Research Progress of Neural Synapses Based on Memristors

Yamin Li 1,2, Kang Su 1,2, Haoran Chen 1,2, Xiaofeng Zou 1,2, Changhong Wang 1,2, Hongtao Man 1,2, Kai Liu 1,2, Xin Xi 1,2 and Tuo Li 1,2,*

1 State Key Laboratory of High-end Server & Storage Technology, Inspur (Beijing), Electronic Information Industry Co., Ltd., Beijing 100089, China; liyamin@inspur.com (Y.L.); sukang@inspur.com (K.S.); chenhaoran@inspur.com (H.C.); zouxf@inspur.com (X.Z.); wangchh01@inspur.com (C.W.); manht@inspur.com (H.M.); liukaibj@inspur.com (K.L.); xixin01@inspur.com (X.X.)
2 Shandong Inspur Artificial Intelligence Research Institute Co., Ltd., Jinan 250101, China
* Correspondence: lituo@inspur.com

Abstract: The memristor, characterized by its nano-size, nonvolatility, and continuously adjustable resistance, is a promising candidate for constructing brain-inspired computing. It operates based on ion migration, enabling it to store and retrieve electrical charges. This paper reviews current research on synapses using digital and analog memristors. Synapses based on digital memristors have been utilized to construct positive, zero, and negative weights for artificial neural networks, while synapses based on analog memristors have demonstrated their ability to simulate the essential functions of neural synapses, such as short-term memory (STM), long-term memory (LTM), spike-timing-dependent plasticity (STDP), spike-rate-dependent plasticity (SRDP), and paired-pulse facilitation (PPF). Furthermore, synapses based on analog memristors have shown potential for performing advanced functions such as experiential learning, associative learning, and nonassociative learning. Finally, we highlight some challenges of building large-scale artificial neural networks using memristors.

Keywords: memristor; synapse; analog; digital; artificial neural networks

1. Introduction

With the advent of artificial intelligence, big data, and the Internet of Things (IoT), data are growing exponentially, posing significant challenges to data storage and computation. In the traditional von Neumann architecture, the calculation and storage of data are separated. In the face of massive data, there are problems with memory walls [1–3]. The human brain can compute, store, and execute thousands of processes simultaneously with high efficiency and robustness [4]. The brain comprises neurons and synapses, interconnected to form a vast neural network that processes massive amounts of data with low power consumption. Understanding how the brain works can provide valuable insights for creating new structures and systems. Therefore, developing brain-inspired intelligent devices and constructing brain-inspired computing systems are critical to breaking through existing bottlenecks [5–7].

The memristor emerges naturally as the fourth fundamental electronic element after the resistor, capacitor, and inductor [8]. L. Chua first proposed the memristor concept in 1971 [9], and HP researchers produced the prototype device of a memristor in 2008 [10]. Subsequently, memristors have been developed based on the connection and fracture of conductive filaments, the oxidation–reduction reaction, the phase transition, and other resistive mechanisms [11–13]. Regarding physical structure, according to the number of memristor terminals, it can be divided into two-terminal memristors and three-terminal memristors. Compared with the two-terminal memristor, the three-terminal memristor

Citation: Li, Y.; Su, K.; Chen, H.; Zou, X.; Wang, C.; Man, H.; Liu, K.; Xi, X.; Li, T. Research Progress of Neural Synapses Based on Memristors. Electronics 2023, 12, 3298. https://doi.org/10.3390/electronics12153298

Received: 26 June 2023; Revised: 26 July 2023; Accepted: 30 July 2023; Published: 31 July 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
has an additional modulation port, and the two-terminal memristor has a higher similarity with the biological synapse [14,15]. A memristor is an excellent option for creating neural synapses due to its high density (4F; F is feature size) [16], nonvolatile nature, and compatibility with CMOS (complementary metal oxide semiconductor) during fabrication [17]. Memristor-based artificial synapses offer several advantages such as linear weight updating, stable resistance, fast running speed (<1 ns) [18], and low energy consumption (the most advanced consumption of each peak can be as low as few femtojoules or even sub-femtojoule [5]). Synapses in biological systems are responsible for transmitting weighted spikes from presynaptic neurons to postsynaptic neurons. The electrical characteristics of a memristor are highly similar to the physical attributes of neural synapses, and its working mechanism is based on ion migration. A single memristor can simulate the function of neural synapses while reducing chip area and power consumption compared to synapses composed of multiple transistors and capacitors [19]. A memristor can be categorized as digital or analog, depending on its resistance conversion mode. The digital memristor has abrupt resistance, while the analog memristor has gradual resistance. These two types can be used to construct digital and analog synapses.

2. Synapses Based on Digital Memristors

The HP Lab discovered a voltage–current curve consistent with the ideal memristor in the sandwich structure of Pt/TiO2/Pt (metal/insulation layer/metal), as shown in Figure 1a [10]. Since this breakthrough, researchers have developed two-terminal memristors with a similar structure by modifying the metal and insulating layer materials. The two-terminal memristor model is popular because of its versatility and simplicity [20]. Memristors can be categorized into digital and analog types based on the resistance conversion mode. The digital memristor’s resistance changes discretely based on its voltage sweep. Two high- and low-resistance values correspond to the logical ‘0’ and ‘1’ states from a memory perspective, as shown in Figure 1b. The resistance change of a digital memristor has an apparent threshold. If the voltage is scanned forward and goes beyond the positive threshold, its resistance will change from high to low. On the other hand, if the voltage is scanned in reverse and goes beyond the negative threshold, its impedance will vary from low to high. However, the resistance remains unchanged if the voltage is applied between the two thresholds. This characteristic can be used to simulate the synapses of neural networks. Spikes can be transmitted from presynaptic neurons to postsynaptic neurons when there is low resistance, but transmission is prevented when there is high resistance. To create artificial neural networks through hardware, synapses are required to connect numerous neurons to perform complicated calculations [21]. Synapses must complete information transmission, and synaptic weights can be accurately stored and changed under specific circumstances [22]. A memristor has a memory function, particularly a digital one. The memristor’s resistance can be changed by applying a voltage beyond the threshold, similar to altering synaptic weights. When the voltage is below the threshold, the memristor’s resistance remains unchanged and is comparable to weight maintenance. The weight is stored in the memristor cross array. When the neuron gives the input, the output can be obtained according to Ohm’s and Kirchhoff’s laws, which genuinely realize the integration of memory and computation. In addition, the memristor’s nano-size and low-power computation are also the advantages of the synapse array. Digital memristor synaptic arrays can be constructed to meet the needs of large-scale integrated circuits [23]. A digital memristor is used to build a neural synapse array to realize the spiking neural networks (SNNs). The synapse weight in SNNs can be positive, negative, or zero. The synapse weight in Sections 2.1 can only be positive and zero, while those in Sections 2.2–2.4 can be positive, negative, and zero.
2.1. One-Memristor Synapse

Simulating the synaptic function through 1M (one memristor) is the simplest connection method in the neural network. In Figure 2a, one end of the memristor is connected to the presynaptic neuron, while the other is connected to the postsynaptic neuron. The memristor determines the relationship between output and input, and the memristor’s resistance can be used as the weight in the neural network. The memristor’s resistance is related to the input voltage and the initial state. When the initial resistance is low or high, and the applied voltage is between the positive and negative thresholds, the synaptic weight remains at ‘1’ or ‘0’. Due to its small area, low power consumption, and high integration density, it is easy to construct large-scale synaptic arrays. However, when a single memristor is used as a synapse to build a synaptic cross array, as shown in Figure 2b, there is a sneak path problem. When reading/writing the resistance of a memristor and there are memristors with low resistance around, crosstalk will occur, limiting the accuracy of the reading/writing.
2.2. Two-Memristor-Two-Resistor Synapse

In the double-memristor synaptic circuit, a parallel or series topology junction is similar to a single-memristor synapse, which cannot gain negative weight. By adding two resistors to form a 2M2R (two-memristor-two-resistor) synaptic circuit, the weighted state of the synapse can be increased. Its connection mode is shown in Figure 3a. On each branch, a memristor and a resistor are connected in series. The voltage difference between points A and B is used as the output voltage.

![Figure 3. (a) Schematic diagram of two-memristor-two-resistor synaptic circuit. (b) Schematic diagram of four-memristor synaptic circuit.](image)

When the input voltage $V_{in}$ exceeding the threshold is loaded at the input end, the memristor’s resistance will change. When $V_{in}$ is greater than the positive threshold voltage, the resistance of $M_1$ will decrease, and $M_2$ will increase; when $V_{in}$ is less than the negative threshold voltage, the two memristors have opposite changes. According to the voltage division formula, the relationship between the output and the input voltage can be obtained as follows:

$$V_{out} = V_A - V_B = \left( \frac{R_1}{M_1 + R_1} - \frac{R_2}{M_2 + R_2} \right) V_{in} \quad (1)$$

We name the ratio of output to input $g$, which represents the synaptic weight of the circuit.

$$g = \frac{R_1}{M_1 + R_1} - \frac{R_2}{M_2 + R_2} = \frac{R_1M_2 - R_2M_1}{(M_1 + R_1)(M_2 + R_2)} \quad (2)$$

To simplify the model, if two resistance values are equal, the expression for synaptic weight is as follows:

$$g = \frac{R_1 (M_2 - M_1)}{(M_1 + R_1)(M_2 + R_2)} \quad (3)$$

From the expression, it can be seen that the synaptic weight is as follows:

$$\begin{cases} 
g > 0 & M_1 < M_2 
g = 0 & M_1 = M_2 
g < 0 & M_1 > M_2 \end{cases} \quad (4)$$

The formula shows that the 2M2R circuit can achieve positive, negative, and zero synaptic weights by cooperating with a differential circuit. Similarly, replacing $M_2$ with a resistor can also achieve positive, negative, and zero synaptic weights [22].

2.3. Four-Memristor Synapse

The memristor bridge synaptic circuit (four-memristor, 4M) resembles the Wheatstone bridge [25]. Figure 3b shows that it comprises four identical memristors, using the difference between points A and B as the output voltage. The resistance of each memristor depends on its initial resistance and the input voltage. When the input voltage exceeds the threshold, the resistance of the memristor will change. When the input voltage exceeds
the positive threshold, the resistance of $M_1$ and $M_4$ decreases, and the resistance of $M_2$ and $M_3$ increases. The input voltage is $V_{in}$, and the output voltage can be obtained by dividing the voltage formula as follows:

$$V_{out} = V_A - V_B = \left(\frac{M_2}{M_1 + M_2} - \frac{M_4}{M_3 + M_4}\right)V_{in}$$

(5)

Similarly, using $g$ to represent the synaptic weight, the value is:

$$g = \frac{M_2M_3 - M_1M_4}{(M_1 + M_2)(M_3 + M_4)}$$

(6)

It can be seen from the formula of synaptic weight that when $M_2M_3 > M_1M_4$, the synaptic circuit can achieve positive weight; when $M_2M_3 = M_1M_4$, it can gain zero weight; and when $M_2M_3 < M_1M_4$, it can gain negative weight. Compared with the single synaptic circuit, the memristor bridge circuit increases the number of memristors. Still, this synaptic structure can more fully simulate the weight and is more suitable for memristor neural networks. Compared with the 2M2R circuit, it also has four components. However, the memristor is smaller than the resistor, making it ideal for reducing the chip area and suitable for large-scale integration.

2.4. Other Digital Synaptic Structures

Chen et al. designed a synaptic circuit composed of three memristors and two transistors [26], as shown in Figure 4. The input voltage is $V_{in}$, and the output $V_{out}$ is the voltage difference between A and B. The threshold voltage of the transistor is $V_T$. When $V_1 > V_T$ and $V_2 \geq V_T$, transistors $T_1$ is on, and $T_2$ is off. At this time, the output voltage is shown in expression (7):

$$V_{out} = V_A - V_B = V_{in} - \frac{M_3}{M_2 + M_3}V_{in} = \frac{M_2}{M_2 + M_3}V_{in}$$

(7)

![Figure 4. Presented spintronic memristor-based synaptic circuit.](image)

When $V_1 \leq V_T$ and $V_2 < V_T$, transistor $T_2$ is on and $T_1$ is off. At this time, the output voltage is shown in expression (8):

$$V_{out} = V_A - V_B = \frac{M_1}{M_2 + M_1}V_{in} - V_{in} = -\frac{M_2}{M_2 + M_1}V_{in}$$

(8)

When $V_1 > V_T$, $V_2 < V_T$ (or $V_1 \leq V_T$, $V_2 \geq V_T$), transistors $T_1$ and $T_2$ are both in the on (or off) state, and $V_{out} = 0$. Therefore, synaptic weight $g$ can be expressed as expression (9):

$$g = \begin{cases} 
\frac{M_2}{M_2 + M_3} & V_1 > V_T, V_2 \geq V_T \\
0 & (V_1 > V_T, V_2 < V_T) \text{ or } (V_1 \leq V_T, V_2 \geq V_T) \\
-\frac{M_2}{M_2 + M_1} & V_1 \leq V_T, V_2 < V_T
\end{cases}$$

(9)
From the expression of the synapse, it can be seen that this circuit can realize the setting of synaptic weights to positive, negative, and zero. In addition, a neuron based on a memristor is designed. Based on this neuron and synapse, a compact neural network is built. When the random weight change (RWC) algorithm is applied to this network, it results in faster training speed and fewer training errors than traditional RWC.

2.5. Summary of Synaptic Circuit Based on Digital Memristors

Digital synaptic circuits mainly use the nonvolatility and resistance variability of the memristor. It focuses on achieving the completeness of weight symbols (positive, negative, and zero weights) and maximizing linear variation. When using the memristor to construct digital synapses, the biological function of simulating neural synapses cannot be realized. Still, the weight is directly stored in the memristor, realizing the integration of storage and computation, reducing data handling, and improving computing efficiency. According to the variable resistance characteristics of the memristor, the weight value can be changed without replacing the hardware equipment to implement the neural network’s reusability. In addition, the nano-size and low power consumption make it advantageous for building large-scale neural networks, which is conducive to the hardware integration and implementation of neural network models. It can significantly accelerate the neural network training and testing process with corresponding hardware.

3. Synapses Based on Analog Memristors

The resistance of the analog memristor is polymorphic, which can continuously change according to the input voltage and remember the altered resistance value, as shown in Figure 5a. Figure 5b compares the electronic synapses and biological synapses. From the electrical characteristics, the analog memristor is more similar to the features of the biological synapse. Structurally, the two terminal memristors’ top electrodes, the middle insulating layer, and the bottom electrodes correspond to the presynaptic membrane, synaptic cleft, and postsynaptic membrane of the biological synapses. Because the memristor’s size is at the nano-level, it is expected to achieve the same number of brain synapses on VLSI (very-large-scale integration) and build a bionic brain. From the perspective of the working mode, external stimuli affect the transmission capacity of both synapses and memristors. The transmission capacity of synapses can be changed by the flow of certain specific ions like Ca$^{2+}$, Na$, and Cl$^-$ inside and outside the cell membrane. Meanwhile, the memristor is realized by the transport of oxygen or metal ions. Moreover, they have a particular frequency dependence on external stimuli. Neurons will show a refractory period for stimulation with too high a frequency. Similarly, if the memristor is stimulated with a high frequency, its hysteresis curve will become a single value curve, which causes the loss of its memory characteristics. Synapse circuits based on analog memristors are often built by one or several memristors. Then, the characteristics of biological synapses can be simulated, mainly including STDP, SRDP, PPF, and other synaptic plasticity. Experience learning and nonassociation/association learning circuits are also built based on these characteristics. When using analog memristors to realize brain-inspired neural computing systems, it is necessary to make the functions of artificial synapses highly consistent with that of actual brain synapses. In Section 3.1, we discuss the basic functions of brain synapses that can be replicated by artificial synapses, while Sections 3.2–3.4 cover advanced functions.
3.1. Basic Synaptic Characteristics of Analog Memristor

Brain-inspired computing is expected to completely change traditional architecture and lead a new generation of powerful computing. A memristor is considered a potential solution for synapses [28]. The memristor has a natural synaptic-like performance, especially for multi-resistance analog memristors with adjustable weight. It can simulate the brain’s information processing methods for computation, achieving synaptic-like performance such as STM to LTM, STDP, SRDP, and PPF [29–33].

3.1.1. Short-Term Memory to Long-Term Memory

Human memory is divided into short-term memory (STM) and long-term memory (LTM) [34]. Correspondingly, synaptic behavior can be divided into long-term plasticity (LTP) and short-term plasticity (STP) [35]. LTP is believed to be related to learning and memory functions. In contrast, STP is often closely related to the critical computational operations of spatiotemporal information processing in biological systems [36]. STP is a short-term enhancement of neuronal connectivity, generally lasting for a few minutes or less. LTP permanently enhances neuronal connectivity, usually lasting several hours or years [37]. The main principle of implementing STP in devices is establishing an unstable conductive state under voltage conditions and restoring the original stable state after a relaxation process. Like biological memory, forgetting is a natural part of all memories, and the memristor’s weight decreases as a result [38]. However, the transformation from STM to LTM can be achieved through training, thus extending the forgetting time, as shown in Figure 6a. This characteristic is one of the fundamental characteristics of synaptic-like devices [39]. Rahmani et al. designed a double-layer material AIZS/CsCuCl5-based memristor and showed the transition from STM to LTM under the repeated same pulse input [34]. Figure 6b shows that when 10 pulses are applied continuously, the increased current gradually decays once the voltage pulse is removed, showing STP characteristics. When 30 pulses are used, the current increases to a higher level and tends to be stable, as illustrated in Figure 6c. Comparing the conductance at the initial stimulus, final stimulus, and 60 s after removing stimulation by different pulse numbers, the attenuation of conductance is slight by 30 pulses, thus realizing the transition from STM to LTM, as shown in Figure 6d. Memristors based on lignin [27], WOx [40], Ag2S [41], NiO [42,43], and TiOx [44] also simulated the characteristics of the transition from STM to LTM.

3.1.2. Spike-Timing-Dependent Plasticity

STDP is a typical case of long-term plasticity [45]. STDP is a mechanism that regulates the synaptic intensity through the relative time difference between pre- and postsynaptic pulses. It is the theoretical basis for learning, information storage, and the development and perfection of neuronal circuits during brain development [46]. It is considered one of
the primary learning rules of unsupervised learning. The specific rules are that if the presynaptic pulse arrives earlier than the postsynaptic pulse, the synapse triggers LTP, increasing synaptic weight and improving signal transmission efficiency between the pre- and postsynaptic neurons. Otherwise, it will lead to LTD. The effect of enhancement and suppression is related to the time difference between the arrival of the front and back pulses—the smaller the time difference, the more pronounced the enhancement/suppression effect [47]. To follow STDP, the memristor synapse should meet the requirements of conductance gradient and rapid response to a single spike [48]. The analog conductance state with linear and symmetric LTP/LTD is especially essential in reasoning and training hardware accelerators. Formula 10 is the expression of weight change in the memristor.

$$\Delta \omega = \begin{cases} A_+ \exp \left( \frac{-\Delta t}{\tau_+} \right), & \Delta t > 0 \\ A_- \exp \left( \frac{\Delta t}{\tau_-} \right), & \Delta t < 0 \end{cases}$$

(10)

where $A_+$, $A_-$, $\tau_+$, and $\tau_-$ are constants. $\Delta t = t_{\text{post}} - t_{\text{pre}}$, where $t_{\text{post}}$ and $t_{\text{pre}}$ are the time constants for pre- and postsynaptic firings, respectively.

In general, the STDP rule can be realized by applying overlapping pulses at both ends of the memristor. The effective pulse is the difference between the pulses received at both ends. The overlap method can convert the pulse timing information into the pulse amplitude information. When the presynaptic pulse arrives earlier than the postsynaptic pulse, the positive part of the practical pulse amplitude is more significant, making the synapse produce LTP. However, when the presynaptic pulse arrives later than the postsynaptic pulse, the negative part of the practical pulse amplitude is more significant, which can make the synapse produce LTD. The shorter the interval between the arrival of two pulses, the larger the adequate pulse amplitude and the more significant the change in synaptic weight caused. Li et al. designed a synapse based on a sulfur memristor, which can achieve the STDP rule, as shown in Figure 7 [49]. In addition, the STDP rule has been observed in memristors based on CuCrO$_2$ [50], Al$_2$O$_3$/TiO$_2$-x [51], WO$_3$ [52], and TaO$_x$ [53].
Figure 7. STDP implementation. (a) Pulse scheme for realizing STDP. (b) STDP rule for memristor [49].

However, the non-overlapping method can be used for the second-order memristor to implement the STDP rule. The change in the conductivity of the second-order memristor depends on the first variable and the second variable. In neurobiology, the relative time information between pulses is naturally embedded and not dependent on external factors. For example, through the natural decay of Ca$^{2+}$, neurons respond to changes in synaptic response intensity, providing an internal time mechanism [54]. The non-overlapping method is more similar to the working mode of actual biological synapses. For example, different synaptic functions can be achieved by utilizing the internal oxygen holes in the oxide memristor at different time scales. Or, the change in local temperature can be used too. The second variable usually turns into transient dynamics, such as the dynamic decay after excitation, similar to the dynamics of calcium ions in biological synapses. Kim et al. proposed a Ta$_2$O$_5$-based second-order memristor whose conductive channel width depends on the device’s conductivity. In contrast, the device’s local temperature can adjust the ions’ mobility in the oxide. The first pulse is followed by the second, which can alter the local temperature. Temperature affects ions’ mobility, leading to conductivity changes [55]. The second-order memristor models based on different materials have been discovered successively, such as TiO$_2$:Ag [56], WO$_3$–x [57], SrTiO$_3$ [58], and InGaZnO [59].

3.1.3. Spike-Rate-Dependent Plasticity

SRDP is a mechanism that changes synaptic plasticity by applying pulse series [60]. In biological learning, time and frequency are the main encoding methods for information. The stimulation mode of pulse series, which is generated by multiple connected neurons, is more commonly used than single or paired spikes. In synaptic devices, activity is not only related to time but also to frequency. SRDP relies on the Bienenstock Cooper Munro (BCM) rule as its higher-order function. According to the BCM rule, LTP will be generated when the frequency exceeds the threshold. When the frequency is lower than the threshold, LTD is caused. The threshold is a historical activity value related to previous synaptic activity. In addition, the difference between the frequency of presynaptic and postsynaptic pulses can also determine the size and direction of the weight change [49]. Figure 8a shows the synaptic weight shift with the number of pulses and pulse interval in a Cu$^{2+}$-doped KNbO$_3$ memristor [61]. Boppidi et al. designed a CuCrO$_2$-based memristor. The measured conductance is 120 mS and 75 mS at 1 MHz and 200 kHz, respectively. And it can be maintained for 1000 s, exhibiting LTP and LTD characteristics. The conductivity change rate increases as the difference in input frequency between presynaptic and postsynaptic increases [50]. In addition, changes in synaptic weight depend on historical states in the SRDP rule. The same stimulus may cause opposite changes in synaptic weight, as shown in Figure 8b [62]. Synaptic weight increases when subjected to stimulation below 25 Hz.
but decreases with stimulation at 7 Hz. However, with continuous administration of 2 Hz inspiration, the synaptic weight increases again when subjected to 7 Hz stimulation.

Figure 8. (a) Synaptic weight as a function of pulse number and interval [61]. (b) Synaptic weight response to a set of postsynaptic spike trains with the frequency sequence [62].

3.1.4. Paired-Pulse Facilitation

PPF is one of the typical manifestations of STP. The human brain’s STP or STM can generally last from a few seconds to a few minutes. In biology, when two pulses are applied to the synapse, the response current caused by the second pulse will be greater than the first. The smaller the interval between the two pulses, the stronger the enhancement effect. Figure 9a reproduces the PPF phenomenon on guinea pig hippocampal cells [63]. The PPF effect is due to the first peak causing residual calcium ions in presynaptic neurons, which are released when the second peak arrives. This leads to an increase in the overall calcium concentration and the production of EPSC (empirical postsynaptic current). The strength of the effect decreases as the time between the two peaks elongates due to the exponential decay of residual calcium ions. Li et al. observed the PPF phenomenon on the CeO$_2$/Nb–SrTiO$_3$ memristor [62], as shown in Figure 9b. The change in synaptic weight is reflected through conductance, and a longer time interval will lead to a minor conductance change. PPF also be observed in La:HfO$_2$ [64], tungsten oxide [57], Zinc Oxide/Poly (3-hexylthiophene) [65], AlO$_x$/HfO$_x$ [66], SiNx [67], and Pt/HfO$_2$/TaO$_x$/TiN [68] based memristor.

Figure 9. (a) PPF on hippocampal cells in guinea pig [63]. (b) PPF is observed on CeO$_2$/Nb–SrTiO$_3$ heterojunction memristors [62].

3.2. Learning Experience

Psychologist Hermann Ebbinghaus’s research on human memory shows that we can turn STM into LTM by strengthening the stimuli. The first learning can be attained with less time or repetition during the second stimulus. This phenomenon is called the learning experience [64,69,70]. The synaptic weight in devices increases as the number of pulses
increases. However, without external input, the conductivity decreases spontaneously, causing the weight to drop. Nevertheless, it does not decay to the initial state, like human memory follows a forgetting curve [71]. Then, by stimulating again during the decay process with less stimulation, the synaptic weight can be restored to a high level and increase the low memory level. It requires far fewer pulses to reach the high memory level after repeated learning and forgetting. This phenomenon is similar to the experience-based knowledge of biological systems, where memory recovery requires less stimulation over time [32]. Zhao et al. designed a photoelectric memristor based on a zinc oxide/poly (3-hexylthiophene) heterojunction, simulating the learning experience process of the human brain through three cycles of learning and forgetting [65]. After stimulation with 100 continuous light pulses (100 ms, 5 Hz, 1.14 W cm⁻²), synaptic weight gradually increases and spontaneously decays for 100 s, reaching an intermediate state. In the second and third stages of learning, it only takes 16 and 7 pulses to restore the synaptic weight to the level achieved in the first stage, as shown in Figure 10. Additionally, the rate of attenuation is slower. Both electrical and optical pulse stimulation can make the synapse show a learning experience process. The type of stimulation used depends on the physical mechanism of the device’s high- and low-impedance changes. In addition, the p⁺-Si/n⁻-ZnO heterostructure [72], Pt/BiFeO₃/SrRuO₃-based ferroelectric second-order [73], sponge-like double-layer porous oxide [74], and p-NiO/n⁻-ZnO heterostructure memristor [75] both can simulate the experiential learning process of biological brains. Other teams also conduct mathematical modeling and analysis of this process [76].

![Figure 10. Learning experience behavior of the zinc oxide/poly (3-hexylthiophene) heterojunction-based device [65].](image)

### 3.3. Associative Learning

Associative learning is a standard method of self-learning in biological organisms. It allows the nervous system to simultaneously remember the connection between two events [77,78]. Reconstructing associative memory at the behavioral level is of great significance. A method similar to the brain’s self-learning neural morphological system can be designed [79]. Pavlov’s classically conditioned reflex is a typical example of associative learning. In the experiment of Pavlov’s dogs, food as the unconditioned stimulus (US) will cause dogs to drool. Before training, the bell rings, as the conditioned stimulus (CS) cannot make dogs drool [80]. In an experiment, dogs learn to associate the sound of a bell with food by being fed after the bell rings. With repeated training, the dogs drool at the sound of the bell. This process indicates that food and the bell have established associative memory. Subsequently, only the bell ring without food, or the food is given first, and then the bell rings. After several training sessions, the bell no longer causes drooling again, indicating a loss of associative memory between the food and the bell [81]. Guo et al. designed a multi-value analog Sr₀·₉Ba₀·₁TiO₃ₓ (SBT) memristor with threshold characteristics. Based on the memristor, a Pavlov associative memory circuit is built. The memristor is the synaptic part, as shown in Figure 11a [82]. The pre-neuron creates the pre-spike, while the postsynaptic neuron triggers the post-spike. The subtractor then generates a control signal called \(V_{out}\), which is used to regulate the weight of the memristor. There are
three neurons involved in this process. The neurons A, B, and C represent the sounds of ringing, food, and a dog, respectively. If neuron A is stimulated independently, neuron C produces no signals. This means that the dog does not drool when it hears the bell. When presynaptic neuron B is stimulated, neuron C produces spike signals. That is, food as an unconditional stimulus can make the dog salivate. When neuron A is stimulated before neuron B, the synaptic weight between neuron A and neuron C increases, and the conductivity stabilizes at 0.0397 S; at this point, the associative memory of food and ringing tones is established. Afterward, when neuron A is stimulated alone, it can cause neuron C to generate spike signals. The schematic diagram of the process of establishing associative memory is shown in Figure 11b.

Numerous researchers have developed Pavlov circuits utilizing memristors. They have expanded the capabilities of these circuits after identifying their fundamental functions. For example, Sun et al. created a neural network circuit capable of time delay Pavlov associative memory [83]. They also designed a Pavlov associative memory circuit that can recognize multiple neurons [84]. Additionally, Shang et al. developed circuits that can execute the generalization and differentiation of Pavlovian associative memory [85]. In addition to Pavlovian associative courses, there are other associative circuits based on memristors. For example, Liao et al. designed emotional associative memory neural network circuits [86]. Shi et al. proposed online learning bidirectional associative memory network circuits [87]. Memristor-based associative circuits can simulate human associative memory learning behavior. It is significant for constructing neural networks for brain-inspired computing [88–91].

3.4. Nonassociative Learning

Nonassociative learning is distinct from associative learning. It does not necessitate a particular link between stimuli and response, also called simple learning [92]. In biology, nonassociative learning is crucial in how organisms respond to stimuli. It is divided into habituation and sensitivity and plays an essential role in evolutionary adaptation. Habituality can help organisms reduce their response to useless repetitive stimuli and filter out
useless information; Sensitivity can gradually increase in response to strong and harmful reaction used to enhance helpful information. Habituality and sensitivity can last several days or weeks, depending on the stimulus pattern. The schematic diagram of nonassociative memory construction is shown in Figure 12a. Zhao et al. simulated nonassociative learning through a zinc oxide nanowire memristor without any other peripheral control circuit [93]. The memristor showed habituation and sensitization behavior under electrical and light stimulation. The memristor showed habituation under continuous 6 V pulse stimulation and sensitization under −3 V pulse stimulation. At the same time, ultraviolet light can sensitize the memristor. Different UV light intensities can lead to varying degrees of sensitization. Hong et al. designed a memory circuit to achieve nonassociative learning [94]. This circuit includes a synaptic module (Figure 12b), a neuron module, and a feedback module. It can generate an output signal with biological nonassociative learning characteristics, and its amplitude varies with the features of the input signal. Habituation behavior is generated under continuous weak stimuli below 30 mV, while sensitization behavior is generated under harmful stimuli above 30 mV. Sun et al. designed a memristor circuit to implement nonassociative learning under different emotions [95]. Nonassociative learning can improve the efficiency of neural network circuits in information processing and reduce the computational load of the circuit.

![Figure 12](image)

**Figure 12.** (a) Construction process of biological nonassociative memory. (b) Synaptic circuit of memristor for nonassociative memory.

### 3.5. **Summary of Synapses Based on Analog Memristors**

The analog memristor has several stable resistance, and the change in the resistance is related to the amplitude and frequency of the pulse. The analog memristor behaves like a natural synapse and can be used alone or with other electronic elements to create synaptic circuits. Various properties of biological synapses, including basic characteristics and advanced learning and memory skills, can be simulated. Memristors based on multiple materials can simulate one or more attributes of neural synapses. Table 1 summarizes the biological factors that significant memristors can replicate. Neuromorphic computing can simulate the operation of human brains and can be physically realized by memristive arrays. What is more, the memristor’s conductance at each cross-point of the array can be mapped to the analog weight of neural networks, such as CNNs (convolutional neural networks), DNNs (deep neural networks), and RNNs (recurrent neural networks). Scientists aim to create an artificial neural network by replicating the brain’s intricate learning process. This breakthrough technology will overcome current limitations in computing.
and storage and use memristors to develop a brain-inspired chip that can perform natural brain-like computations.

**Table 1.** The biological factors that significant memristors can act.

<table>
<thead>
<tr>
<th>Year</th>
<th>Materials</th>
<th>STM to LTM</th>
<th>STDP</th>
<th>SRDP</th>
<th>PPF</th>
<th>Learning Experience</th>
<th>Associative Learning</th>
<th>Nonassociative Learning</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>ZnO</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Yes</td>
<td>/</td>
<td>[93]</td>
</tr>
<tr>
<td>2022</td>
<td>Zinc Oxide/Poly (3-hexylthiophene)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>[65]</td>
</tr>
<tr>
<td>2022</td>
<td>Carbon</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Yes</td>
<td>/</td>
<td>[96]</td>
</tr>
<tr>
<td>2022</td>
<td>AlOx/HfOx</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>/</td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td>CeO2/Nb−SrTiO3</td>
<td>/</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>/</td>
<td>[62]</td>
</tr>
<tr>
<td>2023</td>
<td>Pd/La:HfO2/La0.33Sr0.67MnO3</td>
<td>Yes</td>
<td>Yes</td>
<td>/</td>
<td>Yes</td>
<td>/</td>
<td>Yes</td>
<td>/</td>
<td>[64]</td>
</tr>
</tbody>
</table>

4. Discussion

Using memristors to achieve in-memory computing and brain-inspired neuromorphic computing systems remains challenging. Building circuits and systems for new types of computing requires reliable and efficient memristors and arrays. The memristor is the cornerstone of the whole system and plays a vital role in overall performance improvement. New memory devices such as resistive random access memory (ReRAM), phase change memory (PCM) [11,97], ferroelectric field-effect transistor (FET) [98,99], and magnetoresistive random-access memory (MRAM) [100,101] can be used to build synapses. Each device has its advantages and disadvantages. Depending on the specific application requirements, one type of device may outperform others. Taking MRAM as an example, it boasts a theoretically permanent lifespan, but it faces challenges related to low switch ratio and large size. On the other hand, the ReRAM offers compatibility with CMOS technology, low power consumption, and analog conductivity modulation. Its most significant advantage is its resemblance to biological synapses on the physical level, making it one of the best choices for constructing neural synapses. However, some issues need to be addressed [102]. Research on memristors is primarily in the scientific research stage, and further large-scale industrialization is necessary. The main challenges faced are discussed.

4.1. Uniformity and Consistency

Different memristor devices exhibit significant differences in conductivity, switching speed, reliability, and other characteristics. Poor uniformity in synaptic circuits can cause performance degradation and instability issues. It is essential to enhance the manufacturing process and ensure better uniformity of instruments. One way to improve the consistency of material properties is to optimize the film deposition and annealing process. Another approach is to utilize adaptive learning algorithms to address performance variations caused by device non-uniformity. Moreover, designing synaptic circuit structures with greater fault tolerance and robustness can also enhance performance.

4.2. Stability and Reliability

Inconsistent resistance values during programming and reading can lead to unreliable data storage. This will affect the accuracy of synaptic weight transmission and hinder its application in analog neuromorphic systems. Factors such as the relaxation effect, diffusion/recombination of the oxygen vacancies, and formation/rupture of multiple weak conductive filaments can all lead to unstable resistance. Numerous researchers have addressed this issue from the perspectives of devices, arrays, systems, algorithms, and more. Studies have found that using Ta/HfO2 hybrid devices with Ta (O) conductive channels in HfO2 substrates can provide stable conductivity states. To further improve stability, movable ions with high activation energy can reduce drift and interference in conductivity states [103]. More solutions are shown in Table 2.
Table 2. Ways to improve the stability and reliability of memristor.

<table>
<thead>
<tr>
<th>Year</th>
<th>Level</th>
<th>Question</th>
<th>Method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>Device</td>
<td>Relaxation effect</td>
<td>Adding thermal-enhanced layers</td>
<td>[104,105]</td>
</tr>
<tr>
<td>2021</td>
<td>Array</td>
<td>Excessive oxygen vacancies generated by the abrupt SET process</td>
<td>Triangular pulse for the programming</td>
<td>[106]</td>
</tr>
<tr>
<td>2019</td>
<td>Algorithm</td>
<td>Drifting conductance</td>
<td>Conductance tracking method</td>
<td>[107]</td>
</tr>
<tr>
<td>2020</td>
<td>Algorithm</td>
<td>Electron hopping assisted by traps (vacancies)</td>
<td>Optimizing the neural network weight noise tolerance</td>
<td>[108]</td>
</tr>
<tr>
<td>2021</td>
<td>System</td>
<td>Limited precision of conductance</td>
<td>The tradeoff of neural network accuracy, size, and current consumption</td>
<td>[109]</td>
</tr>
</tbody>
</table>

4.3. Large-Scale Integration

Building a large-scale neural network requires the integration of a high-density memristor array. Each neuron is connected to an average of 10,000 neurons in the human brain. Therefore, achieving such a significant connection requires a variety of rows and columns involving device size, interconnection, power consumption, and heat dissipation. Creating new integration technologies, such as 3D integration technology, is crucial to improve the performance and number of devices. Furthermore, effective power management and cooling solutions must be developed to increase device density and reduce interconnect latency. The sneak path determines whether the memristor 3D crossbar can usually work, and the sneak path of adjacent units interferes with the programming and reading of the memristor. Existing solutions include using a self-rectifying memristor to build a 3D crossbar or adding unidirectional conductive electronic components (diode, selector, and transistor) to limit the sneak current. In addition, IR drop also affects the performance of 3D crossbars, adversely affecting high operating voltage and power consumption [110]. Therefore, minimizing the length of connecting lines and impedance is equally essential as impedance. The 3D crossbar is a widely researched topic for creating large-scale neural networks that mimic the brain’s computing. Table 3 shows the current development status of the 3D crossbar.

Table 3. The current development status of the 3D crossbar.

<table>
<thead>
<tr>
<th>Year</th>
<th>Material</th>
<th>Scale</th>
<th>Layer</th>
<th>3D Structure</th>
<th>Compatible with CMOS</th>
<th>Analog/Digital</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>Ta/TaOx/TiOx/Ti</td>
<td>4 × 4</td>
<td>2</td>
<td>Stacking</td>
<td>/</td>
<td>Analog (50 states)</td>
<td>[111]</td>
</tr>
<tr>
<td>2016</td>
<td>Pt/AlOx/TiOx/TiN/Pt</td>
<td>10 × 10</td>
<td>2</td>
<td>Stacking</td>
<td>Yes</td>
<td>Analog (16 states)</td>
<td>[112]</td>
</tr>
<tr>
<td>2017</td>
<td>p-Si/SiOx/n-Si (self-rectifying)</td>
<td>8 × 8</td>
<td>5 (max)</td>
<td>Stacking</td>
<td>Yes</td>
<td>Digital (10⁶ ON/OFF)</td>
<td>[113]</td>
</tr>
<tr>
<td>2020</td>
<td>Pt/HfAlOx/TaN</td>
<td>8 × 6</td>
<td>3</td>
<td>Stacking</td>
<td>/</td>
<td>Analog</td>
<td>[114]</td>
</tr>
<tr>
<td>2020</td>
<td>HfOx/TaOx</td>
<td>8 × 32</td>
<td>8</td>
<td>Vertical</td>
<td>/</td>
<td>Digital</td>
<td>[115]</td>
</tr>
</tbody>
</table>

4.4. Endurance

The practical large-scale application of the memristor is affected by its endurance, which is worse than that of traditional CMOS devices. Based on existing literature, the endurance of a memristor depends on its material, ranging from 10 to 1 × 10⁶ cycles, with exceptional performance achieving up to 1 × 10⁹ cycles [116]. For instance, Huang et al. reported the anisotropic resistive switching in quasi-two-dimensional (2D) κ-(BEDT-TTF):Cu[N (CN)₂]Cl electronic memristors with an endurance of 1.24 × 10⁹ cycles [117]. Kumar et al. analyzed micrometer-scale titanium niobium oxide prototype memristors, exhibiting high endurance with over a million cycles [118]. Yang et al. outlined criteria to consider when selecting materials for high endurance and reported the Ta oxide memris-
tor with an endurance of $1.2 \times 10^{10}$ cycles [119]. Apart from materials and processes, electrical stress created by interfacing peripheral circuits also impacts endurance. Strukov developed a model that relates the tradeoff between endurance and writing time in nonvolatile memories [120]. Ravi and Prabaharan proposed a new technique by designing a fault-tolerant adaptable write scheme that can adapt based on the behavior and switching faults [121]. In the domain of neuromorphic computing, memristors offer high-density and low-power synaptic storage for SNN using hardware crossbar arrays. Titirsha proposed a new method to enhance lifetime by considering the endurance differences in each crossbar when mapping machine learning tasks [122]. This ensures that synapses with greater activation are always placed on memristors with greater endurance and vice versa, thereby enhancing the SNN’s overall endurance.

4.5. Fundamentals of Biological Theory

Further, we can explore the principle and mechanism of biological neural systems and apply this knowledge to the design and optimization of memristor synaptic circuits. For example, biological neurons’ and synapses’ complex dynamic behavior can be simulated to achieve more prosperous computational functions and higher energy efficiency. As the understanding of the working mechanism of the brain gradually deepens, brain-inspired computing can be realized at a deeper level.

5. Outlook

The memristor can realize the dynamics in biological neural networks and has excellent potential for efficient neural morphology computing. Therefore, it has attracted extensive research from material scientists, electronic engineers, physicists, and computational scientists. Currently, most hardware approaches to neuromorphic artificial intelligence rely on first-order memristors to simulate biological functions. While low-level memristors can simulate primitive low-level biological complexity, higher-order memristors can more faithfully simulate neuronal and synaptic functions. Presently, there have been some reports on the use of second-order or third-order memristors to build neural networks. This higher-order memristor neural network can leverage the inherent characteristics of the equipment to achieve more complex higher-order biological factors without requiring complex circuits. As a result, it promotes the progress of extraordinarily compact and powerful neural morphological computing [123,124]. Synapses are the most essential part of morphological neural networks, and we focus on using memristors in synapse research. Memristors can be categorized as digital or analog synapses based on resistance values. The functions they can achieve are introduced. Finally, the problems and challenges are summarized. As technology advances, we can expect solutions to various issues to arise gradually. Synapses based on memristors for neural morphology computing have excellent potential for multiple applications.

**Author Contributions:** Conceptualization, Y.L. and T.L.; methodology, K.S.; formal analysis, H.C.; investigation, X.Z.; resources, T.L.; data curation, C.W.; writing—original draft preparation, Y.L.; writing—review and editing, H.M.; visualization, K.L.; supervision, H.C.; project administration, X.X.; funding acquisition, T.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key R&D Plan of Shandong Province, grant number 2021SFGC0801.

**Conflicts of Interest:** The authors declare no conflict of interest.
References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.