IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 9, 2020, accepted November 2, 2020, date of publication November 5, 2020, date of current version November 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036088

Low-Power and High-Density Neuron Device for Simultaneous Processing of Excitatory and Inhibitory Signals in Neuromorphic Systems

SUNG YUN WOO[1], (Member, IEEE), DONGSEOK KWON[1], (Graduate Student Member, IEEE), NAGYONG CHOI[1], (Member, IEEE), WON-MOOK KANG[1], YOUNG-TAK SEO[1], (Graduate Student Member, IEEE), MIN-KYU PARK[1], JONG-HO BAE[2], (Member, IEEE), BYUNG-GOOK PARK[1], (Fellow, IEEE), AND JONG-HO LEE[1], (Fellow, IEEE)

[1]Department of ECE and ISRC, Seoul National University, Seoul 08826, South Korea
[2]Department of Electrical Engineering, Kookmin University, Seoul 02707, South Korea

Corresponding author: Jong-Ho Lee (jhl@snu.ac.kr)

This work was supported in part by the National Research Foundation of Korea under Grant NRF-2016M3A7B4909604, and in part by the Brain Korea 21 plus in 2020.

ABSTRACT A positive-feedback (PF) neuron device capable of threshold tuning and simultaneously processing excitatory (G^+) and inhibitory (G^-) signals is experimentally demonstrated to replace conventional neuron circuits, for the first time. Thanks to the PF operation, the PF neuron device with steep switching characteristics can implement integrate-and-fire (IF) function of neurons with low-energy consumption. The structure of the PF neuron device efficiently merges a gated PNPN diode and a single MOSFET. Integrate-and-fire (IF) operation with steep subthreshold swing ($SS < 1$ mV/dec) is experimentally implemented by carriers accumulated in an n floating body of the PF neuron device. The carriers accumulated in the n floating body are discharged by an inhibitory signal applied to the merged FET. Moreover, the threshold voltage (V_{th}) of the proposed PF neuron is controlled by using a charge storage layer. The low-energy consuming PF neuron circuit (∼0.62 pJ/spike) consists of one PF device and only five MOSFETs for the IF and reset operation. In a high-level system simulation, a deep-spiking neural network (D-SNN) based on PF neurons with four hidden layers (1024 neurons in each layer) shows high-accuracy (98.55%) during a MNIST classification task. The PF neuron device provides a viable solution for high-density and low-energy neuromorphic systems.

INDEX TERMS Neuron device, positive-feedback (PF) device, hardware-based neural networks, semiconductor device reliability, silicon-on-insulator (SOI) technology.

I. INTRODUCTION

Recently, hardware-based neural networks (HNNs) have emerged for use in neuromorphic systems to compute complex data efficiently [1]–[3]. For high performance in HNNs, various synaptic arrays and neuron circuits suitable for efficient architectures and learning algorithms have been researched [4]–[8]. Specifically, both excitatory (G^+) and inhibitory (G^-) synaptic arrays are important to improve the accuracy of HNNs [9]–[11]. Neuron circuits that use large capacitors (≥0.1 pF) and many transistors (≥11 MOSFETs) to process simultaneously signals from these two types of synapses have been reported [11]–[13], resulting in increased power consumption and a larger area. Note that processing these signals simultaneously can reduce memory usage and simplify the peripheral circuitry. To alleviate these hardware burdens, memristor-based [14]–[16] and FET-based neuron devices [18]–[23] with memory functionalities have been studied to mimic neurons. Memristor-based neuron devices with two terminals replace membrane capacitors in neuron

The associate editor coordinating the review of this manuscript and approving it for publication was Chaitanya U. Kshirsagar.

circuits and have the advantage of high density over membrane capacitor and FET-based neuronal devices. Neuron circuits require infinite endurance to generate spikes during integrate-and-fire operation. Existing memristor-based neuron devices are being studied to improve endurance during integrate-and-fire operations [16], [17]. In addition, memristor-based neuron devices require additional circuits such as a differential amplifier to compare the resistance of the memristor to a reference resistance and a circuit so as to reset the memristor after the integrate-and-fire operation. On the other hand, FET-based neuron devices, capable of resolving these issues, can process signals from two types (G^+ and G^-) of synapses sequentially or only one type of signals [18]–[23]. However, FET-based neuron devices processing only one type of signals are paired for excitatory and inhibitory signals, and require additional circuits for logic operation in the neural networks. In neural networks based on neuron devices that process signals from two types of synapses sequentially, the learning and computing speeds can also be slower than conventional neuron circuits. It is also difficult or impossible for the reported neuron devices to tune the threshold voltage to reduce threshold fluctuations of the neuron devices.

In this work, we propose a neuron device capable of steep switching with positive-feedback (PF) and investigate the integrate-and-fire (IF) function of the PF neuron device during the simultaneous processing of excitatory and inhibitory signals. The PF neuron device can implement the IF function of neurons by accumulating (or discharging) electrons into an n floating body by excitatory (or inhibitory) signal and replace a large membrane capacitor in conventional neuron circuits. Moreover, the threshold voltage (V_{th}) of the PF neuron device with a charge storage layer is changed by program and erase state, which then changes the threshold of neurons in neural networks. The threshold tuning ability of the PF neuron device with the charge storage layer is examined to alleviate the degradation of the recognition rate caused by device variations in neural networks using a python simulator.

II. METHODS

A. FABRICATION AND IF OPERATION OF THE PF NEURON DEVICE

Fig. 1 shows 3-D schematic and top views of the PF neuron device and a TEM image of the fabricated PF neuron device. The PF neuron device has a structure that efficiently merges a gated PNPN diode and one FET. The PF neuron device consists of three gates (G1, G2 and G3), an anode, a cathode and a drain. G1 and G2 receive excitatory and inhibitory signals, respectively. By applying program or erase pulse to G3 with the charge storage layer, the V_{th} of the PF neuron device can be controlled. The thicknesses of the gate oxide (T_{ox}) and the $SiO_2/Si_3N_4/SiO_2$ stack are 10 nm and 3/6/9 nm, respectively. The channel width (W), p length (L_p), n length (L_n), and the thickness of the Si body (T_{Si}) are 1 μm, 0.7 μm, 1.1 μm, and 100 nm, respectively.

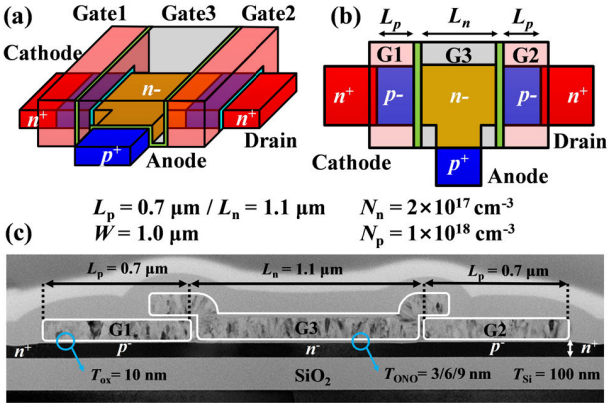

FIGURE 1. (a) 3-D schematic and (b) top views of the neuron device. (c) Cross-sectional TEM image of the fabricated neuron device.

Doping concentrations of p/n floating body (p/n regions) are 1×10^{18} cm^{-3} and 2×10^{17} cm^{-3}, respectively. Fig. 2 shows schematic cross-sectional views of the key fabrication process steps of the proposed PF neuron device. The PF neuron device was fabricated on a silicon on insulator (SOI) wafer. The Si film was patterned as an active region, and boron and phosphorus ions were implanted for the p/n regions, respectively. A SiO_2 layer of 10 nm was deposited as the gate oxide. An n-doped poly Si was deposited and patterned for the gates (G1 and G2) that receive the signals transmitted from the excitatory and inhibitory synaptic devices, respectively. For the memory function, a $SiO_2/Si_3N_4/SiO_2$ stack containing a 6 nm thick Si_3N_4 charge storage layer was deposited. The n^+ poly Si was deposited and patterned on an n region for G3. Boron and arsenic ions were implanted, followed by rapid temperature processing for activation of the anode and cathode. Finally, the back-end process was conducted. Fig. 3 (a) shows an energy band diagram cut along the anode from the cathode in the PF neuron device to explain the PF mechanism. As V_{G1} increases, the electron-injection barrier (V_{e1}) decreases (①) and electrons from the n^+ cathode go into the n region (②). The height of the hole-injection barrier (V_h) in the n region decreases due to the injected electrons (③). Then, holes from the p^+ anode are easily injected into the p region, which further decreases the height of the electron-injection barrier (④). The repeated positive feedback operation enables steep switching characteristics of

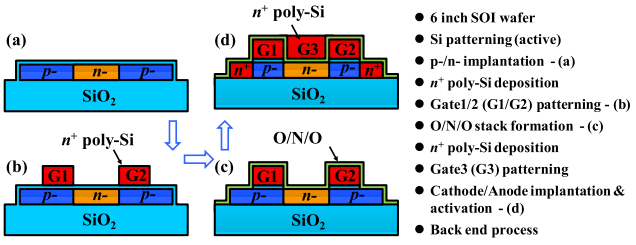

FIGURE 2. Schematic cross-sectional views of the key fabrication process steps of the proposed neuron device.

the PF neuron device. Charges are accumulated in the *n* region by V_{G1} from the excitatory synapse during the processes (②) and (③). On the other hand, electrons accumulated in the *n* region can escape to the drain when the energy barrier (V_{e2}) decreases depending on the V_{G2} voltage, as shown in Fig. 3 (b), which means that the charges are discharged from the neuron by the inhibitory synapses. Note that during the inhibitory operation, the *n* region acts as the source. By charging and discharging electrons in the *n* region, the IF operation of the PF neuron device can simultaneously process the excitatory and inhibitory signals. When the PF neuron device turns off, the electrons accumulated in the *n* region are reduced over time by recombination, which shows a leaky integration. The reduction is related to the retention time, which can be controlled by G3 over the long-term [21], [24]. The V_h is modulated by V_{G3} as shown in Fig. 3 (c), which allows control of the V_{th} of the PF neuron device. Additionally, the V_h can be controlled by the amount of charges in the charge storage layer (Si_3N_4).

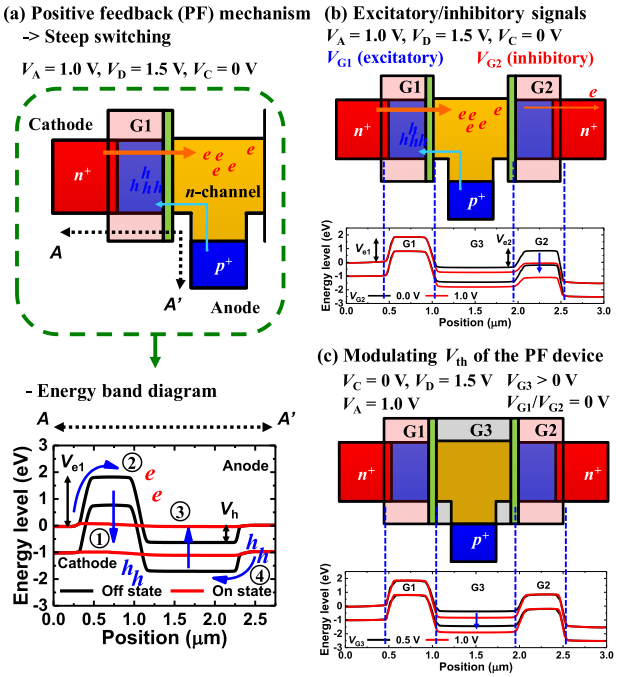

FIGURE 3. Simulated energy band diagrams of the positive feedback (PF) device along the channel direction: (a), (b) IF operation by processing excitatory and inhibitory signals simultaneously. (c) V_{G3} controls the potential well under G3, modulating V_{th} of the neuron device.

III. RESULTS AND DISCUSSION

A. MEASUREMENT OF THE PF NEURON DEVICE FOR IF OPERATION WITH EXCITATORY AND INHIBITORY SIGNALS

Fig. 4 shows the measured I_A-V_{G1} and I_D-V_{G1} curves of the PF neuron device as a parameter of V_{G2}. As V_{G1} increases at a V_{G2} of 0 V, I_A increases sharply by the PF operation, and the subthreshold swing (*SS*) is very steep (<1 mV/dec) as shown in Fig. 4 (a). As V_{G1} increases after the PF operation, I_A remains constant because of diode current flowing from

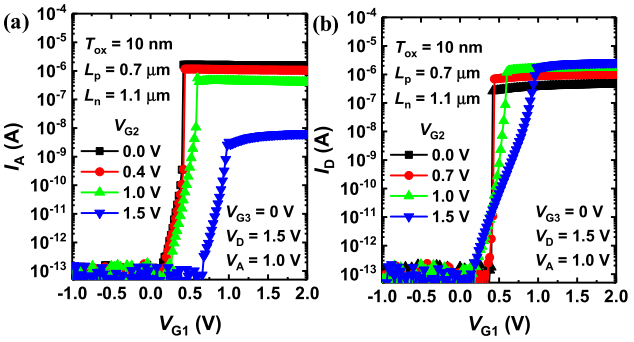

FIGURE 4. (a) Measured I_A-V_{G1} and (b) I_D-V_{G1} curves of the PF neuron device as a parameter of V_{G2}, respectively.

the anode to the cathode region in the PF neuron device at a fixed V_A. As V_{G2} increases, I_D is higher than I_A just before the PF action of electrons and holes in the PF neuron device occurs. The V_{th} of the PF neuron device increases, preventing electrons from accumulating in the *n* region. Fig. 5 shows the measured I_A-V_{G2} and I_D-V_{G2} curves of the PF neuron device as a parameter of V_{G1}. Although V_{G2} increases, I_A and I_D are in an off state at a V_{G1} of 0 V. As V_{G2} increases at a V_{G1} of 0.4 V, I_A decreases and I_D increases because electrons accumulated in the *n* region are discharged to the drain. When I_A and I_D are an on state at a V_{G1} of 0.5 V, the energy band of the PF neuron device is nearly flat due to the PF operation. As V_{G2} increases, electrons accumulated in the *n* region flow to the drain. As a result, the V_h increases, and I_A decreases rapidly by suppressing the PF operation. At the same time, a large amount of current flows instantaneously to the drain. Then, I_D decreases by the increased V_{e1}, which shows a negative resistance at a V_{G1} of 0.5 V as shown in Fig. 5 (b). When the V_D of 1.5 V is higher than the V_A of 1 V, reverse bias is applied between the p^+ anode and n^+ drain. Thus, current cannot flow from the anode to the drain. Fig. 6 (a) and (b) show the measured I_A-V_{G1} curves of the PF neuron device as parameters of V_{G3} and the program/erase (PGM/ERS) operation, respectively. As V_{G3} increases, the hole-injection barrier increases in the *n* region, and the V_{th} of the PF neuron device increases as shown in Fig. 6 (a). When V_{PGM}

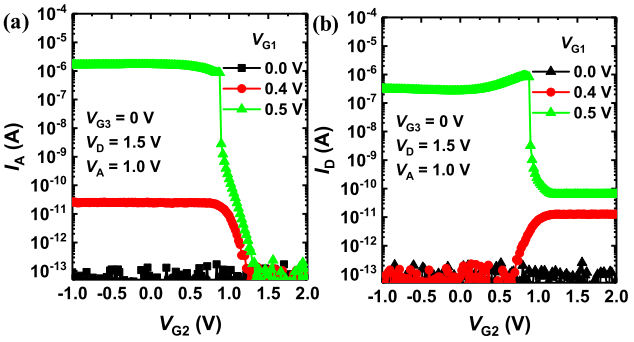

FIGURE 5. (a) Measured I_A-V_{G2} and (b) I_D-V_{G2} curves of the PF neuron device as a parameter of V_{G1}, respectively.

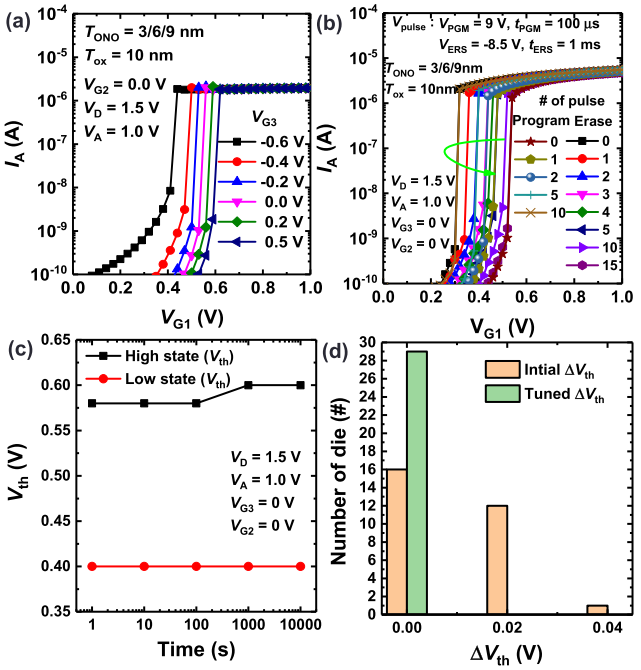

FIGURE 6. (a) Measured I_A-V_{G1} curves of the PF neuron device as a parameter of V_{G3}. (b) Measured I_A-V_{G1} curves of the PF neuron device in program and erase states. (c) V_{th} retention of the PF neuron device in high and low V_{th} states. (d) The number of dies showing the V_{th} difference (ΔV_{th}) between two PF neuron devices on the same die.

(9 V, 100 μs) is applied to the G3 to store electrons in the charge storage layer, the V_{th} of the PF neuron device decreases. Conversely, the V_{th} of the PF neuron device increases by applying V_{ERS} (−8.5 V, 1 ms) to the G3 as shown in Fig. 6 (b). Fig. 6 (c) shows the V_{th} retention characteristic of the PF neuron device in PGM (solid circle symbols) and ERS (solid square symbols) states. The PF neuron device maintains the nonvolatile of two V_{th}s. The G variation of synaptic devices and the V_{th} variation of neuron circuits can affect the spike rate of target neurons, which can degrade the performance of HNNs [25], [26]. Fig. 6 (d) presents the number of dies, showing the difference in the V_{th} of two PF neuron devices on the same die. The V_{th} difference of the two PF neuron devices on the same die is less than 0.04 V. By adjusting the V_{th} of the PF neuron by means of a V_{G3} control strategy or PGM/ERS operation, the V_{th} variation of neurons can be reduced and the accuracy of the spike rate of the target neuron can be improved. Also, it is possible to mimic the homeostasis function of biological neurons, which is essential in spiking neural networks (SNNs) based on spike-timing-dependent-plasticity (STDP) to improve performances [25], [26].

Fig. 7 shows the measured (a) step pulse and (b) pulse transients of the PF neuron device as a parameter of V_{G1}. As V_{G1} increases, the amount of electrons that accumulate in the n region increases. As a result, the turn-on time (t_{on}) is shorter and the PF neuron device fires more rapidly as shown in Fig. 7 (a). In the pulse transients of the PF neuron device,

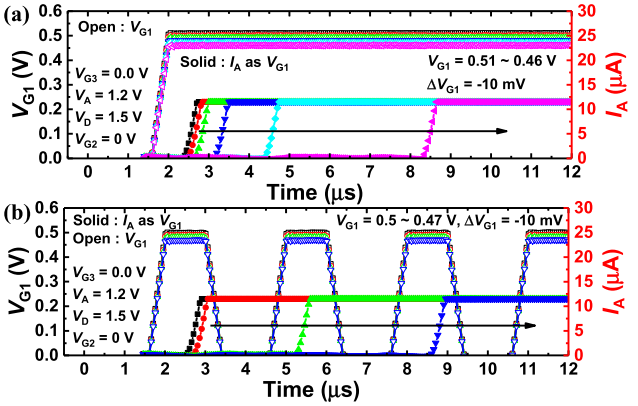

FIGURE 7. Measured I-t plots of the PF neuron device for IF function as a parameter of V_{G1}. (a) step and (b) pulse transient. t_{pulse}, t_{rise} and t_{fall} are 1 μs, 500 ns and 500 ns, respectively.

the pulse width (t_{pulse}), rise time (t_{rise}) and fall time (t_{fall}) applied to G1 are 1 μs, 500 ns and 500 ns, respectively. As V_{G1} increases, the PF neuron device fires upon fewer pulses of V_{G1}. Fig. 8 (a) and (b) show the measured I-t plot of the PF neuron device as parameters of V_{G2} and V_{G3}, respectively. As V_{G2} increases, the amount of electrons that escape from the n region to the drain increases, resulting in a longer t_{on} as shown in Fig. 8 (a). Increasing V_{G3} positively deepens the potential well (hole-injection barrier in the PF neuron device), resulting in a longer t_{on} because more electrons should be accumulated in the n region for fire as shown in Fig. 8 (b). Thus, the amount of accumulated electrons depends on t_{on} at a fixed V_{G1}, and t_{on} becomes longer as V_{G3} increases. Fig. 9 shows the t_{on}s of the PF neuron device for the IF function as parameters of (a) V_{G1} and (b) V_{G2}, respectively. Though a high V_G is applied to the PF neuron device, the t_{on} of the PF neuron device can be delayed due to parasitic capacitors of metal pads and measuring equipment. [27]. In this case, the t_{on}s of the PF neuron device in Fig. 9 was excluded by the delay time (about 500 ns). As V_{G1} (excitatory) increases, the t_{on} becomes exponentially

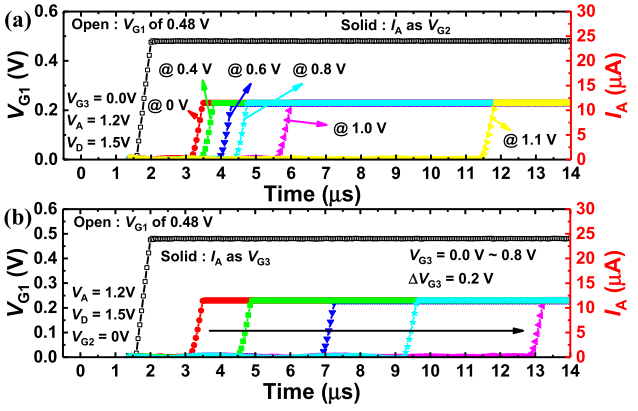

FIGURE 8. Measured I-t plots of the PF neuron device for inhibitory synapse and controllability of V_{th} as parameters of (a) V_{G2} and (b) V_{G3}, respectively.

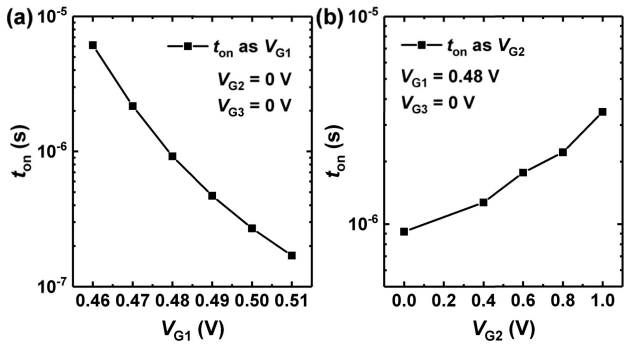

FIGURE 9. Measured turn-on time (t_{on}) of the PF neuron device for integrate-and-fire function as parameters of (a) V_{G1} and (b) V_{G2}, respectively.

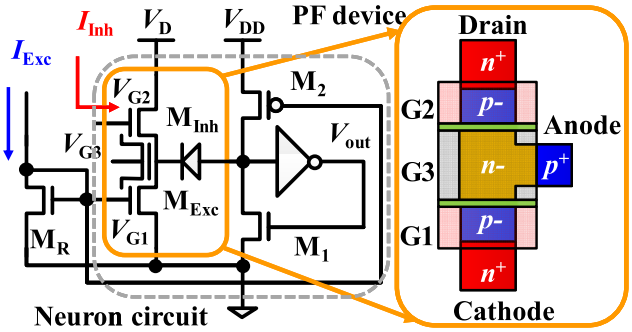

FIGURE 10. A neuron circuit consisting of the proposed PF neuron device, *p*/*n*MOSFETs (M_1 and M_2), and one invertor. Spikes from excitatory and inhibitory synapses are applied to G1 and G2, respectively.

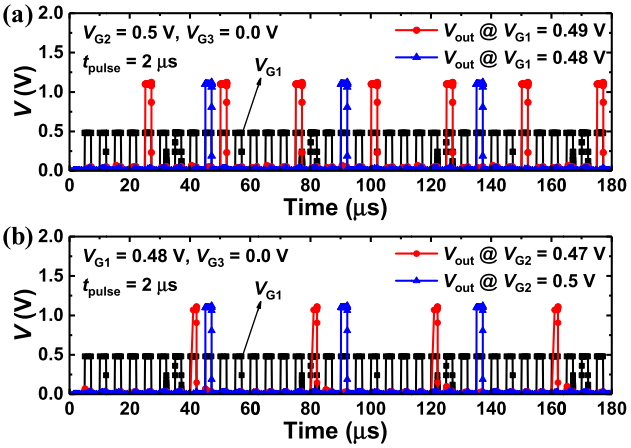

FIGURE 11. Simulated *V*-*t* plot of the PF neuron circuit for IF behavior as parameters of (a) V_{G1} and (b) V_{G2}. Here, V_D and V_{DD} are 1.5 V and 1.1 V, respectively.

short. As V_{G1} (excitatory) increases, the amount of electrons accumulated in the *n* region increases exponentially, causing the t_{on} to short exponentially. Conversely, as V_{G2} (inhibitory) increases, the amount of electrons accumulated in the *n* region decreases exponentially, causing the t_{on} to long exponentially.

B. NEURON CIRCUIT BASED ON PF NEURON DEVICE FOR HNNs

Fig. 10 shows a neuron circuit consisting of the proposed PF neuron device, *p*/*n*MOSFETs, and one invertor. The PF neuron device is represented by two merged *n*MOSFET (M_{Exc} and M_{Inh}) and one diode. Current pulses (I_{Exc} and I_{Inh}) from excitatory and inhibitory synaptic arrays are reflected in G1 and G2, respectively. Note that I_{Exc} is linear with a current flowing through M_{Exc} because M_R and M_{Exc} have the same V_{th} and operate as a current mirror. Just after the PF neuron device fires, V_A becomes 0 V by the invertor and M_1. Electrons in the *n* region are discharged, and the PF neuron device becomes initial state. When V_{G1} is 0 V, V_A becomes high state by the M_2. Also, a current flowing through the M_{Inh} from the drain is determined by I_{Inh}. When I_{Inh} is reflected in G2, electrons integrated in the *n* region are drained by M_{Inh}. Fig. 11 shows simulated *V*-*t* plots of the PF neuron circuit as parameters of (a) V_{G1} and (b) V_{G2}. The neuron circuit based on the PF neuron device is simulated using a mixed-mode option in a TCAD simulator (Sentaurus of Synopsys). In the TCAD simulation, IF and reset operations in the PF neuron circuit are verified. V_D and V_{DD} are 1.5 V and 1.1 V, respectively. As the amplitude of the V_{G1} pulse increases, the spike rate of the PF neuron circuit increases at a fixed V_{G2} of 0.5 V as shown in Fig. 11 (a). As V_{G2} decreases, the firing rate of the PF neuron circuit increases at a fixed amplitude (0.48 V) of the V_{G1} pulse. Fig. 12 shows circuit diagrams and simulated *I*-*t* plots for a comparison of the energy consumption (J/spike) between a conventional neuron circuit and the PF neuron circuit [13]. The conventional neuron circuit consists of capacitors (membrane and refractory) and minimum number of FETs to implement the IF function of neurons. M_1 is used to fully discharge the input node of the invertor, and M_{RESET} is used to reset the membrane potential.

In the conventional neuron circuit, the current supplied from synaptic devices flows into the membrane capacitor (C_{mem}), resulting in a change in a membrane potential (V_{mem}). When the V_{mem} exceeds the V_{th} of the neuron circuit, the neuron circuit fires and generates an output spike. A leakage current flows in the neuron circuit by the relatively slower SS (>60 mV/dec) of conventional CMOS until the V_{mem} exceeds the V_{th} of the neuron circuit, as shown in Fig. 12 (a). On the other hand, in a new neuron circuit based on the proposed PF neuron device, low leakage current (<1 nA) flows in the PF neuron circuit due to steep switching characteristics (SS < 1 mv/dec) before the PF neuron device turns on. Thus, in the PF neuron circuit, the current flows only when the state of V_{out} changes by the fire and reset operation as shown in Fig. 12 (b). The proposed PF neuron circuit (~0.62 pJ/spike) can reduce energy consumption per spike by about 10 times compared to the conventional neuron circuit (~6.14 pJ/spike). To implement a stable IF operation, neuron circuits composed of conventional FETs have been investigated [28], [29]. For input spike integration and output spike firing, a dual mode neuron circuit was reported [28]. Energy consumption of the dual-mode neuron circuit

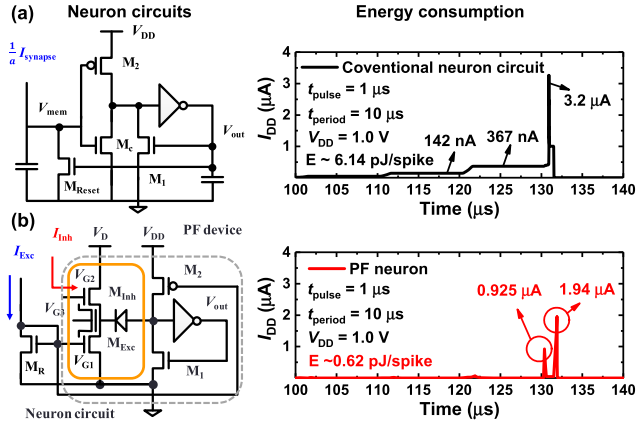

FIGURE 12. Circuit diagrams and simulated *I-t* plots for the comparison of energy consumption between (a) a conventional neuron circuit and (b) PF neuron circuit, respectively.

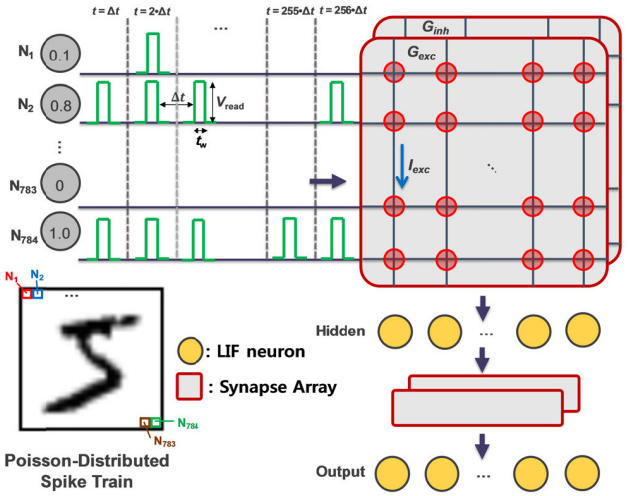

FIGURE 13. A block diagram of deep-spiking neural networks (D-SNNs) for MNIST classification.

is 9.6 pJ/spike/synapse when the load resistance of the dual-neuron circuit is 1 MΩ. The dual-neuron circuit consumed the static current (13 μA) in the integration mode. An improved neuronal circuit consisting of 10 FETs and 2 capacitors (membrane and refractory) has been reported to reduce power consumption [29]. Energy consumption of the improved neuron circuit was reported as 4.2 pJ/spike.

We evaluate the performance of deep-spiking neural networks (D-SNNs) using PF neurons through the MNIST dataset classification accuracy. Fig. 13 shows a block diagram of D-SNN for MNIST classification. Two devices are paired to express the excitatory and inhibitory properties of one synapse. Excitatory and inhibitory synaptic currents are summed through the source or bit-line in the synaptic array. The input of the D-SNN is converted to a Poisson-distributed spike train within 256 steps. The popular model of a leaky integrate-and-fire (LIF) neuron [30] is simply expressed as follows:

$$\tau \frac{dV_{\text{mem}}(t)}{dt} = -V_{\text{mem}}(t) + I_{\text{Exc}}(t) - I_{\text{Inh}}(t), \tag{1}$$

where $V_{\text{mem}}(t)$ is the membrane potential at time t and τ is the time constant of the neuron. $I_{\text{Exc}}(t)$ and $I_{\text{Inh}}(t)$ are the excitatory and inhibitory currents from the pre-synaptic arrays, respectively. Note that the conductance of the synaptic devices plays a role of the weight in the D-SNN. The inputs in the form of voltage spikes are applied to the pre-synaptic arrays and the currents from the arrays flow to the neurons. Solving the linear differential equation (1), $V_{\text{mem}}(t)$ is obtained as

$$V^j_{\text{mem}}(t) = V^j_{\text{mem}}(t-\Delta t)\exp(-\frac{\Delta t}{\tau}) + \frac{\sum_{i=1}^{N_i}\left(G^{i,j}_{\text{Exc}} - G^{i,j}_{\text{Inh}}\right)S^i(t)\,t_{\text{w}}}{C_{\text{mem}}}, \tag{2}$$

where $G^{i,j}_{\text{Exc}}$ and $G^{i,j}_{\text{Inh}}$ are the conductances of synaptic devices representing the excitatory and inhibitory weights respectively, from the i-th pre-neuron to the j-th post-neuron. In addition, t_{w} is the width of the voltage spike and the exponential term determines the ratio of the potential that remains after one time step (Δt). The C_{mem} is a membrane capacitor of the neuron and the $S(t)$ is a voltage spike at time t. If V_{mem} of the neuron exceeds V_{th}, the neuron generates a spike as follows:

$$S^i(t) = g(V^i_{\text{mem}}(t)), \tag{3}$$

$$g(x) = \begin{cases} V_{\text{read}}, & x > V_{th} \\ 0, & else \end{cases} \tag{4}$$

A spike is generated in the form of a voltage pulse with an amplitude of V_{read} and a pulse width of t_{w}. When the neuron generates a spike, V_{th} of the neuron is reset to zero. The amount of charge accumulated in the C_{mem} is determined by the synaptic current (GV_{read}) and the pulse width. Here, the n region of the PF neuron device accumulates electrons, acting as the C_{mem} in the LIF model. The excitatory current from the G_{Exc} array flows into M_{R}, then the drain voltage of M_{R} is applied to V_{G1}, which can copy the excitatory signal through a current mirror circuit consisting of M_{R} and M_{Exc}. Similarly, the inhibitory signal can be copied when a pMOSFET mirror circuit is connected to M_{Inh}. The amount of accumulated electrons in the n region is modulated by the excitatory and inhibitory currents, and the accumulated electrons modulate the height of hole-injection barrier (V_{h}) between the n region and p^+ anode. If the V_{h} is lower than a certain potential (V_{th} in conventional LIF neuron), the PF device turns on due to the PF operation and the neuron circuit generates a spike. In addition, the depth of the potential well in the n region affects the time constant (τ) of the retention time of the accumulated electrons, and V_{G3} can control the depth of the potential well [21], [24]. As a result, the proposed neuron circuit based on the PF device plays a role of the conventional LIF neuron with functionality of adjusting the time constant.

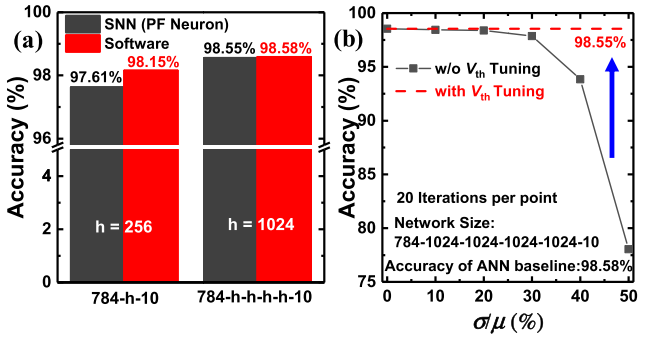

FIGURE 14. (a) Accuracy of MNIST classification in the D-SNN based on the proposed PF neuron and software-based NN. (b) Recovery of accuracy loss by V_{th} turning using G3 of the PF neuron devices in the D-SNN with 4 hidden layers.

To investigate the performance of the D-SNN based on the PF device, synaptic weights are trained with the rectified linear unit (ReLU) activation function using Adam optimizer in Pytorch framework with floating-point operation (software-based networks). In software-based networks, the accuracy rates are 98.15% in a network with 1 hidden layer (256 neurons), and 98.58% in a network with 4 hidden layers (1024 neurons in each layer). Then, trained weights are transferred to the conductance of synaptic devices in the D-SNN [31]. To evaluate the impact of the PF neuron on the performance of D-SNNs, we assume that the trained weights are ideally transferred to the conductance considering V_{read}, t_{w}, and C_{mem} as follows:

$$\frac{t_w V_{read}}{C_{mem}}(G_{exc}^{i,j} - G_{inh}^{i,j}) = w_{train}^{i,j}, \tag{5}$$

where w_{train} is the trained weight in the software-based networks. Whatever the value of the membrane capacitor of PF device is, various type of synaptic devices which meet the range of conductance value can be used. We set τ to 20 μs and Δt to 1 μs; thus, approximately 95% of the membrane potential remains after one time step. The V_{th} is set to 0.5 V in the D-SNNs. The accuracy rates of D-SNNs are 97.61% with 1 hidden layer and 98.55 % with 4 hidden layers, which shows very slight accuracy loss compared to the accuracy rate of software-based networks as shown in Fig. 14 (a). Also, the accuracy degradation caused by the V_{th} variation of PF device is investigated. As shown in Fig. 14 (b), the V_{th} variation of LIF neurons can degrade the accuracy of D-SNN ($\sim$75 % at σ/μ of 50 %). By applying bias or pulses (V_{PGM} or V_{ERS}) to the G3 of the PF neuron device, the V_{th} of PF neuron circuits can be tuned within 5% error by applying the incremental step pulse programming method used in NAND flash memory technology [32], restoring the accuracy of D-SNN to 98.51% as shown in Fig. 14 (b). Since the existing neuron circuit includes many transistors (>11) and a large C_{mem} (0.5 pF) with a footprint of 100 μm^2 for 0.35 μm CMOS technology [12], the proposed neuron circuit (no capacitor, 1 PF device, and 5 MOSFETs) contains much smaller components, so the area occupied is greatly reduced.

TABLE 1. Comparison of key characteristics of the proposed PF neuron and other neuron devices reported for implementing IF function.

	Physics	Pre-fire current	Operation voltage	Integration method	V_{th} tuning ability
FeFET [19]	Polarization switching	$\leq$ 0.5 uA	1.8 V	Sequential (Excitatory (G^+) & Inhibitory (G^-))	X
SOI FET [20]	Impact ionization (Floating Body)	$\leq$ 0.5 uA	$\leq$ 3.5 V	Excitatory (G^+)	X
Biristor [22]	Positive Feedback (Floating body)	< 10 nA	$\leq$ 2.5 V	Excitatory (G^+)	X
SOI FET [23]	Impact ionization (Floating Body)	< 10 nA	$\leq$ 4.5 V	Modulated excitatory (G^+)	X
This work	Positive Feedback (Floating body)	< 1 nA	$\leq$ 1.5 V	Simultaneous (Excitatory (G^+) & Inhibitory (G^-))	O

Table 1 compares the key features our PF device and prior neuron devices. Compared to conventional neuron devices, the proposed neuron devices can simultaneously process excitatory and inhibitory signals by accumulating or discharging electrons and control the threshold of neurons.

IV. CONCLUSION

We have successfully demonstrated that a capacitor-less PF neuron device enables the simultaneous processing of excitatory and inhibitory signals with a small footprint. Thanks to steep switching characteristics (SS < 1 mV/dec) of the PF neuron device, the neuron circuit in this work consumes 0.62 pJ of energy per spike, which is about 10 times less than that of a conventional neuron circuit to implement the IF function. Moreover, the threshold voltage (V_{th}) of the proposed PF neuron device is modulated by adjusting the potential barrier of the n region of the PF device (V_{G3} control or modulating the amount of charge stored in the charge storage layer by the program/erase operation), which controls the firing rate of neurons. The threshold tuning ability of the PF neuron device can implement the homeostasis function of biological neurons and compensate for the V_{th} variation of neurons in HNNs. The high-accuracy (98.55%) of the D-SNN based on the PF neuron device with 4 hidden layers (1024 neurons in each layer) was achieved. The proposed PF neuron device can be a promising solution to replace conventional neuron circuits for high-density and low-power neuromorphic systems.

ACKNOWLEDGMENT

(Sung Yun Woo and Dongseok Kwon contributed equally to this work.)

REFERENCES

[1] S. Yu, "Neuro-inspired computing with emerging nonvolatile memorys," *Proc. IEEE*, vol. 106, no. 2, pp. 260–285, Feb. 2018, doi: 10.1109/jproc.2018.2790840.

[2] E. Z. Farsa, A. Ahmadi, M. A. Maleki, M. Gholami, and H. N. Rad, "A low-cost high-speed neuromorphic hardware based on spiking neural network," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 9, pp. 1582–1586, Sep. 2019, doi: 10.1109/TCSII.2019.2890846.

[3] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, "Fully hardware-implemented memristor convolutional neural network," *Nature*, vol. 577, no. 7792, pp. 641–646, Jan. 2020, doi: 10.1038/s41586-020-1942-4.

[4] M. Chu, B. Kim, S. Park, H. Hwang, M. Jeon, B. H. Lee, and B.-G. Lee, "Neuromorphic hardware system for visual pattern recognition with memristor array and CMOS neuron," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2410–2419, Apr. 2015, doi: 10.1109/TIE.2014.2356439.

[5] H. Wu, P. Yao, B. Gao, W. Wu, Q. Zhang, W. Zhang, N. Deng, D. Wu, H.-S. P. Wong, S. Yu, and H. Qian, "Device and circuit optimization of RRAM for neuromorphic computing," in *IEDM Tech. Dig.*, Dec. 2017, pp. 11.5.1–11.5.4, doi: 10.1109/IEDM.2017.8268372.

[6] H. Kim, S. Hwang, J. Park, and B.-G. Park, "Silicon synaptic transistor for hardware-based spiking neural network and neuromorphic system," *Nanotechnology*, vol. 28, no. 40, pp. 1–10, Sep. 2017, doi: 10.1088/1361-6528/aa86f8.

[7] A. Sengupta, A. Banerjee, and K. Roy, "Hybrid spintronic-CMOS spiking neural network with on-chip learning: Devices, circuits, and systems," *Phys. Rev. A, Gen. Phys.*, vol. 6, no. 6, pp. 1–13, Dec. 2016, doi: 10.1103/PhysRevApplied.6.064003.

[8] D. Ielmini and S. Ambrogio, "Emerging neuromorphic devices," *Nanotechnology*, vol. 31, no. 9, pp. 1–24, Dec. 2019, doi: 10.1088/1361-6528/ab554b.

[9] G. W. Burr, R. M. Shelby, C. di Nolfo, J. W. Jang, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. Kurdi, and H. Hwang, "Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element," in *IEDM Tech. Dig.*, Dec. 2014, pp. 29.5.1–29.5.4, doi: 10.1109/IEDM.2014.7047135.

[10] S. Lim, J.-H. Bae, J.-H. Eum, S. Lee, C.-H. Kim, D. Kwon, B.-G. Park, and J.-H. Lee, "Adaptive learning rule for hardware-based deep neural networks using electronic synapse devices," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 8101–8116, Jul. 2018, doi: 10.1007/s00521-018-3659-y.

[11] S. Hwang, J. Chang, M.-H. Oh, J.-H. Lee, and B.-G. Park, "Impact of the sub-resting membrane potential on accurate inference in spiking neural networks," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Feb. 2020, doi: 10.1038/s41598-020-60572-8.

[12] P. Livi and G. Indiveri, "A current-mode conductance-based silicon neuron for address-event neuromorphic systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2009, pp. 2898–2901, doi: 10.1109/ISCAS.2009.5118408.

[13] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. V. Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen, "Neuromorphic silicon neuron circuits," *Frontiers Neurosci.*, vol. 5, p. 73, May 2011, doi: 10.3389/fnins.2011.00073.

[14] K. Moon, E. Cha, D. Lee, J. Jang, J. Park, and H. Hwang, "ReRAM-based analog synapse and IMT neuron device for neuromorphic system," in *Proc. Int. Symp. VLSI Technol., Syst. Appl. (VLSI-TSA)*, Apr. 2016, pp. 1–2, doi: 10.1109/VLSI-TSA.2016.7480499.

[15] J. Lin, Annadi, S. Sonde, C. Chen, L. Stan, K. V. L. V. Achari, S. Ramanathan, and S. Guha, "Low-voltage artificial neuron using feedback engineered insulator-to-metal-transition devices," in *IEDM Tech. Dig.*, Dec. 2016, pp. 34.5.1–34.5.4, doi: 10.1109/IEDM.2016.7838541.

[16] J. Lin and J.-S. Yuan, "Capacitor-less RRAM-based stochastic neuron for event-based unsupervised learning," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2017, pp. 1–4, doi: 10.1109/BIOCAS.2017.8325169.

[17] M. Barci, L. Perniola, G. Molas, C. Cagli, E. Vianello, M. Bernard, A. Roule, A. Toffoli, J. Cluzel, and B. De Salvo, "Bilayer metal-oxide conductive bridge memory technology for improved window margin and reliability," *IEEE J. Electron Devices Soc.*, vol. 4, no. 5, pp. 314–320, Sep. 2016, doi: 10.1109/JEDS.2016.2567219.

[18] C. Chen, M. Yang, S. Liu, T. Liu, K. Zhu, Y. Zhao, H. Wang, Q. Huang, and R. Huang, "Bio-inspired neurons based on novel leaky-FeFET with ultra-low hardware cost and advanced functionality for all-ferroelectric neural network," in *Proc. Symp. VLSI Technol.*, Jun. 2019, pp. T136–T137, doi: 10.23919/VLSIT.2019.8776495.

[19] J. Luo, S. Wu, Q. Huang, R. Huang, L. Yu, T. Liu, M. Yang, Z. Fu, Z. Liang, L. Chen, C. Chen, and S. Liu, "Capacitor-less stochastic leaky-FeFET neuron of both excitatory and inhibitory connections for SNN with reduced hardware cost," in *IEDM Tech. Dig.*, Dec. 2019, p. 122, doi: 10.1109/IEDM19573.2019.8993535.

[20] S. Dutta, V. Kumar, A. Shukla, N. R. Mohapatra, and U. Ganguly, "Leaky integrate and fire neuron by charge-discharge dynamics in floating-body MOSFET," *Sci. Rep.*, vol. 7, no. 1, pp. 1–7, Aug. 2017, doi: 10.1038/s41598-017-07418-y.

[21] M.-W. Kwon, K. Park, M.-H. Baek, J. Lee, and B.-G. Park, "A low-energy high-density capacitor-less I&F neuron circuit using feedback FET co-integrated with CMOS," *IEEE J. Electron Devices Soc.*, vol. 7, pp. 1080–1084, Sep. 2019, doi: 10.1109/JEDS.2019.2941917.

[22] J.-W. Han and M. Meyyappan, "Leaky Integrate-and-Fire biristor neuron," *IEEE Electron Device Lett.*, vol. 39, no. 9, pp. 1457–1460, Sep. 2018, doi: 10.1109/LED.2018.2856092.

[23] J.-K. Han, M. Seo, W.-K. Kim, M.-S. Kim, S.-Y. Kim, M.-S. Kim, G.-J. Yun, G.-B. Lee, J.-M. Yu, and Y.-K. Choi, "Mimicry of excitatory and inhibitory artificial neuron with leaky integrate-and-fire function by a single MOSFET," *IEEE Electron Device Lett.*, vol. 41, no. 2, pp. 208–211, Feb. 2020, doi: 10.1109/LED.2019.2958623.

[24] N. Navlakha, J.-T. Lin, and A. Kranti, "Improved retention time in twin gate 1T DRAM with tunneling based read mechanism," *IEEE Electron Device Lett.*, vol. 37, no. 9, pp. 1127–1130, Sep. 2016, doi: 10.1109/LED.2016.2593700.

[25] D. Querlioz, O. Bichler, P. Dollfus, and C. Gamrat, "Immunity to device variations in a spiking neural network with memristive nanodevices," *IEEE Trans. Nanotechnol.*, vol. 12, no. 3, pp. 288–295, May 2013, doi: 10.1109/TNANO.2013.2250995.

[26] S. Y. Woo, K.-B. Choi, J. Kim, W.-M. Kang, C.-H. Kim, Y.-T. Seo, J.-H. Bae, B.-G. Park, and J.-H. Lee, "Implementation of homeostasis functionality in neuron circuit using double-gate device for spiking neural network," *Solid-State Electron.*, vol. 165, pp. 1–6, Mar. 2020, Art. no. 107741, doi: 10.1016/j.sse.2019.107741.

[27] K.-B. Choi, S. Y. Woo, W.-M. Kang, S. Lee, C.-H. Kim, J.-H. Bae, S. Lim, and J.-H. Lee, "A split-gate positive feedback device with an integrate-and-fire capability for a high-density low-power neuron circuit," *Frontiers Neurosci.*, vol. 12, pp. 1–13, Oct. 2018, doi: 10.3389/fnins.2018.00704.

[28] X. Wu, V. Saxena, and K. Zhu, "Homogeneous spiking neuromorphic system for real-world pattern recognition," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 5, no. 2, pp. 254–266, Jun. 2015, doi: 10.1109/JETCAS.2015.2433552.

[29] J. Shamsi, K. Mohammadi, and S. B. Shokouhi, "A hardware architecture for columnar-organized memory based on CMOS neuron and memristor crossbar arrays," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 12, pp. 2795–2805, Dec. 2018, doi: 10.1109/TVLSI.2018.2815025.

[30] Y. Wu, L. Deng, G. Li, J. Zhu, and L. Shi, "Spatio-temporal backpropagation for training high-performance spiking neural networks," *Frontiers Neurosci.*, vol. 12, pp. 1–12, May 2018, doi: 10.3389/fnins.2018.00331.

[31] B. Rueckauer, I.-A. Lungu, Y. Hu, M. Pfeiffer, and S.-C. Liu, "Conversion of continuous-valued deep networks to efficient event-driven networks for image classification," *Frontiers Neurosci.*, vol. 11, pp. 1–12, Dec. 2017, doi: 10.3389/fnins.2017.00682.

[32] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis and modeling," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Mar. 2013, pp. 1285–1290, doi: 10.7873/DATE.2013.266.

SUNG YUN WOO (Member, IEEE) received the B.S. degree in electrical engineering from Kyungpook National University, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Seoul National University (SNU), Seoul, South Korea. He is currently with the Inter-University Semiconductor Research Center, SNU. His current research interests include neuromorphic systems and neural networks.

DONGSEOK KWON (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2017. He is currently pursuing the M.S. degree with the Department of Electrical and Computer Engineering, Seoul National University (SNU), Seoul, South Korea. His current research interest includes neuromorphic systems and its application in computing.

NAGYONG CHOI (Member, IEEE) received the B.S. degree in electronic engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University (SNU), Seoul, South Korea, in 2014 and 2020, respectively.

WON-MOOK KANG received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Seoul National University (SNU), Seoul. His research interest includes neuromorphic systems.

YOUNG-TAK SEO (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from Seoul National University (SNU), Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. He is currently with the Inter-University Semiconductor Research Center, SNU. His current research interest includes neuromorphic systems and its application in computing.

MIN-KYU PARK received the B.S. degree in electrical engineering from the University of California at Los Angeles (UCLA), Los Angeles, USA, in 2018. He is currently pursuing the M.S. degree with the Department of Electrical and Computer Engineering, Seoul National University (SNU), Seoul, South Korea. He is currently with the Inter-University Semiconductor Research Center, SNU. His current research interests include nueromorphic systems and neural networks.

JONG-HO BAE (Member, IEEE) received the B.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2011, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Seoul National University (SNU), Seoul, South Korea, in 2018. He was a Postdoctoral Associate with the Inter-University Semiconductor Research Center (ISRC), SNU, from 2018 to 2019, and the University of California at Berkeley, Berkeley, CA, USA, from 2019 to 2020. In 2020, he joined the School of Electrical Engineering, Kookmin University, South Korea, as an Assistant Professor. His current research interests include charge trap memory, and hardware-based neural networks and its applications.

BYUNG-GOOK PARK (Fellow, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University (SNU), Seoul, South Korea, in 1982 and 1984, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. In 1994, he joined as an Assistant Professor with the Department of Electrical and Computer Engineering, SNU, where he is currently a Professor.

JONG-HO LEE (Fellow, IEEE) received the Ph.D. degree in electronic engineering from Seoul National University (SNU), Seoul, South Korea, in 1993. He was a Postdoctoral Fellow with the Massachusetts Institute of Technology, Cambridge, MA, USA, from 1998 to 1999. He has been a Professor with the School of Electrical and Computer Engineering, SNU, since 2009. He is also a Lifetime Member of the Institute of Electronics Engineers of Korea (IEEK).