A COMPETITIVE NETWORK OF SPIKING VLSI NEURONS

Indiveri G.[1], Horiuchi T.[2], Niebur, E.[3], and Douglas, R.[1]

[1]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland

[2]Computational Sensorimotor Systems Laboratory, Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, USA

[3]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, USA

giacomo@ini.phys.ethz.ch timmer@isr.umd.edu niebur@jhu.edu rjd@ini.phys.ethz.ch

Summary

We present a VLSI implementation of a competitive network of integrate and fire neurons, internally configured with local, recurrent excitation and with global feedback inhibition. The VLSI device employs a communication framework for receiving and sending spikes off-chip, implemented using hybrid analog/digital circuits. We characterized the network's main properties by stimulating the neurons with Poisson distributed spike trains. Using the mean-rate interpretation, we demonstrate how the system is capable of exhibiting winner-take-all behaviors. We also show how the VLSI network is able to enhance correlations in the output signals and filter signals on the basis of input correlation, by exploiting the timing of spikes.

Introduction

Networks of integrate and fire (I&F) neurons have been extensively studied using both analytical and numerical simulation tools (Maass and Bishop, 1998). An increasing number of researchers use software implementations of these networks as tools for investigating the effect of spike timing and synchrony on the network's computational properties (Bernander et al., 1994; Niebur and Koch, 1994; König et al., 1995; Niebur and Koch, 1996; Diesmann et al., 1999). The importance of these approaches is underscored by recent experimental results that support the idea that the temporal fine structure of neural spike trains, in particular the correlation between firing times of different neurons, is used for the representation of

sensory input (Stopfer et al., 1997) or internal states, in particular, selective attention (Steinmetz et al., 2000; Fries et al., 2001).

While software simulations are effective for analyzing networks of neurons that use rate codes as the principal representation, detailed simulations of networks of spiking neurons is a CPU-intensive process and can still require a significant amount of simulation time. We present a set of hybrid analog/digital VLSI neural circuits, that allow us to overcome these problems, by exploiting the advantages of both highly parallel analog computation and high-speed asynchronous digital VLSI techniques. Specifically, we present a VLSI device that contains a competitive network of I&F neurons, implemented using analog circuits, and a communication framework for receiving and sending spikes off-chip, implemented using hybrid analog/digital circuits.

The network's architecture consists of a ring of 32 excitatory neurons that project to a global inhibitory neuron. The inhibitory neuron in turn projects its output spikes back to all the excitatory neurons. The excitatory neurons are mutually connected to their nearest neighbors. Each neuron can also be excited or inhibited via additional synaptic circuits that can be accessed by spikes arriving from outside the chip.

Architectures of this type have been shown to exhibit a broad range of interesting computational properties (Hahnloser et al., 2000; Salinas and Abbott, 1996). Hahnloser et al. (2000) present a similar network architecture, but in which the neurons are modeled with simple linear threshold units. In the mean-rate interpretation of neural firing, the network we propose can exhibit a winner-take-all behavior and reproduce other results presented by Hahnloser et al. (2000). In the more realistic spiking interpretation of the network, the system can also produce behaviors such as adaptive thresholding, single-neuron winner-take-all behavior, and synchrony filtering. The temporal dependence of the inhibitory interactions produces a variable selectivity to synchronous events as well as a locking behavior that shifts output spikes closely in time to the input spikes.

The VLSI device

The VLSI device is a 2×2mm^2 chip, implemented using standard 1.2μm CMOS technology. It contains analog circuits implementing the neurons and synapses present in the competitive network, and hybrid analog/digital circuits to implement the input/output communication infrastructure. The hybrid analog/digital interfacing circuits allow the network's neurons to receive and transmit signals in the form of asynchronous spikes. This is ac-
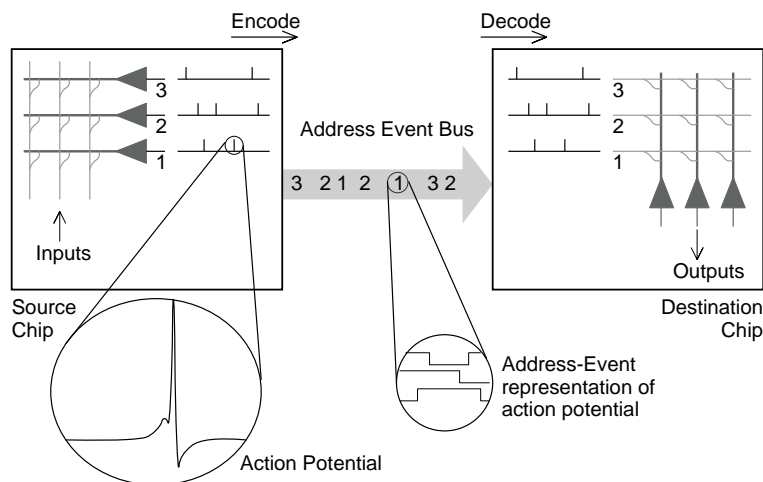
Figure 1: Schematic diagram of an AER chip to chip communication example. As soon as a sending node on the source chip generates an event its address is written on the Address-Event Bus. The destination chip decodes the address-events as they arrive and routes them to the corresponding receiving nodes.

complished by the use of the Address-Event Representation (AER) communication protocol (Mahowald, 1994; Boahen, 1998; Mortara, 1998; Deiss et al., 1998) .

The Address-Event Representation

In this representation, input and output signals ("address events") are sent to (and from) the chip using digital pulses that encode the address of the sending node and carry the analog information in their temporal structure (see Fig. 1). Address-events are stereotyped non-clocked digital pulses. Analog information is carried in the temporal structure of the inter-signal intervals, very much as it is believed to be conveyed in natural spike trains in biological neural systems.

    With appropriate circuitry arbitrating conflicts between simultaneous events, at the *onset* of a spike, the address of the neuron that fired is transmitted quickly through the channel (within 100 to 500 ns) to a circuit on the receiving chip. Because the channel only sends the addresses of active units, the system's bandwidth is devoted entirely to instantaneously active (spiking) units; non-spiking units perform local computations but do not load the inter-chip communication channels.

    An important consequence of using a digital chip-interconnect scheme
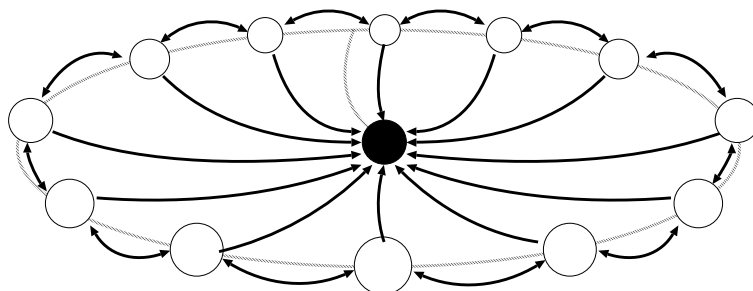
Figure 2: Architecture of the integrate-and-fire ring of neurons chip. Empty circle represent excitatory neurons. The filled circle represents the global inhibitory neuron. The gray line symbolizes inhibitory connections, from the inhibitory neuron to all excitatory neurons. Black arrows denote excitatory connections.

is the relative ease with which these chips are able to interface to existing digital hardware. From the simulation of input spike trains to quickly reconfiguring a network's connectivity via address routers, the flexibility of software can be used to produce a more powerful modeling tool. From the engineering perspective, the translation of our analog signals into a stream of asynchronous spikes not only facilitates communication, it opens up new possibilities for the efficient implementation of both computation and memory in the spike domain.

The competitive network

A schematic diagram of the network architecture is depicted in Fig. 2. The excitatory neurons (empty circles), as well as the inhibitory neuron (filled circle) are implemented using leaky integrate-and-fire neurons, while the synapses are implemented via non-linear integrators.

The network consists of an array of 32 excitatory neurons that project to a global inhibitory neuron, which in turn projects back onto all 32 excitatory neurons. Each excitatory neuron receives its input current from two nearest neighbor excitatory neurons, from the global inhibitory neuron and from both excitatory and inhibitory synapses that can be stimulated from external inputs, through the AER protocol.

The coupling strength of each type of synapse (external excitatory, external inhibitory, internal excitatory-to-excitatory, internal excitatory-to-inhibitory and inhibitory-to-excitatory) and the duration of the synaptic currents can be set by external bias voltages.

As can be seen from Figure 2, we use closed (periodic) boundary conditions for the network: the leftmost neuron receives spikes from the output of the chip's rightmost neuron. Similarly, the rightmost neuron receives spikes from the leftmost neuron on the chip.

Synapses and Integrate-and-Fire Neurons

Voltage pulses (spikes) are integrated into analog currents by a synaptic circuit model. Each synapse is implemented using an extremely compact circuit known as the *current-mirror integrator* (Boahen, 1998). This circuit uses only 4 transistors and one capacitor and allows the user to manipulate both the weight and time constant of the synapse. Once the voltage spikes are converted into currents, the current is injected into the integrate-and-fire neurons.

The circuits we use to implement the neuron are based on those proposed by Mead (Mead, 1989) and by van Schaik (van Schaik et al., 1996) and have been described in detail in (Indiveri, 2000). If a constant DC current is injected into the neuron's membrane capacitance, its membrane voltage V_{mem} increases linearly with time, until it reaches the threshold voltage V_{thr}. When V_{mem} rises above V_{thr}, a positive feed-back loop drives the circuit's output voltage V_{out} to the positive power supply rail, and abruptly increases V_{mem} by a constant factor directly proportional to the power supply voltage and inversely proportional to the membrane capacitance. If V_{out} is high, a reset transistor allows the membrane capacitance to be discharged, decreasing V_{mem} linearly until it reaches V_{thr}. The positive feed-back loop now acts in the opposite direction, bringing both V_{out} and V_{mem} back to zero. The neuron's circuits include a mechanism for setting a *refractory period*, in which the input injection current has no effect on the membrane voltage V_{mem}. As soon as the refractory period is over, the neuron starts integrating the input current again and spikes are generated in the way described above.

Figure 3(a) shows the membrane voltage V_{mem} of one of the neurons of the network, in response to a series of Poisson-distributed pulses sent to an excitatory synapse connected to it. The neuron's threshold voltage V_{thr} was set to 2.5V and its refractory period to approximately 1ms. In Fig. 3(b) we plot the neuron's output spike frequency versus the frequency of the input spikes, for three different values of synaptic weight.
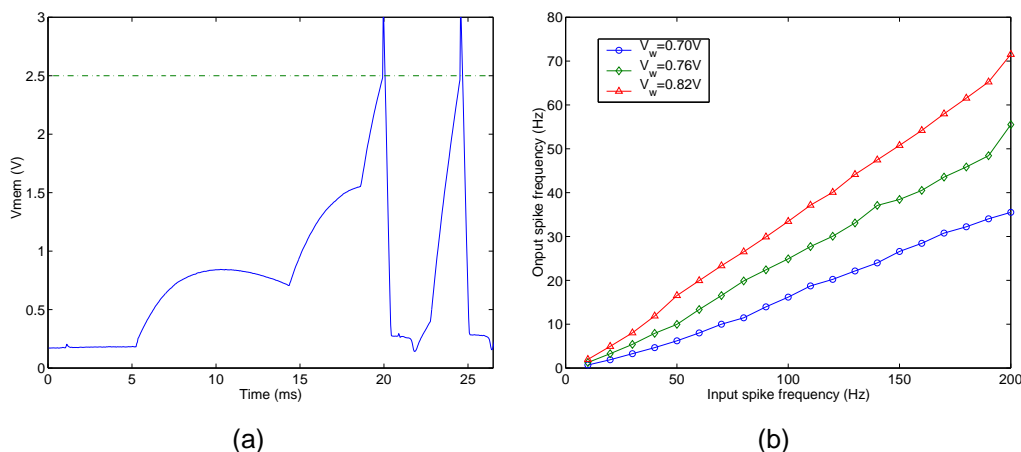
Figure 3: (a) Membrane voltage of an integrate and fire neuron, in response to Poisson-distributed input spikes. The dashed line at 2.5V represents the neuron's threshold voltage. (b) Frequencies of the neuron's output spikes in response to uniformly distributed input spikes. The response gain increases with increasing values of synaptic weight bias settings V_w.

Experimental results

We characterized the behavior of the network by stimulating groups of neurons with Poisson-distributed spike trains generated from a workstation and sent to the chip using the Address-Event Representation. We performed two sets of experiments. The first one was aimed at characterizing the network by analyzing the neuron's *mean firing rate*, in response to different types of input patterns, and as a function of different connectivity conditions. The second sets of experiments was aimed at characterizing the detailed *timing* properties of the network, for example arising from synchronization of output spike patterns or correlations among them.

Mean-Rate Winner-Take-All

To characterize the behavior of the network using the mean-rate representation we stimulate different neurons on the chip with Poisson-distributed spike trains using long EPSPs and no inherent leak, and read out the output spikes via the AER output bus. In order to evaluate the effect of lateral coupling among neurons we stimulated both isolated neurons and groups
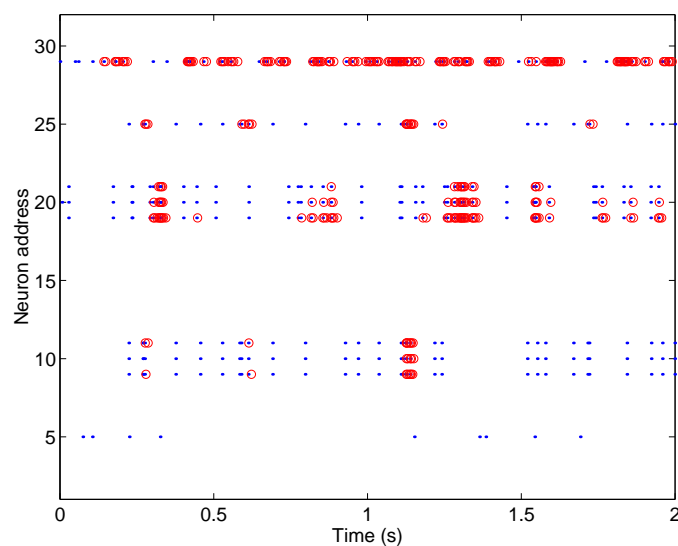
of neighboring neurons. Furthermore, we used different spike-rates for different groups of neurons, to bias the competition. Specifically, we stimulated neuron #5 with spikes-trains with a mean rate of 5Hz; neurons #9, #10, #11 and #25 with synchronous spikes at a rate of 14Hz; neurons #19, #20 and #21 with synchronous spikes at a rate of 25Hz and neuron #29 with spikes at a mean rate of 37Hz.

In Fig. 4(a) we show a raster plot displaying both the occurrences of input spikes (small dots) and output spikes (red empty circles) in the case in which global inhibition is effectively switched off (by setting the strength of the local inhibitory synapses of each neuron to zero). If the strength of the local inhibitory synapses is increased to a relatively high value, the network exhibits a winner-take-all type of behavior: the neuron receiving the strongest excitation has the highest output spike rate and inhibits all other neurons. If we also increase the weight of the local (lateral) excitatory synapses, the network's winner-take-all behavior is even more pronounced due to the recruitment of neighboring neurons. This is seen in Fig. 4(b) where the top panel shows the firing frequency of input spikes for all 32 neurons. The middle trace shows the output frequencies of each of those neurons in the absence of local excitation, and the lower trace the same in the presence of local excitation. Note that the neuron with the highest input, #29 in the simulation generates the most output spikes and that the output of neurons with low firing rates may be completely suppressed. Note, further, that this effect is more pronounced in the case with local excitatory connections and that it spreads to nearest neighbors (#28 and 30) even though they do not receive any input spikes. In this parameter range, the details of spike timing do not play a significant role in determining the winning output.
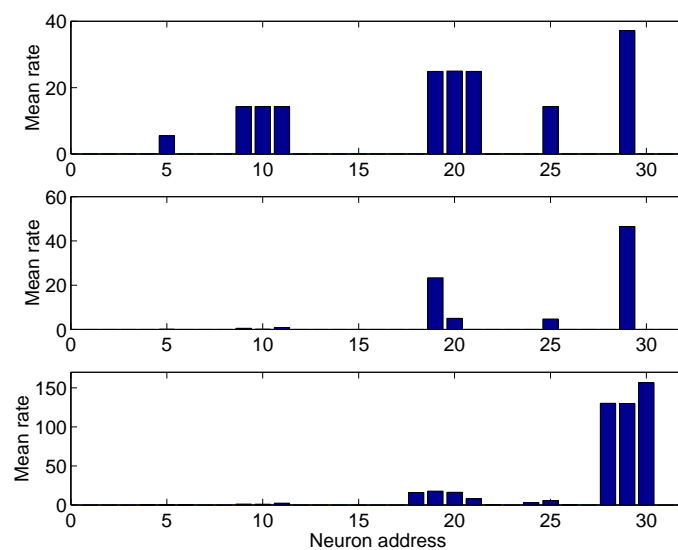
Producing Synchrony

While the amplifying effects of synchronous inputs on leaky integrate-and-fire neurons are well known, good techniques for controlling (inserting) synchrony are less well understood. In the following experiment we show the ability of this network to produce correlated firing combined with the winner-take-all function. To investigate this effect, we simultaneously[1] stimulated three neighboring neurons (neurons #9, #10 and #11, in the example of Fig. 5) with Poisson distributed spike trains at a mean rate of 30Hz. We also stimulate neurons #8, #11, #12 and #13 with independent

[1]"Simultaneous stimulation" in our experimental setup means transmitting address-events with delays of approximately 30μs between each other.

Figure 4: (a) Raster plot of input spike trains (small dots) superimposed onto the output spike trains (empty circles), with global inhibitory feedback turned off (the inhibitory-to-excitatory synaptic weights are set to zero). (b) Histograms of input spike distribution (top trace), output spike distribution of competitive network with global inhibition but no lateral excitation (middle trace) and output spike distribution of competitive network with global inhibition and with lateral excitation (bottom trace)
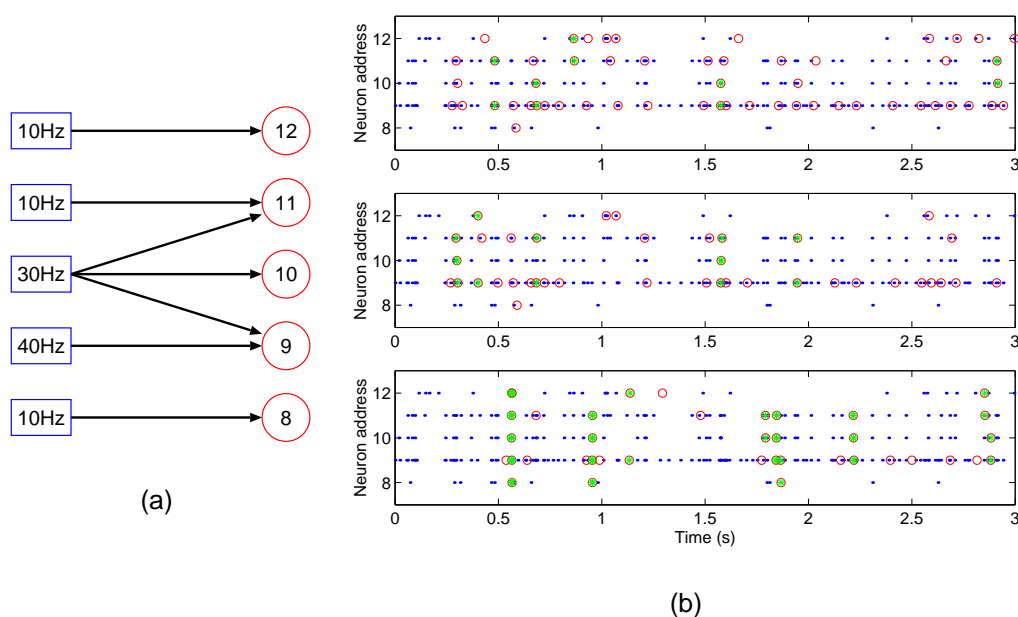
Figure 5: (a) Arrangement of input signals used to stimulate a set of neurons of the network. Each box represents a Poisson distributed spike train source. (b) Raster plots representing input spikes (small dots), output spikes (empty circles), and coincident (within 1ms time window) output spikes (filled circles) for the three network configurations: Without global inhibition (top raster plot), with global inhibition (middle raster plot) and with global inhibition and local excitation (bottom raster plot).

Poisson input spikes at a mean rate of 10Hz, and neuron #9 with an additional independent Poisson spike train with a mean rate of 40Hz (see Fig. 5a). We applied this input stimulus pattern to the network for three different sets of synaptic weights. In the first case we applied the stimuli *without* global inhibition (*i.e.* with the inhibitory synapses set to zero). In a second case we applied the input stimuli *with* global inhibition activated and in the final case we stimulated the network with both global inhibition and local excitation. Figure 5(b) shows raster plots of both input and output spikes for the three different cases. In the absence of global inhibition (top trace of Fig. 5b), the outputs of neurons #9, #10, and #11, which are partially stimulated by the same input, are not well synchronized. As shown by the middle plot of Fig. 5(b), global inhibition partly increases the amount of synchronization between these neurons, and the subsequent addition
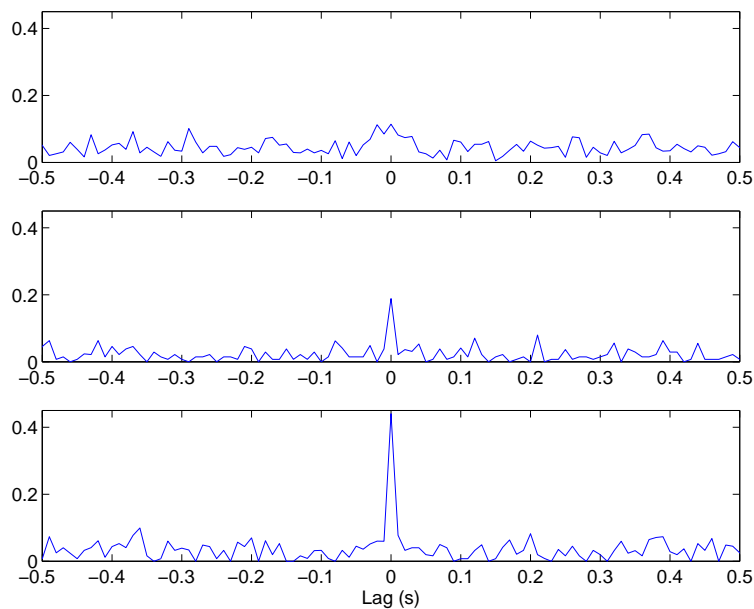
Figure 6: Pairwise cross correlations averaged over neuron pairs 9-10, 9-11 and 10-11. The data of the top trace were computed from the response of the network in the absence of global inhibition. The middle trace corresponds to the case with global inhibition and the bottom trace corresponds to the case with both global inhibition and local excitation turned on.

of local excitation (bottom trace, Fig. 5(b)) increases the synchronization even further.

To quantify these increases in synchronization, we computed the pairwise cross-correlation functions between neuron pairs 9-10, 9-11 and 10-11. Figure 6 shows the average of these three cross-correlation functions for the three cases described above. As shown in the top plot of Fig. 6, the correlation peak around zero lag is weak. Global inhibition, which has a temporal modulatory effect that shortens the time-window of integration, has the effect of narrowing the peak and increasing its amplitude. The combination of global inhibition *and* local excitation has the effect of further increasing the correlation peak, as predicted by theory (Maass and Bishop, 1998).

Conclusions

In this paper we demonstrate a prototype network of spiking neurons with realistic synaptic dynamics configured to operate as a competitive network. While this network can operate in the mean-rate mode to produce winner-take-all behavior, the temporal dependence of inhibition produces a variable selectivity to synchronous events as well as locking the output more closely in time to the input spikes. Beyond synchrony filtering, however, the network connectivity (recurrent local excitation combined with inhibition) also encourages the production of synchronous outputs.

In addition to being a useful tool for investigating the computational properties of spiking networks in real-time , the chip presented in this paper allows us to test these spike-based computational models in real physical sensorimotor systems, by connecting the chip to other neuromorphic sensor chips (e.g. silicon retinas and cochleae) and to robotic actuators. A number of demonstrations (both visual and auditory) that use this chip for spike-based computation are currently under development.

References

Bernander, Ö., Koch, C., and Usher, M. (1994). The effect of synchronized inputs at the single neuron level. *Neural Computation*, 6:622–641.

Boahen, K. (1998). Communicating neuronal ensembles between neuromorphic chips. In Lande, T. S., editor, *Neuromorphic Systems Engineering*, pages 229–259. Kluwer Academic, Norwell, MA.

Deiss, S. R., Douglas, R. J., and Whatley, A. M. (1998). A pulse-coded communications infrastructure for neuromorphic systems. In Maass, W. and Bishop,

C. M., editors, *Pulsed Neural Networks*, chapter 6, pages 157–178. MIT Press.

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402:529–533.

Fries, P., Reynolds, J., Rorie, A., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291:1560–1563.

Hahnloser, R., Sarpeshkar, R., Mahowald, M., Douglas, R. J., and Seung, S. (2000). Digital selection and analog amplification co-exist in an electronic circuit inspired by neocortex. *Nature*, 405:947–951.

Indiveri, G. (2000). Modeling selective attention using a neuromorphic analog VLSI device. *Neural Computation*, 12(12):2857–2880.

Itti, L., Niebur, E., and Koch, C. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

König, P., Engel, A., Roelfsema, P., and Singer, W. (1995). How precise is neuronal synchronization? *Neural Comput.*, 7(3):469–485.

Maass, W. and Bishop, C. M. (1998). *Pulsed Neural Networks*. MIT Press.

Mahowald, M. (1994). *An Analog VLSI System for Stereoscopic Vision*. Kluwer, Boston.

Mead, C. (1989). *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA.

Mortara, A. (1998). A pulsed communication/computation framework for analog VLSI perceptive systems. In Lande, T. S., editor, *Neuromorphic Systems Engineering*, pages 217–228. Kluwer Academic, Norwell, MA.

Niebur, E. and Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, 1(1):141–158.

Niebur, E. and Koch, C. (1996). Control of selective visual attention: modeling the "where" pathway. In Touretzky, D., Mozer, M., and Hasselmo, M., editors, *Advances in neural information processing systems*, volume 8, Cambridge, MA. MIT Press.

Salinas, E. and Abbott, L. (1996). A model of multiplicative neural responses in parietal cortex. *Proc. Natl. Acad. Sci.*, 93:11956–11961.

Steinmetz, P., Roy, A., Fitzgerald, P., Hsiao, S., Johnson, K., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404:187–190.

Stopfer, M., Bhagavan, S., Smith, B., and Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature*, 390:70–74.

van Schaik, A., Fragnière, E., and Vittoz, E. (1996). An analogue electronic model of ventral cochlear nucleus neurons. In *Proceedings of the Fifth International Conference on Microelectronics for Neural, Fuzzy and Bio-inspired Systems; Microneuro'96*, pages 52–59, Los Alamitos CA. IEEE Computer Society Press.