A Back-End, CMOS Compatible Ferroelectric Field Effect Transistor for Synaptic Weights

Mattia Halter[1,2], Laura Bégon-Lours[1], Valeria Bragaglia[1], Marilyne Sousa[1], Bert Jan Offrein[1], Stefan Abel[1], Mathieu Luisier[2], and Jean Fompeyrine[1]
1: IBM Research GmbH - Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland
2: Integrated Systems Laboratory, ETH Zurich, CH-8092 Zurich, Switzerland
*E-mail: att@zurich.ibm.com

***Abstract*— Neuromorphic computing architectures enable the dense co-location of memory and processing elements within a single circuit. Their building blocks are non-volatile synaptic elements such as memristors. Key memristor properties include a suitable non-volatile resistance range, continuous linear resistance modulation and symmetric switching. In this work, we demonstrate voltage-controlled, symmetric and analog potentiation and depression of a ferroelectric $Hf_{0.57}Zr_{0.43}O_2$ (HZO) field effect transistor (FeFET) with good linearity. Our FeFET operates with a low writing energy (fJ) and fast programming time (40ns). Retention measurements have been done over 4-bits depth with low noise (1%) in the tungsten oxide (WO_x) read out channel. By adjusting the channel thickness from 15nm to 8nm, the on/off ratio of the FeFET can be engineered from 1% to 200% with an on-resistance ideally >100kΩ, depending on the channel geometry. The device concept is compatible with a back end of line (BEOL) integration into CMOS processes. It has therefore a great potential for the fabrication of high density, large-scale integrated arrays of artificial analog synapses.**

Keywords—FeFET, hafnium zirconate, ferroelectrics

I. Introduction

Brain-inspired neuromorphic architectures with co-located computation and memory units appear as promising candidates to overcome the Von Neuman bottleneck, in particular crossbar arrays of artificial "synapses" connected to CMOS neurons.[1] Key synapses properties are CMOS compatibility and analog programming capabilities. The recent discovery of ferroelectricity in hafnia composites,[2] a material already present in CMOS lines, has revived interest for integrated ferroelectrics: the programmed states are linked to the partial switching of the ferroelectric polarization, which allows to fine tune the synaptic weights with fast and low-power writing.[3] Ferroelectric synapses have been realized as two-terminal junctions[4] as well as three-terminal ferroelectric field-effect transistors (FeFET),[5,6,7] the later having the advantage of separating the write process (low power write through high impedance gate[3]) and the read process (through source-drain resistance).

Hafnia-based FeFETs were demonstrated mainly as non-volatile memory cells,[7] steep-slope field-effect transistors,[8,9] and artificial neurons,[10] usually implemented on the front end of line (FEOL) and using Si as a channel. Integration in the back end of line (BEOL) can enable larger device area with respect to the size of the ferroelectric domains, which can translate into a larger number of states; moreover, the ability to integrate the synaptic design between any metal level of the BEOL increases the flexibility in the design of neuromorphic circuits. Recently, analog synaptic behavior has been shown in a hafnia based FeFET with indium gallium zinc oxide (IGZO) and poly-Si channels fabricated in the BEOL.[6,11,12,13] Using an oxide channel is expected to alleviate the known issues associated with Si-based FeFETs such as unintended low-k interfacial layers formed at the Si interface. Here, we report on a $Hf_{0.57}Zr_{0.43}O_2$ (HZO) based FeFET utilizing a tungsten oxide (WO_x) channel. Our device is a junction-less transistor where the ferroelectric polarization is used to electrostatically deplete or accumulate free carriers in the thin WO_x channel. We demonstrate the impact of the ferroelectric polarization on the channel resistance, the influence of the channel thickness on the on/off ratio, ferroelectric HZO with a long endurance, the stabilization of multiple differentiable states, a good retention as well as a continuous potentiation and depression.

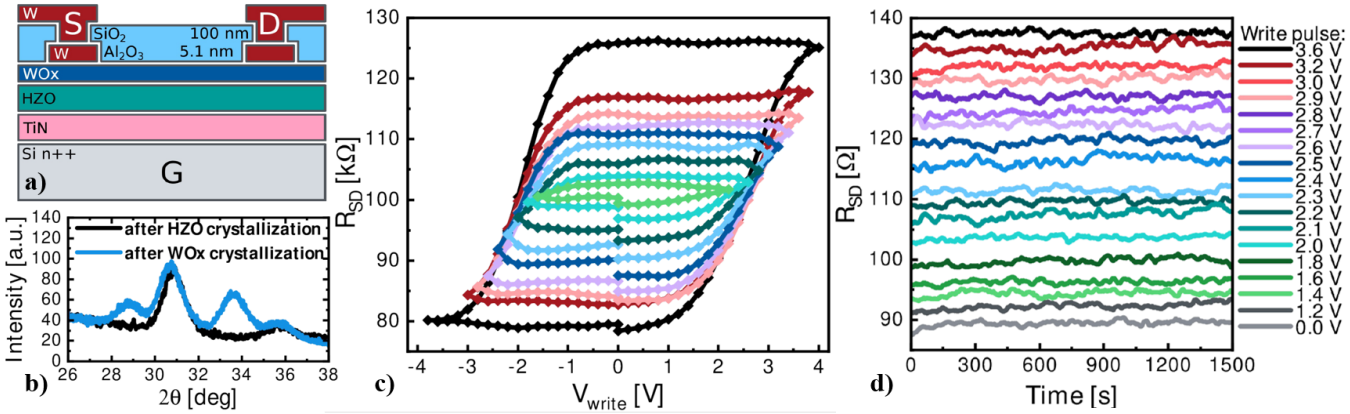

Figure 1: a) Scheme of the FeFET. b) GIXRD scan showing no monoclinic HZO. c) Analog multi-level behavior of a FeFET of 20µm width and 5µm length: the channel resistance (R_{DS}) after the application of 5µs write pulses (V_{write}) of varying amplitudes. Each data point corresponds to a resistance measurement between S and D at $V_{read,D}$ = 200mV. The different curves correspond to different consecutive measurements with reducing V_{write} range. d) Retention measurement for 1500s . $V_{read,D}$ = 200mV was uninterruptedly applied while the current was measured every 5s to determine R_{DS}.

II. Results

FeFET devices were fabricated from a HZO (10nm)/TiN (10nm)/n^+ Si gate stack and an 8 nm thick WO_x channel[20,21] (**Figure 1a**). The HZO layer is crystallized in the orthorhombic/tetragonal phase only (no monoclinic phase is observed, **Figure 1b**) with a reduced thermal budget (~375C) using a ms-Flash Lamp Annealing technique[14]. Then the channel is formed by oxidizing 2.5nm of W. The device is encapsulated by a 5nm Al_2O_3 and a 100nm SiO_2 passivation layer. Contact pads are formed on top of the passivation layers and routed through openings to source and drain. W/WO_x/HZO/TiN/n+*Si* capacitors are also fabricated for electrical characterization: the capacitance per unit area is $C_{OX} = 2.7\mu F/cm^2$ and the polarization is $+P_r = 12.4\mu C/cm^2$.

WO_x is an *n*-type semiconductor. When the HZO ferroelectric polarization points towards (outwards) the interface with WO_x, free-carriers accumulate (deplete) at the interface to screen the electric field, and the resistance (R_{DS}) of the channel decreases (increases): it is a junction-less transistor. Our channel area (20x5 μm^2) is 10^5 larger than the crystalline grain size, which allows analog resistance levels by switching only a subset of the domains. The fraction of the switched ferroelectric domains depends on the amplitude, width, and number of the applied write pulses. the multi-state nature of a 20µm wide and 5µm long FeFET was investigated by applying voltage pulses of varying amplitudes (V_{write}), while keeping a fixed pulse duration of 5µs (**Figure 1c**). When V_{write} varies from −4V to 4V, R_{DS} shows a hysteretic cycle from 80kΩ to 125kΩ with various intermediate states (on/off ≈ 1.55). By reducing the range of V_{write} numerous R_{DS} sub-loops can be accessed. Retention properties are shown in **Figure 1d**: first, an intermediate state was written by a 5µs pulse. Then, a source-to-drain voltage $V_{DS} = 200mV$ was continuously applied for 1500s, while R_{DS} was measured every 5s. The FeFET showed stable retention for 18 differentiable channel resistances (>4bit). The stable retention measurement hints to an absence of depolarization or other screening mechanisms. This is in agreement with a partially depleted channel which is still able to screen the polarization charges. Normalizing the cycle-to-cycle standard deviation for 25 consecutive cycles (**Figure 2a**) by R_{ON} reveals a constant value of about 1%. When fitting the potentiation range from 1V to 3.1V and depression range from −0.9V to −3.0V by linear regression an adjusted residual-square value of 0.952 is obtained. Gaussian process regression (GPR) was used to predict a noise free signal (**Figure 2b**) The average symmetry factor (SF) across the full resistance range is $SF = 0.20$ while the most linear part in the center reaches a very good symmetry factor of $SF = 0.08$ (**Figure 2c**). In addition, potentiation/depression is achieved with a constant pulse amplitude (4.2 V and -3.0 V) and increasing pulse duration from 40 ns to 250 ns.

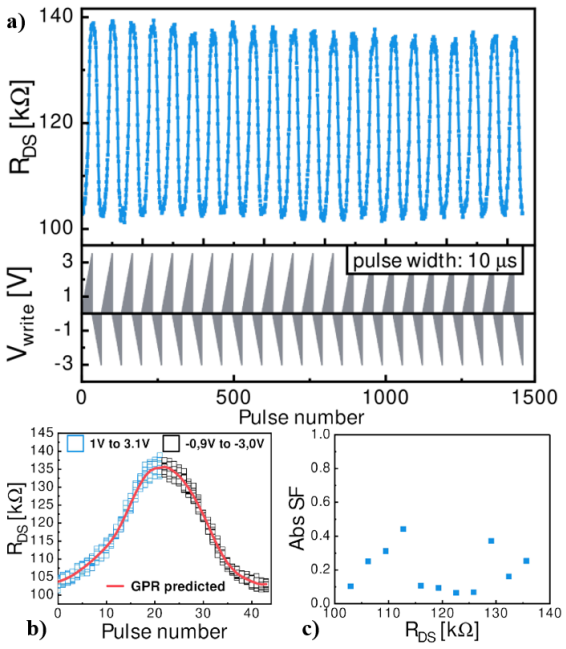

Figure 2: a) channel resistance (R_{DS}) of a 20µm wide and 5µm long FeFET with varying pulse amplitude (V_{write}) and constant pulse width (t_{write}). b) R_{DS} as a function of pulse number and the GPR predicted noise free signal (red). c) Symmetry factor as a function of R_{DS}.

III. Conclusion

We propose a BEOL compatible device that operates by applying the ferroelectric field effect to a thin WO_x channel by using a HZO gate dielectric. Such a device can be used as a synaptic element in hardware-supported neural networks. By utilizing a junction-less transistor design, no high temperature source and drain activation is required. By comparing HZO and HfO_2 based devices, and carefully analyzing capacitor and transistor data, we unambiguously show that the channel resistance is directly coupled to the polarization of the HZO layer and can be programmed in a non-volatile manner. Multilevel states programmed over more than 4-bits depth with a stable retention over 1500s and an almost symmetric potentiation and depression is obtained, together with a low programming energy. The property of the WO_x layer and the geometry of the device can be arranged so that a well-suited resistance range is obtained, favorable to build large scale arrays. The proposed device exhibit therefore promising metrics when considered as a synaptic element for processing cores supporting artificial neural networks.

Acknowledgment

This work is funded by H2020: ULPEC (732642).

References

[1] Likharev, K. *et al.,* N. Y. A. Sci. 2003, 1006, 146–163.
[2] Böscke, T. S. *et al.,* Appl. Phys. Lett. 2011, 99, 0–3.
[3] Slesazeck, S. *et al.,* ICICD Technol. 2018; 121–124.
[4] Chanthbouala, A. *et al.,* Nat. Mater. 2012, 11, 860–864.
[5] Nishitani, Y. *et al.,* J. Appl. Phys. 2012, 111, 1–7.
[6] Jerry, M. *et al.,* IEEE IEDM 2017; pp 6.2.1–6.2.4.
[7] Mulaosmanovic, H. *et al.,* ACS Appl. Mater. Interfaces 2017, 9, 3792–3798.
[8] Sharma, P. *et al.,*. Dig. VLSI Technol. 2017; T154
[9] Krivokapic, Z. *et al.,* IEEE IEDM 2017; p. 15.1.1
[10] Mulaosmanovic, H. *et al.,* Nanoscale 2018, 10, 21755
[11] Kim, M.-K. *et al.,* Nano Lett. 2019, 19, 2044–2050.
[12] Mo, F. *et al.,* 2019 Symp. VLSI Tech. 2019, T42–T43.
[13] Zeng, B. *et al.,* IEEE JEDS. 2019, 7, 551–556.
[14] É. O'Connor *et al.,* APL Materials, 2018, 6, 12