creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Ph.D. DISSERTATION

Compact Neuromorphic System with Four-Terminal Si-Based Synaptic Devices for Spiking Neural Networks

스파이킹 신경 네트워크에 적용 가능한 4단자 실리콘 기반 시냅스 소자를 포함하는 콤팩트 신경모방 시스템

BY

JUNGJIN PARK

February 2017

DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Compact Neuromorphic System with Four-Terminal Si-Based Synaptic Devices for Spiking Neural Networks

스파이킹 신경 네트워크에 적용 가능한 4단자 실리콘 기반 시냅스 소자를 포함하는 콤팩트 신경모방 시스템

指導敎授 朴炳國

이 論文을 工學博士學位論文으로 提出함

2017 년 2 월

서울大學校 大學院

電氣·情報 工學部

朴 正 振

朴正振의 工學博士 學位論文을 認准함

2017 년 2 월

委 員 長 : 이종호 (印)

副委員長 : 박병국 (印)

委 員 : 김재하 (印)

委 員 : 송윤흡 (印)

委 員 : 박완준 (印)

Abstract

As the conventional electronic system originated from Von Neumann's architecture has faced with fundamental physical limit and revealed the weakness in terms of cost and efficiency, we feel the need to introduce a new paradigm of information processing. As an alternative to the conventional system, the neuromorphic system inspired by human neuron network has been developed, recently. The neuromorphic system can be configured based on specific artificial neural networks (ANNs) and the various ANNs have recently appeared on the research field. Among them, spiking neural networks (SNNs) based on temporal human spiking neurons attract a lot of attention with the expectation of their higher information processing ability. The addition of the temporal dimension in the SNNs results in compact representation of large neural networks and has great possibility for solving complicated time-dependent pattern recognition.

In this thesis, we propose a compact neuromorphic system that can work with four-terminal Si-based synaptic devices for spiking neural networks (SNNs) and explain the fabrication method of neuron circuit based on standard CMOS process. The system consists of a synaptic integration part and an action-potential generation part. The synaptic integration part includes the current mirrors connected to the n-channel and p-channel synaptic devices for the expression of excitation and inhibition of biological neurons, respectively. The synaptic device can change its weight using floating-body effect and

charge injection into the floating gate. The number and the weight of the synaptic device affect the amount of the current reproduced from the current mirror. The action-potential generation part includes double-stage inverters and the NMOS with large V_T for generating an asymmetric action-potential. The generated action-potential is transmitted to a next neuron and simultaneously returned to the back gate of the synaptic device for the change of weight based on spike-timing-dependent-plasticity.

As the four-terminal synaptic device can transmit pre-neuron signals and change its weight at the same time, we can constitute the compact neuromorphic system without additional switches or logic operation and emulate the operation of neuron with a minimum number of devices and power dissipation. These advantages of the system are expected to help in the construction of various spiking neural network applications such as unsupervised pattern recognition very efficiently.

Key Words: Neuromorphic, Neuron Circuit, Synaptic Device, Action-Potential, Spike-Timing-Dependent-Plasticity (STDP), Spiking Neural Networks.

Student Number: 2010-20806

Contents

Chapter 1

Introduction

1.1 Emergence of Neuromorphic System

As the conventional electronic system originated from Von Neumann's architecture has faced with fundamental physical limit and revealed the weakness in terms of cost and efficiency [1]-[4], we feel the need to introduce a new paradigm of information processing. As an alternative to the conventional system, the neuromorphic system inspired by human neuron network has been developed, recently [5]-[9]. The neuromorphic system means the analog, digital and mixed-mode analog/digital VLSI and software systems that implement the models of neural systems (for perception, motor control, or multisensory integration) [10]. While the conventional system has serial architecture and fixed wiring among components, the neuromorphic system enables the parallel information processing with high efficiency and low power consumption and has the adaptive, defect-tolerant characteristic with flexible modulation of conductance [11]-[12]. As shown in Fig. 1.1, the von Neumann machine may be more advantageous in terms of machine complexity in the case of simple calculations with small environmental complexity, but as environmental

complexity increases, such as recognition and reasoning, the neuromorphic machine can work with a relatively simple machine configuration.

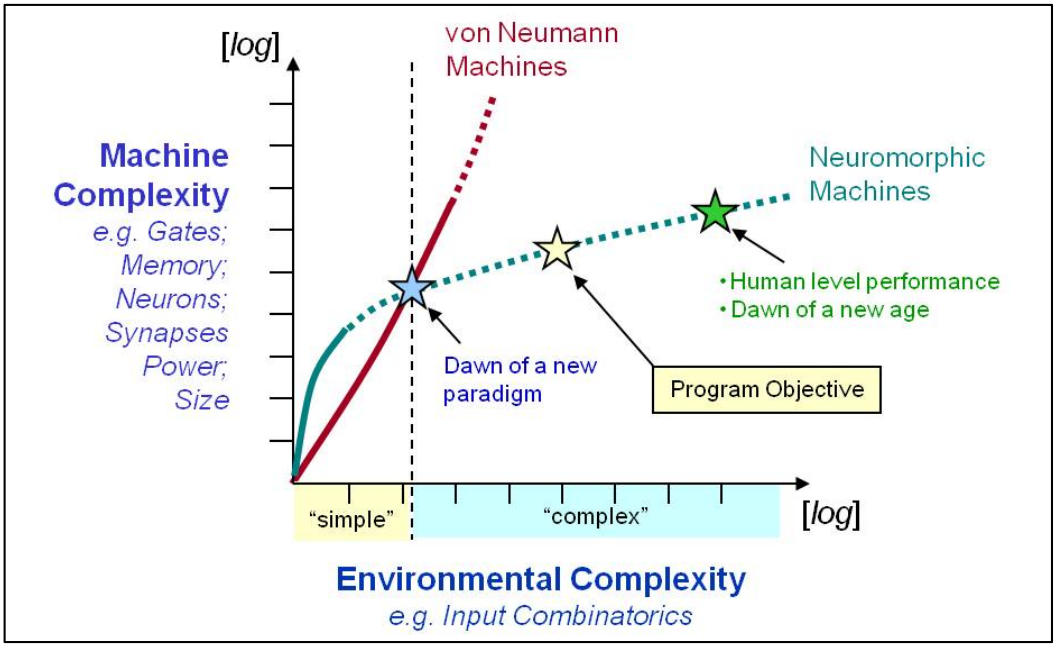


Fig. 1. 1. Von Neumann machines vs. Neuromorphic machines.

1.2 Human Neural Network and Artificial Neural Network

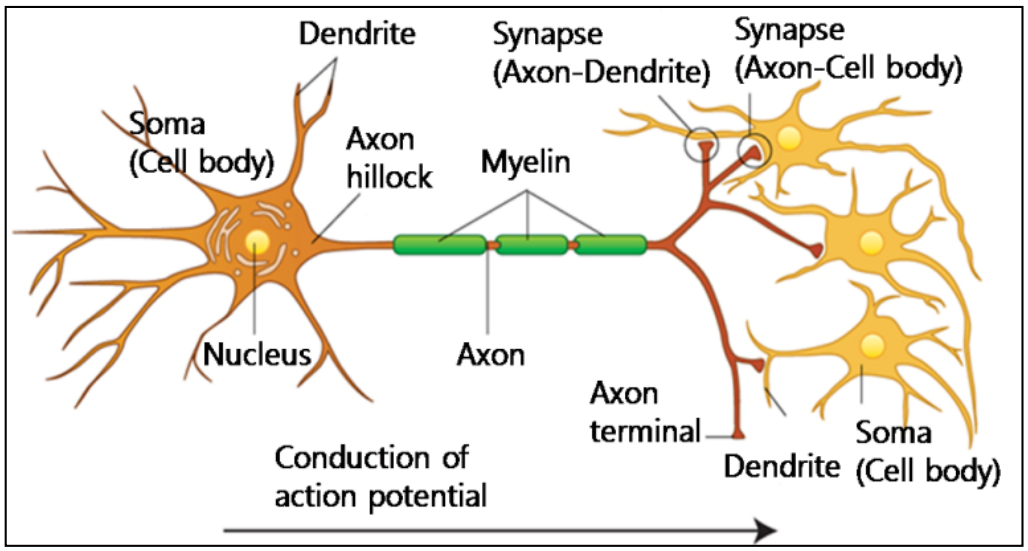


Fig. 1. 2. Diagram of human neural network.

Fig. 1.2 shows the diagram of human neural network. The neural network has axons and dendrites. The axons are path for transmitting action-potentials and the dendrites receive the action-potentials from other neurons through synapse connections. Each neuron may have more than 1000 synapse connections with other neurons. The synapse can modulate own weight through the formation of memory and affects the firing of neuron. Especially, the long-term weight of the synapse depends on the spike-timing-dependent-plasticity (STDP) mechanism originated from the timing difference between the input and the output signal of the neuron. In addition, the synapse is classified into excitatory synapse

and inhibitory synapse. The excitatory synapse transmits positive post-synaptic potentials and promotes the firing of the neuron. The inhibitory synapse transmits negative post-synaptic potentials and inhibits the firing of the neuron. The post-synaptic potentials are integrated in the cell body and if the sum of the integrated potential signals exceeds the threshold value, the post-synaptic neuron will fire and generate a new action-potential. This signal transmission model is called “Integrate-and-Fire neuron model” and it became the basis of the perceptron proposed in 1953. The operation mechanism of perceptron is explained well in Fig. 1. 3.

The neuromorphic system can be configured based on specific artificial neural

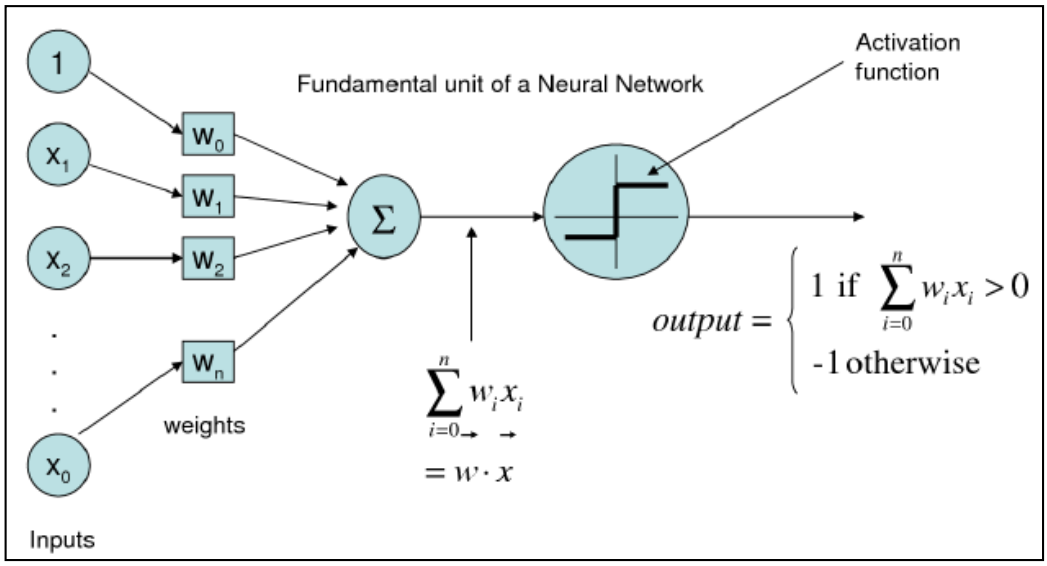


Fig. 1. 3. Operation mechanism of perceptron.

networks (ANNs) algorithm and the various ANNs have recently appeared on the research field. Generally, the ANN consists of an input layer and an output layer, and hidden layers between them. Each layer is connected through synapses. The input and output are real non-negative numbers, and the magnitude of signal corresponds to the magnitude of the number. The synapses that connect each layer repeatedly adjust their weights through an error back-propagation method, and work to find the optimal answer at the output layer. The operation principle of the ANN is summarized in Fig. 1. 4.

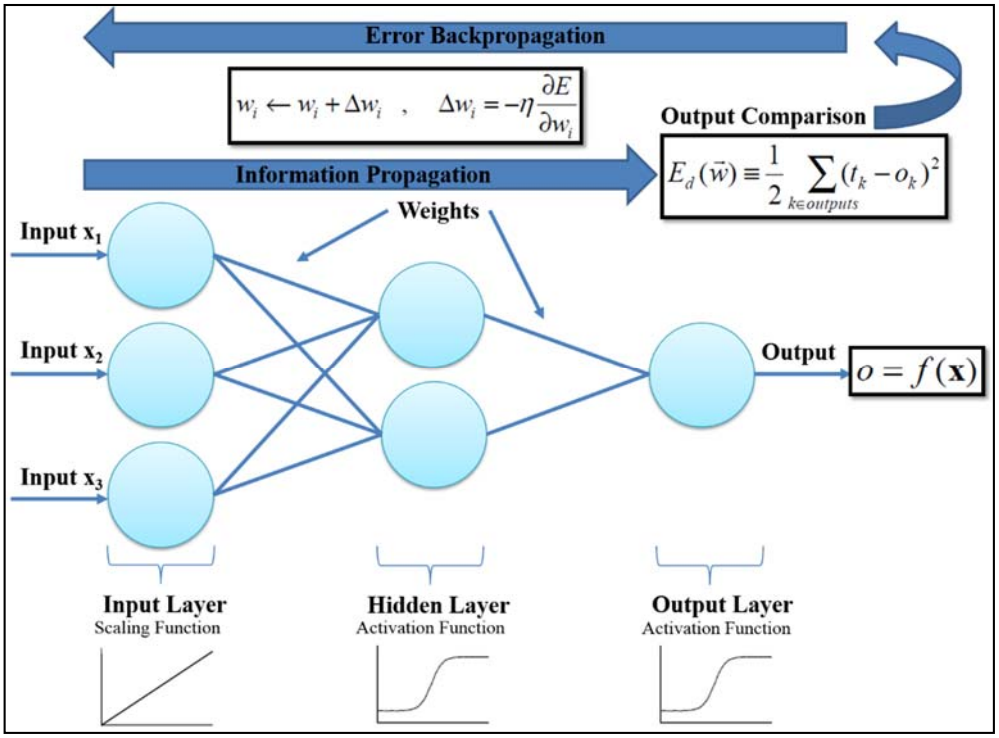


Fig. 1. 4. Origin of artificial neural network.

1.3 Hardware Approach of Neuromorphic System

Recently, special algorithms such as Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) in Fig. 1.5, 1.6 have attracted much attention, and most of them have been applied to the existing von Neumann system. However, since the von Neumann system is a serial architecture in which CPU and MEMORY exchange information, it has a somewhat inefficient structure to apply these algorithms. Therefore, it is necessary to develop a hardware neuromorphic system having a neuron-synapse-neuron structure similar to a real human neural network in order to efficiently apply the neural network-based algorithm. In fact, in 2013, QUALCOMM introduced a hardware-based neuromorphic chip called ZEROTH, and in 2014 IBM developed TrueNorth successively as shown in Fig. 1. 7.

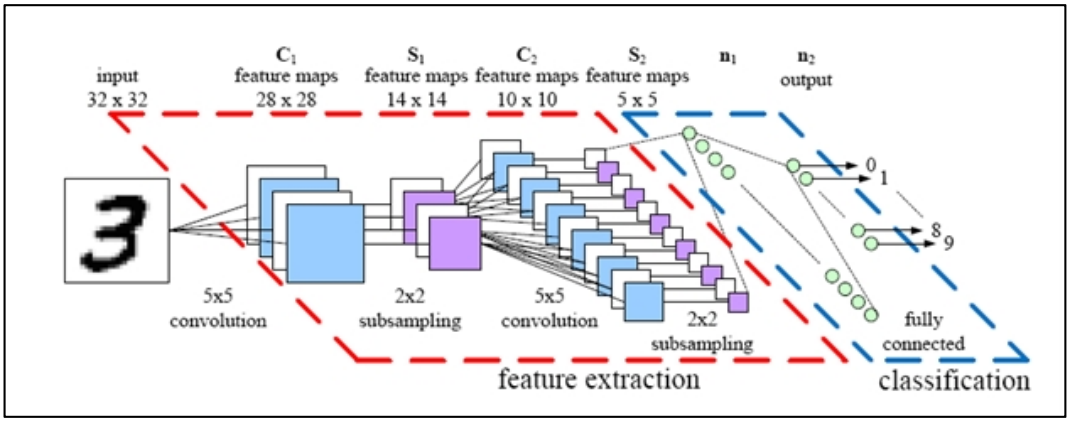


Fig. 1. 5. Conceptual diagram of Convolution Neural Network (CNN).

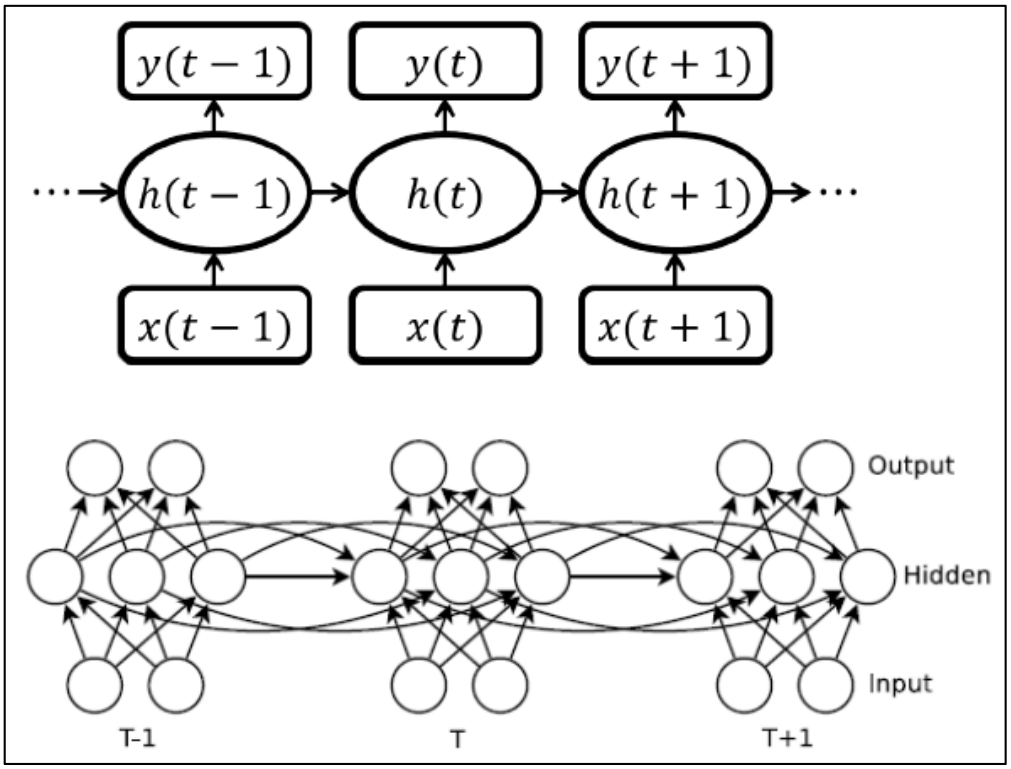


Fig. 1. 6. Conceptual diagram of Recurrent Neural Network (RNN).

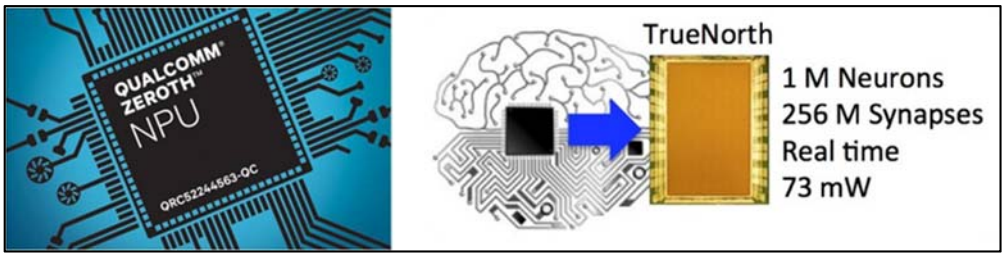


Fig. 1. 7. ZEROTH, QUALCOMM (2013) and TrueNorth, IBM (2014).

1.4 Spiking Neural Network

With the introduction of hardware neuromorphic systems, the need for an artificial neural network that best fits the system has also been raised. Among the many candidates, Spiking Neural Network (SNN) based on temporal human spiking neurons attract a lot of attention with the expectation of their higher information processing ability [13]-[15]. The basis of the SNN is "Integrate-and-Fire neuron model". The addition of the temporal dimension in the SNN results in compact representation of large neural networks and has great possibility for solving complicated time-dependent pattern recognition [16]. The operation mechanism of SNN is explained well in Fig. 1. 8.

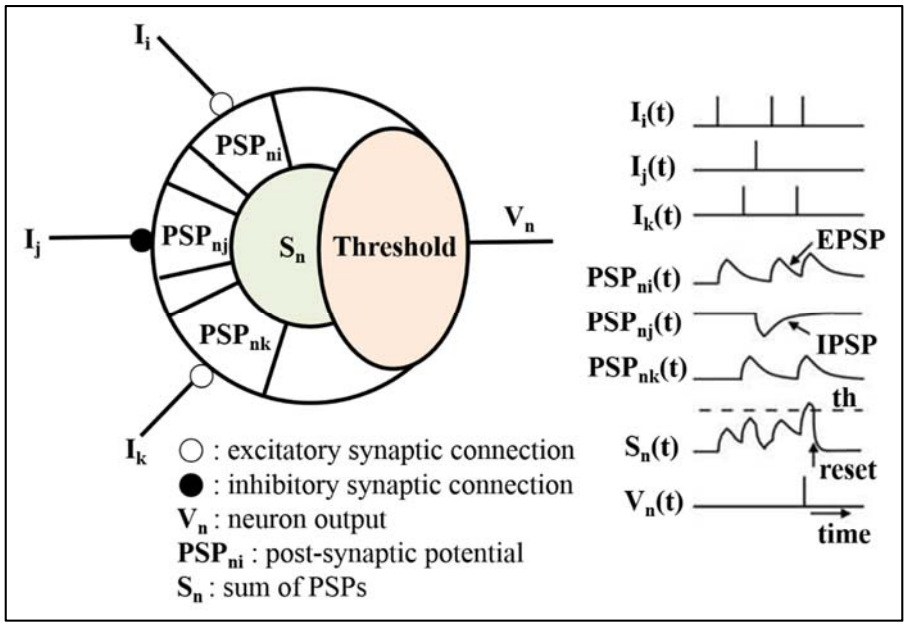


Fig. 1. 8. Operation mechanism of Spiking Neural Network (SNN).

1.5 Previous Research about SNN Applications

In recent years, the various applications of the SNNs have been developed. The applications can be classified according to what kinds of synaptic devices are used. Some applications have utilized several MOSFETs for emulating the operation of synapse [17]-[21] and others have used a resistive switching memory as the synaptic device [22]-[25]. In case of the MOSFET-based synapse, it has good compatibility with CMOS neuron circuit and simple fabrication method but occupies larger area than the resistive-switching memory. Otherwise, the resistive switching memory has the merit of space utilization but has relatively bad compatibility with CMOS and poor reliability. In addition, we can also classify the applications according to which methods such as analog operation [17]-[22] or mixed-mode and digital logic operation [23]-[25] are applied to represent the behavior of neuron. The critical points in designing the hardware neuromorphic systems are: “How can we precisely implement neuron’s mechanism?” and “Can it be operated with minimum power dissipation?” Therefore, we propose a neuromorphic system trying to emulate the signal transmission of the spiking neuron and introduce the four-terminal Si-based synaptic device to configure the compact structure for weight update. Recent applications generally utilize external controllers or switches to express the weight update and the signal transmission in the synaptic device at the same time. Since we use the four-terminal Si-based synapses and connect them to the neuron circuit without the additional switch or logic operation, the proposed system can effectively emulate the neuron’s mechanism.

1.6 Scope of Thesis

In this thesis, we propose a novel neuron circuit for driving the proposed four-terminal Si-based synaptic device in our research group and verify its operation through a simulation. In addition, we examined the operation characteristics of a neuron circuit directly fabricated through a CMOS process, and confirmed whether the proposed neuron circuit with the synaptic devices can construct a hardware neuromorphic system that is efficient in terms of power and space consumption. Finally, by applying the hardware neuromorphic system to the unsupervised pattern recognition, we examined the applicability of this study to various neuromorphic applications in the future.

In chapter 1, the introduction and background knowledge of the study are presented in order to make it easier to understand the topics covered in this study. In chapter 2, we will study the operation principle of a four-terminal Si-based synaptic device, and verify its operating characteristics through simulation tool and actually fabricated devices. In chapter 3, we will study the operation principle of neuron circuit that drives synaptic device, and verify its operation characteristics through simulation tool. In chapter 4, the process of fabricating the neuron circuit through the CMOS process will be examined, and the measurement results of the actually fabricated neuron circuit will be analyzed. In chapter 5, we will examine the application of the proposed hardware neuromorphic system to unsupervised pattern recognition and verify that the system is applicable to various spiking neural network applications. In Chapter 6, we will review the results of this study.

Chapter 2

Si-Based Floating-Body Synaptic Transistor (SFST)

In this chapter, we will look at how the Si-Based Floating-Body Synaptic Transistor works and how it implements short-term memory and long-term memory of biological synapses. In addition, we will verify the operation of the synaptic device through a simulation tool and directly fabricated device, and examine the possibility of autonomous connection with a neuron circuit for the construction of an efficient hardware neuromorphic system. The all simulations were conducted using SILVACO ATLAS TCAD simulator.

2.1 Operation Principle of SFST

We used the Si-based floating-body synaptic transistor (SFST) [26]-[28] in Fig. 2.1 as the synaptic device that has short-term and long-term memory and modulates the conductance of signals. The merit of the Si-based synaptic transistor is that it can be fabricated with CMOS neuron circuit at the same time and the fabrication method of Si-based device is well known. The SFST has the natural transition mechanism between the

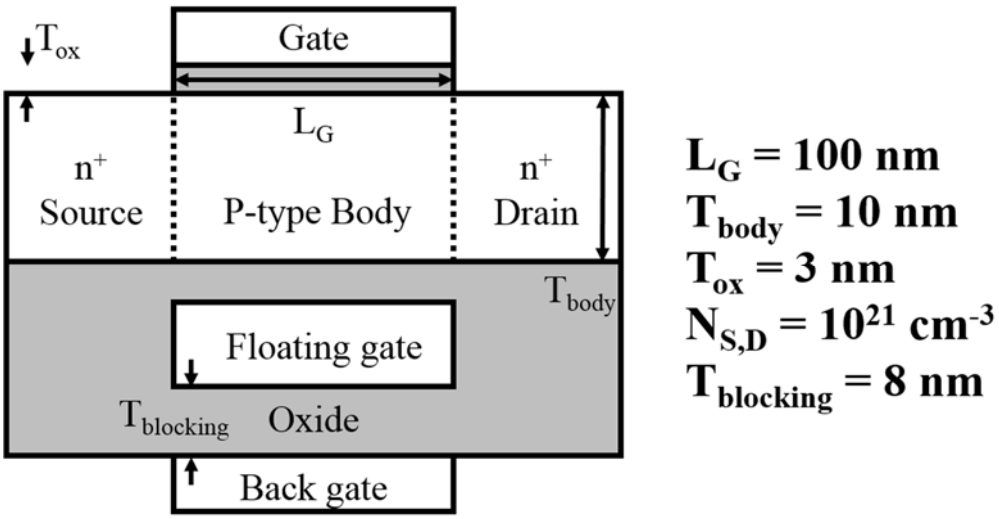
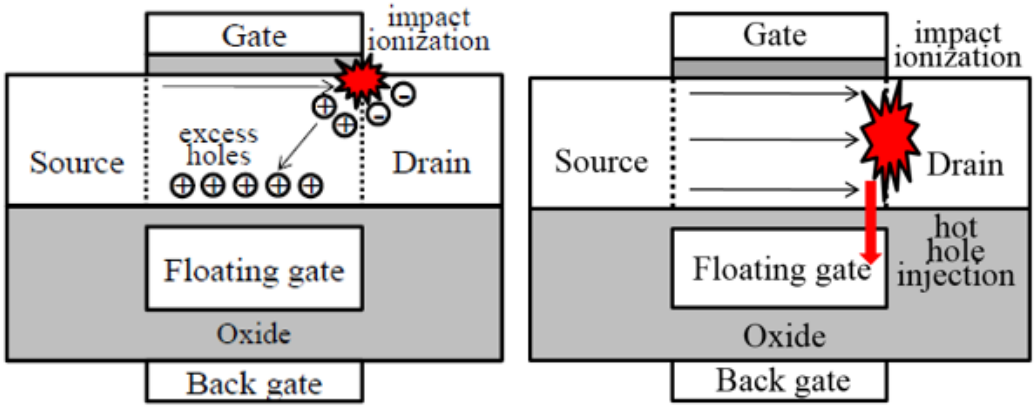


Fig. 2. 1. Si-based floating-body synaptic transistor.

short-term and long-term memory and the relatively easier expression method of STDP with utilizing its four terminals. Above all, since the output pulse from the neuron circuit can be directly connected to the back gate of the synaptic device, the formation of memory in the synaptic device using STDP and the transmission of action-potentials to next neurons can take place at the same time. Therefore, we don't need to set up an additional switch or logic operation.

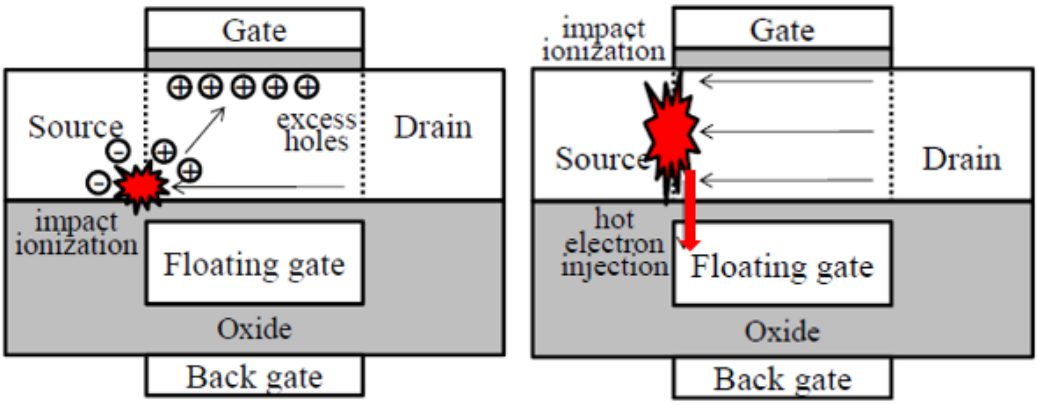
Once the n-channel SFST is appropriately biased (e.g., $V_G = V_{DS} = 2$ V, $V_{BG} = -2$ V) as shown in Fig. 2. 2 (a), the excess holes are generated by impact ionization near the top gate and increase the potential of the body region to accelerate the impact ionization. This positive feedback process lowers the energy barrier between the source and the body and



(a) (b)

Fig. 2. 2. Potentiation in the synaptic device. (a) The excess holes are generated by impact ionization near the top gate and increase the potential of the body region to accelerate the impact ionization. (b) At the point when the source-body junction is forward biased due to the excess holes, the generated hot holes start to enter the floating gate.

the threshold voltage (V_T) of the SFST. However, because the body can't keep the excess holes permanently, the feedback effect can be compared to the short-term potentiation of synaptic weight. At the point when the source-body junction is forward biased due to the excess holes, the impact ionization occurs near the back channel and the generated hot holes start to enter the floating gate as shown in Fig. 2. 2 (b). Then, the decrease of the V_T becomes permanent and the above process corresponds to the short term to long-term transition of the memory in biological system. As the n-channel SFST is reverse-biased (e.g., $V_G = V_{DS} = -2$ V, $V_{BG} = 2$ V) as shown in Fig. 2. 3 (a), the hot electrons are generated



(a) (b)

Fig. 2. 3. Depression in the synaptic device. As the n-channel SFST is reverse-biased, (a) hot electrons are generated near the floating gate and (b) start to enter the floating gate when the overall drain-body junction is forward-biased.

near the floating gate and start to enter the floating gate when the overall drain-body junction is forward-biased, explained in Fig. 2. 3 (b). Since the electrons are injected to the floating gate, the V_T increases and it means the depression of synaptic weight. As the short-term or long-term memory is formed in the synaptic device, the conductance of the synapse is changed from the initial state and it affects the amount of the current reproduced from the current mirrors. As a result, the firing point of the system is changed based on whether the memory is formed in the synaptic device or not.

2.2 Operation Characteristic of SFST from Simulation

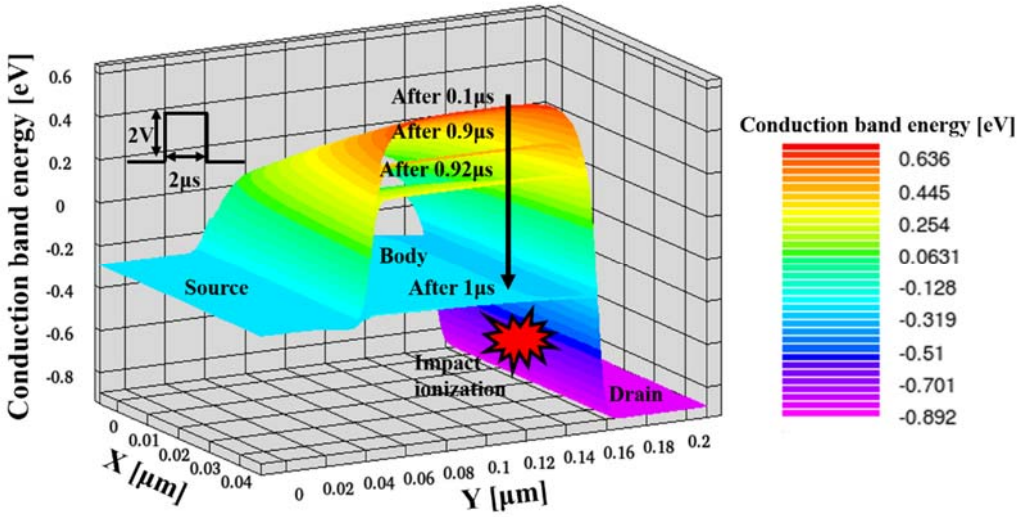


Fig. 2. 4. Change of conduction band energy between source and drain.

Fig. 2. 4 shows the change of conduction band energy between source and drain in the synaptic device. When a pulse having a size of 2 V and a width of 2 μs is applied to the device, the energy barrier between the source and the drain is gradually lowered due to the holes originated from the impact ionization between the body and the drain. As the pulse continues to be applied, the positive feedback between the hole accumulated in the body and the impact ionization becomes more and more intense. After 0.9 μs, the barrier between the source and the drain begins to drop rapidly, and finally the barrier disappears from all parts of the body. Through this process, the conductance of the synaptic device increases in the short term compared with the initial state, and if the holes accumulated in the body

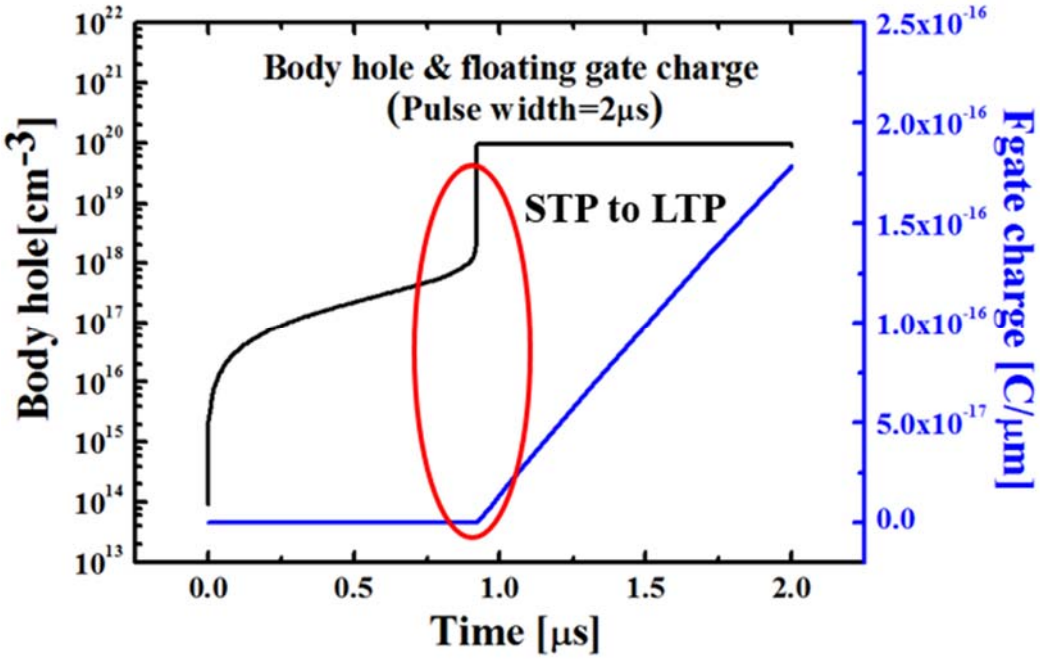


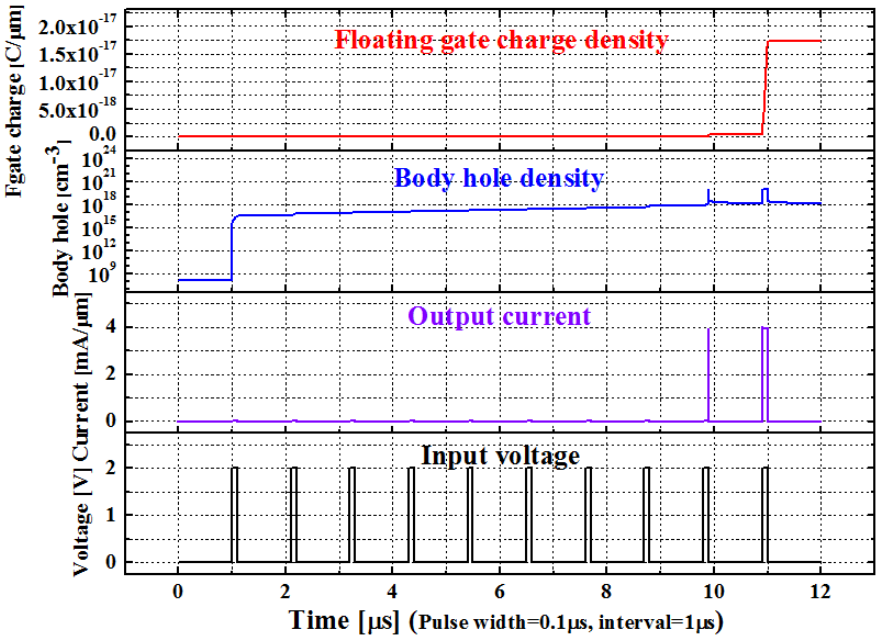
Fig. 2. 5. Change of conduction band energy between source and drain.

are injected into the floating gate due to the voltage applied to the back gate, the long-term increase in conductance can be obtained.

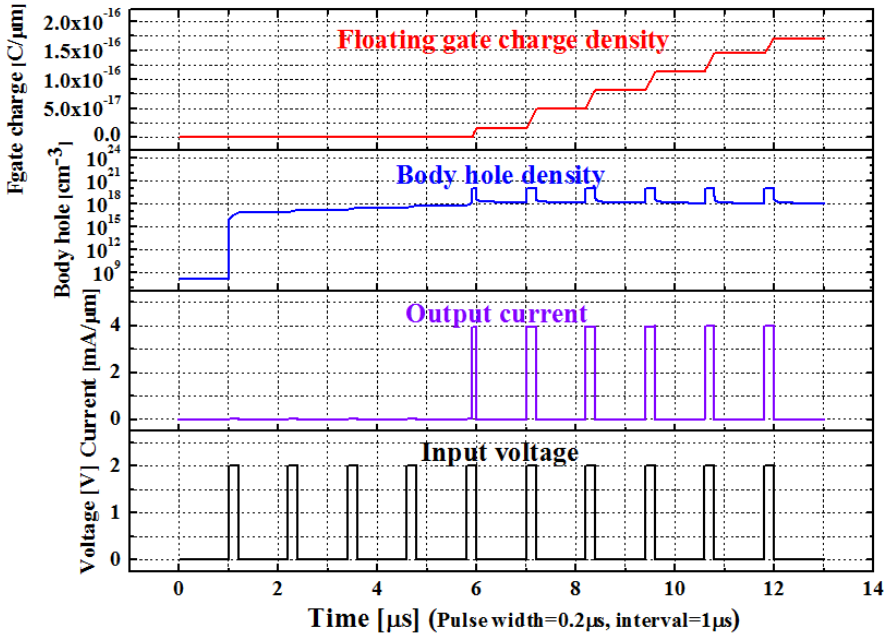
Fig. 2. 5 explains the floating body effect in the Fig. 2. 4 through the behavior of holes accumulated in the body and injected into the floating gate. As the input pulse is applied, the hole accumulated in the body gradually increases, but charge injection to the floating gate does not occur until strong positive feedback occurs across the body. After that, charge injection into the floating gate starts to occur when the hole is explosively accumulated in the body due to the strong feedback between the impact ionization and the holes in the body,

which means the transition from the short-term potentiation to the long-term potentiation.

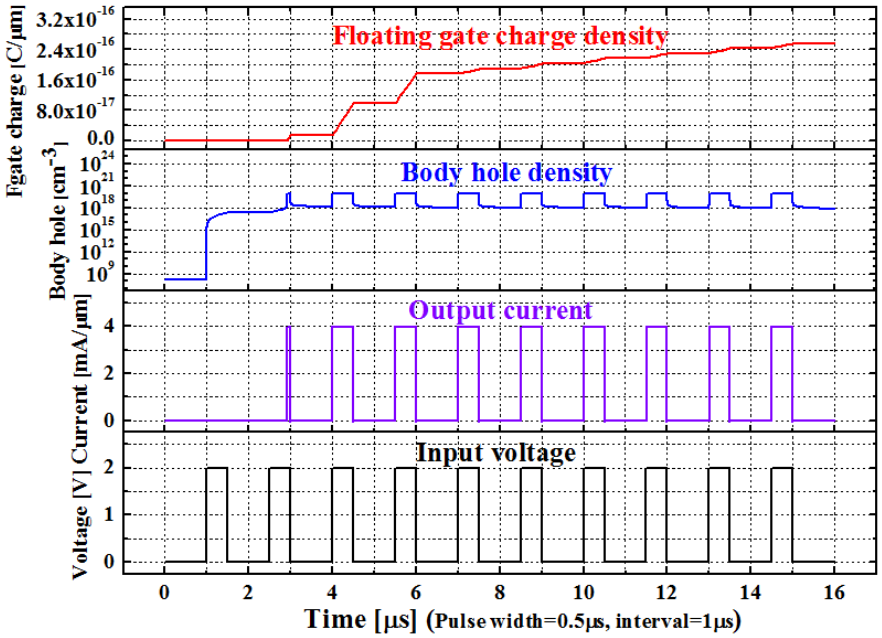
Fig. 2. 6 shows the transient tendency of the hole accumulation in the body from the impact-ionization near the drain and the hole injection into the floating gate when the width of an input pulse is changed. In order to observe the motion of the holes dramatically, we applied the 10 pulses to the synaptic device such as V_G = V_{DS} = 2 V, V_{BG} = -2 V and increased the pulse width by 0.1 μs, 0.2 μs, 0.5 μs while the interval between pulses is fixed to 1 μs. When the pulse width is relatively short as shown in Fig. 2. 6 (a), the triggering point that the holes are explosively generated and injected into the floating gate is retarded because the enough time for the generation of the holes is not secured. As the pulse width



(a)


Floating gate charge density
Body hole density
Output current
Input voltage
Fgate charge [C/μm]
Body hole [cm-3]
Current [mA/μm]
Voltage [V]
Time [μs] (Pulse width=0.2μs, interval=1μs)

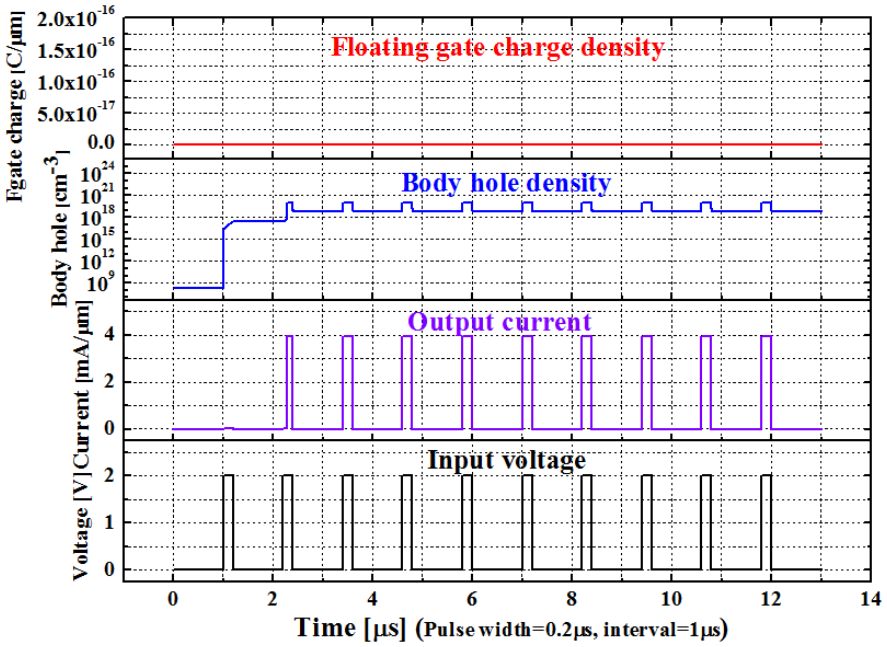
(b)


Floating gate charge density
Body hole density
Output current
Input voltage
Fgate charge [C/μm]
Body hole [cm-3]
Current [mA/μm]
Voltage [V]
Time [μs] (Pulse width=0.5μs, interval=1μs)

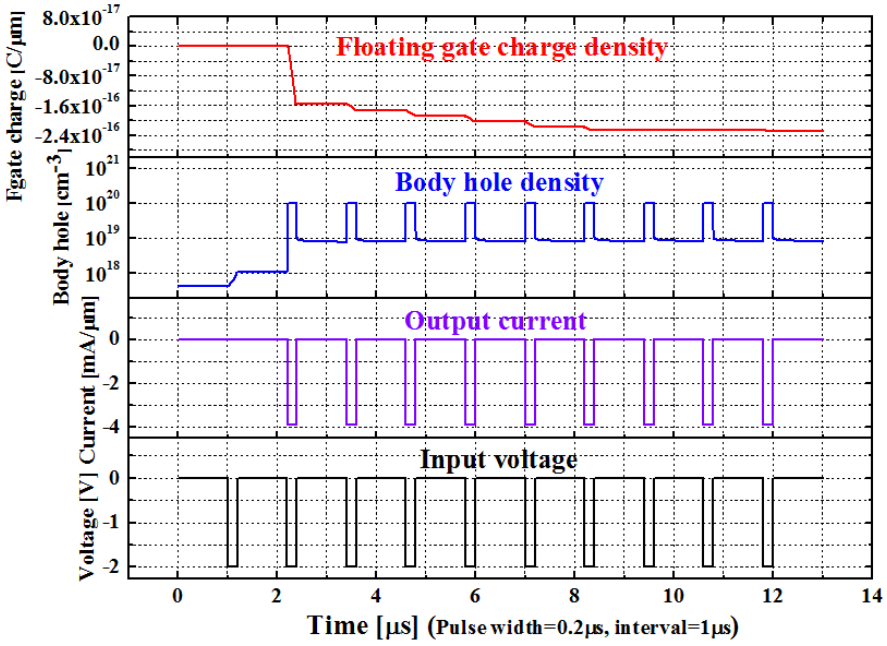
(c)

Fig. 2. 6. The transient tendency of the hole accumulation in the body and the hole injection into the floating gate when the width of an input pulse is changed by (a) 0.1 μs (b) 0.2 μs, (c) 0.5 μs.

becomes longer such as Fig. 2.6 (b), (c), the enough time for the triggering is secured and the triggering point is moved forward. Therefore, when the action potential generated by the neuron circuit is applied to the input of the synaptic device, the duration and the interval of the action potential determine whether short-term memory or long-term memory is generated in the synaptic device.



(a)



(b)

Fig. 2. 7. The transient tendency of the hole accumulation in the body and the charge injection into the floating gate (a) when the conditions of short-term potentiation (V_{BG} = 0 V) and (b) the long-term depression ($V_G = V_{DS}$ = -2 V, V_{BG} = 2 V) are applied to the synaptic device.

As the voltage pulse such as $V_G = V_{DS}$ = 2 V, V_{BG} = 0 V is applied as shown in Fig. 2. 7 (a), the hole increase from the impact-ionization appears without charge injection into the floating gate, which means short-term potentiation (STP). Since we have designed the output of the neuron circuit is directly connected to the back-gate of the synaptic devices,

the charge injection into the floating gate implying long-term memory cannot be formed without the fire of the system. As the voltage pulse such as $V_G = V_{DS} = -2$ V, $V_{BG} = 2$ V is applied as shown in Fig. 2. 7 (b), the hot electrons are generated near the floating gate and start to enter the floating gate when the overall drain-body junction is forward-biased due to the accumulated holes in the body. Since the electrons are injected to the floating gate, the V_T increases and it means the long-term depression of synaptic weight. However, in order for electron injection from the body to the floating gate to occur, there must be a positive feedback where holes are accumulated in the body and the barrier between the body and the drain is reduced. Therefore, it is difficult to implement a short-term depression in the SFST.

In case of the p-channel synaptic device, it can realize the short-term potentiation when the electrons from the impact-ionization are accumulated in the body and the long-term potentiation when the electrons are injected into the floating gate. Since the type of charge that induces the memory effect is different from that of the n-channel synaptic device, there is a slight difference in the magnitude of weight change.

2.3 Spike-Timing-Dependent-Plasticity (STDP) in SFST

The spike-timing-dependent-plasticity (STDP) is the theory that the long-term weight of the synapse is determined by the causal relationship between the pre-signal and post-signal passing through the synapse. If the post-signal occurs after the pre-signal is applied to the synapse, the synapse becomes stronger by judging that the causal relationship between the two signals is established. At this time, the smaller the time difference between the two signals, the stronger the weight of the synapse. Conversely, if a pre-signal is not applied to the synapse but the post-signal is generated through another synapse connection, the strength of the synapse may be weakened by judging that there is no causal relationship

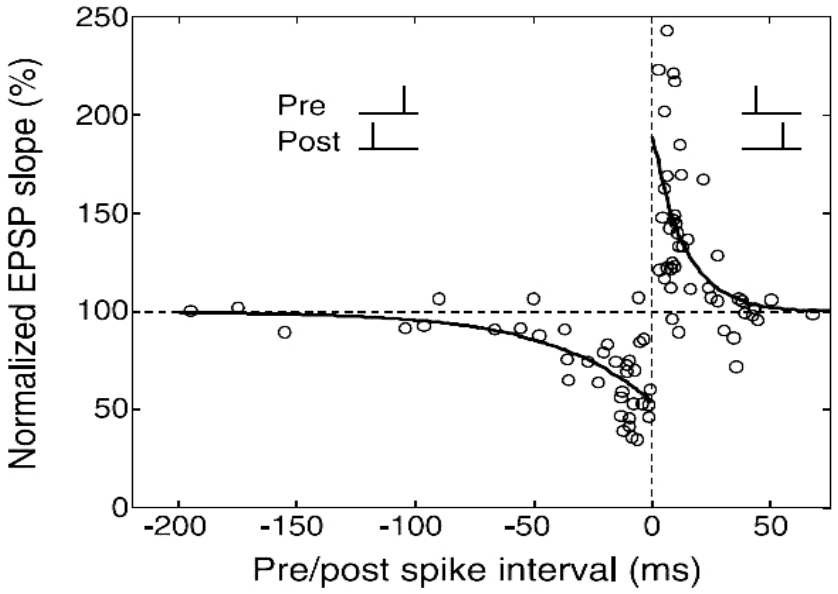


Fig. 2. 8. STDP characteristic of biological synapse.

between the two signals. Similar to the tendency of potentiation, the smaller the time difference between the two signals, the weaker the synaptic weight becomes. Fig. 2. 8 shows the STDP characteristic of biological synapse. As described above, an asymmetric STDP curve can be observed for the time difference between the two signals.

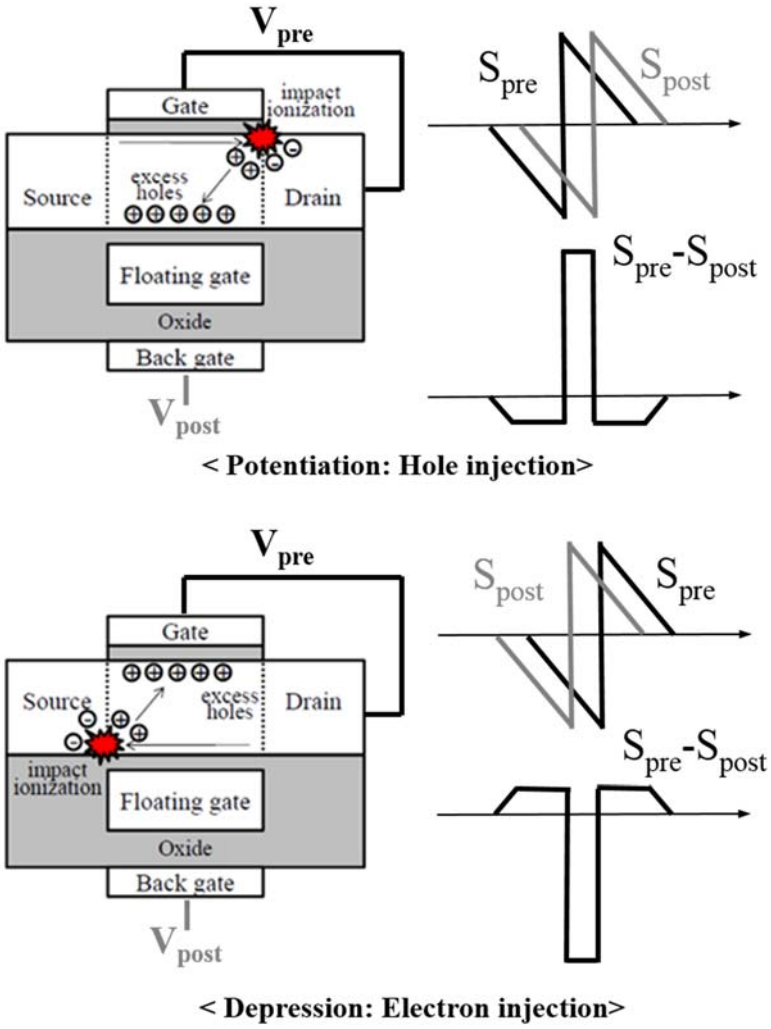
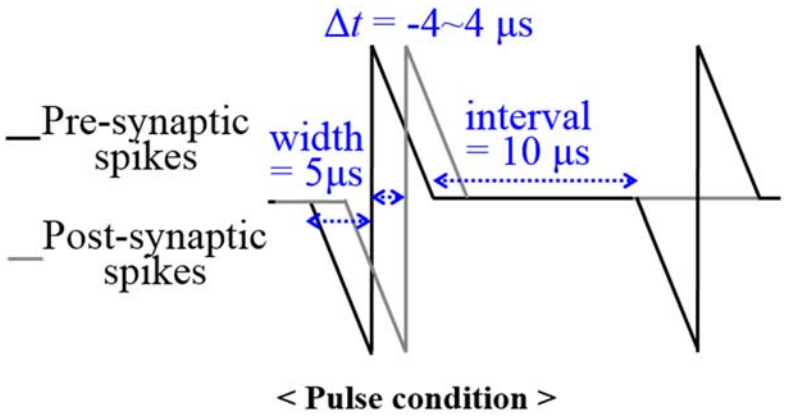
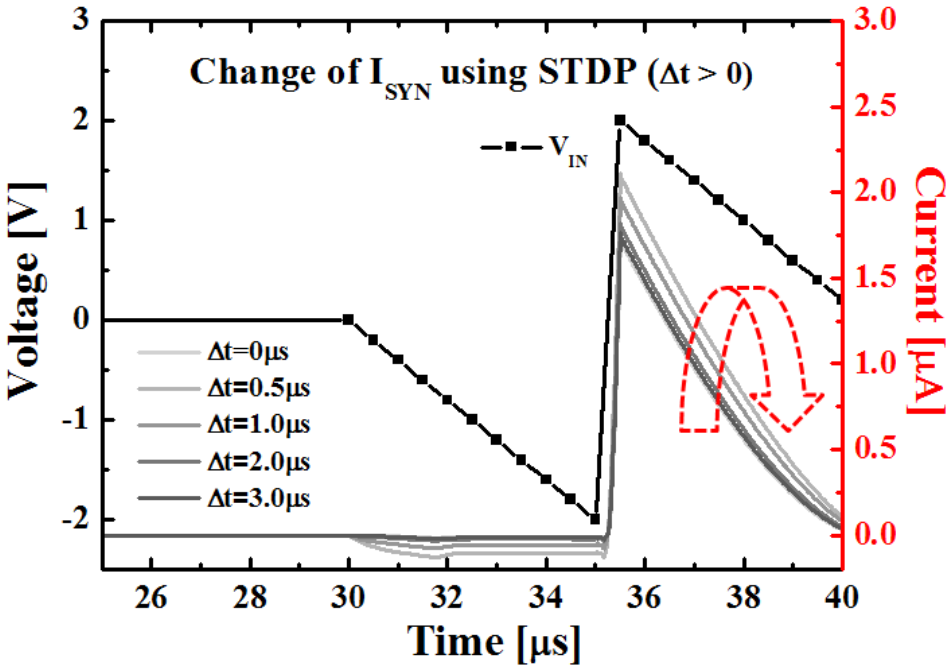


Fig. 2. 9. The principle of implementation of STDP in SFST.

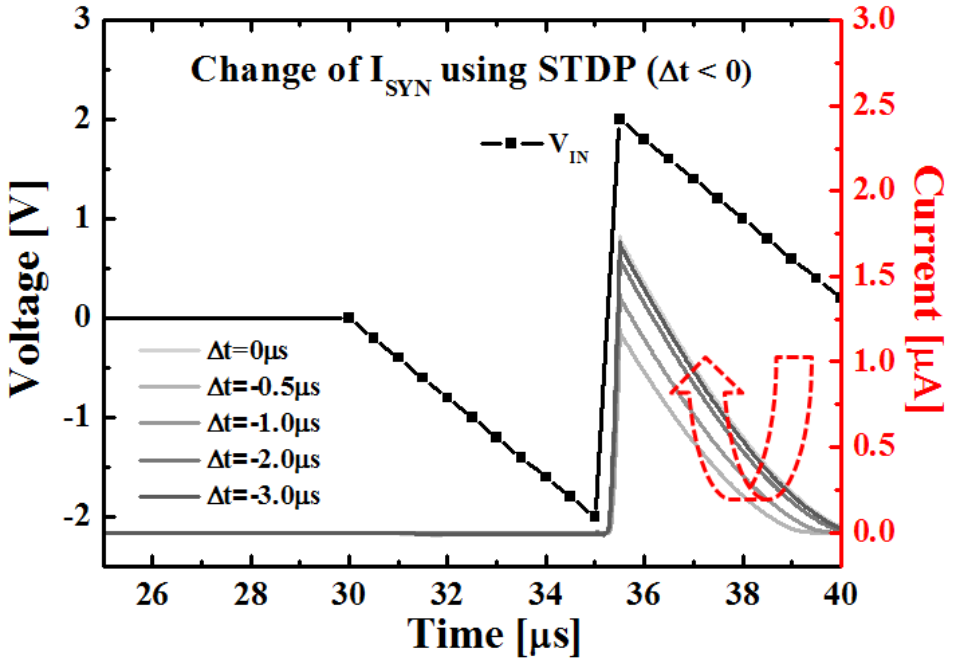
Fig. 2. 9 shows the principle of implementation of STDP in the SFST. In the proposed system, the timing difference between the input action-potential and the output action-potential connected to the back gate of the SFST decides the weight of the long-term memory. For the implementation of the STDP, the input pulse is simultaneously applied to the gate and the drain of the SFST and the output pulse is connected to the back gate of the SFST. Also, even though the action potential in biological neuron has positive value ahead of negative value in time domain, the action-potential in the proposed system has opposite order. As the output pulse is generated immediately after the input pulse is applied, the special situation that maximum positive voltage is applied to the gate and the drain and maximum negative voltage is applied to the back gate takes place. Therefore, more hot holes enter the floating gate and the conductance increases from the initial state. The increase is correspond to stronger long-term potentiation of biological synapse. As the output pulse fires slowly after the input pulse, the potentiation sharply decreases. If there is no causation between the input and the output pulses, the output pulse may fire before the input pulse of synapse. As the timing difference between the input and the output pulse becomes smaller, the special situation that maximum negative voltage is applied to the gate and drain and maximum positive voltage is applied to the back gate takes place. Therefore, more electrons enter the floating gate and the conductance decreases from the initial state. The decrease is correspond to stronger depression of biological synapse. The depression also sharply decreases as the timing difference between the input and the output pulses increases.


Δt = -4~4 μs
Pre-synaptic spikes
Post-synaptic spikes
width = 5μs
interval = 10 μs
< Pulse condition >

(a)


Change of I_SYN using STDP (Δt > 0)
V_IN
Δt=0μs
Δt=0.5μs
Δt=1.0μs
Δt=2.0μs
Δt=3.0μs
Voltage [V]
Current [μA]
Time [μs]
3
2
1
0
-1
-2
3.0
2.5
2.0
1.5
1.0
0.5
0.0
26
28
30
32
34
36
38
40

(b)



(c)

Fig. 2. 10. (a) The condition of input pulse for obtaining STDP characteristics. (b) The change of current when the time difference between input and output pulse is positive or (c) negative.

As mentioned earlier, in order to implement STDP in SFST, pre-signal and post-signal like Fig. 2. 10 (a) are applied to the corresponding terminal of SFST. After the first pre-signal and the post-signal having a certain time difference are applied to the synaptic device at the same time, the current value when the second pre-signal is applied alone is read to observe how the conductance of the device changes. The width and interval of the pulse

applied to the device were 5 μs and 10 μs, respectively, and the time difference between pre-signal and post-signal was changed from -4 μs to 4 μs. As shown in Fig. 2. 10 (b), when the post-signal occurs after the pre-signal, the current increases largely when the time difference between the two signals is 0.5 μs, and the increase is gradually decreased as the time difference increases. On the contrary, as shown in Fig. 2. 10 (c), when the post-signal occurs before the pre-signal, the current decreases largely when the time difference between the two signals is - 0.5 μs, and the decrease is gradually decreased as the time difference increases. Fig. 2. 11 summarizes the experimental results of Fig. 2. 10.

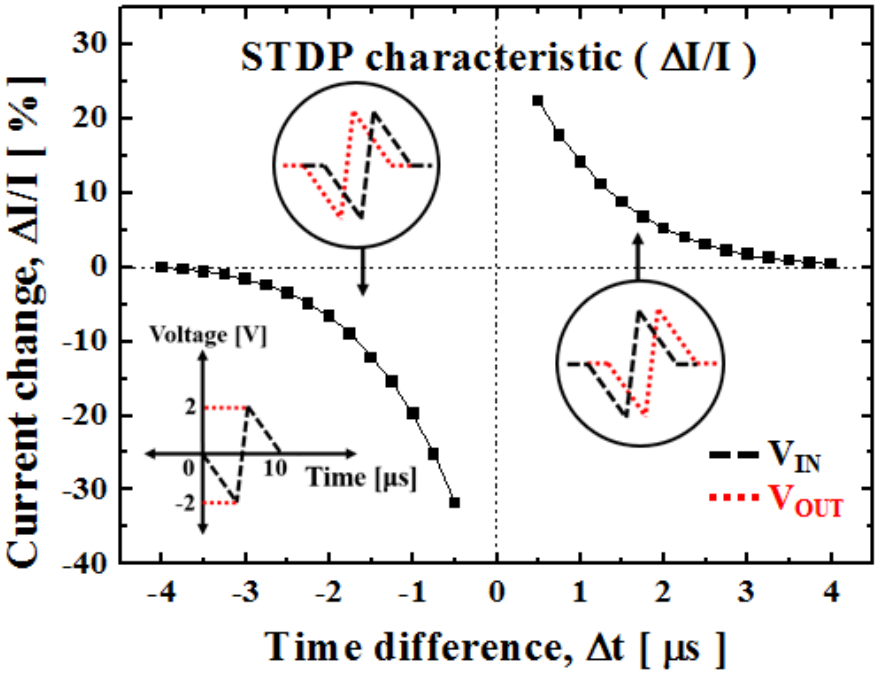


Fig. 2. 11. STDP characteristic that induces current change of SFST.

2.4 Fabrication of SFST

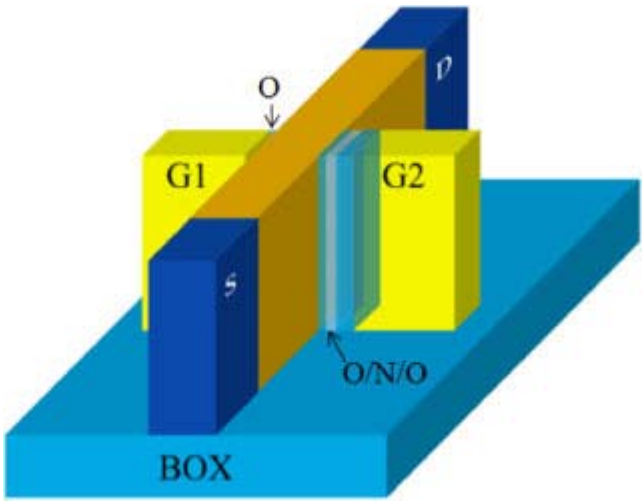


Fig. 2. 12. The schematic of Si-based Floating-Body Synaptic Transistor [33].

Si-based Floating-Body Synaptic Transistor (SFST) was fabricated on an SOI wafer, and compared to other vertically stacked devices, the device was fabricated horizontally to complete a dual gate structure as shown in Fig. 2. 12. The vital steps are two-step chemical mechanical polishing (CMP) processes which enable the first gate (G1) and the second gate (G2) to have different gate stacks from each other and be separated. In case of Oxide/Nitride/Oxide layers designed for charge injection, they were fabricated using a side wall spacer. The process flow for fabricating a dual-gate-type lateral SFST is shown in Fig. 2. 13 and cross sectional HR-TEM image of active fin defined by sidewall and two separated gates is shown in Fig. 2. 14.

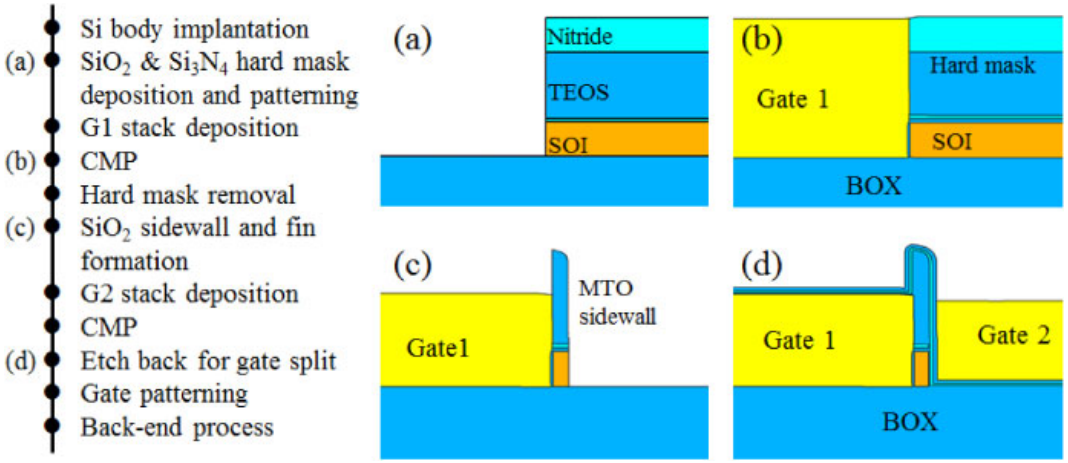


Fig. 2. 13. Fabrication process flow of the asymmetric dual-gate structure [27].

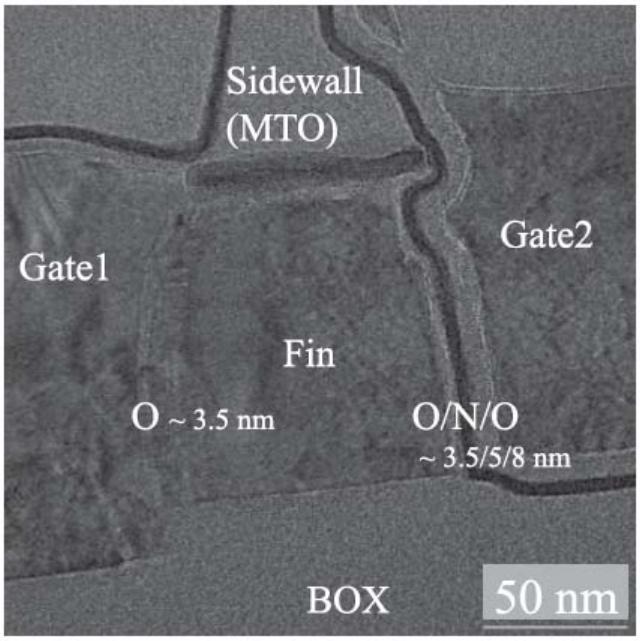


Fig. 2. 14. Cross sectional HR-TEM image of active fin defined by sidewall and two separated gates [27].

2.5 Operation Characteristic of Fabricated SFST

In order to verify the implementation of short-term memory and long-term memory in SFST, which we confirmed through simulation, we measured the operation characteristics of the fabricated SFST. The transient measurements were made by increasing the number of square pulses with a width of 1 μs, and the interval between pulses was increased from 10 μs to 100 μs. The size of the pulse was set to the condition of $V_{G1} = V_D = 2$ V. $V_{G2} = -2$ V, $V_S = 0$ V.

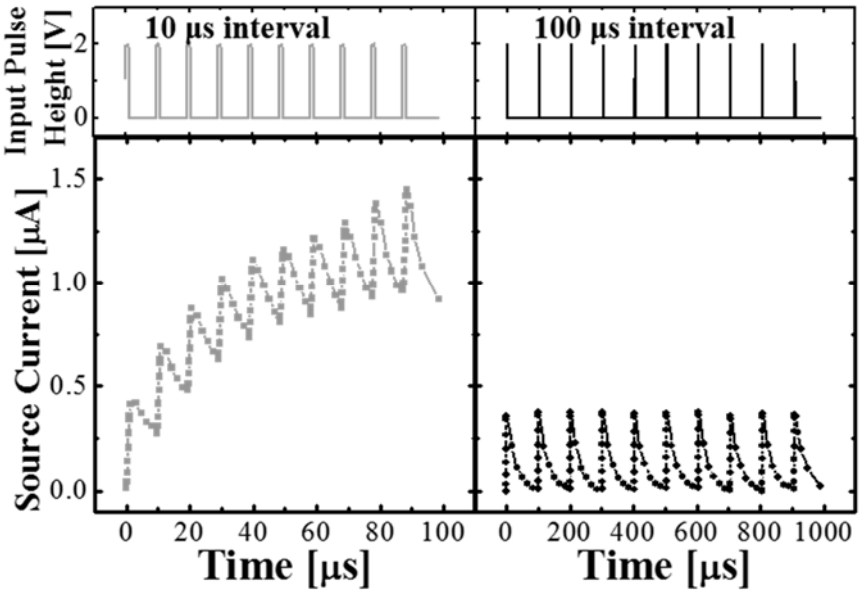


Fig. 2. 15. Measured transient responses of source current when the device learning through several times of input pulses with different interval times (a) 10 μs and (b) 100 μs [27].

As shown in Fig. 2. 15, when 10 pulses are applied to the device at interval of 10 μs, the source current increases while the pulse is applied and the current decreases while the pulse is turned off. However, since the interval between pulses is rather short, the net current continues to increase as the number of applied pulses increases. If the interval increases to 100 μs, the net current does not increase at all during the 10 pulses, since the current increased when the pulse is applied decreases for a relatively long interval time.

Fig. 2. 16 shows the results of a retention measurement under the condition of V_D = 1 V after applying pulses to the device under the same conditions as the experiment in Fig. 2. 14. When the interval between pulses is 10 μs, if the number of applied pulses is less than 6, the increased current is not visible compared to the initial state in retention measurement. However, if more than 7 pulses are applied, the current is increased compared to the initial state even after the pulse is turned off. Therefore, we can confirm that long-term memory is formed in the SFST through the experiment condition. When the interval between the pulses is 100us, no increase in current is observed in the retention measurement, no matter how many pulses are applied. The reason for this difference due to the interval change is that if the interval is relatively short, the positive feedback can be maintained through the hole still in the body of the SFST until the next pulse is arrived, but if the interval becomes longer, since the holes that were present disappear through recombination, the memory effect cannot be formed through the feedback.

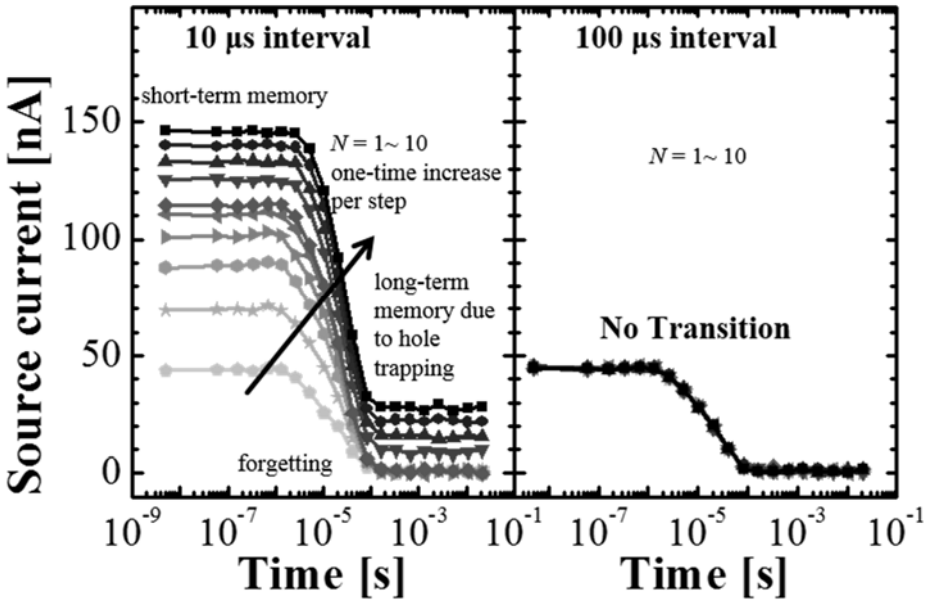


Fig. 2. 16. Measured retention characteristics of source current under the read condition (V_D = 1 V) after several times of input pulses with different interval times as (a) 10 μs and (b) 100 μs.

Chapter 3

Integrate & Fire Neuron Circuit

In this chapter, we present an integrate & fire (I&F) neuron circuit that drives the Si-based floating-body synaptic transistor as described in chapter 2 and transmits the action potential through the construction of spiking neural network. We will look at how the integrate & fire neuron circuit is designed to have the configuration and operation method so that it can be applied to the spiking neural network efficiently with minimal power consumption and without the help of additional circuit elements. In particular, we will concentrate on the strengths of the circuit compared to other previous studies through the generation of asymmetric pulses and autonomic weight update of the synaptic device.

3.1 Operation Principle of I & F Neuron Circuit

We have designed the neuromorphic system including SFST and I&F circuit as shown in Fig. 3. 1. The scale parameters used in the system are disclosed in Table I. The proposed system means one neuron and several synaptic devices can be connected to the input part of the system, respectively. The neuron circuit was constructed by a total of 14 MOSFETs

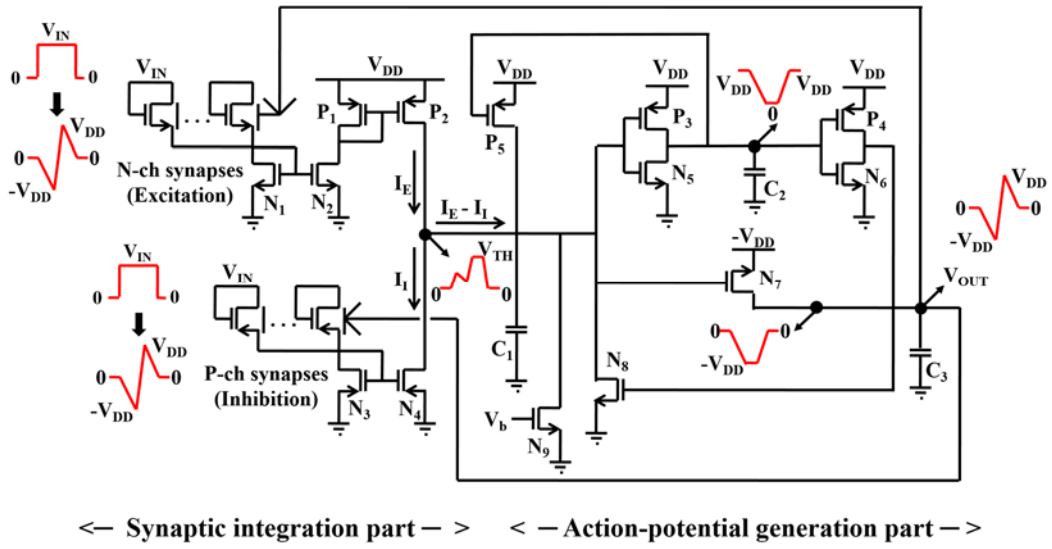


Fig. 3. 1. Schematic diagram of integrate & fire neuron circuit.

TABLE I

SCALE PARAMETERS OF PROPOSED SYSTEM

Device	Width/Length (μm/μm)	V_T (V)	T_{OX} (nm)	Device	Capacitance (pF)
N_1~N_4, N_9	0.02/1	0.6	10	C_1	0.15
N_5	0.04/1	0.6	10	C_2	0.1
N_6	0.01/10	0.6	10	C_3	0.01
N_7	0.05/1	2.5	10		
N_8	0.05/1	0.6	10		
P_1~P_2	0.02/1	-0.6	10		
P_3	0.04/1	-0.6	10		
P_4	0.1/1	-0.6	10		
P_5	0.04/1	-0.2	10		

and 3 capacitors, except for synaptic devices. The system consists of a synaptic integration part and an action-potential generation part. The integration part integrates pre-neuron signals and transmits the signals to post-neurons. As an input signal is applied to the

synaptic device, the device current flows into capacitor C_1 using current mirror. The role of the current mirror is integrating the signals from several synapses and isolating the source terminal of the synaptic device from the integrated potential of the capacitor. The integration part is also divided into an excitation part and an inhibition part. The excitation part including n-channel synapses with double current mirrors generates the current in the direction of increasing the capacitor voltage and the inhibition part including p-channel synapses with a single current mirror generates the current in the reverse direction to the excitation part. As the currents in different directions flow into the same capacitor, the net currents between top and bottom mirrors decide the firing of the neuron circuit. The reason for using different types of synaptic device is that because the excitatory synapse and the inhibitory synapse have opposite STDP rule [10].

As the enough currents from the integration part are flowed to a capacitor C_1, the voltage of the C_1 enough to make output pulse is built up and the system fires with giving feed-back to the synaptic device. The output pulse of the generation part is directly connected to the back gate of the synaptic device in the integration part for expression of STDP. Since STDP is implemented by directly connecting the output of the circuit with the synaptic device without additional circuit elements, the weight-update system can be configured very efficiently. In addition, even though the action potential in biological neuron has positive value ahead of negative value in time domain, the action-potential in the proposed system has opposite order. The same learning rule as the STDP rule that occurs in biological synapse can be applied to the synaptic device through the reversed pulse.

3.2 Operation Characteristics of I&F Neuron Circuit

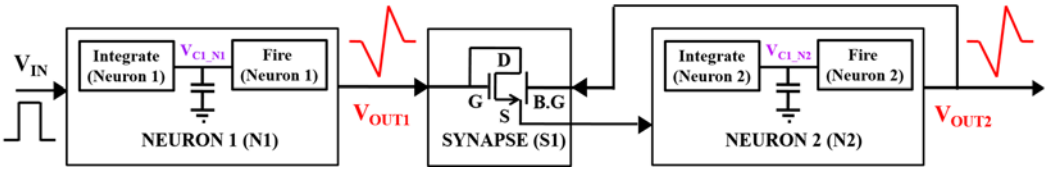
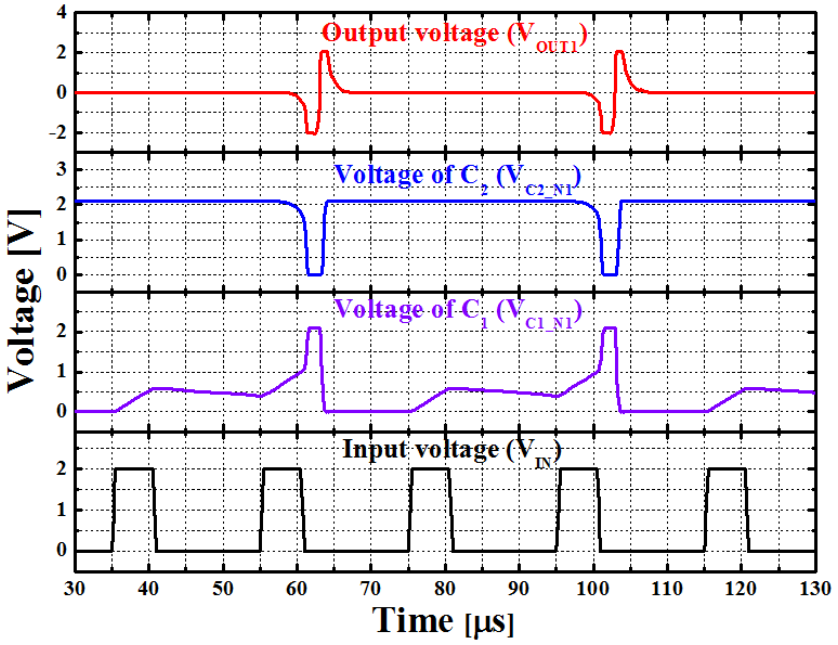


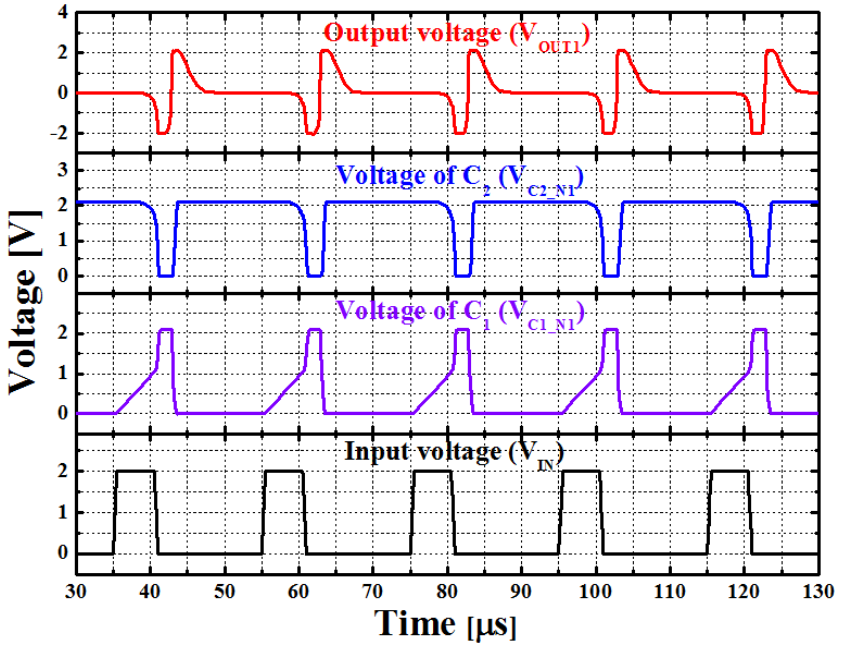
Fig. 3. 2. The synaptic device (S1) receives pre-neuron signal from the neuron 1 (N1) and transmits the signal to the neuron 2 (N2). The post-neuron signal from the N2 is transmitted to other synapses with giving back-propagation signal to the S1.

Fig. 3. 2 shows the connection between the synaptic device and the neuron circuits used in the extraction of output characteristic of the proposed system. The synaptic device (S1) receives pre-neuron signal from the neuron 1 (N1) and transmits the signal to the neuron 2 (N2). The post-neuron signal from the N2 is transmitted to other synapses with giving back-propagation signal to the S1. The initial signal applied to the N1 is square pulse for convenience.

Fig. 3. 3 shows the operation characteristics of the N1. The action-potential generation part in the proposed system creates an action-potential when the node voltage of capacitor C_1 (V_{C1}) exceeds the threshold value. In order to generate the action-potential from negative to positive value, double-stage inverters, the NMOS (N_7) having large V_T value between V_{DD} and V_{IN} +V_{DD} and the feed-back NMOS (N_8) are used. The double-stage inverters


Output voltage (V_{OUT1})
Voltage of C_2 (V_{C2_N1})
Voltage of C_1 (V_{C1_N1})
Input voltage (V_{IN})
Voltage [V]
Time [μs]

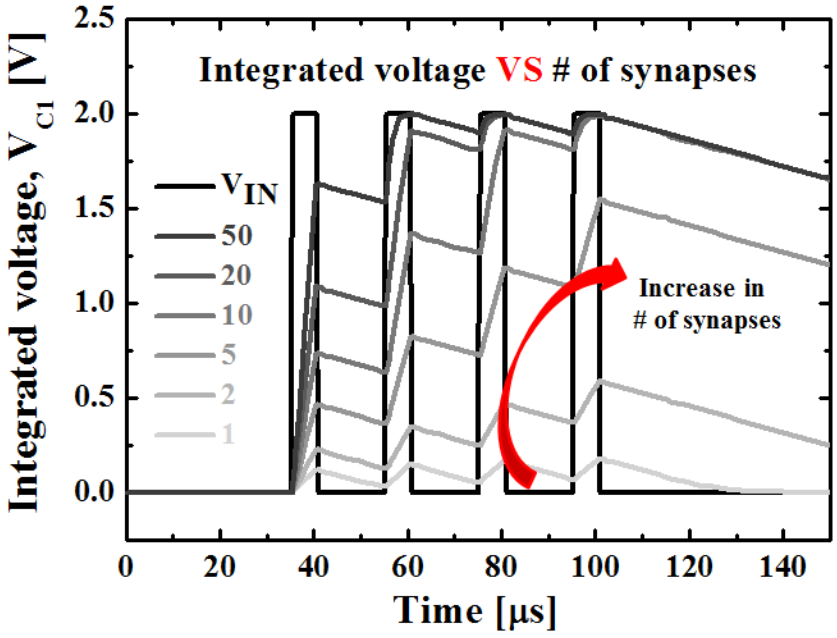
(a)


Output voltage (V_{OUT1})
Voltage of C_2 (V_{C2_N1})
Voltage of C_1 (V_{C1_N1})
Input voltage (V_{IN})
Voltage [V]
Time [μs]

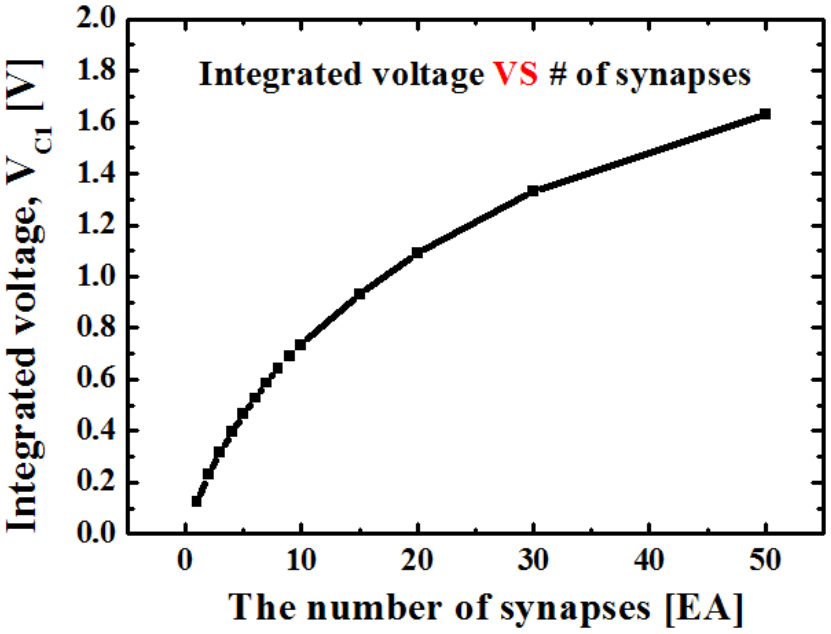
(b)

Fig. 3. 3. The operation characteristic of the proposed system. (a) The second input pulse makes the V_{C1} high enough to exceed the threshold value of the generation part. (b) As the number or the weight of the synaptic devices increases, the V_{C1} rapidly increases and the firing of the neuron takes place earlier.

reverse their input pulse by turns and ensure delay time in order to make the action potential have negative part ahead of positive part. In case of Fig. 3. 3 (a), the second input pulse makes the V_{C1} high enough to exceed the threshold value. When the V_{C1} comes close to the threshold value, the P_5 turns on first and drives up the V_{C1} for stable operation. When the generation part turns on, the negative output of the N_7 appears first and the positive delayed output from the double-stage inverters appears later to switch on the N_8. The N_8 discharges the V_{C1} and returns the system to initial state. In addition, the N_9 is used to emulate the leaky integration of biological systems. Although the negative part of the input action-potential has no direct effect on the increase of the V_{C1}, it can change the conductance of the synaptic device through the combination with the back-gate pulse originated from output node. As the number or the weight of the synaptic devices increases, the V_{C1} rapidly increases and the firing of the neuron takes place earlier as shown in Fig. 3. 3 (b). Changes in the timing of the neuron's firing affect not only the STDP rule that controls the weight of the synapse but also the speed at which the overall system operates, so it is necessary to carefully consider the timing of the neuron's firing when designing the system.


Integrated voltage VS # of synapses
Integrated voltage, V C1 [V]
Time [μs]
V IN
50
20
10
5
2
1
Increase in
# of synapses
2.5
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80
100
120
140

(a)


Integrated voltage VS # of synapses
Integrated voltage, V C1 [V]
The number of synapses [EA]
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50

(b)

Fig. 3. 4. (a) The transient characteristic of integrated voltage (V_{C1}) with increasing the number of synapses. (b) The change of V_{C1} with increasing the number of synapses when one input pulse is applied.

Fig. 3. 4 (a) shows the results of examining the voltage change in C_1 while increasing the number of synapses when four input pulses are applied to the circuit. At this time, in order to check the transition of V_{C1} continuously, we set the circuit not to fire while the pulse was applied. As can be easily expected, V_{C1} increases rapidly as the number of synapses increases. In order to analyze the increase tendency of V_{C1} according to the number of synapses in more detail, we plotted the increase of V_{C1} according to the number of synapses when one pulse was applied as shown in Fig. 3. 5 (b). Initially, the value of integrated V_{C1} increases almost linearly with the number of synapses, but as the number of synapses increases, the increase in V_{C1} gradually slows down. This is due to the fact that the current mirror connecting the synaptic device to C_1 does not operate linearly, and for this reason it is somewhat difficult to increase the number of synapse elements connected to neurons infinitely. This effect can be mitigated as much as possible by adjusting the scale parameters of the synaptic device and the components of the current mirror.

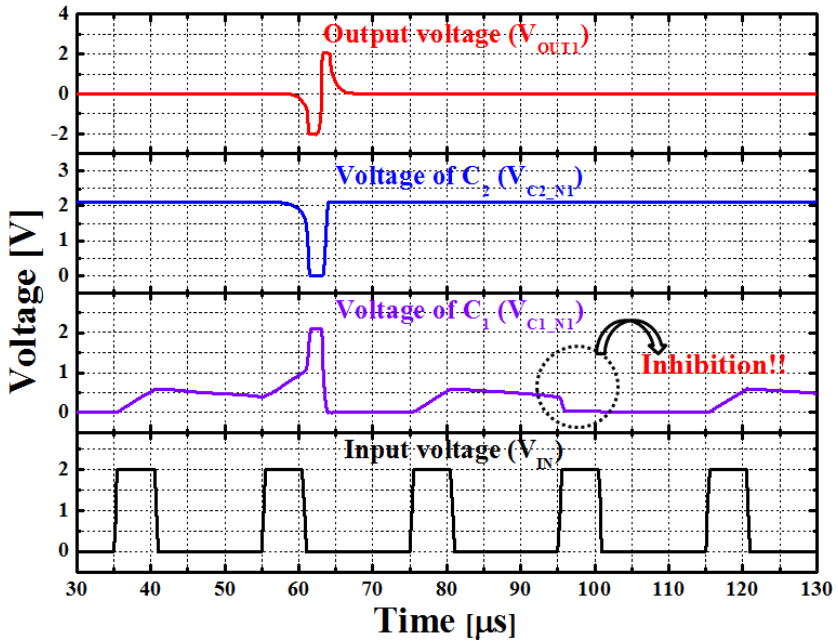


Fig. 3. 5. As the 4th input pulse is applied to the inhibitory synapse at the same time, the V_{C1} is nearly tied up to ground.

Fig. 3. 5 shows that the activation of the inhibition part inhibits the firing of the system. We assumed that the 4th input pulse is applied to the excitatory and inhibitory synapse at the same time for convenience. As the input pulse is applied to the both synaptic devices, the net current from the excitation and inhibition part flows into the C_1 and the V_{C1} is nearly tied up to ground. Through the inhibition, we can select the specific neurons from whole system and cut down on needless power dissipation. In case of the pattern recognition applications [14], the inhibition is used to inhibit the specific neurons that do not need to be fired for removal of noise.

3.3 STDP in the Connection of Synaptic Device and I&F Neuron Circuit

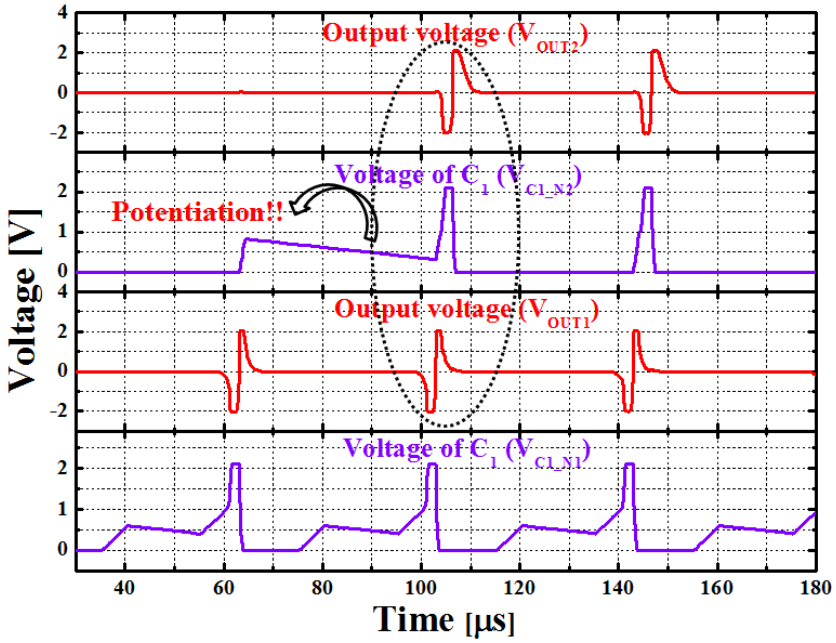
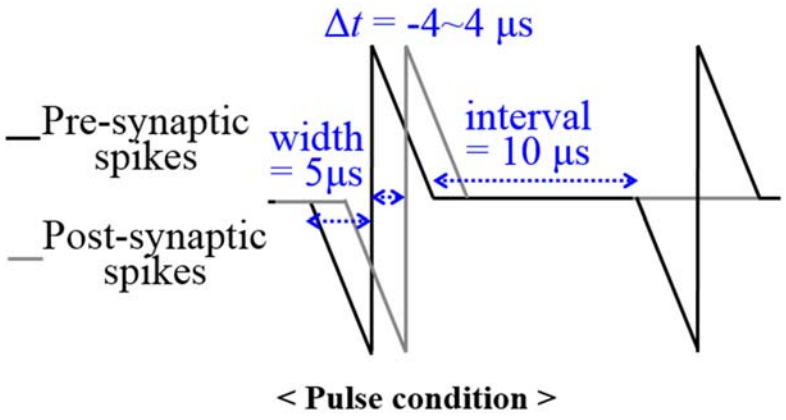


Fig. 3. 6. The weight update originated from the timing difference between V_{OUT1} and V_{OUT2} affects the fire of N2.

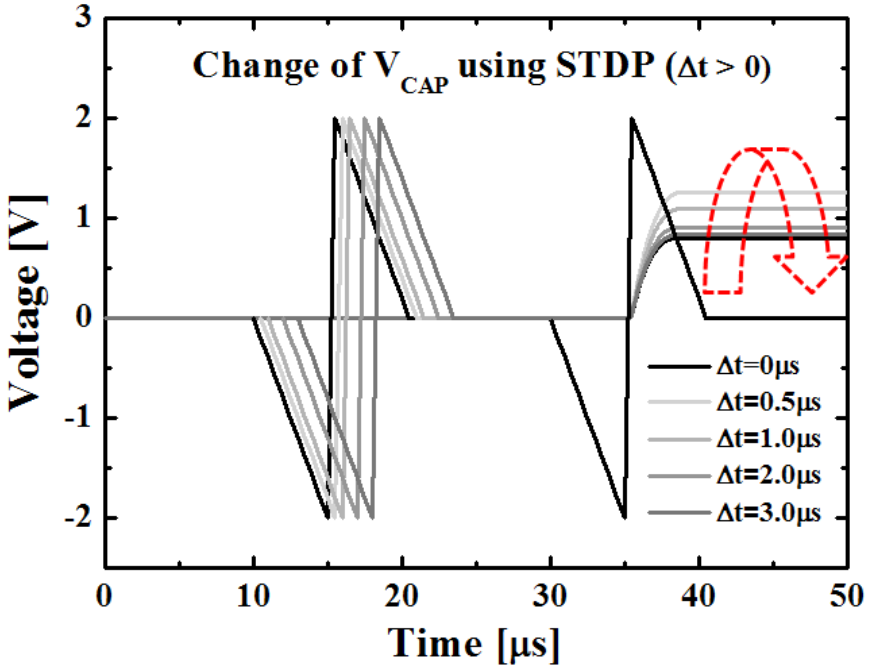
Fig. 3. 6 shows that the weight update of the S1 affects the fire of the N2 in the Fig. 3. 2. When a square input pulse is applied twice, the V_{C1} of the N1 exceeds the threshold point and the N1 fires with generating an asymmetric action potential. The action potential is transmitted to the N2 through the S1 and the V_{C1} of the N2 gradually increases. When the action potential of N1 is applied twice, the N2 fires with generating another action potential.

The action potential is transmitted to next synapses and went back to the back-gate of the S1 for weight update. In case of Fig. 3. 6, because post-signal is slightly applied later than pre-signal, the potentiation occurs in the S1. After the increase of the weight of the S1, only one action-potential is needed to fire the N2.

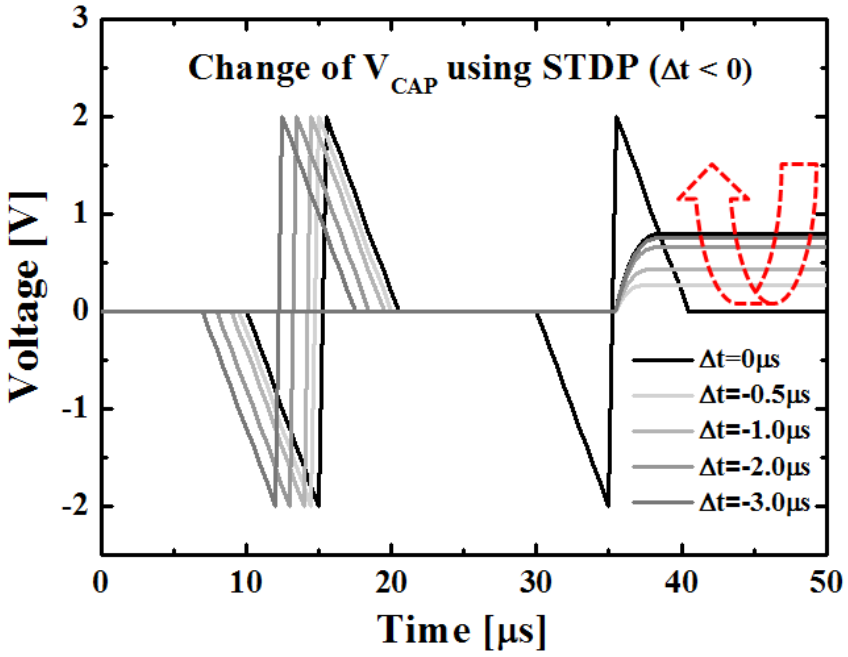
To see the change of V_{C1} while changing the time difference between pre-signal and post-signal more variously, pre-signal and post-signal like Fig. 3. 7 (a) are applied to the corresponding terminal of SFST. After the first pre-signal and the post-signal having a certain time difference are applied to the synaptic device at the same time, the V_{C1} when the second pre-signal is applied alone is read to observe how the conductance of the device changes. The width and interval of the pulse applied to the device were 5 μs and 10 μs, respectively, and the time difference between pre-signal and post-signal was changed from -4 μs to 4 μs. As shown in Fig. 3. 7 (b), when the post-signal occurs after the pre-signal, the V_{C1} increases largely when the time difference between the two signals is 0.5 μs, and the increase is gradually decreased as the time difference increases. On the contrary, as shown in Fig. 3. 7 (c), when the post-signal occurs before the pre-signal, the current decreases largely when the time difference between the two signals is - 0.5 μs, and the decrease is gradually decreased as the time difference increases. Similar to the current change in Fig. 2. 11, when the output pulse is generated slightly after the input pulse, the V_{C1} rapidly increases and the neuron fires well as shown in Fig. 3. 7 (d). Through the results of Fig. 2. 11 and Fig. 3. 7, we can confirm that our system transmits the pre-neuron signal and changes the weight of the synaptic device at the same time.


Δt = -4~4 μs
Pre-synaptic spikes
Post-synaptic spikes
width = 5μs
interval = 10 μs
< Pulse condition >

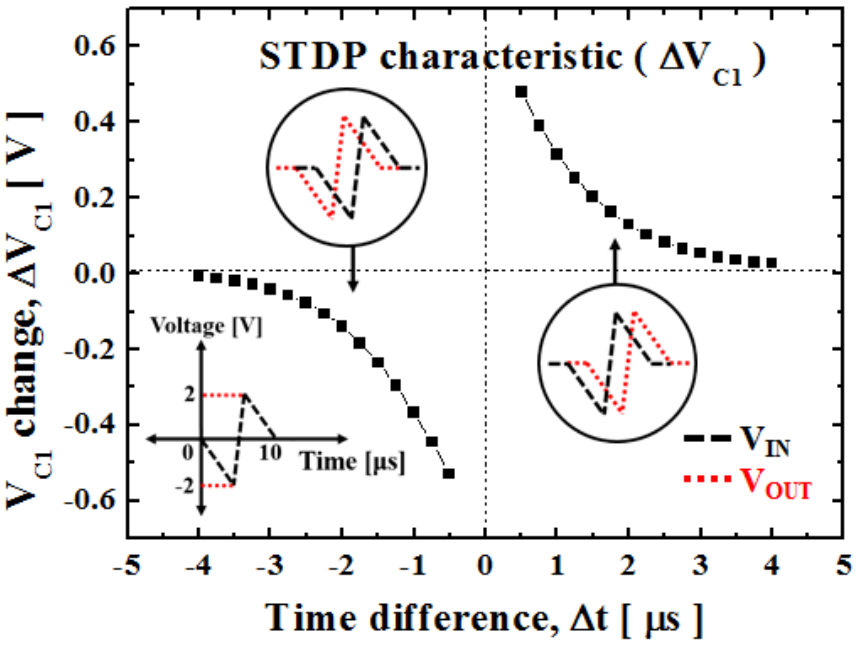
(a)


Change of V CAP using STDP (Δt > 0)
Voltage [V]
Time [μs]
Δt=0μs
Δt=0.5μs
Δt=1.0μs
Δt=2.0μs
Δt=3.0μs

(b)


Change of V_{CAP} using STDP (Δt < 0)
Voltage [V]
Time [μs]
Δt=0μs
Δt=-0.5μs
Δt=-1.0μs
Δt=-2.0μs
Δt=-3.0μs

(c)


STDP characteristic (ΔV_{C1})
V_{C1} change, ΔV_{C1} [V]
Time difference, Δt [μs]
Voltage [V]
Time [μs]
V_{IN}
V_{OUT}

(d)

Fig. 3. 7. (a) The condition of input pulse for obtaining STDP characteristics. (b) The change of V_{C1} when the time difference between input and output pulse is positive or (c) negative. (d) The STDP characteristic that induces the change of V_{C1} of SFST.

3.4 Power Dissipation of Proposed System

Table. II shows the comparison of the proposed system and other neuron models. The power consumption of the systems listed in the table is the value required for the system to generate an action potential and is estimated through the data presented in the paper. Our system uses 14 MOSFETs and 3 capacitors for the emulation of biological neural operation and consumes a total of ~3 pJ for the generation of an action-potential as shown in Fig. 3. 8. Compare to other neuron models in Table. II, the proposed system can emulate the operation of biological neurons efficiently with low power consumption. The most crucial reason for the proposed system to perform this efficient operation is to connect the output pulse directly to the back gate of the synaptic device by using a 4-terminal synaptic device. Therefore, the weight update of synapse can be autonomously processed without any additional circuit elements, and the system can be configured with a minimum number of devices. When using a two-terminal device such as Resistive Random Access Memory (RRAM) or Phase-change Memory (PRAM), the energy consumed by the device itself is small, but it is somewhat inefficient in terms of the system because it requires additional elements to implement the STDP rule.

TABLE II

COMPARISON OF THE PROPOSED SYSTEM AND OTHER NEURON MODELS

Neuron model	Number of transistors used	Power consumption	References
Conductance-based	27-30+	60 µW (> 0.24 µJ)	Mahowald and Douglas, 1991 [29]
Hindmarsh-Rose	90	163.4 µW (> 8.17 µJ)	Lee et al., 2004 [30]
Integrate-and-Fire	18-20	10 µW (~900 pJ)	Indiveri et al., 2006 [20]
Quadratic Integrate-and-Fire	14	8 – 40 µW (8.5 – 9.0 pJ)	Wijekoon and Dudek, 2008 [21]
Log-domain Izhikevich neuron	17+	20 µW (~1 nJ)	Van Schaik et al., 2010 [31]
Memristor-based	1 memristor + 3 transistor	60–110 µW (60 – 110 pJ)	Babacan et al., 2016 [32]
Proposed circuit	14 (w/o synapses)	~3 µW (~3 pJ)	This paper

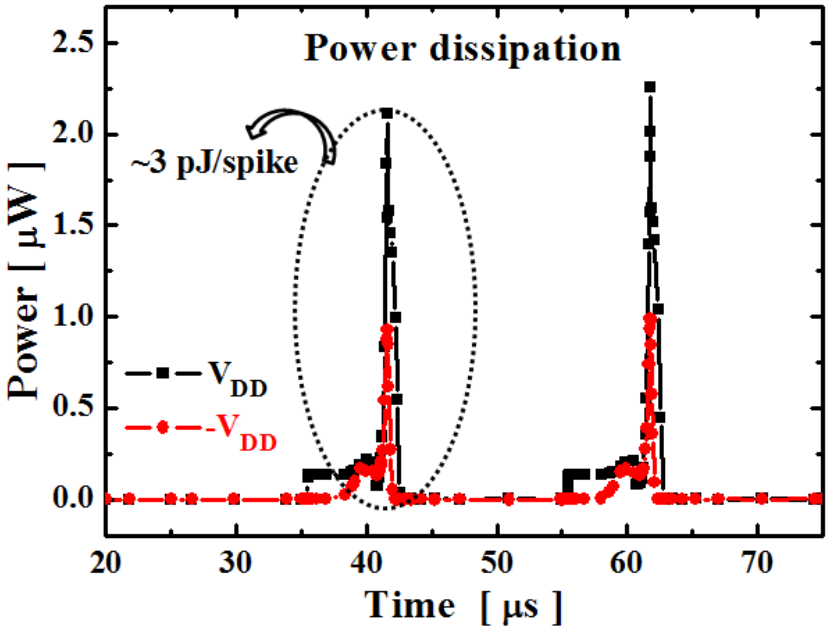
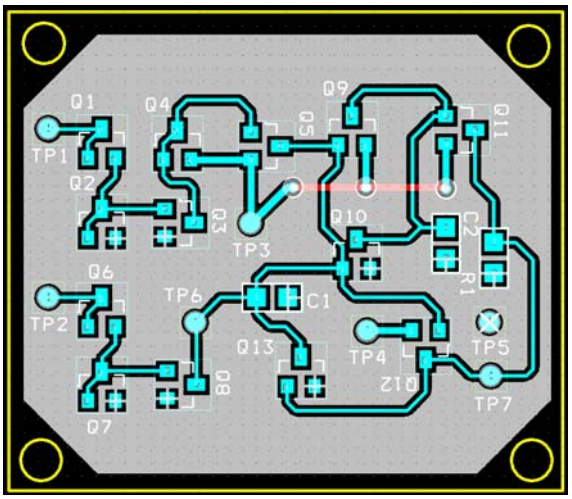
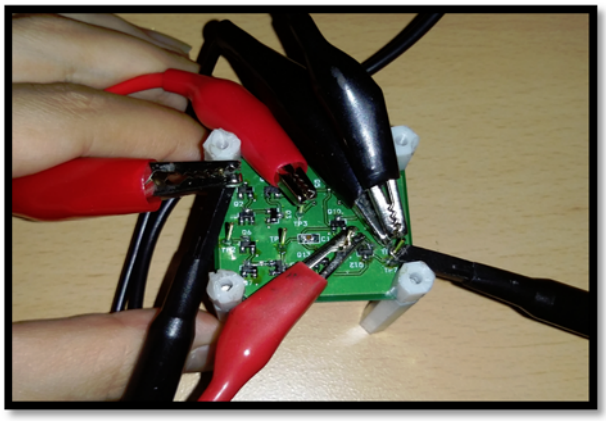


Fig. 3. 8. The power dissipation needed to generate an action potential in the system.

3.5 Implementation of I & F Neuron Circuit on Printed Circuit Board (PCB)



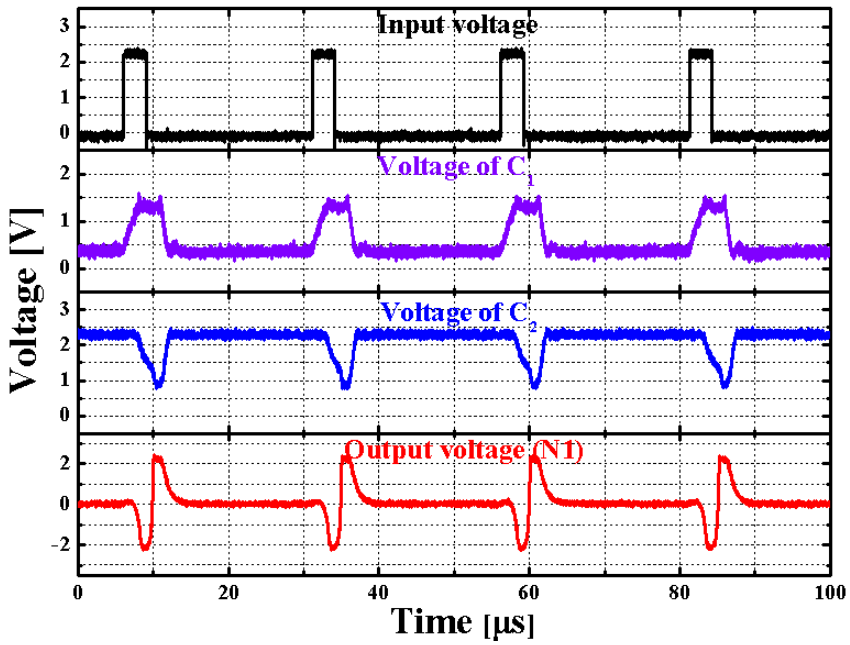
(a)



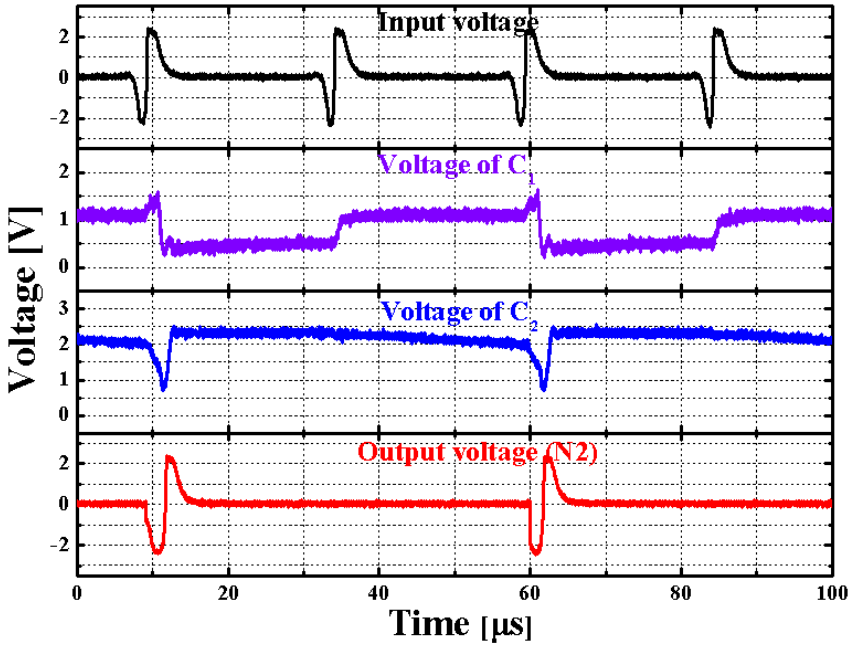
(b)

Fig. 3. 9. (a) Layout of neuron circuit and (b) a scene for measurement of fabricated PCB.

After verifying the operation of the neuron circuit and the synaptic device using simulation tool, we embodied the neuron circuit on printed circuit board (PCB). Fig. 3. 9 (a) shows the layout of neuron circuit for fabricating the PCB and Fig. 3.9 (b) shows a scene for measuring the output characteristics of the fabricated PCB. We utilized a DC supply, pulse generator, oscilloscope for measurement of the PCB. As we applied square pulse train having a size of 2.2 V and a width of 3 μs to a single neuron, the output characteristic such as Fig. 3. 10 (a) appears on the screen of oscilloscope. Similar to the simulation results, as the square pulse is applied, enough charge for generation of action-potential is piled up to capacitor C_1 and the action potential having intended form is made.



(a)



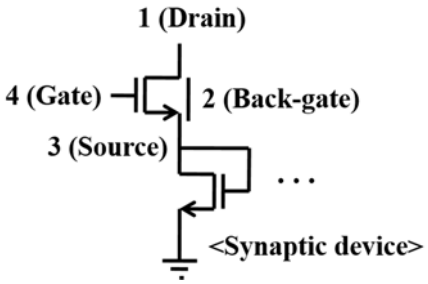
(b)

Fig. 3. 10 (a) Operation characteristic of single neuron circuit and (b) a pair of neuron circuit connected in series.

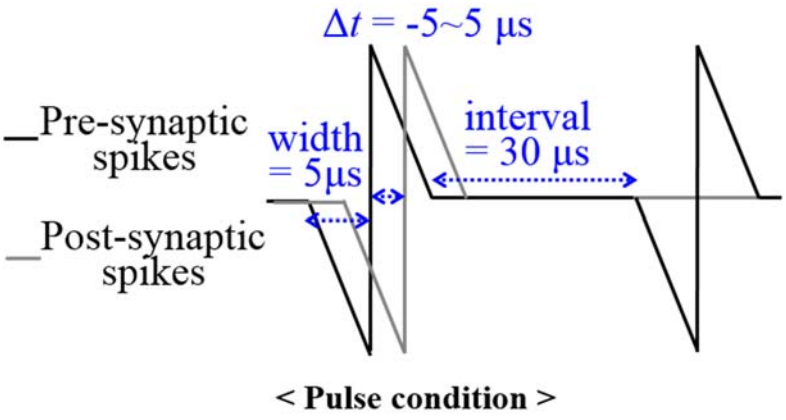
In case of Fig. 3. 10 (b), we connected two neuron circuits in series and the output pulse of the 1st circuit was directly applied to the input of 2nd circuit. When the output pulse from the 1st neuron is applied twice to the input terminal of the 2nd neuron, the voltage of C_1 passes over the threshold point of the 2nd neuron circuit and action potential is generated. We confirmed that the integrate & fire neuron circuit works well through the PCB fabrication and verifying the STDP characteristics through the connection with the synaptic device fabricated on Si wafer is another important experimental goal.

3.6 Measurement of Neuron Circuit on PCB and Synaptic Device on Wafer

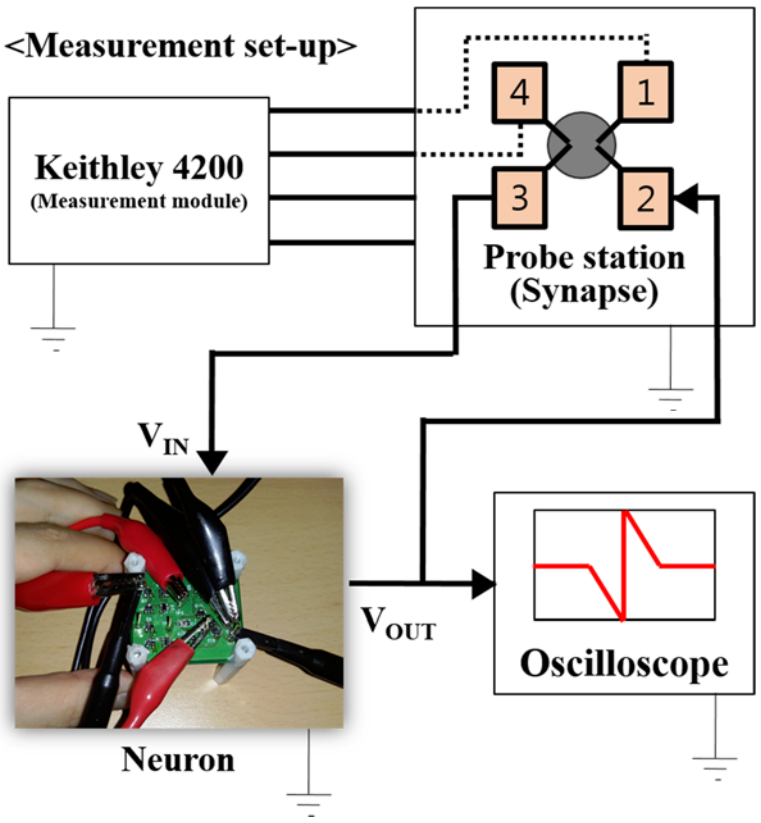
In order to measure the characteristics of a system composed of the neuron circuit fabricated on the PCB and the synaptic device fabricated on the Si wafer, the experimental environment as shown in Fig. 3. 11 was constructed. The synaptic device has a total of four terminals, and the source terminal of the synaptic device has to be connected in series with the NMOS at the bottom as shown in Fig. 3. 11 (a) to form the current mirror of the integration part of the neuron circuit. Since the neuron circuit on the PCB was made with the synaptic device left empty, the source terminal of the synaptic device was directly connected to the neuron circuit on the PCB. The synaptic device's gate terminal and drain terminal were pulsed with a condition corresponding to the pre-synaptic spike of Fig. 3. 11 (b) through a Keithley 4200 measurement module. The output terminal of the neuron circuit was directly connected to the back-gate terminal of the synaptic device. Therefore, the pulse corresponding to the post-synaptic spike was applied directly to the back-gate.



(a)


Δt = -5~5 μs
Pre-synaptic spikes
Post-synaptic spikes
width = 5μs
interval = 30 μs
< Pulse condition >

(b)


<Measurement set-up>
Keithley 4200
(Measurement module)
4
1
3
2
Probe station
(Synapse)
V_IN
V_OUT
Oscilloscope
Neuron

(c)

Fig. 3. 11 (a) The synaptic device connected in series with a NMOS. (b) The condition of input pulse for obtaining STDP characteristics. (c) The completed experimental conditions to measure the STDP characteristics.

The completed experimental conditions is illustrated in Fig. 3. 11 (c). A DC supply was used to apply the supply voltage of the neuron circuit and an oscilloscope was installed to check the point at which the output pulse occurred. We measured the source current change of the synaptic device while adjusting the time difference between -5 μs and 5 μs between the pre-signal generated in the measurement module and the post-signal generated in the neuron circuit on the PCB. The measurement results are shown in Fig. 3. 12, which shows how the conductance of the synaptic device changes compared to the initial state. The STDP curve plotted in Fig. 3. 12 has a shape similar to the STDP characteristic of biological synapse in Fig. 2. 8 and the STDP simulation characteristic shown in Fig. 2. 11 and Fig. 3. 7 (d). When the time difference between the input pulse and the output pulse is positive, the potentiation becomes stronger as the time difference decreases, and the potentiation becomes weaker as the time difference increases. On the contrary, when the time difference between the input pulse and the output pulse is negative, the depression becomes stronger as the time difference decreases, and the depression becomes weaker as the time difference increases. Even when the synaptic device and the neuron circuit are integrated on a single wafer, it can be expected that the synapse weight will be controlled autonomously through the time difference between the input pulse and the output pulse.

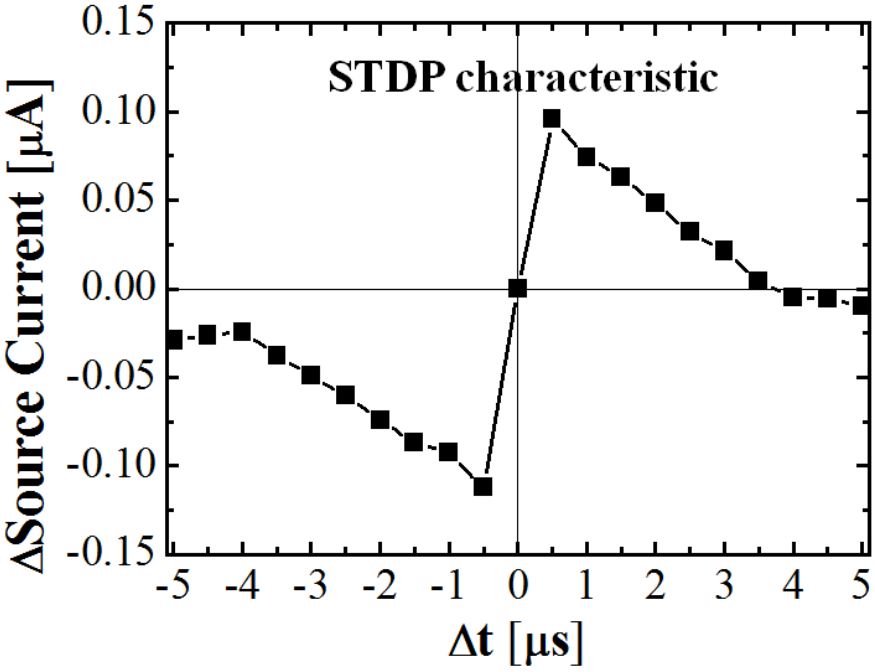


Fig. 3. 12. The measurement results which shows how the source current of the synaptic device changes compared to the initial state.

Chapter 4

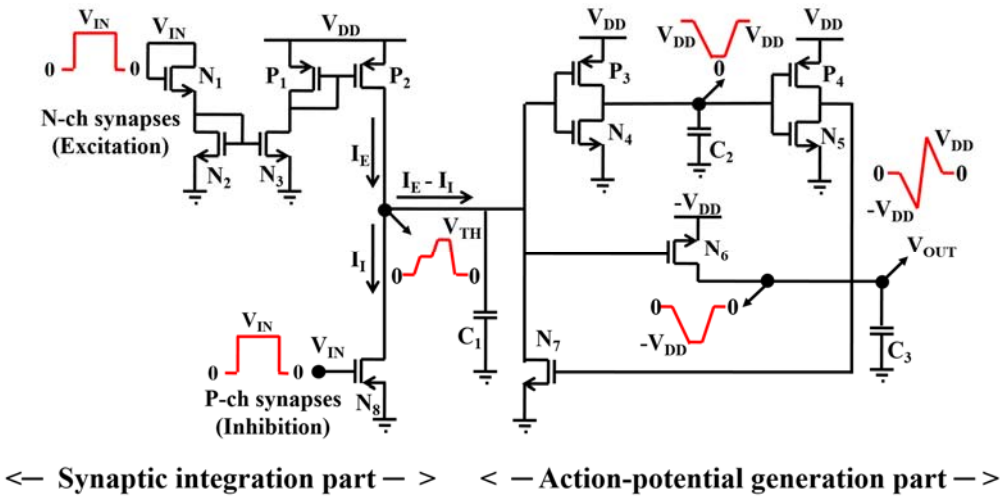
Fabrication of I & F Neuron Circuit

In this chapter, we describe the fabrication process of inverters, current mirrors, capacitors, NMOS, and PMOS, which plays a key role in driving the neuron circuit. The physical parameters of each element constituting the neuron circuit are determined by the simulated values so that the fabricated device can operate stably. After the fabrication of the neuron circuit, we examined whether each element of the fabricated neuron circuit works correctly, and verified that the fabricated neuron circuit has characteristics similar to those obtained through simulation. As a result of analyzing the measured contents, it can be expected that the designed circuit can be realized through the actual process and it is possible to construct one system by integrating with the Si-based synaptic device.

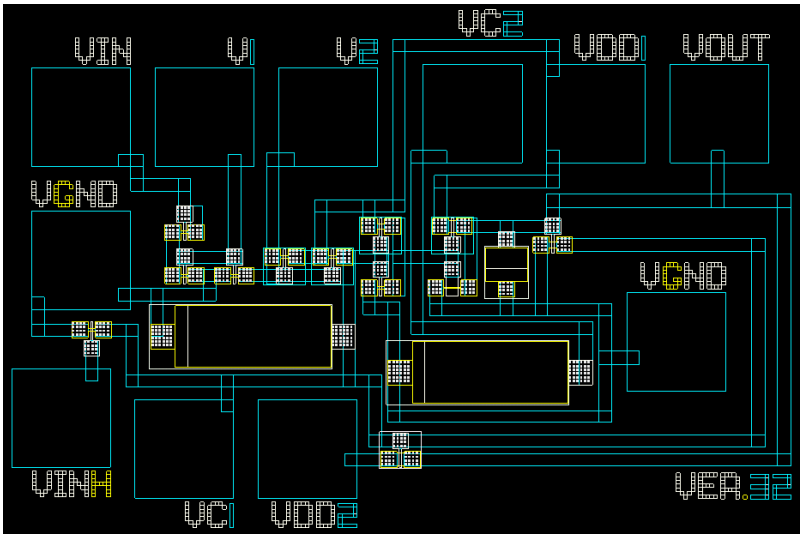
4.1 Mask Layout and Process Flow

The proposed neuron circuit consists of NMOS and PMOS with normal V_T, NMOS with high V_T, and MOS capacitor with high body doping. The neuron circuit to be fabricated includes signal integration using a current mirror and action-potential generation

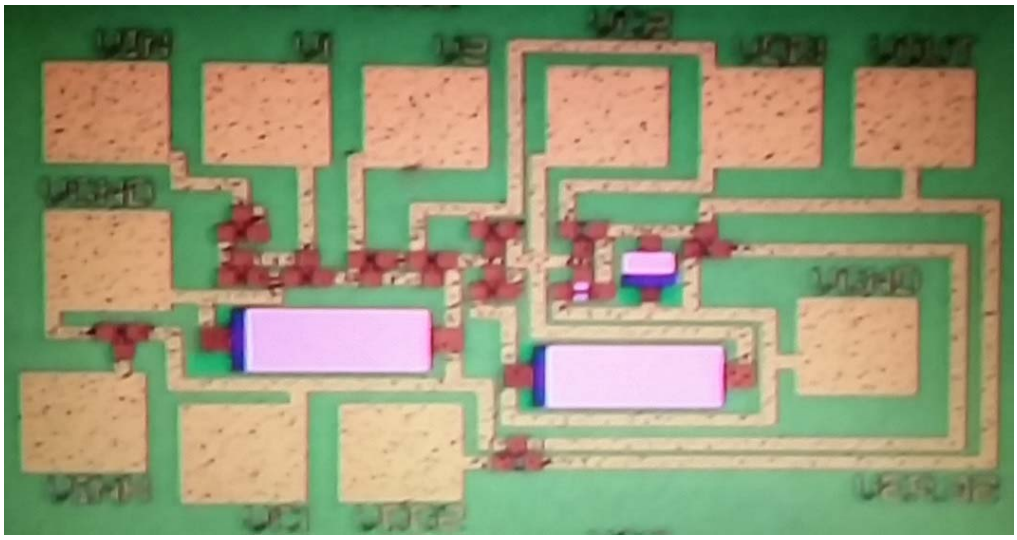
using two inverters, but the process has been simplified as much as possible by using only one metal layer at back-end process. A diagram of simplified neuron circuit for fabrication is shown in Fig. 4. 1 (a). In synaptic integration part, synaptic device is replaced with general NMOS and PMOS. In the case of inhibition part, instead of connecting PMOS and current mirror together, inhibition occurs immediately when PMOS is turned on. In action-potential generation part, most of the configuration is the same as that of the circuit designed in the simulation, but only the PMOS that plays a role of boosting V_{C1} to a very high voltage when the V_{C1} exceeds the threshold point is omitted. When fabricating a circuit including this boosting PMOS, the metal layer must be made in two layers because of the crosstalk between the metal lines connecting the elements. Therefore, we omitted the PMOS to simplify the structure as much as possible and focused on verifying the core function of the neuron circuit.



(a)



(b)



(c)

Fig. 4. 1. (a) Simplified neuron circuit to be made of one layer of metal (b) Completed mask layout for photo process (c) Top view of the fabricated neuron circuit.

Since each element must be doped with a different doping concentration, the mask layout for photo process must be well designed so that the devices are not interfered with each other's process. In addition, since the mask is expensive to manufacture, the fabrication process is simplified and only two quartz masks (126 mm*126mm) divided into four layers are used. The completed mask is shown in Fig. 4. 1 (a) and top view of the fabricated neuron circuit is shown in Fig. 4. 1 (b). The channel length and width of the MOSFET which is a core element of the fabricated neuron circuit, are based on a 2um line width. For stable operation of the circuit, the scale parameters of the MOSFETs are split so that each MOSFET has a line width of at least 0.5 μm to a maximum of 10 μm. The body and gate thicknesses are 100 nm and 150 nm, respectively, and the gate oxide thickness is 10 nm.

A detailed fabrication process flow is shown in Fig. 4. 2. The neuron circuit was fabricated on the SOI wafer, and after the initial cleaning, it entered the process in earnest. (a) First, Si active layer etching is performed to isolate each device and pattern the active layer. Because the buried oxide is located at a thickness of 375 nm below the active layer, Si-etch has been carried out for a sufficient amount of time. (b) After the active layer patterning is completed, buffer oxidation is carried out to a thickness of 10 nm as a preliminary work for body doping of each MOSFET. (c) Body doping is carried out four times in total. The body doping is performed in the order of general PMOS, general NMOS, NMOS having high V_T and MOSCAP with body doping concentrations of $5 * 10^{16}/cm^3$, $1 * 10^{17}/cm^3$, $3 * 10^{18}/cm^3$, $1 * 10^{21}/cm^3$. Each doping process can be performed

independently using the previously prepared photo mask or by using a method of superimposing a high doping concentration on a low doping concentration. (d) Remove the buffer oxide layer laid for the implant and grow the oxide layer to 10 nm for use as gate-oxide. (e) Gate patterning is performed by covering the n+ poly-Si layer used as the gate material with 150 nm and etching the poly-Si layer according to the shape of the mask layout. (f) A 10 nm buffer oxide layer is grown by a pre-process for source and drain implants. In this process, Si residues, which may remain after poly-Si etch, may be removed by oxidation. (g) In the implantation process of the source and the drain, a self-align process is performed to reduce the number of masks by doping the entire surface of the device. Two types of high-level doping with n-type and p-type form the source and drain of NMOS and PMOS, respectively. (h) Rapid thermal annealing (RTA) is carried out so that the dopant can diffuse evenly in the device. The RTA condition is determined by process simulation and the annealing is performed at 950 °C for 90 seconds in this fabrication process. (i) TEOS is deposited as an inter-layer dielectric (ILD) layer to isolate the devices from each other and to block them from the outside at a thickness of 365 nm. (j) A contact hole etch is made to pierce the ILD layer to contact the metal pad with the device. Since metal and device must be in direct contact, over-etch should be done sufficiently to prevent ILD from remaining at the bottom of contact hole. (k) The metal to be used as a pad must be filled in the contact hole in the order of Ti / TiN / Al / TiN, and the metal layers must be filled to a greater thickness than the hole depth. The thickness of the metal layers covered in the actual process is 30 / 30 / 400 / 30nm. (l) A metal pad is formed to measure the characteristics of

the fabricated neuron circuit through the probe station. The metal etch is made in the shape of the metal pad designed on the photo mask, and it is conducted for a sufficient time to isolate the metal pads from each other. (m) After the final step of metal pad formation, the neuron circuit fabricated on the wafer is fab-out to measure the electrical characteristics of.

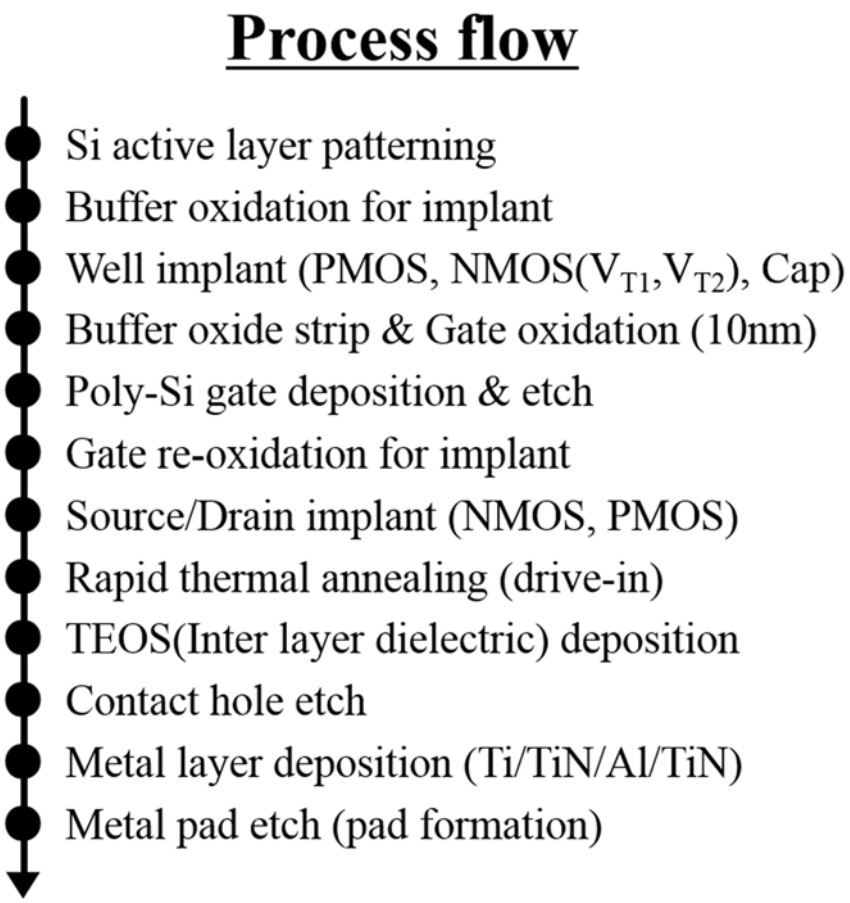


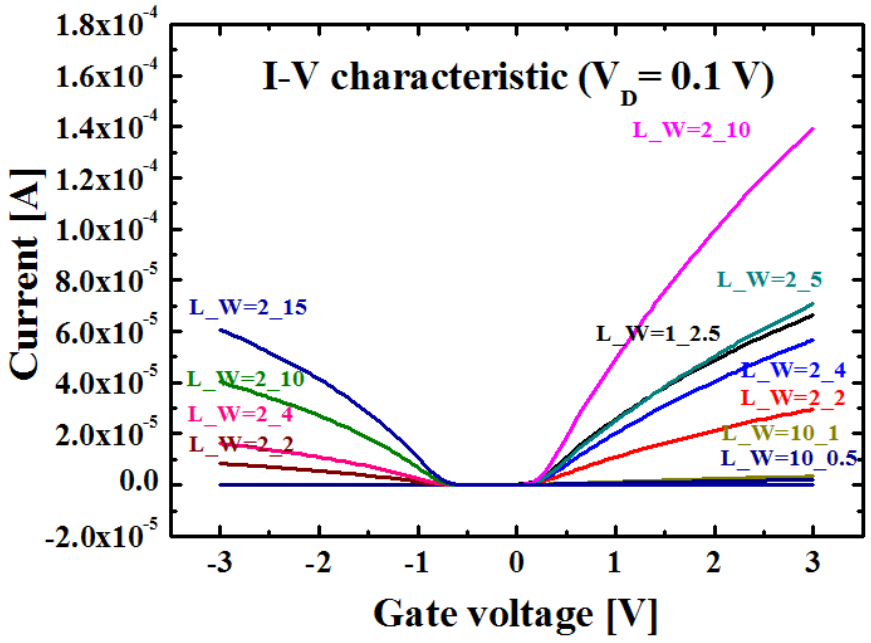
Fig. 4. 2. Process flow for fabrication of neuron circuit.

4.2 Electrical Properties of the Fabricated Neuron Circuit

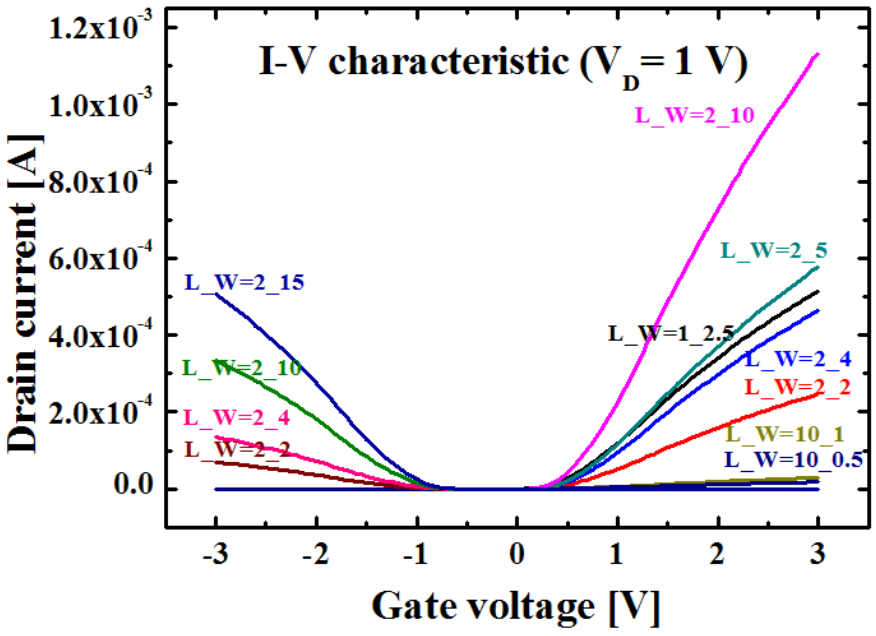
The process was carried out by integrating several neuron circuits in one 6-inch Si wafer, and each neuron circuit was split so that the MOSFETs in the circuit had different scale parameters. Prior to measuring the overall operation of the neuron circuit, we first looked at the operation characteristics of the individual devices that make up the circuit.

The transfer curves of MOSFETs with various channel length and channel width included in the fabricated circuit are shown in Fig. 4. 3. Among the devices shown in Fig. 4. 3, NMOS has body doping of $1 * 10^{17}/cm^3$ and PMOS has body doping of $5 * 10^{16}/cm^3$. In the case of PMOS, buried channels with N^+ gate and P^- body were fabricated to simplify the process steps. The transfer curves show that the V_T of the NMOS and PMOS are 0.3 V and -0.7 V, respectively, which is very similar to the voltage expected in the process simulation. And comparing the transfer curves of devices with different channel length and channel width, it can be seen that the magnitude of the current depends on the L / W ratio of the length divided by the width. Fig. 4.3 (a) and Fig. 4. 3 (b) show the transfer curves at V_D = 0.1 V and 1 V, respectively, and the appropriate current levels and low leakage currents are observed, indicating that the MOSFETs that make up the neuron circuit are well constructed as intended.

Fig. 4. 4 shows the transfer curves of NMOS with intentional high body doping ($3 * 10^{18}/cm^3$) to make high V_T. Since the output voltage of the neuron circuit must have a large negative value, the NMOS with $-V_{DD}$ applied at the source terminal is required, and


I-V characteristic (V_D = 0.1 V)
Current [A]
Gate voltage [V]
1.8x10^-4
1.6x10^-4
1.4x10^-4
1.2x10^-4
1.0x10^-4
8.0x10^-5
6.0x10^-5
4.0x10^-5
2.0x10^-5
0.0
-2.0x10^-5
-3
-2
-1
0
1
2
3
L_W=2_10
L_W=2_5
L_W=2_15
L_W=1_2.5
L_W=2_10
L_W=2_4
L_W=2_4
L_W=2_2
L_W=2_2
L_W=10_1
L_W=10_0.5

(a)


I-V characteristic (V_D = 1 V)
Drain current [A]
Gate voltage [V]
1.2x10^-3
1.0x10^-3
8.0x10^-4
6.0x10^-4
4.0x10^-4
2.0x10^-4
0.0
-3
-2
-1
0
1
2
3
L_W=2_10
L_W=2_5
L_W=2_15
L_W=1_2.5
L_W=2_10
L_W=2_4
L_W=2_4
L_W=2_2
L_W=2_2
L_W=10_1
L_W=10_0.5

(b)

Fig. 4. 3. The transfer curves of MOSFETs with various channel length and channel width included in the fabricated circuit (a) at $V_D = 0.1$ V and (b) $V_D = 1$ V.

since the NMOS should not be turned on until a sufficient voltage is applied to the gate terminal, it should have a large V_T. Fig. 4. 4 shows the transfer curves of two high V_T NMOS with channel length and width of 2 μm / 5 μm and 2 μm / 10 μm, respectively. Due to the high body doping, it can be seen that both NMOS have a V_T of 2 V or more, and the value of current differs according to the ratio of the length to the width and V_D.

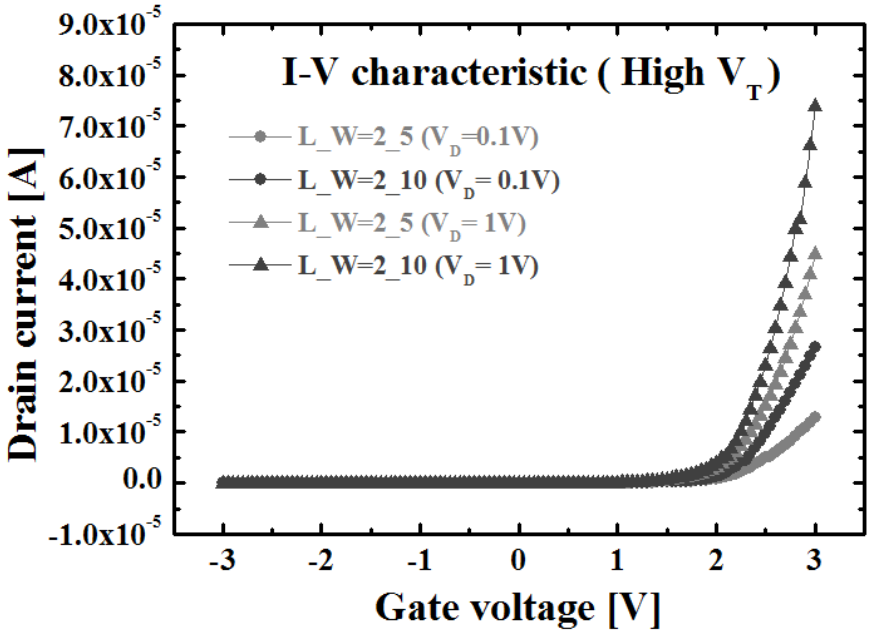


Fig. 4. 4. The transfer curves of MOSFETs with high V_T.

Fig 4. 5 shows the constant capacitance characteristics of MOS capacitors using high body doping. In this neuron circuit process, two types of capacitors were fabricated. C_1 and C_2 shown in Fig. 4.1 (a) were made to have a cross sectional area of 50 μm * 116 μm, and C_3 was made to have a cross sectional area of 34 μm * 17 μm. In a typical MOS capacitor, when the Si-body is doped normally, the capacitance changes according to the voltage applied to both electrodes. However, as the body is doped at a very high concentration, the Si-body has near-metal properties, and a capacitor is created in which the oxide dielectric fills the gate and body with high doping concentrations. Therefore, the capacitors were fabricated in this way as shown in Fig. 4. 5, and constant capacitances of 4.7 pF and 0.6 pF were obtained, respectively.

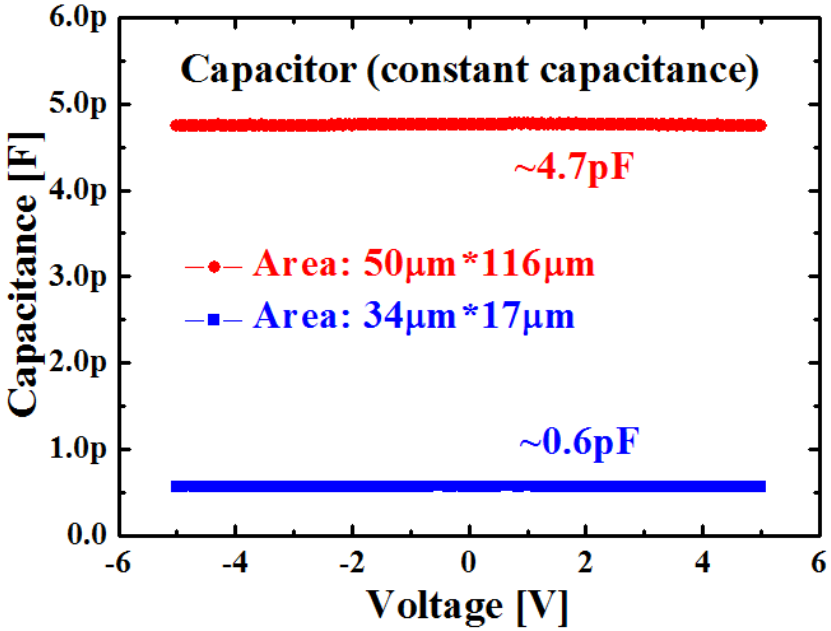
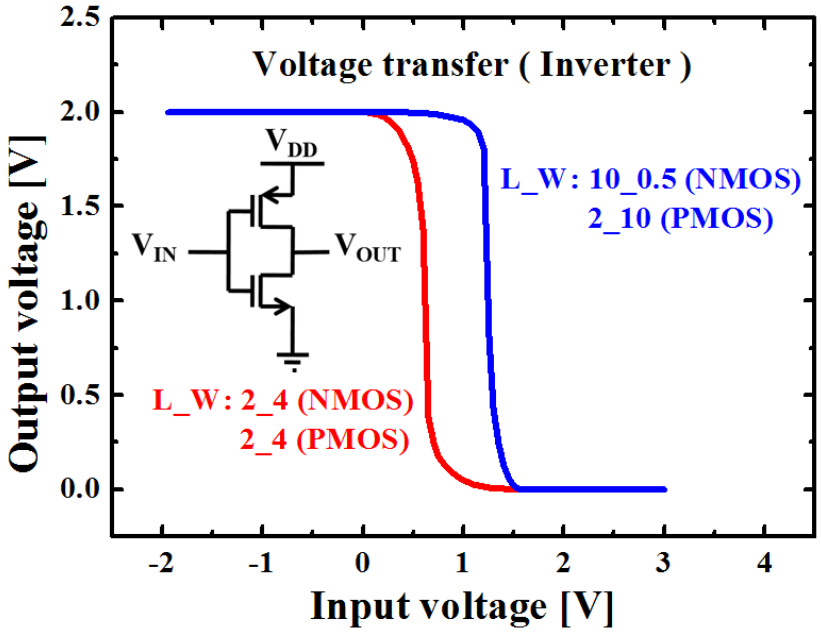
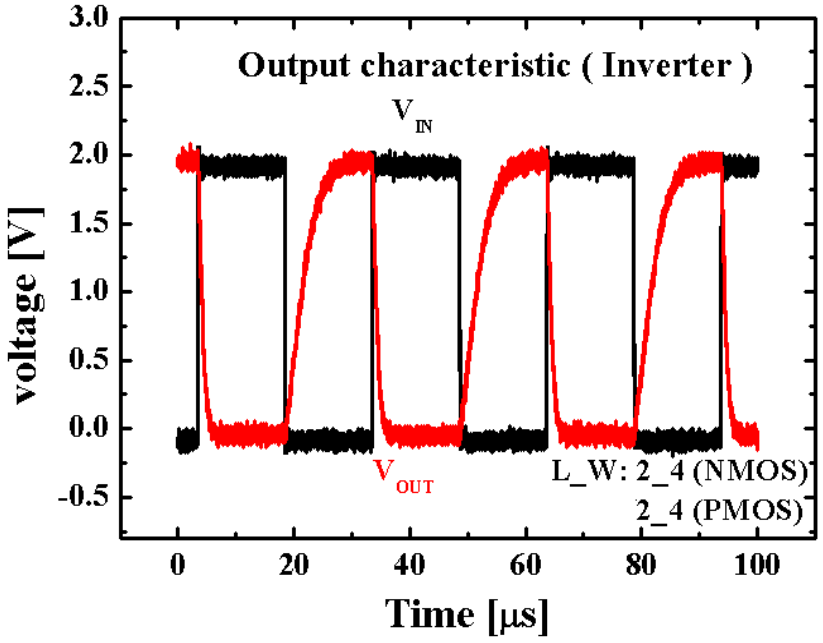


Fig. 4. 5. Constant capacitance characteristics of MOS capacitors using high body doping.

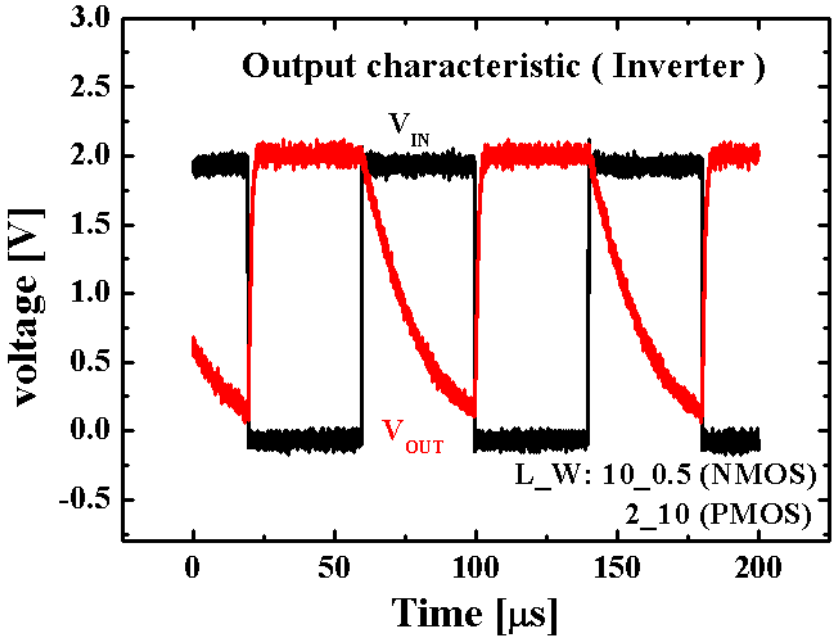
Fig. 4. 6 shows the voltage transfer characteristics of a CMOS inverter composed of NMOS and PMOS and the transient characteristics of the output pulse when the input square pulse is injected. Fig. 4. 6 (a) shows the voltage transfer characteristics of inverters with length / width ratios of NMOS and PMOS of 2 μm / 4 μm, 2 μm / 4 μm and 10 μm / 0.5 μm, 2 μm / 10 μm, respectively. Both inverters maintain a high output voltage until the input voltage exceeds the V_T of the NMOS, and the input voltage falls to the ground as soon as the input voltage exceeds the V_T of the NMOS. The output voltages of the two inverters fall to ground at different points, because the length / width ratios of the NMOS and PMOS that make up the inverters are different. The more current flows through the NMOS relative to the PMOS, the sooner the output voltage falls to ground.



(a)


Output characteristic (Inverter)
V_IN
V_OUT
L_W: 2_4 (NMOS)
2_4 (PMOS)
voltage [V]
Time [μs]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0
20
40
60
80
100

(b)

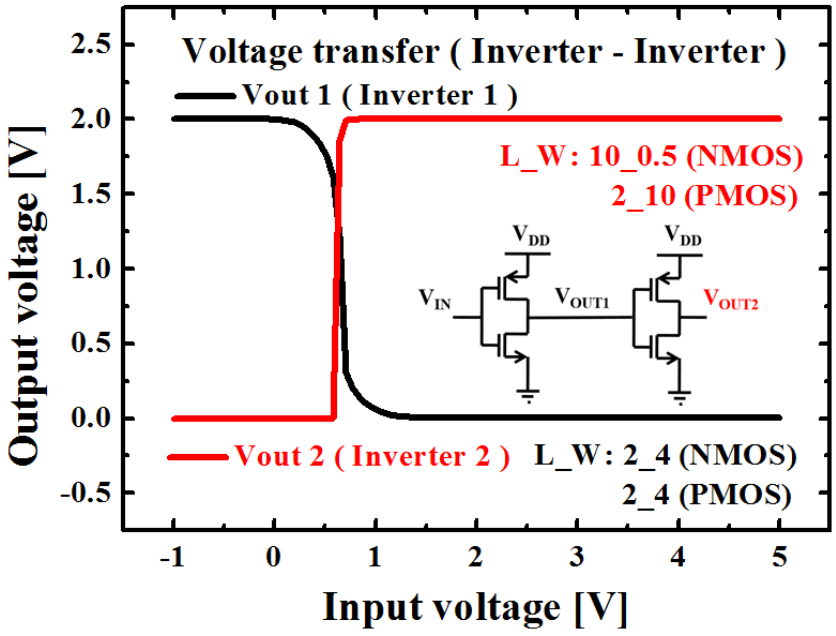

Output characteristic (Inverter)
V_IN
V_OUT
L_W: 10_0.5 (NMOS)
2_10 (PMOS)
voltage [V]
Time [μs]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0
50
100
150
200

(c)

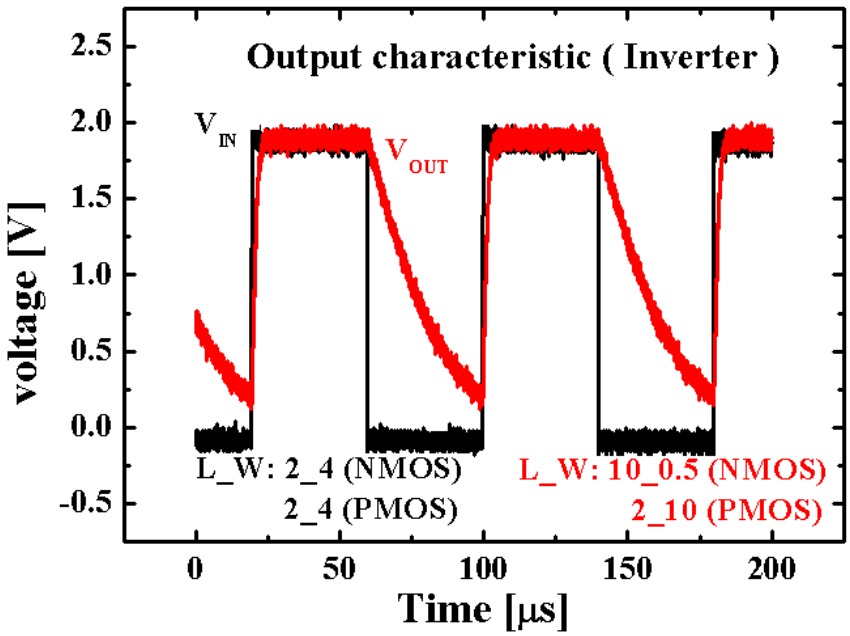
Fig. 4. 6. (a) Voltage transfer characteristics of a CMOS inverter composed of NMOS and PMOS and the transient characteristics of the output pulse when the length/width ratio of NMOS and PMOS are (b) 2 μm / 4 μm, 2 μm / 4 μm and (c) 10 μm / 0.5 μm, 2 μm / 10 μm.

Fig. 4. 6 (b), (c) show the transient characteristics of the output pulse when the length/width ratio of NMOS and PMOS are 2 μm / 4 μm, 2 μm / 4 μm and 10 μm / 0.5 μm, 2 μm / 10 μm, respectively. In Fig. 4. 6. (b), square input pulse with a size of 2 V, a width of 15 μs, and a period of 30 μs was applied and in Fig. 4. 6. (c), square input pulse with a size of 2 V, a width of 40 μs, and a period of 80 μs was applied. As the square input pulse is applied to both inverters, the inverting occurs at the output voltage. However, the rising and falling times of the output pulse are different in both cases. This is similar to that shown in Fig. 4. 6 (a). If the NMOS flows relatively well with respect to the PMOS, the falling time becomes shorter than the rising time. Otherwise, if the PMOS flows relatively well with respect to the NMOS, the rising time becomes shorter than falling time. Using this phenomenon makes it possible to control the shape of the voltages passing through the inverter, and the shape of the action potential generated in the fabricated neuron circuit is also made using this principle.

Fig. 4. 7 shows the voltage transfer curve and the transient characteristic of the output voltage when the two inverters shown in Fig. 4. 6 are connected in series. When two inverters are connected in series, the first input pulse has the same shape as the output pulse generated at the terminal inverter. This repeated inverting seems to have no functional


Voltage transfer (Inverter - Inverter)
Vout 1 (Inverter 1)
L_W: 10_0.5 (NMOS)
2_10 (PMOS)
VDD
VDD
VIN
VOUT1
VOUT2
Vout 2 (Inverter 2)
L_W: 2_4 (NMOS)
2_4 (PMOS)
Output voltage [V]
Input voltage [V]
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1
0
1
2
3
4
5

(a)

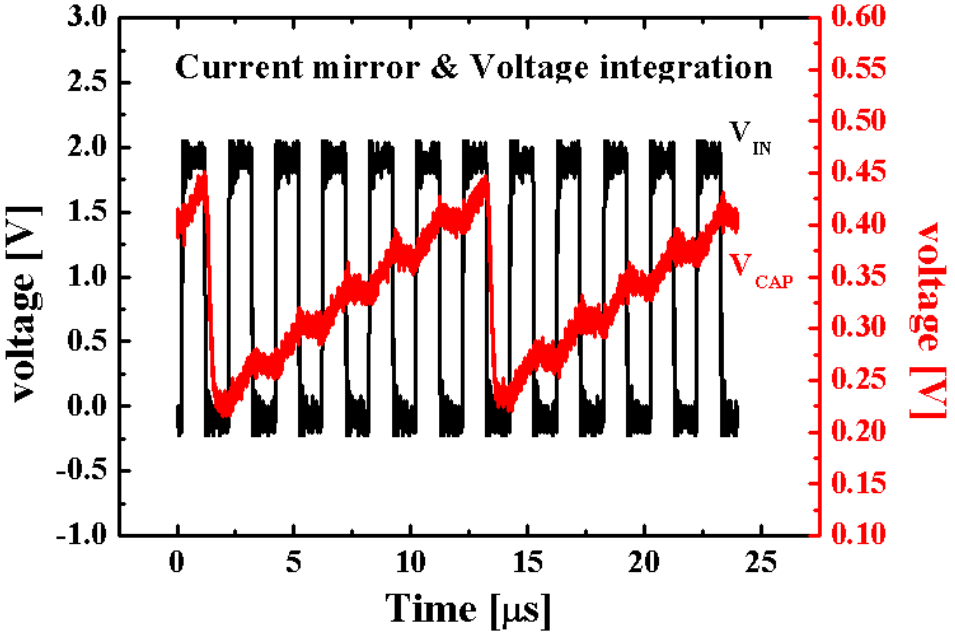

Output characteristic (Inverter)
VIN
VOUT
L_W: 2_4 (NMOS)
2_4 (PMOS)
L_W: 10_0.5 (NMOS)
2_10 (PMOS)
voltage [V]
Time [μs]
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
0
50
100
150
200

(b)

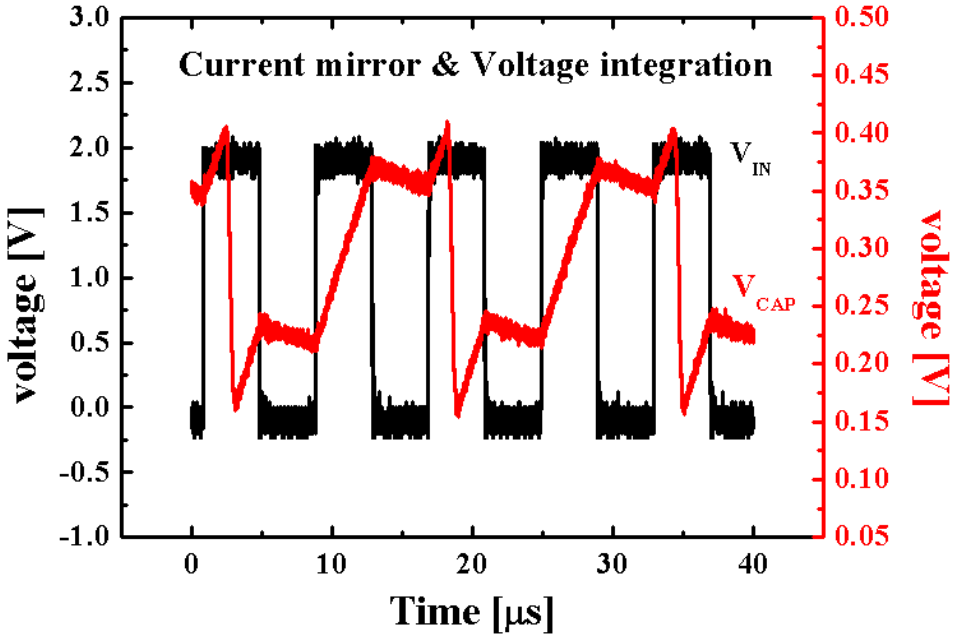
Fig. 4. 7. (a) The voltage transfer curve and (b) the transient characteristic of the output voltage when the two inverters shown in Fig. 4. 6 are connected in series.

significance, but on the proposed neuron circuit it guarantees the delay time required for the output pulse to maintain a negative voltage for a sufficient period of time and adjusts the shape of the pulse when the output voltage is positive. As expected, Fig. 4. 7 (a) shows that inverting occurs sequentially as the initial input voltage increases, and Fig. 4. 7 (b) shows that when a square input pulse with a size of 2 V, a width of 40 μs and a period of 80 μs is applied, the output pulse similar to square pulse is generated. In the case of the output pulse generated Fig. 4. 7 (b), the falling time is considerably smaller than the rising time because the PMOS flows relatively large current than the NMOS in the inverter located at the end.

Fig. 4. 8 shows the integrated voltage at C_1 when the synaptic current copied through the current mirror flows into C_1. While the input pulse is applied to the synaptic device, the voltage slowly begins to accumulate at C_1, and the voltage gradually decreases due to the leakage current while the pulse is turned off. If enough voltage is accumulated in C_1 and exceeds the threshold of the action-potential generation part, an output pulse is generated, which causes N_7 to be turned on in the circuit diagram of Fig. 4. 1 (a), and the voltage accumulated in C_1 falls to ground. Fig. 4. 8 (a) and Fig. 4. 8 (b) show the case where the size of input pulse is the same as 2 V but the pulse width is 1 μs and 4 μs, respectively. In both cases, V_{C1} begins to decrease sharply after about 0.4 V, and this value is the threshold


Current mirror & Voltage integration
V IN
V CAP
voltage [V]
voltage [V]
Time [μs]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0
5
10
15
20
25

(a)


Current mirror & Voltage integration
V IN
V CAP
voltage [V]
voltage [V]
Time [μs]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
10
20
30
40

(b)

Fig. 4. 8. Integrated voltage at C_1 when the synaptic current copied through the current mirror flows into C_1 (a) with the input pulse having a size of 2 V, a width of 1us, a period of 2 μs and (b) a size of 2 V, a width of 4 μs, a period of 8 μs.

value of the action-potential generation part. In the case of Fig. 4. 8 (b) with a large pulse width, action-potential is generated in the middle of pulse application and V_{C1} rapidly decreases. The result of Fig. 4. 8 shows that the operation of the current mirror, the voltage integration of C_1 and the initialization of V_{C1} due to the pulse generation operate correctly in the fabricated circuit.

Fig. 4. 9 shows the final output pulse at the output stage when a pulse is applied to the input stage of the neuron circuit. To create a situation where an output pulse occurs once when an input pulse is injected, we set the input pulse size to 2.55 V, the width to 18 μs, and the period to 50 μs. As shown in Fig. 4. 9, when an input pulse is injected once, an asymmetrical output pulse with the value of -2 V to 2 V and the width of 15 μs is generated, Verification of the operation of discrete devices such as NMOS, PMOS, NMOS with high V_T, capacitor constituting fabricated neuron circuit and inverting operation of 1-stage inverter, 2-stage inverters and voltage integration of current mirror with C_1 and output pulse generation, the neuron circuit, which was directly designed and verified at the simulation stage, is found to operate smoothly even in the fabricated form. If the neuron circuit is integrated with a synaptic device on a single wafer in the future, a complete neuromorphic system can be constructed and it is expected to be a more valuable study.

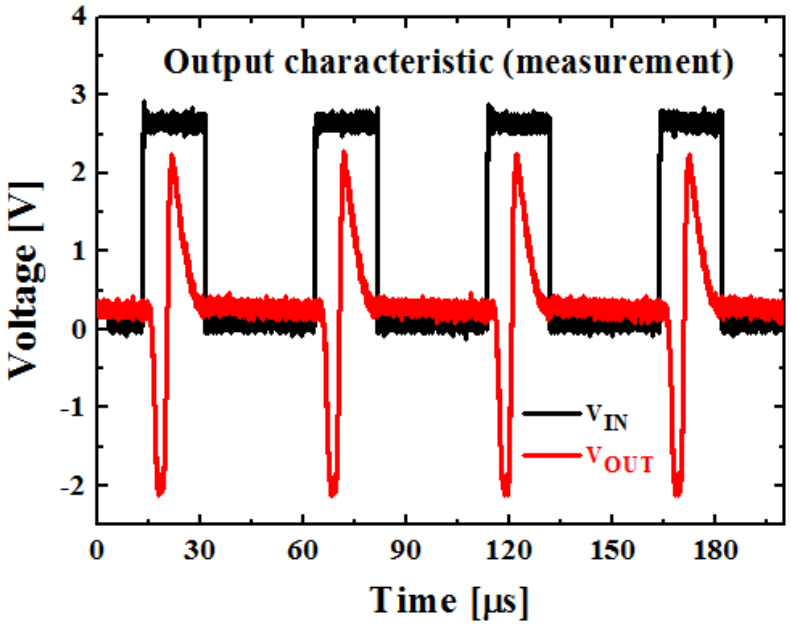


Fig. 4. 9. Output characteristic of the fabricated neuron circuit.

TABLE III

SCALE PARAMETERS OF FABRICATED CIRCUIT

Device	Width/Length (μm/μm)	Body doping (#/cm^3)	TOX (nm)	Device	Capacitance (pF)
N_1~N_3, N_8	2/2	$1*10^{17}$	10	C_1	4.7
N_4	4/2	$1*10^{17}$	10	C_2	4.7
N_5	1/10	$1*10^{17}$	10	C_3	0.6
N_6	10/2	$3*10^{18}$	10		
N_7	10/2	$1*10^{17}$	10		
P_1~P_2	2/2	$5*10^{16}$	10		
P_3	4/2	$5*10^{16}$	10		
P_4	10/2	$5*10^{16}$	10		

Chapter 5

Application to Spiking Neural Networks

In this chapter, we introduce unsupervised pattern recognition using a spiking neural network (SNN) based on the proposed synapse device and neuron circuit. Unsupervised learning is a kind of machine learning and it is mainly used to cluster input objects based on update of synaptic weight without teaching correct answer. In this study, we propose an application that recognizes and classifies number patterns with a total number of cells of 7 * 7 using SMART SPICE circuit simulator. We apply the STDP rule according to timing difference between input spike and output spike in weight update of synapses. It is very meaningful to propose an application based on a spiking neural network composed of pure hardware system by breaking away from the machine learning of general software method.

5.1 Construction of SNN for Unsupervised Pattern Recognition

Spiking neural network uses pulse or spike as input and output, and uses STDP rule to update synaptic weight. These methods are very suitable for driving a hardware-based

neuromorphic system and can be easily applied to the neuromorphic system including a 4-termial Si-based synaptic device and I & F neuron circuit proposed in this study. Since the proposed system can control the synaptic weight autonomously through the STDP rule without additional circuit elements, when the system is combined with a spiking neural network and used for various machine learning applications such as pattern recognition, it can obtain higher efficiency than the software-based machine learning.

Fig. 5. 1 shows the neuron-synapse-neuron connection, the basic unit of spiking neural network in the proposed system. When a square pulse is sufficiently applied to the input of the first neuron (N1), N1 fires and generates an action-potential to be injected into the gate and the drain of synaptic device (S1). The signal passed through S1 becomes the input of the next neuron (N2). When N2 fires, the generated action-potential from N2 is transferred to the next neuron and injected into the back-gate of S1 at the same time. Because of the timing difference between the outputs of N1 and N2, a weight update occurs in S1, and this phenomenon is said to be learning through STDP. As S1 learns different conductance from the previous one, the timing of N2's firing also changes.

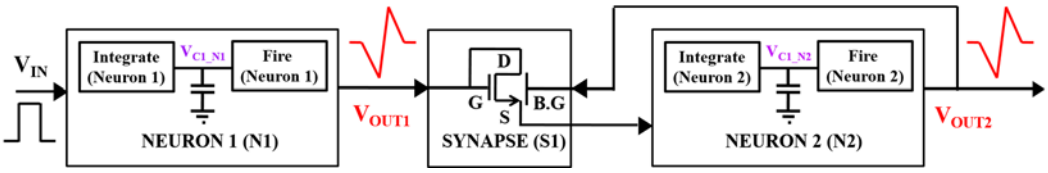
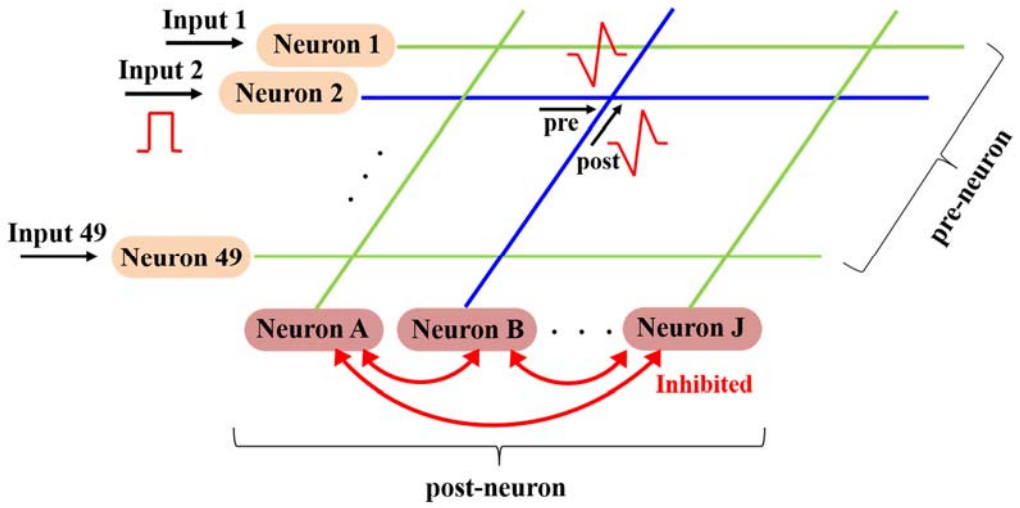


Fig. 5. 1. Neuron-synapse-neuron connection, the basic unit of spiking neural network.

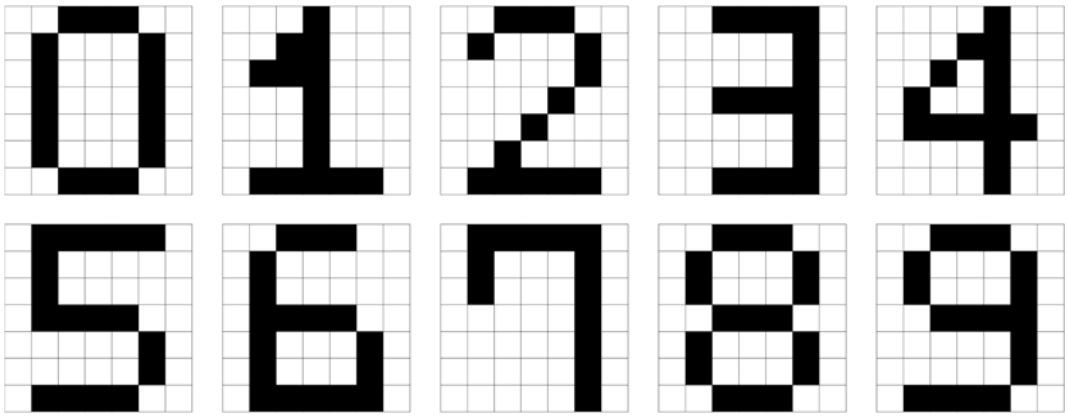
Fig. 5. 2 (a) shows a fully connected 1-layer spiking neural network that recognizes a number pattern from 0 to 9 expressed in 7 * 7 pixels and Fig. 5. 2 (b) shows pattern samples from 0 to 9. Each pixel corresponds to one pre-neuron, and a total of 49 neurons act as input neurons. The numbers from 0 to 9 correspond to a total of 10 post-neurons, and each neuron recognizes one of the numbers from 0 to 9. In this study, we tried to recognize a number pattern expressed in black and white. Therefore, a square pulse is injected into the neuron corresponding to a pattern painted in black, and no square pulse is injected into the neuron corresponding to a white pattern. The neurons injected with square pulse produce an action potential, and the action potentials are injected into all synaptic devices connected to these neurons. The action potentials pass through the synaptic device to enter the inputs of the post-neurons, and since the post neurons are inhibited from each other, only one of the ten post-neurons fires first. Thus, the first firing neurons represent the pattern corresponding to the inputs injected into the pre-neuron and synaptic devices located between pre-neurons injected with input pulses and the firing neurons reinforce their weight by STDP rule. Since only certain synapses associated with a number pattern are enhanced, if another pattern similar to this number comes in later, the pattern is recognized as an exact number and the corresponding neuron is fired.

After learning one number on the network, the neurons corresponding to this learned number should be inhibited from firing for a while. This is because, when learning other numbers, neurons corresponding to previously learned numbers must not be fired through already reinforced synapses. In this way, the post-neurons will be matched with numbers

from 0 to 9 one by one, and after all learning is finished, the post-neuron that has been inhibited through the learning process must be reactivated for the testing process, which will subsequently recognize a random number pattern.



(a)



(b)

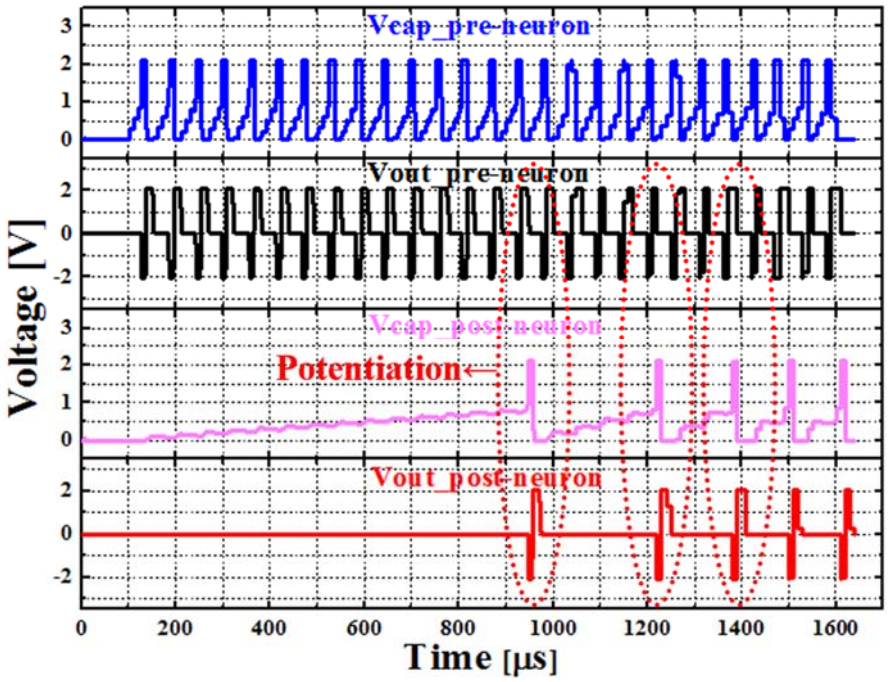
Fig. 5. 2. (a) a fully connected 1-layer spiking neural network that recognizes a number pattern from 0 to 9 expressed in 7 * 7 pixels. (b) Pattern samples from 0 to 9.

5.2 Learning and Testing for Unsupervised Pattern Recognition

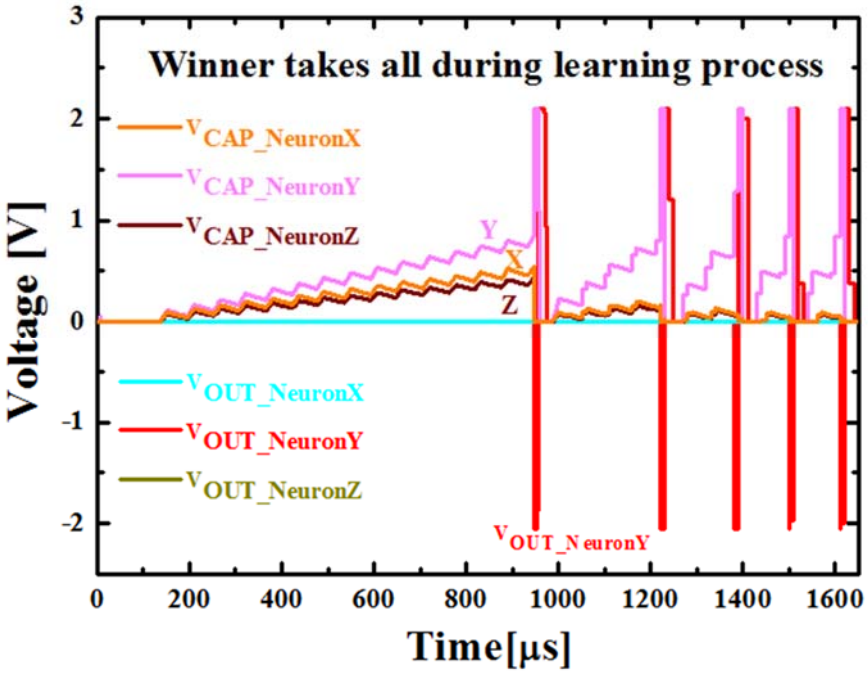
The process of applying spiking neural network to pattern recognition using the proposed system is largely divided into learning and testing. Learning is the process of learning the patterns we are trying to perceive in the neural network, and testing is the process of classifying and recognizing any given pattern into the learned patterns. The learning method applied to the spiking neural network proposed in this paper is an unsupervised learning method. Until a signal is applied to the pre-neurons, it is not known which neuron fires first among several post-neurons. If the input corresponding to a specific pattern is applied to the pre-neuron, the signal is transmitted to the post-neuron through all the synapses connected with the pre-neuron, and the neurons receiving the largest current signal among the several post-neurons fires first. Since the post-neurons are linked to inhibit each other's fire, the fire of the post-neuron inhibits the activation of other neurons based on the winner-takes-all method.

Fig. 5. 3 (a) shows the learning process that the conductance of the synapse is enhanced when pre-neurons and post-neurons connected with each other through specific synapses are fired with a positive time difference. For the experimental results shown in Fig. 5. 3 (a), before the synaptic device is learned, the pre-neuron should fire about 15 times

for the post-neuron to fire. However, when the pre-neuron and the post-neuron are sequentially fired under the STDP, the conductance increases due to the potentiation in the synaptic device, and the number of output pulses of the pre-neurons required for the post neuron to fire is gradually reduced. After learning is finished, the post-neuron fires even if the output from the pre-neuron is delivered twice. As mentioned earlier, in the process of learning one pattern, only one post-neuron is set to fire. Thus, even if there is another post-neuron in which the signal from the pre-neuron is transmitted and the value of the integrated V_{CAP} gradually increases, it eventually fails to fire because the action potential of the firing post-neuron acts as an inhibitory signal as shown in Fig. 5. 3 (b).



(a)

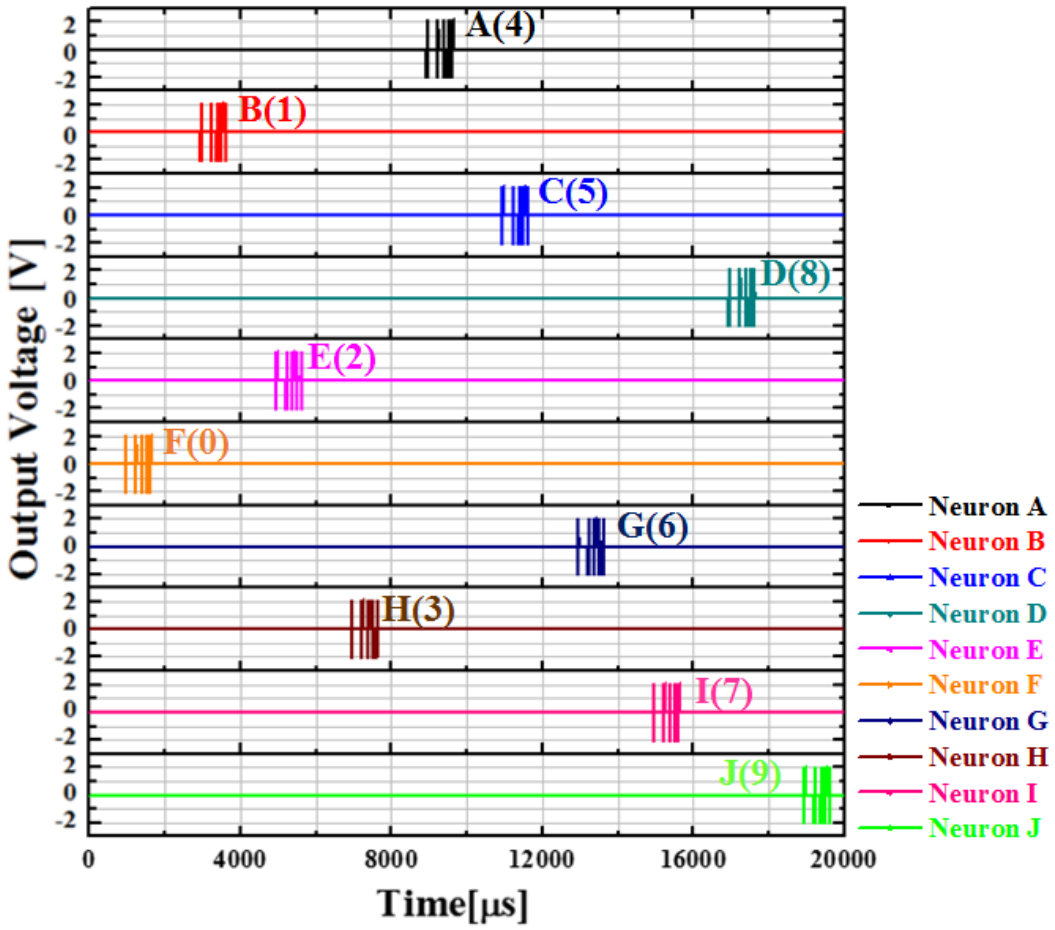


(b)

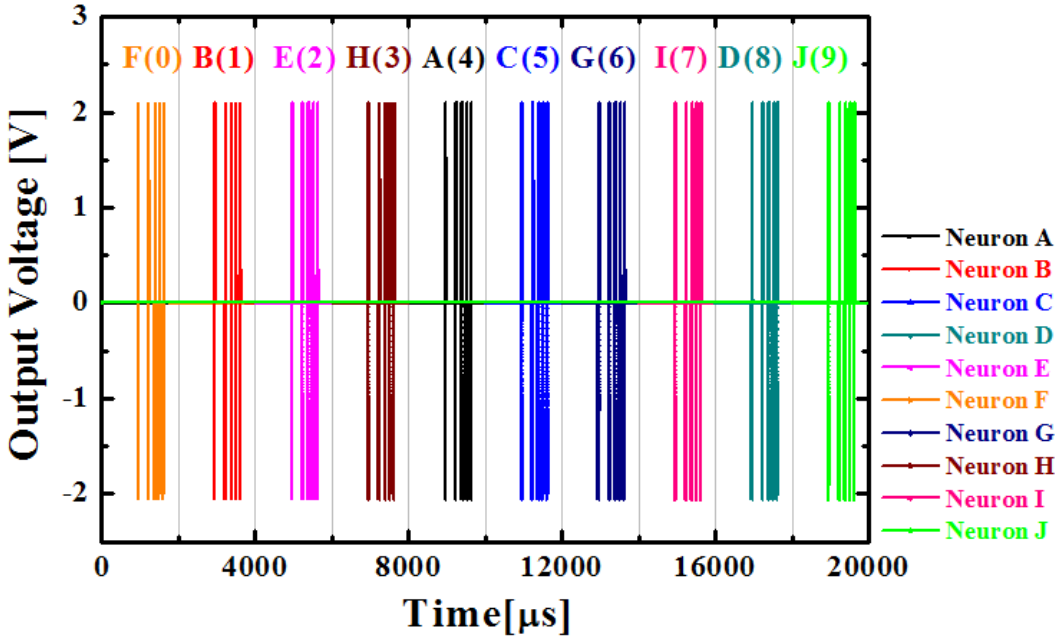
Fig. 5. 3. (a) The learning process that the conductance of the synapse between neurons is enhanced. (b) Action potential of the firing post-neuron acts as an inhibitory signal.

Fig. 5. 4 (a) shows the result of observing the output-action potential of 10 post-neurons when 10 numerical patterns in Fig. 5. 2 (b) are sequentially learned and Fig. 5. 4 (b) summarizes this action potential in the order of occurrence. It can be seen that only one post-neuron fires when learning each pattern, and as the learning progresses, the conductance of the synapse increases and the action-potentials are generated at increasingly rapid intervals. Since the post-neuron corresponding to the learned pattern is set to suppress

own fire while learning another pattern, a total of 10 numerical patterns are uniformly corresponded to 10 post-neurons, respectively. The numerical pattern corresponding to each post-neuron after learning process is listed in Table IV.



(a)



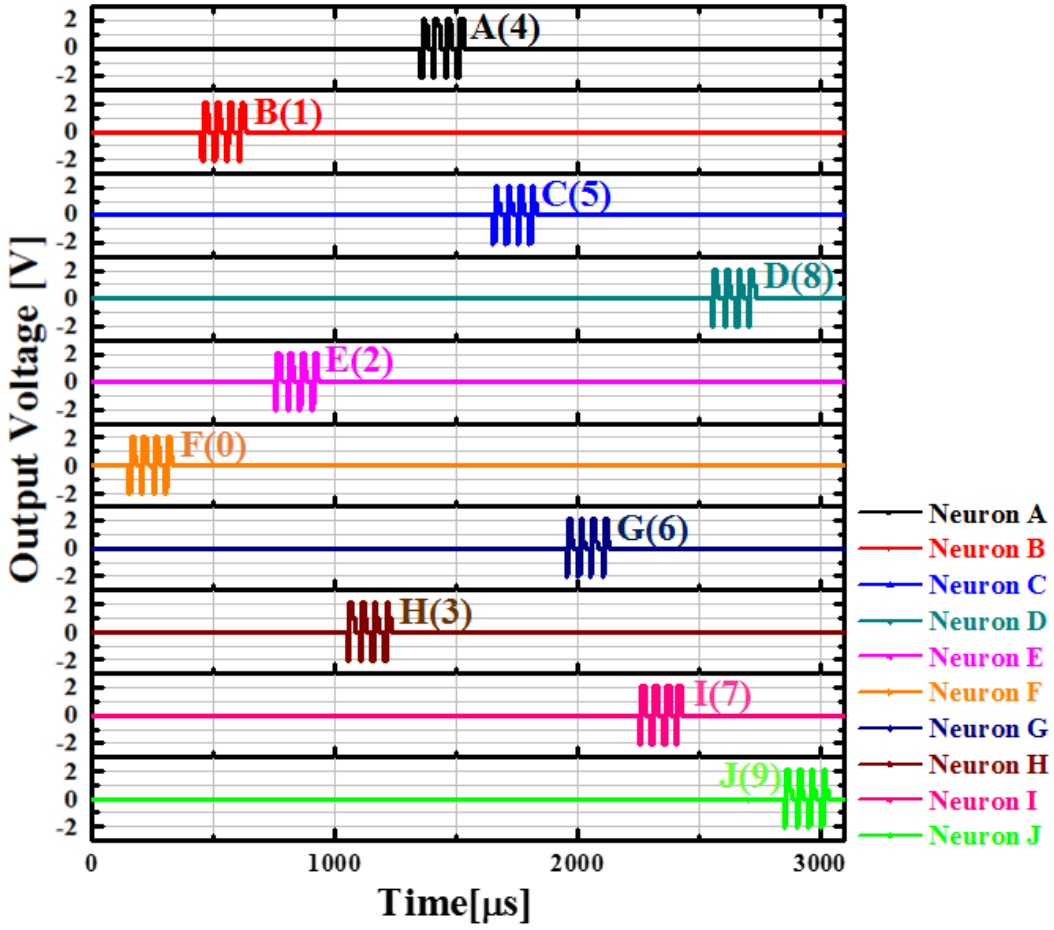
(b)

Fig. 5. 4. (a) Output action-potential of 10 post-neurons when 10 numerical patterns are sequentially learned. (b) The result of (a) is summarized in the order of occurrence.

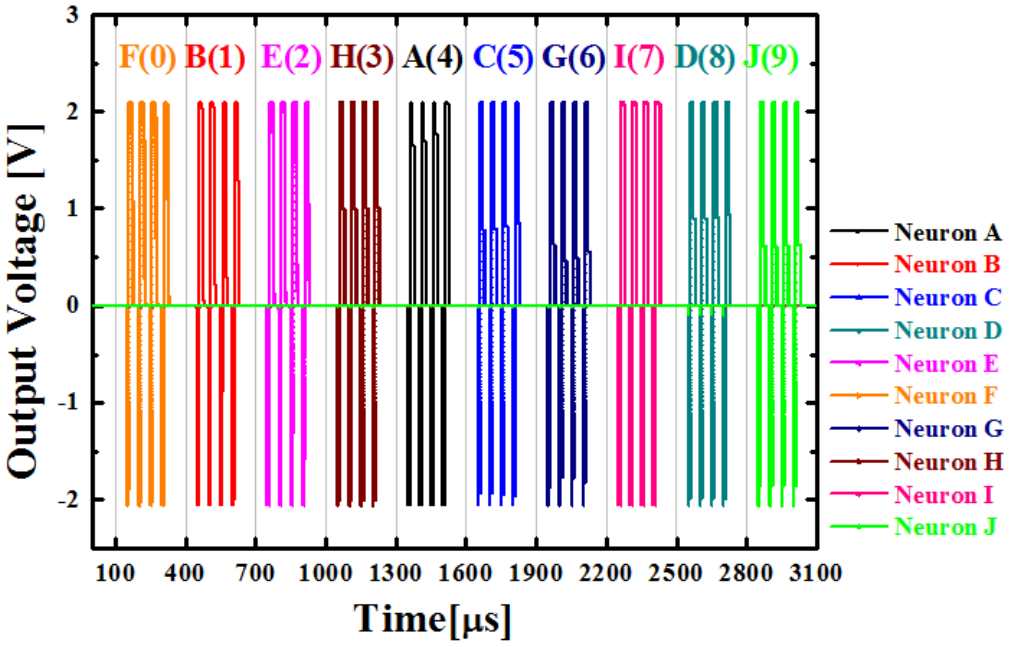
TABLE IV
10 DIGITS CORRESPONDING TO POST-NEURON

Neuron A: 4	Neuron B: 1	Neuron C: 5	Neuron D: 8	Neuron E: 2
Neuron F: 0	Neuron G: 6	Neuron H: 3	Neuron I: 7	Neuron J: 9

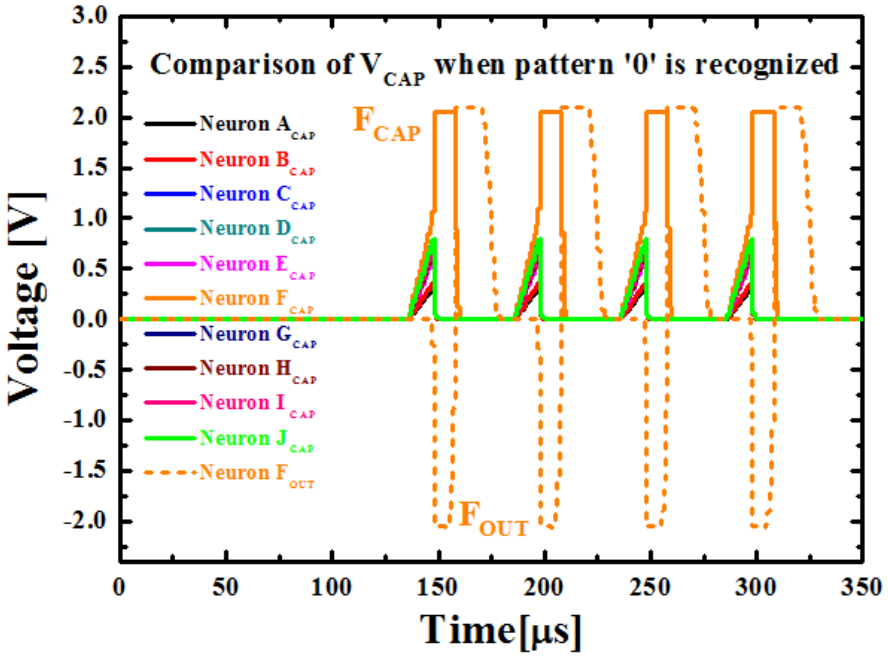
Fig. 5. 5 (a) shows the result of observing the output-action potential of 10 post-neurons when the input pulses corresponding to 10 numerical patterns in Fig. 5. 2 (b) are sequentially injected for recognition and Fig. 5. 5 (b) summarizes this action potential in the order of occurrence. When the input pulses corresponding to each pattern is applied, the signal is transmitted through the synapse strengthened in the learning process, and the



(a)


F(0) B(1) E(2) H(3) A(4) C(5) G(6) I(7) D(8) J(9)
Output Voltage [V]
Time[μs]
Neuron A
Neuron B
Neuron C
Neuron D
Neuron E
Neuron F
Neuron G
Neuron H
Neuron I
Neuron J

(b)


Comparison of V CAP when pattern '0' is recognized
F CAP
F OUT
Voltage [V]
Time[μs]
Neuron A CAP
Neuron B CAP
Neuron C CAP
Neuron D CAP
Neuron E CAP
Neuron F CAP
Neuron G CAP
Neuron H CAP
Neuron I CAP
Neuron J CAP
Neuron F OUT

(c)

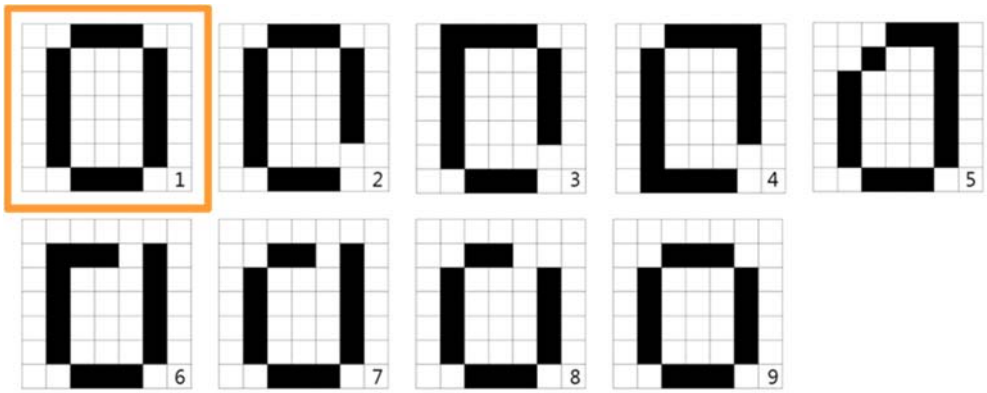
Fig. 5. 5. (a) Output action-potential of 10 post-neurons when the input pulses corresponding to 10 numerical patterns are sequentially injected for recognition. (b) The result of (a) is summarized in the order of occurrence. (c) The difference of V_{CAP} when the pattern is recognized.

strongest current signal flows into the post-neuron corresponding to the specific pattern. Therefore, only one neuron fires and in contrast to Fig. 5. 3 (b) in the learning process, output pulses occur at regular intervals in the testing process after learning. Fig. 5. 5. (c) shows the voltage accumulation in the capacitor located inside the post-neurons when the pattern is recognized. As can be seen in the figure, the capacitor voltage is most concentrated in the post-neuron corresponding to a certain pattern, which causes the neuron to become the first to fire. In other post-neurons, the voltage of the capacitor does not increase significantly because the action potential of the post-neuron, which was first fired, acts as an inhibitory signal.

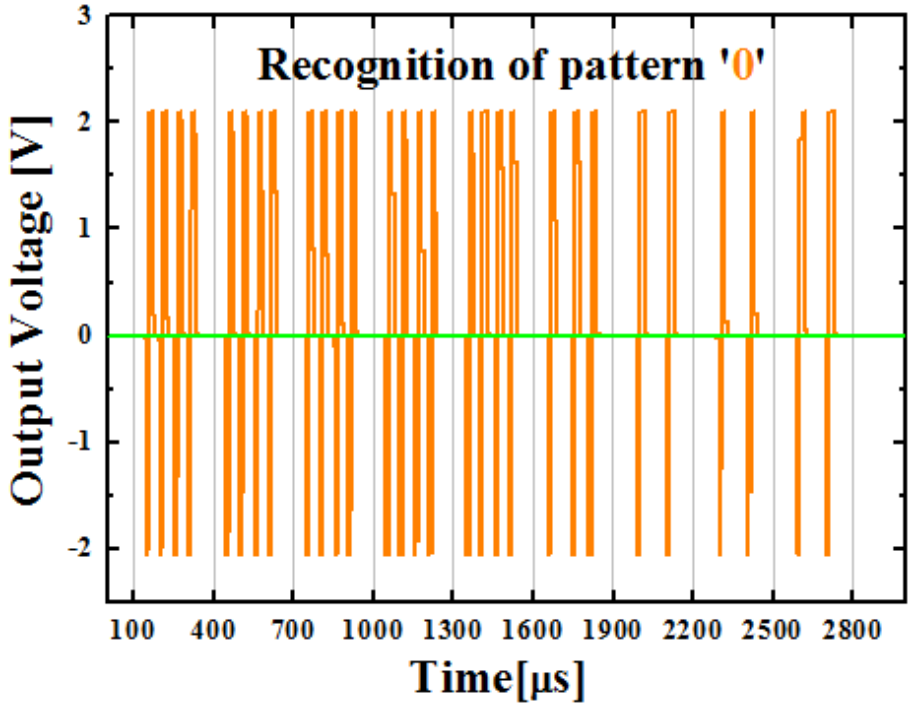
Through the results of Fig. 5. 4 and Fig. 5. 5, we can confirm that the proposed system performs unsupervised pattern recognition based on a spiking neural network and the system needs to be verified how tolerant it is to not only the correctly represented pattern, but also the noise containing pattern. Fig. 5. 6 shows the results of verifying how the proposed system learned by the unsupervised learning method identifies the patterns when recognizing a total of nine '0' patterns including random noise. The nine ‘0’ samples used for recognition is shown in Fig. 5. 6 (a) and each sample is numbered in sequence. The first

pattern in the sample group is the correctly represented pattern and as the number of samples increases, it is set to increase or decrease the amount of noise compared to the first pattern. As shown in Fig. 5. 6 (b), when all '0' patterns containing noise are recognized by the system, post-neuron 'F' only fires alone and the more times the pattern that is different from the correctly represented pattern is recognized, the less the number of fires. This is because when a signal enters a potentiated synapse due to learning, a large amount of current flows into the post-neuron, but the synapse corresponding to the noise is not potentiated and thus cannot transmit a large amount of current.

From the Fig. 5. 7 to Fig. 5. 15, the experimental results of the recognition of numerical patterns including noise from ‘1’ to ‘9’ are shown as in the case of Fig. 5. 6 and the given samples are correctly recognized based on the learned neural network in most cases. In some cases, however, the learned system incorrectly recognizes the noise-added pattern as a different number. In case of Fig. 5. 8, the system recognizes the 8^{th} pattern to not ‘2’ but ‘6’. The 8^{th} pattern in Fig. 5. 8 (a) is a noise-added pattern from pattern '2', but there is a lot of overlap with pattern '6', which makes it possible for the system to recognize it incorrectly. Even when recognizing '3', '5', and '7', if noise is added and the sample becomes similar to other patterns, it may not be recognized correctly. Since the proposed pattern recognition system adopts the unsupervised learning method and the numerical patterns are represented by only 7 * 7 cells, there is also the possibility that these misunderstandings may occur. To overcome these drawbacks, it is necessary to divide the pattern into smaller pieces and to increase the number of layers to be learned.

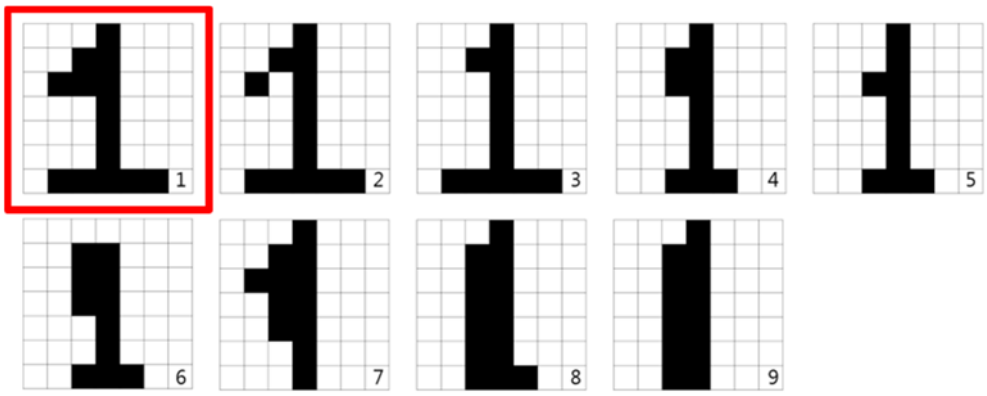


(a)

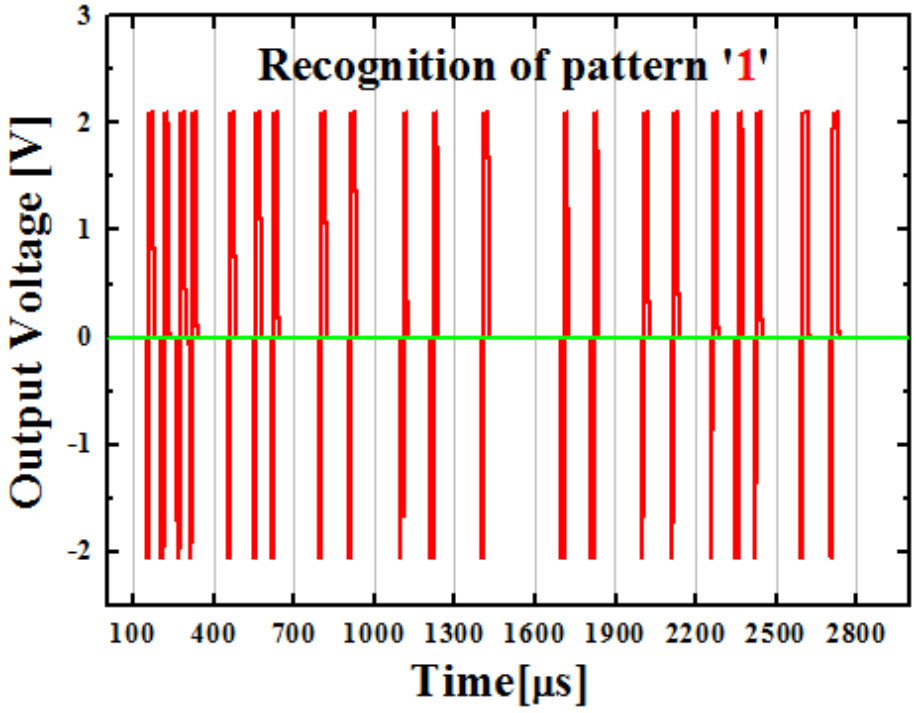


(b)

Fig. 5. 6. (a) The nine ‘0’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘F’ when the ‘0’ samples were recognized in turn.

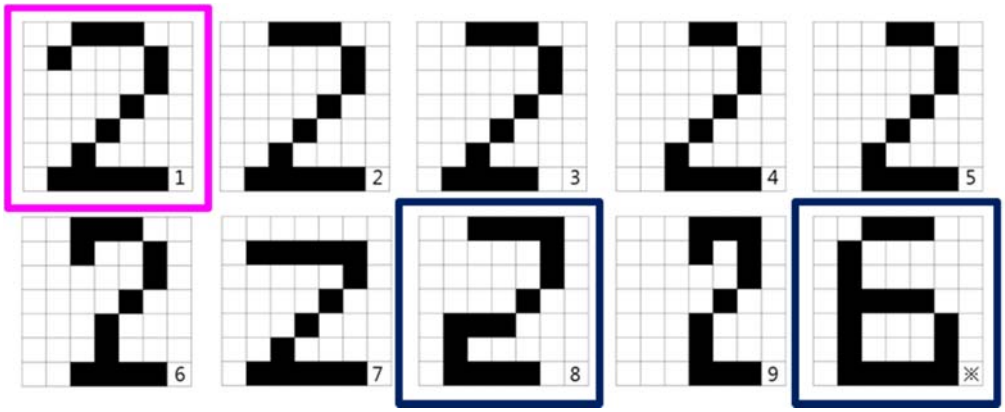


(a)

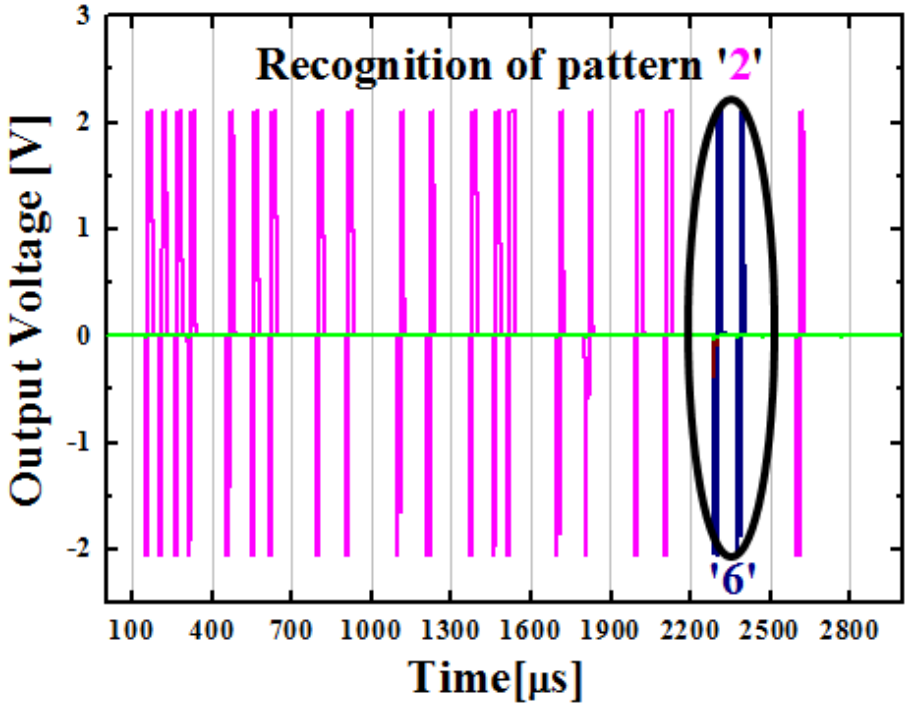


(b)

Fig. 5. 7. (a) The nine ‘1’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘B’ when the ‘1’ samples were recognized in turn.

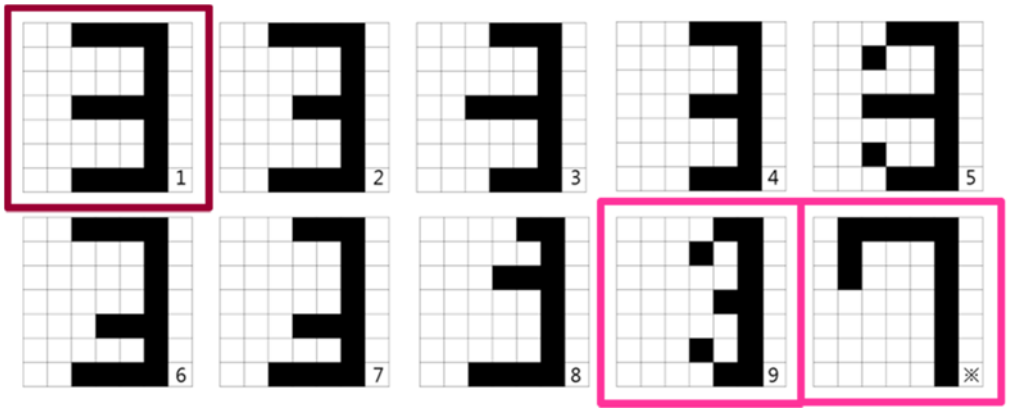


(a)

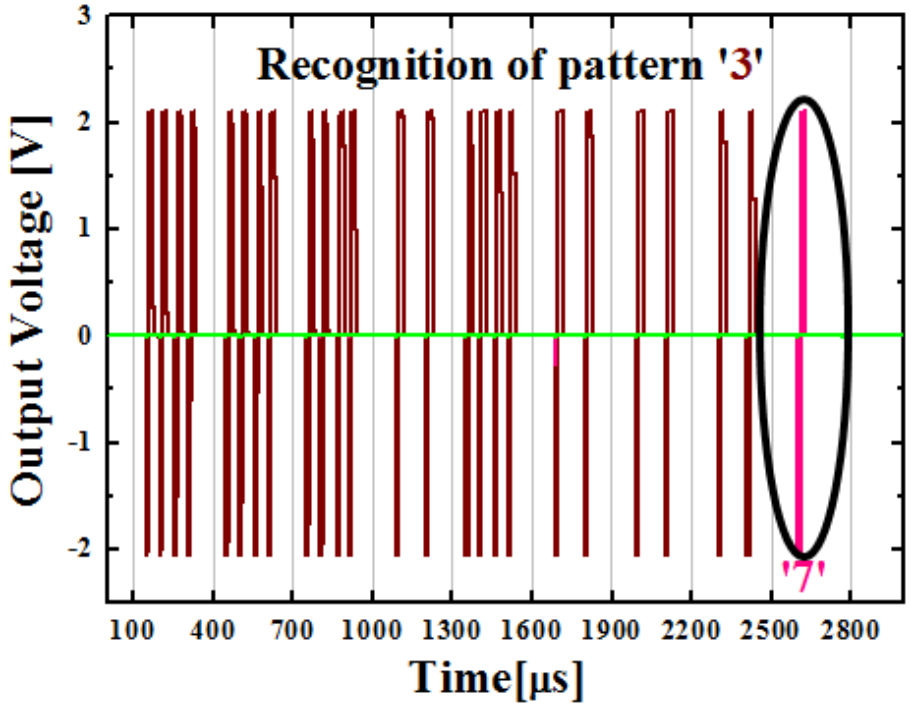


(b)

Fig. 5. 8. (a) The nine ‘2’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘E’ when the ‘2’ samples were recognized in turn.

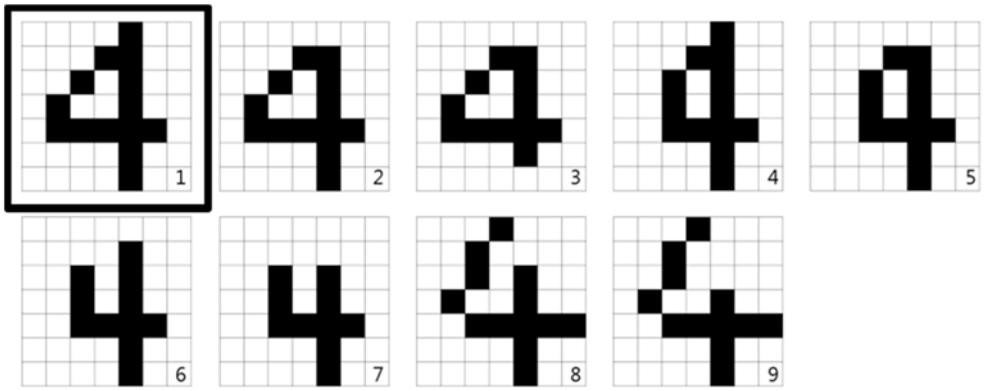


(a)

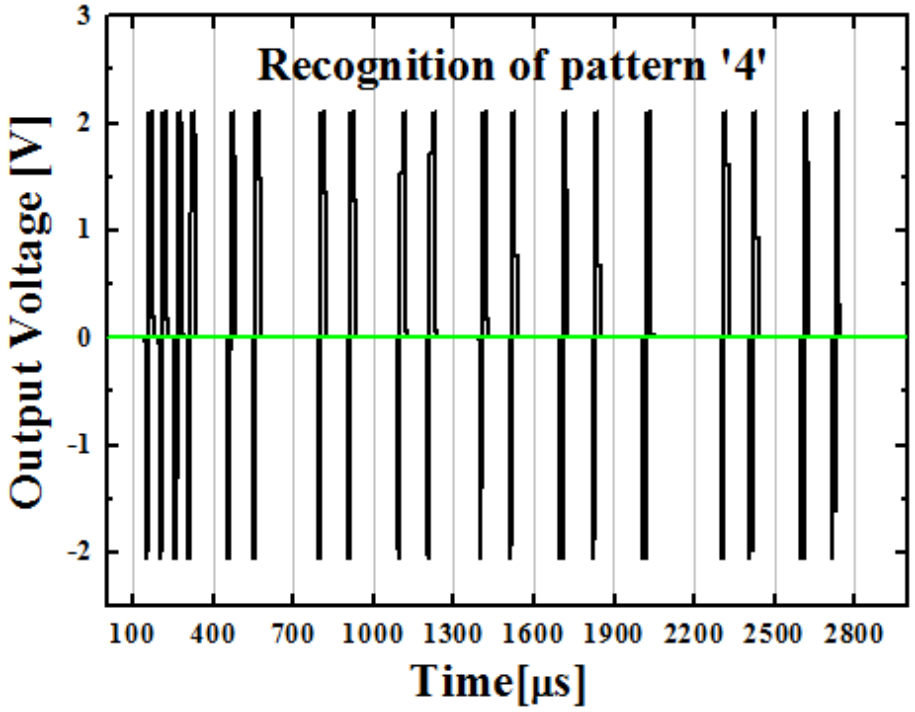


(b)

Fig. 5. 9. (a) The nine ‘3’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘H’ when the ‘3’ samples were recognized in turn.

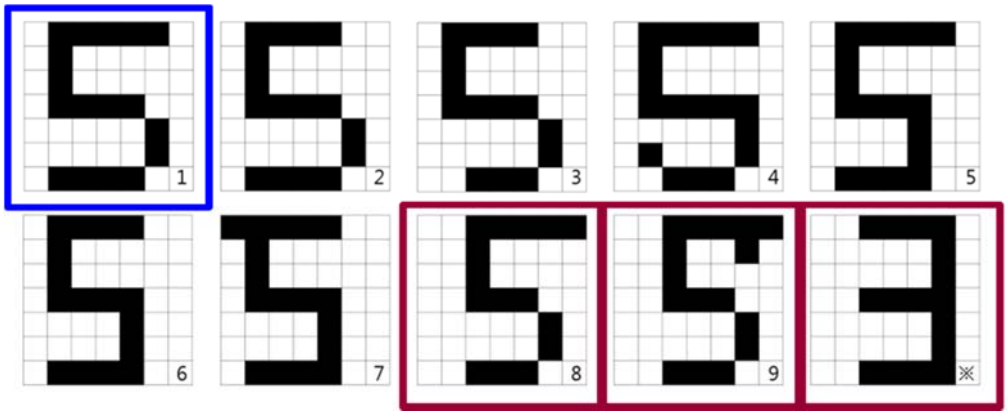


(a)

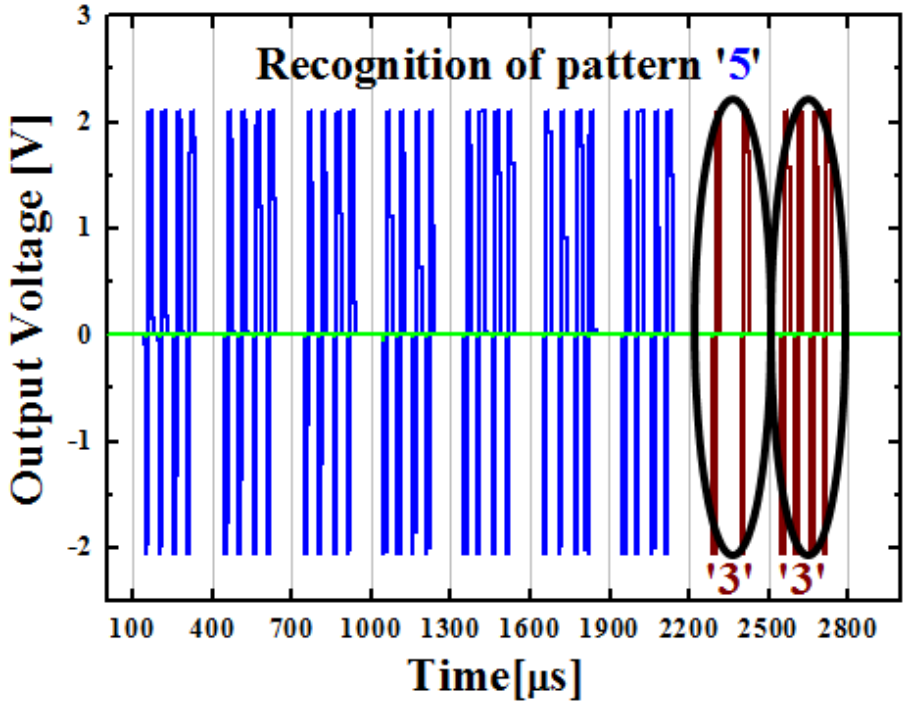


(b)

Fig. 5. 10. (a) The nine ‘4’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘A’ when the ‘4’ samples were recognized in turn.

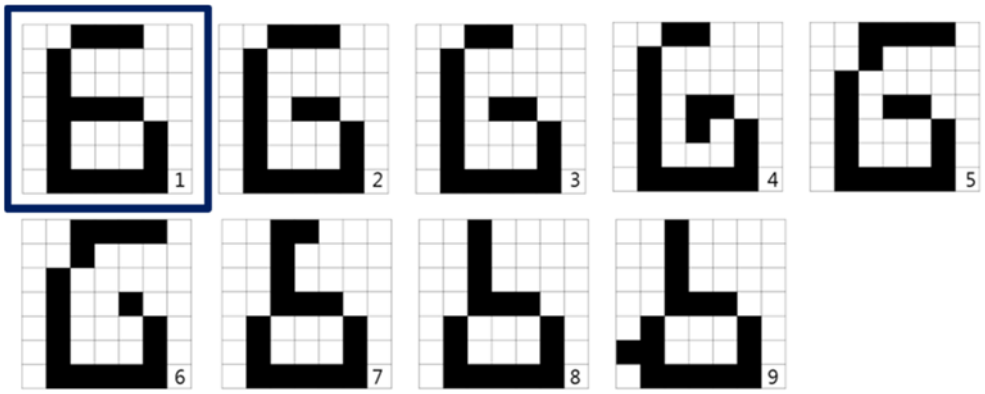


(a)

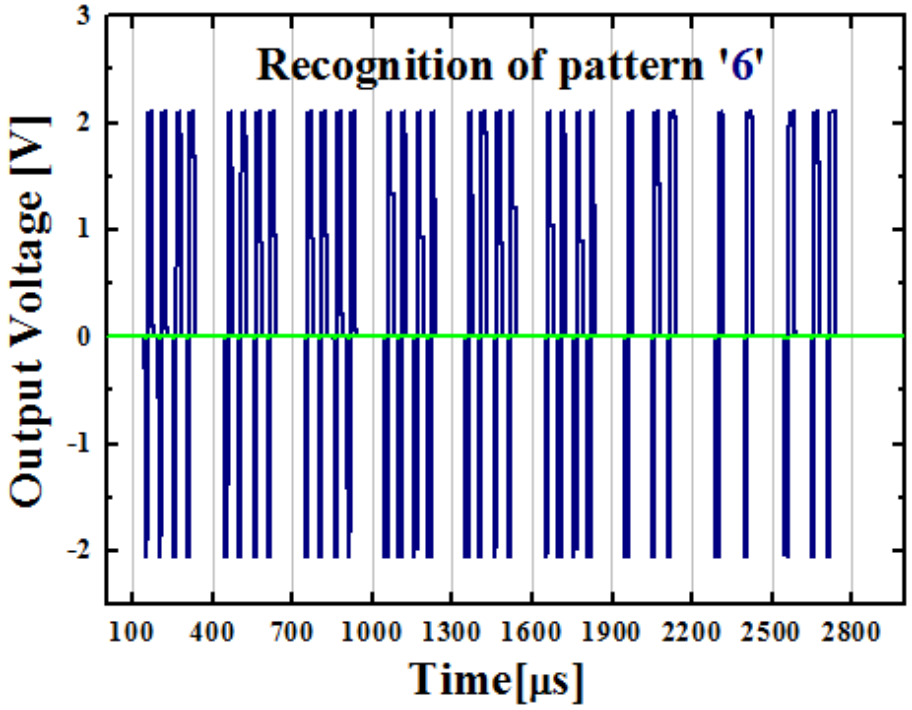


(b)

Fig. 5. 11. (a) The nine ‘5’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘C’ when the ‘5’ samples were recognized in turn.

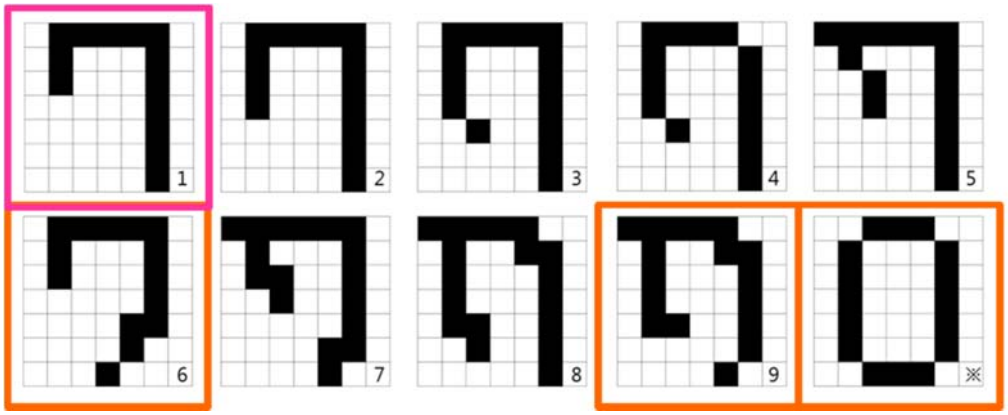


(a)

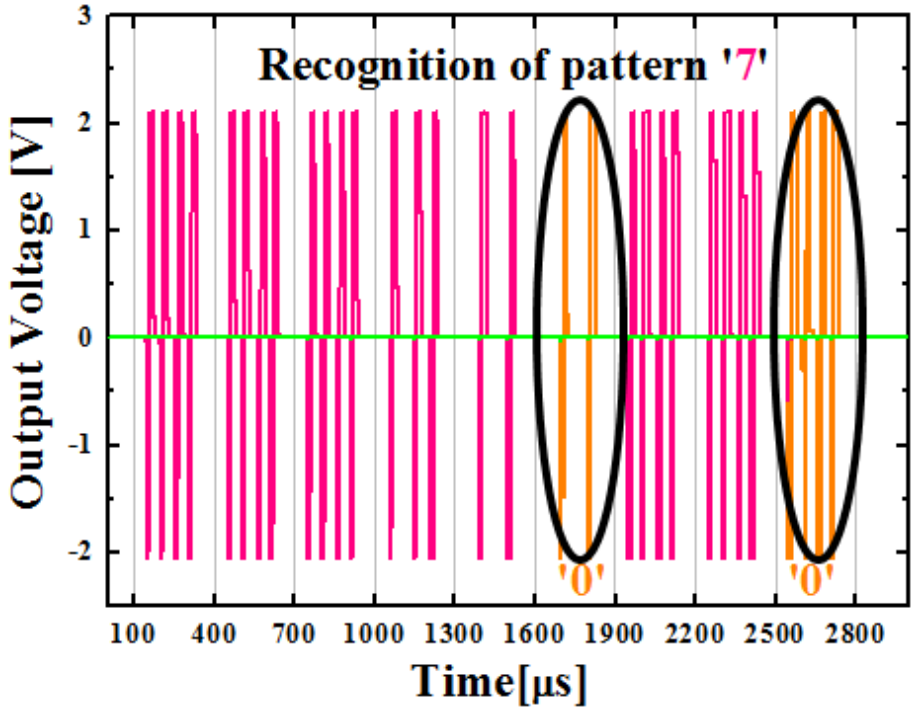


(b)

Fig. 5. 12. (a) The nine ‘6’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘G’ when the ‘6’ samples were recognized in turn.

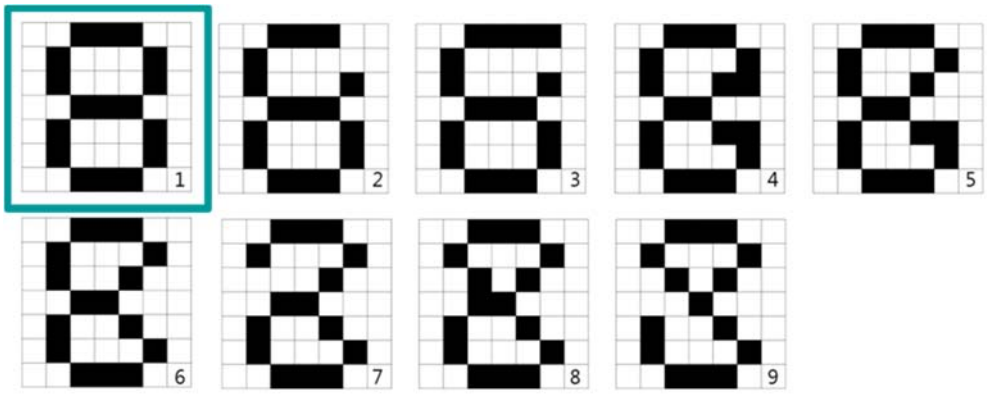


(a)

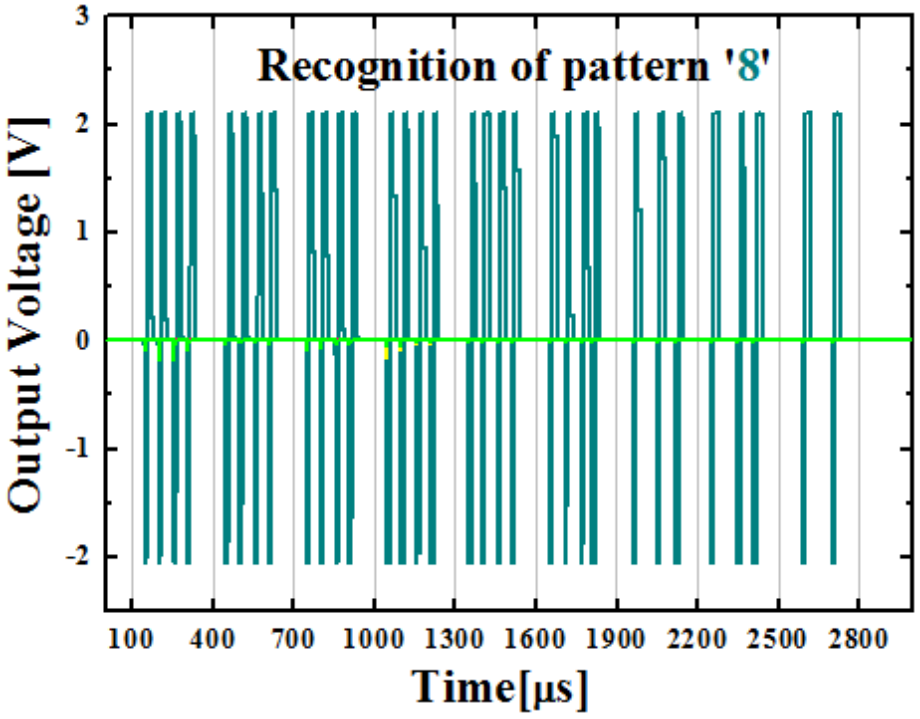


(b)

Fig. 5. 13. (a) The nine ‘7’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘I’ when the ‘7’ samples were recognized in turn.

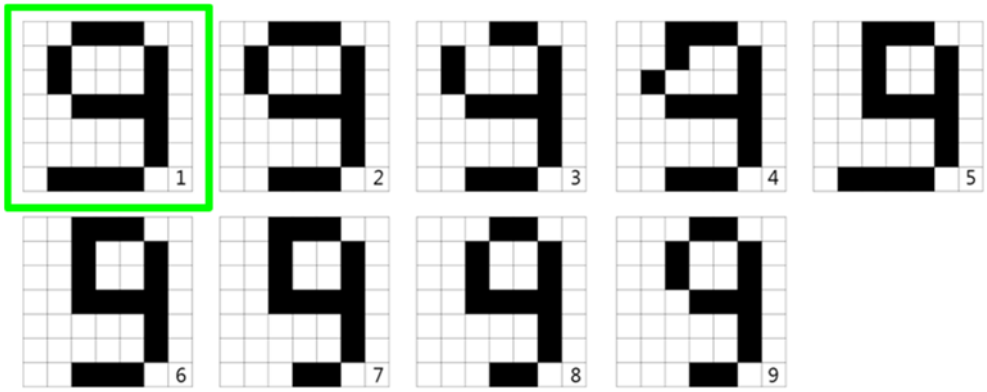


(a)

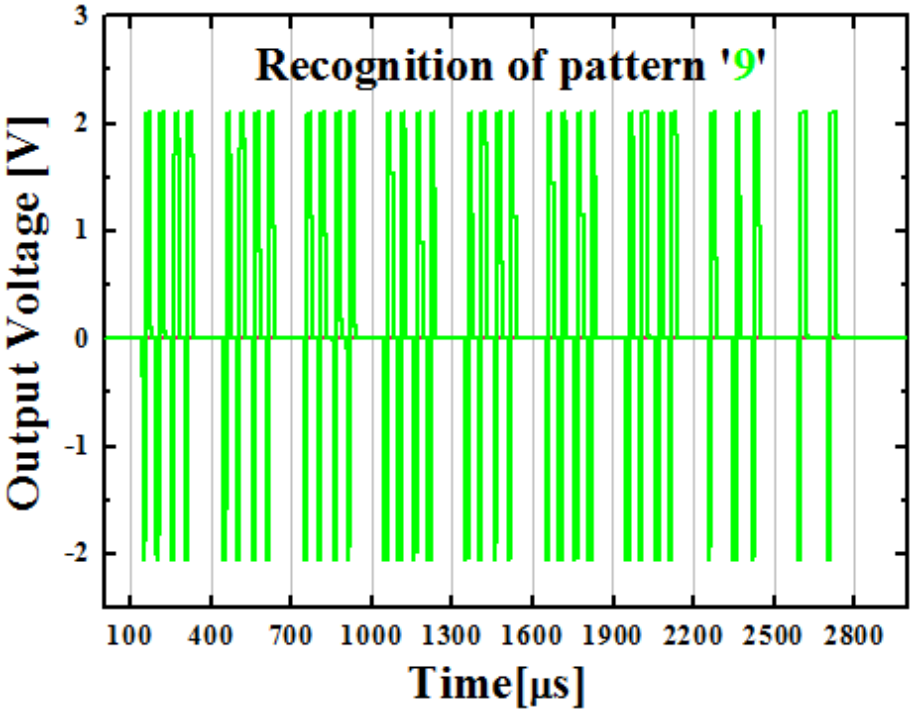


(b)

Fig. 5. 14. (a) The nine ‘8’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘D’ when the ‘8’ samples were recognized in turn.



(a)



(b)

Fig. 5. 15. (a) The nine ‘9’ samples used for recognition of the noise pattern. (b) The behavior of fire of the post-neuron ‘J’ when the ‘9’ samples were recognized in turn.

Chapter 6

Conclusions

In this thesis, the author proposed and fabricated a compact neuromorphic system that can be operated with 4-terminal Si-based synaptic devices, and verified that the system is applicable to spiking neural network applications such as numerical pattern recognition.

In order to configure the neuromorphic system, neurons and synapses constituting the human nervous system must be imitated. Therefore, biological synapses and neurons are imitated through Si-based Floating-body Synaptic Transistor (SFST) and Integrate & Fire (I & F) neuron circuits, respectively, in this research. The SFST has a dual gate and charge trap layer, and can be fabricated with a lateral stack structure using a side-wall spacer. This device can emulate short-term memory and long-term memory of biological synapses using floating body effect and charge injection, respectively. When an asymmetrical signal is applied to the front-gate and the back-gate with a time difference, it can change own conductance based on Spike-Timing-Dependent-Plasticity (STDP) rule.

The (I&F) neuron circuit is divided into a synaptic integration part and an action potential generation part. The synaptic integration part integrates the signal through the current mirror connected with the N-channel and the P-channel synaptic device and transfers it to the capacitor, C_1, in the circuit. The action-potential generation part generates

asymmetric pulses through two consecutive inverters and an NMOS with a large V_T, and is assisted by several MOSFETs that function to rapidly increase or initialize the voltage of C1 (V_{C1}). The generated action potential is directly transferred to the back-gate of the synaptic device located at the entrance of the integration part, and the conductance of the synapse device is autonomously changed by the STDP. The autonomous change of conductance is accomplished without the aid of any external circuitry and logic operations, making it a key factor in constructing a compact neuromorphic system with low power consumption.

The operation of I&F circuit has been proved through the construction of Printed Circuit Board (PCB) and it has been proved that STDP occurs stably in the connection between the PCB and the synaptic device in SOI wafer. Since then, neuron circuit has been fabricated directly through CMOS standard process and the stable operation of NMOS, PMOS, capacitor, one inverter and two consecutive inverters constituting circuit is observed and voltage integration through current mirror and final pulse generation are also verified. In order to investigate the applicability of the proposed system to spiking neural network applications, a 1-layer numerical pattern recognition system is constructed and numerical patterns from ‘0’ to ‘9’ are learned by unsupervised learning method. Although the system recognizes most of the numbers well, since the sample patterns are simply expressed in black and white through 7*7 cells and only 1-layer is constructed, the recognition rate is low for some noise-added patterns. If these disadvantages are improved, the proposed system can be applied to more various applications in the future.

Bibliography

[1] R. Ananthanarayanan, S. K. Esser, H. D. Simon, and D. S. Modha, “The cat is out of the bag: Cortical simulations with 10^9 neurons, 10^{13} synapses,” in Proc. IEEE/ACM Conf. High Perform. Netw. Comput., 2009, pp. 1–12.

[2] 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 Trans. Electron Devices, vol. 58, no. 8, pp. 2729–2737, Aug. 2011.

[3] K. D. Cantley, A. Subramaniam, H. J. Stiegler, R. A. Chapman, and E. M. Vogel, “Hebbian learning in spiking neural networks with nanocrystalline Silicon TFTs and memristive synapses,” IEEE Trans. Nanotechnology, vol. 10, no. 5, pp. 1066–1073, Sep. 2011.

[4] S. Park, J. Noh, M.-l. Choo, A. M. Sheri, M. Chang, Y.-B. Kim, C. J. Kim, M. Jeon, B.-G. Lee, and B. H. Lee, “Nanoscale RRAM-based synaptic electronics: toward a neuromorphic computing device,” Nanotechnology, vol. 24, p. 384009, Mar. 2013.

[5] S. B. Eryilmaz, D. Kuzum, S. Yu, H. -S. P. Wong, “Device and system level design considerations for analog-non-volatile-memory based neuromorphic architectures,” in IEDM Tech. Dig., 2015, pp. 64-67.

[6] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J.-P. Han, C. Miller, K. Hosokawa, C. Lam, “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 IEDM Tech. Dig., 2015, pp. 443-446.

[7] Z. Chen, B. Gao, Z. Zhou, P. Huang, H. Li, W. Ma, D. Zhu, L. Liu, X. Liu, J. Kang, H.-Y Chen, "Optimized Learning Scheme for Grayscale Image Recognition in a RRAM Based Analog Neuromorphic System," in IEDM Tech. Dig., 2015, pp. 467-470.

[8] K. Moon, E. Cha, J. Park, S. Gi, M. Chu, K. Baek, B. Lee, S. Oh, H. Hwang. "High density neuromorphic system with Mo/$Pr_{0.7}Ca_{0.3}MnO_3$ synapse and NbO_2 IMT oscillator neuron," in IEDM Tech. Dig., 2015, pp. 463-466.

[9] D. Lee, J. Park, K. Moon, J. Jang, S. Park, M. Chu, J. Kim, J. Noh, M. Jeon, B. H. Lee, B. Lee, B.-G Lee, H. Hwang, "Oxide based nanoscale analog synapse device for neural signal recognition system," in IEDM Tech. Dig., 2015, pp. 91-94.

[10] C. Zamarreño-Ramos, L. A. Camuñas-Mesa, J. A. Pérez-Carrasco1, T. Masquelier, T. Serrano-Gotarredona and B. Linares-Barranco, "On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex", Front. Neurosci, vol. 5, Mar. 2011.

[11] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, and H. Hwang, "Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device", Nanotechnology, vol. 22, no. 25, p. 254023, May. 2011.

[12] B. Rajendran, Y. Liu, J. Seo, K. Gopalakrishnan, L. Chang, D. J. Friedman, and M. B. Ritter, “Specifications of nanoscale devices and circuits for neuromorphic computational systems”, IEEE Trans. Electron Devices, vol. 60, no. 1, pp. 246–253, Jan. 2013.

[13] D. Querlioz, O. Bichler, P. Dollfus, and C. Gamrat, “Immunity to Device Variations in a Spiking Neural Network with Memristive Nanodevices,” IEEE Trans. nanotechnology, vol. 12, no. 3, pp. 288- 295, May. 2013.

[14] H. Tanaka, T. Morie, K. Alhara, “A CMOS Spiking Neural Network Circuit with Symmetric/Asymmetric STDP Function,” IEICE Trans. Fundamentals, vol. E92-A, no. 7, pp. 1690-1698, Jul. 2009.

[15] A. Afifi, A. Ayatollahi, F. Raissi, “Implementation of biologically plausible spiking neural network models on the memristor crossbar-based CMOS/nano circuits,” in Proc. Euro. Conf. on Circuit Theory and Design, 2009, pp. 563-566.

[16] S. Ghosh-Dastidar, H. Adeli, “Spiking Neural Networks,” Int. Journal of Neural Systems, vol. 19, no. 4, pp.295-308, Aug. 2009.

[17] Y. Ota and B. M. Wilamowski, “Analog implementation of pulse-coupled neural networks”, IEEE Trans. Neural Netw., vol. 10, no. 3, pp. 539-544. May. 1999.

[18] M. Boegerhausen, P. Suter, and S.-C. Liu, “Modeling short-term synaptic depression in silicon,” Neural Comput., vol. 15, pp. 331-348, Feb. 2003.

[19] T. Asai, Y. Kanazawa, and Y. Amemiya, “A subthreshold MOS neuron circuit based on the Volterra system,” IEEE Trans. Neural Netw., vol. 14, pp. 1308-1312, Sep. 2003.

[20] G. Indiveri, E. Chicca, and R. Douglas, “A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity,” IEEE Trans. Neural Netw., vol. 17, pp. 211-221, Jan. 2006.

[21] J. H. Wijekoon and P. Dudek, “Compact silicon neuron circuit with spiking and bursting behaviour,” Neural Netw., vol. 21, pp. 524-534, Mar. 2008.

[22] S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetti, and D. Ielmini, “Spike-timing dependent plasticity in a transistor-selected resistive switching memory,” Nanotechnology, vol. 24, p. 384012, Sep. 2013.

[23] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S. P. Wong, “A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: Experimental characterization and large-scale modeling,” in IEDM Tech. Dig., 2012, pp. 239-242.

[24] S. Park, H. Kim, M. Choo, J. Noh, A. Sheri, S. Jung, K. Seo, J. Park, S. Kim, and W. Lee, “RRAM-based synapse for neuromorphic system with pattern recognition function,” in IEDM Tech. Dig., 2012, pp. 231-234.

[25] K.-H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, and W. Lu, “A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications,” Nano letters, vol. 12, pp. 389-395, Dec. 2011.

[26] H. Kim, J. H. Lee, G. Kim, M.-C. Sun, and B.-G. Park, “Silicon-based floating-body synaptic transistor,” in Proc. Int. Conf. on Solid State Devices and Materials, 2012, pp. 322-323.

[27] H. Kim, J. Park, M.-W. Kwon, J.-H. Lee, B.-G. Park, “Silicon-Based floating-body synaptic transistor with frequency-dependent short- and long-term memories,” IEEE Electron Device Letters, vol. 37, no. 3, pp. 249-252, Mar. 2016.

[28] H. Kim, J.-H. Lee, B.-G. Park, “Asymmetric dual-gate-structured one-transistor dynamic random access memory cells for retention characteristics improvement,” Applied Physics Express, vol. 9, no. 8, pp. 084201-1-084201-4, Aug. 2016.

[29] M. Mahowald, R. Douglas, “A silicon neuron,” Nature, vol. 354, pp. 515–518, Dec. 1991.

[30] Y. J. Lee, J. Lee, Y. B. Kim, J. Ayers, A. Volkovskii, A. Selverston, H. Abarbanel, M. Rabinovich, “Low power real time electronic neuron VLSI design using subthreshold technique,” in Proc. of the Int. Symp. on Circuits and Systems, 2004, pp. 744-747.

[31] A. Van Schaik, C. Jin, A. McEwan, T. J. Hamilton, “A log-Domain implementation of the Izhikevich neuron model,” in Proc. of the Int. Symp. on Circuits and Systems, 2010, pp. 4253–4256.

[32] Y. Babacan, F. Kaçar, and K. Gürkan, “A spiking and bursting neuron circuit based on memristor,” Neurocomputing, vol. 203, pp. 86-91, Mar. 2016.

[33] H. Kim, J. Park, M. –W. Kwon, S. Hwang, B.-G. Park, "Multi-threshold voltages in ultra thin-body devices by asymmetric dual-gate structure," Japanese Journal of Applied Physics, vol. 55, no. 4S, pp. 04ED01-1-04ED01-4, Apr. 2016.

초록

폰 노이만 구조에서 기원한 기존의 전자 시스템이 근본적인 물리적 한계와 직면하게 되고, 비용과 효율 측면에서 약점을 드러냄에 따라, 우리는 정보 처리 분야에 있어서 새로운 패러다임을 도입해야 하는 필요성을 느끼게 되었다. 기존의 시스템에 대한 대안으로, 인간의 신경망에서 영감을 얻은 뉴로모픽 시스템이 최근 들어 개발되기 시작했다. 뉴로모픽 시스템은 특정한 인공 신경망을 기반으로 구성될 수 있으며, 다양한 인공 신경망들이 해당 연구 분야에 등장하고 있다. 그 중에서도, 시간적인 인간의 스파이킹 뉴런을 기반으로 한 스파이킹 신경망이 우수한 정보 처리 능력을 바탕으로 많은 주목을 끌고 있다. 스파이킹 신경망에 더해진 시간적 차원은 방대한 신경망을 간결하게 표현하는 결과를 불러왔으며, 앞으로 복잡한 시간-기반 패턴 인식 문제를 해결하기 위한 큰 가능성을 갖고 있다.

본 논문에서는 스파이킹 신경망에 적용 가능하고 4단자 실리콘-기반 시냅스 모방 소자와 함께 작동 가능한 간결한 뉴로모픽 시스템을 제안하며, 표준 상보성 금속 산화막 반도체 공정을 기반으로 한 뉴런 회로의 제작 방법을 설명한다. 제안된 시스템은 크게 시냅스 집적 부분과 활동-전위 생성 부분으로 나눠지며, 시냅스 집적 부분은 생물학적 뉴런의 흥분과 억제를 각각 표현하기 위해 N-채널 및 P-채널 시냅스 모방 소자와 연결된 전류 거울을 내재하고 있다. 시냅스 모방 소자는 부유-바디 효과와 부유 게이트로의 전하

주입을 이용하여 비중을 변화시킬 수 있으며, 시냅스 모방 소자의 개수와 비중은 전류 거울로부터 재 생성되는 전류의 양에 영향을 미친다. 활동-전위 생성 부분은 비대칭적인 활동 전위를 생성하기 위해 두 단계의 인버터와 높은 문턱 전압을 갖는 N형 금속 산화막 반도체 전계 효과 트랜지스터를 내재하고 있다. 생성된 활동 전위는 다음 뉴런으로 전달되는 동시에 시냅스 모방소자의 후방 게이트로 되돌아가는데, 이는 스파이크-타이밍-기반-유연성을 근거로 하여 시냅스 모방 소자의 비중을 변화시키기 위함이다.

4단자 시냅스 모방 소자는 전-뉴런의 신호를 전달하는 동시에 스스로의 비중을 변화시킬 수 있으므로, 추가적인 스위치나 논리 연산 없이도 간결한 뉴로모픽 시스템을 구성할 수 있게 되며, 최소한의 소자 개수와 전력 소모를 통하여 뉴런의 동작을 모방할 수 있다. 본 시스템의 이러한 장점들은, 매우 효율적으로 패턴 인식과 같은 다양한 스파이킹 신경망의 응용을 구성하는데 도움이 될 것이라고 기대 된다.

주요어: 뉴로모픽, 뉴런 회로, 시냅스 모방 소자, 활동-전위,
스파이크-타이밍-기반-유연성, 스파이킹 신경망

학번: 2010-20806