Integrated Crossbar Array With Resistive Synapses and Oscillation Neurons

Jiyong Woo, Panni Wang, and Shimeng Yu

***Abstract*— In this letter, we fabricate a crossbar array that structurally resembles a column of the neural network, where one neuron is connected with multiple synapses in parallel for on-chip integration. Instead of using complex CMOS neuronal circuit, we integrate a threshold switch at the edge of the crossbar array as a compact oscillation neuron, which converts the weighted sum to an oscillation frequency. When the input vectors are loaded into multiple rows of the array, the oscillation frequency is measured to be proportional to the analog column current. This is the first experimental demonstration of an integrated crossbar array with both synapses and neurons, paving the path to a fully parallel computation and processing using emerging device technologies for the neuromorphic computing.**

***Index Terms*— Resistive synapse, oscillation neuron, crossbar array, neuromorphic computing.**

I. Introduction

ARTIFICIAL intelligence (AI) is currently promoting the development of new industrial applications such as self-driving cars and autonomous drones that need to recognize and classify visual and auditory objects in real time. Unlike the past AIs [1] that sought answers based on stored data and knowledge, state-of-the-art AI such as Google's AlphaGo [2] is able to learn the information on its own through algorithms. Inspired by the brain that processes data from one neuron to the other through a tremendous amount of synapses in parallel, algorithms based on deep neural networks have been developed [3]. These deep learning algorithms heavily rely on the time-consuming vector-matrix multiplications [4]. In this regard, a crossbar array with resistive memory that architecturally speeds up the multiplication in parallel has gain interests for neuromorphic computing [4]–[8]. What is often neglected in the prior works is the neuron node adjacent to the crossbar array. After the analog computation in the array, the read-out current at the end of each bit line (BL) needs to be digitized through the conventional silicon CMOS based neuronal circuit [9]. However, such circuit consisting of tens of transistors with a capacitor obviously occupies a much larger footprint than the BL pitch of the crossbar array. The pitch mismatch problem inevitably causes a single neuron node to be shared with multiple BLs. It means that the read-out currents computed from the synaptic arrays have to be sequentially processed. Recently, a compact threshold switch based neuronal device could potentially get rid of the complex CMOS circuitry and large capacitor, resulting in $>12.5\times$ reduced area through a circuit-level simulation study [10]. However, a single neuronal device with off-chip discrete load resistor has only been experimentally demonstrated so far [11]–[13]. In this work, we aim to integrate the neuronal device with the crossbar array on a single-chip to demonstrate the parallel computation along the BL.

Manuscript received May 7, 2019; revised June 1, 2019; accepted June 5, 2019. Date of publication June 7, 2019; date of current version July 24, 2019. This work was supported in part by NSF-ECCS under Grant No. 1701565, in part by ASCENT, one of the SRC/DARPA JUMP Centers. The review of this letter was arranged by Editor B. S. Doyle. *(Jiyong Woo and Panni Wang contributed equally to this work.) (Corresponding author: Shimeng Yu.)*

The authors are with the School of Electrical Computer and Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: shimeng.yu@ece.gatech.edu).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LED.2019.2921656

II. Experiments

The schematic diagrams of the fabrication were shown in Fig. 1. A thin resistive switching layer with 5-nm-thick HfO_2 was deposited on 12 horizontal TiN word lines (WLs) serving as bottom electrodes (BEs) of 10 μm width. A single Pt BL of 10 μm width as a middle electrode (ME) covered the WLs vertically. As a result, 12 Pt/HfO_2/TiN (top to bottom) resistive memories with size of $10 \times 10 \mu m^2$ were formed at the intersections of the WLs and BL. Then, a 15-nm-thick NbO_2 layer was deposited at the end of the Pt ME line. Another horizontal Pt line of 10 μm width as a top electrode (TE) overlapped the ME line. It determined the active area (10×10 μm^2) of the Pt/NbO_x/Pt threshold switch. Smaller line widths (down to 1 μm) for smaller active device area have been fabricated as well. Fig. 2 show optical microscopic images of the fabricated one-dimensional (1-D) 12×1 crossbar array.

III. Results and Discussion

First, the quasi-DC current-voltage (I-V) characteristics of the resistive memory and threshold switch in the array were evaluated using a probe card connected to a switching matrix. Previous X-ray photoelectron spectroscopy (XPS) analysis of the resistive memory showed that oxygen vacancies due to non-bridging oxygen ions were observed in the HfO_2 film [14]. After an initial forming at about 5 V, when the positive voltage and ground were biased to the BE and ME respectively, the oxygen vacancies were driven towards the Pt ME under the electric field. This resulted in a conductive filament throughout the HfO_2 layer [15], as shown in Fig. 3a. This process led the resistive memory to transition to a low resistance state (LRS). On the other hand, the negative voltage between the BE and ME caused the oxygen vacancies to escape from the filament. The LRS was switched to a high resistance state (HRS) by rupture of the conductive filament. We then programmed all the resistive memories to LRS along the BL by sequentially applying the voltages from cell to another. The median value of the LRS resistances (R_{LRS}) was about 58 kΩ at 1.8 V due

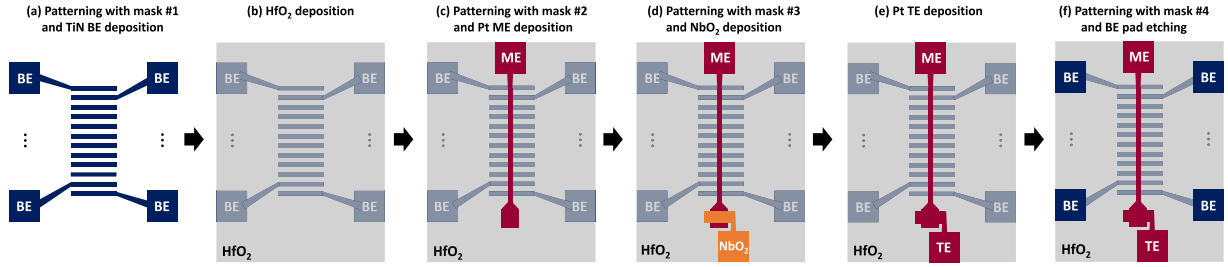

Fig. 1. (a) The TiN BE lines were formed. (b) The HfO_2 layer was deposited by ALD on the entire wafer. (c) The Pt ME line used to monitor the oscillation was located across the 12 TiN BE lines. Up to this step, the Pt/HfO_2/TiN resistive memories with were formed at each cross-point. The (d) NbO_2 and (e) Pt TE line were sequentially deposited by sputter and evaporation at the end of the Pt ME line, resulting in vertically stacked NbO_2 based threshold switch at the edge of the crossbar array. (f) Finally, the HfO_2 layer on top of the BE pads was etched for bottom contact.

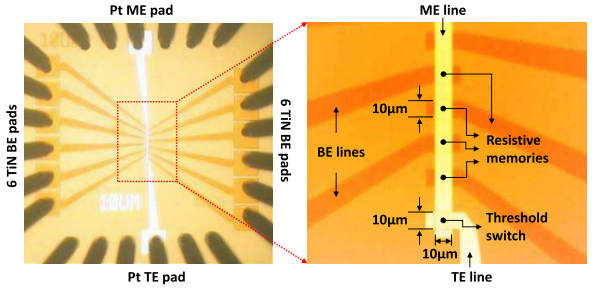

Fig. 2. Optical microscopic images of the 12 × 1 array consisting of 10 × 10 μm^2 sized resistive memories and 10 × 10 μm^2 sized threshold switch.

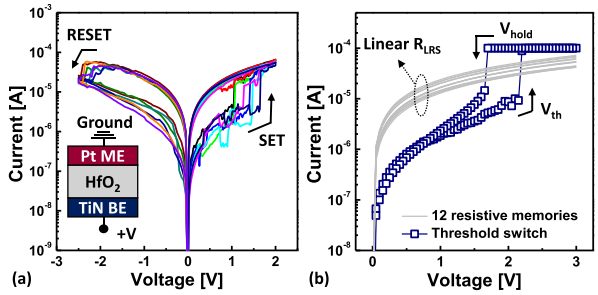

Fig. 3. The quasi-DC I-V traces of the (a) Pt/HfO_2/TiN resistive memory and (b) Pt/NbO_2/Pt threshold switch in the array.

to a self-compliance behavior in the LRS, as shown in gray lines in Fig. 3b. The observed self-compliance behavior has been explained as the chemically mixed layer at the HfO_2 and TiN interface serves as an internal resistor [16] to limit the current flowing through the formed filament. Meanwhile, the ME was used as a common ground for both resistive memory and threshold switch. Applying positive voltage to the TE thus triggered a threshold switching behavior of the NbO_2 layer after a forming process at about 4 V. An off-state was abruptly changed at a threshold voltage (V_{th}) of about 2 V, and on/off ratio of ~20 at 1.8 V was shown under compliance current of 100 μA, as shown in Fig. 3b. The on-state of the threshold switch immediately returned to the off-state when the voltage was reduced below a hold voltage (V_{hold}) of 1.5 V after the voltage source was removed.

Next, input voltage (V_{input}) pulses (6 V, 180 μs) were applied to the BEs in parallel, as shown in Fig. 4. The V_{input} pulse was addressed to only one of the BEs, and the remaining BEs were floating. The V_{input} multiplied by the $1/R_{LRS}$ at the selected resistive memory was expected to be observed as a read-out current along the BL at the grounded TE via the NbO_x. An oscillation was monitored in real time at the ME while the read-out current was flowing. Since the resistance of the off-state (R_{off}) of the threshold

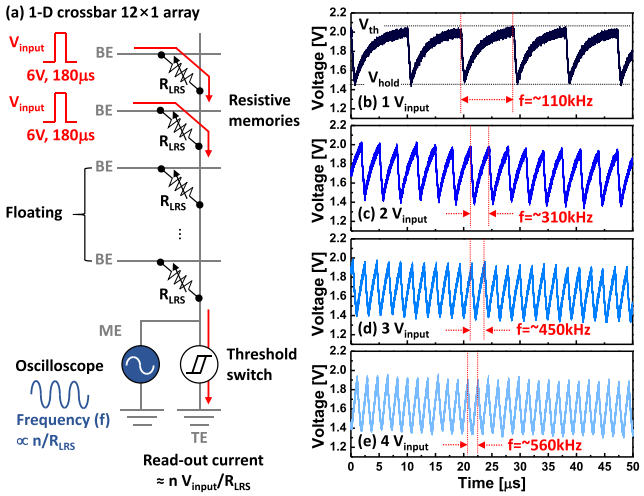

Fig. 4. (a) n V_{input} pulses were provided to the BE pads. (b)–(d) The oscillations with different frequencies were observed depending on the number of V_{input} pulses applied in parallel.

switch was greater than the R_{LRS} (Fig. 3b), most of the voltage initially began to be applied to the threshold switch. Due to the parasitic capacitance, the voltage at the ME that acts as an intermediate node between the two devices gradually increased. As the charged voltage at the ME exceeded the V_{th}, the off-state of the threshold switch was rapidly switched to the on-state. Because the on-state resistance (R_{on}) of the threshold switch was now lowered, the voltage at the ME began to discharge until the voltage remaining on the threshold switch reached the V_{hold}. The reversible transition of the threshold switch repeatedly induced the back and forth of the voltage charging and discharging. It caused the oscillation with a frequency of ~110 kHz in the range of V_{hold} of 1.5 V and V_{th} of 2 V. More importantly, as V_{input} number increased, a larger read-out current corresponding to the equivalently reduced total R_{LRS} was shown in the BL. This resulted in the steadily increased frequencies, as shown in Figs. 4b, 4c, and 4d. It can be further described analytically by solving the equation [10] based on Kirchhoff's Law on the configuration. The charging time (t_{rise}) from V_{hold} to V_{th} can be obtained from the analytical solution as follows:

$$t_{rise} = -R_{rise} \cdot C \cdot log(V_{th} - V_{input} \cdot R_{rise}/R_{LRS})/(V_{hold} - V_{input} \cdot R_{rise}/R_{LRS}) \quad (1)$$

where, the C is the parasitic capacitance to the ME. Similarly, the discharging time can be obtained as follows:

$$t_{fall} = -R_{fall} \cdot C \cdot log(V_{hold} - V_{input} \cdot R_{fall}/R_{LRS})/(V_{th} - V_{input} \cdot R_{fall}/R_{LRS}). \quad (2)$$

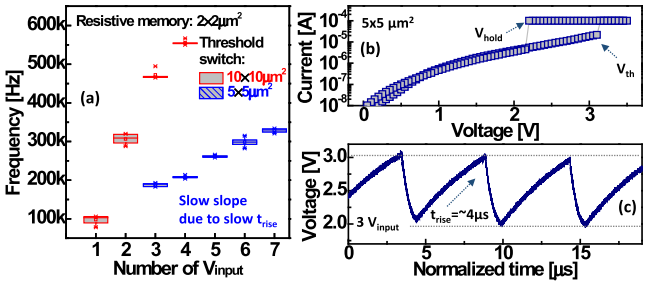

Fig. 5. (a) The oscillation frequency as a function of the number of V_{input} applied in parallel when varying the sizes of the threshold switch. (b) The I-V curve and (c) the oscillation behaivor of the $5 \times 5\ \mu m^2$ sized threshold switch.

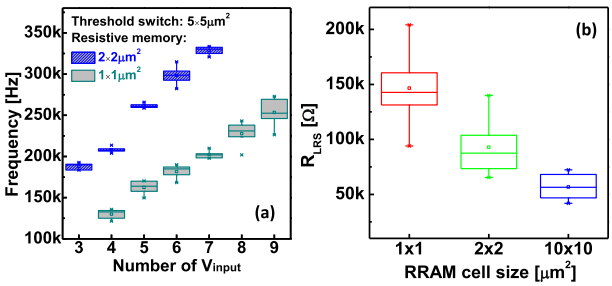

Fig. 6. (a) The oscillation frequency as a function of the number of V_{input} applied in parallel when varying the sizes of the resistive memory. (b) The R_{LRS} as a function of the size of the resistive memory extracted from multiple devices.

The condition for oscillation to occur is that the R_{LRS} of the resistive memories is between R_{off} and R_{on}. In above equations, the R_{rise} and R_{fall} become the R_{LRS} and R_{on} as the charging is through the resistive memory, and the discharging is through the on-state of the threshold switch. Therefore, the t_{rise} is proportional to the R_{LRS}, while the t_{fall} is constant and small due to the small R_{on}. It causes the oscillation of the voltage to have an asymmetric triangular waveform and the oscillation frequency to be determined mainly by the t_{rise}. Note that the charging and discharging are both driven by the first-order RC circuit response. When the number of the resistive memories is small, the charging is slow. The exponential increase as predicted in the RC circuit response is thus shown (Fig. 4b). The charging can be faster, when more resistive memories are involved. However, since the maximum voltage is fixed at the V_{th} of the threshold switch, the voltage rapidly approaches the V_{th} in the initial exponential increase phase, which seems to be linear. If there is no maximum voltage limit to the V_{th}, the initial linear increase trends will saturate as well, showing the exponential tail. The oscillation was no longer observed when more than 4 V_{input} were applied. This is because the sum of the R_{LRS} from 5 resistive memories in parallel becomes too small, which is out of the range between the R_{off} and R_{on}.

As shown in Fig. 5a, the oscillation was observed when 3 to 7 resistive memories were involved in the weighted sum for a smaller-sized threshold switch ($5 \times 5\ \mu m^2$). In addition, the dependence of the frequency as a function of the number of the resistive memories seemed to be less prominent in smaller threshold switch size. It can be explained by the enlarged V_{th} and V_{hold} of the smaller threshold switch, as shown in the I-V curve (Fig. 5b). Through XPS analysis, the deposited NbO_x film was found to have a mixture of NbO_2 and Nb_2O_5 phases and non-bridging oxygen ions [13]. As the area of the NbO_x is reduced, the density of defects is decreased. Based on the percolation theory, it is suggested that the percolation path where the phase transition occurs depends on the density of defects [17]. Because the probability to form the path is lowered, higher V_{th} is required at the smaller-sized of the device. Considering the same C and R_{LRS}, the t_{rise} based on the equation (1) was found to be primarily affected by a logarithmic function representing the ratio of the V_{th} to V_{hold}. The calculated value of the logarithmic function at the small threshold switch was roughly twice due to the increased oscillation's magnitude. This was in good agreement with the experimental results (Fig. 5c), which showed approximately two times slower frequency in the $5 \times 5\ \mu m^2$ sized threshold switch than the $10 \times 10\ \mu m^2$. Furthermore, since the V_{th} was increased, the R_{off} measured at the V_{th} was lowered. It means that more resistive memories should be needed to meet the criteria for the oscillation. The changed switching parameters such as V_{th} and R_{off} of the threshold switch have affected the frequency and the criterion ($R_{off} > R_{LRS} > R_{on}$) for the oscillation. Meanwhile, adjusting the size of the resistive memories could shift the oscillation frequency range (Fig. 6a). As the self-compliance behavior of the resistive memories can be attributed to the interfacial resistance between TiN BE and HfO_2 layer, the R_{LRS} exhibited an area dependency, as shown in Fig. 6b. In the $1 \times 1\ \mu m^2$ sized resistive memory, the summed R_{LRS} from the small number of the resistive memories was too large to be placed in between the R_{off} and R_{on}. Therefore, the oscillation was observed when 4 to 9 resistive memories participated in the weighted sum (Fig. 6a).

The on/off ratio of ~20 in our NbO_2 based threshold switch was small, so that only a limited range of the weighted sum could be identified. Adding a tunneling oxide to NbO_2 may increase the on/off ratio to $> 10^2$ [18]. In addition, when the threshold switching is demonstrated by other mechanisms such as lone-pair electrons of chalcogen atoms [19] or self-dissolvable filament formation [20], the on/off ratio of $10^3 \sim 10^{10}$ can be achieved. Therefore, we expect that the weighted sum in a relatively larger crossbar array with tens to hundreds of synaptic cells can be successfully represented by distinguishable oscillation frequency, if appropriate device engineering is further applied. In prior benchmark [10], the weighted sum task can be performed by the oscillation neuron with energy 5 times less than the CMOS integrate and fire neuron circuit, which also consumes a lot of energy by generating output pulses proportional to the weighted sum. As the neuron becomes compact, the number of BLs shared by a single neuron can be reduced, throughput could be further improved.

IV. Conclusion

We demonstrated the parallel weighted sum operation in the 1-D 12×1 crossbar array with integrated synaptic devices and neuronal device that structurally emulates a part of the neural network. The synaptic weight was stored in each HfO_2 resistive memory, which enables highly dense array to accelerate the vector-matrix multiplication. We then showed that the compact NbO_x based threshold switch processes the sum of the weights from the 12×1 synaptic array by representing the oscillation frequency due to the phase transition mechanism. We also investigated how the frequency and amplitude of the oscillation varied with the sizes of the resistive memory and the threshold switch for future optimization.

REFERENCES

[1] M. Campbell, A. Hoane, Jr., and F. Hsu, "Deep blue," *Artif. Intell.*, vol. 124, nos. 1–2, pp. 57–83, Jan. 2002. doi: 10.1016/S0004-3702(01)00129-1.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–489, Jan. 2016. doi: 10.1038/nature16961.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. doi: 10.1038/nature14539.

[4] M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, J. J. Yang, and R. Williams, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," in *Proc. 53nd ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Jun. 2016, pp. 1–6. doi: 10.1145/2897937.2898010.

[5] G. W. Burr, R. M. Shelby, C. di Nolfo, J. W. Jang, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. Kurdi, and H. Hwang, "Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element," in *IEDM Tech. Dig.*, Dec. 2014, pp. 29.5.1–29.5.4. doi: 10.1109/IEDM.2014.7047135.

[6] M. A. Zidan, J. P. Strachan, and W. D. Lu, "The future of electronics based on memristive systems," *Nature Electron.*, vol. 1, pp. 22–29, Jan. 2018. doi: 10.1038/s41928-017-0006-8.

[7] Z. Wang, S. Joshi, S. Savel'ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, and J. Yang, "Fully memristive neural networks for pattern classification with unsupervised learning," *Nature Electron.*, vol. 1, pp. 137–145, Feb. 2018. doi: 10.1038/s41928-018-0023-2.

[8] F. Bayat, M. Prezioso, B. Chakrabarti, H. Nili, I. Kataeva, and D. Strukov, "Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits," *Nature Commun.*, vol. 9, p. 2331, Jun. 2018. doi: 10.1038/s41467-018-04482-4.

[9] G. Indiveri, B. Linares-Barranco, T. j. Hamilton, A. Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S.-V. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen, "Neuromorphic silicon neuron circuits," *Frontiers Neurosci.*, vol. 5, p. 73, May 2011. doi: 10.3389/fnins.2011.00073.

[10] P.-Y. Chen, J.-S. Seo, Y. Cao, and S. Yu, "Compact oscillation neuron exploiting metal-insulator-transition for neuromorphic computing," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2016, pp. 1–6. doi: 10.1145/2966986.2967015.

[11] N. Shukla, A. Parihar, E. Freeman, H. Paik, G. Stone, V. Narayanan, H. Wen, Z. Cai, V. Gopalan, R. Engel-Herbert, D. G. Schlom, A. Raychowdhury, and S. Datta, "Synchronized charge oscillations in correlated electron systems," *Sci. Rep.*, vol. 4, p. 4964, May 2014. doi: 10.1038/srep04964.

[12] K. Moon, E. Cha, J. Park, S. Gi, M. Chu, K. Baek, B. Lee, S. Oh, and H. Hwang, "High density neuromorphic system with $Mo/Pr_{0.7}Ca_{0.3}MnO_3$ synapse and NbO2 IMT oscillator neuron," in *IEDM Tech. Dig.*, Dec. 2014, pp. 17.6.1–17.6. doi: 10.1109/IEDM.2015.7409721.

[13] L. Gao, P.-Y. Chen, and S. Yu, "NbO_x based oscillation neuron for neuromorphic computing," *Appl. Phys. Lett.*, vol. 111, no. 10, 2017, Art. no. 103503. doi: 10.1063/1.4991917.

[14] R. Fang, Y. G. Velo, W. Chen, K. E. Holbert, M. N. Kozicki, H. Barnaby, and S. Yu, "Total ionizing dose effect of γ -ray radiation on the switching characteristics and filament stability of HfOx resistive random access memory," *Appl. Phys. Lett.*, vol. 104, no. 18, 2014, Art. no. 183507. doi: 10.1063/1.4875748.

[15] H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, "Metal–oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, Jun. 2012. doi: 10.1109/JPROC.2012.2190369.

[16] C. Lee, D. S. Lee, A. Benayad, S. Lee, M. Chang, M.-J. Lee, J. Hur, Y. B. Kim, C. J. Kim, and U.-I. Chung, "Highly uniform switching of tantalum embedded amorphous oxide using self-compliance bipolar resistive switching," *IEEE Electron Device Lett.*, vol. 32, no. 3, pp. 399–401, Mar. 2011. doi: 10.1109/LED.2010.2101044.

[17] J. H. Stathis, "Percolation models for gate oxide breakdown," *J. Appl. Phys.*, vol. 86, no. 10, p. 5757, 1999. doi: 10.1063/1.371590.

[18] M. Kang and J. Son, "Off-state current reduction in NbO_2-based selector device by using TiO_2 tunneling barrier as an oxygen scavenger," *Appl. Phys. Lett.*, vol. 109, no. 20, 2016, Art. no. 202101. doi: 10.1063/1.4967916.

[19] A. Velea, K. Opsomer, W. Devulder, J. Dumortier, J. Fan, C. Detavernier, M. Jurczak, and B. Govoreanu, "Te-based chalcogenide materials for selector applications," *Sci. Rep.*, vol. 7, Aug. 2017, Art. no. 8103. doi: 10.1038/s41598-017-08251-z.

[20] Z. Wang, M. Rao, R. Midya, S. Joshi, H. Jiang, P. Lin, W. Song, S. Asapu, Y. Zhuo, C. Li, H. Wu, Q. Xia, and J. Yang, "Threshold switching of Ag or Cu in dielectrics: Materials, mechanism, and applications," *Adv. Funct. Mater.*, vol. 28, no. 6, 2017, Art. no. 1704862. doi: 10.1002/adfm.201704862.