

# Advantages And Applications of Neural Hardware Based on Synapse Structure

Shouhang Du \*

Yuhang High School, Hangzhou, 310000, China

\* Corresponding Author Email: 201026010136@stu.swmu.edu.cn

**Abstract.** The increasing demand of computing power serves as a motivation for researchers to develop new types of hardware structures, among which the brain-inspired chip hardware show higher potential on computing power and energy consumption over traditional types. Memristors, which serve as a representative example of them, can allow a certain amount of current to pass safely and "remembers" the previous resistance value of the device after a power failure, which can be used as a perfect similar unit of neurons. Memristor can be manufactured by multiple material, such as various kind of metal oxides and have different of structure, classified by the number of memristors in each structure. However, the synaptic structure also faces great challenge to overcome, such as the uniformity of the capacity of each synaptic unit. In conclusion, memristor-based synapses for computing neural morphology have great potential for multiple applications, which can be perfect apply to the artificial intelligence industry.

**Keywords:** synaptic hardware, computing power, memristor.

## 1. Introduction

With the advent of artificial intelligence rooted in deep learning, encompassing domains like cloud recognition, autonomous vehicles, and facial recognition, the conventional hardware infrastructure is finding itself increasingly incapable of shouldering the monumental computational demands. Not only is there a relentless surge in the appetite for computing prowess, but the energy consumption inherent to traditional hardware has also emerged as an urgent predicament demanding resolution. A striking illustration of this conundrum can be discerned in the comparison between Alpha Go and the human brain. Despite Alpha Go's capacity to outperform the human intellect in the intricate game of Go, its artificial intelligence system engulfs an average power consumption that dwarfs that of the human brain by a staggering factor of 115,000. Consequently, it is evident that traditional hardware systems, grounded in data-driven approaches and hardware acceleration, are ill-equipped to sustain long-term development. In contrast, a host of brain-inspired hardware systems beckon with boundless potential, heralding a promising future.

This article covers the research content related to brain-inspired chip hardware, including the theoretical foundation, hardware implementation, and associated challenges and future prospects. Firstly, in the introduction section, the background and objectives of the research are presented. Subsequently, the theoretical basis section introduces the basic concept of biological synapses and the advantages of synaptic hardware, including computing in memory and the application of Spiking Neuron Networks (SNNs). Following that, the brain-inspired chip hardware implementation based on CMOS technology is discussed, as well as the detailed mechanism and materials of digital memristors. The applications of digital memristors on chips are also explored, including one-memristor synapses and two-memristor-two-resistor synapses. Finally, in the challenges and future outlook section, the challenges associated with brain-inspired chip hardware are discussed, along with the prospects for future development. The article concludes by summarizing these contents and presenting a conclusion. In summary, this article investigates the theory and practical applications of brain-inspired chip hardware based on biological synapses, along with related challenges and future trends.

## 2. Theoretical basis

### 2.1. Brief introduction on biological synapse

Synapse is a specific joint structure between neurons, or between neurons and other specific structures. Neurons in the central nervous system connected with other nervous system through synapses, forming a neuronal network, which is the basis of the formation of feelings and thoughts. Synapses between neurons can be divided into chemical synapses and electrical synapses, which the first synapse is the main focus [1]. Communication between chemical synapses depends on the transference of neurotransmitter. When a nerve impulse occurs in the presynaptic cell, the presynaptic membrane releases neurotransmitters. Action potentials are generated in the axon mound of neurons and are transmitted to the presynaptic membrane at a limited speed, leading to the open of voltage-gated calcium ion channels on the presynaptic membrane, forming an inward current of calcium. Calcium ions entering the presynaptic membrane led to the fusion of synaptic vesicles and the presynaptic membrane through a series of chemical reactions, and the release of neurotransmitters. The released neurotransmitters then go through the synaptic cleft and get in to the Ligand-gated channel via its specific receptor. neurotransmitters are divided into excitatory synapses accordingly. and inhibitory synapses. Excitatory synapses excite the next neuron and inhibitory synapses inhibit the next neuron. The typical synaptic structure in biology is shown in figure 1 [1].

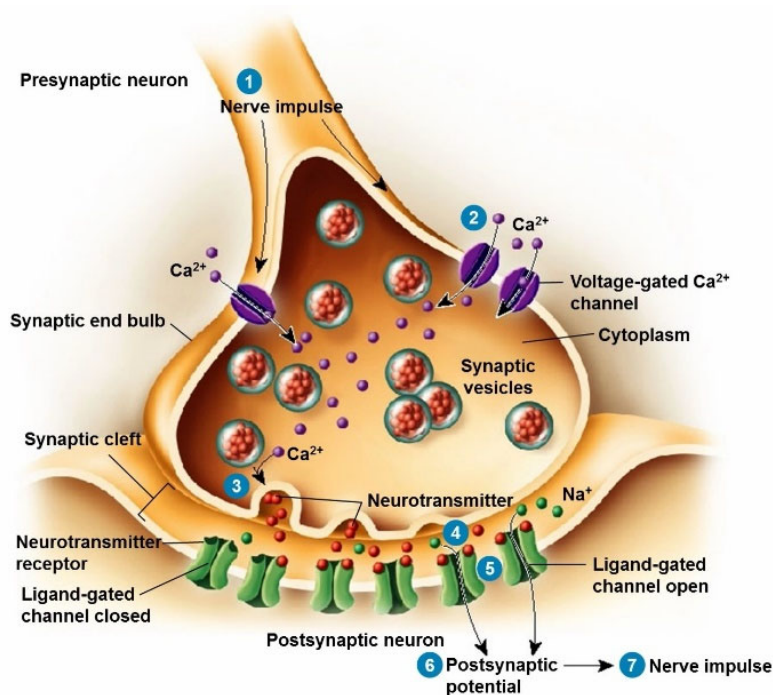


Fig. 1 Typical synaptic structure in biology [1]

### 2.2. Advantages of synaptic hardware

#### 2.2.1 Computing in Memory

Compared to traditional hardware system, synaptic hardware can Embed computing capacity in memory. It is generally accepted that the most striking feature of the human brain is its ability to process, calculate and store information simultaneously. However, what truly distinguishes the human brain from traditional hardware systems is that it continuously adapts to external stimuli, a phenomenon called "synaptic plasticity" [2]. Such advantages could allow the neuron to control the efficiency depending on the amount of information, which the CPU and memory coexist in one single unit is going to deal with.

In the traditional structure, the memory block is design to serve for computing, hence the priority of computing unit and memory block would be considered in the overall structure of the hardware.

Two blocks have to cooperate for data collection, transmission and processing in an optimally coordinated manner. This may result in low CPU performance limited by memory capacity or transfer bandwidth severely, which is so-called the Memory Wall [3]. Although multi-core or many-core can also improve computing power, in the post-Moore era, storage bandwidth has restricted the effective bandwidth of computing systems, and the increase of chip computing capacity has been difficult.

However, the problems can be resolved by using computing in memory structure. Firstly, such combination structure would break the memory wall, by eliminate unnecessary data transference delays and power consumption. The use of storage units can also increase computing capacity by hundreds or thousands of times and reduce costs to run the whole system.

### 2.2.2 Spiking Neuron Networks (SNNs)

Spiking neural networks (SNNs) are a kind of neural network architecture that uses spike-based signaling to communicate information between neurons, rather than the continuous activation values used in traditional artificial neural networks. As for energy efficiency, SNNs can be more energy efficient than traditional artificial neural networks because they use spike-based signaling, which requires less energy to transmit than continuous signals. Besides, SNNs can process information in real-time because they operate on a millisecond timescale, which is closer to the timescale of biological neurons. By adjusting the form of spikes, SNNs can also find the most efficient way to process sophisticated data, such as non-linear data and time series data [4].

## 3. Brain-inspired chip hardware implementation

### 3.1. Brain-inspired chip based on CMOS.

Since 2004, the research concerning Brain-inspired computing has been conducted by governments. Although the synaptic chips are different from the traditional chip's hardware, but the basic structure of the chips is still CMOS-based digital circuits or digital-analog hybrid. To build a combined circuit, simulating the behavior of a single neuron or synapse often requires multiple CMOS devices to form a circuit module. The integrity, power consumption and simulation accuracy are all severely limited. Especially when the size of CMOS has almost been reduced to its physical limit, the number of neurons and synapses in the brain-like neural network that can be constructed by chips relying on advanced technology is still far smaller than the scale of the human brain.[5]

For instance, in August 2014, IBM (International Business Machine Corporation) launched a second-generation brain-inspired chip called "TrueNorth" [6]. It uses Samsung's 28nm process and includes 5.4 billion transistors and 4096 processing cores, equivalent to 1 million programmable neurons and 256 million programmable synapses. Each processing core of "TrueNorth" has about 1.2 million transistors, a small number of which are responsible for data processing and scheduling, while most of the transistors are used for data storage and communication with other cores. In addition, each core has its own local memory, which works much like the coordination between neurons and synapses in the human brain. If 48 TrueNorth chips form a network with 48 million neurons, the intelligence level brought by these 48 chips will be similar to that of an ordinary mouse.

Such sophisticated design actually could make a visible but limited improvement on the efficiency and other aspects theoretically. Afterall, the transistor is not designed to stimulate brain structure from the very beginning.

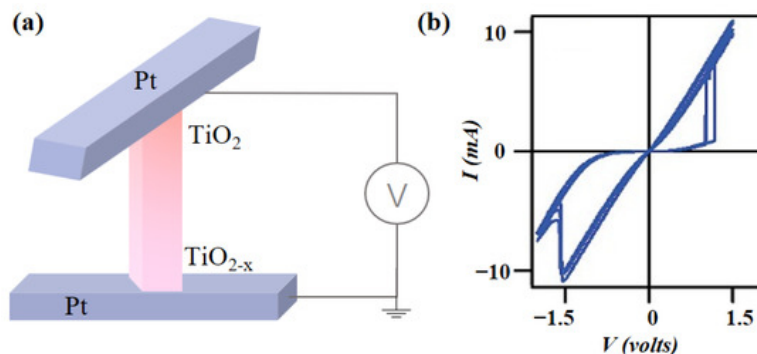
### 3.2. The digital memristor

The memristor is also known as the memory resistor, which is a passive electronic component. Differ from the traditional chip based on CMOS computing, the neuromorphic chip starts from bionics of the basic device. Like a resistor, generates and maintains a steady current through the circuit. But unlike resistors, memristors can "remember" the amount of charge that passed through them even after the power is turned off. By adjusting the current applied to the memristor, the pulse length and amplitude of the voltage, the resistance value of the memristor can be changed. This

property of continuously adjustable high and low resistance states can be used in some storage and neural network applications.

### 3.2.1 Detailed mechanism of memristor apply on the chips

according to the method of resistance memorization by the memristors, it can be divided into One-Memristor Synapse and Two-Memristor-Two-Resistor Synapse. Inside the memristor, '0' and '1' represent for high resistance and low resistance states. Digital memristors have a distinct threshold for resistance changes. If the voltage sweeps forward and crosses the positive threshold, its resistance changes from high to low. The detail is shown in figure 2 [7].



**Fig. 2** (a) Pt/TiO<sub>2</sub>/Pt structure two-terminal memristor. (b) I-V curve of threshold memristor [7]

However, an interesting factor of the digital memristor is that if the potential exerted in between the maximum and minimum threshold the resistance would remain unchanged. This magnificent feature can serve well for mimicking the function of synaptic network. When the resistance is low, the presynaptic neuron can transmit spikes to the postsynaptic neuron, but when the resistance is high, it will not happen. To create artificial neural networks in hardware, synapses connecting large numbers of neurons to perform complex calculations are essential.

### 3.2.2 Materials of memristor

There are diverse types of material that can be observed the memristive effect, such as many metal oxides: nickel oxide, zirconium oxide, zinc oxide, hafnium oxide, titanium oxide, etc.

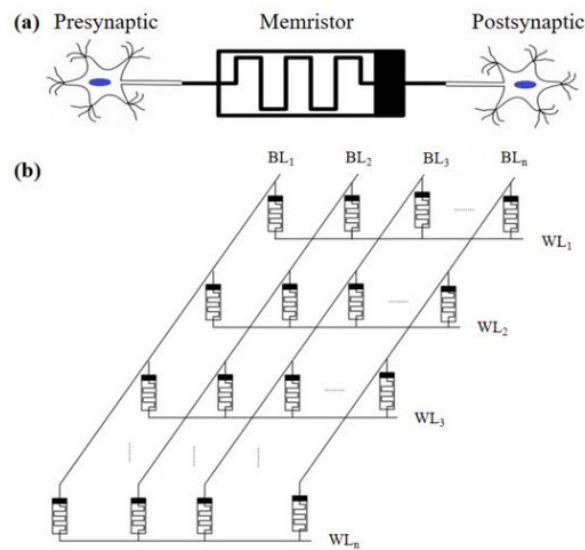
For binary metal oxide memristor materials, the most used is titanium dioxide (TiO<sub>2</sub>) [7]. Compared with other resistive switching materials, binary metal oxides have the advantages of simple structure, easy control of material components, and preparation process compatible with CMOS, so they have attracted more attention from the industry.

In addition to binary metal oxides, common multi-metal oxide memristor materials include quaternary metal oxides such as Pr<sub>x</sub>Ca<sub>1-x</sub>MnO<sub>3</sub>, La<sub>x</sub>Ca<sub>1-x</sub>MnO<sub>3</sub> and La<sub>x</sub>Sr<sub>1-x</sub>MnO and ternary metal oxides such as SrTiO<sub>3</sub>, SrZrO<sub>3</sub> and SrRuO<sub>3</sub>. However, the composition of multi-element metal oxides is relatively complex, and it is difficult to obtain crystal structures with precise chemical ratios [8]. The preparation process of materials and devices is not compatible with the traditional CMOS process, which hinders its development and application to a certain extent.

## 3.3. The digital memristor applications

### 3.3.1 One-Memristor Synapse

It is easy to deduce that the One-Memristor synapse is the simplest way to make each unit connect with each other. Just as the figure 3 shows, one side of the memristor directly link with the front neuron, the other side link with the synapse behind. Furthermore, the memristor can also coordinate the relationship between input and output [9].



**Fig. 3** (a) a single memristor structure. (b) Synaptic cross array composed of single memristor [9]

For instance, there is a corresponding relationship between synaptic weight and initial resistance. ‘1’ represents high resistance and ‘0’ represents low resistance. But the premise is that the applied voltage must be between the positive threshold and the negative threshold. Thanks to its beautifully simple structure, this type of network structure has various advantages such as small size, high integrity as well as low energy consumption. However, potential hazard still exists while using this type of memristor. For example, when building a synaptic cross-array using a single memristor as a synapse, the neighbor memristor would interfere with each other for their different resistance, which may result in inaccurate read or write.

### 3.3.2 Two-Memristor-Two-Resistor Synapse

In dual-memristor synaptic circuits, topological junctions cannot obtain negative weights. But unlike this, by adding two resistors to form a synaptic circuit, you can increase the weighted state of the synapse. The connection method is shown in the figure. There is a memristor and a resistor in series on each branch. The voltage difference between point A and point B is used as the output voltage [9].

## 4. Challenges and the future outlook

### 4.1. Challenges

Using memristors to achieve synaptic computing systems remains challenging. The key factor is the memristor itself, which directly determines the efficiency of the whole system. However, each device has difference in its capacity. The key factor to choose which type of memristor to use is the specific requirements for neural circuit construction [10]. After all, the research on memristors is mainly in the scientific research stage and is still far away from large-scale industrialization.

The different characteristics of memristors are mainly reflected in conductivity, reliability and switching speed. If memristors in the same circuit cannot agree on these characteristics, the overall capacity of the entire system will inevitably decrease. Therefore, better uniformity puts higher requirements on its manufacturing process. Optimizing film deposition and annealing processes are two currently more effective solutions.

### 4.2. The Future outlook

The memristor has great potential for efficient computation of neural morphology. However, Currently, first-order memristors are the hardware method currently used by most neuromorphic artificial intelligence to simulate biological functions. And if memristors are to be able to simulate

neuron and synaptic functions more faithfully, high order memristors must be introduced. But the good news is that some neural networks built using second- or third order memristors are already in the experimental stage. The biggest advantage of this high order memristive neural network is that it does not require complex circuits. It only needs the characteristics of the device itself to complete the tasks that low order memristive neural networks can complete.

## 5. Conclusion

Synapses play a pivotal role in both biological and circuit structures, holding immense significance in the realm of computational neuromorphology. Among the prominent solutions for emulating synapses, memristors shine as a key innovation. This article has shed light on their numerous advantages, delved into the intricacies of their mechanisms, elucidated the achievable functions they enable, and highlighted potential challenges on the horizon. As technology continues to advance at an unprecedented pace, it is inevitable that novel solutions to various problems will emerge. It is in this context that memristor-based synapses stand poised for a transformative role, boasting tremendous potential across a spectrum of applications. Their integration into computational neuromorphic systems represents a promising frontier, offering the prospect of unlocking new avenues of innovation and expanding the horizons of artificial intelligence and neurocomputing.

## References

- [1] Ryan T J, Grant S G N. The origin and evolution of synapses. *Nature Reviews Neuroscience*, 2009, 10(10): 701-712.
- [2] Abbott L F, Nelson S B. Synaptic plasticity: taming the beast. *Nature neuroscience*, 2000, 3(11): 1178-1183.
- [3] Legtchenko S, Williams H, Razavi K, et al. Understanding {Rack-Scale} Disaggregated Storage. 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17). 2017.
- [4] Ghosh-Dastidar S, Adeli H. Spiking neural networks. *International journal of neural systems*, 2009, 19(04): 295-308.
- [5] Cho S W, Kwon S M, Kim Y H, et al. Recent progress in transistor-based optoelectronic synapses: from neuromorphic computing to artificial sensory system. *Advanced Intelligent Systems*, 2021, 3(6): 2000162.
- [6] Nie Y, Ren T, Shi Z. The developments and applications of brain-like computing chips. *International Conference on Algorithms, Microchips and Network Applications*. SPIE, 2022, 12176: 272-285.
- [7] Patil A R, Dongale T D, Kamat R K, et al. Binary metal oxide-based resistive switching memory devices: A status review. *Materials Today Communications*, 2023: 105356.
- [8] Kim J, Ko W, Yoo J M, et al. Structural Insights into Multi-Metal Spinel Oxide Nanoparticles for Boosting Oxygen Reduction Electrocatalysis. *Advanced Materials*, 2022, 34(8): 2107868.
- [9] Li Y, Su K, Chen H, et al. Research Progress of Neural Synapses Based on Memristors. *Electronics*, 2023, 12(15): 3298.
- [10] Lu Y, Alvarez A, Kao C H, et al. An electronic silicon-based memristor with a high switching uniformity. *Nature Electronics*, 2019, 2(2): 66-74.