Simulation of the Spiking Neural Network based on Practical Memristor

Xi Zhu, Yi Sun, Haijun Liu, Qingjiang Li and Hui Xu

School of Electronic Science, National University of Defence Technology, Changsha 410073, China.

Abstract. In order to gain a better understanding of the brain and explore biologically-inspired computation, significant attention is being paid to research into the spike-based neural computation. Spiking neural network (SNN), which is inspired by the understanding of observed biological structure, has been increasingly applied to pattern recognition task. In this work, a single layer SNN architecture based on the characteristics of spiking timing dependent plasticity (STDP) in accordance with the actual test of the device data has been proposed. The device data is derived from the Ag/GeSe/TiN fabricated memristor. The network has been tested on the MNIST dataset, and the classification accuracy attains 90.2%. Furthermore, the impact of device instability on the SNN performance has been discussed, which can propose guidelines for fabricating memristors used for SNN architecture based on STDP characteristics.

1. Introduction

Brain-inspired computing is an emerging field, which can achieve artificial intelligence by means of simulating neural network of human's brain, as it can extend the capabilities of information technology beyond the Von Neumann paradigm [1,2]. In biological systems, synapses are the bridges between neurons [1,3], and they can change their intensity to enhance (synaptic potentiation) or weaken (synaptic depression) the connection between two neurons. This mechanism, used by the brain in learning processes, is called synaptic plasticity [2].

Spiking neural network (SNN) is an important way to achieve brain-inspired computing. And the presence of memristors provides a new physical basis for improving synaptic simulation levels [4-6], and then the possibility of further development of brain-like computing. The nanoscale size of the memristors and low energy consumption can enable high-density distribution of the human brain [7-10]. What's more, memristor conductance can be tuned according to the delay time between two spikes successively applied to the device terminals. This regulation of memristor conductance implements the typical biological learning process named Spike Time-Dependent-Plasticity (STDP) [11-13]. Constructing a SNN using memristors as synapses can take full advantage of the characteristics of memristors for high-density distribution and low power consumption. And the neuron should attain functions such as leaky-integrate-and-fire (LIF), inhibitory period, bidirectional transmission and lateral inhibition [14-15].

In recent years many SNN architectures were developed for pattern recognition tasks which relied on different kinds of models of STDP [16]. A Poissonian spike presentation of the MNIST dataset achieves good performance of classification [17]. A computer simulation achieved classification task on MNIST and had the potential to be implemented using memristors [18], and the immune to device variations was simulated [19]. However, many researchers discussed the SNN architectures based on perfect synaptic devices without considering the characteristics of practical devices and the device yield. So the research on SNN architecture considering practical device characteristics has decent value.

In this work, we firstly analyze the STDP characteristics of Ag/GeSe/TiN device and extract a synaptic data model on it. Then a SNN architecture using the data is proposed, and its performance is demonstrated on the MNIST dataset. What's more, the impact of device variability is tested on the 25*10-scale SNN architecture.

2. Synapse Device

In order to better emulate the STDP characteristics, the Ag/GeSe/TiN memristor is fabricated. First, the Ag/GeSe/TiN device has a 5 μm×5 μm effective area. Vertical lines of TiN (40 nm) acting as bottom electrode were deposited on the SiO2/Si substrate by magnetron sputtering after the first lithography process. Then the solid electrolyte GeSe (50 nm) was grown by magnetron sputtering acting as resistive switching layer after the second lithography process. The GeSe layer was patterned

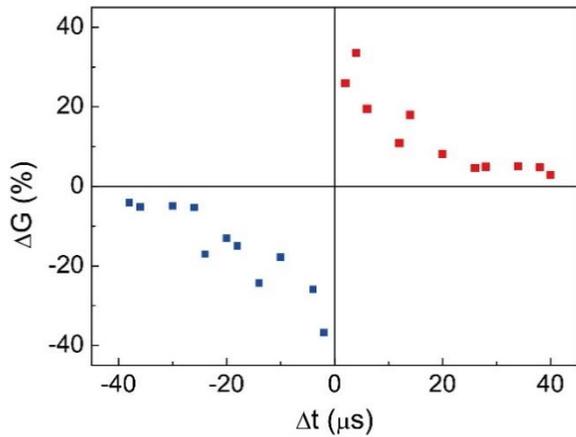to uncover the TiN electrode. After lift-off process, the

Figure 1. STDP characteristics of Ag/GeSe/TiN memristor

substrate with GeSe layer was annealed at 300°C in vacuum environment for 30 seconds. Then the horizontal lines of Ag (40 nm) as top electrode were deposited by e-beam evaporation after the third lithography process.

In order to ensure the validity of the device, all the electrical characteristics of the devices were tested on the Keithley 4200 semiconductor parameter analyser. The ground was applied to the TiN electrode while the Ag electrode was tied to the voltage. The applied signals were equal to the subtraction of pre-spike and post-spike.

To verify the STDP characteristics in Ag/GeSe/TiN memristor, a pair of positive (0.5V,60μs) and negative (-0.5V,60μs) pulses was adopted as spike signal, and we applied the spike signal to the two terminals of the memristor with different Δt. The experiment results are shown in Figure 1. Here Δt represents the interval time between pre- and post-spike. ΔG represents the conductance difference in these two moments. It can be seen that the Ag/GeSe/TiN memristor conductance increases (potentiation, $\Delta G>0$) when the pre-spike appears before the post-spike ($\Delta t>0$). And Vice versa, the memristor conductance decrease (depression, $\Delta G<0$) when the pre-spike follows the post-spike ($\Delta t<0$). What's more, the shorter Δt, the greater the change in the device conductance. It indicates that the Ag/GeSe/TiN device has the STDP characteristics. With 200 pulses, a conductance window of 10 can be achieved, which has been demonstrated to be sufficient for neuromorphic applications [20,21]

Based on the experiment result data of the Ag/GeSe/TiN device, we construct a related synaptic data model, which is used as the synapses of the SNN architecture.

3. Architecture of SNN

According to practical memristor characteristics and model of synaptic data, the architecture of SNN has been proposed. There are two layers in this network including input layer and output layer. Every input neuron has one-to-one correspondence with the pixel in an image. The types in output layer have been divided into 10 blocks. And the number of output neurons n in each block is designed as

appropriate. In this architecture, we choose n=40 in consideration of the classification accuracy and power consumption, as is shown in figure 2. The orange
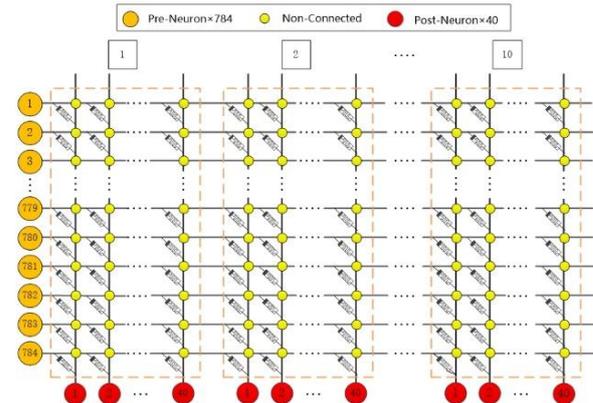

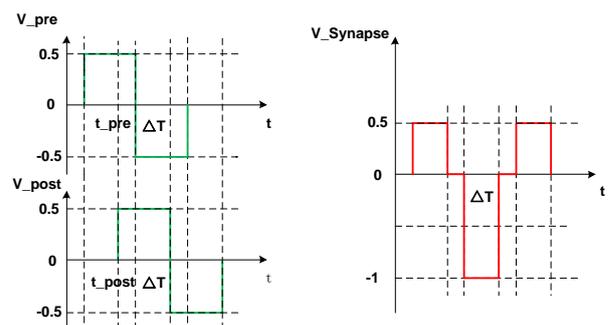**Figure 2.** The architecture of spiking neural network

Figure 3. The spikes on the pre-neurons, post-neurons and synapses

circles represent the pre-neurons and the red circles represent the post-neurons. The row lines are connected with column lines by memristors only.

So there is only one layer of synapses whose weights can be adjusted by the STDP characteristics of Ag/GeSe/TiN device. The concrete rules of adjusting the weights include bidirectional transmission, integrate-and-fire, and lateral inhibition. The number of the synapses is 784*40*10=313,600.

In the training process, the pre-neuron will produce a spike if the related pixel is white in the image at the beginning, and the pre-neurons will produce nothing which is related to a black pixel. Then the post-neurons will receive the spikes transported by the memristors. The post-spike produced by the post-neuron will be overlapped with the pre-spike, as shown in Figure 3. The voltage on the overlapped area will exceed the threshold and change the memristors' conductance. To increase the degree of differentiation, the synapses in n blocks are trained in order.

The testing process is similar to the training process, but there are some differences between testing process and training process. Firstly, the pre-neurons will produce small negative voltage which are related to black pixel. Secondly, the voltage of overlapped area is not big enough for the memristors to change their conductance. Lastly, the post-spikes will be compared with each other, and the block of biggest voltage is the tested label.

4. Simulation and Discussion

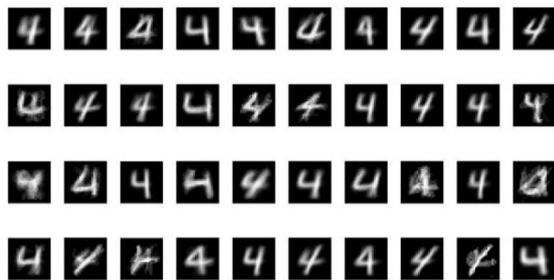### 4.1 Recognizing performance on MNIST dataset

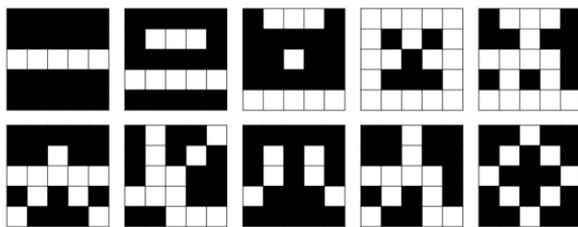**Figure 4.** The images of synaptic weights in block 4

Figure 5. The 5*5-pixel Chinese numerals

In order to test the performance of the architecture, we choose the MNIST benchmark. The MNIST training set (60,000 examples) is designed to train the architecture and the classification accuracy can be attained through testing sets (10,000). Every image has 28*28=784 pixels, which have one-to-one correspondence with the input neurons. The corresponding input neurons receive a pulse, if the pixel value exceeds the threshold we set. And vice versa, if the pixel value can't reach the threshold, the corresponding input neurons receive fixed small negative voltage. Before training, we divide all the training set into 10 types according to their labels. For example, all the training set labeled '4' will be imported into block 4. And then synaptic weights in block 4 will study n major styles of '4' without supervision, as shown in figure 4. The training simulation takes about 43 seconds, and the final classification accuracy is about 90.2%.

4.2 Impact analysis of device variability

The performance of the SNN architecture has been tested on the MNIST dataset and the result is good. However, the memristors can not be adjusted accurately according to the requirements of the algorithm for lack of sufficient awareness of them. In the meanwhile, the result will be so terrible that the influence of memristors' variability can not be analyzed if the architecture is tested on the MNIST dataset considering the device variability. So the similar architecture is adopted to verify the classification accuracy of the group of 5*5-pixel Chinese numerals considering the device variability, as shown in figure 5. the number of pre-neurons is 25, and the number of post-neurons is 10. And the others are the same.

In this section, we try to discuss the impact of device variability with three types, namely maximum resistance R_{max}, minimum resistance R_{min} and device yield D_y. The final resistance is random number between $A*(1-B/2)$ and $A*(1+B/2)$ in Table 1 and Table 2, where A represents R_{max} or R_{min}, B represents fluctuation range Fr.

Table 1 The impact of device variability of synaptic maximum resistance for classification accuracy

R_{max} (kΩ)	F_r				
	0.2	0.4	0.6	0.8	1
60	1	1	1	1	1
80	1	1	1	1	1
100	1	1	1	1	1

Table 2 The impact of device variability of synaptic minimum resistance for classification accuracy

R_{min} (kΩ)	F_r				
	0.5	0.7	0.8	0.9	1
1	1	1	1	1	1
10	1	1	1	0.996	0.99
20	1	0.996	0.994	0.99	0.986

Table 3 The impact of device yield for classification accuracy

T_f	D_y				
	1	0.95	0.9	0.85	0.8
T_o	1	0.988	0.948	0.901	0.83
T_{md}	1	0.989	0.964	0.916	0.878
T_{mi}	1	0.988	0.944	0.912	0.879

In Table 1, it can be seen that the classification accuracy keeps being 1 in spite of the increase of synaptic maximum resistance and synaptic resistance fluctuation range. It shows that the variability of synaptic maximum resistance makes little difference.

Table 2 suggests that the classification accuracy decreases with the increase of synaptic minimum resistance. And so does fluctuation range. The classification accuracy decreases as the synaptic resistance fluctuation range increases. It indicates that the variability of synaptic minimum resistance makes some difference, and the increase of synaptic minimum resistance and fluctuation range reduce the classification accuracy.

In Table 3, we discuss the impact of device failure T_f from three aspects namely opposite T_o, maximum decrease T_{md} and minimum decrease T_{mi} from top to bottom. It can be seen that classification accuracy decreases obviously with the decrease of device yield D_y, and these three types of device failure have similar classification accuracy in the same device yield. It implies that the device yield makes great difference.

The tables show that compared with the maximum resistance, the minimum resistance variability of the synapse device makes more impact, and whatever device yield exerts greater effect under the same circumstances. When a SNN architecture is proposed, keeping the device yield is the first priority, which is followed by the synaptic minimum resistance and fluctuation range. And the last is

the maximum resistance and synaptic resistance fluctuation range.

5. Conclusion

In this paper, SNN architecture is proposed based on data model of the Ag/GeSe/TiN memristor, which has useful STDP characteristic. The architecture has been tested on the MNIST dataset and obtains 90.2% classification accuracy. What's more, the performance of the architecture has been verified considering the device variability. The results show that, compared with the maximum resistance, the minimum resistance variability of the synapse device makes more impact, and whatever device yield exerts greater effect on this pattern recognition task. The conclusion obtained in this work, can be useful for fabricating practical synapses devices to construct feasible SNN.

ACKNOWLEDGEMENTS

References

1. D. Kuzum et al., "Synaptic electronics: materials, devices and applications," *Nanotechnology*, **24**, 382001 (2013)

2. Covi E, Brivio S, Serb A, et al. "HfO2-based memristors for neuromorphic applications," *IEEE International Symposium on Circuits and Systems*, (2016)

3. D. Garbin et al., "HfO2-Based OxRAM Devices as Synapses for Convolutional Neural Networks," *IEEE Trans. Electron Dev.*, **62**, 2494–2501 (2015)

4. Linaresbarranco B, Serranogotarredona T. "Memristance can explain Spike-Time-Dependent-Plasticity in Neural Synapses". *Nature Precedings*, (2009)

5. Markram, H, W. Gerstner, and P. J. Sjöström. "Spike-Timing-Dependent Plasticity: A Comprehensive Overview." *Frontiers in Synaptic Neuroscience* (2012)

6. Markram H, Lübke J, Frotscher M, et al. "Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs". *Science*, **213**, (1997)

7. Choi H, Jung H, Lee J, et al. "An electrically modifiable synapse array of resistive switching memory". *Nanotechnology*, **20**, 15467-15472, 2009

8. Jo S H, Chang T, Ebong I, et al. "Nanoscale Memristor Device as Synapse in Neuromorphic Systems". *Nano Letters*, **10**, 1297, (2010)

9. Kuzum D, Jeyasingh R G D, Lee B, et al. "Nanoelectronic Programmable Synapses Based on Phase Change Materials for Brain-Inspired Computing". *Nano Letters*, **12**, 2179, (2012)

10. Indiveri G, Linares-Barranco B, Legenstein R, et al. "Integration of nanoscale memristor synapses in neuromorphic computing architectures," *Nanotechnology*, **24**, 384010, (2013)

11. S. Ambrogio et al., "Spike-timing dependent plasticity in a transistorselected resistive switching memory," *Nanotechnology*, **24**, 384012, (2013)

12. W. Chan and J. Lohn, "Spike timing dependent plasticity with memristive synapse in neuromorphic systems," in *Proc. of IEEE Int. Joint Conf. Neural Networks (IJCNN)*, 1–6, (2012)

13. T. Serrano-Gotarredona et al., "STDP and STDP variations with memristors for spiking neuromorphic learning systems," *Front. Neurosci.*, **7**, 1–15, (2013)

14. A. Grüning and S. M. Bohte, "Spiking Neural Networks: Principles and Challenges," in *ESANN*, (2014)

15. E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga,"Analog Memristive Synapse in Spiking Networks Implementing Unsupervised Learning," *Frontiers in Neuroscience*, **10**, (2016)

16. Liu, Q., et al. "Benchmarking Spike-Based Visual Recognition: A Dataset and Evaluation." *Front Neurosci.* 10, (2016)

17. Diehl P U, Matthew C. "Unsupervised learning of digit recognition using spike-timing-dependent plasticity". *Frontiers in Computational Neuroscience*, 9:99, (2015)

18. Bill J, Legenstein R. "A compound memristive synapse model for statistical learning through STDP in spiking neural networks". *Frontiers in Neuroscience*, **8**, 412, (2014)

19. Querlioz, D., Bichler, O., & Gamrat, C.. "Simulation of a memristor-based spiking neural network immune to device variations," *International Joint Conference on Neural Networks*, **42**, 1775-1781. (2011)

20. E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga, "HfO2-based memristors for neuromorphic applications," IEEE Int. Symp. Circuits Systems, (2016)

21. D. Garbin, E. Vianello, O. Bichler, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo, and L. Perniola, "HfO2-Based OxRAM Devices asSynapses for Convolutional Neural Networks," IEEE Trans. ElectronDevices, **62**, 2494-2501, (2014)