A REAL-TIME SPIKE-DOMAIN SENSORY INFORMATION PROCESSING SYSTEM

Udayan Mallik[1]*, R. Jacob Vogelstein*[2]*, Eugenio Culurciello*[3]*,*
Ralph Etienne-Cummings[1] *and Gert Cauwenberghs*[1]

[1]Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland
[2]Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland
[3]Dept. of Electrical Engineering, Yale University, New Haven, Connecticut

ABSTRACT

We present an architecture for processing spike-based sensory information in real-time. The system is based on a re-configurable silicon array of general-purpose integrate-and-fire neurons (as opposed to application-specific circuits), which can emulate arbitrary cortical networks. A combined retinal/cortical network has been designed and tested with a neuromorphic silicon retina. Neural activity is communicated between chips at rates of up to 1,000,000 spikes/sec with a bit-parallel Address-Event Representation protocol. This work represents the first step in constructing an autonomous, continuous-time, biologically-plausible hierarchical model of visual information processing using large-scale arrays of identical silicon neurons.

1. INTRODUCTION

Both silicon and biological spiking sensors can generate millions of events per second, all of which must be transmitted to external processors for higher levels of processing. In the brain, this is achieved by extensive connectivity between neural centers with distributed and parallel processing. In artificial neuromorphic systems, spikes can be rapidly transmitted to various locations using the address-event representation (AER) communication protocol [1–3], but relatively few solutions exist for real-time and large-scale processing of this spike-encoded sensory information (however, see [4–11] for some work in this direction).

We are presenting a neuromorphic system that can process sensory information in real-time, up to a rate of about 1,000,000 spikes per second. In previous work, we described an address-event integrate-and-fire array transceiver (IFAT) system that contains up to 9,600 silicon neurons [12], and demonstrated its ability to process visual information produced by a neuromorphic imager [12–15]. However, until now the computations were limited to off-line processing with a computer "in the loop". Here we concentrate on the implementation of an autonomous, real-time, combined retinal/cortical system that can process video information encoded in spikes. Eventually, this approach could lead to an autonomous object recognition device with binocular vision.

The complexity involved and resource allocation necessary for neural computation in higher levels of sensory processing necessitates a multi-chip approach for artificial systems [4, 5, 9, 10]. Software packages, though easily re-configurable, are not suited for real-time processing of spikes. Custom VLSI implementations with dedicated hardware geared towards solving specific tasks offer an attractive alternative. However, unlike software systems, custom VLSI chips have a turnover time between 2-3 months. The re-configurable IFAT system [12] used in this work is a "best of both worlds" solution, as it combines the speed of dedicated hardware with the programmability of software.

The goal of our current research is to implement a biologically-plausible hierarchical model of visual information processing [16] entirely contained within the IFAT. The data described here illustrates outputs from the first stage of this model. Previous work in this area by other groups has relied upon separate chips for each stage of processing, and has advanced to very sophisticated models of retinal processing and the first layer of cortical processing [4–6]. However, implementations of more advanced stages have not been demonstrated in hardware. We believe the ability to perform rapid prototyping and real-time testing of neural architectures with the IFAT will complement the dedicated-chip approach as we work towards a full implementation of neuromorphic vision.

A description of the IFAT and spike-domain imager (called the "Octopus Retina") are provided in Sections 2.1 and 2.2, followed by details of the combined system in Section 2.3. We then discuss a strategy for implementation of the hierarchical visual model in Section 3 and show data from the first stage of processing. Finally, the paper is concluded in Section 4.

2. HARDWARE

2.1. Integrate-and-Fire Array Transceiver (IFAT)

The Integrate-and-Fire Array Transceiver (IFAT) system (Fig. 1) contains up to 9,600 spiking silicon neurons on four custom VLSI chips [17], with re-configurable synaptic connectivity and adjustable neural dynamics. Network architectures are stored in 128 MB of digital RAM, allowing for up to 4,194,304 synapses [12], each of which can specify a different synaptic weight and synaptic equilibrium potential [17]. At the system level, the IFAT acts as an address-event (AE) transceiver, communicating incoming and outgoing events over an asynchronous AE bus. Internally, an FPGA routes incoming spikes to the appropriate neural targets to enable processing at a rate of up to 1,000,000 events per second. The spike input rate of the system is limited by slow access and setup times for the RAM and DACs, respectively. The current setup also utilizes an array of logic-level shifters to communicate between the FPGA and other components on the PCB. This produces additional processing delays.

This work was partially funded by NSF Awards #0120369, #9896362, and IIS-0209289; ONR Award #N00014-99-1-0612; and a DARPA/ONR MURI #N00014-95-1-0409. Additionally, RJV is supported by an NSF Graduate Research Fellowship.

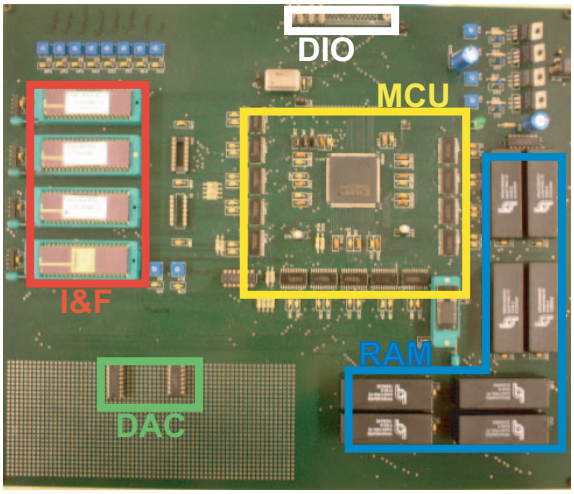

Fig. 1. Printed circuit board (PCB) integrating all components of the IFAT system: custom VLSI integrate-and-fire chips (I&F), digital-to-analog converter (DAC), digital memory (RAM), FPGA microcontroller (MCU), and digital input/output port (DIO).

2.2. Octopus Retina (OR)

The Octopus Retina (OR) is a 60 $\times$ 80 array of integrate-and-fire neurons that translates light intensity levels into inter-spike interval times at each pixel [13, 14]. The design is based on the simple phototransduction mechanism found in the retinae of octopi, although instead of the parallel connectivity found in biology between the retina and cortex, the silicon implementation uses a pipelined and arbitrated read-out scheme and transmits data serially over an address-event bus. Under uniform indoor lighting ($0.1\mathrm{mW/cm^2}$), the OR generates a mean firing rate of 200,000 address events per second (41.7 effective fps), while providing 120dB of dynamic range and consuming 3.4mW.

2.3. Combined System

The novel aspect of this present work, as a milestone in the continuing evolution of the IFAT system, is the integration of IFAT with the OR to form an autonomous, re-configurable, real-time visual information processing system. To interface the two subsystems, we designed a printed circuit board (PCB) that manages connections between up to two octopus retinae, the IFAT, and a personal computer (CPU) (Fig. 2). There are four user-selectable modes of operation: (1) reset mode, (2) RAM loading mode, (3) RAM read-back mode, and (4) transact mode. Depending on the mode, the FPGA on the IFAT enters one of its four state machines and enables/disables the appropriate tri-state buffers on the interface PCB. The buffers determine whether OR_1, OR_2, or the CPU has access to the address-event bus that supplies inputs to the IFAT.

On power-up, the system enters reset mode, which tri-states all the buffers on the PCB and instructs the IFAT to reset all the neurons on the I&F chips. Once reset, the user can choose to load the RAM with a "netlist", which specifies the network topology and synaptic strengths between the OR and IFAT neurons, as well as configuring recurrent connections between IFAT neurons. Because the IFAT uses non-volatile SRAM, it is not necessary to load a netlist each time the system is powered up (while convenient in general, this feature was selected to enable storing learned patterns). After a netlist is programmed, or after on-line modification of an existing netlist, the contents of RAM can be read by the FPGA for verification.

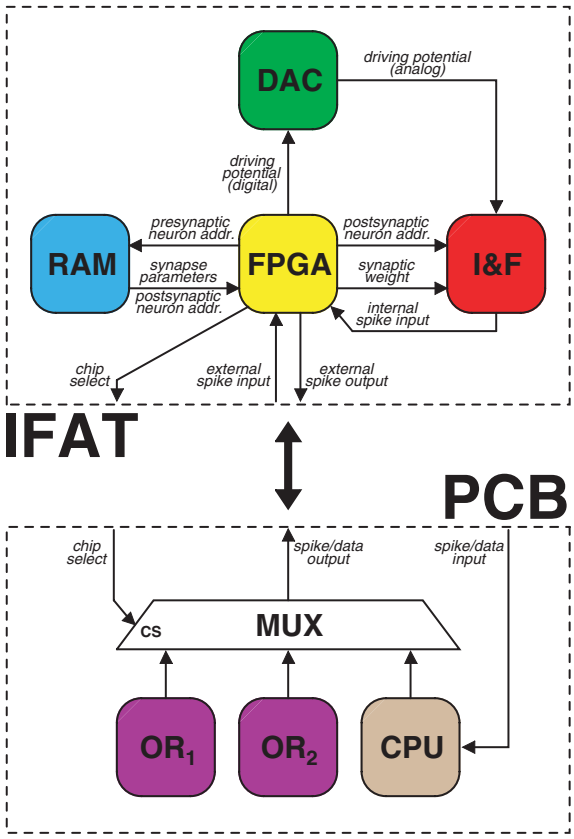

Fig. 2. Block diagram of the combined retinal/cortical system, with IFAT components contained in the upper block. The FPGA receives spikes from either of the two octopus retinae (OR) or the computer (CPU), depending on which is selected by the multiplexer (MUX). Outgoing addresses from the IFAT are sent to the computer for visualization or storage.

Information processing and communication between the ORs and the IFAT occurs in transact mode. Here, the FPGA state machine performs handshaking with up to two ORs and arbitrates between them if events are pending from both. Additionally, this mode allows events to be sent to the IFAT from the CPU, if any external inputs are required. Finally, recurrent connections between I&F neurons on the IFAT can be implemented in this mode.

The netlist stored in RAM contains a look-up table that describes how the FPGA should route events from all of the different sources. Depending on the number of neurons in the overall system, the addressing scheme can include two or more bits of "chip select" (CS) to distinguish similar addresses originating from different locations. When the FPGA receives a request, it enables the appropriate buffer on the PCB, reads the address bits on the incoming AE bus, prepends the CS bits, and stores the binary word as a base address (presynaptic address). It then adds a 'synapse offset' to form a complete 22-bit RAM address, which is used to look up a set of synaptic parameters for the first postsynaptic target. Next,

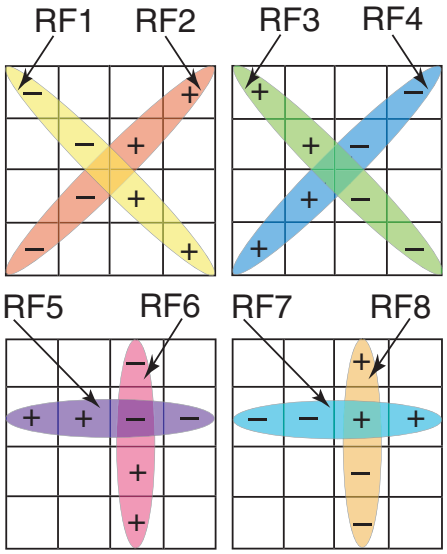

Fig. 3. Orientation selective Receptive Field (RF) kernel compositions in the simple cell network. Top row shows -45° and $+45^\circ$ receptive fields: (RF1/RF2) dark-to-light transitions, (RF3/RF4) light-to-dark transitions. Bottom row shows 0° and 90° receptive fields: (RF5/RF6) light-to-dark transitions, (RF7/RF8) dark-to-light transitions.

the digital synaptic weight is set up on the I&F chips and the analog synaptic equilibrium potential is established by the DAC. The event is then sent to one of the IFAT's silicon neurons by activating its row, column, and chip select lines [17]. Finally, the FPGA increments the offset by one and repeats this process until a reserved word is received from the RAM indicating an end of postsynaptic targets.

3. VISUAL INFORMATION PROCESSING

The long-term goal of this research is to create an autonomous, general-purpose spike processing platform that can be interfaced with any address-event neuromorphic sensor to implement real-time emulations of higher cortical functions. The particular application envisaged by this work is a multi-stage vision system that would receive its input from the ORs and implement silicon facsimiles of cortical simple cells, complex cells, composite feature cells, complex composite cells, and finally, view-tuned cells, according to the model of Riesenhuber and Poggio [16]. Each level of the hierarchy pools inputs from lower levels using either linear summation or a nonlinear maximum operation (MAX), which simultaneously allows for high feature selectivity and invariance [16]. Currently, we have only implemented the first stage of processing in real-time. However, we are working towards a full implementation.

The first stage of processing implements *simple cells*, oriented spatial filters that detect local changes in contrast. We have previously demonstrated a feed-forward network that can perform this function, although it was originally tested off-line using a stream of events captured from the OR and stored on disk [12]. The simple cell network architecture is illustrated in Fig. 3. Briefly, each IFAT cell receives inputs from four consecutive OR neurons oriented in either the vertical, horizontal, or diagonal directions. Within each 4-cell receptive field (RF), two of the inputs are excitatory and two are inhibitory, so that only one of a light-to-dark (+ -) or dark-to-light (- +) transition in the underlying image will provide net excitatory input. Figure 4 shows eight captured frames from the real-time system, each presenting the output of only one of the eight types of simple cells.

In the second stage of processing, similarly-oriented simple cells with nearby RFs must be pooled together using the MAX operator to form *complex cells*, position-invariant spatial filters. Multiple architectures for implementing the maximum operation were proposed in [18], and, using the IFAT, we have verified the operation of a feedback MAX network with an artificially generated input pattern of 30 neurons [19], but we have not yet ported this to the real-time visual system.

The two archetypal network operations described above, linear summation and MAX pooling, form the basis for all stages of processing beyond complex cells. While we have not yet integrated these operations in a single netlist, there are no technical barriers preventing us from doing so. Rather, we are currently limited by the number of I&F neurons available. However, the architecture of the combined system described here is modular [19], in the sense that the controlling IFAT does not know what devices generate incoming addresses, and therefore the system can be extended to include connections for multiple IFATs. In this case, each stage of processing that requires a MAX computation should be contained within one IFAT board to conserve bandwidth and optimize speed, since there are more connections within MAX stages than between stages. To fully implement the Riesenhuber and Poggio model [16], we would also need to train the network to identify specific views of particular objects. Because "learning" in this system equates to modifying the contents of RAM, it can be achieved on-line, as we have demonstrated with a previous version of the IFAT [20].

4. CONCLUSION

We have described a system that processes spike-based sensory information in real-time, using a "general-purpose" re-configurable array of silicon neurons. The system is currently configured for inputs from up to two neuromorphic retinae, but can be extended to include connections from multiple different sensors, additional arrays of neurons, or any other device that communicates with address-events. We are currently working towards a full implementation of a biologically-plausible model of hierarchical visual information processing, but because the system is designed to enable construction of arbitrary neural networks, we invite readers to contact us for collaborative opportunities with different sensory modalities.

5. REFERENCES

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

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

[3] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie, "Silicon auditory processors as computer

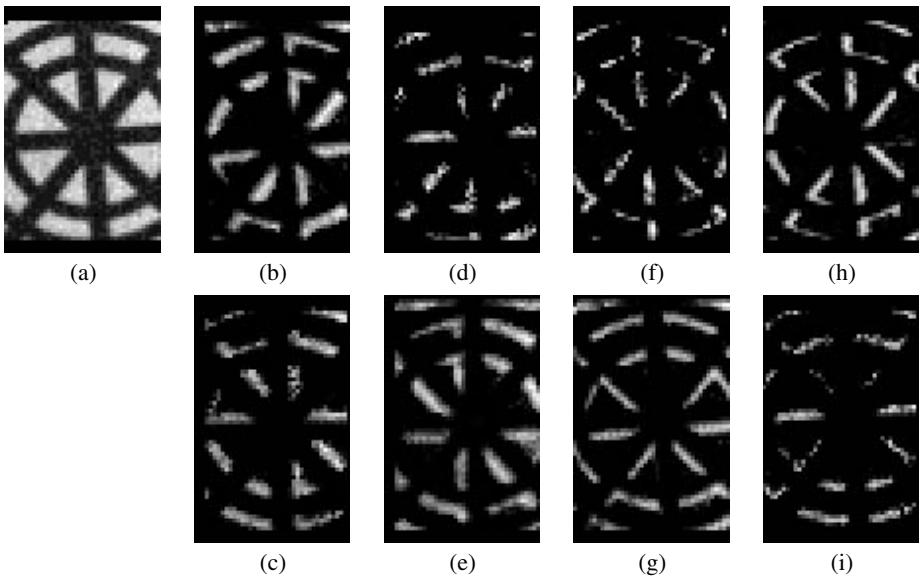

Fig. 4. Video-rate frame captures showing eight different types of simple cell outputs. (a) Original image from the octopus retina, (b–i) output from arrays of IFAT cells implementing receptive fields RF1–RF8 (cf. Fig. 3).

preipherals," *IEEE Trans. Neural Networks*, vol. 4, no. 3, pp. 523–528, 1993.

[4] T. Y. W. Choi, B. E. Shi, and K. A. Boahen, "A multi-chip implementation of cortical orientation hypercolumns," in *ISCAS'04*, vol. 3, May 2004, pp. 13–16.

[5] ——, "An ON-OFF orientation selective address event representation image transceiver chip," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 51, no. 2, pp. 342–353, February 2004.

[6] B. E. Shi, E. K. C. Tsang, and P. S. P. Au, "An ON-OFF temporal filter circuit for visual motion analysis," in *ISCAS'04*, vol. 3, May 2004, pp. 85–88.

[7] E. Chicca, G. Indiveri, and R. J. Douglas, "An event-based VLSI network of integrate-and-fire neurons," in *ISCAS'04*, vol. 5, May 2004, pp. 357–360.

[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, no. 6-7, pp. 781–793, 2001.

[9] C. M. Higgins and C. Koch, "Multi-chip neuromorphic motion processing," in *Proc. 20th Anniversary Conference on Advanced Research in VLSI*, D. S. Wills and S. P. DeWeerth, Eds. Los Alamitos, CA: IEEE Computer Society, 1999, pp. 309–323.

[10] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas, "Orientation-selective aVLSI spiking neurons," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002.

[11] N. Kumar, W. Himmelbauer, G. Cauwenberghs, and A. Andreou, "An analog VLSI chip with asynchronous interface for auditory feature extraction," *IEEE Trans. Circuits and Systems II*, vol. 45, no. 5, pp. 600–606, 1998.

[12] R. J. Vogelstein, U. Mallik, E. Culurciello, G. Cauwenberghs, and R. Etienne-Cummings, "Saliency-driven image acuity modulation on a reconfigurable silicon array of spiking neurons," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005.

[13] E. Culurciello, R. Etienne-Cummings, and K. A. Boahen, "A biomorphic digital image sensor," *IEEE J. Solid-State Circuits*, vol. 38, no. 2, 2003.

[14] ——, "Second generation of high dynamic range, arbitrated digital imager," in *ISCAS'04*, vol. 4, May 2004, pp. 828–831.

[15] R. J. Vogelstein, U. Mallik, E. Culurciello, R. Etienne-Cummings, and G. Cauwenberghs, "Spatial acuity modulation of an address-event imager," in *ICECS'04*, 2004.

[16] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[17] R. J. Vogelstein, U. Mallik, and G. Cauwenberghs, "Silicon spike-based synaptic array and address-event transceiver," in *ISCAS'04*, vol. 5, May 2004, pp. 385–388.

[18] A. J. Yu, M. A. Giese, and T. A. Poggio, "Biophysiologically plausible implementations of the maximum operation," *Neural Computation*, vol. 14, no. 12, pp. 2857–2881, 2002.

[19] R. J. Vogelstein, U. Mallik, and G. Cauwenberghs, "Array of spiking silicon neurons and address-event transceiver," manuscript in preparation.

[20] R. J. Vogelstein, F. Tenore, R. Philipp, M. S. Adlerstein, D. H. Goldberg, and G. Cauwenberghs, "Spike timing-dependent plasticity in the address domain," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003.