A Current Attenuator for Efficient Memristive Crossbars Read-Out

Charanraj Mohan, José M. de la Rosa, Elisa Vianello, Luca Perniola, Carlo Reita,
Bernabé Linares-Barranco and Teresa Serrano-Gotarredona

Abstract—**This paper presents a new current attenuator circuit to scale down the inference currents in memristor based crossbars that drive integrate-and-fire neurons, which subsequently allows to reduce the size of integrating capacitors by several orders of magnitude, making IC integration possible. The proposed circuit uses a linear switch to divide the inference current and scale it down by a factor of about 10^4. The proposed attenuator has been designed in 130nm CMOS technology. Simulation results considering noise, process and temperature variations are shown to validate the presented approach.**

I. Introduction

Memristors prompted an increasing interest among the neuromorphic community since these devices were successfully demonstrated in HP labs [1] as an experimental evidence of the circuit element predicted by Chua in 1971 [2], [3]. Since then, there have been a number of memristive devices developed based on diverse materials and technologies. Among others, Oxide-based Random-Access Memory (OxRAM) structures are one of the most promising approaches to implement synaptic elements in neuromorphic computing systems, because of their low-switching energy and high endurance [4]–[7].

Although OxRAMs have also been demonstrated to operate as analog memories [8]–[10], here we focus on their exploitation as binary (1-bit) memory cells. The OxRAM devices to be used in this work are optimized for binary memory applications [11]. This binary OxRAM (as highlighted in Fig. 1) has to be first formed and then switched between two states: SET or Low Resistance State (LRS)– typically in the order of kΩ – and RESET or High Resistance State (HRS)– in the order of few hundreds of kΩ to MΩ. The resistance states are chosen by applying a control voltage to a selector MOSFET connected in series with the OxRAM in order to avoid sneak path currents, giving rise to a device often referred to as 1T1R, as shown in Fig. 1 [12]. In order to form the filament of the OxRAM, the selector MOSFET is biased with a high positive voltage (≈4V) at the top terminal, the bottom terminal is biased with 0V and the gate terminal voltage (1.5 V to 2.5 V) is adjusted such that the recommended compliance forming current of about 1μA is set. For a RESET operation, the top terminal is biased with 0V, the bottom terminal is biased with a positive voltage (≈4V) and the gate terminal is kept fully ON. For a SET operation the top terminal is biased with a positive voltage (≈2V), the bottom terminal is biased with 0V and the gate is set to limit the current to about $100\mu A$.

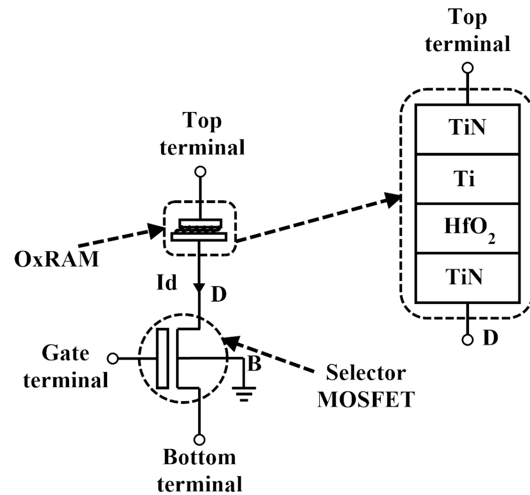

Fig. 1. 1T1R memristive synapse structure

For a READ operation, the top terminal is biased with a small positive voltage (0.2 or 0.3V), the bottom terminal is biased with 0V and the gate is kept fully ON.

Here we consider a memristor-based fully connected feed-forward neural network, which mainly comprises the pre-synaptic neurons, the memristive crossbar and the post-synaptic neurons [13]. As the LRS current across a crossbar line is high due to the typical low resistance during an inference read operation (after a SET operation), an extremely large integrating capacitor would be needed (larger than nFs) at the post-synaptic neuron for reasonable integration speed, making IC integration impossible. Hence, a current attenuator is needed to scale down the read current.

Several current attenuating strategies exist such as the Gilbert's current normalizer circuit [14], MOS-ladder [15] and the Winner Take All (WTA) [16] based current attenuator [17]. This paper considers Gilbert's current normalizer circuit as reference, on which modifications are done to ease attenuation of an inference current. The idea is based on creating a splitting of the inference current by a factor of about two by using a MOS biased in ohmic region. This paper proposes a Modified Current Normalizer (MCN) circuit for attenuating a crossbar read-out line current. Simulation results taking into account the effect of PVT variations are shown to validate the

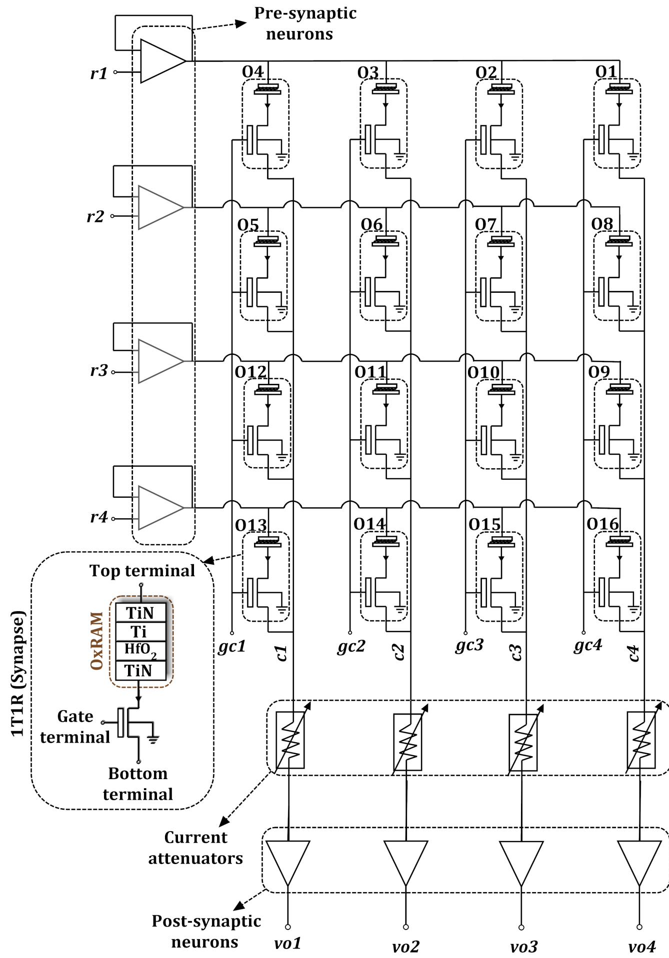

Fig. 2. Scheme of a 4×4 1T1R crossbar with pre-synaptic neurons, current attenuators and post-synaptic neurons.

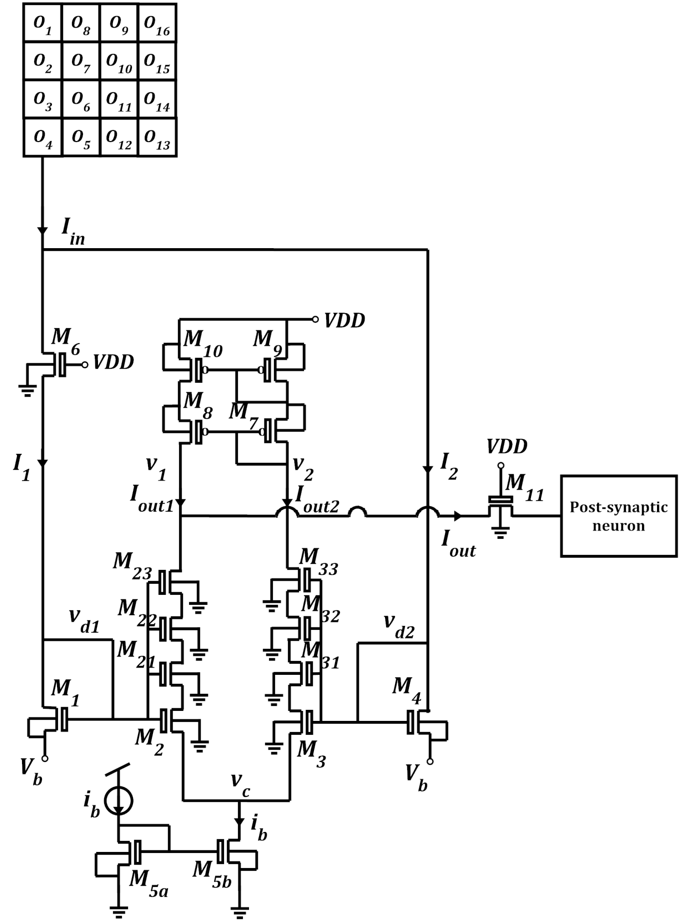

Fig. 3. Details of MCN circuit schematics

proposed circuit technique.

II. 1T1R crossbar with pre-synaptic neurons, current attenuators and post-synaptic neurons

Fig. 2 shows a fully connected 4×4 feed-forward neural network with pre-synaptic neurons, current attenuators and post-synaptic neurons. Circuit elements labeled as $O_{1,2,3,....,16}$ comprise the 1T1R based memristor-selector synaptic devices. Each row has a pre-synaptic neuron, which is made up of a pulse-shaping digital block and an opamp that is finely calibrated [18]. Each column is connected to a current attenuator circuit, followed by a post-synaptic neuron. The post-synaptic neuron comprises a CMOS integrate-and-fire neuron, which integrates the inference current and generates an output spike when a threshold is reached. The OxRAM devices are initially formed to bring them from the Pristine Resistance State (PRS). Later, they can be switched in binary mode, i.e. either SET or RESET [19]. For reading the memristance of the OxRAMs, small read voltages of 0.2 or 0.3 V are applied across the rows, so that the aggregated inference read currents across columns are scaled down and integrated.

III. Proposed MCN circuit

We propose an MCN based attenuator, which mainly includes having an MOS-resistor and a current-mirror at the load on the Gilbert's two-input current normalizer circuit [14]. Fig. 3 shows the proposed MCN circuit connected between a column of a the crossbar and a post-synaptic neuron. It works by inserting an ohmic biased MOS (M_6) which creates a current splitting of the crossbar column current I_{in} by factor of about two. This MOS-resistor creates a small resistance R_{M_6} which is given by

$$R_{M_6} = \frac{1}{\mu_n \cdot C_{ox} \cdot (V_{GS} - V_T)} \cdot \frac{L_{M_6}}{W_{M_6}} \tag{1}$$

where μ_n is the charge-carrier effective mobility, C_{ox} is the gate oxide capacitance, V_{GS} is the gate-to-source voltage, V_T is the threshold voltage, L_{M_6} and W_{M_6} are the length and width of the MOS-resistor. From Fig. 3, and knowing transistors M_1 and M_4 will be biased in strong inversion because of the large input current I_{in}

$$\Delta V_d = v_{d2} - v_{d1} = I_1 R_{M_6} \quad (2)$$

$$I_i = \frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}(v_{di} - V_b - V_T)^2 \quad (3)$$

which results in

$$\frac{I_1}{I_2} = (1 - \frac{\Delta V_d}{v_{d2} - V_b - V_T})^2 = (1 - \frac{\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}} I_1 R_{M_6}}{\sqrt{I_2}})^2 \quad (4$$

I_1 and I_2 are the input currents shown in fig.3. MOSFETs M_1 and M_4 have same size. If we can assume that $R_{M_6} \ll \sqrt{I_2}/(\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}} I_1)$, then

$$I_1 \simeq I_2 \simeq \frac{I_{in}}{2} \qquad with \qquad R_{M_6} \ll \frac{1}{\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}}} \sqrt{\frac{2}{I_{in}}} \quad (5)$$

From eq. (2) we obtain

$$\Delta V_d \simeq \frac{I_{in} R_{M_6}}{2} \quad (6)$$

The differential pair M_2,M_3 in Fig. 3 will be biased in weak inversion, because i_b is intentionally made very small. Consequently,

$$I_{outi} = \frac{2n \cdot \mu_n \cdot C_{ox} \cdot U_T^2 \cdot W_{2,3}}{L_{2,3}} \cdot e^{\frac{v_{di} - v_c}{nU_T}} \quad (7)$$

$$i_b = I_{out1} + I_{out2} \quad (8)$$

where n is the subthreshold slope factor and U_T is thermal voltage. From here

$$\frac{I_{out2}}{I_{out1}} = e^{\frac{\Delta V_d}{nU_T}} = e^x \quad (9)$$

where, from eq. (6) $x = (I_{in} R_{M_6})/(2nU_T)$. Straight forward calculations yield

$$I_{out2} - I_{out1} = i_b \frac{e^x - 1}{e^x + 1} \quad (10)$$

If we can assume that $x \ll 1$ (which is equivalent to $R_{M_6} \ll 2nU_T/I_{in}$), then

$$I_{out} = I_{out2} - I_{out1} \simeq i_b \frac{x}{2} = \frac{i_b R_{M_6}}{4nU_T} I_{in} \quad (11)$$

Therefore, the MCN circuit output current is, under some assumptions for R_{M_6}, proportional to I_{in}, and the proportionality factor can be controlled by bias current i_b. This will allow us to scale down the input current by four orders of magnitude.

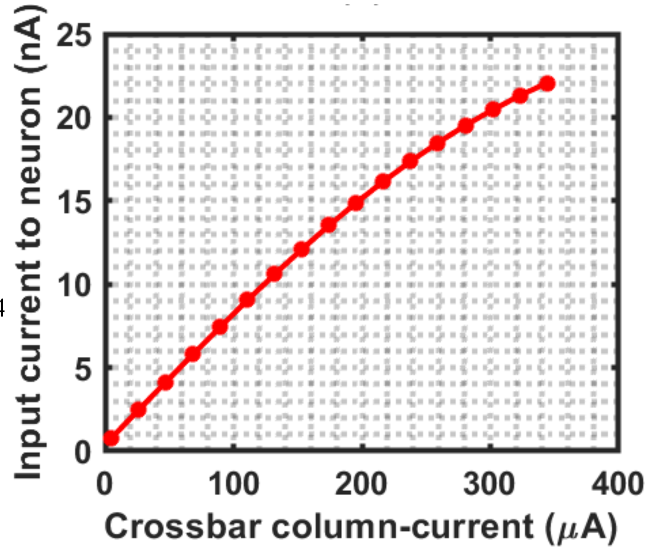

Fig. 4. Minimum to maximum crossbar column inference current Vs Input current to neuron.

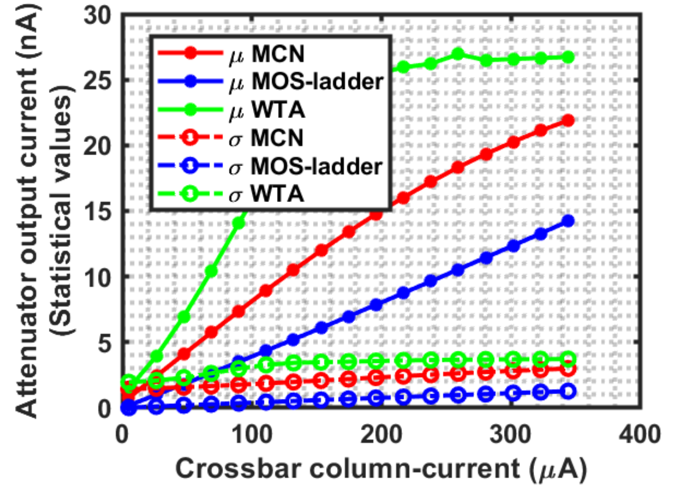

Fig. 5. Comparison of average and standard deviation of output current for different attenuators: MCN circuit, MOS-ladder circuit and WTA circuit considering process and mismatch variations with 100 Monte Carlo runs.

Transistor M_{11} is an optional switch, useful to isolate the post-synaptic neuron for test purposes. Differential MOSFET groups $M_{2,21,22,23}$ and $M_{3,31,32,33}$ are splitted along the length dimension in order to keep them in square shape and exploit inter-digitated layout to minimize mismatch due to gradients.

IV. Simulation Results of the 1T1R Crossbar

This Section shows various simulation results of the proposed approach in comparison with other approaches (MOS-ladder and WTA type circuit) considering mismatch-and-process variations, temperature and input referred noise. The MCN circuit is implemented using a 130nm CMOS technology and its sizing and biasing is shown in Table I. Fig. 4 shows how the output current of the MCN circuit depends on the input current I_{in}. For the 1T1R memristive devices

TABLE I
SIZING AND BIASING OF PROPOSED CIRCUIT, DESIGNED FOR 130NM CMOS

Parameter	Value
Supply voltage	$VDD = 4.8$ V
Size of $M_{1,4}$	$(\frac{W}{L})_{1,4} = (\frac{10\mu m}{1\mu m})$
Size of $M_{2,3}$	$(\frac{W}{L})_{2,3} = (\frac{1\mu m}{1\mu m})$
Size of $M_{21,22,23}$	$(\frac{W}{L})_{21,22,23} = (\frac{1\mu m}{1\mu m})$
Size of $M_{31,32,33}$	$(\frac{W}{L})_{31,32,33} = (\frac{1\mu m}{1\mu m})$
Size of $M_{5,7,8}$	$(\frac{W}{L})_{5,7,8} = (\frac{1\mu m}{1\mu m})$
Size of M_6	$(\frac{W}{L})_{6} = (\frac{8\mu m}{1\mu m})$
Size of $M_{9,10,11}$	$(\frac{W}{L})_{9,10,11} = (\frac{1\mu m}{1\mu m})$
V_b	*0.5 V*
i_b	*25 nA*

we use here and with a read voltage of 0.3V, typical LRS is 13.9 kΩ whose inference read current is 21.5 μA, whereas typical HRS is 1MΩ whose inference read current results in 333nA. Minimum column current is when all the OxRAMs in a column are in HRS and maximum column current is when all the OxRAMs in a column are in LRS. Fig. 4 shows that the MCN output current (input current to neuron) is linear when the inference current stays below about 300μA.

An optimal design procedure is carried out on the MCN circuit to keep mismatch as low as possible. Fig. 5 shows the statistical output current results (both mean-μ and sigma-σ) for 100 Monte Carlo runs of the MCN circuit, a MOS-ladder and a WTA type circuit considering both process and mismatch variations.

Fig. 6 shows how different column inference currents (minimum to maximum) vary due to temperature. We can see that temperature variations have fairly small impact on the output currents (less than 10%).

The input-referred noise for the frequency range 0.01 Hz to 100 MHz is 3.178 nA/$\sqrt{Hz}$ for the MCN circuit, 2.06 nA/$\sqrt{Hz}$ for the MOS-ladder and 1.02 nA/$\sqrt{Hz}$ for the WTA type circuit. Comparison of both area and input referred noise for different attenuators is shown in Fig. 7.

V. CONCLUSIONS

A new current attenuation circuit for OxRAM based crossbar applications has been proposed. The proposed technique allows to attenuate inference currents by several orders of magnitude before sending them to the post-synaptic neurons. Simulation results are provided based on a 130nm CMOS technology, with a scale-down factor of about 10^4.

VI. ACKNOWLEDGEMENTS

This work has been supported in part by the EU H2020 grants 687299 NeuRAM[3], 824164 HERMES, by the Spanish Ministry of Economy and Competitiveness (with support from the European RDF) under contracts TEC2015-63884-C2-1-P (COGNET) and by Lloyds Register Foundation's ICON.

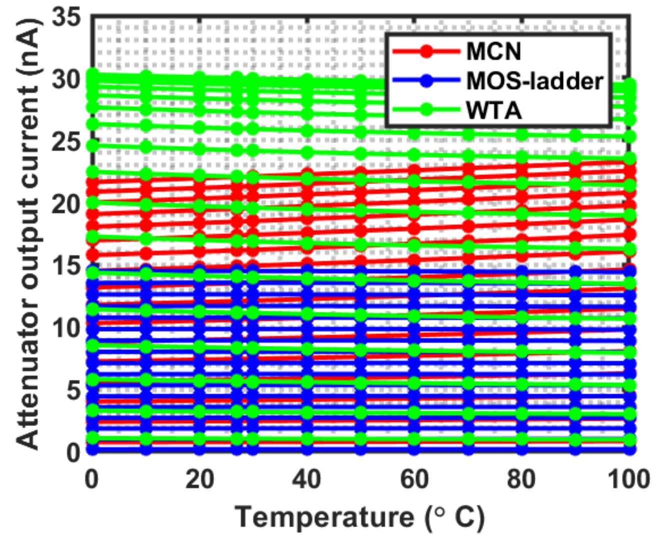

Fig. 6. Temperature variations of output current of MCN circuit, MOS-ladder and WTA circuit for different values of output currents.

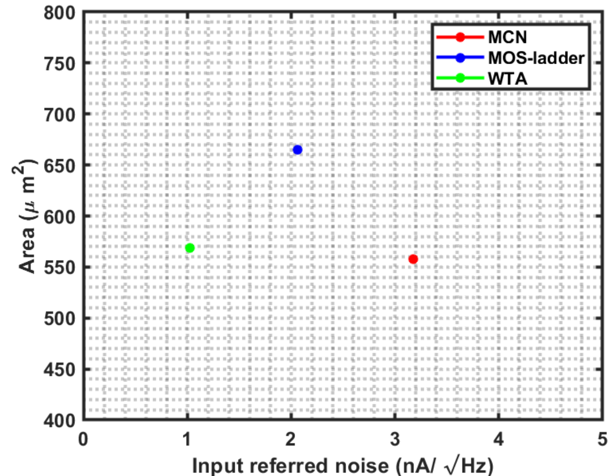

Fig. 7. Area and input referred noise of MCN circuit, MOS-ladder and WTA circuit

REFERENCES

[1] D. B. Strukov, Gregory S. Snider, Duncan R. Stewart and R. Stanley William, "The missing memristor found", *Nature*, 453 (7191), pp. 80-83, 2008.

[2] L.O. Chua, "Memristor- The missing circuit element", *Transactions on Circuits theory IEEE*, vol. CT-18, no. 5, pp. 507-509, Sept. 1971.

[3] L. O. Chua and S. Kang, "Memristive devices and systems", *Proceedings of the IEEE*, vol. 64, no. 2, pp. 209-223, 1976.

[4] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart and R. S. Williams,"Memristive switching mechanism for metal/oxide/metal nanodevices", *Nat. Nanotechnology*.,3, 429, 2008.

[5] R. Waser, R. Dittmann, G. Staikov and K. Szot,"Redox-based resistive switching memories-nanoionic mechanism, prospects, and challenges", *Adv. Mater.*, 21 (25-26), pp. 2632-2663, July 2009.

[6] S. Yu, X. Guan and H. P. Wong,"Conduction mechanism of TiN/HfOx/Pt resistive switching memory: a trap-assisted-tunneling model", *Appl. Phys. Lett.*.,99, 063507-063507, 2011.

[7] S. D. Ha and S. Ramanathan,"Adaptive oxide electronics: a review", *J.Appl.Physics*., 110, 071101, 2011.

[8] J. Frascaroli, S. Brivio, E. Covi, and S. Spiga, "Evidence of soft bound behaviour in analogue memristive devices for neuromorphic computing," *Scientific Report*, vol. 8, Page 7178, 2018.

[9] X. Guo, F. Merrikh-Bayat, L. Gao, B.D. Hoskins, F. Alibart, B. Linares-Barranco, L. Theogarajan, C. Teuscher, and D.B. Strukov, "Modeling and Experimental Demonstration of a Hopfield Network Analog-to-Digital Converter with Hybrid CMOS/Memristor Circuits," *Frontiers in Neuromorphic Engineering. Front. Neurosci.* 9:488, 2015.

[10] S. Stathopoulos, A. Khiat, A. Serb, and T. Prodromakis, "Benchmarking Analogue Performance of Emerging Random Access Memory Technologies," *Proc. of the 2018 IEEE Int. Symp. on Circ. and Syst.* (ISCAS2018), Florence, Italy, 2018.

[11] D. Garbin, E. Vianello, O. Bicher , Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo and L. Perniola,"HfO_2- Based OxRAM Devices as Synapses for Convolutional Neural Networks", *IEEE Transactions on Electron Devices*.,vol. 62, no.8, pp. 2494-2501, August 2015.

[12] Y. Cassuto, S. Kvatinsky and E. Yaakobi,"Sneak-path constraints in memristor crossbar arrays", *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*.,10.1109/ISIT.2013.6620207, October 2013.

[13] C. Zamarreño-Ramos, L. A. Camuñas-Mesa, J. A. Pérez-Carraso, T. Masquelier, T. Serrano-Gotarredona and B. Linares-Barranco,"On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex", *Front Neurosci.*, 17 March 2011.

[14] B. Gilbert,"Current-Mode circuits from a Translinear Viewpoint: A Tutorial", *In: Analogue IC design: the current-mode approach*, Ed. by C. Tomazou, F.J.Lidgey and D.G. Haigh. Stevenage, Herts., UK: Peregrinus, 1990. Chap. 2, pp. 11-91.

[15] K. Bult and J.G.M. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1730-1735, Dec. 1992.

[16] J. Lazzaro, S. Ryckebusch, M. A. Mahowald and C. A. Mead,"Winner-take-all networks of O(N) complexity", *in Advances in Neural Information Processing Systems 1*, Morgan Kaufmann Publishers, San Francisco, CA, 1989.

[17] Manu V. Nair, Lorenz K. Muller, and Giacomo Indiveri,"A differential memristive synapse circuit for on-line learning in neuromorphic computing systems", *Nano Futuresv*, vol. 1, no. 3, 2017.

[18] C. Mohan, L. Camuñas-Mesa, E. Vianello, L. Periniolla, C. Reita, J. M. de la Rosa, T. Serrano-Gotarredona and B. Linares-Barranco, "Calibration of offset via bulk for low-power HfO_2 based 1T1R memristive crossbar read-out system", *Microelectronic Engineering, Elsevier*, vol 198, pp 35-47, 15 October 2018.

[19] D. Garbin *et al.*, "Variability-tolerant convolutional neural network for pattern recognition applications based on OxRAM synapses", *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, pp 28.4.1-28.4.4, December 2014.