Energy Efficient Analog Circuits for Neuromorphic Computing

Sagar Kumashi, Siddharth Kumar Singh, Vikash Sehwag, Mrigank Sharad

***Abstract* — Emerging memristive -crossbar memory technology can be promising for computationally-expensive analog pattern-matching tasks. It provides a new computing paradigm, enabling calculations to be performed in the same chips where data is stored. However, degradation in the performance of the RCM is observed due to parasitic and terminal resistances, which we have demonstrated in this paper. A novel current mode circuit has been proposed to effectively tackle this problem. The proposed circuit is more power efficient in comparison to its voltage mode counterparts.**

I. Introduction

The ever-increasing demand for higher computing capabilities has necessitated the integration of multiple processing cores and larger memory blocks. This in turn has resulted in increasingly busy inter chip links and complex power-hungry input/output (I/O) interfaces for microprocessors. The same is true for on-chip global interconnects such as multi-byte buses, long-distance inter-block links and connection networks for on-chip memory read [1]. With the continued scaling of CMOS technology, the energy efficiency and performance of on-chip global interconnects have degraded because of an increase in the per-unit-length resistance of long metal lines [1]. Therefore, the design of inter chip and on-chip global interconnects has emerged as one of the major challenges in the field of high-speed computing systems.

While solving this conundrum is an open-ended problem, one promising solution appears to be the use of artificial neural networks (ANNs). Inspired by biological networks of our nervous system, artificial neural networks possess the unique trait of housing memory and processing elements in the same block, just like a biological neuron. This altogether eliminates the overhead due to memory read. Moreover, ANNs also inherit other useful features seen in their biological counterparts such as parallelism, generalization, nonlinearity and the ability to decipher imprecise data. Consequentially, ANNs/neuromorphic computing have/has found applications in various fields which appreciate the above mentioned attributes such as pattern matching, predictions and multi variable optimization [1]. Such applications employ highly memory-intensive computing involving the correlation of multidimensional input data with a large number of stored patterns/templates in order to find the best match.

Sagar Kumashi is with the Indian Institute of Technology Kharagpur, Kharagpur, West Bengal – 721302, India.

Siddharth Kumar Singh is with the Indian Institute of Technology Kharagpur, Kharagpur, West Bengal – 721302, India.

Mrigank Sharad is with the Indian Institute of Technology Kharagpur, Kharagpur, West Bengal – 721302, India.

Vikash Sehwag is with Princetion University, Princeton, NJ 08544, USA

In this paper we have discussed the implementation of analog dot product operation for applications such as image recognition by making use of a Resistive Crossbar Memory (RCM) architecture with memristors as building blocks. Thus far designs making use of spintronic devices [2] and analog CMOS circuitry [3] [4] have been proposed. However, the use of multiple analog blocks for large scale RCM leads to power hungry designs due to large static power consumption of such circuits. Memristors act as a strong candidate for replacing existing memory elements because of their non-volatile property and high packing density in a crossbar array. With the appropriate read and write drivers, they can store multilevel values and can achieve higher densities and speeds than static CMOS RAMs. Furthermore, they do not require any energy to maintain the stored data.

The second part of the paper introduces a novel neuron circuit employing current mode processing. Conventionally, neurons have been implemented using analog operational amplifiers where processing is done in voltage mode. The power consumption for such circuits is quite high for large scale RCM because of the involvement of multiple op-amp blocks. This can possibly negate the energy benefits we wish to reap from the RCM. With current mode processing, one can work with lower values of voltages, as has been shown in previous works [5]. This presents the possibility of achieving significant reduction in power consumption.

The paper is organized in the following manner: Section II of the paper discusses the RCM based implementation of dot product operation with memristors functioning as memory blocks. The effect of the input resistance of the neuron on the efficiency of the RCM has been described. Section III provides background on current mode processing and the various advantages it has over voltage mode processing. Section IV describes the proposed neuron design and discusses modifications to the circuit to obtain different transfer characteristics. Section V shows the various results for the RCM simulation as well as the neuron circuit plots.

II. RCM Based Computation

Memristor technology allows a high packing density of storage cells over a simple crossbar structure, which can effectively be modelled as a synapse in a neuron. Fig. 1 shows an RCM network. In a resistive crossbar memory, memristors (usually Ag-Si) having a conductance g_m connects two metal tracks, where one track is vertical while the other is horizontal. The metal tracks themselves contribute to parasitics which can be modelled as a combination of a parasitic resistance R_p between the two nodes where the memristor meets the tracks and parasitic capacitance C_p between each node and ground. The terminal ends of the in-lane tracks are connected to neurons, each having an input resistance of R_t.

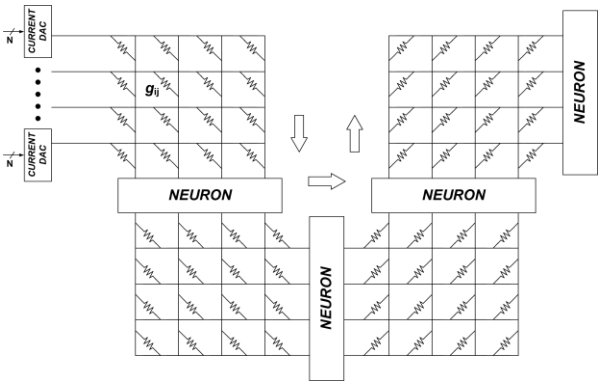

Fig. 1. A Resistive Crossbar Network

If V_i is the input voltage being fed to the i^{th} horizontal track, the current coming out of the j_{th} vertical track can be visualized as the dot product of the input V_i*'s* and the cross-bar conductance values g_{ij}. To make the above operation feasible, the conductance of the last column should be programmed to give us the same total conductance along every row. In this manner, an RCM can directly evaluate correlation between an analog input vector and a number of stored patterns [6].

Due to the parasitic and terminal resistances, one might observe significant deviation in the result of the correlation operation described above. WithIt is desired to achieve the minimum possible terminal resistance such that the terminal voltage will be approximately equal to ground voltage. The effect of each parameter is further discussed on Sec V.

III. Current Mode Processing Scheme

Current mode processing refers to signal transformations in which the computations are carried out in current domain. One of the primary advantages that current mode circuits have over their voltage mode counterparts is the smaller delay that they are capable of providing. This can be inferred from the expression for charging/discharging delay shown below

$$Delay = \frac{C\Delta V_{swing}}{\Delta I_{ch/dch}} \quad (1)$$

Where $\Delta I_{ch/dch}$ denote the charging/discharging current. In voltage mode circuits, one has to reduce ΔV_{swing} in order to shorten the delay. However, the extent to which it can be reduced is limited by the signal to noise ratio requirements of the system. In current mode circuits on the other hand, one reduces the delay by increasing I_{ch}/I_{dch}. These parameters aren't curtailed by any restrictions apart from limits imposed to prevent transistor breakdown, which they rarely exceed. Hence the extent to which delays can be reduced is greater for current mode circuits than voltage mode circuits.

Apart from higher speeds, current mode circuits also offer greater scalability with respect to supply voltage. This follows from the fact that all operations in current mode processing are in current domain and hence aren't affected as much by supply voltage reduction as voltage mode circuits. As a consequence of this, one can expect current mode circuits to also have a lower power consumption too.

IV. Circuit Implementation

A. Sigmoidal Transfer Function

The schematic for the proposed neuron circuit having a sigmoidal transfer function is shown in Fig. 2. The circuit can be divided into two stages: A low impedance current to voltage converter stage followed by a differential amplifier stage.

In the current to voltage converter stage, a feedback path was added between the gate and drain of transistor M_1 in order to reduce the input impedance seen by the RCM. The amplifier A_1 introduced in the feedback path is a pmos-input single stage differential amplifier with moderate gain and high bandwidth. One can compute the effective impedance seen by the RCM due to this stage as

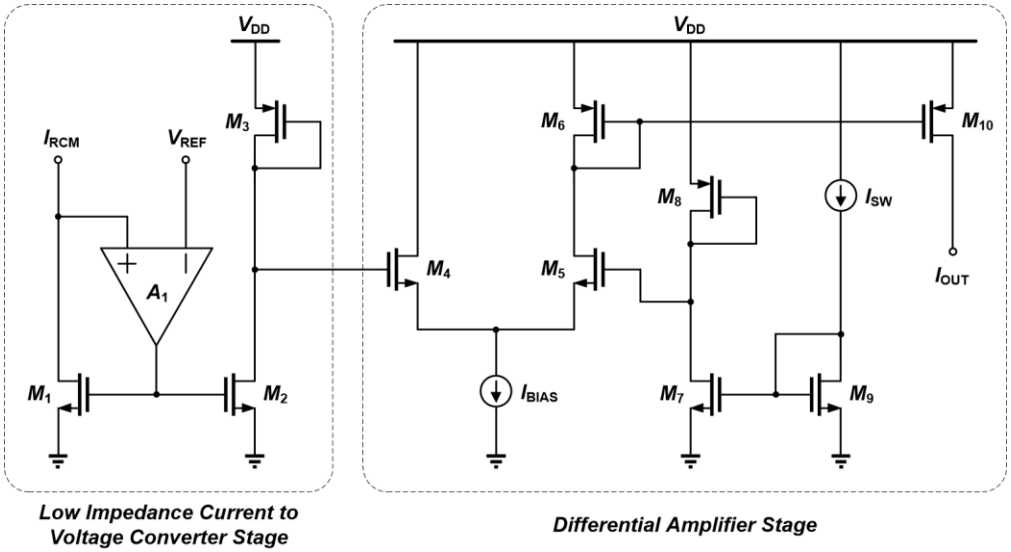

Fig. 2. The Proposed Neuron Circuit

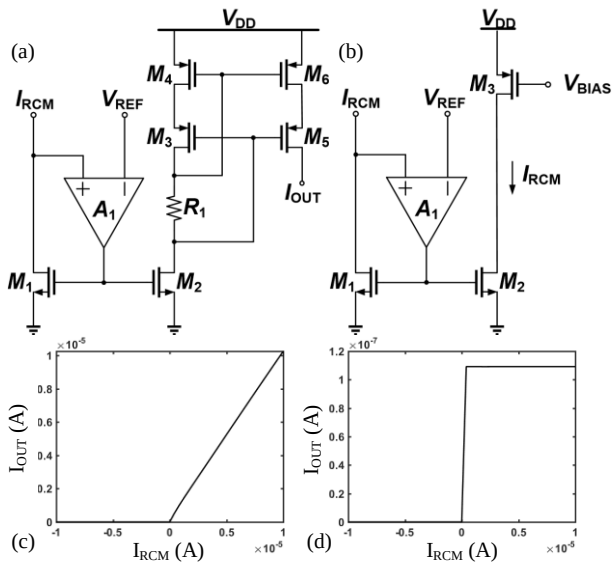

Fig 3. (a) Linear Neuron Circuit (b) Step Neuron Circuit (c) Transfer Characteristics for Linear Neuron Circuit (d) Transfer Characteristics for Step Neuron Circuit

$$r_{in} = \frac{r_{ds}}{1 + g_m r_{ds} A} \tag{2}$$

Where r_{ds} is the small signal resistance of the transistor M_1, g_m is the transconductance of M_1 and A is the gain of the amplifier A_1.

The current flowing into transistor M_1 is mirrored to M_2 which is then converted to voltage by the transistor M_3 by virtue of its diode connection. This voltage is fed to one of the inputs of the differential amplifier stage shown. The other input is at a reference voltage generated from the switching threshold current I_{SW} using the diode connected transistor M_8. The benefit of using the differential amplifier stage for the voltage to current conversion is that it's capable of providing output current saturation for high I_{RCM}, hence giving us the desired sigmoidal transfer characteristics. We can model the sigmoidal curve obtained by the function

$$I_{out} = \frac{I_{bias}}{1 + e^{\kappa(I_{in} - I_\theta)}} \tag{3}$$

Where I_{in} and I_{out} are the input and output currents of the neuron respectively, I_{bias} is the bias current of the differential amplifier stage and κ and I_θ are the parameters of the sigmoidal function. The parameter I_θ corresponds to the value of the input current at which I_{out} reaches half of its maximum value. Since the output switches from low to high for this value of the input current, one can correlate the parameter I_θ and the switching current I_{SW}. In fact the differential pair in the differential amplifier stage implements the $\kappa(I_{in} - I_\theta)$ term in the function. The parameter κ is proportional to the slope of the sigmoid curve at the trip point. This parameter depends on the transconductances of the differential pair transistors M_4 and M_5 as well as the transconductances of transistors M_3 and M_8, which perform the current to voltage operation.

Hence one can obtain the desired sigmoid curve by appropriately adjusting the values of I_{SW} and g_{M4}, g_{M5} of Fig. 2.

B. Alternate Transfer Functions

One can obtain other transfer functions such as linear and step functions by cascading the low impedance current mirror part of the sigmoidal neuron with a second circuit block. The circuit diagrams showing the neuron circuits for linear and step activation functions are shown in Fig. 3. The linear neuron circuit uses a reduced swing cascode current mirror to obtain an I_{OUT} which varies linearly with I_{RCM}. The sizing of transistors pairs M_3-M_5 and M_4-M_6 decides the slope of the function. For the step neuron circuit, the constant bias provided to M_3 shown in Fig. 3(b) helps provide a sharp step like transition. Figures 3(c) and 3(d) show the transfer characteristics of the two circuits.

V. Results and Discussion

A. RCM Modelling Results

The effect of parasitic and terminal resistance on various performance metrics is shown in Fig. 4. In Fig. 4(a), we observe that the bandwidth of the RCM network degrades with increase in the terminal resistance. Figures 4(b) and 4(d) show the deterioration in the accuracy of the dot product operation with increasing g_m (inverse of memristance) and decreasing g_t (inverse of terminal resistance). The effect of increasing terminal resistance on the energy per computation is shown in Fig. 4(c). We observe that beyond a critical R_T, the energy consumption increases linearly with terminal resistance i.e. the input impedance of the neuron circuit. Hence, it's crucial that this is kept as low as possible.

B. Circuit Simulation Results

The transfer characteristics of the sigmoidal neuron circuit is shown in Fig. 5(a). We observe that at input current corresponding to I_{SW}, the value of output current is half of I_{bias}, and hence is equal to I_θ. The variation of the slope of the sigmoid with change in the widths of the input transistors of the differential amplifier stage is shown in Fig. 5(b). We observe an increase in the slope of the sigmoid with increasing width of the differential transistor pair, hence affirming the correlation between the transconductance of the transistor and the slope of the sigmoid. Fig. 5(c) shows the shift in the switching threshold with change in the

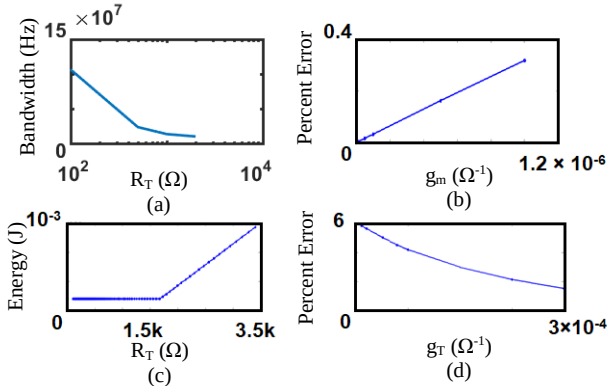

Fig. 4. (a) Bandwidth of RCM vs R_T (b) Percent Error in Correlation vs g_m (c) Energy per computation vs R_T (d)Percent Error vs g_T

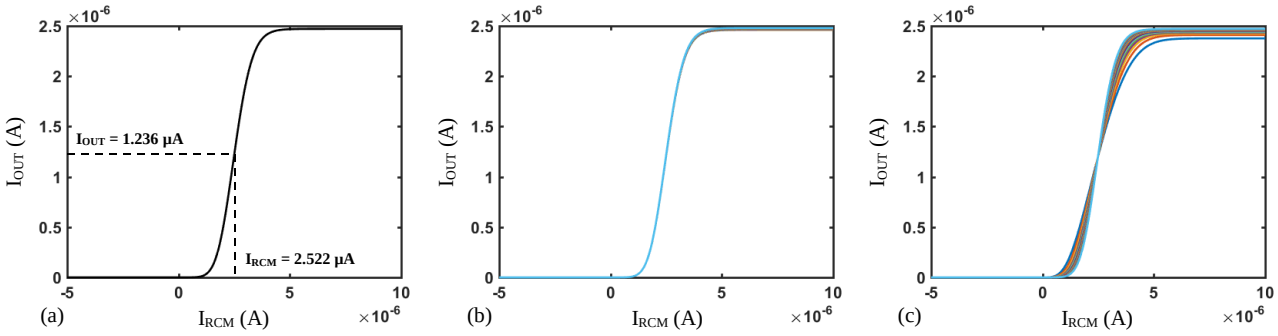

Fig. 5. Transfer Characteristics of the Circuit (a) For Standard Parameters (b)Variation with threshold Voltage (c) Variation with transconductance

threshold voltage of the differential transistor pair. We also observe slight variation in the output current saturation levels.

To observe the deviation in the transfer characteristics of the neuron due to process variation, we ran Monte-Carlo simulation on the value of output current I_{OUT} for different values of I_{RCM} to obtain the variation in the sigmoid parameters κ and I_θ. The mean and standard deviation of the two parameters is listed in Table 1 shown below.

TABLE I. MEAN AND STANDARD DEVIATION OF K AND I_θ

Parameter	*Mean*	*Std Deviation*
Slope Coefficient (κ)	-2188800 A^{-1}	20380.074 A^{-1}
Switching Current (I_θ)	1.216 µA	2.694 nA

The power consumption and bandwidth of the proposed design was compared to a conventional voltage mode neuron having a bandwidth of the same order (shown in Fig. 6) and the results have been tabulated in Table 2. The proposed design has more than 10 times lower power consumption.

TABLE II. POWER CONSUMPTION AND BANDWIDTH COMPARISON

Neuron Design	*Power*	*Bandwidth*
Voltage Mode Circuit	432 µW	3.82 MHz
Proposed Circuit	37.8 µW	11.21 MHz

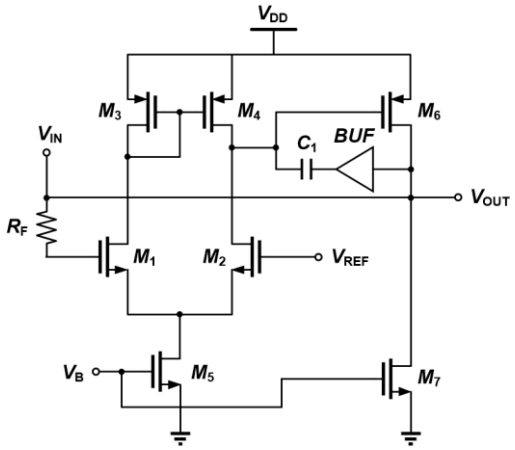

Fig. 6. Voltage mode neuron

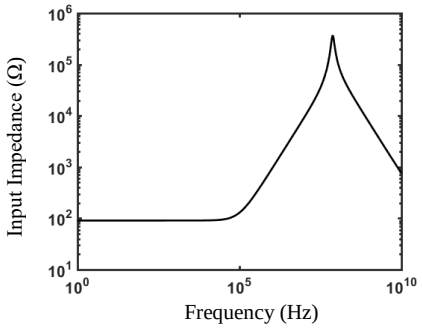

Fig. 7. Frequency variation of the input impedance of the neuron

The frequency response of the input impedance is shown in Fig. 7. We observe a spike in the Bode plot of the input impedance at around 100 Mhz, which is indicative of a zero at that frequency. In [7], it has been shown that the location of this zero depends on the transconductances of the input PMOS and the load NMOS devices in the single stage amplifier A_1 as well as the gate-to-source and gate-to-drain capacitances of the input PMOS.

VI. CONCLUSION

In this paper we have shown the degradation in the performance of the RCM for dot product operations due to the parasitic resistances introduced by the metal tracks which are part of the RCM, as well as the terminal resistance appearing due to the neuron circuit. In general, the percentage error of such computations increases with increasing g_m (decreasing memristance) and decreasing g_T (increasing terminal resistance). To combat this degradation, we proposed a current mode neuron circuit with a low input impedance. The neuron circuit also showed a lower value of power consumption when compared to voltage mode neuron circuits for comparable bandwidths.

REFERENCES

[1] J. D. Owens, W. J. Dally, R. Ho, D. Jayasimha, S. W. Keckler, and L.-S. Peh, "Research challenges for on-chip interconnection networks," *IEEE micro*, vol. 27, no. 5, pp. 96–108, 2007.

[2] D. Fan, M. Sharad, and K. Roy, "Design and synthesis of ultralow energy spin-memristor threshold logic," *IEEE Transactions on Nanotechnology*, vol. 13, no. 3, pp. 574–583, 2014.

[3] B. Mouttet, "Proposal for memristors in signal processing," in *International Conference on Nano-Networks*. Springer, 2008, pp. 11–13.

[4] M. Hu, H. Li, Q. Wu, and G. S. Rose, “Hardware realization of bsb recall function using memristor crossbar arrays,” in *Proceedings of the 49th Annual Design Automation Conference*. ACM, 2012, pp. 498–503.

[5] R. Bashirullah, W. Liu, and R. K. Cavin, “Current-mode signaling in deep submicrometer global interconnects,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 3, pp. 406–417, 2003.

[6] L. Gao, F. Alibart, and D. B. Strukov, “Analog-input analog-weight dot-product operation with ag/a-si/pt memristive devices,” in *VLSI and System-on-Chip, 2012 (VLSI-SoC), IEEE/IFIP 20th International Conference on*. IEEE, 2012, pp. 88–93

[7] H. F. Baghtash and S. J. Azhari, “Very low input impedance low power current mirror,” *Analog Integrated Circuits and Signal Processing*, vol. 66, no. 1, pp. 9–18, 2011.