Journal of Semiconductor Engineering Vol. 2, Issue 3, December, 2021
https://doi.org/10.22895/jse.2021.0004
Received 22 Aug 2021, Revised 27 Oct 2021, Accepted 1 Nov 2021 ISSN(On-line) 2733-6719

Tunneling Field-Effect Transistors for Neuromorphic Applications

Jang Woo Lee, Jae Seung Woo, and Woo Young Choi

Department of Electronic Engineering, Sogang University, Seoul, 04107, Republic of Korea

Corresponding Author: Woo Young Choi (e-mail: wchoi@sogang.ac.kr)

Funding Information: This work was supported in part by the NRF of Korea funded by the MSIT under Grant NRF-2019M3F3A1A02072089, NRF-2021M3F3A2A01037927 (Intelligent Semiconductor Technology Development Program), NRF-2021R1A2C1007931 (Mid-Career Researcher Program), in part by the IITP funded by the MSIT under Grant IITP-2020-2018-0-01421 (Information Technology Research Center Program), in part by the MOTIE/KSRC under Grant 10080575 (Technology Innovation Program), and in part by the Sogang University Research Grant of 202119026.01

ABSTRACT

Recent research on synaptic devices has been reviewed from the perspective of hardware-based neuromorphic computing. In addition, the backgrounds of neuromorphic computing and two training methods for hardware-based neuromorphic computing are described in detail. Moreover, two types of memristor- and CMOS-based synaptic devices were compared in terms of both the required performance metrics and low-power applications. Based on a review of recent studies, additional power-scalable synaptic devices such as tunnel field-effect transistors are suggested for a plausible candidate for neuromorphic applications.

KEY WORDS

Hardware-based neuromorphic computing, synaptic devices, low-power operation, online training, offline training, tunnel field-effect transistors.

1. INTRODUCTION

For a fast operation and higher density, over the past three decades, the shrinking of complementary metal-oxide semiconductors (CMOS) has been achieved by following Moore's law [1]–[3]. Because the scaling of CMOS has been confronted with a limitation of the serious leakage current problem, it also affects conventional serial computing systems. To overcome the power consumption problem of serial processes, which originates from the von Neumann bottleneck, the neuromorphic computing architecture for parallel processing systems has been highlighted by many researchers [4]–[8]. To achieve this goal, many studies on hardware-based neuromorphic computing have focused on mimicking the biological characteristics of synapses using online training, accompanied by spike-time dependent plasticity (STDP) rules [7]. By contrast, it is inevitable for synaptic devices to have reliability problems, seriously affecting hardware-based neuromorphic computing. Thus, the training method, which is not critical to the reliability of synaptic devices, should also be considered for hardware-based neuromorphic computing [8]–[10]. Moreover, numerous artificial synaptic devices have been reported using a memristor, which is one of the most promising candidates [11]. Nevertheless, memristor-based synaptic devices have several drawbacks, such as the requirement of other logic blocks, incompatibility with CMOS technology, vulnerable in the endurance cycle, and a high power consumption [9], [10], [12]. For this reason, CMOS-based memory devices for synaptic transistors have been aggressively reported owing to their better characteristics than those of the aforementioned drawbacks found in memristors [10], [12]. However, the synaptic devices for both the memristor-based and CMOS-based memories still show a higher power consumption than that of the human brain to meet the neurobiological requirement (less than ~10 fJ per single synaptic event) for massive parallelism [13]. In addition, CMOS-based memory suffers from an unscalable power consumption restraint owing to the

Boltzmann tyranny. It is therefore necessary for synaptic devices to lower the operating power consumption compared with conventional memristors or CMOS-based memories.

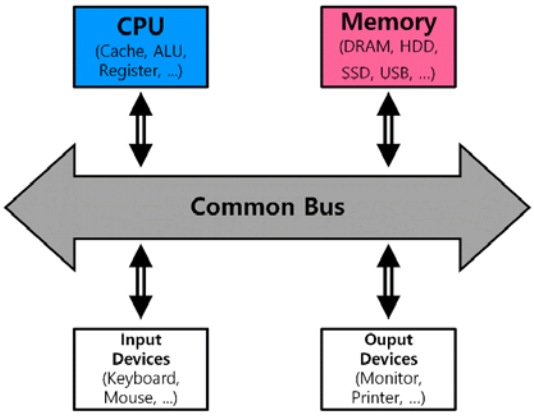

Figure 1. Von Neumann bottleneck owing to the shared common bus between program and data memories.

In this study, two popular training methods for neuromorphic architectures, *i.e.*, online and offline training, are compared in terms of hardware-based neuromorphic computing. Furthermore, memristor- and CMOS-based synaptic devices were compared according to the required performance metrics. Finally, various steep switching devices have been introduced to overcome the fundamental limitations of power scaling in CMOS devices. Among them, it has been suggested that tunnel field-effect transistors (TFETs) are promising candidates for synaptic devices owing to their extremely low on- and off-state currents [14].

2. Fundamental Limitations in Current Computing Systems

The Go match between AlphaGo and Lee Sedol in March 2016 drew significant attention globally and has become a milestone in the quest for artificial intelligence (AI). Consequently, AlphaGo by Google DeepMind astonishingly won 4:1 in a five-game match [15]. This tremendous achievement of AlphaGo originates from adapting the convolutional neural network (CNN) between the policy and value, which significantly reduces the amount of data when calculating based solely on a Monte Carlo tree search (MCTS) [16]. In addition, using a deep neural network (DNN), which means that the number of hidden layers is higher than 2, supervised and reinforcement learning are successfully conducted [16]. However, AlphaGo consists of 1920 CPUs and 280 graphic processing units (GPUs) for this purpose and uses ~100 kW to execute the operations, whereas the human brain only uses ~20 W with ~10^{14} synapses, connecting ~10^{11} neurons. Thus, it is important and problematic for AI to reduce the power consumption when operating a DNN [17]. This problem originates from the von Neumann architecture. The von Neumann architecture is a backbone of the computer architecture in all general computing devices, including central process units (CPUs), memory, and input/output (I/O) devices used to date [18]. However, the von Neumann architecture becomes problematic owing to its sequential process in executing the instructions using the CPU. This is because the CPU operates on a shared common bus between an instruction fetch and a data operation depending on the clock. This is called the von Neumann bottleneck, which constrains the performance of the computing systems, as depicted in Fig. 1 [19], [20]. Thus, as the fundamental cause of the von Neumann architecture, all components are connected to a shared common bus. Moreover, it results in data latency because the dynamic random access memory (DRAM) or flash memory determines the overall data processing time, despite the fast processing speed of the CPU. Thus, it seriously limits the performance of the systems, particularly for the processing of large amounts of data. Therefore, the von Neumann architecture is unsuitable for massive parallel operations. It is inefficient for a von Neumann architecture, which only increases the number of CPUs and the GPU core without fundamentally changing the data path, the common bus of which connects the memory to the cores, to overcome the von Neumann bottleneck. However, the von Neumann architecture still overwhelms the human brain when computing high-speed sequential processes.

3. Background of Neurobiological Characteristics

To address the aforementioned issues, several attempts have been made to realize neuromorphic systems that are completely different from the von Neumann architecture by imitating a biological nervous system [21]–[23]. Furthermore, it is essential to understand a synaptic transmission, which is also referred to as a neurotransmission. Basically, the synapse is defined as the connection between the axon of a presynaptic neuron and the dendrite of a post-synaptic neuron as depicted in Fig. 2. In addition, the strength of the synapses is varied and is called the synaptic weight, which is discussed in detail later. The signals transferred from a presynaptic neuron (synaptic inputs) are integrated at the dendrites of a postsynaptic neuron. All integrated synaptic inputs then determine the membrane potential at the axon hillock of the post-

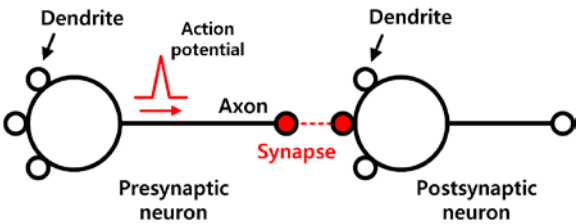

Figure 2. Basic morphology of two neurons: presynaptic and postsynaptic neuron. Action potential: spike shape.

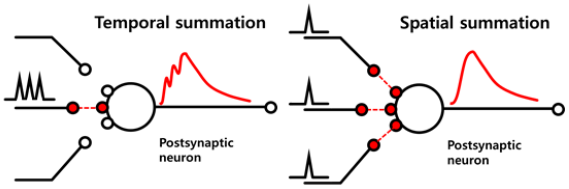

Figure 3. Synaptic integration: Temporal and spatial summations.

synaptic neuron to generate action potentials (spikes) if the membrane potential exceeds the threshold. This is known as synaptic integration or a summation [24]–[26]. Specifically, excitatory signals cause the membrane potential to exceed the threshold and trigger the action potential, whereas inhibitory signals take the membrane potential away from the threshold and are less likely to generate such potential. In addition, the membrane potential is determined by the spatiotemporal integration of neuron inputs, as shown in Fig. 3. The frequency and pattern of action potentials from the same presynaptic neuron are important for triggering the membrane potential of a post-synaptic neuron. In summary, the neuron inputs from a presynaptic neuron are summed to determine the membrane potential of a post-synaptic neuron to determine whether to generate the action potential.

Furthermore, the synapse plays the role of short-term and long-term memory and determines the membrane potential to be triggered by a synaptic transmission [27]. Briefly, weaker stimuli cause short-term memory. In addition, short-term memory is temporally retained from minutes to hours. However, it becomes long-term memory when repeatedly stimulated. Long-term memory is not temporal but is maintained owing to the stronger stimuli. Moreover, neurons physically form new synapses when long-term memory is created. Moreover, long-term memory persists regardless of whether it is used. By contrast, the memory characteristics of the synapse can be mimicked by memory devices for neuromorphic applications.

It is known that spike-timing-dependent plasticity (STDP) is a mechanism for enhancing/depressing the synaptic transmission. In STDP, one of the fundamental learning mechanisms in the biological nervous system, the timing difference of the action potentials between presynaptic and postsynaptic determines the synaptic weight to be potentiated or depressed [28]–[30]. As illustrated in Fig. 4, the spike timing difference $\Delta t = t_{post} - t_{pre}$ determines the STDP learning rule because it results in the causality of the neural connections. If the presynaptic neuron becomes active slightly before the postsynaptic neuron ($\Delta t > 0$), which means that the firing of the former results in the firing of the latter, the input spike strengthens the synaptic weight, leading to long-term potentiation (LTP). By contrast, if the presynaptic neuron becomes active slightly after the postsynaptic neuron ($\Delta t < 0$), which means that the firing of the former has no results in the firing of the latter, the input spike weakens the synaptic weight, leading to long-term depression (LTD). Thus, STDP can be interpreted based on a Hebbian learning rule by adding the concept of causality that the synapse strengthens if the presynaptic neuron results in the firing of the postsynaptic neuron [31]. The STDP function showing the change in synaptic weight, also referred to as the excitatory postsynaptic potential (EPSP), as a function of Δt, was plotted by Bi and Poo [28]. Repeated presynaptic spiking signals before a postsynaptic spiking signal within 50 ms result in LTP, whereas presynaptic spiking signals after postsynaptic spiking signals within 50 ms result in LTD. Furthermore, a high level of change in the synaptic weight is observed as Δt approaches 0 s, which means that the causality between the two neurons is enhanced.

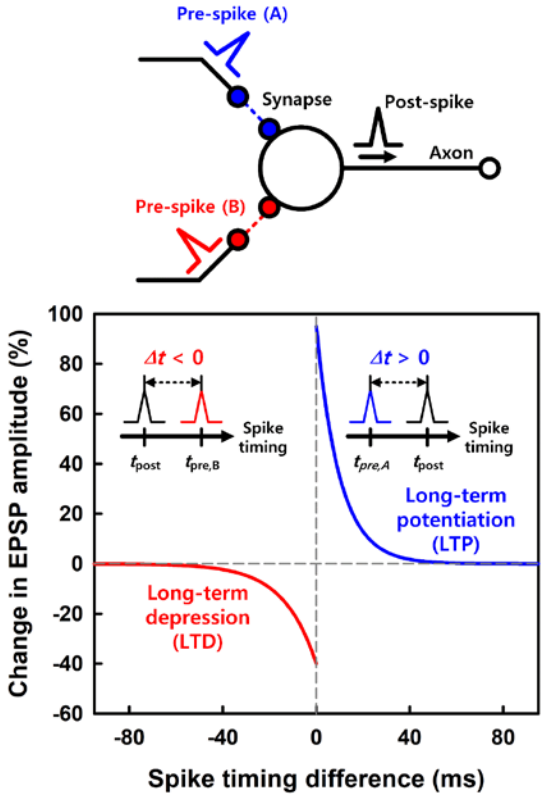

Figure 4. Conceptual illustrations for explaining the STDP learning rules. LTP and LTD characteristics in terms of the change in EPSP amplitude or synaptic weight, with respect to the spike timing difference, Δt.

4. Background of Hardware-Based Neuromorphic Computing

One of the most important features in neuromorphic computing is the exclusion of a clock, which is the signature of a von Neumann architecture [22], [23]. This is because the serial processing of information causes a power consumption problem. During the late 1980s, Mead implemented the first neuromorphic computing chip to mimic the biological nervous system using a complementary metal-oxide semiconductor-based analog circuit [32]. It has attracted considerable attention owing to its low power consumption, massive parallel

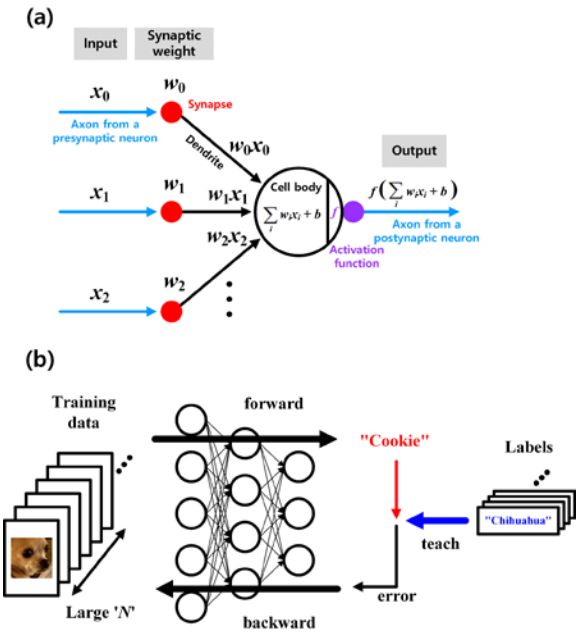

Figure 5. Conceptual illustrations for (a) single perceptron and (b) supervised learning using MLP adapting error back-propagation. In general, ReLU is used as an activation function owing to the lack of gradient vanishing problems.

processing, and false tolerance. However, it has a power efficiency problem, *i.e.*, it consumes more power than the human brain. In addition, the complete absence of the clock in hardware-based neuromorphic systems remains to be improved.

As a pathfinding for the data-centric application, analog-in memory computing (AiMC) has recently been spotlighted as a means of neuromorphic computing [8], [32], [33]. Thus, synaptic devices are considered a major key to implementing neuromorphic systems. Furthermore, it should resemble the memory characteristics and synaptic transmission of biological synapses. Thus, the STDP learning rule is assumed to mimic the biological synapse itself. For this purpose, memristors such as emerging non-volatile memories (eNVMs), including resistive random access memory (RRAM) and phase-change RAM (PCRAM), have been demonstrated as synaptic devices for neuromorphic systems [7], [34]. This is due to their steep on-off transition and gradually changing characteristics during conductance. In addition, high-density crossbar arrays can easily form owing to the simplicity of the device fabrication. The operation mechanism of a memristor as a synapse is based on the threshold voltage for a change in conductance, whether exceeded or not. Given the STDP learning rule, spikes with a low frequency from a presynaptic device cannot change the conductance, whereas spikes with a high frequency from a presynaptic device can exceed the threshold voltage, leading to a change in the conductance [7], [34].

However, there are two types of neural networks that imitate the biological nervous system: artificial neural networks (ANNs) and spiking neural networks

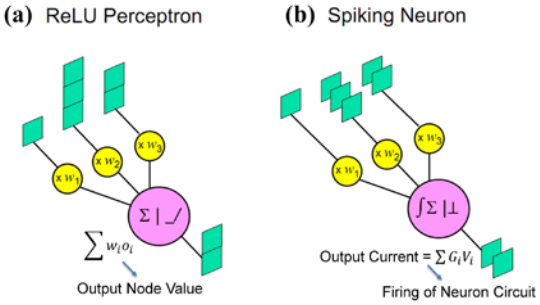

Figure 6. Equivalence between (a) perceptron using ReLU as an activation function for ANN and (b) SNN based on STDP method [42].

(SNNs) [9]. Notably, ANNs use non-spiking (digital) signals, which are different from the spiking neural network (SNN) triggered by the STDP method. For an ANN, the synaptic inputs are encoded by the number of pulses or the voltage level. Moreover, it is based on a mathematical model called a perceptron. Rosenblatt determined that the summation of all neuron inputs (x_i) multiplied by the synaptic weight (w_i) is triggered by the activation function, as shown in Fig. 5 [35], [36]. Typically, most of the demonstrated neuromorphic systems are based on an ANN and feature a multilayer perceptron (MLP) with an error back-propagation method, as depicted in Fig. 5 [37], [38]. In addition, an ANN follows the supervised learning rule owing to an error back-propagation. Specifically, a layer-by-layer error is used to optimize the objective cost function by comparing the errors between the prediction and true label [9]. Thus, an ANN ensures that the learning accuracy and performance are similar to the results of the AI software. The aim of an ANN is to improve the computational efficiency in terms of throughput over power, such as terra operations per second per watt (TOPS/W). An SNN is also a mathematical model; however, it aims to emulate a biologically realistic neural network more closely. In addition, it follows the unsupervised learning rule owing to the STDP. Unfortunately, the learning accuracy and performance of an SNN are worse than those of AI software. By contrast, it has been reported that an ANN

Table 1. Summary of online and offline training characteristics.

	Online training	**Offline training**
Where training occurs	Training occurs inside the system itself	Training occurs in the software AI, outside the system
How training works	Biological learning rule are used as STDP	Transferring (Programming) pre-trained weights from software AI as a form of conductance to the memory device
Performance	Poorer performance than software AI	Very close to software AI

using a rectified linear unit perceptron and an SNN using the STDP method are essentially equivalent, as illustrated in Fig. 6 [39]–[42]. This is proved through the following steps: First, the presynaptic signal levels in the ANN are divided into units. Second, we rearranged them on the time axis by multiplying the synaptic weights. Third, they were integrated through the time axis. Thus, the STDP method is another expression of the ANN.

In addition, there are two types of training (*i.e.*, synaptic weight updating) in a neural network: offline training and online training [9]. Offline training means that the training is conducted using software, *i.e.*, the conventional von Neumann architecture, after which the trained synaptic weights are loaded to the synaptic arrays (or synaptic architecture) of the neuromorphic hardware through one-time programming. An example of offline training is graphically described in [8]. Subsequently, only inference or classification was conducted. Therefore, such inference-only neuromorphic systems are suitable for edge devices, for which the instantaneous change in synaptic weights according to the changing input data is not needed during runtime. By contrast, online training means that the training is conducted during runtime on neuromorphic hardware, *i.e.*, during on-the-fly (instantaneous) training of the synaptic weights. In general, ANNs are closely related to offline training, whereas SNNs follow online training. The characteristics of the two training methods are summarized in Table. 1.

5. Challenges in Neuromorphic Computing

Although it is important to imitate the biologically realistic behavior of the synapse, hardware-based neuromorphic computing is strongly related to the reliability of synaptic devices. Considering that the guaranteed endurance cycles of reported synaptic devices are ~10^6, the instantaneous changes in the synaptic weight of online training contain an inherent limitation [43]. Moreover, online training shows a worse performance than software artificial intelligence (AI) because of their unsupervised learning rule, whereas that of offline training is extremely close to software AI [8], [9], [44]. Thus, online training is a much more challenging training method than offline training. Therefore, offline training is preferred for edge-computing applications. This is because the synaptic weight transfer is conducted through one-time programming, as mentioned before.

Moreover, the current interest in synaptic devices for neuromorphic computing is in reducing the power consumption. To successfully implement massive parallelism such as in the human brain, consisting of ~10^{14} synapses with ~10^{11} neurons, ~10 fJ is required for a single event of a single synaptic device [13], [45]. Thus, a low power consumption of less than 10 fJ with a single synaptic event is strongly demanded for synaptic devices when considering the high density of the synapses. Therefore, to satisfy the neurobiological requirements, the power consumption used in synaptic devices is decreasing [13]. Unfortunately, several synaptic devices under offline training consume much more energy than the neurobiological requirement, despite previous studies demonstrating the functionality of synapses, as shown in Fig. 7. Only memristor-based synaptic devices meet the aforementioned neurobiological requirements [46]. However, memristor-based devices have drawbacks such as incompatibility with a conventional Si-based CMOS process and/or significant material dependence [46]. Thus, it is necessary for synaptic devices to be compatible with conventional Si-based CMOS processes and to satisfy the power constraint of < 10 fJ, *i.e.*, for further low-power operation. The benchmark of the energy consumption (E) in synaptic devices for offline training is evaluated using $E = VIt$, where V is the read voltage, I is the read current, and t is the read pulse width. The energy consumption during the read operation is considered a more important performance parameter of how weekly or energy efficiently the synaptic devices operate. Moreover, read operations for inference occur more frequently than program/erase operations, particularly for offline training. In addition, the required performance metrics for synaptic devices in terms of offline training are summarized in Table 2 [8], [9],

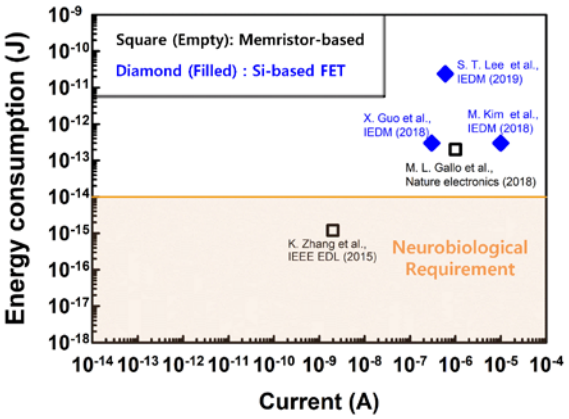

Figure 7. Benchmark of the research trend of the energy consumption in synaptic devices as a function of the synaptic current when considering the neurobiological requirement [6], [8], [10], [12], [46].

Table 2. Summary of the required performance metrics in synaptic devices.

Performance metrics for neuromorphic devices (offline training)	**Desired target**
Energy consumption (during inference)	< 10 fJ / single synaptic event
Dynamic range	> 100
Retention	> 10 years
Endurance	> 10^5

[11], [47].

Table 3. Comparison of the key features exhibited by Si-based CMOS memory devices and emerging memristor devices. Adapted from [43] and [53].

Performance metrics	CMOS-based memory devices			Emerging memristor devices	
	NOR	NAND	SRAM	RRAM	PCRAM
Energy consumption (P/E operation)	>0.1 nJ	~10 fJ	~10 pJ	~1 pJ	~10 pJ
Energy consumption (Read operation)	>10 pJ	>10 pJ	>10 pJ	>10 fJ	>10 fJ
Dynamic range	10^4	10^4	$>10^4$	>10	$>10^2$
Scalability (F^2)	4−6	4−6	120−200	4−6	4−16
write speed	~10 μs	~1 ms	~10 ns	~10 ns	~10 ns
Retention	Long	Long	Long	Medium	Long
Endurance	10^5	10^4	$>10^{15}$	questionable	$>10^5$

Despite its advantages, however, a memristor requires an additional control logic component owing to its 2-terminal feature [7], [8], [48]. Moreover, memristors exhibit challenging issues such as a poor reliability, asymmetric change in conductance, resistance variation, and control logic integration [7], [44], [48]. Integration with the control logic component is problematic because the fabrication process of the memristor is CMOS-incompatible. By contrast, it is necessary for synaptic devices to not only have a low off-state current (I_{off}), but also a low on-state current (I_{on}) with a high I_{on}/I_{off} during a read operation [44]. This is because most of the common neuromorphic systems integrate the currents from different synapses and compare their values according to each synaptic weight. However, the memristor shows a low I_{on}/I_{off} with a high I_{on}, despite a steep switching characteristic [7], [34]. Thus, the low-power applications of most reported memristors remain to be explored for neuromorphic systems. Therefore, CMOS-based memories, such as SONOS-type flash memories, for use in synaptic devices have recently been reported owing to their 3-terminal feature, high reliability, and sufficiently high I_{on}/I_{off} [10], [12], [49]. However, further scaling of the power consumption beyond Moore's law is still necessary because of the unscalable power consumption limit in CMOS devices when considering the AiMC circuit design [10], [44], [49]. Moreover, the CMOS-based NOR-type memory is not energy efficient because it consumes a large amount of energy for program/erase operations owing to the hot carrier injection mechanism, although it shows good data accessibility and latency [50]. However, a CMOS-based NAND-type memory consumes the lowest amount of energy for a program/erase operation owing to its self-boosting mechanism [51], although it exhibits a poor data latency owing to its configuration. In addition, it requires complicated data access schemes even for read operations, which hampers its data accessibility [51]. In hardware accelerators used in a DNN, static RAM (SRAM) is commonly utilized and commercialized to store synaptic weights in CMOS ASIC designs. Nevertheless, an SRAM cell is much larger than any other memory including CMOS-based flashes and eNVMs, showing an area of over 200 F^2 (where F indicates the technology feature size). Thus, it results in severe constraints on the storage and power consumption despite the high speed [52]. The key features of comparing these CMOS-based memory devices and emerging memristor devices are summarized in Table 3 [43], [53]. Overall, it is necessary for CMOS-based memory devices to further scale the energy consumption to meet the neurobiological requirements. Therefore, it is necessary for synaptic devices to show a high reliability, CMOS fabrication compatibility, and power scalability.

However, the logic component for comparing the operations applied in neuromorphic computing is also problematic because of the large size used for comparing the logic blocks, which deteriorates the TOPS/W. Thus, it is necessary to reduce the number of logic components or operation steps of AiMC. It was reasonably reported that a binary neural network (BNN) significantly reduces the number of operation steps, resulting in a downsizing of the computational resources [54], [55]. In a BNN, both synaptic weights and neuron activation are binarized to −1 or +1. Thus, multiplications between activations and weights can be simplified as bit-wise XNOR operations, and the accumulation of the products is equivalent to a bit-counting operation. In addition, MAC is an abbreviation for this multiply-and-accumulate approach. It is known that a BNN is essentially a CNN whose weights and inputs are binarized to −1 or +1. Thus, it is regarded as an extreme quantization case with a less precise CNN. In hardware-based neuromorphic com-

puting, training and inference can be successfully conducted using only 1 bit. This reduces the burden of the synaptic devices because a multi-bit operation is not required, which is affected by the noise margin and/or bit-error rate. However, the most frequently demonstrated neuromorphic computing architectures adapting a BNN are based on a memristor or SRAM [52], [56]. It is therefore necessary to adapt synaptic devices suitable for low-power applications.

6. Necessity of Low-Power Devices for Neuromorphic Computing

Over the past 50 decades, the downscaling of metal-oxide semiconductor field-effect transistors (MOSFETs) has continued to accelerate the complementary MOS (CMOS) manufacturing technology down smaller than 10-nm nodes [1]–[3]. In addition, this scaling trend is based on Moore's law, *i.e.*, the number of transistors on a microprocessor chip doubles every two years, allowing a high operation speed, low power consumption, high chip density, and even lowered cost per transistor to be achieved [1]–[3]. Owing to Moore's law, electronic systems have become widely used for the expansion of semiconductor devices. It has therefore resulted in various types of computing systems, such as handheld devices and portable biosensors, including the Internet of Things (IoT). However, this scaling trend faces a serious leakage current in extremely scaled nanoscale devices, which makes it difficult to maintain Moore's law. Moreover, it causes a static power dissipation, which is generated when transistors are turned off, as one of the most serious problems in current computing systems. As CMOS devices are scaled down, a static (subthreshold) power density rapidly increases toward an active power density [3]. Thus, the static power consumption, which results from the leakage current, is one of the most serious challenges in CMOS devices. Moreover, the scaling trend in CMOS accelerates the increase in leakage current. This is attributed to the short-channel effect [1]–[3]. Thus, by enhancing the gate controllability, an abrupt on-off transition of a MOSFET becomes the key for low-power applications while scaling the V_D. Therefore, it is necessary for MOSFETs and other next-generation devices to have immunity to the short-channel effect. Thus, many researchers have attempted to overcome this problem, such as nano-electromechanical (NEM) devices [57], negative capacitance FETs (NCFETs) [58], impact ionization MOSFETs (I-MOS) [59], and tunnel FETs (TFETs) [60].

By contrast, MOSFETs theoretically have constraints on reducing the leakage current owing to the unscalable characteristic of a subthreshold swing (*SS*), which causes a static power consumption. As shown in Fig. 8, considering the high-energy region in a Fermi-Dirac distribution (Boltzmann tail), a carrier density always exists on the source side of the MOSFETs, which have a higher energy than that of the channel-to-source barrier height. Thus, it results in a static power consumption. In addition, the *SS* is defined as [14]

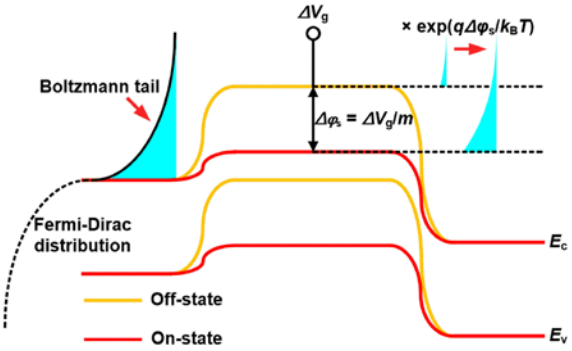

Figure 8. Energy band diagrams for MOSFETs with both off- and on-states, showing the MOSFET mechanism based on thermionic emissions over φ_{bi}.

$$SS = m \times n = \frac{dV_G}{d\varphi_s} \frac{d\varphi_s}{d\left(\log I_D\right)}, \tag{1.1}$$

where V_G is the gate voltage, φ_s is the surface potential, and I_D is the drain current. The m factor is defined as $m = 1 + C_{dep}/C_{gate}$. Moreover, the carrier density in the source of MOSFETs as a function of φ_s, is approximated as below, when considering the Boltzmann tail:

$$Q_S \approx \exp\left(\frac{q\varphi_s}{k_B T}\right), \tag{1.2}$$

where q is the elementary charge, k_B is the Boltzmann constant, and T is the temperature. Finally, as the drain current (I_D) is proportional to Q_s within the subthreshold region, the *SS* of the MOSFETs is defined as follows:

$$SS \approx \ln(10) m \frac{k_B T}{q}. \tag{1.3}$$

From (1.3), the *SS* of the MOSFETs cannot reach lower than 60-mV/dec at room temperature because of their thermionic injection mechanism. This is called *Boltzmann tyranny* [14]. Specifically, there are two possible methods for achieving an *SS* of less than 60 mV/dec. as expressed in Eq. (1.1): first, reducing the m factor to less than 1, and second, reducing the n factor to less than $(k_B T/q)\ln(10)$. The former can be implemented using NEM devices [57] or NCFETs [58]. However, the latter can be realized by changing the carrier injection mechanism, for example, I-MOS [59] or TFETs [60]. Among them, TFETs are considered possible solutions for post-CMOS technology owing to their CMOS-compatible fabrication process [60]. In addition, TFETs show an extremely low I_{off} with an *SS*

of less than 60-mV/dec. Thus, the steep slope characteristic of TFETs in the transfer characteristic (I_{D}–V_{G}) leads to an abrupt on-off transition at room temperature.

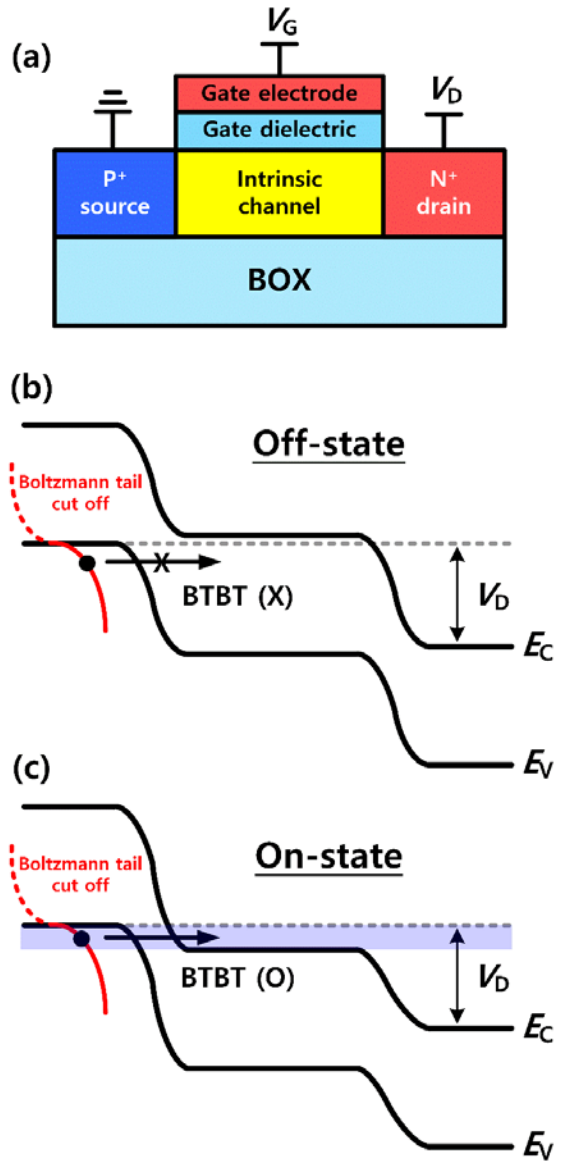

Figure 9. (a) Structure of conventional SOI TFETs. Lateral energy band diagram of TFETs in an (b) off-state and an (c) on-state are also shown. BTBT only occurs in blue-shadowed region, which indicates a tunneling energy window.

The current flowing mechanism of TFETs is based on band-to-band tunneling (BTBT) between the valence band (E_v) and conduction band (E_c). The basic structure and operation mechanism of the n-type TFETs are illustrated in Fig. 9. For simplicity, the tunneling barrier is assumed to be perfectly transparent such that the allowed tunneling energy window has a probability of 1, *i.e.*, $P(E_{c,ch} < E < E_{v,s}) = 1$, and a probability of 0 otherwise. Schematics of the conventional TFETs are depicted in Fig. 9(a). It is assumed that the TFETs are fabricated on a silicon-on-insulator (SOI) wafer, which shows a buried oxide (BOX) located below the SOI layer. The major difference between TFETs and MOSFETs is the symmetry of the doping concentration. Owing to the asymmetric doping concentration, TFETs filter out of the Boltzmann tail, which makes it possible to achieve a level lower than 60 mV/dec by overcoming the Boltzmann tyranny [14]. Unlike MOSFETs, the Boltzmann tail in the TFETs is filtered out by the source forbidden gap for the TFETs regardless of the off- or on-state. Figure 9(b) shows the lateral energy band diagrams of the TFETs in an off-state. In the case of an off-state, because the BTBT path is not formed owing to the channel forbidden gap, the BTBT from the source valence band to the channel conduction band cannot occur in an off-state. Therefore, TFETs exhibit extremely low I_{off} values. By contrast, Fig. 9(c) shows the lateral energy band diagrams of TFETs in an on-state. As V_G increases, the channel conduction band eventually becomes lower than the source valence band. Therefore, BTBT occurs from the source valence band to the channel conduction band in an on-state.

By contrast, the tunneling probability in TFETs is usually calculated based on a Wentzel, Kramers, and Brillouin (WKB) approximation [61]. However, this approach is inappropriate for Si-based TFETs owing to the indirect band gap nature, whereas it is relatively well-fitted in III–V-based direct band gap TFETs [61]. Thus, several efforts have been made to predict the drain current more accurately, even for 2-D transition metal dichalcogenide (TMD) based TFETs [62], [63]. Nevertheless, with advantages such as an extremely low I_{off} and an *SS* of less than 60 mV/dec, TFETs suffer from a severe disadvantage of a low I_{on}. Thus, several studies on improving the I_{on} of TFETs have been reported, such as pocket doping [64], bandgap engineering [65], hetero-gate-dielectric [66], and gate-normal (also referred as vertical- or line-) tunneling [67] techniques. However, despite these efforts, the I_{on} of TFETs is still insufficient to meet the IRDS requirements [68]. In addition, these efforts require complexity during the fabrication process. In particular, the fabrication of III–V materials is still insufficiently mature to adapt to conventional CMOS technology. Only TFETs with Si-based group IV materials, such as Si, SiGe, and Ge, are considered because of their compatibility with conventional CMOS technology, including bandgap engineering [69], [70]. Unfortunately, the increase in I_{on} of TFETs is obtained at the expense of I_{off}, which results in a degradation of the *SS*. It is therefore challenging to replace conventional CMOS technology for logic applications.

By contrast, emerging neuromorphic systems demand not only a low I_{off}, but also a low I_{on} with sufficiently high I_{on}/I_{off} to meet the dynamic range require-

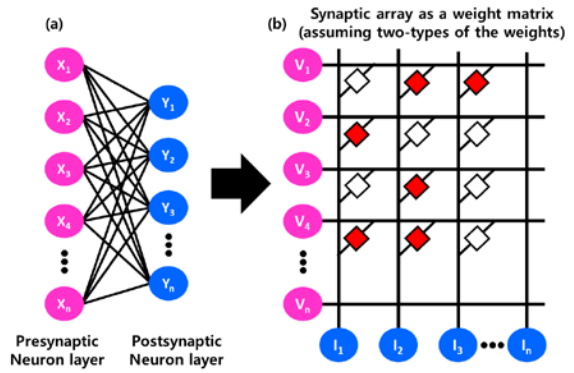

Figure 10. Schematics showing vector-matrix multiplication (VMM). (a) Graphically expressed weight matrix between neuron layers in a neural network. (b) The synaptic array consists of perpendicular rows (voltage inputs) and columns (current outputs) with the memristor-based synaptic devices. The weights in the neural network are mapped to the conductance of the memristor-based synaptic devices. The weighted sum operation is conducted by applying read voltages to all rows and reading out the weighted sum current in all columns. For simplicity, only two types of conductance states are assumed as depicted.

ments of at least 100 [44]. The most common hardware-based neuromorphic systems perform MAC operations by using a vector-matrix multiplication (VMM), which is easily implemented through memory arrays, as illustrated in Fig. 10. As shown in Fig. 10, to deal with the numerous computations, an increase in the size of the memory arrays for VMM is inevitable. Thus, the I_{on} used to evaluate the VMM should be as small as possible for each synaptic device when the system size, *i.e.*, the size of the arrays, increases. However, it has been reported that most synaptic devices for neuromorphic systems are based on RRAM, which has several drawbacks. Thus, CMOS-based synaptic devices have recently been demonstrated to replace RRAM. However, this also has a limitation on the power scaling owing to the Boltzmann tyranny. In addition, the high I_{on} of a MOSFET, which is the preferred characteristic for logic applications, is problematic as a synaptic device for neuromorphic systems. It is therefore necessary to further improve the active power consumption as well as the static power consumption.

7. CONCLUSION

Recent studies on synaptic devices for neuromorphic computing have been reviewed. From the perspective of hardware-based neuromorphic computing, both synaptic devices and training methods should guarantee the required performance metrics, particularly for better reliability. Thus, CMOS-based synaptic devices are promising owing to their highly reliable operation characteristics. However, the inherent power consumption limits of CMOS-based synaptic devices should also be considered when dealing with low-power applications. Therefore, TFETs showing CMOS compatibility with an ultralow-power operation are suggested in this paper. Moreover, they can be a solution to hardware-based neuromorphic computing.

REFERENCES

[1] S. Borkar, “Design challenges of technology scaling,” IEEE Micro, vol. 19, no. 4, pp. 23–29, Jul.–Aug. 1999.

[2] A. M. Ionescu, “Energy efficient computing and sensing in the Zettabyte era: from silicon to the cloud,” *IEEE International Electron Devices Meeting (IEDM)*, pp. 1.2.1-1.2.8, 2017.

[3] E. J. Nowak, “Maintaining the benefits of CMOS scaling when scaling bogs down,” *IBM Journal of Research and Development*, vol. 46, no. 2/3, pp. 169-180, Mar. 2002.

[4] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, pp. 1629-1636, 1990.

[5] D. Ielmini and H.-S. P. Wong, “In-memory computing with resistive switching devices”, *Nature electronics*, vol. 1, pp. 333-343, Jun. 2018.

[6] M. Kim, J. Kim, G. Park, L. Everson, H. Kim, S. Song, S. Lee, and C. H. Kim, “A 68 Parallel Row Access Neuromorphic Core with 22K Multi-Level Synapses Based on Logic-Compatible Embedded Flash Memory Technology,” *IEEE International Electron Devices Meeting (IEDM)*, pp. 352-355, Dec. 2018

[7] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, “Nanoscale Memristor Device as Synapse in Neuromorphic Systems,” *ACS Nanoletters*, vo1. 10, pp. 1297-1301, 2010.

[8] M. L. Gallo, A. Sebastian, R. Mathis, M. Manica, H. Glefers, T. Tuma, C. Bekas, A. Curioni, and E. Eleftheriou, “Mixed-precision in-memory computing,” *Nature electronics*, vol. 1, no. 5, pp. 246-253, Apr. 2018.

[9] S. Yu, “Neuro-Inspired Computing With Emerging Nonvolatile memory,” *Proceedings of the IEEE*, vol. 106, no. 2, pp. 260-285, Feb. 2018.

[10] S. T. Lee, H. Kim, J. H. Bae, H. Yoo, N. Y. Choi, D. Kwon, S. Lim, B. G. Park and J. H. Lee, “High-Density and Highly-Reliable Binary Neural Networks Using NAND Flash Memory Cells as Synaptic Devices,” *IEEE International Electron Devices Meeting (IEDM)*, pp. 927–930, 2019.

[11] D. Kuzum, S. Yu, and H. -S. P. Wong, “Synaptic electronics: materials, devices and applications,” *IOP Nanotechnology*, vol. 24, no. 382001, pp.1-22, Sep. 2013.

[12] X. Guo, F. M. Bayat, M. Bavandpour, M. Klachko, M. R. Mahmoodi, M. Prezioso, K. K. Likharev, and D. B. Strukov “Fast, Energy-Efficient, Robust, and Reproducible Mixed-Signal Neuromorphic Classifier Based on Embedded NOR Flash Memory Technology,” *IEEE International Electron Devices Meeting (IEDM)*, pp. 151-154, 2017.

[13] W. Xu, S. -Y. Min, H. Hwang, and T. -W. Lee, “Organic core-sheath nanowire artificial synapses with femtojoule energy consumption,” *Science Advances Nanotechnology*, vol. 2, no. 6, Jun. 2016.

[14] A. M. Ionescu and H. Riel, “Tunnel field-effect transistors as energy-efficient electronic switches,” *Nature Review*, vol. 479, pp. 329-337, Nov. 2011.

[15] F.-Y. Wang, J. J. Zhang, X. Zheng, and X. Wang, Y. Yuan, X. Dai, J. Zhang, L. Yang, “Where Does AlphaGo Go: From Church-Turing Thesis to AlphaGo Thesis and Beyond,” *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 2, pp.113-120, 2016.

[16] D. Silver, A. Huang, C. J. Maddison, A. G. Laurent Sifre1, G. van den Driesscher, J. Schrittwieser, I. Antonoglou1, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu1, T. Graepel & D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484-489, Jan. 2016.

[17] B. Rajendran, Y. Liu, J. -S. Seo, K. Gopalakrishnan, L. Chang, D. J. Friedman, and M. B. Ritter, "Specifications of Nanoscale Devices and Circuits for Neuromorphic Computational Systems," *IEEE Transactions on Electron Devices*, vol. 60, no. 1, pp. 246-253, Jan. 2013.

[18] J. -V. Neumann, "First Drast of a Report on EDVAC," *IEEE Annals of the History of Computing*, vol. 15, no. 4, pp. 27-75, 1993.

[19] A. Huang, "Architectural considerations involved in the design of an optical digital computer," *Proceedings of the IEEE*, vol. 72, pp. 780-786, 1984.

[20] J. Backus, "Can programming be liberated from the von Neumann style? : a functional style and its algebra of programs," *Communications of the ACM*, vol. 21, no. 8, pp. 613-641, Aug. 1978.

[21] A. E. Pereda, "Electrical synapses and their functional interactions with chemical synapses," *Nature Reviews Neuroscience*, vol. 15, no. 4, p. 250-263, Apr. 2014.

[22] K. Ahmed, "Brain-Inspired Spiking Neural Networks", *Biomimetics*, pp. 1-25, Aug. 2020. doi: 10.5772/intechopen.93435.

[23] S. Yu, B. Gao, Z. Fang, H. Yu, J. F. Kang, and H.-S. P. Wong, "Stochastic learning in oxide binary synaptic device for neuromorphic computing", *frontiers in Neuroscience*, vol. 7, no. 186, pp.1-9, Oct. 2013.

[24] H. Barlow, "Temporal and spatial summation in human vision at different background intensities," *The Journal of Physiology.*, vol. 141, no. 2, pp. 337-350, Apr. 1958.

[25] D. Oertel, "Use of brain slices in the study of the auditory system: spatial and temporal summation of synaptic inputs in cells in the spatial and temporal summation of synaptic inputs in cells in the anteroventral cochlear nucleus of the mouse," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, pp. 328-333, Feb. 1985.

[26] A. T. Gulledge, B. M. Kampa, and G. J. Stuart, "Synaptic Integration in Dendritic Trees", *Journal of Neurobiology*, vol. 64, pp. 75-90, May. 2005.

[27] E. R. Kandel, "The molecular biology of memory storage: a dialogue between genes and synapses," *Science*, vol. 294, pp. 1030-1038, 2001.

[28] G. -Q. Bi and M. -M. Poo, "Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type," *The Journal of Neuroscience*, vol. 18, no. 24, pp. 10464-10472, Dec.1998.

[29] Y. Dan and M. -M. Poo, "Spike timing dependent plasticity of neural circuits," *Neuron*, vol. 44, pp. 23-30, Sep. 2004.

[30] M. C. Van Rossum, G. Q. Bi, and G. G. Turrigiano, "Stable hebbian learning from spike timing learning from spike timing--dependent plasticity," *The Journal of Neuroscience*, vol. 20, no. 23, pp. 8812-8821, Dec. 2000.

[31] D. O. Hebb, "The organization of behavior: A neuropsychological approach," *John Wiley & Sons*, 1949.

[32] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, pp. 1629-1636, 1990.

[33] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices", *Nature electronics*, vol. 1, pp. 333-343, Jun. 2018.

[34] M. Prezioso, F. M. -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 Letter*, vol. 521, pp. 61-64. May. 2015.

[35] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, vol. 65, no. 6, pp. 386-408, 1958.

[36] F. Rosenblatt, "The Perceptron: A Perceiving and Recognizing Automaton," *Report* 85-60-1, Cornell Aeronautical Laboratory, Buffalo, New York, 1957.

[37] M. W. Gardner and S. R. Dorling, "Artificial Neural Networks (The Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences," *Atmospheric environment*, vol. 32, no. 14/15, pp. 2627-2636, 1998.

[38] M. Riedmiller, "Advanced supervised learning in multi-layer perceptrons - From backpropagation to adaptive learning algorithms," *Computer Standards & Interfaces*., vol. 16, pp. 265–278, 1994.

[39] Y. LeCun, Y. Bengio & G. Hilton, "Deep learning," *Nature Review*, vol. 521, pp. 436–444, May. 2015.

[40] A. L. Maas, A. Y. Hannum, A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *International Conference on Machine Learning (ICML)*, 2013.

[41] P. O'Connor and M. Welling, "Deep Spiking Networks," *arXiv*, *(arXiv:1602.08323v2)* 2016.

[42] B. G. Park, "NeuronIC," *The Magazine of the IEEE*, vol. 44, no. 11, 2017.

[43] F. Zahoor, T. Z. A. Zulkifli, and F. A. Khanday, "Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (MLC) Storage, Modeling, and Applications," *Nanoscale Research Letters*, vol. 15, no. 90, pp. 1-26, Apr. 2020.

[44] S. Cosemans, B. Verhoef, J. Doevenspeck, I.A. Papistas, F. Catthoor, P. Debacker, A. Mallik, and D. Verkest, "Towards 10000TOPS/W DNN Inference with Analog in-Memory Computing – A Circuit Blueprint, Device Options and Requirements," *IEEE International Electron Devices Meeting (IEDM)*, pp. 518-521, 2019.

[45] L. Q. Guo, H. Han, L. Q. Zhu, Y. B. Guo, F. Yu, Z. Y. Ren, H. Xiao, Z. Y. Ge, and J. N. Ding, "Oxide Neuromorphic Transistors Gated by Polyvinyl Alcohol Solid Electrolytes with Ultralow Power Consumption," *ACS Applied Materials & Interfaces*, vol. 11, pp. 28352-28358, 2019.

[46] K. Zhang, K. Sun, F. Wang, Y. Han, Z. Jiang, J. Zhao, B. Wang, H. Zhang, X. Jian, and H. -S. P. Wong, "Ultra-Low Power Ni/HfO2/TiOx/TiN Resistive Random Access Memory With Sub-30-nA Reset Current," *IEEE Electron Device Letters*, vol. 36, no. 10, pp. 1018-1020, Oct. 2015.

[47] H. Tsai, S. Ambrogio, P. Narayanan, R. M. Shelby, and G. W. Burr, "Recent progress in analog memory-based accelerators for deep learning," *Journal of Physics D: Applied Physics*, vo. 51, no. 283001, pp. 1-27, Jun. 2018.

[48] S. Park, M. Chu, J. Kim, J. Noh, M. Jeon, B. H. Lee, H. Hwang, B. Lee, and B. G. Lee, "Electronic system with memristive synapses for pattern recognition," *Scientific Reports*, vol. 5, no.

10123, pp. 1-9, May. 2015.
[49] H. Kim, J. Park, M. W. Kwon, J. H. Lee, and B. G. Park, "Silicon-Based Floating-Body Synaptic Transistor With Frequency-Dependent Short- and Long-Term Memories," *IEEE Electron Device Letters*, vol. 37, no. 3, Mar. 2016.
[50] J. S. Meena, S. M. Sze, U. Chand, and T. -Y. Tseng, "Overview of emerging nonvolatile memory technologies," *Nanoscale Research Letters*, vol. 9, no. 526, pp.1-33, Sep. 2014.
[51] K. D. Suh, B. -H. Suh, Y. -H. Lim, J. -K. Kim, Y. -J. Choi, Y. -N. Koh, S. -S. Lee, S. -C. Kwon, B. -S. Choi, J. -S. Yum, J. -H. Choi, J. -R. Kim, and H. -K. Lim, "A 3.3 V 32 Mb NAND Flash Memory with Incremental Step Pulse Programming Scheme," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 11, pp. 1149-1156, Nov. 1995.
[52] J. Kim, J. Koo, T. Kim, Y. Kim, H. Kim, S. Yoo, and J. -J. Kim. "Area-Efficient and Variation-Tolerant In-Memory BNN Computing using 6T SRAM Array," *Symposium on VLSI Circuits Digest of Technical Papers*, pp. C118-C119, 2019.
[53] D. Ielmini and S. Ambrogio, "Emerging Neuromorphic Devices," *IOP Nanotechnology*, vol. 31, no. 092001, pp. 1-24, 2020.
[54] I. Hubara, M. Courbariaux, D. Soudry, R. E. -Yaniv, Y. Bengio, "Binarized Neural Networks," *NIPS*, 2016.
[55] R. Zhao, W. Song, W. Zhang, T. Xing, J. -H. Lin, M. Srivastava, R. Gupta, and Z. Zhang, "Accelerating Binarized Convolutional Neural Networks with Software-Programmable FPGAs," *NIPS*, 2016.
[56] X. Sun, S. Yin, X. Peng, R. Liu, J. -S. Seo, and S. Yu, "XNOR-RRAM: A Scalable and Parallel Resistive Synaptic Architecture for Binary Neural Networks," *IEEE Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pp. 1423-1428, 2018.
[57] W. Y. Choi, T. Osabe, and T. -J. K. Liu, "Nano-Electro-Mechanical Nonvolatile Memory (NEMory) Cell Design and Scaling," *IEEE Transactions on Electron Devices*, vol. 55, no. 12, Dec. 2008.
[58] S. Salahuddin and S. Datta, "Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices" *ACS NanoLetters*, vol. 8, no. 2, pp. 405-410, 2008.
[59] W. Y. Choi, J. Y. Song, J. D. Lee, Y. J. Park, and B. -G. Park, "70-nm Impact-Ionization Metal-Oxide-Semiconductor (I-MOS) Devices Integrated with Tunneling Field-Effect Transistors (TFETs)," *IEEE International Electron Devices Meeting (IEDM)*, 2005.
[60] W. Y. Choi, B. -G. Park, J. D. Lee, and T. -J. K. Liu, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Letters*, vol. 28, no. 8, Aug. 2007.
[61] D. Esseni, M. Pala, P. Palestri, C. Alper, and T. Rollo, "A review of selected topics in physics based modeling for tunnel field-effect transistors," *IOP Publishing Semiconductor Science and Technology*, vol. 32, no.083005, pp. 1-28, Jul. 2017.
[62] M. Luisier and G. Klimeck, "Simulation of nanowire tunneling transistors: From the Wentzel–Kramers–Brillouin approximation to full-band phonon-assisted tunneling," *Journal of Applied Physics*, vol. 107, no.084507, 2010.
[63] I. Huh, S. Park, M. Shin, and W. Y. Choi, "An Accurate Drain Current Model of Monolayer Transition-Metal Dichalcogenide Tunnel FETs," *IEEE Transactions on Electron Devices*, vol. 64, no. 8, pp. 3502-3507, Aug. 2017.
[64] R. Jhaveri, V. Nagavarapu, and J. C. S. Woo, "Effect of Pocket Doping and Annealing Schemes on the Source-Pocket Tunnel Field-Effect Transistor," *IEEE Transactions on Electron Devices*, vol. 58, no. 1, Jan. 2011.
[65] A. S. Verhulst, W. G. Vandenberghe, K. Maex, S. D. Gendt, M. M. Heyns, and G. Groeseneken, "Complementary Silicon-Based Heterostructure Tunnel-FETs With High Tunnel Rates," *IEEE Electron Device Letters*, vol. 29, no. 12, pp. 1398-1401, 2008.
[66] W. Y. Choi, and W. Lee, "Hetero-Gate-Dielectric Tunneling Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 9, Sep. 2010.
[67] J. W. Lee and W. Y. Choi, "Design Guidelines for Gate-Normal Hetero-Gate-Dielectric (GHG) Tunnel Field-Effect Transistors (TFETs)," *IEEE Access*, vol. 8, pp. 67617-67624, Apr. 2020.
[68] International Technology Roadmap for Semiconductors (ITRS). Accessed: Sep. 14, 2019. [Online]. Available: http://www.itrs2.net/itrs-reports.html
[69] S. Blaeser, S. Glass, C. Schulte-Braucks, K. Narimani, N. v. d. Driesch, S. Wirths, A. T. Tiedemann, S. Trellenkamp, D. Buca, Q. T. Zhao, S. Mantl, "Novel SiGe/Si line tunneling TFET with high Ion at low VDD and constant SS," *IEEE International Electron Devices Meeting (IEDM)*, pp. 608-611, 2015.
[70] Y. Kondo, M. Goto, K. Miyano, A. Hokazono, T. Miyata, E. Sugizaki, K. Adachi, T. Ohguro, S. Kawanaka and Y. Toyoshima, "Novel SiGe/Si line tunneling TFET with high I_{on} at low V_{DD} and constant SS," *International Conference on Solid State Devices and Materials (SSDM)*, pp. 746-747, 2013.

AUTHOR BIOGRAPHIES

Jang Woo Lee was born in Seoul, in 1988. He received the B.S. and Ph. D. degrees in 2014 and 2021 in the Department of Electronic Engineering from Sogang University, Seoul, Korea. His current research interests include tunneling field-effect transistors (TFET), hardware-based neuromorphic-computing, and nanoscale novel devices.

Jae Seung Woo was born in Seoul, in 1995. He received the B.S. degree in 2020 from Sogang University, Seoul, Korea. He is currently working toward the M.S. degree in the Department of Electronic Engineering, Sogang University, Seoul, Korea. His current research interests include tunneling field-effect transistors (TFET), hardware-based neuromorphic-computing, and nanoscale novel devices.

Woo Young Choi received the B.S., M.S. and Ph. D. degrees in the School of Electrical Engineering from Seoul National University, Seoul, Korea in 2000, 2002 and 2006, respectively. From 2006 to 2008, he was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA as a post-doctor. Since 2008, he has been a member of the faculty of Sogang University (Seoul, Korea), where he is currently a Professor with the Department of Electronic Engineering. He has authored or coauthored over 320 papers in international journals and conference proceedings and holds 55 Korean/US patents. His current research interests include design, fabrication, characterization and theoretical analysis of nanoscale MOSFETs and emerging extremely-low power logic/analog/memory devices, three-dimensional integration and neuromorphic computing.