Sparse Coding Enables the Reconstruction of High-Fidelity Images and Video from Retinal Spike Trains

Yijing Watkins
Los Alamos National Laboratory
Los Alamos, USA
yazidoudou@gmail.com

Austin Thresher
New Mexico Consortium
Los Alamos, USA
athresher@newmexicoconsortium.org

David Mascarenas
Los Alamos National Laboratory
Los Alamos, USA
dmascarenas@lanl.gov

Garrett T. Kenyon
Los Alamos National Laboratory
New Mexico Consortium
Los Alamos, USA
gkenyon@lanl.gov

ICONS '18, July 23–26, 2018, Knoxville, TN, USA

ACM ISBN 978-1-4503-6544-4/18/07…$15.00
https://doi.org/10.1145/3229884.3229892

ABSTRACT

The optic nerve transmits visual information to the brain as trains of discrete events, a low-power, low-bandwidth communication channel also exploited by silicon retina cameras. Extracting high-fidelity visual input from retinal event trains is thus a key challenge for both computational neuroscience and neuromorphic engineering. Here, we investigate whether sparse coding can enable the reconstruction of high-fidelity images and video from retinal event trains. Our approach is analogous to compressive sensing, in which only a random subset of pixels are transmitted and the missing information is estimated via inference. We employed a variant of the Locally Competitive Algorithm to infer sparse representations from retinal event trains, using a dictionary of convolutional features optimized via stochastic gradient descent and trained in an unsupervised manner using a local Hebbian learning rule with momentum.

We used an anatomically realistic retinal model with stochastic graded release from cones and bipolar cells to encode thumbnail images as spike trains arising from ON and OFF retinal ganglion cells. The spikes from each model ganglion cell were summed over a 32 msec time window, yielding a noisy rate-coded image. Analogous to how the primary visual cortex is postulated to infer features from noisy spike trains arising from the optic nerve, we inferred a higher-fidelity sparse reconstruction from the noisy rate-coded image using a convolutional dictionary trained on the original CIFAR10 database.

To investigate whether a similar approach works on non-stochastic data, we demonstrate that the same procedure can be used to reconstruct high-frequency video from the asynchronous events arising from a silicon retina camera moving through a laboratory environment.

KEYWORDS

Sparse Coding; Compressive Sensing; Retinal Model; Silicon Retina; Spike Train; Event Train

1 INTRODUCTION

In the vertebrate central nervous system, spike trains play a fundamental role as the primary communication mechanism used to represent and transmit information. Understanding how biological neurons learn to infer high-fidelity image content from asynchronous spike trains is thus a fundamental goal of computational neuroscience research. The vertebrate retina provides a particularly accessible example of spike train encoding in the central nervous system. Incident light is focused onto a 2D array of photo-receptors where it is transduced into electrochemical signals via a process that can be formally described by realistic computational models [24]. Employing graded (non-spiking) stochastic release, bipolar cells relay the output of photoreceptor terminals to retinal ganglion cells (RGCs), whose axons comprise the optic nerve. The retina does not simply map light intensity into spikes, however. Rather, the retina consists of two major processing layers that implement a variety of spatiotemporal filtering operations and correspondingly there exist a variety of RGC subtypes which transmit different types of visual information [5, 21, 25]. By exploiting such heterogeneity, spikes arising from RGCs are thought to provide a rich source of information about the visual world to the brain. However, due to stochastic release at bipolar cell synapses and other sources of noise, RGC spike trains are often quite noisy, especially when interpreted as a rate code in which visual information is represented by the total number of spikes in a given interval [9]. On the other hand, it has also been shown that RGCs convey information about spatially and temporally complex stimuli in the relative timing of RGC spikes [9, 14, 23]. Here, we utilize a previously described retinal model that seeks to explain synchronous, stimulus-selective oscillations between RGCs [10, 12, 16]. In particular, we consider how simple cells in the primary visual cortex use learned dictionaries optimized for sparse coding [17] to infer high-fidelity image content from noisy retinal spike trains.

Sparse coding accounts for a variety of experimentally measured linear and non-linear response properties of V1 simple cells [30] and can be implemented in a biologically-plausible manner in terms of lateral inhibition [22]. Algorithmically, sparse coding employs an overcomplete set of non-orthogonal basis functions (feature vectors) to infer a sparse combination of non-zero activation coefficients that most accurately reconstruct each input image. Most relevant for the present study, sparse coding provides a powerful technique for image denoising [7] . Because sparse coding seeks to reconstruct a given image using a small number of previously learned features, added Gaussian or white noise tends to be ignored compared to the original image content. Given that images reconstructed from a rate-coded retinal spike trains are typically noisy, we hypothesized that sparse coding could be used to infer higher-fidelity images from noisy retinal spike trains.

But how to interpret the retinal code? Interpreted as a simple rate code, we can reconstruct noisy retinal images simply by summing all of the spikes from each RGC in a given time interval. Indeed, the initial results presented here assume a simple rate code for interpreting the spikes arising from model RGCs in response to static thumbnail images. But what about the above cited evidence that significant visual information is also encoded in the relative timing of RGC spikes? As there does not yet exist a consensus retinal model that accounts for all aspects of spike timing among RGCs in response to natural video and corresponding experimental data is difficult to obtain, we instead use event trains generated by a silicon retina camera. Analogous to the vertebrate retina, silicon retina cameras also represent and transmit information as trains of discrete events produced whenever the absolute value of the local pixel intensity changes by more than a threshold amount. Here, we use a biologically-plausible implementation of sparse coding to reconstruct visual stimuli from silicon retina event trains as a surrogate for the temporally structured spikes trains generated by RGCs in response to spatially and temporally complex stimuli [2, 18].

Silicon retinal cameras are also useful imaging devices that address some of the key limitations of conventional video cameras. The best state-of-the-art artificial vision devices still suffer from limitations imposed by their frame-based operation, in which visual information is acquired as a series of image frames recorded at a predetermined frame rate. However, things happen between frames and information gets lost. In fact, comparing the performance of biological vision systems to the best state-of-the-art camera, frames do not appear to be a very efficient or useful form of encoding visual information. The second drawback of frame-based visual information operation is redundancy. Each recorded frame conveys the information from all pixels, regardless of whether this information has not changed since the last frame had been acquired. Biological vision systems do not know the concept of a frame, they are controlled and driven by events happening within the scene. Inspired by biological vision systems, silicon retina cameras employ the frameless concept of biological vision to artificial imaging systems. Instead of recording the acquisition of visual information that controlled only limited an array of pixel regardless of whether no changed in the array, it transfers the decision making responsibility to the single pixel to handle its own information individually. Silicon retina only record the events when a small or large pixel

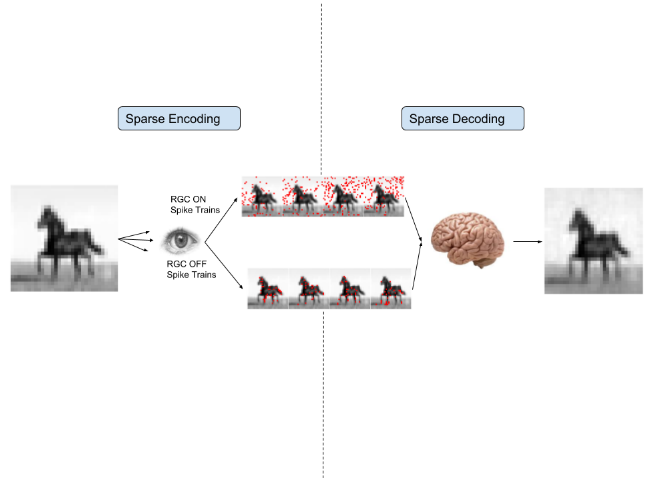


Figure 1: Illustration of the retinal decoding problem. Spikes arising from ON and OFF retinal ganglion cells (RGCs) must be converted back into an image.[11]

intensity has changed within the scene in view which are more efficient and consume less power than comparable digital systems [20]. Thus, there is strong motivation to develop techniques for inferring high-fidelity video from the discrete events generated by a silicon retina camera.

A recent approach to reconstructing high-fidelity images from retinal spike trains used a linear decoder based on measured RGCs kernels followed by a deep neural network to further enhance the image [19]. However, the reconstructed images lose many details compared with the original natural images. The present study is also related to the concept of compressive sensing. Given only a random or fixed subset of pixels as input, sparse coding can identify the minimal set of generators that explains the observed pixels and infer from them the missing pixel values [3, 6, 26, 29].

In this work we apply compressive sensing based sparse coding approaches to enable the reconstruction of high-fidelity images and video from retinal event trains. We employed a variant of the Locally Competitive Algorithm to infer sparse representations from retinal event trains, using a dictionary of convolutional features optimized via stochastic gradient descent and trained in an unsupervised manner using a local Hebbian learning rule with momentum. First, We report that a higher-fidelity sparse reconstruction is inferred from the noisy rate-coded image (The summed spikes from ganglion cell model over a 32 msec time window) by using a convolutional dictionary trained on the original CIFAR10 database. Additionally, we were able to estimate high frame rate video from a low-power, low-bandwidth silicon retina camera by training a dictionary of convolutional spatiotemporal features for simultaneously reconstructing differences of video frames (recorded at 22HZ and 5.56Hz) as well as discrete events generated by the silicon retina (binned at 484Hz and 278Hz).

2 EVENT TRAIN SOURCES

We used two sources of event-based of data streams generated in response to visual stimuli. The Retinal Ganglion Cell model produced spike events corresponding to both ON and OFF ganglion cells when fed images from the CIFAR10 database. The Silicon Retina model encodes video as two streams of discrete binary events, corresponding to supra-threshold increases and decreases in local pixel intensity, respectively.

2.1 Retinal Model

Descriptions of an earlier Matlab based retinal model for generating spiking activity from ON ganglion cells have been published previously [8, 11, 13]. The retinal model used here, implemented in PetaVision [1], added OFF ganglion cells and included gap junction coupling between the ON and OFF small amacrine cells (making them effectively bistratified) but was otherwise anatomically identical (an input file for executing the retinal model and which documents all model parameters is available at: [27]). The sparse reconstruction pipeline architecture is illustrated in Figure 2. Grayscale CIFAR10 images were fed to the cone layer, producing output consisting of ON and OFF ganglion cell spike trains. A sparse coding model (available here: [28]), also implemented in PetaVision, was used to infer the reconstruction of high-fidelity ON and OFF images from the sum of the noisy ON and OFF ganglion cell spike trains.

2.2 Silicon Retina

The silicon retina camera is a form of imaging technology inspired by biological vision [15]. It only measures and transmits event data when the value of a pixel's intensity changes beyond a predefined threshold. The resulting video resembles images run through an edge detection algorithm. This is because intensity changes tend to mostly occur at the edges. This sensitivity to object boundaries allows the silicon retina camera to capture very fast dynamic events with relatively small bandwidth.

The pipeline architecture for sparse reconstruction of events from a silicon retina camera is illustrated in Figure 4. The original silicon retina positive and negative spike events are first produced by a low-power, low-bandwidth silicon retina camera. A sparse coding model is then used to simultaneously reconstruct higher frame rate video from the original silicon retina positive and negative events by training a linked dictionary of convolutional spatiotemporal features for reconstructing silicon retina event trains and static features for reconstructing differences of video frames.

3 SPARSE CODING

Given an overcomplete basis, sparse coding algorithms seek to identify the minimal set of generators that most accurately reconstruct each input image. In neural terms, each neuron is a generator that adds its associated feature vector to the reconstructed image with an amplitude equal to its activation. For any particular input image, the optimal sparse representation is given by the vector of neural activations that minimizes both image reconstruction error and the number of neurons with non-zero activity. Formally, finding a sparse representation involves finding a minimum of the following cost function:

$$E(\overrightarrow{I}, \phi, \overrightarrow{a}) = \min_{\{\overrightarrow{a}, \phi\}} \left[\frac{1}{2} ||\overrightarrow{I} - \phi * \overrightarrow{a}||^2 + \lambda ||\overrightarrow{a}||_1 \right], \qquad (1)$$

In Eq. (1), $\overrightarrow{I}$ is an image unrolled into a vector, and ϕ is a dictionary of feature kernels that are convolved with the sparse representation $\overrightarrow{a}$. The factor λ is a tradeoff parameter; larger λ values encourage greater sparsity (fewer non-zero coefficients) at the cost of greater reconstruction error. Both the sparse representation $\overrightarrow{a}$ and the dictionary of feature kernels ϕ can be determined by a variety of standard optimization methods.

Our approach to compute a sparse representation for a given input image is based on a convolutional generalization of a rectifying Locally Competitive Algorithm (LCA) [4]. Once a sparse representation for a given input image has been found, the basis elements associated with non-zero activation coefficients are adapted according to a local Hebbian learning rule (with a momentum term for faster convergence) that further reduces the remaining reconstruction error. Starting with random basis elements, dictionary learning was performed via Stochastic Gradient Descent (SGD). This training procedure can learn to factor a complex, high-dimensional natural image into an overcomplete set of basis vectors that capture the high-dimensional correlations in the data.

4 APPROACH

In this section we apply compressive sensing based sparse coding approaches to enable the reconstruction of higher fidelity images from model retinal spike trains.

4.1 Sparse Decoder

Our spike train decoder is based on combining sparse coding with ideas from compressive sensing [3, 6].

The architecture is illustrated in Figure 3, where the input layer (top left) is the sum of ON and OFF spike trains. The Reconstructed Image (bottom left) is generated by a sum of feature vectors drawn from an over-complete, non-orthogonal dictionary (bottom right) weighted by a sparse set of non-zero activation coefficients corresponding to the activation of neurons in the V1 layer (top right). V1 activity is driven by the difference (Residual) between the noisy input image (Sum of Spikes) and the Reconstructed Image convolved with the transpose of each feature vector, the results of which drive changes in the values of activation coefficients, resulting in a new Reconstructed Image, a new residual error, and so on.

This iterative process is guaranteed to settle into a local minima of a characteristic energy function that minimizes the least-squares residual reconstruction error while simultaneously minimizing non-zero activation coefficients (basis elements) used in the reconstruction. Once a sparse representation for a given noisy input image has been found, the basis elements associated with non-zero activation coefficients are adapted according to a local Hebbian learning rule (usually modified with a momentum term) that further reduces the remaining reconstruction error relative to the original image. Starting with random basis elements, the above procedure can learn to factor a complex, high-dimensional input stream, such as natural image patches, into an overcomplete set of basis vectors that capture the high-dimensional correlations in the

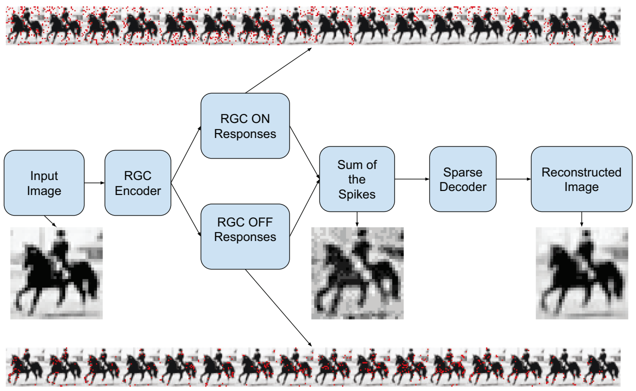


Figure 2: Sparse Reconstruction Pipeline.

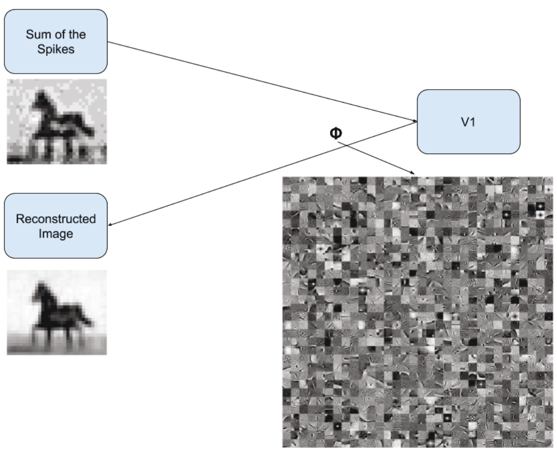


Figure 3: Sparse retinal ganglion cell spike train decoder. We use a convolutional dictionary to infer sparse representations of images constructed the difference of sums of ON and OFF ganglion cell spike trains. The dictionary was trained to minimize the difference the sparse reconstructions of Sum of Spikes input and the original CIFAR10 gray-scale image, in order to promote higher-fidelity sparse reconstructions from the noisy rate-coded images [29] (The summed spikes from ganglion cell model over a 32 msec time window).

data. Here, sparse coding of retinal spikes trains produced higher fidelity reconstructed images.

4.2 Silicon Retina Decoder

The architecture is illustrated in Figure 5, where the input layers (left) are the silicon retina positive and negative event streams. The Residual is the error between the original and the reconstructions of the silicon retina positive and negative event streams. The Reconstructed Images are sparse reconstructed silicon retina positive

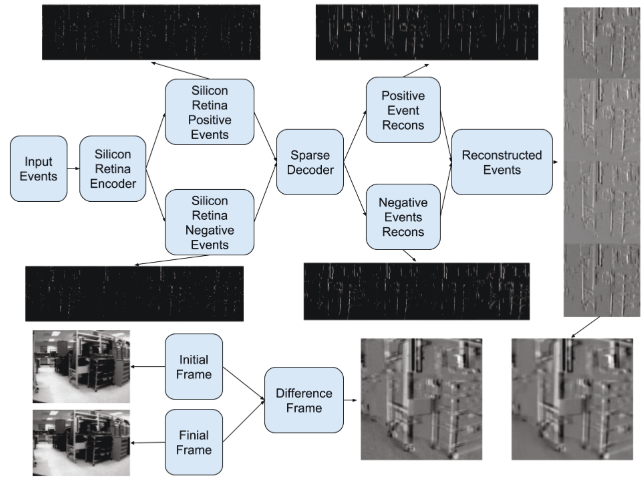


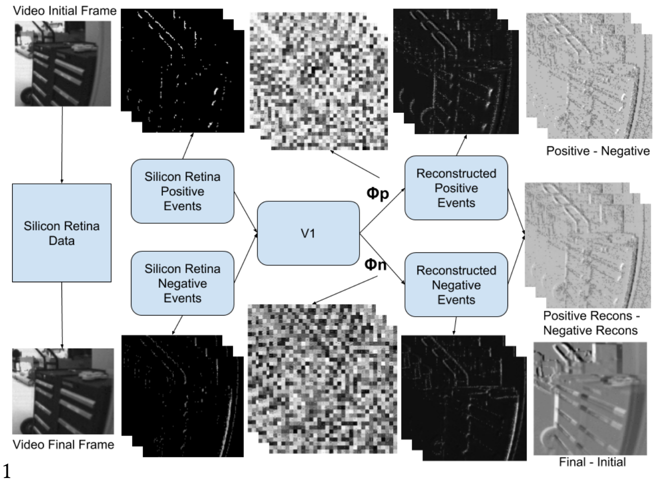
Figure 4: Pipeline for sparse reconstruction of events from a silicon retina camera. Positive and negative events are produced by a low-power, low-bandwidth silicon retina camera. A sparse decoder is used to simultaneously reconstruct higher frame rate video from the silicon retina positive and negative events by training a linked dictionary of convolutional spatiotemporal features and static dictionary based on differences of video frames.

and negative event trains, obtained from sparse linear combinations of spatiotemporal features drawn from an over-complete, non-orthogonal dictionary. Spatiotemporal feature vectors are weighted using a sparse set of non-zero activation coefficients corresponding to the activity of neurons in the V1 layer.

We were able to estimate high frame rate video from a low-power, low-bandwidth silicon retina camera by training a dictionary of convolutional spatiotemporal features for simultaneously reconstructing differences of video frames (recorded at 22HZ and 5.56Hz) as well as discrete events generated by the silicon retina (binned at 484Hz and 278Hz).

5 CONCLUSION

In this work we apply sparse coding to enable the reconstruction of high-fidelity images and video from retinal event trains. We employed a variant of the Locally Competitive Algorithm to infer sparse representations from retinal event trains, using a dictionary of convolutional features optimized via stochastic gradient descent and trained in an unsupervised manner using a local Hebbian learning rule with momentum. We report that a higher-fidelity sparse reconstruction is inferred from the noisy rate-coded image (The summed spikes from ganglion cell model over a 32 msec time window) by using a convolutional dictionary trained so as to better approximate the original CIFAR10 images given the sparse representation of the noisy rate coded images. Additionally, we were able to estimate high frame rate video from a low-power, low-bandwidth silicon retina camera by training a dictionary of convolutional spatiotemporal features for simultaneously reconstructing differences



1

Figure 5: Sparse Silicon Retina Decoder. We use a convolutional spatiotemporal dictionary trained on positive and negative events generated by a silicon retina camera linked with a static convolutional dictionary trained on differences of conventional video frames to reconstruct high frame rate (484Hz) sparse video from the original video (22Hz).

of video frames (recorded at 22HZ and 5.56Hz) as well as discrete events generated by the silicon retina (binned at 484Hz and 278Hz).

REFERENCES

[1] [n. d.]. Petavision. ([n. d.]). https://petavision.github.io/ Software available from petavision.github.io.

[2] Oweiss K. Aghagolzadeh M. 2009. Compressed and Distributed Sensing of Neuronal Activity for Real Time Spike Train Decoding. *IEEE transactions on neural systems and rehabilitation engineering* 17, 2 (2009), 116–127.

[3] E. J. Candes, J. Romberg, and T. Tao. 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* 52, 2 (Feb 2006), 489–509. https://doi.org/10.1109/TIT.2005.862083

[4] Richard G. Baraniuk Christopher J. Rozell, Don H. Johnson and Bruno A. Olshausen. 2008. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation* 20, 10 (2008), 2526–2563.

[5] Pamela Reinagel David K. Warland and Markus Meister. Journal of neurophysiology. Decoding visual information from a population of retinal ganglion cells. *Journal of neurophysiology* 78, 5 (Journal of neurophysiology), 2336âĂŞ2350.

[6] D. L. Donoho. 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52, 4 (April 2006), 1289–1306. https://doi.org/10.1109/TIT.2006.871582

[7] M. Elad and M. Aharon. 2006. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Transactions on Image Processing* 15, 12 (Dec 2006), 3736–3745. https://doi.org/10.1109/TIP.2006.881969

[8] John S. George Bryan J. Travis Garrett T. Kenyon, James Theiler and David W. Marshak. 2004. Correlated Firing Improves Stimulus Discrimination in a Retinal Model. *Neural Computation* 16 11 (2004), 2261–2291.

[9] Michael J. Berry, David Warland, and Markus Meister. 1997. The structure and precision of retinal spike trains. 94 (06 1997), 5411–6.

[10] GARRETT T. KENYON, BARTLETT MOORE, JANELLE JEFFS, KATE S. DENNING, GREG J. STEPHENS, BRYAN J. TRAVIS, JOHN S. GEORGE, JAMES THEILER, and DAVID W. MARSHAK. 2003. A model of high-frequency oscillatory potentials in retinal ganglion cells. *Visual Neuroscience* 20, 5 (2003), 465âĂŞ480. https://doi.org/10.1017/S0952523803205010

[11] Garrett T. Kenyon, Bartlett D Moore, J. A. R. Jeffs, Kate Denning, Greg J. Stephens, Bryan J. Travis, John S. George, James Theiler, and David W. Marshak. 2003. A model of high-frequency oscillatory potentials in retinal ganglion cells. *Visual neuroscience* 20 5 (2003), 465–80.

[12] G. T. Kenyon, B. J. Travis, J. Theiler, J. S. George, G. J. Stephens, and D. W. Marshak. 2004. Stimulus-specific oscillations in a retinal model. *IEEE Transactions on Neural Networks* 15, 5 (Sept 2004), 1083–1091. https://doi.org/10.1109/TNN.2004.832722

[13] G. T. Kenyon, B. J. Travis, J. Theiler, J. S. George, G. J. Stephens, and D. W. Marshak. 2004. Stimulus-specific oscillations in a retinal model. *IEEE Transactions on Neural Networks* 15, 5 (Sept 2004), 1083–1091. https://doi.org/10.1109/TNN.2004.832722

[14] Vaingankar V Wei Y Wang Q Rathbun DL Usrey WM Hirsch JA Sommer FT Koepsell K1, Wang X. 2009. Retinal oscillations carry visual information to cortex. (4 2009), 3–4.

[15] Delbruck T Lichtsteiner P., Posch C. 2008. A 128 X 128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid State Circuits* 43, 2 (2008), 566–575.

[16] George JS Marshak DW Kenyon GT. Miller JA1, Denning KS. 2006. A high frequency resonance in the responses of retinal ganglion cells to rapidly modulated stimuli: A computer model. 23, 5 (September 2006), 779–794.

[17] Bruno A. Olshausen and David J. Field. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. 381 (6 1996), 607 EP.

[18] I. M. Park, S. Seth, A. R. C. Paiva, L. Li, and J. C. Principe. 2013. Kernel Methods on Spike Train Space for Neuroscience: A Tutorial. *IEEE Signal Processing Magazine* 30, 4 (July 2013), 149–160. https://doi.org/10.1109/MSP.2013.2251072

[19] Nikhil Parthasarathy, Eleanor Batty, William Falcon, Thomas Rutten, Mohit Rajpal, E.J. Chichilnisky, and Liam Paninski. 2017. Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 6434–6445.

[20] Christoph Posch, Teresa Serrano-Gotarredona, Bernabé Linares-Barranco, and Tobi Delbrück. 2014. Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output. *Proc. IEEE* 102, 10 (2014), 1470–1484. https://doi.org/10.1109/JPROC.2014.2346153

[21] Fred Rieke, Davd Warland, Rob de Ruyter van Steveninck, and William Bialek. 1999. *Spikes: Exploring the Neural Code.* MIT Press, Cambridge, MA, USA.

[22] Christopher J. Rozell, Don H. Johnson, Richard G. Baraniuk, and Bruno A. Olshausen. 2008. Sparse Coding via Thresholding and Local Competition in Neural Circuits. *Neural Computation* 20, 10 (2008), 2526–2563. https://doi.org/10.1162/neco.2008.03-07-486 arXiv:https://doi.org/10.1162/neco.2008.03-07-486 PMID: 18439138.

[23] Greg Stephens, Sergio Neuenschwander, John George, Wolf Singer, and Garrett T Kenyon. 2006. See globally, spike locally: Oscillations in a retinal model encode large visual features. 95 (11 2006), 327–48.

[24] Hans van Hateren. 2005. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. 5 (02 2005), 331–47.

[25] David K. Warland, Pamela Reinagel, and Markus Meister. 1997. Decoding Visual Information From a Population of Retinal Ganglion Cells. *Journal of Neurophysiology* 78, 5 (1997), 2336–2350. https://doi.org/10.1152/jn.1997.78.5.2336

[26] Adam Watkins, Venkata Naresh Mudhireddy, Haibo Wang, and Spyros Tragoudas. 2014. Adaptive Compressive Sensing for Low Power Wireless Sensors. In *Proceedings of the 24th Edition of the Great Lakes Symposium on VLSI (GLSVLSI '14)*. ACM, New York, NY, USA, 99–104. https://doi.org/10.1145/2591513.2591537

[27] Yijing Z. Watkins. 2018. Ganglion on and off execution. https://github.com/yazidoudou18/Sparse-Retina-Model/blob/master/gan_pvp.lua. (2018).

[28] Yijing Z. Watkins. 2018. Retina Simulation. https://github.com/yazidoudou18/Sparse-Retina-Model/blob/master/Gan.lua. (2018).

[29] O. Iaroshenko Y. Watkins, M. Sayeh and G. Kenyon. 2018. Image Compression: Sparse Coding vs. Bottleneck Autoencoders. In *Proceedings of 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)* (8-10) *(SSIAI '18)*.

[30] Mengchen Zhu and Christopher J. Rozell. 2013. Visual Nonclassical Receptive Field Effects Emerge from Sparse Coding in a Dynamical System. *PLOS Computational Biology* 9, 8 (08 2013), 1–15. https://doi.org/10.1371/journal.pcbi.1003191