Supporting Information

Complementary resistive switching and synaptic-like memory behavior in an epitaxial $SrFeO_{2.5}$ thin film through oriented oxygen vacancy channels

Venkata Raveendra Nallagatla[1], Jihun Kim[2], Kyoungjun Lee[3], Seung Chul Chae[3], Cheol Seong Hwang[2] and Chang Uk Jung[1]**

[1]Department of Physics and Memory and Catalyst Research Center, Hankuk University of Foreign Studies, Yongin 17035, Korea (**email:** cu-jung@hufs.ac.kr)

[2]Department of Material Science and Engineering and Research Institute of Advanced Materials, Seoul National University, Seoul, 08826, Korea. (**email:** cheolsh@snu.ac.kr)

[3]Department of Physics Education, Seoul National University, Seoul 08826, Korea

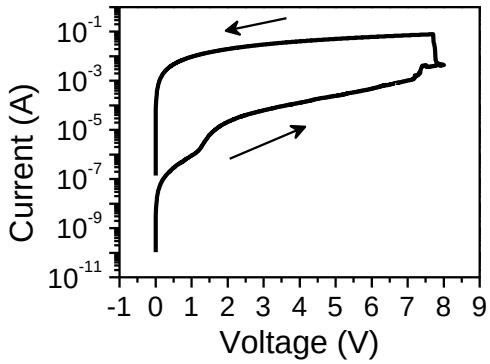

Figure S1. Positive bias electroforming process in $SrFeO_x$ (111) device.

In $SrFeO_x$ (111) device structure, both the Au/BM-$SrFeO_{2.5}$ and the BM-$SrFeO_{2.5}$/$SrRuO_3$ interfaces are Schottky-like contacts because of the high work functions of Au (~5.1 eV) and $SrRuO_3$ (~5.2 eV) metal contacts. Therefore, initial electroforming process can be achieved in both positive and negative bias at the Au TE. For positive electroforming require larger voltage as compare to negative bias electroforming process. This can be understandable considering that the Au TE is the exposed atmosphere; therefore, it can possess an infinite amount of oxygen source in the negative bias. In contrast, epitaxially grown $SrRuO_3$ thin films are underneath of continuous BM-$SrFeO_{2.5}$ switching layer and it is known to be a very stable oxide based metal electrode. As a result, the SrRuO3 BE is less efficient for oxygen supply. Therefore, the devices require large voltages (~ -8 V) for establishing the electroforming process in positive bias (i.e., negative bias at BE), as shown in the above Figure S1.

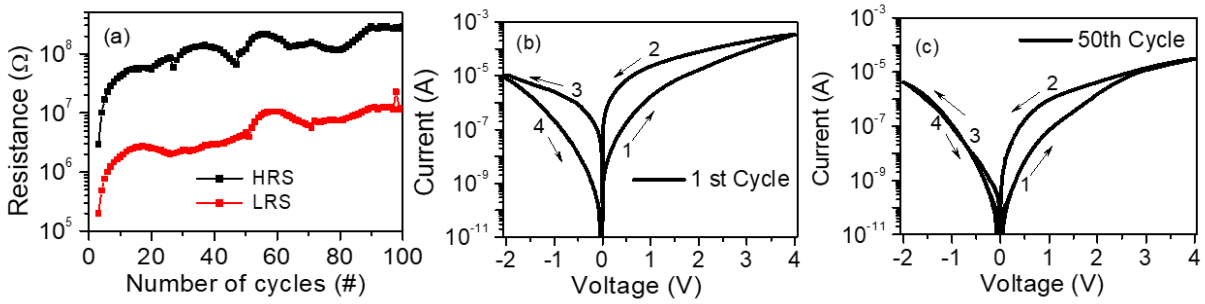

Figure S2. (a) Endurance of the SrFeO*x* (111) device in low-bipolar resistive switching mode, where the resistance readout voltage is 0.25 V. Representative *I–V* curves of (b)1st and (b) 50th cycle.

Figure S2a shows the endurance test of the $SrFeO_x$ (111) device in low-bipolar resistive switching mode up to 100 cycles. Here, the resistance values were measured at 0.25 V. The resistance values of high resistance state and low resistance state were gradually increased with a number of switching cycles. However, the ON/OFF ratio of the device was still maintained more than one order of magnitude up to 100 switching cycles. The initial switching cycle showed asymmetric curves with apparent resistance splitting in forward and reverse voltage sweeping directions for both positive and negative voltage regions, as shown in Figure S2b. At a higher number of cycles, the shape of the hysteresis changed drastically. For example, the resistance splitting of the 50th cycle is almost zero in the negative voltage regain, as shown in Figure S2c. The observed increase of overall device resistance with the number of switching cycles might be related to the variation of Schottky barrier height and width at the top interface (Au/BM-$SrFeO_{2.5}$) during the switching cycles. This Schottky barrier height and width variation can be explained as follows. As prepared BM-$SrFeO_{2.5}$ (111) films were exposed to air before top electrode deposition. Subsequently, surface-oriented OVCs in the BM-$SrFeO_{2.5}$ (111) film were also exposed to air. As a result, oxygen from the air is more likely to be incorporated into the films or at the surface. When a negative bias was applied to the top electrode during the endurance test, those incorporated

oxygen ions might gradually drift towards the bottom interface (BM-$SrFeO_{2.5}$/$SrRuO_3$), while increasing the Schottky barrier height and width at the top interface (Au/BM-$SrFeO_{2.5}$). This might be the reason for the increase of overall device resistance with the number of switching cycles in low-bipolar resistive switching mode.

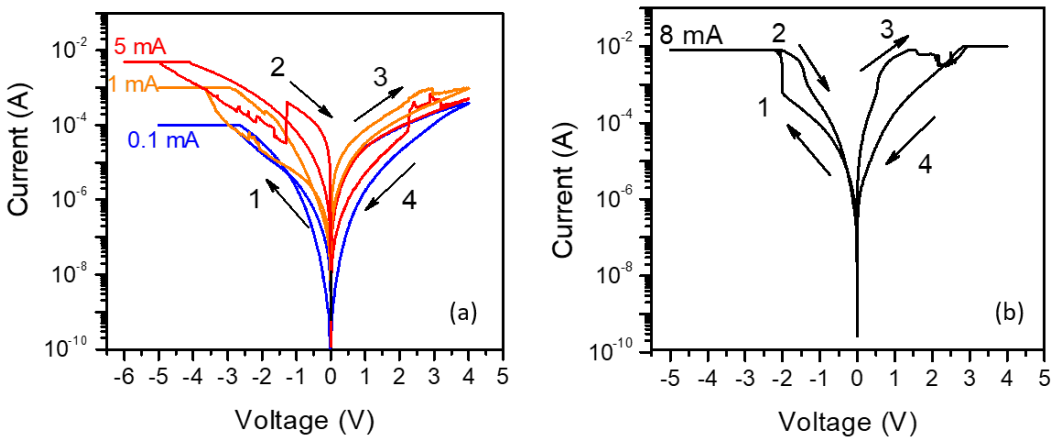

Figure S3. (a) *I-V* characteristics of the $SrFeO_x$ (111) device with different compliance current with a negitive bias. (b) An exemplary complemnatray resistive switching *I-V* curves with 8 mA compliance current limit.

The *I-V* curves with compliance current limit ranging from 0.1 mA to ~5 mA showed no complemnatray resistive switching (CRS) behavior in most of the devices, as shown in Figure S3a. Conversely, CRS behavior was observed when the compliance current limit ranging from 5 mA to 10 mA. An exemplary CRS *I-V* curves with 8 mA compliance current limit is shown in below Figure S3b.

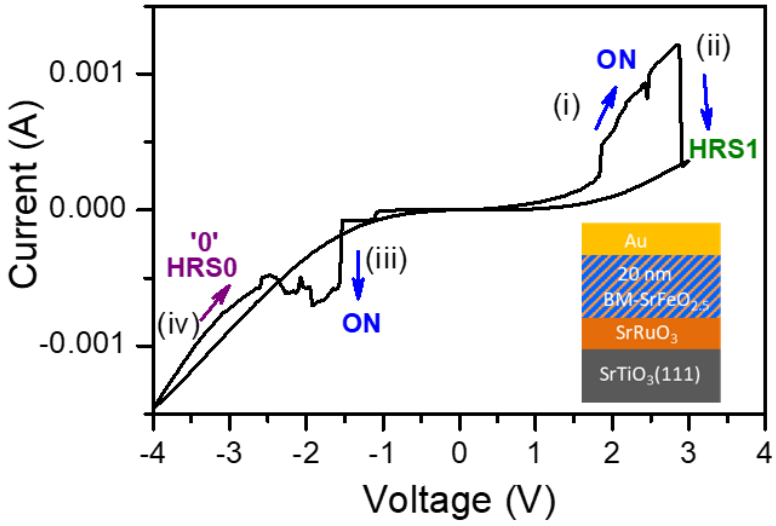

Figure S4. Complementary resistive switching characteristics of a $SrFeO_x$ (111) memory device with a 20 nm thick BM-$SrFeO_{2.5}$ switching layer.

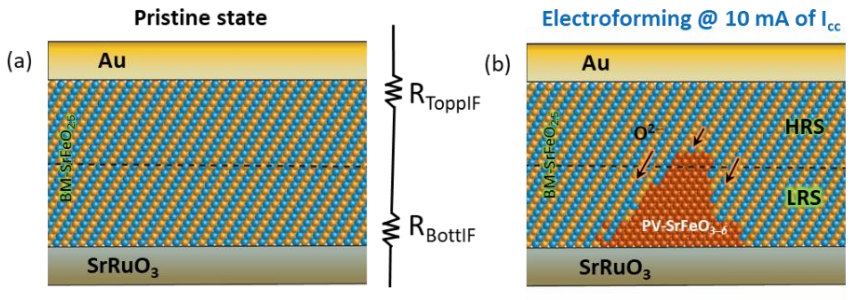

Figure S5. Schematic represenation of electroforming process in $SrFeO_x$ (111) device. (a) Pristine state, and (b) after electroforming with 10 mA of I_{cc} limit.

In $SrFeO_x$ (111) device structure, both the Au/BM-$SrFeO_{2.5}$ and the BM-$SrFeO_{2.5}$/$SrRuO_3$ interfaces are Schottky-like contacts because of the high work functions of Au (~5.1 eV) and $SrRuO_3$ (~5.2 eV) metal electrodes. Thus, the $SrFeO_x$ (111) device was assumed to consist of two resistor regions in a series connection: one at the Au/BM-$SrFeO_{2.5}$ top interface (R_{TopIF}) and the other at the BM-$SrFeO_{2.5}$/$SrRuO_3$-bottom interface (R_{BottIF}), as shown Figure S5a. During the electroforming process in the negative bias region with an I_{cc} of 10 mA, oxygen species from the atmosphere (e.g., O_2 or H_2O) can be incorporated into the BM-$SrFeO_{2.5}$ (111) switching layer and then drift towards the BE interface (BM-$SrFeO_{2.5}$/$SrRuO_3$) through the OVCs. Consequently, the conducting oxygen-rich phase (PV-$SrFeO_{3-\delta}$) accumulated at the BE interface (see Figure S5b), which led to a reduction in the Schottky barrier height at the BE interface. Therefore, the R_{BottIF} was decreased. In contrast, the R_{TopIF} was remained at the high value because of the lack of oxygen ions in the vicinity of the Au/BM-$SrFeO_{2.5}$ top interface due to a strong negative electrical field. Moreover, an application of the I_{cc} (10 mA) limit during the negative bias further limits the amount of oxygen from the air through Au TE. Thus, the low resistance at BE interface ((R_{BottIF}) and the high resistance at the TE interface (R_{TopIF}) induces the overall device resistance to be intermediate. This intermediate resistant state act as an HRS for the CRS process.

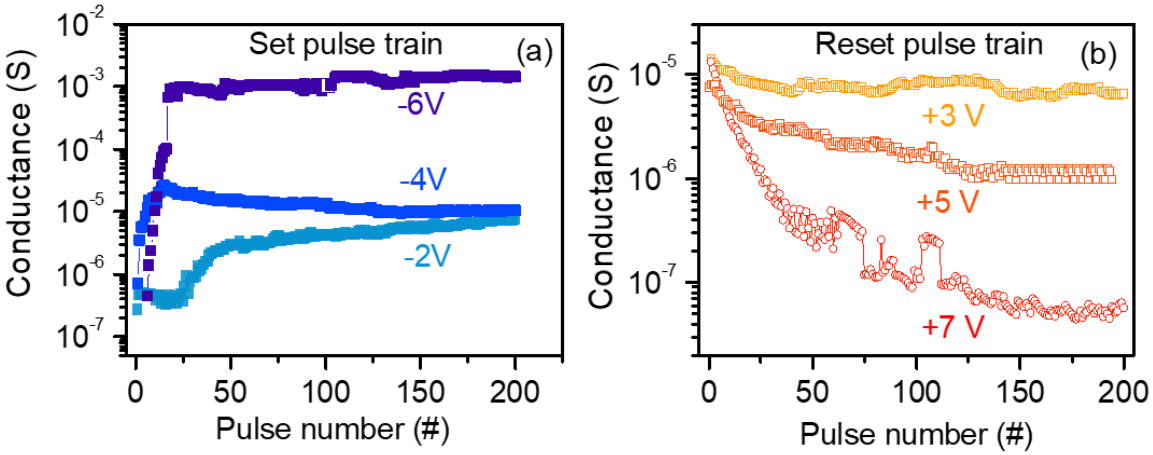

Figure S6. (a) and (b) Pulse amplitude-dependent conductance with 200 set and reset pulses, with a pulse width fixed at 1 ms and pulse amplitude varying from -2 V to -6 V for the potentiation and 3 V to 7 V for the depression, respectively.

The pulse amplitude varied from -2 V to -6 V at a fixed width of 1 ms for the potentiation and from 3 V to 7 V at a fixed width of 1 ms for the depression. Potentiation with a pulse amplitude of -6 V showed high nonlinearity in the conductance change compared to -2 to -4 V, as shown Figure S6a. In contrast, the depression showed a gradual conductance change even at high amplitude (see Figure S6b).

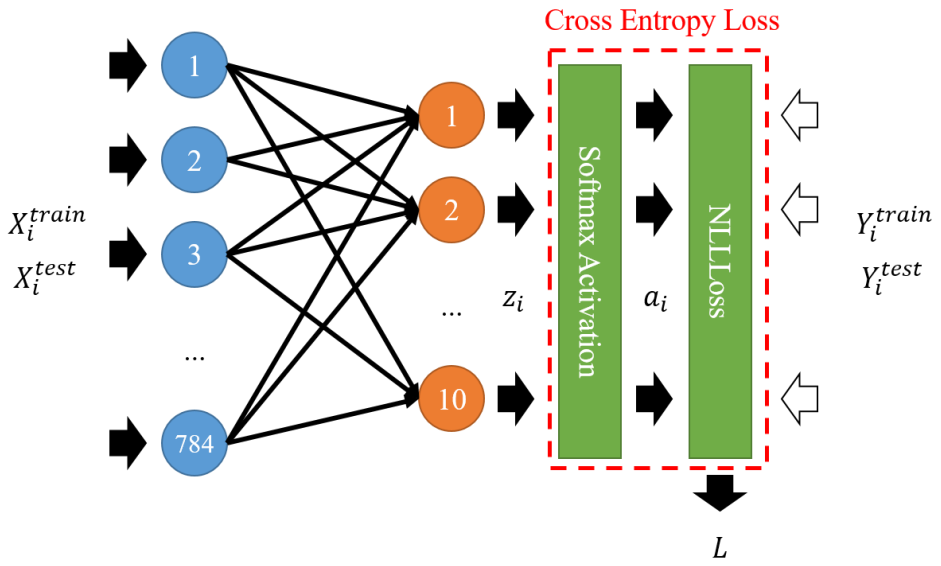

Figure S7. (a) Structure of memristive artificial neural network featuring FC single layer and Cross Entropy Loss

Memristive artificial neural network simulation

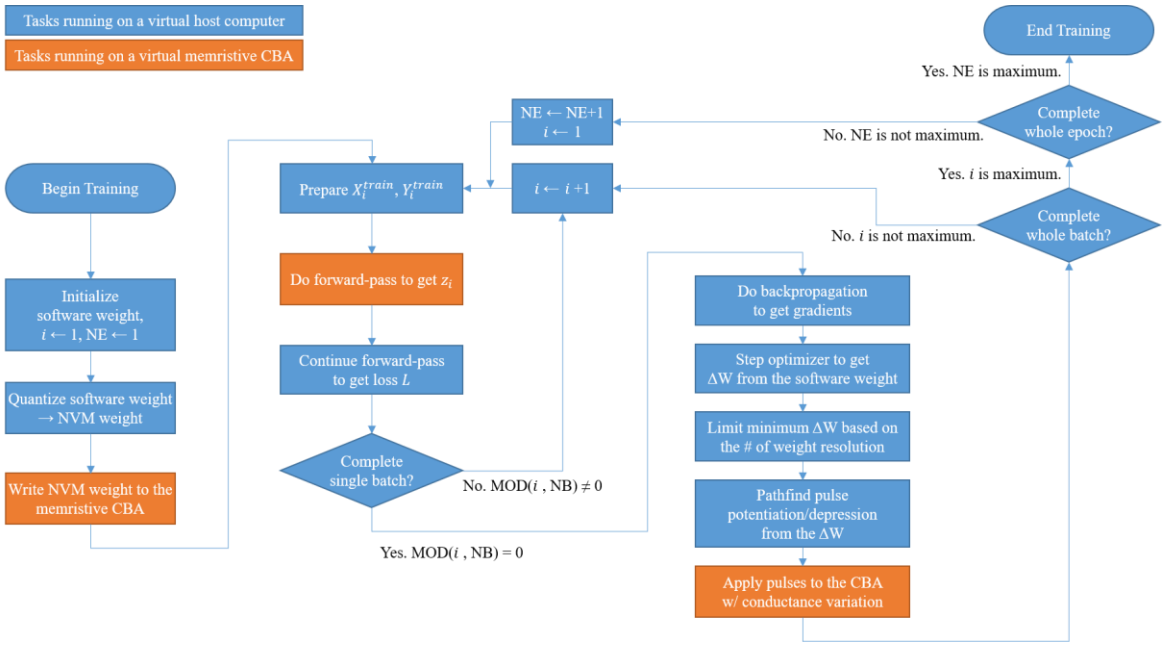

Figure S7. (b) Flowchart of the training algorithm in this work.

Mapping between the device and the simulation framework:

For the simulation in this work, the virtual memristive crossbar array (CBA) is assumed because it has the capability of forwarding operation in parallel by selecting multiple word-lines (WL) and bit-lines (BL) simultaneously. This can be accomplished according to Kirchhoff's current law (KCL), which allows the efficient vector-matrix multiplication by the physical mean in the memristive CBA-based Artificial Neural Network (ANN). Each component of the memristive CBA should be effectively correlated to the software-based ANN to simulate the memristive CBA as ANN. The weight of the ANN can be expressed using the conductance of the memristor. To express a negative weight value using the nonnegative conductance of a memristor, a paired-cell scheme is used by combining two memristor cells, where a weight value is represented by the conductance difference, as shown in Equation S1:

$$W = G^{+} - G^{-} \ldots (S1)$$

The input of the ANN can be expressed using the (read) voltage that fed into the WLs of the CBA. In this simulation, the conductance of a memristor is assumed to be independent of the read voltage. No bias of the ANN is considered in this simulation for simplicity, but several kinds of research treat the bias as a weight with constant input.[1,2] Activation of the ANN is assumed to be accomplished in the virtual host computing system, which also deals with the entire training process and controls read and write operations of the memristive CBA. On-chip supervised learning is assumed during the whole training process. That is, current read for the forward pass and memristor write for weight update are performed in the memristive CBA at each epoch, with the label for the input provided in the host computer. Other tasks, such as loss and gradient

calculation, were also performed at the host computer. The detailed training process is explained in the Training and Optimization section.

MNIST dataset:

The MNIST dataset features 28 × 28 sized grayscaled handwritten numeric digits. It contains 60,000 images for the training and 10,000 images for the test. Each image in the train set and test set were reshaped into a 784 × 1 single column to fit with the one-dimensional WL of the memristive CBA. These vectorized train and test sets are denoted as X_i^{train} and X_i^{test}, respectively. The i represents the data index in each data set. Labels for the train and test sets were prepared using one-hot encoding. For example, the label '3', which corresponds to the handwritten digit '3', is encoded into vectorized value [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]. In this vector, only the n^{th} element is 1, while the others are 0 to express 'n-1' is the correct label for the corresponding input. These one-hot encoded train and test sets are denoted as Y_i^{train} and Y_i^{test}, respectively. This vectorization is performed to fit with the one-dimensional BL of the memristive CBA.

Neural Network Structure:

During the simulation, a fully connected (FC) network with a single layer was used. The layer consists of 784 input neurons and 10 output neurons, which corresponds to the dimensions of X_i^{train}(X_i^{test}) and Y_i^{train}(Y_i^{test}), respectively. After the output neuron, cross-entropy loss is applied, which is a combination of softmax activation (Equation 2) and negative-log likelihood loss

(Equation 3). That is, negative-log likelihood loss is applied after the softmax activation is applied to the output from the FC single layer:

$$a_i = \frac{\exp(z_i)}{\sum_i \exp(z_i)} \dots (2)$$

$$L_i = -\log(a_i) \dots (3)$$

In Equation 2, z_i represents the sum-product of weight and input from the FC layer, which corresponds to the output current from the memristive CBA. In the simulation, a partial loss is generated from the 10 output neurons and averaged into the single loss value L. A detailed outline of the structure is in Figure S7a.

Training and Optimization:

In this work, a batch size of 32 was used during training. That is, each gradient update is performed after 32 times of forward-pass. An adequate number of mini-batch size is important to achieve both robustness for the variation and the minimized number of the read/write access to reach the target test accuracy.[3]

All parameters, including reading currents, previous weights, and inputs, were cached into the virtual host computer to calculate the gradients precisely. Calculated gradients are used in the backpropagation process, which determines the extent of updates for weights to increase the ANN performance.[4]

Gradients are applied using the optimization function named Adam, which is a well-known optimizer that features adaptive learning speed and direction momentum to maximize the efficiency of the gradient update.

Several features were implemented into the simulation that consider the mechanics of the real memristors, which usually contain nonideal factors, such as linearity and asymmetry of the conductance change and endurance problems. First, the conductance change profile by potentiation and depression is used for conductance pathfinding as a form of conductance diamond.[5] This enables the host computer to update to the target conductance with the minimized number of updates and maximum accuracy. Second, the conductance update threshold and adaptive learning rate based on the number of the available states are used to confine the number of updatable cells per iteration. The threshold minimizes the amount of write access to the memristors due to the conductance variation. It also enhances the simulation speed that may benefit the realistic host systems. The adaptive learning rate is introduced to sustain the conductance update with the update threshold. For example, software-based ANN requires a learning rate of 0.002 to maximize the accuracy and reliability of the ANN. In the case of the 5-bit memristive ANN, none of the memristors is updated with the same learning rate due to the update threshold. In this work, 0.02 and 0.012 were used for the 5-bit and 6-bit memristive ANN, respectively. A detailed flowchart of the whole training algorithm is provided in Figure S7b.

References

(1) Prezioso, M.; Merrikh-Bayat, F.; Hoskins, B. D.; Adam, G. C.; Likharev, K. K.; Strukov, D. B. Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors. *Nature* **2015**, *521*, 61–64.

(2) Kim, G. S.; Song, H.; Lee, Y. K.; Kim, J. H.; Kim, W.; Park, T. H.; Kim, H. J.; Min Kim, K.; Hwang, C. S. Defect-Engineered Electroforming-Free Analog HfO_x Memristor and Its Application to the Neural Network. *ACS Appl. Mater. Interfaces* **2019**, *11*, 47063–47072.

(3) LeCun, Y. A.; Bottou, L.; Orr, G. B.; Müller, K.-R. Efficient BackProp; Springer, Berlin, Heidelberg, 2012; pp 9–48.

(4) Kingma, D. P.; Lei Ba, J. *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*.

(5) Burr, G. W.; Shelby, R. M.; Sidler, S.; Di Nolfo, C.; Jang, J.; Boybat, I.; Shenoy, R. S.; Narayanan, P.; Virwani, K.; Giacometti, E. U.; Kurdi, B. N.; Hwang, H. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element. *IEEE Trans. Electron Devices* **2015**, *62*, 3498–3507.