Scaled, Ferroelectric Memristive Synapse for Back-End-Of-Line Integration with Neuromorphic Hardware

Laura Bégon-Lours[1], Mattia Halter[1,2], Francesco Maria Puglisi[3], Lorenzo Benatti[3], Donato Francesco Falcone[1], Youri Popoff[1,2], Diana Dávila Pineda[1] and Bert Jan Offrein[1]*

[1]IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,

[2]ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,

[3]UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy

*lbe@zurich.ibm.com

L. Bégon-Lours, M. Halter, Y. Popoff, D.F. Falcone, D. Dávila Pineda and B. J. Offrein
IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,
E-mail: lbe@zurich.ibm.com

M. Halter, Y. Popoff
ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,

L. Benatti, F.M. Puglisi
UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy

Keywords: ferroelectrics, synaptic weight, back-end-of-line, transport, resistive switching

Abstract

Ohmic, memristive synaptic weights were fabricated with a Back-End-Of-Line compatible process, based on a 3.5 nm $HfZrO_4$ thin film crystallized in the ferroelectric phase at only 400°C. The current density was increased by three orders of magnitude compared to the state-of-the-art. The use of a metallic oxide interlayer, WO_x, allows excellent retention (only 6% decay after 10^6 s) and endurance (10^{10} full switching cycles). The On/Off of 7 and the small device to device variability (<5%) makes them promising candidates for neural networks inference. The synaptic functionality for online learning was also demonstrated: using pulses of increasing (resp. constant) amplitude and constant (resp. increasing) duration, emulating spike-timing (resp. spike-rate) dependent plasticity. Writing with 20 ns pulses only dissipate femtojoules. The cycle-to-cycle variation was below 2%. The training accuracy (MNIST) of a neural network was estimated to reach 92% after 36 epochs. Temperature dependent experiments

revealed the presence of allowed states for charge carriers within the band gap of hafnium zirconate. Upon polarization switching, the screening of the polarization by mobile charges (that could be associated with oxygen vacancies and/or ions) within the ferroelectric layer modifies the energy profile of the conduction band and the bulk transport properties.

1. Introduction

Artificial Neural Networks, by analogy with the brain, consist of collections of interconnected neurons. The information flows from a layer of neurons to another, through vector-matrix multiplications. Tasks such as classification are possible by training the network, i.e. adjusting the matrix elements or "synaptic weights". Analog,[1] in-memory[2,3] as well as neuromorphic[4,5] computers aim at implementing the Vector-Matrix Multiplication (VMM) in the analog domain:[6][7] the "multiply" (through Ohm's law) and "accumulate" (through Kirchhoff's law) operation is performed by a parallel voltage drop through a cross-bar array of non-volatile, programmable resistances. Among the different technologies existing for memristive devices, ferroelectric synaptic weights rely on electrostatic effects: for example Schottky barrier height[8] or width[9] modulation, or electrostatically induced metal-insulator transitions[10]. Neuromorphic computers' learning paradigms (unsupervised[11,12] or supervised[13],[14]) require the ferroelectric memristive devices to show synaptic behavior, i.e. the ability to gradually decrease (depression) or increase (potentiation) their conductance upon voltage pulses emulating the effect on biological synapses of the pre- and post-synaptic spikes emitted by the neurons. Co-integrating ferroelectric materials with CMOS neurons became possible with the discovery of ferroelectricity in hafnia compounds[15]. In the race to miniaturization of Front-End-Of-Line (FEOL) devices, logic functions and discrete multi-level memories are demonstrated at the 28 and even 22 nm node.[16]

Fabrication in the Back-End-Of-Line (BEOL) relaxes the constraint on the device size: several nanometric ferroelectric domains can be contained in the active device area, allowing multi-level and/or analog conductance level updates, as demonstrated for example in field-effect transistors[17] and in Metal-Ferroelectric-Insulator-Metal (MFIM) devices. Based on thick HfO_2 layer, the band diagram of these two-terminals devices is engineered such that the electrons tunnel through the insulator, and partially through the ferroelectric[18,19]. These devices have a large dynamic range, but their non-linearity (which can be circumvented by the use of a logarithmic driver[3]) limit their usage for VMM. In contrast, in Metal-Ferroelectric-Semiconductor (MFS) devices, the electrons see the full thickness of the ferroelectric. In the

Back-End, the process temperature cannot exceed 400°C. 5 to 15 nm thick Hafnia films grown by Atomic Layer Deposition can be crystallized in the ferroelectric phase with a moderate thermal budget by rapid thermal annealing,[20] laser annealing[21] or millisecond-flash lamp annealing[22–24] but they result in low current densities. By decreasing the thickness, the current density increases, but at the cost of a higher thermal budget required for crystallizing the film[25]. In ref[26] 600°C are required to crystallize 5 nm of $HfZrO_4$ (HZO), and in ref[27] 800°C are required for 4, 6 and 8 nm of HZO.

In this work, ferroelectricity was obtained in a TiN/WO_x/ HZO/TiN structure with an HZO film as thin as 3.5 nm crystallized at 400°C, a thermal budget compatible with Back-End-Of-Line. Compared to 5 nm thick devices that were fabricated in the same conditions[24] and to state-of-the-art HfO_2(10 nm)/Al_2O_3 bilayers[18,19], the current density is 3 and 6 orders of magnitude higher, respectively. Their synaptic functionality, endurance and retention properties were characterized. The performance of a neural network based on this technology was estimated. The conduction mechanisms across the device, as well as the mechanisms governing the resistive switching, were explored using structural and ferroelectric characterization, as well as temperature dependent electrical measurements.

2. Ferroelectricity in Ultra-Thin, Back-End Compatible Bilayer

TiN, WO_3, $HfZrO_4$ (HZO) and TiN layers were deposited by Plasma-Enhanced Atomic Layer Deposition on an SiO_2 buffered Si substrate. TiN was chosen as a capping layer to promote the crystallization of HZO in the ferroelectric phase by a mechanical constraint during the annealing.[28] HZO was pre-heated to 400°C, then a 20 millisecond long energy pulse of 90 $J \cdot cm^{-2}$ was applied to crystallize it in the orthorhombic/tetragonal phase, without any fraction of the monoclinic phase, as is confirmed by the Grazing Incidence X-rays Diffraction scan in **Figure 1**. As shown later, the as-grown HZO film shows ferroelectric properties, indicating that the peak observed at 2θ-ω = 30.2° corresponds to the (111) planes of the orthorhombic, ferroelectric phase. The latter was fitted by a Gaussian with a width at half maximum of w = 1.7°. Using a k factor in the range 0.7 ~ 0.94, the in-plane crystallite size was estimated using the Scherrer equation:[29] $d = k\lambda_{k\alpha 1}^{Cu}/(w\cos\theta)$ to be d = 3.8 ~ 5.0 nm. This in-plane crystallite size is slightly larger than the thickness of the HZO layer. The broadening of the (111) peak is attributed to crystallite size and not to the presence of multiple peaks.

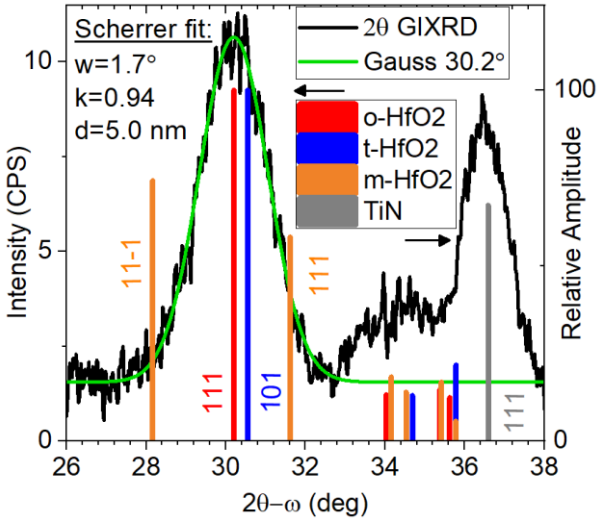


Figure 1: GIXRD scan of the device stack after crystallization (black curve). The peak at 30.2° is fitted by a Gaussian, from which the in-plane crystallite size is estimated at 5.0 nm using the Scherrer formula. The solid lines represent the predicted amplitude for the o-, t-, m- peaks of HZO.

X-ray reflectivity (**Figure S1**) confirms sharp interfaces and an ultra-low HZO thickness of 3.5 nm. To confirm the Back-End-Of-Line compatibility of the crystallization process, the same annealing was performed on MOSFETs (130 nm). The drain current as a function of the gate voltage characteristics of the annealed transistors fell within the spread of the characteristics of the pristine transistors. During the whole process, tools and temperatures compatible with Back-End-Of-Line conditions were used: the devices were defined by optical lithography and Reactive Ion Etching. The HZO, WO_x and bottom TiN layer were etched using Inductive Coupled Plasma with a CF_4 chemistry. The sputtered W metal lines were isolated by a SiO_2 spacer, deposited by Plasma-Enhanced Chemical Vapor Deposition at 300°C, as sketched in **Figure S2**.

3. A Ferroelectric Synaptic Weight

DC Resistive Switching:

Potentiation (depression) of the two-terminal device was first obtained by applying a negative (positive) DC "pulses" of decreasing (increasing) amplitude with respect to the grounded TiN/WO_x electrode. In this DC configuration, the duration of the applied bias was not controlled. After the application of each pulse of amplitude V_{write} (that varies in the -1.4 V to 1.6 V range) the bias was set back to zero, then the conductance was measured at 100 mV, as shown in **Figure 2** (yellow data points). In the High Resistive State (HRS), the HZO polarization points towards the WO_x layer (blue layer in the sketch in **Figure 2**). Upon the application of an

increasing positive bias, the fraction of domains whose polarization points away from the WO_x layer gradually increases. The On/Off ratio, defined as the ratio of the device conductance in the Low Resistive State (LRS) and in the HRS is 7. More than 25 levels can be clearly distinguished. The non-linearity of the long term potentiation and depression was quantified by fitting the normalized data represented in **Figure 2**, by a function of the normalized pulse number $y: x \rightarrow \frac{1-e^{\left(-\frac{x}{A}\right)}}{1-e^{\left(-\frac{1}{A}\right)}}$, as proposed in ref.[30] The parameter A was chosen by minimizing the root mean square error of the fitting. Values of A_{LTP}=0.5 for the long-term potentiation and A_{LTD}=-1 for the long-term depression, respectively, were found. For a device area of 314 μm^2, the resistance in the LRS, R_{on}, is 7 MΩ. The device-to-device variation, measured at 0.1 V as shown in **Figure S3**, is 5%.

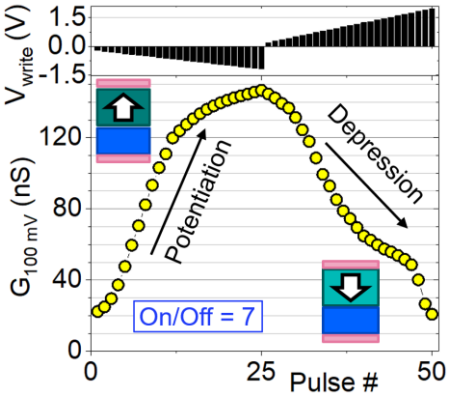


Figure 2: Potentiation and Depression of the device upon DC pulses (the duration is not controlled) of increasing amplitude V_{write}. The On/Off ratio is the maximal conductance divided by the minimal conductance.

In **Figure 3 a)** the I-V characteristics of a device in a 3x3 crossbar are presented. Around the read voltage (100 mV) the characteristics are linear, which is ideal for vector-matrix multiplication. Above, they are highly non-linear: consequently, the effect of sneak-paths during writing in selector-less arrays is limited.[31] The non-linearity, quantified by the I(V)/I(V/2) ratio, gradually increases from 3 at 0.6 V to 18 at 2.0 V, as represented in **Figure S4**.

To assess the effect of sneak paths, the current-voltage characteristic of the same device was measured in the worst-case scenario (all the other devices in the crossbar array were set to the LRS) and in the best-case scenario (all the other devices in the HRS). The data, shown in **Figure 3 a)**, differ by maximum 10% (grey curve). With the parameters listed earlier, the on-line learning accuracy of a selector-less array of 400 input, 250 hidden and 10 output neurons trained

on the MNIST database was simulated using the "MLP+NeuroSimV3.0" framework [32]. As shown in **Figure 3** b), the accuracy reaches 92% after 36 training epochs.

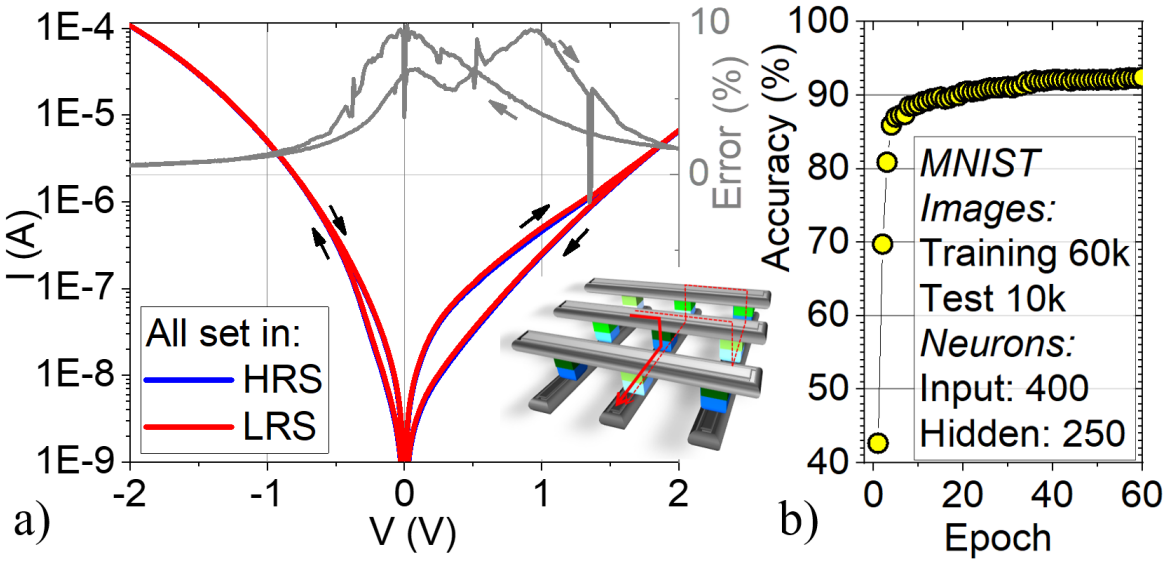


Figure 3: a) I-V of a device in a 3x3 array when the 8 other elements are in HRS (blue line) or in LRS (red line). Grey line: relative error. b) Simulated accuracy of a 1R array of 400 input, 250 hidden and 10 output neurons (MNIST database).

Pulsed weight update:

The conductance was then modulated using a waveform generator. **Figure 4** shows the conductance at V_{read} = 100 mV_{DC}, measured after each write pulse of increasing amplitude (V_{write}) and of constant duration t_{write} = 20 ns. In the inset, the same data is represented as a function of the pulse number within the cycle: the conductance varies by less than 2% from cycle to cycle. The energy dissipated during the writing is minimal for V_{write} = 0.2 V ($E^{0.2V} \approx 10^{-15}$ J) and maximal for V_{write} = -2 V ($E^{-2V} \approx 10^{-12}$ J).

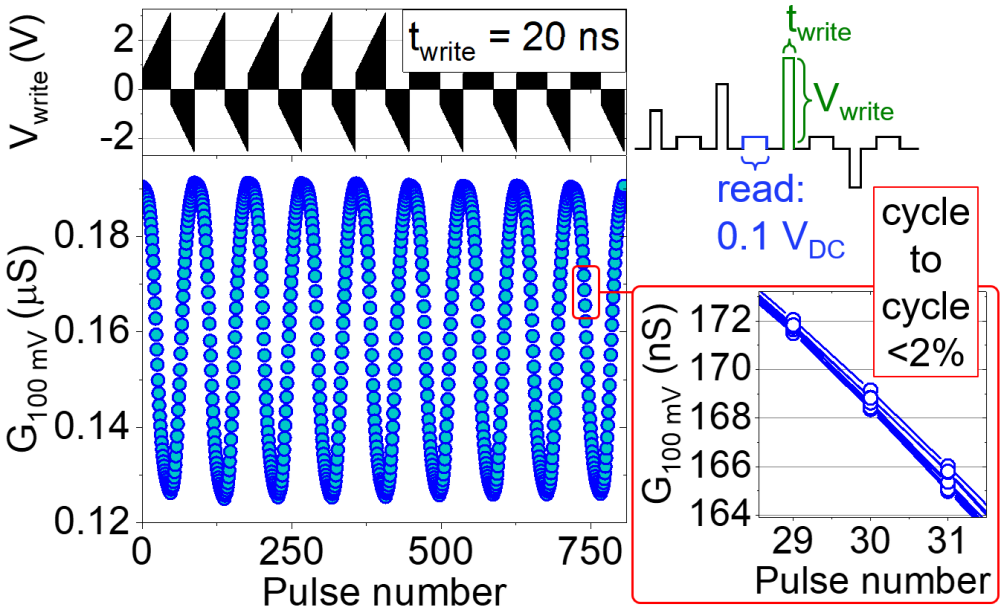


Figure 4: Synaptic Potentiation and Depression with pulses of increasing amplitude. The pulse amplitude V_{write} and the pulse duration t_{write} are defined as in the measurement scheme (top right). The inset shows the cycle-to-cycle variation.

In **Figure 5**, the amplitude is kept constant (-1.4 V for potentiation, 2.0 V for depression) and the pulse width (t_{write}) is increased. With this scheme, 20% of the dynamic range is traversed after the first pulse (20 ns, lower limit of the generator). The circuitry required for the on-chip implementation of this scheme is more accessible than the one required by the scheme with increasing amplitude. Moreover, it emulates spike trains and shows that the proposed devices have potential applications in neuromorphic systems implementing a Spike-Rate-Dependent-Plasticity learning rule.[33] The cycle-to-cycle variation for both schemes is <2%. Cumulative switching (upon applying pulses of constant width and amplitude) was not observed.

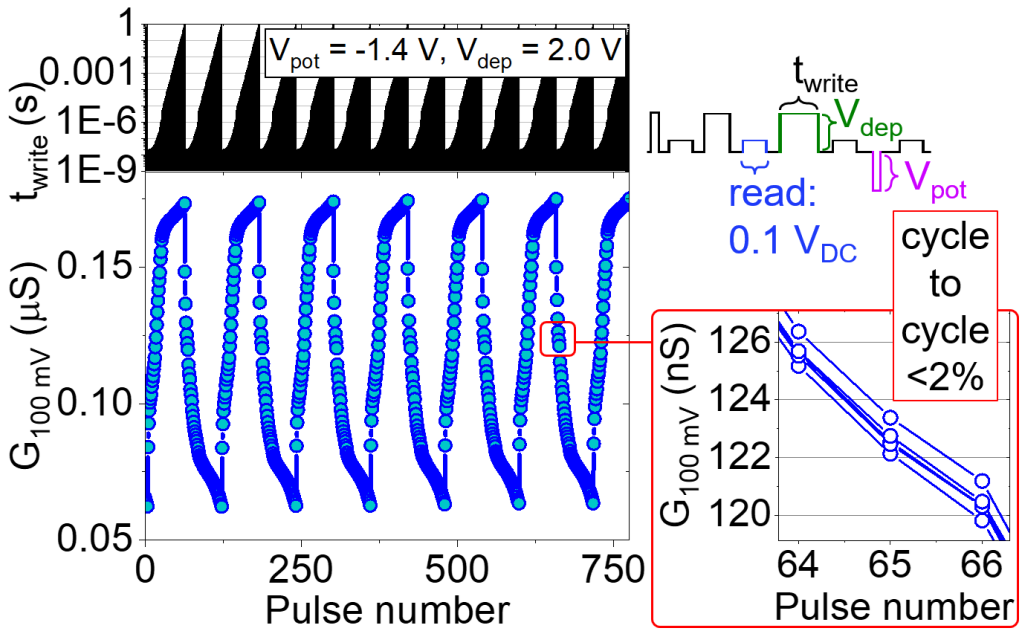


Figure 5: Long Term Potentiation and Depression for pulses with constant amplitudes V_{pot} and V_{dep} (as defined in the measurement scheme) and increasing duration t_{width}. The inset shows the cycle-to-cycle variation.

Retention properties were measured for over ten days: the worst-case scenario corresponds to a 6% decay at 300 h (10^6 s) in the case where the polarization points towards the oxide interlayer (pink triangles in **Figure S5**). This confirms that the devices based on a metal oxide / ferroelectric bilayer showed limited back-switching compared to devices based on a dielectric interlayer: for comparison, optimized junctions based on a SiO_2 / Si:HfO_2 bilayer showed a 30% decay.[34] This observation is consistent with a better screening of the polarization charges by the metallic WO_x, which implies a reduction of the depolarization field across the HZO ferroelectric layer that leads to improved retention characteristics. Finally, the devices showed strong robustness against fatigue, with no dielectric breakdown after more than 10^{10} full switching cycles (+/- 2V at 100 kHz).

4. Investigation of the Resistive Switching Mechanism

Both tungsten oxides[35] and hafnium oxides[36] were used as memristors operating with the displacement of oxygen vacancies and ions. In this work, the current density is constant for circular devices with diameters in the 10 to 100 µm range: **Figure S6** shows the resistances measured at 0.1 V in the HRS and LRS in logarithmic scale, showing a constant On/Off ratio of 7 for diameters in the 10 – 30 µm and a homogeneous conduction across the device: no filaments or conduction at the edges. In addition, for a given polarity, resistive switching was only observed for pulses of increasing amplitude or time. Electroresistance loops (**Figure S7**)

revealed hysteretic behavior: the resistance increases (resp. decreases) while the amplitude increases from 0 to 2V (resp. decreases from 0 to -2V), then remains constant as the amplitude decreases from 2V to 0V (resp. increases from -2V to 0V). This discards current-driven resistive switching mechanisms (governing for example filamentary resistive memories).

Positive Up Negative Down (PUND) experiments were performed to measure the ferroelectric properties of the devices. The waveform is schematized in **Figure S8 a)**. The polarization for a 60 µm capacitor measured at 50 kHz is represented in **Figure S 8b)**. As for the current density, the polarization (per unit area) is independent on the device area (circular devices of diameter 40, 50 and 60 µm), and confirms the ferroelectric properties of the devices ($2P_r \approx 4.5$ µC·cm^{-2}). This value is low compared to 10 nm films (≈20 µC·cm^{-2}) due to the reduced thickness. For a constant voltage range (-1 V to 1 V), the polarization decreases with the frequency of the measurement, as represented in **Figure 6**: this effect reflects the dependence of the resistive switching on the pulse duration observed in **Figure 5.** The polarization switching occurs over a broad range of coercive fields (from +/-0.2 V to above +/-1 V), matching the voltage range where the resistive switching is observed (see **Figure 2**). Such broad distribution is explained by the polycrystalline nature of the HZO film, determined by the X-Ray analysis in **Figure 1**. Finally, we observed that the resistive switching saturates upon increasing field amplitude for both polarities (see the electroresistance loops in **Figure S7**). These observations are consistent with the resistive switching originating from ferroelectric domains switching: it saturates when all the domains are aligned in the same direction, and multiple intermediate configurations are allowed as only the grains with an intermediate coercive field have switched.

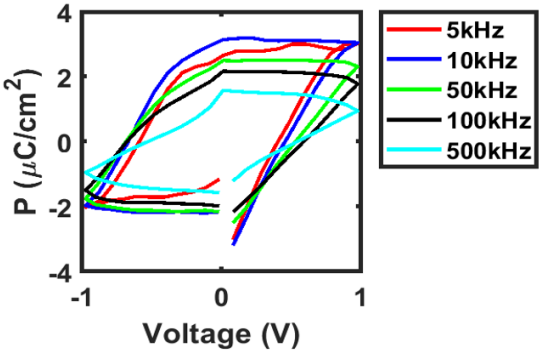


Figure 6: Polarization measured by the PUND method for a 60 µm capacitor, for various frequencies.

In this paragraph, the relative contribution of WO_x to the overall resistance is discussed. WO_3 is an insulator with a band gap of 3.40 eV; oxygen vacancies create donor states that push the Fermi level close to the conduction band, conferring n-type semiconducting properties to

WO_x.[37] For x<2.9, WO_x shows metallic behavior.[38] The situation where WO_x is most likely to contribute to the resistance of the device is when the polarization points away from the WO_x layer, after applying a negative bias on the top (TiN) electrode, causing an electrostatic depletion of electrons in the WO_x. From the resistive switching experiments (for example in **Figure 2**), this corresponds to the LRS of the synaptic weight. This shows that the resistance change in the WO_x layer, is small compared to the resistance change in the HZO layer. In addition, a chip with identical HZO thickness but with a WO_x layer thickness of only one nanometer (compared to two nanometers for the reference stack) was fabricated. Two devices of the same diameter (80 µm) but different WO_x thicknesses shared the same LRS of R_{On} = 1.13 MOhms, showing that the resistance of the WO_x layer was small compared to the resistance of the HZO layer. In the following paragraph, we assume that the electric field drop in the WO_x is small compared to that in the HZO layer.

The conduction mechanisms across the devices were then analyzed with current-voltage (I-V) sweeps between 25 and 55°C, back and forth with no degradation observed upon heating. Although the On/Off ratio remains constant, the resistance decreases with increasing temperature, as shown in **Figure 7** and **Figure S7**. For this reason, direct tunneling[39] and Fowler Nordheim tunneling[40] (respectively proposed in [41] and [42]) were discarded as being the dominant transport mechanism. At low bias, below 75 mV and where the On/Off ratio is maximal, Ohmic conduction is observed, due to the drift of a small number of mobile electrons in the material's conduction band.[43] It is described by:

$$J = \sigma E = \mu q N_C exp\left[\frac{-(E_C - E_F)}{kT}\right] E \quad (1)$$

equivalent to:

$$Log(J) = Log(\mu q N_C/t) + \frac{-(E_C - E_F)}{k} \times \frac{1}{T} + Log(V) \quad (2)$$

where J is the current density, σ the electrical conductivity, µ the electron mobility, q the electronic charge, N_C the carrier concentration at equilibrium, t the sample thickness, E_C - E_F the energy difference between the conduction band and the Fermi level, k the Boltzmann constant, T the absolute temperature and E the electric field across the ferroelectric layer. As discussed in the previous paragraph, the electric field was assumed to drop mainly across the ferroelectric layer resulting in E=V/t where the thickness t is equal to 3.5 nm. Only the non-switching branches of the IV sweeps were analyzed. The product μN_C is assumed independent of the temperature.[43] At each temperature, a linear regression was performed in the Log(V) ↦ Log(J) representation, shown in **Figure 7** a). **Figure 7** c) shows the Arrhenius plot of the intercepts of the linear regression in the Ohmic regime: the presence of the conduction band at

only 0.3 eV above the Fermi level, deducted using Equation (2), indicates the presence of donor states in the band gap of HZO (~5.4 eV[44]). Such states could originate from the presence of oxygen vacancies, but also from hydrogen trapped in the lattice during the atomic layer deposition[45]. From **Figure 7** c) and **Equation (2)**, we found that the resistive switching originates from an increase by one order of magnitude in the μN_C product in the LRS compared to the HRS. At moderate fields (75 ~ 200 mV) the I–V characteristics keep the same diode-like polarity regardless of the polarization direction, showing that the energy band diagram of the TiN/HZO/WO_x junction is not structurally modified upon polarization reversal as for example in refs.[46–48]

At larger bias, we first considered electrode-limited mechanisms: the thermionic emission model[49] (proposed in [50]) fitted reasonably well the data, however the Richardson constant obtained was of the order of 0.1 $A.m^{-2}.K^{-2}$ in the HRS and 100 $A.m^{-2}.K^{-2}$ in the LRS, which is orders of magnitude smaller than the universal Richardson constant[51] (~10^6 $A.m^{-2}.K^{-2}$). Bulk-limited conduction mechanisms were also studied: similarly, it was possible to fit the experimental data with a Poole-Frenkel conduction model (proposed in [52]), but the dielectric constant obtained with this model was unrealistically as high as 60. Thermionic-field emission[53], hopping[54], space-charge-limited[54] fittings were not satisfying. Phonon-mediated trap-assisted-tunneling[55] and simplified trap-assisted-tunneling[56] models, in a single branch approach, were not fitting the data on a sufficient range. Regardless of the conduction mechanisms explored, the corresponding barrier height seen by the electrons on the negative branch (TiN to HZO injection, Φ_b^-) is higher than in the other polarity (Φ_b^+), although the current flowing is larger. It supports the description of the transport by a "Modified Schottky Emission" (MSE) mechanism,[57] which confers both electrode and bulk limited characters to the conduction. The MSE model was previously observed in ZrO_2 thin films,[58] and describes the experimental data well. It is governed by the equation:

$$J = \alpha T^{\frac{3}{2}} E \mu \left(\frac{m^*}{m_0}\right)^{\frac{3}{2}} exp\left[\frac{-q\left(\phi_B - \sqrt{{}^{qE}/_{4\pi\varepsilon_r\varepsilon_0}}\right)}{kT}\right] \qquad (3)$$

equivalent to:

$$Log\left(\frac{J}{T^{\frac{3}{2}}V}\right) = \frac{\alpha}{t}\mu\left(\frac{m^*}{m_0}\right)^{\frac{3}{2}} + \frac{-q\phi_B}{k} \times \frac{1}{T} + \frac{-q\left(\sqrt{{}^{q}/_{t4\pi\varepsilon_r\varepsilon_0}}\right)}{k} \times \frac{1}{T} \times \sqrt{V} \qquad (4)$$

where α = 3×10^{-4} $A{\cdot}s{\cdot}cm^{-3}{\cdot}K^{-3/2}$ is a constant, m_0 the free electron mass, m^* the effective electron mass in HZO, $q\phi_B$ the Schottky barrier height, ε_0 the permittivity in vacuum, and ε_r

the dynamic dielectric constant. Linear regressions in the representations of $Log(J/(T^{3/2}V))$ as a function of $|V|^{0.5}$ were performed (**Figure 7** b).

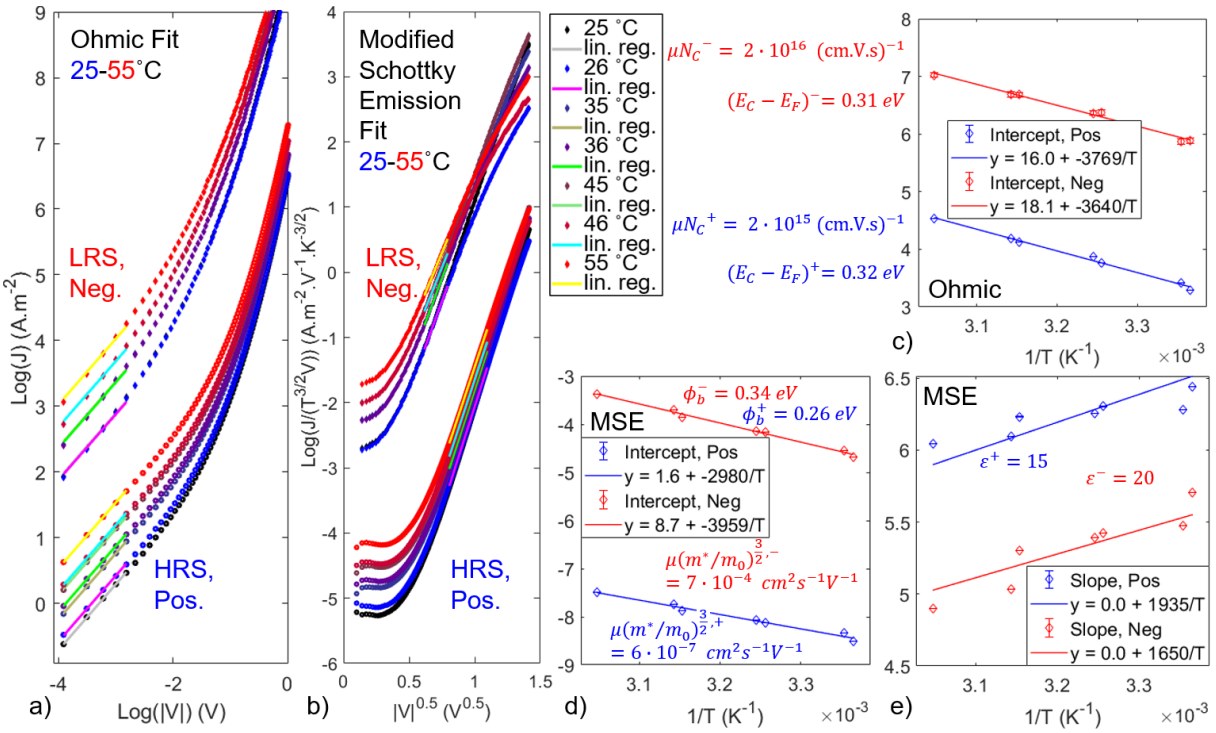


Figure 7: Current-Voltage characteristics in the Ohmic (a) and Modified Schottky Emission (MSE) (b) representations. At each temperature, a linear regression is performed on the negative, increasing branch (LRS, non-switching) and on the positive, decreasing branch (HRS, non-switching). Arrhenius plots allow the calculation of the μN_C and E_C-E_F parameters from the intercepts of the linear regressions in the Ohmic regime (c), the $\mu(m^*/m_0)^{3/2}$ parameter is calculated from the intercepts of the linear regression in the MSE regime (d), and the parameters Φ_B, ε from the slopes of the later (e).

For this mechanism, the barrier heights measured from **Figure 7** d) and Equation (4) are $\Phi_b^- =$ 0.34 eV and $\Phi_b^+ = 0.26$ eV. Consistently with the increase of μN_C measured at low bias, the average product $\mu(m^*/m_0)^{3/2}$ calculated from **Figure 7** d) and **Equation (4)** increases by several orders of magnitude from the HRS to the LRS. On average, the dynamic dielectric constant ε, calculated from **Figure 7** e) and **Equation (4)** also increases.

The models describe HZO as a semiconductor and show that the resistive switching originates from the modification of bulk transport properties upon polarization reversal. Such modification can be explained by the displacement of charged defects within the semiconducting ferroelectric layer upon polarization switching in the vicinities of the interfaces. The displacement of charged defects is driven by the screening of the polarization charges as

observed recently in sub-stoichiometric $BaTiO_{3-x}$ tunnel junctions,[59] but also in a polymer ferroelectric.[60] The simplified analytical models presented above do not capture eventual gradients within the HZO layer: further analysis using compact models simulations[61] may further improve the understanding of the resistive switching in the HZO/WO_x bilayers.

5. Conclusion

In this work, ferroelectricity was obtained in TiN/WO_x/ HZO/TiN structure with an HZO film as thin as 3.5 nm crystallized at 400°C, a thermal budget compatible with Back-End-Of-Line. Thanks to the reduced thickness, the current density was increased by three orders of magnitude compared to the state-of-the-art[62] and the devices operate in the Ohmic regime for read-out, which makes them ideal memristors for analog vector-matrix multiplication. The use of a metallic oxide electrode, WO_x, enables good retention properties: the conductance in the less stable state varies by only 6% after 10^6 s. In comparison, optimized MFIM devices show a 30% decay for the most stable state[34]. The On/Off of 7 and the small device to device variability (<5%) makes them promising candidates for neural networks inference.

In addition, the synaptic functionality was demonstrated using pulses of increasing amplitude and constant duration, emulating spike-timing dependent plasticity. Using pulses as short as 20 ns, the writing energy was in the femtojoule range. Spike-rate dependent plasticity was also emulated, using pulses of constant amplitude (of only 2V) and increasing duration. For both schemes the cycle-to-cycle variation was below 2%. The current-voltage non-linearity in the range of 3~18, allowing limited effects of the sneak paths on addressing a device in passive cross-bar arrays, which was verified on a 3x3 cross-bar. Using the weight update non-linearity parameters of 0.5 and -1, the training accuracy of a neural network based on the proposed synapse on the MNIST dataset was estimated to reach 92% after 36 epochs. In addition, the endurance of the devices exceeds 10^{10} full switching cycles.

The conduction mechanisms across the device, as well as the mechanisms governing the resistive switching, were explored using temperature dependent experiments: the results indicate the presence of allowed states for charge carriers within the band gap of HZO, originating from defects such as oxygen vacancies. Upon polarization switching, the screening of the polarization by mobile charges (that could be associated with oxygen vacancies and/or ions) within the ferroelectric layer modifies the energy profile of the conduction band and the bulk transport properties.

Experimental Section/Methods

Device preparation

A 200 nm thick SiO_2 oxide was grown on Si by thermal oxidation. The active stack was then deposited by Plasma-enhanced atomic layer deposition (PE-ALD): 20 nm of TiN was deposited at 300°C with Tetrakis(dimethylamino)titanium and N_2 as precursors. 2 nm of WO_x was deposited at 375°C with $(BuN)_2W(NMe_2)_2$ and O_2, then 3.5 nm of HZO was deposited at 300°C alternating one cycle with Tetrakis(ethylmethylamino) hafnium (IV) and O_2, and two cycles with Bis (methylcyclopentadienyl) (methyl) (methoxy) zirconium (IV) and O_2. Ten additional nanometers of TiN were deposited. The crystallization was performed with the millisecond flash lamp annealing technique [22]: the sample was preheated to 400°C, then a 20 ms long energy pulse of 90 $J \cdot cm^{-2}$ was applied. A 100 nm thick W metal electrode was then deposited by sputtering. The top electrode, defining the area of the junction, was defined by optical lithography and reactive ion etching (RIE) of the W and top TiN layers. Using this method, the HZO layer acted as an etch stop. The bottom electrode was then defined by optical lithography and inductive coupled plasma reactive ion etching (ICP) of the HZO, WO_x and TiN layers. A 100 nm thick SiO_2 passivation layer was deposited at 300°C by plasma-enhance chemical vapor deposition (PECVD). Vias to the top and the bottom electrode were defined by optical lithography. The SiO_2 layer was etched by RIE, then the HZO and the WO_x were etched by ICP, exposing the TiN layer to air. The etch was immediately followed by the sputtering of 100 nm of W. The first metal lines were then defined by optical lithography and etching by RIE. A 100 nm thick SiO_2 passivation layer was deposited at 300°C by PECVD. Vias to the bottom electrode contacts were defined by optical lithography. The SiO_2 layer was etched by RIE. 100 nm of W was sputtered. The second metal lines were then defined by optical lithography and RIE.

Structural characterization

Grazing-incidence X-ray diffraction (GIXRD) and X-Ray Reflectivity (XRR) measurements were performed on a Bruker D8 Discover diffractometer equipped with a rotating anode generator.

Electrical characterization

Electrical measurements were performed on an Agilent B1500A semiconductor analyzer with a B1530A waveform generator/fast measurement unit (WGFMU). Write pulses were generated by a remote-sense and switch unit (RSU) module close to the probe and applied to the top electrode while the bottom electrode is grounded. The device resistance was measured at

V = +/-100 mV with a high-resolution source-measurement unit (SMU) at the top electrode while the bottom electrode was grounded.

Positive Up Negative Down (PUND) experiments were performed to measure the ferroelectric properties of the devices. The waveform is schematized in **Figure S8 a)**: during each of the P (positive, switching), U (positive, non-switching), N (negative, switching) and D (negative, non-switching) pulses of period T, the current I was measured during a time t_{READ} (that depends on the voltage range spanned and on the frequency) at a voltage V, then the voltage was increased by 0.1 V during a 20 ns long ramp. The polarization was calculated from the measured I-V curves.

Supporting Information

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

Acknowledgements

We thank the Cleanroom Operations Team of the Binnig and Rohrer Nanotechnology Center (BRNC) for their help and support. This work is funded by H2020: FREEMIND (nº 840903), ULPEC (nº732642), BeFerroSynaptic (nº 871737) and UNICO (ANR-19-CHR3-0006).

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

[1] Z. Sun, G. Pedretti, A. Bricalli, D. Ielmini, *Sci. Adv.* **2020**, *6*, eaay2378.
[2] R. Yang, *Nat. Electron.* **2020**, *3*, 237.
[3] R. Berdan, T. Marukame, K. Ota, M. Yamaguchi, M. Saitoh, S. Fujii, J. Deguchi, Y. Nishi, *Nat. Electron.* **2020**, *3*, 259.
[4] Z. Yan, J. Chen, R. Hu, T. Huang, Y. Chen, S. Wen, *Neural Netw.* **2020**, *128*, 142.
[5] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, W. Lu, *Nano Lett.* **2010**, *10*, 1297.
[6] F. J. Kub, K. K. Moon, I. A. Mack, F. M. Long, *IEEE J. Solid-State Circuits* **1990**, *25*, 207.
[7] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, *Nature* **2015**, *521*, 61.
[8] X. Liu, Y. Wang, J. D. Burton, E. Y. Tsymbal, *Phys. Rev. B* **2013**, *88*, 165139.
[9] Z. Xi, J. Ruan, C. Li, C. Zheng, Z. Wen, J. Dai, A. Li, D. Wu, *Nat. Commun.* **2017**, *8*, 15217.
[10] X. Liu, J. D. Burton, E. Y. Tsymbal, *Phys. Rev. Lett.* **2016**, *116*, 197602.
[11] A. Serb, J. Bill, A. Khiat, R. Berdan, R. Legenstein, T. Prodromakis, *Nat. Commun.* **2016**, *7*, 12611.
[12] M. Hansen, F. Zahari, H. Kohlstedt, M. Ziegler, *Sci. Rep.* **2018**, *8*, 8914.
[13] S. R. Nandakumar, I. Boybat, M. Le Gallo, E. Eleftheriou, A. Sebastian, B. Rajendran,

Sci. Rep. **2020**, *10*, 8080.
[14] A. Renner, F. Sheldon, A. Zlotnik, L. Tao, A. Sornborger, in *Proc. Neuro-Inspired Comput. Elem. Workshop*, ACM, Heidelberg Germany, **2020**, pp. 1–3.
[15] T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, U. Böttger, *Appl. Phys. Lett.* **2011**, *99*, 102903.
[16] S. Dunkel, M. Trentzsch, R. Richter, P. Moll, C. Fuchs, O. Gehring, M. Majer, S. Wittek, B. Muller, T. Melde, H. Mulaosmanovic, S. Slesazeck, S. Muller, J. Ocker, M. Noack, D.-A. Lohr, P. Polakowski, J. Muller, T. Mikolajick, J. Hontschel, B. Rice, J. Pellerin, S. Beyer, in *2017 IEEE Int. Electron Devices Meet. IEDM*, IEEE, San Francisco, CA, USA, **2017**, p. 19.7.1-19.7.4.
[17] M. Halter, L. Bégon-Lours, V. Bragaglia, M. Sousa, B. J. Offrein, S. Abel, M. Luisier, J. Fompeyrine, *ACS Appl. Mater. Interfaces* **2020**, *12*, 17725.
[18] B. Max, M. Hoffmann, H. Mulaosmanovic, S. Slesazeck, T. Mikolajick, *ACS Appl. Electron. Mater.* **2020**, *2*, 4023.
[19] R. Berdan, T. Marukame, S. Kabuyanagi, K. Ota, M. Saitoh, S. Fujii, J. Deguchi, Y. Nishi, in *2019 Symp. VLSI Technol.*, IEEE, Kyoto, Japan, **2019**, pp. T22–T23.
[20] M. Lederer, D. Lehninger, S. Abdulazhanov, A. Reck, R. Olivo, T. Kampfe, K. Seidel, in *2021 IEEE Int. Symp. Appl. Ferroelectr. ISAF*, IEEE, Sydney, Australia, **2021**, pp. 1–4.
[21] L. Grenouillet, T. Francois, J. Coignus, S. Kerdiles, N. Vaxelaire, C. Carabasse, F. Mehmood, S. Chevalliez, C. Pellissier, F. Triozon, F. Mazen, G. Rodriguez, T. Magis, V. Havel, S. Slesazeck, F. Gaillard, U. Schroeder, T. Mikolajick, E. Nowak, in *2020 IEEE Symp. VLSI Technol.*, IEEE, Honolulu, HI, USA, **2020**, pp. 1–2.
[22] É. O'Connor, M. Halter, F. Eltes, M. Sousa, A. Kellock, S. Abel, J. Fompeyrine, *APL Mater.* **2018**, *6*, 121103.
[23] L. Bégon-Lours, M. Halter, Y. Popoff, B. J. Offrein, *Phys. Status Solidi RRL – Rapid Res. Lett.* **2020**, pssr.202000524.
[24] L. Begon-Lours, M. Halter, D. D. Pineda, Y. Popoff, V. Bragaglia, A. L. Porta, D. Jubin, J. Fompeyrine, B. J. Offrein, in *2021 5th IEEE Electron Devices Technol. Manuf. Conf. EDTM*, IEEE, Chengdu, China, **2021**, pp. 1–3.
[25] K. Tahara, K. Toprasertpong, Y. Hikosaka, K. Nakamura, H. Saito, M. Takenaka, S. Takagi, in *2021 Symp. VLSI Technol.*, **2021**, p. 2.
[26] Y.-K. Liang, J.-S. Wu, C.-Y. Teng, H.-L. Ko, Q.-H. Luc, C.-J. Su, E.-Y. Chang, C.-H. Lin, *IEEE Electron Device Lett.* **2021**, *42*, 1299.
[27] T. Ali, P. Polakowski, K. Kuhnel, M. Czernohorsky, T. Kampfe, M. Rudolph, B. Patzold, D. Lehninger, F. Muller, R. Olivo, M. Lederer, R. Hoffmann, P. Steinke, K. Zimmermann, U. Muhle, K. Seidel, J. Muller, in *2019 IEEE Int. Electron Devices Meet. IEDM*, IEEE, San Francisco, CA, USA, **2019**, p. 28.7.1-28.7.4.
[28] M. Hyuk Park, H. Joon Kim, Y. Jin Kim, T. Moon, C. Seong Hwang, *Appl. Phys. Lett.* **2014**, *104*, 072901.
[29] B. E. Warren, *X-Ray Diffraction*, Courier Corporation, **1990**.
[30] P. Chen, X. Peng, S. Yu, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2018**, *37*, 3067.
[31] K.-H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, W. Lu, *Nano Lett.* **2012**, *12*, 389.
[32] Y. Luo, X. Peng, S. Yu, in *Proc. Int. Conf. Neuromorphic Syst.*, ACM, Knoxville TN USA, **2019**, pp. 1–7.
[33] S. Li, F. Zeng, C. Chen, H. Liu, G. Tang, S. Gao, C. Song, Y. Lin, F. Pan, D. Guo, *J Mater Chem C* **2013**, *1*, 5292.
[34] S. Fujii, M. Yamaguchi, S. Kabuyanagi, K. Ota, M. Saitoh, in *2020 IEEE Symp. VLSI Technol.*, IEEE, Honolulu, HI, USA, **2020**, pp. 1–2.

[35] T. Chang, S.-H. Jo, K.-H. Kim, P. Sheridan, S. Gaba, W. Lu, *Appl. Phys. A* **2011**, *102*, 857.
[36] Y. Wang, Q. Liu, S. Long, W. Wang, Q. Wang, M. Zhang, S. Zhang, Y. Li, Q. Zuo, J. Yang, M. Liu, **2009**, *21*, 045202.
[37] M. T. Greiner, Z.-H. Lu, *NPG Asia Mater.* **2013**, *5*, e55.
[38] E. Salje, B. Güttler, *Philos. Mag. B* **1984**, *50*, 607.
[39] D. Pantel, M. Alexe, *Phys. Rev. B* **2010**, *82*, 134105.
[40] D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, **2015**.
[41] Y. Wei, S. Matzen, T. Maroutian, G. Agnus, M. Salverda, P. Nukala, Q. Chen, J. Ye, P. Lecoeur, B. Noheda, *Phys Rev Appl.* **2019**, *12*, 031001.
[42] K. Ota, J. Deguchi, S. Fujii, M. Saitoh, M. Yamaguchi, R. Berdan, T. Marukame, Y. Nishi, K. Matsuo, K. Takahashi, Y. Kamiya, S. Miyano, in *2019 IEEE Int. Electron Devices Meet. IEDM*, IEEE, San Francisco, CA, USA, **2019**, p. 6.2.1-6.2.4.
[43] S. M. Sze, K. K. Ng, *Physics of Semiconductor Devices*, Wiley-Interscience, Hoboken, N.J, **2007**.
[44] H. C. Shin, D. Tahir, S. Seo, Y. R. Denny, S. K. Oh, H. J. Kang, S. Heo, J. G. Chung, J. C. Lee, S. Tougaard, *Surf Interface Anal* **2012**, 6.
[45] H. Kim, S. Yun, T. H. Kim, H. Kim, C. Bae, S. Jeon, S. Hong, *Phys. Status Solidi RRL – Rapid Res. Lett.* **2021**, 2100020.
[46] Q. Luo, Y. Cheng, J. Yang, R. Cao, H. Ma, Y. Yang, R. Huang, W. Wei, Y. Zheng, T. Gong, J. Yu, X. Xu, P. Yuan, X. Li, L. Tai, H. Yu, D. Shang, Q. Liu, B. Yu, Q. Ren, H. Lv, M. Liu, *Nat. Commun.* **2020**, *11*, 1391.
[47] T. Choi, S. Lee, Y. J. Choi, V. Kiryukhin, S.-W. Cheong, *Science* **2009**, *324*, 63.
[48] M. C. Sulzbach, S. Estandía, X. Long, J. Lyu, N. Dix, J. Gàzquez, M. F. Chisholm, F. Sánchez, I. Fina, J. Fontcuberta, *Adv. Electron. Mater.* **2020**, *6*, 1900852.
[49] W. Schottky, *Phys. Zeitschr* **1914**, *15*, 872.
[50] H. Y. Yoong, H. Wu, J. Zhao, H. Wang, R. Guo, J. Xiao, B. Zhang, P. Yang, S. J. Pennycook, N. Deng, X. Yan, J. Chen, *Adv. Funct. Mater.* **2018**, *28*, 1806037.
[51] C. R. Crowell, *Solid-State Electron.* **1965**, *8*, 395.
[52] D. S. Jeong, H. B. Park, C. S. Hwang, *Appl. Phys. Lett.* **2005**, *86*, 072903.
[53] C. Hamann, H. Burghardt, T. Frauenheim, *Electrical Conduction Mechanisms in Solids*, Vch Pub, **1988**.
[54] Nevill Francis Mott and Edward A Davis, *Electronic Processes in Non-Crystalline Materials*, Oxford University Press, **2012**.
[55] D. R. Islamov, V. A. Gritsenko, T. V. Perevalov, V. A. Pustovarov, O. M. Orlov, A. G. Chernikova, A. M. Markeev, S. Slesazeck, U. Schroeder, T. Mikolajick, G. Y. Krasnikov, *Acta Mater.* **2019**, *166*, 47.
[56] S. Fleischer, P. T. Lai, Y. C. Cheng, *J. Appl. Phys.* **1992**, *72*, 5711.
[57] J. G. Simmons, *Phys Rev Lett* **1965**, *15*, 967.
[58] F.-C. Chiu, Z.-H. Lin, C.-W. Chang, C.-C. Wang, K.-F. Chuang, C.-Y. Huang, J. Y. Lee, H.-L. Hwang, *J. Appl. Phys.* **2005**, *97*, 034506.
[59] J. Li, N. Li, C. Ge, H. Huang, Y. Sun, P. Gao, M. He, C. Wang, G. Yang, K. Jin, *iScience* **2019**, *16*, 368.
[60] Tian, Bo Bo, Liu, Yang, Chen, Liu Fang, Wang, Jian Lu, *Sci. Rep.* **2015**, *5*, 9.
[61] L. Larcher, A. Padovani, P. Pavan, **2012**, *2*, 6.
[62] L. Begon-Lours, M. Halter, Y. Popoff, Z. Yu, D. F. Falcone, D. Davila, V. Bragaglia, A. La Porta, D. Jubin, J. Fompeyrine, B. J. Offrein, *IEEE J. Electron Devices Soc.* **2021**, 1.

Table of contents

A Back-End-Of-Line, ferroelectric synapse shows excellent cycle-to-cycle and device-to-device variation (2%), retention and endurance (10^{10} cycles). A 10% error is measured in a 3x3 crossbar (92% accuracy predicted on the MNIST dataset). Ohmic conduction at low bias is ideal for multiply and accumulate operation. Temperature measurements reveal the analog resistive switching mechanisms in semiconducting and ferroelectric $HfZrO_4$.

Scaled, Ferroelectric Memristive Synapse for Back-End-Of-Line Integration with Neuromorphic Hardware

Laura Bégon-Lours[1], Mattia Halter[1,2], Francesco Maria Puglisi[3], Lorenzo Benatti[3], Donato Francesco Falcone[1], Youri Popoff[1,2], Diana Dávila Pineda[1] and Bert Jan Offrein[1]*

[1]IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,

[2]ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,

[3]UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy

*lbe@zurich.ibm.com

L. Bégon-Lours, M. Halter, Y. Popoff, D. Dávila Pineda and B. J. Offrein

IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,

E-mail: lbe@zurich.ibm.com

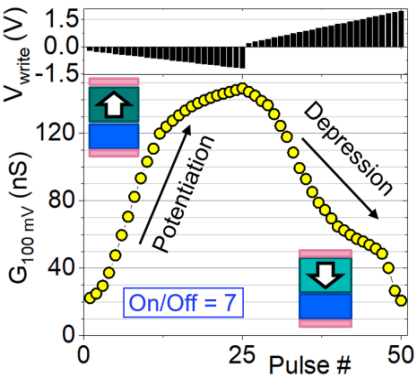
M. Halter, Y. Popoff

ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,

L. Benatti, F.M. Puglisi

UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy

ToC figure :



Supporting Information

Scaled, Ferroelectric Memristive Synapse for Back-End-Of-Line Integration with Neuromorphic Hardware

Laura Bégon-Lours[1], Mattia Halter[1,2], Francesco Maria Puglisi[3], Lorenzo Benatti[3], Donato Francesco Falcone[1], Youri Popoff[1,2], Diana Dávila Pineda[1] and Bert Jan Offrein[1]*

[1]IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,
[2]ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,
[3]UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy
*lbe@zurich.ibm.com

L. Bégon-Lours, M. Halter, Y. Popoff, D. Dávila Pineda and B. J. Offrein
IBM Research Europe - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland,
E-mail: lbe@zurich.ibm.com

M. Halter, Y. Popoff
ETH Zurich - Integrated Systems Laboratory, CH-8092 Zurich, Switzerland,

L. Benatti, F.M. Puglisi
UNIMORE - Dipartimento di Ingegneria "Enzo Ferrari" - 41125 Modena, Italy

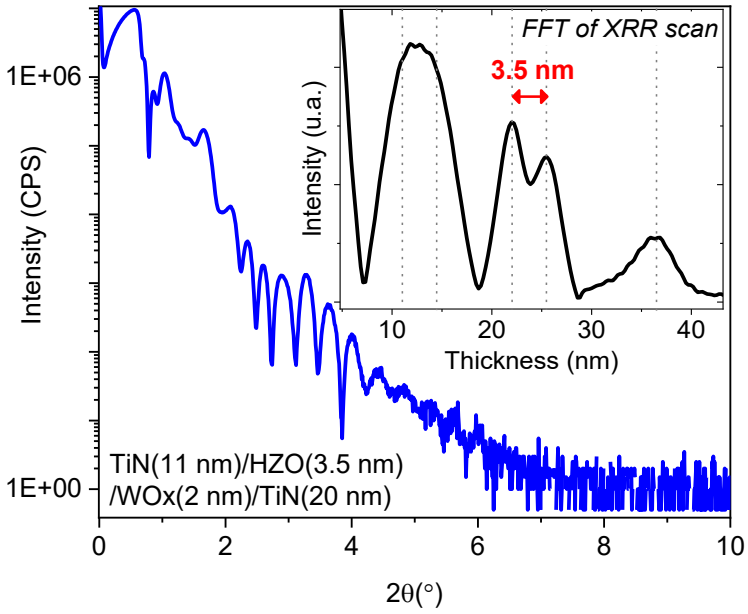


Figure S1: X-Rays Reflectivity measurement of the TiN (11 nm) / HZO (3.5 nm) / WO_x (2 nm) / TiN (20 nm) / SiO_2 (200 nm) / Si stack after ms-Flash Lamp Annealing. The inset shows the Fast-Fourier Transform from which the thickness of 3.5 nm is measured for the HZO layer.

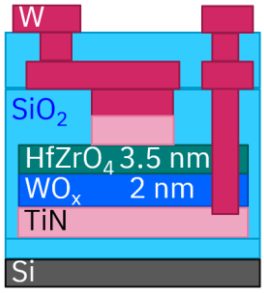


Figure S2: Sketch of the device

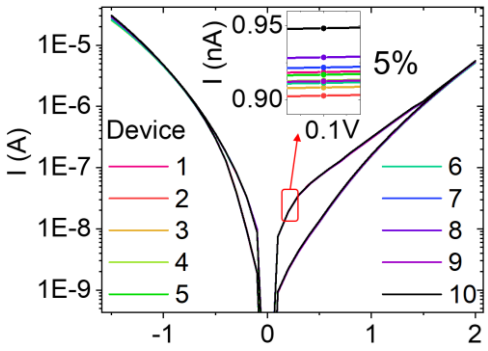


Figure S3: I-V characteristics of ten devices, showing a device-to-device variation of 5% at 0.1 V.

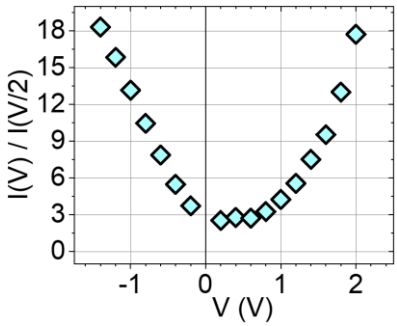


Figure S4: Non-linearity of the I-V characteristics.

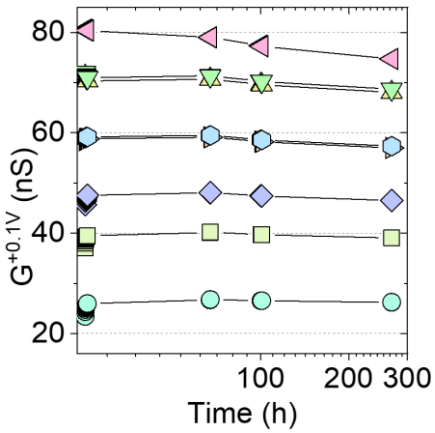


Figure S5: Retention of several devices in the high, intermediate and low resistive states

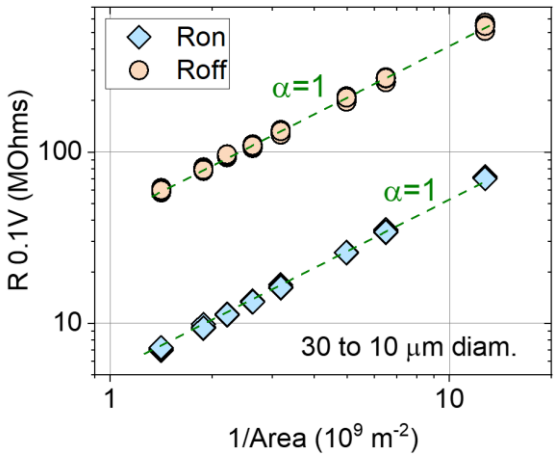


Figure S6: Resistance measured at 0.1 V in the HRS (Roff) and LRS (Ron) as a function of the inverse of the area of the device.

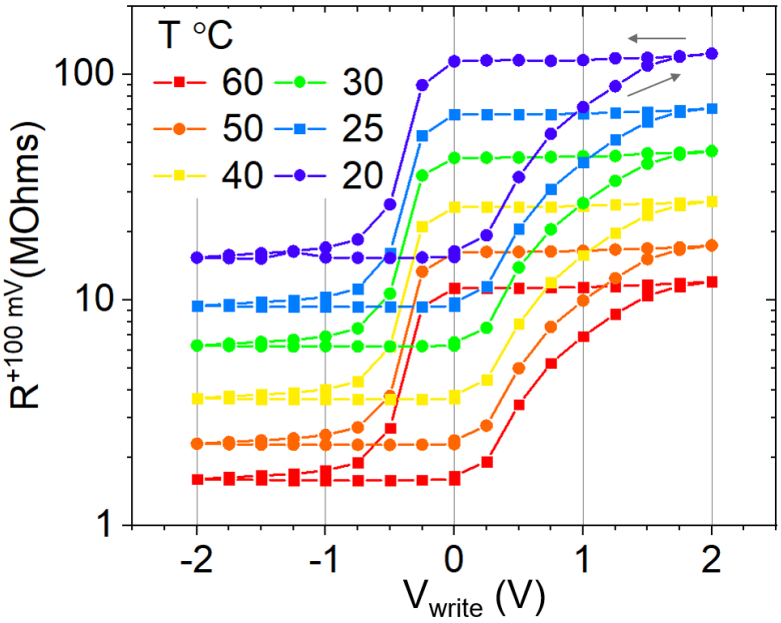


Figure S7: Resistance measured at 100 mV after applying V_{write}, at various temperatures T. The arrows indicate the chronological order.

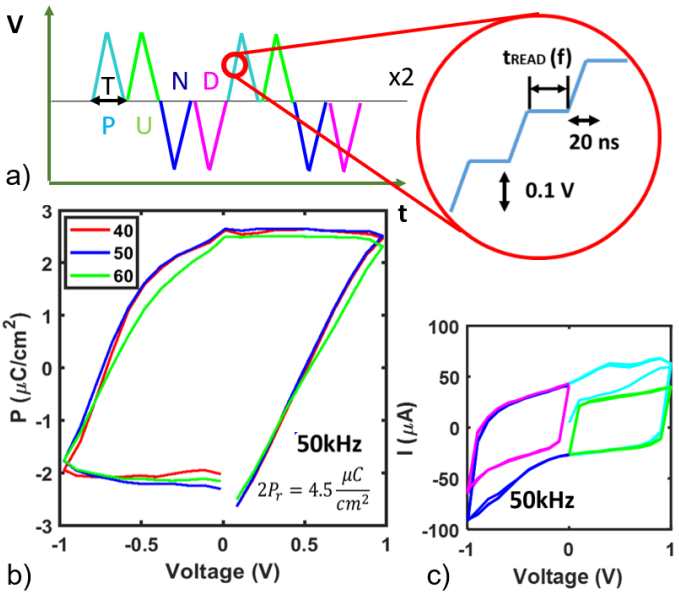


Figure S8: Measure of the ferroelectric polarization by the PUND method. a) P and U, resp. N and D are staircase triangular pulses as detailed in the inset. The polarization in b) is calculated for three capacitors of 40, 50 and 60 µm in diameter, for example for the 60 µm device from the current measured in c): each branch of different color corresponds to a P, U, N or D pulse, as described in a). The frequency (50 kHz in b) and c)) is defined as 1/2T where T is the duration of a pulse as defined in a).