A SILICON RETINA 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

ABSTRACT

The winner-take-all (WTA) circuit is a useful computational circuit for signal processing and learning tasks. By adding spatial coupling between pixels, local regions of competition can be delineated. Recently, we described a normalising circuit which enhances Lazzaro's WTA circuit (1989) only with the addition of a transistor and a global bias to each pixel. This new circuit allows the network to transition between a soft-max and a WTA function through the global bias. The WTA network of Lazzaro together with spatial coupling forms the current-mode silicon retina of Boahen and Andreou (1992). This retina models the center-surround processing performed in biological retinas to enhance responses to spatial contrasts. Here, we show how our normalising network together with spatial coupling performs as a silicon retina. Results are presented from a fabricated circuit in a 2μm CMOS process.

1. INTRODUCTION

The winner-take-all (WTA) function is a useful computation in self-organizing neural networks and signal processing applications. It selects a single winner out of multiple inputs. It has been used in various aVLSI systems for computing stereo,object tracking,and image compression.Lazzaro and colleagues [1] were the first to implement a hardware model of a winner-take-all (WTA) network. Their network consists of N excitatory neurons that are inhibited by a global inhibitory neuron. It computes a single winner, the identity of which is indicated by the outputs of the excitatory cells. Localized winners can be obtained by coupling neurons together through lateral resistive connections. The properties of this network have been enriched with the addition of lateral connections and positive feedback mechanisms [2, 3, 4, 5]. This network is also the basis of the current-mode silicon retina of [7] where the inputs come from photocurrents. This retina models the center-surround processing found in biological retinas to amplify high spatial contrasts.

Recently, we described a network that performs either a soft-max or a winner-take-all function depending on a global parameter [6]. This circuit is similar to Lazzaro's WTA circuit except for the addition of a diode-connected transistor and a global bias. In this work, we describe how our network together with spatial coupling performs with normal input currents or with photocurrents. In the latter case, the network is a silicon retina with a similar architecture to the silicon retina of [7]. Results both from analysis and from the fabricated chip show that there is a smaller dependence of the smoothing space constant of the network on background intensity when compared to the space constant of the retina from [7]. The chip was fabricated in a 2μm CMOS technology process.

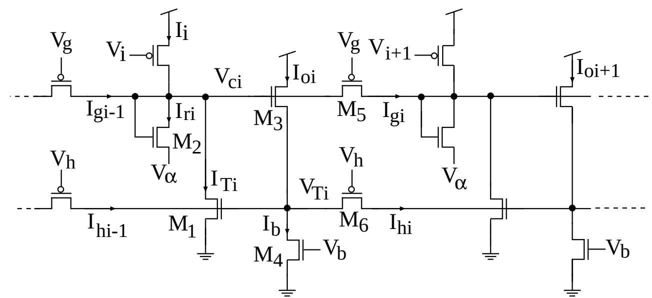

Fig. 1. Circuitry showing coupling between two excitatory neurons in an array of N excitatory neurons (in this paper, N=20). The inhibitory circuit is local to each pixel. The circuit in each excitatory neuron consists of consists of an input current source, I_i, and transistors, M_1 to M_3. The inhibitory transistor is a fixed current source, I_b through M_4. The input to the inhibitory transistor, I_{oi} is normalized with respect to NI_b.

2. CIRCUIT DESCRIPTION

The circuitry for two of the N excitatory neurons and their local inhibitory circuit is shown in Fig. 1. Each excitatory neuron is a linear threshold unit and consists of a pFET that supplies the input current, I_i, and transistors, M_1 to M_4. We ignore the spatial coupling transistors, M_5 and M_6 in each pixel for the moment and short all V_{Ti} nodes such that $V_{Ti} = V_T$. The circuit in Lazzaro's pixel is similar to this circuit but without the transistor, M_2. This transistor, M_2, introduces a rectifying nonlinearity into the system since I_{ri} cannot be negative. The inhibition current, I_{Ti}, to each neuron is determined by the gate voltage, V_{Ti}, which in turn is determined by the input current, I_i through the transistor M_3. In the hard WTA regime, the neuron with the largest I_i sets I_{Ti} for all neurons. Therefore, only the corresponding transistor M_3 in the winning neuron is not in cutoff, and its output current, I_{oi} is equal to the total bias current, NI_b.

The parameter, V_α, determines whether the circuit in Fig. 1 computes the soft-max or WTA function. In the soft-max regime, more than one I_{oi} can be positive and the relative sizes of the I_{oi} is dependent on I_i, V_α and NI_b. We can compute the "active inputs" that set the common voltage, V_T in each neuron and solve for the output current, I_{oi} in terms of I_i:

$$I_{oi} = \frac{I_i}{\sum_j^N I_j}(I_B + I_\alpha N) - I_\alpha \tag{1}$$

where $I_\alpha = I_0\, e^{kV_\alpha/U_T}$, U_T is the thermal voltage, and I_0 is the pre-exponential constant of the subthreshold current equation. In deriving this equation, we assume that the transistors operate in subthreshold and κ (the coupling efficiency of the gate on the channel of a transistor in subthreshold) is equal to 1.

Noting that I_{oi} cannot be non negative for "active" inputs, we get the condition:

$$I_i \geq \frac{I_\alpha \sum_j^N I_j}{I_B + NI_\alpha}. \tag{2}$$

Equation 2 describes the condition under which a neuron stays "active" or its I_{oi} is positive.

2.1. Center-surround property

As previously shown [7], this two-layered network in which the pixels are coupled by diffusors M_5 and M_6, performs a center-surround computation. This computation is analogous to a difference of Gaussians operation to extract local high spatial contrasts. The top layer in Fig. 1 which receives the current inputs performs spatial smoothing on the inputs through the diffusors which are controlled by a global voltage V_g. The bottom layer performs spatial smoothing on the outputs, I_{oi} of the top layer through the diffusors which are controlled by a global voltage V_h. The output of this layer in return inhibits the top network. By setting the diffusor biases V_g and V_h such that the bottom network has a bigger spatial spreading constant than the top network, we can approximate the function of the two-layered network to that of the difference of Gaussians. The equations for the currents

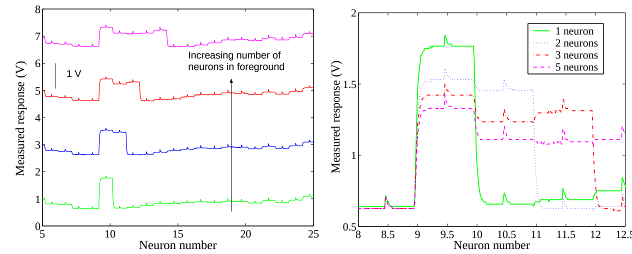

Fig. 2. Response of the network in Fig. 1 for an increasing number of neurons/pixels that receive a higher input current (considered the foreground) than the remaining neurons/pixels. The parameter V_α was set to 0.6V so that the network operated in the soft-max regime. (a) Responses to an increasing number of neurons in the foreground. The traces have been shifted relative to one another for ease of comparison. The lowermost curve is the network response for one neuron that received a larger input current (V_{in}=3.6V) than the remaining neurons (V_{in}=3.7V). The remaining three curves are obtained with an increasing number of neurons in the foreground that received the larger input current. The topmost curve is the network response for five neurons in the foreground. (b) Magnified responses of the foreground neurons. Notice the reduction in the response of the initial sole foreground neuron (the solid curve) as more neurons were added to the foreground. The figures have been adapted from Fig. 5 in [6] with permission.

in both layers of this network are

$$\begin{aligned} I_{Ti} + I_{ri} &= I_i + M\big(\nabla^2 I_{ri}^{1/\kappa}\big) \qquad (3)\\ I_{oi} &= I_b - N\big(\nabla^2 I_{Ti}^{1/\kappa}\big). \qquad (4) \end{aligned}$$

where $M \propto e^{V_\alpha}\, e^{(\kappa_p - 1)V_{dd} - \kappa_p V_g)}$. The spatial constant N is proportional to $e^{(\kappa_p-1)V_{dd}-\kappa_p V_h)}$. Instead of a bi-harmonic operator as in Boahen and Andreou's network, the smoothing function in the top network is a Laplacian operator. The spatial smoothing constant, M and N, does not depend on the input magnitude. Ideally, the space constant should not increase with the input current. However, in reality, the current through the diffusors depend on the κ of the corresponding transistor M_5 in each pixel. (The gate coupling efficiency of the transistor, κ, changes with the magnitude of the current through the transistor.) When the input increases, the larger voltages at the nodes on the top layer lead to an increase in κ (and hence the lateral current). Hence in the network of [7], the spatial smoothing of the photocurrent inputs increases as the background intensity increases. This dependence is unlike that in biological retinas where the spatial smoothing of the inputs increases under low background intensity. In our network, because of the low impedance of the corresponding node, V_{ci}, in each pixel, the change in the voltage at V_{ci} is smaller for

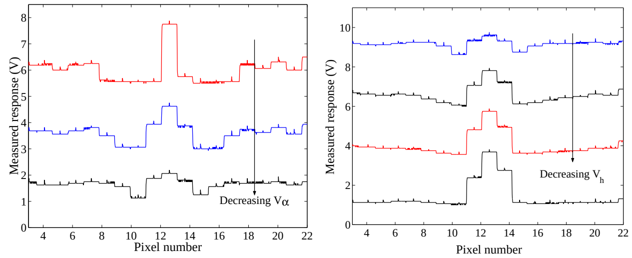

Fig. 3. Response of the network with input photocurrents. The traces have been shifted relative to one another for ease of comparison. A stimulus with a single bright strip was placed over the chip such that only one pixel was stimulated. (a) Network response of the silicon retina for V_α = 0.8V, 0.937V, and 1.2V. The lowestmost trace corresponds to the response of the network when operating in the soft-max regime while the uppermost trace corresponds to the network operation in the WTA regime. (b) Network response for different values of the diffusor voltage (V_h=0V, 0.046V, 0.081V, and 1.53V). V_g=1.28V and V_α = 0.8V.

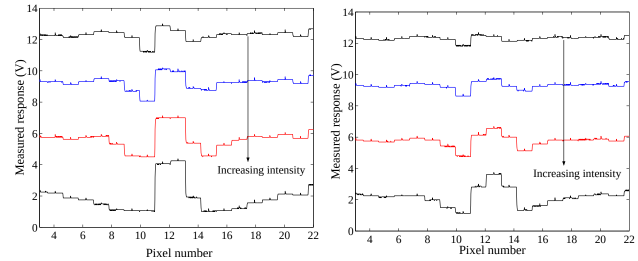

Fig. 4. Response of the network for different light intensities. The traces have been shifted relative to one another for ease of comparison. (a) Response of the silicon retina in the WTA regime (V_α=0.9V). The uppermost curve was obtained at the lowest background intensity. Successive curves were obtained at intensities which are a decade apart. We can see that the spatial smoothing of the network increased with increasing background intensity. (b) Response of retina in the soft-max regime (V_α=0.814V). The space constant was almost invariant over the four decades of background intensity.

the same increase in photocurrent input. Thus, the gain and the smoothing constant of the top layer does not increase as much with the background intensity as in the network of [7].

3. MEASURED RESULTS FROM CHIP

A network consisting of 20 pixels as shown in Fig. 1 was fabricated in a 2μm CMOS process. We describe results from the chip using current source inputs to illustrate how the network can transition between a soft-max function and a winner-take-all function. We then describe results from the same network with photocurrents as the inputs.

3.1. Results from current source inputs

We eliminate the center-surround properties of the network by setting $V_h = 0V$ and $V_g = 5V$. The output currents, I_{oi}, of the neurons were read through an on-chip scanner.These currents were converted to a voltage using an off-chip current sense amplifier and a 22 MΩ resistor.

We describe experiments that show the soft-max property of the network. The details of the other regimes of operation are given in [6]. We set the parameter V_α so that the network operated in the soft-max regime (V_α=0.6V). The input voltages, V_{in} of all the neurons (we called them the background neurons) except for one were set to 3.7V. The input voltage of the remaining neuron (which we call the foreground neuron) was set such that the neuron received a larger input current. The output response of the network for the sole foreground neuron is shown in the lowermost trace in Fig. 2(a). The network response for an increasing number of neurons in the foreground are shown by the remaining traces in Fig. 2(a). As more neurons were added to the foreground, the output current of the initial neuron in the foreground decreased as shown by the magnified responses in Fig. 2(b). The response of the network with a sole neuron (9th neuron) in the foreground is given by the solid curve. The response of this initial neuron decreased with the number of increasing foreground neurons. The responses in Fig. 2 show that the soft-max function of the network. The output currents, I_{oi}, depend on the relative magnitude of the input currents. There is no single winner as in a winner-take-all network.

3.2. Results from silicon retina

The inputs to the network now come from photodiodes (the pFET driven by V_i in each pixel is now switched off) and the network acts as a silicon retina with center-surround properties. By increasing V_α, we can change the operation of the network from the soft-max regime to the WTA regime as shown in Fig. 3(a). In these traces, a stimulus with a single bright strip was placed over the chip such that only one pixel was stimulated. In the uppermost trace where the network operated in the WTA regime, this pixel had the largest response. In the lower traces, the output of this pixel decreased dependent on the relative magnitudes of the photocurrents. The shape of the center-surround kernel of the network can be altered through the relative voltage difference between V_g and V_h as described in Sect. 2.1. By decreasing V_h (this corresponds to an increase in the spatial smoothing in the bottom layer), the inhibitory surround increases as shown in the lowermost trace of Fig. 3(b) and the

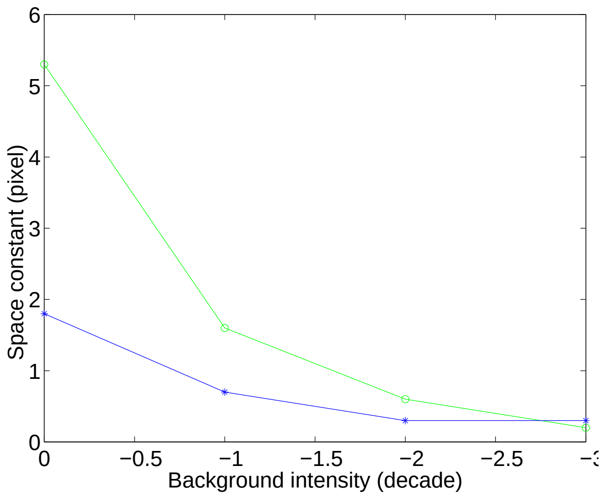

Fig. 5. The network space constant plotted against four decades of background intensity. The ordinate "0" is the brightest intensity. The remaining data points correspond to decreasing background intensity in decades. The space constant was obtained from Fig. 4. The line marked with circles is derived from the winner-take-all data and the line marked with asterisks is derived from the soft-max data. The network space constant is the interpolated pixel value where the retina response drops by 1/e from the peak response.

pixel which saw the bright strip had the largest response.

As we have seen in Fig. 3(a), changing V_α changes the function of the network. When we set V_α so that the network acts as a WTA, it is equivalent to the network of [7]. As pointed out in Sect. 2.1, the spatial spread of the impulse response of the network increases with increasing background intensity. This behavior is shown in Fig. 4(a). The opposite type of dependence is observed in the biological retina. The biological retina has a larger spatial spread under low background intensities because it needs to collect photons from a larger spatial area to get a good S/N ratio response. However, the silicon retina shows the largest spatial spread for the highest background intensity.

By decreasing V_α so that the network operates in the soft-max regime, the spatial space constant is less invariant over four decades of intensity as shown in Fig. 4(b) because of the low impedance nature of V_{ci} in Fig. 1. This smaller invariance is also depicted in Fig. 5 which shows the network space constant over the 4 decades of intensity. This data was derived from Fig. 4.

4. CONCLUSION

We described the response of a silicon retina that displays different spatial smoothing characteristics when tuned for a soft-max computation or a winner-take-all computation. When the network is tuned for a winner-take-all computation, the space constant of the smoothing increases with higher background intensities. This characteristic is undesirable for modelling the smoothing properties of the biological retina. However, if the network is tuned for a soft-max computation, the space constant of the smoothing is almost invariant to background intensity.

5. ACKNOWLEDGMENTS

I thank Rodney Douglas for supporting this work. I also thank Tobias Delbrück for critical reading of the article. This work was supported in part by the Swiss National Foundation Research SPP grant and the U.S. Office of Naval Research.

6. REFERENCES

[1] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of O(n) complexity," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 1, pp. 703–711. Morgan Kaufmann, San Mateo, CA, 1989.

[2] T.G. Morris, T. Horiuchi, and E. Niebur, "Object-based selection within an analog VLSI visual attention system," *IEEE Trans. Circuits and Systems II*, vol. 45, no. 12, pp. 1564–1572, 1998.

[3] J. Choi and B. Sheu, "A high-precision VLSI winner-take-all circuit for self-organizing neural networks," *IEEE Trans. Circuits and Systems II*, vol. 28, no. 5, pp. 576–583, 1993.

[4] J.A. Starzyk and X. Fang, "CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback," *Electronics Letters*, vol. 29, no. 10, pp. 908–910, 1993.

[5] G. Indiveri, "Winner-take-all networks with lateral excitation," *Analog Integrated Circuits and Signal Processing*, vol. 13, no. 1/2, pp. 185–193, 1997.

[6] S.-C. Liu, "A normalizing aVLSI network with controllable winner-take-all properties," *Analog Integrated Circuits and Signal Processing*, vol. 31, no. 1, pp. 47–53, 2002.

[7] K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," *Advances in Neural Information Processing Systems*, vol. 4, pp. 764–772, 1992.