Spike-Based MAX Networks for Nonlinear Pooling in Hierarchical Vision Processing

Fopefolu O. Folowosele[1], R. Jacob Vogelstein[1], and R. Etienne-Cummings[1]
[1]Department of Electrical & Computer Engineering, Johns Hopkins University, Baltimore, Maryland
[fope|jvogelst|retienne]@jhu.edu

***Abstract*— Complex cells in the visual cortex utilize a maximum (MAX) operation to pool the outputs of simple cells to achieve feature specificity and invariance. We demonstrate a biologically-plausible MAX network for nonlinear pooling in hardware, using a reconfigurable multichip address event representation based VLSI system. With this implementation we have shown that we can implement simple and advanced stages of visual processing on the same chip and are one step closer to constructing an autonomous, continuous-time, biologically-plausible hierarchical model of visual information processing using large-scale arrays of identical silicon neurons.**

***Index Terms*—Complex cells, integrate-and-fire array transceiver, MAX network, visual cortex**

I. Introduction

THE massively connected architecture of the mammalian brain allows it to process sensory information in a parallel and distributed manner in real time, to observe, evaluate and react to the environment appropriately. Such an architecture results in a system whose computational abilities are greater than the sum of its components. Combining this architecture with spike communication gives rise to elaborate information coding schemes which the brain uses for communication and processing [1]. The brain is thus able to perform image processing and object recognition tasks with ease – effortlessly recognizing a large number of diverse objects in cluttered scenes [2]. Artificial systems perform poorly in these tasks when compared with the brain. Learning how the brain functions and understanding its information coding schemes will enable us to emulate it in artificial systems, and so bridge the performance gap.

A lot has been learned about the brain through simulations of neural functions in software. However, these simulations run much slower than real time and are thus unable to interact with the environment, limiting the types of studies that can be done. Neuromorphic hardware on the other hand typically operates in real time enabling the design of artificial nervous systems that interact with the environment. Neuromorphic systems use silicon analogs of biological neural elements to emulate the functionality of the brain and communicate spikes rapidly using the address-event representation (AER) communication protocol [3-6].

Because the models of cortical areas are still being studied and are always in flux, a hardwired neuromorphic implementation is often unproductive. A reconfigurable neural array transceiver, such as the ones described in [7-12], is frequently a better alternative as it combines the speed of dedicated hardware with the programmability of software for studying real-time operations of cortical, large-scale neural networks. These neural array transceivers have been used in other vision processing applications [22-24]. In the array transceiver employed here [13], there are no hardwired connections between neurons. Instead, connections are emulated by time-multiplexing action potentials (spikes) onto a fast serial bus, and synapses are implemented with encoders and decoders that monitor the bus and route incoming and outgoing spikes to their appropriate target.

Computational models of the human visual processing system frequently employ a hierarchical structure [14]. One such model of the stages of visual processing is illustrated in Figure 1.

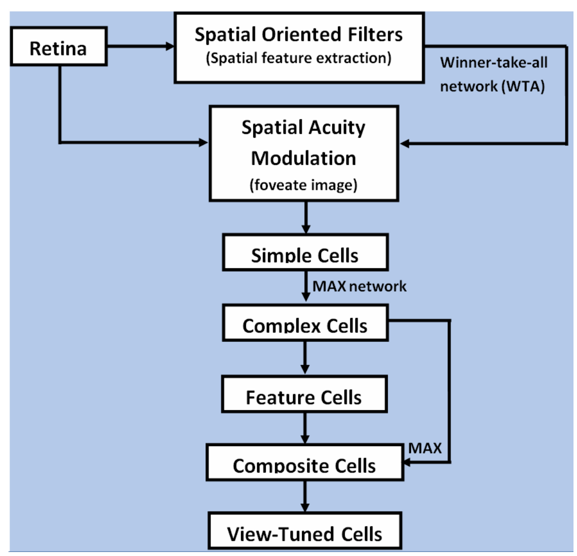

Figure 1. Hierarchical architecture of visual processing as described in [17]. Spatial oriented filters extract spatial features from the retina. The outputs of these filters are combined. A winner-take-all network selects the region with maximum salience and the image is then foveated by spatial acuity modulation. Simple cells with different preferred orientations are then used to process the modulated signal. The outputs of the simple cells are combined with a nonlinear MAX function to form complex cells.

In the model, outputs from the retina are processed through oriented spatial filters that highlight regions of high contrast [16]. A salience detector network then uses this information to focus attention on a region of interest. Data from local spatial filters within the area of interest are combined to create feature cells, composite cells and view-tuned cells [14].

In previous work we have demonstrated the use of the IFAT system to implement the first stages of the model (from the retina to simple cells in Figure 1) [16-18]. In this paper we extend this in hardware to include the complex cells. We demonstrate how the outputs from simple cells with similar preferred orientations and different receptive fields are combined with a maximum (MAX) operation to form complex cells. With the realization of these complex cells, which act as position-invariant oriented spatial filters, we are closer to our goal of a full implementation of neuromorphic vision.

A description of the IFAT system and MAX operation are provided in sections II and III respectively. Results from the implementation of the MAX operation on the IFAT system are discussed in section IV and the paper is concluded in section V.

II. Reconfigurable Hardware

The IFAT system consists of up to 4 integrate-and-fire (I&F) chips with each chip containing an array of 2400 conductance-based integrate and fire neurons [13]. Virtual synapses replace hardwired connections between neurons, with mapping and synaptic parameters (synaptic weight and driving potential) for each synapse stored in 128MB of digital RAM. In addition to the RAM and IFAT chips, the system contains a DAC which converts the driving potential to an analog value and an FPGA to process all the address-event transactions (Figure 2).

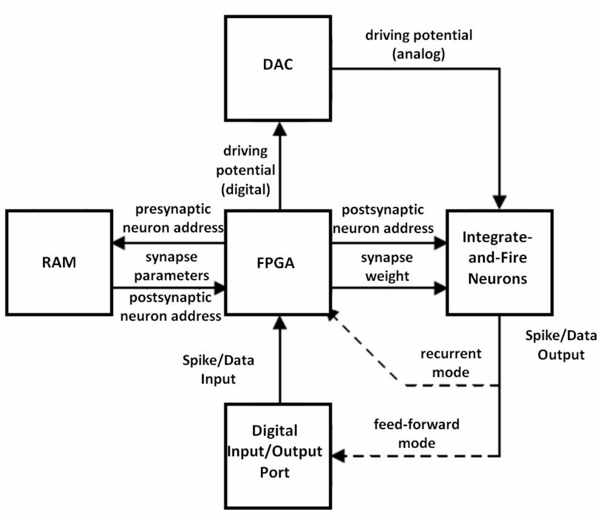

Figure 2. Block diagram of the IFAT system. The FPGA receives spikes via the DIO port. Outgoing spikes are sent to a computer in both feed-forward and recurrent modes. However, in recurrent mode, outgoing spikes are also sent back to the FPGA.

The IFAT acts as an address-event transceiver, communicating incoming and outgoing events over an asynchronous bus. A netlist stored in RAM specifies the network topology and synaptic strength between the presynaptic and postsynaptic neurons. The printed circuit board of the IFAT system with all the components integrated is shown in Figure 3.

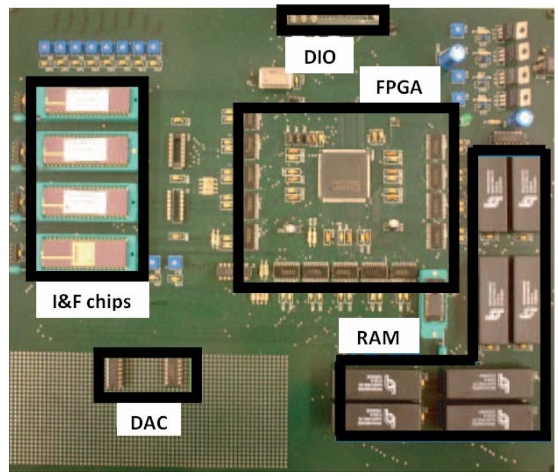

Figure 3: Printed circuit board of the IFAT system with all the components – four integrate-and-fire (I&F) chips, digital memory (RAM), digital to analog converter (DAC) and FPGA.

The FPGA allows for both feed-forward and recurrent transactions. When the FPGA receives a request from an external source, it reads the address bits on the incoming AE bus via the DIO port and stores it as the presynaptic neuron address. The FPGA uses this address as an index into RAM to obtain a set of postsynaptic neuron addresses and the synaptic parameters associated with them. The synaptic weights are set up on the I&F chips and the analog synaptic equilibrium potentials are established by the DAC. The events are then serialized as they are sent to the chip. Output spikes are re-routed to the FPGA to allow for recurrent connections, and are also sent to a computer for monitoring.

III. Max Operation

In the model of visual processing implemented here, outputs from simple cells with similar preferred orientations and different receptive fields are combined with a maximum operation to form complex cells [14, 17]. Pooling by a maximum (MAX) operation achieves both high feature specificity and invariance in a more robust manner than linear summation. The MAX operation is a nonlinear saturating pooling function on a set of inputs such that the output codes the amplitude of the largest input, regardless of the strength and number of the other inputs. It is similar to a winner-take-all (WTA) network [20-21], except that a MAX network conveys the amplitude of the largest input but not its identity while a WTA one reveals the identity but not its precise amplitude.

The MAX network implemented is a biologically-plausible model proposed by Yu et al. [15]. It is a three-layer neural circuit with an input layer *X*, an intermediate layer (which performs a nonlinear transformation) *Y*, and an output neuron *Z* (Figure 4). The inputs cause the output neuron to generate spikes at a rate proportional to the maximum input.

Complex cells in the model take the MAX of similarly-oriented simple cells over a region of space. To generate a complex cell from a 5x5 grid of similarly-oriented simple cells with different receptive fields, for example, will require 25 *X* inputs (i.e. n=25) and a *Z* output where *X* and *Z* are simple and complex cells respectively (see Figure 4).

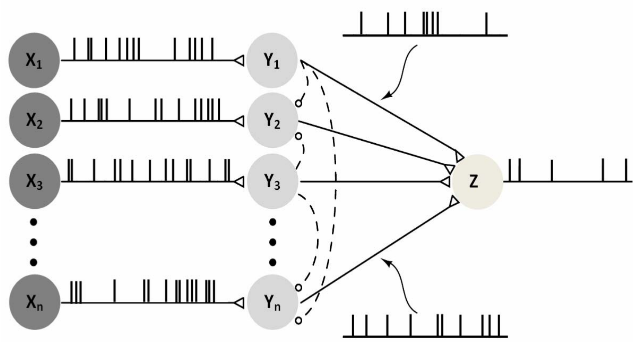

Figure 4. Schematic model of the implemented MAX network [15]. Solid lines with triangular synapses and dashed lines with circular synapses represent excitatory and inhibitory connections respectively. Each intermediate *Y* neuron makes inhibitory connections to all other *Y* neurons, though not all are shown in this figure. The inputs *X* neurons are output spikes from simple cells. The output of the *z* neuron corresponds to complex cell output. Modified from [17].

IV. Method

We generated four images to test the functionality of our complex cell network (Figure 6A). Each image consists of the same five bars with different intensities, but the position of each bar varies in the different images. We converted these images to spikes, mimicking the simple visual transduction pathway of our "octopus" neuromorphic retina chip [18-19], such that higher light intensities generated more spikes. These simulated retinal outputs are then processed by a simple cell network implemented on the IFAT.

Simple cells are oriented spatial filters that detect local changes in contrast. We implemented simple cells with a 4x1 receptive field [17]. Each simple (IFAT) cell receives inputs from four consecutive simulated photodetectors, two of which make excitatory synapses while the other two make inhibitory synapses (Figure 5). The excitatory and inhibitory synaptic weights are balanced so that the simple cells do not respond to uniform light.

We implemented two types of simple cells (Figure 5). One type was oriented in the vertical direction and responded to dark-to-light transitions. The other simple cell type was oriented in the horizontal direction and responded to light-to-dark transitions. The IFAT spikes corresponding to the simple cell outputs for the four test images were recorded and stored on a computer (off-chip). These spikes then formed the *X* inputs to the MAX network.

The MAX operation was implemented on the IFAT to emulate complex cells. This was done by creating both the intermediate *Y* and output *Z* cells on the IFAT system, and streaming outputs from the *X* cells from a computer (Figure 4). Each complex cell was made to take inputs from a 5x5 grid of simple cells (*X*) with the same preferred orientation. Each MAX network requires N intermediate cells for N inputs, plus an additional cell for output. Thus the MAX networks we implemented consisted of 26 cells each. The IFAT system was set in recurrent mode so that a spike from Y_i will be followed by inhibition in the other 24 *Y* cells for that complex cell and will excite its *Z* output. Spikes from the complex cell outputs were logged and stored on a computer.

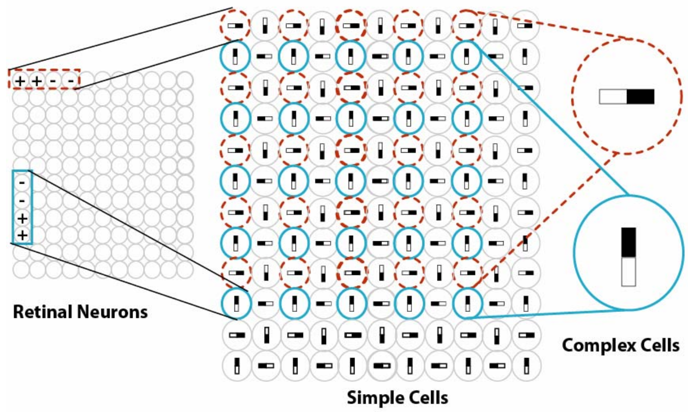

Figure 5. Illustration of the mapping from retinal neurons to simple cells and complex cells. Each simple cell integrates inputs from a 4x1 retinal receptive field. Each complex cell integrates inputs from an array of 5x5 simple cells, covering 10x10 pixels in retinal space. For clarity, the receptive fields of only two simple cells and two complex cells are shown.

V. Results

The four test images are shown in Figure 6A. Horizontally- and vertically-oriented simple cells are in Figures 6B and 6C respectively.

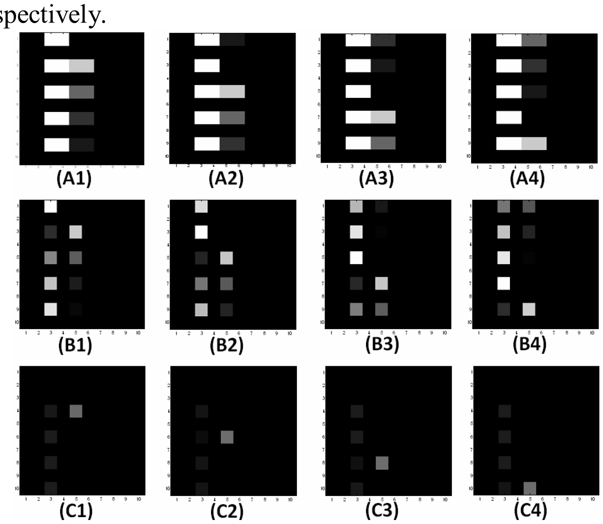

Figure 6. (A1-4) Generated test images. The five bars in all the images are the same, with the difference between images being the positions of the different bars. (B1-4) Horizontally-oriented simple cells that respond to light-to-dark transitions. The position of the simple cell that generated the most spikes represents the cell with the largest light-to-dark transition. (C1-4) Similar to B1-4, except these are vertically-oriented simple cells that respond to dark-to-light transitions.

The ratio between the number of complex cell spikes and the maximum input (the maximum number of spikes in the simple cell with the largest amplitude) was computed for the simple cells in Figures 6B and 6C. The calculation results are shown in Table 1.

The two important properties of a MAX operation are selectivity (output signal depends only on the maximum of the input signals, X_{MAX}) and linearity (output signal depends linearly on X_{MAX}, i.e. $Z/X_{MAX} = k$, where k is constant) [15].

TABLE I
MAX NETWORK COMPUTATION RESULTS

Simple Cells	X_{MAX}	Z	k [Z/X_{MAX}]
B1	5185	355	0.0684
B2	4780	325	0.0679
B3	5022	341	0.0679
B4	4985	341	0.0684
C1	2175	146	0.0671
C2	2027	136	0.0671
C3	2121	145	0.0684
C4	2154	144	0.0668

The ratio k obtained was approximately constant among all the simple cells, with a mean of 0.068 and a standard deviation of 0.0006. This occurs in spite of the differences in the position of the simple cell with the largest amplitude (position invariance) or its intensity (feature specificity). This demonstrates the accuracy of our implementation.

VI. CONCLUSION

We have demonstrated the implementation of a biologically-plausible spike-based nonlinear MAX network on a reconfigurable neural array transceiver. We also demonstrated a hardware implementation of the second stage of processing in a hierarchical model of visual object recognition. More specifically, we implemented silicon facsimiles of complex cells, which use the MAX operator to pool similarly-oriented simple cells with nearby receptive fields.

ACKNOWLEDGMENT

We thank Clyde Clarke and Ndubuisi Ekekwe for helpful discussions, and Gert Cauwenberghs for his help in developing the IFAT hardware.

REFERENCES

[1] D.H. Goldberg, G. Cauwenberghs, A.G. Andreou, "Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons," *Neural Networks*, vol. 14, pp. 781-793, 2001.

[2] M. Riesenhuber, and T. Poggio, "Neural mechanisms of object recognition," *Current Opinion in Neurobiology*, vol. 12, pp. 162-168, 2002.

[3] C. Mead, *Analog VLSI and neural systems*. Reading, MA: Addison-Wesley, 1989.

[4] M. Sivilotti, "Wiring considerations in analog VLSI systems, with application to field-programmable networks," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, 1991.

[5] M. Mahowald, *An analog VLSI system for stereoscopic vision*. Boston, MA: Kluwer Academic Publishers, 1994.

[6] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer peripherals," *IEEE Trans. Neural Networks*, vol. 4, pp. 523-528, 1993.

[7] J.V. Arthur, and K.A. Boahen, "Recurrently connected silicon neurons with active dendrites for one-shot learning," *IEEE International Joint Conference on Neural Networks*, vol. 3, pp. 1699-1704, 2004.

[8] D.H. Goldberg, G. Cauwenberghs, and A.G. Andreou, "Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons," *Neural Networks*, vol. 14, pp. 781-793, 2001.

[9] T. Horiuchi, and K. Hynna, "Spike-based VLSI modeling of the ILD system in the echolocating bat," *Neural Networks*, vol. 14, pp. 755-762, 2001.

[10] G. Indiveri, E. Chicca, and R.J. Douglas, "A VLSI reconfigurable network of integrate-and-fire neurons with spike-based learning synapses," *European Symposium on Artificial Neural Networks*, pp. 405-410, 2004.

[11] B. Taba, and K.A. Boahen, "Topographic map formation by silicon growth cones," *Advances in neural information processing systems*, vol. 13, pp. 173-179, 2003.

[12] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, H. Kolle Riis, T. Delbrück, S.C. Liu, S. Zahnd, A.M. Whatley, R. Douglas, P. Häfliger, G. Jimenez-Moreno, A. Civit, T. Serrano-Gotarredona, A. Acosta-Jiménez, B. Linares-Barranco, "AER building blocks for multi-layer multi-chip neuromorphic vision systems," *Advances in neural information processing systems*, vol. 18, pp. 1217-1224, 2006.

[13] R.J. Vogelstein, U. Mallik, J.T. Vogelstein, G. Cauwenberghs, "Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses," *IEEE Trans. Neural Networks*, vol. 18, pp. 253-265, 2007.

[14] M. Riesenhuber, and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, pp. 1019-1025, 1999

[15] A.J. Yu, M.A. Giese, and T.A. Poggio, "Biophysiologically plausible implementations of the maximum operation," *Neural Computation*, vol. 14, pp. 2857-2881, 2002.

[16] U. Mallik, R.J. Vogelstein, E. Culurciello, R. Etienne-Cummings, G. Cauwenberghs, "A real-time spike-domain sensory information processing system," *In Proceedings of the IEEE ISCAS*, vol. 3, pp. 1919-1922, 2005.

[17] R.J. Vogelstein, U. Mallik, E. Culurciello, G. Cauwenberghs, and R. Etienne-Cummings, "A multichip neuromorphic system for spike-based visual information processing," *Neural Computation*, vol. 19, pp. 2281-2300, 2007.

[18] E. Culurciello, R. Etienne-Cummings, and K.A. Boahen, "A biomorphic digital image sensor," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 281-294, 2003.

[19] E. Culurciello, and R. Etienne-Cummings, "Second generation of high dynamic range, arbitrated digital imager," *IEEE ISCAS*, 2004.

[20] M. Oster, R.J. Douglas, and S-C Liu, "Quantifying input and output statistics of a winner-take-all network in a vision system," *IEEE International Symposium on Circuits and Systems*, pp. 853-858, 2007.

[21] S-C. Liu, and M. Oster, "Feature competition in a s pike-based winner-take-all VLSI network," *IEEE International Symposium on Circuits and Systems*, pp. 3634-3637, 2006.

[22] T.Y. Choi, P.A. Merolla, J.V. Arthur, K.A. Boahen, and B.E. Shi, "Neuromorphic implementation of orientation hypercolumns," *IEEE Transactions on Circuits and Systems*, vol. 52, pp. 1049-1060, 2005.

[23] E. Chicca, A.M. Whatley, P. Lichtsteiner, V. Dante, T. Delbruck, P. Del Giudice, R.J. Douglas, and G. Indiveri, "A multichip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity," *IEEE Transactions on Circuits and Systems,* vol. 54, pp. 981-993, 2007.

[24] P. Merolla, K. A. Boahen, "A recurrent model of orientation maps with simple and complex cells," Departmental Papers, Department of Bioengineering, University of Pennsylvania, PA, 2003.