SCIENCE CHINA
Information Sciences

• **Supplementary File** •

How could imperfect device properties influence the performances of spiking neural networks?

Jingyang CHEN, Zhihao WANG, Tong WANG, Heming HUANG, Zheyuan SHAO, Zhe WANG & Xin GUO*

State Key Laboratory of Material Processing and Die & Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Appendix A Leaky integrate-and-fire (LIF) neuron

The leaky integrate-and-fire (LIF) neuron based on the memristive device of Ag/Ta_2O_5:Ag/Pt is shown in Figure A1. The values of R1, R2 and C are 1 kΩ , 20 Ω and 0.1 μF, respectively. When the voltages applied on the memristive device exceed its threshold voltage, the memristive device switches from the high resistance state (HRS) to the low resistance state (LRS); when the voltages applied on the memristive device is lower than the hold voltage, the memristive device switches from the LRS to the HRS. In the simulation, the threshold voltage follows a normal distribution with a mean value of 0.3 V and a standard deviation of 0.003 V; the hold voltage follows a normal distribution with a mean value of 0.24 V with a standard deviation of 0.005 V. The HRS and LRS of the device are 10^{12} Ω and 10^2 Ω, respectively.

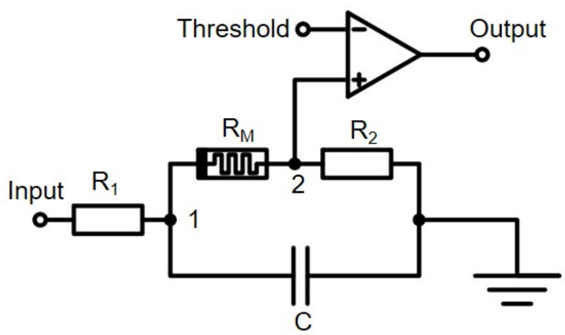

Figure A1 LIF neuron based on the memristive device of Ag/Ta_2O_5:Ag/Pt.

The LIF neuron model can be described as:

$$R_1C\frac{d(V_2(R_M+R_2))}{dt} = V_iR_2 - V_2(R_M+R_1+R_2), \tag{A1}$$

$$V_o = \begin{cases} 1, & V_2 > V_{th} \\ 0, & V_2 \leqslant V_{th} \end{cases}, \tag{A2}$$

where V_i, V_1, V_2 and V_o are the sampling voltage of input, node 1, node 2 and output, respectively.

Appendix B Implementation of the STDP learning rule

Several simple components such as multiplexers (MUXs), timers and basic computing units are needed to realize the STDP learning rule, as shown in Figure B1. Three timers and a MUX are responsible for recording 3 different firing activities. When the input and output neurons fire simultaneously, the MUX receives a control signal "11", and then a pulse is applied on the timer 1 to increase its value. Similarly, when only an input neuron or an output neuron fires, the MUX receives control signals of "10" or "01", and then a pulse is applied on the timer 2 or the timer 3 to add their values. Therefore, after inputting samples, the values of the timers are ΔW_{11}, ΔW_{10} and ΔW_{01}. Afterwards, the computing units add the values of ΔW_{11}, ΔW_{10} and ΔW_{01} to calculate the update strides of each synapse.

* Corresponding author (email: xguo@hust.edu.cn)

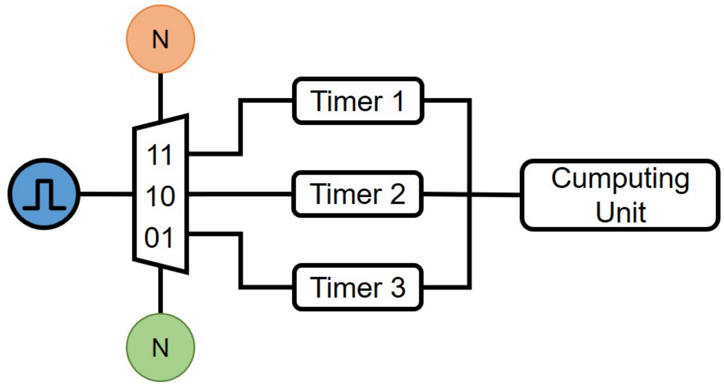

Figure B1 Implementation of the STDP learning rule.

Appendix C Artificial neural network

A fully connected artificial neural network (ANN) is implemented by TensorFlow 2.2, which is commonly used to build ANNs. The structure of the ANN is similar to that of the SNN in the manuscript; both networks are trained by a small dataset (200 samples from the MNIST dataset). In the forward propagation, the training image is expanded into a vector X. The prediction vector Y is obtained by the vector-matrix multiplication (VMM) operation:

$$Y = WX + b, \tag{C1}$$

where W is the weight matrix and b is the bias. Then, a *softmax* function is used as the activation function of the neurons to calculate the probability corresponding to different possible outputs:

$$y_{pred}(i) = \frac{e^{Y(i)}}{\sum\limits_{k=1}^{T} e^{Y(k)}}, \tag{C2}$$

where $Y(i)$ is the i-th element of Y, and T is the number of elements of the vector Y. $y_{pred}(i)$ is the i-th predicted result. The *loss* is calculated following a categorical cross-entropy:

$$loss = \frac{1}{n}\sum_{i=1}^{n} y_{true}(i) \times \Big(-\ln(y_{pred}(i) + 10^{-7})\Big), \tag{C3}$$

where $y_{true}(i)$ and $y_{pred}(i)$ are the label value and the predicted value of the i-th sample, respectively, n is the sample number. After the forward propagation, a standard gradient-based optimization method, *Adam*, is used to minimize the cost function and train the output network.

Appendix D Training SNN without the network verification

Upon eliminating the network verification during training, there is a decline in both the learning rate and the capacity of SNNs to suppress various types of sample noises. Figure D1 shows the conductance updating of the memristive devices in the synaptic array when training the SNN without the network verification. When comparing with Figure 3(C), it is clear from the images that the weights of the features learned in the synaptic array have significantly lower values. The network has the problem of under-learning the sample features, and the network is under-fitting. Furthermore, it can be seen that the synaptic array displays increasingly dispersed features with higher noisy gray values, which may lead to erroneous classification.

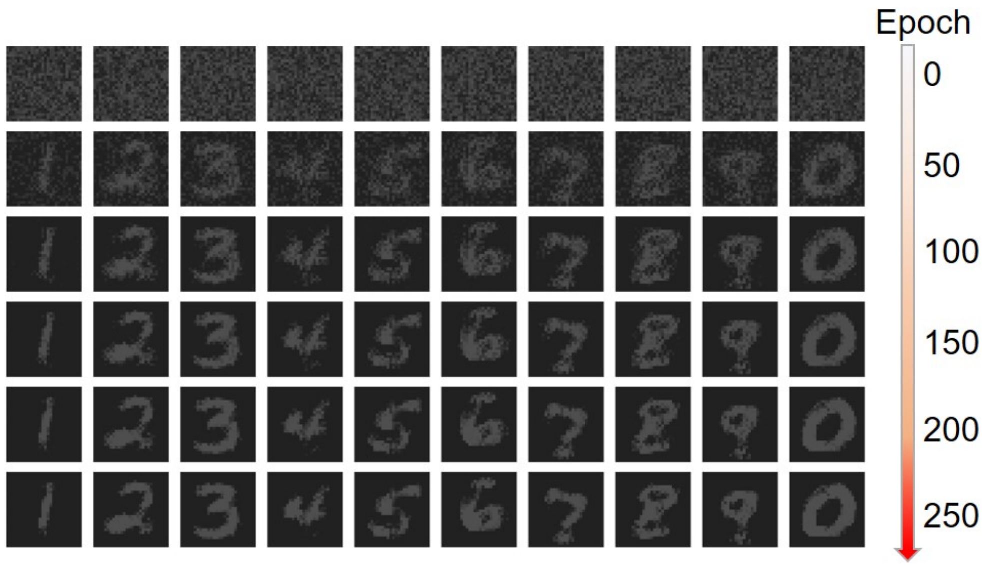

Figure D1 Conductance updating of the memristive devices in the synaptic array when training the SNN without the network verification.

Appendix E Asymmetric ratio and number of conductance states of a memristive device

The asymmetric ratio (AR) of a memristive device is described by Eq. (E1):

$$AR = \frac{\max |G_p(n) - G_d(n)|}{G_H - G_L}, \tag{E1}$$

where $G_p(n)$ is the device conductance after applying the n-th positive spike on it, $G_d(n)$ is the device conductance after applying the n-th negative spike on it; G_H and G_L are the maximum and minimum device conductance, respectively. AR ranges from 0 to 1, describing the relative size of the window between LTP and LTD curves, as shown in Figure E1. A large AR means that the window between the LTP and LTD curves is large, which causes an uneven weight updating during the training period. For example, when the device is at a low conductance state, a positive spike causes a conductance change larger than that at a high conductance state. The large and uneven update stride reduces the classification accuracy of SNNs. When AR approaches 0, the window between the two curves disappears, and the device is identical to an ideal device.

Because the device conductance drifts in the subsequent SET and RESET cycles, some close conductance states are unable to be distinguished. Therefore, conductance states that fluctuate within a specific range (0.23% of the conductance variation range) are combined into one stable conductance state.

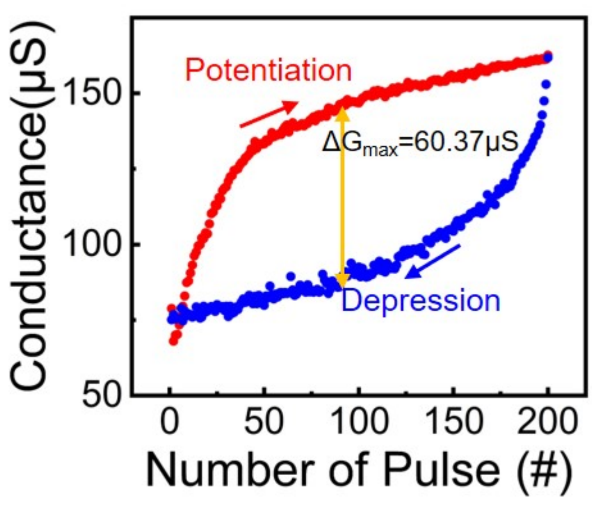

Figure E1 Asymmetric ratio of memristive device.

Appendix F Performances of SNNs when trained with 3000 samples

Figure F1 shows the classification accuracy of the SNNs based on the three different devices trained with 3000 samples. The classification accuracy is improved when the number of samples is increased to 3000. In particular, the SNN based on the Pd/W/WO_3/Pd device achieves a classification accuracy of 80.03%. The highest classification accuracy of the SNNs based on the other two devices are 75.27% and 69.42%, respectively.

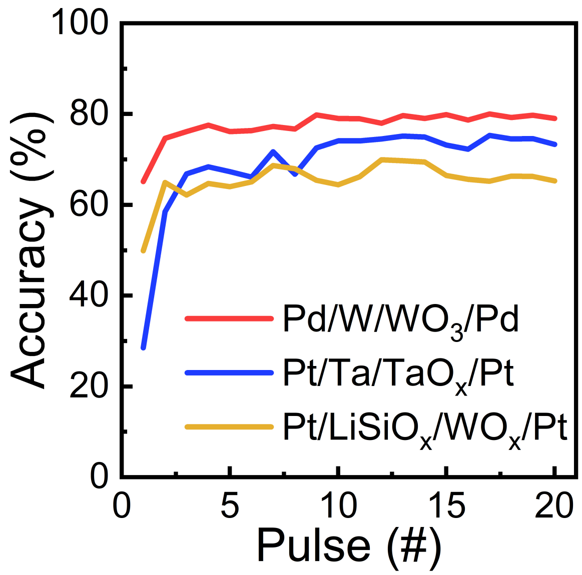

Figure F1 Classification accuracy of the SNNs based on the three different devices trained with 3000 samples.

Appendix G Modelling the device Pd/W/WO_3/Pd

The switching of the memristive device Pd/W/WO_3/Pd is based on the modulation of the filament width [1]. When a forming voltage is applied on the device, mobile ions or ionic defects (primarily oxygen vacancies) migrate and aggregate to generate filaments. As the conductance of the filaments is higher than that of the matrix material, the conductance of the device increases with the forming of filaments. Keeping applying SET spikes, oxygen vacancies continually migrate and aggregate to thicken the filaments, thus the conductance of the device increases, which is defined to be the long-term potentiation (LTP). When a RESET spike is applied on the device, oxygen vacancies migrate and diffuse away from the filaments to the matrix, thinning the filaments and reducing the device conductance, which is defined to be the long-term depression (LTD). The device characteristics can be described by the following equations:

$$I = (1-w)\alpha[1-e^{-\beta V}] + w\gamma, \tag{G1}$$

$$\frac{dw}{dt} = \lambda \sinh(\eta V) - \frac{w}{\tau}, \tag{G2}$$

where Eq. (G1) is the I-V equation including the Schottky term (first term) and the tunnel term (second term). The two conduction channels are parallel, and their relative weights are determined by the internal state variable w; $w = 0$ means fully Schottky-dominated conduction, and the device is at the low conductance state; $w = 1$ means fully Tunnel-dominated conduction, and the device is at the high conductance state. $\alpha, \beta, \gamma, \tau$ and η are all positive parameters. Eqs. (G1) and (G2) are simplified to:

$$G = ae^{-\frac{t}{\tau}} + b, \tag{G3}$$

where G is the conductance of the memristive device Pd/W/WO_3/Pd, a, b and τ are parameters to be fitted. The electrical property of the device Pd/W/WO_3/Pd is shown in Figure G1. The simulation parameters of Eq. (G3) are listed in Table G1.

Table G1 Simulation parameters for the device Pd/W/WO_3/Pd

	a(μS)	τ(μs)	b(μS)
LTP	-87.67	4.03	157.75
LTD	68.40	6.15	75.68

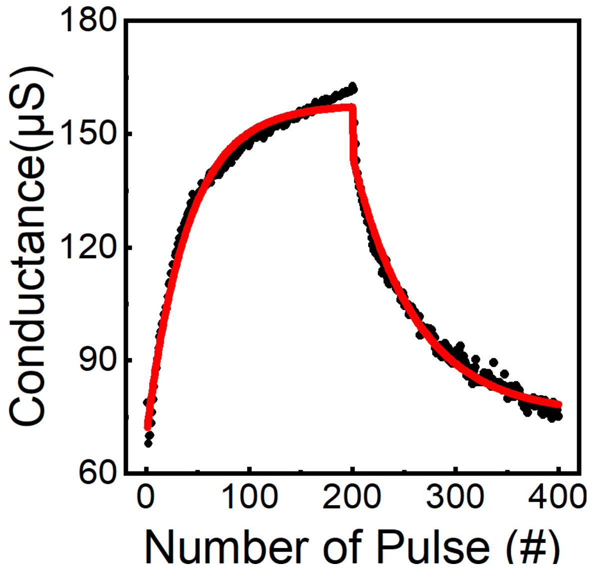

Figure G1 Measured and simulated electrical conductance of the device Pd/W/WO_3/Pd.

References

1 Huang H M, Yang R, Tan Z H, et al. Quasi-Hodgkin-Huxley neurons with leaky integrate-and-fire functions physically realized with memristive devices. Adv Mater, 2019, 31: 1803849