**Materials Science**

**CMOS-compatible compute-in-memory accelerators based on integrated ferroelectric synaptic arrays for convolution neural networks**

Min-Kyu Kim†, Ik-Jyae Kim†, Jang-Sik Lee*

Convolutional neural networks (CNNs) have gained much attention because they can provide superior complex image recognition through convolution operations. Convolution processes require repeated multiplication and accumulation operations, which are difficult tasks for conventional computing systems. Compute-in-memory (CIM) that uses parallel data processing is an ideal device structure for convolution operations. CIM based on two-terminal synaptic devices with a crossbar structure has been developed, but unwanted leakage current paths and the high-power consumption remain as the challenges. Here, we demonstrate integrated ferroelectric thin-film transistor (FeTFT) synaptic arrays that can provide efficient parallel programming and data processing for CNNs by the selective and accurate control of polarization in the ferroelectric layer. In addition, three-terminal FeTFTs can act as both nonvolatile memory and access device, which tackle issues from two-terminal devices. An integrated FeTFT synaptic array with parallel programming capabilities can perform convolution operations to extract image features with a high-recognition accuracy.

**INTRODUCTION**

Convolutional neural networks (CNNs) have been used in various applications, such as speech recognition and image classification (1, 2). CNNs can achieve superior performance in complex image recognition processes through convolution operations that use kernels as filters to extract features from images. However, convolution operations are time-consuming because they involve repeated multiplication and accumulation operations between the input data and the kernel weight. When CNNs are implemented in hardware based on the von Neumann architecture, drawbacks such as increased power consumption and limited data processing speeds arise because large amounts of data need to be transferred between the processing and memory units (3, 4). To overcome these limitations, compute-in-memory (CIM) has been suggested as alternative hardware for CNNs because it enables parallel data processing (5–9). In CIM, parallel data processing can be performed in a crossbar array structure using vector-matrix multiplication (VMM) based on Ohm’s law (for multiplication) and Kirchhoff’s law (for accumulation) (10). For accurate VMM operation, synaptic devices are required, which can precisely adjust the conductance states via analog conductance modulation (5, 11, 12). In previous studies, several CIM devices with a crossbar structure have been demonstrated using two-terminal devices, such as phase-change and resistive-switching memories, as synaptic devices (13–18). However, when CIM is implemented using crossbar arrays based on two-terminal devices, there are issues such as cross-talk, sneak path current, and nonlinear current-voltage characteristics (19–21). The analog conductance modulation characteristics of two-terminal devices are usually achieved by controlling the current of devices during the programming process (22–24). However, in arrays consisting of two-terminal devices, it is difficult to control the current of selected devices during the programming process precisely because of current flowing through the unselected devices (25, 26). In addition, the current flowing through the unselected devices can affect the accuracy of read operation. Therefore, this issue can cause inaccurate conductance modulation in crossbar arrays based on two-terminal devices (26).

Three-terminal devices have the potential to perform accurate conductance modulation in array structure because terminals for program and read operations are separated (19, 27). In an array composed of three-terminal devices, the program and read operations are not affected by the states of the unselected devices. Therefore, three-terminal devices, such as electrochemical transistors and charge-trapping transistors, have been investigated as synaptic devices (10, 17, 19, 20, 28–31). Electrochemical transistors exhibit linear weight updates and low-voltage operation (19, 31). However, the use of active ions (e.g., Li⁺) may be incompatible with complementary metal-oxide semiconductor (CMOS) device fabrication and integration (17, 32). Although charge-trapping transistors are based on mature technology for array integration, they require a high operation voltage and exhibit nonlinear weight update characteristics (16, 20). Among the available three-terminal synaptic devices, ferroelectric transistors based on zirconium-doped hafnium oxide (HfZrOₓ) are advantageous because they have CMOS compatibility, fast operation speed, low operation voltages, and high scalability (33–36). In particular, the conductance of ferroelectric transistors can be precisely modulated by controlling the states of polarization in the ferroelectric layer (33, 37, 38). These properties can be used to implement kernel weights for convolution operations in ferroelectric transistor arrays. Thus, ferroelectric transistor arrays could be used as synaptic arrays for CIM.

In this study, we demonstrate integrated synaptic transistor arrays based on ferroelectric thin-film transistors (FeTFTs) composed of indium zinc oxide (IZO) and HfZrOₓ. In these arrays, FeTFTs are used as synaptic devices. The conductance of the FeTFTs, which represents kernel weight, is linearly controlled by adjusting the states of polarization in the ferroelectric layer. In addition, column- and row-wise parallel programming operations are demonstrated in...
ferroelectric synaptic transistor arrays by the selective control of polarization switching with program-inhibit operations. Using these characteristics, ferroelectric synaptic transistor arrays are trained to implement kernel weights and perform convolution operations, which can be used to extract the features of input images. The four different kernels, which are realized in ferroelectric synaptic transistor arrays, can extract the features of an image with $64 \times 64$ pixels. In addition, neural network simulations based on the weight update characteristics of the ferroelectric synaptic transistor array exhibit an image recognition accuracy of 90.3%. The results of this study provide important information that can contribute to the development of CIM based on ferroelectric synaptic transistor arrays.

RESULTS

Parallel programming of a ferroelectric synaptic transistor array

The ferroelectric characteristics of the HfZrO$_x$ were first confirmed using the ferroelectric capacitor with the molybdenum (Mo)/HfZrO$_x$/tungsten (W) structure (fig. S1). When the range of the voltage sweep increased from $\pm 2$ to $\pm 6$ V, the remnant polarization of the HfZrO$_x$ increased. In a voltage sweep range of $-6$ to $6$ V, a ferroelectric capacitor exhibited positive and negative remnant polarization of 15.1 and $-14.1 \mu C/cm^2$, respectively. These results confirmed that HfZrO$_x$ had ferroelectric characteristics and that its polarization characteristics could be adjusted by changing the amplitude of the applied voltages (39, 40). We fabricated synaptic transistor arrays by integrating FeTFTs with IZO and HfZrO$_x$ (Fig. 1A). The fabrication process of the ferroelectric synaptic transistor array was CMOS compatible and could be achieved at a temperature under 400°C (fig. S2). First, gate lines (GLs) were formed by depositing W on the SiO$_2$/Si substrate. Then, an HfZrO$_x$ layer was deposited as the ferroelectric layer using atomic layer deposition (ALD). Next, a Mo layer was deposited to form the source lines (SLs), source electrodes, and drain electrodes. ALD was used to form an IZO layer, which acted as the channel layer. Then, the devices were annealed at 400°C to induce ferroelectricity in the HfZrO$_x$ layer. Last, an interlayer dielectric was formed, and a Mo layer was deposited to act as the drain lines (DLs). These ferroelectric synaptic transistor arrays were composed of four GLs, four SLs, and nine DLs. In the ferroelectric synaptic transistor array, one column included nine FeTFTs, which had a bottom gate and bottom contact structure. Each column was connected to nine DLs, one GL, and one SL.

The switching characteristics of the FeTFTs were evaluated by measuring the device states after applying GL voltage ($V_{GL}$) pulses with different amplitudes (fig. S3). The device states were confirmed by measuring the DL current ($I_{DL}$) at a DL voltage ($V_{DL}$) of 0.1 V, while sweeping $V_{GL}$ from 0 to $-3.5$ V. Before the measurements, the FeTFTs were erased by applying an erase pulse ($-5$ V, 10 ms) to the GL, while SL and DL were set to 0 V. Subsequently, program pulses...
with a width of 10 ms were applied to the GL. The amplitudes of the program pulses increased from 2.5 to 5 V in increments of 0.25 V. As the amplitudes of the pulses increased, the $I_{DL}$-$V_{GL}$ curves gradually shifted in the negative direction, and the channel conductance increased. The switching characteristics of the FeTFTs were also investigated by applying program pulses with different amplitudes and widths (fig. S4). The pulse width required for conductance modulation decreased when the amplitude of the program pulse increased. These multilevel capabilities may originate from the partial polarization characteristics of the ferroelectric HfZrO$_2$ layer, and these characteristics can be used to realize weights for kernel filters (39).

In ferroelectric synaptic transistor arrays, FeTFTs in the same column share a single GL. Therefore, an unwanted program of unselected devices in the same column (called a program disturbance) can occur during programming of a selected device (41, 42). For selective and parallel programming in ferroelectric synaptic transistor arrays, these issues need to be solved. To demonstrate selective and parallel programming operation, we introduced a program-inhibit operation method. In the ferroelectric synaptic transistor array, 4 x 4 FeTFTs were used to demonstrate parallel programming in multiple columns and rows (Fig. 1B). Four different cases of programming patterns were used. In these cases, four devices were selected in the array. Before the parallel programming, FeTFTs in the array were erased by applying erase pulses (−5 V, 10 ms) to all GLs, while all SLs and DLs were set to 0 V. During the parallel programming of the selected devices, program pulses (4 V, 10 ms) were applied to the selected GLs, and the selected DLs were set to 0 V. In this operation, multiple GLs and DLs were selected to program the devices. Therefore, multiple devices in the arrays could be programmed in parallel. Program-inhibit pulses (2 V, 30 ms) were applied to the unselected DLs and all the SLs to prevent the program disturbance (Fig. 1C). The program-inhibit voltage reduces the difference between the voltages of the GL and the channel layer of the unselected FeTFT, and this can prevent polarization switching in the unselected devices (43). After the parallel programming operation, the states of the devices in the same column could be confirmed simultaneously by measuring the current of the DLs, while a read voltage of 0.1 V was applied to the selected SL. All the DLs and GLs were set to 0 and −2 V, respectively (Fig. 1D) (20). After the parallel programming, the current levels of the selected devices increased, and programming of the unselected devices was prevented (Fig. 1E).

The program disturbance could be successfully prevented using program-inhibit pulses, which was confirmed by repeated program-inhibit operations (fig. S5). Using this parallel programming method, the FeTFTs in the ferroelectric synaptic transistor array could be simultaneously programmed, and program disturbance could be inhibited.

In addition, selective and parallel weight updates can be performed using the above operation methods. To demonstrate these updates, we selected two FeTFTs in the same column (Fig. 2A). During the potentiation and depression operations, multiple update pulses with incremental amplitudes and a width of 10 ms were applied to the selected GL, and the selected DLs were set to 0 V. The amplitudes of the potentiation and depression pulses increased from 2.7 to 3.96 V in a 20-mV step and from −2.7 to −3.96 V in a −20-mV step, respectively. Program-inhibit pulses with incremental amplitudes and a width of 30 ms were applied to the SL and unselected DLs. The amplitude of the program-inhibit pulses for potentiation and depression operations increased from 1.35 to 1.98 V in a 10-mV step and −1.35 to −1.98 V in a −10-mV step, respectively. The conductance of the devices was confirmed by measuring the current of the DLs, while a read voltage of 0.1 V was applied to the selected SL. All the DLs and GLs were set to 0 and −2 V, respectively. Using these methods, the selected devices were selectively and linearly updated. Disturbance from the weight update of neighboring synaptic devices was not observed (Fig. 2B). In addition, the cycle-to-cycle and device-to-device variation characteristics of FeTFTs were investigated. The endurance characteristics of a FeTFT were measured for 30 potentiation and depression cycles (Fig. 2C). During repeated pulse operations, degradation was not observed, and the cycle-to-cycle variation was about 1.3% (Fig. 2D). To evaluate the device-to-device variation, we measured the potentiation and depression characteristics of 36 FeTFTs in the array. The ferroelectric synaptic transistor array showed a device-to-device variation of about 3.7% (Fig. 2E). The small variations in the electrical properties of the ferroelectric synaptic transistor array are required for high accuracy of CIM (12).

With column-wise parallel update method, synaptic devices with different states are hard to be updated because it is difficult to apply the update pulse with different amplitudes to the devices in the same column. This issue can be solved by using row-wise parallel weight update methods. Devices in the same row do not share the same GL, thus update pulses with different amplitudes can be simultaneously applied to devices in the same row. Therefore, we demonstrated a row-wise parallel weight update for devices with different states. To demonstrate row-wise parallel weight update, we selected two FeTFTs in the same row (fig. S6A). During the potentiation and depression operations for each device, different amplitudes of potentiation and depression pulses based on the current device state were applied to the selected GLs, and the selected DLs were set to 0 V. Program-inhibit pulses were applied to the SL and unselected DLs. The conductance of the devices was confirmed by measuring the current of the SLs, while a read DL voltage of 0.1 V was applied to the selected DL. All the SLs and GLs were set to 0 and −2 V, respectively. Using these methods, the devices with different states were simultaneously and linearly updated. Disturbance from the weight update of synaptic devices in the same row was not observed (fig. S6B). Thus, the results showed that the parallel programming to different states could be achieved using column- and row-wise parallel update scheme.

In the ferroelectric synaptic transistor array, an incremental pulse scheme was used for the weight update. The use of an incremental pulse scheme can complicate the circuit design for the generation of pulses with different amplitudes and increase power consumption because of the additional steps for accessing the weight values (44). When an incremental pulse scheme is used, the weight values need to be confirmed before the weight update to determine the amplitude of the update pulse. Thus, the time required for training can be increased when using incremental pulse scheme in FeTFT array. Further study is being done to reduce the training time in the FeTFT array structure. Although there are some issues associated with an incremental pulse scheme, the use of this scheme can produce highly linear weight update characteristics, which are essential requirements for CIM hardware with a high recognition accuracy because the weights of the synaptic devices should be precisely adjusted to the desired values (11). In a ferroelectric synaptic transistor array, symmetric and linear weight update characteristics can be achieved by precisely controlling the state of polarization in the ferroelectric...
layer, which can be used to adjust the weights of the FeTFTs to the desired values. In addition, these linear weight update characteristics can be achieved in parallel weight update using the three-terminal structure of FeTFTs. These parallel weight updates can improve the efficiency of the weight update process as the weights of multiple devices can be updated simultaneously. These results indicate that the ferroelectric synaptic transistor array can be used to realize the efficient CIM hardware with a high accuracy.

**Convolution operation in a ferroelectric synaptic transistor array**

In convolution operations, the intensity value of each pixel, which represents the brightness of the pixel in the input image, is multiplied by the kernel weight (Fig. 3A). The output is the sum of the products of the intensity values of the input pixels and the kernel weights. Therefore, multiplication and accumulation operations are required for convolution operations. To perform convolution operations in the synaptic transistor array, the intensity values of pixels and kernel weights are converted into voltages and the conductance of the synaptic devices, respectively (45). Multiplication of the intensity values of the input pixels by the kernel weights can be achieved using Ohm’s law, and the accumulation can be achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics are required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.

To investigate the potential of ferroelectric synaptic transistor arrays for convolution operations, we performed convolution operations based on a vertical edge kernel for an input image with 6 × 6 pixels. To perform convolution operation, we converted the intensity values of pixels and kernel weights into voltages and the conductance of the synaptic devices, respectively. The multiplication of the intensity values of the input pixels by the kernel weights was achieved using Ohm’s law, and the accumulation was achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics were required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.

To investigate the potential of ferroelectric synaptic transistor arrays for convolution operations, we performed convolution operations based on a vertical edge kernel for an input image with 6 × 6 pixels. To perform convolution operation, we converted the intensity values of pixels and kernel weights into voltages and the conductance of the synaptic devices, respectively. The multiplication of the intensity values of the input pixels by the kernel weights was achieved using Ohm’s law, and the accumulation was achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics were required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.

To investigate the potential of ferroelectric synaptic transistor arrays for convolution operations, we performed convolution operations based on a vertical edge kernel for an input image with 6 × 6 pixels. To perform convolution operation, we converted the intensity values of pixels and kernel weights into voltages and the conductance of the synaptic devices, respectively. The multiplication of the intensity values of the input pixels by the kernel weights was achieved using Ohm’s law, and the accumulation was achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics were required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.

To investigate the potential of ferroelectric synaptic transistor arrays for convolution operations, we performed convolution operations based on a vertical edge kernel for an input image with 6 × 6 pixels. To perform convolution operation, we converted the intensity values of pixels and kernel weights into voltages and the conductance of the synaptic devices, respectively. The multiplication of the intensity values of the input pixels by the kernel weights was achieved using Ohm’s law, and the accumulation was achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics were required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.

To investigate the potential of ferroelectric synaptic transistor arrays for convolution operations, we performed convolution operations based on a vertical edge kernel for an input image with 6 × 6 pixels. To perform convolution operation, we converted the intensity values of pixels and kernel weights into voltages and the conductance of the synaptic devices, respectively. The multiplication of the intensity values of the input pixels by the kernel weights was achieved using Ohm’s law, and the accumulation was achieved using Kirchhoff’s law (20, 46). Thus, linear current-voltage characteristics were required for accurate convolution operations based on the VMM (21, 24, 47, 48). The FeTFTs exhibited linear current-voltage characteristics at multiple states between the erased and programmed states (Fig. S7). The linearity of the current-voltage was defined as $I-V$ linearity = $I_{DL}$ at $V_{DL}$ (0.1 V)/2 $I_{DL}$ at $V_{DL}$ (0.05 V) (47). The $I-V$ linearity of the FeTFTs was about 0.95. These characteristics are important for realizing accurate convolution operation.
values of $3 \times 3$ pixels in the input image into voltages from 0 to 0.1 V, which were assigned to each DL in the ferroelectric synaptic transistor array (Fig. 3B). To realize a kernel with $3 \times 3$ weights in a ferroelectric synaptic transistor array, we used two columns of the array because a differential pair was required to represent both the positive and negative weights in the kernel (Fig. 3C) (14, 20, 49). The channel conductance of the FeTFTs in two columns of the array was adjusted to the value required to realize the weights of the vertical edge kernel, which could extract features corresponding to the vertical edges of the input image (fig. S8). The first and second columns represented the positive- and negative-weight values in the kernel, respectively. Therefore, in a $9 \times 2$ array, a differential of weights could represent a kernel with $3 \times 3$ weights. For the weights of the kernel, precise conductance modulation characteristics are required to adjust the synaptic weight to the desired value (50). In this respect, the analog weight update characteristics of FeTFTs are advantageous in synaptic devices used for convolution operations (51). The FeTFT arrays were programmed to the desired values using the column-wise parallel programming method. When voltages that were converted from intensity values of $3 \times 3$ pixels were applied to DLs of the ferroelectric synaptic transistor array, the output currents of SLs in the columns with positive and negative weights represented the sum of the product of the channel conductance and $V_{DL}$. During convolution operations, read GL voltages of $-2$ V were applied to all
For the differential readout between the output current of positive- and negative-weight columns, a subtractor circuit consisting of two operational amplifiers (op-amps) was used (Fig. 3D). The inverting input terminals of the first and second op-amps were connected to the SLs of the positive- and negative-weight columns, respectively. In this circuit, the differential readout between the output current of positive- and negative-weight columns was converted into the output voltage ($V_{output}$) of the subtractor circuit (S1). A 1-kilohm reference resistance ($R_{ref}$) was used for the first op-amp. A normalization resistance ($R_{norm}$) of 1 megohm was used for the second op-amp. The output terminal of the first op-amp and the inverting input terminal of the second op-amp were connected using the other 1-kilohm $R_{ref}$. The $V_{output}$ of the subtractor circuit was measured at the output terminal of the second op-amp. This represents the differential readout between the output currents of the positive- and negative-weight columns. The ratio between the differential readout and $V_{output}$ can be adjusted according to $R_{norm}$ in the second op-amp. For an input image with 6 × 6 pixels, convolution operation systems composed of ferroelectric synaptic transistor arrays and op-amps exhibited different $V_{output}$ values depending on the features of the image (Fig. 3E). The $V_{output}$ values of convolution operation using the ferroelectric synaptic transistor array exhibited a tendency similar to that of the expected output values shown in Fig. 3A, which confirmed that the ferroelectric synaptic transistor arrays could perform parallel data processing for convolution operations. These accurate convolution operations performed by the ferroelectric synaptic transistor array may be the result of integrated array structure, linear current-voltage behaviors, and analog conductance modulation characteristics of FeTFTs.

Ferroelectric synaptic transistor arrays are advantageous in energy consumption for the convolution operation because weight values of FeTFTs can be realized at a low conductance (~1 μS). When a pulse width of 10 ns is assumed for the convolution operation, the energy consumption of the ferroelectric synaptic transistor array for a single convolution operation based on a kernel with 3 × 3 weights is ~2 fJ [$E = V^2 \times G \times t \times n = (0.1 V)^2 \times 1 \mu S \times 10 \times 18$], where $E$, $t$, and $n$ stands for consumed energy, pulse width, and number of ferroelectric synapses, respectively (6). These characteristics of ferroelectric synaptic transistor arrays have the potential for energy-efficient convolution operation. In addition, the low power consumption of convolution operation using ferroelectric synaptic transistor arrays can improve the energy efficiency of CNN. To evaluate the feasibility of convolution operations using ferroelectric synaptic transistor arrays for image processing, image feature extraction was performed using the Lena image (Fig. 3F) (20, 50, 52). For the image processing, four different kernels, including vertical edge, horizontal edge, mean, and sharpen, were used (fig. S9) (20). To implement these four different kernels, the conductance of the FeTFTs in the 9 × 2 ferroelectric synaptic transistor array was adjusted to the desired values using the column-wise parallel programming method. To demonstrate the image feature extraction operation, a grayscale Lena image with 64 × 64 pixels was used as the input image. The intensity value of each pixel, which represents the brightness of the pixel in the input image, has a range from 0 to 255. The intensity of each pixel was converted into a voltage from 0 to 0.1 V. Then, the voltages that were converted from intensity values of 3 × 3 pixels in the Lena image were sequentially applied to the DLs in the ferroelectric synaptic transistor array and the total current of each column was summed, while all the GLs were set to ~2 V.

**Simulation of CNN based on ferroelectric synaptic transistor arrays**

An eight-layer visual geometry group (VGG-8) network for the Canadian Institute for Advanced Research (CIFAR-10) dataset was simulated to investigate the potential of using ferroelectric synaptic transistor arrays as the hardware for CNNs (12, 53). The VGG-8 network was composed of six convolutional, three max pooling, and two fully connected layers (Fig. 4A). One max pooling layer was used for every two convolutional layers. The size of the input CIFAR-10 images was 32 × 32 × 3 pixels. For the convolutional layers, kernels with 3 × 3 weights were used. When CIFAR-10 images were processed by the first and second convolutional layers, the size of the feature maps was 32 × 32 × 128. After the third and fourth convolutional layer operations, feature maps with a size of 16 × 16 × 256 were obtained. Then, feature maps with a size of 8 × 8 × 512 were acquired by the fifth and sixth convolutional layer operations. Feature maps with a size of 4 × 4 × 512 were obtained after max pooling process. The feature maps were connected to the fully connected layers (12). The convolutional and fully connected layers were used to extract features in the images and classify the images, respectively. The simulation tool used in this work is based on the identical pulse

---

**Fig. 4. CNN based on ferroelectric synaptic transistor arrays.** (A) Schematic illustration of VGG-8 network for the CIFAR-10 dataset. The VGG-8 network is composed of six convolutional, three max pooling, and two fully connected layers. The convolutional and fully connected layers are used to extract features in the images and classify the images, respectively. (B) Weight update characteristics of FeTFTs. For the linear weight update characteristics, incremental pulses (potentiation, 2.7 to 3.96 V with a 20-mV step; depression, −2.7 to −3.96 V with a −20-mV step) are used. (C) Comparison of simulated accuracies of VGG-8 network based on ferroelectric synaptic transistor arrays (blue circles and line) and ideal devices (green squares and line).
scheme to program the synapse array (54). Further simulation studies need to be done to reflect the exact operation scheme and processes including nonidentical pulse operations.

To investigate the synaptic characteristics of FeTFTs, we measured the potentiation and depression characteristics by using an incremental pulse scheme (potentiation, 2.7 to 3.96 V with a 20-mV step; depression, −2.7 to −3.96 V with a −20-mV step). The linearity of weight updates represents the linearity of the curve between the conductance of the FeTFTs and the number of pulses (11). The linearity of the weight update was evaluated using the following equations (12, 55)

$$G_{pot} = B(1 - e^{-\frac{P}{A_{pot}}}) + G_{min} \quad (1)$$

$$G_{dep} = -B(1 - e^{-\frac{P}{A_{dep}}}) + G_{max} \quad (2)$$

$$B = \frac{(G_{max} - G_{min})}{(1 - e^{-\frac{P}{A_{pot}}})} \quad (3)$$

where $P$, $G_{pot}$, and $G_{dep}$ are the number of pulses and conductance after potentiation and depression, respectively. $P_{max}$ is the maximum number of pulses. $G_{max}$ and $G_{min}$ are the maximum and minimum conductance, respectively. $A_{pot}$ and $A_{dep}$ represent the linearity of the potentiation and depression characteristics, respectively (37, 38). In potentiation and depression operations, the FeTFTs showed 64-level conductance states, a $G_{max}/G_{min}$ of 33.1, and a high linearity ($A_{pot} = 0.8139; A_{dep} = 1.1464$) (Fig. 4B). The low $G_{max}$ of FeTFTs in the weight update characteristics can be an advantage in the implementation of large arrays. With the high $G_{max}$ of synaptic devices, the current sum of the synapses may exceed the wire capacity of the arrays, which can lead to a reduction in the read accuracy (19). For an efficient CIM array, $G_{max}$ needs to be lower than 10 $\mu$S (12, 56). In addition, the small cycle-to-cycle and device-to-device conductance variation of the ferroelectric synaptic transistor array are desired characteristics for CIM (Fig. 2, D and E) (12).

The VGG-8 network was simulated using the synaptic characteristics of the ferroelectric synaptic transistor array, including the linearity of weight updates, the number of conductance states, $G_{max}/G_{min}$, the cycle-to-cycle variation, and the device-to-device variation characteristics. In simulations, the VGG-8 network based on a ferroelectric synaptic transistor array achieved 90.3% accuracy for 100 training epochs, which was comparable to the accuracy of 91.0% that the VGG-8 network based on ideal synaptic devices achieved (Fig. 4C) (11, 12). The high accuracy of the VGG-8 network based on a ferroelectric synaptic transistor array could be achieved because of weight update characteristics of the FeTFTs, such as high linearity ($A_{pot} = 0.8139; A_{dep} = 1.1464$), 64-level conductance states, $G_{max}/G_{min}$ of 33.1, and small variation characteristics. In addition, the potentiation and depression characteristics with low conductance variation resulted in a high recognition accuracy of 91% for CIFAR-10 images with VGG-8 network using off-chip training method (57). The change in the conductance of synaptic devices after training can cause degradation of the accuracy (58). Therefore, the retention characteristics of FeTFTs in different conductance states were also evaluated. The average change in the conductance of the FeTFTs was about 4.1%, and no overlap between the states was observed for 10,000 s (Fig. S10). The stable retention characteristics of the FeTFTs can be advantageous for maintaining accuracy (56). The simulation of a VGG-8 network confirmed that ferroelectric synaptic transistor arrays could be used as the effective neuromorphic hardware for CNNs.

**DISCUSSION**

In summary, we integrated FeTFTs based on IZO oxide semiconductors and ferroelectric HfZrO$_x$ to investigate the potential of the ferroelectric synaptic transistor array for use in CIM applications. To implement the weight update characteristics, we controlled the conductance of the FeTFTs by adjusting the polarization of the ferroelectric layer. In ferroelectric synaptic transistor arrays, column- and row-wise parallel programming methods were experimentally demonstrated by the selective control of the polarization switching using program-inhibit operations. Kernel weights for convolution operations were realized in the ferroelectric synaptic transistor arrays using parallel weight update processes. The FeTFTs exhibited linear current-voltage behavior and weight update characteristics, which were essential for synaptic devices used for convolution operations. Accurate convolution operations were demonstrated in the ferroelectric synaptic transistor arrays using the kernel weights in the array. These convolution operations based on ferroelectric synaptic transistor arrays enabled the extraction of the features in an input image with 64 × 64 pixels. To further investigate the potential of the ferroelectric synaptic transistor arrays as CNNs for CIM, we simulated the CNNs using the characteristics of the ferroelectric synaptic transistor arrays. In simulations based on the weight update characteristics of the ferroelectric synaptic transistor arrays, the CNNs achieved an image recognition accuracy of 90.3% for a CIFAR-10 dataset. These results suggest that ferroelectric synaptic transistor arrays based on IZO oxide semiconductors and ferroelectric HfZrO$x$ have the potential to be used as neuromorphic hardware for CNNs.

**MATERIALS AND METHODS**

**Materials**

Hf[N(C$_2$H$_5$)$_3$CH$_3$]$_4$ [tetraakis(ethylmethylamido)hafnium (TEMAH)] and Zr[N(C$_2$H$_5$)$_3$CH$_3$]$_4$ [tetraakis(ethylmethylamido)zirconium (TEMAZ)] were purchased from UP Chemical, Korea. C$_{10}$H$_{28}$NSi$_2$In$_4$ [bis(trimethylsilyl)amidodiethyl indium (INCA-1)] and Zn(C$_2$H$_3$)$_2$ [diethylzinc (DEZ)] were purchased from iChems, Korea. Si wafers with 100-nm-thick thermally grown SiO$_2$ were used as substrates.

**Device fabrication**

The ferroelectric synaptic transistor arrays were fabricated on the SiO$_2$/Si substrate. The photolithography was performed using a mask aligner (MDA-400LJ, Midas System). First, the SiO$_2$/Si substrate was cleaned in acetone, ethanol, and deionized water for 15 min each. W GLs were deposited on the substrate via DC sputtering. The W layer was patterned using the lift-off method. Then, a 24-nm-thick HfZrO$_x$ layer was deposited using the ALD by alternating the ALD cycles of HfO$_x$ and ZrO$_2$ at 280°C. For the ALD process, TEMAH, TEMAZ, and ozone were used as the Hf precursor, Zr precursor, and oxygen source, respectively. Next, Mo SLs and drain electrodes were deposited using the electron beam evaporator. The Mo layer was patterned using the lift-off method. A 10-nm-thick IZO film was deposited at 150°C using INCA-1, DEZ, and ozone as the indium precursor, zinc precursor, and oxygen source, respectively. The IZO layer was patterned using a combination of lithography and wet etching. The channel length and width were 10 and 50 $\mu$m,
respectively. Then, the devices were annealed at 400°C for 10 min in an N₂ environment. A 15-nm-thick HfO₂ layer was deposited as an interlayer dielectric using the ALD. The etching process was done to open the contacts for the SLs, GLs, and drain electrodes. Mo was then deposited and patterned for the formation of DLs using the same method described above.

Characterizations
All the characteristics were measured under ambient conditions and at room temperature. The electrical properties of the ferroelectric synaptic transistor array were measured using a semiconductor parameter analyzer (4200A-SCS, Keithley Instruments) and a switching matrix (707B, Keithley Instruments). The polarization-voltage-curves were measured using a pulse measurement unit (4225-PMU, Keithley Instruments). The polarization-voltage characteristics of the ferroelectric capacitor with a Mo/HfZrOₓ/W structure were measured after 10³ bipolar pulse cycles using voltage pulses with an amplitude of 5 V and a width of 10 μs. The thicknesses of the HfZrOₓ and IZO were measured using an atomic force microscope (NX10, Park Systems). VGG-8 network simulations were performed with a Linux system using C++ code and a Python wrapper (12). For recognition accuracy simulations, the learning rate was 1 for the initial 50 epochs and changed to 0.125 after 50 epochs. In this simulation, the characteristics of the ferroelectric synaptic transistor array, including Gₘax/Gₘin, linearity, the number of states, cycle-to-cycle variation, and device-to-device variation, were considered. For the simulation of the VGG-8 network based on ideal synaptic devices, ideal synaptic characteristics (including a perfectly linear conductance modulation of Gₘax/Gₘin = 100 and 100 conductance states) were used (11).

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abm8537

REFERENCES AND NOTES
1. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature 521, 436–444 (2015).
2. X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, Y. Shi, Scaling for edge inference of deep neural networks. Nat. Electron. 1, 216–222 (2018).
3. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Tabaa, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, D. S. Modha, A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 345, 668–673 (2014).
4. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, H. Bhaskaran, Parallel convolutional processing using an integrated photonic tensor core. Nature 589, 52–58 (2021).
5. R. A. John, R. A. Keshavarzi, S. Datta, The future of ferroelectric field-effect transistor devices. Nat. Electron. 11, 3211 (2020).
6. S. Li, M.-E. Pam, Y. Li, L. Chen, Y.-C. Chien, X. Fong, D. Chi, K.-W. Ang, Wafer-scale 2D hafnium disiloxane based memristor crossbar array for energy-efficient neural network hardware. Adv. Mater. 2103376 (2021).
7. E. J. Fuller, F. E. Gabaly, F. Léonard, S. Agarwal, J. J. Yang, Q. Wu, R. S. Williams, J. J. Yang, X. Qian, Efficient and self-adaptive in-situ learning in multilayer memristor neural network. Nat. Commun. 9, 2385 (2018).
8. C. Li, D. Belkin, Y. Li, P. Yan, M. H. Yu, G. H. Jiang, E. M. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnett, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, Efficient and self-adaptive in-situ learning in multilayer memristor neural network. Nat. Commun. 9, 574–579 (2020).
9. M. J. Marinella, J. J. Yang, A. Salleo, A. A. Talin, Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. Science 364, 570–574 (2019).
10. S. Kim, Y. Lee, H.-O. Kim, S.-J. Choi, Parallel weight update protocol for a carbon nanotube synaptic transistor array for accelerating neuromorphic computing. Nanoscale 12, 2046–2046 (2020).
11. M. Kim, J.-E. Lee, C. Lee, Y. Song, G. Han, J. Seo, D.-W. Kim, Y.-H. Seo, H. Hwang, D. Lee, Multimodal data processing based on nonlinear synaptic devices. J. Electron. Mater. 50, 3471–3477 (2021).
12. A. Prakash, J. Park, J. Song, J. Woo, E. Cha, H. Hwang, Demonstration of low power 3-bit multilevel cell characteristics in a TaOₓ-based RRAM stack via back-gate engineering. IEEE Electron Dev Lett. 36, 32–34 (2015).
13. C. Li, D. Belkin, Y. Li, P. Yan, M. H. Yu, G. H. Jiang, E. M. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnett, Q. Wu, R. S. Williams, J. J. Yang, X. Qian, Efficient and self-adaptive in-situ learning in multilayer memristor neural network. Nat. Commun. 9, 2385 (2018).
14. C. Jung, S. Lim, H. Kim, T. Kim, K. Moon, J. Song, J.-H. Kim, H. Hwang, Effect of conductance linearity and multi-level cell characteristics of TaOₓ-based synapse device on pattern-recognition accuracy of neuromorphic system. Nanotechnology 29, 115203 (2018).
15. K.-H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Sinrivasu, W. Lu, A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. Nano Lett. 12, 389–395 (2012).
16. L. Shi, G. Zheng, B. Tian, B. Dkhil, C. Duan, Research progress on solutions to the sneak path issue in memristor crossbar arrays. Nano Adv. 2, 1811–1827 (2020).
17. S. Seo, J.-J. Lee, H.-J. Lee, H. Hwang, S. Oh, J. J. Lee, K. Heo, J.-H. Park, Recent progress in artificial synapses based on two-dimensional Van der Waals materials for brain-inspired computing. ACS Appl. Electron. Mater. 2, 371–388 (2020).
18. P. Gkoupidenis, D. A. Koutsouras, G. G. Malliaras, Neuromorphic device architectures with global connectivity through electrolyte gating. Nat. Commun. 8, 15448 (2017).
19. H. Han, Y. H. Wei, J. Gong, W. Xu, Recent progress in three-terminal artificial synapses: From device to system. Small 15, 1900695 (2019).
20. S. Yamamoto, G. G. Malliaras, Controlling the neuromorphic behavior of organic electrochemical transistors by blending mixed and ion conductors. ACS Appl. Electron. Mater. 2, 2224–2228 (2020).
21. E. J. Fuller, F. E. Gabaly, F. Léonard, S. Agarwal, J. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, A. A. Talin, Li-ion synaptic transistor for low power analog computing. ACS Appl. Electron. Mater. 2, 371–388 (2020).
22. M. Jerry, P. Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, Ferroelectric FEAT analog synapse for acceleration of deep neural network training, in 2017 IEEE International Electron Devices Meeting (IEDM) (IEEE, 2017), pp. 6.2.1–6.2.4.
23. A. I. Khan, A. Keshavarzi, S. Datta, The future of ferroelectric field-effect transistor technology. Nat. Electron. 3, 588–597 (2020).
24. H. J. Lee, M. Lee, K. Lee, J. Jo, H. Yang, Y. Kim, S. C. Chae, U. Waghmare, J. H. Lee, Scale-free ferroelectricity induced by flat phonon bands in HfO₂. Science 369, 1343–1347 (2020).
25. S. S. Cheema, D. Kwon, N. Shanker, R. dos Reis, S.-L. Hsu, J. Xiao, H. Zhang, R. Wagner, A. Datar, M. R. McCarter, C. R. Serrao, A. K. Yadav, G. Karbsian, C.-H. Hsu, A. J. Tan,
37. M.-K. Kim, J.-S. Lee, Ferroelectric analog synaptic transistors. Nano Lett. 19, 2044–2050 (2019).

38. M.-K. Kim, I.-J. Kim, J.-S. Lee, Oxide semiconductor-based ferroelectric thin-film transistors for advanced neuromorphic computing. Appl. Phys. Lett. 118, 032902 (2021).

39. C. Alessandri, P. Pandey, A. Abusleme, A. Seabaugh, Switching dynamics of ferroelectric Zr-doped HfO₂. IEEE Electron Device Lett. 39, 1780–1783 (2018).

40. D. H. Lee, Y. Lee, K. Yang, J. Y. Park, S. H. Kim, P. R. S. Reddy, M. Materano, H. Mulaosmanovic, K. Ni, X. Li, J. A. Smith, M. Jerry, S. Datta, Write disturb in ferroelectric FETs and its implication for 1T-FeFET AND memory arrays. IEEE Electron Device Lett. 39, 1656–1659 (2018).

41. X. Zhang, M. Takahashi, K. Takeuchi, S. Sakai, 64 kbit ferroelectric-gate-transistor-integrated NAND flash memory with 7.5 V program and long data retention. Jpn. J. Appl. Phys. 51, 04DD01 (2012).

42. M.-K. Kim, I.-J. Kim, J.-S. Lee, CMOS-compatible ferroelectric NAND flash memory for high-density, low-power, and high-speed three-dimensional memory. Sci. Adv. 7, eabe1341 (2021).

43. P. Wang, S. Yu, Ferroelectric devices and circuits for neuro-inspired computing. MRS Commun. 10, 538–548 (2020).

44. S. Liu, K. Li, Y. Sun, X. Zhu, Z. Li, B. Song, H. Liu, Q. Li, A TaO₅-based electronic synapse with high precision for neuromorphic computing. IEEE Access 7, 184700–184706 (2019).

45. M. Hu, C. E. Graves, C. Li, Y. Li, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Q. Xia, J. P. Strachan, Memristor-based analog computation and neural network classification with a dot product engine. Adv. Mater. 30, 1705914 (2018).

46. J. Go, Y. Kim, M. Kwak, J. Song, S. A. Chekol, J.-D. Kwon, H. Hwang, W/WO₃₋ₓ based three-terminal synapse device with linear conductance change and high on/off ratio for neuromorphic application. Appl. Phys. Express 12, 026503 (2019).

47. C. Lee, J.-E. Lee, M. Kim, Y. Song, G. Han, J. Seo, D.-W. Kim, Y.-H. Seo, H. Hwang, D. Lee, Li memristor-based MOSFET synapse for linear I–V characteristic and processing analog input neuromorphic system. Jpn. J. Appl. Phys. 60, 024003 (2021).

48. L. Gao, P. Chen, S. Yu, Demonstration of convolution kernel operation on resistive cross-point array. IEEE Electron Device Lett. 37, 870–873 (2016).

49. C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Davila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, Analogue signal and image processing with large memristor crossbars. Nat. Electron. 1, 52–59 (2018).

50. M. Kwak, J. Park, J. Woo, H. Hwang, Implementation of convolutional kernel function using 3-D TiO₂ resistive switching devices for image processing. IEEE Trans. Electron Devices 65, 4716–4718 (2018).

51. Fractal Coding and Analysis Group, University of Waterloo [online]; https://links.uwaterloo.ca/Repository.html.

52. A. Krizhevsky, ”Learning Multiple Layers of Features from Tiny Images” (2009; www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf).

53. M. Jerry, S. Dutta, A. Kazemi, K. Ni, J. Zhang, P.-Y. Chen, P. Sharma, S. Yu, X. S. Hu, M. Niemier, S. Datta, A ferroelectric field effect transistor based synaptic weight cell. J. Phys. D Appl. Phys. 51, 434001 (2018).

54. P. Y. Chen, X. Peng, S. Yu, NeuroSim: A circuit-level macro model for benchmarking neuro-inspired architectures in online learning. IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. 37, 3067–3080 (2018).

55. X. Peng, S. Huang, Y. Luo, X. Sun, S. Yu, DNN-NeuroSim: An end-to-end benchmarking framework for compute-in-memory accelerators with versatile device technologies, in 2019 IEEE International Electron Devices Meeting (IEDM) (IEEE, 2019), pp. 32.5.1–32.5.4.

56. T.-H. Kim, S. Kim, K. Hong, J. Park, Y. Hwang, B.-G. Park, H. Kim, Multilevel switching memristor by compliance current adjustment for off-chip training of neuromorphic system. Chaos, Solitons Fractals 131, 11587 (2021).

57. W. Zhang, B. Gao, J. Tang, P. Yao, S. Yu, M.-F. Chang, H.-J. Yoo, H. Qian, H. Wu, Neuro-inspired computing chips. Nat. Electron. 3, 371–382 (2020).

Acknowledgments: We thank S. Kim at POSTECH and J. Woo at Kyungpook National University for comments on this manuscript. Funding: This work was supported by Samsung Research Funding and Incubation Center of Samsung Electronics under project no. SRFC-TA1903-05. This work was also supported by the National Research Foundation of Korea (NRF-2016M3D1A1027663, NRF-2019R1A2C2084114, and NRF-2020M3F3A2A01081774). The Sentsaurus TCAD simulator was provided by the Electronic Design Automation (EDA) tool program of IC Design Education Center (IDEC) in Korea. Author contributions: J.-S.L. conceived and directed the research. J.-S.L. and M.-K.K. designed and planned the experiment. J.-S.L. and M.-K.K. performed the experiment and acquired the data. J.-S.L., M.-K.K., and J.-S.L. wrote the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The CIFAR-10 dataset used for image classification is available under the Massachusetts Institute of Technology (MIT) License at www.cs.toronto.edu/~kriz/cifar.html.