

Realizing Behavior Level Associative Memory Learning Through Three-Dimensional Memristor-Based Neuromorphic Circuits

Hongyu An , *Student Member, IEEE*, Qiyuan An, *Student Member, IEEE*, and Yang Yi , *Senior Member, IEEE*

Abstract—Associative memory is a widespread self-learning method in biological livings, which enables the nervous system to remember the relationship between two concurrent events. The significance of rebuilding associative memory at a behavior level is not only to reveal a way of designing a brain-like self-learning neuromorphic system but also to explore a method of comprehending the learning mechanism of a nervous system. In this paper, an associative memory learning at a behavior level is realized that successfully associates concurrent visual and auditory information together (pronunciation and image of digits). The task is achieved by associating the large-scale artificial neural networks (ANNs) together instead of relating multiple analog signals. In this way, the information carried and preprocessed by these ANNs can be associated. A neuron has been designed, named signal intensity encoding neurons (SIENs), to encode the output data of the ANNs into the magnitude and frequency of the analog spiking signals. Then, the spiking signals are correlated together with an associative neural network, implemented with a three-dimensional (3-D) memristor array. Furthermore, the selector devices in the traditional memristor cells limiting the design area have been avoided by our novel memristor weight updating scheme. With the novel SIENs, the 3-D memristive synapse, and the proposed memristor weight updating scheme, the simulation results demonstrate that our proposed associative memory learning method and the corresponding circuit implementations successfully associate the pronunciation and image of digits together, which mimics a human-like associative memory learning behavior.

Index Terms—Memristor, associative memory, artificial neural networks, three-dimensional integrated circuit.

I. INTRODUCTION

BUILDING a neuromorphic computing system with a self-learning capability like the brain has been investigated for a long time [1]. The direct self-learning capability can potentially allow the machines to have the adaptability of performing complex tasks in a dynamic environment, like domestic robotics [2]. Self-learning capability of biological livings comes from associative memory learning [3], which enables them to relate two events that occur simultaneously [3], [4]. Through this learning method, dogs can learn the sound of bells as a sign of food;

people can remember a word representing an object [3], [4]. The investigations on associative memory at the cellular level reveal that the changes in synaptic weight play a critical role in the associative memory [3]. The weight of a synapse, the amount of the chemical neurotransmitters, represents the connection strength between two neurons. With the increase of the connecting strength between neurons, the relationship between two concurrent stimuli is memorized [3].

The emerging memristive devices are an ideal candidate for electronic synapses since their resistance can be programmed gradually, mimicking the changes in synaptic weight [5]–[9]. Some researchers have investigated to employ the memristive synapses in a small-scale associative memory recently [10]–[18]. However, these attempts only associate simple signals together with several neurons (less than ten connecting synapses) [10], [12], [13], [15]–[18]. More importantly, the information carried by these signals is limited [3], [19]. However, the critical step for realizing a self-learning neuromorphic system is to enable the system a capability to associate several pieces of sophisticated information together [19], [20].

In the human brain, the different types of signals, e.g., sound, vision, are processed at different locations through different types of neural networks [3]. Having similar signal processing capabilities, the ANNs can process different types of signals independently [21]–[24], and abstract the input information to the outputs efficiently. For example, the convolutional neural networks (CNNs) are generally used for processing two-dimensional (2D) image signals [23], while recurrent neural networks (RNNs) are more suitable for processing time series signals [24]. In a classification problem, the outputs of these neural networks are a set of scores representing the probabilities of the input belonging to a particular category. Inspired by the working mechanism of ANNs and distributed signal processing methodology in the brain, a novel behavior level large-scale associative neuromorphic architecture has been proposed. Instead of relating pure analog signals together, this architecture associates multiple ANNs together by adding one more layer of the neural network, referred to the associative neural network in this paper.

The proposed architecture encodes the probabilistic scores of the ANNs into the frequencies and magnitudes of spiking signals through several specifically designed Signal Intensity Encoding Neurons (SIENs). The spiking signals would be further imported into the associative neural network for a large-scale analog-based association. In this way, the information

Manuscript received January 23, 2019; revised April 24, 2019; accepted May 25, 2019. This work was supported by the National Science Foundation under Grants NSF 1731928 and NSF 1750450. (Corresponding author: Hongyu An.)

The authors are with the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA (e-mail: hongyu51@vt.edu; anqiyuan@vt.edu; cindy_yangyi@vt.edu).

Digital Object Identifier 10.1109/TETCI.2019.2921787

preprocessed and carried by the ANNs is associated with each other. In this paper, we theoretically discussed the methodology of realizing behavior level large-scale associative memory learning, and the corresponding circuit designs. The detailed contributions can be summarized as:

- 1) Instead of relating signals, a large-scale associative neuromorphic architecture is proposed to associate ANNs together for implementing a behavior level associative memory learning. Particularly we demonstrated an associative memory behavior on learning the pronunciation (auditory signal) and image (visual signal) of digits in this paper.
- 2) In order to associate the outputs data of the ANNs together, several circuitry modules are designed: SIENs encoding the input, e.g., image, audio, into the magnitude and frequency of an analog spiking signal, and a 3D memristive synapse array serving as the associative neural network, and a novel memristor weight updating scheme with no selector devices;
- 3) Compared with other state-of-the-art memristor-based associative memory models (<10 synapses) listed in Table V, the proposed 3D large-scale memristive synapse model successfully relates the signals from 20 neurons together with 100 memristive synapses, realizing a behavior level large-scale associative memory learning.

This paper is organized as follows: Section II introduces the background of associative memory in biology; Section III discusses the proof-of-concept method that realizes a behavior level large-scale associative memory learning; Section IV demonstrates the corresponding circuitry module designs; Section V comprehensively summarizes our work.

II. ASSOCIATIVE MEMORY IN BIOLOGY

Associative memory was initially investigated at a behavior level by Ivan Pavlov through a series of experiments with dogs [3]. In the experiments, Pavlov first rocked the bell and then provided the food to the dog [3]. After a few repetitions, Pavlov noticed that the dog started to salivate when the bell sounded around him even with no food presented. By studying this phenomenon, Pavlov concluded that salivation, normally evoked by a visual input from food, can also be invoked from a disparate signal perception pathway, like auditory sensation.

In Pavlov's study, the dog food is defined as an unconditional stimulus (US) because it would unconditionally evoke the salivation reflection without learning procedures. Meanwhile, the sound of the bell is defined as a conditional stimulus (CS) because its evocative capability is acquired by learning. Pavlov's study reveals that the stimulus signals from two unrelated events can be associated with each other by their concurrence repeatedly. This self-learning behavior by relating concurrent events phenomenon is widely referred to associative memory [3].

The associative memory learning at the cellular level was investigated by Dr. Kandel's research on Aplysia (2000 Nobel Prize) [3]. As illustrated in Fig. 1, the associative memory learning mechanism in Aplysia is simplified into two signal pathways marked in blue and red respectively.

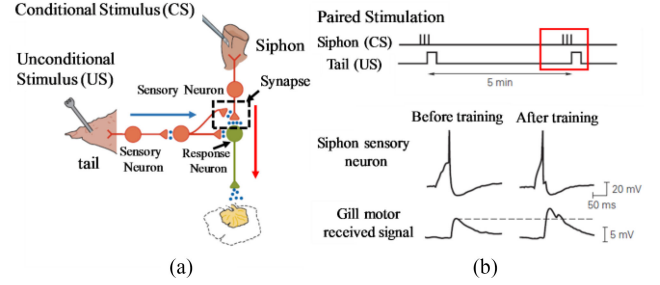


Fig. 1. (a) Conditional stimulus pathway and unconditional stimulus pathway in the cellular level associative memory learning mechanism of Aplysia; (b) a larger magnitude of the received signal at gill motor neuron under paired stimulus from Siphon (CS) and Tail (US) [3].

The US applied on the tail unconditionally evokes the shrinking response of a gill motor neuron. However, the CS from the siphon does not invoke the response of the gill motor neuron alone due to the high signal attenuation effect of the synapse connecting the sensory neuron and response neuron. The higher attenuation effect of the synapse stimulates, the lower received input signal at postsynaptic neuron (motor neuron). This attenuation effect of the synapse comes from the chemical neurotransmitter molecules released from the synapse. When the neurotransmitter arrived at the terminal of the postsynaptic cell, a spiking signal would be stimulated.

The magnitude of the stimulated spiking signal at the postsynaptic cell is highly dependent on the amount of the neurotransmitter received. A larger amount of neurotransmitter molecules stimulates a larger magnitude spiking signal and vice versa. The amount of the neurotransmitter determines the connection strength between neurons, which is widely referred to the "weight" of the synapse.

Normally, the gill motor is unresponsive to siphon stimulation of the siphon before learning. However, by performing a training experiment which consisted of applying a shock to the tail (US) and touching the siphon (CS) simultaneously and repeatedly, the gill motor neuron became more responsive to inputs from the siphon sensory neuron (CS). As depicted in Fig. 1(b), the stimulus from US and CS are paired and overlapped with each other in time that is considered as a trigger condition of associative memory learning at the cellular level [3]. The increased magnitude of the gill motor response results from a stronger synaptic connection induced or imprinted between the sensory neuron of the siphon and the motor neuron of the gill during the associative learning process. This cellular association learning behavior comes from the increment connection strength between the sensory neuron and response neuron due to the repeatedly and simultaneously US and CS.

III. FROM SMALL-SCALE TO LARGE-SCALE ASSOCIATIVE MEMORY LEARNING

In this section, we discuss how to extend the associative memory learning from a cellular level associating pure signal to a behavior level having the capability of associating multiple pieces of sophisticated information together.

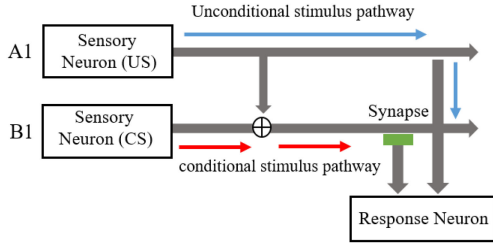


Fig. 2. Cellular level associative memory model with a memristor as the electronic synapse.

In the cellular level associative memory, a larger voltage received at the postsynaptic neuron demonstrates a successful associative memory learning resulting in less attenuation effect of the synapse. Our previous work [25] realizing this less attenuation effect physically through a memristive synapse is illustrated in Fig. 2.

In this model, the cellular level associative neural network is simplified (Fig. 1) into two main signal pathways: conditional and unconditional pathway, respectively. The unconditional pathway directly connects the sensory neuron A1 (US) to the response neuron, while the conditional pathway connects sensory neuron B1 (CS) to the response neuron through a memristive synapse.

On the conditional signal pathway, an analog summation device is used to couple conditional stimulus from neuron B1 and an unconditional stimulus from neuron A1. Initially, the stimulus signal from B1 to response neuron is small due to the attenuation effect caused by the high resistance of the memristor. Furthermore, the magnitude of the spiking signals generated by A1 and B1 are both smaller than the set voltage of the memristor, meaning the signals from A1 and B1 cannot update the weight of memristive synapse alone. Consequently, the associative memory learning cannot be achieved. However, when the neuron A1 and B1 fire simultaneously, their coupled output spiking signals would potentially exceed the set voltage of the memristive synapse, consequently decreasing its resistance. As a result, the magnitude of the signal arriving in response neuron is increased indicating this model perfectly reproduces the cellular level associative memory learning phenomenon in Aplysia.

The main drawback of the cellular level associative memory model is the associated signals can only carry limited sophisticated information restricting the capability of the system from learning more complex information. Nevertheless, the pieces of sophisticated information can be processed by various ANNs. The outputs of an ANN are usually a probabilistic number (score) between “0” to “1”, representing a degree of prediction accuracy. The score indicates the probability of the original import data, e.g., video, voice, belonging to a specific category.

In this way, the information carried by these images, voices, etc., is transformed and embedded into a series of probabilistic scores. Therefore, if these scores are associated together, the information carried by these scores theoretically would be also related together. In this paper, this idea is implemented by us with a large-scale associative neuromorphic architecture illustrated in Fig. 3.

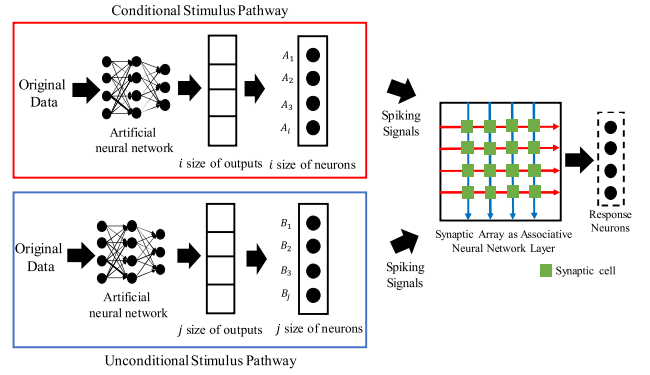


Fig. 3. Proposed large-scale associative neuromorphic architecture partitioned into two pathways constructed by two ANNs.

As illustrated in Fig. 3, the original data is first processed by the ANNs. The information carried by the original data is abstracted into the output scores of ANNs. Then the scores are further imported into the SIENs. Next, SIENs encode the scores into a series of spiking signals whose magnitudes and frequencies corresponding to the values of the scores. The highest scores would be transferred into a spiking signal with the highest peak magnitudes and shortest interval between spikes, accordingly. At last, the spiking signal outputs of SIENs would be delivered to a synaptic array for a large-scale association. The size of the synaptic array is $i \times j$ which are the index of SIENs at two stimulus pathways as illustrated in Fig. 3. The input original data of these two stimulus pathways could be visual and auditory signals, corresponding to the presence of food and sound of bells in the Pavlov's behavior level associative memory learning experiment. In this paper, we associate the visual (image) and auditory data (pronunciation) of digits together.

In the synaptic array, the spiking signals couple and superpose with each other at the synaptic cells described by the following equation:

$$V_{synapse} = V_{coupled} = V_{A_i} + V_{B_j}, \quad (1)$$

where V_{A_i} and V_{B_j} are the output spiking signals from SIEN A_i and B_j , respectively. $V_{synapse}$ is the voltage potential between the terminals of the synapse, which is the coupled signal from V_{A_i} and V_{B_j} . Since the scores from ANNs are different (within the interval [0 1]), the magnitudes of the V_{A_i} (V_{B_j}) are various accordingly. Apparently, the largest spiking signals $V_{synapse_{max}}$ would be generated from the largest signals of SIEN A_i ($V_{A_{max}}$) and B_j ($V_{B_{max}}$). An associative memory learning behavior would occur under the condition of $V_{synapse_{max}} > V_{set}$, where V_{set} is the set voltage of the memristor.

IV. LARGE-SCALE ASSOCIATIVE NEUROMORPHIC ARCHITECTURE IMPLEMENTATIONS

A. 3-D Memristor-Based Synaptic Array

The memristive device, also referred to Resistive Random-access Memory (RRAM), is widely applied as an ideal electronic synapse candidate due to its programmable resistance [13].

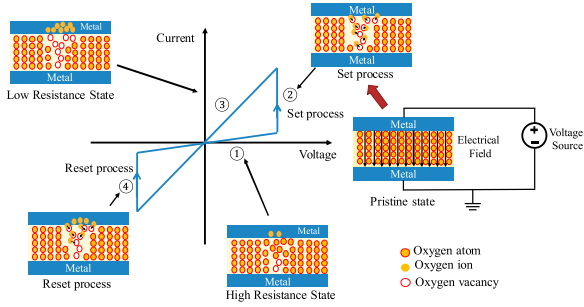


Fig. 4. Illustration of the switching mechanism of a memristor. The memristor has two states (HRS and LRS) marked as (1) and (3), and two transition states (set and reset processes) marked as (2) and (4), respectively. Note that this paper would mainly focus on modeling the set process indicated as a remembering process instead of a biological disremembering process.

The resistance of a memristor is modified with the applied voltage on its terminals exceeds a specific value, called as its set voltage. The resistance modification from the high resistance state (HRS) to the low resistance state (LRS) is defined as a set process. Typically, the memristor is constructed by the metal-insulator-metal configuration. The decrease of the resistance is caused by the formation of the conductive filament in its insulator layer. The increase of synaptic weight, indicating a successful associative memory learning behavior [3], can be realized by programming the resistance of the memristor from its HRS into LRS. Consequently, the received voltage/current of the postsynaptic response neuron would increase, demonstrating the accomplishment of the learning processes [3].

In the metal oxide, the bonding between oxygen ions and metal atoms is breakable. Under the high electric field (>10 MV/cm) stimulated by the applied voltage, some oxygen ions in the metallic oxide would escape from the constraint of the bonding force and drift toward the anode side of a memristor [26]. Fig. 4 demonstrates the switching states of a memristor and the corresponding formation of CFs. The deficiency of oxygen ions leaves the oxygen vacancies or metal precipitates, which would further construct the CFs [27], [28]. As a result, two current paths exist in its LRS. One is through the original oxide and the other is through CFs. These two paths in the parallel lead to the decline of the memristor resistance. In the reset process, the oxygen ions at the interface migrate back into the oxide to refill the oxygen vacancy or re-oxidize the metal precipitates to update the resistance of the memristor back to its HRS.

The memristive synapse in this paper is used for demonstrating a biological-like associative memory mechanism (Fig. 1) indicating the synaptic connection strengthening between neurons as the associative learning accomplishment. This strengthening behavior is modeled as the memristor resistance switching from HRS to LRS. Therefore, this paper would mainly focus on modeling the set process of the memristor without discussing the reset process, which reduces the connection strength between neurons and is considered as a biological disremembering phenomenon [3].

In this paper, the memristive synapse is modeled with the filament growing method [29]. As illustrated in Fig. 4, the resistance

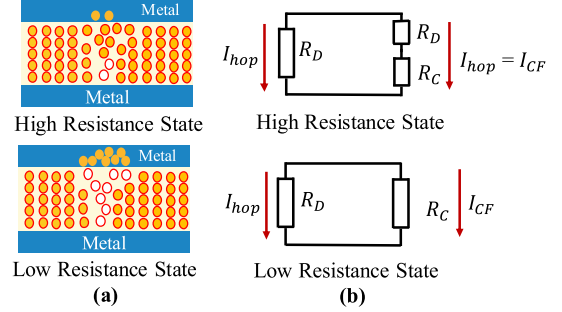


Fig. 5. Current paths of the memristor at the HRS and LRS.

switching between HRS and LRS comes from the construction/deconstruction of the CFs in the metallic oxide. The CFs in the oxide provide an alternative current path with lower resistance. By modeling two current paths with different resistances, notated with R_D (dielectric resistance) and R_C (resistance of CFs), the memristor models at HRS and LRS are illustrated in Fig. 5. Since the disconnection of the memristor only occurs at the interfacial region, the resistance of LRS is actually combined with two cascaded parts, R_D and R_C .

The currents in the CFs and intact oxide region are modeled with metal-like (I_{CF}) and hopping current (I_{hop}), respectively. The resistance of HRS is mainly determined by the R_D with the hopping current I_{hop} , and the resistance at LRS is dominated by the R_D with the current I_{CF} . The current I_{CF} and I_{hop} are governed by the equations in the filament growing method [29]:

$$I_{hop} = I_0 \left(\pi w^2 / 4 \right) \exp(-x/x_T) \sinh(V_{gap}/V_T), \quad (2)$$

$$I_{CF} = \frac{\pi w^2 V_{CF}}{4\rho(x_0 - x)}, \quad (3)$$

where x_0 is the initial value of gap distance. x_T and V_T are the characteristic length and voltage in hopping, respectively. V_{gap} and V_{CF} are the voltage over the gap region and CF region, respectively. In the set process, the w , and x are growing under the stimulus voltage by the following equations:

$$dx/dt = af \exp(-(E_a - \alpha_a ZeE)/k_B T) \quad (4)$$

$$dw/dt = \left(\Delta w + \frac{\Delta w^2}{2w} \right) f \exp\left(-\frac{E_a - \alpha_a ZeE}{k_B T}\right). \quad (5)$$

The parameters in Equ. (4) and (5) are listed in Table I.

Based on the conductive filament evolution concept discussed, a memristor model is developed for the memristive synapse array simulation in our large-scale associative memory learning system. Fig. 6 illustrates the V-I characteristic curve comparison in the set process of our memristor model and the measurement data. As depicted in Fig. 6, the resistance of the memristor model would switches from its HRS (1.6 M Ω) to LRS (64 K Ω) at ~ 3.2 V. the current is at ~ 50 μ A, which matches the measurement data. The current response mismatch above 50 μ A comes from the activated current-compliance for protecting the device on the measurement setting. The detailed parameters of the memristor model are listed in Table I.

TABLE I
PARAMETERS OF THE MEMRISTOR MODEL

Parameter	Descriptions	Values
I_0	Hopping current density in the gap region	$1\text{E}13 \text{ A/m}^2$
ρ	Resistivity of the CF	$2.5\text{E-}4 \Omega/\text{m}^2$
a	Distance between adjacent oxygen vacancy	0.25 nm
f	Vibration frequency of oxygen atom	$1\text{E}13$
x_T	Characteristic length in hopping region	$0.4\text{Ee-}9$
V_T	Characteristic voltage in hopping	0.4
w_0	Initial CF width	$1\text{E-}9$
R_H	High Resistance State	1.6 M Ω
R_L	Low Resistance State	64 K Ω
E_a	Average active energy	1.2eV
α_a	Enhancement factor	0.75 nm
$Z \& e$	Charge number & unit charge	1 & e
k_B	Thermal resistance	$0.86177\text{e-}5$

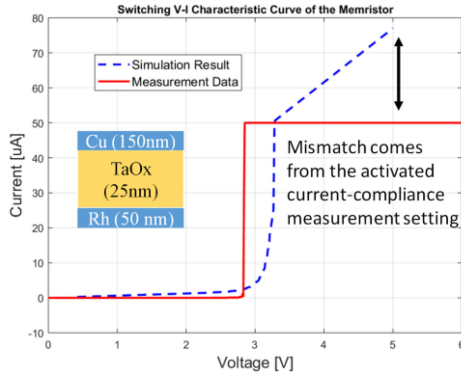


Fig. 6. Set switching V-I characteristic curve of the memristor. The current response mismatch above $50 \mu\text{A}$ comes from the activated current-compliance for protecting the device on the measurement setting.

TABLE II
MEASUREMENT RESULTS OF THE MEMRISTOR

Parameters	Value
V_{form}	4 V
V_{set}	2.85 V
V_{reset}	-3 V
The thickness of Cu layer	150 nm
The thickness of TaO_x layer	25 nm
The thickness of Rh layer	50 nm

The measurement data in Fig. 6 come from the memristive device (Cu/TaOx/Rh) fabricated at the Micro and Nanofabrication Laboratory at Virginia Tech [30]. In the memristor, Copper (Cu) serves as a top metal electrode, oxygen-deficient tantalum oxide (TaOx) as solid electrolyte and Rhodium (Rh) as a bottom electrode. The device has been characterized by monitoring the forming voltage (V_{form}) when conductive filaments (CFs) are being formed initially. The reset voltage (V_{reset}), the set voltage (V_{set}), and the resistance switching characteristic with the applied ramp-shape stimulus having a rate of 2.0 V/s. Table II

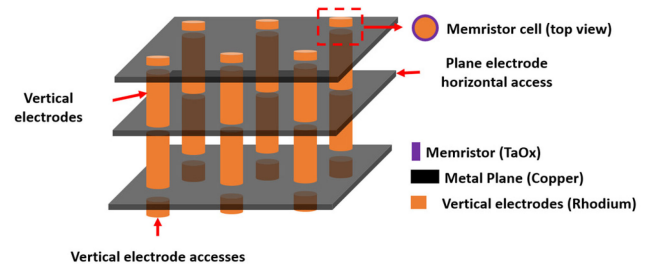


Fig. 7. 3D vertical memristive synapse structure.

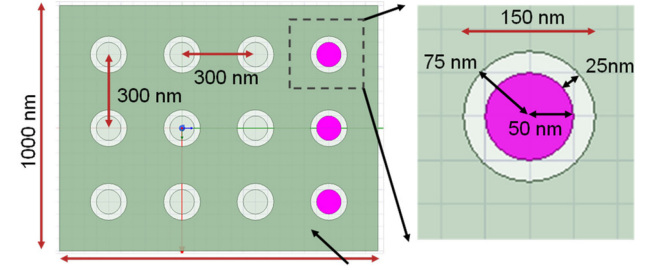


Fig. 8. Top view of the 3D vertical memristive synapse structure.

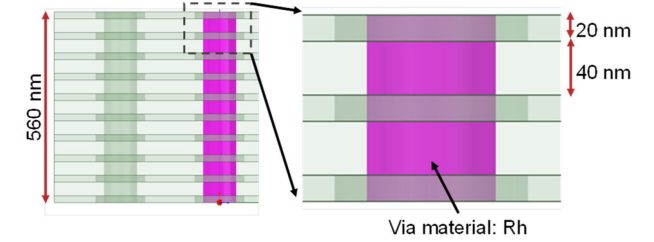


Fig. 9. Side view of the 3D vertical memristive synapse structure.

lists the characteristic parameters of the fabricated memristor. For this device, the set voltage is 2.85 V and the reset voltage is -3 V.

The traditional large-scale memristor array is fabricated in a 2D crossbar configuration which suffers the large design area, power consumption, etc. Therefore, in this paper, a vertical memristor structure is used to offer the following promising benefits, the design area, and power consumption would be reduced by 50% [6] and 35% [31], respectively. Furthermore, a plane is used as the layer access port due to the large resistance attenuation effect of the narrow nanowire on accessing multiple memristors [32].

Figure 7 illustrates our vertical 3D memristive synapse array structure. The geometry of the structure is illustrated in Fig. 8 and Fig. 9. This structure uses vertical planes and monolithic inter-tier vias (MIVs) serving as horizontal and vertical access ports. The MIVs electrode and the plane materials were modeled as copper and rhodium, respectively. The TaOx is used as memristor material sandwiched at the intersection region between the horizontal plane and the vertical MIVs. The 3D vertical memristor structure can be modeled with an array configuration illustrated in Fig. 10. Since the memristor at each layer are connected with each other with a plane metal physically, the port denoted as $Port_P_i$, can access each memristor with the plane resistance denoted as R_{plane} . The resistances of the MIVs is

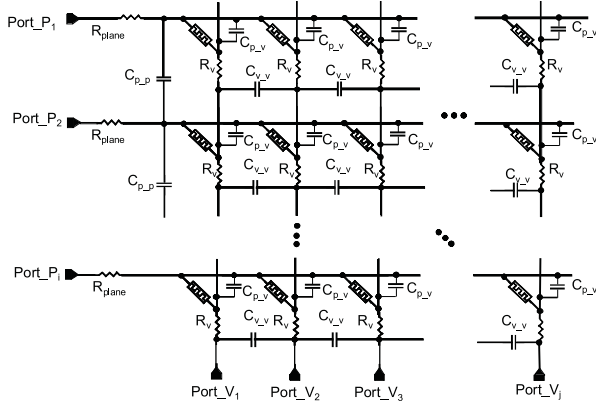


Fig. 10. Model of the vertical memristive synapse array.

TABLE III
PARAMETERS OF OUR VERTICAL 3-D MEMRISTIVE SYNAPSE MODEL

Parameters	Descriptions	Values
R_{plane}	The resistance of the plane	1.179 Ω
R_{via}	The resistance of inter-layer via	0.406 Ω
C_{v-v}	The parasitic capacitance between the vias	1.19 E-8 pF
C_{p-v}	The parasitic capacitance between the plane and the via	7.43 E-6 pF
C_{p-p}	The parasitic capacitance between the planes	7.6 E-5 pF

TABLE IV
GEOMETRY AND MATERIALS OF OUR VERTICAL 3-D MEMRISTIVE SYNAPSE

Parameters/Descriptions	Values
The distance between the MIVs	300 nm
The radius of the MIVs	50 nm
The distance between the MIVs and the anti-pads	25 nm
The size of the plane	1000 nm $\times$ 1300 nm
The distance between the planes	40 nm
The thickness of the plane	20 nm
The material of the plane	Copper
The material of the via	Rhodium
The insulator between the planes	SiO ₂

denoted R_v . The values of the parasitic capacitance between the planes (C_{p-p}), the plane to the via (C_{p-v}), and the MIV to the MIV (C_{v-v}) are listed in Table III. These values are extracted by the ANSYS Q3D Extractor, an industry standard tool for capacitance and resistance computation. The detailed geometry of the 3D vertical memristive synapse structure is listed in Table IV. Due to the extremely small parasitic capacitance ($\sim$ fF), the effect of parasitic capacitance in our design is negligible.

B. Signal Intensity Encoding Neuron Design

In the proposed behavior level large-scale associative memory learning, SIENs would be used to encode the analog input signals into the frequency and magnitude of the spiking signal outputs. As a result, the proposed SIENs implement two unique

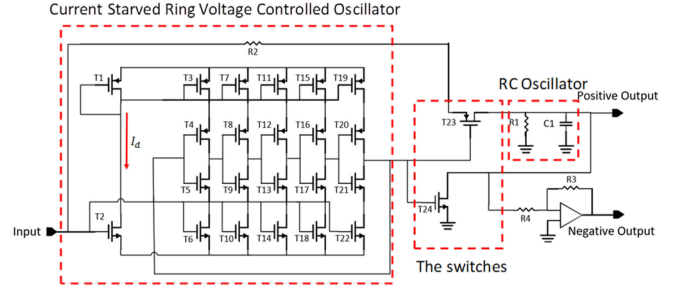


Fig. 11. Signal Intensity Encoding Neuron (SIEN) schematic.

characteristics: input dependent firing frequency/magnitude, and simultaneous excitatory/inhibitory outputs. Although these features widely exist in biological neurons [3], other state-of-the-art neuron designs [33]–[36] lack the realization of these features. The associative memory learning is realized through updating the synaptic weight with a concurrent firing behavior of the sensory neurons at US and CS pathways. The weight updating behavior occurrence depends on whether the magnitude of the coupling signal from the sensory neurons exceeds the set voltage of the memristor (electronic synapse). Thus, the SIENs, as the sensory neurons, are specifically designed to generate a spiking signal, whose magnitude is proportional to the input stimulus (see Equ. (6)–(9)). The model of SIEN is simulated by TSMC 180 nm technology.

As a result, the external stimulus signal with lower magnitude generates the spiking signals with smaller magnitude accordingly, which thus can not trigger the associative memory learning. As introduced in Section III, the coupled spiking signal from neurons A_i and B_j is responsive to updating the weight of memristive synapse. The higher main frequency (smaller intervals between spikes) of the spiking signal would increase the opportunity of superposition of two spiking signals.

As depicted in Fig. 11, there are three central parts of an SIEN: Current Starved Ring Voltage Controlled Oscillator (VCO), a switch pair, and a resistor-capacitor (RC) oscillator. The analog input signal would firstly be imported into the Current Starved VCO to generate an oscillating signal, and its frequency is proportional to the input signal magnitude. Next, this oscillating signal controls a switch pair constructed with a PMOS (positive channel metal oxide semiconductor) transistor and an NMOS (negative channel metal oxide semiconductor) transistor. By controlling the oscillating signal, the switch pair would be charging and recharging the RC oscillator to generate a spiking signal sequence. The frequency of the generated spiking signal sequence by RC oscillator would be proportional to the magnitude of the input analog signal due to the Current Starved VOC controlling the “on” and “off” switching frequency of the switch pair. The neuron firing frequency is determined by the Current Starved VOC with the governing equation [37]:

$$f_{fire} = \frac{I_d}{NC_{total}V_{DD}}, \quad (6)$$

where N is the number of inverter stage, C_{total} is total charging and discharging capacitance of one stage inverter in Current Starved VOC, and V_{DD} is the power supply voltage. The firing

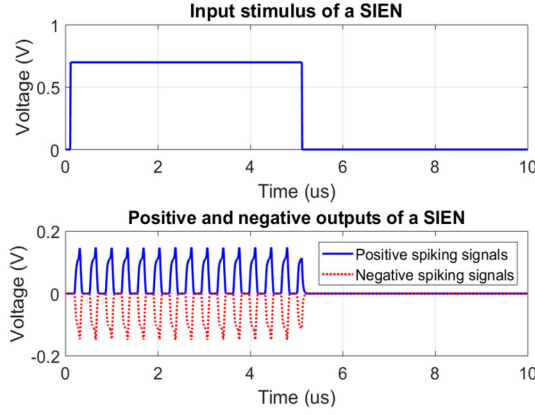


Fig. 12. Positive and negative output spiking signals of an SIEN with 700 mV square wave signal as an input stimulus.

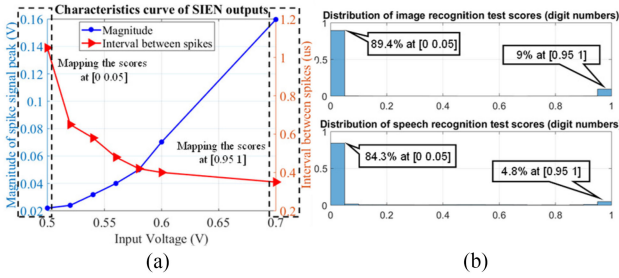


Fig. 13. (a) Characteristics curve of SIEN outputs (b) Distribution of image and speech recognition scores on digits using the datasets: MNIST and Spoken Digit Commands Dataset.

frequency is determined by the current I_d , controlled by the input stimulus as illustrated in Fig. 11.

Moreover, the source terminal of the PMOS transistor in the switch pair is connected to the input signal serving as a charge provider to control the magnitude of the output spiking signal. The effective switching resistances of the PMOS and the NMOS are denoted as R_p and R_n , respectively.

The governing equations of the charging and discharging processes are listed as:

$$V_{charge} = \frac{R_1}{R_1 + R_c} \times V_{input} (1 - e^{-\tau t}), \quad (7)$$

$$V_{discharge} = \frac{R_1}{R_1 + R_p + R_2} e^{-t/(R_d C_1)} \times V_{input}, \quad (8)$$

where R_c equals $R_2 + R_p$, τ is $(R_1 + R_c)/R_1 R_c C_1$, R_d represents $R_1 R_n / (R_1 + R_n)$. The steady-state voltage value of the output is governed by the equation:

$$V_{output} = \frac{R_1}{R_1 + R_p + R_2} V_{input}. \quad (9)$$

Moreover, the SIENs could also generate positive and negative signals simultaneously, which is critical for our novel memristive synapse updating method. Fig. 12 demonstrates the positive and negative output spiking signals of an SIEN with 700 mV square waveform as the stimulus input. The firing response frequency and magnitude corresponding to the different input voltages is illustrated in Fig. 13(a). In this paper, the pronunciations (audio

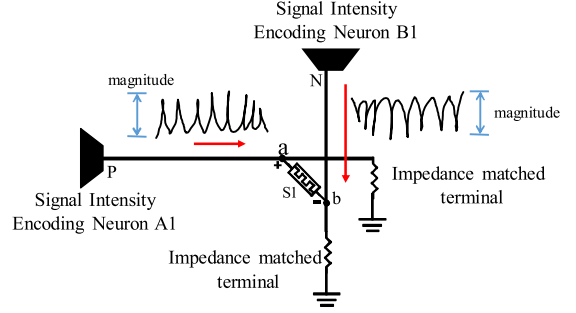


Fig. 14. Novel memristor weight updating scheme.

signal) and images (visual signal) of digits are associated together to produce a behavior level associative memory learning. The SIENs need to map the scores to the frequency and magnitude of their outputs. As depicted in Fig. 13(b), the scores mainly distribute within the intervals $[0, 0.05]$ and $[0.95, 1]$, indicating the lowest and highest scores respectively. This means the input of SIENs will be within two separated ranges, below 0.05 V and above 0.7 V, accordingly, which are marked in Fig. 13(a).

The scores in Fig. 13(b) are generated by using the datasets of Modified National Institute of Standards and Technology database (MNIST) for digit image recognition [38] and Spoken Digit Commands Dataset (SDCD) for digit speech recognition. SDCC is a subset of the Speech Commands Dataset from Google containing 10,000 training and 1,000 test recordings corresponding to spoken digits from 0 to 9 [39].

C. Cellular Level Small-Scale Associative Memory Learning With Novel Memristor Weight Updating Scheme

The cellular level small-scale associative memory model with memristor discussed in Section III (see Fig. 2) requires additional nanowires and adders for the signal coupling, which increases the circuit design area. To address this issue, we proposed a novel memristor weight (resistance) updating scheme without the extra modules of the previous work [25]. Furthermore, the memristor resistance updating behavior of the proposed scheme is controlled by the applied voltage at its two terminals rather than through a selector device [40]–[43]. Thus, the proposed memristor updating scheme makes a nanoscale 3D synaptic array practicable, since the design area of the 3D memristor array is mainly limited by the large selector device, e.g., transistors or diodes [44].

As depicted in Fig. 14, the memristor in this novel scheme receives two opposite polarity signals at its terminals whose voltage potential difference is the stimulus signal for triggering resistance updating of the memristor. The spiking signals from neuron B1 and A1 can be considered as the waveforms propagating in the wires. With the impedance matched terminals, no reflection signals would cause a distortion of the spiking signals. The weight (resistance) of the memristor would be modified when the voltage potential at the terminals exceeds its set voltage.

Figures 15 and 16 illustrate the simulation results of the proposed memristor weight updating scheme. The output spiking signal of SIEN B1 is negative. In Fig. 15, two square

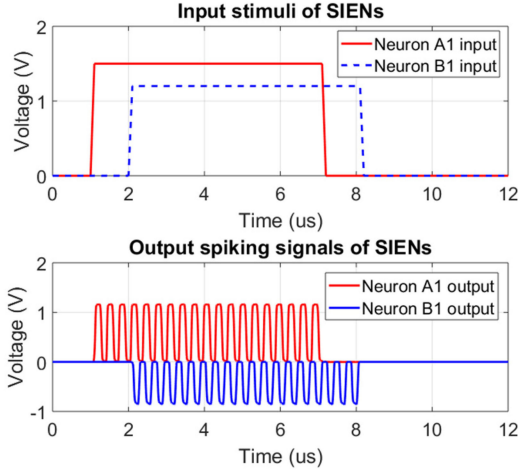


Fig. 15. Input analog signals and output spiking signals of Neuron A1 and Neuron B2.

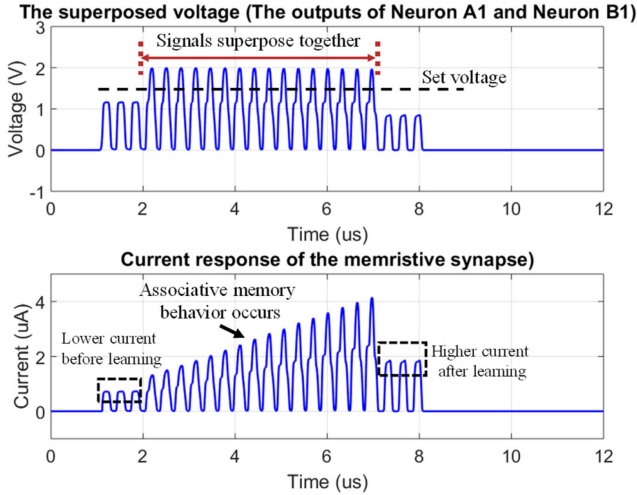


Fig. 16. Voltage potential at terminals of the memristor, which is the superposed voltage from Neuron A and B outputs, and the corresponding current.

inputs of SIENs are not perfectly synchronized and only partially overlapped. At the non-overlapping part, both signals are small, and cannot trigger the memristor switching alone (see Fig. 16).

At the overlapping part, two signals are superposing their peak values with each other. Consequently, the magnitude of the superposed spiking signal will be larger than the set voltage of the memristive synapse, resulting in a resistance modification behavior. As illustrated in Fig. 16, the current after learning is larger than the current before learning indicating a successful associative memory learning behavior.

D. Behavior Level Large-Scale Associative Memory Modeling

By employing the SIENs, 3D memristive synapse array, and the novel memristor weight updating scheme, we produced and mimicked a behavior level large-scale associative memory learning illustrated in Fig. 17. Unlike the cellular level associative

memory with two simple nanowires (Fig. 2), the US and CS pathways in our behavior level large-scale associative memory learning system are constructed by two ANNs that can preprocess and inference the visual and auditory signals respectively. In Fig. 17, the auditory signal and the visual signal of digit number “3” are separately imported into the ANNs for preprocessing. The output is ten scores indicating the probability of the input original data belongs to a specific category. The scores for auditory and visual information of digit 3 are listed in Fig. 17(a). In this paper, we use MNIST [38] and SDCD for the visual and auditory input data, respectively. SDCD is a subset of the Speech Commands Dataset from Google containing spoken digits from 0 to 9 [39]. We can observe that the scores for “3”, marked in red, are highest among other scores. The values of these scores would be further mapped into corresponding spiking signals by SIENs.

Generally, the neural networks of the brain are categorized into training and operating phases [3]. In the operating phase, the topology of the neural network and its synaptic weights are constant, whereas the synaptic weights are changeable in the training phase [3].

As illustrated in Fig. 17(a), the associative memory learning paradigm is divided into two phases: the preprocessing phase and the association phase. the ANNs in the design are used for the operating phase, which means their synaptic weights are trained and fixed. The function of these ANNs is to preprocess the original data from the real world, e.g., visual and auditory signals. The features extracted by them are the image and speech recognition results. Specifically, their outputs indicate the probability of the input (original data) belongs to a specific category.

At the association phase, the prediction scores would be imported into the SIENs to transform the numerical value into a sequence of spiking signals, so that they can be coupled together through the memristive synapse array.

In Fig. 17, the SIENs from visual data is notated as A_i within the unconditional signal pathways. Meanwhile, the sensory neurons (B_j) at conditional signal pathways are connected to the response neurons through a memristive synapse array. Through the SIENs, the largest scores would generate a spiking signal with the largest magnitudes and highest frequencies and vice versa. The memristive synapses connecting the sensory neuron A_i and B_j are notated as $M_{A_iB_j}$. The memristive synapse array for the unconditional pathways (red-dash lines) is modeled by the 3D vertical memristor structure. As illustrated in Fig. 17(a), the memristive associative neural network contains 20 neurons and 100 memristive synapses.

Figure 17(b) and (c) depict the simulation results. With different analog input signals corresponding the scores, the superposed voltage difference at the memristive synapses is different accordingly. The synapse of $M_{A_4B_4}$ has the largest input stimulus due to the corresponding highest scores. Fig. 17(b) illustrates the detailed current response in memristive synapse $M_{A_4B_4}$. When only the auditory signal is provided (no firing behavior in A_i neurons), the current in $M_{A_4B_4}$ is smaller ($< 1\mu A$) than the threshold of the postsynaptic neuron [45], [46]. Meanwhile, as introduced in Section II (Fig. 1), the key condition of a successful associative memory learning is to increase the

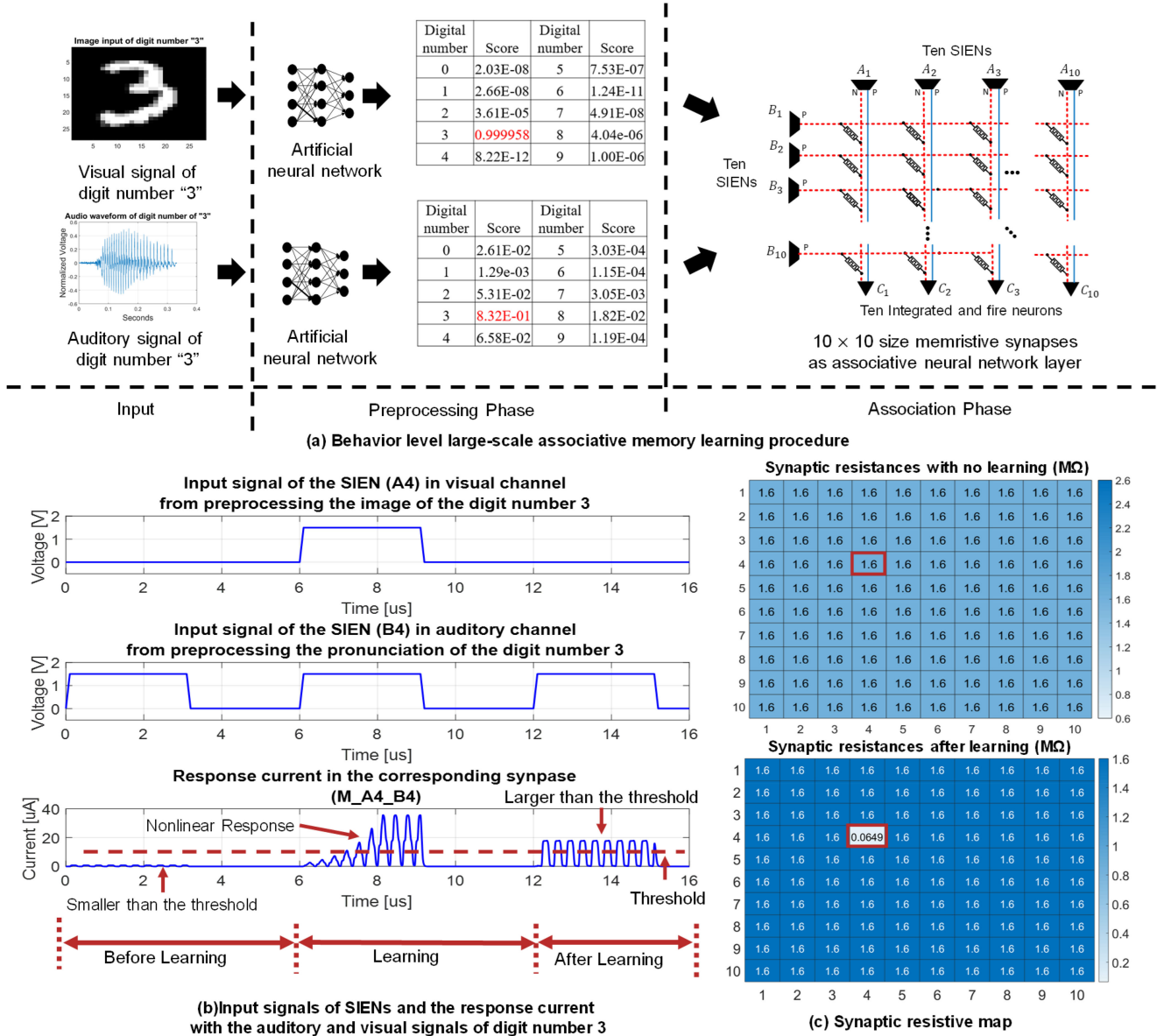


Fig. 17. (a) Behavior level large-scale associative memory learning procedure. (b) the detailed associative memory learning signals at the memristive synapse of M_{A4_B4} . (c) the resistance values of the memristive synapses (HRS and LRS) before and after associative memory learning. The associative memory learning only occurs at M_{A4_B4} marked in the red square.

synaptic connection strength between the sensory neuron and the response neuron so that the received signal at the postsynaptic neuron would exceed its threshold. As a result, the firing phenomena would occur in the postsynaptic neuron. Therefore, the critical design condition of the memristive synapse is its resistance range between the HRS and the LRS should be large enough, so that its response currents before and after learning process are smaller and larger than the threshold of the postsynaptic neuron, respectively. Thus, the effect of the nonlinear updating feature of the memristive synapse on associative memory learning is negligible, as long as its resistance range is sufficiently large as illustrated in Fig. 17(b).

During the learning process, the visual and auditory input are presented simultaneously (firing behavior occur in A_i and B_j neurons), the current in M_{A4_B4} is gradually increasing,

which indicates the resistance reduction of the memristor and the associative memory learning behavior is accomplished.

From Fig. 17(c), we can observe that the memristive synapse of M_{A4_B4} switches from its HRS (1.6 MΩ) to its LRS (64 KΩ). On the contrary, other memristive synapses, connecting the sensory neurons receiving lower input analog stimulus signals, do not switch since the voltage potentials of the spiking signals at their terminals are lower than the set voltage of the memristors.

The simulation results indicate that two outputs of ANNs with the highest probability numbers are associated together to realize a large-scale associative memory learning purpose, which not only relates the pure signals but also associates the large-scale ANNs together. Table V lists the comparison between other state-of-the-art memristor-based associative

TABLE V
COMPARISONS OF SCALES AND ASSOCIATION CAPABILITY WITH OTHER RELATED WORKS

	Scale		Synapse		Neuron	Association Methodology	Association capability
	Neurons	Synapse	Device	Structure			
[18]	6	3	RRAM	2D memristor bridge	Binary neuron model	Hopfield network	Associate signals (cellular level)
[16]	3	1	RRAM	2D	leaky integrate-and-fire	Spike-rate-dependent plasticity	Associate signals (cellular level)
[17]	5	6	RRAM	2D/1R	N/A	N/A	Associate signals (cellular level)
[15]	3	1	RRAM	2D/1R	N/A	N/A	Associate signals (cellular level)
[12]	3	1	RRAM	2D/1R	N/A	Adding	Associate signals (cellular level)
[10]	3	2	RRAM + ADC + digital controller	N/A	Electronic neuron (ADC + microcontroller)	Hebbian rule	Associate signals (cellular level)
[14]	N/A	N/A	PCM	N/A	Integrate-and-fire neurons	Spike timing dependent plasticity	Associate signals (cellular level)
This work	10 + 10	10 × 10	RRAM	3D RRAM structure	SIEN (Ver. 2)	Associate the output of multiple neural networks	Associate visual and audio information together through associate two ANNs together (Behavior level)

memory learning works and our approach mainly in the scale of the learning system and association capability. As we can observe that our approach increases the number of neurons and memristive synapses to 20 and 100, respectively. Unlike other works employing a few memristive synapses, our approach uses an advanced vertical 3D memristive synapse structure. Moreover, the electrical characteristics of the structure are analyzed. At last, the association methodologies are also different. Our design associates two large-scale ANNs together enabling the system to have the capability of learning sophisticated information from the real world, e.g., visual and auditory signals. To our best knowledge, this is the first time of proposing this idea and realizing with memristive devices, making our work has the uniquely innovative contribution to the neuromorphic field.

V. CONCLUSION

In this paper, we proposed and analyzed a novel behavior-level large-scale associative memory learning methodology with the corresponding neuromorphic circuitry designs including SIENs, 3D memristive synapse array, and a synapse updating scheme. Instead of another cellular level associative memory learning methods, our approach successfully associates two large-scale ANNs together, realized by associating the outputs of ANNs with an extra layer of neural network referred to an associative neural network.

The outputs of the ANNs representing the probabilities of the input belonging to a particular category or prediction would be encoded into the magnitudes and frequencies of spiking signals and associated together for the corresponding memristive synapse weight updating. The coupling signal from the two highest values of the outputs of ANNs would decrease the resistance of the memristive synapse from HRS to LRS. The decrease of the synaptic weight demonstrates that the connection between presynaptic and postsynaptic neurons is becoming strong, which

further indicates an accomplishment of successful associative memory behavior.

Through a large-scale simulation with 20 neurons and 100 memristive synapse array, the proposed behavior level associative memory learning system demonstrates the ability to associate the auditory and visual information of digits together like our brain.

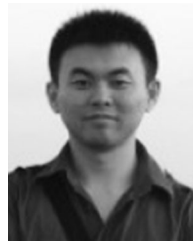
ACKNOWLEDGMENT

The authors would like to present a special appreciation and gratitude to Dr. M. Orlowski, M. S. Mamun, and the Micro and Nanofabrication Laboratory of Virginia Tech [30].

REFERENCES

- [1] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.
- [2] I.-B. Jeong, W.-R. Ko, G.-M. Park, D.-H. Kim, Y.-H. Yoo, and J.-H. Kim, "Task intelligence of robots: Neural model-based mechanism of thought and online motion planning," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 1, no. 1, pp. 41–50, Feb. 2017.
- [3] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. Hudspeth, *Principles of Neural Science*. New York, NY, USA: McGraw-hill, 2000.
- [4] P. I. Pavlov, "Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex," *Ann. Neurosci.*, vol. 17, no. 3, pp. 136–141, 2010.
- [5] S. Kim, H. Kim, S. Hwang, M.-H. Kim, Y.-F. Chang, and B.-G. Park, "Analog synaptic behavior of a silicon nitride memristor," *ACS Appl. Mater. Interfaces*, vol. 9, no. 46, pp. 40420–40427, 2017.
- [6] S. Kim *et al.*, "Neuronal dynamics in HfOx/AlOy-based homeothermic synaptic memristors with low-power and homogeneous resistive switching," *Nanoscale*, vol. 11, no. 1, pp. 237–245, 2019.
- [7] Y. F. Chang *et al.*, "Beyond SiOx: An active electronics resurgence and biomimetic reactive oxygen species production and regulation from mitochondria," *J. Mater. Chem. C*, vol. 6, no. 47, pp. 12788–12799, 2018.
- [8] Y. F. Chang *et al.*, "Demonstration of synaptic behaviors and resistive switching characterizations by proton exchange reactions in silicon oxide," *Sci. Rep.*, vol. 6, Feb. 2016, Art. no. 21268.

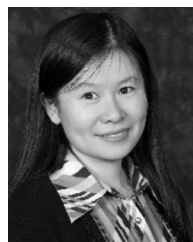
- [9] H. Liang *et al.*, "Memristive neural networks: A neuromorphic paradigm for extreme learning machine," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 1, pp. 15–23, Feb. 2019.
- [10] Y. V. Pershin and M. Di Ventra, "Experimental demonstration of associative memory with memristive neural networks," *Neural Netw.*, vol. 23, no. 7, pp. 881–886, 2010.
- [11] D. Kuzum, R. G. Jeyasingh, and H.-S. P. Wong, "Energy efficient programming of nanoelectronic synaptic devices for large-scale implementation of associative and temporal sequence learning," in *Proc. IEEE Int. Electron Devices Meeting*, 2011, pp. 30.3.1–30.3.4.
- [12] M. Ziegler *et al.*, "An electronic version of Pavlov's dog," *Adv. Funct. Mater.*, vol. 22, no. 13, pp. 2744–2749, 2012.
- [13] D. Kuzum, S. Yu, and H. P. Wong, "Synaptic electronics: Materials, devices and applications," *Nanotechnology*, vol. 24, no. 38, 2013, Art. no. 382001.
- [14] S. B. Eryilmaz *et al.*, "Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array," *Frontiers Neurosci.*, vol. 8, p. 205, 2014.
- [15] K. Moon *et al.*, "Hardware implementation of associative memory characteristics with analogue-type resistive-switching device," *Nanotechnology*, vol. 25, no. 49, 2014, Art. no. 495204.
- [16] X. Liu, Z. Zeng, and S. Wen, "Implementation of memristive neural network with full-function pavlov associative memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 9, pp. 1454–1463, Sep. 2016.
- [17] X. Hu, S. Duan, G. Chen, and L. Chen, "Modeling affections with memristor-based associative memory neural networks," *Neurocomputing*, vol. 223, pp. 129–137, 2017.
- [18] J. Yang, L. Wang, Y. Wang, and T. Guo, "A novel memristive Hopfield neural network with application in associative memory," *Neurocomputing*, vol. 227, pp. 142–148, 2017.
- [19] H. An, Z. Zhou, and Y. Yi, "Opportunities and challenges on nanoscale 3D neuromorphic computing system," in *Proc. IEEE Int. Symp. Electromagn. Compat. Signal/Power Integrity*, 2017, pp. 416–421.
- [20] H. An, K. Bai, and Y. Yi, "The roadmap to realize memristive three-dimensional neuromorphic computing system," in *Advances in Memristor Neural Networks-Modeling and Applications*. London, U.K.: IntechOpen, 2018.
- [21] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [23] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1995.
- [24] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 6645–6649.
- [25] H. An, Z. Zhou, and Y. Yi, "Memristor-based 3D neuromorphic computing system and its application to associative memory learning," in *Proc. IEEE 17th Int. Conf. Nanotechnol.*, 2017, pp. 555–560.
- [26] H. S. P. Wong *et al.*, "Metal-oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012.
- [27] M. Janousch, G. I. Meijer, U. Staub, B. Delley, S. F. Karg, and B. P. J. A. M. Andreasson, "Role of oxygen vacancies in Cr-doped SrTiO₃ for resistance-change memory," *Adv. Mater.*, vol. 19, no. 17, pp. 2232–2235, 2007.
- [28] G.-S. Park, X.-S. Li, D.-C. Kim, R.-J. Jung, M.-J. Lee, and S. J. A. P. L. Seo, "Observation of electric-field induced Ni filament channels in polycrystalline NiOx film," *Appl. Phys. Lett.*, vol. 91, no. 22, 2007, Art. no. 222103.
- [29] H. Li, P. Huang, B. Gao, B. Chen, X. Liu, and J. Kang, "A SPICE model of resistive random access memory for large-scale memory array simulation," *IEEE Electron Device Lett.*, vol. 35, no. 2, pp. 211–213, Feb. 2014.
- [30] M. Al-Mamun, S. W. King, and M. K. Orlowski, "Impact of the heat conductivity of the inert electrode on rram memory cell performance and endurance," *ECS Trans.*, vol. 85, pp. 207–212, 2018.
- [31] M. Swaminathan, "Electrical design and modeling challenges for 3D system integration," presented at the DesignCon, Santa Clara, CA, USA, 2012.
- [32] H. An, M. A. Ehsan, Z. Zhou, and Y. Yi, "Electrical modeling and analysis of 3D synaptic array using vertical RRAM structure," in *Proc. 18th Int. Symp. Quality Electron. Des.*, 2017, pp. 1–6.
- [33] E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Netw.*, vol. 14, no. 6, pp. 1569–1572, Nov. 2003.
- [34] H. Lim *et al.*, "Relaxation oscillator-realized artificial electronic neurons, their responses, and noise," *Nanoscale*, vol. 8, no. 18, pp. 9629–9640, 2016.
- [35] S. Dutta *et al.*, "Dynamics, design, and application of a silicon-on-insulator technology based neuron," *MRS Adv.*, vol. 3, nos. 57/58, pp. 3347–3357, 2018.
- [36] C. D. Schuman *et al.*, "A survey of neuromorphic computing and neural networks in hardware," 2017, arXiv:1705.06963.
- [37] R. J. Baker, *CMOS: Circuit Design, Layout, and Simulation*. Hoboken, NJ, USA: Wiley, 2008.
- [38] Y. LeCun, L. Bottou, Y. Bengio, and P. J. P. O. T. I. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [39] P. J. Warden, "Speech commands: A public dataset for single-word speech recognition," 2018, arXiv:1804.03209.
- [40] S. Kim *et al.*, "Dual functions of V/SiOx/AlOy/p++Si device as selector and memory," *Nanoscale Res. Lett.*, vol. 13, no. 1, Aug. 23 2018, Art. no. 252.
- [41] C.-Y. Lin *et al.*, "Attaining resistive switching characteristics and selector properties by varying forming polarities in a single HfO₂-based RRAM device with a vanadium electrode," *Nanoscale*, vol. 9, no. 25, pp. 8586–8590, 2017.
- [42] M. Guo *et al.*, "Unidirectional threshold resistive switching in Au/NiO/Nb: SrTiO₃ devices," *Appl. Phys. Lett.*, vol. 110, no. 23, 2017, Art. no. 233504.
- [43] C.-C. Hsieh, Y.-F. Chang, Y.-C. Chen, D. Shahrjerdi, and S. K. Banerjee, "Highly non-linear and reliable amorphous silicon based back-to-back Schottky diode as selector device for large scale RRAM arrays," *ECS J. Solid State Sci. Technol.*, vol. 6, no. 9, pp. N143–N147, 2017.
- [44] B. Hudec *et al.*, "3D resistive RAM cell design for high-density storage class memory—A review," *Sci. China Inf. Sci.*, vol. 59, no. 6, 2016, Art. no. 061403.
- [45] H. Jiang *et al.*, "Cyclical sensing integrate-and-fire circuit for memristor array based neuromorphic computing," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2016, pp. 930–933.
- [46] K. Bai and Y. Yi, "DFR: An energy-efficient analog delay feedback reservoir computing system for brain-inspired computing," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 14, no. 4, 2018, Art. no. 45.



Hongyu An received the M.S. degree in electrical engineering from Missouri University of Science and Technology, Rolla, MO, USA. He is currently working toward the Ph.D. degree with The Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. His research interests include artificial intelligence



Qiyuan An received the M.S. degree in computer engineering from Syracuse University, Syracuse, NY, USA. He is currently working toward the Ph.D. degree with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. His research interests include emerging deep learning algorithms/systems and energy-efficient and high-performance implementations of deep learning and artificial intelligence systems.



Yang Yi (M'09–SM'16) received the B.S. and M.S. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA. She is an Assistant Professor with the Bradley Department of Electrical Engineering and Computer engineering, Virginia Tech, Blacksburg, VA, USA. Her research interests include very large-scale integrated circuits and systems and neuromorphic architecture for brain-inspired computing systems.