# Synapse Device Based on Charge-Trap Flash memory for Neuromorphic Application

Yu Jeong Park, Yoon Kim
Dept. of Nanoenergy Engineering,
BK21 Plus Nanoconvergence
Technology Division
Pusan National University
Pusan, KOREA
pyoojeng@naver.com,
yoonkim@pusan.ac.kr

***Abstract*—This paper proposes a synaptic device based on charge trap flash (CTF) memory that has good CMOS compatibility and excellent reliability compared to other synaptic devices. Using hot-electron injection (HEI) and hot-hole injection (HHI), we designed the operation method to fulfill incremental-step-pulse program (ISPP). To demonstrate the ability of the device for neuromorphic, the device simulation (TCAD) and the MATLAB simulation were performed simultaneously. We also implemented the multi-level operation.**

***Keywords*—Synapse device, charge trap flash (CTF) memory, incremental-step-pulse program (ISPP), neuromorphic**

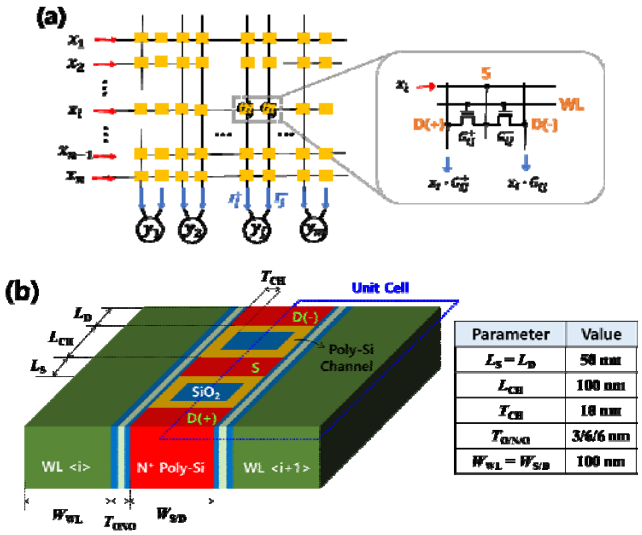
## I. Introduction

Recently, neuromorphic systems have been spotlighted to overcome the existing computing systems based on the von Neumann architecture [1-5]. The term "neuromorphic" refer to artificial neural system that mimics neurons and synapses of the biological nervous system [3]. In the case of the biological nervous system, it is composed of neurons and synapses. A neuron generate a spike when a signal received from the pre-neuron exceeds a threshold, and the generated spike is transmitted to the post-neuron. A synapse refers to the junction between neurons, and each synapse has its own synaptic weight, which is the connection strength between neurons [6]. In neuromorphic system, synaptic weight can be represented by the conductance of synapse device.

To implement neuromorphic system, it is essential to develop a synapse device that has small cell size, low-energy consumption, multi-level operation, symmetric and linear weight change, high endurance and CMOS compatibility [5]. Various memory devices, such as static random-access memory (SRAM) [7], resistive random-access memory (RRAM) [8], phase change memory (PCM) [9], floating gate-memory (FG-memory) [10], have been proposed to implement the synapse operation. However, the above-mentioned devices have critical limitations [5].

In this paper, we propose a synaptic device based on a charge trap flash (CTF) device, which has already been commercialized in NAND flash memory [11-12]. Compared to other memory devices, CTF devices have good CMOS compatibility and excellent reliability [5]. The conductance of CTF device can be modulated by the hot-electron injection (HEI) and hot-hole injection (HHI) mechanisms. To obtain excellent synaptic behavior, studies have been conducted on the pulse scheme for HEI and HHI. Also, the feasibility of the synaptic device is verified through a pattern recognition application with the Modified National Institute of Standards and Technology (MNIST) database.

This research was supported by the MOTIE (Ministry of Trade, Industry & Energy (10080583) and KSRC (Korea Semiconductor Research Consortium) support program for the development of future semiconductor devices.




| Parameter | Value |
|---|---|
| $L_S = L_D$ | 50 nm |
| $L_{CH}$ | 100 nm |
| $T_{CH}$ | 18 nm |
| $T_{ONO}$ | 3/6/6 nm |
| $W_{WL} = W_{S/D}$ | 100 nm |

Fig. 1. (a) Proposed synapse crossbar array to act inhibitory and excitatory in the same time. (b) Synaptic device based on a CTF device.

## II. Synapse Device

Generally, the synaptic behaviors include the potentiation operation to increase the synaptic weight by excitatory synapse part and the depression operation to decrease the synaptic weight by inhibitory synapse part. In this section, we proposed the synapse device based on a CTF device and the operation method for the synaptic behavior.

### A. Synaptic device based on the CTF memory

An artificial neural network (ANN) with a single $n \times m$ can be implemented by a crossbar array of memory device pairs shown in Fig. 1(a). The synapse weight of each synapse can be represented by different two conductances: $w_{ij} = G_{ij}^{+} - G_{ij}^{-}$, where $w_{ij}$ is the weight of the synapse device from neuron $i$ to $j$. By using a pair of devices, it can be represented the negative and positive weight at the same time [14]. The total output current is explained by the sum of each vertical current as follows:

$$y_j = I_j^{+} - I_j^{-} = \sum_{i=1}^{n} x_i \cdot G_{ij}^{+} - \sum_{i=1}^{n} x_i \cdot G_{ij}^{-} = \sum_{i=1}^{n} x_i \cdot (G_{ij}^{+} - G_{ij}^{-}) = \sum_{i=1}^{n} x_i \cdot w_{ij}$$

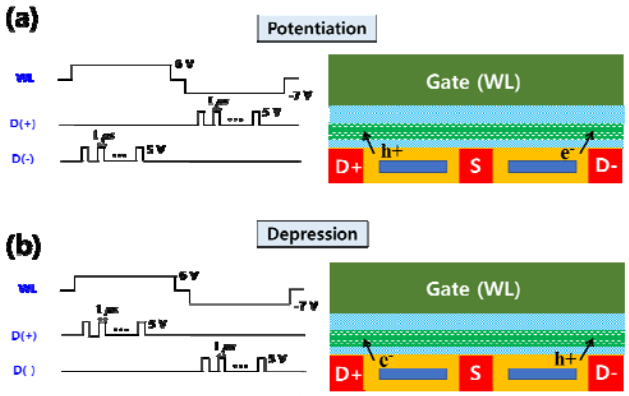



Fig. 2. (a) Potentiation method by successive drain pulse train. (b) Depression method by successive drain pulse train.

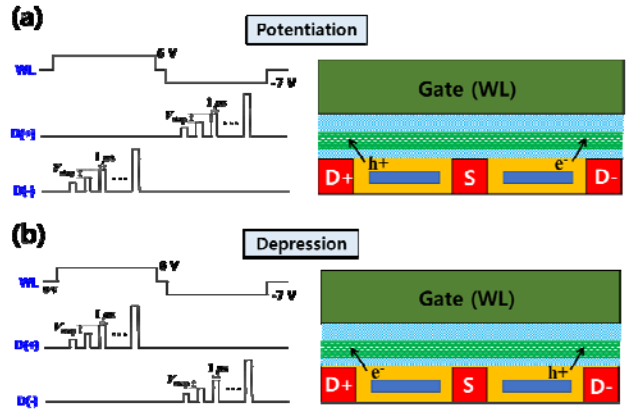



Fig. 3. (a) Potentiation method by ISPP pulse train. (b) Depression method by ISPP pulse train.

Fig. 1(b) shows the proposed device based on CTF device with a silicon nitride layer as a storage layer. It is combined two MOSFET transistors sharing a source (S) and gate. When the voltage is applied to a source, the drain currents of D(+) and D(-) flow into the neuron circuit and is determined by $G_{ij}^{+}$ and $G_{ij}^{-}$. Through this architecture, the part of excitatory can be represented by $G_{ij}^{+}$ and inhibitory can be represented by $G_{ij}^{-}$ simultaneously. Each conductance is determined by threshold voltage according to the amount of trapped charge in each storage layer. A technology computer-aided design (TCAD) simulation (Synopsys Sentaurus) was used to demonstrate the synaptic operation of the device [15].

### B. *Successive-Pulse Programing*

HEI and HHI are used as a charge injection mechanism and for the conductance modulation of the proposed synapse device. HEI occurs when positive voltage is applied to the drain under the positive gate bias, which means the conductance decreases and the threshold voltage increases by injecting the electron [16]. On the other hand, if the drain voltage is applied under the negative gate bias, the conductance is increased and the threshold voltage is decreased by HHI as injecting the hole. In addition, HHI operates at lower voltage and faster than Fowler-Nordheim (FN) tunneling [17-18]. As a result, the potentiation can be performed simultaneously by HHI in the D(+) region and HEI in the D(-) region, as shown in Fig. 2(a). The depression operation is performed by HEI in the D(+) region and HHI in the D(-) region, as shown in Fig. 2(b).

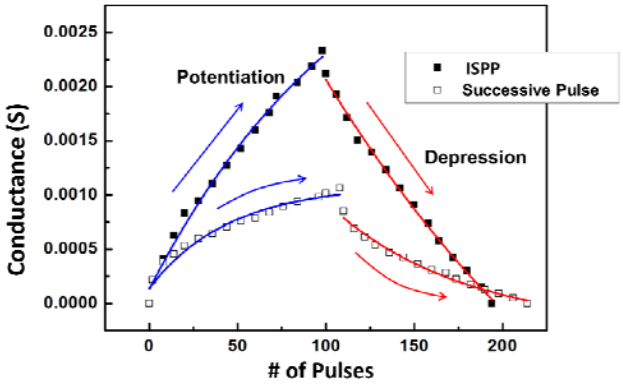



Fig. 4. Gradual changes of conductances (synaptic weight) by successive pulse programming and ISPP.

### C. *Incremental-Step-Pulse Programming*

The Incremental-Step-Pulse Programming(ISPP) is used for the program scheme of NAND flash memory [12]. The program pulse is increased by a constant value $V_{step}$ after each program step as shown in Fig. 3. Using a TCAD device simulation, we compared the conductance modulation characteristics of the successive-pulse programming and the ISPP. As shown in Fig. 4, the ISPP scheme shows better synaptic behavior than the successive-pulse scheme. The ISPP scheme showed that the conductance changes linearly according to the number of applied pulses. Also, the range of available conductances (memory window) can be further increased [19]. Consequently, The ISPP scheme is more suitable than the successive-pulse scheme for the operation method of the proposed synapse device.

### D. *Pattern Recognition Simulation*

To demonstrate the functionality of the proposed device, the single-layer ANN system was simulated as shown in Fig. 1 (a). The MNIST database is a large database of handwritten digits, which contains about 60,000 learning images and 10,000 test images. A total of 784 input nodes and 10 output nodes are used for MNIST pattern recognition. At this time, 784 input nodes represent $28 \times 28$ black pixels of the learning image, and 10 output nodes represent ten digits (0-9). We also used a rectifier linear unit (ReLU) as a activation function, which is one of the popular activation functions. Because the lack of vanishing gradient problems appear in other ones such as sigmoid or hyperbolic tangent functions [20-22]. The learning process is as follows: the error is calculated by supervised learning, and the target of conductance change is determined by the gradient descent method. After that, the synaptic weight value is updated based on a fitted equations for the conductance modulation characteristics with the successive-pulse scheme and the ISPP scheme.

Fig. 5(a) shows the MATLAB simulation result of the recognition accuracy using the 10,000 untrained samples based on the number of trained samples is plotted in Fig. 4. The synaptic weight maps are illustrated in Fig. 5(b). The successive-pulse scheme and the ISPP scheme show accuracy of 79.83 % and 85.9 %, respectively. This result indicates that synaptic devices should have linear conductance modulation characteristic for the better performance of a neuromorphic system [23].

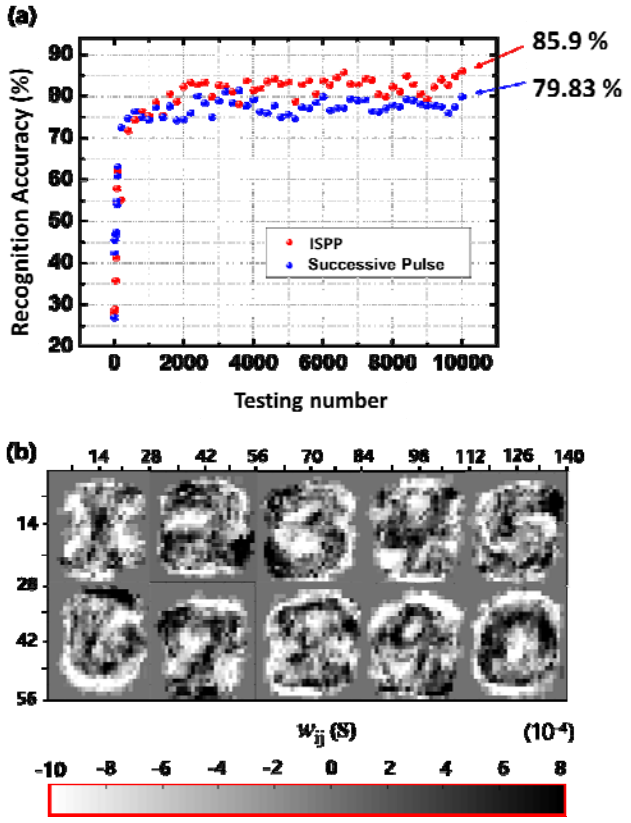



Fig. 5. (a) Recognition accuracy as a function of the number of trained samples. (b) Synaptic weight map of the ISPP after training 10,000 samples.

We have also analyzed the conductance modulation characteristics according to the $V_{step}$ of the ISPP scheme and its effect on the pattern recognition rate. As illustrated in Fig. 6(a), a smaller $V_{step}$ allows for fine conductance modulation, which means that the number of conductance level can be increased. Consequently, the ISPP scheme with the smaller $V_{step}$ exhibits better pattern recognition rate as shown in Fig. 6(b).

## III. Conclusion

We have proposed a synapse device based on a CTF memory device. The operation method was designed by using HEI and HHI to apply ANN systems. The synapse behavior was operated by TCAD simulation, and we verified the ability of synaptic device through a MATLAB simulation with MNIST database. Finally, we obtained the high accuracy and implemented the multi-level operation by ISPP.

## Acknowledgment

This research was supported by the MOTIE (Ministry of Trade, Industry & Energy (10080583) and KSRC (Korea Semiconductor Research Consortium) support program for the development of future semiconductor devices.

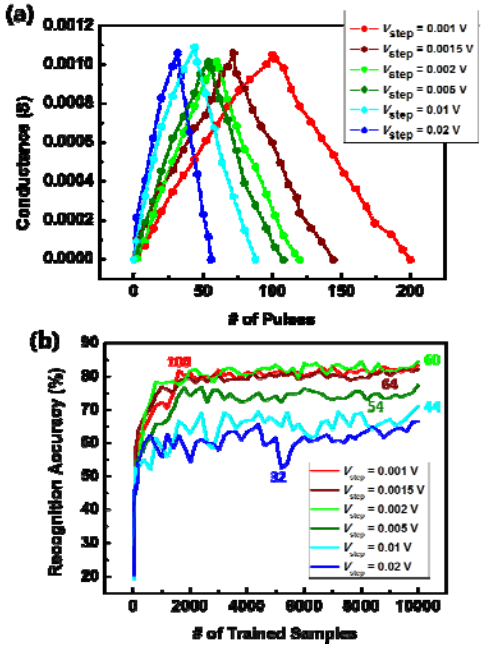



Fig. 6. (a) The gradual conductance change by applying various $V_{step}$. (b) Recognition accuracy as a function of the number of the trained samples. The number in the figure is the pulse number from minimum conductance to maximum conductance called the conductance level.

## References

[1] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S. P. Wong, "A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation," *Advanced Materials*, vol. 25, pp. 1774-1779, March 2013.

[2] X. Liu, M. Mao, B. Liu, H. Li, Y. Chen, B. Li, Y. Wang, H. Jiang, M. Barnell, Q. Wu, and J. Yang, "RENO: A high-eficient reconfigurable neuromorphic computing accelerator design," in *52nd Design Automation Conference (DAC)*, June 2015, pp. 1-6.

[3] C. A. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, pp. 1629-1636, Oct. 1990.

[4] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, "Neuromorphic computing using non-volatile memory," *Advances in Physics: X*, vol.2, no. 1, pp. 89-124, Dec 2016.

[5] H.-S. Choi, D.-H. Wee, H. Kim, S. Kim, K.-C. Ryoo, B.-G. Park S. H. Park, and Y. Kim, "3-D Floating-Gate Synapse Array with Spike-Time-Dependent Plasticity," *IEEE Transactions on Electron Devices*, vol. 65, no. 1, pp. 101-107, Jan 2018.

[6] P. D. Roberts, and C. C. Bell, "Spike timing dependent synaptic plasticity in biological systems," *Biol. Cybern*, vol. 87, pp. 392-403, June 2002.

[7] F. Akopyan, J. Sawada, A. Cassidy, R. Alveraz-lcaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G.-J. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, and D. S. Modha, "Truenorth: Design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537–1557, Oct 2015.

[8] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S. P. Wong, "An Electronic Synapse Device Based on Metal Oxide Resistive Switching Memory for Neuromorphic Computation," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2729–2737, Aug 2011.

[9] N. Panwar, D. Kumar, N. K. Upadhyay, P. Arya, U. Ganguly, and B. Rajendran, "Memristive synaptic plasticity in $Pr_{0.7}Ca_{0.3}MnO_3$ RRAM by bio-mimetic programming," in *Proc. 72nd Annu. Device Res. Conf.*, Jun 2014, pp. 135–136.

[10] C. Diorio, P. Hasler, B. A. Minch, and C. A. Mead, "A floating-gate MOS learning array with locally computed weight updates," *IEEE*

*Transactions on Electron Devices*, vol. 44, no. 12, pp. 2281–2289, Dec 1997.

[11] M. Kang and Y. Kim, "Natural Local Self-Boosting Effect in 3D NAND Flash Memory," *IEEE Electron Device Letters*, vol. 38, no. 9, pp. 1236–1239, Sep. 2017.

[12] Y. Kim, J. Y. Seo, S.-H. Lee, and B.-G. Park, "A New Programming Method to Alleviate the Program Speed Variation in Three-Dimensional Stacked Array NAND Flash Memory," *Journal of Semiconductor Technology and Science*, vol. 14, no. 5, pp. 566–571, Oct. 2014.

[13] T. Cho, Y.-T. Lee, E.-C. Kim, J.-W. Lee, S. Choi, S. Lee, D.-H. Kim, W.-G. Han, Y.-H. Lim, J.-D. Lee, J.-D. Choi, and K.-D. Suh, "A Dual-Mode NAND Flash Memory: 1-Gb Multilevel and High-Performance 512-Mb Single-Level Modes," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 11, pp. 1700-1706, Nov 2001.

[14] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. CCassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668-673, 2014.

[15] Sentaurus Device User Guide, Ver. I-2013.12, Synopsys Inc., Mountain View, CA, USA.

[16] C. Diorio, P. Hasler, B. A. Minch, and C. A. Mead, "A Single-Transistor Silicon Synapse," *IEEE Transactions on Electron Devices*, vol. 43, no. 11, pp. 1972-1980, Nov. 1996.

[17] P. Hasler, B. A. Minch, and C. Diorio, "Adaptive circuits using pFET floating-gate devices," in *Proc. 20th Annu. Device Res. Conf.*, March 1999, pp. 215-229.

[18] H. Kim, S. Cho, M.-C Sun, J. Park, S. Hwang, and B.-G. Park, "Simulation Study on Silicon-Based Floating Body Synaptic Transistor with Short- and Long-Term Memory Functions and Its Spike Timing-Dependent Plasticity," *Journal of Semiconductor Technology and Science*, vol. 16, no. 15, pp. 657-663, Oct. 2016.

[19] H.-T. Lue, T.-H. Hsu, Y.-H. Hsiao, S.-C. Lai, E.-K. Lai, S.-P. Hong, M.-T. Wu, F. H. Hsu, N. Z. Lien, C.-P. Lu, S.-Y. Wang, J.-Y. Hsieh, L.-W. Yang, T. Yang, K.-C. Chen, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "Understanding STI edge fringing field effect on the scaling of charg-trapping (CT) NAND Flash and modeling of incremental step pulse programming (ISPP)," in *Electron Devices Meeting (IEDM)*, Dec. 2009, pp. 1-4.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.

[21] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013.

[22] A, Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2013.

[23] G. W. Burr, R. M. Shelby, S. Sidler, C. d. Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, and H. Hwang, "Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element," *IEEE Transactions Electron Devices*, vol. 62, no. 11, pp. 3498–3507, July. 2015.