

## PROJECT SUMMARY

### OVERVIEW

The efficient coding hypothesis, which posits that sensory systems maximize the mutual information between natural stimuli and neural activity subject to energetic and other biological constraints, is one of the most successful predictive principles in neuroscience, particularly in early vision. In addition to offering explanations for the shape and spatial organization of retinal receptive fields, recent work by the PIs has demonstrated its ability to explain key aspects of inter-mosaic coordination, including the proliferation of functionally distinct retinal ganglion cell types as the number of neurons available for encoding is increased. Nonetheless, early efficient coding models required numerous simplifying assumptions for analytical tractability, and even modern computational models have limited flexibility. This proposal seeks to extend efficient coding models previously developed by the PIs in three key directions: 1) inclusion of multiple chromatic channels to mimic different photoreceptor types; 2) inclusion of structured noise in both retinal ganglion cell inputs and outputs; 3) multi-layer cascade models capable of capturing successive transformations of sensory information. For each extension, model predictions will be tested against (and refined by) data from *ex vivo* multielectrode array recordings of the retina across multiple species, pushing the boundaries of efficient coding as a hypothesis for the organization of early vision.

### INTELLECTUAL MERIT

Despite a century of theoretical effort, first-principle explanations remain a rarity in neuroscience. While efficient coding has been incredibly successful, optimization-based models of efficient coding address only a limited range of phenomena. As a result, the theory has not nearly been pushed to its limits. By developing a more flexible modeling framework for efficient coding and testing its predictions against retinal data across multiple species, we aim to answer fundamental questions like: Why do retinal ganglion cells come in dozens of types, each with its own mosaic? What should each type encode, how should their receptive fields be organized, and how does this change as the noise in the system is varied? Why are some retinal ganglion cell types linear (X-like) and some nonlinear (Y-like)? Can efficient coding theory predict more than just center-surround receptive fields? Answering these questions based on simple requirements of information processing as shaped by biological constraints would represent a decisive step forward in our understanding of the retina. More importantly delineating cases where efficient coding fails will identify aspects of sensory processing where new ideas are needed to bring theory in line with experiments.

### BROADER IMPACTS

While this proposal centers on efficient coding in the context of vision, the models we propose are data-agnostic and could readily be applied to other sensory encoding problems. These models, along with unique natural movie data that we will collect, will help disseminate the methods we develop to the broader vision and machine learning fields. More generally, because the retina is the most sophisticated visual sensor in existence and yet uses a fraction of the power of a cell-phone camera, a better understanding of its organization and information processing properties can inform the design of new, lower-power sensors and computer vision systems. Finally, by providing cross-training in both retinal physiology and theoretical modeling to members of the team and incorporating the resulting scientific insights into online teaching materials, this project will help to train the next generation of neuroscientists working at the theory-experiment interface.

## INTELLECTUAL MERIT

### INTRODUCTION

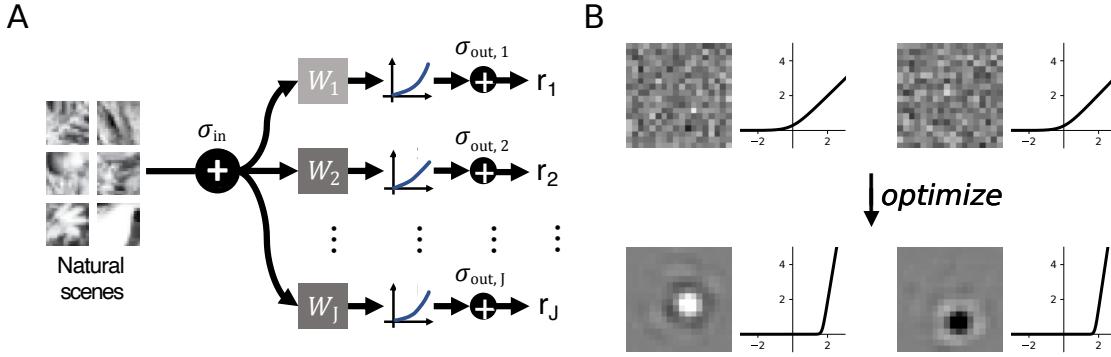
The fundamental challenge facing theoretical neuroscience is the lack of a concise, parsimonious set of principles from which to make predictions about brain structure and function. Indeed, the current project of the field can be described as a search for such principles [1]. In particular, these guiding ideas should be effective proxies for the results produced by a long process of evolution, formulations capable of producing quantitative predictions that can be tested against experiments.

Perhaps the most successful such candidate, particularly in early sensory systems, is efficient coding (EC) theory [2]. EC, along with closely related ideas like predictive coding [3, 4] and the information bottleneck [5–7], simply posits that nervous systems are optimized to make efficient use of incoming information. More specifically, EC asserts that sensory systems seek to maximize the mutual information between incoming information and neural responses.

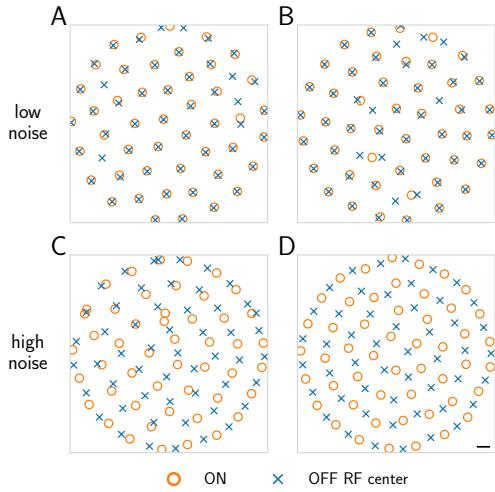
This simple principle has been remarkably successful in explaining many features of early sensory processing, including coding properties of the auditory nerve [8–10] and tuning of the vestibular system to natural stimuli [11–13]. Yet the most stunning success of this idea has been in vision, where it accounts for the structure of center-surround receptive fields [14, 15], color coding [16–18], predictive encoding of motion [6], diversity in sensory populations [19–22], and retinal mosaics [22, 23]. Nevertheless, **most theoretical studies of EC in vision have been limited to highly simplified models that can be solved in closed form or simulated inexpensively**. This limitation stems from the difficulty of estimating and optimizing mutual information [24], which typically requires idealized stimulus sets and highly simplified encoding models for tractability. As a result, current EC models of retinal processing yield relatively few cell types with quite simple receptive field structures (e.g., pseudolinear, center-surround, and space-time separable). This deviates significantly from what is observed in the mammalian retina, which contains ~ 45 cell types at the output layer with complex and nonlinear receptive field structures (e.g., subunit rectification, direction selectivity, looming sensitivity, and edge detection).

We hypothesize that EC would be capable of predicting much more about the organization of the retina if it were used to optimize neural network architectures that more accurately resembled the retina — for example, architectures containing hidden layers to produce subunit rectification and models that allow receptive fields to be unconstrained in space-time. The advent of deep learning models and their attendant advances in large scale optimization present an opportunity to test this hypothesis. However, recent efforts at applying deep learning to models of the retina have taken a *data-driven* rather than *theory-driven* approach. These data-driven approaches have been able to capture many features of retinal processing [25–29], but they do not embody predictions from first-principles like EC. Recent work from our group and others has used EC to formulate relatively simple optimization models, but these require careful training, and unrealistic constraints (e.g., space-time separable receptive fields) that limit their ability to predict important aspects of retinal function [21–23, 30]. Thus, testing first-principles like EC **presents a critical need for flexible and scalable models that can generate predictions in more complex scenarios**. In particular, these models should accommodate multiple information channels, multiple layers, and multiple diverse sources of noise. Building such models and testing their predictions against experimental data is the central aim of this proposal.

### PREVIOUS WORK



**Figure 1: Model concept and training.** **A.** Model schematic. After pixelwise noise  $\sigma_{\text{in}}$  is added to a natural image patch, the result is filtered by a linear filter  $w_j$ , passed through a nonlinearity, and combined with output noise  $\sigma_{\text{out}}$  to produce a firing rate  $r_j$ . **B.** Receptive fields are initialized as random filters but assume characteristic center-surround shapes after training. Parameters of the nonlinearities are likewise learned. Figure adapted from [30].



**Figure 2: Efficient coding models predict mosaic coordination.** Optimized receptive field center locations for ON and OFF cells under four different noise regimes. Optimizing the model in Figure 1 with low noise generally produces aligned ON and OFF mosaics (**A, B**), while optimizing with higher noise levels results in anti-alignment (**C, D**). Figure adapted from [30, 31].

Previous work from our group in modeling EC in the retina built on the optimization framework of [23] (Figure 1). We modeled RGCs, which form the output layer of the retina, as rate-based linear-nonlinear (LN) units: images  $\mathbf{x}$  were subject to neuron-specific linear filters  $\mathbf{w}_j$ , the results of which were fed through a nonlinear function:

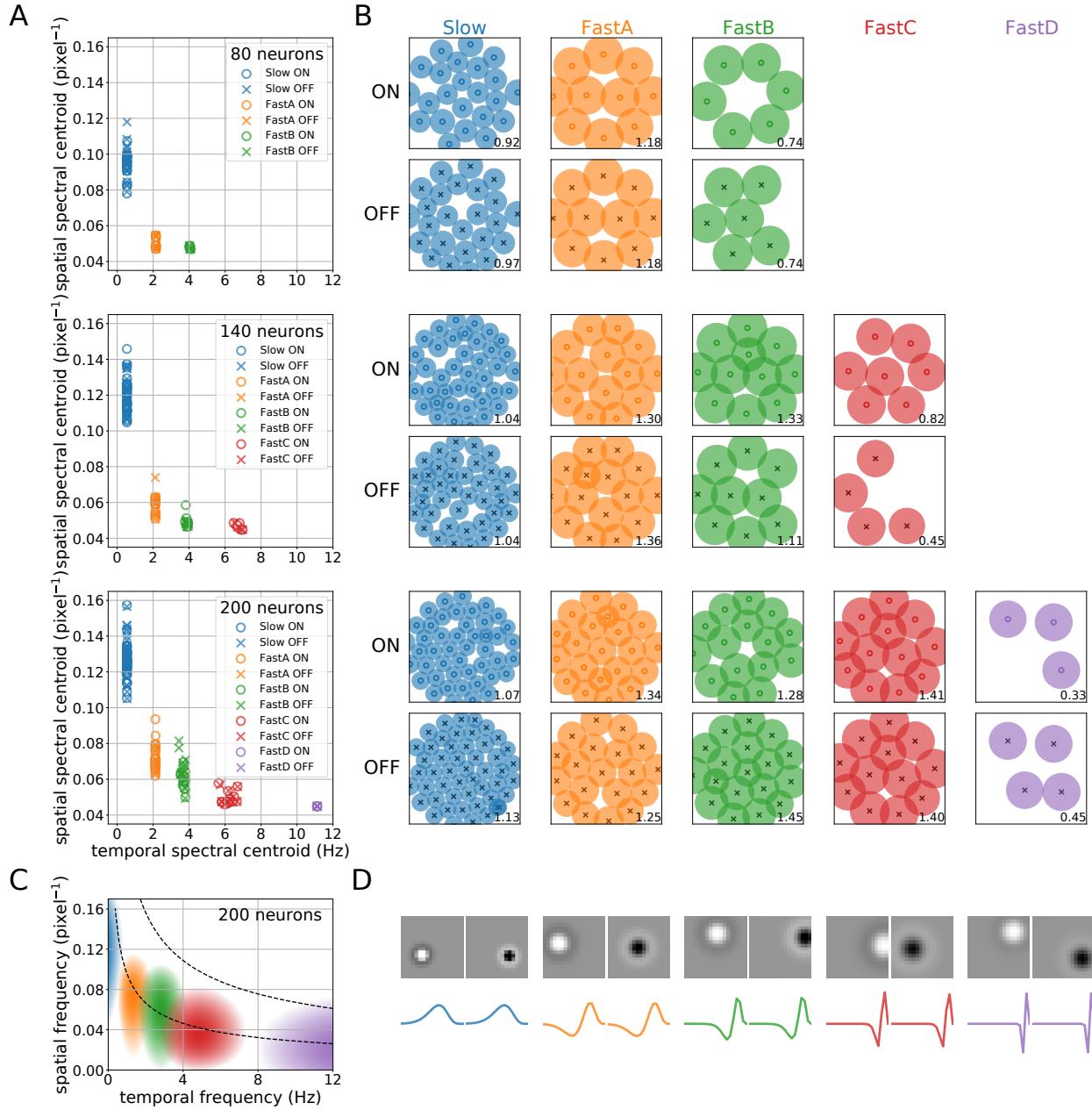
$$r_j(\mathbf{x}) = \gamma_j \cdot \eta(\mathbf{w}^\top \cdot (\mathbf{x} + \mathbf{n}_x) - \theta_j) + n_{r,j}, \quad (1)$$

where  $\gamma_j$  and  $\theta_j$  are the gain and threshold of unit  $j$ ,  $\mathbf{n}_x$  is a Gaussian pixelwise input noise,  $n_{r,j}$  is Gaussian firing rate noise, and  $\eta(\cdot)$  is the output nonlinearity (in our case, the softplus function  $\eta(y) = \log(1 + e^{\beta y})/\beta$ ). As in [23], this model was trained by maximizing a Gaussian approximation to the mutual information between images  $\mathbf{x}$  and output firing rates  $\mathbf{r}$ , subject to the constraint that mean firing rates across all stimuli are constant ( $\mathbb{E}r_j = 1$ ), a proxy for the energetic costs of spiking. More specifically, we solved the optimization problem

$$\max_{\mathbf{W}, \gamma, \theta} \mathbb{E}_{\mathbf{x}} \left[ \log \frac{\det(\mathbf{G}(\mathbf{x})\mathbf{W}^\top(\mathbf{C}_x + \mathbf{C}_{\text{in}})\mathbf{WG}(\mathbf{x}) + \mathbf{C}_{\text{out}})}{\det(\mathbf{G}(\mathbf{x})\mathbf{W}^\top\mathbf{C}_{\text{in}}\mathbf{WG}(\mathbf{x}) + \mathbf{C}_{\text{out}})} \right] \quad \text{subject to } \mathbb{E}_{\mathbf{x}} r_j(\mathbf{x}) = 1, \quad (2)$$

where  $\mathbb{E}_{\mathbf{x}}$  is an average over natural images, filters for each neuron are the normalized columns of  $\mathbf{W}$  ( $\|\mathbf{w}_j\| = 1$ ),  $\mathbf{G}$  is a (image-dependent, diagonal) matrix of local gains, and  $\mathbf{C}_x$ ,  $\mathbf{C}_{\text{in}}$ , and  $\mathbf{C}_{\text{out}}$  are the covariance matrices of natural images, input noise, and output noise, respectively.

We found that, as in [23] optimizing (2) reliably produced mosaics of RGCs with ON and OFF center-surround receptive fields when trained on natural images [30]. Moreover, as levels of noise varied, the relative alignments of these mosaics shifted in a phase transition: at low noise levels, ON and OFF mosaics were aligned, whereas at high noise levels, mosaics were anti-aligned (Figure 2, [30]). This accorded with previous results from our collaboration, which established that such anti-alignment between mosaics indeed occurs in both rodent and primate retinas [31]. Importantly, our model did not *presuppose* a mosaic configuration: receptive fields were free to move or change shape, and this flexibility was critical in comparing with experimental data.



**Figure 3: Distinct cell types emerge as the number of available neurons varies.** **A.** Spatial versus temporal center frequencies of learned filters as a function of number of neurons in the model (successive rows). **B.** Mosaic allocations of the learned linear filters. New mosaics only begin to emerge when previous mosaics are “filled.” **C.** Power spectral density of a typical filter for each mosaic for  $J = 200$  neurons. As new mosaics emerge, previous mosaics further specialize in temporal frequency. **D.** Temporal and spatial filters for the learned mosaics. As seen in **C**, higher spatial frequencies are associated with lower temporal frequencies and vice-versa.

We subsequently extended this model to consider the encoding of natural movies (instead of just natural images), with each RGC's receptive field (RF) formed by the outer product of a spatial and temporal filter. Similar to previous work that had *assumed* a particular mosaic structure [21], we found that mosaics of ON and OFF RFs with distinct spatial and temporal frequencies naturally emerged from EC [22] (**Figure 3**). More specifically, as we varied the number of RGCs available for encoding a region of space, new mosaics with larger RFs and higher-frequency temporal preferences continued to emerge, indicating that considerations of coding capacity can help explain the multiplicity of functional cell types within the retina. More importantly, the pairing of large RFs with high temporal frequencies (differentiating behavior) and small RFs with low temporal frequencies (integrating behavior) accords with known physiological results in primate retina [32, 33]. Here again, our use of a model that did not, *a priori* assume a fixed number of mosaics or RF spacing was critical in establishing this result, highlighting the importance of a *flexible, optimizable framework* for generating EC predictions.

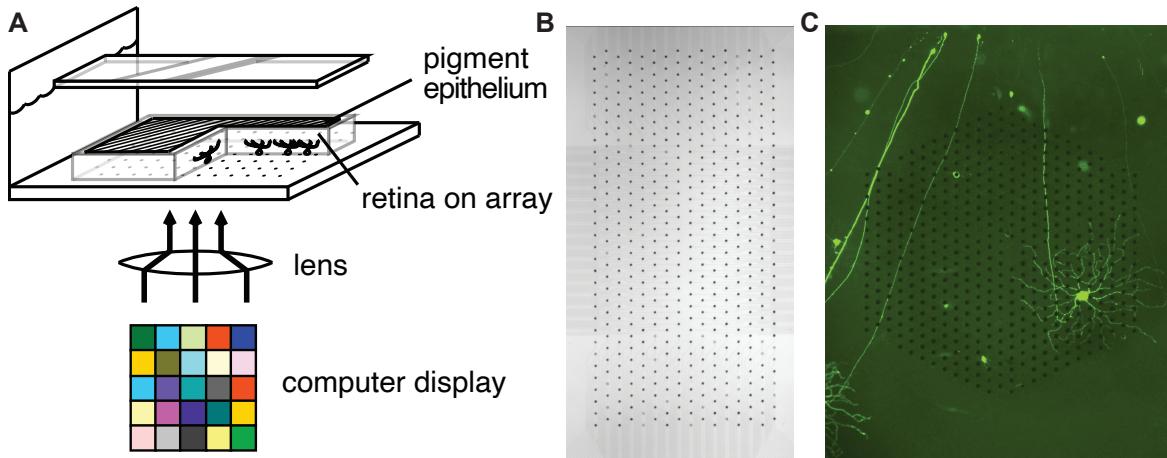
## PROPOSED RESEARCH

As reviewed above, the class of flexible EC models we have previously used is capable of generating predictions for both natural images and natural movies, replicating key facets of functional organization in the retina. However, these models still fall far short in capturing other critical aspects of visual processing in the retina. In particular: 1) these models deal only with single-channel (monochrome) images; 2) they assume noise that is independent and uncorrelated across both space and time, as well as across RGCs; 3) they are single-layer models, making predictions only about the output (RGC) layer of the retina.

Our goals in this proposal are to address each of the limitations listed above, producing a more flexible class of models that can be used to investigate the predictions of EC for retinal encoding across a variety of stimulus sets and neural circuit architectures. We will use prior knowledge of the retina (e.g., the spectral sensitivity of the cones) and new measurements we make (i.e., correlated noise across diverse populations of RGCs) to constrain these optimizations. Furthermore, we will test predictions from our optimized models experimentally by using large-scale electrode arrays [31] to measure the receptive field structures and mosaic organizations of cell types not previously examined within an EC framework, such as direction-selective RGCs and looming-sensitive RGCs. Thus, we will take advantage of a tight feedback loop between experiments, theoretical modeling, and predictions about subsequent experiments to refine our understanding of EC theory. The overarching objective is to understand precisely the limits of EC in capturing retinal function: if the mammalian retina contains  $\sim$ 40 RGC types, how many of these types are predicted by EC, what is the accuracy of these predictions, and which cell types are not predicted by EC? This knowledge would greatly advance our understanding of EC and its limits, as well as point the way toward additional objective functions (i.e., reaction time) must be added to EC as constraints to understand and model retinal processing.

## Experimental methods

**Multielectrode array experiments:** EC optimizations will be compared to measurements of spatiotemporal receptive field structures made using a large-scale multielectrode array (MEA) system in the Field Lab [31, 34, 35]. Animals are euthanized and retinas isolated and placed on the array (**Figure 4**). This system allows for measuring the activity of hundred of RGCs simultaneously, providing state-of-the-art information about the organization of retinal signaling in different species [31, 36]. We will utilize primate, squirrel and rat retinas. The rationale for primate (macaque) is they will be useful for **Extensions 1, 2, and 3** (see below). Primate retinas will be harvested from animals at Duke University from Dr. Marc Sommer. (The Field lab retains an MEA system at



**Figure 4: Schematic of multielectrode array measurements from retina.** **A.** A sample of isolated retina is placed RGC side down on electrode array. Pigment epithelium is kept attached to maintain high-sensitivity photoreceptor responses. Image from computer display (DLP lightcrafter, Texas Instruments) is focused onto the photoreceptors through the mostly transparent electrode array. Retina is perfused in oxygenated Ames media and can be kept alive and responsive for 10-20 hours post mortem. **B.** Photograph of one array design with 60  $\mu\text{m}$  pitch between electrodes; whole array spans  $\sim 1 \times 2 \text{ mm}$ . **C.** Second array geometry with 30  $\mu\text{m}$  electrode pitch, with living retina over the array. Retina contains a sparse set of RGCs expressing green fluorescent protein, with one fully filled fluorescent cell over the array for sense of scale.

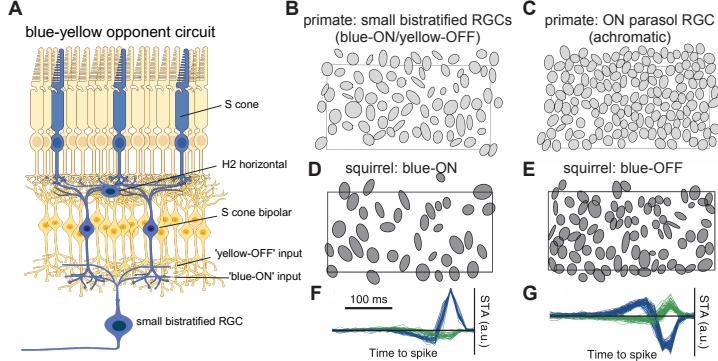
Duke to support an unrelated, NIH/NEI-funded collaboration with Dr. Marc Sommer that utilizes macaque retinas.) The rationale for squirrel retinas is that they will be particularly useful for **Extension 1** (see below). The rationale for rat retinas is that they will be useful for **Extensions 2 and 3** (see below).

**Collection of Natural Movies:** To collect natural multi-spectral movies that capture the spatial and temporal activations of different photoreceptor types in natural environments, we will expand on a previously developed camera system specifically developed for the mouse retina [37]. We will begin with the simplest system: 2 aligned cameras. Bandpass spectral filters will be placed in front of the cameras to mimic the spectral sensitivities of short and middle wavelength cones in squirrel and primate retinas. We will collect movies from the forest surrounding Duke University and the Santa Monica mountain range north of UCLA. Once this system is operational, we will build extensions that include three and eventually four cameras to generate natural movies that can be used in **Extension 1**.

#### Extension 1: Multi-channel inputs (Years 1–2)

**Rationale:** While our previous models used only a single (luminance) channel, many species, including primates, have two or more photoreceptor types with differing spectral sensitivities that give rise to color vision. The most common arrangement in mammals is dichromatic with a sparse array ( $\sim 5\text{--}10\%$ ) of cones that express a short (S) wavelength sensitive opsin, while the rest of the cones express a middle (M) wavelength sensitive opsin. Correspondingly there is a conserved ‘blue-yellow’ circuit in the mammalian retina (Figure 5) that begins with a retinal bipolar cell type that synapses exclusively with the S cones. This S-cone bipolar cell is an ‘ON’ bipolar cell (it responds to increment of light intensity) that provides excitatory input to a specialized, chromatically opponent RGC, called the small bistratified RGC. This RGC type, as the name implies, has a bistratified dendritic field, with one tier receiving input from the S-cone bipolar cell, and the other tier receiving input from OFF bipolar cells that carry M cone signals (and ‘L’ cone signals in primates). Thus, these small bistratified RGCs are chromatically opponent, responding to increments of ‘blue’ light and decrements of ‘yellow’ light. There is also a specialized horizontal cell

in this pathway that creates an M-cone mediated surround for the S-cones. Thus, this blue-yellow opponency arises at the first synaptic layer in the mammalian retina. We begin our investigation of multi-channel EC optimizations with this pathway because it is so well conserved across species. Focusing on the two channel case will also facilitate comparisons between EC predictions and the measured receptive fields as well as mosaic properties of blue-yellow opponent RGCs.



**Figure 5: Blue-Yellow opponent circuits in the mammalian retina.** A. Illustration of blue-yellow circuit. Short (S) wavelength-sensitive cone inputs (top) are collected by S (ON) cone bipolar cells (middle). These bipolar cells provide input to one (lower) layer of dendrites in the small bistratified cell (bottom). The second (upper) layer of dendrites receives OFF bipolar cell input from longer wavelength sensitive cones. This feedforward circuit generates a ‘blue-ON/yellow-OFF’ receptive field in the small bistratified cell that is enhanced by lateral inhibition from the H2 horizontal cell. B. Mosaic of blue-ON/yellow-OFF receptive fields from primate retina. Rectangle show outline of electrode array from Figure 4B and each ellipse shows the 1-standard deviation contour of a two-dimensional Gaussian fit to the spatial receptive field. Note the receptive fields approximately tile space. Large gaps are most likely cells that were not recorded on the electrode array. C. Same as B, but receptive field mosaic of simultaneously measured ON parasol RGCs over the electrode array. D. Mosaic of blue-ON/yellow-OFF RGCs from squirrel retina. E. Mosaic of blue-OFF/yellow-ON RGCs from squirrel retina; measurements simultaneous with those in D. F. Temporal receptive fields measured from spike-triggered averages; same data as in D. Right side is time of spike, and traces show average time course of blue and green display primaries prior to the spike. Data show the cells are strongly driven by increments in blue light and weakly driven by decrements in green light. G. Same as F, but for blue-OFF cells in E. Data are from [38, 39].

features of color vision from information-theoretic arguments, but all were forced to rely on strong simplifying assumptions, particularly the use of decorrelation (or ICA) objectives as proxies for mutual information. Moreover, these studies gave significantly less attention to the spatial structure of color-opponent receptive fields, and none to their temporal properties. Finally, all models were *linear* encoding models, which are insufficient for explaining several critical aspects of mosaic organization [30]. **Our goal is to determine whether EC principles are sufficient to explain key aspects of color vision, including the chromatic structure of receptive fields and their mosaic organization. More specifically: are these mosaics coordinated in any way with achromatic RGC mosaics?**

**Research design:** We propose to address these challenges by generalizing the model (1, 2) [22, 30, 31] to two-channel linear filters  $w_{cj}$ , where  $j$  again indexes neuron and  $c$  color channel. As before, the model is trained to maximize mutual information between inputs  $x$  and output firing rates  $r$ , this time using natural images where each channel encodes photoreceptor (M and S) activations. Once again, we will train models using stochastic gradient descent on minibatches of image patches drawn from a public dataset in which the channels correspond to estimated L, M,

From a mathematical perspective, inputs from these two cone types can be thought of as encoding multi-channel images (pixels  $\times$  pixels  $\times$  channels). Prior work to explain chromatic opponency from EC followed previous ideas from Barlow and considered decorrelation as an EC strategy, arguing that significant correlations between photoreceptor responses should inevitably give rise to color opponency [17, 40]. This line of reasoning was subsequently extended beyond second-order independence (decorrelation) and Gaussian images to full independence (minimum mutual information) and natural images by [18, 41], who used ICA to learn an efficient image basis set. This approach predicted color opponency without the need to assume decorrelation. Later, [16], building on a previous EC model [14, 15], derived predictions for the *shapes* of chromatic RFs in a two-receptor model. Each of these approaches was able to explain key

and S cone activations [42]. We will optimize on just the M and S channels in these datasets. We will consider both models in which receptive fields are initially random unparameterized filters and those in which these images are parameterized as differences of Gaussians [30]. In addition, during the middle stages of training, we will temporarily jitter the center locations and narrow their widths, as this has been shown to improve model training [30, 31]. After training, we will examine the extent to which these models successfully reproduce known features of retinal physiology, including the existence of monochromatic/luminance and blue-yellow opponent receptive fields, each organized into mosaics. Furthermore, we will examine the effects of both input and firing rate noise on the relative organization of these mosaics [30], as well as the effects of different relative noise levels in each cone channel [16]. This is relevant because S cones exhibit lower noise levels than M cones [43]. Our prior work demonstrates that for the emergence and investigation of mosaic structure, a *nonlinear* transformation from filtered signal to firing rate is *crucial*, along with the use of natural scene input statistics [22, 23, 30], both of which are correctly and uniquely accounted for in our model.

For optimizing purely in the spatial domain (optimizing to images), a preexisting database will likely suffice [42]. However, this database does not contain movies, which is important for optimizing multi-channel EC models in space and time. Furthermore it only contains three color channels that are matched to the spectral sensitivities of the L, M, and S cones in the primate retina. Other species have cone types with different spectral sensitivities, and some species (e.g., birds) are frequently tetrachromatic. Thus, we propose building a video camera system with three (eventually four) aligned cameras with filters in front of each camera that mimic the spectral absorptions of the different opsins. The design will follow previous work using two cameras [44]. We will then film motion through natural environments using this camera system. We can superimpose eye movements measured from primates or other species onto these movies [45, 46]. While these movies will be terrestrial and thus not mimic the natural visual environments across all species (e.g., they will exclude fish), they will allow us to test how EC predictions depend on varying the wavelength content and the number of channels in a set of natural movies.

We will then compare predictions from EC to the receptive fields and mosaic organization of chromatically opponent blue-yellow RGCs in the primate retina (macaque and marmoset) and ground squirrel retina [39]. We will perform large-scale multi-electrode array measurements from primate and squirrel retinas, while presenting blue-yellow checkerboard noise: the activity of a ‘blue’ primary will be modulated independently of a ‘yellow’ primary at each checker across space and time. Our video display is based on a digital light projector (DLP) with a digital mirror device (DMD) and LED light sources [37]. The LEDs are modular in the system, so the primary wavelength can be chosen to maximize the contrast between the S and M cones (M+L cones in primates). We will use the checkerboard noise to map the spatiotemporal RFs of blue-yellow opponent and achromatic center-surround RFs. We will make the following comparisons:

- 1. Individual RF structure:** We will examine the extent to which the structure of individual receptive fields matches that of the EC predictions. Specifically, blue-ON/yellow-OFF RFs exhibit a yellow-OFF region that is nearly, but not precisely, spatially coextensive with the blue-ON region [38]. This differs significantly from that of ‘achromatic’ channels in the primate retina, the so-called ON and OFF parasol RGCs, which exhibit surround receptive fields that are much larger than the centers [32, 38]. Additionally, the yellow-OFF response exhibits longer temporal integration than the blue-ON response. Are these spatiotemporal features of the blue-yellow retinal pathway predicted by EC?

- 2. Within RGC Type Mosaic Organization:** We will compare the mosaic organization predicted by EC with the measured organization of blue-ON/yellow-OFF receptive fields in primate

and squirrel and blue-OFF/yellow-ON pathways in squirrel. We will also examine how these structures are impacted by changes in the input and output noise parameters in the model.

**3. Cross-Type RGC Mosaic Coordination:** Our previous work indicates that EC predicts ON and OFF mosaics encoding similar visual features that are aligned or anti-aligned depending on the input and output noise [30, 31]. This prediction was for black-white images. We will likewise examine the predictions of our model for coordination between blue-yellow opponent mosaics in the ground squirrel. We will also examine predictions for mosaic coordination *across* achromatic and chromatic channels: for example, ON-parasol RGCs and blue-ON/yellow-OFF RGCs in the primate retina, or ON-alpha RGCs and blue-ON/yellow-OFF RGCs in the squirrel retina.

**Beyond Dichromacy:** Many primates are trichromatic, and birds and fish are frequently tetra-chromatic. Thus, multichannel optimizations of EC are likely to provide important insights into the theoretical underpinnings of color vision across the animal kingdom. Once we have achieved our goals with examining the predictions of EC with respect to dichromatic processing, we anticipate expanding these predictions to three- and four-channel optimizations, with the spectral sensitivities matched to the cone opsins of other species. We can then ask if optimal mosaic organizations or intermosaic coordination depend on the number of chromatic channels available and how these results depend on noise. One potential prediction from the biology under high noise conditions (e.g., a starlit night) is that chromatically opponent channels collapse to achromatic channels, as evidenced by the switch between day and night vision observed almost universally across the animal kingdom.

**Pitfalls and alternative approaches:** Though the multi-channel model is a straightforward generalization of our earlier work, it may be possible that optimization will produce incomplete mosaics or spurious receptive fields. In this case, we may approximate receptive fields using a trainable Gaussian basis set in each channel. This will reduce the complexity of the optimization problem without assuming the chromatic composition of receptive fields. In turn, if this approach fails, we could separately learn the chromatic structure and spatial organization of mosaics in separate models, which would still allow us to reproduce known results and study relative spatial arrangements of the various cell types.

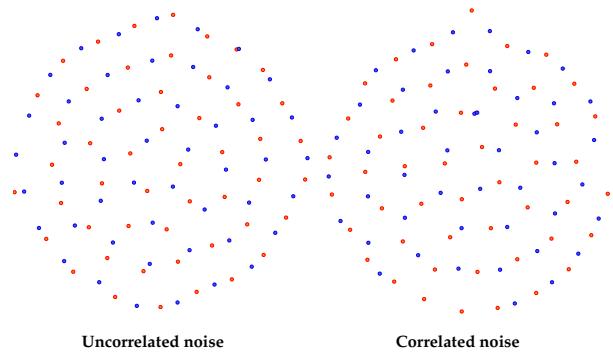
#### Extension 2: Correlated noise (Years 2–3)

**Rationale:** While it is widely appreciated that the correlation structure of natural images is essential for understanding the organization of retinal function under EC assumptions, and many EC models have considered the effects of differing noise inputs on coding strategies [47, 48], most models of retinal activity have assumed uncorrelated noise in the retina [e.g., 14–18, 23, 30, 48], despite the fact that shared neural variability beyond that suggested by stimulus correlations has been widely observed [34, 49–52]. Indeed, failing to account for correlated noise significantly diminishes the performance of retinal decoding models [34]. Thus, it is essential that EC models correctly account for the difference between stable features of the visual world (signal) and distortions of their transduced representations in the early visual system (noise).

Furthermore, different types of RGCs exhibit striking differences in their noise correlations. Two RGC types, each forming their own mosaics, will display large differences in the amplitude of noise correlations between neighboring RGCs [34, 53, 54]. Furthermore, these noise correlations change in both amplitude and temporal structure with adaptation state [34, 55]. The strength and temporal structure of the noise correlations reflect aspects of the presynaptic circuits that feed the RGCs such as amount of shared photoreceptor input and gap junction coupling [53, 56–58]. Correlations can be introduced in the input noise by gap junction coupling between photoreceptors and by horizontal cell feedback to photoreceptors. These input noise correlations can be further

increased by gap junction coupling within the retina, such as between amacrine cells [59]. In general this coupling is dynamically modulated by light adaptation and neuromodulators such as dopamine [60, 61], and coupling is stronger at low light levels and weaker at higher light levels. Finally, gap junctions also exist between many (but not all) RGC types, and their conductances can be modulated by light level [57, 62]. All of this indicates that signal integration, and perhaps noise correlations themselves, are under the direct control of adaptational processes, raising the question whether this, too, can be explained by EC.

We propose to generalize both input ( $C_{in}$ ) and output ( $C_{out}$ ) noise sources in our model (1, 2) beyond the independent Gaussian case. Specifically, we will consider: **1)** Gaussian input noise with spatial and temporal correlation scales  $\rho$  and  $\tau$ ; **2)** input noise with long-range (e.g., Cauchy or  $t$ -distributed) correlations; **3)** input noise with large, sparse, short-range distortions (random occlusions, flashes, etc.). In doing so, we aim to test the robustness of the learned retinal codes against these more drastic perturbations, asking how receptive fields and mosaics change in response to less reliable input. In particular, **we hypothesize that**, as with independent Gaussian noise, receptive fields widen in response to larger spatial and temporal correlations, performing more averaging. In addition, we expect RF shapes to expand when occlusions are small (and can be averaged out) and contract when they are large, preserving fine-scale information.



**Figure 6: Models with correlated noise inputs form regular mosaics.** As in models with uncorrelated noise inputs (left), models with RGC inputs correlated over some spatial scale  $\rho$  also form regular mosaics of ON and OFF receptive fields. Blue and orange dots mark the centers of OFF and ON RFs, respectively.

case with entries  $c_{ij} = C_0 \exp(-\|\mathbf{z}_i - \mathbf{z}_j\|/\rho)$  for pixels at spatial locations  $\mathbf{z}_i$  and  $\mathbf{z}_j$ . As Figure 6 shows, the learned mosaics in the case of both uncorrelated noise ( $\rho = 0$ ) and correlated noise exhibit the same type of anti-aligned mosaic coordination reported in [30, 31]. Both models required approximately the same training time and produced center-surround receptive fields. The next steps are highlighted in **Resarch Design**.

**Research design:** We will train models of the form outlined above (1, 2) using noise correlations  $C_{in}$  and  $C_{out}$  that generalize their previous diagonal forms. Initially, we will train these models on monochrome natural images distorted by a variety of noise models including correlated Gaussian, long-tailed, and structured (random, sparse “blobs”) of characteristic spatial scale  $\rho$ . We will then examine the effects of these perturbations as a function of  $\rho$  on receptive field shape and size, as well as on mosaic organization and alignment [30, 31]. Model training will take place as described in **Exension 1**. In addition, we will train similar models on natural *movies* as described in our previous work [22], this time including similar temporal correlations of characteristic duration  $\tau$ .

Likewise, we will examine the effects of output noise correlations on EC predictions by con-

Moreover, we will also consider the case of correlated noise among RGC outputs ( $C_{out}$ ). Here, we will ask whether it is optimal for RGCs that share correlated noise to be located within the same mosaic or different mosaics, and the extent to which the answer depends on whether correlated noise is restricted to be among nearby neurons [63–65]. The latter can be induced by making  $C_{out}$  a function of RF centers, so that output noise correlations vary systematically across the visual field.

**Preliminary studies:** We have successfully generalized our previous models [30, 31] to consider spatial correlations in the input matrix  $C_{in}$ . More specifically, we considered the

straining  $C_{\text{out}}$  to match *measured* noise correlations across diverse RGC types using our large-scale multi-electrode array [34, 54, 64, 65]. We will do this in two ways: **First**, we will constrain only the *form* of these noise correlations, with nearby RGC centers sharing distance-dependent correlations *within mosaics*. We will then examine how this shared output noise structure affects receptive field shapes, mosaic structure, and coordination across mosaics. **Second**, we will also consider the ability of correlated RGC *inputs* to explain the measured correlations among RGCs within and across types. Above, we have used  $C_{\text{in}}$  to model noise statistics in natural images, but these noise terms, as inputs to the RGC filters, are mathematically equivalent to *noise within the retina itself*. As such, we can ask how well observed RGC output correlations can be explained by different assumed forms of  $C_{\text{in}}$ . Ultimately, in **Extension 3** below, we will consider multi-layer models in which noise in the input images is distinct from noise in intermediate layers of processing.

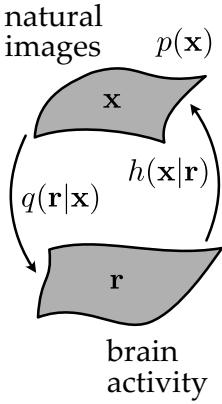
Finally, we propose to compare these predictions with experimentally measured RFs from RGCs at different input light levels. Previous work has demonstrated not only that gap junction coupling between photoreceptor is modulated by light levels, but so are correlations among RGCs [54, 55]. Thus, both the effective  $C_{\text{in}}$  and  $C_{\text{out}}$  parameters in our model can be changed by manipulating light levels. Moreover, previous theoretical studies [14, 15] found that EC predictions for optimal RF shapes were recapitulated within the *same* RGCs *across light levels*. In other words, RF shapes measured across differing light levels appear to correspond to predictions from EC models at differing levels of noise. This suggests that dynamic changes in RF structure across light levels can be framed as solutions to a meta-optimization problem in which RGCs are parameterized as *noise-dependent* linear filters that efficiently code across a wide range of light levels.

**Pitfalls and alternative approaches:** As shown above (**Figure 6**), preliminary results suggest that our models easily generalize to the case of correlated Gaussian noise (in both  $C_{\text{in}}$  and  $C_{\text{out}}$ ), so we do not anticipate difficulties with model optimization. However, if models fail to train at higher noise levels, we may implement a two-stage training procedure in which models are first trained on independent Gaussian noise, followed by a subsequent period of training in which the correlation scales are gradually increased. This should encourage models to find solutions “near” biological configurations and allow us to track the emergence of deviations in RF structure and organization as the perturbations are ramped up. A second potential challenge is accurately measuring short-timescale spike correlation because spike shapes can superimpose and interfere with spike sorting. Fortunately, recently developed spike sorting algorithms, such as YASS [66, 67], explicitly handle the spike super position problem. We have tested YASS and indeed, more spikes are found per neuron and short-time scale correlations tend to increase. Thus, we will use spike sorters that explicitly handle the superposition problem to minimize the impact on estimating correlated spiking among RGCs.

### Extension 3: Multi-layer encoding (Years 4–5)

**Rationale:** Previous studies from our group and others [21–23, 30, 31] have optimized single-layer encoding models according to EC. These models have produced first-principled predictions that are often in striking agreement with the observed functional organization of RGC receptive fields. However, because the model being optimized is just a single layer, they lack the complexity required to generate key features of RGC receptive fields such as subunit rectification (**Figure 8**) [68]. ‘Subunits’ are widely recognized as being central to the complex processing that takes place in the retina [69]. By contrast, multi-layer convolutional neural network models fit to retinal data [25–27, 29] have produced state-of-the-art matches to these data, including predicting many surprising nonlinearities observed in retinal processing. While illuminating, these models are divorced from a first principles understanding of early visual processing. Here, we propose to

bridge this divide by investigating multi-layer models of EC, asking how multi-stage feedforward circuits should be organized to maximize information extraction from visual scenes.



**Figure 7: Probabilistic framework for multi-layer encoding.** Natural images  $x \sim p(x)$  are mapped to brain activity  $r$  via a parameterized encoding model  $q(r|x)$ . This relationship is approximately inverted by the learned mapping  $h(x|r)$ , which is a deep neural network.

additional regularization terms that would be present in the ELBO are those necessary for ensuring that  $q(r|x)$  also approximates the Bayesian posterior when (3) is optimized over  $\phi$ ; they do not involve  $h(x|r)$ .

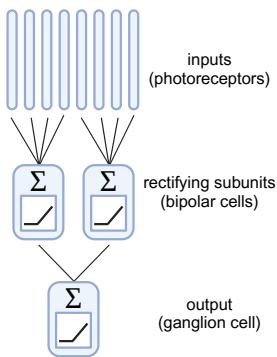
Using this framework, we will consider two specific multi-layer models that address distinct questions: **First**, we take  $q$  to be a two-layer feedforward network with a first layer comprising linear weights with parameterized output nonlinearities, followed by a second layer similar to (1). That is, the first layer models a set of computational subunits [68, 69] (Figure 8), while the second layer again models RGCs. Here, we can ask about both the optimal form of these subunit nonlinearities and how they contribute to the formation of receptive fields in the RGCs, as well as the relationship between noise correlations among first-layer units and output correlations among second-layer units. **Second**, we consider an encoding network that is a cascade of two linear-nonlinear (LN) models of the form (1). This model corresponds to the populations of RGC and LGN cells, expanding EC predictions past the retinal output layer. Crucially, though we focus first on the two-layer case, this model naturally scales to deeper networks and does not require assumptions of Gaussian noise in the encoding model, allowing us to consider direct modeling of spike counts. In each case, we will parameterize the inverse mapping  $h$  from brain activity to images as Gaussian:  $h_\theta(x|r) = \mathcal{N}(\mu_\theta(r), \varepsilon^2 \mathbb{1})$ .

Once again, we will train networks as previously described, performing stochastic gradient descent on the objective (3) coupled with an augmented Lagrangian method to enforce the firing rate constraints  $\mathbb{E}r_j = 1$  within each layer for LN cells (1) [30, 31]. Units with a saturating nonlinearity will not require this constraint. We will verify that this model correctly recapitulates previous results for RGCs [23, 30] when trained on monochromatic images. We will also examine if trained receptive fields exhibit subunits that locally rectify input [68, 71, 72], a hallmark of bipolar cells in the retina.

**Research design:** The key technical challenge to optimizing mutual information between network outputs ( $r$ ) and natural image inputs ( $x$ ) lies in the fact that (2), our previous training objective, depends strongly on Gaussian approximations in a single-layer model to approximate the posterior  $p(x|r)$ . In [23], it was argued that computing this posterior (and thereby  $H(X|R)$ ) was a simpler route to mutual information  $I(X; R)$  than computing both  $H(R|X)$  (easy) and  $H(R)$  (hard). We propose to continue this approach but use a more flexible method—a neural network—to approximate  $p(x|r)$ . More specifically, following the literature on variational autoencoders (VAEs) [70], let  $p(x)$  be the (unknown, sampled) distribution over natural images,  $q_\phi(r|x)$  be the multi-layer EC model parameterized by  $\phi$ , and  $h_\theta(x|r)$  be a *reconstruction* model (parameterized by  $\theta$ ) that is a valid conditional probability density (Figure 7). In this case, we can consider the optimization problem in which we minimize the negative log likelihood of this reconstruction:

$$\min_{\theta, \phi} \mathbb{E}_{x \sim p(x)} \mathbb{E}_{r \sim q_\phi(r|x)} [-\log h_\theta(x|r)] \text{ subject to } \int h_\theta(x|r) dx = 1. \quad (3)$$

It is straightforward to show that the solution to the optimization over  $\theta$  is given by  $h_*(x|r) \propto p(x)q(r|x)$ , i.e., the posterior distribution, and in this case, (3) is equal to  $H(X|R)$ . (Note: the expression (3) is closely related to the evidence lower bound (ELBO) for VAEs with a crucial difference: the additional regularization terms that would be present in the ELBO are those necessary for ensuring that  $q(r|x)$  also approximates the Bayesian posterior when (3) is optimized over  $\phi$ ; they do not involve  $h(x|r)$ .)



**Figure 8:** Circuit schematic showing rectifying subunits. Photoreceptor inputs (top) are locally summed and then passed through a rectifying nonlinearity, approximating the operation performed by retinal bipolar cells. Subunit outputs are then summed at a second stage and again passed through a rectifying nonlinearity, approximating RGC output [68, 69].

We will compare predictions of this model to data from primate, squirrel, and rat retinae. Across species, some RGC types are X-like in their response properties, meaning that they exhibit little to no subunit rectification: midget RGCs in the primate retina and brisk sustained RGCs in the rat retina are two examples [71, 73]. Other RGCs exhibit strong Y-like responses, meaning strong subunit rectification classically measured using contrast reversing gratings and extracting the amplitude of the second harmonic response to the sinusoidal modulation of the grating contrast in time [74]. Prior EC optimizations, because they were performed on single-layer retinal models, have never predicted subunit rectification in an RGC type, nor have they predicted that some RGC types will be X-like and others Y-like. We will compare predictions of the multi-layer EC to the degree of subunit rectification across the major RGC types with center-surround RFs in these species to determine the extent to which EC can explain this aspect of retinal coding. Furthermore, prior experiments indicate that blue-ON/yellow-OFF RGCs in the primate retina are X-like [75]. Finally, combining **Extension 1** with **Extension 3**, we will likewise test whether multi-channel, multi-layer models will produce X-like cone-opponent RGC types and Y-like achromatic RGC types.

In combination with **Extension 2** above, we will also consider the effects of various noise distributions on optimizations of this multi-layer model of the retina, particularly on mosaic alignment, its upstream precursors and downstream readouts. That is, we will ask how anti-aligned mosaics could arise through differing connectivity patterns with the hidden layer in the network (which would model subunits) and how this information might in turn be utilized in downstream areas such as the lateral geniculate nucleus or primary visual cortex.

**Pitfalls and alternative approaches:** Should we find that our multi-layer model proves difficult to train end-to-end, we will make use of greedy layerwise training to incrementally build models. This is partially justified on theoretical grounds by the data processing inequality [76], which states that for a chain of variable dependencies  $X \rightarrow Y \rightarrow Z$ ,  $\mathcal{I}(X; Z) \leq \mathcal{I}(X; Y)$ : one cannot extract more information in total than is passed through each link in the chain. Thus, for networks with unique (or nearly unique) optima, layerwise training should produce nearly the same result as end-to-end training. Given our previous success in training such models (and the Pearson Lab’s previous experience with similar VAE models [77–79]), we do not anticipate additional difficulties in multiple rounds of such optimizations. Moreover, after establishing the shapes taken by unconstrained receptive fields, we may additionally parameterize receptive field shapes matching these prior results (e.g., differences of Gaussians) for faster convergence.

## RESULTS FROM PRIOR SUPPORT

NEI	R01-EY031396	Field (PI)	8/2020 – 3/2024
<i>Receptive field coordination across mosaics of diverse retinal ganglion cell types in the mammalian retina</i>			
Roles: Field, PI; Pearson, Co-I	\$1,440,000	(total costs)	

Throughout the brain, neural circuits are composed of diverse cell types. In the retina, cell types are spatially organized into mosaics, in which the receptive fields of each type approxi-

mately tile space. This organization ensures that the computations performed by each cell type occur uniformly across the visual field, with no gaps or gluts in processing. This exquisite coordination within each cell type to uniformly cover space raises the question, “Is there coordination across cell types?” In general, this coordination could manifest as either a tendency toward alignment, or anti-alignment, between two mosaics of receptive fields. Our central hypothesis is that mosaics of receptive fields are intricately coordinated across retinal cell types and that this coordination reflects fundamental organizing principles for how the retina processes natural scenes. We demonstrate with preliminary data that across four retinal ganglion cell (RGC) types in the rat retina, mosaics of receptive fields are intricately coordinated. Specifically, mosaics of ON and OFF RGC types that signal similar visual features are consistently anti-aligned. The objectives of this proposal are to build upon these observations to (1) understand the mechanisms that underlie inter-mosaic coordination (Aim 1), (2) determine how and when this coordination develops (Aim 2), and (3) to determine the significance of inter-mosaic coordination of visual as well as how extensively mosaics are coordinated across additional RGC types (Aim 3). This proposed research is significant because it will uncover an entirely new phenomenon in the vertebrate retina: inter-mosaic coordination across diverse cell types with diverse receptive field properties. It will also reveal either new developmental mechanisms, or new roles for previously established mechanisms in coordinating mosaics across RGC types. It will also make novel predictions for how downstream neurons could pool over retinal inputs to produce orientation and direction tuned responses. Finally, this work is significant because it extends the theoretical basis (e.g. ‘efficient coding theory’) of how we understand the organization of retinal processing. The proposed research is innovative because it applies a recently developed analytical framework to large-scale population measurements of RGC receptive fields with the goal of understanding the contributions of cell position and synaptic specificity to inter-mosaic coordination. Furthermore, the work is conceptually innovative because it shows receptive field mosaics are coordinated and identifies the benefits of coordination to vision. The expected outcome of this research is a novel set of mechanisms and developmental process that yield functional coordination across distinct cell types in the retina. We anticipate these discoveries as being fundamental and impactful to understanding the retina and sensory processing. **We note the Aims of this previously funded NIH R01 grant are completely distinct from the goals of this CRCNS application. Specifically, the previous grant was focused on the biological mechanisms that generate inter-mosaic coordination, while this grant compares predictions of efficient coding theory to the biology.**

- Roy, S. *et al.* Inter-mosaic coordination of retinal receptive fields. *Nature* **592**, 409–413 (2021)
- Jun, N. Y. *et al.* Scene statistics and noise determine the relative arrangement of receptive field mosaics. *Proceedings of the National Academy of Sciences* **118** (2021)
- Jun, N. Y. *et al.* Efficient coding, channel capacity, and the emergence of retinal mosaics. *Advances in neural information processing systems* **35**, 32311–32324 (2022)
- Ruda, K. *et al.* The functional organization of retinal ganglion cell receptive fields across light levels. *bioRxiv*, 2022–09 (2022)

BRAIN/NIDA 1RF1-DA056376 Pearson (PI)  
*Real-time mapping and adaptive testing for neural population hypotheses*  
Role: PI \$1,400,000 (total costs)

9/2022 – 8/2025

Recent advances in neural recording technologies have made it possible to study increasingly large and diverse subsets of neurons, producing a growing interest in the collective computational

properties of neural populations. Ideally, causally testing these population hypotheses requires timing and selecting experimental manipulations based on the current state of neural dynamics, but technical limitations have rendered this difficult in practice. However, recent work on real-time preprocessing and modeling of neural data has demonstrated that up-to-the minute estimates of neural population dynamics are indeed possible, opening the door to adaptive experiments in which the design of the task changes based on incoming data. The goal of this proposal is to disseminate these advances to the widest possible audience of systems neuroscientists by: 1) Designing and validating new methods for mapping neural states and behavior online. 2) Developing algorithms for optimally timing and selecting experimental manipulations based on these instantaneous neural and behavioral states. 3) Making *improv*, our platform for adaptive experiments, easier to install, use, and configure for diverse model organisms and hardware setups. By allowing researchers to test ideas online, such tools will facilitate rapid “drill-down” from the whole brain to the local circuit levels, maximizing statistical efficiency in limited experimental time and providing stronger causal inferences for neural population hypotheses, with broad implications for systems neuroscience. **This software platform is publicly developed under the [pearsonlab GitHub organization](#) and distributed under a permissive (MIT) open source license.**

- Draeflos, A. *et al.* *improv*: A flexible software platform for adaptive neuroscience experiments. *bioRxiv* (2021)

## BROADER IMPACTS

**Benefits to other areas of neuroscience:** As discussed above, one of the key factors limiting progress in neuroscience is the lack of simple theoretical principles for organizing our thinking about nervous system function. What’s more, of the few candidate principles we have, even fewer are capable of making concrete, testable predictions about experimental data. Efficient coding (EC) is one of the most successful of such principles, and our proposed model extensions make EC predictions tractable for a much wider variety of signals and systems. For instance, our multichannel encoding model (**Extension 1**) is primarily targeted at color vision, but it shares this feature with high-dimensional sensory systems like olfaction with hundreds of sparsely active input channels. Likewise most sensory systems deal with correlated noise, both in their sensory inputs and in their internal (downstream) representations; the techniques in **Extension 2** will be applicable to understanding the consequences of correlated noise. Finally, our proposed framework for multi-layer models (**Extension 3**) is quite general, and could be further generalized to incorporate recurrence, decoding, and other features across a wide variety of neural circuits. Thus, we expect that the model training techniques developed as part of this proposal will impact neuroscience research far beyond the retinal physiology and visual neuroscience communities.

In addition, as with previous work (repositories [nayoungjun/efficient-image](#) and [pearsonlab/efficientcoding](#)), we will make all model code, training details, and data analyses publicly accessible on GitHub under a permissive (MIT) open source license. Electrophysiology data comprising hundreds of neurons and dozens of cell types per recording will likewise be made available through public repositories (DANDI), and we plan to release our training dataset of cone-filtered natural movies as a public benchmark dataset.

**Clinical and technological applications:** Despite the impressive successes of graphics processing units (GPUs) in facilitating the deep learning revolution, models based on deep networks are incredibly energy-hungry, requiring hundreds of thousands of kilowatt-hours to train a state-of-the-art large language model [82]. By contrast, the retina is the most energy-efficient visual sensor we know, consuming a tiny fraction of the power of commercial cameras while seamlessly adapt-

ing to orders of magnitude greater variation in luminance. Thus, understanding the core design principles by which the retina extracts information from the visual scene can provide valuable inspiration for the design of both visual sensors (hardware) and computer vision systems (software). Moreover, as we argued earlier in this section, EC principles are likely to generalize across sensory systems, providing ideas and motifs that could be replicated by sensors for other natural signals, including those for which no biological analogs exist.

Similarly, efforts to produce an artificial retinal prosthesis, remain hampered by the mismatch between the information provided by available sensors and the signals expected by the rest of the brain. Here again, performing efficient compression under real-world conditions is an enormous engineering challenge. Models like ours, which mimic the *functional* properties of the retina, can thus be used for comparison with simulated retinal implants, asking how well the latter perform in relation to optimal systems under power constraints. In this way, EC can likewise become a driving force in helping design the next generation of visual prosthetics.

**Quantitative training:** Quantitative skills are among the most essential for modern neuroscientists to acquire, and yet few undergraduate majors require classes in statistics or computer science. Dr. Pearson has been active in helping to remedy this gap, teaching courses in programming and neural data analysis for which the materials are publicly available (hosted on GitHub at [jmxpearson/neural-data-analysis-book](https://jmxpearson/neural-data-analysis-book)). In these efforts, topics in visual neuroscience are among the most effective at connecting with students, since programming exercises based on image processing offer students clear visual feedback against which to test their understanding. Likewise, Dr. Field will begin teaching a course targeted to advanced undergraduates and graduate students at UCLA on Visual Neuroscience and Computation that will blend learning about visual processing with coding and optimization exercises to model neurons and circuits. Efficient coding (EC) is a two week module in this 10-week course. He has previously taught overlapping material at Duke University and he guest lectures at a summer course on Visual Neuroscience at Cold Spring Harbor Labs that draws students internationally. We thus propose to develop teaching modules based on EC that can be added to Dr. Pearson's freely available online course materials.

**Fostering diversity, inclusivity, and accessibility:** The PIs are committed to nurturing talent across the widest possible cross-section of society, particularly among those historically underrepresented in science. Including students matriculating Fall 2024, Dr. Pearson's lab will be >40% female in a field (computational neuroscience) that is ~ 80% male, with 66% of the lab falling into categories defined as underrepresented by NIH. For his mentoring efforts, he has been recognized by the Duke University School of Medicine with both its teaching (2020) and early-career mentoring (2022) awards. Furthermore, he is an active participant in outreach activities, including as a speaker at Duke's first and second annual "Brain Careers Day" in partnership with North Carolina's HBCUs and as the instructor for a one-week short course to be delivered to rural high school students in summer 2024. Dr. Field has been the lead advisor or co-advisor for six graduate students, all working in computational neuroscience, all six of whom are women. Of the eight post-baccalaureate researchers who have worked in Dr. Field's lab, four have come from underrepresented groups in STEM. While at Duke, Dr. Field ran graduate admissions for the Neurobiology Program for 7 years, during which the fraction of matriculating students from underrepresented groups exceeded the fraction in the application pool. Furthermore, Dr. Field served on the Diversity, Equity and Inclusion Committee for the Department of Neurobiology at Duke University, and he is active in the equivalent committee in the Department of Ophthalmology at UCLA and is responsible for organizing an outreach event to California State Universities, many of which are Hispanic-Serving Institutions (HSIs).

## COORDINATION PLAN

Successful completion of the research plan outlined above requires the close collaboration and complementary expertise of both PIs, Dr. Field in retinal physiology and data analysis and Dr. Pearson in theoretical modeling and model fitting. In addition to the fact that Dr. Field has a long history of successful collaboration with theorists [34, 64, 71, 83–86] and Dr. Pearson with experimentalists [e.g., 77–79, 87–89], both have a significant history (>5 years) of collaboration with each other, including shared funding, a jointly advised graduate student, and multiple publications [22, 30, 31]. Since Dr. Field’s move to UCLA, they have continued to meet weekly via Zoom with Dr. Pearson’s graduate student Mr. St-Amand. Moreover, their previous student, Dr. Jun, was remote through most of her PhD, giving the team extensive experience with productive remote collaboration.

**Roles of the PIs:** In their previous collaborative efforts, the PIs have worked closely to jointly supervise *all* aspects of the work, including weekly meetings, supervision of computational experiments and analysis of results, and drafting of manuscripts for publication. However, **Dr. Pearson** will bear primary responsibility for model design, fitting, and debugging, mathematical derivations and analysis, and theoretical conclusions, while **Dr. Field** will be responsible overseeing electrode array experiments, building the camera system for natural movie generation, generating the natural movies, alignment of modeling assumptions with retinal physiology, comparison of the models with retinal data, and interpretation of the results in the context of early vision.

**How the project will be managed:** As noted above, both PIs already maintain a standing weekly meeting (via Zoom) with a graduate student, Mr. St-Amand, to discuss results and plan next steps. Future meetings will also include Dr. Field’s postdoc Dr. Gallego so that all members of the team can coordinate activities on a regular basis. This will ensure that the planned data analysis and theoretical modeling are aligned. For day-to-day interactions, the PIs have for years made extensive use of a Slack workspace, along with shared document repositories using Box and Overleaf. In addition, the collaboration has a history of open sourcing its code on GitHub (repositories `nayoungjun/efficient-image` and `pearsonlab/efficientcoding`), making the models developed by their groups available to the broader community.

Finally, we will facilitate cross-training opportunities for team members via budgeted site visits, members of the Duke team hosted at UCLA and the UCLA team at Duke in alternating years. These meetings will allow Dr. Pearson and Mr. St-Amand to enhance their firsthand knowledge of the experimental preparations used in retinal recording and for Dr. Field and Dr. Gallego to dedicate in-person time to intensive model development and vision casting for the project.

**Leadership plan:** Each PI will be responsible for fiscal management of the portion of the award dedicated to his institution, as well as day-to-day management of members of the team at their respective sites. As described above, team members will meet weekly via Zoom to plan analyses and computational experiments, discuss results, and review progress toward publication. Dr. Field will supervise all data collection, storage, and analysis, while Dr. Pearson will supervise theoretical derivations, model building, and debugging. Both PIs will co-supervise drafting of figures, results, and conclusions for eventual publication.

**Conflict resolution:** In the case of conflicts among trainees, the PIs will mediate the dispute, if necessary with the help of a neutral third party agreed upon by both trainees. If disputes arise between the PIs that they find themselves unable to resolve, they will submit these disagreements to a committee consisting of PIs’ department chairs plus an agreed-upon third member, none of whom will have any intellectual or financial connection to the award.

**Intellectual property:** While we do not anticipate the generation of intellectual property beyond the models discussed above, any legally protected IP will be subject to the terms of the PIs' employment. More specifically, any commercialization or patent efforts will be jointly coordinated through the relevant technology transfer offices at the PIs' institutions.

## BIBLIOGRAPHY & REFERENCES CITED

1. Bialek, W. *Biophysics: Searching for Principles* (Princeton University Press, 2012).
2. Barlow, H. B. Possible principles underlying the transformation of sensory messages. *Sensory communication* **1**, 217–234 (1961).
3. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2**, 79–87 (1999).
4. Jiang, L. P. & Rao, R. P. in *Oxford research encyclopedia of neuroscience* (2022).
5. Tishby, N., Pereira, F. C. & Bialek, W. The information bottleneck method. *arXiv preprint physics/0004057* (2000).
6. Palmer, S. E., Marre, O., Berry, M. J. & Bialek, W. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences* **112**, 6908–6913 (2015).
7. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences* **115**, 186–191 (2018).
8. Lewicki, M. S. Efficient coding of natural sounds. *Nature neuroscience* **5**, 356–363 (2002).
9. Smith, E. C. & Lewicki, M. S. Efficient auditory coding. *Nature* **439**, 978–982 (2006).
10. Gervain, J. & Geffen, M. N. Efficient neural coding in auditory and speech perception. *Trends in Neurosciences* **42**, 56–65 (2019).
11. Carriot, J., Jamali, M., Chacron, M. J. & Cullen, K. E. Statistics of the vestibular input experienced during natural self-motion: implications for neural processing. *Journal of Neuroscience* **34**, 8347–8357 (2014).
12. Mitchell, D. E., Kwan, A., Carriot, J., Chacron, M. J. & Cullen, K. E. Neuronal variability and tuning are balanced to optimize naturalistic self-motion coding in primate vestibular pathways. *Elife* **7**, e43019 (2018).
13. Carriot, J., McAllister, G., Hooshangnejad, H., Mackrouss, I., Cullen, K. E. & Chacron, M. J. Sensory adaptation mediates efficient and unambiguous encoding of natural stimuli by vestibular thalamocortical pathways. *Nature communications* **13**, 2612 (2022).
14. Atick, J. J. & Redlich, A. N. Towards a theory of early visual processing. *Neural computation* **2**, 308–320 (1990).
15. Atick, J. J. & Redlich, A. N. What does the retina know about natural scenes? *Neural computation* **4**, 196–210 (1992).
16. Atick, J. J., Li, Z. & Redlich, A. N. Understanding retinal color coding from first principles. *Neural computation* **4**, 559–572 (1992).
17. Buchsbaum, G. & Gottschalk, A. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences* **220**, 89–113 (1983).
18. Lee, T.-W., Wachtler, T. & Sejnowski, T. J. Color opponency is an efficient representation of spectral properties in natural scenes. *Vision Research* **42**, 2095–2103 (2002).
19. Gjorgjieva, J., Sompolinsky, H. & Meister, M. Benefits of pathway splitting in sensory coding. *Journal of Neuroscience* **34**, 12127–12144 (2014).
20. Gjorgjieva, J., Meister, M. & Sompolinsky, H. Functional diversity among sensory neurons from efficient coding principles. *PLoS computational biology* **15**, e1007476 (2019).

21. Ocko, S., Lindsey, J., Ganguli, S. & Deny, S. *The emergence of multiple retinal cell types through efficient coding of natural movies* in *Advances in Neural Information Processing Systems* (2018), 9389–9400.
22. Jun, N. Y., Field, G. & Pearson, J. Efficient coding, channel capacity, and the emergence of retinal mosaics. *Advances in neural information processing systems* **35**, 32311–32324 (2022).
23. Karklin, Y. & Simoncelli, E. P. *Efficient coding of natural images with a population of noisy linear-nonlinear neurons* in *Advances in neural information processing systems* (2011), 999–1007.
24. Paninski, L. Estimation of entropy and mutual information. *Neural computation* **15**, 1191–1253 (2003).
25. McIntosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S. & Baccus, S. Deep learning models of the retinal response to natural scenes. *Advances in neural information processing systems* **29** (2016).
26. Tanaka, H., Nayebi, A., Maheswaranathan, N., McIntosh, L., Baccus, S. & Ganguli, S. From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. *Advances in neural information processing systems* **32** (2019).
27. Maheswaranathan, N., Kastner, D. B., Baccus, S. A. & Ganguli, S. Inferring hidden structure in multilayered neural circuits. *PLoS computational biology* **14**, e1006291 (2018).
28. Wang, S., Hoshal, B., de Laittre, E., Marre, O., Berry, M. & Palmer, S. Learning low-dimensional generalizable natural features from retina using a U-net. *Advances in neural information processing systems* **35**, 11355–11368 (2022).
29. Maheswaranathan, N., McIntosh, L. T., Tanaka, H., Grant, S., Kastner, D. B., Melander, J. B., Nayebi, A., Brezovec, L. E., Wang, J. H., Ganguli, S., et al. Interpreting the retinal neural code for natural scenes: From computations to neurons. *Neuron* **111**, 2742–2755 (2023).
30. Jun, N. Y., Field, G. D. & Pearson, J. Scene statistics and noise determine the relative arrangement of receptive field mosaics. *Proceedings of the National Academy of Sciences* **118** (2021).
31. Roy, S., Jun, N. Y., Davis, E. L., Pearson, J. & Field, G. D. Inter-mosaic coordination of retinal receptive fields. *Nature* **592**, 409–413 (2021).
32. Croner, L. J. & Kaplan, E. Receptive fields of P and M ganglion cells across the primate retina. *Vision research* **35**, 7–24 (1995).
33. Field, G. D. & Chichilnisky, E. Information processing in the primate retina: circuitry and coding. *Annual review of neuroscience* **30**, 1–30 (2007).
34. Ruda, K., Zylberberg, J. & Field, G. D. Ignoring correlated activity causes a failure of retinal population codes. *Nature communications* **11**, 4605 (2020).
35. Yao, X., Cafaro, J., McLaughlin, A. J., Postma, F. R., Paul, D. L., Awatramani, G. & Field, G. D. Gap junctions contribute to differential light adaptation across direction-selective retinal ganglion cells. *Neuron* **100**, 216–228 (2018).
36. Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., et al. Functional connectivity in the retina at the resolution of photoreceptors. *Nature* **467**, 673–677 (2010).
37. Franke, K., Maia Chagas, A., Zhao, Z., Zimmermann, M. J., Bartel, P., Qiu, Y., Szafko, K. P., Baden, T. & Euler, T. An arbitrary-spectrum spatial visual stimulator for vision research. *elife* **8**, e48779 (2019).

38. Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M. & Chichilnisky, E. Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *Journal of Neuroscience* **27**, 13261–13272 (2007).
39. Sher, A. & DeVries, S. H. A non-canonical pathway for mammalian blue-green color vision. *Nature neuroscience* **15**, 952–953 (2012).
40. Barlow, H. B. Critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond., B* **212**, 1–34 (1981).
41. Wachtler, T., Lee, T.-W. & Sejnowski, T. J. Chromatic structure of natural scenes. *JOSA A* **18**, 65–77 (2001).
42. Doi, E., Inui, T., Lee, T.-W., Wachtler, T. & Sejnowski, T. J. Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural computation* **15**, 397–417 (2003).
43. Rieke, F. & Baylor, D. A. Origin and functional impact of dark noise in retinal cones. *Neuron* **26**, 181–186 (2000).
44. Qiu, Y., Zhao, Z., Klindt, D., Kautzky, M., Szatko, K. P., Schaeffel, F., Rifai, K., Franke, K., Busse, L. & Euler, T. Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations. *Current Biology* **31**, 3233–3247 (2021).
45. Rajashekhar, U., Cormack, L. K., Bovik, A. C. & van der Linde, I. DOVES: a database of visual eye movements. *Spatial Vision* **22**, 161–177 (2009).
46. Samonds, J. M., Geisler, W. S. & Priebe, N. J. Natural image and receptive field statistics predict saccade sizes. *Nature neuroscience* **21**, 1591–1599 (2018).
47. Brinkman, B. A., Weber, A. I., Rieke, F. & Shea-Brown, E. How do efficient coding strategies depend on origins of noise in neural circuits? *PLoS computational biology* **12**, e1005150 (2016).
48. Röth, K., Shao, S. & Gjorgjieva, J. Efficient population coding depends on stimulus convergence and source of noise. *PLoS computational biology* **17**, e1008897 (2021).
49. Ala-Laurila, P., Greschner, M., Chichilnisky, E. & Rieke, F. Cone photoreceptor contributions to noise and correlations in the retinal output. *Nature neuroscience* **14**, 1309–1316 (2011).
50. Vidne, M., Ahmadian, Y., Shlens, J., Pillow, J. W., Kulkarni, J., Litke, A. M., Chichilnisky, E., Simoncelli, E. & Paninski, L. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of computational neuroscience* **33**, 97–121 (2012).
51. Simmons, K. D., Prentice, J. S., Tkačík, G., Homann, J., Yee, H. K., Palmer, S. E., Nelson, P. C. & Balasubramanian, V. Transformation of stimulus correlations by the retina. *PLoS Computational Biology* **9**, e1003344 (2013).
52. Grimes, W. N., Hoon, M., Briggman, K. L., Wong, R. O. & Rieke, F. Cross-synaptic synchrony and transmission of signal and noise across the mouse retina. *Elife* **3**, e03892 (2014).
53. Mastronarde, D. N. Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to X-and Y-cells. *Journal of Neurophysiology* **49**, 303–324 (1983).
54. Greschner, M., Shlens, J., Bakolitsa, C., Field, G. D., Gauthier, J. L., Jepson, L. H., Sher, A., Litke, A. M. & Chichilnisky, E. Correlated firing among major ganglion cell types in primate retina. *The Journal of physiology* **589**, 75–86 (2011).
55. Mastronarde, D. N. Correlated firing of cat retinal ganglion cells. II. Responses of X-and Y-cells to single quantal events. *Journal of Neurophysiology* **49**, 325–349 (1983).

56. Ala-Laurila, P., Greschner, M., Chichilnisky, E. & Rieke, F. Cone photoreceptor contributions to noise and correlations in the retinal output. *Nature neuroscience* **14**, 1309–1316 (2011).
57. Hu, E. H., Pan, F., Völgyi, B. & Bloomfield, S. A. Light increases the gap junctional coupling of retinal ganglion cells. *The Journal of physiology* **588**, 4145–4163 (2010).
58. Völgyi, B., Pan, F., Paul, D. L., Wang, J. T., Huberman, A. D. & Bloomfield, S. A. Gap junctions are essential for generating the correlated spike activity of neighboring retinal ganglion cells. *PloS one* **8**, e69426 (2013).
59. Bloomfield, S. A. & Dacheux, R. F. Rod vision: pathways and processing in the mammalian retina. *Progress in retinal and eye research* **20**, 351–384 (2001).
60. Bloomfield, S. A. & Völgyi, B. The diverse functional roles and regulation of neuronal gap junctions in the retina. *Nature Reviews Neuroscience* **10**, 495–506 (2009).
61. Roy, S. & Field, G. D. Dopaminergic modulation of retinal processing from starlight to sunlight. *Journal of pharmacological sciences* **140**, 86–93 (2019).
62. Völgyi, B., Abrams, J., Paul, D. L. & Bloomfield, S. A. Morphology and tracer coupling pattern of alpha ganglion cells in the mouse retina. *Journal of comparative neurology* **492**, 66–77 (2005).
63. Schneidman, E., Berry, M. J., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
64. Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M. & Chichilnisky, E. The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience* **26**, 8254–8266 (2006).
65. Shlens, J., Field, G. D., Gauthier, J. L., Greschner, M., Sher, A., Litke, A. M. & Chichilnisky, E. The structure of large-scale synchronized firing in primate retina. *Journal of Neuroscience* **29**, 5022–5031 (2009).
66. Lee, J. H., Carlson, D. E., Shokri Razaghi, H., Yao, W., Goetz, G. A., Hagen, E., Batty, E., Chichilnisky, E., Einevoll, G. T. & Paninski, L. YASS: yet another spike sorter. *Advances in neural information processing systems* **30** (2017).
67. Lee, J., Mitelut, C., Shokri, H., Kinsella, I., Dethé, N., Wu, S., Li, K., Reyes, E. B., Turcu, D., Batty, E., et al. YASS: Yet Another Spike Sorter applied to large-scale multi-electrode array recordings in primate retina. *BioRxiv*, 2020–03 (2020).
68. Hochstein, S. & Shapley, R. Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *The Journal of physiology* **262**, 265–284 (1976).
69. Gollisch, T. & Meister, M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* **65**, 150–164 (2010).
70. Kingma, D. P., Welling, M., et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning* **12**, 307–392 (2019).
71. Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., Paninski, L., Simoncelli, E. P., et al. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *Elife* **4**, e05241 (2015).
72. Freedland, J. & Rieke, F. Systematic reduction of the dimensionality of natural scenes allows accurate predictions of retinal ganglion cell spike outputs. *Proceedings of the National Academy of Sciences* **119**, e2121744119 (2022).
73. Ravi, S., Ahn, D., Greschner, M., Chichilnisky, E. & Field, G. D. Pathway-specific asymmetries between ON and OFF visual signals. *Journal of Neuroscience* **38**, 9728–9740 (2018).

74. Crook, J. D., Peterson, B. B., Packer, O. S., Robinson, F. R., Troy, J. B. & Dacey, D. M. Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. *Journal of Neuroscience* **28**, 11277–11291 (2008).
75. Chichilnisky, E. & Baylor, D. Receptive-field microstructure of blue-yellow ganglion cells in primate retina. *Nature neuroscience* **2**, 889–893 (1999).
76. Cover, T. M. & Thomas, J. M. *Elements of information theory* (John Wiley & Sons, 2005).
77. Goffinet, J., Brudner, S., Mooney, R. & Pearson, J. Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife* **10**, e67855 (2021).
78. Singh Alvarado, J., Goffinet, J., Michael, V., Liberti III, W., Hatfield, J., Gardner, T., Pearson, J. & Mooney, R. Neural dynamics underlying birdsong practice and performance. *Nature* **599**, 635–639 (2021).
79. Brudner, S., Pearson, J. & Mooney, R. Generative models of birdsong learning link circadian fluctuations in song variability to changes in performance. *PLOS Computational Biology* **19**, e1011051 (2023).
80. Ruda, K., Rudzite, A. M. & Field, G. D. The functional organization of retinal ganglion cell receptive fields across light levels. *bioRxiv*, 2022–09 (2022).
81. Draelos, A., Nikitchenko, M., Sriworarat, C., Sprague, D., Loring, M. D., Pnevmatikakis, E., Giovannucci, A., Naumann, E. A. & Pearson, J. M. *improv*: A flexible software platform for adaptive neuroscience experiments. *bioRxiv* (2021).
82. Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243* (2019).
83. Doi, E., Gauthier, J. L., Field, G. D., Shlens, J., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., Mathieson, K., Gunning, D. E., et al. Efficient coding of spatial information in the primate retina. *Journal of Neuroscience* **32**, 16256–16264 (2012).
84. Cafaro, J., Zylberberg, J. & Field, G. D. Global motion processing by populations of direction-selective retinal ganglion cells. *Journal of Neuroscience* **40**, 5807–5819 (2020).
85. Duncker, L., Ruda, K. M., Field, G. D. & Pillow, J. W. Scalable Variational Inference for Low-Rank Spatiotemporal Receptive Fields. *Neural computation* **35**, 995–1027 (2023).
86. Dyballa, L., Rudzite, A. M., Hoseini, M. S., Thapa, M., Stryker, M. P., Field, G. D. & Zucker, S. W. Population encoding of stimulus features along the visual hierarchy. *Proceedings of the National Academy of Sciences* **121**, e2317773121 (2024).
87. Mohl, J. T., Pearson, J. M. & Groh, J. M. Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. *Journal of Neurophysiology* **124**, 715–727 (2020).
88. Yoo, S. B. M., Hayden, B. Y. & Pearson, J. M. Continuous decisions. *Philosophical Transactions of the Royal Society B* **376**, 20190664 (2021).
89. Subramanian, D., Pearson, J. M. & Sommer, M. A. Bayesian and Discriminative Models for Active Visual Perception across Saccades. *eneuro* **10** (2023).