

Article

# Solving Overlapping Pattern Issues in On-Chip Learning of Bio-Inspired Neuromorphic System with Synaptic Transistors

Hyungjin Kim <sup>1,\*</sup> and Byung-Gook Park <sup>2,\*</sup>

<sup>1</sup> Department of Electronic Engineering, Yeungnam University, Gyeongsan 38541, Korea

<sup>2</sup> Inter-University Semiconductor Research Center and Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea

\* Correspondence: hyungjin@yu.ac.kr (H.K.); bgpark@snu.ac.kr (B.-G.P.)

Received: 23 November 2019; Accepted: 18 December 2019; Published: 21 December 2019



**Abstract:** Recently, bio-inspired neuromorphic systems have been attracting widespread interest thanks to their energy-efficiency compared to conventional von Neumann architecture computing systems. Previously, we reported a silicon synaptic transistor with an asymmetric dual-gate structure for the direct connection between synaptic devices and neuron circuits. In this study, we study a hardware-based spiking neural network for pattern recognition using a binary modified National Institute of Standards and Technology (MNIST) dataset with a device model. A total of three systems were compared with regard to learning methods, and it was confirmed that the feature extraction of each pattern is the most crucial factor to avoiding overlapping pattern issues and obtaining a high pattern classification ability.

**Keywords:** neuromorphic system; on-chip learning; overlapping pattern issue; pattern recognition; synaptic device; spiking neural network

## 1. Introduction

Even though computing systems based on von Neumann architecture still dominate computer architecture, this architecture is considered inefficient for dealing with big data in the training of deep neural networks (DNNs) because of its serial signal processing [1]; therefore, a totally new computing system is required for the next generation of artificial intelligence. Recently, a bio-inspired neuromorphic system based on a spiking neural network (SNN) has been widely investigated because of its power-efficiency and parallel signal processing properties [2–5]. With regard to its application, the neuromorphic system, which is a hardware implementation of an artificial neural network, has been utilized mostly for pattern recognition [6–10], but also as a denoising auto encoder [11], for color image reconstruction [12], and for speech recognition [13]. In addition, various kinds of electronic devices have been studied as an artificial synaptic device, a crucial building block for constructing neuromorphic systems, including resistive switching materials [14–17], phase change materials [18–20], ferroelectric materials [21,22], and transistors [23–25]. Among them, transistor-based synaptic devices are considered as having better reliability characteristics and device variation for very-large-scale integration (VLSI) implementation of neural networks compared to their counterparts.

In our previous works, we reported a synaptic transistor with an asymmetric dual-gate structure as having short- and long-term memories and spike-timing dependent plasticity (STDP) characteristics [26–28], and its fabrication method [29]. In this work, a system-level study of a SNN for pattern recognition is presented with a binary modified National Institute of Standards and Technology (MNIST) handwritten dataset. The necessity of an inhibitory synaptic component is analyzed in order

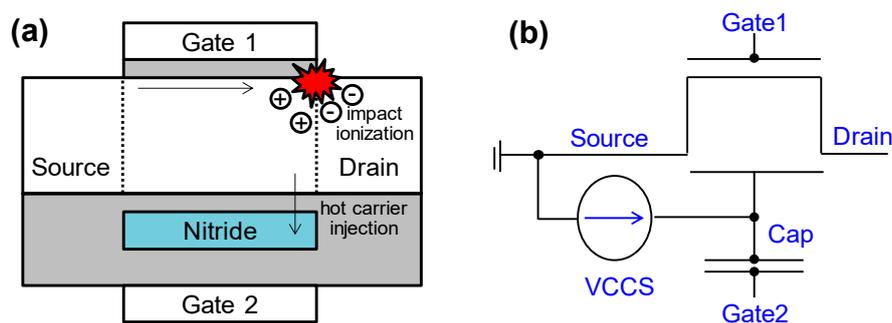
to solve an overlapping pattern issue when it comes to pattern recognition for on-chip learning of bio-inspired neuromorphic systems in the form of SNNs.

## 2. Device Model of Synaptic Transistor for System-Level Study

A schematic view of weight modulation in the synaptic transistor is illustrated in Figure 1a. As the pre-synaptic spikes are applied to the first gate (G1) and the drain, excess holes are generated by impact ionization and accumulate in the floating body region. The impact generation region expands as a result of the positive feedback between impact generation rate and accumulated holes. Afterwards, newly generated hot carriers near the second gate (G2) are injected into the nitride layer depending on the second gate voltage ( $V_{G2}$ ). The device is potentiated and depressed when holes and electrons are stored in the nitride layer because of the threshold voltage ( $V_T$ ) change. These weight modulation characteristics of the synaptic transistor are incorporated into a device model with a voltage-controlled current source (VCCS) [30] based on the gate current caused by hot carrier injection [31], as shown in Figure 1b. The VCCS delivers the second gate current ( $I_{G2}$ ) to the nitride layer, which is modeled by the gate current flowing by hot carrier injection as a function of  $V_{G2}$  as per the following equation:

$$I_{G2} \text{ represented by VCCS} = \alpha \cdot (V_{G1} - V_T)^2 \cdot V_{G2}^2 \cdot \exp(-1/V_{G2}) \quad (1)$$

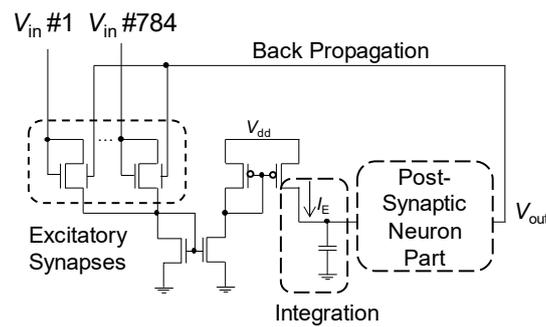
where  $\alpha$  is a fitting coefficient. The type and number of injected carriers are determined depending on  $V_{G2}$  so that the amount of  $V_T$  change ( $\Delta V_T$ ) per each pre- and post-synaptic spike is calculated, providing good agreement with the measured data [28].



**Figure 1.** (a) Schematic view of weight modulation in the synaptic transistor with dual-gate structure. (b) Device model of the synaptic transistor with a voltage-controlled current source (VCCS).

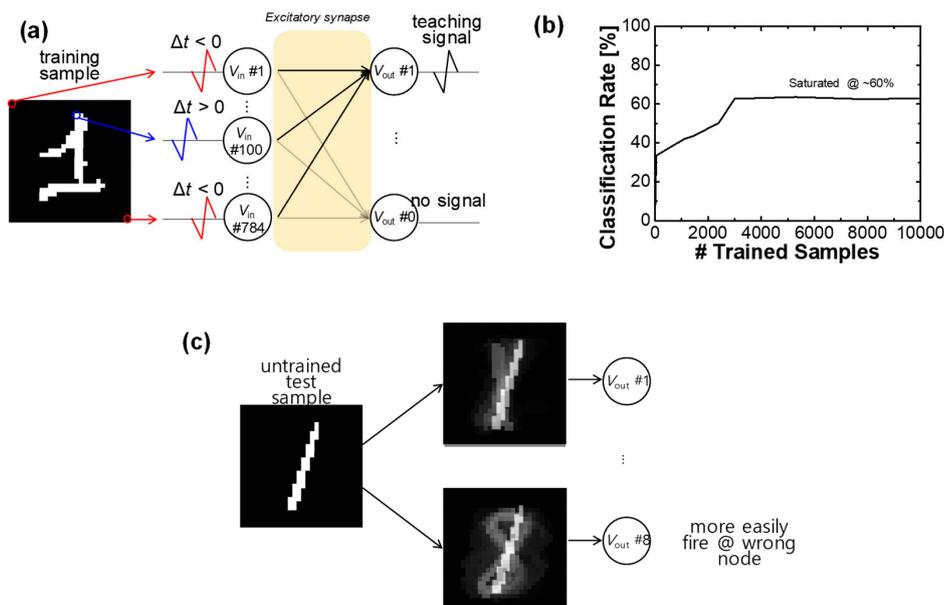
## 3. Results and Discussion

With the help of the developed device model, the performance of the SNN composed of the synaptic transistors was studied with regard to pattern recognition. A  $784 \times 10$  single-layer SNN was constructed to train and test  $28 \times 28$  binary MNIST images (60,000 training images and 10,000 testing images). A total of 784 synaptic transistors were connected to each output node as shown in Figure 2. Charges were integrated at a capacitor node while pre-synaptic spikes were applied to each synaptic transistor, and a post-synaptic neuron circuit generated post-synaptic spikes at the output node when the node voltage of the capacitor exceeded  $V_T$  of the neuron circuit [32]. The spike generation rate of each post-synaptic neuron circuit was considered as the intensity of the output node; therefore, the system was considered successful in pattern recognition when the answer node fired most among all the output nodes during test operation. The reason why recognition accuracy was calculated in this manner is that the weight sum of transferred currents ( $I_E$ ) to the output node, which is the most congruous to the testing sample, was expected to be the largest owing to the potentiated synaptic transistors in the shape of the digit, leading to high current flows.



**Figure 2.** Single-layer spiking neural network (SNN) for pattern recognition with synaptic transistors and neuron circuit.

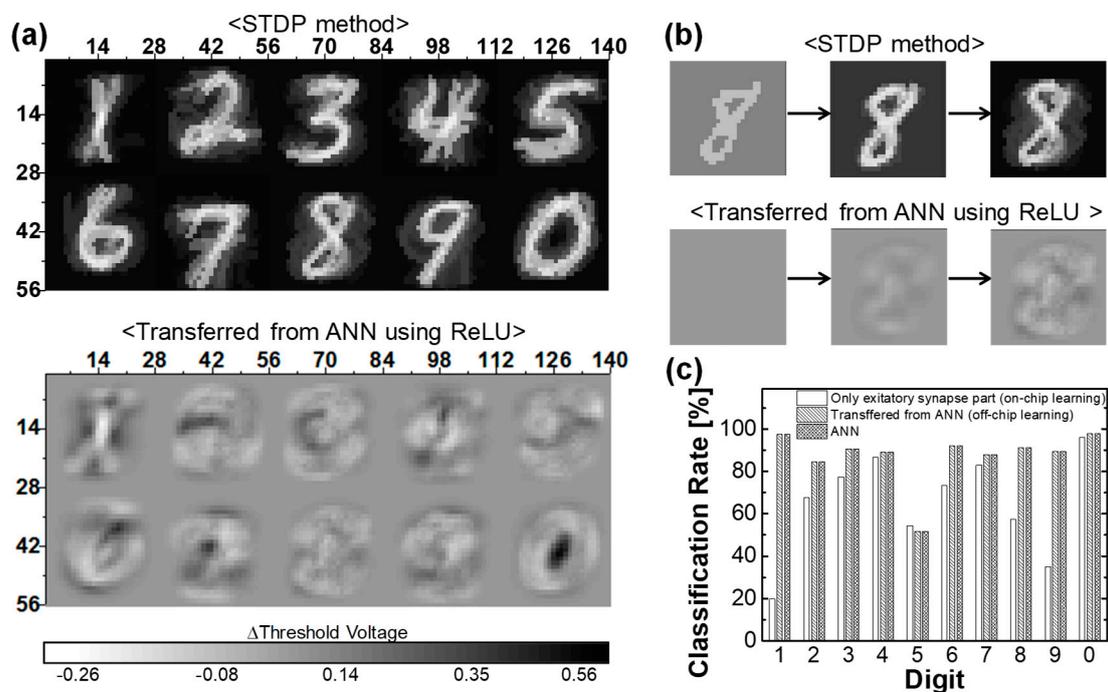
Figure 3a shows how the system was trained using the binary MNIST images and STDP characteristics. The pre-synaptic spikes were applied to the corresponding synaptic transistors with different timing depending on their colors: black with  $\Delta t = 0.5 \mu s$  and white with  $\Delta t = -0.5 \mu s$ , compared to a teaching signal which was given to the output node matching to the digit of the training sample. Therefore,  $V_T$  was increased (depression) for the synaptic transistors representing black pixels (background) and decreased (potentiation) for white pixels (handwritten digit). Figure 3b shows the classification rate of the SNN with untrained testing samples as a function of the number of trained samples.



**Figure 3.** (a) Learning method using the spike-timing dependent plasticity (STDP) rule. (b) Classification rate depending on the number of trained samples. (c) Overlapping pattern issue.

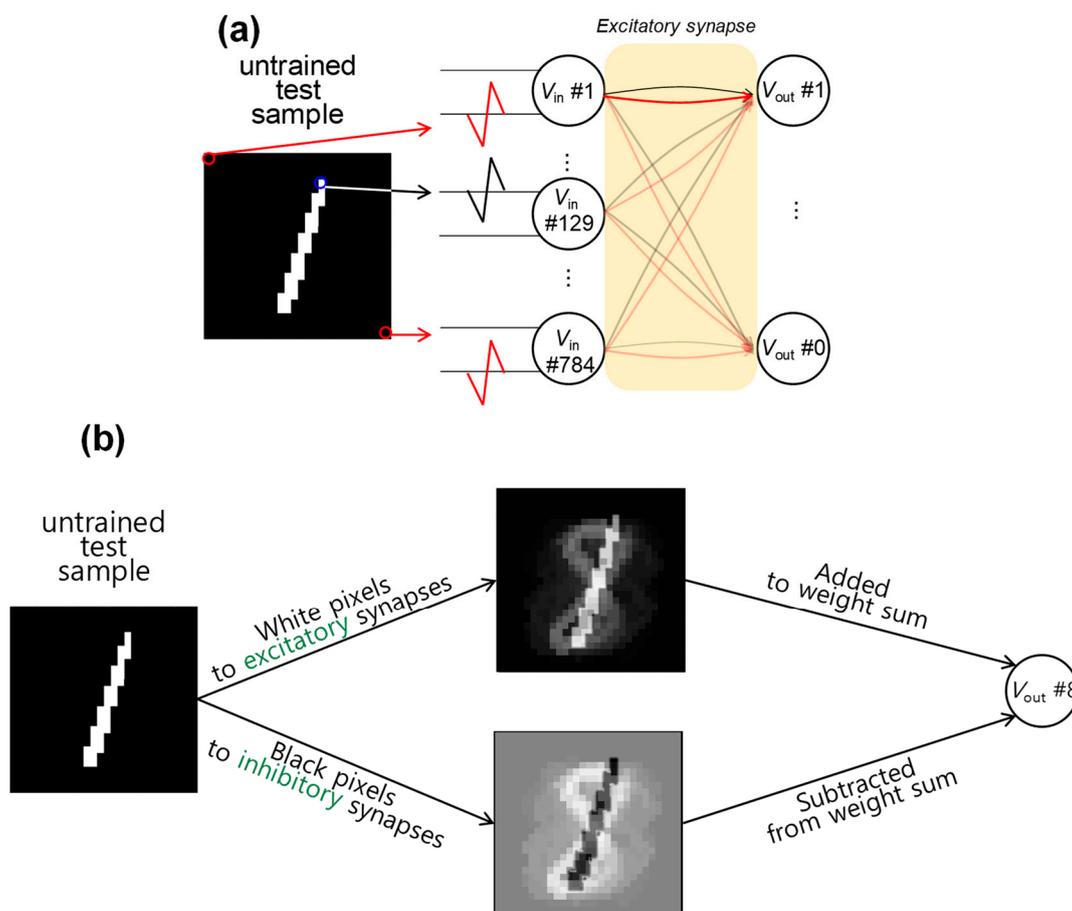
The accuracy rate became rapidly saturated due to the nonlinear weight modulation characteristics coming from the hot carrier injection model. The more electrons or holes were trapped in the nitride layer, the less likely were additional electrons or holes to be injected due to the potential inhibition by the already stored ones. The saturated accuracy rate of over 3000 trained samples was about 60%, which is quite low compared to other SNN systems because of the overlapping pattern issue. Figure 3c describes how the overlapping pattern issue degrades the classification rate. The output nodes having more white pixels in their weight maps, such as eight or zero, have a higher probability to fire, even though they do not match the digits of test samples, leading to a low recognition rate of the ones that have less white pixels (such as digit 1).

Figure 4a compares two weight maps in the form of  $\Delta V_T$  at the same scale, which were learned through the STDP rule and transferred from an artificial neural network (ANN) through off-chip learning; here, the synaptic weights of the ANN were converted to the ones of the SNN to be proportional to their square roots, so that the transferred  $I_E$  can be in line with the weight sum of the ANN with a rectified linear unit (ReLU), which is one of the most popular activation functions in ANNs because of the lack of vanishing gradients problems compared to other ones, such as sigmoid or a hyperbolic tangent [33–35]. The former looks like carving digits to the synaptic devices, whereas the latter is well characterized by the features of each digit. That is why the hardware-based SNN has a poor accuracy of 60% because its weight map does not reflect the characteristics of each digit. In the case of the STDP method, the  $V_T$  is modulated only according to whether a training sample is the answer or not, and the amount of  $V_T$  change is determined by the amount of already stored carriers in the nitride layer. However, the amount of  $V_T$  change is adjusted in the case of the ANN according to a backpropagation algorithm. Illustrated in Figure 4b are the transformation processes of the weight maps for digit 8 as the training progressed for the two cases. The carved pattern on the weight map by the STDP method becomes clearer in the direction in which it can fire frequently by digit input samples; however, the transferred weight map from the ANN exhibits its unique features in fine detail, so that all the weight maps have higher classification accuracies, which means that even a narrower memory window of the synaptic transistors can provide a higher accuracy when the weight map reflects the unique characteristics of the training images which are supposed to be classified. Figure 4c plots the classification rates for each digit depending on the methods. The poor accuracies, especially digit 1 and digit 9, have been highly improved by adopting the transferred synaptic weights, leading to 87.6% of the total accuracy. In addition, the most noteworthy thing is that the classification rates of the transferring method and the ANN itself are almost the same for every single digit. It is believed that the SNN using the transferred weight maps and the ANN with ReLU are equivalent in their operations in the respect that the intensity of the output nodes can correspond to the firing rate [36].



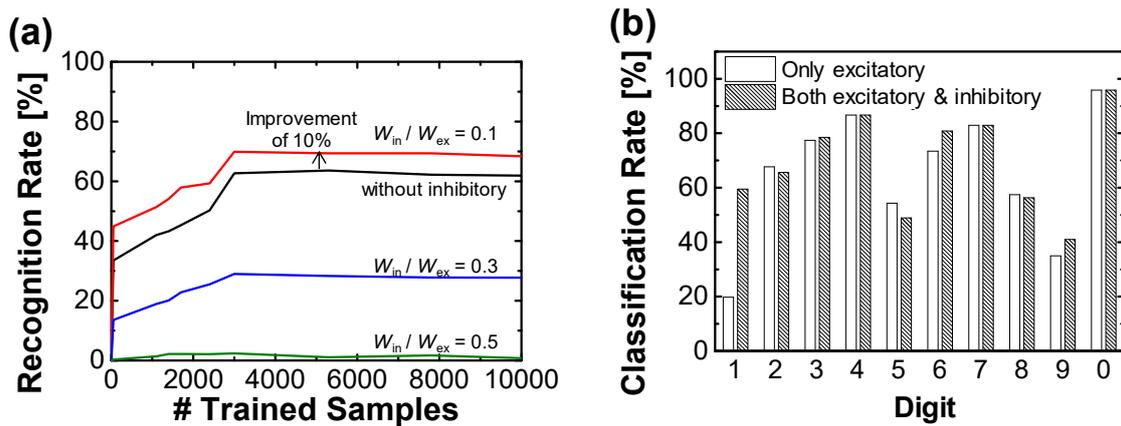
**Figure 4.** Transferred synaptic weights from artificial neural networks. (a) Comparison of the weight maps learned by the STDP method and transferred from an artificial neural network (ANN). (b) Training progress of each weight map. (c) Comparison of the classification accuracy of each digit for the STDP method, transferred synaptic weights method, and ANN.

In order to reduce the classification error caused by the overlapping pattern issue discussed above, the inhibitory synaptic devices with the same weight maps as the excitatory ones are added as shown in Figure 5a. As in the previous method, the input signals are applied to the excitatory synapses corresponding to their own pixels in the case of the white pixels; at the same time, the input signals are applied to the inhibitory synapses in the case of the black pixels. This change in the manner of classification leads to the result that if the testing samples cover not only their own digits but also other digits, the remaining parts contribute to a subtraction of the weight sum by the current flows ( $I_I$ ) through the inhibitory synaptic transistors as shown in Figure 5b. The overlapping pattern issue can be significantly solved in this way because it mainly comes from the contribution of the remaining parts to undesired firings.



**Figure 5.** (a) Illustration of the classification with the addition of inhibitory synaptic transistors. (b) How the inhibitory synaptic transistors solve the overlapping pattern issue.

Figure 6a shows the accuracies as a function of the number of training samples for various ratios between the channel widths of the excitatory synaptic transistors ( $W_{ex}$ ) and the inhibitory ones ( $W_{in}$ ). The accuracy is improved by 10% at  $W_{in}/W_{ex} = 0.1$ ; however, it starts decreasing after that and reaches the bottom (nearly 0% instead of 10%) when  $W_{in}/W_{ex} = 0.5$ . This is because the output nodes cannot fire when  $W_{in}$  is too wide. The number of black pixels is larger than that of white pixels and  $I_I$  is higher than  $I_E$  in most testing samples when  $W_{in}/W_{ex}$  exceeds 0.5. Figure 6b compares the classification rate of each digit for those two SNN systems. It is noteworthy that the accuracies of the digits which have a small number of white pixels, such as one, is significantly enhanced from 19 to 60%, while the accuracies of other digits maintain similar values. It is confirmed that the addition of an inhibitory synapse part can effectively solve the misclassified cases stemming from the overlapping pattern problem.



**Figure 6.** (a) Classification rates after adding the inhibitory synapse part. (b) Classification accuracy of each digit depending on the learning method and system structure.

#### 4. Conclusions

In conclusion, we presented a system-level study regarding pattern recognition with the help of a device model. The device model was developed with a VCCS based on measured data and gate current by hot carrier injection. A total of three SNN systems were constructed and analyzed using binary MNIST images. A SNN with only the excitatory synaptic transistors trained under the STDP rule had a poor classification rate with 60% of the total accuracy because of the pattern overlapping issue. This dramatically improved to 87.6% in the case of a SNN with transferred synaptic weights from an ANN using ReLU. The difference between those two systems was whether the region representing the unique features of each digit was potentiated or the handwritten digit region was just carved. The addition of inhibitory synaptic transistors with the same weight maps improved the classification accuracy by 10% by solving the overlapping pattern problem, which comes from the fact that the output nodes having more white pixels tend to fire to unmatched training samples. These results lead us to conclude that these SNN systems and learning methods provide a framework for future studies about hardware-based neuromorphic systems using both excitatory and inhibitory synaptic devices for pattern recognition applications.

**Author Contributions:** H.K. and B.-G.P. conceived the device structure and modeling and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the NRF funded by the Korean government under Grant 2019M3F3A1A03079821, 2019R111A3A01061262, 2019R1A6C1030008 and in part by the Nano-Material Technology Development Program through the NRF funded by the Ministry of Science, ICT and Future Planning (2016M3A7B4910348).

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

- Backus, J. Can programming be liberated from the von Neumann style?: A functional style and its algebra of programs. *Commun. ACM* **1978**, *21*, 613–641. [[CrossRef](#)]
- Jeong, D.S.; Kim, K.M.; Kim, S.; Choi, B.J.; Hwang, C.S. Memristors for Energy-Efficient New Computing Paradigms. *Adv. Electron. Mater.* **2016**, *2*, 160090. [[CrossRef](#)]
- Merolla, P.A.; Arthur, J.V.; Alvarez-Icaza, R.; Cassidy, A.S.; Sawada, J.; Akopyan, F.; Jackson, B.L.; Imam, N.; Guo, C.; Nakamura, Y.; et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **2014**, *345*, 668–673. [[CrossRef](#)] [[PubMed](#)]
- Misra, J.; Saha, I. Artificial neural networks in hardware: A survey of two decades of progress. *Neurocomputing* **2010**, *74*, 239–255. [[CrossRef](#)]

5. Maass, W. Networks of spiking neurons: The third generation of neural network models. *Neural Netw.* **1997**, *10*, 1659–1671. [[CrossRef](#)]
6. Cohen, E.; Malka, D.; Shemer, A.; Shahmoon, A.; Zalevsky, Z.; London, M. Neural networks within multi-core optic fibers. *Sci. Rep.* **2016**, *6*, 29080. [[CrossRef](#)] [[PubMed](#)]
7. Hwang, S.; Kim, H.; Park, J.; Kwon, M.-W.; Baek, M.-H.; Lee, J.-J.; Park, B.-G. System-level simulation of hardware spiking neural network based on synaptic transistors and I&F neuron circuits. *IEEE Electron Device Lett.* **2018**, *39*, 1441–1444.
8. Seo, M.; Kang, M.-H.; Jeon, S.-B.; Bae, H.; Hur, J.; Jang, B.C.; Yun, S.; Cho, S.; Kim, W.-K.; Kim, M.-S.; et al. First demonstration of a logic-process compatible junctionless ferroelectric finfet synapse for neuromorphic applications. *IEEE Electron Device Lett.* **2018**, *39*, 1445–1448. [[CrossRef](#)]
9. Park, Y.J.; Kwon, H.T.; Kim, B.; Lee, W.J.; Wee, D.H.; Choi, H.-S.; Park, B.-G.; Lee, J.-H.; Kim, Y. 3-D stacked synapse array based on charge-trap flash memory for implementation of deep neural networks. *IEEE Trans. Electron Devices* **2018**, *66*, 420–427. [[CrossRef](#)]
10. Sung, C.; Lim, S.; Kim, H.; Kim, T.; Moon, K.; Song, J.; Kim, J.-J.; Hwang, H. Effect of conductance linearity and multi-level cell characteristics of TaO<sub>x</sub>-based synapse device on pattern recognition accuracy of neuromorphic system. *Nanotechnology* **2018**, *29*, 115203. [[CrossRef](#)]
11. Kim, H.; Hwang, S.; Park, J.; Yun, S.; Lee, J.-H.; Park, B.-G. Spiking neural network using synaptic transistors and neuron circuits for pattern recognition with noisy images. *IEEE Electron Device Lett.* **2018**, *39*, 630–633. [[CrossRef](#)]
12. Shabairou, N.; Cohen, E.; Wagner, O.; Malka, D.; Zalevsky, Z. Color image identification and reconstruction using artificial neural networks on multimode fiber images: Towards an all-optical design. *Opt. Lett.* **2018**, *43*, 5603–5606. [[CrossRef](#)] [[PubMed](#)]
13. Truong, S.N.; Ham, S.-J.; Min, K.-S. Neuromorphic crossbar circuit with nanoscale filamentary-switching binary memristors for speech recognition. *Nanoscale Res. Lett.* **2014**, *9*, 629. [[CrossRef](#)] [[PubMed](#)]
14. Prezioso, M.; Merrih-Bayat, F.; Hoskins, B.; Adam, G.; Likharev, K.K.; Strukov, D.B. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **2015**, *521*, 61–64. [[CrossRef](#)]
15. Ambrogio, S.; Balatti, S.; Milo, V.; Carboni, R.; Wang, Z.-Q.; Calderoni, A.; Ramaswamy, N.; Ielmini, D. Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM. *IEEE Trans. Electron Devices* **2016**, *63*, 1508–1515. [[CrossRef](#)]
16. Park, J.; Kwak, M.; Moon, K.; Woo, J.; Lee, D.; Hwang, H. TiO<sub>x</sub>-based RRAM synapse with 64-levels of conductance and symmetric conductance change by adopting a hybrid pulse scheme for neuromorphic computing. *IEEE Electron Device Lett.* **2016**, *37*, 1559–1562. [[CrossRef](#)]
17. Kim, S.; Kim, H.; Hwang, S.; Kim, M.-H.; Chang, Y.-F.; Park, B.-G. Analog synaptic behavior of a silicon nitride memristor. *ACS Appl. Mater. Interfaces* **2017**, *9*, 40420–40427. [[CrossRef](#)]
18. Kim, S.; Ishii, M.; Lewis, S.; Perri, T.; BrightSky, M.; Kim, W.; Jordan, R.; Burr, G.W.; Sosa, N.; Ray, A.; et al. NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning. In Proceedings of the 2012 International Electron Devices Meeting, San Francisco, CA, USA, 7–9 December 2012; pp. 443–446.
19. Tuma, T.; Le-Gallo, M.; Sebastian, A.; Eleftheriou, E. detecting correlations using phase-change neurons and synapses. *IEEE Electron Device Lett.* **2016**, *37*, 1238–1241. [[CrossRef](#)]
20. Kuzum, D.; Jeyasingh, R.G.; Lee, B.; Wong, H.-S.P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.* **2011**, *12*, 2179–2186. [[CrossRef](#)]
21. Oh, S.; Kim, T.; Kwak, M.; Song, J.; Woo, J.; Jeon, S.; Yoo, I.K.; Hwang, H. HfZrO<sub>x</sub>-based ferroelectric synapse device with 32 levels of conductance states for neuromorphic applications. *IEEE Electron Device Lett.* **2017**, *38*, 732–735. [[CrossRef](#)]
22. Mulaosmanovic, H.; Ocker, J.; Müller, S.; Noack, M.; Müller, J.; Polakowski, P.; Mikolajick, T.; Slesazeck, S. Novel ferroelectric FET based synapse for neuromorphic systems. In Proceedings of the 2017 Symposium on VLSI Technology, Kyoto, Japan, 5–8 June 2017; pp. T176–T177.
23. Wang, J.; Li, Y.; Liang, R.; Zhang, Y.; Mao, W.; Yang, Y.; Ren, T.-L. Synaptic computation demonstrated in a two-synapse network based on top-gate electric-double-layer synaptic transistors. *IEEE Electron Device Lett.* **2017**, *38*, 1496–1499. [[CrossRef](#)]
24. Wan, X.; Yang, Y.; Feng, P.; Shi, Y.; Wan, Q. Short-term plasticity and synaptic filtering emulated in electrolyte-gated IGZO transistors. *IEEE Electron Device Lett.* **2016**, *37*, 299–302. [[CrossRef](#)]

25. Shi, J.; Ha, S.D.; Zhou, Y.; Schoofs, F.; Ramanathan, S. A correlated nickelate synaptic transistor. *Nat. Commun.* **2013**, *4*, 2676. [[CrossRef](#)] [[PubMed](#)]
26. Kim, H.; Park, J.; Kwon, M.-W.; Lee, J.-H.; Park, B.-G. Silicon-based floating-body synaptic transistor with frequency dependent short-and long-term memories. *IEEE Electron Device Lett.* **2016**, *37*, 249–252. [[CrossRef](#)]
27. Kim, H.; Cho, S.; Sun, M.-C.; Park, J.; Hwang, S.; Park, B.-G. Simulation study on silicon-based floating body synaptic transistor with short-and long-term memory functions and its spike timing-dependent plasticity. *J. Semicond. Technol. Sci.* **2016**, *16*, 657–663. [[CrossRef](#)]
28. Kim, H.; Hwang, S.; Park, J.; Park, B.-G. Silicon synaptic transistor for hardware-based spiking neural network and neuromorphic system. *Nanotechnology* **2017**, *28*, 405202. [[CrossRef](#)]
29. Kim, H.; Sun, M.-C.; Hwang, S.; Kim, H.-M.; Lee, J.-H.; Park, B.-G. Fabrication of asymmetric independent dual-gate FinFET using sidewall spacer patterning and CMP processes. *Microelectron. Eng.* **2018**, *185*, 29–34. [[CrossRef](#)]
30. Kim, S.; Lee, S.-H.; Kim, Y.-G.; Cho, S.; Park, B.-G. Highly compact and accurate circuit-level macro modeling of gate-all-around charge-trap flash memory. *Jpn. J. Appl. Phys.* **2016**, *56*, 014302. [[CrossRef](#)]
31. Sonoda, K.; Tanizawa, M.; Shimizu, S.; Araki, Y.; Kawai, S.; Ogura, T.; Kobayashi, S.; Ishikawa, K.; Eimori, T.; Inoue, Y.; et al. Compact modeling of a flash memory cell including substrate-bias-dependent hot-electron gate current. *IEEE Trans. Electron Devices* **2004**, *51*, 1726–1733. [[CrossRef](#)]
32. Park, J.; Kwon, M.-W.; Kim, H.; Hwang, S.; Lee, J.-J.; Park, B.-G. Compact neuromorphic system with four-terminal si-based synaptic devices for spiking neural networks. *IEEE Trans. Electron Devices* **2017**, *64*, 2438–2444. [[CrossRef](#)]
33. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
34. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
35. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
36. Diehl, P.U.; Neil, D.; Binas, J.; Cook, M.; Liu, S.-C.; Pfeiffer, M. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In Proceedings of the 2015 International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; pp. 1–8.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).