Master's Thesis

Evaluation of Neuromorphic Hardware using Oxide Semiconductor by Simulation and Real Chips

Hiroya Ikeda
Program of Information Science and Engineering
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Prof. Professor Yasuhiko Nakashima
Computing Architecture Lab. (Division of Information Science)

Submitted on March 17, 2020

A Master's Thesis
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Hiroya Ikeda

Thesis Committee:

Professor Yasuhiko Nakashima
(Supervisor, Division of Information Science)
Professor Kazushi Ikeda
(Co-supervisor, Division of Information Science)
Associate Professor Takashi Nakada
(Co-supervisor, Division of Information Science)
Assistant Professor Tran Thi Hong
(Co-supervisor, Division of Information Science)
Assistant Professor Renyuan Zhang
(Co-supervisor, Division of Information Science)

Evaluation of Neuromorphic Hardware using Oxide Semiconductor by Simulation and Real Chips*

Hiroya Ikeda

Abstract

As research on artificial intelligence is advancing, neural networks that model the operation of brain neural circuits are used for computations such as machine learning using big data and combinatorial optimization. Although the performance of machine learning has been improved by deep learning using a multilayer neural network in which many nodes are stacked, there are problems that large amount of resources are required in terms of calculation amount and power consumption. A new technology that solves the above problems is neuromorphic computing, which is implemented by hardware by imitating not only the mechanism of brain firing but also its structure. By implementing neural networks at hardware level, large-scale integration and low power consumption can be achieved, and the above problems can be solved. Further, by using amorphous In-Ga-Zn-O (α-IGZO) of an oxide semiconductor that enables fabrication at room temperature and large-scale integration for a synapse of a network, neural network of similar performance with less computational resources and power consumption can be constructed.

In this paper, in order to evaluate the operation of the neuromorphic hardware, it was simulated using an associative memory algorithm. Moreover, assuming that the deviation of initial resistance of the oxide semiconductor is 20%, in the simulation of 3×3 pixels, it was found that the correction ratio of the associate memory of more than 80% can be obtained when the input pattern has difference

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 17, 2020.

pixel of about 30% compared to the memorized pattern. In the simulation of 6×6 pixels, it was found that the correction ratio of the associate memory of more than 80% can be obtained when the input pattern has difference pixel of about 20%. In the simulation of 9×9 pixels, it was found that the correction ratio of the associate memory of more than 80% can be obtained when the input pattern has difference pixel of about 10%.

In addition, operation verification was performed using the created neuromorphic hardware chips. There are 3 types of deposition structures for the created chips. The first one was that only argon gas was injected during sputtering, and the deviation of the initial resistance of the chip was 27%. It was confirmed that the operation of associate memory of 1 letter was possible in the experiment using the chip. The second one was that argon gas and oxygen were injected during sputtering, and the deviation was improved to 23%. This is because the influence from the external environment could be suppressed. It was confirmed that the operation of associate memory of 1 letter was possible. The last one was that light is irradiated on the deposition area of the chip having the second deposition structure, and the deviation was improved to 21%. This is because the free carriers generated by the light terminated non-uniform defects in the deposition area.It was confirmed that the operation of associate memory of 1 letter and 2 letters was possible.

In conclusion, the operation of the chip was confirmed, which can be created at low temperature and can be said to be a prototype of neuromorphic hardware using highly integrated oxide semiconductor synapses. However, the cellular neural network used in this study is limitedly used for autoassociative memory. In order to make the application versatile, use of networks having tight connection between neurons, such as Hopfield network is considered.

Keywords:

Neuromorphic hardware, Oxide semiconductor, Simulation, Associate memory, Characteristic deviation, Real chip evaluation

酸化物半導体を用いた ニューロモルフィックハードウェアの シミュレーションと実機による評価*

池田　裕哉

内容梗概

人工知能に関する研究が発展する中, ビッグデータを用いた機械学習や組み合わせ最適化などの演算に脳の神経回路の動作をモデル化したニューラルネットワークが活用されている. ノードを多く重ねた多層ニューラルネットワークを用いる深層学習によって, 機械学習の高性能化が進んだが, その一方で計算量や消費電力などで多大なリソースを必要とする場合があり問題となっている. このような問題を解決する新たな技術として, 脳の発火のメカニズムだけではなく構造も模倣することでハードウェア実装するニューロモルフィックコンピューティングが挙げられる. ニューラルネットワークをハードウェアレベルで実装することで大規模集積化・低消費電力化が可能となり, 上記の問題を解決することが可能になる. 室温での作製が可能であり, 大規模集積化も可能となる酸化物半導体であるアモルファス In-Ga-Zn-O (α-IGZO) をネットワークのシナプスに用いることでより少ない計算リソースと消費電力で, 同じパフォーマンスのニューラルネットワークを構成することが可能となる.

本論文では, 実際のデバイスで作製するにあたって, ニューロモルフィックハードウェアの動作を評価するために, 文字連想記憶を用いたアルゴリズムを使って動作シミュレーションを行った. そのシミュレーションにより、ネットワークの隠れ層は 1 層が最適であることを明らかにした。また、酸化物半導体の初期抵抗値の偏差が 20%であると仮定した場合、3×3 ピクセルのシミュレーションでは記憶パターンに比べ入力パターンのおよそ 30%に差異がある場合でも 80%以上もの高い連想成功率が得られることを明らかにした。また、6×6 ピクセルのシミュレー

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和 2 年 3 月 17 日.

ションでは入力パターンのおよそ 20%に差異がある場合でも 80%以上もの高い連想成功率が得られることを明らかにした。また、9×9 ピクセルのシミュレーションでは入力パターンのおよそ 10%に差異がある場合でも 80%以上もの高い連想成功率が得られることを明らかにした。

さらに、作製したニューロモルフィックチップを使用し、実機での動作を試みた。作製したチップの成膜構造の種類は 3 種類ある。1 つ目はスパッタリングの際に、アルゴンガスのみを導入したものであり、その成膜構造で作製したチップの初期抵抗値の偏差は 27%となった。そして、その成膜構造を持つチップを使った実験では 1 文字の連想記憶が可能であることを確かめることが出来た。2 つ目はスパッタリングの際に、アルゴンガスと酸素を投入したものであり、その成膜構造で作製したチップの初期抵抗値の偏差は 23%にまで改善された。これは外部環境からの影響を抑制することが出来たからである。そして、その成膜構造を持つチップを使った実験では 1 文字の連想記憶が可能であることを確かめることが出来た。3 つ目は 2 つ目の成膜構造で作製したチップの成膜部分に光を照射したものであり、その成膜構造で作製したチップの初期抵抗値の偏差は 21%にまで改善された。これは光により発生したフリーキャリアが膜中の不均一な欠陥を終端したためである。そして、その成膜構造を持つチップを使った実験では 1 文字と 2 文字の連想記憶が可能であることを確かめることが出来た。

結論として、低温で作製可能な高集積化された酸化物半導体シナプスを使用したニューロモルフィックハードウェアのプロトタイプと言えるチップの動作が確認出来た。ただし、本研究で用いたセルラニューラルネットワークは自己想起型の連想記憶に限定的に利用される。用途に汎用性を持たせるためにはホップフィールドネットワーク等のニューロン同士の密な結合を持つネットワークの利用が検討される。

キーワード

ニューロモルフィックハードウェア, 酸化物半導体, シミュレーション, 連想記憶, 特性ばらつき, 実機検証

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

In recent years, artificial intelligence has been used in various fields, and is expected to enrich people's lives with further development[1, 2]. Neural networks aiming at imitating the characteristics of the information processing functions of the brain and using them mathematically are typical technologies for realizing artificial intelligence[3, 4, 5, 6]. Their advantages are self-organization, self-learning ability, parallel and distributed computing, and fault tolerance. A neural network has many neurons and synapses. A lot of information can be processed by connecting them. In fact, the human brain is thought to consist of about 10 billion neurons and about 100 trillion synapses[7], and it is possible to process complex information with these many neurons and synapses. In other words, it is necessary to connect more neurons and synapses to improve the performance of the neural network.

However, traditional neural networks are complex software running on high-performance hardware[8, 9], the machine size is very large, and the power consumption is incredibly large for not only the hardware operation itself but also the air conditioner for cooling it. In addition, since it is executed on a conventional Neumann computer in which the entire process stops when only one physical device fails, the advantages such as parallel distributed computing and fault tolerance cannot be obtained. Furthermore, since the computing architecture is not optimized for neural networks, the extra circuits occupy a large area, making it difficult to incorporate them into independent devices such as IoT.

In order to solve such problems, research on LSI (Large Scale Integration) implementation of neural network models is being conducted[10]. It means implementing a program-based neural network on hardware that was software running on high-performance hardware. If the CPU (Central Processing Unit) mounted on the computer performs a heavy load process that outputs the results in real time, the processing performance of the CPU will be reduced. Recently, new products of CPUs and microcomputers with high processing performance have been released, but the processing speed has reached its limit due to the complexity and sophistication of software. On the other hand, the LSI can process

only a specific program, but its processing efficiency is much higher than CPU or microcomputer.

One of the researches on neural network hardware is called neuromorphic computing, which mimics the firing mechanism of biological neurons and the structure of brain neural circuits.

1.2 Purpose

The purpose of this research is to realize a neuromorphic hardware that performs associative memory, which is one of the information processing functions of the brain, by fabricating a neural network with circuits and devices capable of low power consumption and high integration. Therefore, it was proposed that a neural network using oxide semiconductor synapses is used, which can be created at low temperature and which can be highly integrated by three-dimensionalization. In addition, the effect of variation in characteristics of oxide semiconductors on neural networks was evaluated using associative memory algorithms by simulations and actual hardware chips.

1.3 Composition

In this paper, first, the background and purpose of this research were described in Chapter 1. Next, Chapter 2 describes related research. Then, Chapter 3 describes the elemental technologies for realizing the proposed method, and Chapter 4 describes the proposed method. Next, Chapter 5 explains the algorithm of the associative memory simulation method performed, and Chapter 6 evaluates and discusses the results of associative memory simulation. Then, Chapter 7 describes the experimental method using real hardware chips, and Chapter 8 evaluates and discusses the experimental results. Finally, Chapter 9 described the conclusion of this research.

2. Related Research

2.1 Neumann-type Computing

In recent years, with the spread of information terminals, the amount of information data handled by people has been explosively increasing [11]. In addition to the data growth, how to process and use big data including unstructured and missed data is a social issue because it can not be managed [12]. The conventional computers are built on the Neumann-type architecture, which store data and programs to be processed in main memory beforehand [13]. The central processing unit (CPU) calls and executes program instructions and data from the main memory in order. The configurations of the CPU and main memory are physically separate, and they are connected through a transmission path through which each electronic circuit exchanges data. The versatility of the computer can be improved with these configurations by checking the program and results step by step, because the operation instruction and data to be processed are read from the memory at each step, and the operation result is also written one by one. However, because there is a limit to the amount of data that can be transmitted through this transmission path, performance is degraded when data increases so much, and the conventional computers that process data in a fixed procedure can not handle big data very well.

Moreover, progress of semiconductor manufacturing technology have improved computational speed and efficiency of the sequential computing models [14]. However, growth has been slowed owing to the limits of semiconductor miniaturization and rising manufacturing costs. Therefore, there is a growing demand for refinements of the Neumann-type architecture and creation of a non-Neumann-type one [15].

In spite of the disadvantages of the Neumann-type computers, there are many researches, because they are general-purpose computational models that can be applied to a variety of problems. However, the application of artificial intelligence (AI) technology is expanding, and calculations such as machine learning and combinatorial optimization using big data require a large amount of resources in terms of calculation amount and power consumption. Machine learning is supported by a large number of parameters, and calculating these parameters

requires a unit with high computational power. In current machine learning, a graphical processing unit (GPU) known as a processor specialized for image processing and field programmable gate array (FPGA) capable of programming logic circuits suitable for data and processing content to be handled are used as hardware platforms [16, 17]. However, it is itself still a serious problem that a large amount of resources in terms of calculation amount and power consumption are required as abovementioned.

2.2 Non-Neumann-type Computing

Instead of the Neumann-type computing, different-type computing is considered to be effective to solve the abovementioned problems. These are collectively referred to as non-Neumann computing. Some non-Neumann computing has an architecture that essentially does not have a von Neumann bottleneck. For example, it is possible to realize machine learning by providing primitive processing capability to memory devices that were simply used to temporarily store data, and combining them in large quantity, or brain-type computing that mimics a neural network [18]. In these non-Neumann computing, memory and processing units are composed near without physical distance, and there is no need to communicate between the logic unit and memory. Therefore, it is not necessary to exchange instructions and data between the processor and memory, which is a problem of the Neumann-type computer, and it is possible to respond to big data in a timely manner.

Among non-Neumann-type computing, brain-type computing is expected as a new technology that solves problems which cannot be handled by existing computers. Brain-type computing has an architecture that mimics a neural network. The neural networks and neuromorphic computing are attracting attention. A neural network is a computational model that uses artificial neurons simplifying the operation of biological neurons as nodes and changes the connection strength among the nodes by learning [19]. This is a model based on statistical methods that mimics only brain function. A neural network is composed of an input layer, an output layer, and a hidden layer, and a synaptic weight that determines the connection strength between neurons existings between the layers. The input data is weighted and the sum of these values is input to the activation function

to obtain the output of this neuron. It propagates output data from the input layer to the output layer. Neural networks are attracting attention to improve the performance of machine learning because the accuracy of pattern recognition has been greatly improved by deep learning using a multi layered neural network with many nodes. Large amount of input informations are distributed to and stored in the circuit elements, and can be processed simultaneously by interaction between them. The processing time of brain-type computing is equal to the response time of the circuit elements, and the brain-type computing can perform the processing that requires several tens of steps with the CPU of a conventional digital computer in a few steps [20]. Therefore, it can be said that the brain-type computing is an approach that is very promising as an excellent computing except the Neumann-type computing.

2.3 Neuromorphic Computing

In the previous researches, research was conducted to realize virtual neural networks that are effective for pattern recognition and image processing using software [4, 5, 6, 19]. However, software-implemented neural networks have the problem of being large in size and power consumption. Therefore, researches to realize neuromorphic computing built on actual hardware are actively being carried out [21, 22].

Neuromorphic computing is a hardware-implemented approach that mimics the neuron and synapse functions and firing mechanisms of the brain [23, 24]. It is different from a technical approach to improve the calculation performance of product-sum operation using the GPU and FPGA based on the conventional Neumann type computer. Its basic components are neurons and synapses. A neuron performs non-linear processing and is realized by a CMOS digital arithmetic circuit or an analog arithmetic circuit. The synapse is responsible for determining the strength of the connection between adjacent neurons and storing the strength of the connection with analog or digital memory.

The neuromorphic computing has the same advantages as the brains of a living bodies because it mimics the neural network of the brains at the hardware level. First, the machine size is very small: The machine size of Watson, which is famous for the conventional neural networks, is known to be about 10 refrigerators. On

the other hand, neuromorphic computing that mimics the brain can be realized with a chip size as large as that in a personal computer [25]. Next, there is the advantage of low power consumption: Although Watson consumes about 100 kW, neuromorphic computing is expected to operate at only 20 W [26]. Finally, it has the advantage of being robust. The human brain is known to lose 100,000 neurons every day, but it can still maintain the necessary functionality. Given the above advantages, neuromorphic computing certainly seems to have better performance than traditional neural networks. Some neuromorphic computing is well known, such as Qualcomm's Zeroth processor [28], TeraDeep's nn-X [29], IBM's TrueNorth [30], and Kyushu Institute of Technology's integrated brain system [31], but these large scale integrated (LSI) chips are hybrid systems of the Neumann-type computers and neuromorphic computing. For example, only a few neuron elements are prepared and many neuron elements are virtually emulated using time sharing algorithm, discrete values are stored in digital memory and continuous values like analog memory are approximated, etc. Therefore, the advantages of neuromorphic computing can be obtained only in limited ways. The neuromorphic chip "Loihi" reported in the paper [25] is made of digital circuits, with 128 neuromorphic cores arranged in a mesh, 130,000 neurons and 130 million synapses. Moreover, it is equipped with a synaptic load updating function of a single chip. The neuromorphic chip "Neurogrid" reported in the paper [27] has a structure in which digital and analog circuits are mixed. Neurogrid implements all circuits other than the axon axis with analog circuits, which makes it extremely energetically efficient and can simulate millions of neurons in real time using several watts. In addition, since the conventional semiconductor manufacturing technology is used, only a two dimensional structure can be obtained in principle whereas three dimensional structures are used in the brains, many processing elements are prepared, and the connection structure is simple.

As an example of a neuromorphic system, we will explain IBM's TrueNorth [30]. TrueNorth chip is the largest of IBM's chips at least at that time with 5.4 billion transistors, 1 million programmable neurons, and 256 million programmable synapses and runs with as little as the power of 70 mW. However, it is still far from a human brain with 10 billion neurons and 100 to 150 trillion synapses. If implementing with TrueNorth is considered, the 100 trillion synapses in a human

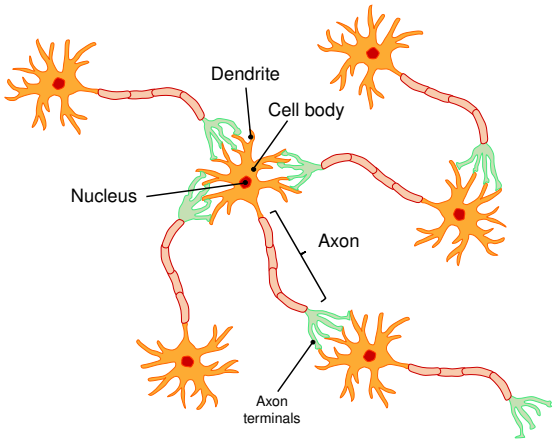


Figure 1. Brain neural network

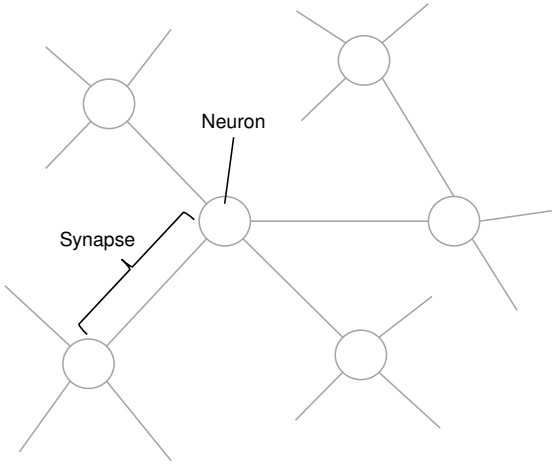


Figure 2. Neuromorphic computing model that mimics brain neural network

brain is about 400,000 times of the 256 million synapses in TrueNorth. That is, in order to reproduce a human brain, the 400,000 TrueNorth chips and power consumption of 70 mW x 400,000 = 28 kW are required. Therefore, although it is possible to achieve it using a huge server, it can be understood that embedding into an independent device is not realistic.

Research on brain-type computing is being carried out in conjunction with the evolution of brain science that analyzes the mechanisms and functions of the brain through the establishment of machine learning methods in recent deep learning, and the progress of computer science including electronics. In this paper, it was proposed the realization of neuromorphic hardware with neuromorphic computing was proposed aiming at imitating human brain neural network.

Our proposed neuromorphic hardware is a complete analog computer made using oxide semiconductor α-IGZO. The synaptic weight value is recorded as an analog value on a single chip, and the synaptic load is updated. Consequently, higher energy efficiency can be expected. In this proposal, synapses can be stacked three-dimensionally by using oxide semiconductors, and it can be expected to develop neuromorphic hardware with high integration.

3. Fundamental Technology

3.1 Neuron Circuit

Fig. 3 shows a configuration of a neuron circuit, where one inverter is composed of two transistors. In this case, the total number of required transistors is only four, and it is suitable for high integration. The neuron circuit of this configuration exhibits characteristics similar to a sigmoid function, and has the simple property of maintaining the state of voltage and firing or changing stably in accordance with a certain threshold.

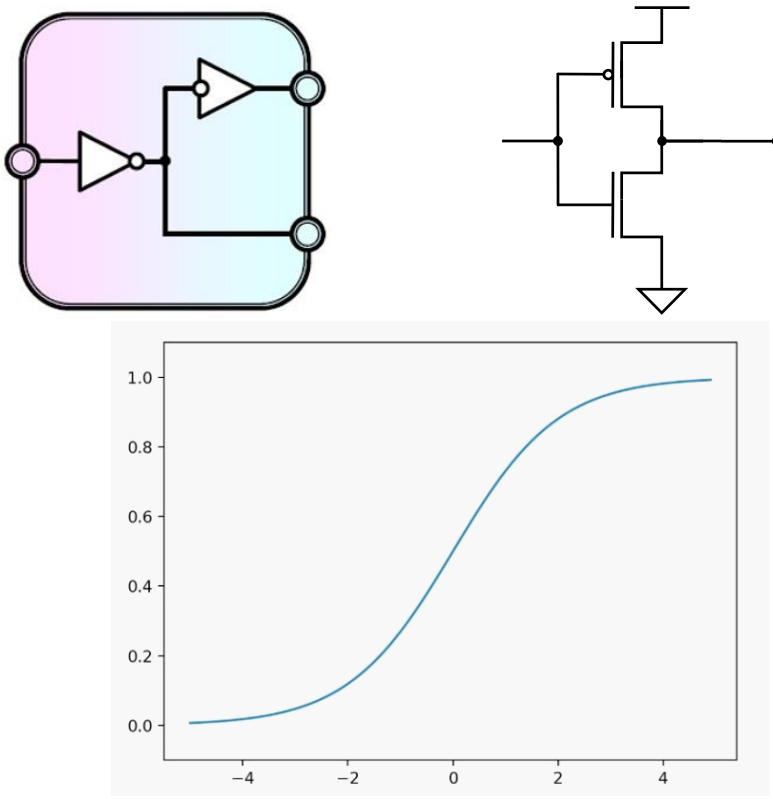


Figure 3. Neuron circuit

3.2 Oxide Semiconductor Device

Oxide semiconductors exhibit excellent performance even when deposited in the amorphous phase at low temperatures [33], and in addition, oxide semiconductors can be potentially formed using printing methods, which facilitate achieving a three-dimensional layer structure for neuromorphic computing. α-IGZO is a typical example among these [34], and recently, it is used as thin film transistors (TFTs) for flat panel display (FPDs) such as high resolution liquid crystal displays (LCDs) and organic light emitting diode (OLEDs) [35]. Fig. 4 shows the α-IGZO device [36]. α-IGZO thin film is deposited on a silicon wafer with metal electrodes using radio frequency (RF) magnetron sputtering, where the sputtering target is IGZO ceramic whose composition is In:Ga:Zn = 1:1:1 and the sputtering gas is Ar, and the thickness of the α-IGZO thin film is 70 nm. These manufacturing conditions are the same as those for the TFTs with the highest performance.

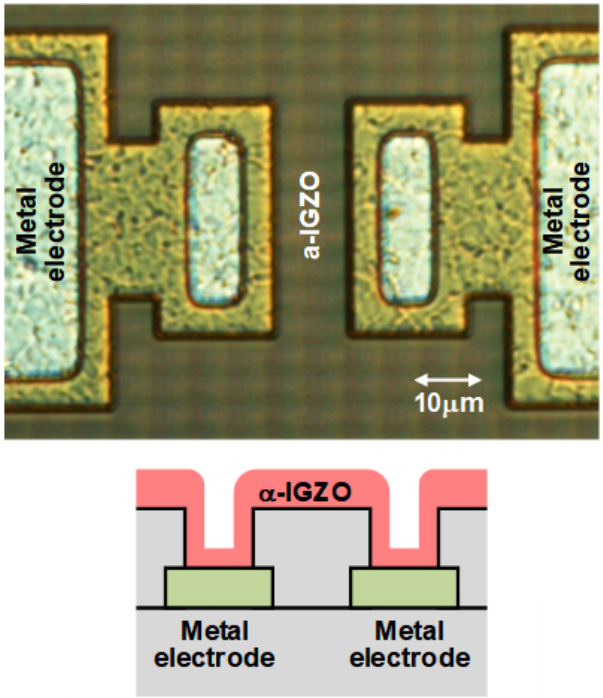


Figure 4. α-IGZO device

Fig. 5 shows a current flowing through the α-IGZO device. Here, 1.8 V was applied to the left and right metal electrodes, and the current flowing through the α-IGZO element was measured. It was found that the conductance decreases continuously when the current flows, and the decrease in conductance is irreversible but relatively slow. This characteristic can be used to make use of the modified Hebbian learning rules described later. The function required for the synapses in the learning phase is that the connection strength properly changes depending on the state of two connected neurons, and maintains. The conductance can be regarded as the connection strength because how easily the signal is transmitted is dependent on it. Moreover, when the two neurons connected to the device are in different states, there is a voltage deference, current flows, deteriorating occurs, and the conductance decrease, which corresponds to that the connection strength decreases. On the other hand, when the two neurons connected to the device are in the same states, there is no voltage deference, no current flows, no deteriorating occurs, and no conductance decrease, which corresponds to that the connection strength maintains. This learning method is different from Hebb's learning rule [37], which raises the connection strength of synapse when the neighboring neurons are both in firing states. In the literature [38], this learning method is called the modified Hebb's learning rule, where it was confirmed that AND, OR, and XOR circuits are remembered based on this learning rule using poly-Si TFTs. We use the oxide semiconductor α-IGZO as synapses, which is a device with a characteristic that the conductance deteriorates when current flows [35].

Here, we describe the size and power consumption when α-IGZO thin films are used as synapses and the number of α-IGZO is equal to the number of synapses in the human brain. When α-IGZO is used in the synaptic element of our proposed neuromorphic hardware, the size of the α-IGZO thin film is roughly 1μm^3. If the number of α-IGZO is equal to the number of synaptic elements in the human brain, 100 trillion α-IGZO synapses are required. If I do not consider the size of the neuron element here, the total size of the neuromorphic hardware is thought to be as small as 100cm^3. This is because the number of neuronal elements is much smaller than the number of synaptic elements as shown in previous studies. In addition, since the power consumption of one α-IGZO thin film is 1pW, the

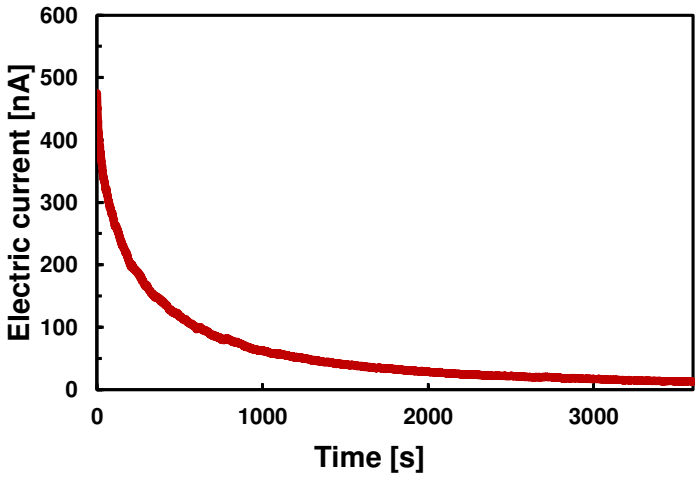


Figure 5. Currrent flowing through the α-IGZO device

power consumption of the entire neuromorphic hardware is thought to be as small as 100W. This power consumption is only about 5 times that of the human brain. Besides, a learning function is naturally included because the connection strength of the network can be changed independently. This is done along the modified Hebbian learning rule described later. The specs of the TrueNorth chip in the previous research was compared with that of the neuromorphic hardware which is proposed under the assumption of the same number of synapses as the human brain. The size is 172 m^2, and our proposed neuromorphic hardware is 100 cm^3. Our proposal does not take into account of neurons size, so it cannot be clearly compared. However, since α-IGZO can be stacked three-dimensionally, it has a high degree of integration. In terms of power consumption, TrueNorth is 28kW, which is 14,000 times the power consumption of humans, whereas our proposal is 100W, which is only five times the power consumption of humans. Finally, our proposed neuromorphic hardware has a learning function because the synaptic connection strength can be changed independently by changing the resistance value of the synapse. This is a major difference in the conventional neuromorphic hardware, which does not have its own learning function. Thus, our neuromorphic hardware has the potential for solving the size and power consumption problems of the conventional neuromorphic hardware and achieving the research objectives.

In this research, neuromorphic hardware using oxide semiconductor synapse

and cellular neural network was developed and evaluated by using logic simulation. In particular, by using the oxide semiconductor synapses of amorphous α-IGZO, which can be manufactured at low temperature and highly integrated in three dimensions, it is possible to construct a larger scale neural network. In order to realize neuromorphic hardware, electronic elements corresponding to neurons and synapses are required. The synapse that connects neuron circuits in neuromorphic hardware is usually composed of resistive elements. The resistance value needs to be rewritten according to learning. Therefore, there is a research report in which a device works as a nonvolatile resistance change memory device such as memristor or ReRAM (Resistive Random Access Memory) that is used for a synapse circuit [39, 40, 41, 42]. The memristor operates as a non-volatile memory in which the resistance value changes stepwise each time the voltage is applied, and each intermediate resistance value is retained even when the voltage is cut off. ReRAM is also referred to as a resistance change type memory, and is a non-volatile memory that stores, as data, the resistance value of an element changed by applying a voltage to a metal oxide sandwiched between electrodes.

Currently, there are several research reports on the development of synaptic devices using memristor and ReRAM structures [39, 40, 41, 42]. Examples thereof include titanium oxide (TiO_2), hafnium oxide (HfO_2), and tantalum oxide(TaOx). Firstly, TiO_2 and HfO_2 are used in neuromorphic hardware as synaptic devices with a memristor structure [39, 40]. TiO_2 enables monolithic three-dimensional stacking by sputtering at 300°C or less during deposition. However, for α-IGZO, the sputtering temperature during deposition can be about room temperature, and it enables monolithic three-dimensional stacking. In addition, the titanium oxide-based memristor used in the former research has a more complicated structure than α-IGZO because it uses multiple material. α-IGZO functions as a variable resistance element with a single material. It does not have a complicated structure and can be easily manufactured. HfO_2 is used for neuromorphic hardware as an analog variable resistance element, like TiO_2. The HfO_2-based memristor used in the latter research has a stacked structure of titanium nitride (TiN) and titanium (Ti) in addition to the oxide hafnium dioxide. Compared to α-IGZO, the material is expensive. Next, we explain researches on TiO_2 as a device with ReRAM structure. TiO_2 has a high relative dielectric constant and

excellent insulating properties. It is highly stable in resistance change and has been developed as an analog resistance change element for neuromorphic applications. Previous researches have included TiO_2-based ReRAMs that use Ta_2O_5 / TaOx as the oxide materials [41, 42]. Compared to α-IGZO, the structure is complicated, and it is important to adjust the ratio of the material composition in order to obtain stable operation. In order to realize a neural network at the hardware level, it is necessary to select an optimal device and structure, such as α-IGZO.

3.3 Cellular Neural Network

The configuration of the entire cellular neural network we produce is shown in Fig. 6. In the cellular neural network, adjacent neurons are connected by synapses, and each neuron is connected to top, bottom, left, right and diagonal neurons [43, 44]. The cellular neural network is suitable for pattern recognition and image processing because its structure is similar to that of retinal cells. Furthermore, since the connection of neurons through synapses is not complicated,

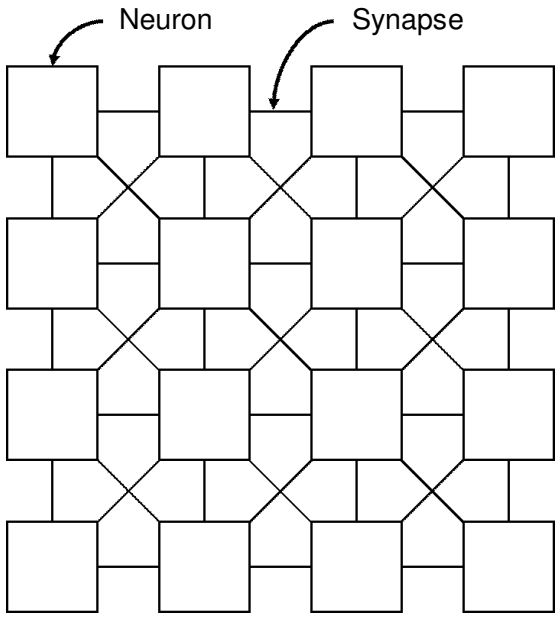


Figure 6. Model of cellular neural network

it can be easily scaled up when the hardware is created. However, there are concerns that the number of synapses is small and the efficiency of learning is low, and it is important to confirm the operation on simulations or real machines.

4. Proposed Method

In this capter, we propose a method to configure neuromorphic hardware using some novel technologies. An oxide semiconductor is used to configure the neural network that is easy to implement in hardware, and then how the network learns is shown.

4.1 Oxide Semiconductor Synapse

The function required for the synapse during learning is to connect neurons and the connection strength changes to an appropriate level depending on the state of the two neurons, and to maintain that strength. At the phase of correct, it is necessary to send a signal weighted by the connection strength to the signal from the neuron to the corresponding neuron. Therefore, the device used as a synapse was selected to have the property of deteriorating when a current flows and increasing the resistance value. When the resistance value of the synapse increases, it becomes difficult to transmit the signal from the neuron, and it can be said that the connection strength is weakened. When two connected neurons are in the same state, the connection strength is maintained. When the two neurons are in different states, electricity flows, so the synapse deteriorates and the connection strength decreases. This learning method differs from Hebb's learning rule[37], which raises the synaptic connection strength when the next neuron fires due to the firing of the previous neuron. In literature[32], this learning method was called modified Hebb's learning rule, and it was confirmed that learning was performed based on this learning rule using poly-Si TFTs, so that OR circuits and XOR circuits could be learned. Amorphous In-Ga-Zn-O (α-IGZO) oxide semiconductors is used as synapses. By annealing at a low temperature after forming a film, it becomes a device with the property of deteriorating when a current flows[35]. Although it has the characteristic that the resistance value of the deteriorated part decreases when the current is applied while applying light, it is unstable, so this time only the characteristic of deterioration is used.

Fig. 7 shows the synapse connection between the neurons, where the synapse connection includes a concordant synapse and discordant synapse. The concordant synapse connects the positive output of the previous neuron and the input

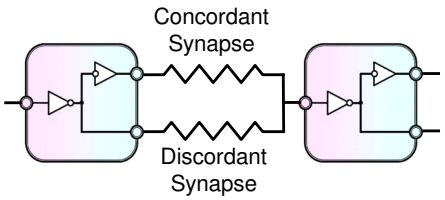


Figure 7. Synapse connection

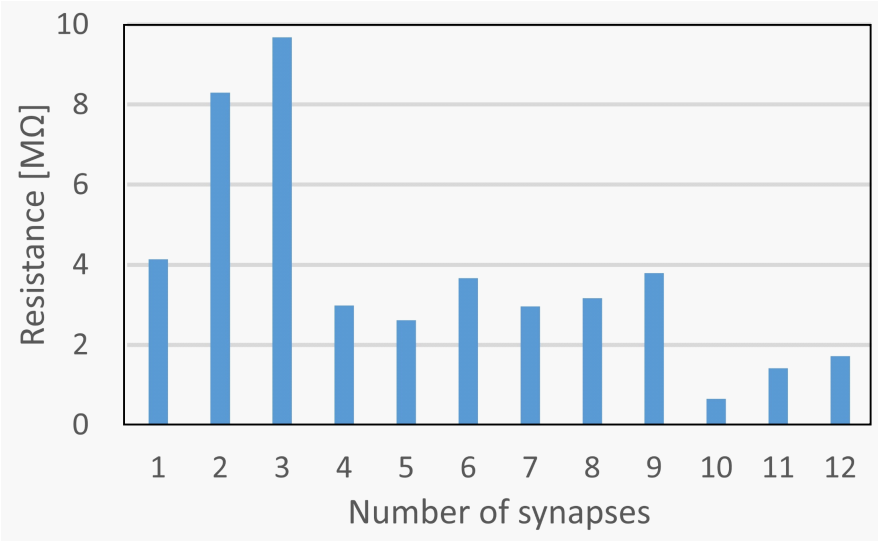


Figure 8. Initial resistance variation of the oxide semiconductor synapses

of the next neuron and operates to make both neurons in the same state. On the other hand, the discordant synapse connects the negative output of the previous neuron and the input of the next neuron and operates to make both neurons into different states. These two synapses can be used to increase or decrease the net connecting strength like an actual synapse. For example, deterioration of the concordant synapse relatively increases the net connecting strength.

It is known that the initial resistance of the oxide semiconductor has variations when it is fabricated. Fig. 8 shows the results of the measured resistance of 12 α-IGZO samples. The results show that the mean is 3.75 MΩ and the standard deviation is 2.5 MΩ, and although many samples are about 3 MΩ to 4 MΩ, the standard deviation becomes large because two samples exceed 8 MΩ. Therefore, the influence of the initial resistance variations of the oxide semiconductor on the

performance of the neuromorphic hardware was evaluated by simulation.

4.2 Network Configuration

Fig. 9 shows the connection of a neuron and eight adjacent neurons. The red arrow indicates the direction of current flow, and the size of the arrow roughly indicates the amount of current flow. The currents flowing from the output of the adjacent previous eight neurons are weighted through two types of synapses, concordant synapse and discordant synapse, according to the connecting strength of each synapse, and all the currents are summed and input to the next neuron. Depending on the input current, the next neuron becomes either firing state or stable state, and the state is maintained by the neuron circuit. Furthermore,

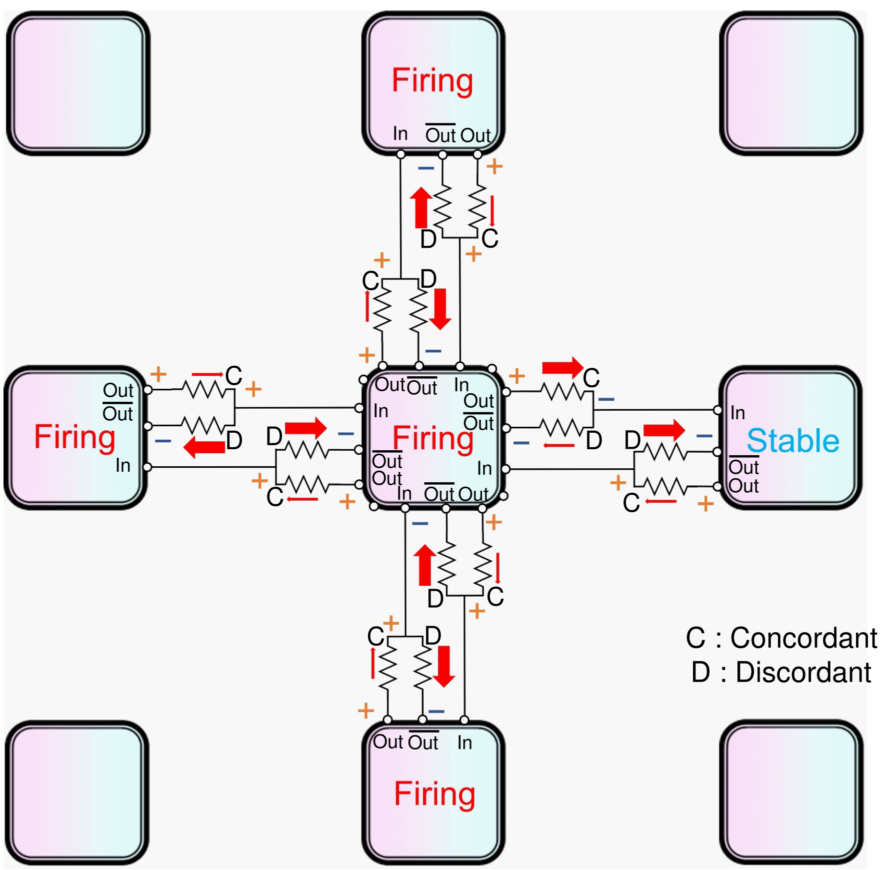
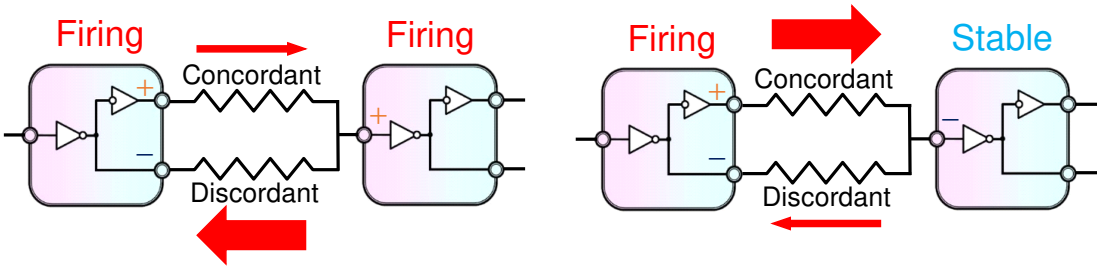


Figure 9. Connection of a neuron and eight adjacent neurons

in this cellular neural network, since the synapse is connected not only in the direction from the previous neuron to the next neuron but also interactively, there are four synapses between each neurons. Therefore, it is expected that the number of the synapses will increase and the accuracy of the neural network will be improved.

4.3 Learning Principle

Hebbian learning rules are typical rules in biological and artificial neural networks [37]. In Hebbian learning rules, the synapse connection strength increases when the states of both neurons connected to the synapse are the firing states and maintains or decreases otherwise. On the other hand, in the modified Hebbian learning rule [38], the synapse connecting strength is expressed by the relationship between those of the concordant synapse and discordant synapse. The synaptic connection strength changes depending on the state of adjacent neurons, which is similar to Hebbian learning rules. The way of changes in the synaptic connection strength according to the state of adjacent neurons is described below. Fig. 10(a) shows the connection of neurons when the neighboring neurons are in the same state. In this case, large amount of current flows from the input of the next neuron through the discordant synapse connected to the negative output of the previous neuron. As a result, the discordant synapses deteriorate, namely, the deterioration of discordant synapse decreases the electrical conductance, and the synaptic connection strength relatively increases. Fig. 10(b) shows the connection of neurons when the neighboring neurons are in different states. In this case,



(a)Increased synaptic connections (b)Decreased synaptic connections

Figure 10. Modified Hebbian learning rule(Only one direction)

large amount of current flows from the positive output of the previous neuron through the concordant synapse connected to the input of the next neuron. As a result, the concordant synapses deteriorate, namely the deterioration of concordant synapse decreases the electrical conductance, and the synaptic connection strength relatively decreases.

The advantage of modified Hebbian learning rules is that it is not necessary to add circuits to control synapse connection strength because the synapse connection strength is automatically controlled using local currents and voltages. Therefore, the modified Hebbian learning rules make it easy to implement neural networks at the hardware level.

5. Simulation Method

In this research, to evaluate neuromorphic hardware, it is evaluated by simulation assuming an associative memory algorithm. This section describes the associative memory algorithm and its simulation output method.

5.1 Associate Memory Algorithm

One of the functions of the human brain is associative memory. For example, it is easy for humans to look at a part of a photograph of a familiar person or pet and remember their faces. This feature has been addressed by many researchers in the field of neural networks on the theme of associative memory.

Fig. 11 shows a cellular neural network. Here, we used 3×3 neurons as input / output (I/O) and hidden neurons, where one hidden neuron was each placed between all the I/O neurons. The learning efficiency is improved by increasing the hidden layer, but if the hidden layer is increased too much, it becomes difficult for the input data to reach the output layer. In the example of Fig. 11, the character "T" is input to the I/O neuron, the neuron in the firing state is shown in red, and the neuron in the stable state is shown in white.

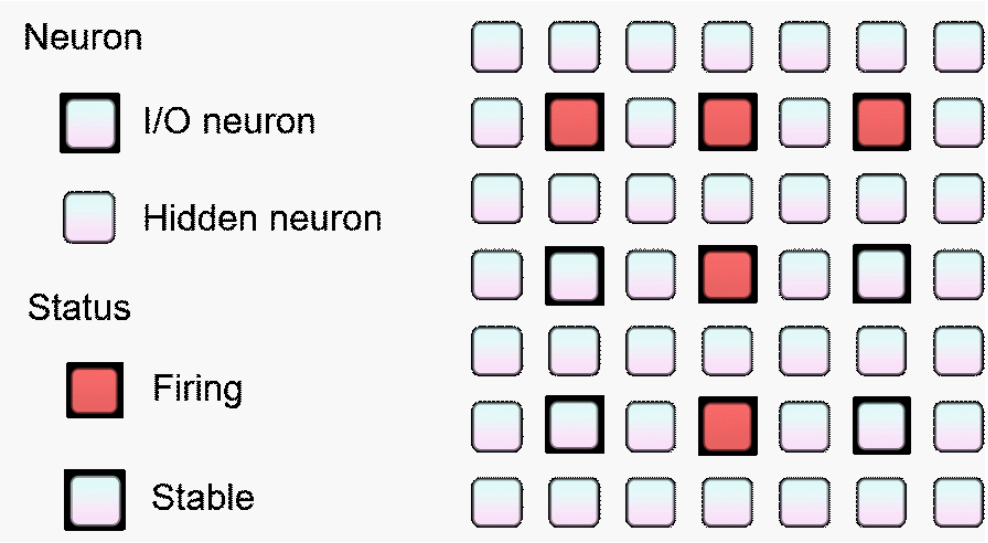


Figure 11. Associate memory algorithm

5.2 Output Method

Initial values are given to each neuron and synapse for simulation, where a voltage of 0.9 V is input to the neuron as an initial value, which is a half of the voltage sourse of 1.8 V. Then, the resistance value is input to the synapse in consideration of its characteristic variation that occurs when the oxide semiconductor is actually manufactured. Fig. 12 shows a letter pattern to be learned. The letter pattern is input, namely, the corresponding I/O neurons are set in the firing or stable state, where 1.8 V is input to the neuron in firing state, and 0 V is input to the neuron in stable state. After that, the states of all neurons including I/O neurons and hidden neurons are determined according to the connection strength of oxide semiconductor synapses between the neighboring neurons. Next, current flows through the oxide semiconductor synapse according to the states of the neurons, as a result, the synapse is deteriorated, and the synapse connection strength is reduced. After inputting the letter pattern to be learned in this way, a slightly different letter pattern is input to the I/O neuron, and we check if the letter pattern output matches the memorized letter pattern. If they match, the sumulation is successful, whereas if they do not match, the same procedure is repeated.

Fig. 13 shows a successful example of the associate memory simulation. When learning is completed, the letter pattern of Fig. 13(a) is input, and after waiting for a while, the letter pattern of Fig. 13(b) is output, which means that the simulation is successful.

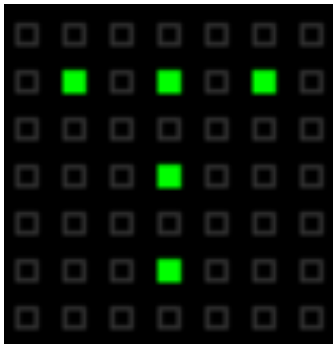
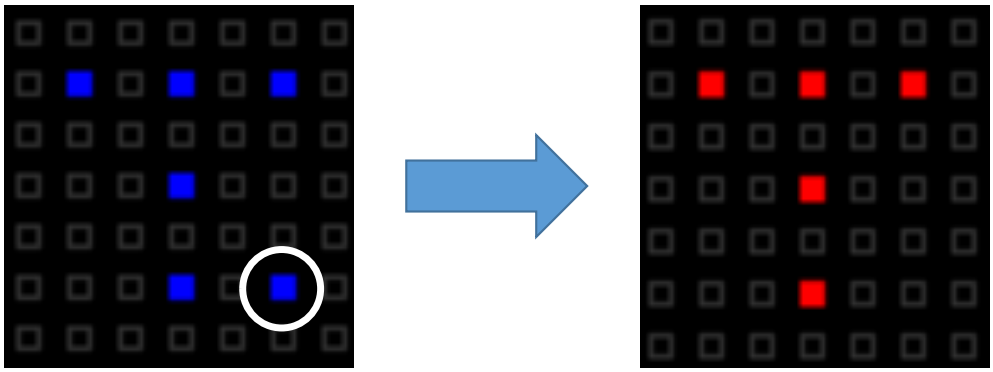


Figure 12. Memorized pattern



(a)Input pattern (b)Associated pattern

Figure 13. Example of the successful associate memory simulation

6. Simulation Results and Discussion

This chapter describes the results of three types of simulations using the associate memory algorithm described in chapter 5. First, we evaluated whether it depends on the number of hidden layers in the network. Second, we evaluated whether it depends on the number of letters memorized. Finally, we evaluated whether it depends on the number of pixels of the input pattern at the phase of association .

6.1 Dependence on the Number of Hidden Layers

In this section, the dependence of the hidden layer is evaluated using simulation with an associative memory algorithm. From the results, we confirmed which hidden layer number is most suitable for the assumed network.

Fig. 14 shows the model pattern of the associative memory simulation for the letter "T" of 3×3 pixels. A precise simulation is done using 9 kinds of input patterns with 1 different pixel. Fig. 15 shows the associated pattern when the number of hidden layers is set to 1, 3, 5, and 7 and the simulation using the above model is performed.

Fig. 16 shows the result of simulating the dependence of the number of hidden layers. The horizontal axis indicates the standard deviation of the initial resistance, and the vertical axis indicates the rate at which the memorized pattern in the associative memory matches the associated pattern. This is the success ratio

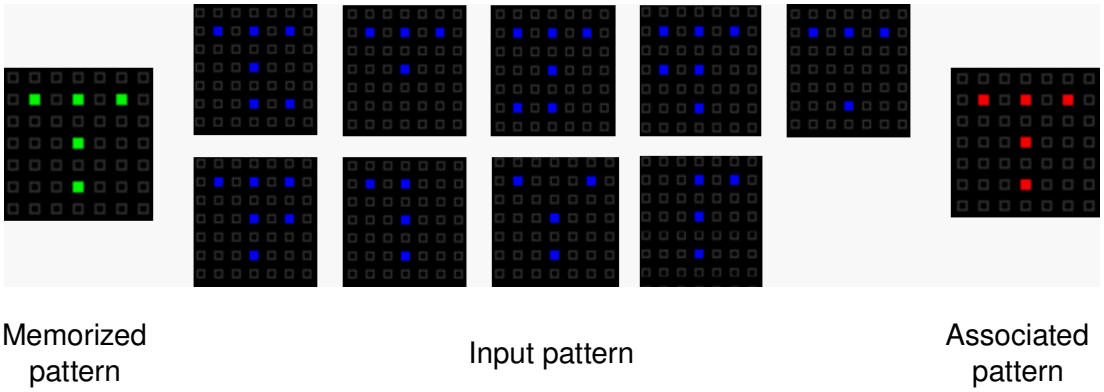


Figure 14. Model patterns for hidden layer dependence

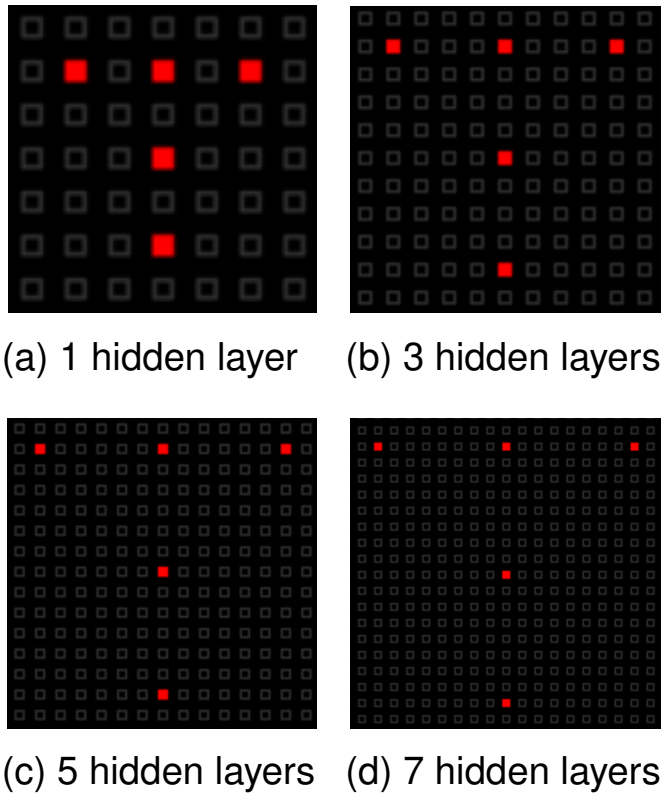


Figure 15. Associated patterns

of associative memory.

These results show that increase the number of hidden layers did not increase the correction rate very much. It can be seen that the correction ratio when the hidden layer is 7 is slightly higher than the others. However, considering the scale of the network proposed this time, if the hidden layer is larger than 1, there is a risk that the scale of the circuit will increase, and if the hidden layer is larger than 1, the shape as a letter will be lost. Therefore, it was found that the number of hidden layers suitable for the proposed neuromorphic hardware was 1.

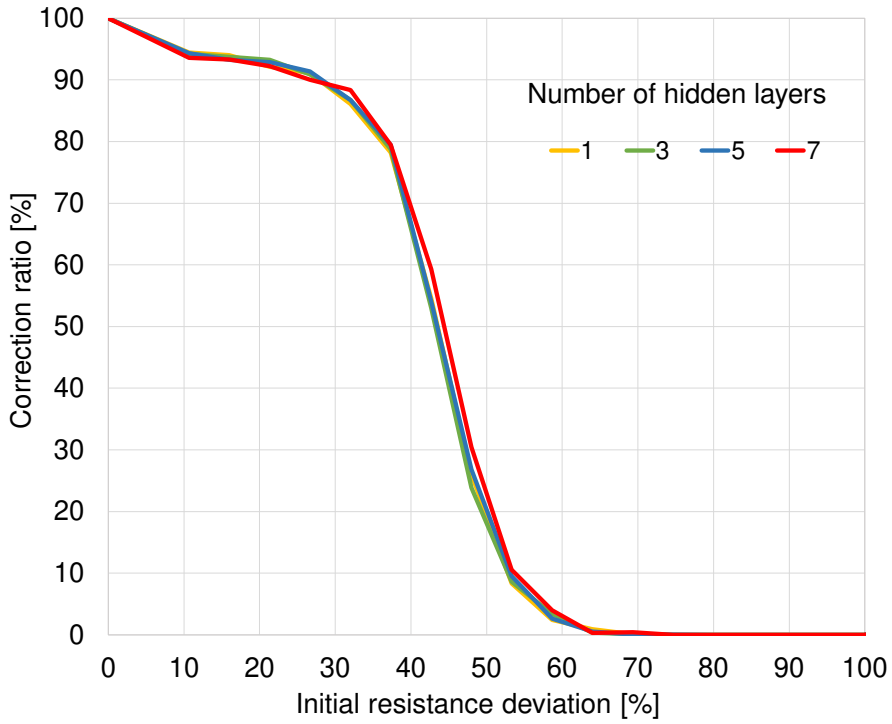


Figure 16. Dependence on the number of hidden layers

6.2 Dependence on the Number of Letters

In this section, the dependence of the number of letters to be memorized on a simulation is evaluated using the associate memory algorithm. Fig. 17 shows the model pattern of associative memory simulation for 3×3 pixel letters "T" and "L". We made a precise simulation using an input pattern with one different pixel.

Fig. 18 shows the result of simulation the dependence on the number of letters. The horizontal axis is the standard deviation of the initial resistance, and the vertical axis is the correction ratio of the associative memory. From these results, it was found that when the number of letters was increased from one to two, the correction ratio of associative memory was significantly reduced. This is what happened by memorizeed two letters at the same time. In order to make practical use of neuromorphic hardware, it is better that the size of the associative memory is large. Therefore, in experiments using real chip, the goal is to achieve 2 letters associative memory.

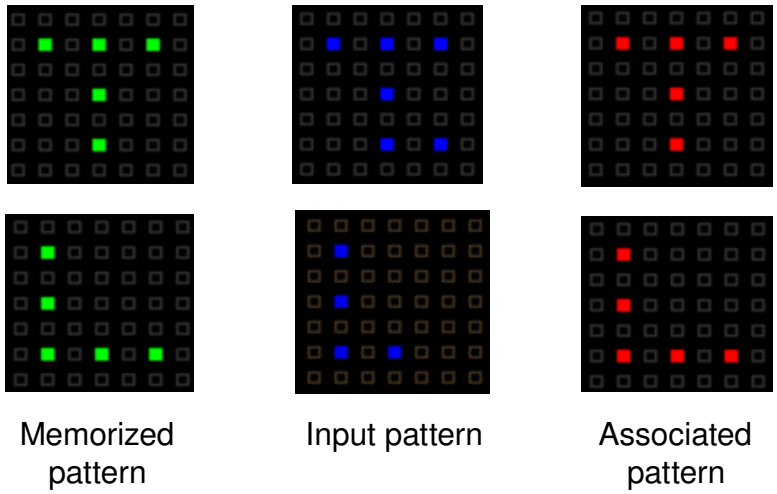


Figure 17. Model patterns for number of letters dependence

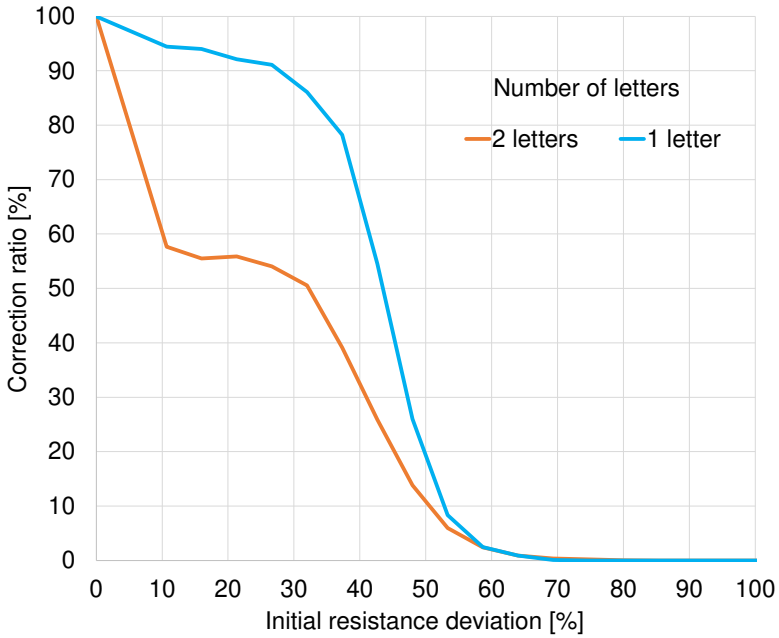


Figure 18. Dependence on the number of letters

6.3 Dependence on the Ratio of Different Pixels

In this section, the dependence on the ratio of different pixels of the input pattern is evaluated using a simulation with an associative memory algorithm. First,

Fig. 19 shows the model pattern of the associate memory simulation for the letter "T" of 3×3 pixels. A precise simulation is done using the input pattern in which the ratio of pixels from 0% to 50% is inverted.

Fig. 20 shows the result of simulating the dependence on the ratio of different pixels of 3×3 pixels input pattern. The horizontal axis is the standard deviation of the initial resistance value, and the vertical axis is the correction ratio of the associate memory. It was found that a high correction ratio of 80% or more can be obtained if the ratio of different pixels in the input pattern is up to 30% even

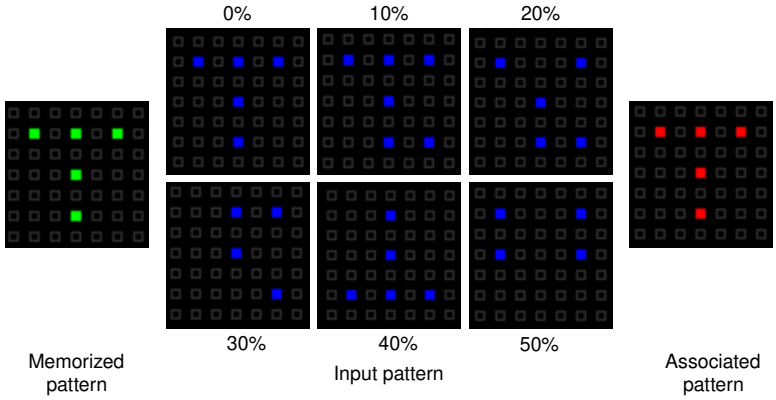


Figure 19. Model patterns for the ratio of different pixels (3×3 pixels)

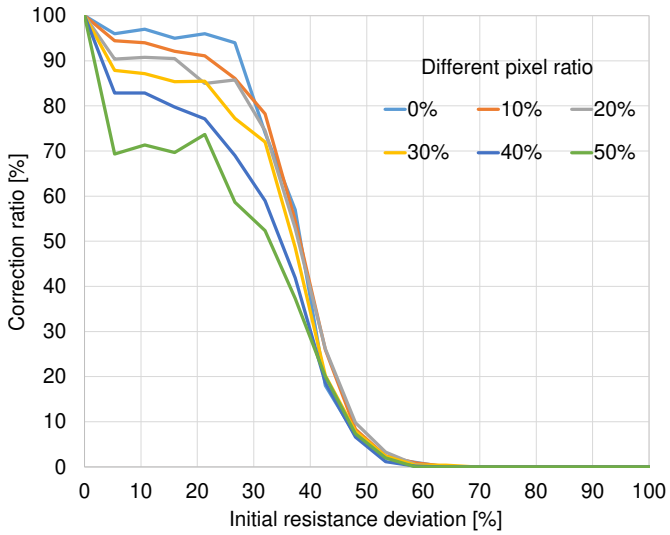


Figure 20. Dependence on the ratio of different pixels (3×3 pixels)

if the deviation of the initial resistance is 20%.

Next, Fig. 21 shows the model pattern of the associate memory simulation for the letter "0" of 6×6 pixels. As with the 3×3 pixels, I made a precise simulation using the input pattern in which the ratio of pixels from 0% to 50% is inverted.

Fig. 22 shows the result of simulating the dependence on the ratio of different pixels of 6×6 pixels input pattern. It was found that a high correction ratio of 80% or more can be obtained if the ratio of different pixels in the input pattern

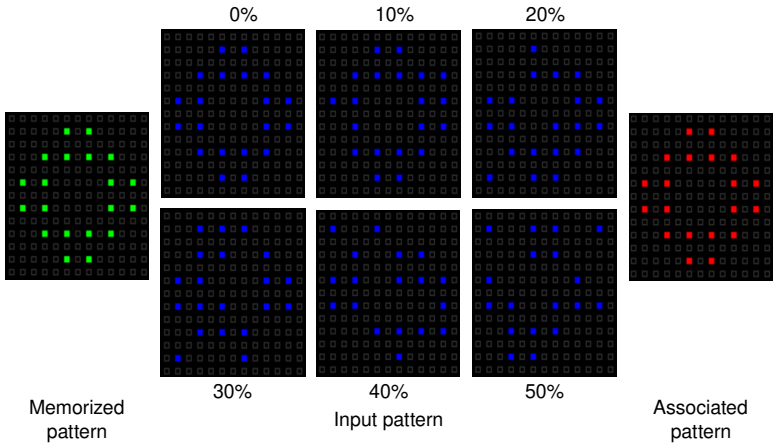


Figure 21. Model patterns for the ratio of different pixels (6×6 pixels)

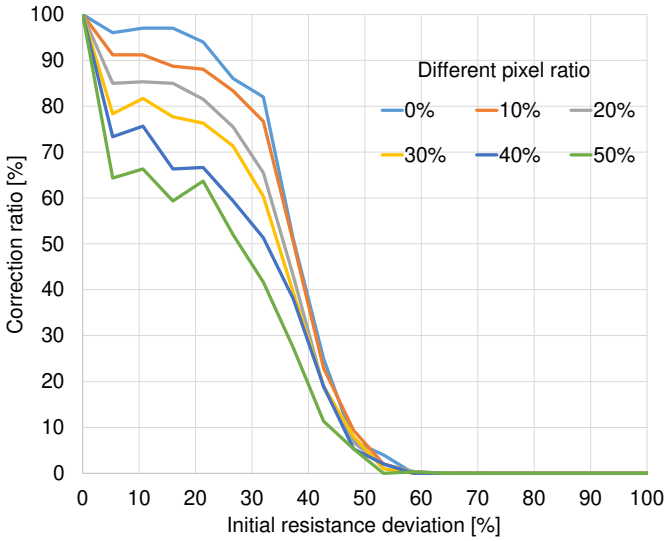


Figure 22. Dependence on the ratio of different pixels (6×6 pixels)

is up to 20% even if the deviation of the initial resistance is 20%.

Finally, Fig. 23 shows the model pattern of the associate memory simulation for the letter "月" of 9×9 pixels. As with the 3×3 pixels and 6×6 pixels, I made a precise simulation using the input pattern in which the ratio of pixels from 0% to 50% is inverted.

Fig. 24 shows the result of simulating the dependence on the ratio of different pixels of 9×9 pixels input pattern. It was found that a high correction ratio of

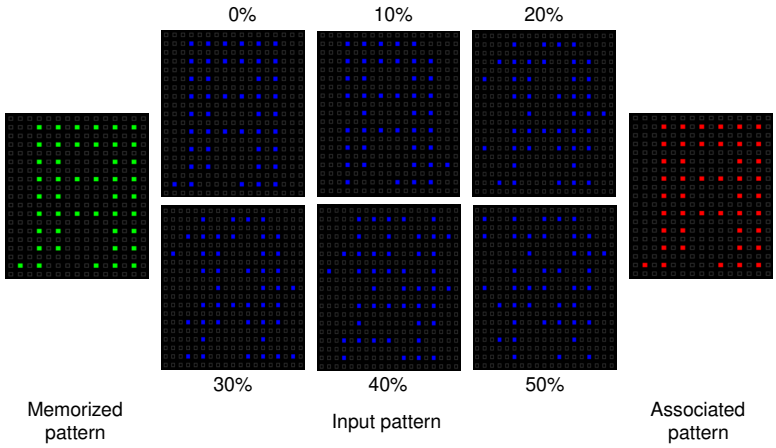


Figure 23. Model patterns for the ratio of different pixels (9×9 pixels)

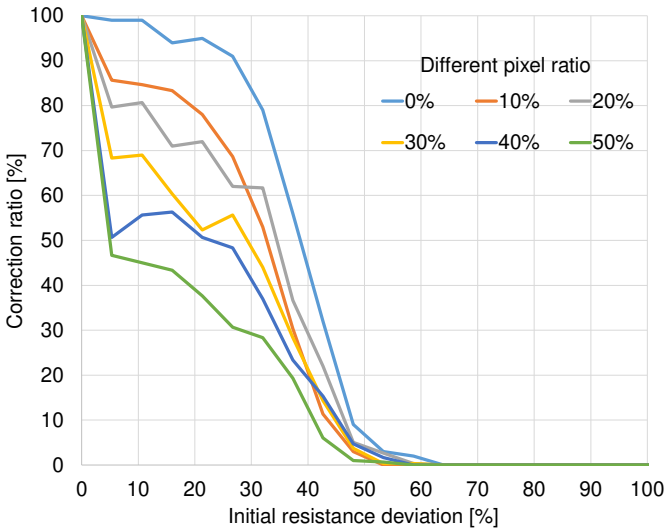


Figure 24. Dependence on the ratio of different pixels (9×9 pixels)

80% or more can be obtained if the ratio of different pixels in the input pattern is up to 10% even if the deviation of the initial resistance is 20%.

Compared to the result of 3×3 pixels, 6×6 pixels and 9×9 pixels, it can be seen that correction ratio has been decreased by increasing the size of target letter. However, if the pixel ratio of the input pattern differs by 10% to 30%, it is difficult even for us humans to judge what letter it is. From these results, we found that the proposed neuromorphic hardware can associatively memorize even letters having different shapes depending on the writer such as handwritten letters. Therefore, it is possible to construct an information processing system that can flexibly and robustly respond to input data that contains a lot of noise and ambiguity.

7. Experimental Method

In this chapter, it is described that the proposed neural network was created by hardware based on the evaluation by simulation and evaluated by using the associate memory algorithm.

7.1 Real Chips

The above simulation results show that the optimal hidden layer of the proposed network is 1 and that at least 2 letters learning is possible. Therefore, we created the neuromorphic hardware chip as shown in Fig. 25. The figure on the right shows the connection method between the top layer wiring (blue line) and the oxide semiconductor synapse placed thereon. The oxide semiconductor synapse is placed on the entire chip, and the dark yellow portion is an oxide semiconductor α-IGZO that operates as synapse. With this part, it is possible to connect to neurons located in eight directions through synapses. Fig. 26 shows a real neuromorphic hardware chip that is not a model. Its size is about 2.5 cm in length and width. In section 7.2, the specific manufacturing method of real chips is described.

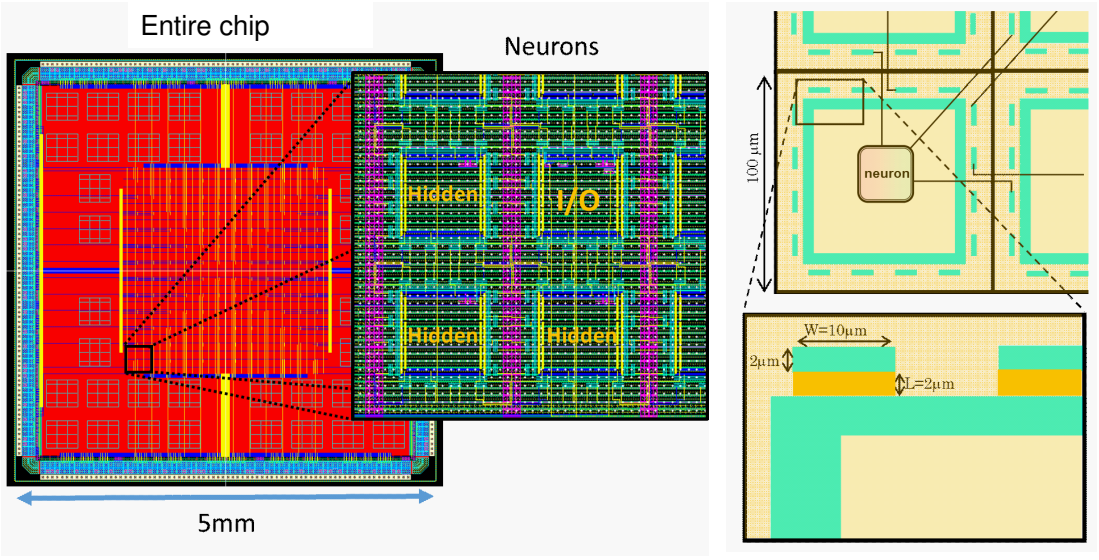


Figure 25. Real chip model

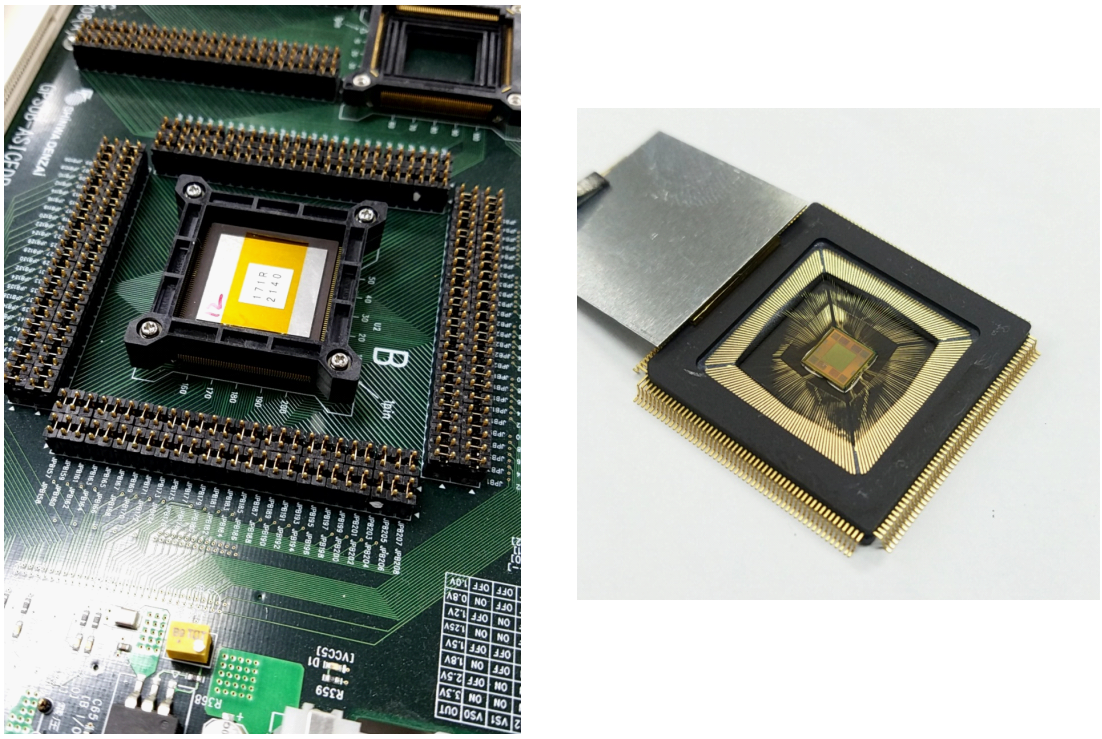


Figure 26. Real chip

7.2 Deposition Method and Deposition Conditions

7.2.1 Deposition Method : Sputtering

Fig. 27 shows the oxide semiconductor deposition method used in this experiment. Sputtering method is a type of physical vapor deposition (PVD) used for deposition. Sputtering phenomenon is that particles of the target are scattered by striking ions at high speed toward target. Using this phenomenon, a target material is formed on a substrate. First, a substrate and target(IGZO) are placed in vacuum chamber (1. in Fig. 27), and argon(Ar) gas is let in vacuum chamber. Next, a voltage is applied using the target side as cathode and the substrate side as anode. After that, argon is discharged to become positive ions, which are attracted to the cathode target (2. in Fig. 27). Argon ions clashed with the target, and the surface of target is scattered to the particle level (3. in Fig. 27). Finally, those scattered adhere to the substrate (4. in Fig. 27). The reason why sputtering is performed in a vacuum is to turn argon into ions, to reach target

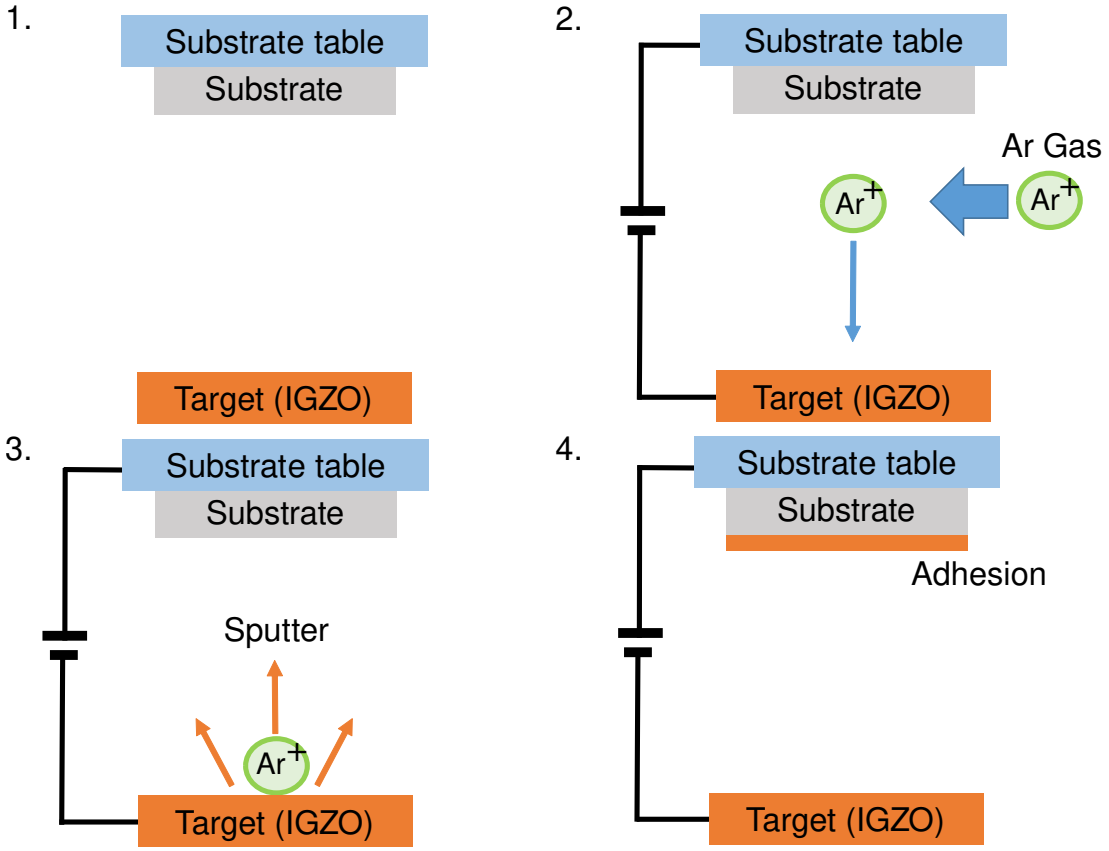


Figure 27. Deposition method

that have been sputtered to the substrate, and to form film with few impurities.

7.2.2 Deposition Conditions

Here we described the conditions when α-IGZO was deposited to create the neuromorphic hardware chip. Table 1 shows the deposition conditions of the created chips.

Chip No.1-6 shown in table 1 are created with the deposition structure shown in Fig. 28 (Example: the deposition structure of Chip No.1). It can be seen that deviation of the initial resistance of IGZO is changed by forming films with different deposition times. It is necessary to determine the deposition time at which the deviation value becomes small. In the chips using this deposition structure, the deviation could be improved to 27%.

Table 1. Deposition conditions of Chip No.1-6

Chips	Deposition Conditions		
No.	Deposition Time	Gas Flux[sccm]	Deviation[%]
1	25min	Ar : 20	197
2	12min	Ar : 20	27
3	10min	Ar : 20	29
4	7min30sec	Ar : 20	112
5	6min15sec	Ar : 20	139
6	5min	Ar : 20	130

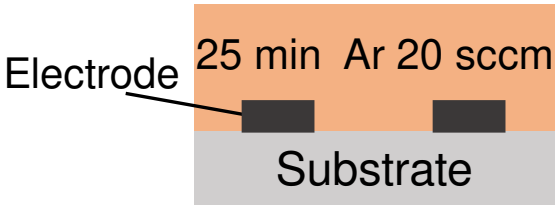


Figure 28. Deposition structure of chip.1-6

Next, chips are created by changing the composition of gas used for sputtering. These deposition conditions are shown in Table 2. Chip No.7-10 shown in Table 2 are created with the deposition structure shown in Fig 29 (Example: the deposition structure of Chip No.9).

From Fig. 29, only argon gas was let for the first 10 minutes and then oxygen is added only for 5 minutes. After 15 minutes, the film is formed only with argon without oxygen for 10 minutes. Oxygen was added during the deposition process to achieve uniform characteristics. When oxygen is present in all the films, it becomes difficult for the current to flow. This is because the influence from the external environment could be suppressed. Therefore, as shown in Fig 29, the film is separated into layers with oxygen and layers without oxygen. The film is deposited on No.7 chip without changing the composition of Gas, only changing the deposition time. The film is deposited on No.8 chip without changing the gas composition of Gas, only the deposition structure. It was deposited with argon gas and oxygen for only 5 minutes. As the result, the deviation of the

Table 2. Deposition conditions of Chip No.7-10

Chips	Deposition Conditions		
No.	Deposition Time	Midle Layer Gas Flux[sccm]	Deviation[%]
7	10 + 5 + 10 min	Ar : 20, O_2 : 0.1	107
8	5min	Ar : 20, O_2 : 0.1	66
9	10 + 5 + 10 min	Ar : 20, O_2 : 0.2	101
10	10 + 5 + 10 min	Ar : 20, O_2 : 0.4	23

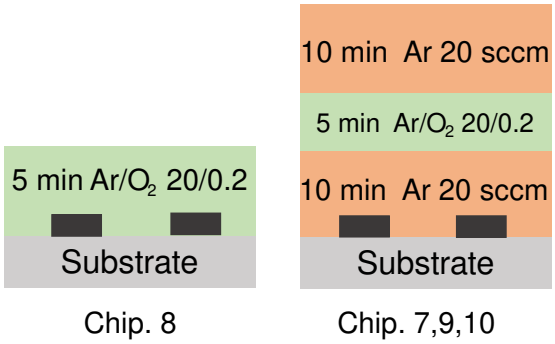


Figure 29. Deposition structure of chip.7-10

initial resistance was improved from 107% to 66%. Next, The film is deposited on No.9 and No.10 chips without changing the deposition time, only changing the gas composition. As a result, the deviation of the initial resistance was greatly improved from 101 % to 23 %.

Finally, Chip No.11 is created by irradiating the IGZO film with light, without changing the deposition conditions. This is because it has been found that when light is irradiated to the deposition area during the chip creating, the deviation of the initial resistance is made uniform. Table 3 shows the deposition conditions for Chip No.11. The deposition structure is the same as that of Chip No.9 shown in fig. 29.

Fig. 30 shows the initial resistance of Chip No.11 when light is not irradiated, Fig. 31 shows the initial resistance of Chip No.11 when irradiated with light. From the graph of Fig. 30, it was found that the deviation of the initial resistance when light is not irradiated was 52%. Conversely, from the graph of Fig. 31, it

Table 3. Deposition conditions of Chip No.11

Chips	Deposition Conditions		
No.	Deposition Time	Midle Layer Gas Flux[sccm]	Deviation[%]
11	25min	Ar : 20, O_2 : 0.2	21

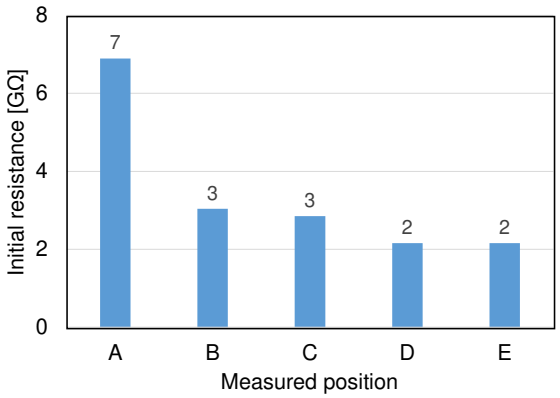


Figure 30. Initial resistance of a chip without light

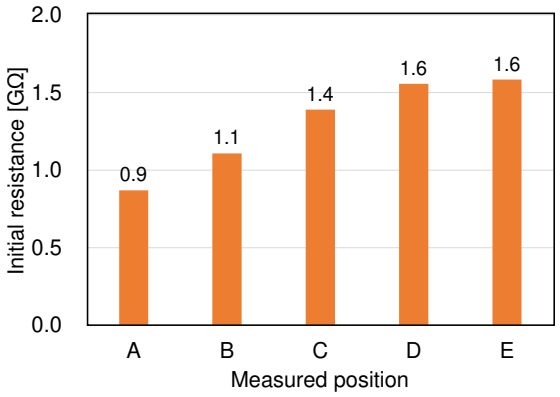


Figure 31. Initial resistance of a chip with light

was found that the deviation of the initial resistance when irradiated with light was 21%. This is because the free carriers generated by the light terminated non-uniform defects in the deposition area. Therefore, the chip irradiated with light is used for this experiment.

8. Experimental Results and Discussion

In this chapter, the created neuromorphic hardware chip is evaluated by using an associative memory algorithm. Fig. 32 shows the model pattern of the associative memory experiment for 3×3 pixel letters "T" and "L".

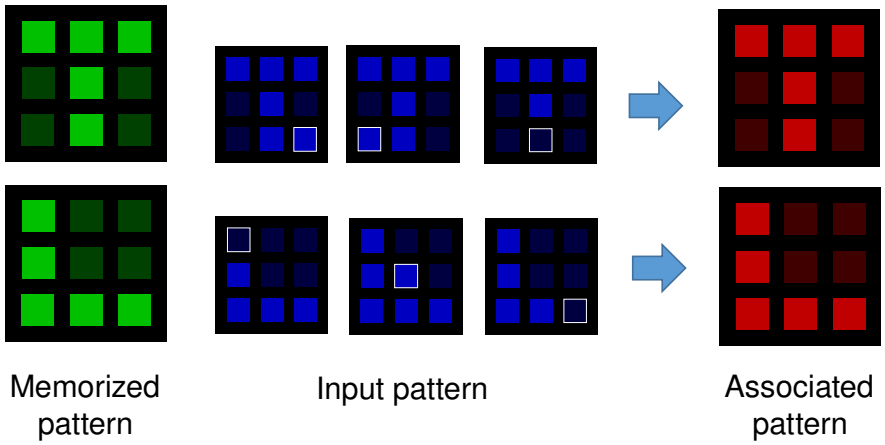


Figure 32. Model patterns for associate memory experiment

8.1 Experimental Results for Each Chip

Table.4 shows the experimental results of Chip No.1-6. Chip No.2 with deviation of 27%, Chip No.3 with deviation of 29%, and Chip No.6 with deviation of 130% succeeded in 1 letter associate memory experiment. From these results, it was found that the experiment of the associate memory would succeed if the deviation of initial resistance was small. However, no 2 letters associative memory was successful.

Fig. 33 shows the experimental result of Chip No.3 which is a successful example. As this figure shows, 3 types of input patterns in the associative memory experiment was successful.

Next, Table 5 shows the experimental results of Chip No.7-10. Chip No.8 with deviation of 66% succeeded in 1 letter associate memory experiment. However, no experiment of 2 letters associative memory was successful. Chip No.10 had small initial resistance deviation of 23%, but the wires in the chip were broken. Therefore, the experiment could not be performed well.

Table 4. Experimental results of Chip No.1-6

Chips	Experimental Results	
No.	1 Letter	2 Letters
1	×	×
2	✓	×
3	✓	×
4	×	×
5	×	×
6	✓	×

✓ : success × : not success

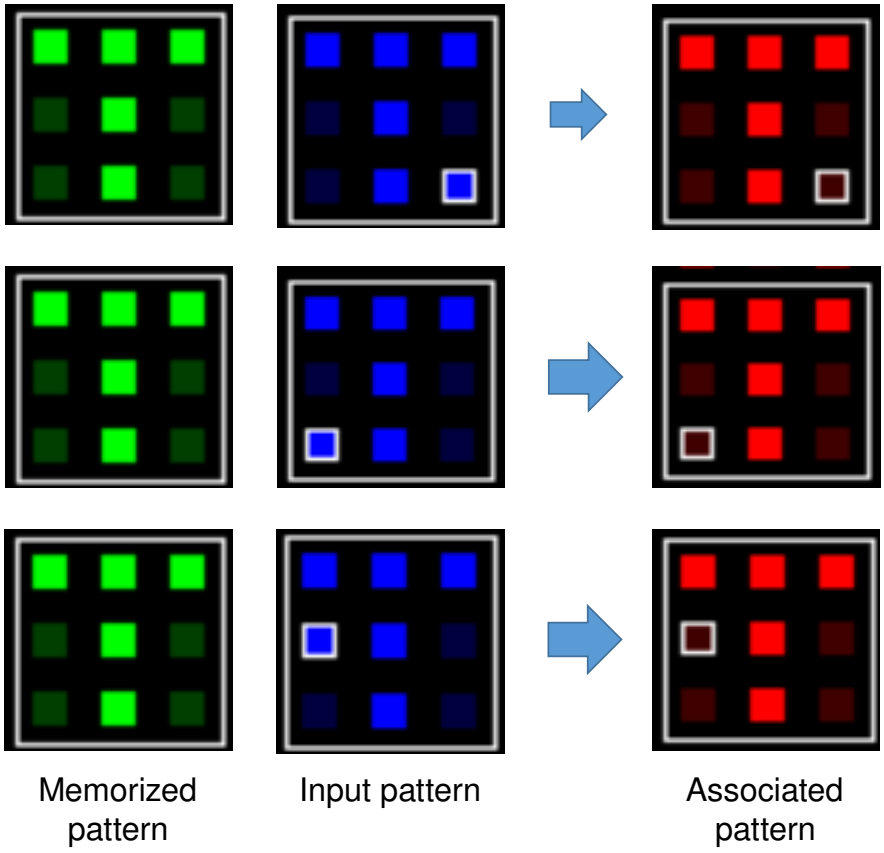


Figure 33. Successful experiment : Chip No.3

Table 5. Experimental results of Chip No.7-10

Chips	Experimental Results	
No.	1 Letter	2 Letters
7	×	×
8	✓	×
9	×	×
10	×	×

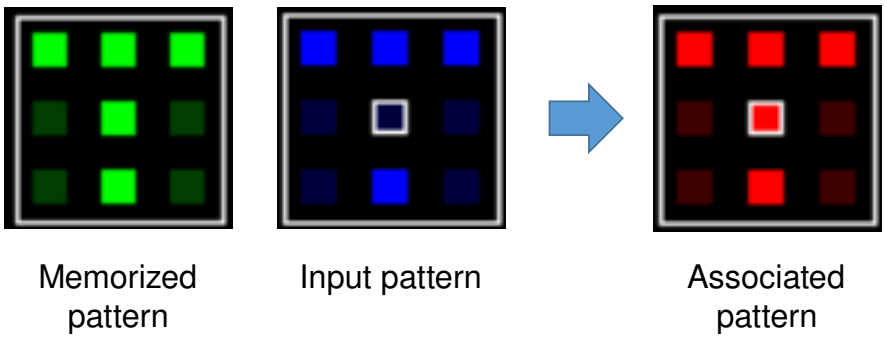


Figure 34. Successful experiment : Chip No.8

Fig. 34 shows the experimental result of Chip No.8 which is a successful example. As this figure shows, the associate memory experiment of the 1 type of input pattern was successful.

Finally, Table 6 shows the experimental results of Chip No.11. 1 letter and 2 letters associate memory experiment of Chip No.11 with deviation of 21% succeeded.

Fig. 35 shows the experimental result of Chip No.11 which is a successful example. As this figure shows, the associate memory experiment of 1 letter and 2 letters using 1 types of input patterns was successful.

Table 6. Experimental results of Chip No.11

Chips	Experimental Results	
No.	1 Letter	2 Letters
11	✓	✓

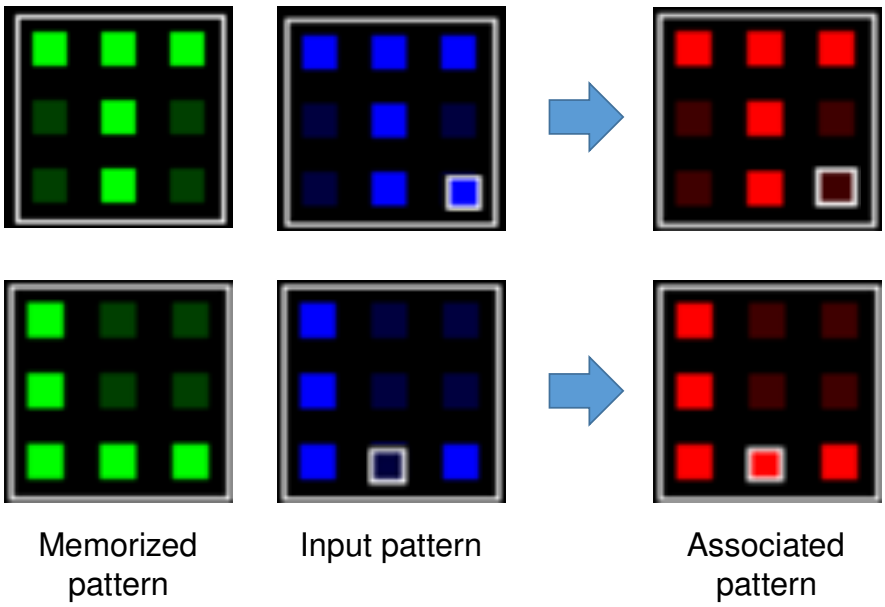


Figure 35. Successful experiment : Chip No.11

8.2 Comparison between simulation and experimental results

Fig. 36 shows comparison between the results of this experiment and the results of 2 letters associative memory simulation. The horizontal axis is the standard deviation of the initial resistance, the left vertical axis is the correction ratio in the associate memory simlation and the right vertical axis is the number of letters successful in the associate memory experiment. From this graph, It was found that the number of successful experiments increased as the deviation of the initial resistance value decreased to 20%. When the deviation of the initial resistance value is 21%, it can be seen that 2 letters associate learning was succeeded in the experiment using the neuromorphic hardware chip. From this result, it was found

that associative learning was successful if the deviation of the initial resistance of the oxide semiconductor α-IGZO was small.

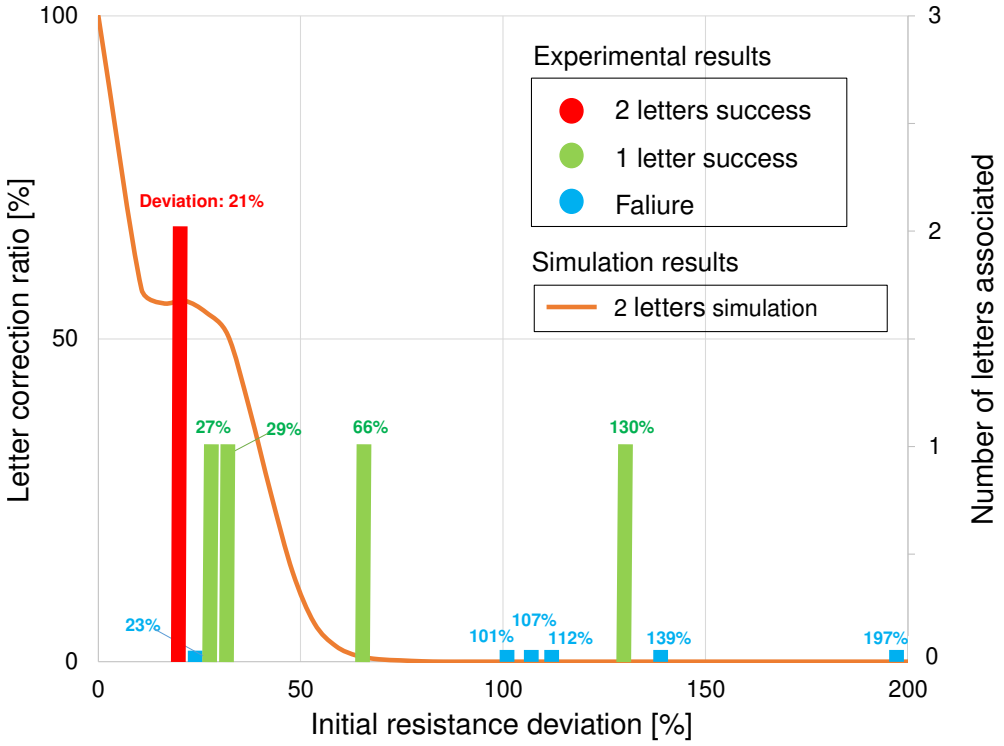


Figure 36. Comparison between experimental and simulation results

9. Conclusion

In this paper, the operation of a hardware-based neuromorphic chip using oxide semiconductor α-IGZO was confirmed by simulation using an associative memory algorithm. First, in simulation of the dependence on the number of hidden layers in the network, I found that the correction ratio of associative memory did not change as the number of hidden layers increased. Moreover, considering the distortion of shape of the target letter due to the increase in the number of hidden layers, it was concluded that it is optimal to have 1 hidden layer.

In simulation of the dependence on the number of associative letters, I found that the correction ratio for 2 letters associative learning was lower than that for 1 letter associative learning. From this result, the goal of this research was to make 2 letters associate learning successful in experiments using real chips.

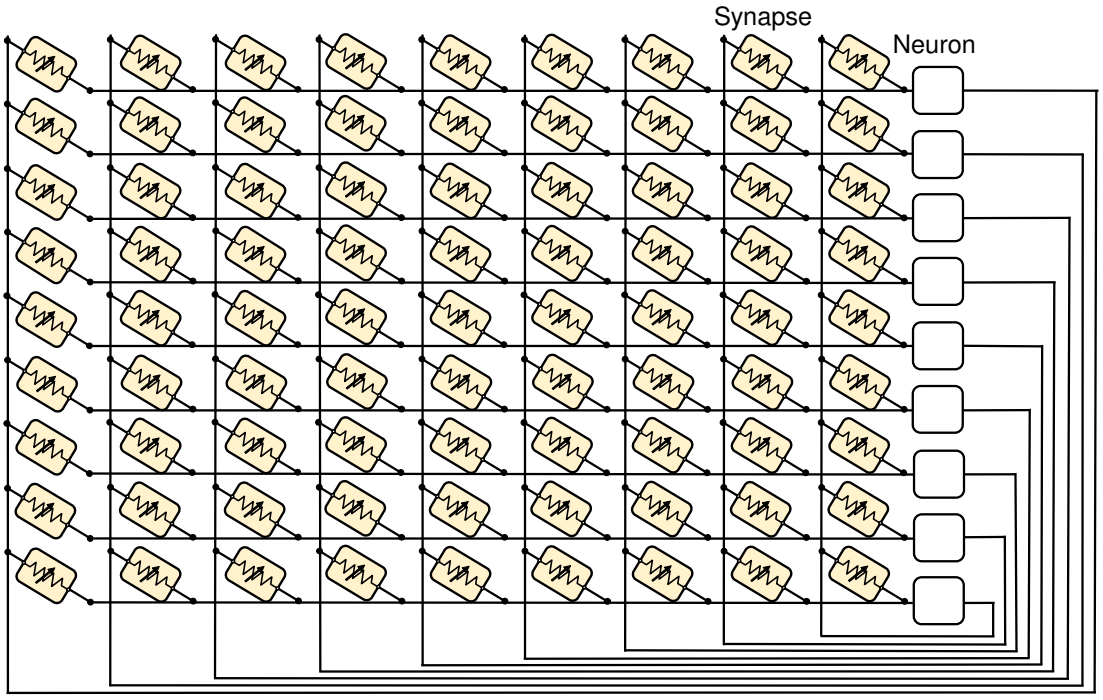
In 3×3 pixels simulation on the dependence of the ratio of different pixels on the input pattern, it was found that high correction ratio of 80% or more could be obtained if the initial resistance deviation is less than 20% when the ratio of different pixels of the input pattern is up to 30%. Next, in 6×6 pixels simulation on the dependence of the ratio of different pixels on the input pattern, it was found that high correction ratio of 80% or more could be obtained if the initial resistance deviation is less than 20% when the ratio of different pixels of the input pattern is up to 20%. Next, in 9×9 pixels simulation on the dependence of the ratio of different pixels on the input pattern, it was found that high correction ratio of 80% or more could be obtained if the initial resistance deviation is less than 20% when the ratio of different pixels of the input pattern is up to 10%. From these results, it was found that the proposed neuromorphic hardware can flexibly and robustly respond to input data having a lot of noise and ambiguity.

In addition, from the evaluation results of the simulation, we created real neuromorphic hardware chips using actual devices. The fabricated chip has 3 types of deposition structures. One of them uses only argon gas during sputtering. These chips with initial resistance deviation of 27%, 29%, and 130% succeeded in 1 letter associate memory experiments. Moreover, the other one of 3 types of deposition structures uses argon gas and oxygen gas during sputtering. These chips with initial resistance deviation of 66% succeeded in 1 letter associate memory experiments. Furthermore, the last one is subject to the irradiation of the light

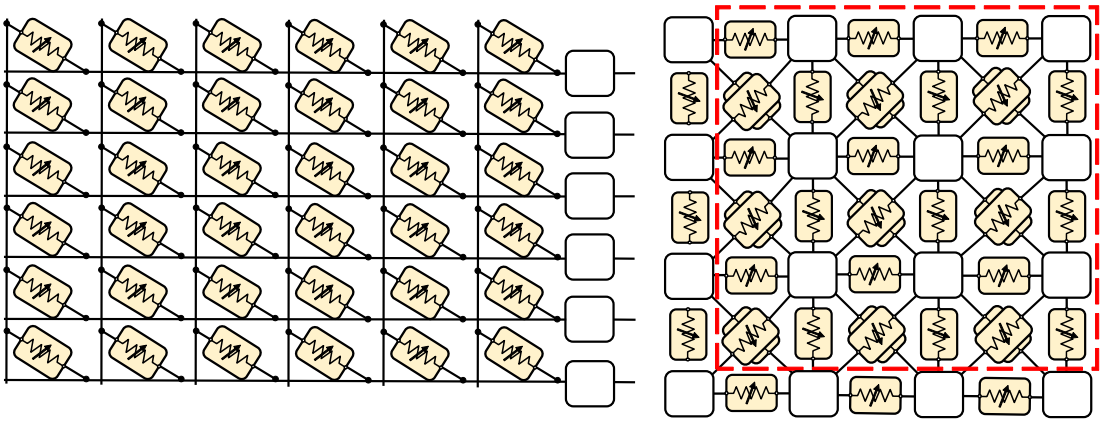
to the deposition area of the chip. This chip with initial resistance deviation of 21%, succeeded in 1 letter and 2 letters associate memory experiments.

Finally, I could confirm the operation of the chip that can be created at low temperature and can be said to be a prototype of neuromorphic hardware using highly integrated oxide semiconductor synapses. In the future, it is urgently necessary to produce oxide semiconductor synapses more accurately and to improve the manufacturing environment for wiring to chips.

In this paper, we proposed a cellular neural network in which adjacent neurons are connected. Here, the cellular neural network proposed in this research and the Hopfield network in which all neurons are connected by synapses is compared. Fig. 37 (a) shows a Hopfield network with 9 neurons, (b) shows a unit area Hopfield network and cellular neural network. The red dotted line indicates the unit area of the cellular neural network. The hopfield network has overwhelming number of synapses compared to cellular neural network. Therefore, it is considered that the learning efficiency of the hopfield network is better. However, considering the size of networks per unit area, a cellular neural network does not require long-distance wiring when the network is expanded. Therefore, the power loss can be reduced, the number of neurons is increased, and the integration efficiency is improved. However, the cellular neural network is limitedly used for the autoassociative memory. The autoassociative memory is a storage method as shown in this research, in which information is expressed as a pattern and is memorized in a dispersed manner, and when a part of the pattern is given, the entire original pattern is corrected. On the other hand, the Hopfield network is hetero associative memory. It can perform a function similar to pattern recognition for reading out corresponding pattern from plurality of patterns. It is expected that Hopfield networks using crosspoints will be used for various purposes in future research.



(a)Hopfield network with 9 neurons



(b)A unit area Hopfield network and cellular neural network

Figure 37. Comparison of hopfield network and cellular neural network

Acknowledgements

研究活動において, 多大なるご助言やご指導を頂きました本学の中島康彦教授に深く感謝の意を表します. 本論文をご精読頂き研究発表を通じて貴重なご意見を頂いた本学の池田和司教授に深く感謝いたします. また学会発表やミーティング, 講義において支援とご助言を頂いた中田尚准教授, TRAN Thi Hong 助教, 張任遠助教に深く感謝いたします. 本研究を行うにあたり, ひとかたならぬご尽力と的確なご指導を頂きました客員教授の木村睦教授には感謝の念に堪えません. また実験にあたっては龍谷大学木村睦研究室の院生, 学部生の皆様にご協力頂きました. ここに感謝致します. そして, 2 年間を共に過ごし支えて頂いた同輩の岩本淳氏, 新谷隆太氏, 西本宏樹氏の協力と励ましが無ければ, 私がこの修論を完成させることはなかったでしょう. 心から感謝致します. 研究室での生活を日頃からサポートして頂いた後輩の滝下雄太氏, 野村武司氏, 本田卓氏をはじめとするコンピューティング・アーキテクチャ研究室の皆様には大変お世話になりました. 最後に, 大学院での生活を支えてくださった家族に心より感謝致します. 本当にありがとうございました.

References

[1] J. McCarthy, M. L. Minsky, N. Rochester and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," *Dartmouth Conference*, 1956.

[2] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach, Pearson Education," *Prentice Hall*, 2009.

[3] J.E.Dayhoff, "Neural Network Architectures: An introduction," *Van Nstrand Reinhold*, 1990.

[4] R. Hecht-Nielsen, "Neurocomputing," *Addison-Wesley Reading*, 1990.

[5] S. Becker, "Self-Organizing Neural Network that discovers Surfaces in Random-Dot Stereograms," *Nature*, No. 355, pp. 161–163, 1992.

[6] J. V. Stone, "Neural-Network Models: Predicting Spontaneous Recovery of Memory," *Nature*, No. 414, pp. 167–168, 2001.

[7] V.B. Moutcastle, "The Cerebral Cortex," *Harvard College*, 1998.

[8] Y. Lecun and L. Bottou and Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[9] Krizhevsky, Alex and Sutskever, Ilya and E. Hinton, Geoffrey, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Information Processing Systems*, Vol. 25, 2012.

[10] T. Morie, S. Ishikawa, "Intelligent Image Recognition Technology and Brain-like LSI Implementation," *The Journal of the Institute of Electronics, Information and Communication Engineers*, Vol. 94, No. 6, pp. 459–463, 2011.

[11] Ministry of Internal Affairs and Communications, "The Arrival of the Big Data Use Year" *White Paper Information and Communications*, pp. 52–62, 2017.

[12] A. Nagamatsu, “Current Issues Related to Big Data Utilization in Japan,” *Journal of Information and Management*, Vol. 36, No. 4, pp. 3–11, 2016.

[13] J. von Neumann, “First Draft of a Report on the EDVAC,” *Philadelphia: University of Pennsylvania*, 1945.

[14] TOKYO OHKA KOGYO CO.LTD., “TOK’s Microfabrication Processing Technology and the Future of Computers,” *CSR Report 2017*, pp. 13–18, 2017.

[15] G. E.Moore, “Cramming More Components Onto Integrated Circuits,” *Proceedings of the IEEE*, No. 86(1), pp. 82–85, 1998.

[16] J. Nickolls, W. J. Dally, “The GPU Computing Era,” *IEEE Micro*, No. 30(2), pp. 56–69, 2010.

[17] Altera Corp, “Arria 10 Core Fabric and General Purpose I/Os Handbook,” *Arria 10 Handbook*, 2016.

[18] IBM, “IBM Scientists Demonstrate In-memory Computing with 1 Million Devices for Applications in AI,” https://www.ibm.com/blogs/research/2017/10/ibm-scientists-demonstrate-memory-computing-1-million-devices-applications-ai/, Accessed in May 2019

[19] J. E. Dayhoff, “Neural Network Architectures, An Introduction,” *Van Nostrand Reinhold*, 1990.

[20] G. Matsumoto, Y. Shigematsu, M. Ichikawa, “Information science from brain reserch,” *Journal of Information Processing Society of Japan*, Vol. 38, No. 5, pp. 369–375, 1997.

[21] Zidong Du, Daniel D, Ben-Dayan Rubin, Yunji Chen, Liqiang He, Tianshi Chen, Lei Zhang, Chengyong Wu, and Olivier Temam, “Neuromorphic Accelerators: A Comparison Between Neuroscience and Machine-learning Approaches,” *Proceeding of the 48th Interbational Symposium on Microarchitecture*, pp. 494–507, 2015.

[22] J. Schemmel, et al.,"Live demonstration: A scaled-down version of the brainscales wafer-scale neuromorphic system," *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, p. 702, 2012.

[23] Mead C, "Analog VLSI and Neural Systems," *Addison-Wesley Publishing*, 1989.

[24] Y. Arima, K. Mashido, K, Okada, T. Yamada, A. Maeda, H. Kondoh, S. Kayano, "A 336-neuron, 28 k-synapse, self-learning neural network chip with branch-neuron-unit architecture," *IEEE Journal of Solid State Circuits*, Vol. 26, No. 11, pp. 1637–1644, 1991.

[25] M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imai, S. Jain, Y. Liao, C. K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, Guruguhanathan, Venkataramanan, Y. H. Weng, A. Wild, Y. Yang, H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, Vol. 38, No. 1, pp. 82–99, 2018.

[26] Yiran Chen, Qing Wu, "Neuromorphic Computing: A SoC scaling path for the next decades," *IEEE International SOC Conference*, 2012.

[27] B. V. Benjamin, et al., "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," *Proc. IEEE*, Vol. 102, No. 5, pp. 699–716, 2014.

[28] Qualcomm, "Introducing Qualcomm Zeroth Processors: Brain Inspired Computing," `https://www.qualcomm.com/news/onq/2013/10/10/introducing-qualcomm-zeroth-processors-brain-inspired-computing`, Accessed in May 2019.

[29] Vinayak Gokhale, Jonghoon Jin, Aysegul Dundar, Berin Martini, Eugenio Culurciello, "A 240 G-ops/s Mobile Coprocessor for Deep Neural Networks (Invited Paper)," *Computer Vision and Pattern Recognition Workshops*, 2014.

[30] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G. Nam, B. Jackson, D, S, Modha, "Truenorth: Design and Tool Flow of a 65 mw 1 Million Neuron Programmable Neurosynaptic Chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and systems*, Vol. 34, No. 10, pp. 1537–1557, 2015.

[31] Information processing model by neuron's spike firing timing and its LSI implementation, `http://www.brain.kyutech.ac.jp/~morie/topics/spiking.shtml`, Accesed in March 2019.

[32] T. Kasakawa, "An Artificial Neural Network at Device Level using Simplified Architecture and Thin-Film Transistors," *IEEE Trans. Electron Devices*, No. 57, pp. 2744–2750, 2010.

[33] John. F. Wager, "Transparent electronics," *Science*, Vol. 300, No. 5623, pp. 1245–1246, 2003.

[34] Kenji Nomura, Hiromichi Ohta, Akihiro Takagi, Toshio Kamiya, Masahiro Hirano, Hideo Hosono, "Room-Temperature Fabrication of Transparent Flexible Thin-Film Transistors using Amorphous Oxide Semiconductors," *Nature*, Vol. 432, pp. 488–492, 2004.

[35] S. Imai, M. Kimura, "Degradation Evaluation of α-igzo TFTs for Application to AM-OLEDs," *Electron Device Letters*, Vol. 31, pp. 963–965, 2010.

[36] M. Kimura, K. Ikushima, D. Yamakawa, H. Yamane, Y. Nakashima, "Real Neuromorphic System using LSI Chip and Thin-Film Devices," *International Conference of Neuromorphic Systems 2019 (ICONS2019)*, 2019.

[37] D. O. Hebb, "The Organization of Behavior: A Neuropsychological Theory," *Science Education*, Vol. 34, No. 5, pp. 336–337, 1950.

[38] T. Kasakawa, H. Tabata, R. Onodera, H. Kojima, M. Kimura, H. Hara, S. Inoue, "An Artificial Neural Network at Device Level using Simplified Architecture and Thin-Film Transistors," *IEEE Transactions on Electron Devices*, Vol. 57, No. 10, pp. 2744–2750, 2010.

[39] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, "Training Andoperation of An Integrated Neuromorphic Network Based on Metal-Oxide Memristors," *Nature*, Vol. 521, pp. 61–64, 2015.

[40] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, S. Spiga, "HfO2-Based Memristors for Neuromorphic Applications," *In Proceedings of the 2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, Montreal, QC, Canada, pp. 22–25, 2016.

[41] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Shimakawa, et al., "An 8 Mb Multi-Layered Cross-Point ReRAM Macro with 443 MB/s Write Throughput," *IEEE J. Solid-State Circuits 2013*, No. 48, pp. 178–185, 2013.

[42] Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, K. Kawai, S. Muraoka, S. Mitani, S. Fujii, K. Katayama, M. Iijima, et al, "Highly Reliable TaOx ReRAM and Direct Evidence of Redox Reaction Mechanism," *In Proceedings of the 2008 IEEE International Electron Devices Meeting*, pp. 15–17, 2008.

[43] L. O. Chua and L. Yang, "Cellular Neural Networks: Theory," *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 10, pp. 1257–1272, 1988.

[44] L. O. Chua and L. Yang, "Cellular Neural Networks: Applications," *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 10, pp. 1273–1290, 1988.

Publication List

Journal Paper

1. Hiroya Ikeda, Hiroki Yamane, Yuta Takishita, Mutsumi Kimura, and Yasuhiko Nakashima "Influence of Characteristic Variation of Oxide Semiconductor and Comparison of the Activation Function in Neuromorphic Hardware", *Nonlinear Theory and Its Applications(NOLTA), IEICE*, Vol. E11-N, No. 2, Apr. 2020, to be published

International Conference

2. Hiroya Ikeda, Hiroki Yamane, Mutsumi Kimura, Yuki Shibayama and Yasuhiko Nakashima "Evaluation of Neuromorphic Hardware using Cellular Neural Networks and Oxide Semiconductors", *2019 IEEE International Parallel and Distributed Processing Symposium Workshop on Advances in Parallel and Distributed Computational Models(IPDPS-APDCM)*, pp. 603–608, 2019.

3. Hiroya Ikeda, Hiroki Yamane, Yuki Shibayama, Mutsumi Kimura, and Yasuhiko Nakashima "Evaluation of Letter Reproduction System using Cellular Neural Network and Oxide Semiconductor Synapses by Logic Simulation", *CANDAR 2018 : The Sixth International Symposium on Computing and Networking*, pp. 145–147, Nov. 2018.

Co-Auther

4. 山川 大樹, 柴山 友輝, 大西 祐輝, 池田 裕哉, 中島 康彦, 木村 睦 "アモルファス In-Ga-Zn-O 薄膜シナプスの可塑性", 薄膜材料デバイス研究会 第 16 回研究集会, pp. 65-67, 2019 年 11 月

5. 木村 睦, 田中 遼, 山川 大樹, 柴山 友輝, 池田 裕哉, 滝下 雄太, 中島 康彦 " 局所学習則と薄膜デバイスを用いるリアルニューロモーフィックシステム - 文字学習の動作確認 -", 第 80 回応用物理学会秋季学術講演会, 20a-F211-7, pp. 19-26, 2019 年 9 月

6. 木村 睦, 生島 恵典, 杉崎 澄生, 田中 遼, 山川 大樹, 山根 弘樹, 池田 裕哉, 中島 康彦 “局所学習則と薄膜デバイスを用いるリアルニューロモーフィックシステム”, 第 66 回応用物理学会春季学術講演会, 11p-W810-13, pp. 18-24, 2019 年 3 月.