Retinomorphic Channel Design and Considerations

Jonah P. Sengupta*, Andreas G. Andreou*
*Department of Electrical & Computer Engineering, Johns Hopkins University
Email: {jsengup1, andreou} @jhu.edu

2023 57th Annual Conference on Information Sciences and Systems (CISS) | 978-1-6654-5181-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/CISS56502.2023.10089617

***Abstract*—By extrapolating functionality from the retina, retinomorphic engineering has yielded devices that have shown promise to alleviate the challenges presented in modern computer vision tasks. An incredible amount of work has been devoted in recent years to the development and deployment of these event-based vision sensors in applications requiring low-latency, energy-efficient, high dynamic range sensing solutions. However, not much work has been devoted to the area, energy, and speed analysis of the various encoding and decoding mechanisms necessary for sensory pipelines. This paper outlines an empirical framework that presents a clear tradeoff between the various methodologies to transduce physical information in to spikes (encoding) and reconstruct said stimuli from the incident events (decoding). Software-based models of these methodologies were constructed to evaluate the accuracy of stimuli reconstruction for a variety of input profiles. As a result, it is shown that an optimized retinomorphic architecture for a specific set of system-driven cost metrics requires a heterogenous fabric of encoders with a composition of 95% temporal contrast pixels and 5% intensity encoder assuming a temporal jitter of 1μs. Much like the composition of ganglion cells in magno- and parvo-cellular pathway, this multi-modal solution provides the most time, area, and power efficient method to convey visual data.**

***Index Terms*—neuromorphic vision, system design, information theory, retinomorphic engineering**

I. Introduction

Neural circuits in the inner and outer plexiform layers of the retina provide the necessary building blocks to efficient encode visual stimuli. The outer plexiform layer responses consists of a horizontal cell shunting inhibition, pooled average receptor response and local contrast enhancement. This spatial, bandpass response reduces transmitting spikes corresponding to redundant, low-frequency background while attenuating contributions from high-frequency, uncorrelated fixed pattern noise. Bipolar cell connectivity and amacrine cell feedback in the inner plexiform layer amount to the transient response of the α-ganglion cells. However, output activity of the retina does not consist of a singular encoding modality, but rather a mix of sustained and transient responses. As will be suggested in this work, multi-modal encoding provides the means to maximized information transmission while reducing energy consumption.

Retinomorphic engineering, coined by Kwabena Boahen in 1996 [1], has sought to duplicate a vast array of the retina's characteristics and processing capabilites in silicon-based sensor arrays. Boahen goes on to remark that retinomorphic systems should abide by many of the same characteristics of the retina: local automatic gain control, bandpass spatiotemporal filtering, and adaptive quantization and bandwidth utilization. The results of such efforts are known as **silicon retina**.

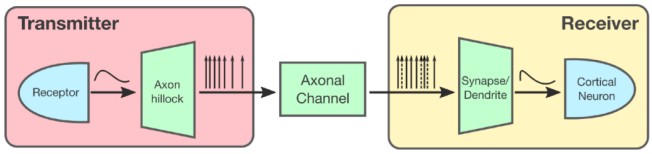


Fig. 1: Retinomorphic Communication Pipeline: a transmitter (enclosed in light red) preconditions and encodes a sensory signal into a spike train which is decoded by the receiver (yellow) where the signal is ultimately reconstructed. Analog domain blocks are blue and digital are in green. Spikes are removed and timing is disturbed after propagating through the axonal channel, leading to decoding errors, represented as dotted spikes.

Many retinomorphic sensors have been reported to efficiently solve the pressing problems in machine vision [2]–[6]. However, these sensors neglect the important considerations that can be gleaned from the retina. For one, modern arrays only implement encoder functionality that mirrors that seen in the magnocellular pathway [7]. This neglects the rich information offered in the parvocellular pathway which is comprised of sustained responses from color sensitive, densely packed midget ganglion cells. Multi-modality is essential to not only detect changes in the scene, but reconstruct high spatial frequencies and wide-spectrum intensity profiles. Furthermore, retinomorphic sensors are designed without consideration of downstream decoding mechanisms. These latter architectures, if not effectively implemented with respect to speed and area, will reduce the gains realized by the retinomorphic sensor.

This work highlights the different encoding and decoding mechanisms used for retinomorphic communication. It outlines the benefits and shortfalls of each in the context of architecture speed, area, and power to rationalize the need for the multi-modal approach seen in the mammalian retina. Section II provides an overview of the channel model used along with some other system considerations. Sections III and IV provided an overview of encoding and decoding methodologies respectively. Section V presents numerical analysis of the retinomorphic channel before concluding in Section VI.

II. Retinomorphic Channels

The basic structure of a biological computation and communication pipeline is seen in Figure 1. A sensory receptor is responsible for transducing physical phenomenon into an analog signal which is then filtered using spatiotemporal information. These early processing stages are responsible

978-1-6654-5181-9/23/$31.00 ©2023 IEEE

for encoding the signal based its temporal characteristics, spatial prevalence, or absolute magnitude. Analog encoding neuron responses are integrated onto the soma of an output neuron whose spike activity is proportional to its inputs and local, axonal parasitics. Stochastic dynamics of the axon subsequently cause the loss of spikes and temporal jitter. This lossy spike train is decoded via synapses which project onto the dendrite of a receiver neuron. Final reconstruction takes place utilizing the inverse operation used to encode the stimuli. However, system output is corrupted by analog noise, dropout of spikes, and temporal jitter which skewed spike timings. High-level, hardware analysis of this framework is presented in this chapter and software/hardware implementations will be explored in subsequent chapters of this work.

Encoding methodologies and design trade-offs with respect to power, area, and speed are provided here. However, basic rationale for retinomorphic sensing arrays need to be provided. System switching activity provides a simple metric to compare hardware platforms. For a retinomorphic architecture to be more appealing from a energy standpoint, array switching activity needs to be less than that of a frame-based platform. This trade-off, built off the analysis from [1], can be outlined as such

$$f_{spk,arr} < f_{aps} \quad (1)$$

$$N_{row}N_{col}\widehat{f}_{spk}\log_2(N_{row}N_{col}-1) < N_{row}N_{col}f_{fps}M \quad (2)$$

$$(\hat{f}_{act}\alpha - \hat{f}_{bk}(1-\alpha))\log_2(N_{row}N_{col}-1) < f_{fps}M \quad (3)$$

$$(\hat{f}_{act}\alpha - \frac{\hat{f}_{act}}{\gamma}(1-\alpha))\log_2(N_{row}N_{col}-1) < f_{fps}M \quad (4)$$

where the retinomorphic array output bandwidth, $f_{spk,arr}$, and frame-based sensor throughput, f_{aps} can be deconstructed into relation of average pixel firing rate for stimuli driven events, f_{act}, and frame-rate, f_{fps}. Parameter α denotes the probability a pixel fires with rate f_{act} given local encoding and signal processing, while γ denotes the ratio between this active frequency and a resting firing rate. Output bits from a retinomorphic system are AER driven thus they denote spatial locality of event activation, $\log_2 N_{pix} - 1$. Spike timings and firing rates encode stimuli information. In contrast, spatial location is inherently represented in the scanning approaches of frame-based sensors, but information is encoded as a digital value of size M. By equating $f_{act} = f_{aps}$, the maximum retinomorphic activity level tolerated for given array size (N_{pix}) and attenuation factor γ is

$$\alpha < \frac{\gamma}{\gamma-1}\left(\frac{M}{\log_2(N_{pix}-1)} - \frac{1}{\gamma}\right) \quad (5)$$

Figure 2 is a plot of this maximum tolerated pixel activity, α, for a set of array sizes and attenuation factors, γ. When the resting spike rate goes to 0 and γ diverges to infinity, the maximum switching activity converges to $\frac{M}{\log_2(N_{pix}-1)}$. For an array size of 10^6, the activity is bounded to a value of 40.17% meaning a pixel can only be in an active state for less than 40% of the time and various stimuli profiles. Therefore, in-pixel adaptation and spatiotemporal encoding need to be utilized to reduce activity rates.

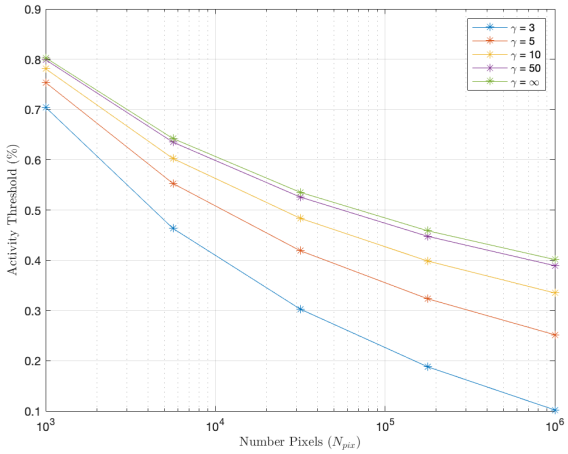


Fig. 2: Activity level threshold, α, for retinomorphic platform: average pixel frequency such that frame-based platforms are inferior from an energy standpoint. Metrics are extracted using increasing array sizes and baseline activity attenuation, γ.

III. Encoding Schema

As a baseline mechanism, absolute intensity can be encoded into a spike train by integrating input synaptic current or photocurrent onto a comparator node which fires a pulse when the potential has exceeded a threshold. Therefore, the simple integrate-and-fire pixel encodes absolute intensity into a pulse-frequency modulated, spike train with rate:

$$f_{spk,iaf} = \frac{i_{enc}}{V_{th}C_{mem}} \quad (6)$$

where $i_{enc} = i_{ph}$ is the photocurrent from the sensing element, V_{th} is the requisite threshold the potential reaches to fire a spike, and C_{mem} is the membrane capacitance. However, it can be seen from the expression that high intensity background profiles would result in high firing rates. Therefore, as seen in the outer plexiform layer, spatial filtering to reduce this redundancy. Spatial contrast is defined as

$$i_{enc,scon} = i_{norm}\left|\frac{i_{ph} - i_{bck}}{i_{bck}}\right| \quad (7)$$

where the ratio of difference between input, i_{ph}, and background intensity, i_{bck}, and the latter is known as Weber Contrast which is converted into a current response using i_{norm}. Output response i_{scon} can be similarly integrated and transduced into a spike train using the IAF neuron encoder. A potential downfall of this encoding methodology is the presence of a sustained response to high contrast stimuli. To address this, transient stimuli can be encoded

$$i_{enc,tcon} = i_{ph} - (i_{ph} * H_{lpf}) = \frac{di_{ph}}{dt} \quad (8)$$

where $i_{enc,tcon}$ represents the temporal contrast response, or temporal high-pass response, of the transient encoder. This

consists of computing the difference between the input current and a temporal average.

IV. Decoding Schema

After the stimuli current from the photodiode is processed using spatiotemporal filtering chain, output encoder current is transduced into a spike stream. These address events travel along varying degrees of the hardware communication chain before arriving at a software or hardware-realized decoder architecture. Function of these decoders are similar independent of implementation medium: events are received, processed, and used as the means to reconstruct the input stimuli or accomplish another application task. These events can arrive via direct 1-to-1 connection realized via 3D integration [8] or via address-driven selection after optional computation-driven remapping [9], [10]. Decoding elements can be organized into two subclasses: modules that realize linear filters and ultimately integrate the events or elements that use the precise timing between spikes as the basis for computation. The former are classified as linear decoders or rate-code decoders as they utilize linear filters to accumulate spikes over a temporal window. Common linear decoding filters are of low-pass type since spike integration is used to reconstruct the signal. This can be accomplished using, among others, a first-order RC filter or box filter with rectangular window. The former, exponential filter impulse response can be expressed as

$$g_{exp}(t) = A_g e^{\frac{-t}{\tau_e}} \tag{9}$$

uch that τ_e is the pole of the low-pass filter and A_g is the filter gain. First-order nature of the filter has lent itself to straightforward implementation in analog hardware as a model of a synapse [1], [10]–[12]. A rectangular filter counts the number of spikes in a sliding temporal of width t_r. Filter impulse response is expressed as

$$g_{rect}(t) = A_g(1 - u(t - \tau_r)) \tag{10}$$

Instances of uniform, moving window synaptic responses are uncommon in the brain and are not straightforward to implement in analog or digital hardware.

Alternatively, a non-linear, temporal-code decoder uses the instantaneous interspike timing to reconstruct the encoded signal [13]. This temporal-coded method utilizes local memory to encode recent spike timings and computational elements to calculate the intervals, $t'_{isi,i}$:

$$t_{isi,i} = t_{spk,i} - t_{spk,i-1} \tag{11}$$

Ultimately, decoder reconstruction of the input stimuli is achieved by observing the spike rate-transfer function from the IAF neuron:

$$\hat{i}(t) = \frac{C_{mem}V_{thr}}{t_{isi,i}} = \frac{Q_{thr}}{t_{isi,i}} \tag{12}$$

where $\hat{i}(t)$ is the reconstructed stimulus.

V. Results and Analysis

A simulation framework of the communication channel presented in Figure 1 was implemented to analyze the trade-off of encoder and decoder designs with respect to error rates and power consumption in the presence of channel noise. The intention of this effort was to inform the composition of future retinomorphic arrays by maximizing performance over these system parameters.

Two noise effects were included to capture the stochastic, mixed-signal implementation of retinomorphic arrays. First, random spike dropout was used to model the impact of fixed-pattern noise which affects neuron firing rates [2], pixel-level temporal noise which shunts event activation [14], or receiver subsampling methodologies to compensate for buffer utilization [15]. In the analysis model, dropout events, $D(p_d)$ is a vector of size N_{spks} where each element is a Bernoulli random variable with $p = P_d$. For instance, a $p_d = 0$ represents a lossless channel with no dropout events. In tandem, a temporal jitter was added to the spike timings generated by the neuron backend in the encoders. This is a manifestation of some combination of fixed pattern, shot, flicker, and thermal noise, all of which are present with varying degrees in the analog encoder circuits. After the injection of these noise mechanism, each spike timing $t_{i,n}$becomes:

$$t_{i,n} = t_i * (1 - D_i) + \mathcal{N}(0, \sigma_j^2) \tag{13}$$

where t_i is the pre-noise spike time of spike i, D_i is the dropout random variable for spike i, and the latter term represents the temporal jitter which is a Normal random variable with standard deviation σ_j. In the presence of these noise mechanisms, the channel performance was captured in terms of mean-squared error:

$$\hat{e}(\mathbb{F}_{enc}, \mathbb{G}_{dec}, p_d, \sigma_j) = \frac{1}{T_{stim}} \int_0^{T_{stim}} (\hat{i}(t) - i(t))^2 \tag{14}$$

which is a function of the encoding ($\mathbb{F}_{enc}$) and decoding $\mathbb{G}_{dec}$ schemes as well as the noise parameters defined above. This metric encapsulates the difference between the reconstruction $\hat{i}(t)$ and a input stimulus, $i(t)$ over the period T_{stim}.

After applying the decoding method, an inverse mapping of the analog encoding needs to be applied in order to complete the reconstruction. Since the absolute intensity, IAF pixel does not apply any spatial or temporal processing, the backend reconstruction $\hat{i}_{iaf}(t)$ is complete post-decoding. However, for the spatial encoder, the background intensity (i_{bck}), contrast polarity (s_{sc}), and normalization current (i_{norm}) are used to produce the reconstructed current:

$$\hat{i}_{scon}(t) = \frac{i_d(t)s_{sc}i_{bck}}{i_{norm}} - i_{bck} \tag{15}$$

Since the TCON encoding is performing a temporal derivative on the input stimuli, reconstruction consists of doing the temporal integration of this decoder output whose sign is dictated by the temporal contrast polarity, s_{tc}:

$$\hat{i}_{tcon}(t) = \hat{i}_{tcon}(t-1) + s_{tc}(i_d(t)) \tag{16}$$

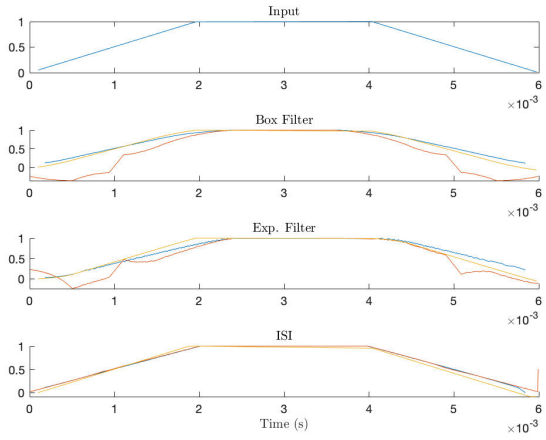
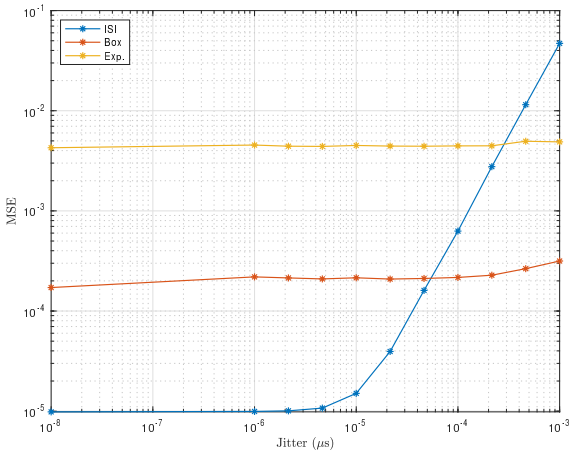


Fig. 3: Reconstruction performance of encoder-decoder pairs: SCON-orange, TCON-yellow, IAF-blue. First panel - input stimuli. Second panel - reconstruction using box filter. Third panel - reconstruction using exponential filter. Fourth - reconstruction using interspike interval

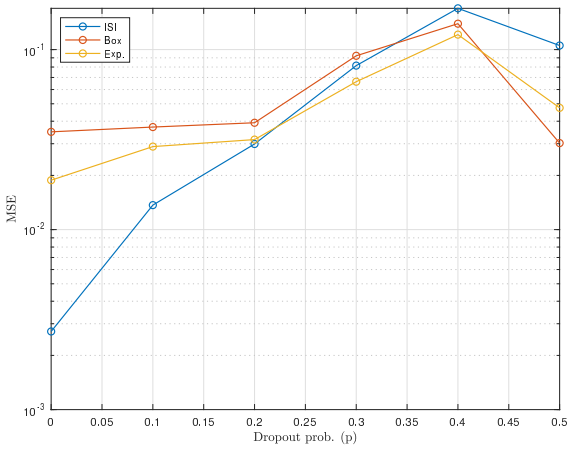
Figure 3 depicts the reconstruction results for the three encoder methods using the three decoders without the presence of noise. Reconstruction error for the SCON method stems from the latency introduced by the rate-coded decoders as a τ_e and τ_r time is needed to accumulate a high spike rate that results from high, negative contrast at the advent of the stimulus. In addition, the edges of the stimulus are delayed in the reconstruction using the rate-coded approaches due to the latency of each filter, $\tau_e = 250\mu s$ and $\tau_r = 1ms$ thereby highlighting a downside of utilizing linear methods.

Figure 4 outlines the MSE of the different decoder methods versus temporal jitter and spike dropout rates respectively. The accuracy was extracted for each encoder method and averaged for each decoder technique to provide a curve that can be utilized for analysis. For jitter profiles with $\sigma_j < 10\mu s$, the ISI decoder proves to have the smallest error for the given stimulus. A similar relationship is seen with spike streams with small dropout rates. Following intuition, the ISI accuracy deteriorates when dropout rate and jitter increase while linear decoder accuracy maintains. This is because the latter methods effectively average noise within a temporal window of τ width.

Figure 5 is a plot of MSE for TCON and IAF encoder methods averaged over all decoder techniques. These plots highlight a distinct trade-off between the encoder approaches. The TCON pixel is strictly worse for all jitter profiles given dropout rate of 0 %. However, with a jitter distribution with deviation of $1\mu s$, the IAF pixels become worse from an error perspective with dropout rates within 0.1 and 0.25. These results also highlight a potential shortfall of the TCON pixel. When the channel is noisy and induces stochastic temporal jitter or lossy transmission, reconstruction accuracy suffers.



(a) MSE as a function of jitter, $p_d = 0$



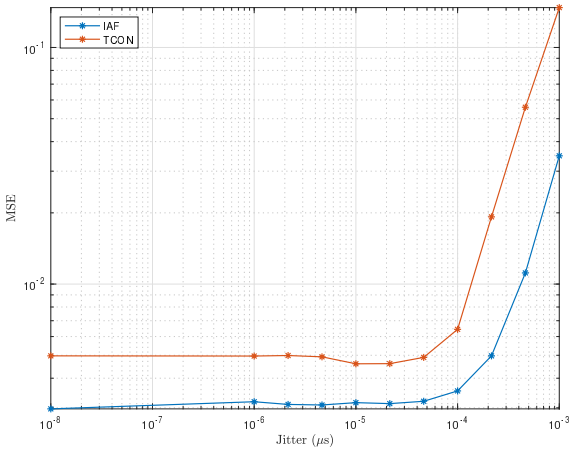
(b) MSE as a function of dropout, $\sigma_j = 1\mu s$

Fig. 4: MSE with respective to decoder method (tc=temporal code, rect=box filter, exp=exponential filter). The top and bottom panel outlines accuracy of each method with respect to temporal jitter and spike dropout rates respectively.

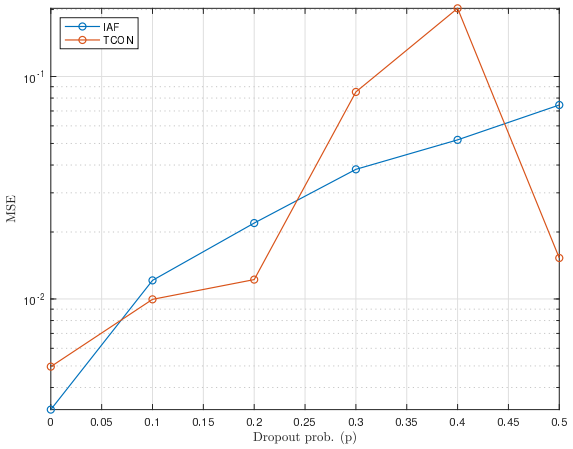
This is due to the TCON pixel's high attenuation factor (γ) within a static scene.

Encoder-dependent, noise artifacts, which are a result of circuit complexity, can be analyzed using this simple channel model. For instance, as mentioned in Sections III and IV, TCON and box filter analog implementations are complex hence one could expect higher jitter and dropout rates, say $\mu_j = 10\mu s$ and $p_d = 0.1$. This can be compared to a IAF pixel with ISI decoder. In this configuration, lack of encoding reduces complexity-driven noise and low jitter, but increases the change of channel congestion which leads to higher dropout rates in the readout. For a $\sigma_j = 0.01\mu s$ and $p_d = 0.2$, the $\hat{e}_{IAF,ISI} > \hat{e}_{TCON,BOX}$ despite lower complexity.

Error and circuit complexity is not the only constraint



(a) MSE as a function of jitter, $p_d = 0$



(b) MSE as a function of dropout, $\sigma_j = 1\mu s$

Fig. 5: MSE with respective to encoder method (IAF and TCON). The top and bottom panel outlines accuracy of each method with respect to temporal jitter and spike dropout rates respectively.

for retinomorphic systems. In large arrays that are active in embedded contexts, low power consumption is paramount. Power of each encoder methodology can be approximated by multiplying a nominal energy per spike by output spike rate. The former is a result of circuit complexity and design realization while latter is dependent on encoder attenuation factor, γ ($\gamma_{tcon} >> \gamma_{iaf}$). Table I outlines the qualitative rankings of the encoder and decoder methods given the aforementioned analysis and discussion.

Construction of retinomorphic systems should exist on a continuum where a multi-modal encoder/decoder approach can be taken to optimize composition with respect to costs and element performance. Such a relationship can be conceptualized

TABLE I: Module Rankings: (e,d) denote the ranking for encoders and decoders respectively. Decoder implementation in the analog domain was assumed ([10], [11], [13])

Module	Area	Power	Speed	Accuracy
IAF	1e	3e	1e	1e
SCON	2e	2e	2e	2e
TCON	3e	1e	3e	3e
Rate-Rect	3d	3d	2d	1d
Rate-Exp.	2d	1d	2d	1d
Temp-ISI	1d	2d	1d	2d

using the following ratios

$$\mathcal{R}_e = \frac{1}{C_e}\frac{\hat{e}_{iaf}}{\hat{e}_{tcon}} \tag{17}$$

$$\mathcal{R}_p = \frac{1}{C_p}\frac{P_{tcon}}{P_{iaf}} \tag{18}$$

where the C_e is the cost or weight of MSE given our system design and C_p is the same for the power constraint. These ratios capture disparity between system extremes: for values of $R_{e,p} << 1$, there is a massive gulf in between the performance of the pixels with respect to each metric. Therefore, in order to construct a balanced system with mixing coefficient $\alpha_r = 0.5$, then more cost will need to be associated with the alternative metric. The following expression captures this relationship between the above ratios and array composition

$$\alpha_r = min(\frac{\mathcal{R}_e}{\mathcal{R}_p}, 1) \tag{19}$$

When $\mathcal{R}_e > \mathcal{R}_p$ and power becomes the dominant metric, $\alpha_r = 1$ and the array will be fully composed of TCON pixels. An example 2D mapping of α_r with respect to constraint costs $C_{e,p}$ is shown in Figure 6. In this example, $\mathcal{R}_e = \frac{\mathcal{R}_p}{10}$, thus a large power cost (a power constrained system) is needed relative to the error cost. For instance, α_r=0.5 and the array composition is balanced with a cost relation of $C_p = 20C_e = 2$. However, with a system that requires higher error sensitivity and $C_e \to 2$, $\alpha_r \approx 0.1$ and the retinomorphic system is almost entirely IAF pixels despite their higher power consumption. The mixing ratio seen between transient, β and sustained, α ganglion cells is also depicted as a line inset within the figure. Interestingly, the retina ganglion cells are composed of 93% of these transient variety which confirms the fact that evolution also optimized the neural layer with respect to power costs instead of accuracy [7], [16]. Parallelized output and hierarchy seen in the latter stages of the visual pathway are then used to compensate for the lower accuracy of the retina to reconstruct the world around us.

VI. Conclusion

An overview and numerical analysis of retinomorphic channel elements and performance has been presented. Three encoding and decoding methodologies were introduced. The 9 unique combinations of these techniques were applied to transmit and reconstruct an example stimulus. Reconstruction accuracy in the presence of temporal jitter and spike dropout was used to compare performance. Power was approximated

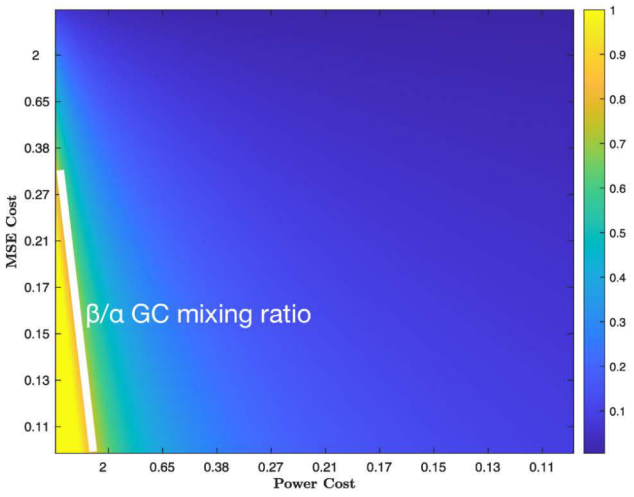


Fig. 6: Surface map of mixing coefficients for a variety of error and power costs: MSE and Power were extracted for the IAF and TCON streams using a ISI decoder and $i_{ph,max} = 10nA$. Encoder metric ratios: $R_e = 0.0037$, $R_p = 0.0395$. Channel error profile: $P_{drop} = 0$ and $\sigma_j = 1\mu s$. Inset line is the mixing ratio seen between β and α-ganglion cells in the retina.

by computing the product of energy consumption per spike and the pixel output frequency.

With respect to encoder performance, all methods were assumed to have comparable speed since they all leverage the IAF neuron for spike encoding. However, it was seen that the TCON pixel was superior with respect to power for the given stimulus, but performed poorly with respect to accurate reconstruction in the presence of extrinsic channel noise. With respect to encoder performance, the linear decoder methods were strictly better when reconstructing the signal in the presence of noise, but incurred a speed penalty as they require τ seconds to accumulate a result. Furthermore, ISI and Exp. decoder methods lend themselves to compact analog circuit implementation. Encoding power and accuracy metrics were then leveraged to understand how to architect an array of pixels given system constraints.

Future work is needed to expand the multi-modal analysis performed in Section V to decoder techniques. These can be analyzed with respect to speed and accuracy. Furthermore, additional stimuli should be used to understand architecture composition in alternate scenarios (e.g: mostly static/mostly dynamic scenes). By doing so, this framework can be utilized to inform the next generation of vision systems which can realize the effective communication of visual information seen in the retina.

ACKNOWLEDGMENT

This work was funded by the Defense Advanced Research Projects Agency (DARPA) Microsystems Technology Office (MTO) ReImagine project under Contract HR0011-17-C0071 and in part by the Northrop Grumman Faculty Award to support graduate student research (GT).

REFERENCES

[1] K. Boahen, "Retinomorphic vision systems," in *Proceedings of Fifth International Conference on Microelectronics for Neural Networks*. IEEE, 1996, pp. 2–14.

[2] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128× 128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[3] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-based vision sensors: bioinspired cameras with spiking output," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1470–1484, 2014.

[4] C. Brandli, R. Berner, M. Yang, S. Liu, and T. Delbruck, "A 240 × 180 130 db 3 µs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

[5] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, "5.10 a 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86µm pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 112–114.

[6] B. Son, Y. Suh, S. Kim, H. Jung, J.-S. Kim, C. Shin, K. Park, K. Lee, J. Park, J. Woo *et al.*, "4.1 a 640× 480 dynamic vision sensor with a 9μm pixel and 300meps address-event representation," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 2017, pp. 66–67.

[7] H. Kolb, E. Fernandez, and R. Nelson, "Webvision: the organization of the retina and visual system [internet]," 1995.

[8] M. A. Marwick and A. G. Andreou, "Retinomorphic system design in three dimensional soi-cmos," in *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2006, pp. 4–pp.

[9] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, L. Camuñas-Mesa, R. Berner, M. Rivas-Pérez, T. Delbruck *et al.*, "Caviar: A 45k neuron, 5m synapse, 12g connects/s aer hardware sensory–processing–learning–actuating system for high-speed visual object recognition and tracking," *IEEE Transactions on Neural networks*, vol. 20, no. 9, pp. 1417–1438, 2009.

[10] T. Serrano-Gotarredona, A. G. Andreou, and B. Linares-Barranco, "Aer image filtering architecture for vision-processing systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 46, no. 9, pp. 1064–1071, 1999.

[11] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.

[12] K. A. Zaghloul, *A silicon implementation of a novel model for retinal processing*. University of Pennsylvania, 2001.

[13] D. H. Goldberg and A. G. Andreou, "Spike communication of dynamic stimuli: rate decoding versus temporal decoding," *Neurocomputing*, vol. 58, pp. 101–107, 2004.

[14] R. Sarpeshkar, "Analog versus digital: extrapolating from electronics to neurobiology," *Neural computation*, vol. 10, no. 7, pp. 1601–1638, 1998.

[15] J. P. Sengupta, M. Villemur, P. O. Pouliquen, P. Julian, and A. G. Andreou, "Embedded processing pipeline exploration for neuromorphic event based perceptual systems," *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, p. 1–5, May 2022.

[16] H. Kolb, "How the Retina Works," *American Scientist*, vol. 91, pp. 1–8, Feb. 2003.