A Ferroelectric Tunnel Junction-based Integrate-and-Fire Neuron

P. Gibertini†, L. Fehlings†, S. Lancaster‡, Q. T. Duong†, T. Mikolajick‡, E. Covi†, Q. T. Duong†, and V. Deshpande§
†NaMLab gGmbH, Dresden, Germany ‡Technical University of Dresden, Dresden, Germany §Helmholtz Zentrum Berlin, Berlin, Germany ∗Free University Berlin, Berlin, Germany †These Authors contributed equally to this work
Corresponding Authors: erika.covi@namlab.com, veeresh.deshpande@helmholtz-berlin.de

Abstract—Event-based neuromorphic systems provide a low-power solution by using artificial neurons and synapses to process data asynchronously in the form of spikes. Ferroelectric Tunnel Junctions (FTJs) are ultra low-power memory devices and are well-suited to be integrated in these systems. Here, we present a hybrid FTJ-CMOS Integrate-and-Fire neuron which constitutes a fundamental building block for new-generation neuromorphic networks for edge computing. We demonstrate electrically tunable neural dynamics achievable by tuning the switching of the FTJ device.

Index Terms—Integrate-and-Fire Neuron, FTJ, HZO, neuromorphic computing, edge computing.

I. INTRODUCTION

Event-based neuromorphic systems where data is processed asynchronously as spikes are an attractive solution for low power edge computing. The basic computing blocks in such systems are artificial neurons and synaptic circuits. While there has been significant effort to develop CMOS neurons and synapse circuits [1], the recent developments in emerging devices provides an opportunity for integration with CMOS technology to design neuron and synaptic circuits that can be power and area efficient compared to their CMOS only counterparts [2]. Among the various emerging devices, ferroelectric tunnel junctions (FTJ) feature ultra-low switching energy [3]. Hafnium Zirconium Oxide (HfO2ZrO2O2; HZO) based FTJ devices can be integrated on the CMOS back-end-of-line enabling hybrid FTJ-CMOS circuits [4]. Despite growing interest in synaptic applications of HZO FTJs [5], neuron circuits have not been much explored. The HZO FTJ device features thresholding behavior in resistance change which is useful for neuron circuits. In this context, we present a hybrid FTJ-CMOS Integrate-and-Fire neuron circuit. The circuit utilizes an HZO-based bilayer FTJ device that realizes the pulse integration. The circuit can become a building block for low-power neuromorphic systems.

This work was supported in part by the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No 871737 and in part by the European Research Council (ERC) through the European’s Horizon Europe Research and Innovation Programme under Grant Agreement No 101042585. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Fig. 1: (a) Schematic of the FTJ stack. (b) DC-IV measurements of the FTJ in ON and OFF state, measured up to 1.5 V. Schematics of the corresponding band structures are shown in the inset.

II. GRADUAL SWITCHING PROPERTIES OF FTJ DEVICES

Fig. 1a shows the HZO based bilayer FTJ device considered in this work. Details of the FTJ device fabrication can be found in [6]. The operation principle of the devices is shown through band diagrams in the insets of Fig. 1b. When the ferroelectric polarization (blue/red arrows) points towards the Al2O3 layer, an electric field is induced which results in the ON state defined by a high tunneling current. When the polarization points away, the band bending results in a larger tunneling barrier width and an OFF state with a low tunneling current. The polarization is switched ON (set) by applying 5 V, and switched OFF (reset) by applying -5 V. The polarization can be partially switched by using lower voltage amplitudes or shorter pulse times. The currents of these intermediate states are directly proportional to the amount of switched polarization [7]. In order to study gradual switching under identical set or reset pulses (representing spikes) for the neuron integration function, the switching under multiple square pulses with fixed amplitude and width was characterized. Variable numbers of voltage pulses were applied to the device and the switched polarization was measured by switching the device back with a single, triangular -5 V, 500 µs pulse [7]. Fig. 2 shows the back-switching current after gradual switching with pulses of 4 V.
amplitude and 10 µs width. Based on the non-switching current measured on a second back-switching pulse, the switching current and accordingly the switched polarization can be measured, which is illustrated by the shaded area in Fig. 2. Complete switching was measured under a triangular voltage pulse with amplitude 5 V and width 500 µs. An evaluation of the switched polarization for different pulse widths and an amplitude of 3.5 V (Fig. 3a), as well as for different voltages with a pulse width of 10 µs (Fig. 3b) supports the viability of the gradual switching approach employed in the neuron circuit.

III. FTJ COMPACT MODEL

For the circuit simulations, a compact model based on a HZO ferroelectric capacitor was used. Here we provide a summary of the salient aspects of the model. For a more detailed description, please see [8]. We do not model a two-layer structure, rather we take the Al₂O₃ layer into account by adapting the value of the effective coercive field $E_C$ to represent the capacitive divider we have in our FTJ. The HZO layer consists of a mix of ferroelectric, antiferroelectric, and paraelectric phases due to the fabrication process. Therefore, the total capacitance of the FTJ is modeled as the parallel of a ferroelectric capacitor and a dielectric one (Fig. 4b).

The FTJ model is based on the Preisach model of ferroelectric hysteresis and has been calibrated to fit the characteristics of the device discussed in Section II. The Preisach model describes the evolution of the polarization $P$ as a function of applied electric field $E$ [9]. We calculate the asymptotic polarization for a certain voltage following Miller et al. [10] as:

$$P(E, k, P_{off}) = k \cdot P_{sat} \cdot \tanh \left( \frac{E_{eff} \pm E_C}{2\delta} \right) + P_{off}$$ (1)

where $k$ and $P_{off}$ are a scaling factor and the offset polarization, respectively, used to account for unsaturated loops [11], $P_{sat}$ is the saturation polarization, $E_{eff}$ the field applied to the ferroelectric layer, and $\delta$ is:

$$\delta = E_C \left( \ln \frac{1 + P_r/P_{sat}}{1 - P_r/P_{sat}} \right)^{-1}$$ (2)

where $P_r$ is the remanent polarization. The switching kinetics is added to the model by using the static polarization $P$ as in (1) and adding a time dependence to find the dynamic polarization $P_{dyn}$:

$$P_{dyn} = P \frac{\Delta t + P_{old} \cdot \tau_{PE}}{\tau_{PE} + \Delta t}$$ (3)

$\Delta t$ is the actual time-step of the simulation, $P_{old}$ is the polarization calculated in the previous time-step, and $\tau_{PE}$ is

$$\tau_{PE} = \tau_p 10^{\alpha_E (\frac{E_{eff}}{E_C} - 1)}$$ (4)

where $\tau_p$ and $\alpha_E$ are fitting parameters.

Each change in the polarization results in a polarization current, $I_{pol}$, modeled through the ferroelectric capacitor (Fig. 4b):

$$I_{pol} = \frac{dP_{dyn}}{dt} \cdot A_{tot}$$ (5)
TABLE I: Values of the parameters used to calibrate the compact model.

| Parameter | Value   | Parameter | Value   |
|-----------|---------|-----------|---------|
| $E_C$     | 3.3 MV/cm | $\tau_p$  | 10 $\mu$s |
| $k$       | 0.5     | $\alpha_E$| 0.25   |
| $P_{sat}$ | 20 0000 $\mu$C/cm$^2$ | $A_{tot}$  | 3.14 $\times$ $10^{-4}$ cm$^4$ |
| $P_{v}$   | 19 9997 $\mu$C/cm$^2$ | $R_{A0}$   | 110 $\mu$A |
| $V_{P0}$  | 0.36 V  | $\Delta V_P$| 0.06 V |

The very low leakage current in FTJ devices (see Fig. 1b) allows very energy efficient operations, but poses a challenge in reading the memory state. The proposed circuit, shown in Fig. 5, adopts a variation of the 2 transistor-1 capacitor (2T1C) readout scheme ($T_1$ to $T_4$ in Fig. 5) which uses a read transistor $T_4$ to amplify the signal from the FTJ [12]. The circuit was designed in a standard 180 nm CMOS technology. Transistor $T_3$ is added to prevent any disturbance to reach the gate of $T_4$ during the writing phase. Two inverters with tunable switching threshold ($T_4$ to $T_7$) are used in order to obtain sharp pulses starting from a slow changing input signal [13]. The PMOS gate voltages ($V_{p1}$ and $V_{p2}$) are used to tune the switching threshold of the inverters. During the writing operation (green and purple shades in Fig. 6a), transistors $T_1$, $T_2$ are turned ON through $V_{WL}$, while the pass transistor $T_3$ is OFF. Voltage pulses, negative to reset and positive to set, are applied at the plate line ($PL$) while the bit line ($BL$) is kept at a constant voltage. The reading operation (orange shade in Fig. 6a) is divided in two different phases, pre-charge and integration. During the pre-charge phase (2 $\mu$s, evidenced by the overlapping of the orange shade with the green / purple one in Fig. 6a), transistors $T_1$, $T_2$, and $T_3$ are turned ON, thus charging node $n_1$ to $V_{BL}$. $V_{BL}$ is chosen near the switching voltage of the first inverter, in the high-gain region. Then the integrating phase begins. $T_1$, $T_2$ are turned OFF so that $BL$ is disconnected, leaving $n_1$ floating, and a read voltage ($V_{read}$) is applied to $PL$. Under these conditions, the FTJ discharges, i.e., the voltage at $n_1$ ($V_{n1}$) increases, because of the FTJ leakage current. Indeed, the device can be modeled as an RC circuit where the capacitance is constant and the resistance is proportional to the polarization (see Sect. III). The final $V_{n1}$ after a read phase of 100 $\mu$s depends on the resistance of the FTJ. If it is higher than the switching voltage of the first inverter, $V_{out}$ rises and the neuron fires, as shown in Fig. 6a. Fig. 6b shows $V_{n1}$ during each read phase shown in Fig. 6a. The slope of $n_1$ depends on the state of the FTJ and the difference between the final voltage at the end of the reading phase after the first and the last pulse is of almost 20 mV.

It is worth noting that the signals which provide the full functionality of the neuron (e.g., bias and pulse generators) have to be generated outside the neuron itself, although on the same chip. However, the additional circuit can be shared among several neurons when using an asynchronous, event-based approach. Each single neuron can process event occurring at a minimum time distance of 150 $\mu$s. The limit on the bandwidth imposed by the necessary integration time of the single neuron is overcome in scaled systems with a large number of neurons [14].

The choice of $V_{read}$ is critical, as there is only a small window where a readable leakage current is obtained (see Fig. 1b) without modifying the polarization. Therefore, a voltage of 1.5 V was chosen. Other important parameters for the performance of the neuron are the $BL$ voltage $V_{BL}$ and the integration, i.e., read, time. Decreasing the integration time and/or reducing $V_{BL}$ slows down the neuron response, i.e., increases the number of pulses needed before the firing event.

The reset can be performed by applying a strong negative pulse (-5 V for 10 $\mu$s) or by applying 5 V at the $BL$ while...
keeping the $PL$ at 0V. The behavior of the neuron, i.e., the number of pulses needed to fire, can be modulated by changing the parameters of the set pulse. In accordance to the experimental data (Fig. 3), we can observe in Fig. 7 that a change in the set pulse amplitude or time width results in a different number of pulses needed before the firing event. It is observed that higher voltages / longer times lead to firing with a lower number of pulses. From the perspective of circuit design, it would be more convenient to exploit the time- rather than voltage-dependence of the FTJ. However, the FTJ shows a better modulation with voltage amplitude than time width. A change in voltage of 2.5 V results in a gradual reduction of pulses before fire, from 43 to 1 for 10 μs pulses. When fixing the voltage at 3 V and modulating the pulse width, a sharp reduction is instead observed between $10^{-6}$ and $10^{-5}$ s, from 43 to 10 pulses before fire. This hardly changes for pulse widths over the next two orders of magnitude. One way to improve flexibility in the neural behavior and cope with process variability, both at circuit and at device level, whilst still keeping the design compact can be by changing the values of $V_{p1}$ and $V_{p2}$ to change the switching thresholds of the inverters. More specifically, a higher voltage at the gate of $T_5$, $T_7$ lowers the switching threshold of the inverters.

V. CONCLUSIONS

In this work, a novel hybrid FTJ-CMOS neuron circuit design and simulations are presented. The integration function in the neuron can be achieved by accumulative polarization switching behavior, which has been demonstrated experimentally in FTJ devices under application of identical pulses. The read current through the FTJ is amplified by the CMOS circuits and leads to the firing event. We simulated the pulse number needed for firing and its dependence on voltage pulse amplitude and width. The possibility of electrically tuning the neural dynamics by controlling the switching of the FTJ makes

the proposed neuron suitable as a fundamental building block in new-generation neuromorphic computing systems.

REFERENCES

[1] E. Covi et al. Front. in Neurosci., p. 429, 2021.
[2] E. Covi et al. in IEEE Int. Conf. Flexible and Printable Sensors and Systems (FLEPS), 2022.
[3] B. Max et al. J. Electron Devices Soc., vol. 7, pp. 1175–1181, 2019.
[4] V. Deshpande et al. Solid-State Electronics, vol. 186, p. 108054, 2021.
[5] B. Max et al. ACS Appl. Electron. Mater., vol. 2, no. 12, pp. 4023–4033, 2020.
[6] S. Lancaster et al. in IEEE ESSDERC, 2022 (Accepted).
[7] S. Lancaster et al. Appl. Phys. Lett., vol. 120, no. 2, p. 022901, 2022.
[8] S. Slesazeck, “D2.4 Provide initial FTJ SPICE model for adoption in Task 5.3.” http://www.beferrosynaptic.eu/deliverables, 2021.
[9] F. Preisach Zeitschrift für physik, vol. 94, no. 5, pp. 277–302, 1935.
[10] S. Miller et al. J. Appl. Phys., vol. 83, no. 12, pp. 6463–6471, 1999.
[11] B. Jiang et al. in IEEE Symp. VLSI Tech., pp. 141–142, 1997.
[12] S. Slesazeck et al. in IEEE Int. Memory Workshop (IMW), 2019.
[13] V. Kornjuk et al. Frontiers in Neuroscience, vol. 10, p. 212, 2016.
[14] S. Moradi et al. IEEE Trans. Biomed. Circuits Syst., vol. 12, pp. 106–122, 2017.