

## Design of Complementary Low-Power CMOS Architectures for Looser-take-all and Winner-take-all

Nicolas Donckers<sup>1</sup>, Carlos Dualibe<sup>1,2</sup>, Michel Verleysen<sup>1</sup>

<sup>1</sup> Université Catholique de Louvain, Microelectronics Laboratory,  
3 place du Levant, B-1348 Louvain-la-Neuve, Belgium

<sup>2</sup> Universidad Católica de Córdoba, Laboratorio de Microelectrónica,  
Cmno. a Alta Gracia KM 10, 5000 Córdoba, Argentina

E-mail: {donckers, dualibe, verleysen}@dice.ucl.ac.be

### Abstract

*A novel architecture for winner-take-all (WTA) and looser-take-all (LTA) circuits is proposed. As compared with other realisations, the LTA does not require input subtraction from a reference, which decreases accuracy and input dynamics. The architectures have been designed using the  $g_m/I_D$  methodology. As it will be shown, this method allows a rapid new dimensioning when specifications are modified. Both the WTA and the LTA can operate with low voltage supply, and show better speed characteristics (delay and raising time) for a 6 bits accuracy and a typical consumption of 50  $\mu$ W/cell than previous realisations.*

### 1 Introduction

Winner-take-all's (WTA) and looser-take-all's (LTA) are analogue computation cells selecting the highest (or lower) analogue value among a set of candidates [1,2]. These functions are widely used [3,4] in neural network computations (such as Kohonen's maps, vector quantization, classification's algorithms, etc.) , and in other applications (for example in fuzzy logic) including an interface between parallel analog computations and digital processing.

Many WTA circuits have been published since Lazarro [5]. Their performances can be measured in terms of speed, accuracy, delay, power consumption, etc. A fairly good review of several circuits can be found in [6]. Nevertheless, little effort has been done towards low-power WTA cells.

Moreover, most LTA circuits proposed in the literature are based on WTA, where inputs are subtracted from a fixed reference to achieve the desired computation; this implies a loss of precision and input dynamics due to the analog subtraction. A specific architecture for LTA has been proposed in [7-8], but it was not designed for low-power applications.

This paper proposes complementary low-power architectures for WTA and LTA circuits, functioning with low voltage supply. Furthermore, the LTA circuit does not need input signals subtraction from a reference. Simulations presented in this paper were realised using a 3  $\mu$ m SOI (Silicon-On-Isolator) technology.

### 2 WTA and OTA

Figure 1a shows the classical architecture of an OTA [9] and Figure 1b shows the architecture of a 2-cells winner-take-all [1]. The equivalence of the 2 architectures is obvious if we split  $M5$  (Figure 1a) into 2 transistors. A WTA has to find the highest input voltage among a set of candidates. This is just the behaviour of an OTA when it goes to saturation.

The output voltage corresponding to the highest input voltage goes to the positive voltage supply while the second output goes to the negative voltage supply. In the following, we will generalise this architecture using more than 2 cells. We will show that OTA structure is a fairly good solution to design low-power WTA and LTA.

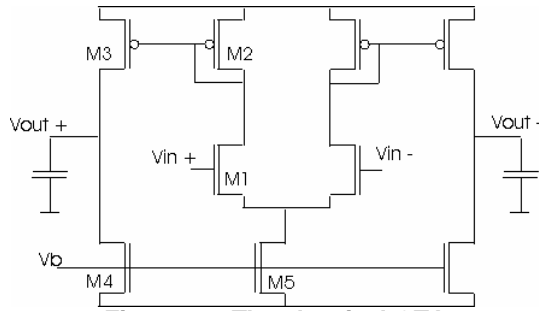


Figure 1a: The classical OTA

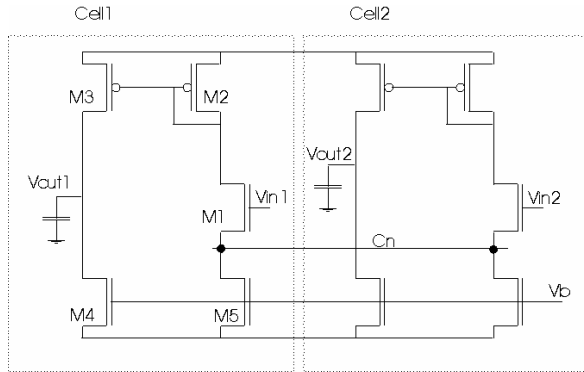


Figure 1b: The basic WTA

### 3 The WTA

#### 3.1 The basic cell

The principle of the WTA can be explained using Figure 1b. Transistor  $M1$  converts the input voltage into a current. When  $V_{in1}$  increases, current in  $M1$  increases, as well as current in  $M3$  through the current mirror formed by  $M3$  to  $M2$ , which makes  $V_{out1}$  increase.

This is however not sufficient to achieve the WTA effect. The principle is that differences between voltage inputs should be amplified, rather than the inputs themselves. The cell with the highest input should have a negative effect on the output voltage of all other cells, while the effect on all other cells on the "winner" output should be positive. This goal is carried out by connecting together node  $C_N$  from all cells. The current flowing from this node through all transistors  $M5$  connected in parallel is constant (at first order); any current increase in one of the transistors  $M1$  will force the current in the remaining  $M1$  transistors to decrease, what in turn decreases  $V_{out}$  in these cells.

Nevertheless, there may be particular situations where several inputs voltage become so high that any further increase in  $M1$  transistors' currents can not be supplied by  $M5$  transistors. Thus, those  $M1$

transistors will work in the triode region. As a consequence, several "winners", rather than only one, may appear.

One idea to avoid this problem is to adjust adaptively  $M5$  current sources using the *current steering* circuit depicted in figure 2, which is added to each cell of the WTA.

In each cell, when  $V_{out}$  increase, current in  $Ms1$  increases, as well as in  $Ms3$ . This makes the current in  $M1$  decrease (since the sum of these last two currents is a constant), the consequence being that the output voltages of all cells will decrease. The size of the transistor  $Ms4$  (and therefore the amount of current steering) is set in order to have, at the equilibrium, only one output voltage going to high. Connecting together the steering cells through a common node (the source of  $Ms1$ ) enhances the amplifier effect of the current steering circuit.

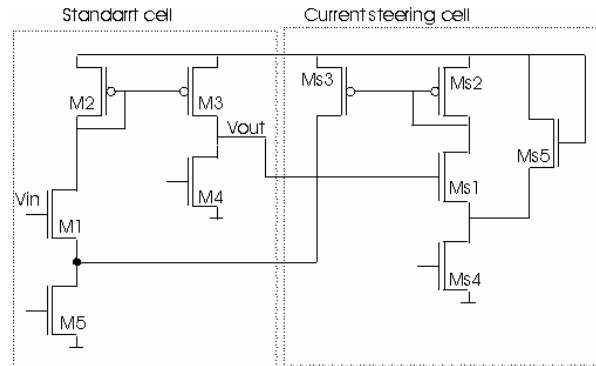


Figure 2: WTA using a current steering cell

#### 3.2 The cascoded cell

The product of the transconductance by the output impedance classically gives the gain of an amplifier. The gain has to be as high as precision is needed.

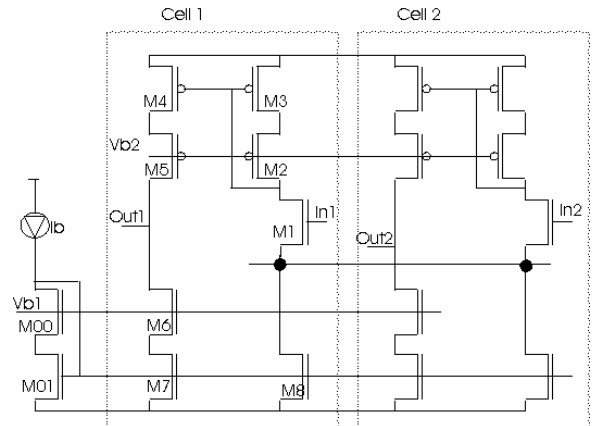


Figure 3: WTA with cascoded output

A fairly good solution to increase the gain is obtained by increasing the output impedance using a cascode technique. Fig. 3 shows the basic cell of our WTA (without steering cell for clarity).

It can be noticed that the cascode current mirror has an unusual configuration for a minimal loss of output dynamic, which in turn allows the circuit to work with a low voltage supply. The steering cell we will use in the simulations is the same as in Figure 2.

A similar core of circuit has been previously presented in [11].

## 4 Dimensioning principles

### 4.1 Characteristics of the circuits

The static gain of an amplifier using the basic cell of Fig. 3 is classically given by [9]:

$$A_{v0} = \left[ \frac{g_m}{I_D} \right]_1 \cdot \left[ \frac{g_m}{I_D} \right]_5 \cdot n \cdot V_{ea_4} \cdot V_{ea_{out}} \quad (1)$$

$$\text{with } V_{ea_{out}} = \frac{V_{ea_5} \cdot V_{ea_6}}{V_{ea_5} + V_{ea_6}}$$

Where  $V_{ea}$  is the Early voltage,  
 $g_m$  is the transconductance,  
 $I_D$  is the drain current,  
 $n$  is the body effect factor.

According to [7], the  $g_m/I_D$  is an essential parameter which depends only on the inversion degree. We decide to impose values to these parameters in order to fix the transistor's region of operation.

An other important parameter of the circuit is the GBW (Gain Bandwidth product) giving an idea about the speed of the circuit. In our case we have:

$$GBW = \frac{g_{m1}}{2 \cdot \pi \cdot C_L} \quad (2)$$

where  $C_L$  is the load capacitance.

Transistors sizes can be found using equations (1) and (2) according to the method described bellow. The value of the load capacitance has been set to 0.1pF. In neural networks implementations, WTA or LTA is the last cell of the circuit and is connected to the digital world through a buffer witch input capacitance can be estimated to 0.1pF.

### 4.2 The EPFL model of the MOS transistor

The model of the MOS transistor presented in [10] gives a curve (Figure 4) presenting the  $g_m/I_D$  as a function of the normalised current  $I$  given by:

$$I = \frac{I_D}{\mu \cdot C_{ox} \cdot \frac{W}{L}} \quad (3)$$

where  $C_{ox}$  is the oxide capacitance.

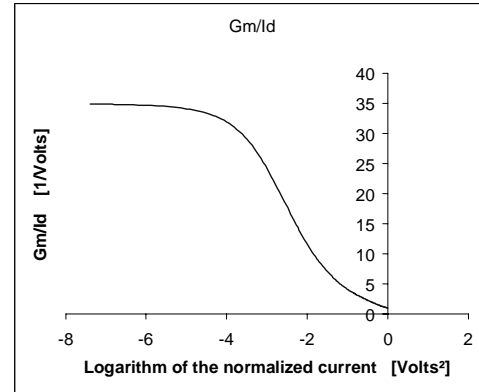


Figure 4: Characteristic curve of the MOS transistor

This function is strictly decreasing so there is a univoque relation between  $I$  and the  $g_m/I_D$ . Using this curve and the relations given by the analytical analysis of the circuit, we can derive a powerful dimensioning method.

### 4.3 Hand calculations

First, we choose the values of  $g_m/I_D$  for the different transistors in order to fix their region of operation. These values are reported in Table 1, where the  $g_m/I_D$  of transistor  $M1$  is set to a greater value than the others in order to improve the gain.

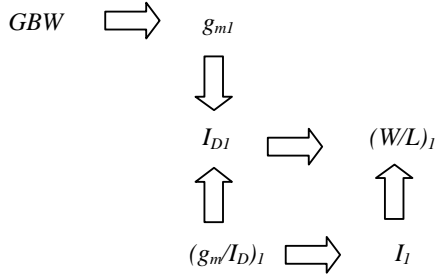
Note that all transistors are working in Moderate Inversion which is a well-known technique for analogue design optimisation giving rise to a trade-off between current consumption and die silicon area.

M1	M2	M3	M4	M5	M6	M7	M8
30	25	25	25	25	25	25	25

Table 1:  $g_m/I_D$  of the transistors of the cascoded cell

Secondly, we fix the value of the  $GBW$ : 100 MHz. Then, the dimensions are computed using the following relations:

GBW imposes the value of  $g_{m1}$  (since  $C_L$  comes from data specifications). The values of Table 1 allow to compute  $I_{D1}$  and  $I_1$  (the normalised current in  $M1$ ). Knowing these elements, we can compute the value of  $W/L$  using the relation between  $I_1$  and  $I_{D1}$ . With these elements and the topology of the circuit, it is simple to deduce the drain current of all other transistors and, knowing the  $g_m/I_D$ , the  $W/L$ . These operations can be summarised as:



The values obtained are:

	M1	M2	M3	M4	M5	M6	M7	M8
W/L	128	85	85	85	85	35	35	35

**Table 2: Size of the transistors**

The steering circuit has been sized using the same algorithm. The results we obtained are:

	Ms1	Ms2	Ms3	Ms4	Ms5
W/L	200	85	85	170	35

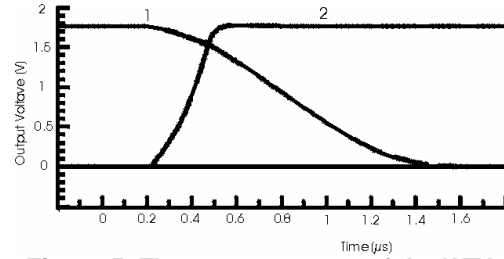
**Table 3: Dimensions of the steering cell**

## 5 Performances

Figure 5 shows the step response of an 8 cells winner-take-all with current steering. Signal 1 is the output voltage of the previous winner cell and signal 2 is the output voltage of the now one's. The simulation takes the parasitic capacitances into account. Table 4 summarizes the main characteristics of the circuit.

Simulations predict an accuracy of 8 bits. However, due to speed and matching considerations only 6 bits are attainable.

Raising time has been measured as the time interval between 10% and 90% of the full output voltage range. Precision is defined as the ratio between the smallest voltage input step which makes the circuit switch to another state and the full input voltage dynamics.



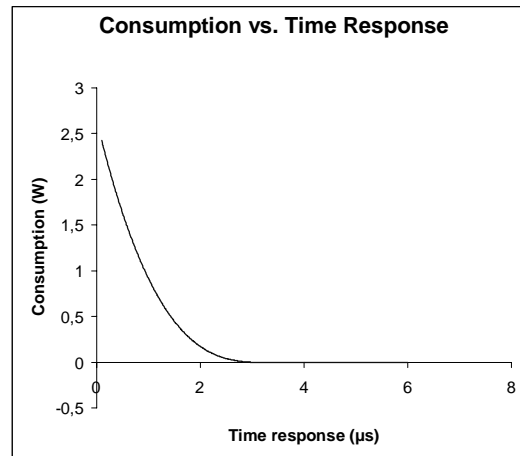
**Figure 5: The step response of the WTA**

Supply voltage	1,8 V
Power consumption	50 $\mu$ W/cell
Raising time	500 ns (typical) 200 ns (min.) 3,5 $\mu$ s (max.)
Precision	6 bits
Delay	300 ns (typical) 150 ns (min.) 1 $\mu$ s (max.)

**Table 4: Performances of the WTA**

Cell delay is defined as the time interval between the input voltage step and the output voltage reaching 90% of its full range. This delay of course strongly depends on the voltage differences between input signals; the maximum delay is measured when the voltage step which makes the circuit switch to another state is 1 LSB. Notice that possible static offsets were not take into account.

Figure 6 shows the relationship between the power consumption and the raising time of the circuit. By decreasing the  $g_m/I_D$  (i.e.: increasing the drain current) the raising time can decrease. Thanks to the design methodology, the circuit can be easily dimensioned for others specifications and characteristics.



**Figure 6: Consumption vs. Time Response**

## 6 From WTA to LTA

Our circuit splits the WTA function into two parts. The basic cell realises a comparison of all input voltages while the steering cell ensures the WTA function. This observation allows us to realise a LTA using the same technique. To validate this hypothesis, we realised a LTA using the same basic cell as for the WTA and transformed the steering cell by replacing all the NMOS transistors by PMOS transistors and *vice-versa* (and of course mirroring the schematics). In terms of performances, this is a very bad solution because the steering cell has to maintain artificially all the output voltages to the positive voltage supply (except for the looser) while the basic cell tries to maintain only one output (the winner) at the positive voltage supply. But it shows that it is possible to realise a LTA using this architecture.

An interesting solution to this problem has been found by applying to the basic cell the same transformation as to the steering cell. One cell of the obtained circuit is shown at Figure 7.

## 7 The LTA

The LTA as to select the lower analogue values among a set of candidates. As for the WTA this can be done using several conventions. One of these is to put the output voltage corresponding to the lower input at the positive voltage supply and the others outputs to the negative voltage supply. Due to the architecture of our LTA (Figure 7), we will choose an other convention, witch will naturally be inverted regarding to one used for the WTA. The output corresponding to the lower input voltage will goes to negative voltage supply, all others going to positive voltage supply.

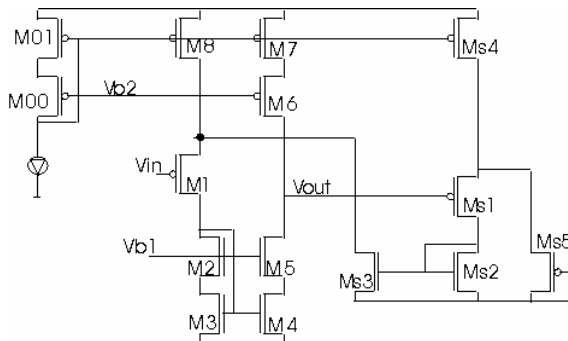


Figure 7: The loser-take-all

Obviously, due to the similarity, the LTA has been dimensioned using the same technique as for the WTA. The dimensions of the transistors are reported

in Table 5. The performances of the circuits under the same conditions as for the WTA are summarised in Table 6.

	M1	M2	M3	M4	M5	M6	M7	M8
W/L	300	35	35	35	35	85	85	85

Table 5: Dimensions of the LTA

supply voltage	1.8 V
power consumption	50 $\mu$ W / cell
raising time	400 ns (typical), 200 ns (min), 1,6 $\mu$ s(max).
precision	6 bits
delay	1 $\mu$ s (typical), 800 ns (min), 2 $\mu$ s (max).

Table 6: Performances of the LTA

The comparison between Tables 4 and 6 shows that the WTA and LTA performances are similar, which is an improvement over previously published solutions.

## 8 Conclusion

This paper presents a new circuit and a new methodology for the realisation of winner-take-all and loser-take-all cells.

We proposed a top-down methodology to dimension the circuit. Moreover, it is easy to make a new dimensioning if some of the specifications must be changed.

The main advantages of the architecture are its low supply voltage and low power consumption. The LTA has the supplementary advantage that it does not involve current subtraction, what always decreases the accuracy, and the possible input range (which is a crucial point with low supply voltage). This circuit is mainly intended to be used in analog neural networks and other analog processors.

## References

- [1] Joongho Choi and Bing J. Sheu "A High-precision VLSI Winner-take-all circuit for self-organizing neural networks" IEEE JSSC vol. 28 n°5 may 1993
- [2] Andreas Demosthenous, Sean Smedley, and John Taylor "A CMOS Analog winner-take-all Networks for Large-Scale Applications" IEEE Transactions on circuits and systems I, vol. 45 n°3, pp. 300-303, march 1998

- [3] Damien Macq, Michel Verleysen, Paul Jespers, *Fellow IEEE*, and Jean-Didier Legat, “*Analog implementation of Kohonen Map with On-Chip Learning*” IEEE Transactions on Neural Networks, vol. 4, n°3, May 1993.
- [4] Joydeep Ghosh, Ajat Hukkoo, *Member IEEE*, and Anjun Varma, *Member IEEE*, “*Neural Networks for Fast Arbitration and Switching Noise Reuction in Large Crossbars*” IEEE Transactions on circuits and systems, vol. 38, n°8, August 1991.
- [5] J. Lazzaro, S. Ryckebush, 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, pp. 703-711
- [6] Z. Sezguin Günay and Edgar Sanchez-Simencio. “*CMOS winner-take-all circuits : a detail comparaisn*”, Proceedings of ISCAS'97, vol.1, pp. 41-44, Hong-Kong, June 1997.
- [7] P. Thissen “*Architectures de circuits massivement parallèles pour la classification par méthodes neuronales*”, pp. 149-152, Ph. D. thesis (in French), Université Catholique de Louvain, Belgium, 1996.
- [8] M. Verleysen, P. Thissen, J.-L. Voz, J. Madrenas, “*An analog processor architecture for a neural network classifier*”, IEEE Micro vol. 14, n°3, June 1994.
- [9] J-P Eggermont, D. De Ceuster, D. Flandre, B. Gentinnes, Paul G. A. Jespers, J-P Collinge, “*Design of SOI CMOS Operational Amplifier for Applications up to 300°C*” IEEE JSSC vol. 31 n°2 February 1996
- [10] Christian C. Enz, François Krummenacher and Eric A. 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, pp 86-114 (1995)
- [11] Elfadel and Wyatt, Advances in Neural Information Processing System 6, p. 882 Cowan, Tesauro and Alspector, Morgan Kaufman.