Ph.D. DISSERTATION

Low Power Synaptic Device using Positive Feedback Field Effect Transistor with Charge Trap Layer

양성 피드백 전계 효과 트랜지스터를 활용한 저전력 시냅스 소자

BY

KIM SUHYEON

AUGUST 2020

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

Ph.D. DISSERTATION

Low Power Synaptic Device using Positive Feedback Field Effect Transistor with Charge Trap Layer

양성 피드백 전계 효과 트랜지스터를 활용한 저전력 시냅스 소자

BY

KIM SUHYEON

AUGUST 2020

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

Low Power Synaptic Device using Positive Feedback Field Effect Transistor with Charge Trap Layer

양성 피드백 전계 효과 트랜지스터를 활용한 저전력 시냅스 소자

지도교수 박 병 국

이 논문을 공학박사 학위논문으로 제출함

2020년 8월

서울대학교 대학원

전기 · 정보 공학부

김 수 현

김수현의 공학박사 학위 논문을 인준함

2020년 8월

위 원 장:	이 종 호	(인)
부위원장:	박 병 국	(인)
위 원:	박 철 홍	(인)
위 원:	조 성 재	(인)
위 원:	김 형 진	(인)

Abstract

The neuromorphic system has been widely used and commercialized in many fields in recent years due to its potential for complex problem solving and low energy consumption. The basic elements of this neuromorphic system are synapse and neuron circuit, in which synapse research is focused on emerging electronic devices such as resistive change memory (RRAM), phase-change memory (PCRAM), magnetoresistive random-access memory (MRAM), and FET-based devices.

Synapse is responsible for the memory function of the neuromorphic system, that is, the current sum quantization with the specific weight value. and the neuron is responsible for integrating signals that have passed through the synapse and transmitting information to the next synapse. Since the synapse element is the largest portion of the whole system, It consumes most of the power of the entire system. So low power implementation is essential for the synapse device. In order to reduce power consumption, it is necessary to lower the off-current leakage and operate on low voltage. To overcome the limitation of MOSFETs in terms of $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio, small sub-threshold swing and power consumption, various devices such as a tunneling field-effect transistor (TFET), negative capacitor field-effect transistor (NCFET), ferroelectric field-effect transistor (FeFET), and feedback field-effect transistor (FBFET) have been stud-

ied.

Another important factor in synapse devices is the cost aspect. The deep learning technology that made Alpha-go exist is also an expensive system. As we can see from the coexistence of supercomputers and personal computers in the past, the development of low-cost chips that can be used by individuals, in the end, is inevitable. Because a CMOS compatible process must be possible since the neuron circuit is needed to fabricate at the same time, which helps to ensure mass productivity. FET-based devices are CMOS process compatible, which is suitable for the mass production environment.

A positive FBFET (Feedback Field Effect Transistor) device has a very low subthreshold current, SS (subthreshold swing) performance, and I_{ON}/I_{OFF} ratio at the low operating voltage. We are proposing the synaptic device with a positive FBFET with a storage layer.

From the simulation study, the operation method is studied for the weight modulation of the synaptic device and electrical measurement confirms accumulated charge change by program and erase condition each. These results for the synaptic transistor in this dissertation can be one of the candidates in low power neuromorphic systems.

keywords: Neuromorphic system, Positive feedback field effect transistor, FBFET, Dual gate thyristor, low-power synapse, subthreshold swing

student number: 2016-30204

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Limitation of von Neumann Architecture computing

The conventional computing system has reached the level of solving almost all problems in nature based on CMOS logic and memory using von Neumann architecture. [1] The result was successful in sending humans to the moon and recreating the Big Bang. The evolution of semiconductors supported it by enabling faster computing through the development of device technology. But von Neumann also recognized the processing speed bottleneck, which is from existing of data transferring between the memory system and CPU. The system performance is is determined by the speed of transferring between CPU and storage through BUS. [2, 3, 4] Fig. 1.1

To solve these problems, parallel computing, in-memory computing, and neuro-

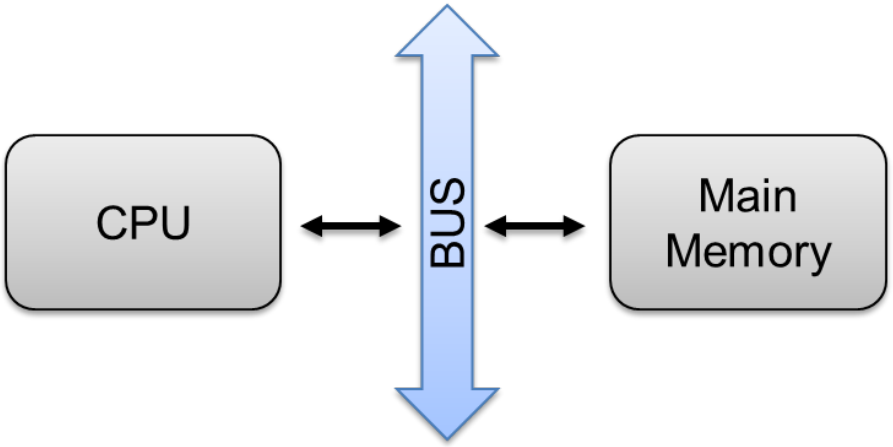


Figure 1.1: Basic structure of a classical von Neumann architecture.

morphic computing have been proposed. [6, 7, 8] Parallel computing achieved performance improvement. However, It also has a limitation to improve speed by Amdahl's law. Fig. 1.2 Currently, in-memory computing and neuromorphic computing are being studied. Especially, neuromorphic computing become a hot topic because it is inspired by the biological system and estimated to improve power consumption and performance.

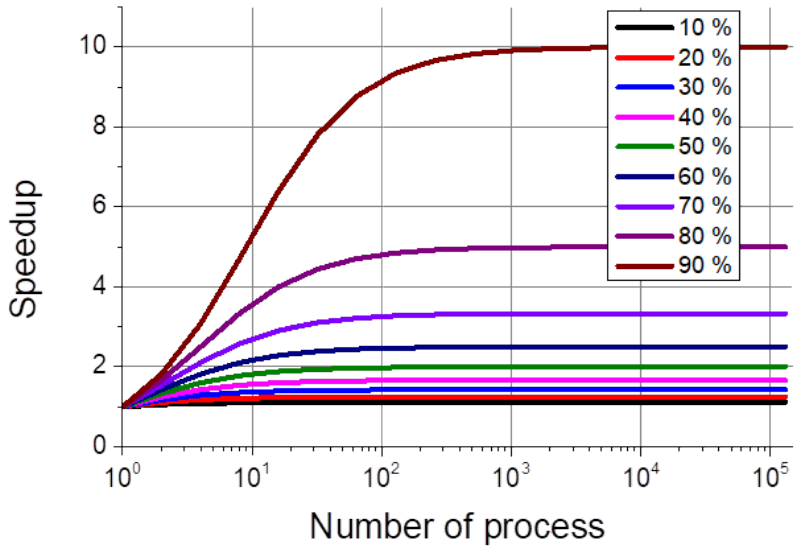


Figure 1.2: Amdahl's law and strong scaling.

1.2 Biological Synapse

The biological nervous system consists of neurons and synapses. synapse is responsible for receiving and transmitting signals between neurons. This is called neurotransmission. The signal transmitted through the synapse is integrated through the neuron, and the neuron sends the signal to the connected post-synapse. These connections are continuously connected and form a neural network. [5] Fig. 1.3

The essential role of the synapse is to determine the frequency and pattern of the signal when sending the received signal to the next neuron. One synapse amplifies the signal, and another synapse attenuates the signal. The characteristic value of each

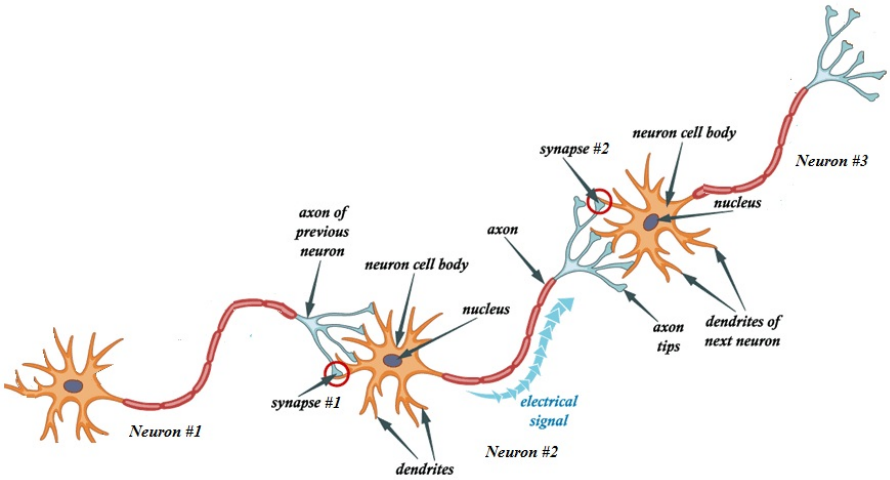


Figure 1.3: Series of neurons and synapses

synapse is called synaptic weight. With this combination of synapses and neurons, humans can recognize and remember the information. From this mechanism, synapse has to remember a certain weight value. This behavior is similar to a memory device. And neuron is corresponding to logic circuit.

Biological synapse transmits signals in a chemical way. The electrical signal transmitted to the axon terminal releases a neurotransmitter and transmits information to the dendritic spine of the next neuron by the receptor. This synapse is classified as an inhibitory synapse that reduces the membrane potential of the synapse and the excitatory synapse that amplifies information. Each synapse changes their strength through the learning mechanism by changing the receptor structure. This connected system is explained by the neural network model.

1.3 Spiking Neural Network (SNN)

The neural network system was created as a digital model by utilizing the existing semiconductor system. The digital system, called the artificial neural network (ANN) system, has been well known in computing systems of von Neumann architecture through Deep-learning algorithms such as a fully connected network (FCN), convolutional neural network (CNN), recurrent neural network (RNN) and so on. [9, 10, 11, 12, 13] This digital system contributed to solving the problem of natural language processing and pattern recognition. This system calculates the weighted sum of the input data with each synapse's weight value. This weighted sum calculation is vector dot production. So we called it vector-matrix multiplication (VMM). This operation can greatly increase the calculation speed by using parallel computing. Therefore, a parallel processor such as a GPU is often used to calculate the ANN model. Matrix multiplication required a large number of parallel devices. This system requires many synapses and circuits. Therefore, big computing power is mandatory. Then, power consumption and cost problems are inevitable. To solve this fundamental problem, a hardware-based analog neural network model was studied. [14, 15, 16, 17, 18, 19, 20, 21, 22, 23] Fig. 1.4 Existing GPU-based digital architecture is classified into a synapse device and neuron circuit similar to the biological neural network system. The VMM operation is performed in the logic part based on stored weight value in memory. This memory

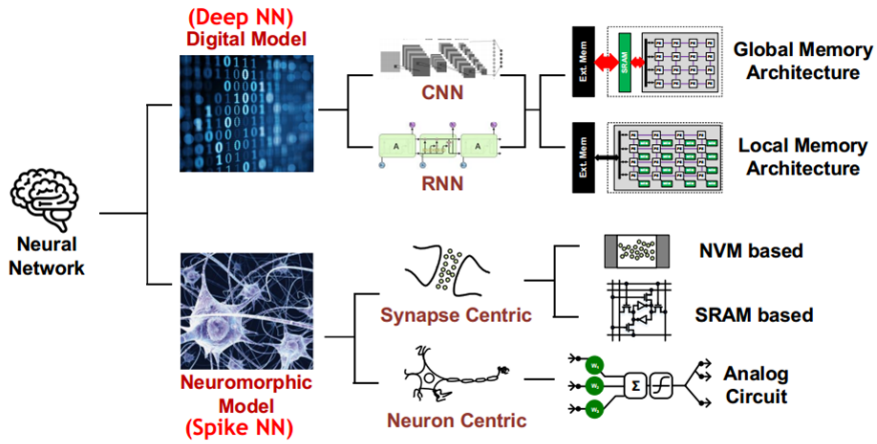


Figure 1.4: Neural network system for digital and analog model [24]

part architecture is replaced by the memorable device in SNN which called synapse device. The synaptic device has a role which is information connector and memory of synaptic weight. And the neuron circuit takes the place of analog CMOS circuit. As a synapse device candidate, all kinds of the memory devices are being studied from existing non-volatile memory to emerging devices.

Existing devices are NAND-Flash RAM, NOR-Flash RAM, SRAM, DRAM, and emerging devices are Resistive-switiching RAM (RRAM), Ferro-electric FET (Fe-FET), phase-change RAM (PCRAM), and magnetic RAM (MRAM) is being studied. [25, 26, 27, 28, 29, 30, 31] Fig. 1.5 Each device has a pros and cons as is.

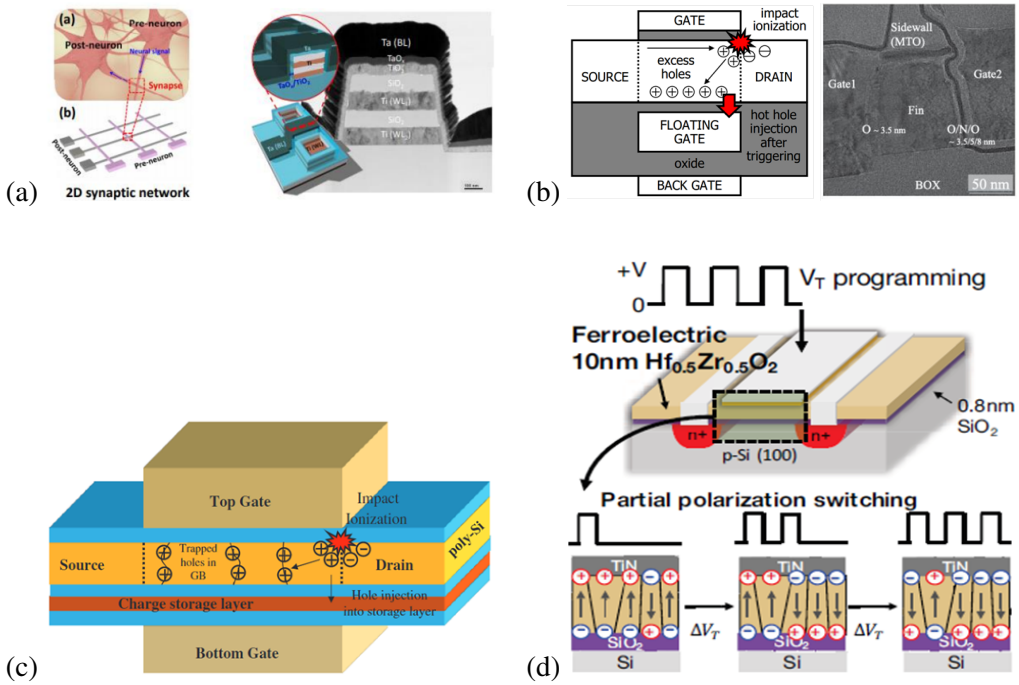


Figure 1.5: Various synaptic Devices (a) Resistive Switching RAM (RRAM) [26] (b) Dual gate silicon FET [27] (c) Poly-silicon based synapse [28] (d) Ferroelectric FET [29]

1.4 Requirements of synaptic device

As described in the previous paragraph, the neuromorphic system must have both synapse part and neuron part. We can make each part separately. But it's not efficiency in terms of cost. The neuron circuit is same as general CMOS circuit that follows a standard semiconductor process. But if additional special process for synapse device, there should be performance degradation, product variation, and cost overhead issues. This is one of most important requirements for synapse mass production. In addition,

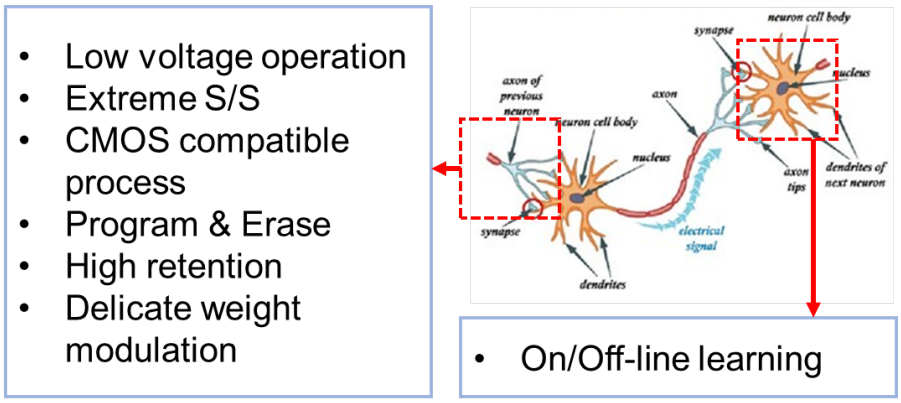


Figure 1.6: Synaptic device requirements

a low current operation is necessary because the synapse device occupies most of the entire neural network system.[32]

In addition, weight modulation which is a synapse-specific function is required to convert the pre-synaptic signal to a specific size and frequency signal according to each weight size. High-level current quantization property and low current variability can improve system accuracy. These are whole memory related characteristic. That is function of program (PGM) and erase (ERS). And high retention property is also required because energy consumption is quite big in PGM and ERS. An on line learning function is also required.

Most of all, low current characteristic is the important factor. Even ANN has shown great achievement in many ways. Super computing power is mandatory. [33] Fig. 1.6

1.5 Advantage of Feedback Field-effect transistor (FBFET)

FBFETs are representative devices that can meet the requirements of synaptic devices which are mentioned in previous chapter. FBFETs has attracted attention as a future low-power device with its ultimate sub-threshold characteristic, high on/off ratio, low leakage , and low voltage operation. There are many types of FBFEETs such as a dual gated thyristor, gated diode, Z^2-FET, and Z^3-FET. [34, 35, 36, 37] These devices are already used widely for the various fields in the semiconductor business.

FBFETs are operating with very low on current I_{ON} because off current I_{OFF} can be suppressed. [38, 39, 40, 41] and FBFETs can be fabricated with the standard CMOS mass production method. These two advantages are the most attractive point. That is, FBFETs can be fabricated with neuron circuits with low operating voltage. If we can add function of memory in FBFETs, there must be a change to use synaptic device in powerful ways.

1.6 Outline of the Dissertation

The purpose of this dissertation is to propose the novel synaptic device using FBFETs with the storage layer, and the main targets are presentations of the structure of FBFET synapse, the operation method as synaptic device, PGM-ERS conditions with inhibition cases in array scheme, fabrication method, and array structure with this device. Based on the previous discussion, The remainder of this dissertation is organized as follows. Chapter 2 introduces the basic operation physics of FBFETs and method as the synaptic device and covers verifying operation by TCAD simulation. In chapter 3, the process flow for the fabrication is described and measurement data of the fabricated device with basic feedback operation, program erase operation, and hysteresis reduction method will be proposed. and further will be discussed about the temperature compensation method which is one of the weakest points in FBFETs. FBFETs have a high sensitivity to temperature because it has basically diode operation. In chapter 5, a device model developed by the current equation from measured data will be discussed for SPICE simulation for VMM. Chapter 5 concludes this dissertation with a summary.

Chapter 2

Positive Feedback FET with storage layer

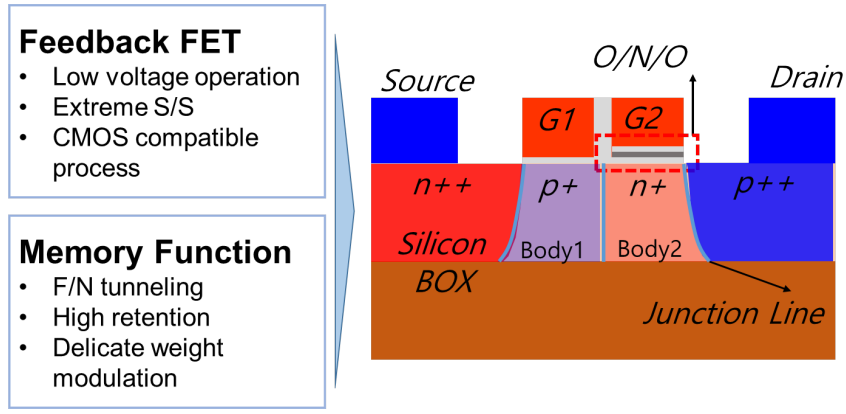
Fig.2.1 is a device structure. The designed FBFET is a n^{++}-p^{+}-n^{+}-p^{++}-doped double gates structure. The barrier height of the 1st channel region is adjusted by p-type implantation dose. The well depth of the 2nd channel region can be self-aligned implantation after 1st Gate patterning. A self-aligned implantation can define a precise junction position. The storage layer(O/N/O) is deposited after 2nd channel implantation. Only one mask is added compared to a conventional CMOS process. The 1st Gate controls the potential barrier of the electron. The 2nd Gate is used for a program and erase operation. The depth of well on the 2nd channel is controlled by the amount of trapped electron or trapped hole charge in the storage layer which consists of Silicon Oxide/Silicon Nitride/Silicon Oxide (ONO) structure between the 2nd Gate and the n+ doped channel. The specific parameters which are used for the simulation are dis-

closed in Table Fig.2.1. We use a reasonable value that can match the feature size of the device which is used as a reference.

Parameter	Silicon Value
Gate oxide thickness	$0.009\mu m$
Tunnel oxide thickness	$0.003\mu m$
Silicon Nitride thickness	$0.007\mu m$
Body thickness	$0.1\mu m$
Gate length (1, 2)	$0.4\mu m$
Body1 (p)/Body2 (n) doping concentration	$2 \times 10^{18} cm^{-3}$
Source (n)/drain (p) doping concentration	$1 \times 10^{21} cm^{-3}$

Table 2.1: Parameters of dual gate charge storable FBFET synapse for TCAD simulation.

(a)



(b)

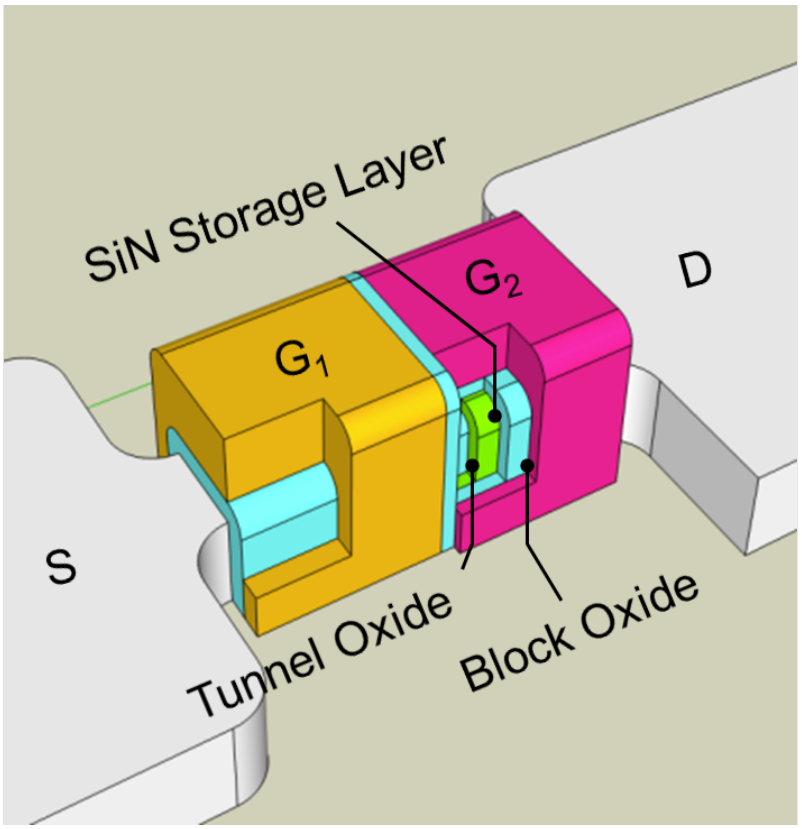


Figure 2.1: Schematic illustration (a) 2D scheme with doping profile (b) 3D illustration

2.1 Normal operation Principle of FBFET

The 1st Gate controls the electron's potential barrier height, and 2nd Gate adjusts the potential well depth for charge integration. When the n+ doped source is grounded and a positive voltage is applied to the 2nd Gate, any charge can not move through channel by each potential barrier. Once the positive voltage is biased to the 1st Gate and drain, electrons are injected from the source and accumulate in the n-type potential well under 2nd Gate. These electrons lower the potential barrier height of the valence band on the drain side. This barrier lowering allows the hole in the drain to be injected into the p-doped body under the 1st Gate. In the same way, the accumulated holes lower the potential barrier of the band at the source side for electron injection. As a result, the feedback loop can be formed, which makes the potential barrier height be very low and then the electron and hole currents increase dramatically at some point. The drain current is not changed by changing 1st Gate voltage level because the FBFET acts like a forward-biased p-n diode after the energy band is almost flat and the feedback loop is formed. Fig.2.2

But this mechanism which is triggering by 1st Gate voltage control has a serious problem. Because FBFET can not be turned off status by 1st Gate voltage control only. This device stays on-state once it turned on. The only way is to set the drain voltage to 0V to go off-state. From the viewpoint of the synapse device, when the 1st Gate input

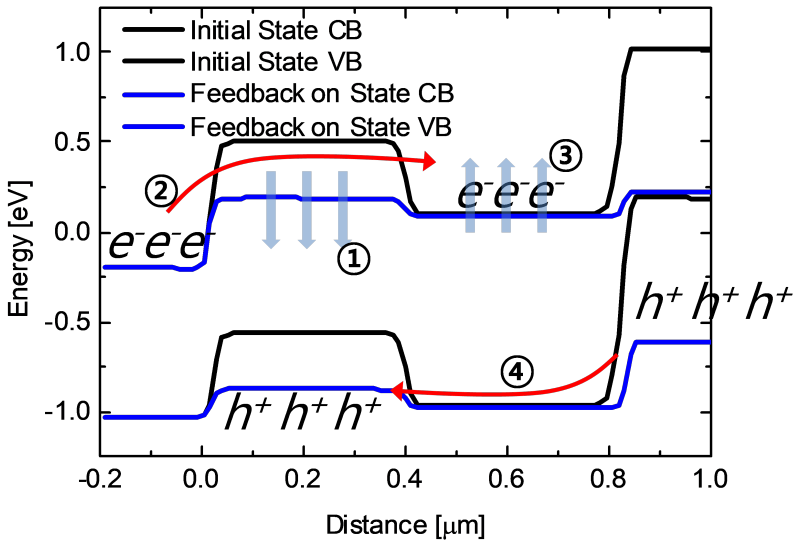


Figure 2.2: Band diagram of FBFET by 1st Gate input signal. ① As 1st Gate voltage increases, potential barrier of body1 for electron is lowered. ② Electron from source moves to body2 through potential barrier in body1. ③ Electron is accumulated in body2. It raises the energy band level. ④ Potential barrier is lowering by accumulated electrons in body2. Then holes can move body1. Positive feedback occurs by this flow.

signal disappears, the drain bias voltage must also be synchronized to 0V. This method is not suitable for asynchronous SNN systems. Therefore, we need to find a way to control the device with one input signal.

2.2 Operation Mechanism by Drain Input Pulse

The method using 1st Gate bias voltage can control feedback more intuitively. As mentioned earlier, you cannot create an asynchronous SNN system in this way. Therefore, in order to use it as a synaptic device, it is necessary to control the device by adjusting the drain bias voltage. When a positive drain bias is applied to 1st Gate while maintaining an appropriate bias voltage, the hole barrier between the body2 region and the drain is lowered, and the hole over the barrier is accumulated in body1, thereby lowering the conduction band energy of electrons. When the injected hole has a sufficient lifetime to accumulate in body1, it continuously lowers an electron barrier height in the body1 region, which is a condition that can make feedback. Fig.2.3

Weight modulation is required for use as a synapse device. When the amount of charge is adjusted in the O/N/O storage layer under 2nd Gate, an electric field is created according to the amount of trapped charge, and this electric field changes the channel potential. Eventually, the energy of the valence band can be controlled by changing the potential applied to the channel. When the electron is trapped, the band energy can be linearly raised. Since the body2 well depth becomes shallow, the device can quickly reach the feedback state. Fig.2.4 2.5

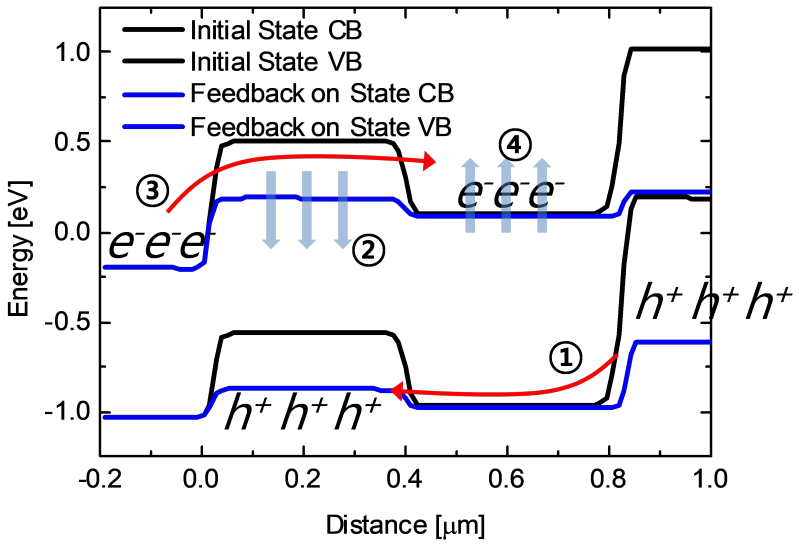


Figure 2.3: Band diagram of FBFET by drain input signal. ① Hole moves to body1 through barrier of 2nd channel. Accumulated holes make potential barrier lower on body1. ② Electron from source moves to body2 over potential barrier in body1. ③ Electron is accumulated in body2. It raises the band energy. ④ Potential barrier is lowering by accumulated electrons in body2. Positive feedback occurs by this flow

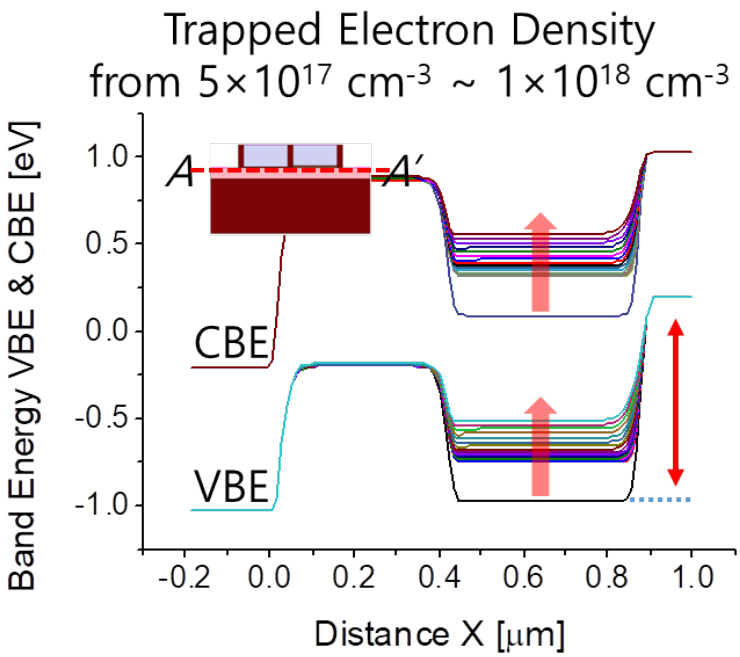


Figure 2.4: Valence band energy increases by amount of trapped electron

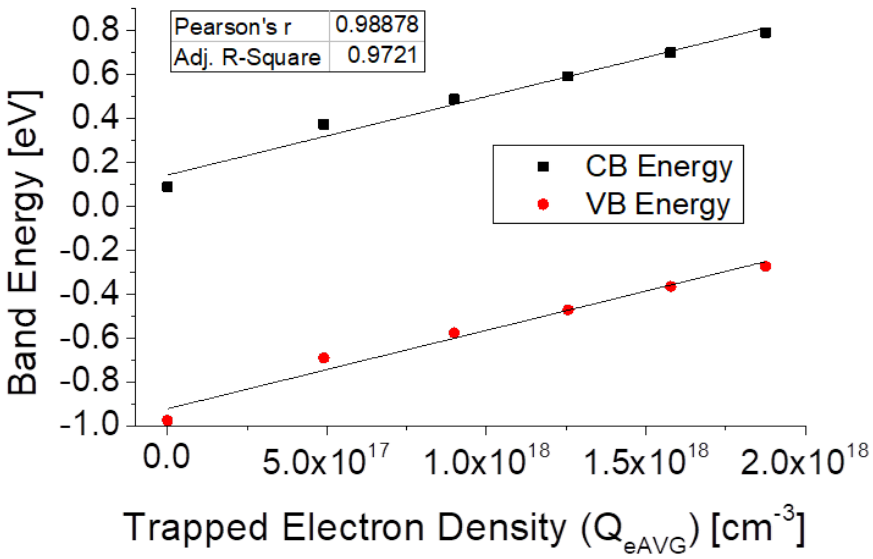


Figure 2.5: Linear correlation is between band energy and trapped density

2.3 Weight Modulation Mechanism

In ANN, the weighted sum is defined as vector-matrix multiplication. The weight value of the post neuron is the sum product of the input value and weight of each node. In a typical SNN, the input signal is defined by the same spike frequency, and the weight value of each synapse is defined by the device's threshold voltage (V_{TH}). When V_{TH} is large, less current flows, and when V_{TH} is small, a large current flows and the amount of charge accumulated in the input capacitor of the Integrate-and-Fire neuron circuit becomes the weighted sum. FBFET synapse can be used current modulation by controlling device's turn-on/turn-off time. As soon as the device turns on at the same input pulse, the number of charges accumulated in the membrane capacitor is the maximum. Even if FBFET synapse is off with input signal spike, no charge is not accumulated in membrane capacitor.the relationship between off-state time and the amount of accumulated charge is linear relation. Fig.2.6

Fig. 2.7 Shows the weighted sum modulation of the FBFET synapse. When the pulse width is 400 μsec, if the time when the synapse is off is set to t_{OFF}, the amount of charge accumulated in the membrane capacitor is calculated as follows.

$$i = C_{mem}\frac{dv_{mem}}{dt} \tag{2.1}$$

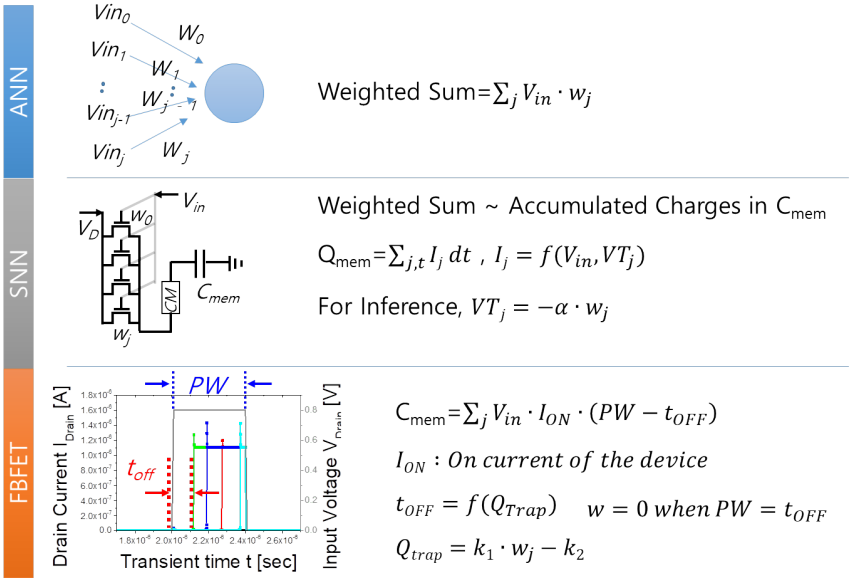


Figure 2.6: Weighted sum definition for each system ANN, SNN, FBFET synapse

$$\int \frac{dv_{mem}}{dt} dt = \int_0^{t_{OFF}} dt + \int_{t_{OFF}}^{PW} \frac{I_{ON}}{C_{mem}} dt \tag{2.2}$$

$$\Delta v_{mem} = -k_1 t_{OFF} + k_2 \tag{2.3}$$

$$k_1 = \frac{I_{ON}}{C_{mem}}, k_2 = \frac{I_{ON}}{C_{mem}} PW \tag{2.4}$$

Eventually, the change in membrane voltage is inversely proportional to t_{OFF}.

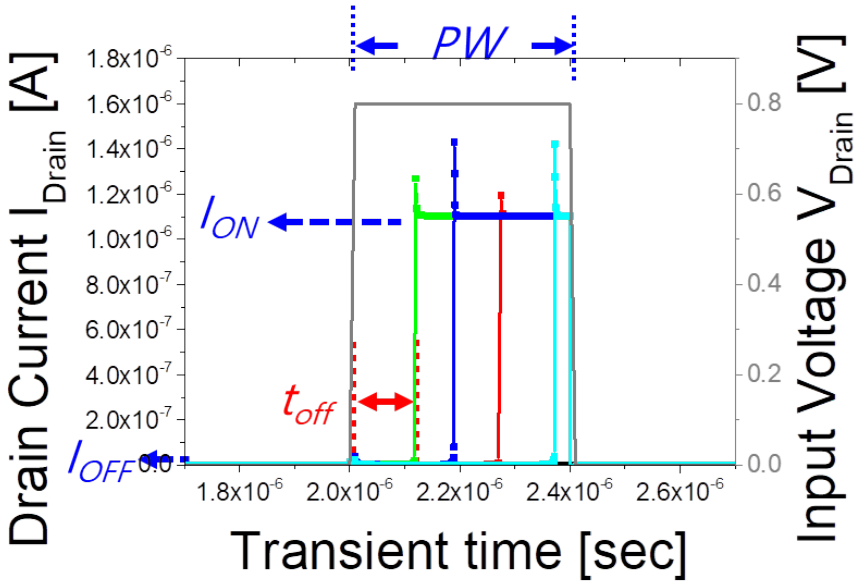


Figure 2.7: Weighted sum definition for FBFET synapse

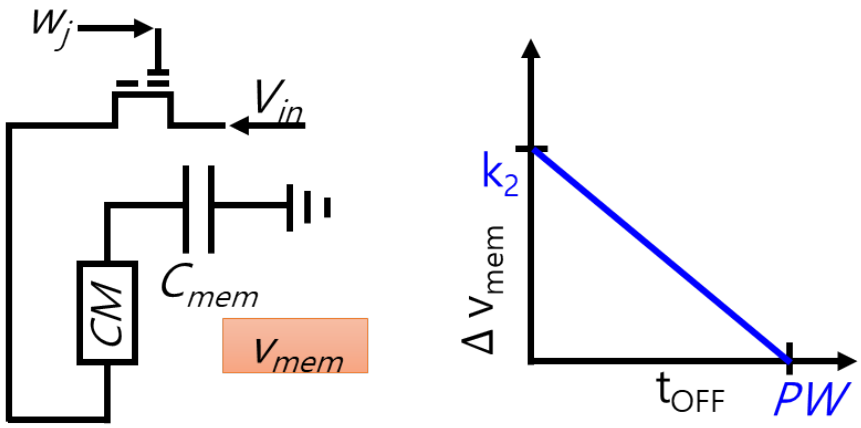


Figure 2.8: Membrane voltage for Integrate-and-Fire neuron circuit

2.4 TCAD Simulation Result for Weighted Sum

Fig. 2.9 is a transient simulation result for the behavior of the FBFET synapse according to the amount of electron trapped in the nitride storage layer. Device t_{OFF} decreases as the amount of trapped electrons increases. This is device V_{TH} is corresponding to turn-on time. As a result, when the drain voltage is 0.8V, it shows that it can be modulated in the sub-μA current. It can be seen that the amount of accumulation charge accumulated in the membrane capacitor of the neuron is proportional to the trapped charge. At a charge density of 1.0×10^{17} to $7.0 \times$S 10^{17}, the operation window is around 79% and has a relatively large modulation window. In some cases, it is possible to enlarge the operation window using a larger trap.

It is physically possible to trap electrons close to 1.0×10^{19}. So, it is possible to secure the operation window using high trapped electrons, but the amount of trapped electrons is limited in accumulation charge modulation. As Fig. Fig. 2.11 shows, the accumulation charge is saturated when the trapped electron density goes above 1.0×10^{18}. The reason for this is shown in Fig. 2.12. The band energy changes linearly up to about 1.0×10^{18}, but if it is larger than that, the well depth is not sufficient, so that the time for electrons to accumulate cannot be secured. To solve this, basically, there is a method of lowering the band energy height by increasing the doping concentration of the body2 part, and by trapping holes in the storage layer, the band energy can be

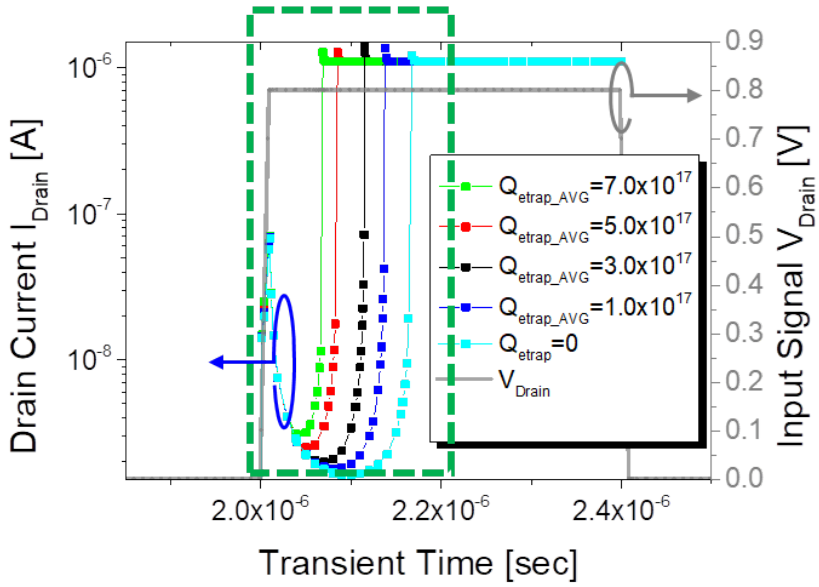


Figure 2.9: Turn-on time modulation by amount of trapped electron

lowered to secure the operation window.

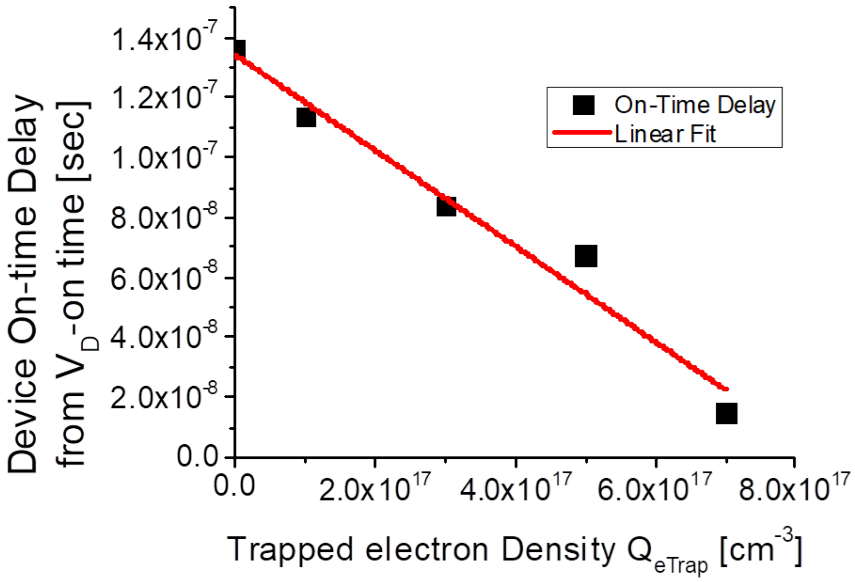


Figure 2.10: Turn-on time modulation by amount of trapped electron time wise. Delay time from time of V_D-on

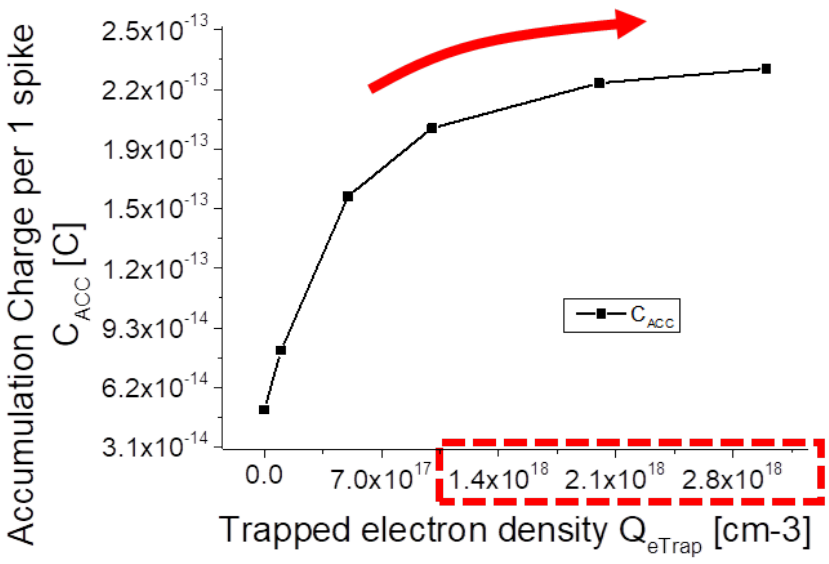


Figure 2.11: Accumulation charge for higher trapped electron

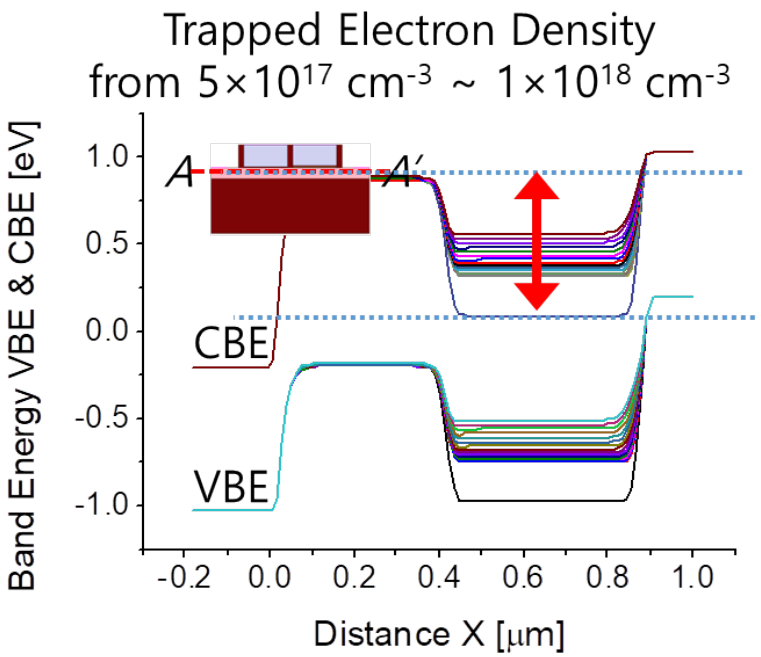


Figure 2.12: Band diagram for high trapped electron case

2.5 TCAD Simulation Result for Program and Erase

In order to control the weight of the FBFET synapse, the program operation is needed to control the number of electron traps in the nitride layer through FN tunneling. Conversely, when a negative weight is required, an electron trap can be removed or a hole can be injected through an erase operation. Fig. 2.13 is the result of the program operation simulation. The parameters used in the simulation are in Table. 2.2 The trapped electron are increases as the program pulse are given. Conversely, Trapped electrons are reduced through the erase operation. This is the same as the operation method of NAND device under mass production, and Incremental Step Pulse Programming (ISPP) and Incremental Step Pulse Erasing (ISPE) can be used for accuracy and efficiency of the program and erase. The erase efficiency is relatively low. Fig. 2.14 This can be solved by applying a higher erase voltage, but there is a limit to hole injection. Basically, the body2 area is an environment with many donors through n-channel doping, and the hole is extremely small. Because of this, it is considered that the erase efficiency is low. To solve this, it can be solved by applying a positive drain bias during erase and supplying a hole through the drain. Fig.2.15 As a result of the simulation, it was confirmed that the erase efficiency increased about 100 times when 5 V drain bias was applied.

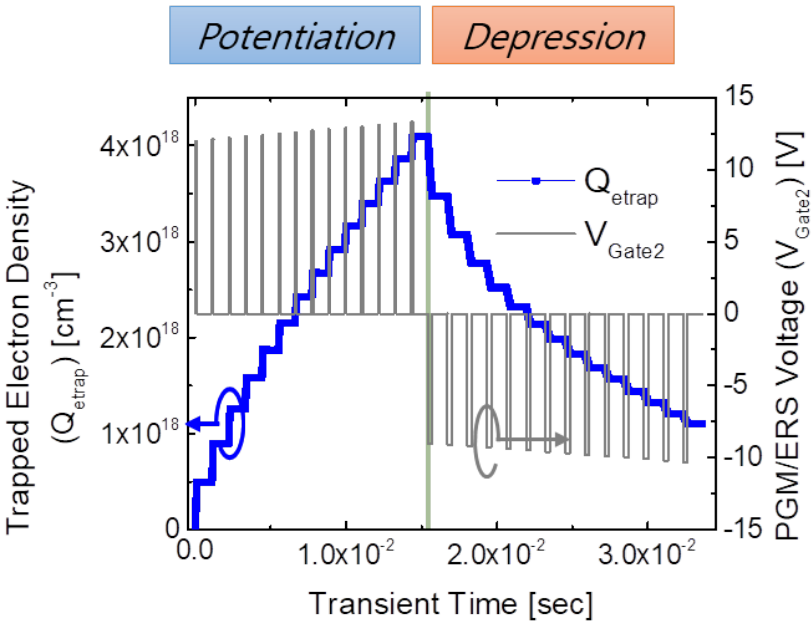


Figure 2.13: Potentiation by electron program

	Pulse Width	V_{Gate2}	V_{drain}	V_{source}
PGM condition	100 μ sec	12 V	0 V	0 V
ERS condition	300 μ sec	-10 V	0 V	0 V

Table 2.2: Condition of program and erase operation

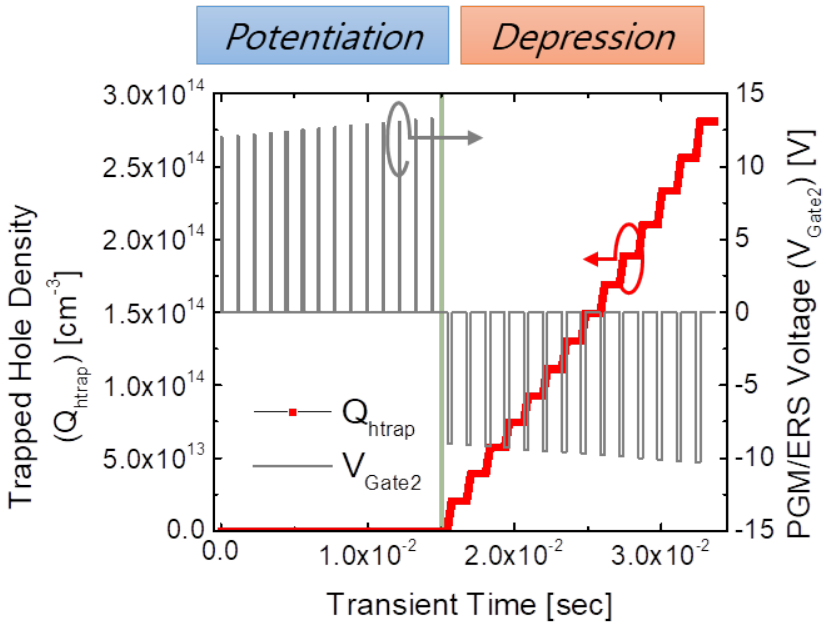


Figure 2.14: Depression by erase operation

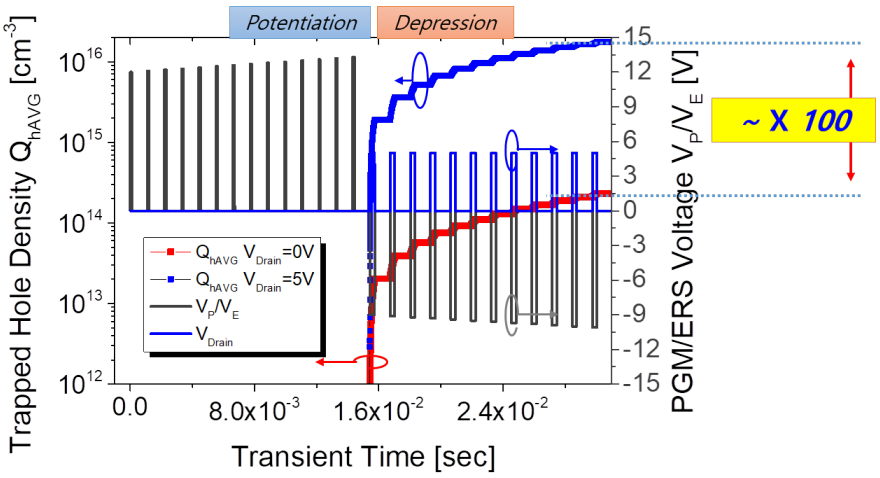


Figure 2.15: Hole erase efficiency by applying positive bias at drain

2.6 Array structure and Inhibition scheme

In order to use it as a synaptic device, the structure and operation method of the array must be clear. In Fig. 2.16 Through the peripheral circuit, the pre-synaptic input signal shares the drain input node of the synapse, and 2nd Gate used as the word line for the program and erase is connected horizontally. The source terminal is connected in the horizontal direction and connected to the post-synaptic output signal and sent to the neuron circuit. Fig.2.17 is a 3D array schematic. As shown in the figure, a large penalty is expected to control gate1 for each device. Therefore, gate1 should be connected in common to reduce device variation due to floating. In the case of Z^2-FET, gate1 is intrinsic, but device variation may occur in this case. At the same time, it is possible to implement a more accurate junction profile through the self align doping method of the body1 region.

The inhibition method is very important for the program and erase operations in an array structure. A specific device is needed to control to project the correct weight. In this part, the result of ANN weight should be the same as possible, so the neural network test through inference is the same. Fig.2.18 Indicates the case of the program on the target synapse S_{00}. At this time, the program voltage is applied to gate2 and the program voltage is also applied to S_{10} connected in the horizontal direction. In this state, both synapse S_{00} and S_{10} are programmed. At this time, if the same voltage as

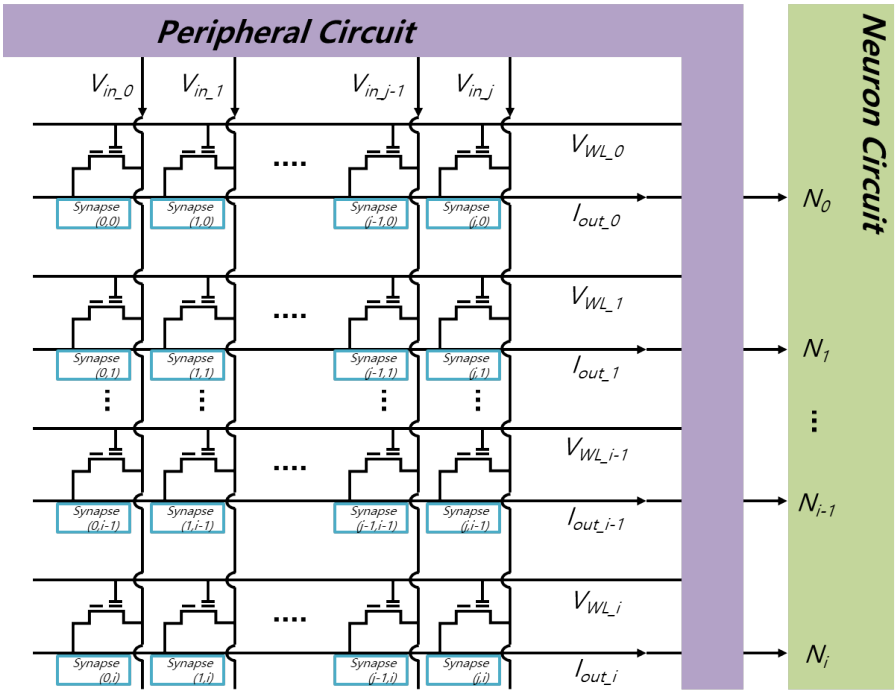


Figure 2.16: Synaptic array schematic

the program voltage is applied to the drain voltage, it has the same potential as the voltage applied to gate2. In this case, as can be seen from the band diagram, the slope of the band becomes smooth in S01. Therefore, the program is inhibited. Fig. 2.18 (c) shows that electron trapped charge increases only in the case of S_{00} which is the target synapse as a result of the simulation.

Similarly, Fig.2.19 Indicates the erase case. When the erase voltage is applied to the drain, in the same way, it can be confirmed that the erase is inhibited in the case of S_{10}, which must be inhibited.

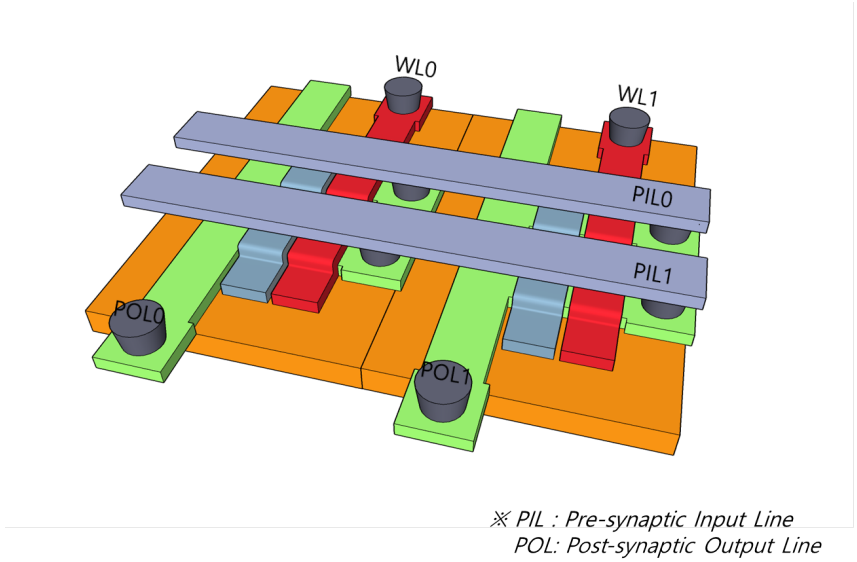
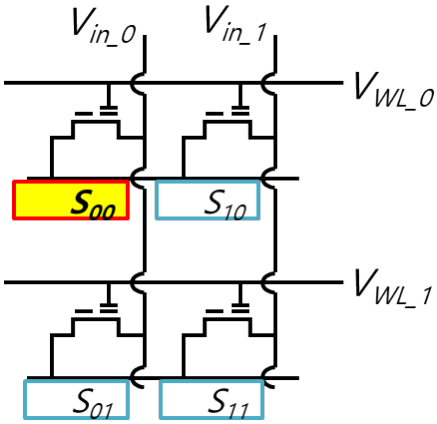


Figure 2.17: 3D array schematic in bird-eye view

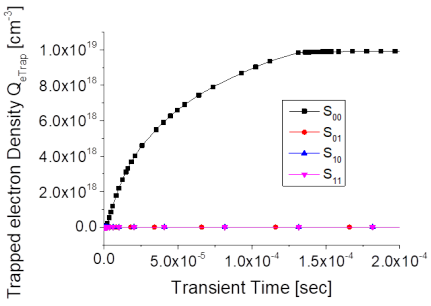


(a)

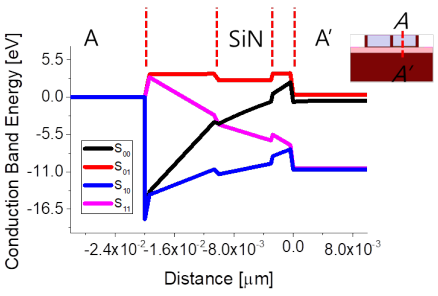
※ Unit : V Target Synapse

	PGM				ERS			
	V_{in_0}	V_{in_1}	V_{WL_0}	V_{WL_1}	V_{in_0}	V_{in_1}	V_{WL_0}	V_{WL_1}
S_{00}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{10}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{01}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{11}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0

(b)

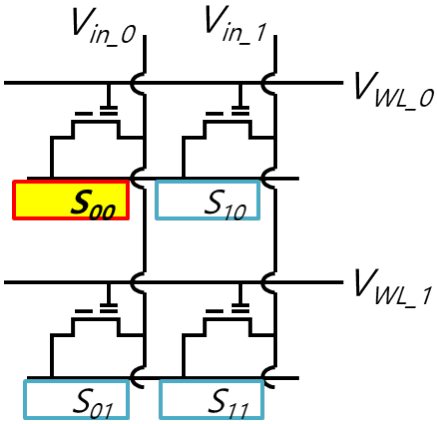


(c)



(d)

Figure 2.18: Inhibition scheme in program operation (a) 2 by 2 array (b) table of operation voltage (c) Trapped electron density for each case (d) Band diagram for each case

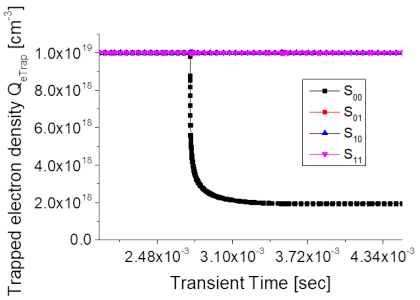


(a)

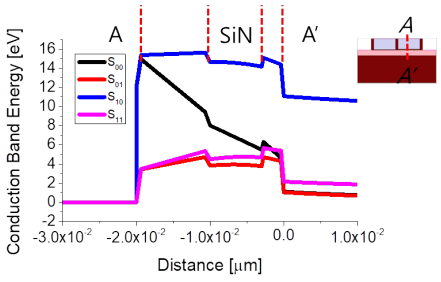
※ Unit : V Target Synapse

	PGM				ERS			
	V_{in_0}	V_{in_1}	V_{WL_0}	V_{WL_1}	V_{in_0}	V_{in_1}	V_{WL_0}	V_{WL_1}
S_{00}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{10}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{01}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0
S_{11}	0	V_{PGM}	V_{PGM}	0	0	V_{ERS}	V_{ERS}	0

(b)



(c)



(d)

Figure 2.19: Inhibition scheme in program operation (a) 2 by 2 array (b) table of operation voltage (c) Trapped electron density for each case (d) Band diagram for each case

Chapter 3

Fabrication and Measurement

In this chapter, process flow of FBFET synapse is explained with the schematic chart and final output device picture. Next, the electrical characteristics of fabricated devices are discussed.

3.1 Fabrication process of FBFET synapse

Actual process experiments are performed for precisely designed features. Each step was matched to the target value using a monitoring sample and confirmed using an optical microscope and an elipso-meter. Fig. 3.1 shows the fabrication process of the proposed FBFET synapse. The entire process was done with CMOS compatible process and only one mask was added for dual-gate structure. Active was defined by patterning on a silicon-on-insulator (SOI) wafer, and implantation was performed with BF^{2+} at a dose of $1 \times 10^{13}\ cm^{-2}$. $1 \times 10^{13}\ cm^{-2}$ doping is the concentration of body1 doping, and the channel doping value of NMOSFET is expected to have a threshold voltage close to 0cF. 1st Gate oxide was subjected to dry oxidation $950°$C with a 10 nm target. After 1st Gate oxide formation, n^+ doped polysilicon $1000\mathring{A}$ was deposited to form 1st Gate and 1st Gate was defined through patterning. Body2 was subjected to n-type As^+ doping at a dose of $2 \times 10^{13}\ cm^{-2}$. The corresponding concentration was determined through simulation and a relatively high doping concentration was used to ensure sufficient well depth. Buffer oxide was removed by HF wet etching and dry oxidation was performed at $800°$C for 30 sec for 3 nm tunneling oxide. After that, 8 nm of Si3N4 as a storage layer was deposited using LPCVD, and then 10 nm of blocking oxide was deposited using MTO. Finally, for the formation of 2nd Gate, It was completed through patterning after deposition of n+ doped poly 1000

A. 3×10^{15} cm^{-2} high doping of As^{+} source and BF^{2+} drain was performed by ion implantation. Finally, a contact was formed through a back-end-of-line (BEOL) process. Fig. 3.1 Shows the entire process flow as a 3D process simulation process. 3D simulation was performed before the process to improve process accuracy. You can check the pad of each source / drain / 1st Gate / 2nd Gate with shows SEM image. Fig. 3.3 The show s TEM image of the fabricated double gate FBFET synapse transistor. It can be confirmed that the profile is correct compared to the target.

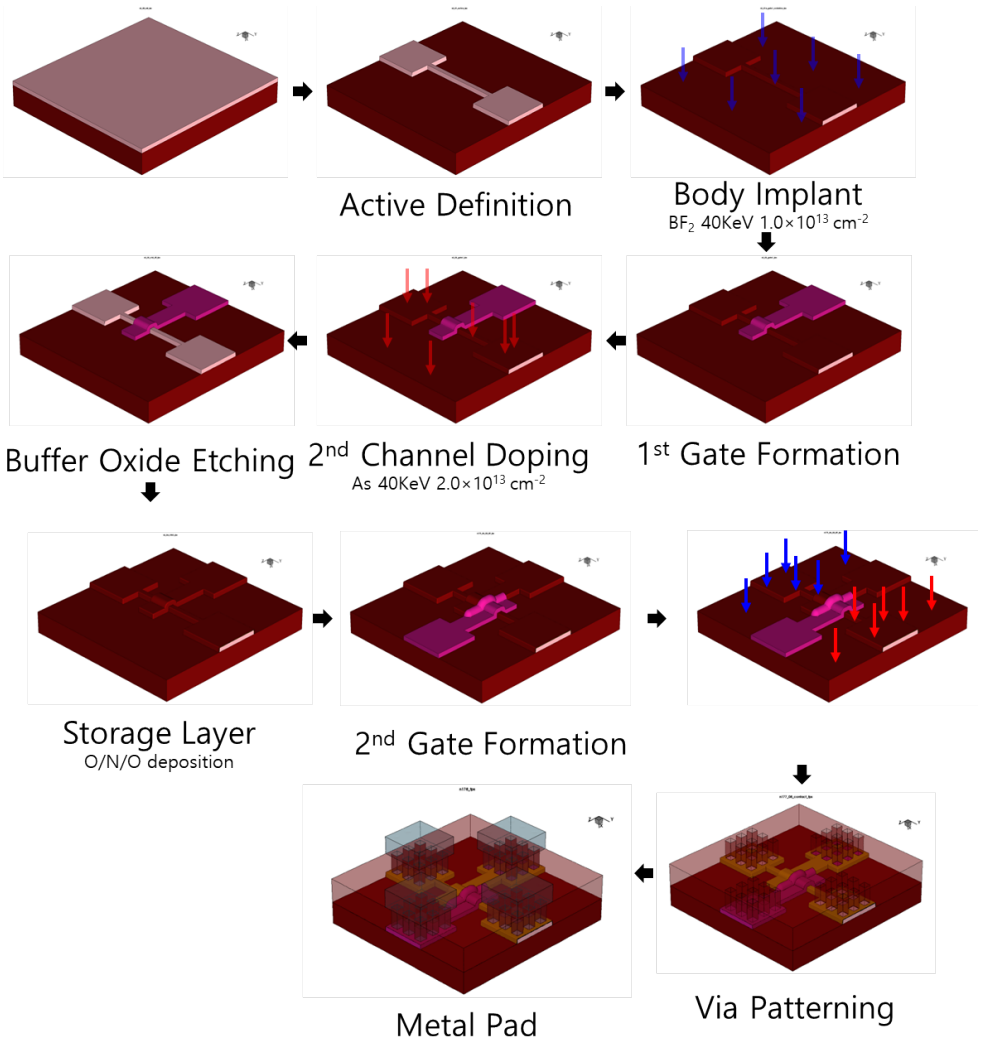


Figure 3.1: Bird-eye view of process flow for fabrication by 3D process simulation tool

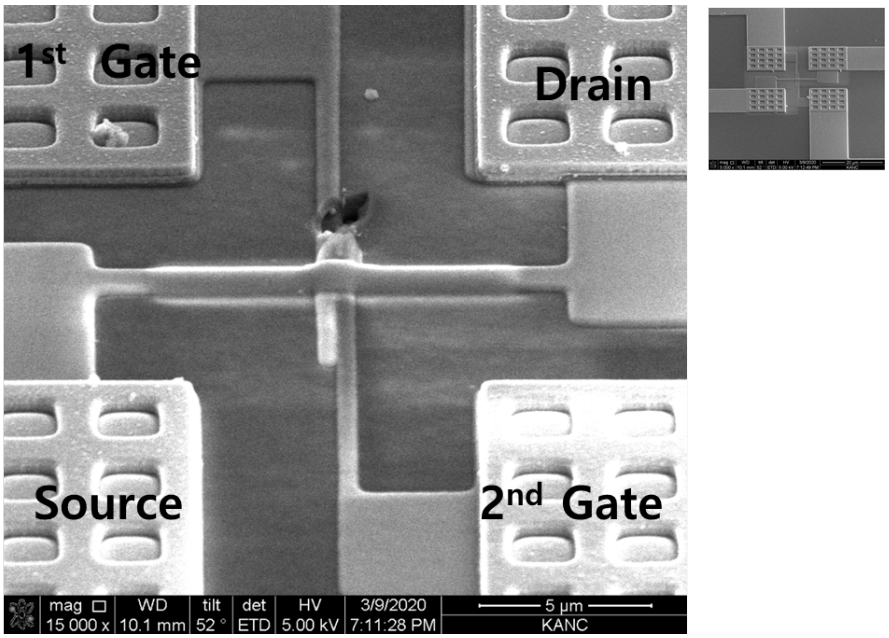


Figure 3.2: SEM image in bird-eye view

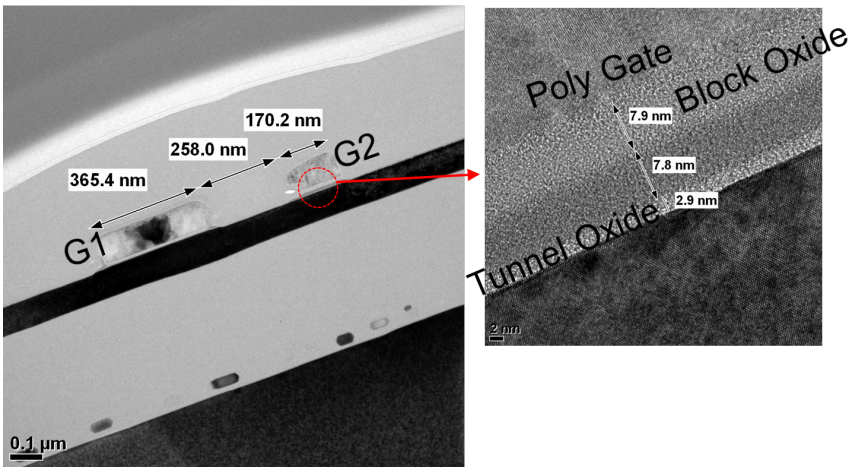


Figure 3.3: Cross sectional TEM image of FBFET synapse

3.2 Measurement result

Fig. 3.4 shows the measurement result of drain current vs. 1st Gate voltage curves according to the 2nd Gate voltage. As mentioned in introduction part, FBFET has a extreme steep switching characteristic with low leakage current. The leakage current is around $1.0 \times 10^{-13} \mu$A. and on current is around $1.0 \times 10^{-7} \mu$A. 6 ordered On/Off ratio is measured with a sub-threshold slope (SS) of less than 0.55 mV/dec. One of biggest issue in semiconductor device is a limitation to reduce threshold voltage. This device can have close to 0 V threshold voltage based this measurement data.

To verify V_T modulation by 2nd body potential barrier difference, varies 2nd Gate applied voltage from 1V to 3V. This DC measurement result can be confirmed this fabricated device is working in the low voltage operation environment.

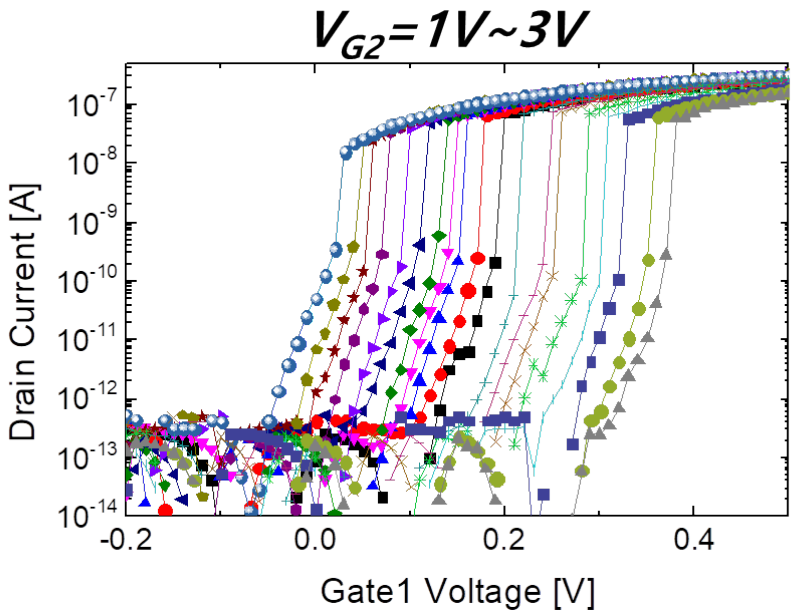


Figure 3.4: Low voltage operation measurement of FBFET Synapse (V_{G1} Operation)

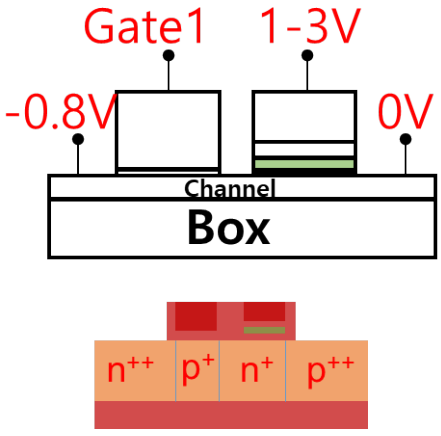


Figure 3.5: Condition for gate1 control measurement

Fig. 3.6 shows the measurement scheme of PGM transient. 1st read measurement performs and 12 V 100μsec pulse was applied for each PGM. To remove hysteresis effect on this measurement, 500μsec delay is placed after PGM operation. Based on measurement, over 500μsec delay time is needed to avoid hysteresis distortion.

Fig. 3.7 is the overlapped transient measurement data by synchronizing input pulse time. Initial delay time is largest in the data. But Device on-time delay is reducing by number of PGM pulse. That is, t_{OFF} value is decreasing by increasing trapped electron amount on the storage layer. At the same time, delay time difference is also decreasing as t_{OFF} close to 0 sec. This is predicted from simulation due to small barrier height between body1 and body2.

Low voltage operation is a strong point for FBFETs device. But we use 2.7 V drain input signal. Then This device has μA level operation current. This is a reason from the variation on fabrication. Around 250μ space is measured by TEM image in Fig. 3.3. This space make floating region on read operation environment. Then, It acts as a resistance component. To achieve the feedback loop, over 2.5 V drain voltage is required for this fabricated device.

For the program operation, electrons are moving from body2 region to silicon nitride trap layer by FN-tunneling. Low PGM efficiency can be estimated because body2 region is floating. But even all contact is grounded except for the 2nd Gate, program

operation is working well. This is because relative high n+ doped is applied on body2.

So we can regard that there is enough electron for program in body2 area.

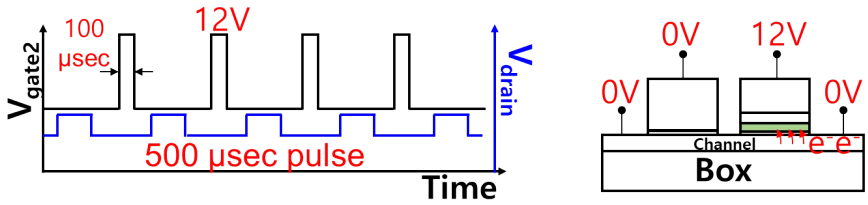


Figure 3.6: Program pulse scheme, $100\mu secprogrampulse$ and $500\mu sec$ pulse width for read. source, 1st Gate, and drain are grounded.

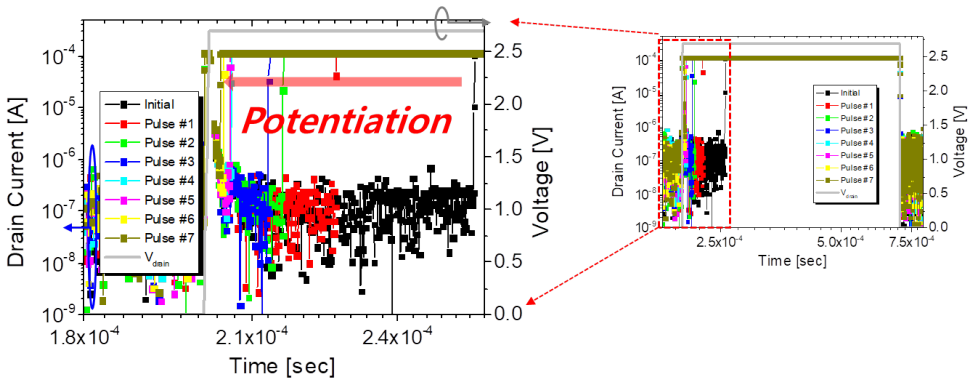


Figure 3.7: Potentiation measurement result

Fig. 3.8 shows the transient measurement result of ERS operation. 15 V and 1 msec pulse was applied for each PGM relatively high bias voltage than program operation. t_{OFF} value is increasing by each ERS pulse. t_{OFF} is larger than nominal reference t_{OFF}. It means that holes are accumulated in storage layer in Fig. 3.9

To secure erase performance, 1 V drain voltage is applied. Low erase efficiency is estimated due to lack of hole on body2, which is n+ doped. So erase efficiency has secured by supplying holes from drain side.

Trapped electron on storage layer can not easily erased. So It can move to body2 by applying high erase voltage -15 V. Then It is difficult to see threshold rolling effect on this situation. It means that delicate erase operation may not be possible.

Non-linear characteristic is observed with several PGM pulse in Fig. 3.10. It is also shown same result as simulation. To avoid the accumulated charge saturation, doping concentration amount adjusting is needed. and still accumulation charge amount is not linear by number of pulse. This problem can be solved by Incremental Step Pulse Program (ISPP) method.

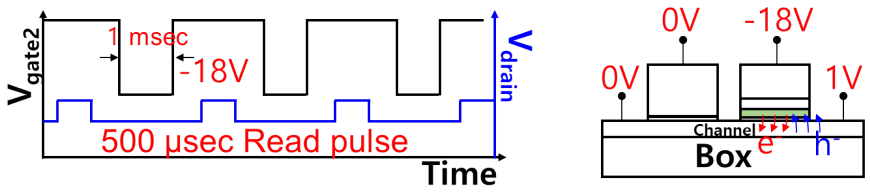


Figure 3.8: Depression measurement condition scheme. To secure high erase efficiency, holes can be supplied from drain side by applying 1 V on drain.

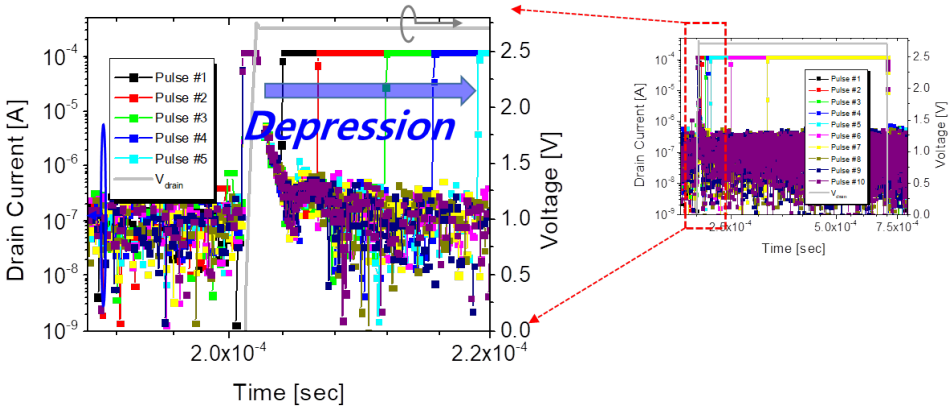


Figure 3.9: Depression measurement result

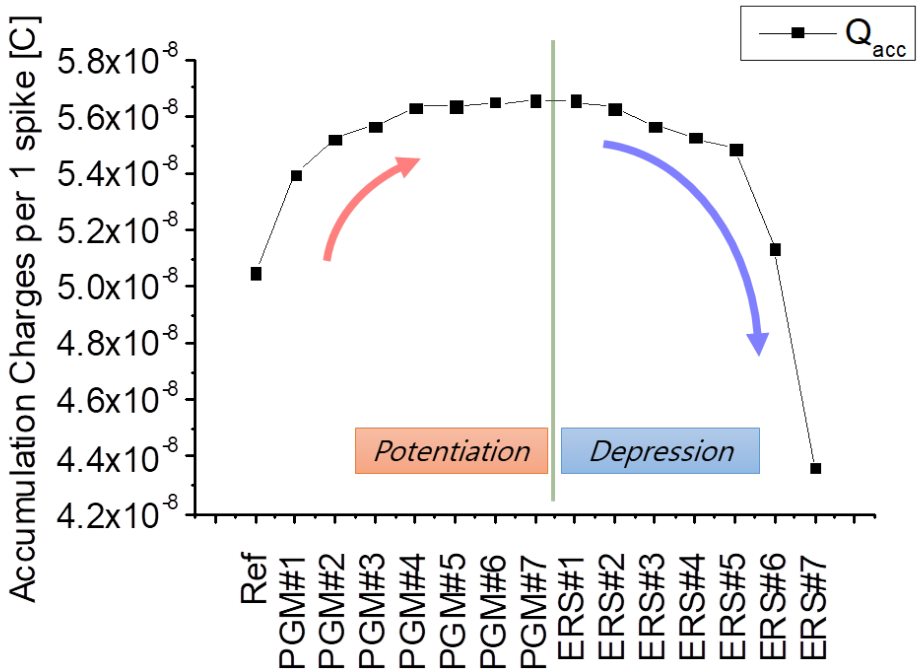


Figure 3.10: Accumulated charge by PGM/ERS pulse

3.3 Hysteresis Reduction

The hysteresis of FBFETs is well known as one of biggest problem. [42] When FBFETs device is on, carrier is occupied on channel region. Even device is turned off, those carriers still remained. It can change potential barrier height until whole carriers are disappeared as each carrier life-time. So threshold voltage V_T is also affected, usually V_T is decreased.

This hysteresis characteristic may be used in applications such as short term memory. However, where FBFETs are used as synapse devices, this hysteresis creates a big problem. Because the pre-synaptic signal is unpredictable, in some cases, a lot of current can flow as expected.

To avoid this problem, It is needed to set the default delay time of the input signal to be greater than the hysteresis time. It is necessary to measure the hysteresis of this synapse device. Fig. 3.11

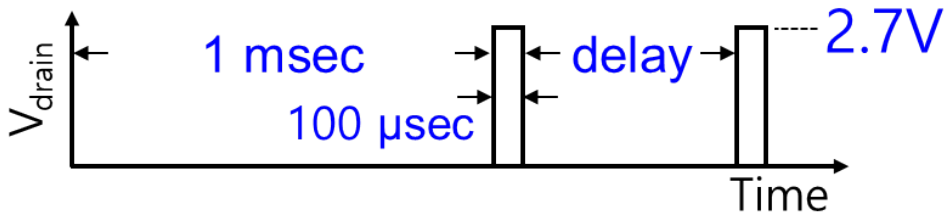


Figure 3.11: hysteresis measurement scheme

First, input read signal of 100μsec pulse width is applied for an equilibrium state with a delay of 1 msec. After that, test input signal is applied according to the delay time split.

Red and green line is the reference output curve. The black line is the data when the delay time is 100μsec. As can be seen from the figure, it can be seen that the device is turned on about 1μsec earlier. It can be confirmed that there is hysteresis with the carrier remaining in the channel. On the other hand, if a delay of 500μsec is given, the result similar to the reference data is displayed.

This result concludes that the pre-synaptic signal must have a delay of about 500μsec to guarantee system accuracy. Basically, it must be an additional burden to the latency problem, which is known as a weakness of SNN. To solve this problem, the following input signal profile is proposed.

The problem is the carriers remaining in the channel. If the carrier is removed at the beginning of the pre-synaptic signal, on-time variation can be removed. So, adding -1 V of 1μsec to the front of the input signal of the synapse will remove the remaining holes toward the drain. Fig. 3.13

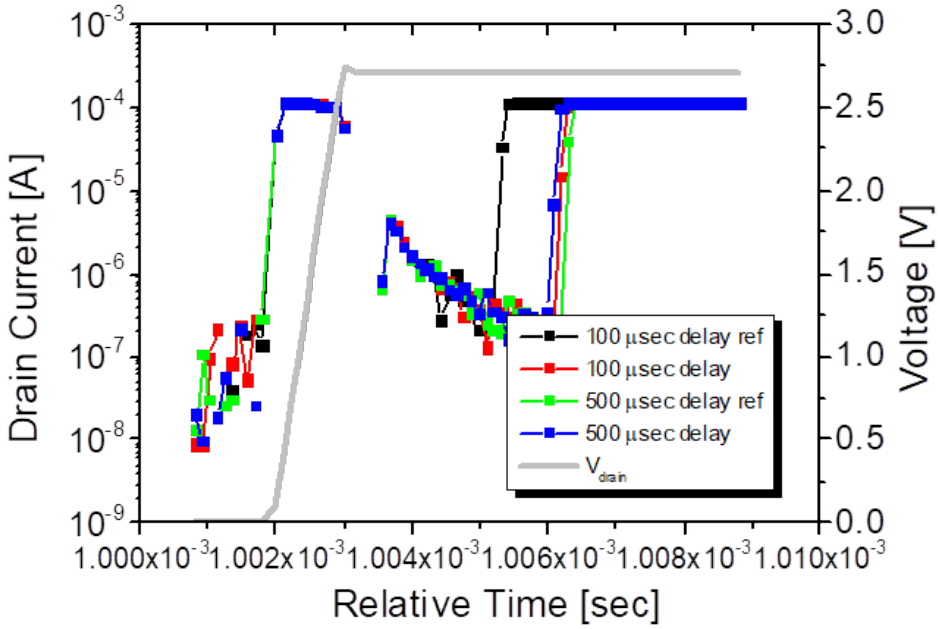


Figure 3.12: Hysteresis measurement result, red and green line is represented for reference lien with enough delay time. black line is for 100μsec delay

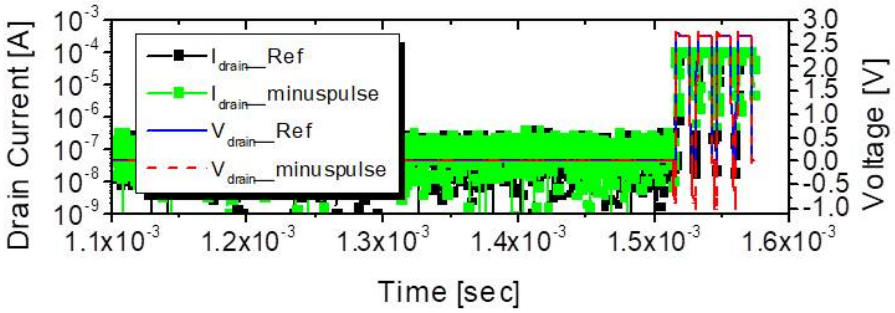


Figure 3.13: Hysteresis reduction measure scheme

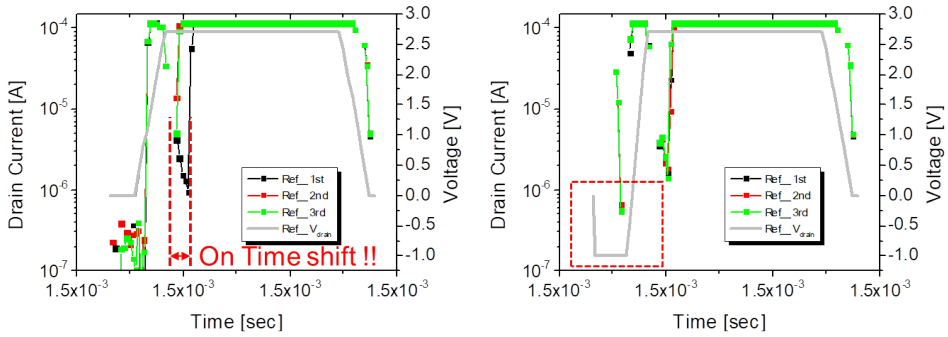


Figure 3.14: comparison of hysteresis reduction, on-time shift is not shown with distorted input signal

Fig. 3.14 is a comparison of data using a modified input signal and existing hysteresis. You can see that the on time shift that actually existed has disappeared.

3.4 Temperature Compensation method

Another FBFET synapse problem is that the current change with temperature is large. Basically, in the proposed FBFET synapse, diode current flows when the device is turned on. Diode current is greatly influenced by temperature. The carrier accumulation rate and the carrier life time are complicatedly affected to device characteristic. Fig. 3.15 Shows that the properties of the synaptic device greatly change according to the same input signal as the temperature changes. It can be seen that not only the on-current is greatly changed, but also the time at which feedback occurs, that is, the on time delay is shifted by 0.1 msec as 100 K degree changes.

To solve this problem, 1st Gate can be used. Basically, 1st Gate can reduce the variation of device by applying a specific voltage bias in the normal case. This is because, if the body1 area is floating, the charge barrier height can be changed in some cases. This 1st Gate bias can eventually shift the global threshold voltage of the devices. When a minus voltage is applied to the 1st Gate, the electron barrier increases, so the threshold voltage becomes large. Using this effect, a temperature compensation method is proposed. Fig. 3.16

Fig. 3.16 is a chart showing the amount of accumulation charge according to temperature and 1st Gate bias voltage. As the temperature increases, the amount of accumulation charge increases as the bias voltage of the 1st Gate decreases. Assuming that

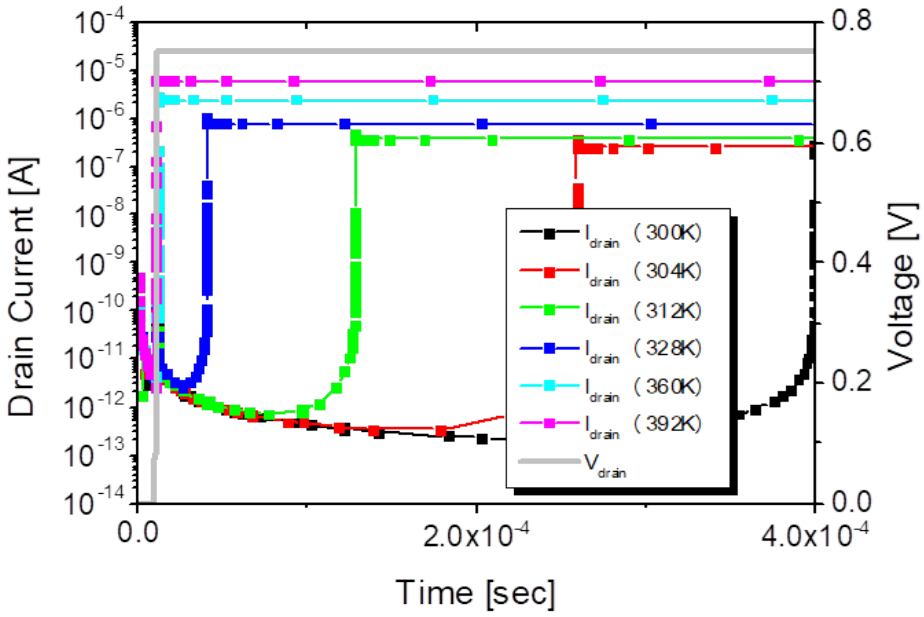


Figure 3.15: FBFET synapse characteristics by temperature.

the amount designed through the initial inference was 300 K, in the situation of 310 K, the 1st Gate voltage can be lowered as shown in the figure to ensure accuracy.

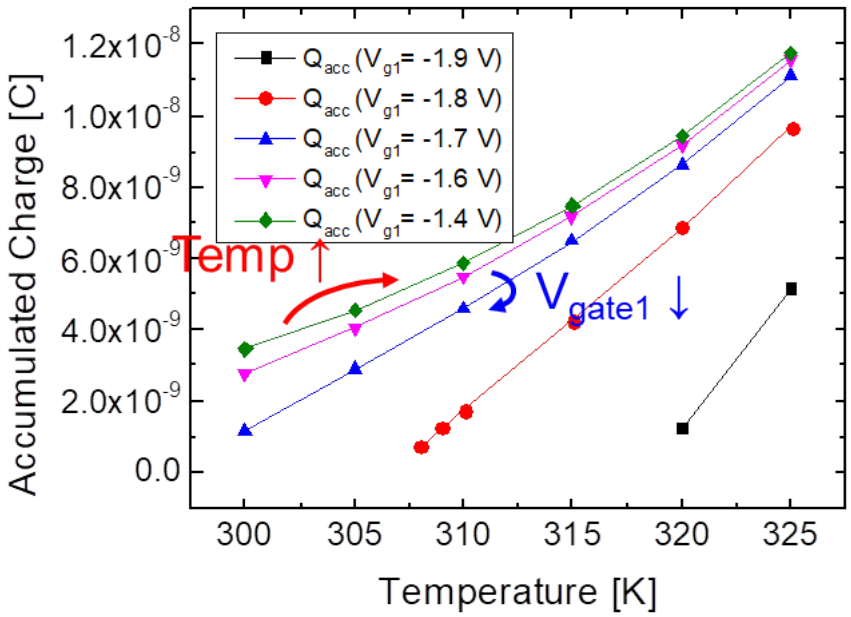


Figure 3.16: Temperature compensation method by applying bias voltage in 1st Gate based on chip temperature measuring

Chapter 4

Modeling and High level simulation

In order to perform high level simulation using fabricated FBFET synapse, it is necessary to develop a model based on measurement data. Although simulation can be performed using Python or MATLAB, high level SPICE simulation was performed to check more realistic system characteristics.

4.1 Compact modeling for SPICE

To implement FBFET synapse with SPICE, conventional N-type MOSFET (NMOS), capacitor, and current source with switch were used. As shown in the Fig. 4.1 , when the V_{Drain} input comes in, the potential of C_{body1} increases, and when the potential reaches a certain threshold voltage value, the SW1 switch turns on and the calculated diode current flows.

The threshold voltage of the FBFET synapse, that is, the amount of trapped charge, is determined by the size of the capacitor. As the capacitor size decreases, the threshold voltage also decreases, corresponding to a program phenomenon in which charge is trapped.

This is a comparison of the model and actual measurement data. Fig. 4.2 This model is the result of confirming the characteristic of synapse according to weight through SPICE simulation. in Fig. 4.3

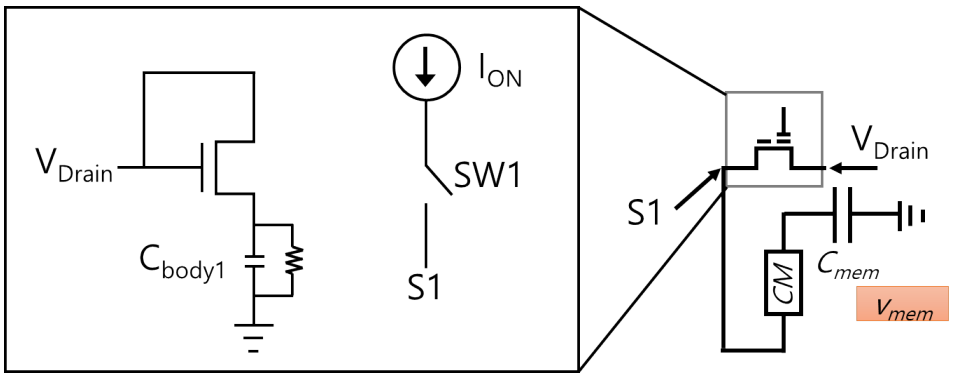


Figure 4.1: FBFET synapse model for SPICE.

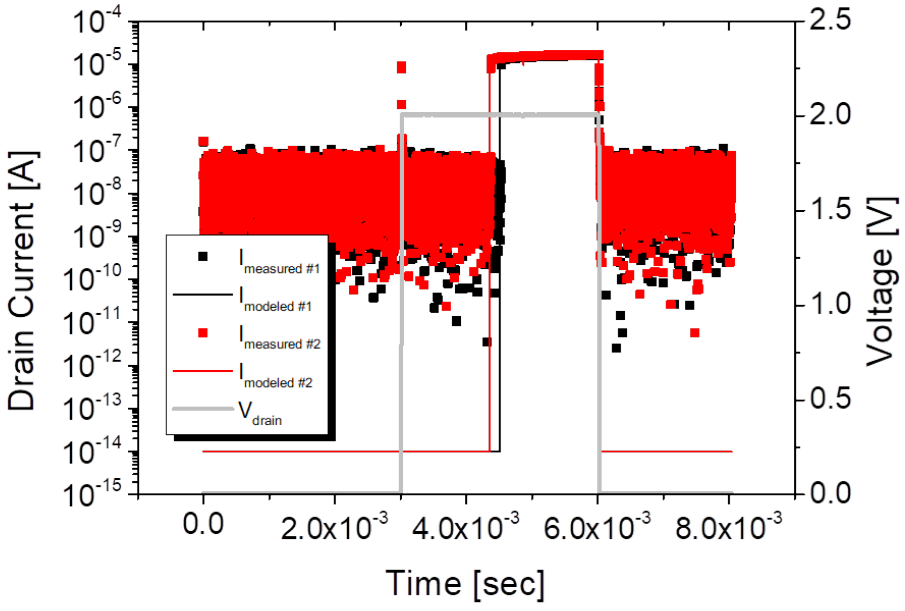


Figure 4.2: Comparison between SPICE model result and measured data

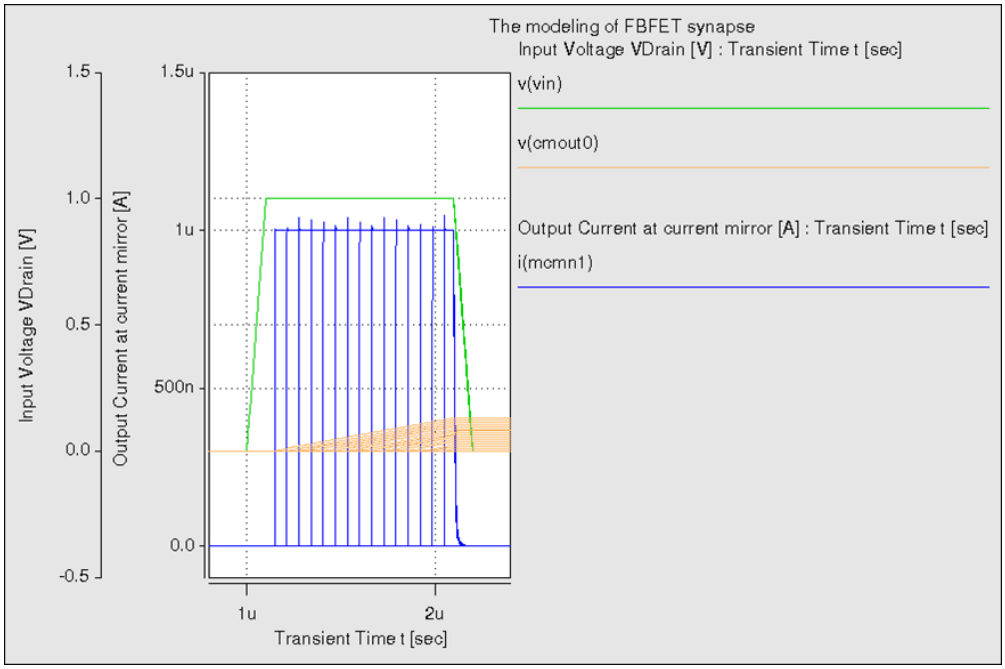


Figure 4.3: Membrane voltage simulation by SPICE

4.2 SPICE simulation for VMM

Based on FBFET synapse device, we perform the VMM simulation by HSPICE. The specific sampling methods which are used are disclosed in Table 4.4. The weighted sum value is defined as the dot product of input and weight value which are generated from 0 to 1 randomly. The input signal is rate coded with max 100 pulses. 19 steps are used for assigning weight value by adjusting the capacitance of modeled synapse each. As shown in Fig. 4.5 . the R^2 value is 0.98.in Fig. 4.7 Which correlation loss may come from a FBFET synapse model issue. When the current source is on by the switch simultaneously, impulse affects to generate an overshoot.

Samples	200 samples for each weighed sum value (0-1,1-2,2-3,3-4,4-5,5-6,6-7) Weighted Sum=$\sum \frac{InputValue_j}{100} \cdot \frac{w_j}{16}$
Input Values	Rate coded 1V pulse (0-100), max 100 pulses for 1 ms Each pulse has 1 μsec pulse width
Synapse Weight	16 steps capacitor value in modeled synapse $C_{synapse} = (1.0e-6 + w \times 1.0e-7)F, w = 0 - 2^4$

Figure 4.4: Environments of VMM simulation for HSPICE

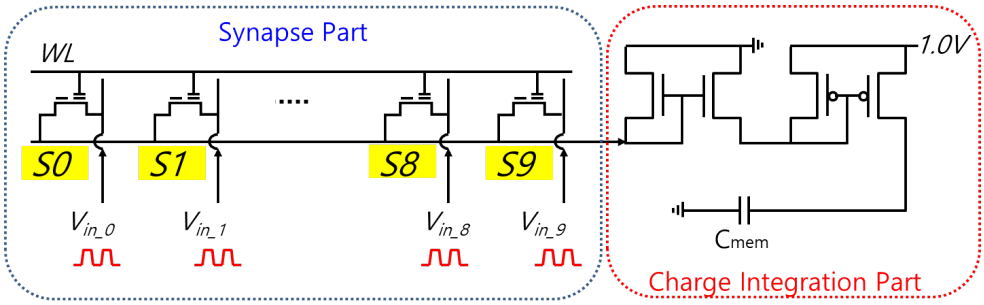


Figure 4.5: Synapse array circuit for VMM simulation

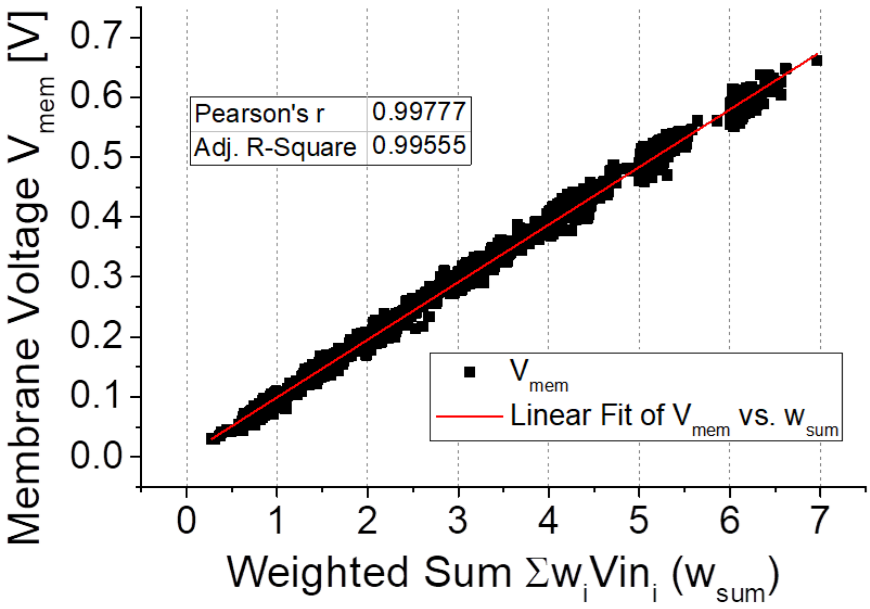


Figure 4.6: Correlation chart of membrane voltage and weighted sum value. 100% correlation is ideal. But some loss from SPICE model.

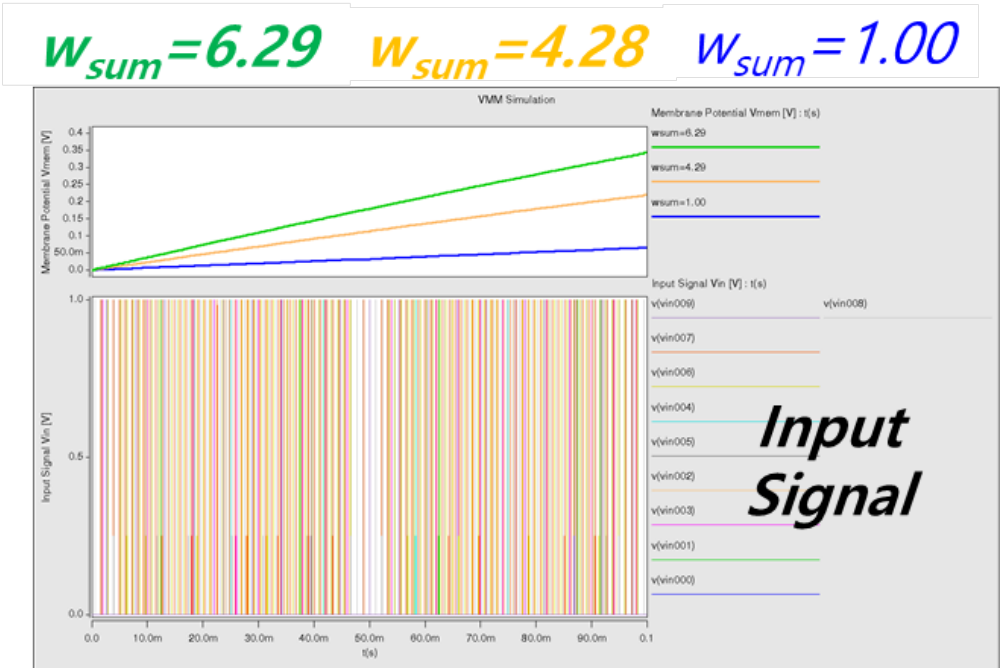


Figure 4.7: Transient membrane voltage output sample for 3 cases.

Chapter 5

Conclusion

5.1 Review of Overall Work

In this dissertation, a synaptic device using FBFET was proposed. The possibility of low power synaptic device was presented through the device design that enables the high retention using FN-tunneling and low voltage operation characteristic of FBFET . In addition, we confirmed the proposed operation scheme by electrical measurement based on fabricated device characteristics. Finally, defined circuit model which created by fitting with measured device data, VMM simulation was verified on HSPICE simulation tool.

5.2 Future work

As a simulation result, a low voltage feedback operation using a drain input signal was possible. However, as a result of actual measurement, drain voltage bias of 2.5V or more was required. This is due to the gap between the 1st Gate and the 2nd Gate, as mentioned earlier. With the proposed fabrication process, the problem inevitably arises because the overlay between two Gates cannot be perfect. In order to solve this, it is necessary to introduce a self-align process. Additional self-aligned process flow studies are needed. And we only verified VMM with high level simulation. Inference verification using MNIST is also necessary. It can give more confidence level for synaptic device.

Bibliography

[1] J. von Neumann, “First draft of a report on the EDVAC,” *IEEE Annals of the History of Computing*, vol. 15, no. 4, pp. 27–75, 1993, doi: 10.1109/85.238389.

[2] A. Huang, “Architectural considerations involved in the design of an optical digital computer,” *Proceedings of the IEEE*, vol. 72, no. 7, pp. 780–786, Jul. 1984, doi: 10.1109/PROC.1984.12938.

[3] J. Backus, “Can programming be liberated from the von Neumann style a functional style and its algebra of programs,” *Commun. ACM*, vol. 21, no. 8, pp. 613–641, Aug. 1978, doi: 10.1145/359576.359579.

[4] G. M. Amdahl, “Validity of the single processor approach to achieving large scale computing capabilities”, *Proceedings of spring joint computer conference.*, Apr. 1967, https://dl.acm.org/doi/10.1145/1465482.1465560.

[5] E. R. Kandel, “The molecular biology of memory storage: a dialogue between genes and synapses”, *Science*, Vol. 294, Issue 5544, pp. 1030-1038, 2001

[6] F.-Y. Wang et al., “Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond,” *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 2, pp. 113–120, Apr. 2016, doi: 10.1109/JAS.2016.7471613.

[7] M. J. Puckelwartz et al., “Supercomputing for the parallelization of whole genome analysis,” *Bioinformatics*, vol. 30, no. 11, pp. 1508–1513, Jun. 2014, doi: 10.1093/bioinformatics/btu071.

[8] K. Riesen and H. Bunke, “IAM Graph Database Repository for Graph Based Pattern Recognition and Machine Learning,” *in Structural, Syntactic, and Statistical Pattern Recognition*, Berlin, Heidelberg, 2008, pp. 287–297, doi: 10.1007/978-3-540-89689-033.

[9] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back, “Face recognition: a convolutional neural-network approach,” *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, Jan. 1997, doi: 10.1109/72.554195.

[10] K. Funahashi and Y. Nakamura, “Approximation of dynamical systems by continuous time recurrent neural networks,” *Neural Networks*, vol. 6, no. 6, pp. 801–806, Jan. 1993, doi: 10.1016/S0893-6080(05)80125-X.

[11] L. Duan and Z. Guo, “New results on periodic dynamics of memristor-based recurrent neural networks with time-varying delays,” *Neurocomputing*, vol. 218, pp. 259–263, Dec. 2016, doi: 10.1016/j.neucom.2016.08.049.

[12] R. Pascanu, T. Mikolov, and Y.Bengio, ”On the difficulty of training recurrent neural networks.” *In: International Conference on Machine Learning*, pp 1310-1318, 2013.

[13] A. Krizhevsky, I. Sutskever and G. E. Hinton, ”Imagenet classification with deep convolutional neurl networks.” *In: Advances in neural information processing systems*. pp. 1097-1105, 2012.

[14] V. Douence, A. Laflaquière, S. Le Masson, T. Bal, and G. Le Masson, “Analog electronic system for simulating biological neurons,” in *Engineering Applications of Bio-Inspired Artificial Neural Networks*, Berlin, Heidelberg, 1999, pp. 188–197, doi: 10.1007/BFb0100485.

[15] G. Indiveri, E. Chicca, and R. Douglas, “A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity,” *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, Jan. 2006, doi: 10.1109/TNN.2005.860850.

[16] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Netw*, vol. 111, pp. 47–63, Mar. 2019, doi: 10.1016/j.neunet.2018.12.002.

[17] S. Mitra, S. Fusi, and G. Indiveri, "A VLSI spike-driven dynamic synapse which learns only when necessary," *in 2006 IEEE International Symposium on Circuits and Systems*, May 2006, p. 4 pp.-, doi: 10.1109/ISCAS.2006.1693200.

[18] S. Park and S. Yoon, "Analysis of the Training Performance and Time of SNN by STDP Algorithms and Spike Temporal Interactions," *KIISE Transactions on Computing Practices*, vol. 24, pp. 482–486, Sep. 2018, doi: 10.5626/KTCP.2018.24.9.482.

[19] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Netw*, vol. 111, pp. 47–63, Mar. 2019, doi: 10.1016/j.neunet.2018.12.002.

[20] J. Wu, Y. Chua, M. Zhang, Q. Yang, G. Li, and H. Li, "Deep Spiking Neural Network with Spike Count based Learning Rule," *in 2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, pp. 1–6, doi: 10.1109/IJCNN.2019.8852380.

[21] P. U. Diehl, D. Neil, J. Binas, M. Cook, S.-C. Liu, and M. Pfeiffer, “Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing,” *in 2015 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2015, pp. 1–8, doi: 10.1109/IJCNN.2015.7280696.

[22] E. Covi et al., “Spike-driven threshold-based learning with memristive synapses and neuromorphic silicon neurons,” *J. Phys. D: Appl. Phys.*, vol. 51, no. 34, p. 344003, Jul. 2018, doi: 10.1088/1361-6463/aad361.

[23] S. Hwang et al., “System-Level Simulation of Hardware Spiking Neural Network Based on Synaptic Transistors and I F Neuron Circuits,” *IEEE Electron Device Letters*, vol. 39, no. 9, pp. 1441–1444, Sep. 2018, doi: 10.1109/LED.2018.2853635.

[24] Hoi-Jun Yoo, ”Deep Learning Processors: Turning Challenges into Opportunities”, *Short Course in 2019 International Symposium on VLSI Technology and Circuits*, 2019

[25] S. Kim et al., ”A Novel Vector-matrix Multiplication (VMM) Architecture based on NAND Memory Array”, em JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE, VOL.20, NO.3, JUNE, 2020

[26] I.-T. Wang, Y.-C. Lin, Y.-F. Wang, C.-W. Hsu, and T.-H. Hou, “3D synaptic architecture with ultralow sub-10 fJ energy per spike for neuromorphic computation,” *in 2014 IEEE International Electron Devices Meeting*, Dec. 2014, p. 28.5.1-28.5.4, doi: 10.1109/IEDM.2014.7047127.

[27] Hyungjin Kim, Jung Han Lee, Garam Kim, Min-Chul Sun, and Byung-Gook Park, ”Silicon-based floating-body synaptic transistor,” *International Conference on Solid State Devices and Materials*, pp. 322-323, Sep. 2012

[28] Myung-Hyun Baek, Taejin Jang, Min-Woo Kwon, Sungmin Hwang, Suhyeon Kim, and Byung-Gook Park, ”Polysilicon-Based Synaptic Transistor and Array Structure for Short/Long-Term Memory,” *Journal of Nanoscience and Nanotechnology*, Vol. 19, no. 10, pp. 6066-6069, Oct. 2019

[29] M. Jerry et al., “Ferroelectric FET analog synapse for acceleration of deep neural network training,” *in 2017 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, p. 6.2.1-6.2.4, doi: 10.1109/IEDM.2017.8268338.

[30] G. W. Burr et al., “Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element,” *IEEE Transactions on Electron Devices*, vol. 62, no. 11, pp. 3498–3507, Nov. 2015, doi: 10.1109/TED.2015.2439635.

[31] A. Vincent, J. Larroque, N. Locatelli, N. Ben Romdhane, O. Bichler, C. Gamrat, et al., "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems", *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 2, pp. 166-174, Apr. 2015.

[32] J. Joshi et al., "A biomimetic fabricated carbon nanotube synapse for prosthetic applications," *in 2011 IEEE/NIH Life Science Systems and Applications Workshop (LiSSA)*, Apr. 2011, pp. 139–142, doi: 10.1109/LISSA.2011.5754178.

[33] J. X. Chen, "The Evolution of Computing: AlphaGo,", *Computing in Science Engineering*, vol. 18, no. 4, pp. 4–7, Jul. 2016, doi: 10.1109/MCSE.2016.74.

[34] K.-B. Choi et al., "A Split-Gate Positive Feedback Device With an Integrate-and-Fire Capability for a High-Density Low-Power Neuron Circuit," *Front. Neurosci.*, vol. 12, 2018, doi: 10.3389/fnins.2018.00704.

[35] S.-M. Joe, H.-J. Kang, N. Choi, M. Kang, B.-G. Park, and J.-H. Lee, "Diode-Type NAND Flash Memory Cell String Having Super-Steep Switching Slope Based on Positive Feedback," *IEEE Transactions on Electron Devices*, vol. 63, no. 4, pp. 1533–1538, Apr. 2016, doi: 10.1109/TED.2016.2533019.

[36] J. Wan, C. Le Royer, A. Zaslavsky, and S. Cristoloveanu, "Progress in Z2-FET 1T-DRAM: Retention time, writing modes, selective array operation, and

dual bit storage," *Solid-State Electronics*, vol. 84, pp. 147–154, Jun. 2013, doi: 10.1016/j.sse.2013.02.010.

[37] H. El Dirani et al., "Sharp-switching band-modulation back-gated devices in advanced FDSOI technology," *Solid-State Electronics*, vol. 128, pp. 180–186, Feb. 2017, doi: 10.1016/j.sse.2016.10.008.

[38] A. Padilla, Chun Wing Yeung, C. Shin, C. Hu, and Tsu-Jae King Liu, "Feedback FET: A novel transistor exhibiting steep switching behavior at low bias voltages," *in 2008 IEEE International Electron Devices Meeting*, Dec. 2008, pp. 1–4, doi: 10.1109/IEDM.2008.4796643.

[39] S.-M. Joe, H.-J. Kang, N. Choi, M. Kang, B.-G. Park, and J.-H. Lee, "Diode-Type NAND Flash Memory Cell String Having Super-Steep Switching Slope Based on Positive Feedback," *IEEE Transactions on Electron Devices*, vol. 63, no. 4, pp. 1533–1538, Apr. 2016, doi: 10.1109/TED.2016.2533019.

[40] M. Kim, Y. Kim, D. Lim, S. Woo, K. Cho, and S. Kim, "Steep switching characteristics of single-gated feedback field-effect transistors," *Nanotechnology*, vol. 28, no. 5, p. 055205, Dec. 2016, doi: 10.1088/1361-6528/28/5/055205.

[41] C. Lee, J. Sung, and C. Shin, "Understanding of Feedback Field-Effect Transistor and Its Applications," *Applied Sciences*, vol. 10, no. 9, Art. no. 9, Jan. 2020, doi:

10.3390/app10093070.

[42] K. Lee, S. Woo, J. Cho, H. Kang, and S. Kim, “게이트 절연막과 게이트 전극 물질의 변화에 따른 피드백 전계효과 트랜지스터의 히스테리시스 특성 확인,” em JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE, vol. 22, no. 2, pp. 488–490, Jun. 2018.

초 록

신경망 모방 시스템은 폰 노이만 구조의 계산 시스템이 가지는 약점인 복잡한 인식 문제를 해결과 에너지 소비의 효율성의 가능성으로 수년간 많은 분야에서 연구되고 있고 일부는 상용화 단계에까지 이르렀다. 이 신경 모방 시스템은 시냅스 모방 소자와 뉴런 회로로 이루어 지는데 시냅스 모방 소자는 신호전달과 기억 기능을 담당하고 있다.

시냅스는 전체 신경모방 시스템에서 가장 큰 부분을 차지 한다. 따라서 시스템내 대부분의 전력 소비가 시냅스 부분에서 일어나게 되므로 저전력 구현이 필수적인 요소다. 이런 이유로 저전력 소자에 특화된 소자인 터널 전계 효과 트랜지스터 (TFET), 네거티브 커페시터 전계효과 트랜지스터 (NCFET), 강유전계 효과 트랜지스터 (Fe-FET) 및 피드백 전계 효과 트랜지스터 (FBFET) 등이 연구되고 있다.

이런 다양한 소자중에 현재의 상보형 금속-산화물-반도체 (CMOS) 공정을 그대로 사용할 수 있는 피드백 전계 효과 트랜지스터는 뉴런 회로와 동시에 제작이 필요한

신경망 모방 시스템에서 대량 생산 가능성에 있어서 매우 유리하다.

본 논문에서는 이 피드백 전계 효과 트랜지스터를 기반으로 하고 NAND 플래시 메모리 구조에서 사용하는 파울러 노르다임 터널링(Fowler-Nordheim tunneling)을 방식으로 차치 트랩 층에 시냅스 소자의 가중치를 기억하는 방식의 시냅스 장치를 제안하고 있다.

해당 소자의 저전력 특성과 구동 방법을 테크놀로지 컴퓨터 지원 설계 (TCAD) 시뮬레이션을 사용하여 유효성을 확인 하였고, 서울대 반도체 공동 연구소 (ISRC) 의 CMOS 공정을 사용하여 소자를 제작하였고 전기적 특성 측정을 통해 제안된 방법을 확인 및 검증 하였다.

주요어: 인공지능 신경망, 양성 전계효과 트렌지스터, 2중 게이트 구조의 싸이리스터, 저전력 시냅스 소자

학번: 2016-30204