Supplementary Information

Prospects of Analog In-Memory Computing using Ferroelectric Tunnel Junctions

Mattias Borg[1,2], Christos Papadopoulos[1], Alec Guerin[1], Robin Athle[2], Saeed Bastani[1]

[1] Ericsson Research, Ericsson AB, 22 362 Lund, Sweden

[2] Electrical and Information Technology, Lund University, 22 100 Lund, Sweden; NanoLund: Centre for Nanoscience, Lund University, 22 100 Lund, Sweden

Access resistance

The bottom electrode is common for all FTJ devices on the chip and is accessed through a device with large area located close to the device under test (DUT), which is intentionally broken by applying a voltage beyond the breakdown voltage of the HZO film. After breaking the device acts as an ohmic resistor in series with the DUT. The access resistance R_{access} is approximately 1000 Ω, which although being negligible compared to the DUT resistance (>100 MΩ) will dominate the RC constant of the circuit. The capacitance of a DUT with radius r of 25 µm, HZO thickness of 45 Å, and $\varepsilon_r = 25$ is approximately $C = \frac{\varepsilon_0 \varepsilon_r \pi r^2}{d} = 97\ pF$,. In addition, the cables between the parameter analyzer and probes add an additional capacitance of 20 pF. All in all one can estimate the magnitude of $R_{access}C \sim$ 100 ns. Circuit simulations confirm (unsurprisingly) that this has a significant effect on the shape of voltage pulses with similar duration or shorter (Figure S1).

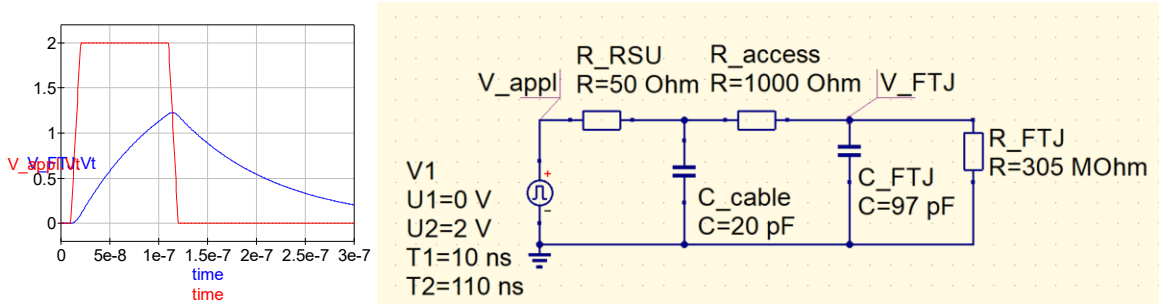

Figure S1. (left) Simulated voltage response of the measurement circuit (right), for a rectangular signal voltage pulse of 100 ns duration.

Tunneling electroresistance (TER) at V_{read} = 0.3 V

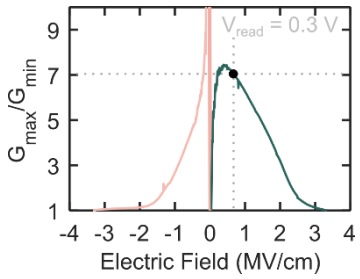

*Figure S2. G_{max}/G_{min} = TER extracted from Figure 1**d**, highlighting the maximum TER > 7 close to the chosen V_{read} = 0.3 V.*

Acquisition of PUND measurement data in Figure 3d.

The data of Figure 3d was obtained by using the PUND method. In this method a pulse train consisting of five triangular voltage pulses, as shown in Figure S3**a**:

1. A preset pulse that poles the ferroelectric film into its negative polarity state.
2. A positive (P) pulse that initiates ferroelectric switching into the positive polarity state, resulting in a polarization switching current, among other contributions.
3. Another positive pulse (U), that in theory will give the same response as for P apart from the polarization switching current, i.e. including any capacitive and leakage currents.
4. A negative (N) pulse similar to P, but initiating ferroelectric switching towards the negative polarity state again.
5. Another negative pulse (D) like (U).

The ferroelectric current density response (Figure S3**b**) is then separated from other contributions by taking the difference P-U and N-D, which is plotted as current density versus electric field, after normalizing to the area of the device and thickness of the ferroelectric film.

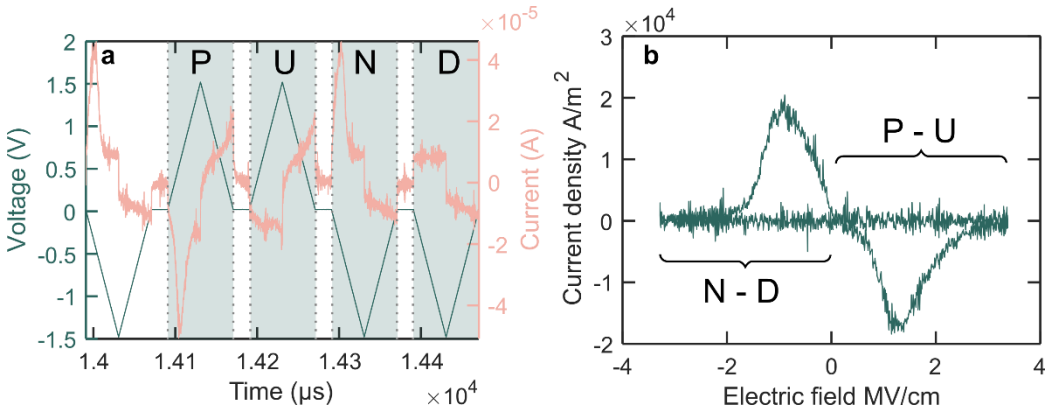

Figure S3. a) Voltage train applied in a PUND measurement, starting with a preset pulse followed by four pulses labeled P, U, N and D. Also shows the current response to the applied voltage pulses. b) Resulting graph of current density versus electric field obtained by subtracting the current response of U from P and D from N.

Table S1: Model parameters in AIHWKIT

	BALANCED	EXTREME	CONSTRICTED
Resistive Device class	`SoftBoundsDevice`	`SoftBoundsDevice`	`LinearStepDevice`
`dw_min`	0.06	0.9	0.02
`gamma`	1.15	2	N/A
`dw_min_std`	0.01	0.01	0.044

Table S2: Operation parameters for the different modes in compact model

	BALANCED	EXTREME	CONSTRICTED
Pulse duration (ns)	10	1000	0.5
Peak voltage (V)	0.9	0.9	0.9
S_{min}	0.12	0.00	0.40
S_{max}	0.66	1.00	0.58

2B-1A EXTREME RPU

A RPU consisting of three FTJ memristors operated in EXTREME mode can be used as a combination of digital and analog memory elements, forming a 2-bit + 1 analog (2B-1A) RPU. In a cross bar array, the three FTJs would be arranged on the same word line but the corresponding bit lines would be split into three parts each with a local ADC. Two of the devices would be operated as binary elements, representing the most significant bits of the weight. After the ADC the left-most signal will be bit-shifted before added to the signal from the bit line of the secondary binary devices. Again, the joined signal will be bit-shifted, before added to the third analog signal. This is described in Figure S4.

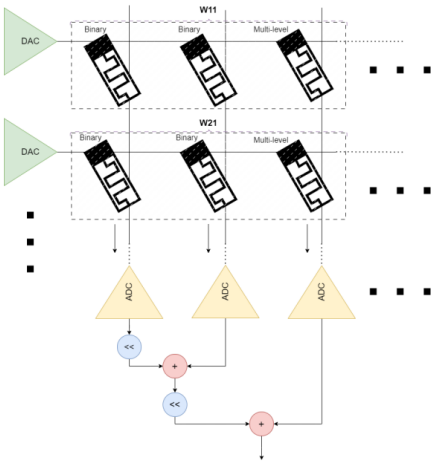

Figure S4. Schematic of a 2B-1A arrangement using three FTJ memristors per RPU.

Drift model

Conductance drift in the memristor device hardware can be very detrimental on the inference performance of a model mapped to hardware. The weights lose their initial values and as a result the activation of neurons is not proper any more. In the case of the FTJ device, experimental data were used to model the drift effect, as shown in Figure S4. The conductance evolution is empirically modeled as:

$$g_{drift} = g_{progr}\left(\frac{t}{t_c}\right)^{-v} \qquad (1)$$

Where v is a drift coefficient and tc is the time since the weight was programmed. This equation was evaluated for a range of programmed conductance states covering the whole dynamic range of an FTJ device (diameter 50 μm). As is clear from Figure S5**a** the drift is almost negligible and extracted drift coefficients based on fitting of Eq. 1 to the measurement data is shown in Figure S5**b** as function of the target conductance values, with an average drift coefficient $|\langle v \rangle| = 2 \times 10^{-5}$.

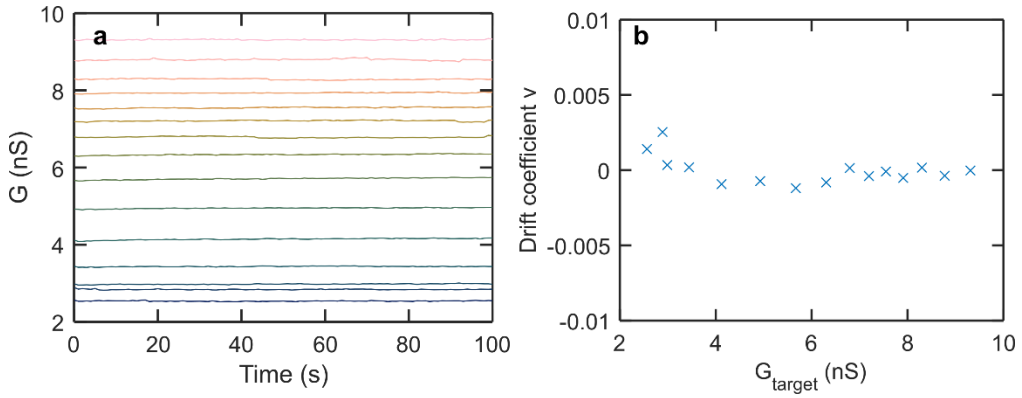

Figure S5. a) Retention of programmed states throughout the dynamic range over 100s. b) Fitted drift coefficients as function of the target conductance.

IR drop and conductance drift

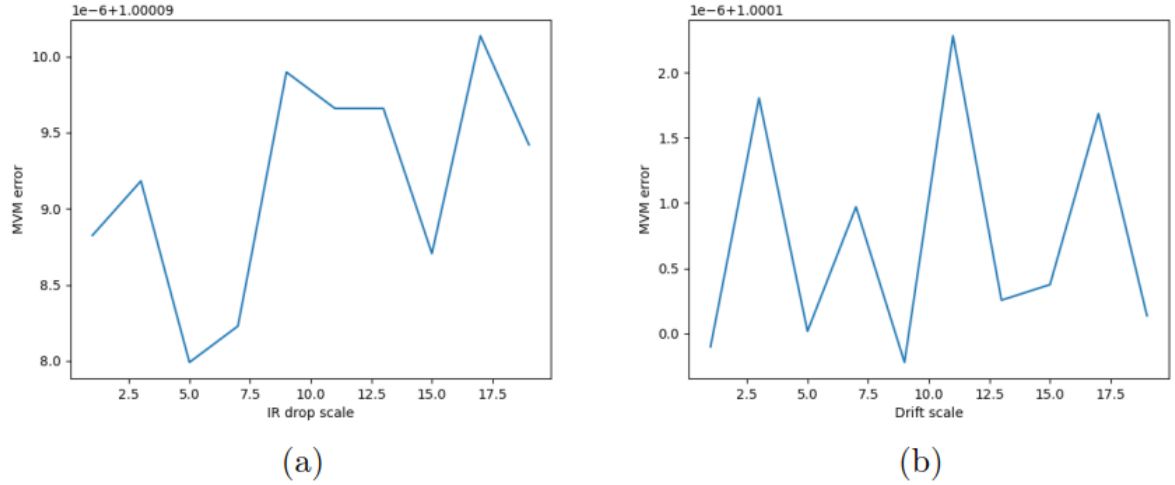

Figure S6. Effect of (a) IR drop scaling, and (b) Drift magnitude, on the error of MVM operations using a BALANCED mode FTJ RPU. The x-axis indicates the multiple of parameter scaling from the baseline value.