

A CURRENT-MODE CMOS LOSER-TAKE-ALL WITH MINIMUM FUNCTION FOR NEURAL COMPUTATIONS

Nicolas Donckers^{*,1}, Carlos Dualibe^{*,**} and Michel Verleysen^{*,2}

^{*}Université catholique de Louvain – Microelectronics laboratory
Place du Levant, 3 - B-1348 Louvain-la-Neuve – Belgium

^{**}Universidad Católica de Córdoba, Laboratorio de Microelectrónica,
Cmno. a Alta Gracia KM 10, 5000 Córdoba, Argentina

ABSTRACT

A novel architecture for loser-take-all functions is proposed. Inputs and outputs of the circuit are currents, which make the circuit appropriated for low-voltage neural hardware computation. In contrast to most existing realisations the circuit does not require subtraction from a fixed reference what decreases accuracy and input dynamic. Moreover, in addition to the loser, it also outputs the minimum input current.

The circuit was synthesized using a SOI (silicon on insulator) technology and optimised to work with 1.5V voltage supply showing improved speed and accuracy for a very low power consumption (Typically 5 μ W per cell when the input current is 1 μ A).

1. INTRODUCTION

Loser-take-all are analogue computation cells pointing out the lowest analogue value among a set of candidates. They are widely used in hardware implementation of neural networks (such as Kohonen maps, vector quantization, classification algorithms, etc.).

Since Lazzaro [1], a lot of architectures have been proposed to compute the winner-take-all (WTA) function and the closely related loser-take-all (LTA) [2][3][4]. Some of the proposed architectures use winner-take-all (selecting the highest value among these inputs) to compute the loser-take-all function by subtracting input values from a fixed reference. The analogue subtraction implies a loss of accuracy and limits the input dynamic to the value of the fixed reference.

The circuit of Shoi and Cheu [2] uses a 2-stages operational amplifier. Based on this circuit, we proposed [5] dual architectures to compute the LTA and WTA functions. These structures do not require any subtraction to compute the LTA and can work under low voltage supply (1.8V). Their main drawback is their need for voltage inputs. Current inputs are better suited to the LTA function since it is usually preceded by a sum.

Frequently in neural computations, as well as in fuzzy logic, it is mandatory to know also the analogue value of the minimum. The circuit described below implements the LTA and the MIN

functions. It has been sized to work in the weak inversion region with a low voltage supply. Nevertheless, we use the g_m/I_D methodology which enables the designer to extend the synthesis over all regions of operation of the MOS transistor, while meeting an optimal solution for the constraints of speed, accuracy, and consumption imposed by the specifications.

2. THE PROPOSED ARCHITECTURE

2.1 LTA computation

In the circuit depicted in Figure 1, we kept the same architecture than in Lazzaro's WTA[1]: a set of current-controlled voltage-sources (Cell *i*) are connected in parallel to a common node N_C and fight to impose their own voltage. But in contrast to [1], and due to the source-follower connected PMOS transistor M103 in each cell, the common node N_C will follow the lowest voltage source rather than the highest. In this way, the lowest input controlling current will be prompted as is just required for a LTA.

In Figure 1 Cell *i* is repeated for each input. Current inputs are illustrated by ideal current sources for clarity but their actual circuit, shown in Figure 2, has to be considered. In each Cell *i* the controlling loop is made up by the common-source connected transistor M102 whose drain controls the gate of transistor M103 through another source follower stage (M101 and I_p). Transistor M101 must be inserted to shift down the drain DC voltage level of transistor M102 and adapt it to the DC level required at the gate of transistor M103.

As long as current $I_{m,i}$ becomes smaller, the drain voltage of transistor M102 falls and the gate voltage of transistor M103 follows the same trend. The cell with the smallest input current defines the voltage at node N_C . Current I_0 will thus be sunk through the corresponding transistor M103 by the diode-connected transistor M104 whose gate voltage drop can be also used as digital output for the LTA. Transistors M107 in all other cells with higher input currents remain in the triode region with a high drain voltage level so that transistors M103 in the corresponding cells switch off.

¹ N. Donckers is working towards the Ph. D. degree under FRIA fellowship.

² M. Verleysen is a research associate of the Belgian FNRS.

The MIN computation is carried out at the common cell by transistor M002, being saturated and having its gate tied to the node Nc. Therefore, its drain current is a mirror of the input current at transistor M102 in the loser cell.

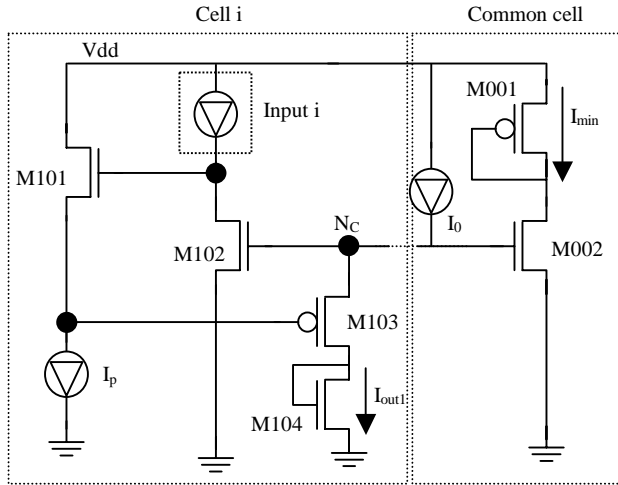


Figure 1. The basic cell and the common cell of the LTA. All cells are connected through the node Nc.

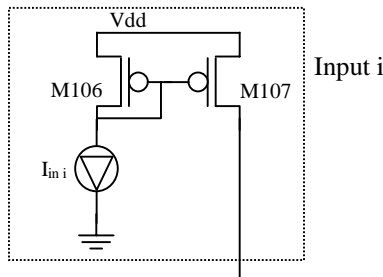


Figure 2. Cell i input circuit.

3. SMALL SIGNAL ANALYSIS AND SYNTHESIS

In spite the circuit performs a non linear operation, small signal linear analysis provides insight concerning its accuracy and speed behavior. For this purpose, we assume a circuit with only two inputs whose values are close enough so that we can represent the circuit by its small signal equivalent.

Looking at the DC gain and bandwidth of the obtained transfer function, we can have an idea about the precision and dynamic of the circuit. Moreover, all above mentioned features can be expressed in terms of the g_m/I_D of transistors which, in turn, is used to lead the design.

3.1 Synthesis methodology

We have used the sizing methodology proposed in [6] and the model of the MOS transistor proposed in [7][9]. The key parameter is the ratio g_m/I_D , where g_m is the transconductance and I_D the drain current. It has been shown that this ratio only depends on the inversion degree of the transistor and is almost independent on the transistor size. Typical values of this

parameters are $1V^{-1}$ to $10V^{-1}$ for strong inversion, $10V^{-1}$ to $25V^{-1}$ for moderate inversion and more than $25V^{-1}$ for weak inversion. The design methodology introduces a parameter named the *normalised current* (I). It is defined as the drain current I_D divided by the product $\mu C_{ox}(W/L)$.

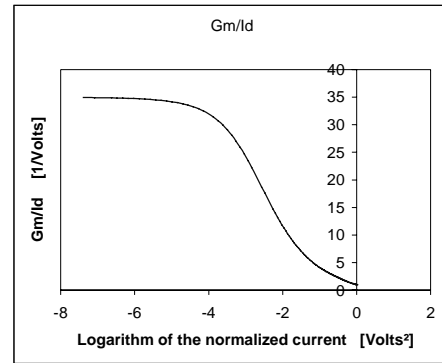


Figure 3. The characteristic curve of the MOS transistor.

The g_m/I_D ratio is used as a free variable parameter whose value is adjusted according to the policy driving the synthesis. For instance, if more speed is required transistors must be biased near strong inversion and the g_m/I_D has to be lowered. The curve of Figure 3 shows g_m/I_D versus the normalised current of a NMOS SOI transistors in our technology.

3.2 Frequency response analysis

Our main goal in this paper is to present a novel architecture that computes the loser-take-all and the minimum function. Nevertheless, to illustrate the use of the presented methodology, minimization of the power consumption will drive the synthesis.

In order to limit the consumption current I_0 has been set to $1\mu A$. The transfer function of the circuit shows 3 poles and 2 zeros. For similar reasons $g_m/I_D = 26V^{-1}$ is a good choice for transistor M101. A small g_m/I_D for M102 would be needed to achieve good performances in terms of speed. But low power consumption demands to work in weak inversion. A g_m/I_D of $24V^{-1}$ is a good compromise.

Figure 4 shows the value of the transition frequency of the circuit as a function of the g_m/I_D of transistor M103. g_m/I_D of the other transistors are fixed according to the previous discussion. The value of $25V^{-1}$ is chosen to maximize the bandwidth.

Simulations show that in strong inversion the position of the dominant pole depends on the value of I_p , whereas, in weak inversion it is independent from I_p . It can be set to $1\mu A$ optimising also the power consumption.

3.3 Transistor sizes

Once the g_m/I_D of each transistor and currents have been chosen sustained by the above discussion we can easily find the transistor sizes. With the known g_m/I_D we compute the normalized current I according the curve of Figure 3. Since the drain currents are already fixed, the size of the transistors are found to be equal to:

$$\frac{W}{L} = \frac{I_D}{\mu C_{ox} I}$$

It is useful to notice that if parameters (as the drain current, the g_m/I_D , etc.) have to be changed, the g_m/I_D methodology makes it possible to compute the new sizes of the transistors in a few simple steps, avoiding the use of complex formulas and working directly with measured curves and technological data.

Table 2 displays the values found for transistor sizes.

| | M101 | M102 | M103 | M104 | M106 | M107 |
|---|------|------|------|------|------|------|
| W | 63 | 42 | 123 | 41 | 126 | 126 |
| L | 3 | 3 | 3 | 3 | 3 | 3 |

Table 2. Sizes of the transistors (in μm)

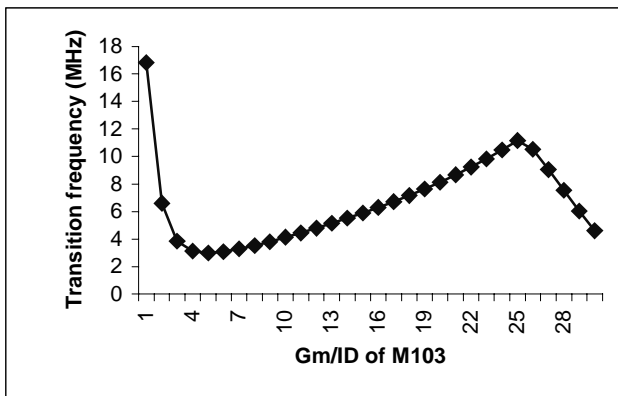


Figure 4. Transition frequency of the circuit vs g_m/I_D of M103

3.4 Improvements

In order to improve the speed and accuracy performances of the circuit (without significant changes in the power consumption), a second feedback loop has been introduced. Figure 5 shows the final version of the circuit. Transistor M003 has been added to the common cell replicating the minimum current in each Cell i through mirror transistors M105. In the loser cells this feedback forces the drain current of M102 to decrease (and the gate voltage of M101 to increase) significantly faster than before, speeding up the system.

The accuracy of the minimum function is also improved. From Figure 5 we can realize that the loser cell together with the common cell remain configured as a variant of an enhanced-Wilson current mirror with cascoded output.

4. PERFORMANCES

The performance of the circuit is quantified in terms of speed, accuracy and power consumption. Thanks to the SOI technology and to the design methodology, the circuit works optimally under 1.5V of voltage supply. Figure 6 shows a transient simulation of the circuit to prove the functionality. Overshoots presented at the outputs can be easily eliminated using fast Schmidt triggers connected to digital outputs.

4.1 Accuracy

Assuming that input currents can range from 0 to 15 μA the circuit exhibits between 6 and 7 bits of precision. But actually, due mainly to mismatch, a precision greater than 6 bits is difficult to obtain unless special layout techniques are applied.

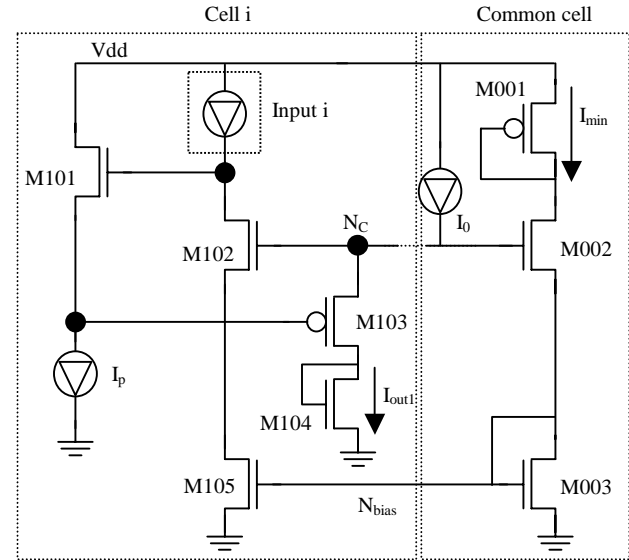


Figure 5. The final version of our circuit

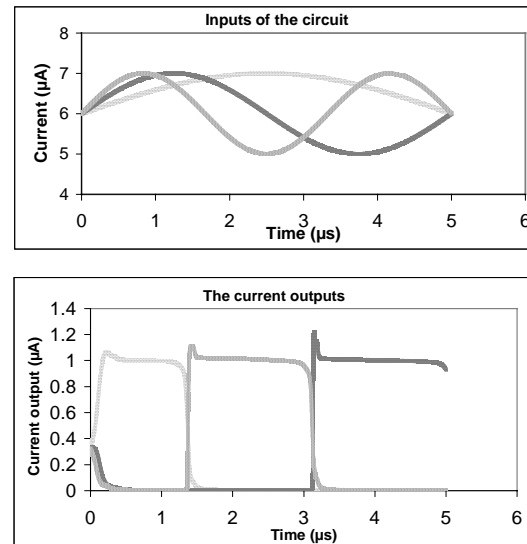


Figure 6. Transient simulation of the circuit.
Up: inputs. Down: Outputs

4.2 Speed and power consumption

To measure the speed characteristics of the circuit, we apply 2 constant inputs and a current step to the inputs of a 3-cells circuit. Two quantities are measured: the total delay and the

raising time. The total delay is defined as the time interval between the input current step and the output current reaching 90% of its full range. The raising time is defined as the time interval between 10% and 90% of the full output current range. The worst case has been obtained when the difference between the loser and another current is 1 LSB (0.06 μA) and the input currents are between 0 and 1 μA .

| | Typical | Worst case |
|--------------|--------------------|--------------------|
| Total delay | 0.35 μs | 0.82 μs |
| Raising time | 0.01 μs | 0.04 μs |

Table 3. Speed performances of the circuit

4.3 The minimum function

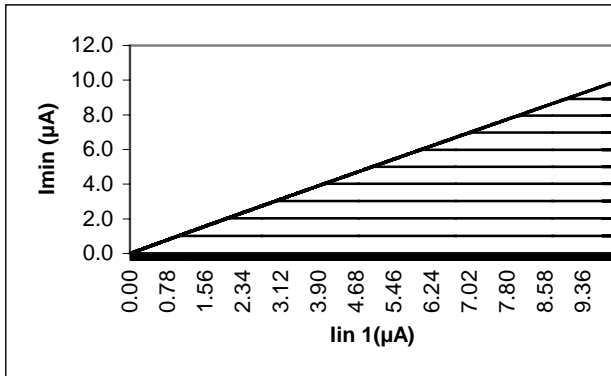


Figure 7. The minimum function.

Figure 5 shows a nested DC simulation performed for minimum function using a two-inputs circuit. Both inputs were swept from 0 to 10 μA , one by 10 nA steps and the other by 1 μA steps.

Note from the figure the sharp definition of knees at each switching level what can be taken as a qualitative measure of accuracy. This last was estimated by further zooming in at the knees and it was found to be near 6 bits.

4.4 Comparison with other LTA & WTA

As compared with other realizations, our LTA presents several advantages. Most of the previous circuits work under 5 V voltage supply and their power consumption is in the range 0.1 mW – 1 mW per cell [3]. The circuit presented in [5] works under 1.8 V of power supply but it uses voltages as input variables, which is not suited to neural computations. In terms of accuracy, our circuit shows similar performances as in [5], [8] (see [10] for a more extensive work) and [2]. Finally, none of the previous mentioned realizations are able to compute the analogue minimum function.

5. CONCLUSION

This paper presents a novel architecture for the loser-take-all and minimum functions. The circuit is a direct LTA of $O(n)$ complexity without need for subtraction. It can work efficiently

under low voltage supply. Speed, power consumption and accuracy specifications can be straightforwardly met using the presented g_m/I_D methodology with real technological data. This circuit is mainly intended to be used in analogue neural networks and fuzzy logic processors.

6. REFERENCES

- [1] 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 system, vol. 1, D. S. Touretsky, Ed. Los Altos, CA: Morgan Kaufman, 1989, pages 703-711
- [2] J. Choi and B. J. Sheu, "A High-Precision Winner-Take-All Circuit for Self-Organizing Neural Networks" *IEEE JSSC*, vol.28, n°5, may 1993
- [3] Z. Sezguin Günay and E. Sanchez-Simencio, "CMOS winner-take-all circuits: a detail comparison", *Proceedings of ISCAS'97*, vol. 1 pages 41-44, Hong-Kong, June 1997
- [4] S. Smedley, J. Taylor, M. Wilby, "A Scalable High-Speed Current-Mode Winner-Take-All Network for VLSI Neural Applications", *IEEE Transaction on circuits and systems I*, vol.42, n°5, pages 289-291, march 1995
- [5] N. Donckers, C. Dualibe, M. Verleysen, "Design of Complementary Low-Power Architectures for Loser-take-all and Winner-take-all" *Proceedings of the 7th International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, MicroNeuro'99 Grenada, Spain, 1999, pages 360-365
- [6] J-P Eggermont; D. De Ceuster, D. Flandre, B. Gentinnes, Paul G. A. Jespers and J-P Collinge, "Design of SOI CMOS Operational Amplifier for Applications up to 300°C" *IEEE JSSC*, vol.31, n°2, February 1996
- [7] Christian C. Enz, F. Kruppenacher and E. Vittoz, "An Analytical MOS Transistor Model Valid in All Regions of Operations and Dedicated to Low-Voltage and Low-Current Applications", *Analog Integrated Circuits and Signal Processing* 8, pages 86-114, 1995
- [8] T. Serrano and B. Linares-Barranco, "A Modular Current-Mode High-Precision Winner-Take-All circuit" *IEEE Transaction on Circuits and Systems II*, vol. 42, n°2, February 1995
- [9] C.G.Montoro, et al, "A Current-Based MOSFET Model for IC Design", Chapter 2 in 'Low-Voltage/Low-Power Integrated Circuits and Systems', E.Sanchez-Sinencio and A.G.Andreou (Eds.), IEEE Press, 1999
- [10] Serrano-Gotarredona T, Linares-Barranco B, « A high-precision current-mode WTA-MAX circuit with multichip capability », *IEEE JSSC* 33: (2) 280-286 FEB 1998