

Winner-take-all networks with lateral excitation

GIACOMO INDIVERI

giacomo@klab.caltech.edu
Computation and Neural Systems Program
California Institute of Technology
Pasadena, CA 91125, U.S.A.

Received ??. Revised ??.

Abstract.

In this paper we present two analog VLSI circuits that implement current mode winner-take-all (WTA) networks with lateral excitation. We describe their principles of operation and compare their performance to previously proposed circuits. The desirable properties of these circuits, namely compactness, low power consumption, collective processing and robustness to noisy inputs make them ideal for system level integration in analog VLSI neuromorphic systems. As application example, we implemented a circuit that employs an adaptive photoreceptor array as the input stage to the WTA network for edge enhancement.

Keywords: Analog VLSI, Current mode circuits, WTA networks, Edge detection, Visual attention

1. Introduction

The analog VLSI current mode winner-take-all (WTA) circuit, originally presented in [8] is a good example of a very well designed architecture. It is able to process globally all the signals of an input array, it uses a very limited amount of transistors per input node and it operates in parallel, with strictly local interconnections. This architecture has been extensively and successfully used in a wide variety of applications [3] [6] [7]. More recently, interesting modifications to the original circuit have been proposed in [4] and [12]. In both cases, the authors added to each element of the WTA network a local feedback circuit to obtain a hysteretic behavior in the selection/de-selection of the winning node: every time a new winner is selected the local feedback circuit adds a constant bias current to its input. The circuit will then de-select the winner when either its input current becomes lower than other inputs by a factor greater than the bias current or when the whole network is reset. From a functional point of view, this operation enhances the resolution of the network and eliminates instability problems, providing a robust mechanism that withstands the selection of other

potential winners unless they are stronger than the selected one by a set amount. The authors of [4] proposed a scheme for distributing locally the hysteretic component so that the winning input would be able to shift between adjacent locations maintaining its winning status, without having to reset the network. Following a similar approach, in this paper we present two novel variants of the original WTA network that also use a feedback circuit to provide hysteresis, but that have a different scheme for implementing lateral excitation. The two variants differ in the way the output signal is read: the first one has a discrete output, with only the winning element active and all others inactive; the second one represents a generalized version of the WTA architecture, with all its output elements active simultaneously, which behaves like a non-linear filter. These new circuits can be used as handy building blocks for VLSI models of attention mechanisms [3] [11] and, more generally, for a larger set of neuromorphic analog VLSI architectures [5].

2. Circuit descriptions

2.1. Discrete output WTA

Figure 2 shows the single elements of the modified winner-take-all circuit described in [4] next to the discrete output variant here proposed. As shown, the circuits are remarkably similar, yet their operating conditions and functional behaviors are quite different.

In the circuit of Fig. 2(b) laterally connected transistors M2 implement a “diffusor network” operating in weak-inversion [1] [14], used to distribute the hysteretic component of the winner’s output (*i.e.* the feedback current I_b flowing through transistor M3) to neighboring units. Lateral excitation is independent of the intensity of the winner’s input current; on the other hand, in the circuit in Fig. 2(a), laterally connected transistors are used to implement a different type of diffusor network that distributes the sum of both input current and hysteretic component to neighboring units. This operation, while laterally spreading the hysteretic current, simultaneously performs smoothing on the input data. The circuit here proposed will thus tend to favor areas that have a higher *average* input activity rather than selecting the single input with maximum intensity. This is instrumental in eliminating errors arising from salt-and-pepper noise and is helpful in eliminating errors that arise from offsets and device mismatches typical of analog VLSI technology. Moreover, the circuit in Fig. 2(a) has a

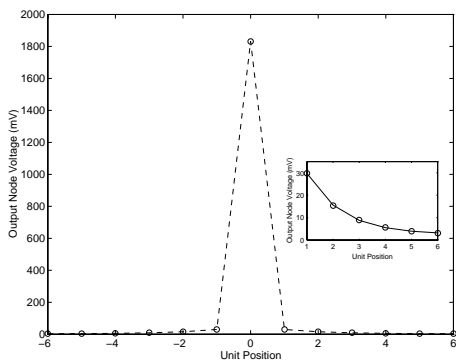


Fig. 1. Voltage distribution at the input nodes, obtained through circuit simulations of an array of 13 WTA elements with $I_b = 5\text{nA}$, $V_r = 0.80\text{V}$, $I_0 = 10\text{nA}$ and all other input currents null. The inset is a zoom-in of the data for units 1 to 6 fitted with the exponential function $f(x) = 51.39 \cdot e^{-0.73(x-0.13)} + 2.53$.

discrete output which is convenient for use with centroid circuits [2] [13] for encoding the winner’s spatial position in the array, whereas the WTA network proposed in [4] has multiple outputs that follow the current distribution imposed by the diffusor network: one is maximum for the winning element and the others decrease exponentially with distance. Another significant difference between the two circuits is in the way the bias voltages of the laterally connected transistors M2 are set: in order to correctly operate the diffusor network of the circuit in Fig. 2(b) the gate voltages V_r need to be set at values higher than the power supply voltage V_{dd} ; this problem does not occur for the circuit in Fig. 2(a) for which the voltages V_r are typically in the range 0.5V to 1V.

The circuit of Fig. 2(a) works as follows: if n is the winning node, V_n is set so that transistor M3 supplies all the bias current I_b and $V_{i \neq n}$ are all set so that the currents generated by the transistors with those gate voltages are approximately null. The bias current is hence copied only in the winning element through the current mirror M4-M5 and diffused, along with part of the input current, to neighboring elements through the diffusor network. The diffusor network is implemented by transistors M1, M2 and all the equivalent ones belonging to the other elements of the array.

Figure 1 shows the voltage distribution of the input nodes of a 13 element array, for a case in which there is an input current in the center and no input at all other nodes. The diffused current flows out of the winning element because $V_n > V_{n \pm 1} > V_{n \pm 2}$, and so forth. As shown in the figure, the winner-take-all network forces a discontinuity in the voltage distribution at the winning node. For all the elements that are more than one node away from the winning one, the voltage distribution has a traditional resistive-network form, which can be approximated by the equation

$$V = V_0 e^{-\alpha|x|}$$

where x represents the distance from the input node and α represents the space constant of the diffusor network, defined as the rate at which signals die out with distance from the source [10].

The diffusor network here proposed is a *current mode* one, hence also the current distribution will

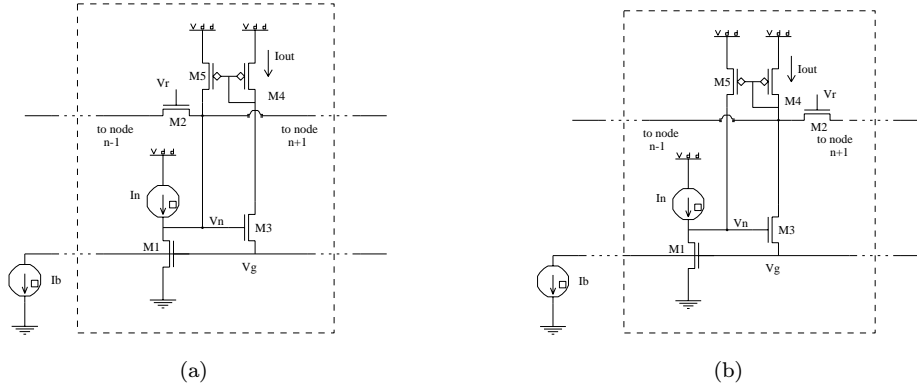


Fig. 2. Elements of the modified WTA circuits. (a) Version described in this paper; (b) Previously proposed version. The two circuits differ in the way the transistors M2 are connected. The power supply voltage V_{dd} is set to 5V. Input currents typically range from picoamperes to microamperes. The laterally connected transistors operate in weak-inversion.

follow a similar profile. For such a network the space constant is defined as:

$$\alpha = e^{-\frac{\kappa}{2U_T}(V_r - V_g)}$$

where κ is the subthreshold slope coefficient, U_T is the thermal voltage and V_r and V_g are the gate voltage of the laterally connected transistors and the common node voltage (gate voltage of transistors M1 in Fig. 2) respectively.

The amount of lateral excitation can thus be controlled by changing the bias voltage V_r (see Fig. 7(a)) or by having input currents of different intensities (see Fig. 7(b)). Specifically, since the common node voltage V_g increases logarithmi-

cally with the winning input current (for transistors operating in weak-inversion), the space constant α increases linearly with input current intensity. As a consequence, if the winning input is relatively strong (high confidence), the lateral excitation area is confined to a small neighborhood around the winning node. If, on the other hand, the winning input is relatively weak (low confidence), the lateral excitation area is wider.

Having not instrumented transistor M1 of Fig. 2(a) (and all other equivalent transistors in the array), the data shown in Fig. 7 was obtained through circuit simulations, by means of which it was possible to measure the current flowing through the transistors M1 without affecting the behavior of the circuit. Experimental data on the output nodes of the circuits has been obtained for a 25 element WTA network, implemented on a 2.3mm by 2.3mm chip using analog CMOS $2\mu\text{m}$ technology. Fig. 3 shows measurement results for a case in which 3 input units are active and all others are null. The input current on unit 8 is gradually swept from 70 nA to zero and back; initially the WTA network selects unit 8 as the winner. As soon as the current I_8 decreases to values lower than 61 nA the network selects unit 10 as the winner, despite its input current being lower than the current on unit 16 (follow dashed line on Fig. 3 from left to right). This is a consequence of the effect of lateral excitation. Due to the same effect, for the particular values of bias current I_b and gate voltage V_r used in this

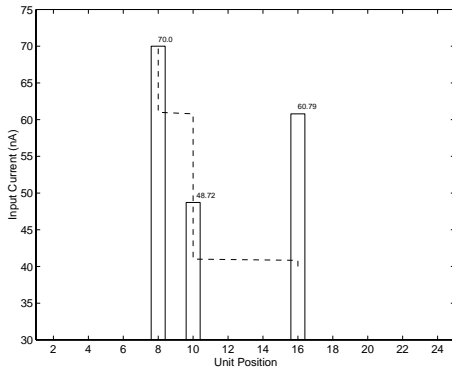


Fig. 3. Chip data measurements for a 25 element WTA network with $V_r = 0.75\text{V}$ and $V_b = 0.9\text{V}$, where V_b is the gate voltage of a $4\mu\text{m}$ by $4\mu\text{m}$ transistor used to generate the bias current I_b . The dashed line shows the selection of the winner as the input current on unit 8 is swept from 70 nA to zero.

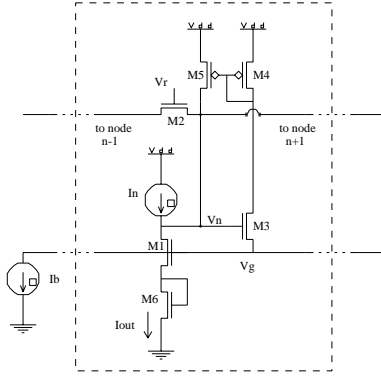


Fig. 5. Circuit diagram of an element of the generalized WTA architecture with lateral excitation and analog output. The circuit differs with the one of Fig.1(a) by one transistor (M6).

experiment, the network will switch to selecting unit 16 as the winner, only when I_8 decreases to a value lower than approximately 41 nA. The network will then switch back to selecting unit 8, neglecting unit 10, when I_8 increases back to values greater than 61 nA. This experiment, while indirectly demonstrating the hysteretic behavior of the network, shows how it tends to select elements close to the previously selected winner when the average input activity around the winner is high, and how it tends to function as a traditional WTA network when the winner is an isolated input.

An application of the discrete-output WTA architecture, that exploits this property, can be found in [7]: the authors used this architecture as the last computational stage of a *focus of expansion* detection chip. Such system was designed for selecting heading direction in case of translatory ego-motion and tracking it in time. The authors chose to use this variant of the WTA architecture in order to account for the *a-priori* assumption that the heading direction position shifts smoothly in space.

2.2. Generalized analog output WTA

The basic element of the generalized WTA architecture with lateral excitation and analog output is shown in Fig. 5. The circuit is identical to the one shown in Fig. 2(a) except for the addition of the extra diode-connected transistor M6, which is used to read the output current.

For a case similar to the one of Fig. 7, in which there is only one active input in the center of the WTA array, the behavior of the network is similar to the one of the circuit previously described (see Fig. 6). For more realistic cases though, in which there is a structured pattern of input values, the architecture behaves like a non-linear filter that enhances the input with maximum amplitude and smoothes the rest of the data. Specifically the enhancement effect and the smoothing effect are superimposed: the diffusor network implemented with transistors M1 and M2 performs the smoothing operation on the input data while the winner-take-all network adds the hysteretic feedback current to the input with maximum intensity. The net result is that of having output currents that correspond to the sum of the smoothed input data with the smoothed hysteretic current.

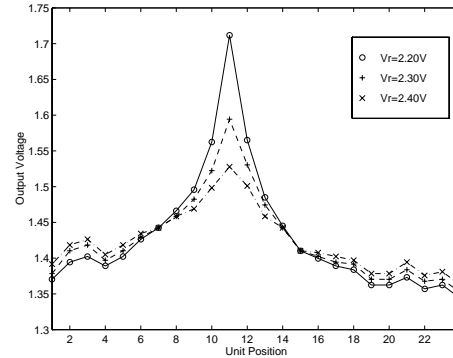


Fig. 6. Chip data measurements for a 25 element generalized WTA network, with $I_b = 21\text{nA}$, $V_{11} = 0.9\text{V}$ and $V_{i \neq 11} = 0.75\text{V}$. The output current has been converted into voltage using an off-chip sense amplifier.

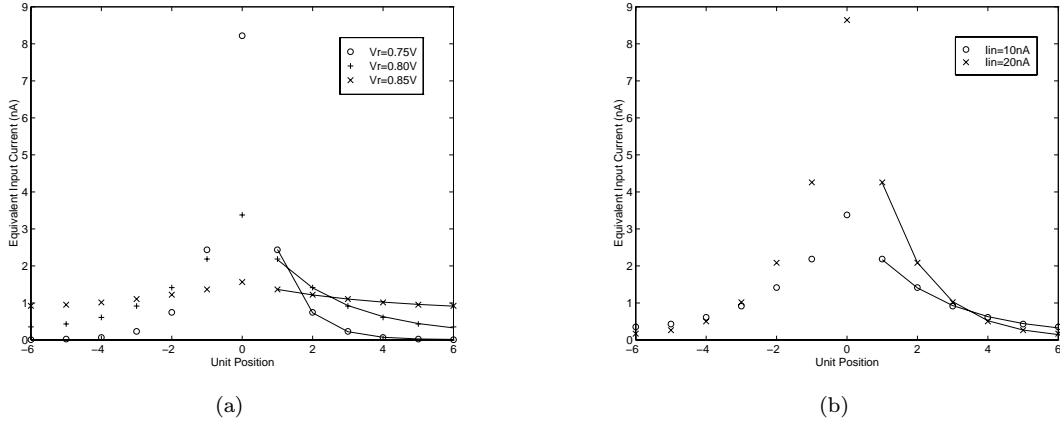


Fig. 7. Sum of input current and hysteric current flowing through transistors M1 of the array. (a) Simulation results for different values of V_r with $I_b = 5\text{nA}$, $I_0 = 10\text{nA}$ and all other input currents null; (b) simulation results for different values of I_0 with $I_b = 5\text{nA}$ and $V_r = 0.80\text{V}$. Both sets of data are fitted with exponential functions that have realistic space constant coefficients, for positions more than one unit away from the winner.

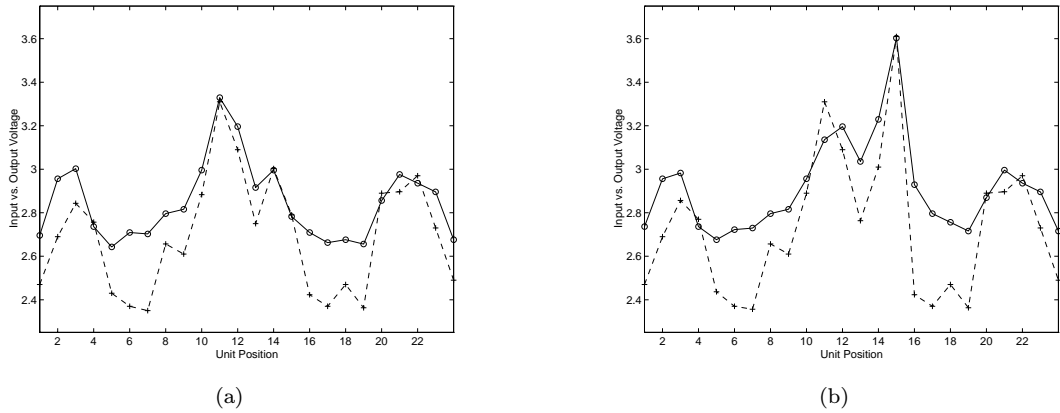


Fig. 4. Data measurements of input versus output currents for a 25 element generalized WTA network with $V_r = 2.25\text{V}$ and $I_b = 21\text{nA}$. The solid line represents the output of the network, while the dashed line represents its input. In (a) the input on unit 15 is set to $V_{15} = 0.89\text{V}$ and the input to unit 11 (set to $V_{11} = 0.91\text{V}$) is the maximum input value; in (b) unit 15 (set to $V_{15} = 0.95\text{V}$) is the maximum value, with the rest of the input values unchanged. Note how the output of unit 15 is enhanced (and the rest of the data smoothed); a normal smoothing network would have decreased this value, possibly making it loose its winning status.

Figure 4 shows measurement results, obtained from a 25 elements generalized WTA network. In the example shown all nodes of the array have random input values set by external potentiometers that control the gates of the input transistors. Initially (Fig. 4(a)) unit 11 has maximum input. As shown the output corresponds to a smoothed version of the input with the winning node enhanced. Subsequently (Fig. 4(b)) the winning input shifts from position 11 to position 15. The WTA tracks

the winning input, modifying the rest of the data accordingly. For the data shown, the bias current I_b lies within the range of the input currents. By modifying the value of I_b it is possible to intensify (or weaken) the strength of the enhancement effect over the smoothing effect, thus emphasizing (or de-emphasizing) the winner-take-all nature of the circuit.

3. Application example

As an application example, we have designed a system in which the input to the WTA network is provided by a 1-D silicon retina as proposed by [9]. The circuit diagram of Fig. 8 shows one basic element of the overall 1-D array. The bottom part of the circuit implements the adaptive photoreceptor circuit with spatial coupling between pixels described in [9]. The current mirror implemented by the two p-type transistors in the central part of the circuit amplifies the photodiode light induced current. The top part of the circuit implements both types of WTA architectures proposed: by connecting V_{sel} to V_{dd} we will implement the discrete-output variant of the WTA architecture described in section 2.1 and obtain positive output currents sourced from the p-type output transistor, whereas by connecting V_{sel} to ground we will implement the generalized WTA architecture described in section 2.2 and obtain negative output currents sunk from the n-type output transistor.

The spatio-temporal filtering properties of the silicon retina designed allow the circuit to extract edges at different spatial frequencies. By changing the bias voltages V_{gp} and V_{rp} it is possible to set the spatial-frequency tuning of the filter (see Fig. 9). The WTA network will then select the edge with spatial-frequency that elicits the strongest response and track it as it moves. Depending on the value of V_{sel} , the output of the chip will correspond either to a single active pixel in the output array (representing the winner) or to a smoothed contour map of the input image with the strongest edge enhanced.

A chip containing 25 pixel, fabricated using a $2\mu\text{m}$ n-well CMOS process provided by MOSIS, is currently being tested.

4. Conclusions

We have presented two circuits that implement current-mode winner-take-all networks with lateral excitation. Their behavior was analyzed and compared to the one of similar circuits previously proposed. As an application example of these WTA networks, we have presented an analog CMOS VLSI system, with on-chip adaptive photoreceptors, for selecting the most salient edges

present in an image. As demonstrated, the circuits are extremely compact and suitable for integration with analog VLSI neuromorphic systems.

Acknowledgments:

I would like to thank Christof Koch for his support, Ernst Niebur and Tonia Morris for their constructive comments and Timothy Horiuchi for the insightful discussions on winner-take-all circuits. This work was supported in part by the Center for Neuromorphic Systems Engineering as a part of the National Science Foundation Engineering Research Center Program; and by the California Trade and Commerce Agency, Office of Strategic Technology. Fabrication of the integrated circuits was provided by MOSIS.

References

1. K.A. Boahen and A.G. Andreou. A contrast sensitive silicon retina with reciprocal synapses. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in neural information processing systems*, volume 4. IEEE, MIT Press, 1992.
2. S. P. DeWeerth. Analog VLSI circuits for stimulus localization and centroid computation. *Int. J. of Comp. Vision*, 8(3):191–202, 1992.
3. S.P. DeWeerth and T.G Morris. Analog VLSI circuits for primitive sensory attention. In *Proc. IEEE Int. Symp. Circuits and Systems*, volume 6, pages 507–510. IEEE, 1994.
4. S.P. DeWeerth and T.G Morris. CMOS current mode winner-take-all circuit with distributed hysteresis. *Electronics Letters*, 31(13):1051–1053, June 1995.
5. R. Douglas, M. Mahowald, and C. Mead. Neuromorphic analogue VLSI. *Annu. Rev. Neurosci.*, (18):255–281, 1995.
6. T. Horiuchi, W. Bair, B. Bishofberger, J. Lazzaro, and C. Koch. Computing motion using analog VLSI chips: an experimental comparison among different approaches. *International Journal of Computer Vision*, 8:203–216, 1992.
7. G. Indiveri, J. Kramer, and C. Koch. System implementations of analog VLSI velocity sensors. *IEEE Micro*, 16(5):40–49, October 1996.
8. J. Lazzaro, S. Ryckebusch, M.A. Mahowald, and C.A. Mead. Winner-take-all networks of $O(n)$ complexity. In D.S. Touretzky, editor, *Advances in neural information processing systems*, volume 2, pages 703–711, San Mateo - CA, 1989. Morgan Kaufmann.
9. S. Liu and K. Boahen. Adaptive retina with center-surround receptive field. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in neural information processing systems*, volume 8. MIT Press, 1996.

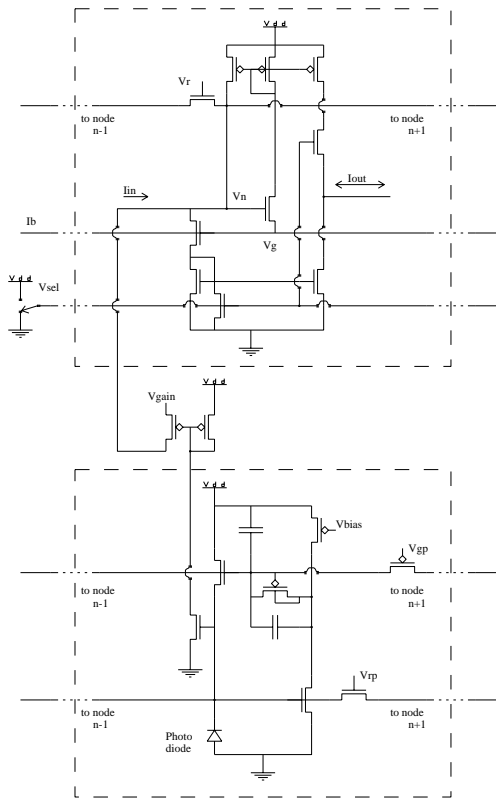


Fig. 8. Basic cell for a 1-D adaptive retina chip connected to the WTA architecture. The bottom part of the figure contains the adaptive photoreceptor circuit. The top part of the figure contains a circuit that implements both types of WTA networks described in section 2 (depending on the value of V_{sel}). The current mirror in the middle part of the figure is used to amplify the output current of the adaptive photoreceptor.

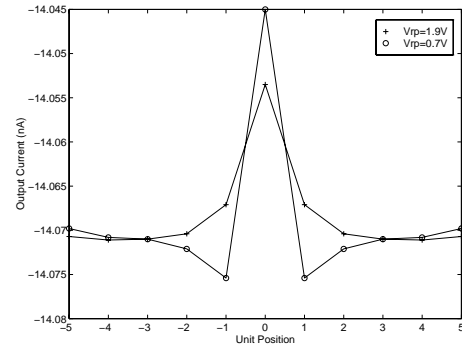


Fig. 9. Simulation results of a 1-D array of adaptive photoreceptor circuits with spatial coupling. The curves plotted represent the spatial impulse response of the system, obtained by setting all photodiode currents of the array to 100pA, except for the current of unit zero, which was set to 300pA. The photoreceptor bias voltage was set to $V_{bias} = 0.4V$, the p-type coupling transistor gate voltage was set to $V_{gp} = 0.9V$ and the n-type coupling transistor gate voltage was varied. As shown, it is possible to obtain center-surround convolution kernels with different frequency selectivities by changing the values of these gate voltages.

10. C.A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA, 1989.
11. E. Niebur and C. Koch. Control of selective visual attention: modeling the “where” pathway. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in neural information processing systems*, volume 8. MIT Press, 1996.
12. A. Starzyk, J. and X. Fang. CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback. *Electronic Letters*, 29(10):908–910, May 1993.
13. M. Tartagni and P. Perona. Computing centroids in current-mode technique. *Electronics Letters*, 29(21):1811–1813, October 1993.
14. E.A. Vittoz and X. Arreguit. Linear networks based on transistors. *Electronics Letters*, 29(3):297–298, February 1993.