APL Materials

Tailoring the synaptic properties of a-IGZO memristors for artificial deep neural networks

Cite as: APL Mater. **10**, 011113 (2022); https://doi.org/10.1063/5.0073056
Submitted: 27 September 2021 • Accepted: 06 January 2022 • Published Online: 25 January 2022

Maria Elias Pereira, Jonas Deuermeier, Pedro Freitas, et al.

Tailoring the synaptic properties of a-IGZO memristors for artificial deep neural networks

Cite as: APL Mater. 10, 011113 (2022); doi: 10.1063/5.0073056
Submitted: 27 September 2021 • Accepted: 6 January 2022 •
Published Online: 25 January 2022

Maria Elias Pereira,[1] Jonas Deuermeier,[1] Pedro Freitas,[2] Pedro Barquinha,[1] Weidong Zhang,[2] Rodrigo Martins,[1] Elvira Fortunato,[1] and Asal Kiazadeh[1,a)]

AFFILIATIONS

[1] i3N/CENIMAT, Department of Materials Science, NOVA School of Science and Technology and CEMOP/UNINOVA, NOVA University Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal
[2] Liverpool John Moores University, Faculty of Engineering and Technology, School of Engineering, Liverpool, United Kingdom

Note: This paper is part of the Special Topic on Materials Challenges for Nonvolatile Memory.
[a)]**Author to whom correspondence should be addressed:** a.kiazadeh@fct.unl.pt

ABSTRACT

Neuromorphic computation based on resistive switching devices represents a relevant hardware alternative for artificial deep neural networks. For the highest accuracies on pattern recognition tasks, an analog, linear, and symmetric synaptic weight is essential. Moreover, the resistive switching devices should be integrated with the supporting electronics, such as thin-film transistors (TFTs), to solve crosstalk issues on the crossbar arrays. Here, an a-Indium-gallium-zinc-oxide (IGZO) memristor is proposed, with Mo and Ti/Mo as bottom and top contacts, with forming-free analog switching ability for an upcoming integration on crossbar arrays with a-IGZO TFTs for neuromorphic hardware systems. The development of a TFT compatible fabrication process is accomplished, which results in an a-IGZO memristor with a high stability and low cycle-to-cycle variability. The synaptic behavior through potentiation and depression tests using an identical spiking scheme is presented, and the modulation of the plasticity characteristics by applying non-identical spiking schemes is also demonstrated. The pattern recognition accuracy, using MNIST handwritten digits dataset, reveals a maximum of 91.82% accuracy, which is a promising result for crossbar implementation. The results displayed here reveal the potential of Mo/a-IGZO/Ti/Mo memristors for neuromorphic hardware.

INTRODUCTION

Artificial intelligence (AI) is currently the key feature on innovative technologies for smart systems, pushing for breakthroughs in numerous fields, ranging from healthcare to security solutions. However, the AI requirements of fast processing and low energy consumption are slowing its further development since the currently applied von Neumann's architecture cannot fulfill both demands. Therefore, a new type of compatible hardware is demanded, and neuromorphic computation is an exciting alternative.

Neuromorphic computation allows power-efficient systems with high density information, in-memory computation, and parallel data processing by preserving the massive parallelism observed in the human brain.[1] Endowed with intelligent functions, neuromorphic systems can execute adaptive learning algorithms and perform tasks ranging from real-time big data analysis to pattern recognition. A great potential candidate for this technology is the resistive switching (RS) device or memristor.[2]

The memristor is a non-linear two-terminal device with the capability of being reduced to the nanoscale, whose conductance level depends on present and past external inputs. Inputs are through its two terminals, resembling a biological synapse where neurons convey data through electrical or chemical pulses. In fact, the memristor can simulate a variety of synaptic functions, the most important being plasticity.

In the biological synapse, plasticity implies the reinforcement or impairment of the connection between two neurons by potentiation or depression steps, respectively, in which the update of their synaptic weight takes place. This behavior can be simulated by the memristor in which the synaptic weight is related to its conductance state. When the memristor adapts its resistance level from the low resistance state (LRS) to the high resistance

state (HRS), in the reset process, it is simulating the synaptic weight decrease through depression. Similarly, when the memristor is in the set event, adapting its resistance from the HRS to the LRS, it is mimicking potentiation and increasing the synaptic weight.

RS devices in crossbar arrays have been proposed for deep neural network (DNN) hardware, a high-performance algorithm for classification and recognition applications.[3,4] Common DNNs require iterative multiply–accumulate operations of high-precision weights (>6 bits),[5] and therefore, a linear current change with respect to the repeated training pulses and a symmetric rate at potentiation and depression translates into the most accurate learning. However, most RS devices exhibit a natural non-linear response to consecutive identical spiking. Hence, several approaches have been reported on the pulse scheme modulation that resulted in an improved linearity and symmetry,[6] for instance, applying incremental pulse voltage or width[7] or current pulses[8] or a heating spike before set/reset pulses.[9]

Moreover, to develop an integrated analog controlled weight storage on-chip technology, for a large-scale energy-efficient DNN, the RS devices should be integrated with supporting electronics. An active element such as the transistor solves crosstalk issues on the crossbar arrays[10] that occur due to the interference of neighboring cells.[11] They can also act as an on-chip current compliance to the memristor input and/or compose the additional signal conditioning circuit or support electronics.

The development of a RS device that is eligible for circuit integration and meets all the requirements for DNNs depends on the selection of the RS material and the engineering of top and bottom electrodes since the modulation of the plasticity characteristics can be controlled by the oxygen concentration in oxides and reactivity of the electrodes.[6]

Several materials have been proposed for the RS layer, such as titanium oxide,[8] hafnium dioxide,[12] tantalum oxide,[13] aluminum oxide,[14] and others.[15,16] However, most of the devices are conductive filament (CF)-based memristors. For neuromorphic computing, CF-type memristors imply some disadvantages related to their natural abrupt switching behavior,[17] which typically translates into a digitalized plasticity characteristic, displaying either full potentiation or full depression.[18] The most common solutions involve multiple devices and complex added circuits to represent one synapse or multilevel cell (MLC) devices capable of some discrete conductance levels.[19,20] Nevertheless, these solutions add complexity to the neuromorphic system, and the filament formation and destruction within each device often leads to poor reproducibility.[17]

Therefore, a RS device with analog type of switching, where a gradual conductance change occurs on both set and reset sides, is preferable for the implementation of an integrated system coupled with synaptic functions.[17] It is due to interfacial RS properties that result in a high reproducibility and low cycle-to-cycle (C2C) variation.[21] Amorphous oxide semiconductors (AOSs) have been suggested as active layers for analog memristors[22,23] since the electrical conductivity of the films can be meticulously controlled by the oxygen/cation composition,[24] enabling a dynamic response to input signals.

Indium-gallium-zinc-oxide (IGZO) is a transparent AOS, employed on thin-film transistors (TFTs) in display technology,[25,26] that greatly favors cost-efficiency and a high integration density on a single IC without the need of interfacing with diverse technologies (i.e., CMOS devices). The a-IGZO-based RS device is the ideal candidate for the memristor integration with TFTs, and in fact, it has been previously reported in a single memristor,[27–31] including a study by our group,[32] and also integrated with TFTs.[33–35] However, a fully integrated circuit has not been demonstrated where both transistor semiconductor and memristor RS layer share the one and same processing step as well as the electrode materials, which would imply a significant decrease in the total lithography mask count, improved interconnectivity, and drastic cost reduction. The reason behind this gap is the fact that TFTs should be optimized for a high stability and low leakage and the RS device should be optimized for a defect-enabled switching ability with high on/off ratio, which results in a contradictory film optimization where compromises must be made. In fact, most of the studies on IGZO memristors that focus on artificial neural network applications show that a simulation of pattern recognition accuracies either does not report on the crosstalk effect[30–36] or use different IGZO layers and/or electrodes for the TFT.[34]

In our previous work, we reported on the a-IGZO-based memristor with Mo as both top and bottom electrodes. In these devices, the conductivity state is controlled by a potential barrier and the switching entails adjustments on the barrier height. The a-IGZO memristor has a forming-free and area-dependent performance and enables analog control of resistance states.[32]

Here, we demonstrate the modulation of the a-IGZO-based plasticity characteristics by applying non-identical spiking schemes. The pattern recognition accuracy, using the MNIST handwritten digits dataset, was tested and revealed a maximum of 91.8% accuracy using consecutive pulses with a linear increase in voltage amplitude and width, which presents itself as very promising. Moreover, we optimize the memristor a-IGZO composition for a lower oxygen content for compatibility with TFT semiconductors, aiming for a forthcoming integration with the a-IGZO-based TFT with Mo as gate, source, and drain electrodes.[37] The results displayed here reveal the potential of Mo/a-IGZO/Ti/Mo RS devices for crossbar and/or TFT integration for neuromorphic hardware systems.

EXPERIMENTAL

The RS devices were fabricated on glass substrates, formerly cleaned in repeated ultrasonic baths of acetone and isopropanol and rinsed with deionized water and dry nitrogen. In Figs. 1(a) and 1(b), a schematic of one RS device in a cross-point structure and a micrograph of a 4 μm^2 device are presented, respectively.

In Fig. 1(c), the material structure is displayed—Mo/a-IGZO/Ti/Mo. For the bottom and top electrodes, radio-frequency (RF) magnetron sputtering was used to deposit a 70 nm thick Mo layer in an AJA ATC-1800 system with a flow rate of 50 SCCM of Ar, a sputtering power of 175 W, and a deposition pressure of 1.7 mTorr. For the active layer, IGZO thin films were deposited by RF magnetron co-sputtering from three binary ceramic oxide targets. The sputtering powers used on each target were In_2O_3: 121 W, Ga_2O_3: 100 W, and ZnO: 50 W.

Since this study has envisioned an upcoming memristor integration with TFTs, two different conditions were used regarding the oxygen content during the IGZO film deposition: one that we will

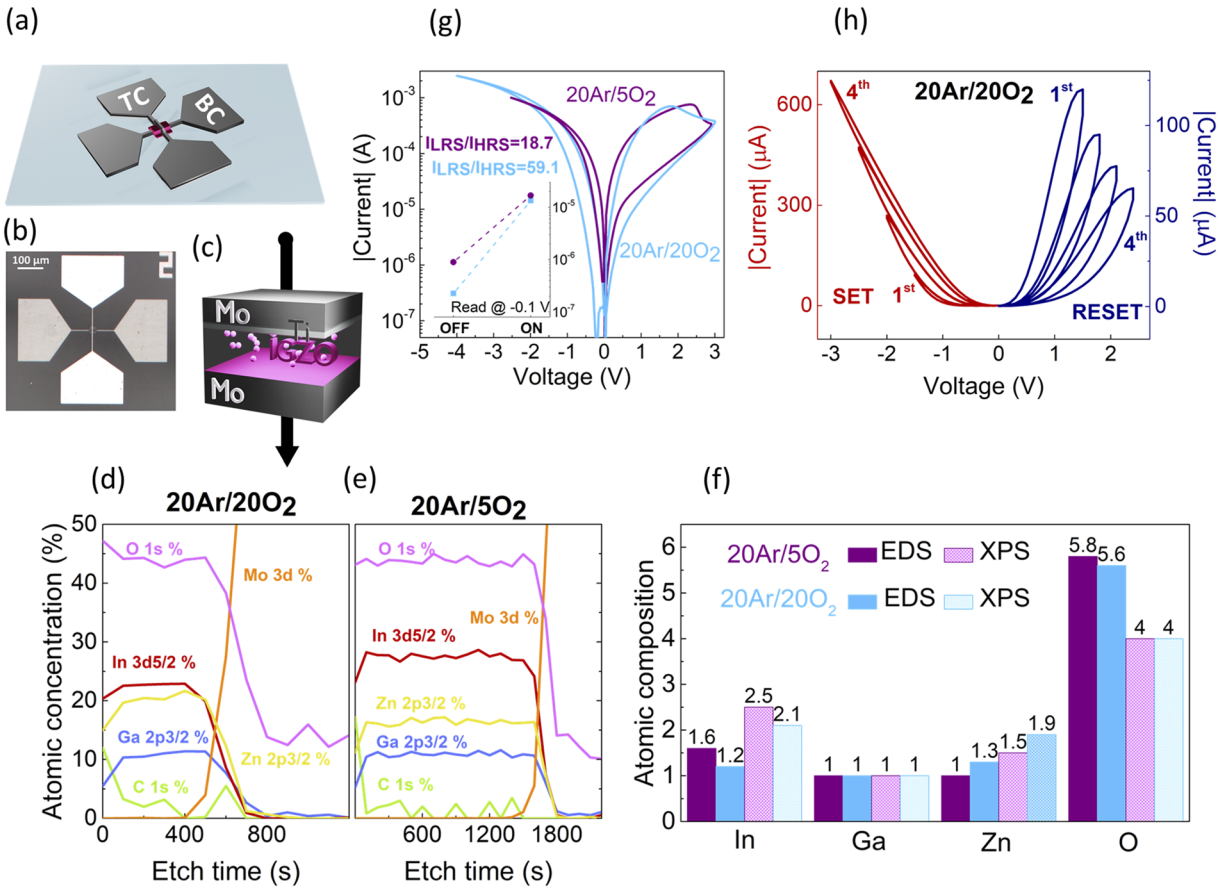

FIG. 1. (a) Schematic illustration of the cross-point structure of the memristors with the bottom contact (BC) and top contact (TC) hinted; (b) micrograph of one 4 μm^2 device; (c) schematic illustration of the Mo/a-IGZO/Ti/Mo material structure; and XPS argon cluster depth profiles of (d) 20Ar/20O_2 and (e) 20Ar/5O_2 films without the top contact. The relative atomic concentrations are displayed with respect to etch time. (f) Atomic composition of In, Ga, Zn, and O of both IGZO films analyzed by EDS and XPS for comparison; (g) I–V characteristic of one set and one reset for comparison of both devices—20Ar/5O_2 and 20Ar/20O_2—with the inset of read current at −0.1 V after each set and reset displaying on/off ratios of 59.1 and 18.7, respectively; and (h) I–V characteristic displaying analog behavior by a gradual increase of voltage sweep span for set and reset. The order of the measurements is displayed highlighting the first (1st) sweep and the last (4th) for both set and reset.

refer to as 20Ar/5O_2, which used a flow rate of 20 SCCM of Ar and 5 of O_2, being the most similar with the IGZO deposition for TFTs[38] and another, for comparison, 20Ar/20O_2, which used a flow rate of 20 SCCM of Ar and 20 of O_2, as previously reported.[32] In Fig. S1 of the supplementary material, the transfer curves of TFTs fabricated using both IGZO films are presented. The IGZO deposition pressure was kept constant at 2.3 mTorr. The thicknesses were 40 nm in the case of 20Ar/20O_2 and 65 nm for 20Ar/5O_2, confirmed by a profilometer. E-beam evaporation in a homemade apparatus was used to deposit a thin 6 nm layer of Ti between the IGZO and Mo top layers.

The patterning of the bottom Mo electrode was achieved by reactive ion etching in a Trion Phantom 3 system, using SF_6, whereas the IGZO active layer and Ti/Mo top electrode patterning was done via lift-off. An annealing step of 150 °C was performed on a hot plate for 60 min after the IGZO patterning and again after the top electrode patterning, to better simulate the co-fabrication of TFTs that require annealing for improved stability and performance.[39]

All the electrical characterization of the devices was done using a Keithley 4200 SCS semiconductor analyzer connected to a Janis ST-500 probe station. The DC sweeps and the pulses were applied to the top electrode, while maintaining the bottom electrode connected to ground.

The XPS argon cluster depth profiles were performed with a Kratos Axis Supra, using a monochromatic Al Ka source running at 300 W. The analysis area was limited to 110 μm in diameter by an aperture, and the analyzer was set to a pass energy of 80 eV. Argon clusters of ~500 atoms with a kinetic energy of 10 keV were employed for etching with a step duration of 100 s. CasaXPS Version 2.3.19PR1.0 was used for data analysis. Further energy dispersive spectroscopy (EDS) analysis was undertaken, using a Carl Zeiss AURIGA CrossBeam FIB-SEM workstation, to confirm the XPS results.

RESULTS AND DISCUSSION

In Figs. 1(d) and 1(e), the results from the XPS argon cluster depth profile are shown. In order to maximize the signal from the a-IGZO and to obtain a clear analysis of the bottom contact interface, the IGZO films have been measured without the Mo/Ti top contact.

The bottom interface is critical for the conduction since it represents the barrier for electron injection due to an oxidation of the Mo at the interface.[32] The modulation of the Schottky barrier controls the resistive switching.[22,40] Concerning the top electrode, a thin Ti layer was added due to its well-known oxygen getter effect.[41] Ti reacts with the switching oxide by extracting oxygen ions, which increases the donor concentration and, therefore, boosts the conductance of the Ohmic contact and creates a highly conductive interface region in a-IGZO memristors.[42] Contrasting with the Schottky barrier at the Mo bottom electrode, an asymmetry is built in the oxide, which results in a more pronounced non-linear profile and memristors with higher on/off ratios.[43]

The difference in IGZO thickness due to the different oxygen amounts in the process gas affects the etch times until the Mo bottom contact is reached, being higher for the $20Ar/20O_2$ sample. Moreover, the $20Ar/5O_2$ sample is likely more resistive to the etching due to a higher film density caused by the increased kinetic energy of ions in the process gas, compared to $20Ar/20O_2$.[44,45] An important observation is that the oxygen amount is identical in the two samples, around 44 at. %, despite the different levels of oxygen flow. However, the cation ratio of IGZO is strongly altered. The sample $20Ar/20O_2$ has an average bulk In:Ga:Zn atomic composition of 2.1:1.0:1.9, whereas $20Ar/5O_2$ has a composition of 2.5:1.0:1.5. The ratio was normalized to the Ga concentration since both samples have an identical Ga concentration of 11 at. %. It can be excluded that this difference is caused by damage induced by the argon cluster beam because the trend is the same for the films prior to the etching. In a previous study, the In/Ga ratio had been found to be critical for the a-IGZO resistivity. More In content leads to a more conductive material.[24]

To confirm these results, EDS was performed on the films deposited on silicon. The results are shown in Fig. 1(f), where the comparison of the atomic composition obtained by XPS and EDS techniques is shown. EDS shows the same trend in the cation ratios as it was revealed by XPS cluster depth profiling. The $20Ar/20O_2$ film presents less In and more Zn, which is in accordance with previous reports.[46] This can be explained by an increased bombardment of O_2 on the film itself while growing. During the deposition, O_2 gas is inserted in the sputtering system near the substrate, while Ar gas is inserted near the target. The additional oxygen bombardment can lead to the removal of the material from the film, which has resulted in a lower density in previous studies.[47] The elements with the weaker bonds in the film will be removed preferentially, such as In compared to Ga.

The oxygen content obtained with both techniques is given in Fig. 1(f). The absolute amounts differ between techniques, but the identical oxygen content in the two samples is confirmed. The fact that the oxygen amount in the films remains unaltered can infer that with the smallest oxygen flow (5 SCCM), there is already a saturation of this element in the film since oxide targets are being used. In addition, it needs to be mentioned that the oxygen amount at the surface, determined from XPS prior to argon cluster etching, is indeed higher in the $20Ar/20O_2$ sample, about 46% compared to 44%. However, this does not represent the bulk oxygen content, which is confirmed by the EDS results. Since the devices are controlled by the barrier properties at the bottom contact, the surface of the IGZO is of minor relevance for the resistive switching.

In Fig. 1(g), the comparison between the set and reset of $20Ar/5O_2$ and $20Ar/20O_2$ is shown and in the inset; the read current after each measurement at −0.1 V is provided. The 4 μm^2 devices work in the bipolar switching mode; set occurs in negative polarity, while reset happens under positive polarity. The $20Ar/20O_2$ device has a higher on/off ratio compared to the $20Ar/5O_2$ device due to the lower conductivity of the semiconductor in the pristine state and the increased Schottky barrier height. The impact of the In/Ga ratio on conductivity in IGZO is well known.[24] In IGZO Schottky diodes, the barrier height has been found to be sensitive to the In/Ga ratio:[48] lower In/Ga ratios lead to higher barriers. This is explained by the strong Ga–O bond (compared to In–O and Zn–O), which suppresses oxygen vacancies.[49] Note that the depletion width is also affected by the In/Ga ratio, being smaller for higher values of In/Ga.[48] The principle behind the resistive switching is a change of Schottky barrier profile,[32] which is in agreement with other reports with rectifying and analog resistive switching properties.[40]

The typical analog behavior of a $20Ar/20O_2$ device is shown in Fig. 1(h), where successive voltage DC sweeps on negative and positive polarities were applied to the top contact, which resulted in the device conductance being gradually increased and decreased, respectively. To demonstrate a gradual set process, four voltage sweeps were carried out starting from 0 to −1.5 V during the first measurement until 0 to −3 V for the last measurement. The results are displayed in red in Fig. 1(h), where the gradual increase of the device conductance state is clear. For the reset process, the same methodology was used for positive polarities with the first voltage sweep being from 0 to 1.5 V and the last sweep being from 0 to 2.4 V, displayed in dark blue in Fig. 1(h). In this case, there is a gradual decrease of the conductance state. All measurements were performed with a voltage step of 0.1 V, the speed was set at the normal mode, and the integration time was in auto setting. Both hold and delay times were zero.

Several devices with different areas were fabricated, and a micrograph of a representative part of a sample is shown in Fig. S2(a) of the supplementary material. Generally, in AOS materials, the switching mechanism is area-dependent [Figs. S2(b) and S2(c) of the supplementary material]. All the results presented throughout this paper are with respect to 4 μm^2 devices.

To simulate the synapse plasticity, through potentiation and depression tests, the device characteristics were studied using an identical pulse train. Here, potentiation implies facilitating the connection between two neurons, and it is simulated by applying negative pulses to the top contact and increasing the current state of the device. Depression, which is constraining the neuron's connection, is replicated by applying positive pulses to the top contact and, therefore, decreasing the device current state.

An optimal condition of pulse amplitude and width should be chosen for a gradual current state increase/decrease. In Figs. 2(a) and 2(b), the 100 pulse potentiation tests for the $20Ar/5O_2$ memristor with different pulse amplitudes for the same width and different pulse widths for the same amplitude are presented, respectively. The read step is always performed immediately after each pulse and

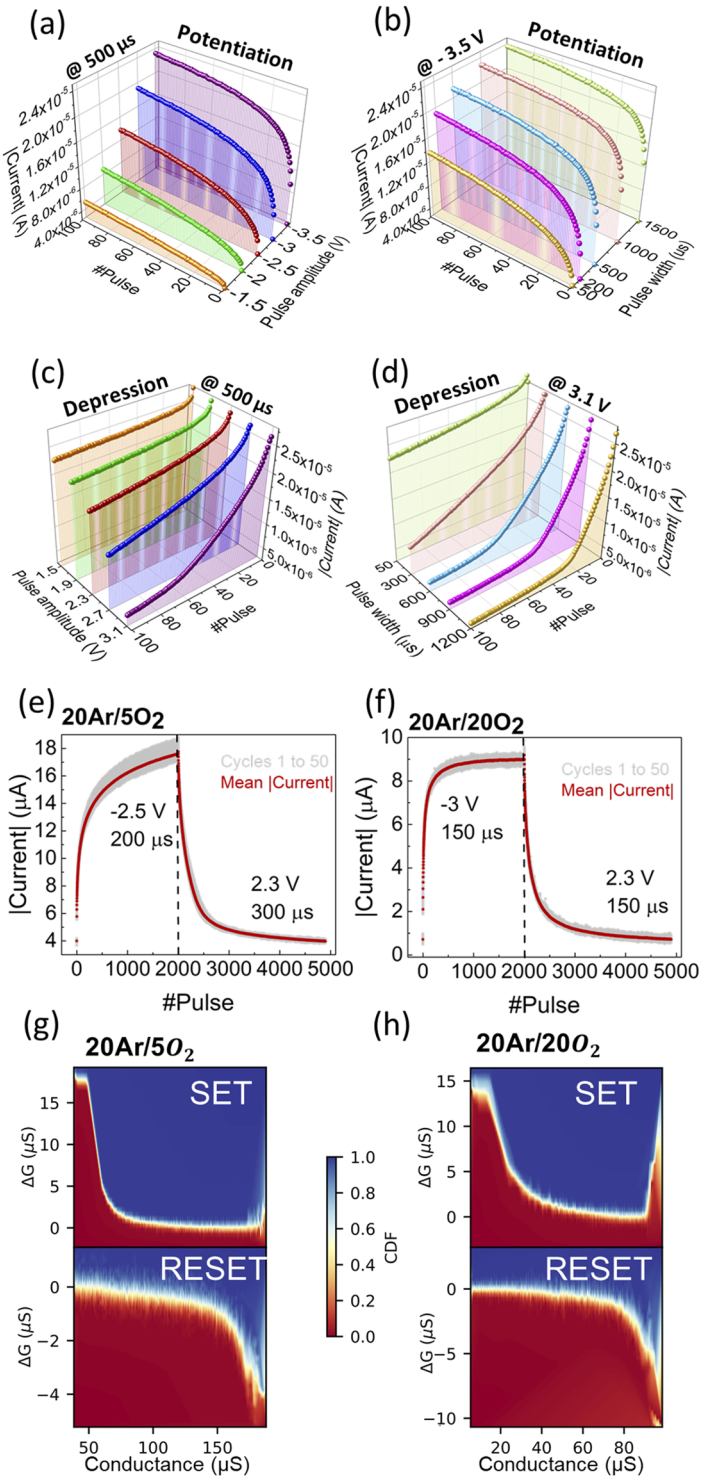

FIG. 2. 100 identical pulse potentiation tests, for the memristor with low oxygen ($20Ar/5O_2$), with (a) different amplitudes, same width and (b) different widths, same amplitude; and depression tests with identical pulses of (c) different amplitudes, same width and (d) different widths, same amplitude. Read current at −0.1 V; 50 cycles of 5000 pulses of potentiation and depression with the mean current in red for (e) the memristor with low oxygen ($20Ar/5O_2$) and (f) the memristor with high oxygen ($20Ar/20O_2$); and experimental Cumulative Distribution Function (eCDF) lookup tables of the conductance change (ΔG) within the 50 cycles of potentiation/depression as a function of conductance (G) for the set and reset processes for the (g) $20Ar/5O_2$ and (h) $20Ar/20O_2$ devices.

at −0.1 V. All the conditions result in a non-linear curve and the difference lies on the current state achieved, which should be the maximum possible for a high on/off ratio. Evidently, the higher the pulse amplitude and width, the higher the current reached but also the faster the current state gets saturated. For a gradual increase, a pulse scheme in which the current state does not reach a saturated value is preferable as it is for lower pulse amplitudes/widths. One can also conclude that the pulse amplitude is the principal parameter that controls the ON state reached, while the pulse width controls the gradualness of the increase.

The same tests for depression can be found in Figs. 2(c) and 2(d) for the $20Ar/5O_2$ memristor, where 100 positive pulses with different pulse amplitudes for the same width and different pulse widths for the same amplitude are shown, respectively. Here, the gradual decrease of the current is desirable as it is, for example, on 3.1 V at 300 μs pulses. Therefore, one can conclude that by controlling the pulse width, small adjustments to the current state are accomplished, while for larger adjustments, the pulse amplitude should be tuned. The identical pulse training study for $20Ar/20O_2$ devices can be found in Fig. S3 of the supplementary material.

To evaluate C2C variability, 50 cycles of potentiation/depression tests were carried out using 5000 pulses. Considering the previous conclusions and the high number of pulses for this test, the programmed pulse amplitudes and widths chosen for $20Ar/5O_2$ device were −2.5 V at 200 μs for potentiation and 2.3 V at 300 μs for depression, as can be seen in Fig. 2(e). For $20Ar/20O_2$ devices, consecutive pulses of −3 V at 150 μs for potentiation and of 2.3 V at 150 μs for depression were applied, as shown in Fig. 2(f). In Figs. 2(g) and 2(h), the experimental Cumulative Distribution Function (eCDF) lookup tables of the conductance change (ΔG) within the 50 cycles of potentiation/depression as a function of conductance (G) for the set and reset processes are presented for the $20Ar/5O_2$ and $20Ar/20O_2$ devices, respectively. Here, the colormap of the CDF indicates the probability that ΔG is less than or equal to the value on the y axis for a given conductance on the x axis. For the lowest C2C variability, a straight line with ΔG = 0 should be presented with red below zero and blue above zero with this case colormap. C2C variation is more pronounced during the set operation of the $20Ar/5O_2$ device since the current state is not yet saturated. When comparing both devices, one can see that the LRS of $20Ar/5O_2$ is higher than that of the $20Ar/20O_2$ devices. On the other hand, the HRS of the $20Ar/20O_2$ device is lower than that of the $20Ar/5O_2$ device, providing a higher on/off ratio.

However, an ideal artificial synapse should present a linear and symmetric current increase/decrease under consecutive applied pulses for a maximum accuracy in pattern recognition.[6] The linearity condition comes from the fact that the device should be able to achieve as many distinct conductance states as possible according to the high weight precision requirement, and the symmetry between set and reset enables the implementation of the fastest and most efficient programming method of the memristor arrays. In the $20Ar/5O_2$ device, the current state does not saturate so prominently as in the $20Ar/20O_2$ memristor. That has to do with the pulse scheme parameters that were chosen for the highest on/off ratio possible.

Different pulse schemes were tried to modulate the potentiation/depression characteristics to improve linearity and symmetry and are shown in Fig. 3(a). Scheme A consists of identical pulses. Pulse schemes B and C are non-identical spikes with either increasing pulse amplitude or width, respectively. In pulse scheme D, a consistent increase in pulse amplitude and width in each pulse is applied. For all conditions, the read process is a pulse of 20 μs at −0.1 V.

In Fig. 3(b), potentiation and depression results for the $20Ar/5O_2$ memristor are presented for each pulse scheme. Scheme A, the purple curve, shows an exponential response where a rapid current increase/decrease takes place within a few pulses and then reaches a saturated current regime, as previously discussed. Scheme B, the red curve, provides a slower current increase/decrease during the first few pulses, and it becomes a linear change after −1.9 V for potentiation and 2.5 V for depression. On the other hand, scheme C, the blue curve, shows a current saturation after a 800 μs pulse for potentiation and 890 μs pulse for depression.

For scheme D, a combination of both B and C schemes was tested. The pulse amplitude was linearly increased from −1.9 to −4 V for potentiation and 2.5–3 V for depression, as scheme B suggested. On depression, the pulse width was also linearly increased according to the results of scheme C, from 1 to 890 μs. For potentiation, since −1.9 V was the starting pulse amplitude, a 1 μs pulse would not be sufficient for a linear current increase at the first few pulses, and therefore, a shift on the width interval was implemented. Instead of using a pulse width interval from 1 to 800 μs as pulse scheme C suggested, an interval of 400–1200 was used, maintaining the difference between first and final pulses width applied. The results correspond to the green curve where a linear and symmetrical response to the input pulses from beginning to end is presented. It is noted that the proposed pulse scheme D is rather complex, involving an increase in pulse amplitude and width in each pulse. Nevertheless, in order to deal with non-linearity issues in an efficient pulse scheme, design of 1T1R is proposed,[50] *which* is effectively implantable with the IGZO condition, as mentioned in the current report for both transistor and memristor devices.

To compare the linearity and symmetry on the synaptic plasticity tests that resulted from each pulse scheme, a non-linearity parameter α was extracted,[51] using the following equation:

$$G = \begin{cases} \left(\left(G_{LRS}^{\alpha} - G_{HRS}^{\alpha}\right) \times w + G_{HRS}^{\alpha}\right)^{1/\alpha} & if \quad \alpha \neq 0, \\ G_{HRS} \times \left(G_{LRS}/G_{HRS}\right)^{w} & if \quad \alpha = 0, \end{cases} \tag{1}$$

where G is the conductance change, G_{LRS} and G_{HRS} are low resistance state (LRS) and high resistance state (HRS) conductances, respectively, and w is an internal variable of the synaptic weight, which varies from 0 to 1; w increases or decreases as potentiation or depression pulses are applied. α is the non-linearity parameter that controls potentiation (α_p) or depression (α_d) characteristics. The closer α is to 1, the more linear is potentiation/depression. Furthermore, the smaller the difference between the values of α for potentiation and depression, the more improved is the symmetric response.

Figure 3(c) shows the values of α_p and α_d for all schemes. As it can be seen, scheme D has α_p = 1.23 and α_d = 0.83, numerically revealing closeness to optimal linear and symmetric synaptic characteristics. The identical study for the $20Ar/20O_2$ memristor is presented in Fig. S4 of the supplementary material.

Using pulse scheme D, 50 cycles of potentiation and depression were carried out on both $20Ar/5O_2$ device and $20Ar/20O_2$ memristor, shown in Figs. 3(d) and 3(e), respectively, to evaluate

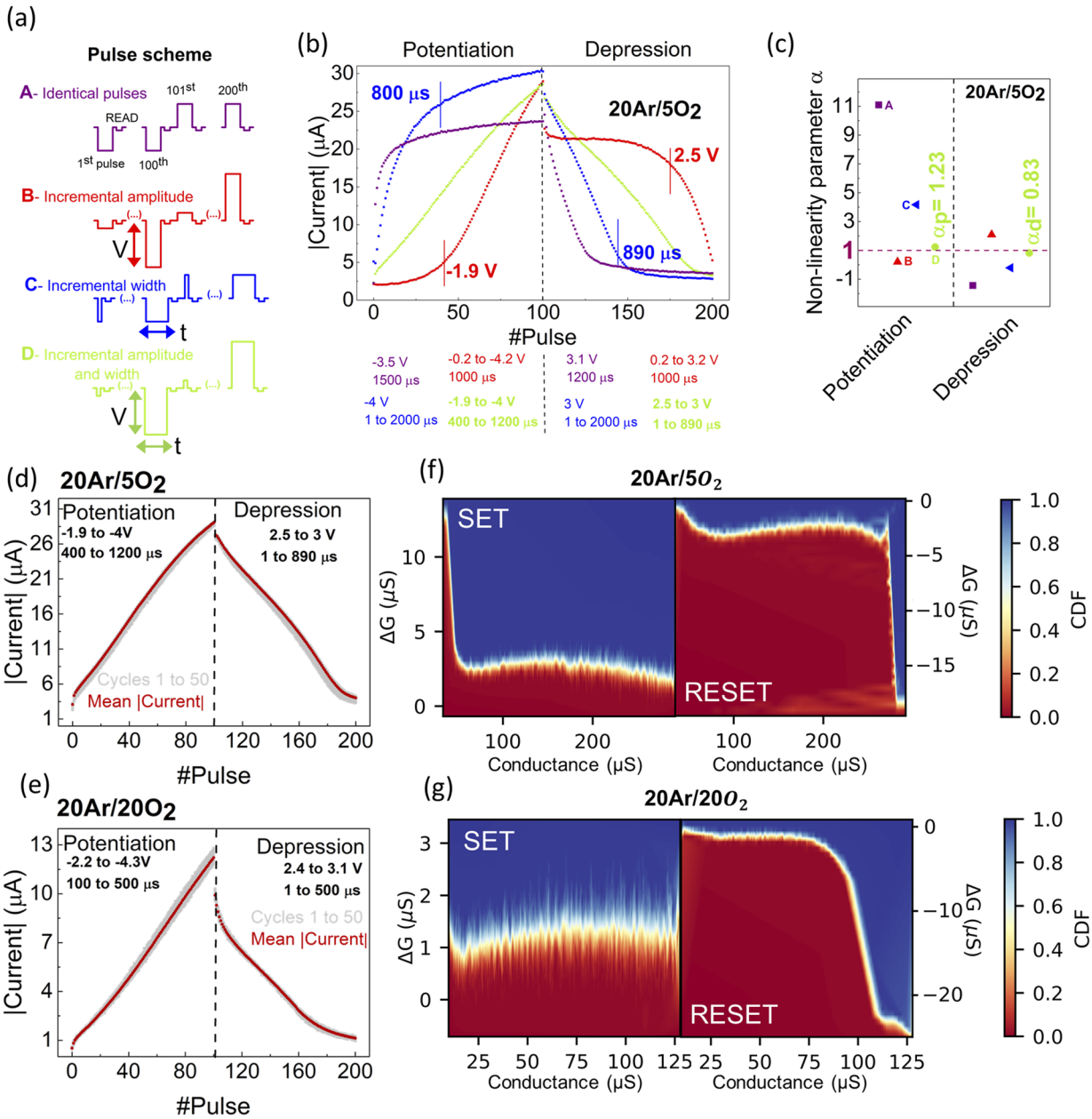

FIG. 3. (a) Schematic illustration of pulse schemes A (identical pulses), B (incremental amplitude pulse), C (incremental width pulses), and D (incremental amplitude and width pulse); (b) potentiation and depression results for $20Ar/5O_2$ of the different pulse schemes with the numerical description of each scheme at the bottom of the graph. The conditions that limit linearity are within the graph; (c) non-linearity parameter for potentiation and depression for each pulse scheme displayed on (b), with α_p = 1.23 and α_d = 0.83 of pulse scheme D highlighted; (d) 50 cycles of potentiation and depression for the linear pulse scheme D of the $20Ar/5O_2$ memristor and (e) optimized pulse scheme for the $20Ar/20O_2$ memristor. Experimental Cumulative Distribution Function (eCDF) lookup tables of the conductance change (ΔG) within the 50 cycles of potentiation/depression as a function of G for the set and reset processes for the (f) $20Ar/5O_2$ and (g) $20Ar/20O_2$ devices. Read current is −0.1 V for all the tests.

the linearity stability under incremental amplitude/with pulses. It is important to note that several devices were tested for the modulation of the pulse scheme and behave identically within their category, $20Ar/5O_2$ and $20Ar/20O_2$, as it can be evaluated in Fig. S5 of the supplementary material, where potentiation and depression tests under the same non-identical pulse schemes for five different devices are presented, proving the reproducibility of the fabrication process and the reliability of the proposed scheme for a linear and symmetric synaptic characteristic of the Mo/a-IGZO/Ti/Mo structure.

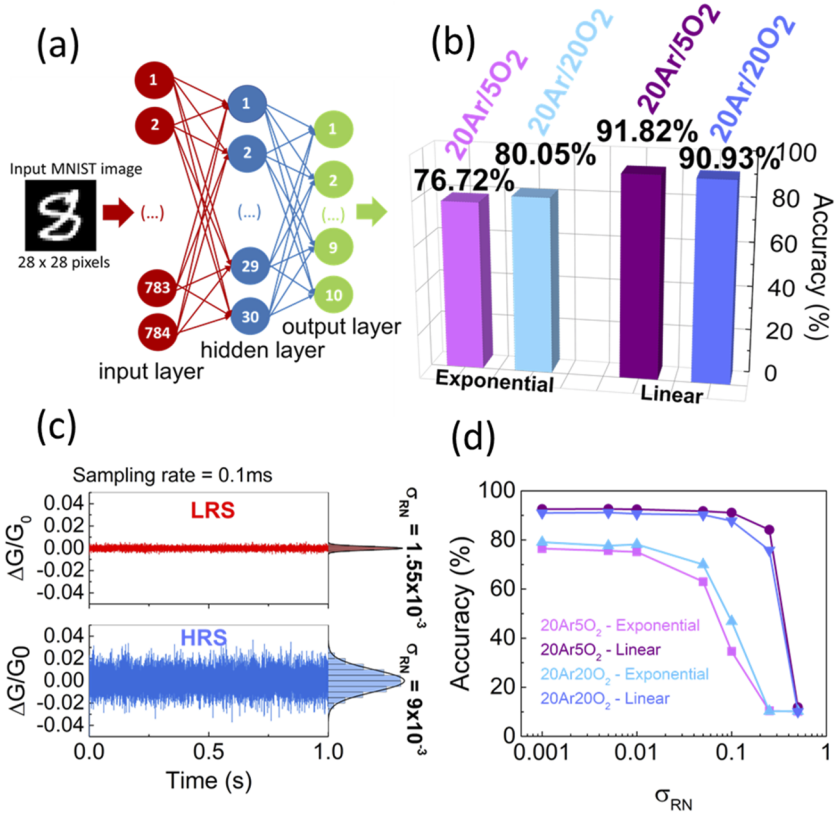

FIG. 4. (a) Schematic illustration of a three-layer perceptron neural network (the number of neurons used on the input, hidden, and output layers is 784, 30, and 10, respectively); (b) accuracies for pattern recognition achieved for the exponential natural response on potentiation and depression using identical spiking [Figs. 2(e) and 2(f)] for both memristors [20Ar/5O_2 (76.72%) and 20Ar/20O_2 (80.05%)] and for the linear response achieved by pulse scheme D with incremental pulse amplitude and width [Figs. 3(d) and 3(e)], also for both memristors [20Ar/5O_2 (91.82%) and 20Ar/20O_2 (90.93%)]; (c) normalized low-frequency read noise (σ_{RN}) measured for 1 s, using a sampling rate of 0.1 ms/sample for the HRS and LRS; and (d) impact of σ_{RN} on the training accuracy of both devices (20Ar/5O_2 and 20Ar/20O_2).

In Figs. 3(f) and 3(g), where the eCDF lookup tables of ΔG within the 50 cycles as a function of G for set and reset of both types of devices are presented; one can confer that the higher variability during cycles is presented for the 20Ar/20O_2 devices. Nevertheless, it can be stated that both memristors have low variability between C2C, which proves the potential of the analog a-IGZO-based RS device. The on/off ratio is higher for 20Ar/20O_2 devices; nonetheless, an on/off ratio of more than 10 might be enough for a highly accurate pattern recognition[52] in a 1T1R active crossbar, considering that crosstalk would be minimized.

To assess the effect of the synaptic characteristics on pattern recognition accuracy, a simulation of a three-layer (input, hidden, and output layers) perceptron neural network was performed by CrossSim[53] using a handwritten digit dataset (MNIST). The number of neurons on the input, hidden, and output layers is 784, 30, and 10, respectively, as shown in Fig. 4(a). Each neuron on the input layer is connected to all neurons in the hidden layer through synapses and each neuron on the hidden layer is connected to all output neurons, and they communicate by conveying the conductance states. Figure 4(b) presents the pattern recognition accuracies obtained with the linear response shown in Figs. 3(d) and 3(e) and the exponential response displayed in Figs. 2(e) and 2(f) for the 20Ar/5O_2 and the 20Ar/20O_2 memristors, respectively.

The exponential response reaches higher accuracies than expected, given the non-linearity presented (79.72% and 80.05%), due to the high number of pulses applied (5000 pulses) and low C2C variability. However, the linear response is the one with the excellent performance, disclosing accuracies of 91.82% and 90.93% for the 20Ar/5O_2 and 20Ar/20O_2 memristors, respectively. For the 20Ar/5O_2 device, the default test provided by CrossSim that uses the 25%–75% range of the on/off window was performed. For the 20Ar/20O_2 device, since the depression test was not as linear at the end of the test, an adjustment was made for the 10%–60% range instead (Fig. S6 of the supplementary material).

These results reveal that with at least 6 bit precision (default for CrossSim) and disregarding any read noise induced errors, an on/off ratio of 1 order of magnitude does not pose as a drawback in achieving high accuracies, and instead, C2C variability and linearity play a much more important role.

The impact of nonlinearity, asymmetry, cycle-to-cycle variability, and noise can be accurately considered in the CrossSim.[53] Since the dynamic range of devices is different and for the 20Ar/5O_2 device is smaller (i.e., on/off ratio > 10), additionally, the normalized low-frequency read noise (σ_{RN}) was measured for 1 s, using a sampling rate of 0.1 ms/sample [Fig. 4(c)], where σ_{RN} is the standard deviation of ΔG/G0.

The impact of σ_{RN} on the training accuracy was then evaluated by injecting noise into each weight's simulated conductance state, each time a read operation occurs, using the following equation:

$$G = G_0 + N(\sigma), \quad \sigma = \sigma_{RN} \times G_{Range}. \tag{2}$$

Figure 4(d) shows the impact of σ_{RN} on the training accuracy of both devices ($20Ar/5O_2$ and $20Ar/20O_2$) following both of their programming methods (exponential and linear responses). A noise threshold of $\sigma_{RN} = 0.01$ and $\sigma_{RN} = 0.1$ exists for the exponential and linear responses, respectively, where the training accuracy starts to degrade. However, the highest measured σ_{RN} for these devices falls below the lowest threshold, indicating that the impact of read noise on training of both devices can be considered negligible.

CONCLUSIONS

In conclusion, amorphous IGZO-based memristors with Mo as the bottom contact and Ti/Mo as the top contact were fabricated. The IGZO film was optimized for a lower oxygen content, and annealing steps were integrated on the fabrication processes for compatibility with the TFT semiconductor aiming for a forthcoming integration with the IGZO-based TFT with Mo as gate, source, and drain electrodes.

Our proposed methodology is proven to be reproducible and reliable and uses conventional patterning strategies and no noble-metals, being a cost-effective approach, compatible with room temperature-based processes, allowing for the implementation of transparent and flexible substrates, such as polymer and paper, which is a core feature for IoT applications.

The devices have an electroforming-free and an area-dependent analog resistive switching, which are very appealing properties for neuromorphic hardware applications. The modulation of the a-IGZO memristor plasticity characteristics was achieved by applying non-identical spiking schemes and a linear and symmetrical potentiation and depression was obtained, proven by the calculated non-linearity parameter. The pattern recognition accuracy, using the MNIST handwritten digits dataset, was tested and revealed a maximum of 91.82% accuracy using consecutive pulses with incremental voltage amplitude and width, which presents itself as a very promising result for neuromorphic systems.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional results on the TFT transfer curves using both IGZO compositions and the I–V sweeps for set and reset for various memristor areas. The experimental data on $20Ar/20O_2$ devices, namely, on the pulse scheme modulation for both exponential and linear responses, and details on the CrossSim simulation are also shown. Device-to-device variability for five different devices is presented for both IGZO conditions.

ACKNOWLEDGMENTS

This research was funded by FEDER funds through the COMPETE 2020 Programme and National Funds through the FCT—the Portuguese Foundation for Science and Technology—under the scope of the Project No. UIDB/50025/2020-202, doctoral Grant No. DFA/BD/8335/2020, and the project "NeurOxide" Reference No. PTDC/NAN-MAT/30812/2017. We would also like to thank Dr. Ana Rovisco for her support with the EDS measurements. This work also received funding from the European Community's H2020 program [Grant Agreement Nos. 716510 (ERC-2016-StG TREND), 787410 (ERC-2019-AdG DIGISMART), 952169 (SYNERGY, H2020-WIDESPREAD-2020-5, CSA), and 101008701 (EMERGE, H2020-INFRAIA-2020-1)] and the EPSRC of U.K. (Grant Nos. EP/M006727/1 and EP/S000259/1).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts of interest to disclose.

Author Contributions

M.E.P., P.F., and A.K. contributed to conceptualization; M.E.P. contributed to device fabrication; M.E.P. and A.K. contributed to methodology; M.E.P. and A.K. performed validation; M.E.P., J.D., P.F., and A.K. performed formal analysis; M.E.P. and A.K. wrote the original draft; M.E.P, J.D., P.F., P.B., W.Z., R.M., E.F., and A.K. contributed to writing—review and editing; A.K. contributed to supervision; and A.K., W.Z, R.M., and E.F contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, Nat. Rev. Phys. **2**, 499 (2020).
[2] D. Ielmini, Z. Wang, and Y. Liu, APL Mater. **9**, 050702 (2021).
[3] S. H. Tan, P. Lin, H. Yeon, S. Choi, Y. Park, and J. Kim, APL Mater. **6**, 120901 (2018).
[4] R. Hasan, T. M. Taha, and C. Yakopcic, in *International Joint Conference on Neural Networks* (IEEE, 2017), pp. 3527–3534.
[5] S. Yu, Z. Li, P. Y. Chen, H. Wu, B. Gao, D. Wang, W. Wu, and H. Qian, in *IEEE International Electron Devices Meeting* (IEEE, 2017), pp. 16.2.1–16.2.4.
[6] K. Moon, S. Lim, J. Park, C. Sung, S. Oh, J. Woo, J. Lee, and H. Hwang, Faraday Discuss. **213**, 421 (2019).
[7] S. Oh, T. Kim, M. Kwak, J. Song, J. Woo, S. Jeon, I. K. Yoo, and H. Hwang, IEEE Electron Device Lett. **38**, 732 (2017).
[8] J. Park, M. Kwak, K. Moon, J. Woo, D. Lee, and H. Hwang, IEEE Electron Device Lett. **37**, 1559 (2016).
[9] Y. J. Jeong, S. Kim, and W. D. Lu, Appl. Phys. Lett. **107**, 173105 (2015).
[10] A. Chen, IEEE Trans. Electron Devices **62**, 2857 (2015).
[11] A. Kiazadeh and J. Deuermeier, *Advances in Non-Volatile Memory and Storage Technology*, 2nd ed. (Elsevier, 2019), pp. 519–538.
[12] F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M. Von Witzleben, U. Böttger, R. Waser, and S. Hoffmann-Eifert, APL Mater. **7**, 091105 (2019).
[13] C. Sung, S. Lim, H. Kim, T. Kim, K. Moon, J. Song, J.-J. Kim, and H. Hwang, Nanotechnology **29**, 115203 (2018).
[14] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, and H. Hwang, IEEE Electron Device Lett. **37**, 994 (2016).
[15] H. J. Kim, M. Kim, K. Beom, H. Lee, C. J. Kang, and T.-S. Yoon, APL Mater. **7**, 071113 (2019).
[16] K. Sun, J. Chen, and X. Yan, Adv. Funct. Mater. **31**, 2006773 (2021).
[17] Y. Xi, B. Gao, J. Tang, A. Chen, M.-F. Chang, X. S. Hu, J. V. D. Spiegel, H. Qian, and H. Wu, Proc. IEEE **109**, 14–42 (2021).
[18] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, Sci. Rep. **7**, 5288 (2017).
[19] L. Song, X. Qian, H. Li, and Y. Chen, in *IEEE International Symposium on High Performance Computer Architecture* (IEEE, 2017), pp. 541–552.

[20]C. Yakopcic, M. Z. Alom, and T. M. Taha, in *International Joint Conference on Neural Networks* (IEEE, 2016), pp. 963–970.
[21]W. Zhang, B. Gao, J. Tang, X. Li, W. Wu, H. Qian, and H. Wu, Phys. Status Solidi RRL **13**, 1900204 (2019).
[22]C. Silva, J. Martins, J. Deuermeier, M. E. Pereira, A. Rovisco, P. Barquinha, J. Goes, R. Martins, E. Fortunato, and A. Kiazadeh, Electron. Mater. **2**, 105 (2021).
[23]M. Kimura, R. Sumida, A. Kurasaki, T. Imai, Y. Takishita, and Y. Nakashima, Sci. Rep. **11**, 580 (2021).
[24]A. Olziersky, P. Barquinha, A. Vilà, C. Magaña, E. Fortunato, J. R. Morante, and R. Martins, Mater. Chem. Phys. **131**, 512 (2011).
[25]A. Santos, B. Tiwari, J. Martins, A. Santa, K. Chapagai, P. Bahubalindruni, and P. Barquinha, in *2018 International Flexible Electronics Technology Conference, IFETC* (IEEE, 2018), pp. 1–6.
[26]H. Hosono, Nat. Electron. **1**, 428 (2018).
[27]B. Dang, K. Liu, J. Zhu, L. Xu, T. Zhang, C. Cheng, H. Wang, Y. Yang, Y. Hao, and R. Huang, APL Mater. **7**, 071114 (2019).
[28]V. Dumitru, C. Besleaga, and O. N. Ionescu, IEEE J. Electron Devices Soc. **8**, 695 (2020).
[29]L. Zhang, Z. Xu, J. Han, L. Liu, C. Ye, Y. Zhou, W. Xiong, Y. Liu, and G. He, J. Mater. Sci. Technol. **49**, 1 (2020).
[30]D. Kim, J. T. Jang, E. Yu, J. Park, J. Min, D. M. Kim, S.-J. Choi, H.-S. Mo, S. Cho, K. Roy, and D. H. Kim, ACS Appl. Electron. Mater. **2**, 2390 (2020).
[31]Y. Shibayama, Y. Ohnishi, T. Katagiri, Y. Yamamoto, Y. Nakashima, and M. Kimura, IEEE Electron Device Lett. **42**, 1014 (2021).
[32]M. Pereira, J. Deuermeier, R. Nogueira, P. A. Carvalho, R. Martins, E. Fortunato, and A. Kiazadeh, Adv. Electron. Mater. **6**, 2000242 (2020).
[33]C.-C. Chang, P.-T. Liu, C.-Y. Chien, and Y.-S. Fan, Appl. Phys. Lett. **112**, 172101 (2018).
[34]J. T. Jang, D. Kim, W. S. Choi, S.-J. Choi, D. M. Kim, Y. Kim, and D. H. Kim, ACS Appl. Electron. Mater. **2**, 2837 (2020).
[35]K. J. Heo, H. S. Kim, J. Y. Lee, and S. J. Kim, Sci. Rep. **10**, 9276 (2020).
[36]J. T. Jang, J. Min, Y. Hwang, S.-J. Choi, D. M. Kim, H. Kim, and D. H. Kim, IEEE Access **8**, 192304 (2020).
[37]P. Barquinha, A. M. Vila, G. Gonçalves, L. Pereira, R. Martins, J. R. Morante, and E. Fortunato, IEEE Trans. Electron Devices **55**, 954 (2008).
[38]W.-T. Chen, S.-Y. Lo, S.-C. Kao, H.-W. Zan, C.-C. Tsai, J.-H. Lin, C.-H. Fang, and C.-C. Lee, IEEE Electron Device Lett. **32**, 1552 (2011).
[39]C. Peng, S. Yang, C. Pan, X. Li, and J. Zhang, IEEE Trans. Electron Devices **67**, 4262 (2020).
[40]C. Baeumer, C. Schmitz, A. Marchewka, D. N. Mueller, R. Valenta, J. Hackl, N. Raab, S. P. Rogers, M. I. Khan, S. Nemsak, M. Shim, S. Menzel, C. M. Schneider, R. Waser, and R. Dittmann, Nat. Commun. **7**, 12398 (2016).
[41]K.-H. Choi and H.-K. Kim, Appl. Phys. Lett. **102**, 052103 (2013).
[42]J. Rosa, A. Kiazadeh, L. Santos, J. Deuermeier, R. Martins, H. L. Gomes, and E. Fortunato, ACS Omega **2**, 8366 (2017).
[43]S. Hoffmann-Eifert and R. Dittmann, *Advances in Non-Volatile Memory and Storage Technology* (Elsevier, 2019), pp. 171–202.
[44]S. Maidul Haque, P. R. Sagdeo, S. Balaji, K. Sridhar, S. Kumar, D. Bhattacharyya, D. Bhattacharyya, and N. K. Sahoo, J. Vac. Sci. Technol. B **32**, 03D104 (2014).
[45]J. Provine, P. Schindler, Y. Kim, S. P. Walch, H. J. Kim, K.-H. Kim, and F. B. Prinz, AIP Adv. **6**, 065012 (2016).
[46]Y.-S. Lee, W.-J. Chen, J.-S. Huang, and S.-C. Wu, Thin Solid Films **520**, 6942 (2012).
[47]P. Barquinha, R. Martins, L. Pereira, and E. Fortunato, *Transparent Oxide Electronics: From Materials to Devices* (Wiley, 2012), pp. 9–61.
[48]J.-W. Kim, T.-J. Jung, and S.-M. Yoon, J. Alloys Compd. **771**, 658 (2019).
[49]K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, Nature **432**, 488 (2004).
[50]A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, and A. J. Kenyon, Adv. Intell. Syst. **2**, 2000085 (2020).
[51]J.-W. Jang, S. Park, G. W. Burr, H. Hwang, and Y.-H. Jeong, IEEE Electron Device Lett. **36**, 457 (2015).
[52]S. Y. Pai-Yu Chen and X. Peng, in *2017 IEEE International Electron Devices Meeting, IEDM* (IEEE, 2018), pp. 6.1.1–6.1.4.
[53]S. Agarwal, S. J. Plimpton, D. R. Hughart, A. H. Hsia, I. Richter, J. A. Cox, C. D. James, and M. J. Marinella, in *International Joint Conference on Neural Networks* (IEEE, 2016), pp. 929–938.