Check for updates

ADVANCED SCIENCE Open Access
www.advancedscience.com

Proposition of Adaptive Read Bias: A Solution to Overcome Power and Scaling Limitations in Ferroelectric-Based Neuromorphic System

Ryun-Han Koo, Wonjun Shin, Seungwhan Kim, Jiseong Im, Sung-Ho Park, Jong Hyun Ko, Dongseok Kwon, Jae-Joon Kim, Daewoong Kwon, and Jong-Ho Lee**

Hardware neuromorphic systems are crucial for the energy-efficient processing of massive amounts of data. Among various candidates, hafnium oxide ferroelectric tunnel junctions (FTJs) are highly promising for artificial synaptic devices. However, FTJs exhibit non-ideal characteristics that introduce variations in synaptic weights, presenting a considerable challenge in achieving high-performance neuromorphic systems. The primary objective of this study is to analyze the origin and impact of these variations in neuromorphic systems. The analysis reveals that the major bottleneck in achieving a high-performance neuromorphic system is the dynamic variation, primarily caused by the intrinsic 1/*f* noise of the device. As the device area is reduced and the read bias (V_{Read}) is lowered, the intrinsic noise of the FTJs increases, presenting an inherent limitation for implementing area- and power-efficient neuromorphic systems. To overcome this limitation, an adaptive read-biasing (ARB) scheme is proposed that applies a different V_{Read} to each layer of the neuromorphic system. By exploiting the different noise sensitivities of each layer, the ARB method demonstrates significant power savings of 61.3% and a scaling effect of 91.9% compared with conventional biasing methods. These findings contribute significantly to the development of more accurate, efficient, and scalable neuromorphic systems.

1. Introduction

The von Neumann architecture, which has served as the foundation of computing architecture for decades, is countering challenges owing to the increasing prevalence of data-centric cognitive tasks such as computer vision, natural language processing, and autonomous driving.[1–3] These data-centric cognitive tasks require frequent and inefficient transfers between the processor and memory, which consume significant time and power. Neuromorphic computing, which mimics the structure of the human brain, is considered a promising solution.[1–6] The human brain contains many synapses that interact to compute and store data simultaneously. Therefore, for the successful implementation of neuromorphic systems, the appropriate physical devices must be selected to mimic biological synapses.

Numerous device candidates are available for the implementation of neuromorphic systems. The key difference between them is the representation of the degree of activation of the biological synaptic weights. For instance, resistive random-access memory (ReRAM) stores information as a resistance,[7–10] phase-change memory (PCM) stores it as a phase change,[11–14] magnetic tunnel junction (MTJ) stores it magnetically,[15–18] and ferroelectric (FE)-based memory stores it in a polarization state.[19–23] To realize highly integrated, low-power, high-performance neuromorphic systems, synaptic devices require high integration, low current density, high reliability, and high operating speed. However, PCM and MTJ suffer from high power consumption owing to their high current density, and ReRAMs suffer from reliability issues when learning frequently.[10,14,18] In contrast, ferroelectric tunnel junctions (FTJs), which have the advantages of fast operation speed, low current density, high reliability, process simplicity, and high integration ($4F^2$, F: Minimum feature size), are considered the strongest candidates for implementing synapses in neuromorphic computing systems.[20–25] This movement gained significant traction following the breakthrough discovery of ferroelectricity in hafnium oxide (HfO_2) thin films in 2011, which was fully compatible with conventional CMOS processes. Subsequently,

R.-H. Koo, W. Shin, S. Kim, J. Im, S.-H. Park, J. H. Ko, D. Kwon, J.-J. Kim, J.-H. Lee
Inter-University Semiconductor Research Center
Department of Electrical and Computer Engineering
Seoul National University
Seoul 08826, South Korea
E-mail: jhl@snu.ac.kr
D. Kwon
Department of Electrical Engineering
Hanyang University
Seoul 04763, South Korea
E-mail: dw79kwon@hanyang.ac.kr
J.-H. Lee
Ministry of Science and ICT
Sejong 30109, South Korea

© 2023 The Authors. Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202303735

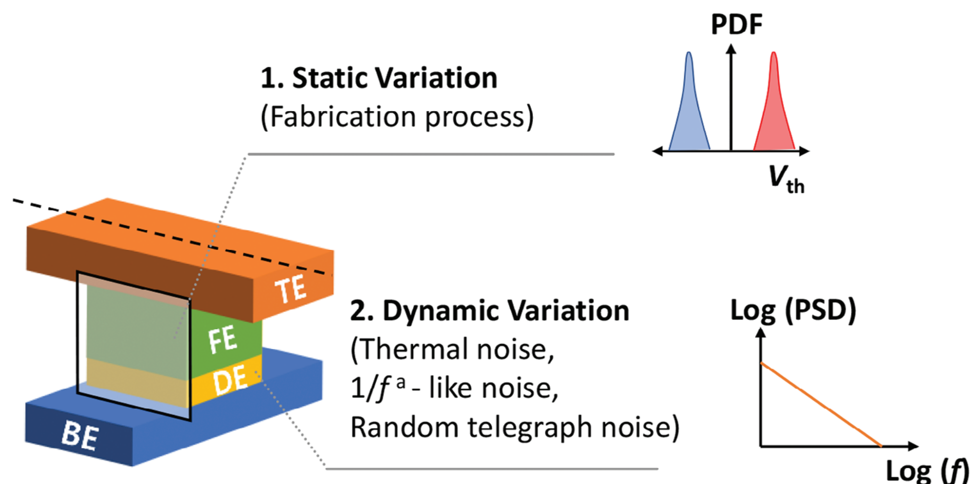


Figure 1. Schematic diagram of the variation in the HfO_2-FTJ variation sources. a) Static variation, which remains constant over time. The device fabrication process induced device-to-device static variation. Inset shows a schematic diagram of the probability density function (PDF) versus threshold voltage (V_{th}) of different memory states. b) Dynamic variation, which changes continuously over time. Device-intrinsic read noise, such as thermal noise, $1/f$-like noise, and random telegraph noise (RTN), is the source of dynamic variation. The inset shows a schematic diagram of power spectral density (PSD) versus frequency (f).

numerous studies have reported the possibility of implementing neuromorphic systems using HfO_2-FTJs.[22–25]

To implement neuromorphic systems based on electrical devices such as HfO_2-FTJ, reliability is a crucial challenge that limits the performance and accuracy of large-scale arrays. Therefore, several studies have analyzed the impact of device reliability, such as nonlinearity, asymmetry, dynamic range, endurance, retention, read/write disturbances, and variations in neuromorphic systems.[26–29] Among these, variation becomes critical as neuromorphic systems are integrated at large scales. This can be analyzed from three perspectives. First, at the single-device level, implementing multiple synaptic conductances reduces the number of synaptic devices used in the system and enhances network performance. However, device variation makes it difficult to distinguish between different conductances.[30–32] Second, at the system level, the human brain has more than 10^{15} synapses, and neuromorphic systems that imitate them increase the number of synapses exponentially as the amount of data to be addressed increases. Because it is impossible for such a large number of synapses to have the same electrical properties, system-level variation is introduced, which is a critical issue for large-array operations in practice.[33–35] Finally, to achieve highly integrated neuromorphic systems, the synaptic elements are gradually scaled down, leading to increased variation at both the single-device and system levels. Therefore, it is necessary to properly evaluate the impact of reliability issues, particularly variations, as we move toward the implementation and miniaturization of neuromorphic systems.[36,37]

To understand how variations affect neuromorphic systems, a deeper understanding of the types of device variations is required. Here, it is imperative to differentiate between variation, which encompasses random fluctuations in device parameters, and cycle-induced degradation or retention degradation, which are non-random, directional phenomena with distinct causes and effects. This distinction is made because the essence of variation lies in the random distribution of a particular parameter. In contrast, cycle-induced variation or retention degradation does not exhibit randomness; instead, it follows a specific degradation pattern. For instance, retention characteristics degrade in a specific direction, where the memory window narrows or closes. Additionally, cycle-induced degradation leads to the collapse of the memory window, which is not a random occurrence but rather a deterministic phenomenon with a clear cause-and-effect relationship. **Figure 1** shows the different sources of variation in the HfO_2-FTJs. These variations are divided into two categories based on their behavior over time. Static variation maintains a constant origin that remains consistent over time, whereas the source of dynamic variation undergoes continuous changes as time progresses. During device fabrication, various factors such as SiO_2 and HfO_2 film thickness, trap density, and size of ferroelectric domains in the FE layer vary.[25,38,39] These are expressed as device-to-device variations, classified as static variations, in that they do not change over time. However, there is variation due to device intrinsic read noise, such as thermal noise, $1/f$-like noise, and random telegraph noise (RTN), which is classified as dynamic noise because it changes over time.[40–49]

Because different neuromorphic system structures have different sensitivities to each variation, careful consideration of the system architecture is required to evaluate the impact of these variations on neuromorphic systems.[29] Neuromorphic systems can be classified into two categories based on their training methods. In off-chip training, the software-trained weight map is pre-encoded to the conductance of the physical devices.[50,51] In contrast, on-chip training is based on real-time system-level conductance adjustment by calculating the error between the target value and the vector-matrix multiplication (VMM) result.[51,52] These two methods have fundamental differences in their learning processes; therefore, the effects of these variations are also different. From the above discussion, the performance of neuromorphic systems must be re-evaluated in different training scenarios (off-chip and on-chip), considering their static and dynamic variations, as well as other non-idealities of physical devices.

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

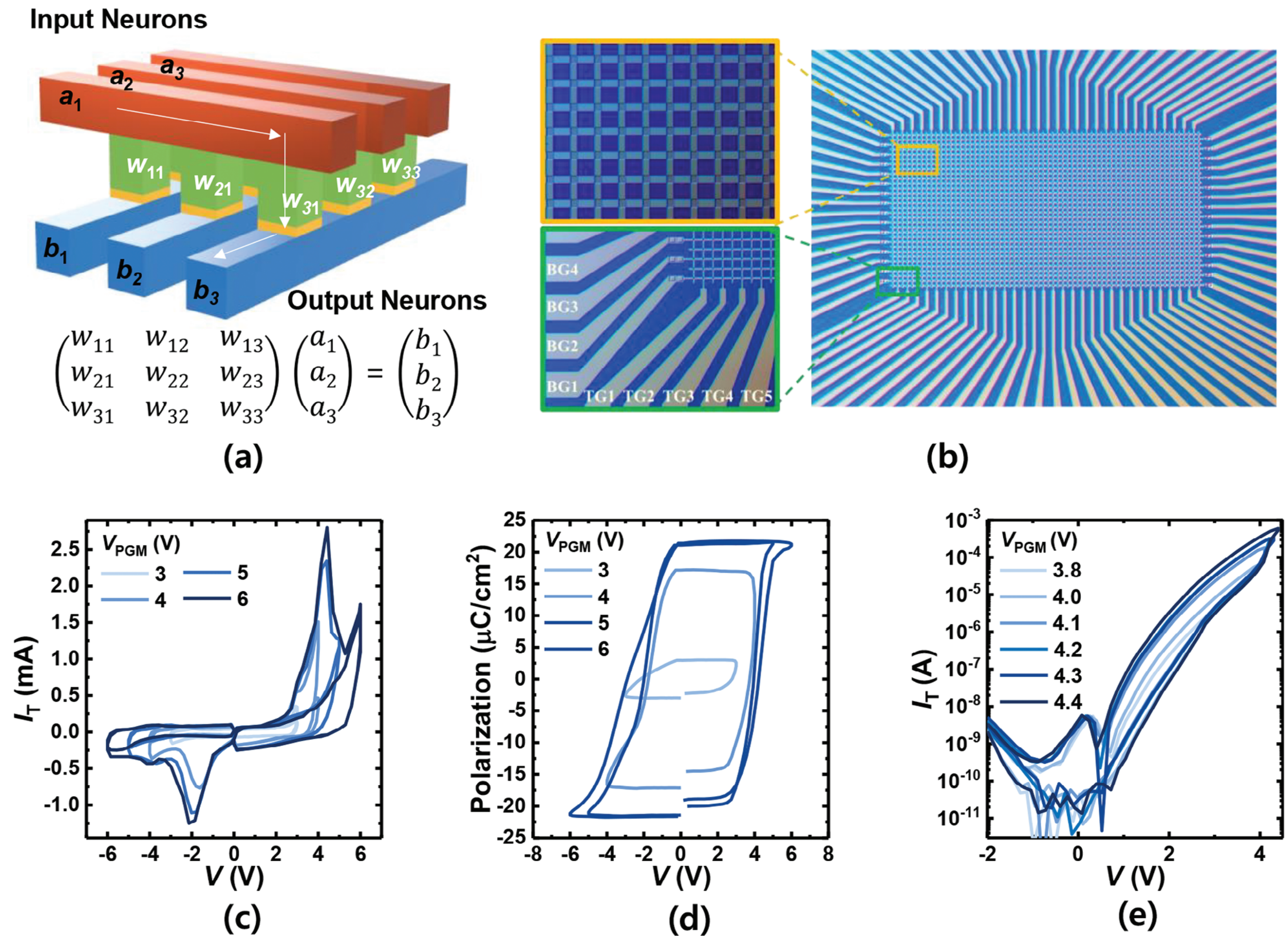


Figure 2. Electrical characteristics of HfO_2-FTJ. a) Schematic diagram of the matrix array for a neuromorphic system based on HfO_2-FTJ. b) Optical microscopic image of the 36 × 24 HfO_2-FTJ matrix array. c) PUND measurement results with different V_{PGM}. For PUND measurement, 10 µs ramp time, and 1 µs hold time is used. d) Polarization versus V as a parameter of V_{PGM}. e) Double sweep I_T versus V as a parameter of V_{PGM}.

Furthermore, it is necessary to propose methods that reduce variations at the device level and minimize their impact at the system level, as highly integrated hardware neuromorphic systems are being developed on a large scale.[30–37]

In this study, we examine the effects of static and dynamic variations on a hardware neuromorphic system based on HfO_2-FTJs. A 32 × 24 HfO_2-FTJ array is fabricated to analyze the static and dynamic variations. In addition to the electrical characteristics of the device, low-frequency noise (LFN) spectroscopy is used to understand the causes of variation. By rigorously studying the origin of the LFN of the HfO_2-FTJs, we propose an innovative approach to mitigate the dynamic variation at the device level: Applying a high read bias (V_{Read}) to the device. This reduces the dynamic variation, which increases with device downsizing, potentially overcoming the scaling limit of FTJs. However, at the system level, increasing V_{Read} leads to higher power consumption, resulting in a trade-off between power consumption and synaptic density. To address this issue, we propose a novel biasing method termed “adaptive read bias” (ARB) to mitigate the impact of dynamic variation at the system level. By utilizing the ARB method, we achieve a 61.3% reduction in power consumption and a 91.9% scaling effect without compromising the accuracy of the overall neuromorphic system. Therefore, the ARB method provides a general solution for implementing power- and area-efficient neuromorphic systems.

2. Results and Discussion

2.1. Electrical Characteristics of the TiN/HfO_2/SiO_2/n^+-Si FTJ

Figure 2a shows a schematic of the synaptic array used for neuromorphic computing. The crossbar array of the FTJ can be used as the matrix component of an artificial neural network consisting of a two-terminal artificial synapse connecting the pre- and post-synaptic neurons. The FTJ array mimics synaptic transmission by multiplying the input neuron signal (encoded as the voltage applied to the device) by the corresponding weight (conductance) and passing the multiplication product (resulting current) to the output neuron. Dense, fast, and power-efficient VMM computation is thus inherently possible at the physical level using Ohm’s

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

and Kirchoff's laws.[1–6] Figure 2b shows an optical microscopy image of a 32 × 24 crossbar array with HfO_2-based FTJ nodes. The FTJ cells are positioned at the intersections of perpendicular Ti/TiN/Al/TiN electrode lines. The fabrication process is illustrated in Figure S1, Supporting Information. Figure S2, Supporting Information shows a cross-sectional transmission electron microscopy (TEM) image of the (metal-ferroelectric-insulator-semiconductor) MFIS HfO_2-FTJ, where the thicknesses of the TiN, HfO_2, and SiO_2 are 100, 6, and 1.2 nm, respectively. Table S1, Supporting Information shows the area of the fabricated device, the area where power spectral density (PSD) measurements are performed, and the area of the theoretically studied device.

To emulate the learning mechanisms of the biological brain effectively, a hardware neuromorphic system requires additional functionality beyond the VMM operations mentioned above. The postsynaptic current of each device cell should be gradually modified based on the history of the applied pulse spikes from the presynaptic neuron, enabling learning and memory functions.[12,53] To achieve a gradual conductance change (multiple conductance levels) in response to input spikes, FE-based devices exploit the partial polarization of the FE layer. In Figure 2c, the positive-up-negative-down (PUND) measurement results of the FTJ demonstrate the existence of ferroelectric switching in the fabricated HfO_2 layer owing to the polarization current. Figure 2d shows the relationship between the polarization (P) and voltage applied to the top gate (V) as a parameter of the maximum program voltage (V_{PGM}). As V_{PGM} increases, the remnant polarization (P_r) also increases, indicating partial polarization property.[54] Figure 2e shows the DC double-sweep tunneling current (I_T) versus V as a function of V_{PGM}. Because of the partial polarization of the FE layer, the tunneling electroresistance (TER) ratio increased with increasing V_{PGM}. These results show that when an input voltage is applied to synaptic devices for learning, not all domains of the ferroelectric layer are switched at once, but rather, they are switched in part, demonstrating that multiple conductance levels can be achieved for neuromorphic devices.

In addition to multiple conductances, nonlinearity plays a critical role in the implementation of high-performance neuromorphic systems.[29] Nonlinearity determines the rate at which the conductance changes with the number of voltage inputs, thereby affecting the training accuracy. If nonlinearity is significant, the updated value may deviate from the intended value, resulting in detrimental effects on the learning speed and accuracy of the neuromorphic system. In FE-based devices, nonlinearity depends on how the input voltage pulse is applied to update the conductance; therefore, we experiment with two schemes: Identical and incremental pulse schemes. **Figure 3**a shows the dependence of P_r on V_{PGM} (black symbols) and T_{PGM} (red symbols). The change in P_r with V_{PGM} is related to the conductance change in the incremental pulse scheme and the P_r change with T_{PGM} is related to the constant pulse scheme. In general, the constant pulse scheme is easier to implement, whereas the incremental pulse scheme provides higher accuracy.[55–57] Figures 3b,c show the behavior of long-term potentiation (LTP) and long-term depression (LTD) using identical and incremental pulse schemes, respectively. Figure S3, Supporting Information shows P–V, endurance, retention, and LTP-LTD characteristics of the HfO_2-FTJ array. Figure S4, Supporting Information shows the programming operation scheme in the HfO_2-FTJ array. The nonlinearity of each update scheme can be extracted as follows.[29]

$$W_{LTP} = g_{min} + \alpha \left(1 - e^{\beta x}/N\right) \tag{1}$$

$$W_{LTD} = g_{max} - g_{min} - \alpha \left(1 - e^{\beta x}/N\right) \tag{2}$$

$$\alpha = \left(g_{max} - g_{min}\right) / \left(1 - e^{-\beta}\right) \tag{3}$$

where g_{max} and g_{min} are the maximum and minimum conductances, respectively, and β is the nonlinear factor. N is the total number of pulses applied during the LTP (LTD). While the identical pulse scheme has a positive β for both LTP and LTD, the incremental pulse scheme has a negative β for LTP and a positive β for LTD. This is due to the partial polarization characteristics of the FE layer. As shown in Figures 3b,c, the first pulse in the identical pulse scheme induces a larger polarization switching compared to the incremental pulse scheme, resulting in a convex upward LTP curve with positive β. However, in the incremental pulse scheme, the conductance increases with V_{PGM}, forming a downward convex LTP curve, and β is negative. This negative β in LTP is known to be a good factor for neuromorphic computing because of its symmetry with the LTD characteristics.[56,57] Now, we compare the learning accuracy of an identical pulse scheme and an incremental pulse scheme. Figure S4, Supporting Information shows the average learning accuracy for the ten learning curves using identical and incremental pulse schemes. In both cases, the learning accuracy increased with some fluctuations in the epoch. The cause of this fluctuation is that the intended weight update value is distorted by the nonlinearity of the synaptic devices.[29] In the case of an incremental pulse, the scheme exhibits less fluctuation and faster convergence. This is because the mismatch between LTP and LTD is less than that in an identical pulse scheme. Figure 3d shows the accuracy of the identical (red box plot) and incremental (blue box plot) pulse schemes for the 40th (early stage) and 100th (late stage) epochs. The learning curves are shown in Figure S5, Supporting Information. As expected, the accuracy of the 40th epoch with the incremental pulse scheme is higher than that of the identical pulse scheme owing to the fast convergence speed with symmetric LTP-LTD characteristics. However, in the 100th epoch, the identical pulse scheme fully converged, and there was no difference in the final accuracy. Although a difference occurred in the convergence speed, we believe that the identical pulse method is suitable for on-chip learning because there is no significant difference in the final accuracy, and the hardware and software costs are very low compared to those of the incremental pulse method. To implement a neuromorphic system using an incremental pulse scheme, it is necessary to store the historical data of the applied pulses in each device, which requires significant memory and computing power. This becomes even more challenging considering that even a light MobileNet model consists of over one million parameters.[58–60] Consequently, employing an incremental pulse scheme in real-world applications is not feasible. Consequently, this study relies on simulations using an identical pulse scheme to examine the impact of the variations.

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

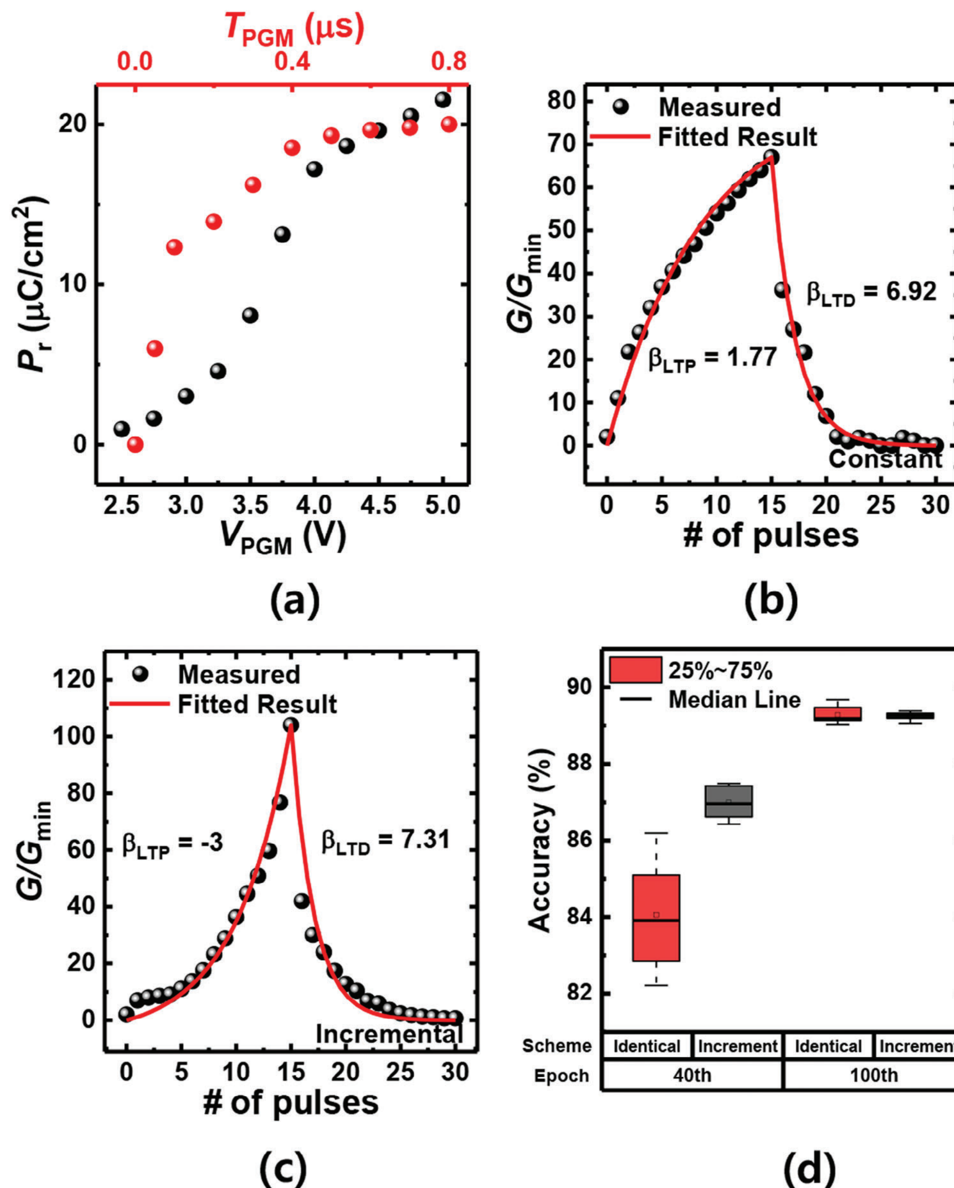


Figure 3. Synaptic characteristics of HfO_2-FTJ. a) P_r versus V_{PGM} (black symbol) and T_{PGM} (red symbol). For the P_r versus V_{PGM} measurement, T_{PGM} is fixed at 1 μs. For the P_r versus T_{PGM}, V_{PGM} is fixed at 5 V. b) LTP-LTD characteristics of HfO_2-FTJ with an identical pulse scheme. An identical pulse of 5 V, 1 μs is applied. The red line shows the fitting results of nonlinearity. c) LTP-LTD characteristics of HfO_2-FTJ with incremental pulse scheme. Program pulses with increasing amplitude of 4.0 to 5.5 V, 1 μs are applied. The red line shows the fitting result of the nonlinearity. d) Accuracy of the network trained with identical (red box plot) and incremental pulse schemes (blue box plot) for the 40th and 100th epochs.

2.2. Origin of HfO_2-based FTJ Variation and Its Impact on the Hardware Neuromorphic System

In addition to nonlinearity and asymmetry, we analyze the effect of variations (static and dynamic variations) on hardware neuromorphic systems depending on the learning methods (on-chip and off-chip learning methods). Figure S6, Supporting Information shows a schematic diagram of the simulation methodology. First, we analyze the effects of static variation on the neuromorphic system. We measured 100 devices with areas of 100 × 100, 75 × 75, 50 × 50, and 35 × 35 μm^2. **Figure 4**a shows the DC double-sweep I_T versus V as a function of the area. As the area is scaled down, I_T decreases because the current density remains constant. Notably, the FTJ has a very low current density compared with other 2-terminal memories (ReRAM, PCM, and MTJ), highlighting its potential for extremely low-power operation of hardware neuromorphic systems. Figure 4b shows the correlation between the area and the normalized current variation (σ_{IT}/I_T) of the FTJs in the low-resistive state (LRS) and high-resistive state (HRS). As the device is scaled down, the fabrication-induced variation increases because variables such as the interface trap density, HfO_2 film thickness, and number of domains become less uniform.[36,37] These fabrication process-induced variations are categorized as static because their origins are established

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

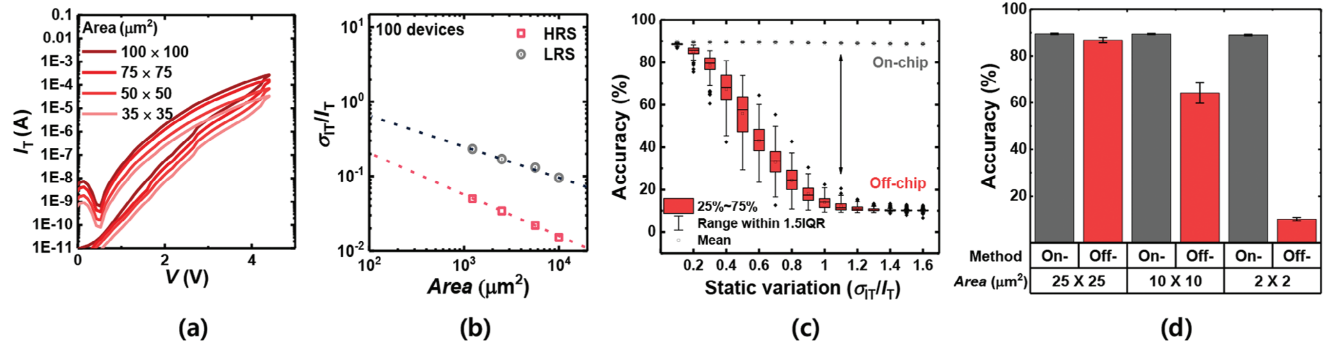


Figure 4. Impact of the static variation on the hardware neuromorphic system. a) Double sweep I_T versus V as a parameter of area. b) Correlation between the area and normalized current variation (σ_{IT}/I_T) of 100 HfO_2-FTJs in LRS and HRS. c) Accuracy after performing on-chip and off-chip learning in the presence of static variation. d) Final accuracy of the network with respect to learning method and area of the devices.

during the fabrication process itself. Depending on how the neuromorphic system is trained, its sensitivity to static variation also varies; therefore, we investigate the effect of static variation using on-chip and off-chip training methods. Figure 4c show the accuracy after off-chip and on-chip training versus the static variation. The results shows that static variation degraded the accuracy of the off-chip training method, whereas the accuracy of the on-chip training remained constant. This significant difference arises from the consideration of static variations during training. In off-chip learning, the weight map trained by the software is pre-encoded to the conductance of physical devices. Therefore, if a static variation exists, the parameter weights learned from the external computer are not correctly transferred, resulting in accuracy degradation. In contrast, on-chip training is based on a real-time system-level conductance adjustment by calculating the error between the target value and the VMM result. Therefore, unlike off-chip training, on-chip training is robust to this static variation because the training is conducted in an environment where there is already device-to-device variation, and the weight values are determined to account for this variation.[29,61]

Figure 4d shows the results of neural network simulations with different training schemes as the FTJs are scaled down. When the off-chip training is applied, the accuracy is maintained up to 25 × 25 μm², but as the device is scaled down, the accuracy decreases. At 2 × 2 μm², it returns to its initial state without training (accuracy ≈ 10%). However, in the case of on-chip training, there is no decrease in accuracy as the device is scaled down. In addition, on-chip training, which updates the weight and performs VMM computation on the same hardware platform, does not require extracting or encoding the weight between the software and hardware, ultimately improving the area/energy efficiency.[50,51] Therefore, on-chip training is essential for implementing size- and power-efficient neural networks with further downscaling.

Subsequently, we analyze the impact of intrinsic device noise (dynamic variation) on an on-chip neuromorphic system. In semiconductor devices, intrinsic device noise such as $1/f$ noise and random telegraph noise (RTN) causes dynamic variation. This is because these noise types originate from random processes generated by carrier trapping or fluctuations in barrier height, and their behavior is influenced by various defects unique to each process.[40–49,62–65] **Figure 5**a shows the normalized transient I_T ($\frac{I_T - I_{avg}}{I_{avg}}$, I_{avg} is the average I_T) counts in the FTJ in the a—1) HRS and a—2) LRS. In all cases, I_T follows a Gaussian distribution, and the standard deviation increases with device downscaling. To reveal the physical origin of these current fluctuations, the power spectral density (PSD) is measured for different areas (100 × 100, 75 × 75, 50 × 50, and 35 × 35 μm²). Figure S5, Supporting Information shows the setup for LFN spectroscopy. Figure 5b shows the normalized PSD ($S_{IT}/I_T{}^2$) as a function of the device area in the b—1) HRS and b—2) LRS. As shown in Figure 5b, the FTJs exhibit $1/f^{\gamma}$ ($\gamma = -\Delta ln(S)/\Delta lnf$) noise behavior in all cases. Furthermore, the γ value of FTJs is close to one regardless of the area. In addition, $S_{IT}/I_T{}^2$ increases with device downscaling. By analyzing the I_T–V characteristics at different temperatures and performing LFN spectroscopy, we demonstrate that the $1/f$ noise characteristics of FTJ in the HRS originate from shot noise at low V_{Read} and barrier height fluctuation (BHF) at high V_{Read}.[24] The detailed procedure is shown in Figure S8 and Text S1, Supporting Information. The shot noise owing to the direct tunneling current can be expressed as follows.[66,67]

$$S_{IT} = 2qFI_T \tag{4}$$

where q is the elementary charge and F is the Fano factor related to the time constant of the trapping and detrapping processes.

$1/f$ noise generated from the BHF can be expressed as follows.[68]

$$\frac{S}{I_T^2} \propto \frac{1}{fA} exp\left(\frac{qV_{bi}}{k_B T}\right) \tag{5}$$

where A is the area and V_{bi} is the built-in potential.

By extracting the trap energy level during current conduction and LFN spectroscopy, it is revealed that the $1/f$ noise of the FTJ in the LRS was originated from the PF emission noise in all V_{Read}. The detailed procedure is shown in Figure S9 and Text S2, Supporting Information. When $1/f$ noise is generated from the PF emission, $S_{IT}/I_T{}^2$ is expressed as follows.[69]

$$\frac{S}{I_T^2} = \frac{\delta^2 N_D}{E\epsilon_{ins}^2 WL} \frac{q^2 R}{f} \tag{6}$$

© 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

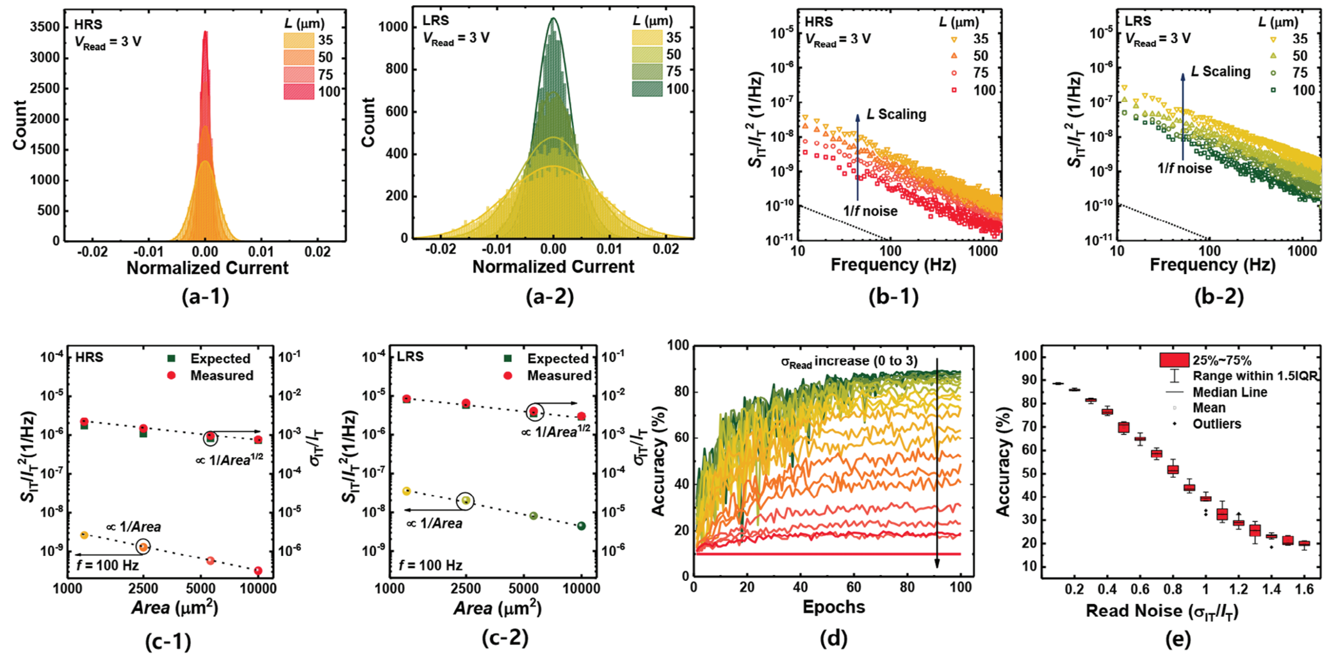


Figure 5. Area dependence of the dynamic variation and its impact on the hardware neuromorphic system. a) Normalized transient current counts in the HfO_2-FTJ in a—1) HRS and a—2) LRS. b) S_{IT}/I_T^2 versus f as a parameter of the device area in b—1) HRS and b—2) LRS. c) S_{IT}/I_T^2 (left axis) and σ_{IT}/I_T^2 (right axis) with respect to the device area in c—1) HRS and c—2) LRS. d) Learning accuracy of the VGG-11 network based on HfO_2-FTJs with identical pulse scheme as a parameter of dynamic variation (0 to 3). e) Final accuracy obtained during the learning process (10 times) with respect to σ_{IT}/I_T.

where δ denotes the fitting parameter, N_D denotes the trap density, E denotes the electric field, ϵ_{ins} denotes the permittivity of the insulator layer, and R denotes the ratio of trap time constants. In Equations (5) and (6), S_{IT}/I_T^2 is inversely proportional to the area, so the S_{IT}/I_T^2 increases as the device is downscaled. The left axis of Figure 5c shows the relationship between S_{IT}/I_T^2 and area. Note that S_{IT}/I_T^2 is sampled at 100 Hz. In both the LRS and HRS, S_{IT}/I_T^2 is proportional to $1/A$, which is consistent with theoretical predictions. The right axis of Figure 5c shows the relationship between the normalized standard deviation of I_T (σ_{IT}/I_T) and area. The measurement results show a $1/A^{1/2}$ dependency. If the random fluctuation of I_T originated from the intrinsic device noise, not resistance degradation or external noise, it is possible to calculate σ_{IT}/I_T from the plot of S_{IT}/I_T^2 versus frequency. The σ_{IT} can be calculated by integrating the spectrum S according to the following equation.[70–72]

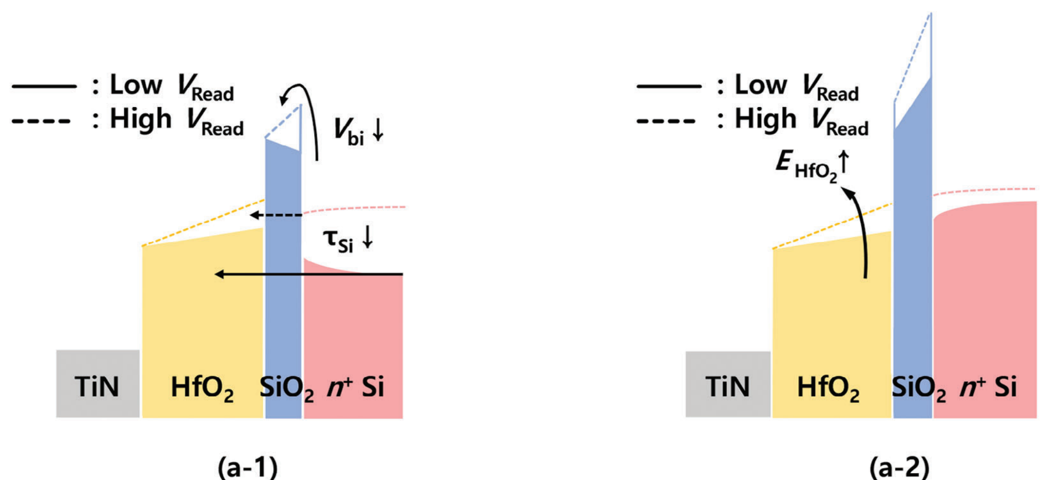
$$\sigma = \int_{f_{min}}^{f_{max}} S df \tag{7}$$

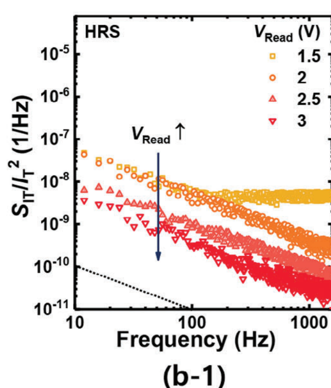
where $f_{min} = t^{-1}$, $f_{max} = 2t_s^{-1}$, and t_s is the sampling time. We calculate the expected values σ_{IT}/I_T using the measurement results of S_{IT}/I_T^2 and depict them as green points in Figure 5c. In both the LRS and HRS, the expected σ_{IT}/I_T and measured σ_{IT}/I_T show excellent consistency, demonstrating that the random fluctuation of current originated from device intrinsic noise, not resistance degradation or external noise.[71] Because a neuromorphic system utilizes currents to represent the results of VMMs, random fluctuations in the current degrade the accuracy of the neuromorphic system. In particular, device intrinsic noise, whose magnitude is increased with area scaling, may serve as a major bottleneck in the implementation of dense, power-efficient neuromorphic systems. Figure 5d shows the learning accuracies of the VGG-11 networks based on FTJs with identical pulse schemes as a parameter of dynamic variation. Unlike static variation, accuracy degradation due to dynamic variation cannot be recovered, even in the case of on-chip learning. This is because, unlike static variation, dynamic variation changes with time; therefore, it cannot be compensated for during the learning process. Even if learning is performed for a specific noise direction, a completely different random noise fluctuation occurs in the next time step, such that the accuracy is saturated even if learning is fully performed. Figure 5e shows the final accuracy obtained during the learning process (10 iterations) with respect to σ_{IT}/I_T. As mentioned previously, dynamic variation is a source of accuracy degradation. We also perform a simulation of VGG-11 networks based on 1×1 μm^2 FTJs ($V_{Read} = 2$ V) using an identical pulse scheme. Thus, there is a serious accuracy degradation from 90% accuracy to 48.94% in the presence of intrinsic noise. If we use a larger device to reduce the accuracy degradation, the device should scale up to 17×17 μm^2 to achieve an accuracy drop of less than 3%. Therefore, it is crucial to determine a method to reduce the intrinsic noise of a device to implement an area-efficient hardware neuromorphic system. The key to reducing the intrinsic device noise of an FTJ is to clearly understand the noise mechanism.

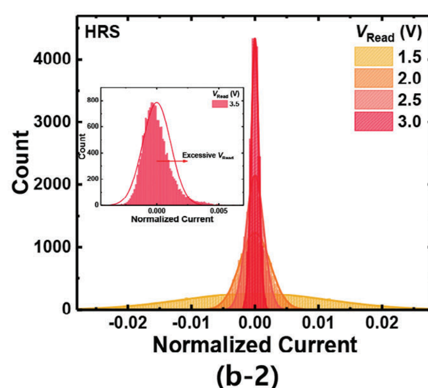
Figure 6a shows a schematic energy band diagram of the HfO_2-FTJ. The dotted line indicates the band diagram obtained when a high V_{Read} is applied. As V_{Read} increased, V_{bi} and the time constant for electron trapping from Si (t_{Si}) decreased in the HRS

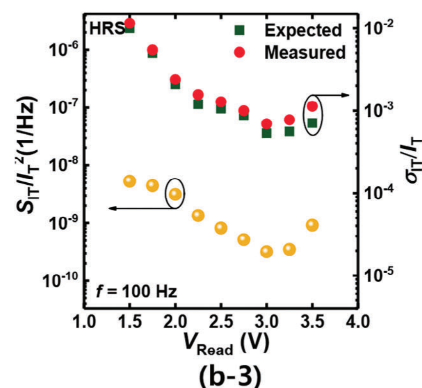
 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

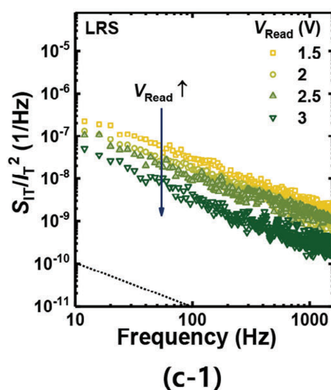
21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

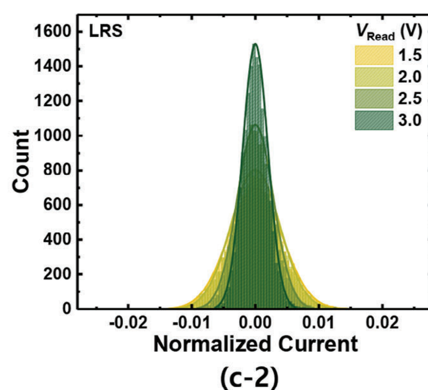


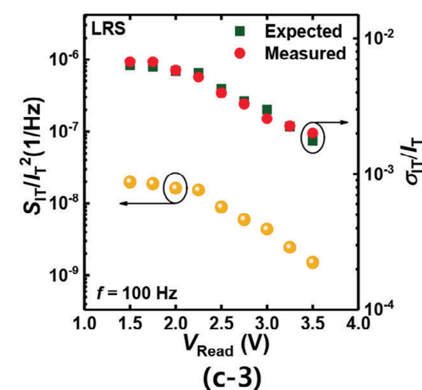


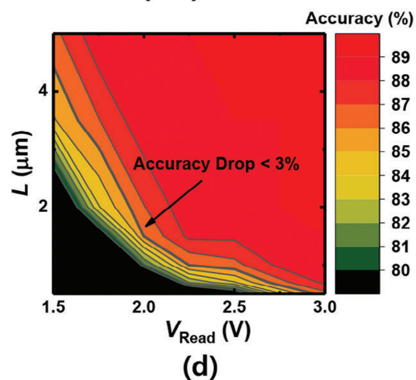


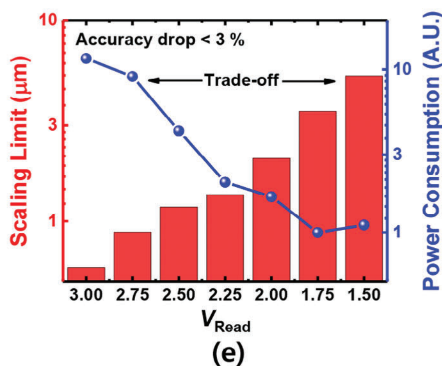


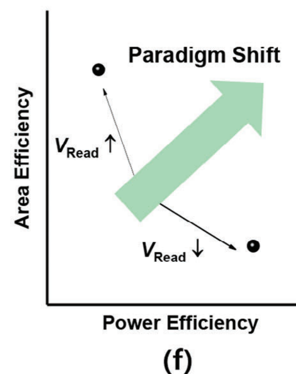












© 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

device, and the E-field in the HfO_2 layer (E_{HfO2}) increased in the LRS device. Figure 6b–1 shows S_{IT}/I_T^2 of the FTJ in the HRS as a function of V_{Read}. The behavior of S_{IT}/I_T^2 changes from shot noise to $1/f$ noise as V_{Read} is increased from 1.5 to 2.0 V. τ_{Si} is compared to the time constant for electron detrapping to TiN (τ_{TiN}), so τ_{Si} determines F. By increasing V_{Read}, the value of τ_{Si} decreases, which also decreases F, as shown in Figure 6b–1 ($V_{Read} = 1.5$ to 2.0 V). Furthermore, when shot noise is minimized owing to a high V_{Read}, the presence of $1/f$ noise owing to the BHF becomes apparent.[29] Thus, S_{IT}/I_T^2 at $V_{Read} = 2.0$ V shows $1/f$ noise characteristics. When S_{IT}/I_T^2 follows the $1/f$ noise from the BHF, the magnitude of the noise decreases as V_{bi} decreases with increasing V_{Read}. Figure 6b–2 shows the normalized transient I_T counts in the FTJ of the HRS as a function of V_{Read}. As V_{Read} increases, σ_{IT}/I_T decreases, similar to the trend observed in Figure 6b–3. Note that as V_{Read} changes from 1.5 to 2.0 V, σ_{IT}/I_T decreases rapidly as the LFN changes from shot noise to $1/f$ noise. The inset of Figure 6b–2 shows the normalized transient I_T counts at V_{Read} = 3.5 V. Unlike when V_{Read} ranges from 1.5 to 3.0 V, the I_T counts do not follow a Gaussian distribution, indicating that resistance degradation rather than fluctuation occurs at high V_{Read}s.[71] The left axis of Figure 6b–3 shows the sampled S_{IT}/I_T^2 versus V_{Read} (yellow symbols), and the right axis shows the measured σ_{IT}/I_T. S_{IT}/I_T^2 decreases with increasing V_{Read}. Furthermore, σ_{IT}/I_T decreases with increasing V_{Read} from 1.5 to 3.0 V. Note that σ_{IT}/I_T increases with increasing V_{Read} above 3.0 V due to resistance degradation.

Figure 6c–1 shows the $S_{IT}/I_T^2$of of the FTJ in the LRS as a function of V_{Read}. As deduced from Equation (5), as V_{Read} increases, E_{HfO2} increases and S_{IT}/I_T^2 decreases. In addition, unlike the LFN in the device in the HRS, S_{IT}/I_T^2 exhibits a $1/f$ noise trend regardless of V_{Read}. Figure 6c–2 shows the normalized transient I_T counts in the FTJ in the LRS as a function of V_{Read}.[29] Once again, as V_{Read} increases, σ_{IT}/I_T decreases, as shown in Figure 6c–1. Notably, the decrease in σ_{IT}/I_T with respect to V_{Read} is much slower in the LRS than in the HRS. This is due to the origin of the LFN. S_{IT}/I_T^2 is proportional to e^{Vbi} in HRS, while S_{IT}/I_T^2 is proportional to $1/E_{HfO2}$ in LRS. Thus, S_{IT}/I_T^2 decreases much faster with increasing V_{Read} in the HRS device. The left axis of Figure 5c–3 shows the sampled S_{IT}/I_T^2 versus V_{Read} (yellow symbols), and the right axis shows the measured σ_{IT}/I_T. Again, S_{IT}/I_T^2 decreases with increasing V_{Read}.

Consequently, we present methods to reduce the dynamic variation of the FTJ in both the HRS and LRS by applying a high V_{Read} bias. By increasing $V_{Read,}$ the device size can be further reduced. Figure 6d shows the relationship between the FTJ area, V_{Read}, and accuracy of the neuromorphic system. The thick green line shows the condition for reducing the accuracy to less than 3%. The red bar in Figure 6e shows the device scaling limit with respect to V_{Read} to achieve an accuracy degradation of less than 3%. Increasing V_{Read} reduces dynamic variation, allowing further device downscaling. In particular, increasing $V_{Read} = 1.5$ to 3.0 V results in an area downscaling from 5.29 × 5.29 to 0.59 × 0.59 μm^2. The blue line and symbols in Figure 6e represent the power consumption of the neuromorphic system with respect to V_{Read}. The power consumption and scaling limit exhibit opposite tendencies. As V_{Read} increases, the power consumption of the entire neuromorphic system increases if the device scaling limit improves. Notably, there is a trade-off between the scaling limit and power consumption. Increasing V_{Read} allows the device to be scaled down by reducing the dynamic variation, but the power consumption increases, as shown in Figure S10, Supporting Information. Therefore, as summarized in Figure 6f, we have shown how the control of V_{Read} can improve either the power efficiency or area efficiency of a neuromorphic system. However, owing to the inherent trade-off between these two attributes, it is impossible to optimize them simultaneously. This is a fundamental challenge in the realization of area- and power-efficient neuromorphic systems, and novel paradigms must be explored to overcome this obstacle.

2.3. Demonstration of an Area- and Energy-Efficient Hardware Neuromorphic System with Adaptive Read Bias

In the conventional neuromorphic system, the same V_{Read} is applied to all layers. However, previous studies reported that different layers in a neuromorphic system have different sensitivity to variation.[73,74] Therefore, in this paper, we propose a novel method called ARB to adopt different V_{Read} for each layer, enabling additional headrooms for area downscaling and power reduction. Specifically, by applying a low V_{Read} to the noise-robust layer, power consumption can be reduced and area downscaling can be achieved. Conversely, applying a high V_{Read} to the noise-sensitive layer enhances the accuracy of the overall system. To successfully implement ARB, it is necessary to determine the noise sensitivity of the different layers. **Figure 7**a shows the accuracy when a selected layer is biased with a high V_{Read} (low read noise), whereas the other layers are biased with a low V_{Read} (high read noise). The green dotted line indicates the baseline accuracy when the low V_{Read} case is applied to all layers (66.3%). Reducing noise in the convolution (C)1 layer shows the best accuracy recovery, resulting in 15%p accuracy recovery. Furthermore, the C2 to C5 layer shows moderate accuracy recovery (5.7 to 3.4%p), whereas the C7 to the fully connected (F)3 layer shows almost no accuracy recovery (1.9%p to 1.4%p). This trend is due to the characteristics of a convolutional neural network (CNN), in which feature extraction occurs in the convolutional layer. The earlier the convolutional layer, the more important the features that are extracted.[73,75] Therefore, layers C1 to C4 are sensitive to noise, and it is necessary to reduce the dynamic variation by applying a high V_{Read} to these layers to increase the accuracy. Figure 7b shows the accuracy when the selected layer is biased with a low

Figure 6. Understanding the origin of LFN in HfO_2-FTJ and presenting methods to reduce the dynamic variation for device downscaling. a) Schematic energy band diagram of HfO_2-FTJ. The dotted line indicates the application of a high V_{Read}. b–1) S_{IT}/I_T^2 of the FTJ in HRS and c–1) LRS as a parameter of V_{Read}. b–2) Normalized transient I_T counts in HRS and c–2) LRS as a parameter of V_{Read}. b–3) S_{IT}/I_T^2 (left axis) and measured σ_{IT}/I_T (right axis) with respect to device area in HRS and c–3) LRS. d) Relationship between HfO_2-FTJ area, V_{Read}, and accuracy of the neuromorphic system. The green line indicates the condition for accuracy degradation to be less than 3%. e) Device scaling limit (red bar) and power consumption (blue line) with respect to V_{Read} to achieve accuracy degradation of less than 3%. f) Schematic representation of the trade-off between area and power efficiency. The paradigm shift needs to achieve both metrics.

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

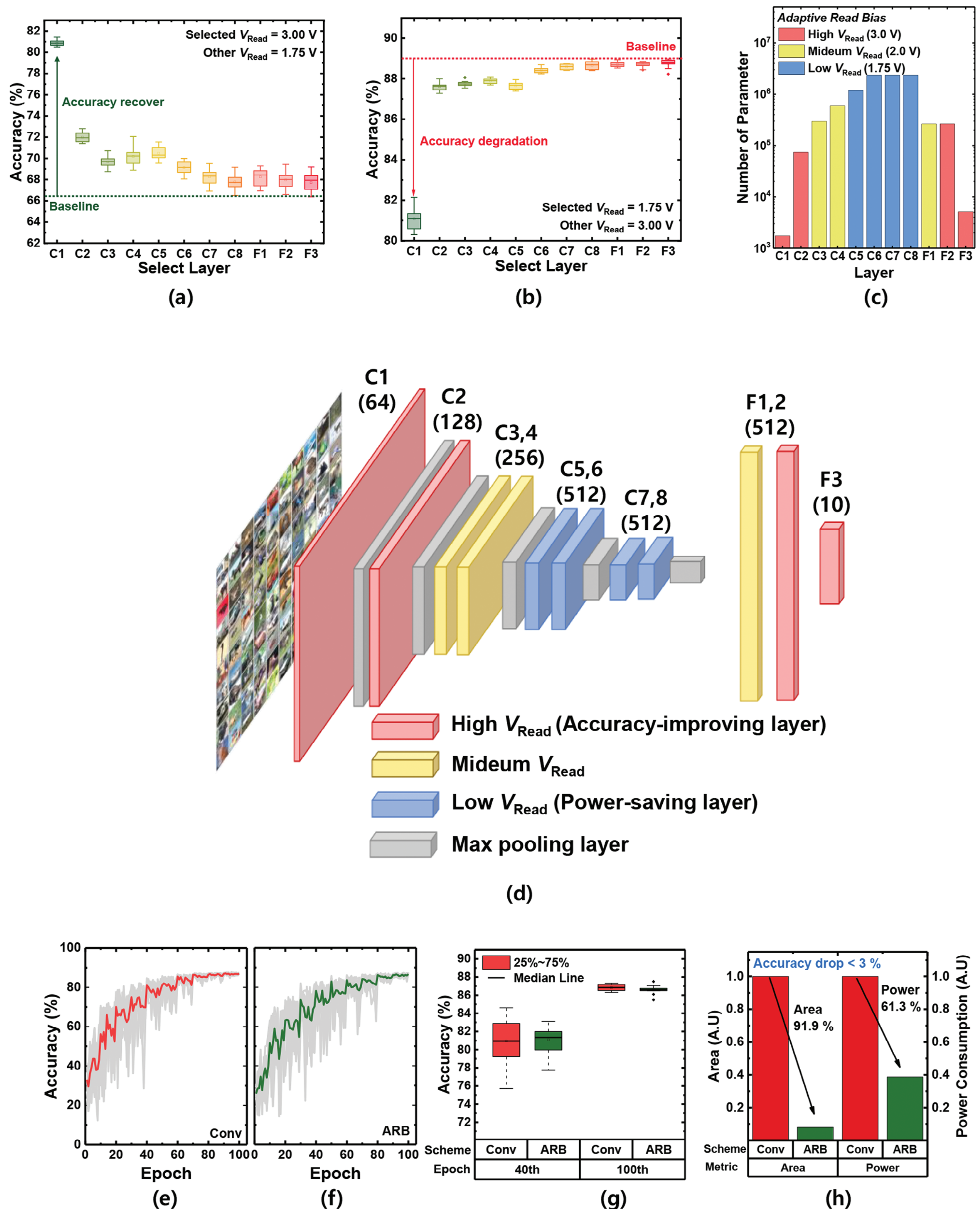


Figure 7. Proposed ARB methods. a) Accuracy when a selected layer is biased with a high V_{Read} and another layer is biased with a low V_{Read}. b) Accuracy when the selected layer is biased with a low V_{Read} and other layers are biased with a high V_{Read}. c) Number of parameters in each layer of the VGG-11 network. d) Schematic diagram of the overall ARB methods. e) Averaged learning accuracy for the conventional biasing scheme and f) ARB scheme. g)

© 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

V_{Read} (for high read noise), and the other layers are biased with a high V_{Read} (for low read noise). The red dotted line indicates the baseline accuracy when a high V_{Read} is applied in all layers (89%). Increasing noise in the C1 layer shows dramatic accuracy degradation, resulting in a decrease in accuracy by 7.91%p. The C2 to C5 layers show moderate accuracy degradation (1.35 to 1.11%p), and the C7 to F3 layers show almost no degradation (0.41 to 0.25%p). Therefore, these layers can be used as power-saving layers with a low V_{Read}, thereby reducing power consumption without sacrificing accuracy. Moreover, to select the layer that serves as an accuracy-improving layer and a power-saving layer, it is important to consider the number of weight parameters in each layer as well as the noise sensitivity. This is because the number of parameters is related to the power consumption of the entire neuromorphic system. If the noise sensitivity is similar between the two layers, it is preferable to choose a layer with fewer parameters to improve accuracy while maintaining low power consumption. The number of parameters in each layer is shown in Figure 7c. The number of parameters of layers C1, C2, F1, F2, and F3 is relatively small. Among them, layers C1 and C2 are the most noise-sensitive, as shown in Figure 7a. Therefore, it is desirable to apply a high V_{Read} to the C1 and C2 layers to improve the accuracy, which we refer to as the accuracy-improving layers. In contrast, layers C5–C8 have the highest number of parameters. In addition, these layers are noise-robust, as shown in Figure 7b. Therefore, applying a low V_{Read} to these layers can achieve significant power savings while maintaining an acceptable level of accuracy; these layers are called power-saving layers. In addition, for system stabilization, layers F2 and F3 are designated as accuracy-improving layers. The rationale underlying this decision is that if the middle layer in the network becomes noisy, the subsequent layers must adapt and counteract this noise. However, if the last layer becomes noisy, the system may become unstable because it may not be able to effectively compensate for the noise. As shown in Figure 7b, when a large noise is introduced to layers F2 and F3, the average decrease in accuracy is small; however, there is a significant variation in the accuracy and presence of outliers, indicating system instability. Therefore, it is preferable to apply a high V_{Read} to layers F2 and F3 to ensure their stabilization and overall network performance, given their relatively small number of parameters. Figure 7d shows the overall ARB scheme of the system. Figures 7e,f show the averaged learning accuracy for the conventional bias scheme (V_{Read} = 1.75 V, area = 3.53 × 3.53 μm^2, which shows the lowest power consumption in Figure 6e) and ARB methods (area = 1 × 1 μm^2). Figure 7g shows the accuracies of the conventional (red) and ARB (green) methods for the 40th (early stage) and 100th (late stage) epochs. The results show that the ARB method has a convergence speed, final accuracy, and stability similar to those of the conventional method. Figure 7h shows the power consumption and area of the ARB and conventional methods. Although the two biasing methods exhibit similar performances (Figure 7g), ARB has 61.3% lower power consumption and 91.9% better density than the conventional method. Therefore, a breakthrough for implementing area- and power-efficient neuromorphic systems is presented in this study. Notably, the proposed ARB method can be applied not only to FTJs, but also to many other devices with $1/f$ noise behavior (ReRAM, PCM, and FET), thus providing a general solution for neuromorphic systems.[40–49,76,77]

3. Conclusion

We proposed an ARB method to address the performance and scaling limitations in ferroelectric-based neuromorphic devices, focusing on mitigating the effects of nonideal factors such as static and dynamic variations. Our investigation identified dynamic variation as a key challenge in achieving high accuracy in neuromorphic systems. To address the accuracy degradation caused by the intrinsic noise of the device, we analyzed the LFN of the FTJs. Our results show that the intrinsic noise of FTJs increases with decreasing device size and V_{Read}, which limits the implementation of both area- and power-efficient neuromorphic systems. To overcome this issue, the ARB method was used, which applies different read biases to each layer of a neuromorphic system. This method takes advantage of the fact that each layer in a neuromorphic system has a different performance sensitivity to the noise of the device, and that noise is related to the read bias. This allows a noise-robust layer to be used as a power-saving layer and a noise-sensitive layer to be used as an accuracy-improving layer. Our proposed adaptive read biasing method demonstrated 61.3% power savings and 91.9% scaling improvements compared with the conventional biasing method. These results contribute significantly to the development of accurate, efficient, and scalable ferroelectric neuromorphic systems. The ARB method can also be applied to neuromorphic systems using other noisy devices.

4. Experimental Section

Fabrication Process of the HfO2-FTJ: Figure S1 (Supporting Information) shows the fabrication process of the undoped HfO_2 FTJ array. 300 nm SiO_2 was grown on Si substrate via wet oxidation and deposited 100 nm n^+ polysilicon by low-pressure chemical vapor deposition (LPCVD) for the bottom electrode of the FTJ. Then, 1.2 nm SiO_2 was grown by chemical oxidation using an APM solution (H_2O: H_2O_2: NH_4OH = 5:1:1, 80 °C, 20 min) for the DE layer of FTJ. 6 nm undoped HfO_2 was deposited by thermal atomic layer deposition (ALD) at 340 °C for the FE of FTJ. Tetrakis (hafnium) was used as the Hf precursor, and ozone was used as the oxidant. Then, 100 nm TiN was deposited by DC sputtering at 200 °C for the top electrode. Photolithography and dry etching were used to define the TiN top gate. To induce ferroelectricity of the HfO_2 layer, post-metal annealing (PMA) (N_2 ambient, 150 sccm, 5 torr) at 800 °C for 30 s was performed. A 300 nm SiO_2 layer was deposited with tetraethylorthosilicate (TEOS) via LPCVD to form an interlayer dielectric (ILD). After the contact etching of the ILD, metallization was performed with a Ti/TiN/Al/TiN (30/30/300/30 nm) stack.

Device Spectroscopy Analysis of HfO2-FTJ: Figure S2 (Supporting Information) shows the material characteristics of the fabricated HfO-FTJ. Transmission electron microscopy (TEM) images of the fabricated FTJ (TiN/HfO_2/SiO_2/Si) showed the thicknesses of the SiO_2 and HfO_2 layers. The composition ratio of Hf:O in the HfO_2 layer was determined by X-ray photoelectron spectroscopy (XPS) analysis. Grazing incidence X-ray

Accuracy of the conventional biasing method (red box plot) and ARB scheme (green box plot) for the 40th and 100th epochs. h) Power consumption and area of the network using the conventional biasing method (red box plot) and ARB (green box plot) to achieve an accuracy degradation of less than 3%.

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

diffraction (GIXRD) analysis of the as-deposited HfO_2 and after PMA was performed. It was confirmed that the PMA at 800 °C crystallized the HfO_2 to induce the o-phase, which generated the ferroelectricity.

Electrical Measurement of the HfO_2-FTJ: To measure the electrical characteristics of the HfO_2-FTJ devices, a semiconductor parameter analyzer (B1500A, Keysight) and a waveform generator/fast measurement unit (WGFMU) module of the B1500 (Keysight) were used.

Figure S7 (Supporting Information) shows the measurement setup for LFN spectroscopy, where a low-noise amplifier (SR570) and signal analyzer (35670A) are utilized to measure the LFN of the FTJ. The output current of the HfO_2-FTJ was connected to SR570, which converted the current fluctuation into a low-noise voltage fluctuation. The voltage fluctuation was converted to the power spectral density at 35670A. The floor of the noise level was lower than 10^{-23} A^2 Hz^{-1}, which showed that the current noise of the FTJ measured in the subsequent experiments was a reliable result.

Software-Based Simulation: Simulations for the classification of CIFAR-10 were performed using Python software. The nonideal characteristics of the FTJ, such as nonlinearity, asymmetry, and variation, were reflected in the simulation. The power consumption during learning and inference was calculated using Python.

Calculation of Power Consumption: The power consumption of individual cells was calculated using the formula.[78]

$$E_{cell} = V_{Read}^2 \times G_{cell} \times T_{read} \quad (8)$$

where E_{cell} is the energy consumption of a single cell, G_{cell} is the conductance of the cell, and T_{read} is the width of the read pulse. By summing the energy consumed by all cells, the overall energy consumption of the array in the neuromorphic system could be calculated. Furthermore, the neuron circuit's operational energy was estimated by the following equation.[78]

$$E_{TIA} = n_{neuron} \times T_{read} \times 0.34\text{mW} \quad (9)$$

where E_{TIA} is the energy consumption of neurons, and n_{neuron} is the number of neuron circuits. By adding E_{cell} and E_{TIA}, the synaptic array's total power consumption was calculated.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

R.-H. K. and W.S. contributed equally to this work. This research was supported in part by the National Research Foundation (NRF) funded by the Korean Ministry of Science and ICT under Grant Nos. 2022M3I7A2085479, in part by Samsung Research Funding & Incubation Center of Samsung Electronics under Project Number SRFC-TA2103-02, and in part by the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2023.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

ferroelectric, hafnium oxide, low-frequency noise, neuromorphic, power-efficient

Received: June 8, 2023
Revised: October 11, 2023
Published online:

[1] M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **2018**, *1*, 22.
[2] C. Wu, T. W. Kim, H. Y. Choi, D. B. Strukov, J. J. Yang, *Nat. Commun.* **2017**, *8*, 752.
[3] J. J. Yang, D. B. Strukov, D. R. Stewart, *Nat. Nanotechnol.* **2013**, *8*, 13.
[4] L. Danial, E. Pikhay, E. Herbelin, N. Wainstein, V. Gupta, N. Wald, Y. Roizin, R. Daniel, S. Kvatinsky, *Nat. Electron.* **2019**, *2*, 596.
[5] J. K. Han, M. Kang, J. Jeong, I. Cho, J. M. Yu, K. J. Yoon, Y. K. Choi, *Adv. Sci.* **2022**, *9*, 18.
[6] T.-Y. Wang, J.-L. Meng, Z.-Y. He, L. Chen, H. Zhu, Q.-Q. Sun, S.-J. Ding, P. Zhou, D. W. Zhang, *Nanoscale Horiz.* **2019**, *4*, 1293.
[7] J. Jang, S. Gi, I. Yeo, S. Choi, S. Jang, S. Ham, G. Wang, *Adv. Sci.* **2022**, *9*, 22.
[8] W. Wan, R. Kubendran, C. Schaefer, S. B. Eryilmaz, W. Zhang, D. Wu, S. Deiss, P. Raina, H. Qian, B. Gao, S. Joshi, H. Wu, H.-S. P. Wong, G. Cauwenberghs, *Nature* **2022**, *608*, 504.
[9] J. F. Kang, B. Gao, P. Huang, H. T. Li, Y. D. Zhao, Z. Chen, C. Liu, L. F. Liu, X. Y. Liu, in *2015 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Washington, DC **2015**.
[10] S. Yu, in *2014 IEEE Int. Symp. on Circuits and Systems (ISCAS)*, IEEE, Melbourne, VIC **2014**.
[11] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G. W. Burr, N. Sosa, A. Ray, J.-P. Han, C. Miller, K. Hosokawa, C. Lam, in *2015 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Washington, DC **2015**.
[12] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, B. DeSalvo, in *2011 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Washington, DC **2011**.
[13] M. Xu, X. Mai, J. Lin, W. Zhang, Y. Li, Y. He, X. Miao, *Adv Funct Mater* **2020**, *30*, 50.
[14] Y. Abbas, S. M. Ansari, I. Taha, H. Abunahla, M. U. Khan, M.'d Rezeq, H. M. Aldosari, B. Mohammad, *Adv Funct Mater* **2023**, 2214615.
[15] A. W. Stephan, S. J. Koester, *IEEE Trans. Electron Devices* **2020**, *67*, 487.
[16] A. F. Vincent, J. Larroque, N. Locatelli, N. Ben Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J.-O. Klein, S. Galdin-Retailleau, D. Querlioz, *IEEE Trans Biomed Circuits Syst* **2015**, *9*, 166.
[17] D. Zhang, L. Zeng, Y. Qu, Y. Zhang, M. Wang, W. Zhao, Y. Wang, in *2015 IEEE Int. Symp. on Circuits and Systems (ISCAS)*, IEEE, Lisbon, Portugal **2015**.
[18] M. Lederer, T. Kampfe, T. Ali, F. Muller, R. Olivo, R. Hoffmann, N. Laleni, K. Seidel, *IEEE Trans. Electron Devices* **2021**, *68*, 2295.
[19] T. Mikolajick, M. H. Park, L. Begon-Lours, S. Slesazeck, *Adv Mater* **2022**, *35*, 2206042.
[20] Z. Li, T. Wang, J. Yu, J. Meng, Y. Liu, H. Zhu, Q. Sun, D. W. Zhang, L. Chen, *Adv. Electron. Mater.* **2022**, *8*, 2200951.
[21] W. Shin, R. H. Koo, K. K. Min, D. Kwon, J. J. Kim, D. Kwon, J. H. Lee, *IEEE Electron Device Lett.* **2022**, *44*, 2.
[22] B. Max, M. Hoffmann, H. Mulaosmanovic, S. Slesazeck, T. Mikolajick, *ACS Appl. Electron. Mater.* **2020**, *2*, 12.
[23] 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, J. Grollier, *Nat. Mater.* **2012**, *11*, 860.
[24] R. H. Koo, W. Shin, K. K. Min, D. Kwon, D. H. Kim, J. J. Kim, J. H. Lee, *IEEE Electron Device Lett.* **2022**, *44*, 164.

© 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[25] W. Shin, J. H. Bae, D. Kwon, R. H. Koo, B. G. Park, D. Kwon, J. H. Lee, *IEEE Electron Device Lett.* **2022**, *43*, 6.
[26] M. Payvand, M. V. Nair, L. K. Müller, G. Indiveri, *Faraday Discuss.* **2019**, *213*, 487.
[27] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. Le Gallo, K. Moon, J. Woo, H. Hwang, Y. Leblebici, *Adv. Phys.: X* **2017**, *2*, 89.
[28] Y. H. Lin, C. H. Wang, M. H. Lee, D. Y. Lee, Y. Y. Lin, F. M. Lee, C. Y. Lu, *IEEE Trans. Electron Devices* **2019**, *66*, 3.
[29] M. Zhao, B. Gao, J. Tang, H. Qian, H. Wu, *Appl. Phys. Rev.* **2020**, *7*, 011301.
[30] M. Rao, H. Tang, J. Wu, W. Song, M. Zhang, W. Yin, Y. Zhuo, F. Kiani, B. Chen, X. Jiang, H. Liu, H.-Y. Chen, R. Midya, F. Ye, H. Jiang, Z. Wang, M. Wu, M. Hu, H. Wang, Q. Xia, N. Ge, J. Li, J. J. Yang, *Nature* **2023**, *615*, 823.
[31] M. Ismail, C. Mahata, O. Kwon, S. Kim, *ACS Appl. Electron. Mater.* **2022**, *4*, 3.
[32] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, H. Hwang, *IEEE Electron Device Lett.* **2016**, *37*, 8.
[33] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, H. S. P. Wong, *Adv. Mater.* **2013**, *25*, 12.
[34] N. K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, J. Joshua Yang, *Adv. Mater. Technol.* **2019**, *4*, 4.
[35] G. H. Lee, T. H. Kim, M. S. Song, J. Park, S. Kim, K. Hong, H. Kim, *Chaos Solitons Fractals* **2022**, *157*, 111999.
[36] S. Majumdar, H. Tan, Q. H. Qin, S. Van Dijken, *Adv. Electron. Mater.* **2019**, *5*, 3.
[37] P. Y. Chen, S. Yu, *IEEE Trans. Electron Devices* **2015**, *62*, 12.
[38] W. Shin, J. H. Bae, D. Kwon, R. H. Koo, B. G. Park, D. Kwon, J. H. Lee, *IEEE Electron Device Lett.* **2022**, *43*, 6.
[39] S. Kim, K. Lee, M. H. Oh, J. H. Lee, B. G. Park, D. Kwon, *IEEE Electron Device Lett.* **2021**, *42*, 6.
[40] P. Dutta, P. M. Horn, *Rev. Mod. Phys.* **1981**, *53*, 497.
[41] A. A. Balandin, *Nat Nanotechnol* **2013**, *8*, 549.
[42] M. Niemann, H. Kantz, E. Barkai, *Phys. Rev. Lett.* **2013**, *110*, 14.
[43] K. L. Schick, A. A. Verveen, *Nature* **1974**, *251*, 599.
[44] R. H. Koo, W. Shin, K. K. Min, D. Kwon, J. J. Kim, J. H. Lee, *IEEE Electron Device Lett.* **2023**, *44*, 6.
[45] M. S. Song, T. H. Kim, H. Hwang, S. Ahn, H. Nili, H. Kim, *Adv. Int. Syst.* **2023**, *5*, 2200358.
[46] S. Pazos, T. Becker, M. A. Villena, W. Zheng, Y. Shen, Y. Yuan, M. Lanza, *Adv Funct Mater* **2023**, 2213816.
[47] W. S. Choi, M. S. Song, H. Kim, D. H. Kim, *Micromachines* **2022**, *13*, 1870.
[48] S. Choi, Y. Yang, W. Lu, *Nanoscale* **2014**, *6*, 400.
[49] R. Soni, P. Meuffels, A. Petraru, M. Weides, C. Kügeler, R. Waser, H. Kohlstedt, *J. Appl. Phys.* **2010**, *107*, 024517.
[50] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. Paul Strachan, M. Barnell, Q. Wu, R. S. Williams, J. Joshua Yang, Q. Xia, *Nat Commun* **2018**, *9*, 2385.
[51] X. Wu, V. Saxena, K. Zhu, S. Balagopal, *IEEE Trans Circuits Syst II Express Briefs* **2015**, *62*, 11.
[52] F. Alibart, E. Zamanidoost, D. B. Strukov, *Nat Commun* **2013**, *4*, 2072.
[53] N. K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, J. Joshua Yang, *Adv. Mater. Technol.* **2019**, *4*, 1800589.
[54] D. Kim, Y. R. Jeon, B. Ku, C. Chung, T. H. Kim, S. Yang, C. Choi, *ACS Appl Mater Interfaces* **2021**, *13*, 44.
[55] S. Oh, T. Kim, M. Kwak, J. Song, J. Woo, S. Jeon, I.n K. Yoo, H. Hwang, *IEEE Electron Device Lett.* **2017**, *38*, 732.
[56] L. Wang, X. Wang, Y. Zhang, R. Li, T. Ma, K. Leng, K. P. Loh, *Adv Funct Mater* **2020**, *30*, 45.
[57] T. Lu, X. Zhao, H. Liu, Z. Yan, R. Zhao, M. Shao, T. L. Ren, *IEEE Trans. Electron Devices* **2023**, *70*, 5.
[58] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, H. Adam, *ArXiv* **2017**, arXiv:170404861.
[59] W. Wang, Y. Li, T. Zou, X. Wang, J. You, Y. Luo, *Inf Syst* **2020**, *2020*, 7602384.
[60] Y. Li, G. Yuan, Y. Wen, J. Hu, G. Evangelidis, S. Tulyakov, J. Ren, *Adv Neural Inf Process Syst* **2022**, *35*, 12934.
[61] F. Alibart, E. Zamanidoost, D. B. Strukov, *Nat. Commun.* **2013**, *4*, 2072.
[62] R. Müller, Generation-recombination noise. ′In Noise in Physical Systems: Proceedings of the Fifth International Conference on Noise, Bad Nauheim, Fed. Rep. of Germany, March 13–16, 1978 Springer, Berlin, Heidelberg, **1978**.
[63] M. V. Haartman, M. Östling, *Low-Frequency Noise in Advanced MOS Devices*, Springer, Dordrecht, The Netherlands **2007**.
[64] A. N. Nazarov, I. Ferain, N. D. Akhavan, P. Razavi, R. Yu, J. P. Colinge, *Appl. Phys. Lett.* **2011**, *98*, 9.
[65] Y. Yuzhelevski, M. Yuzhelevski, G. Jung, *Rev. Sci. Instrum.* **2000**, *71*, 1681.
[66] A. Reklaitis, L. Reggiani, *Phys. B* **1999**, *272*, 279.
[67] J. Lee, G. Bosman, K. R. Green, D. Ladwig, *IEEE Trans. Electron Devices* **2003**, *50*, 12.
[68] D. Kwon, W. Shin, J. H. Bae, S. Lim, B. G. Park, J. H. Lee, *IEEE Electron Device Lett.* **2021**, *42*, 3.
[69] C. T. Angelis, C. A. Dimitriadis, I. Samaras, J. Brini, G. Kamarinos, V. K. Gueorguiev, T. E. Ivanov, *J. Appl. Phys.* **1997**, *82*, 4095.
[70] D. Kang, S. Jang, S. Choi, S. Kim, *Semicond. Sci. Technol.* **2021**, *36*, 12.
[71] T. Gong, D. Dong, Q. Luo, X. Xu, J. Yang, J. Yu, M. Liu, *IEEE Electron Device Lett.* **2021**, *42*, 3.
[72] H. Kim, J. H. Bae, S. Lim, S. T. Lee, Y. T. Seo, D. Kwon, B.-G. Park, J. H. Lee, *Neurocomputing* **2020**, *378*, 189.
[73] R.-H. Koo, W. Shin, S. Ryu, K. Lee, S.-H. Park, J. Im, J.-H. Ko, J.-H. Kim, D. Kwon, J.-J. Kim, D. Kwon, J.-H. Lee, *IEEE Electron Device Lett.* **2023**, *44*, 1623.
[74] R. Chauhan, K. K. Ghanshala, R. C. Joshi, *2018 First Int. Conf. on Secure Cyber Computing and Communication (ICSCCC)*, IEEE, Jalandhar, India **2018**.
[75] S. Albawi, T. A. Mohammed, S. Al-Zawi, in *2017 Int. Conf. on Engineering and Technology (ICET)*, IEEE, Antalya, Turkey **2017**.
[76] J.-M. Peransin, P. Vignaud, D. Rigaud, L. K. J. Vandamme, *IEEE Trans. Electron Devices* **1990**, *37*, 2250.
[77] S. Vodapally, C. G. Theodorou, Y. Bae, G. Ghibaudo, S. Cristoloveanu, K.-S. Im, J.-H. Lee, *IEEE Trans. Electron Devices* **2017**, *64*, 3634.
[78] S. Yu, H. Jiang, S. Huang, X. Peng, A. Lu, *IEEE Circuits Syst. Mag.* **2021**, *21*, 31.

 © 2023 The Authors. Advanced Science published by Wiley-VCH GmbH

21983844, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/advs.202303735 by Hanyang University Library, Wiley Online Library on [12/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License