A Brain-inspired in-Memory Computing System for Neuronal Communication via Memristive Circuits

Xiaoyue Ji, *Student Member, IEEE*, Zhekang Dong, *Member, IEEE*, Chun Sing Lai, *Senior Member, IEEE*, and Donglian Qi, *Member, IEEE*

Abstract—**Brain-inspired approaches can efficiently analyze the activities of biological neural networks and solve computationally hard problems with energy efficiencies unattainable with von Neumann architectures, indicating a significant improvement in the understanding of neuronal communications and functionalities. Here, we present a brain-inspired multimodal signal processing system with organic memristor arrays that can potentially integrate the signal sensory, storage and computation. To facilitate the multimodal signal processing system design, we used four components. First, we present a multimodal signal sensory module mainly responsible for multimodal (iconic, echoic, olfactory, muscular and gustatory) signal collection, fusion and storage. Second, a high-density cross-point memristive synapse array is constructed after fabrication of the albumin protein memristor to realize the dense connectivity between layers of computing, data storage, and communication. Third, considering the structure and function of the brain region, we demonstrate a general learning module for hierarchy learning, which can recognize and imagine multimodal information. Finally, the necessary peripheral circuit module (consisting of winnerless competition function circuit, analogue-to-digital converter, digital-to-analogue converter, pulse modulator, etc.) is designed. Notably, our system can capture massive amounts of data every second and perform in situ processing of multimodal signals. This study is expected to help achieving the deep integration of nano materials into neuromorphic computing systems and energy-efficient integrated circuits.**

Index Terms—**Multimodal signal processing, in-memory computing, neuronal communication, memristive circuits**

I. INTRODUCTION

Brain-inspired computing systems are designed to improve computing efficiency by simulating human brains as shown in Fig. 1. Brain-inspired approaches have attracted more and more attention in recent years since they provide new opportunities to achieve the goal of general intelligence [1]. In contrast, since traditional computing systems are always based on complementary metal oxide semiconductor (CMOS) transistors, the internal communication may suffer from two main limitations (i.e., the storage capacity and transmission efficiency), which leads to high energy consumption and computational burden, i.e., many orders of magnitude higher than human brains [2]. Therefore, the drawbacks of traditional computing systems are an increasing problem. From the perspective of the device, leakage currents become a problem when the channel length and the gate dielectric thickness of a transistor get closer to the scaling limit [3]. With respect to the architecture, the data transfer between processors and memory units significantly reduces both speed and energy efficiency (referred to as the 'von Neumann bottleneck') [4]. Meanwhile, the performance mismatch between the memory and processing units leads to great latency (also called the 'memory wall') [5]. Furthermore, the traditional computing systems based on centralized processing framework are accurate for repeated tasks and artificial functionalities; while distributed ones, which learn from the hierarchical brain architecture, are more adaptive in the data processing and analyzing. Finally, traditional communication approaches need to transmit and store large amounts of raw recording data, followed by extensive processing offline, posing significant challenges to the hardware and preventing real-time analysis and feedback.

In-memory computing is beginning to replace traditional approaches, which shows the ability to cache countless amounts of data and constantly conduct computing [6]. New computing systems aided by the in-memory computing concept have been established upon many emerging beyond-CMOS devices and nanotechnology [6], which offer promising solutions to the energy consumption and speed problems. Memristor, a two-terminal electronic element that directly relates electrical charge to flux, is a potential candidate for the in-memory system. It was first proposed by L. O. Chua in 1971 [7] and was further associated with physical devices by R. Stanley Williams and his team from Hewlett-Packard Labs in 2008 [8]. Due to its intrinsic dynamics, analogue behaviour, nonvolatility, high speed, low power, high density, great scalability, etc., it is particularly suitable to act as the synapse in brain-inspired computing architectures. These memristive synapse arrays exhibit high degrees of parallelism and enable multiply-accumulate (MAC) operations, which are the primary calculations in information processing and analyzing. This analogue

Xiaoyue Ji and Donglian Qi are with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, P. R. China; Zhekang Dong (corresponding author) is with the School of Electronic Information, Hangzhou Dianzi University, Hangzhou, 310018, P. R. China and also with the Zhejiang Provincial Key Lab of Equipment Electronics Hangzhou, 310018, P. R. China; Chun Sing Lai (corresponding author) is with the Department of Electronic and Electrical Engineering, Brunel University London, UB8 3PH, London, UK.
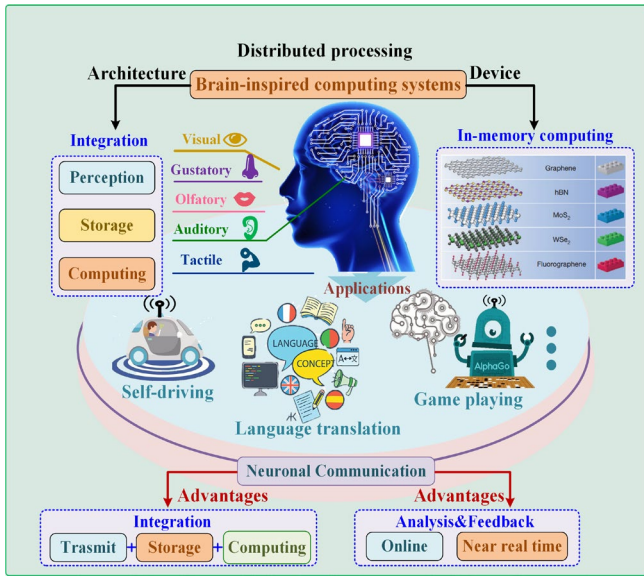
Fig. 1 Brain-inspired in- memory computing system.

in-memory computing effectively avoids the huge energy and time overheads associated with frequent data shuttling in von Neumann systems. Nowadays, inspired by the sensory processing and perceptual learning mechanism of human beings, in-memory computing systems supported by machine learning algorithms have been equipped with the capability to perceive and process the visual, tactile, auditory, olfactory, and gustatory data, as well as to integrate the first two types of information [9]. While the existing in-memory computing systems still suffers from some limitations given below:

From the perspective of the sensory modality, the sensory devices in the existing in-memory computing paradigm are required to have self-adaptation and self-identification characteristics for efficient information processing. Current investigations mainly focus on a single type of sensory processing.

From the perspective of the computing devices, more reliable and eco-friendly computing devices are required. Using several common organic materials in computing devices raises a lot of expectation for future applications due to their flexibility, durability, and degradability.

From the perspective of the learning algorithms, there is still a lack of brain-inspired in-memory computing system combining the different senses, as well as the hierarchical learning mechanism to recognize and imagine multimodal information [10].

The main contributions of this study are as follows:

1) Different to existing in-memory computing systems, we present a brain-inspired in-memory computing system that can potentially integrate the multimodal signals sensory, storage and computation.

2) To overcome the limitations of the 'von Neumann bottleneck' and 'memory wall', we construct a high-density cross-point memristive synapse array after fabrication of the albumin protein memristor that can reduce transmission cost and energy consumption while improving computational efficiency in communication.

3) We design a hierarchical general learning module, which aims at achieving recognize and imagine multimodal information indicating a significant improvement in the understanding of neuronal communications and functionalities.

II. Brain-inspired in-Memory Computing System Architecture and Module Design

Brain-inspired computing is one of the most important approaches to understand neuronal communications and functionalities [1]. Our motivation is to construct a brain-inspired in-memory computing system based on memristive circuits, aimed at solving computationally hard problems with energy efficiencies unattainable for von Neumann architectures. To facilitate understanding of the brain-inspired in-memory computing system design, we describe it using four components, as shown in Fig. 2:

- **Multimodal Signal Sensory Module:** This module is consisted of a sensor family (visual sensor, tactile sensor, auditory sensor, olfactory sensor, and gustatory sensor), a coupled auto-en/decoder, and buffer circuit for multisensory data sense, fusion, and storage, respectively.

- **High-Density Cross-Point Memristive Synapse Array:** This synapse array based on albumin protein memristor is primarily employed for performing MAC operations as per Kirchhoff's law.

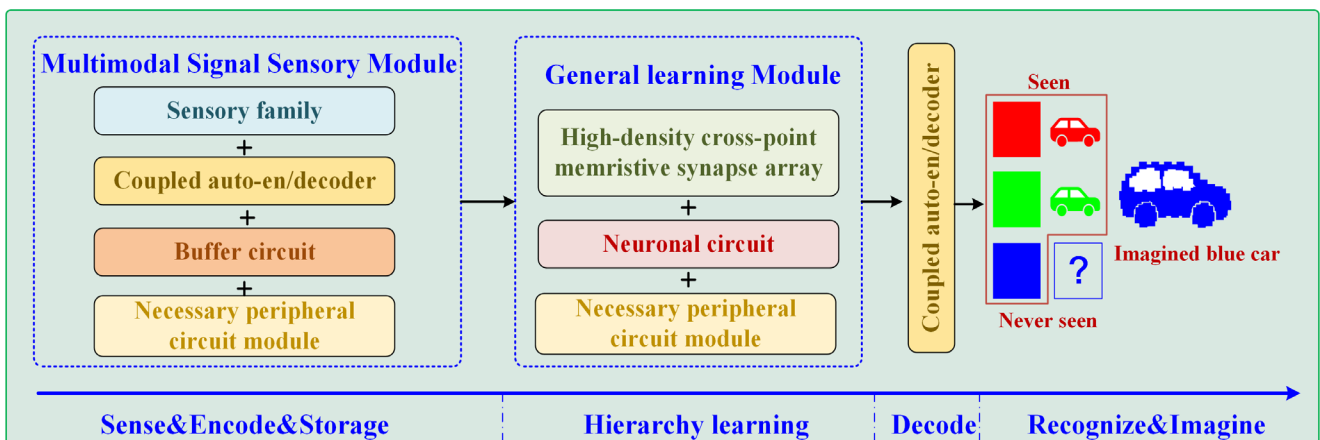- **General Learning Module:** Based on the

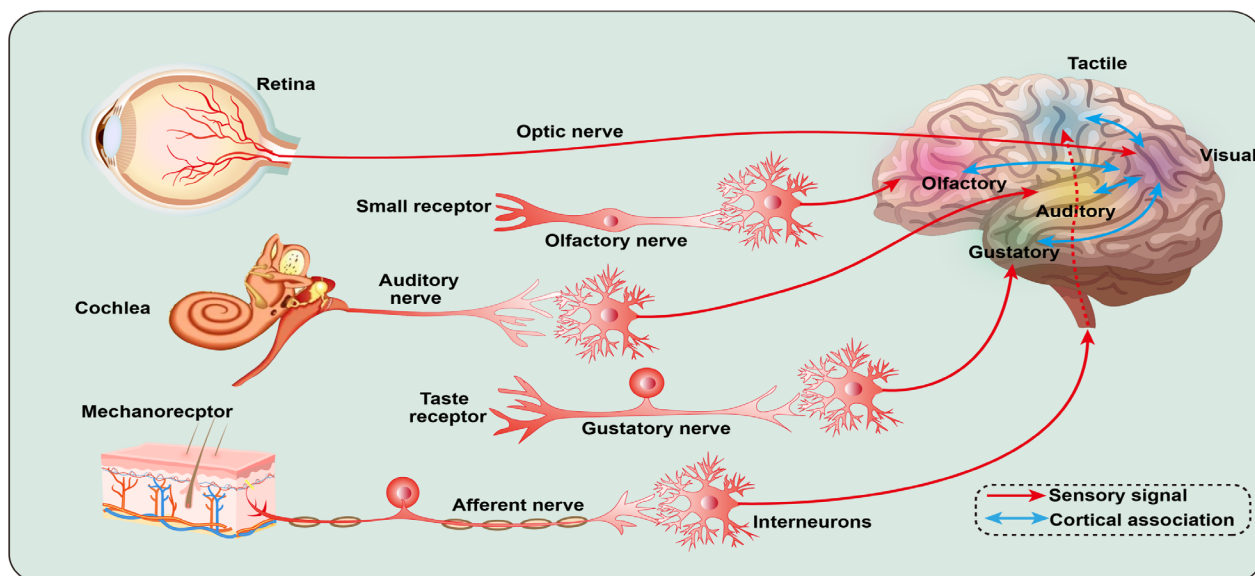Fig. 2 Schematic of the brain-inspired in-memory computing system.

Fig. 3 Schematic of the human environmental perception system [11].

proposed memristive synapse array and neuron circuit, a general learning module is implemented, which can be used to process and analyze the complex multisensory information.

- **Necessary Peripheral Circuit Module:** The necessary peripheral circuit module is consisted of analogue-to-digital converters (ADCs), digital-to-analogue converters (DACs), a winnerless competition function circuit, a pulse modulator, etc.

III. Multimodal Signal Sensory Module

The human environmental perception system is achieved by multiple primary sensors, such as photoreceptor cells in the retina, mechanoreceptors in the skin, cochleae in the ears, and olfactory and taste receptors in the nose and tongue [9]. Specifically, the encoded sensory information carried by spike trains of sensory neurons can be transmitted to the corresponding brain cortices. The association of sensory information from different receptors takes place in various parts of the cerebral cortex, where nerve impulses from the visual, auditory, somatosensory, gustatory, and olfactory association areas are integrated and processed with multisensory functions (as shown in Fig. 3). Therefore, humans can simultaneously perceive and process different types of information in a very small perceptive field and complex environments. However, current investigations into human perceptual systems mainly focus on a single type of sensory processing, and there is still lack of research on brain-inspired multimodal signal processing systems. Hence, we develop a multimodal signal sensory module for the real-time fusion of different sensory experiences, such as visions, touches, sounds, smells, tastes, etc. (as shown in Fig. 4).

In Fig. 4, our multimodal signal sensory module utilizes a sensor family to detect environmental information and convert this information into voltage signals via Si-based photodetectors (visual), AgNWs/PDMS-based pressure sensors (tactile), sound detectors (auditory), CNFET-based receptors (olfactory) and ZnS based receptors (gustatory).

Since there is a potential relation among the voltage signals captured from the sensor family, it is necessary to extract their common features when performing the representation and fusion. A coupled auto-en/decoder is further adopted to handle the voltage signals, as shown in the middle of Fig. 4. Notably, the auto-encode and the auto-decode can be deemed as a reciprocal process, which can be implemented by the same framework and executed in the same hardware configuration. Specifically, the coupled auto-encoder consists of multiple basic auto-encoders with the same architecture, which is used to convert the multimodal signals into 39-dimensional feature vectors. The basic auto-encoders are coupled on the hidden layers where similarity is evaluated, so as to capture voltage signal correlations. That way, the integrated voltage information from the sensor family can be acquired using the coupled auto-encoder, implying successful multimodal data fusion by the feature extraction process. Next, the fused voltage signals are coded into spike trains based on the temporal coding principle for neuronal data processing. Spike encoding not only further prevents voltage degradation and parasitic resistance, but also has greater data volumes than voltage amplitude coding. A buffer circuit composed of a random storage unit and a time factor calculation unit is designed to record the historic spike trains, as shown in Fig. 4. The former is used to store the encoded spike trains, and the latter is used to read the pulse sequence in the random storage unit and generate pulses according to the attenuation factor as input to the hierarchical general learning module. The buffer circuit can help produce reconfigurable spike duration and has a high capacity for bitwise shift operations.

IV. High-Density Cross-Point Memristive Synapse Array

A. Preparation of Ag/Egg Albumen/ITO Memristor

Natural proteins offer near-perfect structural and functional advantages for a huge range of smart electronic device applications, as well as for
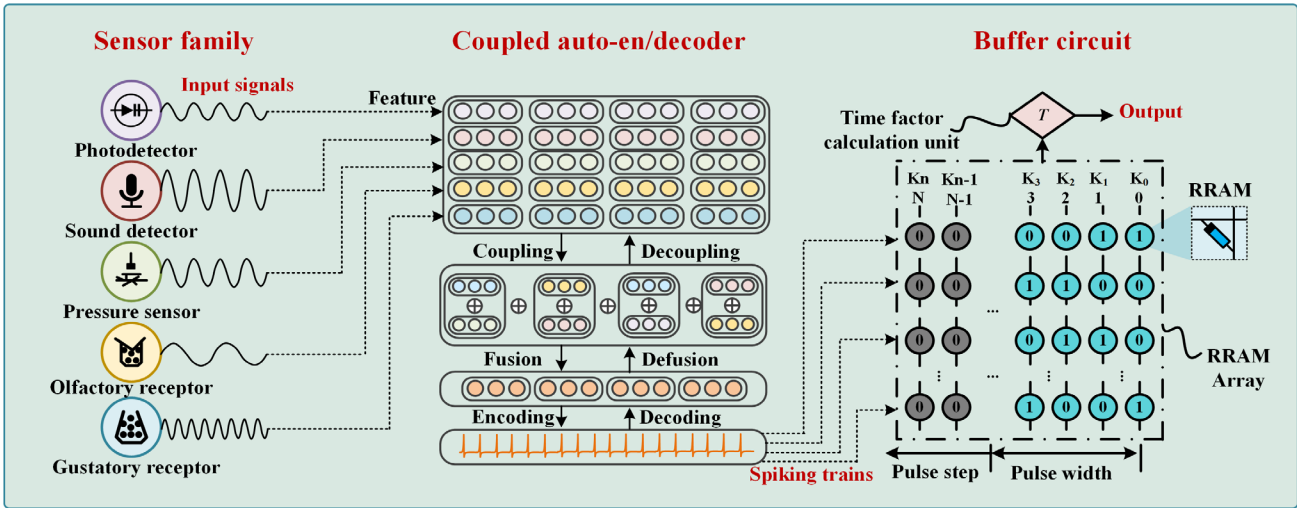
Fig. 4 Schematic of the multimodal signal sensory module.

fundamental research [11]. In this study, an Ag/Egg Albumin/ITO memristor was fabricated based on the sol–gel and the magnetron sputtering methods, where the former was used to prepare the egg albumin functional layer, and the latter was used to synthesize the Ag and ITO electrodes. The fabrication process can be summarized in the following steps:

Step 1: H_2O_2 solutions were prepared at 5%, 10%, 15%, and 30%, respectively.

Step 2: Native egg albumin solutions and H_2O_2 solutions of different concentrations were mixed in a 1:10 volume ratio and then stirred for 30 minutes in a continuous manner with a magnetic stirrer.

Step 3: A white flocculate appeared with stirring, which was filtered to fabricate the precursor.

Step 4: The ITO plastic substrate was placed on a spin coater in vacuum. The prepared precursor was continuously spin-coated on the substrate at 4,500 rpm for 60 seconds.

Step 5: The ITO substrate was transferred to a muffle furnace and annealed at 97 °C in ambient atmosphere for 3 hours, and then the albumin-based substrate was formed.

Step 6: Magnetron sputtering was used to prepare the Ag electrode (diameter: 200μm, thickness: 120nm) on the albumin-based substrate. After that, the Ag/Egg Albumin/ITO memristor was fabricated.

Devices with resistive switching memory are marked by a radical switch between a high resistive state (HRS) and a low resistance state (LRS), which

is achieved by a continuous bias voltage sweep. In this study, we used the electrochemical workstation CHI-600D to test the I–V and resistance curves of an Ag/Egg Albumin/ITO memristor at room temperature, as shown in the right of Fig. 5. As for the concentration of the H_2O_2 solution, at 5%, the resistive switching effect is inconspicuous, and there is no obvious SET and RESET process. As the concentration increases to 10%, the device exhibits a significant resistive switching effect, and the SET and RESET processes are also obvious. For a 15% H_2O_2 solution, the resistive switching effect is weak and its resistance switch ratio is drastically reduced. With a 30% concentration of H_2O_2 solution, the device exhibits the resistive switching effect comparable to those with the 15% concentration. Therefore, an optimized memory cell was developed using 10% hydrogen peroxide modified egg albumin as a function film at room temperature. To further study the stability of the resistive switching memory effects of Ag/Egg Albumin/ITO, a 0.2V reading voltage was applied in both states (HRS and LRS) for 900 switching cycles. The HRS/LRS ratio was about 10^4 and remained stable over the measurement time, indicating that the prepared memristor has good stability and endurance.

B. Memristive Synapse Array

In this scheme, the memristive synapse array was primarily employed for performing analogue MAC
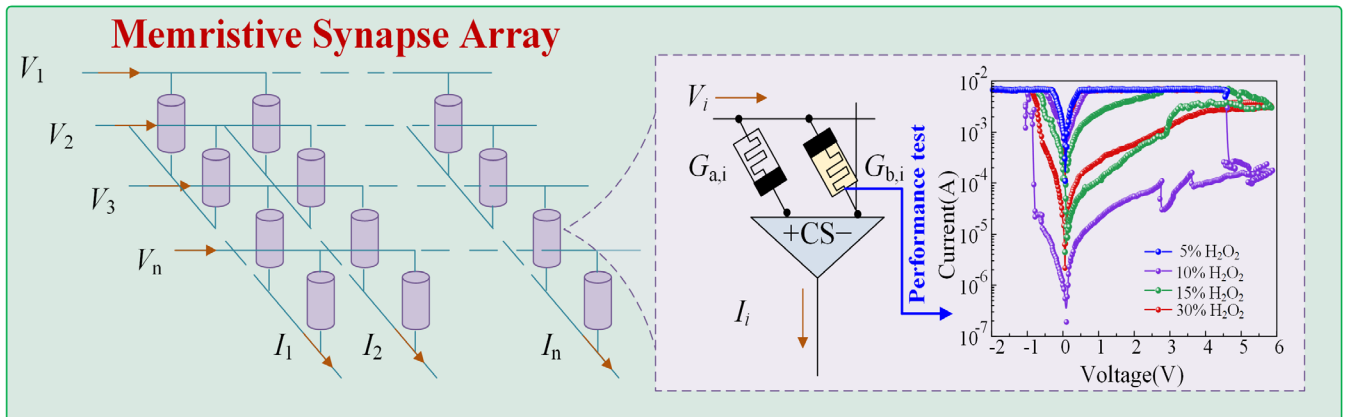

Fig. 5 Design of a high-density cross-point memristive synapse array after fabrication of an Ag/Egg Albumin/ITO Memristor.

operations via Kirchhoff's law. In the left of Fig. 5, each memristive synapse in the array (small solid purple cylinder) can be implemented by combining a pair of Ag/Egg Albumin/ITO memristors interconnected with reversed polarity. The different weights can be achieved by synapse modulation, where the solid blue triangle symbol, represents the current subtractor (CS) utilizing a differential amplifier circuit. As for synapse modulation, if the input voltage V_i is less than the memristor threshold voltage, the memristor state will be maintained. Conversely, if the applied voltage V_i is greater than the threshold voltage, synapse weights can be programmed. When a positive voltage is applied, the conductance $G_{b,i}$ decreases, while $G_{a,i}$ increases, resulting in an increase in synapse weight W_i. However, with negative voltage, the synapse weight shows the opposite. In the above programming process, since the polarity of the two memristors is reversed, their changes are always reversed too, regardless of whether the applied signal is positive or negative. This not only accelerates the weight programming operation, but also ensures that positive, zero, and negative synapse weights can be all obtained during successive programming without setting them in advance. After the synapse modulation operation, the output (i.e., the current) can be achieved by Ohm's Law.

C. Neuronal Circuit

In this scheme, the neuronal circuit was mainly used to achieve the functions of current processing and activation [12]. An appropriate neuron circuit for implementing the sigmoid activation function is designed, where the input signal to the neuron circuit is the summed current from the memristive synapse array. According to the current input value, a voltage V_{output} was generated at the output node. Transistors were biased using only one biasing voltage V_{bias} for proper functioning of the neuronal circuit. Notably, the output voltage shows a positive correlation to the input current, which causes a change in operation region of transistors. The sigmoid activation function can be obtained by modifying these operation regions.

V. General Learning Module

Hierarchical chunking mechanism of human brain can help enhancing our memory performance and offer us an ideal approach to achieve memory dynamics in neuromorphic circuits. In this section, we demonstrate a general learning module for hierarchy learning, which can recognize and imagine multimodal information.

The structure of the proposed hierarchical general learning module involves the multi-layered memristive neural networks. For each layer n, we can observe that the fully connected neurons form a chunk and then link with the sub-chunks in layer $n + 1$, with their communication signals denoted by dotted arrows. In the hierarchical structure, we define the chunks in layer n as parent chunks (PCs), and their connected sub-chunks in layer $n + 1$ are named as child chunks (CCs). A winner neuron in a PC can communicate the corresponding CC based on the dynamical principle of winnerless competition (WLC) [13]. Notably, each winner neuron is in a metastable state, and it will switch from one neuron to another through neuronal communication, forming a sequential memory trace.

Based on the proposed memristive synapse array and neuron circuit, a hierarchical general learning module is implemented (as shown in Fig. 6). Whether a PC or a CC, it mainly includes memristive synapse arrays and serval neuron circuits. For each column, a MAC operation is performed in the memristive synapse array and its cascaded amplifier. At each time step, the neuronal outputs are stored in temporal registers and then feedback into the network as synaptic inputs. The clock signal is utilized for the timing sequence of the whole network ($k \rightarrow k + 1...$). In the chunk circuit, contain digital and analogy signals, the conversion between digital and analogy signals is realized by DAC and ADC. Moreover, each chunk circuit can be constructed in a hierarchy so as to form a complete general learning module. In each PC, the winner neuron in each time step is determined by a winner neuron activator, which helps communicate the excitatory signal to the corresponding CC circuit. When the CC receives the signal, the clock will be triggered, and its iterative neuronal dynamics will form a memory trace. There are basically two stages for the general learning
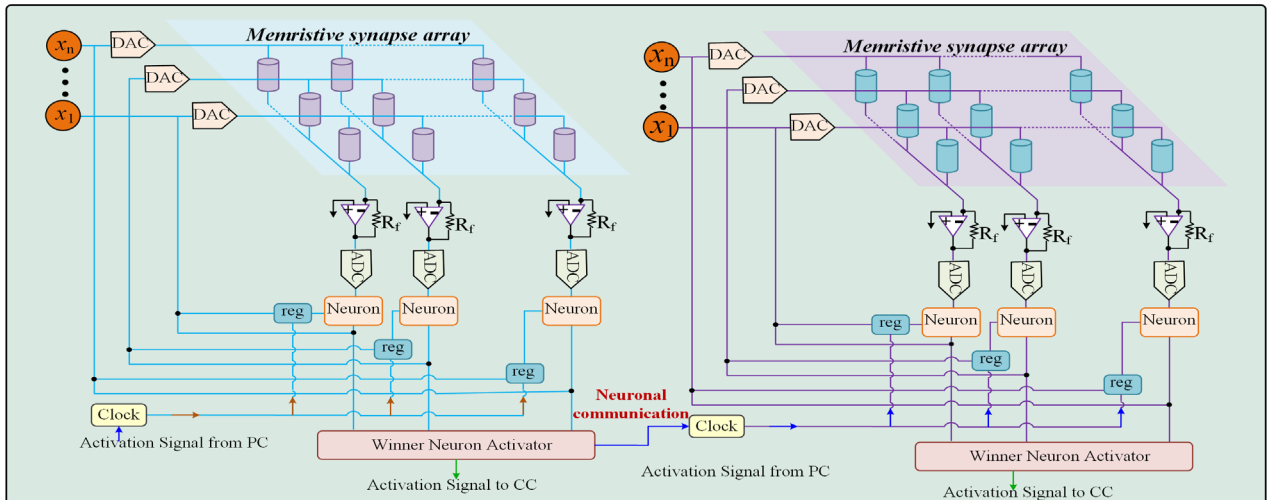

Fig. 6 Design of the general learning module.

module to process information, namely the write (synapse modulation) stage and the read (neuronal processing) stage. For the write stage, the memristive synapse array is largely controlled by the pulse modulator block. While the weight calculator block is adequate for calculating the theoretical synapse weight based on a pre-defined chunking sequence. The pulse modulator block creates the pulse train (potentiation or depression pulses), which is required to get the desired memristor conductance. At the read stage, when the pulse modulator is inactivated, data flows from the memristive synapse array to the neuron block, followed by feedback. It is worth noting that the scale of the input voltages should be small enough, so that the state of trained memory will not change during the read process.

When the spike trains from the multimodal sensory module is given to the general learning module, the model demonstrates a memory trace where the temporary winner neurons among different chunks are involved, and the sequential memory is manifested. Fig. 6 demonstrates the successful reproduction of multimodal signals upon hearing. In this experiment, we trained the proposed system to produce the representations with auto-encoded visual, tactile, auditory, olfactory, and gustatory information when hearing the sound of a car horn with different loudness and frequency. During training, we used 2000 sets of audio signals, including with 'truck', 'train', 'car' random music fragments, as input under the proposed system. Once the training phase is completed, hierarchical general learning module provides a sequential memory spike trains as the input of the coupled auto-decoder. The test results are obtained after learning and decoding, demonstrating the potential of overall sematic imagination.

VI. Conclusion

The main hindrance for traditional communication systems produces high energy consumption, transmission cost, and computational burden to process various and complex cognitive in the daily life. In this article, we presented a novel brain-inspired in-memory computing system for neuronal communication via memristive circuits. To manage the complexity of brain-inspired in-memory computing system, we describe it using four components: 1) multimodal signal sensory module, 2) high-density cross-point memristive synapse array, 3) general learning module, 4) necessary peripheral circuit module.

Firstly, the multimodal signal sensory module is proposed using sensor family, coupled auto-en/decoder, and buffer circuit to store, integrate, and memorize multisensory information at the hardware level. Then, a high-density cross-point memristive synapse array is constructed after the fabrication of Ag/egg albumen/ITO memristor, which can realize the dense connectivity between layers of computing, data storage, and communication. Finally, the hierarchical general learning module is designed, which enables to recognize and imagine multimodal information. This article serves to inspire research to design, analyze, and applicate novel memristic circuits and investigate new techniques that perform signal processing functions in the neuromorphic computing domain in order to support the advancement of

nanomaterials science and the modern circuit theory.

VII. Future work

Brain-inspired in-memory computing is still in an infancy stage with abundant opportunities and challenges. Innovations are critical not only to the devices but also to the circuits and systems.

At the device level, more reliable and eco-friendly memory devices and memristive arrays are required. To further improve device performance, the physical mechanism should be more thoroughly analyzed and comprehended.

At the circuit level, existing memristive synapse circuits suffer from many limitations, i.e., the operating mechanism is still obscure, the cognition function modeling is not good, and the fault diagnosis system is still in progress. Therefore, the novel memristive synapse circuit will possess biological synaptic features, both in cognition function and basic structure, and meanwhile, is suitable for neuromorphic computing system.

At the system level, the learning algorithms of brain-inspired in-memory computing system are still under development. With better understanding of neuronal communications and functionalities, general learning algorithms should be designed to promote hardware development as well.

VIII. Acknowledgment

References

[1] Y. Zhang, Z. Wang, J. Zhu, Y. Yang, M. Rao, W. Song, and J. Yang, "Brain-inspired computing with memristors: Challenges in devices, circuits, and systems," *Applied Physics Reviews*, vol. 7, no. 1, pp. 011308, 2020.

[2] Y. Zhao, Z. Fan, Z. Du, T, Zhi, L. Li, Q. Guo, and Y. Chen, "Machine Learning Computers with Fractal von Neumann Architecture," *IEEE Transactions on Computers*, vol. 69, no. 7, pp. 998-1014, 2020.

[3] S.Dutta, H, Jeong, Y. Yang, V. Cadambe, T. M. Low, P. Grover, and G. D. Couples, "NONSILICON, NON-VON NEUMANN COMPUTING—PART II," *Proceedings of the IEEE*, vol. 108, no. 8, 2020.

[4] F. Zhou, and Y. Chai, "Near-sensor and in-sensor computing," *Nature Electronics*, vol. 3, no. 11, pp. 664-671, 2020.

[5] N. Jouppi, C. Young, N. Patil, and D. Patterson, "Motivation for and evaluation of the first tensor processing unit," *IEEE Micro*, vol. 38, no. 3, 10-19, 2018.

[6] D. Ielmini, and H. S. P. Wong, "In-memory computing with resistive switching devices," *Nature Electronics*, vol. 1, no. 6, pp. 333-343, 2018.

[7] L. Chua, (1971). "Memristor-the missing circuit element," *IEEE Transactions on circuit theory*, vol. 18, no. 5, pp. 507-519, 1971.

[8] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80-83, 2008.

[9] C. Wan, P. Cai, M. Wang, Y. Qian, W. Huang, and X. Chen, "Artificial sensory memory," *Advanced Materials*, vol. 32, no. 15, pp. 1902434, 2020.

[10] W. Jiao, M. Lyu, and I. King, "Real-time emotion recognition via attention gated hierarchical memory network," *In Proceedings of*

the AAAI Conference on Artificial Intelligence, vol. 34, no. 05, pp. 8002-8009, April. 2020.

[11] G. Zhou, Z. Ren, L. Wang, B. Sun, S. Duan, and Q. Song, "Artificial and wearable albumen protein memristor arrays with integrated memory logic gate functionality," *Materials Horizons*, vol. 6, no. 9, pp. 1877-1882, 2019.

[12] G. Khodabandehloo, M. Mirhassani, and M. Ahmadi, "Analog implementation of a novel resistive-type sigmoidal neuron," *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 20, no. 4, pp. 750-754, 2011.

[13] H. Lin, and S. Liu, "Circuit dynamics in lobster stomatogastric ganglion based on winnerless competition network," *Physica A: Statistical Mechanics and its Applications*, vol. 560, pp. 125107, 2020.

BIOGRAPHIES

Xiaoyue Ji (ji.xiaoyue@zju.edu.cn) received the B.E. degree in electronics and information engineering in 2016 degree from the School of Electrical Engineering, Harbin Engineering University, China. She is currently working toward the Ph.D. degree from the School of Electrical Engineering, Zhejiang University, China. Her research interests cover memristor and memristive system, artificial neural network, the design and analysis of nonlinear systems based on memristor and computer simulation.

Zhekang Dong (englishp@hdu.edu.cn) received the B.E. and M.E. degrees in electronics and information engineering in 2012 and 2015, respectively, from Southwest University, Chongqing, China. He received the Ph.D. degree from the School of Electrical Engineering, Zhejiang University, China, in 2019. Currently, he is an associate professor in Hangzhou Dianzi University, Hangzhou, China. He is also a Research Assistant (Joint-Supervision) at The Hong Kong Polytechnic University. His research interests cover memristor and memristive system, artificial neural network, the design and analysis of nonlinear systems based on memristor and computer simulation.

Chun Sing Lai (chunsing.lai@brunel.ac.uk) received the BEng in electronic and electrical engineering from Brunel University London, UK, and DPhil in engineering science from the University of Oxford, UK in 2013 and 2019, respectively. Dr Lai is currently a Lecturer at the Department of Electronic and Electrical Engineering, Brunel University London, UK. His current interests are in data analytics, power system optimization, energy system modelling, and energy economics for low carbon energy networks and energy storage systems.

Donglian Qi (qidl@zju.edu.cn) received the Ph.D. degree from the School of Electrical Engineering, Zhejiang University, China, in 2002. She is currently a Full Professor and a Ph.D. Advisor with Zhejiang University. Her recent research interest covers intelligent information processing, chaos system, and nonlinear theory and application.