Applied Physics Letters ARTICLE scitation.org/journal/apl

Winner-takes-all mechanism realized by memristive neural network

Cite as: Appl. Phys. Lett. **115**, 243701 (2019); doi: 10.1063/1.5120973
Submitted: 22 July 2019 · Accepted: 29 November 2019 · Published Online: 10 December 2019

View Online Export Citation CrossMark

J. J. Wang,[1] Q. Yu,[1] S. G. Hu,[1] Yanchen Liu,[1] Rui Guo,[1] T. P. Chen,[2] Y. Yin,[3] and Y. Liu[1,a)]

AFFILIATIONS

[1]State Key Laboratory of Electronic Thin Films and Integrated Devices, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[3]Graduate School of Engineering, Gunma University, 1-5-1Tenjin, Kiryu, Gunma 376-8515, Japan

[a)]yliu1975@uestc.edu.cn

ABSTRACT

Winner-takes-all (WTA), an important mechanism in neural networks of recurrently connected neurons, is a critical element of many models of cortical processing. However, few WTA neural networks have been realized physically, especially by memristor networks. In this work, we have designed and implemented a neural network with memristor-based synapses to realize the WTA in a neural system. Neuronal self-excitatory, excitatory, and inhibition by other neurons have been demonstrated. Competitions between two neurons, among three neurons, and between two groups of neurons are realized based on the memristive neural network. The winner neuron or winner group can suppress the other neuron(s) or other group(s) of neurons and dominate the neuronal firing. This work paves the way for further realization of complex models of cortical processing with memristive neural networks.

Neuromorphic engineering is an interdisciplinary discipline that takes inspiration from physics, biology, computer science, and mathematics to build an artificial neural network. The physical architecture of an artificial neural system is often analogous to that of biological nervous systems. Artificial neural networks (ANNs) implemented with traditional complementary metal-oxide-semiconductor (CMOS) integrated circuits (ICs), which provide a certain level of intelligence, have been reported.[1–4] However, constructing CMOS circuits as the synapses and neurons is at the expensive costs of large chip area and power consumption.[5] A critical issue in constructing a large scale artificial neural network is to look for a suitable device structure that can be used to build neurons and synapses. CMOS circuits have been employed to mimic the synaptic dynamics, but three terminal devices bear limited resemblance to the biocounterparts. A memristor, which was first reported in 2008,[6] was used for training and operation of an integrated neuromorphic network.[7] By using a memristor, the complexity and power consumption of artificial neural networks can be remarkably decreased; at the same time, it is a two-terminal device, resembling the membrane structures in neurons and synapses of biological neural systems.[8–10]

There have been many research works on realization of artificial neurons and synapses for artificial neural networks. Here are some examples. Teresa *et al.* exploited AER (Address Event Representation) technology and crossbar synapses based on memristors to realize a vision sensoring and processing system.[11] Tuma *et al.*[12] proposed an artificial neuron that used a phase-change memristor (PCM) to realize an integrate-and-fire functionality with stochastic dynamics. Cruz-Albrecht *et al.*[13] presented ultralow power neurons and synapses by using device subthreshold regime properties. Hu *et al.*[14] proposed a Hopfield network using HfO_2 memristors and peripheral devices to realize the associative memory that is capable of retrieving a piece of data upon presentation of partial information from that piece of data. In our previous works, we reported the memristive BP (Back-propagation) neural network to predict house price in American towns[15] and memristive neurons to realize convolutional neural networks (CNNs).[16]

The winner-takes-all (WTA) mechanism in neural networks of recurrently connected neurons is a critical element of many models of cortical processing. Oster *et al.*[17] utilized a simplified Markov model of the spiking network to examine analytically the ability of a spike-based WTA network to discriminate the statistics of inputs ranging from stationary regular to nonstationary Poisson events. Yu *et al.*[18] realized emergent inference by the hidden Markov model in Spiking Neural Networks (SNNs) through WTA. However, few of the researchers have physically built a network to realize WTA in SNN;[19] in particular, few research has been done on the realization of WTA with memristive neural networks.

In this work, we have constructed a spiking neural network using HfO_2 memristors to realize WTA functionality. Several WTA competition modes have been examined experimentally. The neurons in the network can inhibit each other, i.e., when one neuron wins competition, the other neurons are suppressed. In group competition, neurons excite each other in the same group, but they inhibit the neurons in other groups. Eventually, one group wins, while the other groups are suppressed. This study provides a method to realize the WTA network by using memristor-based synapses, which can be applied in more complex models of cortical processing. Figure S1 shows the photograph of the overall system.

The detailed memristor device structure and performance used in this work were reported in our previous work.[14] In this work, the Leaky-integrate-fire (LIF) model[20] is adopted for the artificial neurons. The equation of the LIF model is expressed as follows:

$$
\begin{aligned}
C_{mem}\frac{dv(t)}{dt} &= -\frac{v(t)}{R} + i_{in} \quad t \in \left\{ t \neq t_f + t_p | t_f \in t^{(f)} \right\} \\
v(t_f + t_p) &= 0 \quad t_f \in t^{(f)} \\
t^{(f)} &= \{ t | v(t) = v_{th}, \lim_{\Delta t \to 0^+} v(t + \Delta t) > v_{th} \} \\
v_{spike}(t) &= \begin{cases} v_h & \text{when } v(t) \geq v_{th} \\ v_l & \text{otherwise} \end{cases},
\end{aligned}
\tag{1}
$$

where t is the time; t_f and t_p are the time when the neuron fires a spike and pulse width of the neuron spike, respectively; $v(t)$ is the membrane potential; $v_{spike}(t)$ is the neuron spike voltage; C_{mem} is the capacitance of the membrane capacitor; R is the resistance of the leaky resistor; v_{th} is the threshold voltage; i_{in} is the input current; and v_h and v_l are the high level and low level of the neuron spike, respectively. In the LIF model, the differential equation in (1) describes the relationship between membrane voltage and time during the integration process and the neuron firing process. When $v(t)$ approaches v_{th}, it costs t_p(time) setting $v(t)$ back to zero. $t^{(f)}$ is a set of time of the positive edge of spikes. v_{spike} only depends on membrane voltage $v(t)$. When $v(t) > v_{th}$, v_{spike} outputs v_h; otherwise, v_{spike} outputs v_l.

Figure 1(a) shows the circuit schematic of a neuron based on the LIF model. When the input current is more than leaky current, its membrane potential increases. When the neuron's potential is above the threshold, the neuron fires a spike. At that time, the neuron's potential returns to its resting potential. The characteristics of the neuron under current input are shown in Fig. 1(b). With the increase in the input current, the firing frequency of neurons increases. Figure 1(c) shows the time domain neuronal membrane voltage. When the membrane voltage is above the threshold voltage, the neuron fires a spike.

A memristor-based synapse is designed in this work as shown in Fig. 2(a). The synaptic weight is proportional to the conductance and is stored in the memristor. In our model, the presynaptic current is proportional to the weight and the integration of the neuron spike. In the synapse, a linear transconductance amplifier, containing an operational amplifier and an MOSFET, converts a spike voltage to a spike current, and then, a capacitor integrates the current. The conductance of the memristor determines the transconductance of the amplifier,

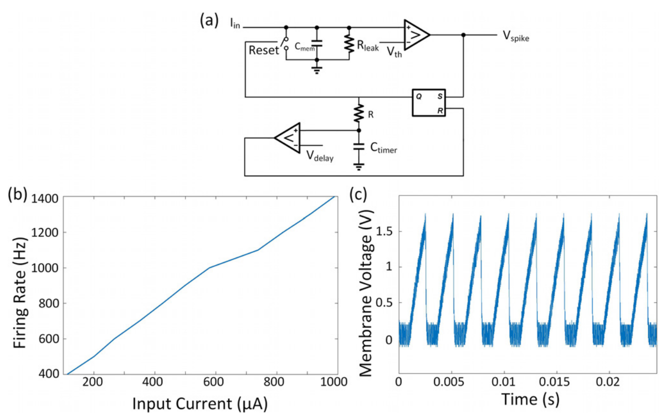

FIG. 1. (a) Circuit schematic of the electronic neuron based on the LIF model, (b) neuron firing rate as a function of input current, and (c) neuronal membrane voltage as a function of time.

and thus, the memristor reflects the synaptic weight. The output of neurons is connected to postsynapses with synaptic weight w_{ij}, and the output of synapses is fed back to the input of neurons by w_{ji}. The switches are controlled by external signals to obtain positive or negative synaptic weights. The ANN consists of a set of interconnected artificial neurons and synapses. Figure 2(a) shows the circuit schematic of a synapse based on the memristor. Figure 2(b) presents the input spikes from the preneuron. Figure 2(c) shows the voltage on the synaptic capacitor, indicating the membrane potential change. Figure 2(d)

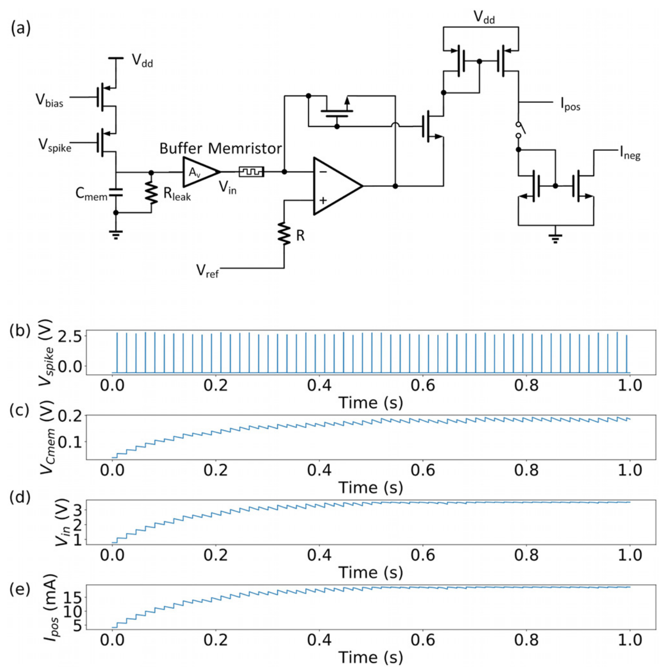

FIG. 2. (a) Circuit schematic of a synapse based on the memristor, (b) the input spikes from the preneuron, (c) the voltage waveforms of the synaptic capacitor, (d) voltage waveforms on the left side of the memristor, indicating the state of the memristor, and (e) the output current of the synapse when the synapse excites.

shows the voltage waveforms on the left side of the memristor, indicating the state of the memristor. Figure 2(e) shows the output current of the synapse when the synapse excites.

The competition between neuron N_1 and neuron N_2 is illustrated in Fig. 3. The configuration is shown in Fig. 3(a). S_{ij} means the synapse, while w_{ij} means the magnitude of weight for the synapse. It should be noted that the synapse from Neuron N_i to Neuron N_j is named as S_{ij} and the weight of S_{ij} is named as w_{ij}. As shown in Fig. 3(a), Neuron N_1 is excited itself by Synapse S_{11} whose weight is w_{11}, while it inhibits neuron N_2 by Synapse S_{12} whose weight is w_{12} at the same time. The operation of neuron N_2 is similar to that of N_1. To ensure the fairness of competition, w_{11} was programmed equal to w_{22}. w_{12} was programmed to a value smaller than w_{21}. Then, we can observe N_2 wins as shown in Fig. 3. To achieve its target weight, the conductance of a relevant memristor is adjusted by applying step-like voltage pulses with appropriate voltage magnitude, pulse width, and pulse number to the memristor. As shown in Fig. 3(b), the firing rate and magnitude of N_1 start to decrease at $\sim$0.02s and keep decreasing gradually. After $\sim$0.05s, N_1 stops firing spikes, indicating that it loses in the competition. In contrast, as can be observed in Fig. 3(c), the firing rate and magnitude of N_2 remain unchanged, i.e., N_2 remains in its initial state. It is obvious that N_2 wins the competition.

Multineuron competition has also been examined. Figure 4 shows the competition of three neurons. The configuration is shown in Fig. 4(a). In this configuration, a WTA neural network is constructed with three neurons and nine synapses. Each neuron has four inputs, including the fixed initial synaptic current, excitatory current from the neuron itself, and inhibitory current from other neurons.

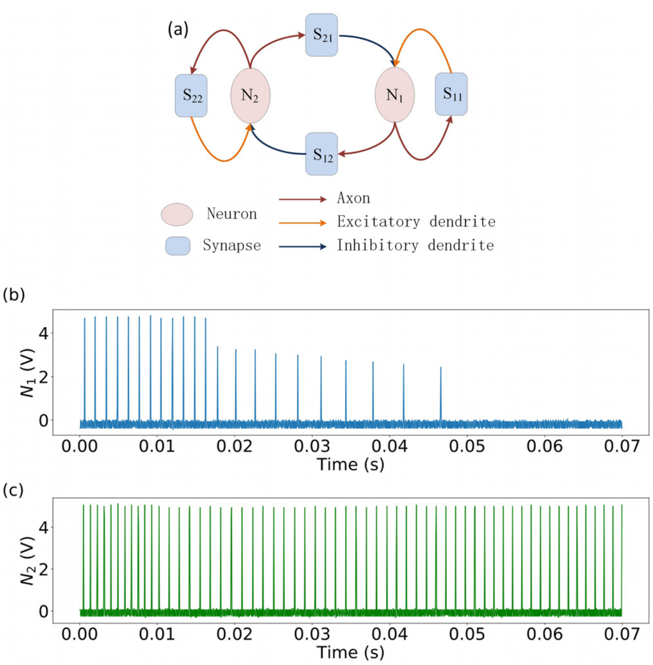

FIG. 3. (a) Schematic illustration of competition between two neurons; (b) firing characteristics of N_1; and (c) firing characteristics of N_2. Detailed circuits are presented in Fig. S2.

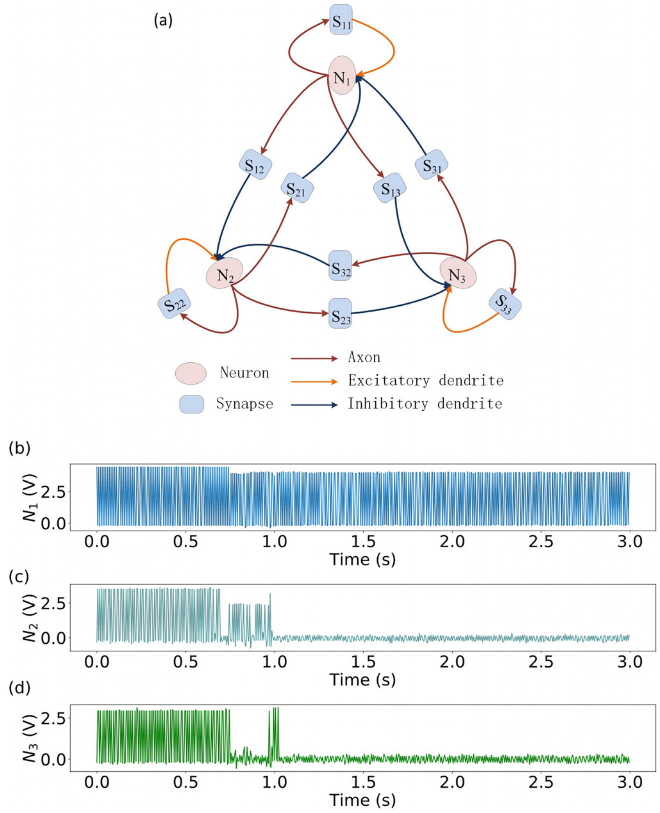

FIG. 4. (a) Schematic illustration of competition among three neurons (N_1, N_2, and N_3) and (b)–(d) waveforms of N_1, N_2, and N_3. Detailed circuits are presented in Fig. S3. Weights for w_{21}, w_{31}, and w_{32} are very close to zero.

The self-excitatory weights w_{ii} are set as close as possible to ensure fairness in the competition. The competition result is shown in Figs. 4(b)–4(d). As can be observed, the firing rate of N_2 and N_3 decreases gradually, while N_1 keeps firing. This means that eventually N_1 wins the competition.

Figure 5 shows the competition between two groups of neurons. Group 1 (G_1) consists of neurons N_1 and N_2, while group 2 (G_2) consists of neurons N_3 and N_4. Neurons in the same group excite each other. On the other hand, Neuron N_2 inhibits G_2 via N_3 and N_4 inhibits G_1 via N_1. In the initial stage with the same applied initial current, the two groups maintain their firing rates and magnitudes. After removing the initial currents, the two groups begin to compete. N_1 competes with G_2 directly and stops firing earlier, while N_2 is inhibited indirectly and stops firing after a longer time, as shown in Figs. 5(b) and 5(c). During the competition, the firing rate of neurons in Group G_1 decreases gradually, while the firing rate of neurons in G_2 increases gradually. Eventually, the neurons in G_1 stop firing, while the neurons in G_2 keep firing.

The videos for two-neuron competition, three-neuron competition, and group competition can be found in the supplementary material. By setting the weights, we can select the winner/winner group. If equal weights are set for all neurons/neuron groups, the winner may be randomly determined by the noise during the competition

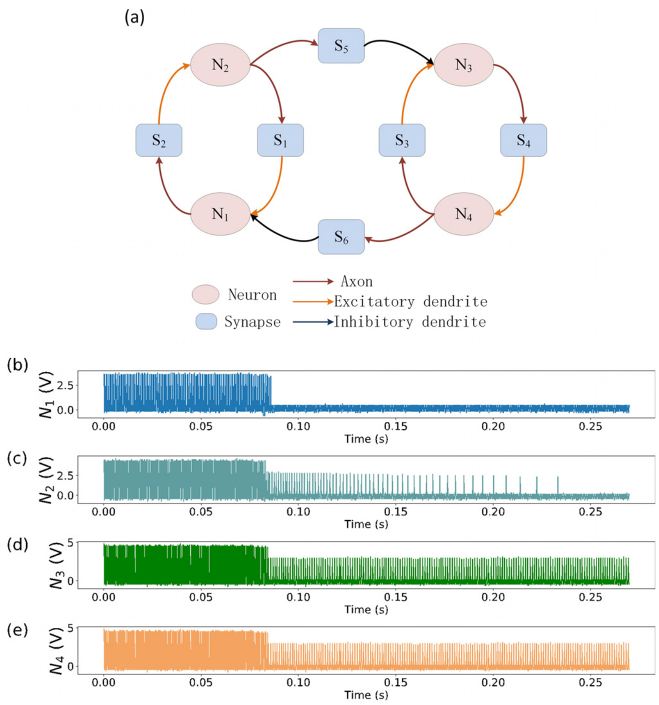

FIG. 5. Schematic illustration of competition between two groups of neurons (a) and firing characteristics for N_1(b), N_2(c), N_3(d), and N_4(e). Detailed circuits are presented in Fig. S4.

process. A larger and extendable network can be designed based on this work. As shown in Fig. S5(a), a larger and extendable network can be realized based on our design. Figure S5(b) shows an example of WTA in a ten-neuron network based on the extendable memristive network. Neuron N_8 wins the competition among the 10 neurons. As shown in Fig. S6(c), the WTA network has a reasonably high tolerance against the variation of resistance of the memristors. As can be observed from the figure, even when the memresistance changes by up to ∼70%, the WTA outcome still does not change. This means that the network is quite robust.

In this paper, a memristor-based WTA neural network is reported. Two-neuron competition, three-neuron competition, and group competition have been demonstrated with the proposed memristive neural network. The study paves the way for emulating more functions of biological neural systems using memristive artificial neural networks.

See the supplementary material for photographs of the memristor-based WTA system, circuit schematics for various types of WTA competitions, illustration of an extendable WTA network, and the firing rate of neuron as a function of variation of resistance. Videos for various competitions are also provided.

This work was supported by NSFC under Project Nos. 61774028 and 61771097 and the Fundamental Research Funds for the Central Universities under Project No. ZYGX2016Z007.

REFERENCES

[1]G. Liu, D. Kong, S. Hu, Q. Yu, Z. Liu, T. P. Chen, Y. Yin, S. Hosaka, and Y. Liu, “Smart electronic skin having gesture recognition function by LSTM neural network,” Appl. Phys. Lett. **113**, 084102 (2018).
[2]D. Kong, S. Hu, Y. Wu, J. Wang, C. Xiong, Q. Yu, Z. Shi, Z. Liu, T. Chen, Y. Yin *et al.*, “Realization of a power-efficient transmitter based on integrated artificial neural network,” IEEE Access **6**, 68773–68781 (2018).
[3]D. Kong, S. Hu, J. Wang, Z. Liu, T. Chen, Q. Yu, and Y. Liu, “Study of recall time of associative memory in a memristive hopfield neural network,” IEEE Access **7**, 58876–58882 (2019).
[4]K. M. Cherry and L. Qian, “Scaling up molecular pattern recognition with dna-based winner-take-all neural networks,” Nature **559**, 370 (2018).
[5]M. Verleysen, B. Sirletti, A. Vandemeulebroecke, and P. G. Jespers, “A high-storage capacity content-addressable memory and its learning algorithm,” IEEE Trans. Circuits Syst. **36**, 762–766 (1989).
[6]D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, “The missing memristor found,” Nature **453**, 80 (2008).
[7]M. Prezioso, F. Merrikh-Bayat, B. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, “Training and operation of an integrated neuromorphic network based on metal-oxide memristors,” Nature **521**, 61 (2015).
[8]I. E. Ebong and P. Mazumder, “Cmos and memristor-based neural network design for position detection,” Proc. IEEE **100**, 2050–2060 (2012).
[9]A. Thomas, “Memristor-based neural networks,” J. Phys. D: Appl. Phys. **46**, 093001 (2013).
[10]P. Wijesinghe, A. Ankit, A. Sengupta, and K. Roy, “An all-memristor deep spiking neural computing system: A step toward realizing the low-power stochastic brain,” IEEE Trans. Emerging Top. Comput. Intell. **2**, 345–358 (2018).
[11]R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco, “A neuromorphic cortical-layer microchip for spike-based event processing vision systems,” IEEE Trans. Circuits Syst. I **53**, 2548–2566 (2006).
[12]T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, “Stochastic phase-change neurons,” Nat. Nanotechnol. **11**, 693 (2016).
[13]J. M. Cruz-Albrecht, T. Derosier, and N. Srinivasa, “A scalable neural chip with synaptic electronics using cmos integrated memristors,” Nanotechnology **24**, 384011 (2013).
[14]S. G. Hu, Y. Liu, Z. Liu, T. P. Chen, J. J. Wang, Q. Yu, L. J. Deng, Y. Yin, and S. Hosaka, “Associative memory realized by a reconfigurable memristive hopfield neural network,” Nat. Commun. **6**, 1–5 (2015).
[15]J. Wang, S. Hu, X. Zhan, Q. Luo, Q. Yu, Z. Liu, T. P. Chen, Y. Yin, S. Hosaka, and Y. Liu, “Predicting house price with a memristor-based artificial neural network,” IEEE Access **6**, 16523–16528 (2018).
[16]J. Wang, S. Hu, X. Zhan, Q. Yu, Z. Liu, T. P. Chen, Y. Yin, S. Hosaka, and Y. Liu, “Handwritten-digit recognition by hybrid convolutional neural network based on HfO 2 memristive spiking-neuron,” Sci. Rep. **8**, 12546 (2018).
[17]M. Oster, R. Douglas, and S.-C. Liu, “Computation with spikes in a winner-take-all network,” Neural Comput. **21**, 2437–2465 (2009).
[18]Z. Yu, S. Guo, F. Deng, Q. Yan, K. Huang, J. K. Liu, and F. Chen, “Emergent inference of hidden markov models in spiking neural networks through winner-take-all,” IEEE Trans. Cybern. **PP**, 1–8 (2018).
[19]S. Wenke, A. Rush, T. Bailey, and R. Jha, “Novel spiking neural network utilizing short-term and long-term dynamics of 3-terminal resistive crossbar arrays,” in *IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)* (IEEE, 2017), pp. 432–435.
[20]A. N. Burkitt, “A review of the integrate-and-fire neuron model: I. homogeneous synaptic input,” Biol. Cybern. **95**, 1–19 (2006).