

Received May 7, 2020, accepted May 12, 2020, date of publication May 20, 2020, date of current version June 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995907

A SPICE Model of Phase Change Memory for Neuromorphic Circuits

XUHUI CHEN¹, HUIFANG HU¹, XIAOQING HUANG¹, WEIRAN CAI²,
MING LIU^{3,4}, (Fellow, IEEE), CHUNG LAM⁵, XINNAN LIN¹, (Member, IEEE),
LINING ZHANG⁶, (Senior Member, IEEE), AND MANSUN CHAN^{7,8}, (Fellow, IEEE)

¹Shenzhen Key Laboratory of Advanced Electron Device and Integration, ECE, Peking University Shenzhen Graduate School, Shenzhen 518055, China

²Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen 518061, China

³Key Laboratory of Microelectronics Devices and Integration Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China

⁴University of Chinese Academy of Sciences, Beijing 100049, China

⁵Jiangsu Advanced Memory Technology Corporation Ltd., Huai'an 223302, China

⁶School of Electronic and Computer Engineering, Peking University, Shenzhen 518055, China

⁷Shenzhen Research Institute, HKUST, Shenzhen 518057, China

⁸Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong

Corresponding authors: Xinnan Lin (xnlm@pkusz.edu.cn), Lining Zhang (lnzhang@ieee.org), and Mansun Chan (mchan@ust.hk)

This work was supported in part by the Natural Science Foundation of China under Grant 61704144, in part by the Shenzhen Science Plan under Grant JCYJ20170810163407761, Grant JCYJ20170818114156474, and Grant JCYJ20180305125340386, and in part by the Croucher Foundation under Project CAS18EG01.

ABSTRACT A phase change memory (PCM) model suitable for neuromorphic circuit simulations is developed. A crystallization ratio module is used to track the memory state in the SET process, and an active region radius module is developed to track the continuously varying amorphous region in the RESET process. To converge the simulations with bi-stable memory states, a predictive filament module is proposed using a previous state in iterations of nonlinear circuit matrix under a voltage-driven mode. Both DC and transient analysis are successfully converged in circuits with voltage sources. The spiking-time-dependent-plasticity (STDP) characteristics essential for synaptic PCM are successfully reproduced with SPICE simulations verifying the model's promising applications in neuromorphic circuit designs. Further on, the developed PCM model is applied to propose a neuron circuit topology with lateral inhibitions which is more bionic and capable of distinguishing fuzzy memories. Finally, unsupervised learning of handwritten digits on neuromorphic circuits is simulated to verify the integrity of models in a large-scale-integration circuits. For the first time in literature an emerging memory model is developed and applied successfully in neuromorphic circuit designs, and the model is applicable to flexible designs of neuron circuits for further performance improvements.

INDEX TERMS Neuromorphic circuits, phase change memory, SPICE model, spike-time-dependent plasticity, spiking neural networks.

I. INTRODUCTION

Traditional computers are based on the Von Neumann structure, which results in a spatial separation between the processor and the memory. Due to the existence of the memory wall problem [1], most of the time is consumed during data migration from memory to CPU when processing big data, which even becomes a major factor affecting computer speed and power consumption [2]. However, in the biological neural brain, the storage unit and the logic unit are

integrated, and data does not need to be transferred between them. When appropriate connections are established between different neurons, a neural network with storage and computing functions can be formed. In the 1980s, Carver Mead, a computer scientist at the California Institute of Technology, first proposed the concept of "neuromorphic" [3], which aims to use the characteristics of analog circuits to imitate the human nervous system, and finally create chips that simulate the human brain.

The implementation of neuromorphic circuits depends on non-volatile memory. Among the many new non-volatile memories, PCM is considered to be one of the next generation

The associate editor coordinating the review of this manuscript and approving it for publication was Shiping Wen.

of mainstream non-volatile memories due to its advantages such as high read speed, high write speed, long life, stable storage, simple process, and multi-level storage. In order to improve the yield of the chip, circuit simulation needs to be performed on the chip before the actual chip manufacturing, and device models are essential for circuit simulation [4]–[6].

Neural networks use brain-like topologies for information processing and are popular because they can effectively solve practical problems (such as speech recognition, face recognition, etc.) [7]–[9]. Compared with artificial neural networks (ANN), the spiking neural network (SNN) contains timing information, has stronger biological interpretation, and has the advantages of event-driven and low power consumption. Considering its advantages, SNN is considered as the next generation neural network [10], [11].

Currently, Some researchers implement neural network system-level simulation without actual synapse models [12]–[16], and other reseachers [17] is directly manufacturing PCM-based neural network chips. However, in the design and simulation of neuromorphic circuits based on PCM synapses, an effective PCM model is needed. Modeling SNN requires more detail time tracking model which is seldom implemented in previous PCM model.

It is common that the resistance change in a PCM cell is usually expressed in term of the input current [18], [19]. However, the input in a real circuit and the solution method in a circuit simulator usually take voltage as an input other than current. As a result, a voltage-driven model is more important in practical situation to mimic the behavior of a PCM cell. Moreover, current PCM models mostly focus on the application of the memory [20]–[22], neglecting the change of its analog characteristics, such as the partial RESET characteristic. To better implement neuromorphic ciucuits, a voltage-driven PCM SPICE model that can reflect the analog characteristics is urgently needed. It is reported that the lateral inhibition mechanism of the human brain can distinguish fuzzy memories [23], [24]. In order to improve the bionics and practicality of neuromorphic circuits, the lateral inhibition function in neuromorphic circuits also needs to be considered.

In this work, a time tracking PCM model suitable for neuromorphic circuits will be developed, and the functionality and integrity of the PCM model will be verified in large-scale-integration circuits simulation.

II. SPICE MODELING OF PCM

At present, the mushroom structure of PCM is widely used in academia and industry. As shown in figure 1, a phase change material (GST) is sandwiched between top electrode and bottom electrode. In figure 1, TE is top electrode, BE is bottom electrode, z_{act} is the radius of the active region, W_G is the width of the top electrode, W_{BE} is the width of the bottom electrode, H_G is the thickness of the phase change layer, H_{BE} is the height of bottom electrode and r_{BE} is the radius of bottom electrode. The PCM shown in figure 1 is in a partial SET state, and the active region has been partially

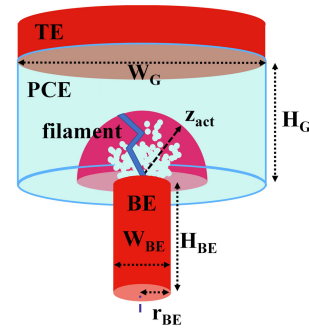


FIGURE 1. Schematic diagram of commonly used PCM mushroom structure.

crystallized, but a stable conductive path has not yet been formed. The filament shown in the figure 1 is a dynamic conductive path. When the electric field inside the PCM exceeds a certain switching threshold, dynamic conductive filaments appear. This will significantly reduce the resistance and make the PCM enter a unstable low resistance state. When the voltage drops and the electric field is below the switching threshold, the PCM returns to the high resistance state. This process is called ovonic threshold switch (OTS) effect [25].

A. MODELING THE GRADUAL SET PROCESS

Memory is thought to be encoded by changes in synaptic strength [26]. To achieve long-term potentiation (LTP), synaptic strength needs to be continuously enhanced. The continuously adjustable nature of PCM is very similar to the non-linear adjustment of biological synapses. Continuous enhancement of PCM synaptic conductance can achieve LTP. It is reported in [27] that PCM is suitable for multi-level storage, and the continuously adjustable feature of PCM has been experimentally proven by reseachers of IBM [17].

In the process of PCM synapse modeling, the dynamic state variable C_f is used to track the crystal fraction of the active area of PCM. According to the phase change dynamics the crystal fraction is determined by the differential Johnson-Mehl-Avrami (JMA) equation [28], which resembles a RC circuit equation format. As shown in (1), a RC circuit equation is used to represent the JMA equation, in which the node voltage V_{Cf} represents the C_f [29].

$$C_{SET} \cdot \frac{dV_{Cf}}{dt} = I_{SET} - \frac{V_{Cf}}{R_{SET}} \quad (1)$$

In (1), C_{SET} is equivalent capacitance, I_{SET} is equivalent current and R_{SET} is equivalent resistance. By setting a SET pulse of appropriate width and height, a slow increase in the crystallization ratio C_f can be achieved. Figure 2 shows the circuit and simulation results. In each set of simulations, the pulse height remains unchanged. After each pulse is applied, the C_f of the PCM is recorded. In the next set of simulations, the pulse height is maintained at another value. The results show that within a certain range, the higher the pulse height, the faster the crystallization.

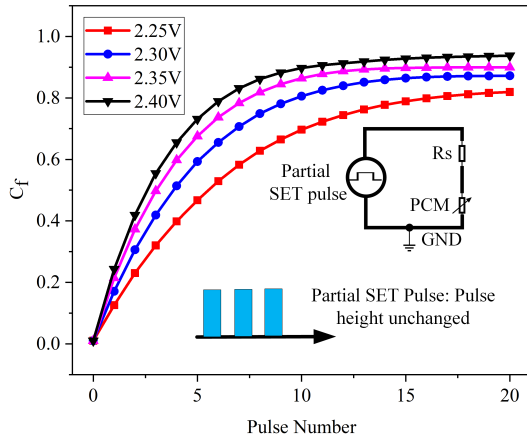


FIGURE 2. Circuit schematic of modeling the gradual SET process and corresponding simulation results. In each set of simulations, the height and width of the pulses are the same. E.g., all 20 pulses have a pulse height of 2.25V and a pulse width of 10ns. In this work, pulses are applied one by one, and the crystalline ratio C_f of the PCM is recored after each pulse.

B. MODELING THE GRADUAL RESET PROCESS

In addition to LTP, long-term depression (LTD) is also widely found in the nervous system and is considered to be an important basis for learning and memory [30]. The continuous weakening of PCM synaptic conductance can achieve LTD.

During PCM synapse modeling, the dynamic variable z_{act} is used to track the active area radius of the PCM. z_{act} is determined by the internal temperature of the active area during the RESET operation. Reference [20] gives an equation for calculating the radius of the active area. The radius depends on the maximum temperature inside the active area. However, there is no detailed description of the temperature calculation, which cannot be used directly in the continuous partial RESET process. In order to calculate the dynamic active area radius, the maximum temperature of each partial RESET process needs to be recorded. The local maximum temperature T_{max} including timing and spatial information is written as (2). Where T_{melt} is the melting temperature, $S_{tranMelt}(T)$ is a smoothing function about the melting temperature, and $S_{tranMax}(T)$ is a smoothing function about the maximum temperature, as shown in (3),(4) respectively.

$$T_{max}(t + \Delta t) = T_{melt} \cdot S_{tranMelt}(T) + [1 - S_{tranMelt}(T)] \cdot [1 - S_{tranMax}(T)] \cdot T \quad (2)$$

$$S_{tranMelt}(T) = \frac{1}{1 + \exp\left(\frac{T - T_{melt}}{\tau_{melt}}\right)} \quad (3)$$

$$S_{tranMax}(T) = \frac{1}{1 + \exp\left(\frac{T - T_{max}}{\tau_{max}}\right)} \quad (4)$$

In (3), τ_{melt} is melting temperature conversion time constant, and in (4), τ_{max} is maximum temperature conversion time constant. Finally, z_{act} is modified as (5), where H_G is the phase change layer height of PCM, and T_{top} is the temperature of the top electrode (default is room temperature).

$$z_{act}(t + \Delta t) = z_{act}(t) \cdot S_{tranMelt}(T) + [1 - S_{tranMelt}(T)] \cdot [1 - S_{tranMax}(T)] \cdot \frac{T_{max} - T_{melt}}{T_{max} - T_{top}} \cdot H_G \quad (5)$$

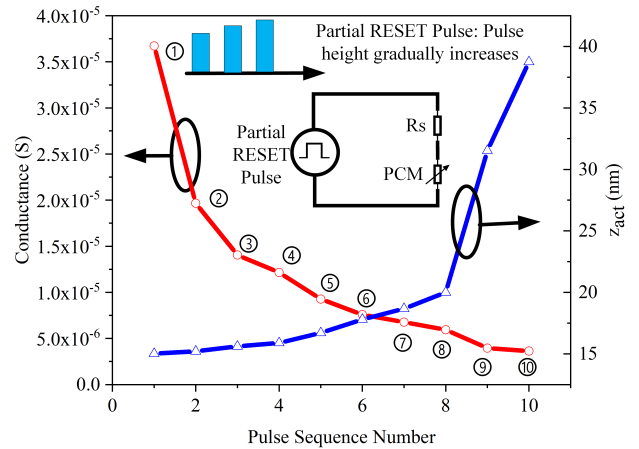


FIGURE 3. The embedded diagram is the circuit schematic of this process. The red line is the simulation result of PCM conductance change with pulse number during gradual RESET process. The blue line is the simulation result of the active area radius. In this process, each pulse causes the internal temperature of the PCM to exceed the melting temperature. The internal temperature increases as the pulse height increases, which causes the radius of the active region to become larger, and finally causes the conductance of PCM to decrease.

Figure 3 shows the circuit and simulation results of gradual RESET process. In this process, each time a pulse is applied, the corresponding resistance value is recorded. The pulse width is the same, but the height gradually increases. The increase in pulse height causes the internal temperature of the PCM to increase, resulting in an increase in the radius of the active area. This makes the amorphous region larger, and the conductance of the PCM further decreases.

C. SPICE MODELING METHODS

The structure-based PCM resistance model can be expressed as (6) [20]. In the equation, R_C is the crystalline resistance outside the active area, and g_f represents the dynamic conductance of the active area filament. R_A represents the static resistance of the active area and it can be calculated by conformal mapping, as shown in (7) [20].

$$R_{PCM} = R_C + \frac{R_A}{1 + g_f \times R_A} \quad (6)$$

$$R_A = \frac{\rho_{act}}{\pi} \frac{H_{REff} (k'_{wRE})}{W_{REff} \left(\sqrt{1 - k_{wRE}^2} \right)} \times \left[\frac{1}{z_{act} + 2W_{BE}} + \frac{2 \ln\left(\frac{2z_{act}}{W_{BE}}\right)}{\sqrt{4z_{act}^2 - W_{BE}^2}} \right] \quad (7)$$

In (7), H_{REff} and W_{REff} are the effective height and width after mapping the hemisphere shaped top electrode to a line parallel to the bottom electrode. k_{wRE} and k'_{wRE} are the modulus of the complete elliptical integral of the first kind for W_{REff} and H_{REff} that satisfy the condition $k_{wRE}^2 + k_{wRE}'^2 = 1$. The effective resistivity of the active region ρ_{act} is described as follows [31]:

$$\rho_{act} = \rho_{amo} + C_f (\rho_{cry} - \rho_{amo}) \quad (8)$$

Therefore, R_A depends on the active area radius z_{act} and the crystal fraction C_f .

The circuit schematic of voltage-driven PCM simulation analysis is shown in figure 4(a). As shown in figure 4(b), the load line of the voltage source intersects with the resistance characteristic curve of the high-impedance PCM at point B(V_2, I_2), and intersects with the resistance characteristic curve of the low-impedance PCM at point A(V_1, I_1). Supposing the PCM is in a high-resistance state at the beginning. During the SET process, the initial steady-state point in the circuit is point B. Since $V_2 > V_{th}$ at this time, OTS effect occurs and a low-resistance filament appears in the active area, causing PCM to enter a temporary low resistance state. At this time, the steady state point in the circuit is point A. However, since $V_1 < V_{th}$, the filament disappears, and PCM returns to high-resistance state, so the steady-state point in the circuit returns to point B. This situation will cause the voltage oscillation phenomenon across the PCM, jumping back and forth between V_1 and V_2 .

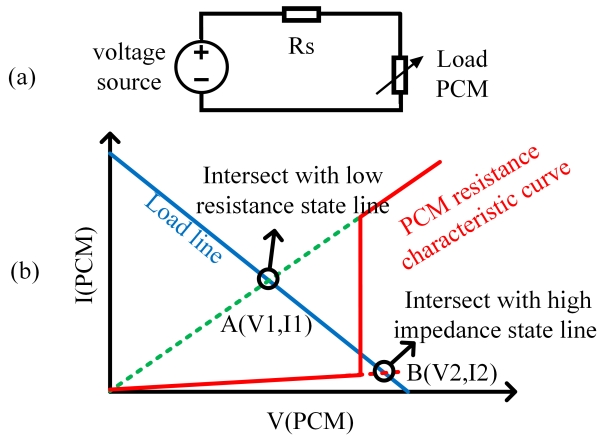


FIGURE 4. (a) Schematic diagram of voltage-driven PCM simulation analysis. (b) Schematic diagram of bistable points of PCM. The solid red line indicates the SET operation of the PCM from a high-impedance state to a low-impedance state, and the dashed green line indicates a low-impedance PCM state. The expression of power load line is $\frac{V}{V_{OC}} + \frac{I}{I_{SC}} = 1$, and the expression of PCM resistance characteristic curve is $V = I \times R_{PCM}$. Where V_{OC} is open-circuit voltage, I_{SC} is short-circuit current and R_{PCM} is nonlinear.

In order to solve the voltage oscillation problem caused by the bistable state in the PCM model, a new conductivity equation of the filament that depends on the previous state is proposed, as shown in (9).

$$g_f = S_{tranGf} \times g_{fOld} \times \exp\left(\frac{-\Delta t}{\tau_{lifetime}}\right) \times S_{tranfss} + (1 - S_{tranGf}) \times g_{fCal} \quad (9)$$

In this equation, g_{fOld} is the filament conductance of the previous state. Δt is the time difference between the current state and the previous state. $\tau_{lifetime}$ is the disappearance time constant of the filament. g_{fCal} is the calculate conductance of the current state, as shown in (10), where m_f is the conductivity coefficient of the filament, V_{th} is the high threshold voltage

of filament and τ_{st} is the smoothing transition time constant. S_{tranGf} is a smoothing function which depends on the filament conductances of the previous state and the calculate conductance, as shown in (11). $S_{tranfss}$ is a smoothing function which depends on the lower threshold voltage of filament, as shown in (12).

$$g_{fCal} = \frac{m_f \times (1 - C_f)}{z_{act}} \times \frac{1}{1 + \exp\left(-\frac{V - V_{th}}{\tau_{st}}\right)} \quad (10)$$

$$S_{tranGf} = \frac{1}{1 + \exp\left(\frac{g_{fCal} - g_{fOld}}{\tau_{st}}\right)} \quad (11)$$

$$S_{tranfss} = \frac{1}{1 + \exp\left(\frac{V - V_{fss}}{\tau_{st}}\right)} \quad (12)$$

During the SET process, the OTS effect occurs when $V > V_{th}$, the PCM changes from a high resistance state to a low resistance state, and the steady state point in the circuit changes from point B to point A. When calculating the conductance of the current state, even if $V_1 < V_{th}$, since the g_f will not immediately become 0, the circuit's steady state point will remain near point A. The PCM performs the SET process until a phase change occurs, from amorphous state to full crystalline state. At this time, the PCM is in a stable low-impedance state, and the steady-state point of the circuit is maintained at point A (V_1, I_1).

D. PCM MODEL FUNCTIONAL VERIFICATION:DC ANALYSIS AND TRANSIENT ANALYSIS

The circuit diagram and the simulation results of DC analysis are shown in figure 5. The voltage source scans forward, and the initial state of the PCM is a high-impedance state. When the voltage across the PCM is smaller than the threshold voltage ($V < V_{th}$), PCM is in a high-impedance state, so the current is very small. When $V > V_{th}$, a filament appears. The resistance is in a temporary low-resistance state and the current increases. In the DC simulation, each state point is considered to be a steady state, so the PCM undergoes a

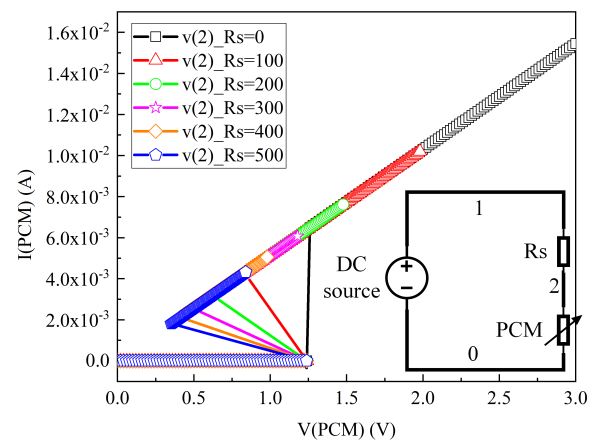


FIGURE 5. The resistance characteristic curve of PCM when the series resistance R_s takes different values in the DC forward scan. Embedded figure is DC analysis circuit schematic.

SET process under a large current and sufficient Joule heat, from an amorphous state to a crystalline state. Because the conductance of the PCM changes during the SET process, when different static resistances R_s are connected in series, the voltage across the PCM will jump in the circuit, showing different degrees of snapback. The larger R_s , the more obvious the snapback phenomenon.

In order to further verify the practicability of the PCM model, a transient analysis simulation is performed. The schematic diagram of the simulation circuit and the simulation results are shown in figure 6. In the transient simulation analysis, the SET process is performed first, and the snapback phenomenon can be seen from the figure 6(c). Figure 6(e) shows the decay of filament conductance over time. After the SET process is performed, the PCM changes from a high resistance state to a low resistance state, as shown in figure 6(f). The following is the RESET process. After the RESET, the resistance value of the PCM returns to the high-resistance state.

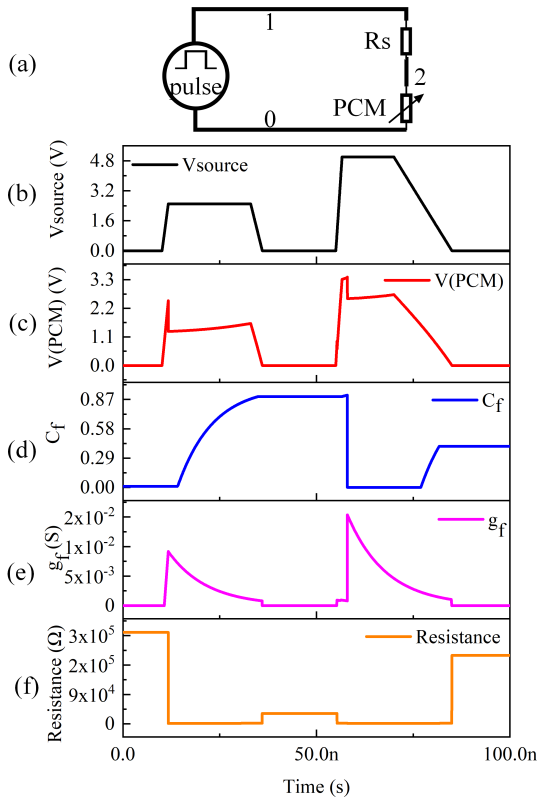


FIGURE 6. (a) Trans analysis simulation circuit schematic. (b) Voltage of the pulse source, first perform SET operation and then RESET operation. (c) Voltage across PCM. (d) Change of crystallization ratio (C_f). (e) Change of filament conductance (g_f). (f) Change of PCM resistance.

III. BUILDING CIRCUITS BASED ON PCM MODELS TO MIMIC BIOLOGICAL STDP

In the synaptic design of neuromorphic circuits, researchers of IBM proposed a 2T1R (2 Transistors and 1 Resistance) structure [17]. Compared with the traditional 1T1R structure, the 2T1R structure can reduce the capacitance from 1 μF

to 1 pF, which can not only reduce the layout area of the capacitor, but also reduce the power consumption during neural network training.

Since the leaky-integrate-and-fire (LIF) neuron circuit model has a simple structure and has basic biological neuron properties [32], it is often used to mimic biological neurons. The model is described as (13), where $I(t)$ is signal input current, $V_m(t)$ is cell membrane potential, and R_m is cell membrane equivalent impedance.

$$I(t) - \frac{V_m(t)}{R_m} = \frac{dV_m(t)}{dt} \quad (13)$$

Multiple neurons are connected by synapses to form a neuromorphic circuit suitable for learning. Figure 7(a) shows a pre-neuron, post-neuron, and synaptic structure diagram between them. This structure implements a LIF neuron circuit model. The synaptic connection of this neural circuit diagram uses a 2T1R structure. The capacitance on the post-synaptic neuron represents the biological cell membrane potential. If the pre-synaptic neuron does not generate LIF word line (LIF WL) pulses, the voltage of this capacitor will be slowly charged and returned to a high level, at which time the neuron is at resting state. When the pre-synaptic neuron generates a LIF WL pulse, the pre-synaptic neuron will generate a STDP word line (STDP WL) pulse at the falling edge of the LIF WL pulse and act on the gate of the transistor M2. At this

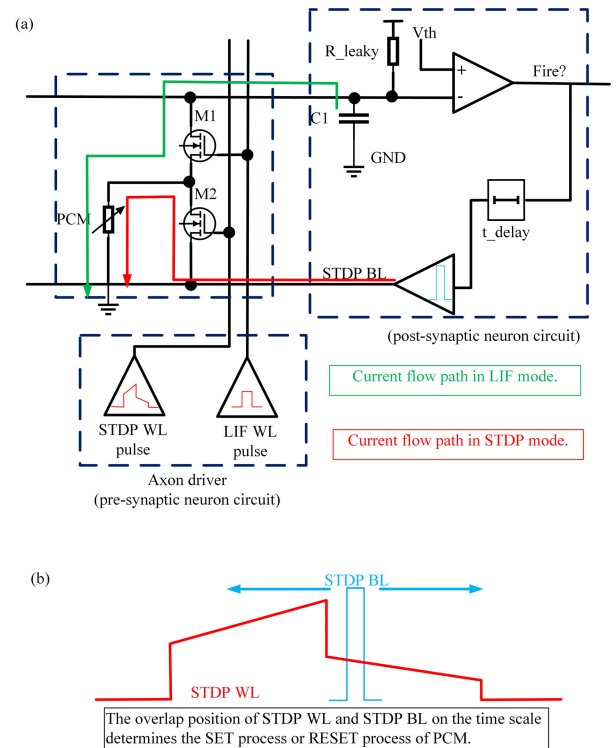


FIGURE 7. (a) Schematic diagram of neuron circuit structure. When in LIF WL mode, the discharge path of the capacitance of the post-synaptic neuron is $C1 \rightarrow M1 \rightarrow \text{PCM} \rightarrow \text{GND}$. When in STDP WL mode, the current path through the PCM synapse is $\text{STDP BL} \rightarrow M2 \rightarrow \text{PCM} \rightarrow \text{GND}$. (b) Overlap schematic of STDP WL pulse and STDP BL pulse. Red: from pre-synaptic neuron, applying on the gate of transistor M2; Blue: from post-synaptic, applying on the drain of transistor M2 [17].

time, the neural circuit is in LIF mode, and the capacitance of the post-synaptic neuron will leak through the PCM. The leakage speed depends on the resistance of the PCM. When the voltage of the capacitor is lower than the threshold V_{th} , the post-neuron will fire. This will propagate LIF WL pulses to the neurons in the next layer, and after a certain time delay, STDP bit line (STDP BL) pulses will be generated to put the neuromorphic circuit into STDP mode. The overlap of the STDP BL pulse and the STDP WL pulse in time determines the resistance value of the transistor M2, which affects the current flowing through the PCM. As shown in figure 7(b), when the left side of STDP WL pulse overlaps with STDP BL pulse, the current flowing through the transistor M2 is large enough to cause a partial RESET process of the PCM. If the right side of STDP WL pulse overlaps with STDP BL pulse, the current flowing through the transistor M2 can only cause the partial SET process of the PCM.

In order to verify the STDP function of this neuromorphic circuit, a 1-neuron-2-synapses neuromorphic circuit was constructed and simulated on the SPICE platform. The simulation results are shown in figure 8. The simulation time is within 1 period ($1\mu s$). The LIF WL pulse of pre-synaptic neuron1 arrives first, and the LIF WL pulse of pre-synaptic neuron2 then appears. Since the LIF WL pulses of pre-synaptic

neuron1 cause post-synaptic neuron fire, it is considered to contribute to post-synaptic neuron fire. At this time, the right side of STDP WL pulse of pre-synaptic neuron1 overlaps with STDP BL pulse, and the partial SET process occurs in synapse1. In contrast, the LIF WL pulse of pre-synaptic neuron2 arrives after the post-synaptic neuron fires, and it is considered that the LIF WL pulse does not contribute to the post-synaptic neuron fire. Therefore, the left side of STDP WL pulse of pre-synaptic neuron2 overlaps with STDP BL pulse, which causes a partial RESET process.

For synapses, the trigger time of the LIF WL pulse of the pre-synaptic neuron is t_{pre} , and the trigger time of the post-synaptic neuron is t_{post} . If $\Delta t = t_{post} - t_{pre} > 0$, the synapse will undergo the SET process. The contact conductance increases, i.e. the synaptic weight W (neuron connection strength) increases, $\Delta W > 0$; if $\Delta t = t_{post} - t_{pre} < 0$, the synapse undergoes a RESET process, and the synaptic conductance decreases, i.e. the synaptic weight W decreases, $\Delta W < 0$. Figure 9 shows the simulation results of the STDP test based on the 2T1R structure. The relationship curve between the PCM synaptic weight changes relatively ($\frac{\Delta W}{W}$) and the relative time ($\Delta t = t_{post} - t_{pre}$) is similar to the shape of experimental statistics of [17]. The synapse has strong spike-time-dependent plasticity, which shows that the neuromorphic circuit constructed in this paper is suitable for neural network applications.

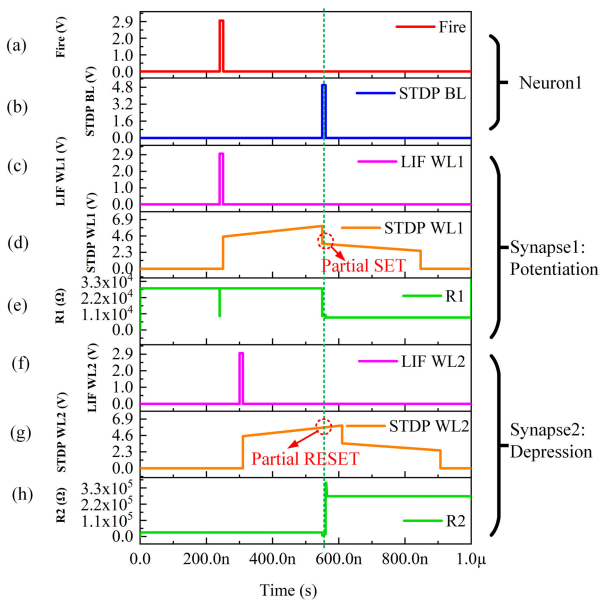


FIGURE 8. Simulation results of STDP test for neuromorphic circuit (simple 1 neuron with 2 synapses). (a) When $V < V_{th}$, the neuron will be activated and output the Fire signal. (b) STDP BL of post-synaptic neuron:when the neuron is activated, STDP BL pulse is generated after a delay. (c) LIF WL of pre-synaptic neuron1. (d) STDP WL of pre-synaptic neuron1: at this time, the right side of STDP WL pulse overlaps with STDP BL pulse, resulting in a partial SET operation of PCM. (e) The resistance value R1 is calculated by (6). After the partial SET operation occurs, the increase in C_f causes the resistance of the PCM to decrease. (f) LIF WL of pre-synaptic neuron2. (g) STDP WL of pre-synaptic neuron2: at this time, the left side of STDP WL pulse overlaps with STDP BL pulse, resulting in a partial RESET operation of PCM. (h) The resistance value R2 is calculated by (6). After a partial RESET operation, both z_{act} and C_f change. C_f is affected by the RESET cooling rate, and z_{act} is affected by the maximum temperature. As a result, the partial RESET process causes the resistance of the PCM to increase.

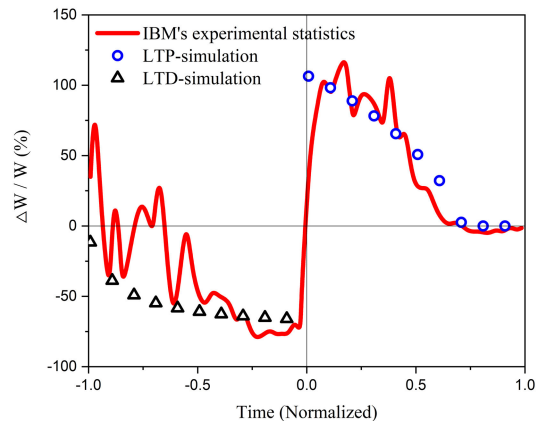


FIGURE 9. STDP comparison of simulation results based on PCM model and experimental statistics of [17]. $\Delta W/W$ is the PCM synaptic weight changes relatively, $Time$ is normalized ($t_{post} - t_{pre}$).

IV. CIRCUIT DESIGN BASED ON PCM MODEL: NEW NEUROMORPHIC CIRCUIT STRUCTURE FOR LATERAL INHIBITION

The circuit implementation of the lateral inhibition mechanism between neurons in the same layer is shown in figure 10. M3 is a NMOS transistor, and M4 is a PMOS transistor. When in LIF mode, the Pctrl wire transmits a high level, and the Nctrl wire transmits a high level. At this time, M3 is turned on and M4 is turned off. The charge on the capacitor C1 will flow to ground through M3->M1->PCM->GND. When there is a neuron fire, the module “Inhibition and Refractory Control”

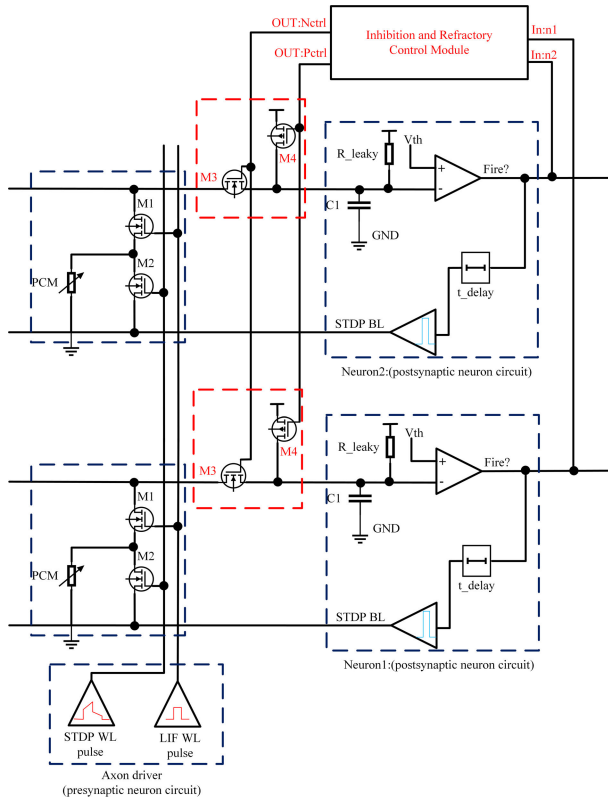


FIGURE 10. Circuit diagram of the neuromorphic circuit for lateral inhibition. M3 is NMOS and M4 is PMOS. The “Inhibition and Refractory Control” macro model is used to control the opening or closing of M3 and M4 to achieve refractory and lateral inhibition of neuron 1 and neuron 2.

will detect the fire signal. At this time, the Pctrl wire transmits a low level and the Nctrl wire transmits a high level. The neuron that just fired entered the refractory period, and will return to LIF mode after the refractory period. Neurons without fire in the same layer are affected by lateral inhibition. The capacitor C1 (cell membrane potential) is charged to a high level and is in a resting state.

V. LEARNING CHARACTERISTICS OF NEUROMORPHIC CIRCUIT FOR UNSUPERVISED LEARNING

In order to verify the effectiveness of the PCM model used in neuromorphic circuit, unsupervised learning is performed in this neuromorphic circuit. The unsupervised learning process is shown in figure 11(a). In this spiking neural network, the input is a 7×7 picture. Figure 11(b) shows the schematic of the spiking neural network circuit architecture used in this paper. In this spiking neural network, the input layer contains 49 input neurons (pre-synaptic neuron), and the output layer contains 2 neurons (post-synaptic neuron). Each post-synaptic neuron is connected to 49 pre-synaptic neurons through 49 synapses, and the strength of the synaptic connection is reflected in the conductance of PCM. The closer the neurons are connected, the greater the conductance of PCM.

The encoding method used in this paper is time-to-first-spike encoding [33]. This encoding method is described by (14). In this equation, $Time_{LIF\ WL}$ represents the time

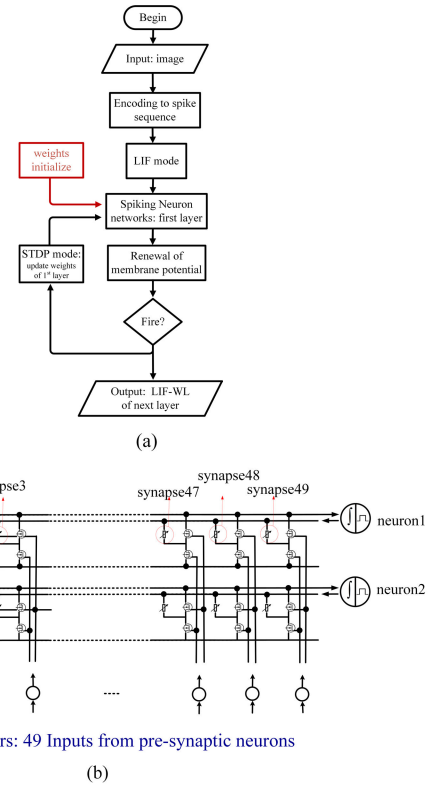


FIGURE 11. (a) Unsupervised learning flowchart for a spiking neural network (b) A schematic diagram of the circuit architecture of a spiking neural network, where 49 input neurons correspond to 49 pixels of the image.

of pulse emission, $Value_{pixel}$ represents the numerical value of the corresponding pixel, and $Width_{STDP\ WL}$ represents the waveform width of STDP WL pulse.

$$Time_{LIF\ WL} = \left(1 - \frac{Value_{pixel}}{255}\right) \times \frac{Width_{STDP\ WL}}{2} \quad (14)$$

Time-to-first-spike encoding can be further described as figure 12. The left side of the figure is a schematic diagram of a 7×7 pixel matrix, and each grid represents a pixel. Each picture with a size of 7×7 can be regarded as such a two-dimensional matrix, and the values in the matrix represent the numerical value of the corresponding pixels. The 2-dimensional matrix is converted into a 1-dimensional vector in the manner shown in the figure, and a vector of length 49 is obtained, which corresponds to 49 input neurons. The higher numerical value of the pixel, the earlier the neuron emits pulses; the lower numerical value of the pixel, the later the neuron emits pulses.

This paper uses the neuromorphic circuit structure shown in figure 7 to build a 49×2 neural network, which is shown in figure 11(b). Then assigning initial values to the weights of the neural network randomly and uniformly. The neural network contains 2 post neurons, each of which has 49 synapses. In the “MNIST” dataset, the picture “0” is randomly selected, and the picture is compressed into a picture with a size of 7×7 using a bilinear interpolation

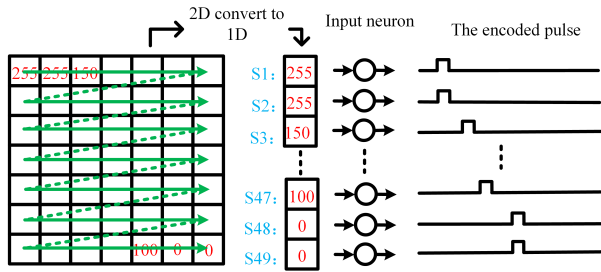


FIGURE 12. The coding form of the input data of spiking neural networks, using time-to-first-spike encoding method.

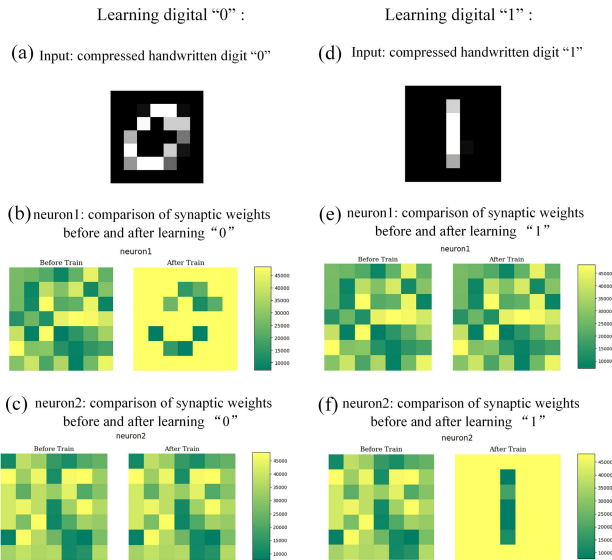


FIGURE 13. Example of spiking neural network used for unsupervised learning. (a) The input picture of the neural network when learning “0”, using linear interpolation to compress the MNIST handwritten digit “0” from a 28×28 pixels matrix to 7×7 pixels matrix, where each pixel corresponds to an input neuron. (b) When learning picture “0”, the synaptic weight of neuron 1 is compared before and after training. (c) When learning picture “0”, neuron 2 has no synaptic weight changes due to the lateral inhibition. (d) The input picture of the neural network when learning “1”, where each pixel corresponds to an input neuron. (e) When learning picture “1”, neuron 1 has no synaptic weight changes due to the lateral inhibition. (f) When learning picture “1”, the synaptic weights of neuron 2 are compared before and after learning.

method, as shown in figure 13(a). The whiter the color of a pixel, the higher numerical value of the pixel, and the earlier the time corresponding to the pulse emission. In contrast, the darker the color of a pixel, the later the pulse is emitted.

The picture shown in figure 13 (a) is used as the input of the neural network to perform unsupervised learning. The change of synaptic conductance of neuron 1 before and after learning is shown in figure 13 (b). After learning, neuron 1 can automatically learn the information of the input picture without the teacher’s supervision. The synaptic conductance of input neurons with high pixel numerical value increases, while the synaptic conductance of input neurons with low pixel numerical value decreases. If the next input picture is similar to the previous input picture, the neuron will be more sensitive and fire earlier. Similar to biological neuron, the connections between neurons will change under the stimuli of the external environment, making the organism more sensitive to external

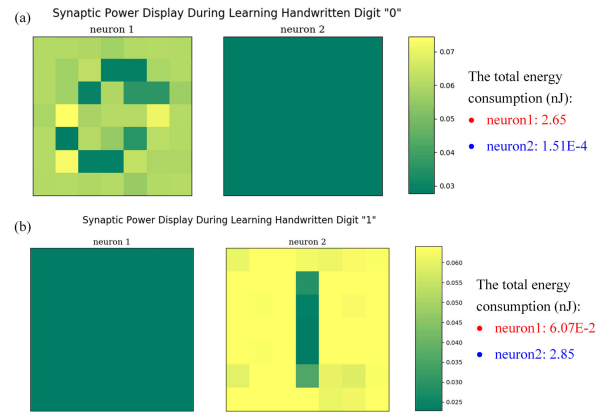


FIGURE 14. (a) Schematic diagram of the energy consumption of each synapse (49 synapses) connected by a post-synaptic neuron when learning the handwritten number “0”. The unsupervised learning time is 1 period. (b) Schematic diagram of the energy consumption of each synapse (49 synapses) connected by a post-synaptic neuron when learning the handwritten number “1”.

stimuli. When learning picture “0”, neuron 1 fires first. Due to the lateral inhibition, neuron 1 does not fire, so the synaptic weight of neuron 1 does not change after learning, as shown in figure 13 (c).

As shown in figure 13(d), when the input picture “1” is given to the neural network, unsupervised learning of “1” in neural network is performed. The synapse of neuron 1 before and after learning is shown in figure 13(e). Due to the effect of lateral inhibition, there is no change in the synaptic weight of neuron 1. When learning picture “1”, the synaptic conductance of neuron 2 is shown in figure 13(f), which shows that neuron 2 can learn the information of number “1”.

In unsupervised learning of handwritten digit “0”, the power consumption of neuron 1 and neuron 2 is shown in figure 14(a). Since neuron 1 fires first, neuron 2 does not fire because of lateral inhibition. Therefore, neuron 1 consumes more power than neuron 2. In the unsupervised learning of handwritten digit “1”, since neuron 2 fires first, neuron 1 does not fire due to the effect of lateral inhibition, so the power consumption of neuron 2 is greater.

VI. CONCLUSION

This paper presents a time tracking PCM model with analog characteristics, which is suitable for neuromorphic circuits. The state variables C_f and z_{act} are used to track the crystal fraction and active area radius, respectively. The improved equation of z_{act} can reflect not only spatial information but also temporal information, which is suitable for continuous changes in synaptic conductance. Using the conductance of the filament’s previous state, the convergence of the voltage-driven model is improved. This paper validates the DC analysis and transient analysis of the model, and validates STDP on the 2T1R structure of researchers from IBM, further illustrating the availability of the model. In order to improve the bionics of neuromorphic circuits and distinguish fuzzy memories, a method of implementing lateral inhibition mechanism in neuromorphic circuits has been proposed. It is the

first time in literature to discuss the learning characteristics of a spiking neural network based on PCM at the level of circuit simulation. Unsupervised learning of the handwritten digit “0” and the handwritten digit “1” verifies the integrity of models in a large-scale-integration circuits.

The PCM model proposed in this work provides a first demonstration on the methodology to model the time evolution of a PCM device, which is of great significance for the simulation of neuromorphic circuits. There are still a lot of possible refinements in term of the device physics and efficiency enhancement. For example, the time evolution of the filament formation and the crystallization process of active region.

REFERENCES

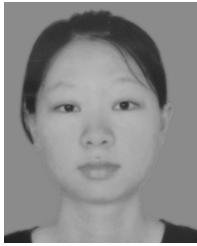
- [1] W. A. Wulf and S. A. McKee, “Hitting the memory wall: Implications of the obvious,” *ACM SIGARCH Comput. Archit. News*, vol. 23, no. 1, pp. 20–24, Mar. 1995.
- [2] D. Ielmini and H.-S.-P. Wong, “In-memory computing with resistive switching devices,” *Nature Electron.*, vol. 1, no. 6, pp. 333–343, Jun. 2018.
- [3] M. A. Sivilotti, M. R. Emerling, and C. A. Mead, “VLSI architectures for implementation of neural networks,” *Neural Netw. Comput.*, vol. 151, no. 1, pp. 408–413, 1986.
- [4] L. W. Nagel and D. Pederson, “Spice (simulation program with integrated circuit emphasis),” Dept. EECS, Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB/ERL M382, Apr. 1973.
- [5] R. W. Dutton, “Modeling and simulation for VLSI,” in *IEDM Tech. Dig.*, 1986, pp. 2–7.
- [6] K. M. Cham, S.-Y. Oh, J. L. Moll, K. Lee, and P. Vandevoorde, *Computer-Aided Design and VLSI Device Development*, vol. 53. New York, NY, USA: Springer, 2012.
- [7] A. S. Miller, B. H. Blott, and T. K. Hames, “Review of neural network applications in medical imaging and signal processing,” *Med. Biol. Eng. Comput.*, vol. 30, no. 5, pp. 449–464, Sep. 1992.
- [8] Y. Cao, Y. Cao, S. Wen, T. Huang, and Z. Zeng, “Passivity analysis of delayed reaction-diffusion memristor-based neural networks,” *Neural Netw.*, vol. 109, pp. 159–167, Jan. 2019.
- [9] S. Wang, Y. Cao, T. Huang, Y. Chen, P. Li, and S. Wen, “Sliding mode control of neural networks via continuous or periodic sampling event-triggering algorithm,” *Neural Netw.*, vol. 121, pp. 140–147, Jan. 2020.
- [10] S. Ghosh-Dastidar and H. Adeli, “Spiking neural networks,” *Int. J. Neural Syst.*, vol. 19, no. 4, pp. 295–308, Aug. 2009.
- [11] S. R. Nandakumar, I. Boybat, M. Le Gallo, A. Sebastian, B. Rajendran, and E. Eleftheriou, “Supervised learning in spiking neural networks with MLC PCM synapses,” in *Proc. 75th Annu. Device Res. Conf. (DRC)*, Jun. 2017, pp. 1–2.
- [12] I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, and E. Eleftheriou, “Neuromorphic computing with multi-memristive synapses,” *Nature Commun.*, vol. 9, no. 1, p. 2514, Dec. 2018.
- [13] S. R. Nandakumar, I. Boybat, M. Le Gallo, E. Eleftheriou, A. Sebastian, and B. Rajendran, “Supervised learning in spiking neural networks with phase-change memory synapses,” 2019, *arXiv:1905.11929*. [Online]. Available: <http://arxiv.org/abs/1905.11929>
- [14] S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, “A phase-change memory model for neuromorphic computing,” *J. Appl. Phys.*, vol. 124, no. 15, Oct. 2018, Art. no. 152135.
- [15] Y. Cao, S. Wang, Z. Guo, T. Huang, and S. Wen, “Synchronization of memristive neural networks with leakage delay and parameters mismatch via event-triggered control,” *Neural Netw.*, vol. 119, pp. 178–189, Nov. 2019.
- [16] S. Wang, Y. Cao, T. Huang, Y. Chen, and S. Wen, “Event-triggered distributed control for synchronization of multiple memristive neural networks under cyber-physical attacks,” *Inf. Sci.*, vol. 518, pp. 361–375, May 2020.
- [17] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J.-P. Han, C. Miller, K. Hosokawa, and C. Lam, “NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning,” in *IEDM Tech. Dig.*, Dec. 2015, pp. 1–17.
- [18] Y. Wei and X. Lin, “An improved SPICE model of phase-change memory (PCM) for peripheral circuits simulation and design,” in *Proc. Int. Symp. Integr. Circuits (ISIC)*, Dec. 2014, pp. 132–135.
- [19] R. A. Cobley, C. D. Wright, and J. A. V. Diodado, “A model for multilevel phase-change memories incorporating resistance drift effects,” *IEEE J. Electron Devices Soc.*, vol. 3, no. 1, pp. 15–23, Jan. 2015.
- [20] H. Hu, D. Liu, X. Chen, D. Dong, X. Cui, M. Liu, X. Lin, L. Zhang, and M. Chan, “A compact phase change memory model with dynamic state variables,” *IEEE Trans. Electron Devices*, vol. 67, no. 1, pp. 133–139, Jan. 2020.
- [21] C. Pigot, M. Bocquet, F. Gilibert, M. Reyboz, O. Cueto, V. D. Marca, P. Zuliani, and J.-M. Portal, “Comprehensive phase-change memory compact model for circuit simulation,” *IEEE Trans. Electron Devices*, vol. 65, no. 10, pp. 4282–4289, Oct. 2018.
- [22] R. A. Cobley and C. D. Wright, “Parameterized SPICE model for a phase-change RAM device,” *IEEE Trans. Electron Devices*, vol. 53, no. 1, pp. 112–118, Jan. 2006.
- [23] C. Blakemore, R. H. Carpenter, and M. A. Georgeson, “Lateral inhibition between orientation detectors in the human visual system,” *Nature*, vol. 228, no. 5266, p. 37, 1970.
- [24] C. Espinoza, S. J. Guzman, X. Zhang, and P. Jonas, “Parvalbumin⁺ interneurons obey unique connectivity rules and establish a powerful lateral-inhibition microcircuit in dentate gyrus,” *Nature Commun.*, vol. 9, no. 1, p. 4605, Dec. 2018.
- [25] I. V. Karpov, S. Kostylev, and C. C. Kuo, “Ring heater for a phase change memory device,” U.S. Patent 7 863 596, Jan. 4, 2011.
- [26] R. C. Malenka, “Synaptic plasticity in the hippocampus: LTP and LTD,” *Cell*, vol. 78, no. 4, pp. 535–538, Aug. 1994.
- [27] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, and B. DeSalvo, “Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction,” in *IEDM Tech. Dig.*, Dec. 2011, p. 4.
- [28] K. C. Kwong, “Phase change memory model development based on cell geometry,” Dept. Electron. Commun. Eng., Hong Kong Univ. Sci. Technol., Hong Kong, 2012.
- [29] D. Liu, L. Zhang, X. Lin, and M. Chan, “A smooth and continuous phase change memory SPICE model for improved convergence,” in *Proc. IEEE 2nd Electron Devices Technol. Manuf. Conf. (EDTM)*, Mar. 2018, pp. 86–88.
- [30] M. Ito, “Long-term depression,” *Annu. Rev. Neurosci.*, vol. 12, no. 1, pp. 85–102, 1989.
- [31] Y. Chen, K. C. Kwong, X. Lin, Z. Song, and M. Chan, “3-D resistance model for phase-change memory cell,” *IEEE Trans. Electron Devices*, vol. 61, no. 12, pp. 4098–4104, Dec. 2014.
- [32] L. F. Abbott, “Lapicque’s introduction of the integrate-and-fire model neuron (1907),” *Brain Res. Bull.*, vol. 50, nos. 5–6, pp. 303–304, Nov. 1999.
- [33] C. Shoushun and A. Bermak, “A low power CMOS imager based on time-to-first-spike encoding and fair AER,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2005, pp. 5306–5309.



XUHUI CHEN is currently pursuing the master’s degree with Peking University, China. His research interests include neuromorphic computing and phase change memory modeling.



HUIFANG HU is currently pursuing the master’s degree with Peking University, China. Her research interest includes phase change memory modeling.



XIAOQING HUANG is currently pursuing the master's degree with Peking University, China. Her research interests include device physics and modeling.



XINNAN LIN (Member, IEEE) is currently an Associate Professor with the Shenzhen Key Laboratory of Advanced Electron Device and Integration, ECE, Peking University Shenzhen Graduate School, China. His research interests include device physics and modeling, and power electronics.



WEIRAN CAI is currently pursuing the bachelor's degree with Shenzhen University, China. His research interest includes neuromorphic computing.



LINING ZHANG (Senior Member, IEEE) is currently an Assistant Professor with the School of Electronic and Computer Engineering, Peking University. His research interests include device physics and modeling, electronic design automation, and neuromorphic computing.

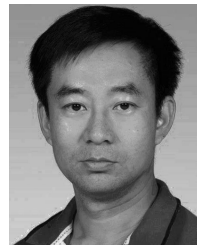


MING LIU (Fellow, IEEE) is currently a Professor with the Key Laboratory of Microelectronics Devices and Integration Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China, and also with the University of Chinese Academy of Sciences, Beijing.



CHUNG LAM received the B.Sc. degree from the Polytechnic University of New York, in 1978, and the M.Sc. and Ph.D. degrees from the Rensselaer Polytechnic Institute, in 1987 and 1988, respectively, all in electrical engineering. Since 1978, he joined IBM, where he has taken responsibilities in various disciplines of semiconductor research, development, and manufacturing, including circuit and device designs as well as process integrations for memory and logic applications in the IBM

Microelectronic Division. In 2003, he transferred to the IBM Research Division, T. J. Watson Research Center. He was named as an IBM Distinguished Engineers, in 2007. Since leaving IBM, he has served as the Chief Technology Officer of Jiangsu Advanced Memory Technology Corporation Ltd.



MANSUN CHAN (Fellow, IEEE) is currently a Professor with the HKUST Shenzhen Research Institute, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, China.

...