 Extended Abstracts of the 2019 International Conference on Solid State Devices and Materials, Nagoya, 2019, pp993-994

Modeling of Embedded Split-Gate Flash Memory for Pattern Classification

Lu-Rong Gan, Chen Wang, Lin Chen*, Hao Zhu, Qing-Qing Sun and David Wei Zhang
State Key Lab. of ASIC and System, School of Microelectronics, Fudan University, Shanghai, 200433, China
Phone: +86-21-55664324 *E-mail: linchen@fudan.edu.cn

Abstract

We have simulated the LTP/LTD performance of 90nm split-gate SuperFlash® memory cell by Sentaurus TCAD tools and constructed a two-layer perceptron to do pattern recognition based on the 59 available current state values derived from the LTP/LTD simulation data. The recognition accuracy achieves 100% which is consistent with the original network. These results are useful for the extraction of more current states and the application in complex analog neuromorphic networks.

1. Introduction

Non-volatile floating gate memory has been used as a synaptic weight storage cell in analog neuromorphic networks for more than 30 years [1]. Compared with the "synaptic transistor" made by standard CMOS technology (~10^3 F^2, where F is the minimum feature size and large area leads to large time delay and power consumption) [2-4], the non-volatile floating gate memory can be scaled down to F ~ 20nm after highly optimized [5], and the current state of a single cell can be individually and accurately adjusted in the array, which is very suitable for the application of adjustable synapse in analog neuromorphic networks.

The advanced 3rd generation SuperFlash® cell (ESF3) adopt the split gate design, compared with the stacked gate cell, it shows very fast, efficient, and low power Source-Side CHE program [6]. It uses Poly-to-poly FN tunneling erase, low power read, without the over-erase problem and allow simpler program-erase algorithms design [7]. Currently, ESF3 mainly serves the general single-chip microcomputer as well as the automotive and smart card markets. Because of the excellent performance and reliability, this paper researches a hot application field for embedded ESF3-90 -- analog neuromorphic networks.

2. Device structure and electrical characteristics

This simulation is based on the process flow designed for embedded ESF3-90 technology. The logic process is based on 90nm LP process with 1.2V core and 3.3V I/O. The process and electrical characteristics of the device have been researched by Sentaurus TCAD tools. The diagram of the device structure is shown in Fig. 1. In neuromorphic networks, long-term potentiation (LTP) and long-term depression (LTD) are the material basis for learning and memory. For the LTP/LTD operation of ESF3 cell, the corresponding bias conditions are given in Table I. Five channel voltage-pulse signals are applied to device synchronously. For the LTP operation, ESF3 cell is erased by using a series of 100 identical voltage-pulse signals (V_{EG}=13.5V) with 0.01-ms peak width at half height (PWHH) and I_d gradually increases with the

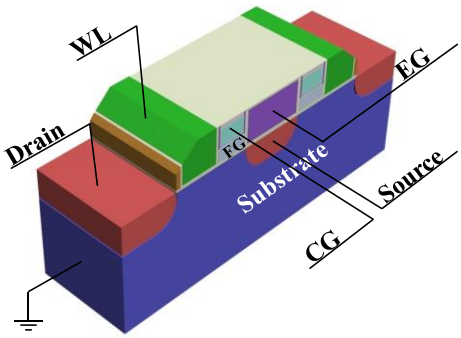

Fig.1 The structure of ESF3 memory cell.

Table I Operation Voltage of ESF3 Memory Cell.

	WL	Drain (BL)	Source	EG	CG
Erase	0V	0V	0V	13.5V	0V
Program	1V	0V	4.5V	4.5V	10V
Read	3.3V	1V	0V	0V	3.3V

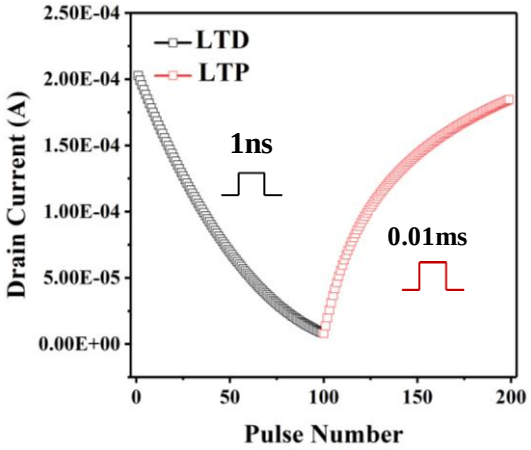

Fig.2 LTP/LTD behaviors of the device using the operation voltage scheme in Table I.

increase in the pulses number, while for the LTD operation, EFM is programmed by using a series of 100 identical voltage-pulse signals (V_{WL}=1V, V_{Source}=4.5V, V_{EG}=4.5V, and V_{CG}=10V) with 1-ns PWHH and I_d gradually decreases with the increase in the number of pulses. The weight update behavior is shown in Fig. 2. Transient current is read by applying 2-ns voltage-pulse signals (V_{WL}= 3.3V, V_{BL}=1V, V_{CG}=

3.3V) behind each LTP/LTD voltage-pulse signal. If needed, we can change the PWHH and/or the amount of the LTP/LTD voltage-pulse signals to achieve more precise adjustment.

3. Weight mapping and pattern recognition

For the neuromorphic classifier, when the network is trained, the weights are constantly updated as the number of training epoch increases. And the differences between weights are important because they show that the network can remember the characteristics of the input objects very well, which is the key metric for more accurate pattern classification.

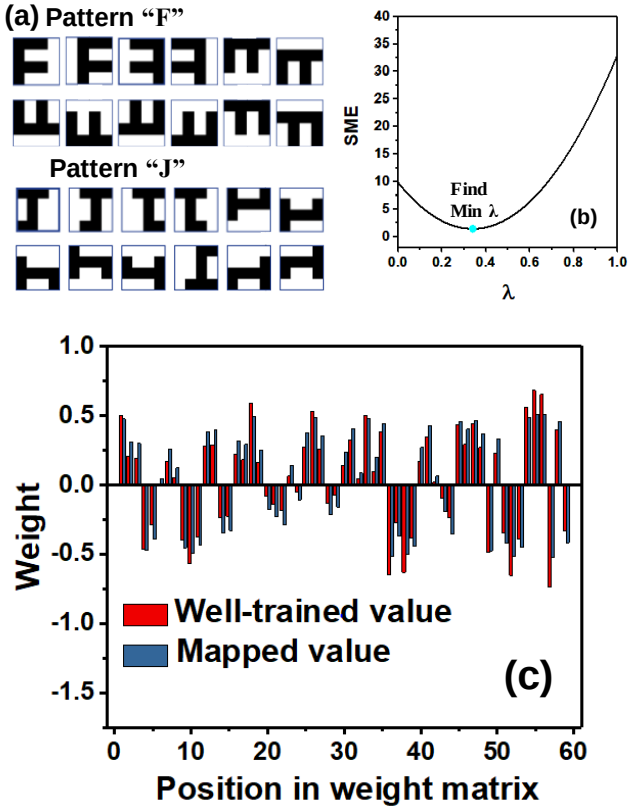

Fig.3 Multilayer perceptron classifier: (a) Input training set. (b) Find the optimal mapping coefficient λ to minimize squared matching error by scaling the 59 available current state values from the LTP data. (c) Weight distributions of the well-trained values and mapped values in the two-level weight matrix of the perceptron.

We construct a two-layer fully connected perceptron (one hidden layer), which has 16 inputs representing 4×4 black-and-white pixels of the input images, three hidden layer neurons, and two output neurons. The input layer and the hidden layer each contain a bias. The activation function uses rectified linear unit (ReLU), and the perceptron training uses traditional backpropagation algorithm. There are (4×4+1) ×3+ (3+1) ×2 = 59 synaptic weights between different layers of neurons. We simply use two ESF3 cells to represent a signed floating-point number (one of which is used to represent symbols), so the network needs 118 ESF3 cells in total. Fig. 3(a) is the input training set for two-class patterns, stylistically representing the letters "F" and "J" (the test set is the same). After 100 epochs, the accuracy of network classification reaches 100%. We extract the trained weights and take absolute values to map the first 59 current values of LTP data in Fig. 2. Here, we define a mapping coefficient λ, by finding an optimal λ that minimizes the square sum of the difference of the trained weights and the scaled current values [8]. Then, we add the corresponding symbols to the current values after mapped and put them back into the original network to get a new network. Since the network is designed according to the number of current states in LTP/LTD simulation, and the number of synaptic weights is less than the total number of LTP/LTD current states in Fig. 2. Thus, for convenience, instead of assigning the original well-trained floating-point numbers to the nearest memory current values in the mapping process [8], we do one-to-one mapping first, although this simple approach reduces the recognition accuracy. If the recognition accuracy of the matched network differs greatly from that of the original network, we will perform a more accurate matching by applying more complex matching rules.

Finally, the same test set is tested through the network, and the recognition accuracy is still 100%. It shows that the fault-tolerant ability between the LTP/LTD current states of the ESF3-90 memory cell is sufficient for this classification task. Thus, in further exploration, we can extract more current states, and combine the corresponding programming algorithm to realize more application research of complex analog neuromorphic networks.

4. Conclusions

The LTP/LTD behaviors simulation of the ESF3-90 memory cell has shown good synaptic performance. In the experiment of weight mapping, the LTP/LTD data is applied to a two-layer fully connected perceptron to do a simple recognition task, and the recognition accuracy of the new mapped network has achieved 100%, which is same to the original well-trained network. So, ESF3-90 memory cell can be as a kind of multi-valued memory applied to the analog neural networks thus saving a lot of memory resources to store weights and intermediate values.

Acknowledgement

The authors would like to acknowledge the financial support in part by the NSFC (61704030 and 61522404), Shanghai Rising-Star Program (19QA1400600), the Program of Shanghai Subject Chief Scientist (18XD1402800), and the Support Plans for the Youth Top-Notch Talents of China.

References

[1] C. Mead, Analog VLSI and Neural Systems. Addison Wesley (1989).
[2] C. Diorio, *et al.*, IEEE Trans. Elec. Devi. **43** (1996) 1972.
[3] S. Chakrabartty, *et al.*, IEEE J. Solid-St. Circ. **42** (2007) 1169.
[4] J. Hasler, *et al.*, Front. Neurosci. **7** (2015) 118.
[5] X. Guo, *et al.*, *IEEE International Electron Devices Meeting* (2017) 151.
[6] Y. Tkachev, et *al.*, *IEEE International Memory Workshop* (2017).
[7] D. Shum, *et al.*, *IEEE International Memory Workshop* (2017).
[8] B. Yan, *et al.*, *IEEE International Electron Devices Meeting* (2017) 270.