A High-Precision Current-Mode WTA-MAX Circuit with Multichip Capability

Teresa Serrano-Gotarredona and Bernabé Linares-Barranco

***Abstract*—This paper presents a circuit design technique suitable for the realization of winner-take-all (WTA), maximum (MAX), looser-take-all (LTA), and minimum (MIN) circuits. The technique presented is based on current replication and comparison. Traditional techniques rely on the matching of an N transistors array, where N is the number of system inputs. This implies that when N increases, as the size of the circuit and the distance between transistors will also increase, transistor matching degradation and loss of precision in the overall system performance will result. Furthermore, when multichip systems are required, the transistor matching is even worse and performance is drastically degraded. The technique presented in this paper does not rely on the proper matching of N transistors, but on the precise replication and comparison of currents. This can be performed by current mirrors with a limited number of outputs. Thus, N can increase without degrading the precision, even if the system is distributed among several chips. Also, the different chips constituting the system can be of different foundries without degrading the overall system precision. Experimental results that attest these facts are presented.**

***Index Terms*— Analog circuits, analog computation, current-mode circuits, maximum circuits, transistor mismatch, winner-take-all.**

I. Introduction

WINNER-TAKE-ALL (or looser-take-all) and MAX (or MIN) circuits are often fundamental building blocks in neural and/or fuzzy hardware systems [3]–[5]. Given a set of N external inputs $(I_1, I_2, \cdots, I_i, \cdots, I_N)$, their operation consists in determining which input i presents the largest (or smallest) value, or what is this maximum (or minimum) value, respectively. If a winner-take-all (WTA) or MAX circuit is available, a looser-take-all (LTA) or MIN circuit is obtained by simply inverting the input $(-I_1, -I_2, \cdots, -I_i, \cdots, -I_N)$.[1] Hence, this paper will only concentrate on WTA and MAX circuits.

In literature, the physical implementation of these systems has been tackled through two main approaches: 1) systems of $O(N^2)$ complexity: their connectivity increases quadratically with the number of inputs [6]–[10] and 2) systems of $O(N)$ complexity: their connectivity increases linearly with the number of inputs [1], [2]. In a system of $O(N^2)$ complexity, as shown in Fig. 1(a), there is one cell per input; each cell has an inhibitory connection (black triangle) to the rest of the cells and an excitatory connection (white triangle) to itself. Therefore, the system has N^2 connections. Each cell j receives an external input I_j. The cell that receives the maximum input will turn all other cells OFF and will remain ON. If the system is a WTA circuit, each cell has a binary output that indicates whether the cell is ON or OFF. In a MAX circuit, the winning cell will copy its input to a common output. Under some circumstances[2] it is possible to convert the $O(N^2)$ topology of Fig. 1(a) into an $O(N)$ one, as shown in Fig. 1(b). In these cases, a global inhibition term is computed. Each cell contributes to this global inhibition, and each cell receives the same global inhibition. Note that now, each cell contributes to inhibit itself. Consequently, the excitatory connection that each cell has to itself must be increased to compensate for this fact. Typical $O(N)$ WTA circuits reported in literature [1], [2][3] correspond to the topology shown in Fig. 1(c). In such circuits there are also N cells, each receiving an external input I_j. Each cell connects to a common node, through which a global property (for example, a current) is shared between all cells. The amount of that global property taken by each cell depends (nonlinearly) on how much its input I_j deviates from an "average" of all inputs. Usually this "average" is not an exact linear average, but is somehow nonlinearly dependent on all inputs. The cell with the maximum input I_j takes most (or all) of the common global property, leaving the rest with little or nothing. Due to the way this global property is shared and how the "average" is computed, the operation of these circuits relies on the matching of transistor threshold voltages of an array of transistors [1] and/or other transistor parameters. The number of transistors in the array equals, at least, the number of inputs N of the system. If the WTA or MAX circuit has such a large number of inputs so that it must be distributed among different chips, the matching of threshold voltages (and/or other transistor parameters) will degrade significantly, and the overall system will lose precision in its operation.

This paper presents an $O(N)$ complexity circuit technique [which can be represented by the topology in Fig. 1(b)] for implementing either WTA and/or MAX circuits, based on current-mode principles. The resulting circuit does not rely on the matching of an N-size transistor array, but on precise local current replication and comparison. The circuit can be distributed among several chips, as is sometimes demanded by neural and/or fuzzy systems [11], while not degrading its precision, as shown in the section on experimental results.

II. Current-Mode Implementation of WTA-MAX Operation

A mathematical model that realizes the WTA-MAX operation and which is suitable for an $O(N)$ current-mode-based circuit implementation is presented as follows. Consider a system of N cells, such that each cell j produces an output

Manuscript received March 4, 1996; revised July 1, 1997.

The authors are with the National Microelectronics Center (CNM), Ed. CICA, 41012 Sevilla, Spain.

Publisher Item Identifier S 0018-9200(98)00728-8.

[1] Optionally, a common offset term may be added.

[2] If the inhibition that goes from cell i to cell j does not depend on j.

[3] The circuit in [2] processes voltage input signals, while the circuit in [1] and in this paper processes current input signals.

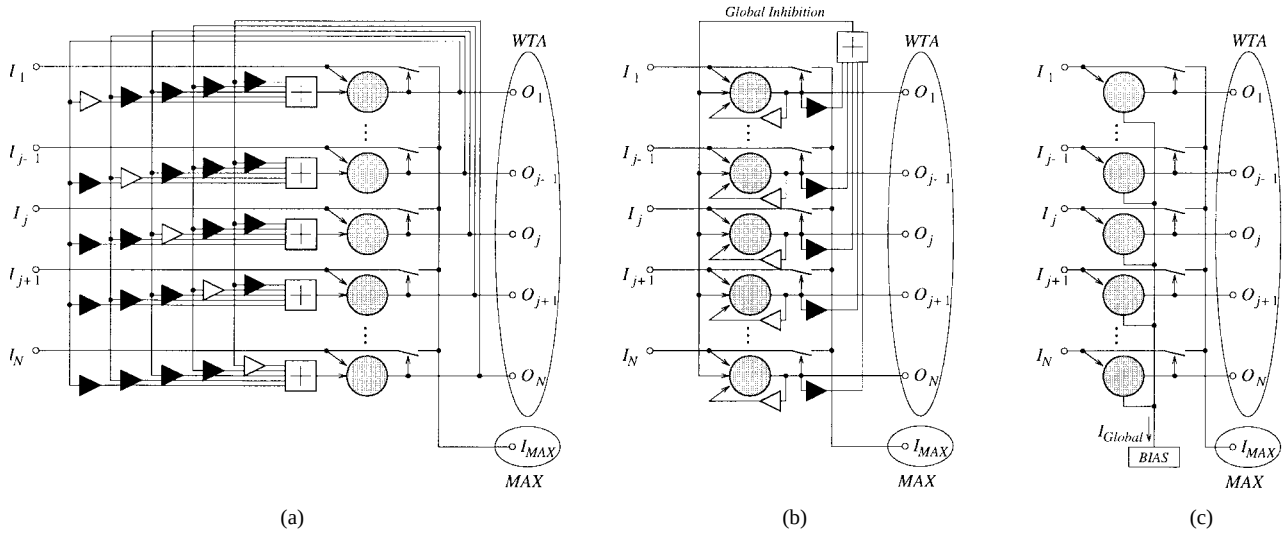

Fig. 1. WTA topologies. (a) WTA of $O(N^2)$ complexity, (b) transformation to $O(N)$ complexity, and (c) typical topology of $O(N)$ WTA hardware implementation. Black triangles represent inhibitory connections, white triangles excitatory connections, and shaded circles are generally nonlinear time-dependent processing elements whose outputs become (after a transient) either "0" or "1."

current $I_{oj} = \alpha_j H(I_j - I_o)$ with $j = 1, \cdots, N, H(\cdot)$ is the step function, I_j is the external input to the jth cell, and

$$I_o = \sum_{j=1}^{N} I_{oj} = \sum_{j=1}^{N} \alpha_j H(I_j - I_o). \tag{1}$$

Fig. 2 graphically represents functions $f_1(I_o) = \Sigma_j \alpha_j H(I_j - I_o)$ and $f_2(I_o) = I_o$. Their intersection provides the solution to (1). If $\alpha_j > 0$ $(\forall j)$, (1) has a unique equilibrium point S. Furthermore, if $\alpha_j \geq I_j$ $(\forall j)$, the value of I_o at the equilibrium point S is $I_o|_S = \max\{I_j\}$ and the cell that drives a nonzero output $I_{oj} \neq 0$ is the winner. If each input I_j is changed to $I_L - I_j$, where I_L is an upper bound for all input, $0 \leq I_j \leq I_L$ $(\forall_j)$, an LTA and/or MIN circuit results. Fig. 3(a) shows a current-mode circuit that implements the operation of one cell for the case $\alpha_j \equiv I_j$. It consists of a two-output current mirror, a digital inverter, and a MOS transistor. Each cell j receives two input currents, I_j and I_o, and delivers one output current I_{oj}. The inverter acts as a current comparator. If $I_o > I_j$, the inverter output v_{oj} is low, the MOS transistor is OFF, and I_{oj} is zero. If $I_o < I_j$, the inverter output v_{oj} is high, the MOS transistor is ON, and $I_{oj} = I_j$. Fig. 3(b) depicts the transfer curve of this unit cell. Fig. 3(c) and (d) shows the detailed schematic of the fabricated cells, one in the double-poly MIETEC 2.4-μm technology and the other in the single-poly ES2 1.0-μm technology, respectively. Fig. 4 shows the complete WTA or MAX circuit. It consists of N unit cells and an additional N-output current mirror. The function of the N-output current mirror is to deliver the sum of currents $I_o = \Sigma_j I_{oj}$ to each of the N unit cells. Replication of current I_o must be very precise. If the number of unit cells N is too large, or if the circuit has to be distributed among several chips, high precision in I_o replication cannot be guaranteed by a single current mirror with N outputs. In this case, replication of current I_o must rely on several mirrors with a smaller number of outputs but with guaranteed precise replication. Fig. 5 shows an arrangement to distribute the circuit of Fig. 4

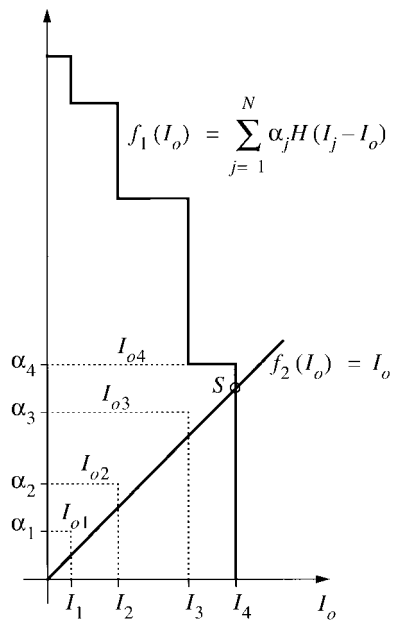

Fig. 2. Graphic representation of the solution of (1).

among several chips. The fact that current I_o can be replicated many times without relying on the matching of a large array of transistors is the main advantage of this WTA and MAX (or LTA and MIN) circuit technique over other implementations.

III. System Stability Analysis

Let us assume that the dynamics of each cell [see Fig. 3(a)] can be modeled by the following first-order nonlinear differential equation:

$$C_c \dot{v}_{xj}(t) + G_c(v_{xj}(t) - v_M) + I_j = I_o(t) = \sum_{j=1}^{N} I_{oj} \tag{2}$$

where C_c is the total capacitance available at node v_{xj}, G_c is the total conductance at this node, and v_M is the inverter

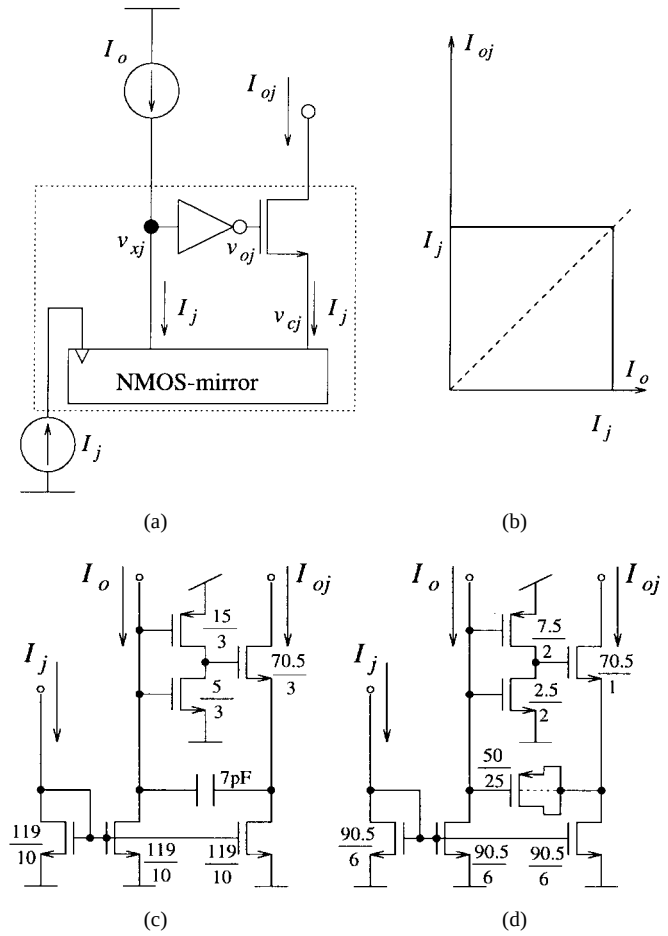

Fig. 3. WTA unit cell: (a) simplified schematic, (b) transfer curve, (c) circuit diagram of cell fabricated in the MIETEC 2.4-μm technology, and (d) circuit diagram of cell fabricated in the ES2 1.0-μm technology.

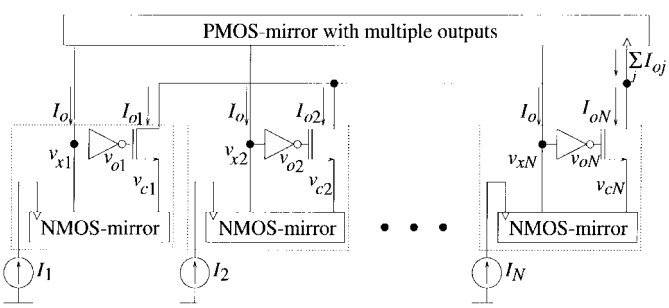

Fig. 4. Diagram of the WTA circuit.

trip voltage. Let us also assume that the output current of a cell is given by

$$I_{oj}(t) = I_j U(v_M - v_{xj}(t)) \tag{3}$$

where $U(\cdot)$ is a continuous and differentiable approximation to the step function. For example, we can define $U(\cdot)$ as the sigmoidal function $U(x) = 1/(1+e^{-x/\varepsilon})$ where ε is positive and nonzero but close to zero. Now consider (2) for two nodes, j and w. Let w be the node that eventually should become the winner. If we subtract (2) for the two nodes j and w, then

$$C_c[\dot{v}_{xj}(t) - \dot{v}_{xw}(t)] + G_c[v_{xj}(t) - v_{xw}(t)] = I_w - I_j. \tag{4}$$

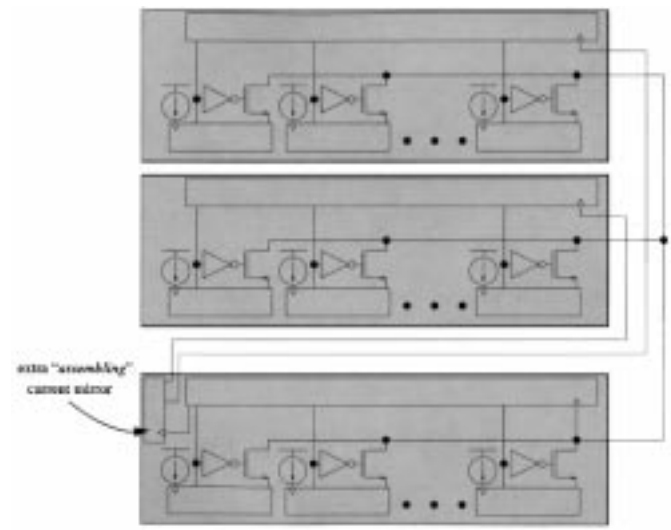

Fig. 5. Strategy to assemble several chips.

Equation (4) has the following solution:

$$\begin{aligned} v_{xj}(t) - v_{xw}(t) &= \frac{I_w - I_j}{G_c} + \left[v_{xj}(0) - v_{xw}(0) - \frac{I_w - I_j}{G_c}\right] e^{-\frac{t}{\tau_c}}, \\ \tau_c &= \frac{C_c}{G_c}. \end{aligned} \tag{5}$$

After a few time constants τ_c, the difference between the two node voltages will remain constant and equal to their difference at the equilibrium point. Therefore, if we can obtain the expression for $v_{xw}(t)$, applying (5) would obtain $v_{xj}(t)$ for the rest of the nodes.

Consider now (2) for node w, and substitute (3) into it

$$C_c\dot{v}_{xw}(t) + G_c(v_{xw}(t) - v_M) + I_w = \sum_j I_j U(v_M - v_{xj}(t)). \tag{6}$$

Since $v_{xj}(t)$ is given by (5), after a few time constants τ_c (6) becomes

$$\begin{aligned} C_c\dot{v}_{xw}(t) &= G_c(v_M - v_{xw}(t)) - I_w \\ &\quad + \sum_j I_j U\left(v_M - v_{xw}(t) - \frac{I_w - I_j}{G_c}\right). \end{aligned} \tag{7}$$

This first-order differential equation has stable equilibrium points if $(d\dot{v}_{xw}/dv_{xw})|_{\text{equilibrium point}} < 0$. Deriving (7) with respect to v_{xw} results in

$$C_c \frac{d\dot{v}_{xw}}{dv_{xw}} = -G_c - \sum_j I_j U'(\cdot). \tag{8}$$

Since G_c, I_j, and $U'(\cdot)$ are always positive, (8) is always negative for all possible values of v_{xw} (including its unique equilibrium point). Consequently, (7) represents the dynamics of a stable system.[4] This discussion assumes that the N-output current mirror presents no delay. This is not very realistic, however it can be shown [12] that the circuit is still stable

[4]The stability proof given in [20] for this system is not correct because it implicitly assumes symmetric interconnection weights between the cells, which is not true.

when assuming the N-output current mirror presents a delay modeled by first-order dynamics.

Performing electrical simulations of the circuit in Fig. 4 reveals that the previous stability analysis is a good approximation as long as the equilibrium point does not lie in the transition region of any of the N sigmoidal functions $U(\cdot)$. This can only be guaranteed if $\alpha_j = I_j$ and the two largest inputs I_j and I_w are sufficiently different. If $\alpha_j > I_j$ or (with $\alpha_j = I_j$) if two or more inputs I_j are maximum and very similar, the equilibrium point of the system (see Fig. 2) will be in the transition region of some sigmoids $U(\cdot)$. In these cases, transistor parasitic elements that have been neglected in the analysis of Section IV may render unstable behavior. Consequently, some kind of compensation is necessary. Under unstable conditions the system exhibits the following characteristics (observed through electrical simulations with HSPICE).

a) Only the cells j whose sigmoid functions $U(\cdot)$ must be in their transition region at the equilibrium point are unstable. The rest of the cells behave as if the system had reached its equilibrium point.
b) The unstable cells present oscillations (presence of complex conjugate poles).
c) In the case of $\alpha_j = I_j$ and with two or more equal maximum inputs, the steady-state oscillating waveforms at these cells become the same, regardless of their initial conditions.

This last observation suggests that a stability analysis could be performed by simply considering one cell in the system, which represents the parallel connection of all unstable cells, as shown in Fig. 6(a). On the other hand, since the unstable cells have the equilibrium point in the transition region of their sigmoid $U(\cdot)$, we can linearize these sigmoids for the stability analysis. Therefore, let us consider the small signal equivalent circuit shown in Fig. 6(b), where the circuitry comprised by dashed lines represents the parallel of all cells with equal and maximum input. The rest of the circuitry models the N-output current mirror (or set of current mirrors) responsible for distributing the global current I_o among the N cells. The minimum set of dynamic elements needed for the system to present unstable oscillating behavior are parasitic capacitors C_c, C_p, and C_g (observed through electrical simulation). Performing small signal analysis on the circuit in Fig. 6(b), it can be shown that the stability condition for this circuit is approximately [12]–[13]

$$AM < \frac{C_c}{g_n}\left(\frac{g_{mp}}{C_p} + \frac{g_{mn}}{C_g}\right) \tag{9}$$

where M is the number of cells with equal and maximum input. This condition is not easy to satisfy since A must be large for proper operation, M may become large, and it is not trivial to make the right hand side of (9) very large. Stability compensation can be achieved by introducing capacitor C_A, as shown in Fig. 6(c). By small signal analysis of this circuit, it can be shown that the stability condition for this circuit is [12]

$$C_A > \frac{g_n}{g_{mn}/C_g + g_{mp}/C_p}. \tag{10}$$

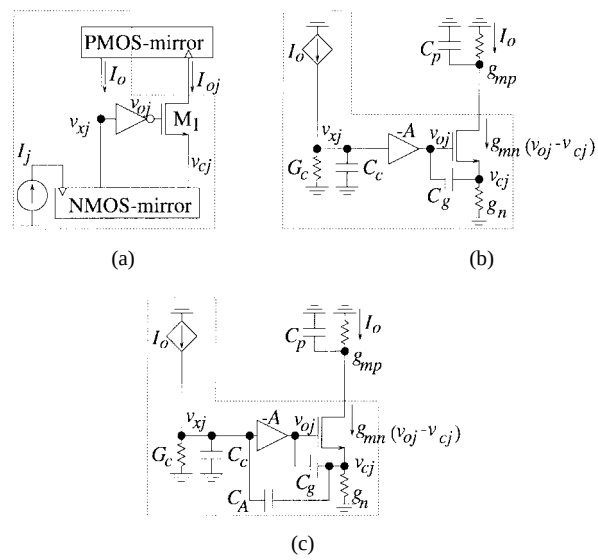

Fig. 6. (a) Parallel connection of unstable cells, (b) uncompensated small signal equivalent circuit, and (c) compensated small signal equivalent circuit.

Note that now the stability condition does not depend on gain A and is easier to fulfill. However, now capacitor C_A degrades the settling speed of the system. Capacitor C_A acts as a Miller capacitance. Since the dc gain from node v_{xj} to node v_{cj} is approximately $-A$ (i.e., the negative of the slope of $U(\cdot)$), there will be an effective Miller capacitance of value $(A+1)C_A$ in parallel with the original C_c capacitor. If the sigmoid is not in its transition region, $A = 0$, but if the sigmoid is in its transition region, A can be very large. Therefore, for compensated cells, (7) must be changed to

$$\begin{aligned}&[C_c + C_A + U'(v_M - v_{xw})C_A]\dot{v}_{xw}\\ &\quad = G_c(v_M - v_{xw}) - I_w + \sum_j I_j U\left(v_M - v_{xw} + \frac{I_w - I_j}{G_c}\right).\end{aligned} \tag{11}$$

If the winning cell is in its transition region, $U'(v_M - v_{xw}) \neq 0$ and a large capacitance $C_c + (A+1)C_A$ is present at node v_{xw}. Otherwise, $U'(v_M - v_{xw}) = 0$ and the effective capacitance is only $C_c + C_A$.

IV. Experimental Results

A WTA-MAX system with $N = 10$ competing cells has been designed and fabricated in two different CMOS technologies. The first prototype has been integrated in a double-metal single-poly 1.0-μm CMOS technology (ES2), and the other in a double-metal double-poly 2.4-μm CMOS process (MIETEC). Both technologies were available through the European silicon foundry service, EUROCHIP. Circuit schematics and transistor sizes of the unit cells are shown in Fig. 3(c) and (d) for the MIETEC 2.4-μm and ES2 1.0-μm CMOS processes, respectively. Sizes of the PMOS current mirroring transistors were 175 μm$\times$ 4 μm and 151 μm $\times$ 2.5 μm for the MIETEC and ES2 prototypes, respectively.

If the circuit is going to be used as a MAX circuit, all current mirrors must provide good replication precision. They need to have small systematic errors and small random deviations

TABLE I
CURRENT-MODE WTA PRECISION MEASUREMENTS

Technology	number of used chips	I_P			
		10μA	100μA	500μA	1mA
ES2_1.0μm	1	2.00%	1.07%	0.58%	0.56%
ES2_1.0μm	2	2.35%	1.03%	0.59%	0.57%
MIETEC_2.4μm	1	1.94%	0.98%	0.70%	0.69%
MIETEC_2.4μm	2	2.15%	1.17%	0.96%	0.87%
MIETEC_2.4μm ES2_1.0μm	2	2.24%	1.05%	0.73%	0.74%

[14], so that the resulting value of current I_o resembles the maximum among all inputs as close as possible. However, if the circuit is going to be used as a WTA circuit, requirements are not that severe. If inside one single chip, a WTA performs the same even if the current mirrors have appreciable systematic errors. Since systematic errors are common with respect to all inputs, the system can still determine which input is maximum. On the other hand, random mismatch errors in the current mirrors must be kept small because these errors change randomly from one input to another. Reducing random errors implies using larger transistor sizes. Reducing systematic errors implies using more elaborate current mirror topologies that either reduce their output conductance (using cascode [15], regulated cascode [16], or gain-boosting [17] techniques), decrease their input impedance [18], or both [19]. The application we had in mind when we developed this circuit was a WTA for a multichip real time clustering system [11]. Consequently, it was not critical that the final value of I_o be an exact replica of the maximum of the inputs. Therefore, we used a simple three-transistor current mirror (without any output conductance or input impedance decreasing technique) for the two-output NMOS current mirror of each cell. However, we used active input current mirrors [18] for the N-output PMOS current mirror and for the extra NMOS assembling current mirror (see Fig. 5). These current mirrors assure fixed voltages at their input nodes. This was necessary because if the system is distributed among several chips, the presence of the assembling current mirror would break the symmetry between some of the inputs, making systematic errors affect these inputs differently. The following presents proper system operation of a WTA circuit in one single chip, in two chips of the same technology, and in two chips each of a different technology. As will be shown, the dc behavior of the system is not degraded when the operation is distributed among several chips. In the remainder of this section we will detail experimental measurements related to the precision of a WTA and its speed response.

A. Operation Precision

The dc transfer curves of the system have been measured for different input current levels and for different system configurations. Fig. 7 shows 30 transfer curves when the competing cells are inside the same chip. Each curve is obtained by randomly selecting a pair of input cells i and j applying a constant input current $I_i = I_P$ to the first, and

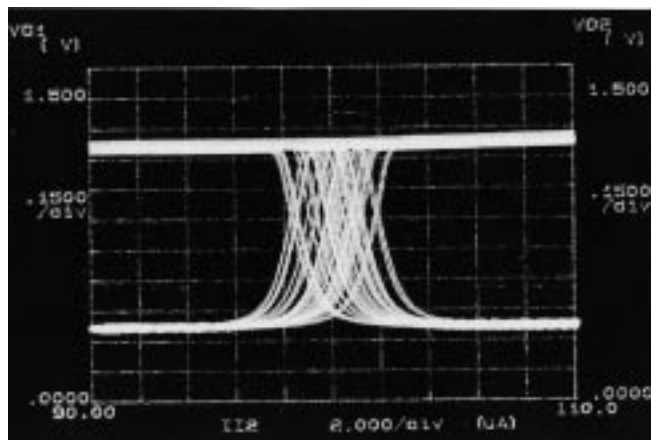

Fig. 7. Transfer curves of the WTA implemented in a ES2 1.0-μm chip for an input current level of 100 μA.

sweeping the input current of the second I_j from $0.9 \times I_P$ to $1.1 \times I_P$. The figure represents the two inverter output voltages v_{oi} and v_{oj} versus the current I_j. For each pair of cells i and j, we measured the value of I_j at the point where $v_{oi} = v_{oj}$. Let us call this value I_M. Thirty curves were measured for each value of I_P, resulting in 30 values of I_M. The difference between the mean of these 30 I_M values and I_P is a measure of the systematic error of I_M. Let us call it $\varepsilon(I_P)$. The variance of the 30 I_M values represents the random error of I_M. Let us call it $\sigma(I_P)$. In the case of Fig. 7, corresponding to a WTA inside one single chip fabricated in the ES2 1.0-μm CMOS technology with $I_P = 100$ μA, we measured a random deviation of $\sigma(I_P) = 1.04\%$ and a systematic error of $\varepsilon(I_P) = 0.03\%$.

Table I contains the measured total error (defined as $\sigma(I_P)+\varepsilon(I_P)$) for three decades of change in I_P. The table shows results for the cases of WTA's inside one chip, assembled using two chips of the same technology, and assembled with two chips of different technologies. Note that the precision degradation is very small when the system is distributed among two chips, regardless of whether the chips are of the same technology or not. This is the main advantage of this WTA-MAX circuit with respect to others reported in literature [1], [2].

B. Operation Speed

Delay measurements were performed as follows. Only two input signals were made nonzero. Let us call them I_1 and I_2. Current I_1 was made constant and equal to $I_{\rm IN}$, while current

TABLE II
MEASURED DELAY TIMES FOR ONE-CHIP WTA'S

I_{IN}	ΔI_{IN}	ES2_1.0μm				MIETEC_2.4μm			
		t_{d1}	t_{d2}	t_{d3}	t_{d4}	t_{d1}	t_{d2}	t_{d3}	t_{d4}
10μA	2μA	6.132μs	4.824μs	2.616μs	3.282μs	3.796μs	2.988μs	3.842μs	5.208μs
10μA	10μA	1.574μs	1.369μs	1.258μs	1.408μs	2.076μs	1.818μs	1.977μs	2.284μs
50μA	10μA	1.289μs	1.023μs	677ns	805ns	1.037μs	909ns	1.058μs	1.178μs
50μA	50μA	364ns	319ns	308ns	342ns	502ns	471ns	455ns	532ns
100μA	20μA	943ns	801ns	222ns	254ns	594ns	587ns	641ns	637ns
100μA	100μA	191ns	167ns	131ns	153ns	276ns	249ns	273ns	281ns
500μA	100μA	161ns	147ns	125ns	128ns	166ns	144ns	112ns	130ns
500μA	200μA	59ns	68ns	48ns	9ns	104ns	95ns	108ns	102ns

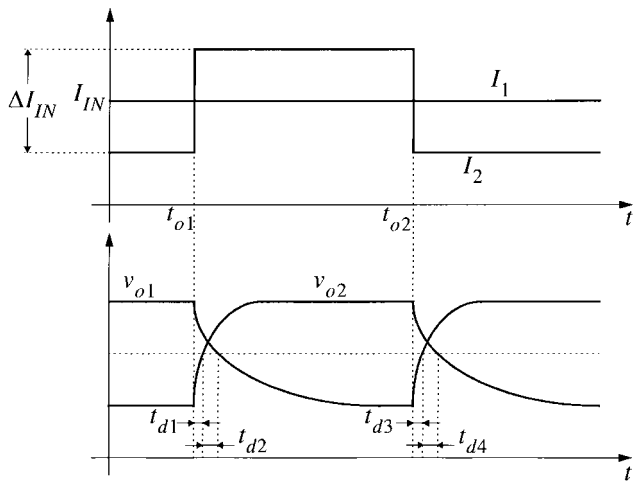

Fig. 8. Transfer curves when two ES2 1.0-μm chips are assembled and for an input current level of 10 μA.

I_2 changed in a pulse between values $I_{\rm IN} - 0.5\Delta I_{\rm IN}$ and $I_{\rm IN} + 0.5\Delta I_{\rm IN}$, as shown in Fig. 8(a). The pulse starts at time t_{o1} and ends at time t_{o2}. Waveforms v_{o1} and v_{o2} have the shape depicted in Fig. 8(b). Four different delay times were measured. For the system response caused by a rising edge in I_2, time t_{d1} is the delay between time t_{o1} and the instant at which voltage v_{o2} crosses the 50% value of its range. Delay t_{d2} is the same for output voltage v_{o1}. For the system response caused by a falling edge in I_2, time t_{d3} is the delay between time t_{o2} and the instant at which voltage v_{o1} crosses the 50% value of its range. Delay t_{d4} is the same for output voltage v_{o2}. Measurements were performed for $I_{\rm IN}$ values of 10 μA, 100 μA, and 500 μA, and for $\Delta I_{\rm IN}$ equal to $0.2I_{\rm IN}$ and $I_{\rm IN}$. Table II shows the measured delay times for those cases where the system is inside one single chip. Table III shows the delay times measured when a WTA is assembled using two chips of the ES2 1.0-μm process. Note that, in general, speed is degraded for a two-chip WTA. When the system is scaled up (increasing the number of inputs and chips) its speed will be further decreased. However, as long as current levels are maintained, its precision is preserved. Note that when increasing the number of inputs, the current levels can be maintained, because in the steady state (for one single winner) there is only one two-output NMOS mirror ON

TABLE III
MEASURED DELAY TIMES FOR A TWO-CHIP WTA

I_{IN}	ΔI_{IN}	ES2_1.0μm			
		t_{d1}	t_{d2}	t_{d3}	t_{d4}
10μA	2μA	16.8μs	6.50μs	5.40μs	3.40μs
10μA	10μA	3.2μs	2.35μs	1.86μs	1.80μs
100μA	20μA	470ns	480ns	750ns	590ns
100μA	100μA	235ns	230ns	270ns	240ns
500μA	100μA	154ns	134ns	375ns	150ns
500μA	200μA	150ns	104ns	136ns	110ns

and the PMOS mirror(s) drive its corresponding input current. On the other hand, for stability, (10) has to be satisfied: by increasing the number of chips, capacitance C_p will increase; however if instead of (10) the following condition is imposed:

$$C_A > \frac{g_n}{g_{mn}} C_g \tag{12}$$

the system will remain stable no matter how large C_p is.

V. CONCLUSION

A WTA-MAX circuit design technique based on current-mode signal processing has been proposed. The precision of the circuit relies on the proper replication and comparison of currents. This maintains good precision for circuits with a large number of inputs and when the circuit is distributed among several chips. Stability analysis of the proposed circuit has been addressed and stability conditions derived. A stability compensation scheme has been proposed. Two prototypes, for two different technologies, have been designed, fabricated, and tested. Proper performance has been experimentally verified for both prototypes, as well as for circuits assembled with different chips, even if each chip is of a different technology. The performance of this WTA-MAX circuit as compared to previous implementations [1], [2] is similar for both precision and speed. Actually, for speed performance, worst results would be expected with the proposed circuit since it needs stability compensation. The advantage of the present circuit is that it does not loose precision when used in multichip systems. In order to achieve this with previous implementations [1],

[2], some on-chip calibration schemes would be needed to compensate for interchip systematic transistor mismatch errors.

REFERENCES

[1] J. Lazaro, R. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of $O(N)$ complexity," *Advances in Neural Inform. Processing Syst.*, vol. 1, pp. 703–711, 1989.

[2] J. Choi and B. J. Sheu, "A high-precision VLSI winner-take-all circuit for self-organizing neural networks," *IEEE J. Solid-State Circuits*, vol. 28, May 1993.

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: IEEE Press, Macmillan, 1994.

[4] E. Sánchez-Sinencio and C. Lau, *Artificial Neural Networks: Paradigms, Applications, and Hardware Implementations*. New York: IEEE Press, 1992.

[5] J. C. Bezdec and S. K. Pal, *Fuzzy Models for Pattern Recognition*. New York: IEEE Press, 1992.

[6] S. A. Elias and S. Grossberg, "Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks," *Biological Cybernetics*, vol. 20, pp. 69–98, 1975.

[7] T. Kohonen, *Self-Organization and Associative Memory*, 3rd ed. Berlin: Springer-Verlag, 1989.

[8] A. L. Yuille and D. Geiger, "Winner-take-all-mechanisms," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: MIT Press, pp. 1056–1060, 1995.

[9] Y. He, U. Cilingiroglu, and E. Sánchez-Sinencio, "A high-density and low-power charge-based hamming network," *IEEE Trans. VLSI Syst.*, vol. 1, pp. 56–62, Mar. 1993.

[10] B. Linares-Barranco, E. Sánchez-Sinencio, A. Rodríguez-Vázquez, and J. L. Huertas, "A modular T-Mode design approach for analog neural network hardware implementations," *IEEE J. Solid-State Circuits*, vol. 27, pp. 701–713, May 1992.

[11] T. Serrano-Gotarredona and B. Linares-Barranco, "A real-time clustering microchip neural engine," *IEEE Trans. VLSI Syst.*, vol. 4, pp. 195–209, June 1996.

[12] T. Serrano-Gotarredona, "VLSI neural categorizers," Ph.D. Dissertation, University of Seville, Dec. 1996.

[13] T. Serrano-Gotarredona and B. Linares-Barranco, "Experimental results on the current-mode WTA-MAX circuit with multichip capability," in *1997 IEEE Int. Symp. Circuits and Systems (ISCAS'97)*, Hong Kong, 1997, vol. 1, pp. 561–564.

[14] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, 1989.

[15] P. E. Allen and D. R. Holberg, *CMOS Analog Design*. New York: Holt-Rinehart and Winston, 1987.

[16] D. Sackinger and W. Guggenbuhl, "A high-swing, high-impedance MOS cascode circuit," *IEEE J. Solid-State Circuits*, vol. 25, pp. 289–298, Feb. 1990.

[17] K. Bult and G. J. G. M. Geelen, "The CMOS gain-boosting technique," *Analog Integrated Circuits and Signal Processing*, vol. 1, pp. 119–135, 1991.

[18] D. G. Nairn and A. T. Salama, "A ratio-independent algorithmic analog-to-digital converter combining current mode and dynamic techniques," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 319–325, Mar. 1990.

[19] T. Serrano and B. Linares-Barranco, "The active-input regulated-cascode current mirror," *IEEE Trans. Circuits Syst.–I*, vol. 41, pp. 464–467, June 1994.

[20] ——, "A modular current-mode high-precision winner-take-all circuit," *IEEE Trans. Circuits Syst.–II*, vol. 42, pp. 132–134, Feb. 1995.