AN EVENT–BASED VLSI NETWORK OF INTEGRATE–AND–FIRE NEURONS

Elisabetta Chicca, Giacomo Indiveri and Rodney J. Douglas

Institute of Neuroinformatics, University of Zurich and ETH Zurich
Winterthurerstrasse 190, CH–8057 Zurich, Switzerland

ABSTRACT

The growing interest in pulse–based neural networks is encouraging the development of hardware implementations of massively parallel, distributed networks of Integrate–and–Fire (I&F) neurons. We have developed a mixed–mode (analog/digital) VLSI device that comprises a reconfigurable network of I&F neurons and adaptive synapses. The synapses receive input spikes and the neurons transmit output spikes (events) using an asynchronous Address–Event Representation (AER). In this paper we describe the network architecture, present experimental data demonstrating the characteristics of the single elements on the chip, and show that a competitive network configuration has Winner–Take–All (WTA) behavior and produces spike synchronization.

1. INTRODUCTION

Networks of I&F neurons have been shown to exhibit a wide range of useful computational properties, including feature binding, segmentation, pattern recognition, onset detection, input prediction, *etc.* [1]. These types of networks are very well suited for VLSI implementation. Recent and growing interest in pulse–based neural networks, together with the emergence of a standard that allows VLSI neurons to communicate using asynchronous pulse–frequency modulated events (spikes), have led to the development of a large number of VLSI implementations of networks of I&F neurons (see the ISCAS04 invited session on spiking neural networks). The asynchronous communication protocol is based on the Address–Event Representation [2, 3]. In this representation input and output signals are real–time, digital events that carry analog information in their temporal structure (interspike intervals). Each event is represented by a binary word encoding the address of the sending node. On–chip arbitration schemes are used to handle event "collisions" (cases in which sending nodes attempt to transmit their addresses at exactly the same time). Systems containing more than two AER chips can be assembled using additional off–chip arbitration. These off–chip arbiters can also use lookup–tables and processing elements to remap, time–stamp and perform digital operations on address–events [2, 4].

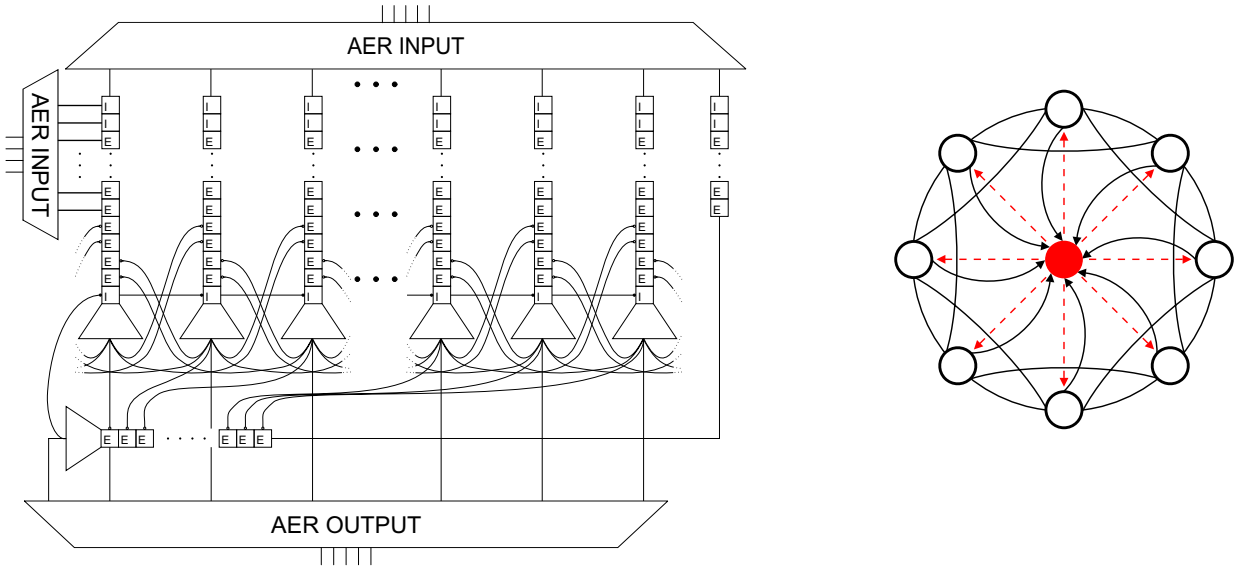
In this paper we present an AER chip comprising a network of I&F neurons and dynamic synapses. The I&F neuron circuit is described and fully characterized in [5]. The circuits implementing the synapses are of two types [6, 7]. Both types integrate input spikes, producing biologically plausible dynamics: one type is compact and exhibits short–term depression, while the other is larger, but can exhibit either short–term depression or facilitation, depending on its parameter settings. The parameters that control the synaptic dynamics and their strength are global and can be set by external voltage references.

The AER input synapses and AER output neurons offer the possibility of implementing networks of arbitrary topology when the device is interfaced to a dedicated PCI–AER board [4], able to log, monitor, map and generate address–events. In addition to externally addressable AER synapses, we included synaptic circuits with hard–wired on–chip connectivity to implement a competitive network topology. The circuits used on this chip are to a large extent technology independent and the network could be scaled up to arbitrary size.

2. NETWORK ARCHITECTURE

The architecture of the VLSI network of I&F neurons is shown in Fig. 1(a). It is a two–dimensional array containing a row of 32 neurons, each connected to a column of afferent synaptic circuits. Each column contains 14 AER excitatory synapses, 2 AER inhibitory synapses and 6 locally connected (hard–wired) synapses. When an address–event is received, the synapse with the corresponding row and column address is stimulated. If the address–events routed to the neuron with the corresponding column address integrate up to the neuron's voltage threshold for spiking, then that neuron generates an address–event which is transmitted off–chip. Arbitrary network architectures can be implemented using off–chip look–up tables and routing the chip's output address–events to one or more AER input synapses. The synapse address can belong to a different chip, therefore,

This work was supported by the EU grant ALAVLSI (IST–2001–38099).



(a) (b)

Fig. 1. (a) Chip architecture. Squares represent excitatory (E) and inhibitory (I) synapses, trapezoids represent I&F neurons. The I&F neurons can transmit their spikes off–chip and/or to locally connected synapses (see text for details). (b) Schematic representation of the connectivity pattern implemented by the internal hard–wired connections (closed boundary condition). Empty circles represent excitatory neurons and the filled circle represents the global inhibitory neuron. Solid/dashed lines represent excitatory/inhibitory connections. Connections with arrowheads are monodirectional, all the others are bidirectional.

arbitrary multi–chip architectures can be implemented.

Synapses with local hard–wired connectivity are used to realize a competitive (WTA) network with nearest neighbor and second nearest neighbor interactions (see Fig. 1): 31 neurons of the array send their spikes to 31 local excitatory synapses on the global inhibitory neuron; the inhibitory neuron, in turn, stimulates the local inhibitory synapses of the 31 excitatory neurons; each excitatory neuron stimulates its first and second neighbors on both sides using two sets of locally connected synapses. The first and second neighbor connections of the neurons at the edges of the array are connected to pads. This allows us to leave the network open, or implement closed boundary conditions (to form a ring of neurons [8]), using off–chip jumpers.

All of the synapses on the chip can be switched off. This allows us to inactivate either the local synaptic connections, or the AER ones, or to use local synapses in conjunction with the AER ones. In addition, a uniform constant DC current can be injected to all the neurons in the array. The amplitude of this current can be set through a global bias voltage.

The chip was implemented using a standard $0.8\ \mu m$ CMOS technology. The layout of the whole array, including the AER input and output sections covers an area of about $1.1 \times 1.9\ mm^2$. The layout of one column of the array, including the I&F neuron, the 16 AER synapses and the 6 local ones covers an area of about $31 \times 1500\ \mu m^2$. Only about 6% of this area is occupied by the neuron ($31 \times 86\ \mu m^2$).

In theory this network can scale up to any arbitrary size, both in terms of the number of neurons, and the number of synapses. In practice the network size is limited by the AER bandwidth available. If we consider a network of neurons configured via the PCI–AER board with 30% connectivity (a typical figure used in modeling studies), in which (typically) only 10% of the neurons fire at a mean rate of 100Hz, the speed of the (non–optimized) AER circuits implemented on the current chip limits the maximum number of possible neurons to approximately 1000. Using the same $0.8\ \mu m$ CMOS technology used for the current device, an array of 1000×300 I&F neurons and synapses would require a silicon area of approximately $31 \times 20\ mm^2$.

3. EXPERIMENTAL RESULTS

To verify the correct behavior of the circuits on the chip we injected the same DC current to all the neurons in the array and measured the network's response properties as a function of different configuration parameters (such as the strengths of different synaptic weights). In these experiments we did not stimulate the neurons via the AER synapses (that have been tested previously and shown to function correctly).

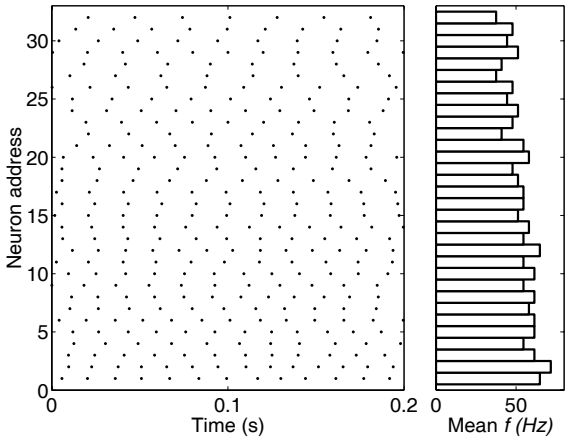


Fig. 2. Network response to homogeneous constant input current with all synaptic connections disabled. Left panel: raster plot of the network activity. Right panel: mean output frequencies. The differences in mean output frequency are due to device mismatch effects both in the input transistors and in the I&F neuron circuit elements.

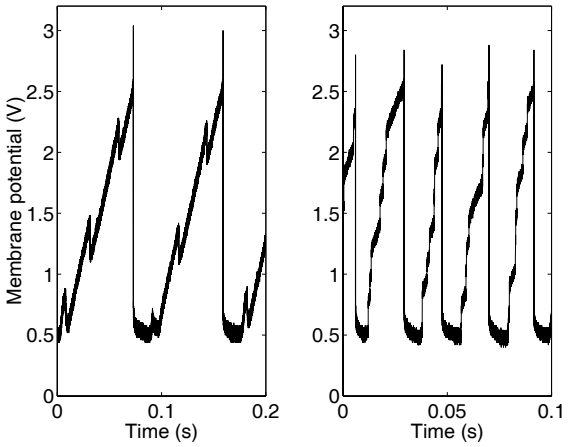


Fig. 3. Membrane potentials. Left panel: membrane potential of one of the excitatory I&F neurons in the network. The neuron integrates a constant DC current while receiving inhibitory spikes from the global inhibitory neuron. Right panel: membrane potential of the global inhibitory neuron. This neuron integrates the same constant DC current while receiving excitatory inputs from all the active excitatory neurons in the array.

3.1. Basic experiments

Initially we performed a set of basic experiments to test the functionality of the main building blocks of the chip: the neurons; the synapses; and the AER sections.

In a first experiment we switched off all the local hard–wired connections, injected a constant DC current to all the neurons and monitored their spiking activity using the PCI–AER board. In Fig. 2 we show a raster plot of the expected regular firing observed. The differences in mean firing rate are due to device mismatch effects both in the input transistors and in the I&F neuron circuit elements.

In a second experiment, we tested the competitive network topology (without lateral interactions) by switching on the connections in both directions between the excitatory neurons and the global inhibitory neuron. In this case, in addition to the constant DC current, the excitatory neurons integrate inhibitory inputs that tend to decrease their output firing rates, while the global inhibitory neuron integrates its excitatory inputs that increase its mean firing rate. The membrane potential of all the neurons in the array can be measured through an on–chip voltage scanner, which allows either all the neurons in parallel to be probed (multiplexed in time) or only one neuron at a time. In Fig. 3 we show the membrane potential of one of the excitatory I&F neurons in the network, next to the membrane potential of the global inhibitory neuron.

3.2. Network behavior

In these sets of experiments we activated the network's hard–wired connections to implement two different types of competitive networks with lateral connections. In both cases we activated the hard–wired connections from the excitatory neurons to the inhibitory one and the connections from the inhibitory neuron to the excitatory ones, stimulated the network by injecting a constant DC input current to all the neurons, and used the PCI–AER board to monitor the network spiking activity.

In the first experiment, symmetric nearest neighbor lateral connections were activated. Even in this extremely simplified case, with constant homogeneous inputs and symmetric connectivity, the network was able to produce a classical WTA behavior. Although all neurons should receive the same input current, due to device–mismatch effects, one neuron wins the competition and suppresses, through the inhibitory neuron, all other ones, while exciting nearest neighbors (see Fig. 4). As the coupling between neurons was set to be relatively strong, the excitatory and inhibitory neurons synchronized their spiking activity.

In the second experiment we activated both first and second neighbor excitatory connections. When the strength of these connections is asymmetric and global inhibition is strong enough, the network generates a traveling wave of activity, as shown in Fig. 5. Global inhibition allows the winning neurons to suppress all the others and the asymmetric lateral excitation propagates the activity in one direction. The neurons at the edge of the array were connected to form a ring [8], so that the wave could propagate cyclically through the array.

In both experiments the spiking activity of the neurons

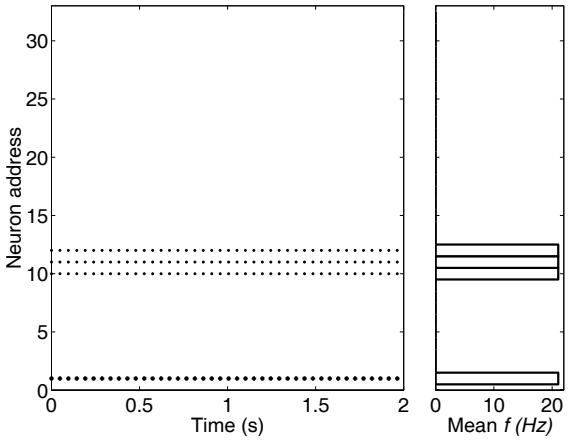


Fig. 4. WTA behavior. Left panel: raster plot of the network activity in response to a constant DC input current with lateral excitatory (first neighbor) connections, excitatory to inhibitory connections and global inhibition activated. The neuron with address 1 is the global inhibitory neuron. Right panel: mean output frequencies.

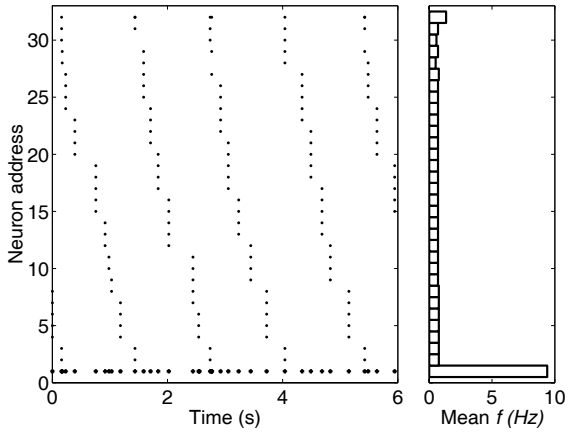


Fig. 5. Traveling wave. Left panel: raster plot of the network activity in response to a constant DC input current with asymmetric excitatory first and second nearest neighbor connections and with global inhibition. The neuron with address 1 is the global inhibitory neuron. Right panel: mean output frequencies.

is highly synchronized. This is a consequence of the parameters used in these experiments. These are extreme cases, used to characterize the architecture with its hard–wired competitive network topology, in which the input is a simple homogeneous constant current, and the strength of the connections is set to relatively high values, to amplify the small differences in neuronal activity due to mismatch parameters.

4. DISCUSSIONS AND FUTURE WORK

In this paper we have presented a reconfigurable VLSI array of AER neurons and synapses with additional on–chip connectivity that implements a competitive network topology. We demonstrated the correct behavior of the main blocks present on the device and showed how, even using the simplest possible input stimulus, the local hard–wired competitive network can give rise to interesting behaviors such as WTA functionality, spike synchronization and traveling wave generation. So far, we have used a simple constant DC current as the input to the network. In future work, we will take advantage of the PCI–AER board to stimulate the network with Poisson distributed spike trains, or with address–events generated by neuromorphic vision [9] or auditory sensors. Having multiple instances of the same synaptic circuit for each neuron will allow us to explore the effect of adaptation to several competing stimuli. We will use this device to implement real–time models of selective attention systems [5]; we will also study the network's ability to generate traveling waves, persistent activity (even after the input stimulus is removed), binding by synchronization, and other behaviors observed in cortical circuits.

5. REFERENCES

[1] W. Maass and C. M. Bishop, Eds., *Pulsed Neural Networks*, The MIT Press, 1999.

[2] S. R. Deiss, R. J. Douglas, and A. M. Whatley, "A pulse-coded communications infrastructure for neuromorphic systems," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds., chapter 6, pp. 157–178. MIT Press, 1998.

[3] K. A. Boahen, "Communicating neuronal ensembles between neuromorphic chips," in *Neuromorphic Systems Engineering*, T. S. Lande, Ed., pp. 229–259. Kluwer Academic, Norwell, MA, 1998.

[4] V. Dante and P. Del Giudice, "The PCI–AER interface board," in *Report on the 2001 Workshop on Neuromorphic Engineering*, 2001, pp. 99–103.

[5] G. Indiveri, "Modeling selective attention using a neuromorphic analog VLSI device," *Neural Computation*, vol. 12, no. 12, pp. 2857–2880, Dec 2000.

[6] E. Chicca, G. Indiveri, and R. J. Douglas, "An adaptive silicon synapse," in *Proc. IEEE International Symposium on Circuits and Systems*. IEEE, May 2003.

[7] C. Rasche and R. Hahnloser, "Silicon synaptic depression," *Biological Cybernetics*, vol. 84, no. 1, pp. 57–62, 2001.

[8] R. Hahnloser, R. Sarpeshkar, M. Mahowald, R. J. Douglas, and S. Seung, "Digital selection and analog amplifi cation coexist in an electronic circuit inspired by neocortex," *Nature*, vol. 405, pp. 947–951, 2000.

[9] J. Kramer, "An integrated optical transient sensor," *IEEE Trans. on Circuits and Systems II*, vol. 49, no. 9, pp. 612–628, Sep 2002.