Analog Integrated Circuits and Signal Processing, 31, 47–53, 2002

A Normalizing aVLSI Network with Controllable Winner-Take-All Properties

SHIH-CHII LIU

Institute of Neuroinformatics, University of Zurich and ETH Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
E-mail: shih@ini.phys.ethz.ch

Received June 28, 2000; Revised June 26, 2001; Accepted July 11, 2001

Abstract. We describe an aVLSI network consisting of a group of excitatory neurons and a global inhibitory neuron. The output of the inhibitory neuron is normalized with respect to the input strengths in a manner that is useful in any system where we wish the output signal to code only the strength of the inputs, and not be dependent on the number of active inputs. The circuitry in each neuron is equivalent to that in Lazzaro's winner-take-all (WTA) circuit [1] with one additional transistor and a voltage reference. As in Lazzaro's circuit, the outputs of the excitatory neurons code for the neuron with the largest input. The novel feature is that multiple winners can be chosen (soft-max). By varying one parameter, the network can operate in a soft-max regime or a WTA regime. We show results from two different fabricated networks.

Key Words: winner-take-all circuit, normalizing circuit, analog VLSI, neuronal networks

1. Introduction

The winner-take-all (WTA) function is a useful computation in self-organizing neural networks [2] and signal processing applications. It selects a single winner out of multiple outputs. It has been used in various aVLSI systems for computing stereo [3], object tracking [4–7] and image compression [2]. Lazzaro and colleagues [1] were the first to implement a hardware model of a WTA network comprising multiple excitatory neurons that are inhibited by a global inhibitory neuron. The network computes a single winner, the identity of which is indicated by the outputs of the excitatory neurons. Localized winners can be obtained by coupling adjacent neurons through resistive lateral connections. Variants of this network that include lateral connections, self-amplication through positive feedback mechanisms, and a cascade configuration have been implemented [4,5,8,9]. Similar networks of coupled excitatory and inhibitory neurons that exhibit soft-max and WTA properties have been used to model different types of cortical processing [10–12]. Such a network has also been used to model the gain-control properties of direction-selective cells in the fly visual system [13].

In this work, we describe a network of multiple excitatory neurons and one inhibitory neuron that performs either a soft-max computation (there is no single winner) or a WTA computation (there is only one winner). In the soft-max regime, the outputs of the excitatory neurons code the relative input strengths: They depend on the number of inputs, the relative input strengths and two parameter settings. They are also normalized with respect to a constant bias current. The global inhibitory signal can also be used as an output. This output saturates with an increasing number of active inputs: The saturation level is independent of the number of inputs and depends only on the input values. The circuity implemented for each neuron is equivalent to that in Lazzaro's WTA network with an additional transistor and a global parameter bias. This bias determines the regime of operation of the network.

The outputs of the excitatory neurons can also code the absolute strength of the inputs by using a variant of this network. Results from two fabricated networks of 20 neurons show the different regimes of operation.

2. Circuit Description of Network

The generic architecture of a recurrent network with excitatory neurons and a single inhibitory neuron is shown in Fig. 1. The excitatory neurons each receive an external input, e_i, and the inhibitory neuron receives

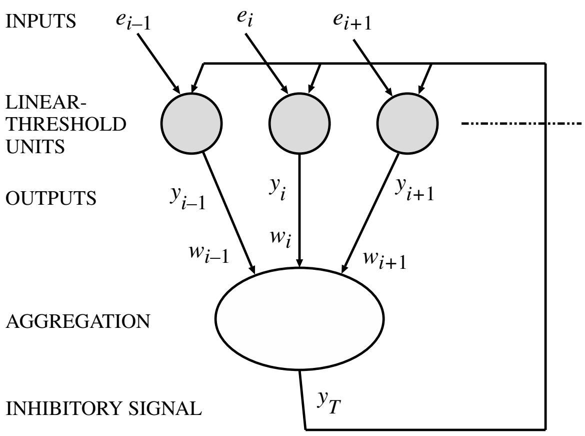


Fig. 1. Architecture of a recurrent network which consists of N linear threshold excitatory neurons (shaded circles) and one global inhibitory neuron (open ellipse). The inputs to the excitatory neurons are described by e. The global inhibitory signal, y_T, to the excitatory neurons, depends on the weights, w, and output states, y, of the neurons.

inputs y_i (weighted by w_i) from the excitatory neurons. The output of the inhibitory neuron, y_T, in turn, inhibits the excitatory neurons.

The circuitry for two excitatory neurons and one inhibitory neuron is shown in Fig. 2. Excitatory neuron 1 which consists of transistors M_1 to M_3, receives an input current, I_1. The state of the neuron is represented by the current, I_{r1} (or the voltage, V_{r1}). Each excitatory neuron is a linear threshold unit because I_{r1} cannot be negative. The inhibitory current, I_T, to each neuron is determined by the output currents, I_{o1} and I_{o2}. These currents sum to the bias current, I_b, supplied by transistor M_4. Note that the current I_T cannot exceed the largest input current. The global parameter V_a determines the regime of operation of the network. In the WTA regime, only one of the I_{oi} currents is equal to I_b and the remaining output currents are zero. In the soft-max regime, more than one of the output currents will be non-zero and the relative magnitudes of these currents depend on V_a. In the next two subsections, we derive the dependence of the output currents and the inhibitory current on V_a and the input currents, I_i.

2.1. Dependence on V_a

The inhibitory current I_T in each neuron is determined by the voltage V_T. Using Kirchhoff's current law at V_T and assuming that the transistors operate in weak inversion, we can solve for V_T in terms of I_b and V_{ri}. The voltage V_{ri} is determined by the input current, I_i, and I_T. We can indirectly compute the dependence of I_T on I_b:

$$I_b = \left(\sum_i^N I_i - N I_T\right)\left(\frac{I_\alpha}{I_T}\right)^{\frac{1}{\kappa}}$$

The parameter κ is the coupling efficiency from the gate to the channel of a transistor in subthreshold, N is the number of "active" excitatory neurons (that is, neurons whose $I_i > I_T$), and $I_\alpha = I_0 e^{\kappa V_a/U_T}$. Assuming that $\kappa = 1$, we can solve for I_T directly:

$$I_T = \frac{I_\alpha \sum_i^N I_i}{I_b + I_\alpha N} = \frac{\sum_i^N I_i}{I_b/I_\alpha + N} \tag{1}$$

This equation shows that I_T is normalized to the number of "active" inputs.

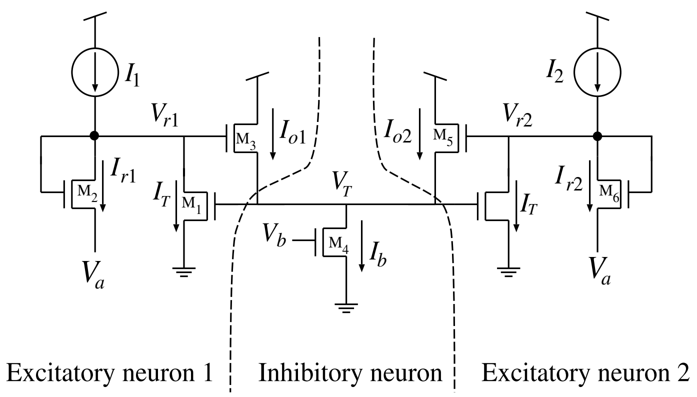


Fig. 2. Circuitry for two excitatory neurons and the global inhibition neuron, M_4. The circuit in each excitatory neuron consists of an input current source, I_1, and transistors, M_1 to M_3. The inhibitory transistor is a source of a fixed current, I_b. The output currents I_{o1} and I_{o2} are normalized with respect to I_b. The width and length of all transistors in the excitatory neuron circuit are 7.2 μm.

We solve for the output currents I_{oi} by using the translinear principle on transistors M_2 to M_6:

$$I_{oi} = \frac{I_{ri}}{\sum_j^N I_{rj}} I_b = \frac{I_i - I_T}{\sum_j^N (I_j - I_T)} I_b \tag{2}$$

Both equations (1) and (2) are valid only when the currents, I_{ri}, are finite and the network is operating in the soft-max regime. In this regime, V_a is less than V_{ri}. If we increase V_a, eventually all the I_{ri} currents go to zero and we can disregard the diode-connected transistors. The network reduces to that of Lazzaro's network and displays the normal WTA response where only one I_{oi} is nonzero. In the soft-max regime, the node V_{ri} in each neuron is a low-impedance (or low-gain) node. In the WTA regime, this node becomes a high-impedance (or high-gain) node: Input current differences are greatly amplified. The gain at node V_{ri} depends on the drain conductances of the transistors and is determined by the Early voltage. For high gain, we can increase the Early voltage of the transistors by making transistor M_1 and the input transistor that supplies I_{in} long.

In this circuit, the output currents I_{oi} are normalized with respect to I_b. If we replace the current source transistor M_4 by a diode-connected transistor, the output currents reflect the relative magnitudes of the input currents. This situation was analyzed in [14].

2.2. Dependence on I_i

In the soft-max regime, the number of "active" neurons that contribute to I_T depends on the relative strengths of the input currents, the parameter V_a, and the bias current I_b. To derive the conditions under which a neuron i is "active," we use equations (1) and (2), and solve for I_{oi} as a function of I_i:

$$I_{oi} = \frac{I_i}{\sum_j^N I_j}(I_b + I_\alpha N) - I_\alpha \tag{3}$$

Noting that I_{oi} must not negative for "active" inputs, we obtain

$$I_i \geq \frac{I_\alpha \sum_j^N I_j}{I_b + N I_\alpha} \tag{4}$$

We look at a specific case of N excitatory neurons, where an α fraction of the neurons receive an input current of magnitude βI_i ($\beta \geq 1$) and the remaining neurons receive an input current of magnitude I_i. Using equation (4), we know that the latter neurons are "active" when the following condition is met:

$$\alpha N(\beta - 1) \leq I_b / I_\alpha \tag{5}$$

The relative magnitudes of the input strengths and the relative number of neurons with input βI_i determine whether the other neurons are "active."

3. Chip Results

A network consisting of 20 excitatory neurons and an inhibitory neuron as shown in Fig. 2 was fabricated in a 1.2 μm CMOS process. The results from this circuit are described in Sections 3.1 and 3.3. A modified network (Fig. 3) consisting of 20 excitatory neurons that are coupled together at the nodes V_{ri} and V_T by horizontal diffusors [15] or pseudoconductances [16] (M_5 and M_6) was fabricated in a 2 μm CMOS process.

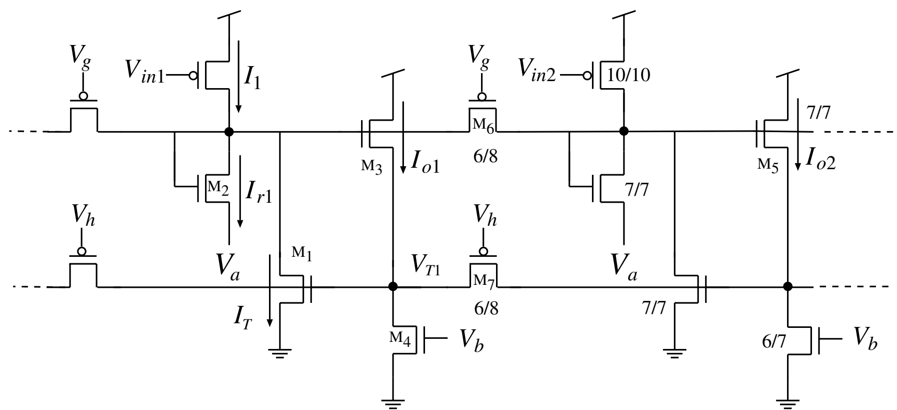


Fig. 3. Network of 20 excitatory neurons that are coupled together by diffusors or pseudoconductances (M_6 and M_7). The inhibitory transistor, M_1, is local to each neuron. The sizes of the transistors are in units of micrometers. This circuit was fabricated in a 2 μm CMOS technology.

Each neuron has its own current source transistor, M_4. The diffusors act as lateral resistors and are biased by V_g and V_h, respectively. This network allows for localized regions of competition. The results from this chip are described in Section 3.2.

3.1. Results from 2-Input Interaction

We looked at the interaction between two input neurons in the different regimes of the network in Fig. 2. The input current to each neuron is supplied by a pFET whose gate voltage is V_{in}. This voltage was set to the same value in two neurons; the remaining neurons receive zero input. We varied V_a (thus changing the regime of operation of the network) and measured the output currents of the neurons. The measured currents, I_{o1} and I_{o2} as a function of V_a are shown in Fig. 4(a). The four curves correspond to four different values of V_{in}. Currents I_{o1} and I_{o2} were equal at a low value of V_a as expected in the soft-max regime. As V_a increased, the ratio of the output currents started to deviate from 1. One of two outputs starts to account for more of the bias current I_b because of a small mismatch between the two input currents. Eventually this output current goes to I_b as V_a was increased further. The value of V_a when the output currents start to deviate from each other depends on the magnitude of the input current (V_{in}). As V_a approaches V_{ri} of the winning neuron, we use Kirchhoff's current law at node V_T and obtain

$$\kappa V_a = V_T + \kappa V_b \tag{6}$$

Because V_T depends on the input current, V_a increases for decreasing V_{in} (increasing input current).

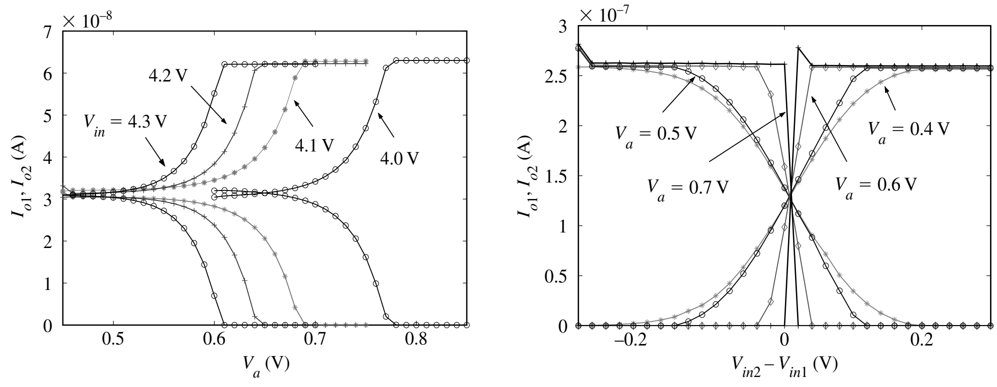
The different regimes of network operation corresponding to the different values of V_a can also be seen by measuring I_{o1} and I_{o2} while varying the differential voltage between the two inputs as shown in Fig. 4(b). Here, V_{in2} was swept differentially around a fixed input voltage of $V_{in1} = 4.3$ V for four parameter settings of V_a. As V_a was increased from 0.4 V to 0.7 V, the differential linear input range decreases from around 400 mV (soft-max regime) to 20 mV (WTA regime).

3.2. Results from Multi-Input Interaction

We show here the interaction between multiple inputs in a network where the pixels are coupled together with diffusors as shown in Fig. 3. Instead of measuring the output currents I_{oi}, we converted these currents into a voltage through an on-chip scanner [17], an off-chip current sense amplifier and a 22 MΩ resistor.

In this experiment, we demonstrate the normalizing behavior of the network in the soft-max regime. The input current of one neuron (which we call the foreground neuron) was set to a higher value ($V_{in} = 3.6$ V) than that of the remaining background neurons ($V_{in} = 3.7$ V). Even though the network allows for local regions of competition, we set the biases for the lateral diffusors, V_h and V_g to 0.153 V and 1.27 V respectively so that the neurons compete for a constant bias current.

The output voltages of the neurons as a function of the number of neurons in the foreground are shown in Fig. 5(a). The four curves correspond to the measured



(a) (b)

Fig. 4. Response of two neurons in the network shown in Fig. 2. The parameter V_a determines whether the network operates in the soft-max regime or the WTA regime. (a) Output currents I_{o1} and I_{o2} as functions of V_a for a subthreshold bias current and $V_{in} = 4.0$ V to 4.3 V. (b) Output currents as functions of the differential input voltage, $V_{in2} - V_{in1}$, with $V_{in1} = 4.3$ V.

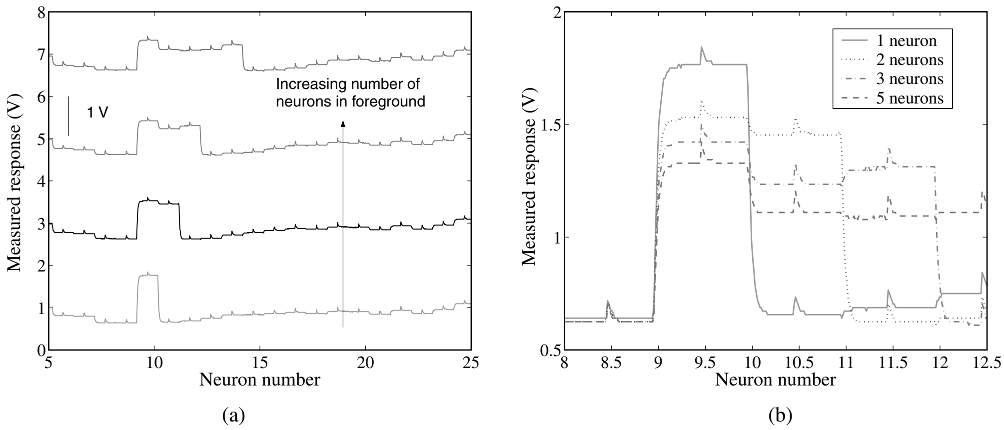


Fig. 5. Response of the network shown in Fig. 3 for an increasing number of neurons (the foreground) that received a larger input current than the remaining neurons. V_a was set so that the network operated in the soft-max regime ($V_a = 0.6$ V). The output currents of the neurons were converted to voltages through an off-chip sense amplifier and a 22 MΩ resistor. (a) Traces corresponding to different numbers of neurons in the foreground have been shifted relative to one another for ease of comparison. The lowermost curve is the network response for a single neuron that received a larger input current ($V_{in} = 3.6$ V) than the remaining neurons ($V_{in} = 3.7$ V). The remaining three curves were the measured output voltages of the neurons when an increasing number of foreground neurons received the larger input current. The topmost curve is the network response for five foreground neurons. (b) Magnified responses of the output voltages of the foreground neurons. The curves show the reduction in the output voltage (solid curve) of the 9th neuron (the initial sole foreground neuron) as more neurons with the larger input current are added. These responses illustrate the normalizing behavior of the network in this regime.

output currents for 1, 2, 3, and 5 foreground neurons. As more neurons were added to the foreground, the output voltage of the initial sole neuron in the foreground decreased as shown by the magnified superimposed curves in Fig. 5(b). The output voltage of the foreground neuron was normalized to the increased number of neurons sharing the same input current.

The network response in the two operating regimes for two spatially separated groups of four neurons whose input currents are higher ($V_{in} = 3.5$ V) than those of the remaining neurons ($V_{in} = 3.7$ V) is shown in the next experiment. The response of the network in the soft-max regime is shown in the lowermost curve in Fig. 6. There are multiple winners as illustrated by the similar output voltages of the neurons in the two groups. The output voltages of the winners are slightly different because of the mismatches in the neuron circuitry. The coefficient of variation (standard deviation/mean) of the actual current outputs was around 3%. As we increased V_a, the network transitions to a WTA regime and only one of the neurons in the two groups wins as shown in the topmost trace of Fig. 6.

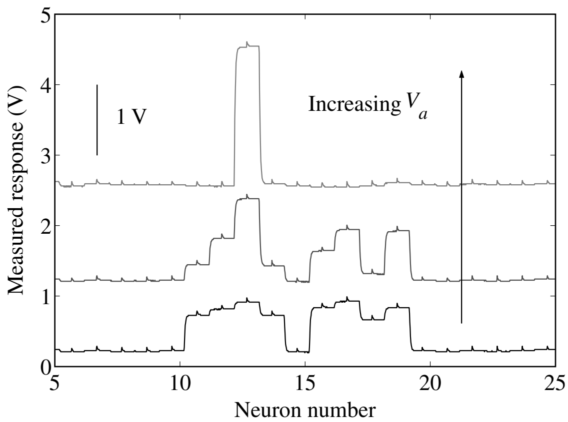


Fig. 6. Response of network in Fig. 2 with two spatially separated groups of four foreground neurons: These groups receive a higher input current ($V_{in} = 3.5$ V) than the remaining neurons ($V_{in} = 3.7$ V). The neurons in the network share a constant bias current. The three curves correspond to three values of V_a (0.85 V, 1.0 V, and 1.1 V). The curves have been shifted relative to one another for ease of comparison. The lowermost curve shows the response of the network operating in the soft-max regime ($V_a = 0.85$ V). The topmost curve shows the response of the network when operated in the WTA regime. One neuron wins and takes all the bias current.

3.3. Response of Common-Node Voltage, V_T

The common-node voltage V_T of the circuit in Fig. 2 reflects the inhibitory current to the excitatory neurons. It codes the strengths of the inputs independent of the number of inputs. In these experiments, we measured V_T of the fabricated circuit as we increased

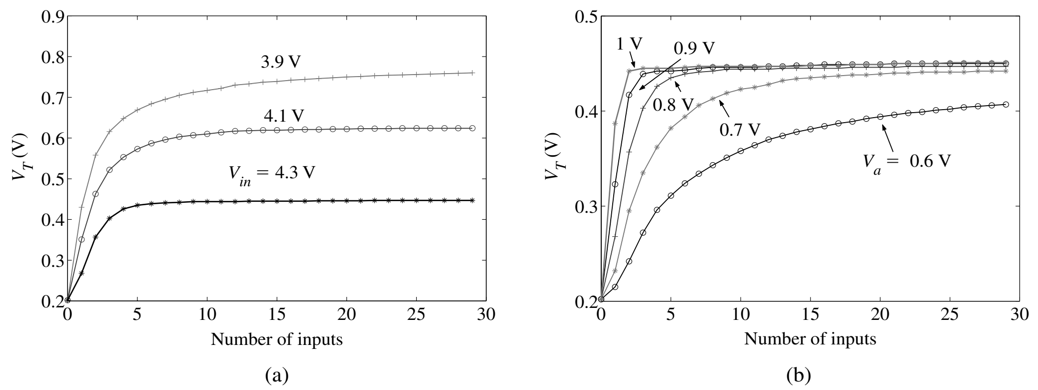


Fig. 7. (a) Common-node voltage V_T as a function of the number of "active" neurons with the same input current. The network was operated in the soft-max regime ($V_a = 0.8$ V, $V_b = 0.7$ V). The voltage V_T codes the input strengths independent of the number of "active" neurons. The saturation value of V_T increases with the input current. (b) The number of neurons at which V_T saturates depends on V_a. The different curves correspond to V_a ranging from 0.6 V to 1 V and $V_{in} = 4.3$ V. As V_a increases, the curve saturates earlier.

the number of the neurons that receive an input current. These measurements (Fig. 7(a)) show that this voltage initially increased and eventually saturated as more neurons received the same input current. This response is described by equation (1). The experiment was repeated for two other input voltages; the value at which V_T saturates depends on the input voltage.

The number of inputs at which V_T saturates is dependent on the ratio, I_b/I_α (described by equation (1)). By holding V_{in} constant and varying V_a (thus I_α), we see that the lowermost curve in Fig. 7(a) saturates at different points as plotted in Fig. 7(b).

4. Conclusion

We described a normalizing aVLSI network with controllable winner-take-all properties. By varying a parameter, the network can transition between a soft-max regime or a winner-take-all regime. A recent aVLSI network by Hahnloser [18] also displays soft-max properties. This network does not exhibit winner-take-all properties unless the neurons receive additional self-excitation. The inhibitory signal is generated via a diode-connected transistor rather than a current source and the neuron circuit uses more transistors. Our network is useful in a signal processing task that requires either soft-max or winner-take-all computation. The global inhibitory signal codes the relative magnitudes of the input strengths in the soft-max regime. The network can be used to model the gain control properties of the direction-selectivity in the fly visual system and the normalizing properties of cortical processing.

Acknowledgments

I thank Rodney Douglas for supporting this work; and Tobias Delbrück and Adrian Whatley for proofreading this document. This work was supported in part by the Swiss National Foundation Research SPP grant and the U.S. Office of Naval Research.

References

1. Lazzaro, J., Ryckebusch, S., Mahowald, M. A. and Mead, C. A., "Winner-take-all networks of O(n) complexity." In: D. Touretzky (ed.), *Advance in Neural Information Processing Systems.* Morgan Kaufmann: San Mateo, CA, 1, pp. 703–711, 1989.
2. Fang, W., Sheu, B., Chen, O. and Choi, J., "A VLSI neural processor for image data-compression using self-organization network." *IEEE Trans. Neural Network* 3(3), pp. 506–518, 1992.
3. Mahowald, M. and Delbruck, T., "An analog VLSI implementation of the Marr-Poggio stereo correspondence algorithm," in *Proc. Intl. Neural Network Soc.*, Boston, MA, 1988.
4. Indiveri, G., "Winner-take-all networks with lateral excitation." *Analog Integrated Circuits and Signal Processing* 13(1/2), pp. 185–193, 1997.
5. Morris, T., Horiuchi, T. and Niebur, E., "Object-based selection within an analog VLSI visual attention system." *IEEE Trans. on Circuits and Systems II* 45(12), pp. 1564–1572, 1998.
6. Horiuchi, T. and Niebur, E., "Conjunction search using a 1-D, analog VLSI-based attentional search/tracking chip." in *Conference on Advanced Research in VLSI,* Atlanta, GA, 20–24 March 1999.

7. Indiveri, G., Whatley, A. M. and Kramer, J., "A reconfigurable neuromorphic VLSI multi-chip system applied to visual motion computation," in *Proceedings of the Seventh Int. Conference on Microelectronics for Neural, Fuzzy and Bio-inspired Systems; Microneuro'99*. IEEE Computer Society: Los Alamitos, CA, pp. 37–44, 1999.
8. Choi, J. and Sheu, B., "A high-precision VLSI winner-take-all circuit for self-organizing neural networks." *IEEE Trans. Circuits and Systems II* 28(5), pp. 576–583, 1993.
9. Starzyk, J. and Fang, X., "CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback." *Electronics Letters* 29(10), pp. 908–910, 1993.
10. Amari, S. and Arbib, M., "Competition and cooperation in neural nets." In: J. Metzler (ed.), *Systems Neuroscience*. Academic Press: New York, pp. 119–165, 1977.
11. Grossberg, W., "Nonlinear neural networks: Principles, mechanisms, and architectures." *Neural Networks* 1, pp. 17–61, 1988.
12. Salinas, E. and Abbott, L., "A model of multiplicative neural responses in parietal cortex." in *Proc. Natl. Acad. Sci. USA* 93, pp. 11956–11961, 1996.
13. Reichardt, W., Poggio, T. and Hausen, K. "Figure-ground discrimination by relative movement in the visual system of the fly, Part II: Towards the neural circuitry." *Biological Cybernetics* 46, pp. 1–30, 1983.
14. Liu, S.-C., "A winner-take-all circuit with controllable soft max property." In: S. A. Solla, T. K. Leen, K.-R. Müller (eds.), *Advances in Neural Information Processing Systems*. MIT Press: Cambridge, MA, 12, pp. 717–723, 2000.
15. Andreou, A. and Boahen, K., "Translinear circuits in subthreshold MOS." *Analog Integrated Circuits and Signal Processing* 9, pp. 141–166, 1996.
16. Vittoz, E. A., "Pseudo-resistive networks and their applications to analog collective computation," in *Proceedings of MicroNeuro '97*, Dresden, pp. 163–173, 1997.
17. Mead, C. and Delbrück, T., "Scanners for visualizing activity of analog VLSI circuitry." *Analog VLSI and Signal Processing* 1, pp. 93–106, 1991.
18. Hahnloser, R., "About the piecewise analysis of networks of linear threshold neurons." *Neural Networks* 11, pp. 691–697, 1998.

Shih-Chii Liu received the B.Sc. degree in electrical engineering from Massachusetts Institute of Technology in 1983, and the M.S. degree in electrical engineering from University of California, Los Angeles in 1988. She received the Ph.D. degree in the Computation and Neural Systems Program from California Institute of Technology in 1997. She is currently a Ueberassistentin at the Institute of Neuroinformatics, ETH/UNIZ in Zurich, Switzerland.

Dr. Liu was with Gould American Microsystems from 1983 to 1985, and with LSI Logic from 1985 to 1988. She worked at Rockwell International Research Labs from 1988 to 1997. Her research interests include neuromorphic modeling of visual and cortical processing, networks for behavior generation, and hybrid analog/digital signal processing.