS}}} \left( \sqrt{((G_{\text{LRS}} - G_{\text{HRS}}) \times w + G_{\text{HRS}}^2)} \right) - G_{\text{HRS}} \times \frac{G_{\text{LRS}}}{G_{\text{HRS}}} \right)
\]

\[G_{\text{LRS}}\] represents low resistance state (LRS) conductance whereas \[G_{\text{HRS}}\] represents high resistance state (HRS) conductance. There is a parameter \(\alpha\) that controls the enhancement (ap) or suppression (ad) characteristics as well as a internal variable \(w\) which verifies from 0 to 1. In the experiment, \(w\) changes due to the two different pulse were applied to the memristor array devices, if \(\alpha > 1\), the enhancement and suppression characteristics of the device model are sunken, and concave-up if \(\alpha < 1\) [5].

The 3D stacking structure is also of interest, B. Gao. et al. show the Gd-doped HfO2 memristor device which composed of TiN, SiO2, Pt, and TiON with a vertical structure. In the synaptic training, as shown in Fig. 4a and Fig. 4b, the SET conversion and RESET conversion caused by repetitive SET pulses (+0.17V/1 ns) and RESET pulses (-0.13V/1 ns), respectively, in which their device working as binary synapse or analog synapse. In a neuromorphic visual system for pattern recognition, the first layer, which representing the retina, contains 32×32 neurons, and it connected with second layer, which representing the primary visual cortex, through
out 32×32×n. In addition, depending on the light intensity of the input mode, the neurons in the first layer are inspired and send a pulse to the second layer of neurons (representing the visual cortex) through synapse. Cortical neurons sum the input current and integrate it, also, the neurons with the Maximum input current are emitted firstly, inhibiting the emission of others. Then, to regulate the weight of the synapse, the winner neuron sends a pulse back to all retinal neurons. In the imitation, based on the experimental results, the relative resistance fluctuation (δR = R) was set to 9%. Besides, it recognizes the validity as a function of relative resistance change [12], [45], [46].

Fig. 3. Memristor array. (a) The SET and RESET operation of 1T1R cell. (b) When the DC SET operation has been applied, the Low bias conductance serves as a function of $V_{g-set}$ for three different cells. (c) After single SET and RESET square pulses with duration for 100 ns and amplitude of 1.1 V and 1.4 V separately, low bias conductance was be adapted as a function of $V_{g-set}$. (d) Conductance serves as synaptic weight. Reproduced from [12], Institute of Physics (a-c).

Fig. 4. 3D stacking structure for memristive device. Synaptic training characteristics of 3D ReRAM devices. (a) The RESET conversion caused by repetitive RESET pulses (-0.13V/1 ns), in which their device working as analog synapse. (b) The SET conversion caused by repetitive SET pluses (+0.17V/1 ns), in which their device working as binary synapse. Reproduced from [21], Institute of Physics (a-b).

The current neural network operation mode requires a lot of neurons, and the sparse coding algorithm [47] can reduce the use of neurons. Sparse coding is an unsupervised learning method, which can find more effectively through the competition between neurons on the chip. The memristor chip is trained to use few neurons to successfully find key features from some photos, thus reducing energy consumption. Additionally, there are key challenges involving the use of memristors for building hardware DPEs, [21], [48] for examples, the nonlinear memristor dynamics, the negative impact and damage of the working environment noise, and the requirements and influence of the switching sensitivity of the memristor itself on the initial state. The operation and maintenance of the DPE are affected by controllable factors as well as uncontrollable factors, especially the conductance programming whose calculation steps requiring accurate values.

III. ADVANTAGES OF USING MEMRISTOR NETWORK

A. Shustanov et al. used CNN (Convolutional Neural Networks) to design for real-time traffic sign recognition. [49] The design system used speed signals received from the experimental vehicle, which allows the first researcher to accurately calculate the presence of surrounding obstacles. On the other hand, researchers could accurately predict the proportion of adjacent frames and real-time precise coordinates. Therefore, the efficiency and accuracy of the system work was greatly increased. However, the complexity of computing remains a challenge. Here, they considered the limitation of the time to process a single frame, used CNN to implement the classification of local objects, and developed a method for positioning. As it can be seen, artificial neural network is an excellent choice for solving pattern recognition issues. As for calculation method, which is described as follows:

$$a_j^i = \sigma(\sum_k a_k^{i-1}w_{jk})$$

where $a_j^i$ is the $j^{th}$ neuron in the $i^{th}$ layer, $w_{jk}$ represents the weight of the synapse, and connects the $j^{th}$ neuron in the $i^{th}$ layer alongside the $k^{th}$ neuron in the $i-1$th layer. The
training process used a gradient-based minimization method (also known as backpropagation) to minimize the cost function. In classification problems, the most common cost function was the cross entropy:

$$H(p, q) = -\sum_i Y(i) \log y(i)$$

For the existence of problem of the gradient disappearance, it is difficult to have a deep network with sigmoid activation. In order to minimize the impact of this problem, researchers adapted an activation function of ELU:

$$ELU(x) = \begin{cases} \exp(x) - 1, & x \leq 0 \\ x, & x > 0 \end{cases}$$

Also, they used TensorFlow for training. During the experimental training phase, for each iteration, the training network processed a batch of 50 images from the training data set, accurately calculating intermediate accuracy with every 100 iterations and using 50 images batch from the test data set. Finally, the experiment showed 99.94% correct positioning and detected dangerous traffic signs [50].

![Image](image.jpg)

Fig. 5. Application of memristor for recognition. (a) Histogram of device conductance at the initial state, which was measured after training epoch 21, and after epoch 54. (b) The corresponding averages/standard deviations of the weights. (c) RESET switching: Develop of memristor's conductance (measured at 0.1 V) under the effect of 500 μs pulse trains of several magnitudes. (d) SET switching: this figure shows the conductance change of a fixed range pulse repeatedly in application to the same device. Reproduced from [33], Springer Nature Ltd. (a-d).

However, if the training time is long, a large amount of training data are needed, which does not guarantee the better results and complete reliable. Also, it is inclined to fall into local minimum and does not acquire incremental learning ability, the stored information interferes with each other and degrades. CNN needs to program thousands of codes to complete, which is very troublesome. It requires check debug, complicated circuit and high labor cost. But the characteristics of the memristor synapse, combining neural network computing with a memristor, can fuse the units which are combined with data storage and processing, greatly increasing the parallelism of data transmission and processing. M. Prezioso. et al used an integrated neuromorphic network based on metal-oxide memristor to recognize alphabet Z, V and N [8]. In Fig. 5a and Fig. 5b, the experiment showed that weight is expressed by the change of a single conductance. Because the physical difference of each memristive device, the conductance of the initial state is not same, and the different characterization is reflected in the color difference of the panels, so the voltage is used for the memristive device on the node for initialization. This procedure was repeated three times, with different pulse magnitudes, for both set and reset transitions (See in Fig. 5c and Fig. 5d). When the alphabet was identified, they were respectively coded into voltage pulse signals, and then the array was stimulated with pulses to obtain a conductivity map. For example, when 0.1 V was used to represent z, the initial $V_0$ was 1%, and the target weight w is 40%, and then the measured $V_1$ was 40%, so $|V_0-V_1|=39\%$, infinitely close to 40%, indicating a match.

IV. CONCLUSIONS AND PERSPECTIVES

In the Big Data Era, the current system based on Von Neumann is facing challenges. The invoking between computing and storage consumes too much time and energy. In order to break through this pattern, we need a more efficient system. The human brain is an ideal chip, the synapses in the brain are verified and related to learning and memory, achieving learning storage implemented in one structure. However, the current artificial neural network based on algorithms does not break through the von Neumann architecture. And transistors and flash memory do not have any progress space. Therefore, for the purpose of solving the above two issues, the field of memristor is booming. The nanoscale size, low power, non-volatile and memory functions of the memristor make it widely used in the fields of resistive random access memory, neural networks, cross-array, signal processing and circuit design. Especially in the field of biological function simulation. Biggest advantage of the nano-level memristor device acting as a synapse of the neuromorphic system is that the memristor can maintain the state of the internal resistance by applying voltage and current. At this stage, the single memristor device has made breakthroughs in STP, LTP, STDP, etc. through metal oxides. The practical application of memristor-based arrays has also begun, for example, E. J. Merced-Grafals et al. study about the test and operation of an integrated neural morphology network based on metal oxide memristors. Also, the memristor on physical structure is a mapping of neural networks evolved from a multi-layer perceptron. The combination of neural network and memristor will greatly improve the data transmission and processing capabilities.

In recent years, to mimic the human brain function, scientists have worked painstakingly. Software-based neural network has been put into operation, but its computational storage of large-scale data is lacking in adaptability to complex inputs, and its huge cost also brings difficulties. Now scientists shift the focus to hardware research inspired by the human brain, neural network computing is a
reasonable solution. One of the keys for successful operation is to realize the synaptic function of electronic devices. First of all, the key components should be small enough to meet the basic requirements of synapses. The other one is to realize the operation and storage as one unit, breaking through the structure of Von Neumann, to achieve energy saving and performance improvement. However, in order to satisfy both aspects, the material of the electronic synapse may take into account the strong plasticity and durability, the plasticity acts on the space occupation. In practical applications, it is necessary to ensure the heat resistance of a single memory when calculating and storing in the same unit. Of course, what is needed is not only the possibility of a single device implementing neurons in terms of storage and conduction, but also the integration of nanoscale integrated circuits.[51, 52] The development of hardware-based artificial neural networks to mimic brain operations is the best choice. The memristor of 3D stacking structure breaks through the limitations of horizontal integration. Researchers have more choices in the vertical direction and circuit design. Consequently, it is more complicated, not only the interference of the horizontal current, but also the effect of the vertical coil.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
Ye-Guo Wang conducted the research and wrote the paper; Ye-Guo had approved the final version.

REFERENCE


Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Ye-Guo Wang was born in Shandong, China on November 13, 1998. He is currently a fourth-grade student. He is studying at Qingdao University of Science and Technology, No.99 songling Rd, Qingdao, Shandong, P.R. China, 260061.