Extended Abstracts of the 2015 International Conference on Solid State Devices and Materials, Sapporo, 2015, pp1160-1161

TMO-ReRAM based synaptic device for neuromorphic computing

Jinfeng Kang, Bin Gao, Zhe Chen, Peng Huang, Yudi Zhao, Lifeng Liu, Xiaoyan Liu

Institute of Microelectronics, Peking University, Beijing 100871, China
E-mail: kangjf@pku.edu.cn

Abstract

Transition oxide-based resistive switching devices (TMO-ReRAM) have emerged as the leading candidate to realize the synapse functions for neuromorphic computing. This paper will address the design and optimization of TMO-ReRAM based synaptic devices. The impacts of the synaptic devices features on the performances of neuromorphic visual system are discussed. The possible solutions are presented to suppress the intrinsic variation of the TMO-ReRAM based synaptic devices to achieve high recognition accuracy and efficiency of neuromorphic visual systems.

1. Introduction

Brain-inspired neuromorphic computing is an attractive paradigm that complements the von Neumann architecture. Developing hardware based artificial neural network directly imitate human's brain is the best way to realize neuromorphic computing with low cost and low energy consumption [1]. To realize the synaptic function in the artificial neuron network, transition-metal oxide based resistive random access memory device (TMO-ReRAM) is utilized due to its nonvolatile data storage and data computing capability and the simple device structure [2]. However, specific device design is still required to fight against parameter variability, uncontrolled switching behavior, and other non-ideal effects on the neuromorphic computing systems. In this paper, we will introduce and review our work on the design and optimization of TMO-ReRAM based synaptic devices to meet the requirements of neuromorphic systems.

2. Device Optimization

A neural network is consisted of multi-layered neurons and synapses connecting each neuron, as illustrated in Fig.1. Cross-point ReRAM array works as the synapses and the resistive switching behavior represents the modification of synaptic weighting process. Although binary synapse can be used for pattern recognition [3], multilevel switching ability is crucial for ReRAM to change different weightings. However, the generation of oxygen vacancies is usually related to an avalanching process, resulting in the formation of clustered oxygen vacancy filaments and abrupt switching behavior. To solve this problem, we developed a defect engineering based methodology to control the generation and recovery of oxygen vacancies, as illustrated in Fig.2 [4]. After optimization design by the methodology, both set and reset processes show gradual multilevel switching behaviors, as shown in Fig.3 [4].

Fig.4 shows the synaptic training process of the Gd doped-HfO_x devices optimized by the methodology [5]. Identical pulses are applied on the devices to imitate exoteric stimulations, and the resistance changes of the devices under the pulse operations. Set and reset processes correspond to the long-term potentiation (LTP) and long-term depression (LTD), respectively.

3. Design of Neuromorphic System

Fig.5(a) illustrates the established neuromorphic visual system with a winner-take-all architecture (Fig.5(b)) and TMO-ReRAM based synapses [2]. The neurons in the 1st layer (representing retina) fire according to the light intensity of the input pattern, and send pulse to the neurons in the 2nd layer (representing visual cortex) through synapses. The cortex neurons sum and integrate the input currents, and the neuron with the largest input current fires first, which inhibits all the other neurons from firing. Then the winner neuron sends pulse back to all the retina neurons to modulate the weighting of synapses. The task for the system is to classify the face orientation of different persons in the camera images, as shown in Fig.5(c) [6].

Fig.5(d) shows the conductance map of different faces on ReRAM arrays. Due to the resistance fluctuation during training process as shown in Fig.6(a) [7], noises can be observed on the conductance map, which leads to the degradation of pattern recognition accuracy. Fig.6(b) shows that the recognition accuracy decreases with the variation. One way to solve the accuracy degradation issue is to use the mean value of several devices. Fig.6(b) shows that the accuracy increases significantly with the number of devices [7]. To realize such a function, we developed a 3D ReRAM array with vertical devices, as illustrated in Fig.7. Each device can be randomly accessed based on the designed training scheme [8]. ReRAM devices on different layers with the same pillar work as one synaptic cell. These devices receive the same training pulses at the same time. Fig. 8 shows the measured training process on the fabricated two layered array (see inset). Two devices on the same pillar are trained and read at the same time. Significant accuracy improvement can be achieved [6].

To reduce energy consumption during training process, the generation of oxygen vacancies should be carefully controlled to achieve thin filament. Fig.9 shows that when the initial resistance is set to ~1MΩ, the energy per spike can be reduced to less than 0.3pJ [6].

4. Summary

Device level and system level design methodologies are provided to realize high accuracy pattern recognition on the ReRAM based neuromorphic visual systems. The achievements boost the application of neuromorphic computing.

Acknowledgement: This work is supported in part by the 973 Program (2011CBA00600) and NSFC Program (61334007, 61404006). The authors would like to thank Prof. H.-S. Philip Wong at Stanford University and Prof. S.M. Yu at Arizona State University for their great contributions on this work.

References: [1] P. A. Merolla et al, Science 345, 668 (2014). [2] S. Yu et al, Adv. Mater. 25, 1774 (2013). [3] S. Yu et al, Front. Neurosci. 7, 186 (2013). [4] B. Gao et al, Trans. Electron Devices 60, 1379 (2013). [5] B. Gao et al, Progress In Natural Science: Materials International 25, 47 (2015). [6] B. Gao et al, ACS Nano 8, 6998 (2014). [7] B. Gao et al, ICSICT, O39-07 (2014). [8] B. Gao et al, Trans. Electron Devices 61, 1377 (2014).

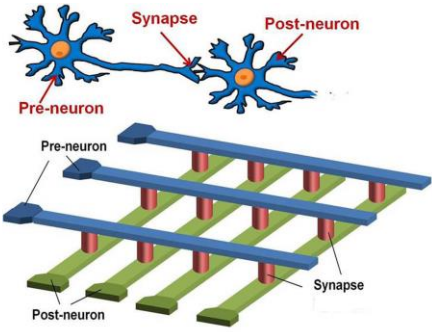

Fig. 1 Schematic view of neural network and TMO-ReRAM based artificial neural network. A neural network is consisted of neurons and synapses. ReRAM array works as synapses to modify the connect strength between pre- and post- neurons.

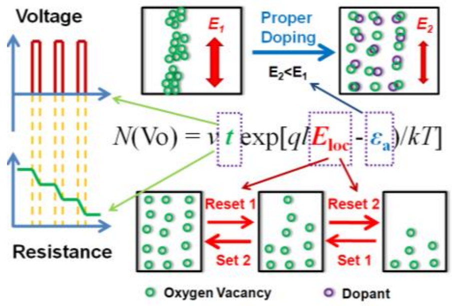

Fig. 2 Schematics of methodology to achieve controlled oxygen vacancy distribution and multilevel switching behavior based on defect engineering. Innovative cell design and operation scheme implementation are developed.

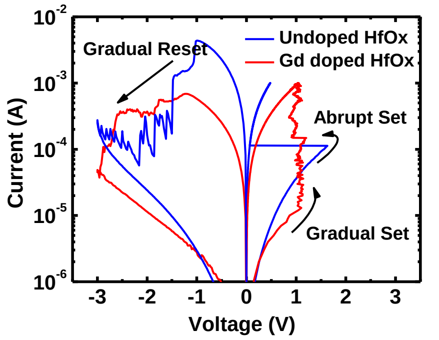

Fig. 3 Typical resistive switching I-V curve of TMO-ReRAM cell before and after optimization. For undoped HfO_x device, abrupt set process is observed. Whereas, after optimization by doping, multilevel set and reset processes are both easy to realize.

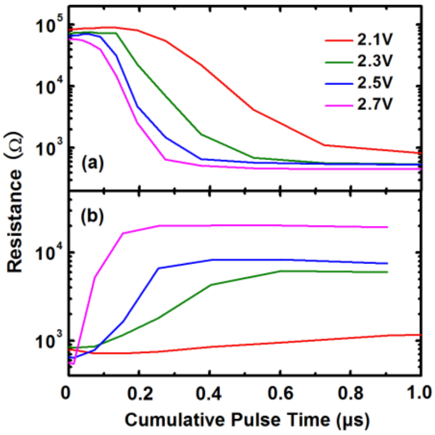

Fig. 4 Measured resistance as a function of cumulative pulse time for TiN/Gd:HfOx/Pt ReRAM device during (a) set and (b) reset process. Multilevel synaptic training processes are observed. Different amplitudes of pulses are used. Training speed increases with pulse amplitude.

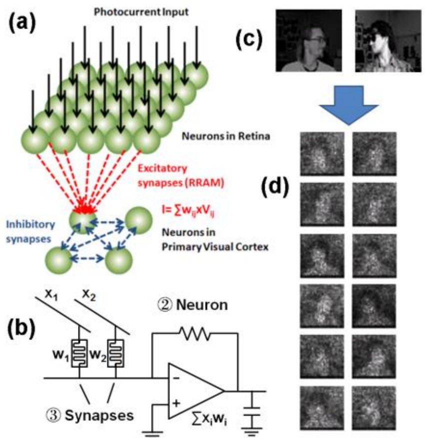

Fig. 5 (a) Schematic of a neuromorphic visual system consisting two layers of neurons. (b) An implementation of neuron circuit with TMO-ReRAM synapses. (c) Training samples used in the visual system. (d) Simulated final resistance maps of TMO-ReRAM with each cortex neuron after training.

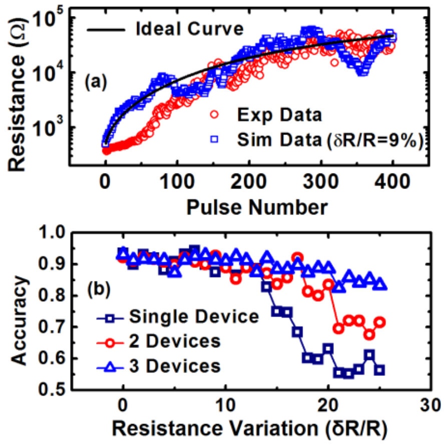

Fig. 6 (a) Measured and simulated multilevel reset training process by 400 identical consecutive pulses. (b) Simulated pattern recognition accuracy of the visual system by paralleling several TMO-ReRAM devices working as a synapse together. A significant improvement is observed as the device number increasing.

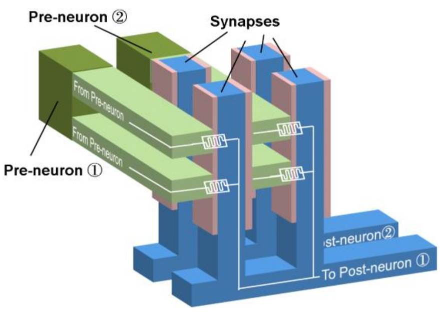

Fig. 7 Schematic of 3D vertical ReRAM array architecture to immunize resistance variation during synaptic training process. The devices with the same pillar electrode can be viewed as a synapse. More devices can be introduced as a combined synapse by increasing horizontal layers.

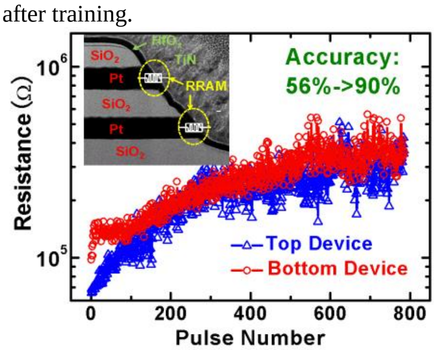

Fig. 8 Measured training process of the top and bottom ReRAM devices in the 3D vertical array by applying 800 identical consecutive pulses. Pulses applied on the two devices with the same vertical electrode at the same time. Inset: TEM image of the fabricated 3D vertical ReRAM cells.

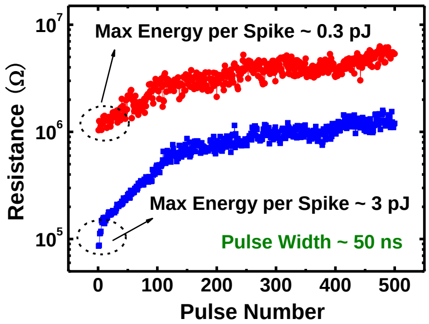

Fig. 9 Measured training process for the fabricated 3D vertical synaptic device with different initial resistance states achieved by different current compliances during the previous set cycles. If starting at 1 MΩ, the maximum energy consumption per spike drops below 1 pJ.