SPATIAL ACUITY MODULATION OF AN ADDRESS-EVENT IMAGER

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

[1]Department of Biomedical Engineering
[2]Department of Electrical and Computer Engineering
Johns Hopkins University, Baltimore, MD 21218
{*jvogelst,umallik1,euge,gert,retienne*}*@jhu.edu*

***Abstract*— We have implemented a foveated vision system that uses an 80 × 60 address-event neuromorphic image sensor and varies its effective spatial acuity throughout the visual field to use bandwidth more efficiently. Low resolution pixels are created by pooling spikes from neighboring photoreceptors using overlapping Gaussian kernels. The pooling is effected by routing address-events through a memory-based projective field mapping, and new pixels are created by summing inputs on an array of integrate-and-fire neurons. The fovea can be relocated without physically moving the imager by performing a simple computation in the address domain. We provide an example of the system in operation and compute a figure of merit to quantify foveation savings.**

I. INTRODUCTION

Recent advancements in the field of autonomous, battery-powered sensor networks have allowed for collection of enormous amounts of information from the environment at an extremely low energy cost. Unfortunately, the transmission of information and the subsequent processing cannot typically be performed with such efficiency. Furthermore, most of the data extracted from the environment is not relevant, particularly in the case of wide-angle image sensors in a monitoring sensor network. It is clear that a reduction of data would be useful, and is probably necessary if these networks are to be deployed with a useful battery life.

To reduce the amount of data transmitted, we use a foveated vision system. Similar to the human retina, the system produces a small region of high acuity (called the *fovea*) surrounded by areas of progressively lower resolution. It can emulate uniformly high resolution by moving the fovea to different parts of the visual field; in most cases, however, the sensor will fixate a target of interest and move the fovea only if the target moves or another salient object appears.

One way to perform spatial acuity modulation for an imager would be to implement a fovea on-chip by varying the pixel size, i.e. create a high-acuity region surrounded by areas of lower acuity, like the human retina [1]. However, this solution would require physically moving the imager through space in order to relocate the fovea, which is undesirable due to the increased size, power requirements, and control systems required to do this. Therefore, we have implemented a relocatable "virtual fovea" by processing the outputs of an address-event image sensor through an integrate-and-fire array transceiver (IFAT) [2], which can easily handle a large volume of address-events.

The IFAT is a reconfigurable array of integrate-and-fire neurons that uses a memory-based (RAM) projective field mapping to route incoming address-events to different target neurons. For this work, we configured it to emulate a foveated retina. By modifying the contents of its RAM, or, equivalently, adding constant offsets to incoming events (Section IV), the high-acuity region can be moved to different parts of the visual field. Depending on the size of the virtual fovea and the width of the spatial filters implementing the low-acuity regions, the stream of address-events generated by the IFAT can require significantly lower bandwidth than the original image. For example, our 16 × 16 pixel fovea produces a 75% decrease in event rate (Section IV). Furthermore, if an address domain feature detection algorithm is applied to the retinal data stream, the IFAT could perform "content-dependent" foveation [3]. This would allow for high-resolution scrutiny of specific regions of interest, while transmitting a generally low-resolution image over the communication channel. This capability is desirable for smart, low-bandwidth surveillance networks.

The remainder of this paper is divided into four sections. Section II introduces address-events. Section III briefly discusses the operation of the address-event imager and the IFAT. Section IV presents the experimental setup and results. Finally, Section V concludes the paper.

II. ADDRESS-EVENTS

The "Address-Event Representation" (AER) communication protocol uses time-multiplexing on a fast data bus to emulate extensive parallel connectivity between neurons. Originally proposed to implement a one-to-one connection topology [4], AER is now the common language of neuromorphic chips and has been extended to allow convergent and divergent connections [5]–[10].

A typical AE system has one group of neurons that transmits events and another group that receives events (although recurrent connections are also common), and each neuron is assigned unique address. To form an AE "connection" between

0-7803-8715-5/04/$20.00

cells (called a *synapse*), the sending neuron first places its address on the AE bus and asserts an asynchronous handshaking request signal. When the receiving group is ready, it observes the address, activates the appropriate target neuron(s), and asserts the handshaking acknowledge signal, completing the transaction.

In a traditional AE system, targets are determined using a hardwired decoding scheme. However, more complex or flexible routing is possible by storing a look-up table in RAM with a list of recipients corresponding to each sender's address [5], [9], [10]. The IFAT system used for this project employs such a memory-based solution.

III. Hardware

A. Address-Event Imager

Our neuromorphic retina is a second-generation, fully arbitrated 80 $\times$ 60 address-event imager based on the simple phototransduction chain of the octopus retina [11]. Each pixel in the "Octopus Retina" (OR) chip (Figure 1a) operates like an integrate-and-fire neuron, integrating photons collected by an n-type photodiode on a 0.1pF capacitor until the voltage exceeds a threshold. When this occurs, the pixel generates a request to be serviced by the arbitration circuitry on the periphery of the array (Figure 1b). When the arbiter acknowledges the pixel, it places the pixel's address on the address-event bus and resets the integration capacitor. The intensity of light at a given pixel is therefore proportional to the inverse of the time between successive events.

Unlike serially-scanned imager arrays, in which every analog pixel is read in sequence, the OR requires individual pixels to request access to the output bus after they have collected a sufficient number of photons. By implementing a pipelined read-out scheme and an efficient arbitration tree, the octopus retina produces a mean event rate of 200,000 address-events per second (41.7 effective fps) under uniform indoor light (0.1 mW/cm^2) while providing 120dB of dynamic range and consuming 3.4mW. At full-speed, the imager produces 40 million address-events per second (8,300 effective fps) and consumes 71mW.

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

The IFAT system functions as a reconfigurable array of integrate-and-fire neurons. Its primary components are a fast digital microprocessor, 128 MB of RAM, and up to four custom aVLSI chips that each implement 2,400 integrate-and-fire neurons (Figure 2). The microprocessor controls a bidirectional AER bus and performs handshaking with both the aVLSI chips and external neuromorphic systems. When an external device sends a spike to the IFAT, a "chip identifier bit" is added to the sending neuron's address and the result is used as an index into the RAM. Each 32-bit wide RAM entry stores two main pieces of information about a single AER synapse: the receiving neuron's address and the synapse parameters (see below). These data are then used to activate the appropriate neuron on the aVLSI chips. A special "stop code" in the RAM data field indicates that the sending neuron has no more connections; otherwise, an offset is added to the RAM index and another synapse is activated. If an incoming AE results in a spike from one of the on-board integrate-and-fire neurons, the same sequence of events occurs and the RAM data can either specify a recurrent connection or an external target. In this way, arbitrary patterns of connectivity can be implemented, limited only by the memory's capacity.

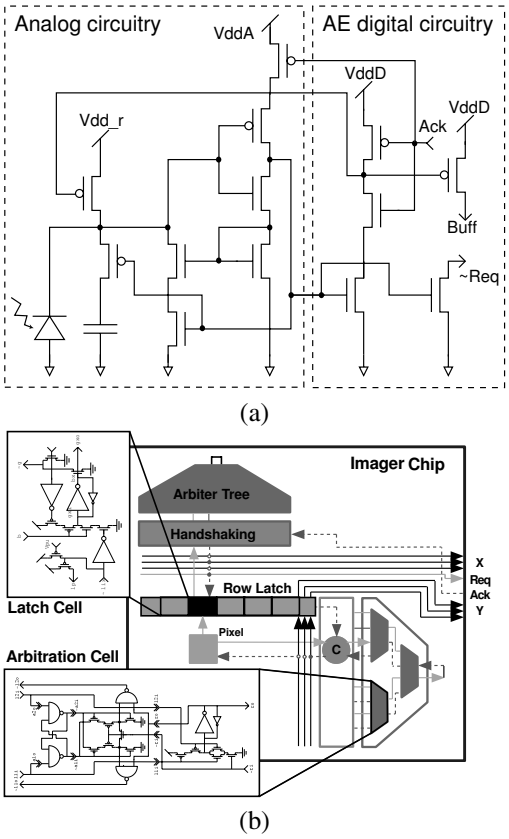


Fig. 1. (a) Octopus Retina pixel schematic. (b) Row and column arbitration architecture. Both figures from [12].

The 2,400 integrate-and-fire neurons on every custom aVLSI chip are identical and each implement a simple conductance-like model of a single, "general-purpose" synapse (Figure 3a) [2]. The synapse has two internal parameters that can be dynamically modulated for each incoming event: the synaptic driving potential (E) and the synaptic weight (W0-W2). Additionally, the IFAT microprocessor can control two external synapse parameters: the number of events to send and the probability of sending an event. By varying these parameters, it is possible to emulate a large number of different kinds of synapses.

IV. Experimental Setup and Results

The goal of these experiments was to locate a feature on a target image and then modulate the spatial acuity of the Octopus Retina around that location in order to reduce its bandwidth requirements and minimize the amount of extraneous information transmitted. There are at least two ways to implement this functionality—the center of vision could

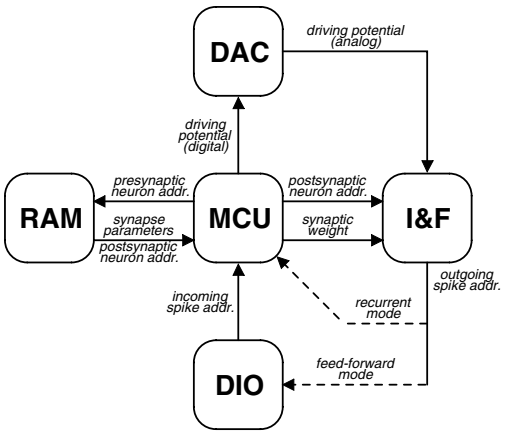


Fig. 2. Block diagram of IFAT system. Incoming and outgoing address-events are communicated through the digital I/O port (DIO), with handshaking executed by the microcontroller (MCU). The digital-to-analog converter (DAC) is controlled by the MCU and provides the synaptic driving potential ('E' in Figure 3a) to the integrate-and-fire neurons (I&F), according to the synapse parameters stored in memory (RAM).

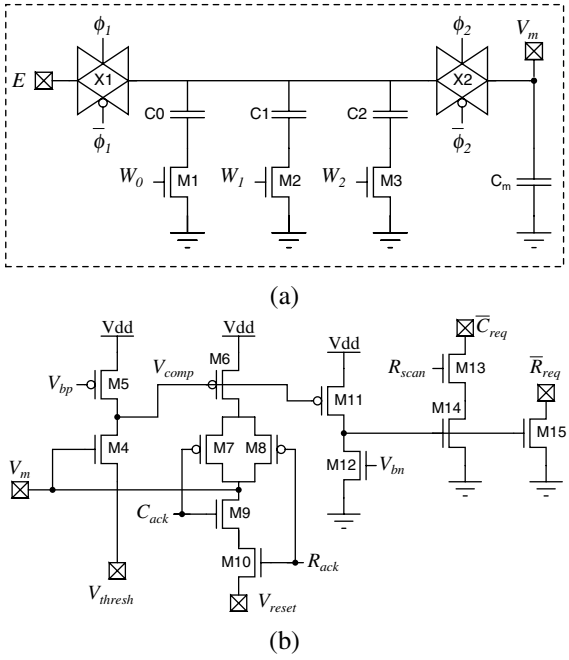


Fig. 3. (a) "General-purpose" synapse contained within each integrate-and-fire neuron. (b) Spike-generation and address-event handshaking circuitry. Modified from [13].

be scanned across the visual field, simultaneously allowing a feature-detection algorithm to search the high-resolution area; or the feature-detection algorithm could process the image before foveation and direct the center of vision to the location of interest. In this paper we have taken the second approach. The following sections describe the connectivity between OR and IFAT neurons, our strategy for moving the center of vision, the analysis, and our results.

A. Network Connectivity

The address space for events coming into the IFAT is 14-bits wide and can be visualized as a 128 $\times$ 128 pixel visual field. The center of the visual field contains a 16 $\times$ 16 high-resolution fovea surrounded by concentric rings of progressively lower resolution. In the high-resolution region, there is a one-to-one mapping of incoming addresses to IFAT neurons, and every synapse is assigned the highest connection strength possible. Within the inner-most concentric ring, inputs from 4 $\times$ 4 blocks of neighboring pixels converge onto single IFAT neurons, with each block overlapping the next by half the center width. The synaptic weights for these connections are arranged like a 2-D Gaussian filter, with stronger connections for the center pixels and weaker connections for the peripheral pixels. Overall, the connection strengths from pixels arranged in 4 $\times$ 4 blocks are lower than those for the one-to-one connections in order to equalize the total amount of excitation received by each IFAT neuron. The outer two concentric rings have similar connection patterns but larger block sizes: 8 $\times$ 8 and 16 $\times$ 16. In total, the 16,384 possible inputs to the IFAT converge onto 1,650 integrate-and-fire neurons.

By enlarging the visual field to 128 $\times$ 128 pixels even though the OR is only 80 $\times$ 60 pixels, it is possible to change the location of the center of vision. By default, all 4,800 OR pixels will fall in the upper-left region of the IFAT visual field, placing the high-resolution region over the bottom-right portion of the captured image. However, adding a constant offset to either the x or y component of the address moves the fovea to the right or up, respectively. Between the minimum and maximum offsets, the fovea can traverse 64 $\times$ 84 pixels.

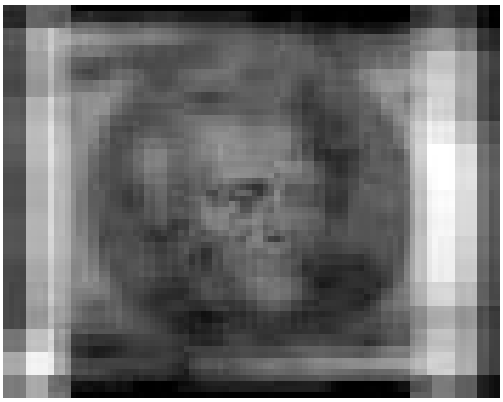
B. Analysis and Results

A test image, reconstructed from the output of the OR, and its foveated counterpart, reconstructed from the output of the IFAT, are shown in Figure 4. The location of the center of vision was determined by using a simple template matching algorithm to locate eyes on the OR image. Whereas the original image was constructed from approximately 100,000 address-events, the foveated image was built from approximately 25,000 address-events, 75% fewer. Additional savings can be achieved by increasing the threshold of the IFAT neurons or globally reducing the weights between OR and IFAT neurons. If these parameters are modulated in response to ongoing network activity, it is possible to achieve a form of automatic gain control, so that both dim- and bright-light images occupy the same bandwidth.

An even greater reduction in bandwidth and power are achieved with larger imagers. Assuming that the 16 $\times$ 16 fovea remains fixed and that successive concentric rings use overlapping blocks that increase in size by factors of two (i.e. the innermost ring is composed of 4 $\times$ 4 blocks, the next ring is composed of 8 $\times$ 8 blocks, etc.), the approximate number of pixels M in a foveated imager with k concentric rings is: $M = 16^2 + 2^{(k+2)} - 4$, which is much fewer than the N^2 pixels in the unfoveated version.

Figure 5 shows an analysis of the energy costs of transmitting images over a wireless link with and without foveation

(a)



(b)

Fig. 4. Test image (a) before and (b) after foveation. Center of vision is Jackson's left eye.

[14]. For the purposes of this computation, a frame is assumed to consist of approximately two events from every pixel, the minimum data required to estimate light intensity from inter-spike intervals. While the energy cost increases linearly with the number of pixels in an unfoveated system, a foveated system uses almost the same energy as image size increases. In the case of the 4,800 pixel OR, the energy savings are low, but significant savings are possible when larger sensors are considered.

V. Conclusion

We have presented a system that can modulate the spatial acuity of an address-event imager by routing spikes through a reconfigurable array of integrate-and-fire neurons and specifying convergent connections between neighboring pixels. The resulting stream of events occupies less bandwidth and transmits more relevant information over the communication channel. Future work will focus on implementing feature detection algorithms in the address domain to enable dynamic repositioning of the fovea, and on actively modulating the threshold of the integrate-and-fire neurons to provide automatic gain control.

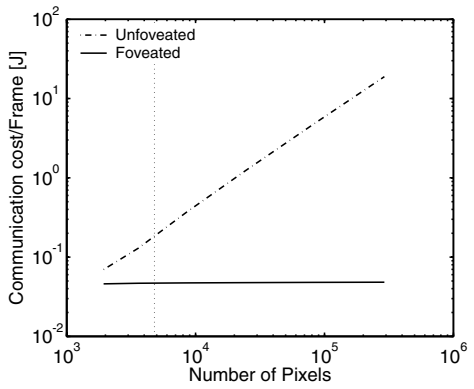


Fig. 5. Energy costs of transmitting images over a wireless link. The 4,800 pixels contained in the OR is marked with a dotted line.

Acknowledgments

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.

References

[1] R. Etienne-Cummings, J. V. der Spiegel, and P. Mueller, "A foveated silicon retina for two-dimensional tracking," *IEEE Trans. Circ. Sys. II*, vol. 47, no. 6, 2000.

[2] R. J. Vogelstein, U. Mallik, and G. Cauwenberghs, "Silicon spike-based synaptic array and address-event transceiver," *ISCAS'04*, 2004.

[3] M. Massie, C. Baxter, J. P. Curzan, P. McCarley, and R. Etienne-Cummings, "Vision chip for navigating and controlling micro unmanned aerial vehicles," *ISCAS'03*, vol. 3, pp. 786–789, 2003.

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

[5] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circ. Sys.-II*, vol. 47, no. 5, pp. 416–434, 2000.

[6] C. M. Higgins and C. Koch, "Multi-chip neuromorphic motion processing," in *Proc. 20th Ann. Conf. Adv. Res. VLSI*, D. S. Wills and S. P. DeWeerth, Eds. Los Alamitos, CA: IEEE Computer Society, 1999, pp. 309–323.

[7] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas, "Orientation-selective aVLSI spiking neurons," in *Adv. Neural Info. Proc. Sys. 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA: MIT Press, 2002.

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

[9] G. Indiveri, A. M. Whatley, and J. Kramer, "A reconfigurable neuromorphic VLSI multi-chip system applied to visual motion computation," *Proc. MicroNeuro'99*, Apr. 1999.

[10] 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.

[11] E. Culurciello, R. Etienne-Cummings, and K. Boahen, "High dynamic range, arbitrated address event representation digital imager," *ISCAS'04*, 2004.

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

[13] 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 *Adv. Neural Info. Proc. Sys. 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003.

[14] C. Bandera and P. D. Scott, "Foveal machine vision systems," *Proc. IEEE International Conf. on Systems, Man and Cybernetics*, vol. 2, pp. 596–599, 1989.