JOURNAL OF APPLIED PHYSICS **124**, 151903 (2018)

Perspective: A review on memristive hardware for neuromorphic computation

Changhyuck Sung, Hyunsang Hwang, and In Kyeong Yoo[a)]
Center for Single Atom-based Semiconductor Device and Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea

(Received 28 April 2018; accepted 31 August 2018; published online 5 October 2018)

Neuromorphic computation is one of the axes of parallel distributed processing, and memristor-based synaptic weight is considered as a key component of this type of computation. However, the material properties of memristors, including material related physics, are not yet matured. In parallel with memristors, CMOS based Graphics Processing Unit, Field Programmable Gate Array, and Application Specific Integrated Circuit are also being developed as dedicated artificial intelligence (AI) chips for fast computation. Therefore, it is necessary to analyze the competitiveness of the memristor-based neuromorphic device in order to position the memristor in the appropriate position of the future AI ecosystem. In this article, the status of memristor-based neuromorphic computation was analyzed on the basis of papers and patents to identify the competitiveness of the memristor properties by reviewing industrial trends and academic pursuits. In addition, material issues and challenges are discussed for implementing the memristor-based neural processor. *Published by AIP Publishing.* https://doi.org/10.1063/1.5037835

I. INTRODUCTION

As computer performance has continued to improve, artificial intelligence (AI) has attracted renewed attention. Along with the development of machine learning, AI services are growing as a new industry catering to bit data resources.[1] Semiconducting materials are at the base of the value chain for the computer hardware on which these advances rely.

Artificial Neural Network (ANN) algorithms offer fast computations by mimicking the neuronal network of brains.[2] A weight matrix is used in neural networks (NNs) for parallel processing that makes computing faster. Most of the commercially available AI chips are actually accelerators[3] and not neuromorphic processors. Some companies pursue the development of Graphics Processing Unit (GPU)-based accelerators, Field Programmable Gate Arrays (FPGAs), or Application Specific Integrated Circuit (ASICs) for effective AI services such as pattern recognition. However, the chip price of FPGAs, for example, is still relatively high, and hardware competition will focus on fast computation, low power consumption, small footprint size,[4] as well as low manufacturing cost.

The memristor has attracted much attention because of its potential to have linear multilevel conductance states[5,6] for vector-matrix multiplication (output = weight × input), corresponding to parallel processing. However, software also continues to improve with central processing unit (CPU) and GPU resulting in higher speeds. Therefore, it is important for memristors to be positioned properly within the value chain of hardware.

A bottom up approach is considered when identifying the value chain, which starts from materials and extends to AI service levels, and a top down approach is considered in reverse. It is time to evaluate memristor's value using both approaches because unforeseen consequences may arise by one of the approaches. For example, in the long short term memory (LSTM) of the recurrent neural network (RNN) algorithm where a forget gate is used, a fast weight has been proposed that does not require erasing weight for the forget process.[7] This means that the synaptic weight need not be nonvolatile. Such short term memory opens a new opportunity for memristors. The fast weight, however, may motivate a new DRAM-based product, too, for DRAM may be used as fast weights. In such a competitive landscape, it is necessary to analyze both threat and opportunity factors of memristors to take suitable and best action.

Some review articles on resistive switching material-based neuromorphic computation have presented useful guidelines. Yu has reviewed algorithms, architectures, and material properties in broad view.[8] Kuzum, Yu, and Wong dealt with material issues that are appropriate for biological synapse characteristics.[9] Burr *et al.* reviewed hardware from an implementation viewpoint.[10] In addition to these review points, it is desired to review the effectiveness of the memristor-based hardware in training and learning. Take back-propagation, for instance, IBM fabricated transposable 8T SRAM in TrueNorth to run a back-propagation algorithm. We need to review how transposable resistive switching random access memory (RRAMs) are being studied for on-chip training and learning. We also need to understand trends of device development to identify the competitiveness of memristors in comparison with other candidates. It is our intention to propose a direction to explore and improve the properties of memristors through this review.

II. MEMRISTOR-BASED NEUROMORPHIC COMPUTATION

The following technologies are being studied and under development as candidates for next generation computers.[11]

[a)]Author to whom correspondence should be addressed. Electronic mail: inyoo@postech.ac.kr

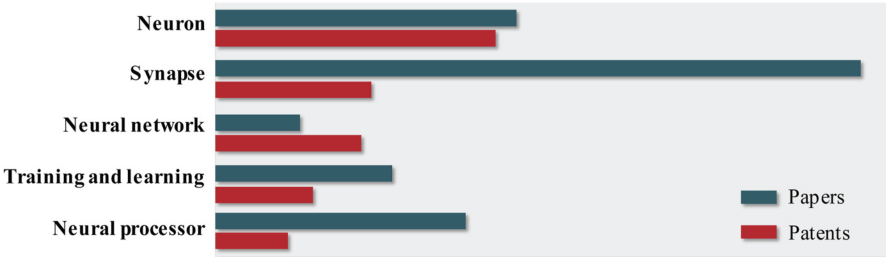

FIG. 1. Publications on memristor-based neuromorphic computation.

Neuromorphic computing and open platforms are motivated by "beyond Moore's law" and machine learning.[4] Google's Tensor Processing Unit (TPU) is one of the open platforms for deep learning.

- Reconfigurable Logic
- Memory-Centric Processing
- Silicon Photonics
- Neuromorphic Computing
- Quantum Computing
- Analog Computing
- Open Platforms

One of the main functions of accelerators is matrix multiplication. The main computation part in Google's TPU is a matrix multiply unit.[12] Memristors are suitable for the node of matrix multiplication because of their multilevel resistance. However, memristors should be suitable for supervised training/learning on chip in order to predominate over the CMOS-based neural network.

A. Memristor terminology

We need to agree on the terminology of the memristor before discussing the memristor-based hardware. A memristor is "a contraction for memory resistor."[13] It has two properties, a charge-controlled memristor, $v(t) = M[q(t)]i(t)$, and a flux-controlled memristor, $i(t) = W[\varphi(t)]v(t)$.[13] Therefore, memristor material shows the relationship between memristance and memductance, $M(q) = 1/W(\varphi)$. Biolek *et al.* reported that the HP's memristor was not a true memristor but a type of a current controlled memristive system (CCMrS).[14] Vongehr and Meng published that memristors were not yet found.[15] Serrano-Gotarredona *et al.* defined the memristor as a "two-terminal electronic device which is similar to a resistor, but whose resistance changes dynamically as the device is being used."[16] In this paper, we follow the definition of Serrano-Gotarredona *et al.* that includes resistive switching.

The resistive switching includes threshold switching and memory switching with several switching mechanisms. Threshold switching has two types, that is, current controlled negative resistance (CCNR) and voltage controlled negative resistance (VCNR).[17] It was reported that memory switching is driven by power.[18]

B. Neuromorphic computing devices

Since neuromorphic computation imitates a biological brain, each part of the neuronal network is modeled and implemented into hardware to run machine learning algorithms. There are neural processors fabricated by full CMOS technologies based on neuromorphic models. Memristor based-neuromorphic hardware is also studied in relation to both off and on-chip learning. For example, there is a single spike and oscillating spikes generated by utilizing memristor's threshold switching.[19,20] The memristive synaptic weight stems from the memory switching property.[21–24] There are also memristive logic[25] and memristor based-recognition chip.[26] Memristor-based neurons, memristor synaptic weights, and memristor-based training/learning are reviewed in Sec. II.

C. Papers and patents related to memristor-based neuromorphic computation

The patent analysis tool, LexisNexis PatentStrategies™, was used in searching patents related to a memristor-based neural processor. Appropriate patents were selected from the searched data. Each patent was sorted into the fields of neuron, synaptic weight, neural network, training/learning, and neural processor. Published papers were also selected and sorted similar to that of the patent search. When searching patents with a keyword of "memristor based neural processor," memristor, memristive materials, or resistive switching materials occupy a large part of the patent scope. Narrowing the scope of the patent by using "memristor neuromorphic computation, memristor neural network, or memristor neural circuit" gives rise to a relative distribution as shown in Fig. 1. The number of patents and papers is updated every moment and it is practically impossible to show exact numbers, so that the relative sizes were made as done in Schuman's review article.[4]

The reason that memristor synapse papers are dominating is that many memristor memory papers deal with synaptic weight, showing that researchers are giving top priority in achieving synaptic properties. Publication numbers of patents and papers are nearly equal in memristor neuron and memristor-based training/learning. This is because many of these papers were also filed into patents. The patent filing numbers decrease as the topic moves from the device level to the system level.

The trends of memristor-based neuromorphic computation are summarized as follows when considering papers, patents, and company status:

- The research on memristive memory (storage) has been expanded to synaptic weights.
- The portion of materials and devices is large in the neuromorphic patent portfolio.
- The neural processor (or AI chip) becomes specialized or dedicated by FPGA or ASIC.

TABLE I. Comparison of CMOS neurons and memristor-based neurons.

Models		Configuration	Power or Energy/spike	Reference
CMOS	Leaky integrate-and-fire	18–20 transistors	0.3–1.5 μW, 2850 pJ/spike	Indiveri[b]
		16 transistors[a]	40.2 pW, 0.4 pJ/spike	Cruz-Albrecht *et al.*[c]
		14 transistors	4.3 pJ/spike	Shamsi *et al.*[d]
	Morris-Lecar	9 transistors[a]	4 fJ/spike	Sourikopoulos *et al.*[e]
	Hindmarsh-Rose	90 transistors	163.4 μW	Lee *et al.*[f]
	Wijekoon	14 transistors	8–40 μW	Wijekoon-Dudek[g]
	Babacan	1 memristor emulator (operational transconductance amplifier OTA + multiplier) + 3 transistors	60–110 μW	Babacan[h]
CMOS + Memristor	Saxena	Memristor emulator (8 transistors[a])	14 fJ–1.4 pJ/spike	Saxena *et al.*[i]
	(Oscillatory)	1 memristor + 1 magnetic junction + CMOS circuit	3.3 μW, 150 pJ/junction	Mizrahi *et al.*[j]
	(Stochastic neurons)	1 memristor + CMOS circuit	249 fJ/single write@50% switching probability	Wijesinghe *et al.*[k]

[a]Number of transistors was estimated according to circuits in each reference. Memristor-based neuron in this table is defined as CMOS circuits that include memristive parts.
[b]Ref. 39.
[c]Ref. 40.
[d]Ref. 41.
[e]Ref. 42.
[f]Ref. 43.
[g]Ref. 44.
[h]Ref. 45.
[i]Ref. 46.
[j]Ref. 49.
[k]Ref. 50.

- As the amount of data increases, deep learning algorithms are effective and deep neural network (DNN) is getting applied even to mobile services.

Data scattering and reliability issues in memristor synaptic weights are obstacles for commercialization of memristor-based hardware. This is because switching data itself shows an intrinsic statistical variance. For example, the number of conducting paths occurring during switching shows Poisson distribution,[18] where randomness cannot be controlled. No reliable nonvolatile linear multilevel memristor has been reported yet. A group of memristors was tried to make a multi-bit or multilevel synaptic weight,[27,28] which may contribute to reducing multilevel data variance.

D. Memristor neurons

Neuron models are classified into a biologically plausible model and a biologically inspired model. The bio-plausible model mimics a biological neural system, and the bio-inspired model exploits the characteristics of a biological neural system. Many transistors are needed when fabricating neurons by CMOS technology only. Memristor-based neurons were proposed to replace some CMOS devices to simplify circuits. Memristor-based Hodgkin-Huxley,[29] memristor-based Morris and Lecar,[30] memristor-based FitzHugh-Naguma,[31] and memristor-based Hindmarsh-Rose[32] were reported by simulating their signals. There is also a memristor-based simple spiking model and integrate-and-fire.[33] Al-Shedivat *et al.* simulated a memristor-based stochastically spiking neuron.[34] They proposed the enhanced analytical model of the memristors. Shamsi *et al.* designed an analog modular neuron based on the memristor.[35] This memristor was also simulated with a linear model for the Pt/TiO_2/Pt device. Mehonic and Kenyon observed the threshold voltage spiking/instability by applying threshold current into SiO_2 which is unipolar switching memory.[36] Pantazi *et al.* incorporated phase-change memristors into the architecture implementing the integrate-and-fire functionality of the neurons as well as the plasticity of the synaptic elements.[37] The sets of level-tuned neurons demonstrated selectivity related to the input signals. They presented how the single-neuron building block of a spiking neural network (SNN) can be realized with nano-scale phase-change devices in all-memristive configuration; however, open issues remain to be addressed related to interconnectivity and the integration of the memristive components in a neuromorphic processor chip.[37] Teimoori *et al.* used memristor logic to fabricate integrate-and-fire neurons, by replacing resistors of CMOS neurons with memristors to obtain a single pulse or pulse train.[38] It is noted that a CMOS transistor can emulate the memristor,[46–48] but memristors cannot fully replace CMOS transistors or CMOS circuits. Instead, a hybrid approach is employed in which a memristor models a biological synapse, while CMOS circuits model neuronal dynamics as Mehonic suggested.[36]

Some authors demonstrated energy consumption of CMOS neurons. Table I presents CMOS neurons and memristor-based neurons with power consumption and/or energy consumption per neuron spike. A combination of a memristor node with CMOS circuits and memristor node with transistors is also presented.

Some power consumption should be noted in Table I. The Wijekoon-Dudek model using 14 transistors produces all types of spiking and bursting like Babacan's method but consumes

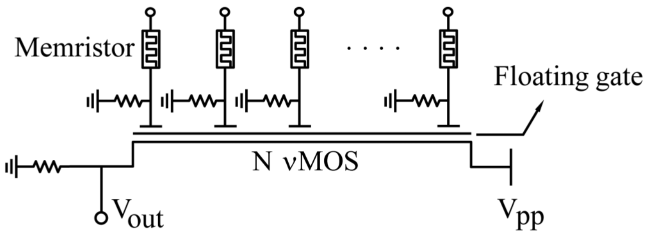

FIG. 2. A linear multilevel synaptic device that takes advantage of floating gate. This may be an example of activation function devices.

40% of the power of Babacan's memristor neuron. Therefore, trade-off among chip size, computing speed, and power consumption should be considered, and this may be determined by application. Deng *et al.* analyzed energy consumption under different learning stages.[51] Perhaps, the most promising neuron in Table I is Sourikopoulos' Morris-Lecar from an energy viewpoint. But, speed and chip size should also be considered in hardware architecture depending on AI applications and trade-offs may be required for a specific chip design.

A memristor can be used for output signals as well as input signals. There is a patent (Fig. 2) that generates multi-level synaptic weight signal using threshold switching of the memristor.[52] The neuron MOS (νMOS) transistor, which is the original concept of the patent CN103324979, was introduced earlier for parallel processing.[53] The linear multilevel synaptic weight is achieved in this device when memristors connected to gates of the νMOS transistor are used as a group of single bit memories.

Memristors may be used for various devices in neural networks, i.e., neurons and synapses as well as neuronal circuits. Al-Shedivat *et al.* generated a spike by applying the memristor to neurons and synapses and ran the winner-take-all (WTA) algorithm in the SNN[34] and determined the synaptic weight by Spike Time Dependent Plasticity (STDP) learning. It is true that the memristor replaces some of the CMOS neural circuits, but the memristor becomes competitive only when it significantly improves neural network performance or reduce chip size compared to CMOS neural networks. The performance of memristor-based neural networks has been predicted mainly by simulation. It is therefore desired to demonstrate a breakthrough in memristor characteristics.

E. Memristor synapses

The challenging issues in memristor synaptic weights are nonvolatility, linearity, and multilevel. However, the results satisfying these three properties simultaneously have not yet been obtained. A number of patents and papers on neuron, synapse, architecture, training, and learning have been published with many efforts to have analog memory characteristics. Figure 3 shows a CMOS integrated-and-fire neuron generating a neuron pulse.[54] The memristor is placed between the input neuron and the output neuron, and a memristor-based synaptic weight crossbar is formed. Each memristor in the crossbar is trained by a STDP learning rule.

The synaptic weight is determined in STDP learning by the difference between pre-neuron spiking time and post-neuron spiking time. The synaptic weight gives nonlinear values in this case, and it is generally applied to unsupervised training/learning with the winner-take-all (WTA) algorithm such as position detection.[55] Zheng and Mazumder have therefore proposed to develop a hardware friendly algorithm rather than to develop hardware to fit the algorithm and demonstrated weight dependent supervised STDP learning.[56,57]

The linearity of synaptic weight is highly required in deep learning where vector matrix multiplication (VMM) is applied for parallel processing in the neural network. In general, the pulse train is applied to the input node and a linear increase in potentiation and a linear decrease in depression of conductance are required for the memristor in VMM processing. Symmetry between potentiation and depression is also crucial for learning in the neural network. Table II summarizes the mechanisms that determine some memristive switching types of materials. No unipolar switching has been reported so far that shows multilevel switching during potentiation and depression simultaneously. The bipolar switching, even though it may not be symmetric, gives multi-levels in both potentiation and depression. Organic materials, magnetic materials, and other oxides such as ZnO may show synaptic properties; however, full information of multi-level, symmetry, and/or on-off ratio was not reported.

The unipolar switching may have the same mechanism as that of bipolar switching in some cases. NiO_x, for example, has unipolar switching characteristics, but bipolar switching and anti-bipolar switching have been observed so

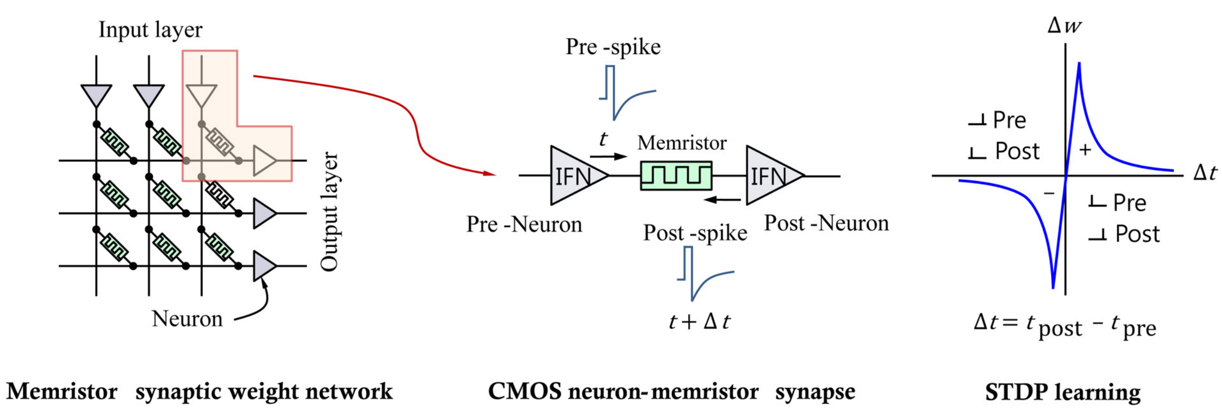

FIG. 3. Crossbar SNN architecture with memristor synapses, a synapse connected between two spiking neurons showing pre-synaptic spike and post-synaptic spike, and graphical depiction of a bio-inspired pair-wise STDP-learning rule. Partially adapted from Ref. 54.

TABLE II. Multilevel memristive synaptic materials. Off current level and current resolution of multilevel should be considered in field operation. Values with asterisk are reported in references. Other values were estimated based on data in each reference. Incremental voltage pulses are applied to ferroelectric switching for making multilevel while constant voltage pulse train is applied to other materials for potentiation and depression.

Source of switching mechanisms	Materials system	Multi-level	Symmetry between potentiation and depression	On-off ratio	Off-current	Ref.
Electrochemical filament-based resistive switching	Ag/Pd/SiGe	100	Symmetric	100*	10 nA	58
	Ag/AgInSbTe	50	Asymmetric	2	800 μA	21
	Ag/Si	100	Symmetric	10	5 nA	6
Oxygen vacancy filament-based resistive switching	HfO_2/AlO_x	40	Symmetric	3*	1 μA	22
	TaO_x/HfO_x	100	Asymmetric	5	1 μA	59
	SiO_2/TaO_x	300	Symmetric	2	40 μA	23
	Ta_2O_5/TaO_x	20	Asymmetric	2	40 μA	60
Interface resistive-based switching	$Ta/TaO_x/TiO_2/Ti$	50*	Asymmetric	2	7 nA	61
	Mo/PCMO	32*	Asymmetric	15	500 pA	62
	Al/Mo/PCMO	100	Asymmetric	100	10 pA	24
	Mo/TiO_x	64*	Asymmetric	20	1 nA	63
	WO_x	100	Asymmetric	100	20 nA	64
	TiO_x/TiO_y	100	Symmetric	10	1 nA	65
Ferroelectric tunneling	BTO/LSMO	100	Asymmetric	10	10 μA	66
Ferroelectric switching	HZO	32*	Symmetric	45*	1 μA	67

that unipolar switching is found to be a part of these double curves.[68] This can be described by a schematic switching model as shown in Fig. 4. There are many switching mechanisms from soft breakdown through nano-filaments in resistive switching materials. In Fig. 4, a typical filamentary model was applied as an example. A similar unipolar switching with double bipolar switching may be presented.

The analog switching material may be volatile, but when pulse rate, width, and voltage are optimized, the pulse train can achieve linear potentiation even to volatile synaptic weight.[64] Therefore, on chip training/learning may be performed during the period when retention loss occurs slowly. However, precise modulation of the device conductance over a wide dynamic range may be necessary with linearity to maintain high network accuracy. In such a synapse, the synaptic weight may be represented by the combined conductance of multi-cells.[27] Irmanova and James designed 10 levels of synaptic weight by combining three sub-memristor cells where each memristor of the cell is placed into sub-cells.[28]

F. Memristor-based learning

Schuman *et al.* commented that perhaps the most popular on-line, unsupervised learning mechanism in neuromorphic systems is STDP.[4] STDP-based unsupervised learning has been proposed mainly for binary synapses,[69] and Covi *et al.* proposed an HfO_2-based analog memristor as a synaptic element which performs STDP within a small spiking neuromorphic network operating unsupervised learning for character recognition.[69] Zheng and Mazumder also pointed out that many of the spiking neural networks (SNNs) do not have the capability to conduct on-chip learning.[56] The training is performed in advance using a computer or server for off-chip learning, and then memristor synaptic weight information is stored separately. In this case, the weight information can be stored sequentially in columns or rows of the weight matrix so that the control circuit can be simplified as compared to that for simultaneous storage. Inference by unsupervised training can be effective in mobile AI services where simplicity and speed are crucial. Recently, there have been reports of both on-chip unsupervised learning[70] and on-chip supervised learning by simulation.[71] We must be able to update the weight by accessing each synaptic weight randomly, independently, and directly during on-chip learning in order to perform learning in real time as soon as data arrive. Synaptic weights should be accessed simultaneously for perfect random access. But this operation requires more circuit lines. For example, in the case of the 2×2 1T-1R synapse arrays as shown in Fig. 5(a), it is possible to access one cell, two cells, and four cells at the same time and randomly, but it is impossible to access three cells simultaneously. Thus, a separate and additional word line or bit line is required for each cell for perfect random access. This makes the circuit overhead increase. Since the circuit overhead should be minimized in order to reduce the chip size, we have to accept some degree of sequential processing when updating the synaptic weights even during *in situ* on-chip learning. Consequently, the data processing must be fast for on-chip learning, but fast processing also increases the circuit overhead.

1. STDP learning

Both unsupervised STDP learning[72–74] and supervised STDP learning have been reported.[56,57,75,76] Pedretti *et al.* presented unsupervised STDP learning with a memristor synapse where synaptic weights are updated by STDP.[72] They discussed applications of unsupervised techniques such as data clustering and anomaly detection. Ly *et al.* trained the neural network with a stochastic STDP.[73] In this work, a visual pattern extraction application, they fully connected the network of Leaky-Integrate and Fire (LIF) neurons and RRAM-based synapses.

Nishitani *et al.* reported that STDP supervised learning can be performed using ferroelectric memristors.[77] This is

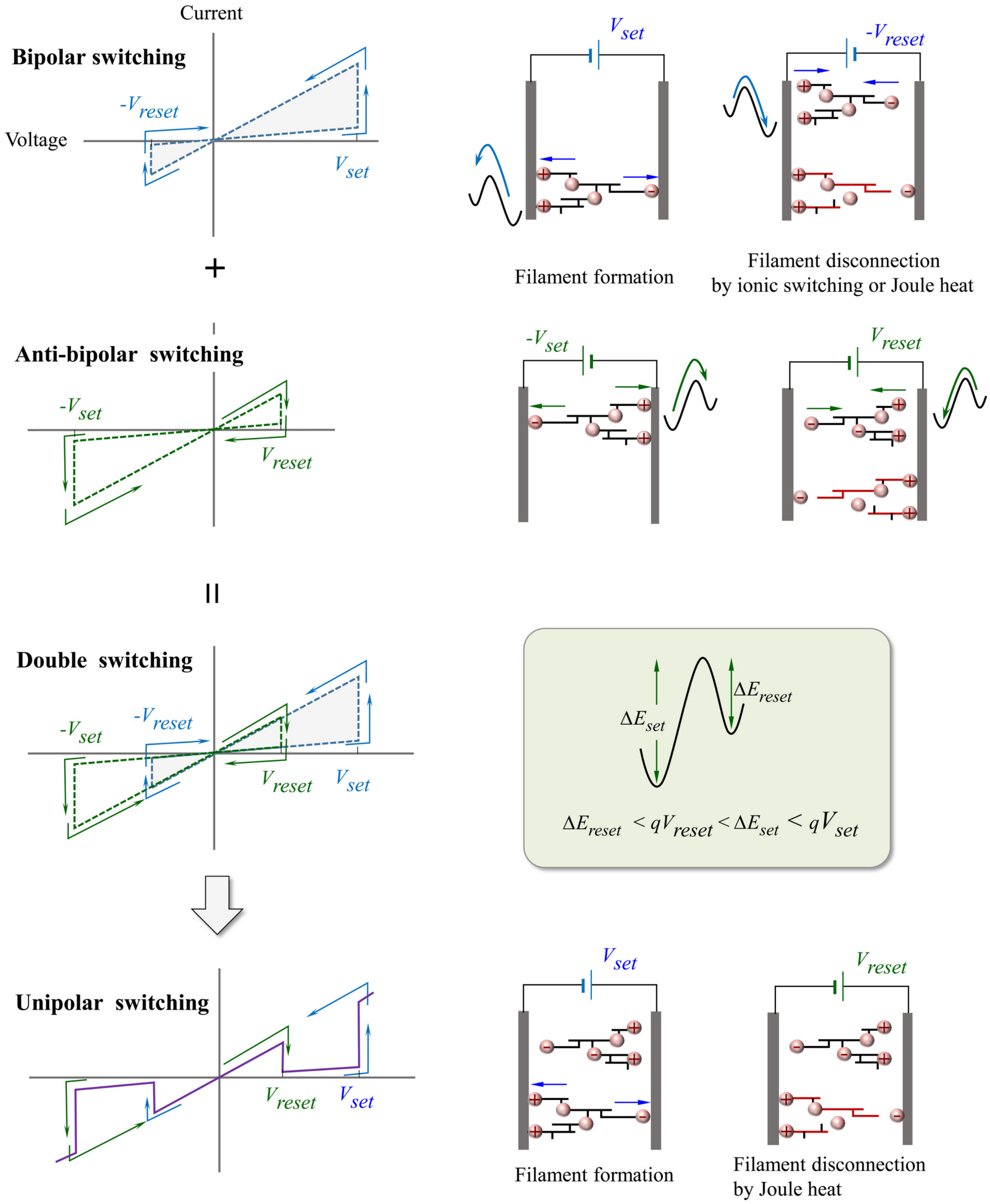

FIG. 4. Analysis of correlation between double switching curve and unipolar switching curve in NiO_x thin films.

because the ferroelectric has polarization and polarization reversal property. The ferroelectric is polarized in the positive direction during forward propagation and in the opposite direction during backpropagation. Positive polarization corresponds to excitatory postsynaptic potential (EPSP), and negative polarization corresponds to inhibitory postsynaptic potential (IPSP). This bi-stable synaptic weight improves the dynamic range of weight as discussed in Sec. III A.

2. Backpropagation circuits

The back-propagation algorithm is carried out when correcting errors in the neural network during supervised learning. Select transistors are connected to synaptic weights to update them randomly in hardware. This operation should be possible not only in forward propagation but also in backward propagation. Figure 6(a) shows the features how to access synaptic weights during both forward propagation and backpropagation. It is seen that the weight

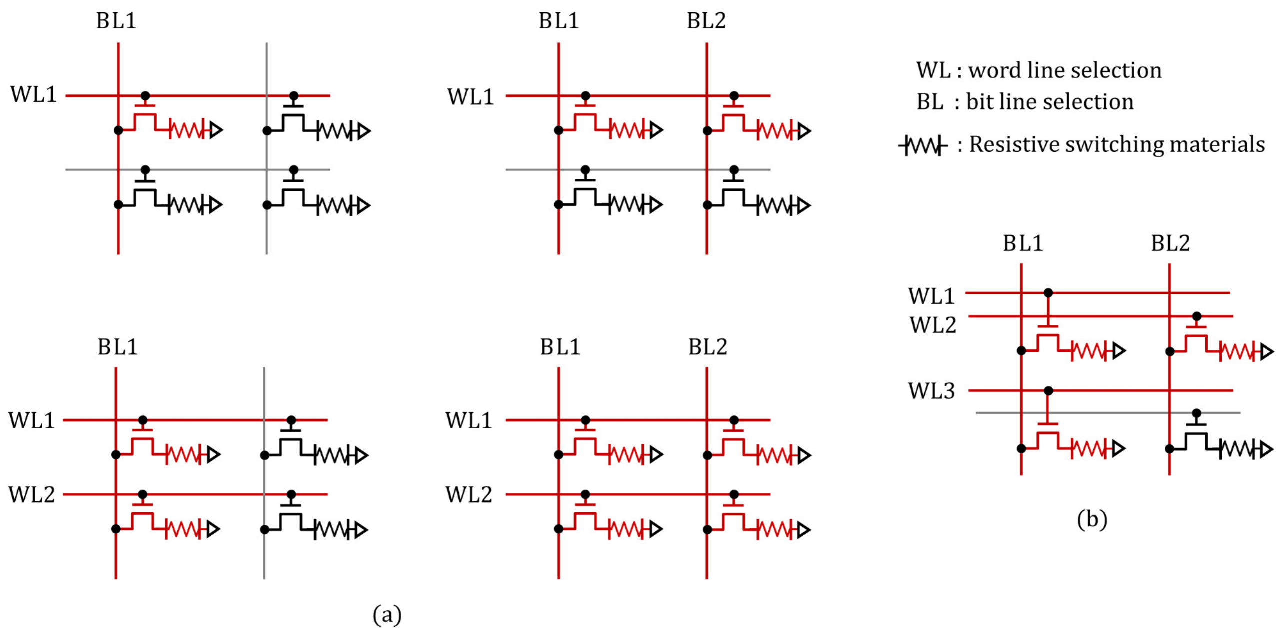

FIG. 5. Random access of synaptic weights. Simultaneous random access requires circuit overhead.

matrix of the forward direction and that of the backward direction make a transpose relationship, $\boldsymbol{W}$ and $\boldsymbol{W}^{\mathbf{T}}$. Each weight should be randomly accessible in both forward and backward directions for both matrices. In the case of 1T-1R memory such as RRAM, a bit line and a plate line are placed in parallel as shown in Fig. 6(b), and the word line is perpendicular to both the bit and plate lines. Thus, random access is possible in both forward and backward directions in the memory array. However, the input bit line and the output line are placed perpendicular to each other in

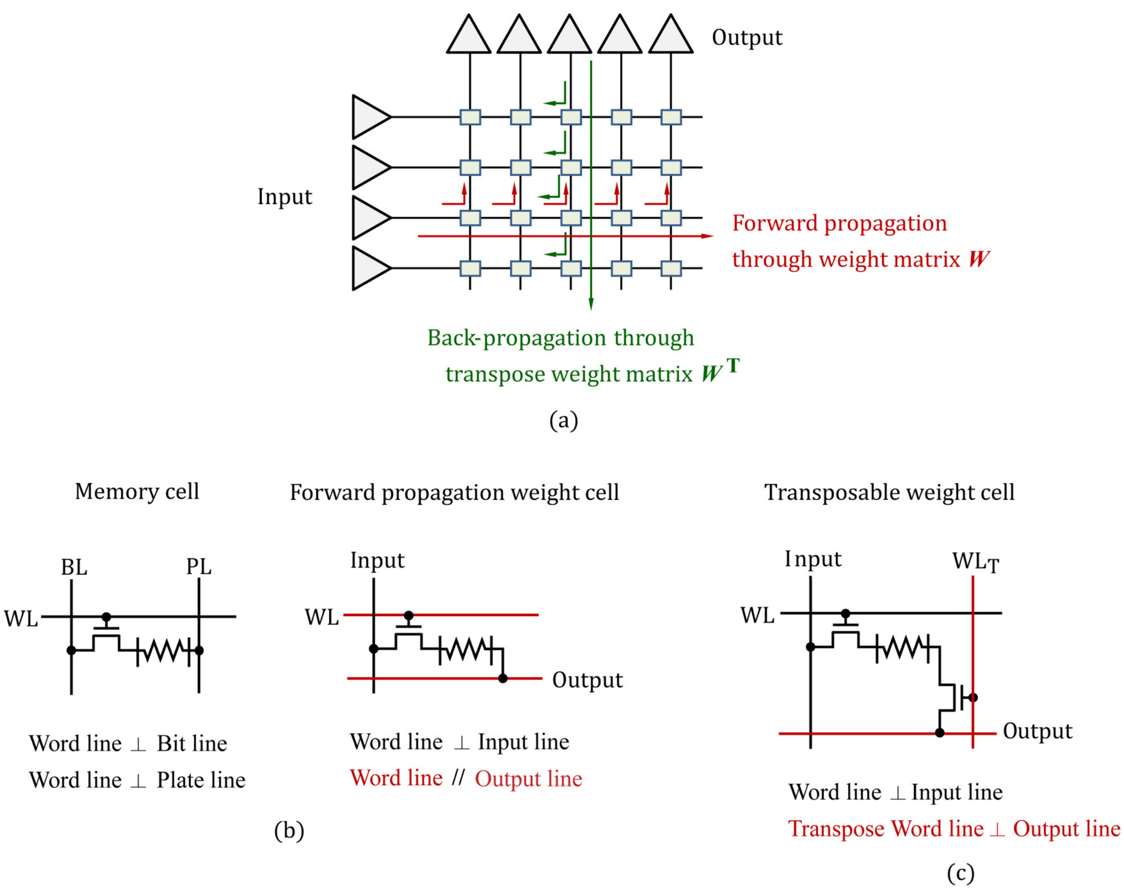

FIG. 6. Transposable synaptic weight 2T-1R for backpropagation circuit.

the neural network. Thus, an additional transpose word line WL_T is required that is perpendicular to the output line for backward propagation as shown in Fig. 6(c). IBM connected two select transistors to the SRAM synaptic weight to enable backpropagation and called it transposable memory.[78] In the case of the memristor, we use two select transistors making 2T-1R as shown in Fig. 6(c), which we call transposable synaptic weight. Another IBM patent gives an example of transposable weight, a phase change material (PCM), with two select transistors.[79]

As for energy consumption between off-chip unsupervised STDP and backpropagation, Deng *et al.* analyzed energy consumption by simulation for various memristive networks under different learning strategies.[51] An order of nJ energy consumption for STDP and μJ for neural network is estimated in their article.

G. Memristor-based neuromorphic chip

Shafiee *et al.* analyzed power consumption in memristor-based vector matrix multiplication.[80] Bayat *et al.* demonstrated a classifier equipped with memristor perceptron.[81] knowm® has also released a classifier product using anti-Hebbian and Hebbian rules using a binary switching.[82] Memristor-based neuromorphic computation shows limitations, especially in the dynamic range of the synaptic weight during on-chip learning. The wide bit-width of synaptic weight is required even in off-chip learning for best performance. It is practically impossible for memristor to match with 16 bit width or 64 bit width of synaptic weight which is not so unusual in software-based learning. Accordingly, data compression or pruning techniques have been proposed by preventing the AI function from being damaged when receiving learning information on a mobile device.[83] In addition, there is an example of fine tuning technique that performs on-chip learning for *in situ* optimization.[81] Bayat *et al.* trained perceptron to classify a stylized letter pattern using four alternative approaches as shown in Table III. In their demonstration, some stages of *in situ* training were assisted by an external computer. Table III compares the pros and cons of combinations in on-chip learning and off-chip learning. They pointed out that a potential drawback of a defect-aware *ex situ* scheme is that the chip-specific precursor training may not be suitable for some applications, e.g., when training takes too much time. In the light of such limitations, the mobile neural processor may become a specialized and a dedicated ASIC, but reconfiguration may be required to some extent.

H. Discussion on Sec. II

Analog property of the memristor was applied to the Hodgkin-Huxley neuron at first, and multilevel RRAM became one of the candidates for synaptic weight. Yu summarized[8] the guidelines of synaptic weight properties such as linearity, bit-width, nonvolatility, lifetime, etc. Even though memristors including resistive switching materials, metal-insulator transition (MIT) materials, and others have potentials to make memristor neurons, memristor synapses, and even memristor logic, satisfactory candidates have not yet been developed. Instead, most memristor-based neuromorphic computing is demonstrated mainly by simulation. Nevertheless, the multilevel conductance property of memristor still motivates the development of new algorithms and chip architectures in addition to the material property itself. That is the reason why materials science and engineering such as switching mechanisms should be studied more rigorously in order to control the conductance level, even at the quantized scale, for example.

Accuracy, speed, size, and power will be issued continuously in AI chips for applications. The top priority for mobile AI chip may be speed and low power, for now. Then, the mobile device will take over minimal AI functions with the help of the main server or computer in training and learning as suggested by Bayat *et al.*[81] Unsupervised learning is useful and has many applications; however, supervised learning is also one of the social needs when considering various AI services. Then, memristor-friendly algorithms such as Mazumder's weight dependent STDP[57] may become one of the main streams in the near future.

TABLE III. Training approaches to cope with imperfect hardware.[81]

Training approach	Training steps	Pros	Cons
Ex situ	Step 1. Precursor training Step 2. Weight import to HW	Lowest HW overhead	• Poor imperfection tolerance/fidelity • Off-line learning
Defect-aware ex-situ	Step 1. HW test Step 2. Precursor training Step 3. Weight import to HW	Best imperfection tolerance/fidelity Low HW overhead	• Poorly scalable step 1 (HW test) • Off-line learning • Chip specific training
In situ	*In situ* training on HW	Suitable for on-line learning	• High HW overhead • Sub-optimal fidelity • Long training times • Chip-specific training
Hybrid	Step 1. Precursor training Step 2. Weight import to HW Step 3. *In situ* training on HW	Best imperfection tolerance/fidelity for on-line learning	• High HW overhead

III. COMPETITIVENESS OF MEMRISTOR

Memristor may be able to make neurons and synaptic weights, but there are competing technologies available. CMOS-based neural processors rely on software and store the weight information in a separate storage, and they are reliable in neuromorphic computing. Therefore, in order to have dominating competitiveness of memristors, characteristics such as multilevel weight with reliability which cannot be obtained in any other competing technologies should be secured.

A. Memristor synaptic weight

Synaptic weight may have negative values during training and learning. These bi-weights are caused by EPSP and IPSP. It is usual for software to assign a dynamic range of weight having both positive and negative weight values. It can also use floating point with an unlimited weight bit width. However, a memristor has limited fixed point of weight with a narrow bit-width. It cannot have a negative value resistance, either. A floating gate transistor can have a positively induced channel when it is charged with electrons, but it cannot be charged with positive charges to make a negatively induced channel. Thus, it has been proposed that a group of memristors be used to make a bi-polarity weight. For example, a run-time programmable complementary bi-polarity synapse crossbar was reported in Ref. 57. The memristor bridge synapse using four memristors can have positive, negative, and zero weight values.[41,84]

On the contrary, ferroelectrics show intrinsic bi-stable memory due to positive and negative polarization. This makes bi-weight in a simpler cell. When a ferroelectric is deposited on the gates of both n-type metal–oxide–semiconductor (NMOS) and p-type metal-oxide-semiconductor (PMOS) transistors, the direction of the current flowing through the channel is determined by the polarization direction of the ferroelectric so that a positive weight and a negative weight can be distinguished. Figure 7(a) illustrates the working principle of bi-stable ferroelectric synaptic weight.[85] As for memristors such as resistive switching materials, circuits and operational scheme of the complementary crossbar are more complicated than that of the intrinsic bi-stable synaptic weight of the ferroelectric transistor.[56,86] For example, a positive voltage or negative voltage can be applied to the ferroelectric synaptic weight directly on ferroelectric transistors. But one of the memristor pairs should be set at the "off" state, when the other is written at a certain weight value.[56,87] Therefore, four memristors are required to make a memristive bi-weight, while two ferroelectric transistors are required to make a ferroelectric bi-weight as shown in Fig. 7.

Nonvolatility and relatively high weight bit-width are strong properties of ferroelectrics.[67,88,89] TFT type ferroelectric bi-weight was also patented for stacked structure of high density synaptic weight.[90] Even though the ferroelectric shows a nonvolatile bi-stable multilevel synaptic weight, this is still nonlinear. Fatigue in the ferroelectric is also a concern for reliability. Even though fatigue of some ferroelectric materials has been overcome by using a conductive interlayer between electrode and ferroelectrics, new interlayer materials may be required for synaptic ferroelectric materials such as HfO_x and HZO($HfZr_xO_y$).

B. Memristor vector matrix multiplication (VMM)

A conductance-based VMM using resistors [Fig. 8(a)] is registered in US 9,934,463.[91] A capacitance-based VMM scheme [Fig. 8(b)] was also filed earlier (US 5,146,542).[92] Memristors are elements of conductance-based VMM architecture. The DC power consumption issue was pointed out for conductance-based VMM. That is why capacitance-based VMM began to be considered recently[90] because it guarantees low power consumption with linearity. But parasitic capacitance such as bit line capacitance is unavoidable and one of the essential issues.

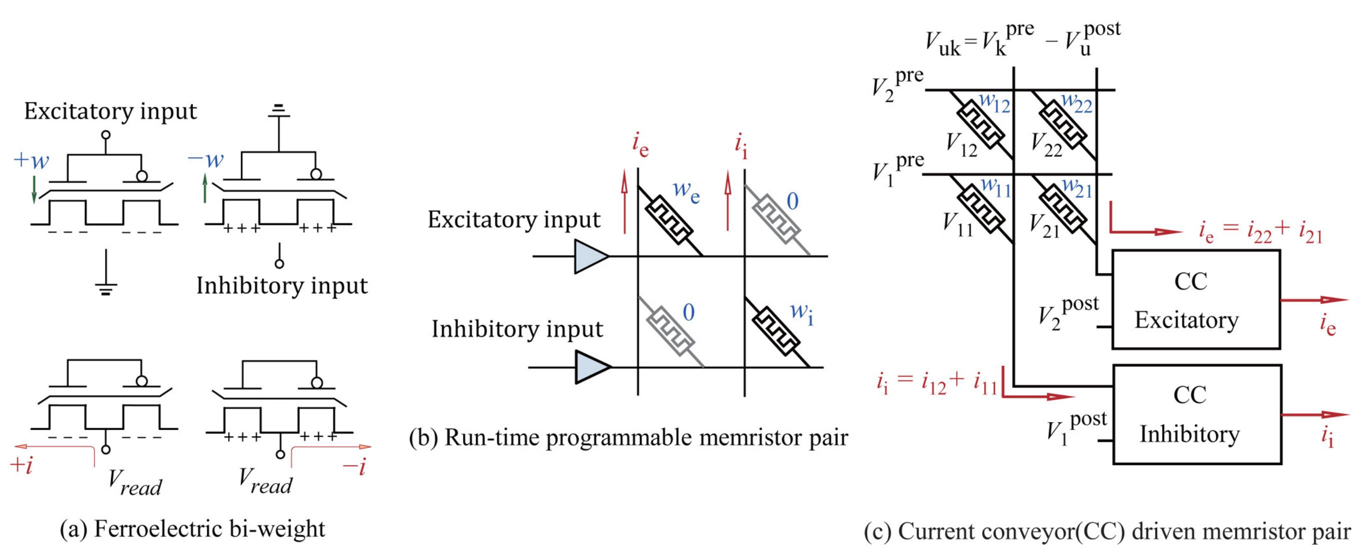

FIG. 7. Synaptic weights for excitatory and inhibitory inputs. (a) Operation of bi-stable ferroelectric synaptic weight. Polarity of weight is distinguished by current flow direction. Two ferroelectric transistors make a bi-weight.[85] [(b) and (c)] Memristor pair for a bi-weight. Excitatory and inhibitory currents flow separately in the same direction. Four memristors are required to make a bi-weight. Partially adapted from Refs. 57 and 86.

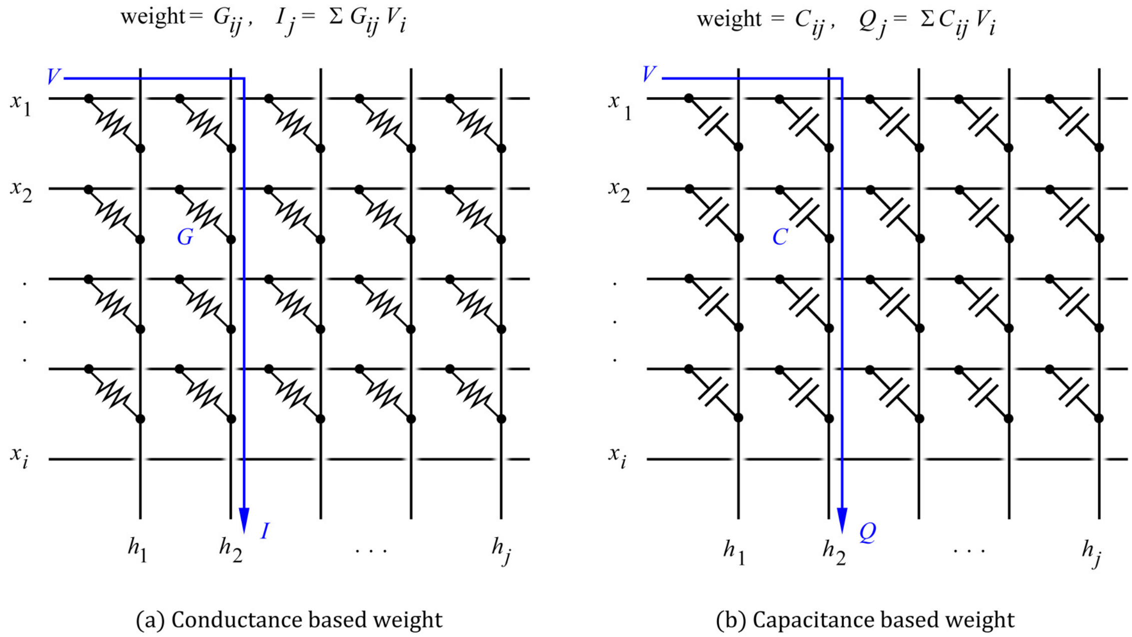

FIG. 8. Synaptic weights for vector matrix multiplication.

Just as the memristor needs multilevel, so the capacitor also needs multilevel. However, nonvolatile multilevel capacitance cannot be achieved so that the capacitor is charged by applying a pulse train to make multilevel-valued capacitance. In this case, AI services should be carried out in a short period of time before capacitors are discharged.[93] The ferroelectric synaptic weight is also capacitive switching and power consumption can be avoided during the write process (training). It is conductance-based multiplication when reading weight values of the ferroelectric synaptic weight, and parasitic capacitance can be avoided. Therefore, once ferroelectric weight is controlled linearly, it can be used as a nonvolatile multilevel synaptic weight, up to 5 bit-width, according to the report of Jerry *et al.*[67] If the bi-weight scheme is applied to the above Jerry's HZO synaptic weight, it will cover a dynamic range of ±5 bit (or 6 bit, 64 levels). But the ferroelectric synaptic weight is a three-terminal device in contrast to the memristors including the ferroelectric tunneling junction[66] that are two-terminal devices, which leads to sacrifice in chip size.

C. Memristor stacked crossbar

Both conductance-based VMM and capacitance-based VMM require high capacity structure such as stacked cross-point. It is the same case in storage. 1S-1R (1 selector-1 resistor) or 1D-1R (1 diode-1 resistor) synaptic weights are stacked, layer by layer, to make stacked cross-point structures. This structure is, therefore, a horizontal cross-point. The vertical cross-point in Fig. 9(a) is fabricated in a way similar to the NAND process. However, the vertical cross-point stack is suitable for 1D-1R only.

Figure 9(a) shows the capacitors connected to the diode line with conducting paths. It guarantees the linearity of the capacitive synaptic weight with sufficient multilevel with a wide dynamic range.[94] In this structure, an insulator or a

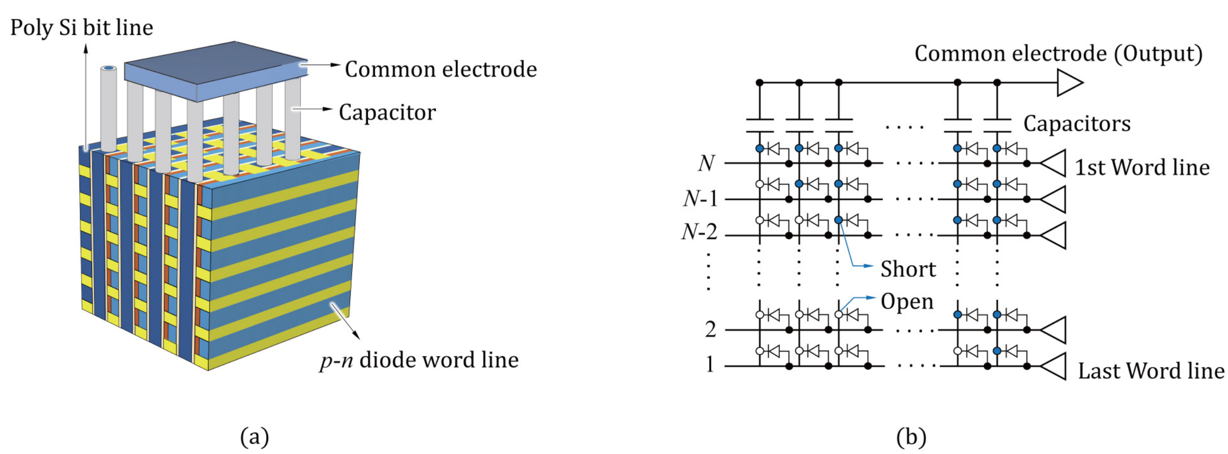

FIG. 9. Multiple capacitor based synaptic weights with vertical cross-point structure.

unipolar switching memristor is deposited between the vertical line and the horizontal line, and the point where the intersection of the vertical line and the horizontal line breaks down or the low resistance state (LRS) is made to allow current to flow. As shown in Fig. 9(b), the number of conducting points is formed on each horizontal word line, from 0 to N, in order. When selecting the corresponding word line during on-chip learning, as many capacitors are charged as the number of the vertical bit lines connected to the selected word line. This allows designating and updating the capacitive synaptic weight. Although this structure guarantees the linearity of the synaptic weight, the final weight information needs to be stored separately after training and learning. The circuit overhead is large, too. It is a matter of course that overhead can be reduced if nonvolatile linear multilevel memristors replace capacitors in Fig. 9.

D. Discussion on Sec. III

Three-terminal ferroelectric synaptic weight shows nonvolatile multilevel bi-weight compared to the two-terminal memristor. No detailed issues such as integration process and reliability on ferroelectric have been reported yet. However, special circuits that generate incremental voltage pulses, fatigue proof interlayer, and semi-conductive oxides may be required in order to realize stacked ferroelectric synaptic weights. Capacitive "write" and conductive "read" of ferroelectric synaptic weight are also attractive; however, it is hard to get linear weight values by applying incremental pulses. Therefore, the ferroelectric friendly algorithm may need to be developed.

Vertical cross-point synaptic weight may guarantee a linear wide dynamic range of synaptic weight, but it takes large space when integrated into the chip. But, this device will be useful when it is equipped with large systems such as a server. NAND and DRAM compatible process can be applied to the vertical cross-point structure. It is noted that the vertical cross-point matrix itself is storage. Memristors such as resistive switching materials are facing challenges to overcome issues of nonvolatility, multilevel, and linearity as well as lifetime. Therefore, hybrid structure such as vertical cross-point memristor synaptic weight may be one solution.

IV. SUMMARY

Neuromorphic computing was motivated by beyond Moore's law and machine learning leading to parallel distributed processing. Memristor-based vector matrix multiplication was proposed to satisfy this need. A parallel distributed architecture is also required even in a mobile application for fine-tuning when supervised learning is indispensable. However, since the parallel distributed processing makes the chip bulky, it has limited scalability. As a result, making multilevel with a wide synaptic weight bit-width is a fundamental breakthrough that overcomes the limit of scaling down. A development guideline of memristor synaptic weight is nonvolatility, linearity, and multilevel.

Resistive memory switching based on bipolar switching and unipolar switching has been found to coexist in the same switching material so that more detailed physical interpretation is necessary for memory switching mechanisms. Analog switching is not limited to threshold switching. Charge trap materials can also be used as an analog switching node.[95] It has been used as a floating gate, but it can be used for neuron and short term memory node. Thus, it is desired to develop new charge trap materials with various de-trap rates with corresponding physical interpretation.

No perfect memristor-based neuron is developed yet except for node such as MIT threshold switching and magnetic tunneling switching while CMOS-based memristors have been emulated. This implies that memristor will be adopted to the CMOS circuit for specific and special functionality such as analog switching node that may reduce circuit overhead.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers and the Associate Editor for their valuable comments and suggestions on improving the quality of this article. We also would like to thank David Seo for his sincere proofreading. This work was supported by the National Research Foundation of Korea (No. NRF-2018R1A3B1052693).

[1] J. Marous, "Artificial intelligence needs a strong data foundation," The Financial Brand, August 22, 2017, see https://thefinancialbrand.com/67039/ai-hierarchy-of-data-needs/.
[2] C. Mead, "Neuromorphic electronic systems," Proc. IEEE **78**, 1629 (1990).
[3] Z. Du, D. D. Ben-Dayan Rubin, Y. Chen, L. Hel, T. Chen, L. Zhang, C. Wu, and O. Temam, "Neuromorphic accelerators: A comparison between neuroscience and machine-learning approaches," in *Proceedings of 48th International Symposium on Microarchitecture (MICRO), Waikiki, HI*, 5–9 December 2015 (IEEE, 2015), pp. 494–507.
[4] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell. M. E. Dean, G. S. Rose, and J. S. Plank, "A survey of neuromorphic computing and neural networks in hardware," e-print arXiv:1705.06963, see https://arxiv.org/abs/1705.06963.
[5] M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, J. J. Yang, and R. S. Williams, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," in *Proceedings of the 53rd Annual Design Automation Conference (DAC), Austin, TX*, 5–9 June 2016 (ACM, 2016), pp. 1–6.
[6] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," Nano Lett. **10**, 1297 (2010).
[7] J. Ba, G. Hinton, V. Mnih, J. Z. Leibo, and C. Ionescu, "Using fast weights to attend to the recent past," Adv. Neural Inf. Process. Syst. **29**, 4331–4339 (2016).
[8] S. Yu, "Neuro-inspired computing with emerging nonvolatile memory," Proc. IEEE **88**, 260 (2018).
[9] D. Kuzum, S. Yu, and H.-S. P. Wong, "Synaptic electronics: Materials, devices and applications," Nanotechnology **24**, 382001 (2013).
[10] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sances, I. Boyat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, "Neuromorphic computing using non-volatile memory," Adv. Phys. X **2**, 89 (2017).
[11] Summary Report of the Advanced Scientific Computing Advisory Committee (ASCAC), "Future high performance computing capabilities," Subcommittee, December 19, 2017, see https://science.energy.gov/~/media/ascr/ascac/pdf/meetings/201712/ASCAC-Future-HPC-report.pdf.
[12] N. P. Jouppi *et al.*, "In-datacenter performance analysis of a tensor processing unit," in Proceedings of the 44th International Symposium on Computer Architecture (ISCA), Toronto, 24–28 June 2017.
[13] L. Chua, "Memristor—the missing circuit element," IEEE Trans. Circ. Theory **18**, 507 (1971).
[14] D. Biolek, Z. Biolek, and V. Biolkova, "SPICE modeling of memristive, memcapacitative and meminductive systems," in *Circuit Theory and Design*, *Antalya, Turkey*, 23–27 August 2009 (IEEE, 2009), pp. 249–252.

[15]S. Vongehr and X. Meng, "The missing memristor has not been found," Sci. Rep. **5**, 11657 (2015).

[16]T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco, "STDP and STDP variations with memristors for spiking neuromorphic learning systems," Front. Neurosci **7**, 2 (2013).

[17]V. P. Malinenko, A. L. Pergament, O. V. Spirin, and V. I. Nikulshin, "Threshold and memory switching in oxides of molybdenum, niobium, tungsten, and titanium," J. Sel. Top. Nano Electron. Comput. **2**, 45 (2014). see, http://jstnec.petrsu.ru/files/pdf/3064_en.pdf (or https://www.semanticscholar.org/paper/Threshold-and-Memory-Switching-in-Oxides-of-%2C-%2C-%2C-Malinenko-Pergament/0abe04065cf96f8d780dd24eeabbf3c32ea7485e).

[18]I. K. Yoo, B. S. Kang, Y. D. Park, M. J. Lee, and Y. Park, "Interpretation of nanoscale conducting paths and their control in nickel oxide (NiO) thin films," Appl. Phys. Lett **92**, 202112 (2008).

[19]C. Yakopcic, T. M. Taha, G. Subramanyam, and S. Rogers, "Multiple memristor read and write circuit for neuromorphic applications," in International Joint Conference on Neural Network (IJCNN), San Jose, CA, 31 July–19 August 2011.

[20]M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," Nat. Mater. **12**, 114 (2013).

[21]J. J. Zhang, H. J. Sun, Y. Li, Q. Wang, X. H. Xu, and X. S. Miao, "AgInSbTe memristor with gradual resistance tuning," Appl. Phys. Lett. **102**, 183513 (2013).

[22]J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, and H. Hwang, "Improved synaptic behavior under identical pulses using AlO_x/HfO_2 bilayer RRAM Array for neuromorphic systems," IEEE Electron Device Lett. **37**, 994 (2016).

[23]Z. Wang, M. Yin, T. Zhang, Y. Cai, Y. Wang, Y. Yang, and R. Huang, "Engineering incremental resistive switching in TaO_x based memristors for brain-inspired computing," Nanoscale **8**, 14015 (2016).

[24]K. Moon, A. Fumarola, S. Sidler, J. Jang, P. Narayanan, R. M. Shelby, G. W. Burr, and H. Hwang, "Bidirectional non-filamentary RRAM as an analog neuromorphic synapse, Part I: $Al/Mo/Pr_{0.7}Ca_{0.3}MnO_3$ material improvements and device measurements," IEEE J. Electron Devices Soc. **6**, 146 (2018).

[25]A. K. Mann, D. A. Jayadevi, and A. P. James, "A survey of memristive threshold logic circuits," IEEE Trans. Neural Netw. Learn. Syst. **28**, 1734 (2017).

[26]H. Yu, L. Ni, and H. Huang, "Distributed in-memory computing on binary memristor-crossbar for machine learning," in *Advances in Memristors, Memristive Devices and Systems* (Springer, 2017), pp. 275–304.

[27]I. Boybat, M. L. Gallo, Nandakumar S. R., T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, and E. Eleftheriou, "Neuromorphic computing with multi-memristive synapses," Nat. Commun. **9**, 2514 (2018).

[28]A. Irmanova and A. P. James, "Neuron inspired data encoding memristive multi-level memory cell." Analog Integrated Circuits and Signal Processing **95**, 429 (2018).

[29]F. Corinto, A. Ascoli, and Sung-Mo Kang, "Memristor-based neural circuits," in IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, China, 19–23 May 2013.

[30]A. Amirsoleimani, M. Ahmadi, and A. Ahmadi, "STDP-based unsupervised learning of memristive spiking neural network by Morris-Lecar model," in International Joint Conference on Neural Network (IJCNN), Anchorage, AK, 14–19 May 2017.

[31]J. Jhang and X. Liao, "Synchronization and chaos in coupled memristor-based FitzHugh-Nagumo circuits with memristor synapse," Int. J. Electron. Commun. (AEÜ) **75**, 82 (2017).

[32]B. Bao, A. Hu, H. Bao, Q. Xu, M. Chen, and H. Wu, "Three-dimensional memristive Hindmarsh-Rose neuron model with hidden coexisting asymmetric behaviors," Hindawi Complexity **2018**, 3872573 (2018).

[33]S. Lashkare, S. Chouhan, T. Chavan, A. Bhat, P. Kumbhare, and U. Ganguly, "PCMO RRAM for integrate-and-fire neuron in spiking neural networks," IEEE Electron Device Lett. **39**(4), 484 (2018).

[34]M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. Salama, "Memristors empower spiking neurons with stochasticity," IEEE J. Emerg. Sel. Top. Circuits. Syst. **5**, 242 (2015).

[35]J. Shamsi, A. Amirsoleimani, S. Mirzakuchaki, and M. Ahmadi, "Modular neuron comprises of memristor-based synapse," Neural Comput. Appl. **28**, 1 (2017).

[36]A. Mehonic and A. J. Kenyon, "Emulating the electrical activity of the neuron using a silicon oxide RRAM cell," Front. Neurosci. **10**, 57 (2016).

[37]A. Pantazi, S. Woźniak, T. Tuma, and E. Eleftheriou, "All-memristive neuromorphic computing with level-tuned neurons," Nanotechnology **27**, 355205 (2016).

[38]M. Teimoori, A. Ahmadi, S. Alirezaee, S. V. Al-Din Makki, and M. Ahmadi, "A novel memristor based integrate-and-fire neuron implementation using material implication logic," in *IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax, NS*, 3–6 May 2015 (IEEE, 2015), pp. 1176–1179.

[39]G. Indiveri, "A low-power adaptive integrate-and-fire neuron circuit," in IEEE International Symposium on Circuits and Systems (ISCAS), Bangkok, Thailand, 25–28 May 2003.

[40]J. M. Cruz-Albrecht, M. W. Yung, and N. Srinivasa, "Energy-efficient neuron, synapse and STDP integrated circuits," IEEE Trans. Biomed. Circuits. Syst. **6**, 246 (2012).

[41]J. Shamsi, K. Mohammadi, and S. B. Shokouhi, "A hardware architecture for columnar-organized memory based on CMOS neuron and memristors crossbar arrays," IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **99**, 1 (2018).

[42]I. Sourikopoulos, S. Hedayat, C. Loyez, F. Danneville, V. Hoel, E. Mercier, and A. Cappy, "A 4-fJ/spike artificial neuron in 65nm CMOS technology," Front. Neurosci **11**, 123 (2017).

[43]Y. J. Lee, J. Lee, Y. B. Kim, J. Ayers, A. Volkovskii, A. Selverston, H. Abarbanel, and M. Rabinovich, "Low power real time electronic neuron VLSI design using subthreshold technique," in IEEE International Symposium on Circuits and Systems (ISCAS), Vancouver, BC, Canada, 23–26 May 2004.

[44]J. H. B. Wijekoon and P. Dudek, "Compact Si neuron circuit with spiking and bursting behavior," Neural Netw. **21**, 524 (2008).

[45]Y. Babacan, F. Kaçar, and K. Gürkan, "A spiking and bursting neuron circuit based on memristor," Neurocomputing **203**, 86 (2016).

[46]V. Saxena, X. Wu, and K. Zhu, "Energy-efficient CMOS memristive synapses for mixed-signal neuromorphic system-on-a chip," in IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018.

[47]Y. Babacan and F. Kaçar, "Memristor emulator with spike-timing-dependent-plasticity," Int. J. Electron. C **73**, 16 (2017).

[48]S. Acciarito, A. Cristini, L. D. Nunzio, G. M. Khanal, and G. Susi, "An a VLSI driving circuit for memristor-based STDP," in Ph.D. Research in Microelectronics and Electronics (PRIME), Lisbon, Portugal, 27–30 June 2016.

[49]A. Mizrahi, T. Hirtzlin, A. Fukushima, H. Kubota, S. Yuasa, J. Grollier, and D. Querlioz, "Neural-like computing with populations of superparamagnetic basis functions," Nat. Commun. **9**, 1533 (2018).

[50]P. Wijesinghe, A. Ankit, A. Sengupta, and K. Roy, "An all-memristor deep spiking neural computing system: A step towards realizing the low power, stochastic brain," e-print arXiv:1712.01472, see https://arxiv.org/abs/1712.01472.

[51]L. Deng, D. Wang, Z. Zhang, P. Tang, G. Li, and J. Pei, "Energy consumption analysis for various memristive networks under different learning strategies," Phys. Lett. A **380**, 903 (2016).

[52]CN 103324979 Programmable threshold value circuit.

[53]T. Shibata and T. Ohmi, "An intelligent MOS transistor featuring gate-level weighted sum and threshold operations," in International Electron Devices Meeting (IEDM), Washington, DC, 8–11 December 1991.

[54]X. Wu, V. Saxena, and K. Zhu, "A CMOS spiking neuron for dense memristor-synapse connectivity for brain-inspired computing," in International Joint Conference on Neural Network (IJCNN), Killarney, 12–17 July 2015.

[55]I. E. Ebong and P. Mazumder, "CMOS and memristor-based neural network design for position detection," Proc. IEEE **100**, 2050 (2012).

[56]N. Zheng, and P. Mazumder, "Online supervised learning for hardware-based multilayer spiking neural networks through the modulation of weight-dependent spike-timing-dependent plasticity," IEEE Trans. Neural Netw. Learn. Syst. **29**, 4287 (2017).

[57]N. Zheng, and P. Mazumder, "Learning in memristor crossbar-based spiking neural networks through modulation of weight dependent spike-timing-dependent plasticity," IEEE Trans. Nanotechnol. **17**, 520 (2018).

[58]S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, Pai-Yu Chen, H. Yeon, S. Yu, and J. Kim, "SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations," Nat. Mater. **17**, 335 (2018).

[59]W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, and H. Qian, "Improving analog switching in HfO_x-based resistive memory with a thermal enhanced layer," IEEE Electron Device Lett. **38**, 1019 (2016).

[60]S. Kim, S. Choi, and W. Lu, "Comprehensive physical model of dynamic resistive switching in an oxide memristor," ACS Nano **8**, 2369 (2014).

[61]I.-T. Wang, C.-C. Cahng, L.-W. Chiu, T. Chou, and T.-H. Hou, "3D $Ta/TaO_x/TiO_2/Ti$ synaptic array and linearity tuning of weight update for hardware neural network applications," Nanotechnology **27**, 365204 (2016).

[62]K. Moon, E. Cha, J. Park, S. Gi, M. Chu, K. Baek, B. Lee, S. Oh, and H. Hwang, "Analog synapse device with 5-b MLC and improved data retention for neuromorphic system," IEEE Electron Device Lett. **37**, 1067 (2016).
[63]J. Park, M. Kwak, K. Moon, J. Woo, D. Lee, and H. Hwang, "TiO_x-based RRAM synapse with 64-levels of conductance and symmetric conductance change by adopting a hybrid pulse scheme for neuromorphic computing," IEEE Electron Device Lett. **37**, 1559 (2016).
[64]R. Yang, K. Terabe, Y. Yao, T. Tsuruoka, T. Hasegawa, J. K. Gimzewski, and M. Aono, "Synaptic plasticity and memory functions achieved in a WO_{3-x}-based nanoionics device by using the principle of atomic switch operation," Nanotechnology **24**, 384003 (2013).
[65]K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, and H. Hwang, "Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device," Nanotechnology **22**, 254023 (2011).
[66]A. Chanthbouala, V. Garcia, R. O. Cherifi, K. Bouzehouane, S. Fusil, X. Moya, S. Xavier, H. Yamada, C. Deranlot, N. D. Mathur, M. Bibes, A. Barthélémy, and J. Grollier, "A ferroelectric memristor," Nat. Mater. **11**, 860 (2012).
[67]M. Jerry, P.-Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, and S. Datta, "Ferroelectric FET analog synapse for acceleration of deep neural network training," in International Electron Devices Meeting (IEDM), San Francisco, CA, 2–6 December 2017.
[68]I. Yoo, M. Lee, D. Seo, and S. Kim, "Interpretation of set and reset switching in nickel oxide thin films," Appl. Phys. Lett. **104**, 222902 (2014).
[69]E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga, "Analog memristive synapse in spiking networks implementing unsupervised learning," Front. Neurosci. **10**, 482 (2016).
[70]R. Hasan, T. M. Taha, and C. Yakopcic, "On-chip training of memristor based deep neural networks," in International Joint Conference on Neural Network (IJCNN), Anchorage, AK, 14–19 May 2017.
[71]A. Zyarah, N. Soures, L. Hays, R. Jacobs-Gedrim, S. Agarwal, M. Marinella, and D. Kudithipudi, "Ziksa: On-chip learning accelerator with memristor crossbars for multilevel neural networks," in IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, 28–31 May 2017.
[72]G. Pedretti, V. Milo1, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, "Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity," Sci. Rep **7**, 5288 (2017).
[73]D. R. B. Ly, A. Grossi, T. Werner, T. Dalgaty, C. Fenouillet-Beranger, E. Vianello, and E. Nowak, "Role of synaptic variability in spike-based neuromorphic circuits with unsupervised learning," in IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018.
[74]V. Ntinas, I. Vourkas, A. Abusleme, G. Ch Sirakoulis, and A. Rubio, "Experimental study of artificial neural networks using a digital memristor simulator," IEEE Trans. Neural Netw. Learn. Syst **1**, 5098 (2018).
[75]Y. Zeng, K. Devincentis, Y. Xiao, Z. I. Ferdous, X. Guo, Z. Yan, and Y. Berdichevsky, "A supervised STDP-based training algorithm for living neural networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 15–20, April 2018.
[76]A. Tavanaei, and A. S. Maida, "BP-STDP: Approximating backpropagation using spike timing dependent plasticity," e-print arXiv:1711.04214, see https://arxiv.org/abs/1711.04214.
[77]Y. Nishitani, Y. Kaneko, and M. Ueda, "Supervised learning using spike-timing-dependent plasticity of memristive synapses," IEEE Trans. Neural Netw. Learn. Syst. **26**, 2999 (2015).
[78]US 4,845,669 Transposable memory architecture.
[79]US 8,275,727 B2 Hardware analog-digital neural networks.
[80]A. Shafiee, A. Nag, N. Muralimanohar, R. Balasubramonian, J. P. Strachan, M. Hu, R. S. Williams, and V. Srikumar, "ISAAC: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars," in *ACM/IEEE 43rd Annual International Symposium on Computer Architecture*, *Seoul, Korea*, 18–22 June 2016 (IEEE, 2016), pp. 14–26.
[81]F. M. Bayat, M. Prezioso, B. Chakrabarti, I. Kataeva, and D. Strukov, "Memristor-based perceptron classifier: Increasing complexity and coping with imperfect hardware," in *International Conference on Computer Aided Design (ICCAD)*, *Irvine, CA*, 13–17 November 2017 (IEEE, 2017), pp. 549–554.
[82]See https://knowm.org/ahah-computing/ for information about classifier product.
[83]S. Han, H. Mao, W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," e-print arXiv:1510.00149, see https://arxiv.org/abs/1510.00149.
[84]H. Kim, M. P. Sah, C. Yang, T. Roska, and L. O. Chua, "Memristor bridge synapses," Proc. IEEE **100**, 2061 (2012).
[85]IP 160018N-P2616167 Weighting device and method of the same, POSTECH.
[86]G. Lecerf, J. Tomas, and S. Saïghi, "Excitatory and inhibitory memristive synapses for spiking neural networks," in IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, China, 19–23 May 2013.
[87]Y. Kaneko, Y. Nishitani, and M. Ueda, "Ferroelectric artificial synapses for recognition of a multishaded image," IEEE Trans. Electron Dev. **61**, 2827 (2014).
[88]S. Oh, T. Kim, M. Kwak, J. Song, J. Woo, S. Jeon, I. K. Yoo, and H. Hwang, "$HfZrO_x$-based ferroelectric synapse device with 32 levels of conductance states for neuromorphic applications," IEEE Electron Device Lett. **38**, 732 (2017).
[89]H. Mulaosmanovic1, J. Ocker, S. Müller, M. Noack, J. Müller, P. Polakowski, T. Mikolajick, and S. Slesazeck, "Novel ferroelectric FET based synapse for neuromorphic systems," in Symposium on VLSI Technology, Kyoto, Japan, 5–8 June 2017.
[90]I. Yoo, J. Lee, S. Seo, H. Hwang, "Weighting device and method of the same," P2016167-01-KR.
[91]US 9,934,463 Neuromorphic computational system(s) using resistive synaptic devices.
[92]US 5,146,542 Neural net using capacitive structures connecting output lines and differentially driven input line pairs.
[93]Y. Li, S. Kim, X. Sun, P. Solomon, T. Gokmen, H. Tsai, S. Koswatta, Z. Ren, R. Mo, C. C. Yeh, W. Haensch, and E. Leobandung, "Capacitor-based cross-point array for analog neural network with record symmetry and linearity," in Symposium on VLSI Technology, Honolulu, HI, 18–22 June 2018.
[94]KR IP 17001N Capacitance-based multilevel synapse device and method, POSTECH.
[95]X. Zou, H. Ong, L. You, W. Chen, H. Ding, H. Funakubo, L. Chen, and J. Wang, "Charge trapping-detrapping induced resistive switching in $Ba_{0.7}Sr_{0.3}TiO_3$," AIP Adv. **2**, 032166 (2012).