Inference of Nonlinear Spatial Subunits by Spike-Triggered Clustering in Primate Retina

Nishal P. Shah[1], Nora Brackbill[2], Colleen Rhoades[3], Alexandra Tikidji-Hamburyan[4, 6, 7], Georges Goetz[4,6 7], Alan Litke[5], Alexander Sher[5], Eero P. Simoncelli[8,9], E.J. Chichilnisky[4,6,7]

1. Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
2. Department of Physics, Stanford University, Stanford, CA 94305, USA
3. Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
4. Department of Neurosurgery, Stanford University, Stanford, CA 94305, USA
5. Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz, CA 95064, USA
6. Department of Ophthalmology Stanford University, Stanford, CA 94305, USA
7. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA
8. Center for Neural Science, New York University, New York, NY 10003, USA
9. Howard Hughes Medical Institute, Chevy Chase, MD 20825, USA

Abstract

Integration of rectified synaptic inputs is a widespread nonlinear motif in sensory neuroscience. We present a novel method for maximum likelihood estimation of nonlinear subunits by soft-clustering spike-triggered stimuli. Subunits estimated from parasol ganglion cells recorded in macaque retina partitioned the receptive field into compact regions, likely representing bipolar cell inputs. Joint clustering with multiple RGCs revealed shared subunits in neighboring cells, producing a parsimonious population model. Closed-loop subunit validation was then performed by projecting white noise into the null space of the linear receptive field. Responses to these null stimuli were more accurately explained by a model with multiple subunits, and were stronger in OFF cells than ON cells. Presentation of natural stimuli containing jittering edges and textures also revealed greater response prediction accuracy with the subunit model. Finally, the generality of the approach was demonstrated by application to V1 data.

Introduction

Simple models of neural response, along with methods for fitting and testing them with experimental data, have proven essential for understanding neural circuits. An important example is the use of linear and linear-nonlinear cascade models, which, despite their simplicity, have been central in elucidating the properties of the visual system (Hubel & Wiesel, 1962; Jones & Palmer, 1987) and, in particular, retinal ganglion cells (RGCs). Purely linear models provided the first quantitative predictions of image representation and functional cell type subdivisions in the retina, using harmonic stimuli and parameter estimation approaches from the theory of shift-invariant linear systems (Enroth-Cugell & Robson 1966; Enroth-Cugell & Pinto,

1970; Movshon et al., 1978). Subsequently, linear-nonlinear-Poisson (LNP) and related cascade models were introduced to account for nonlinear spike generation not captured by linear models. Along with white noise visual stimuli (Marmarelis & Naka, 1972; Korenberg & Hunter 1986, Hunter and Korenberg 1986, Korenberg et al. 1989, Sakai 1992) and algorithms for parameter estimation (Chichilnisky, 2001; Arcas & Fairhall, 2003; Pillow et al., 2008; Keat et al., 2001), these models permitted more precise characterization of different types of RGCs and many of their properties (Field et al., 2007; Field et al., 2010; Chander & Chichilnisky, 2001; Baccus and Meister, 2002; Nirenberg & Meister, 1997).

However, as understanding of neural computations has evolved, the limitations of LNP and related models have also become evident, revealing a need for improved models and associated tools for parameter estimation. An important example is the identification of spatial subunits within the receptive fields of RGCs (Hochstein & Shapley 1976, Demb et al., 2001), which are thought to reflect bipolar cells outputs that are rectified and summed to drive RGC light response. This kind of nonlinear spatial computation cannot be captured by LNP models. In the retina, many studies have revealed the importance of subunits in producing responses to fine textures, motion and natural scenes (Ölveczky et al., 2003; Gollisch and Meister, 2008; Schwartz & Rieke 2011; Schwartz & Rieke 2012; Turner & Rieke 2016; Freeman et al., 2015; Real et al., 2017), as well as some of the underlying mechanisms (Demb et al., 2001; Kuo et al., 2016). Similarly, complex cells in V1 are thought to combine rectified inputs of simple cells (Hubel & Wiesel, 1959, 1962; Adelson & Bergen, 1985; Vintch et al., 2015), producing position-invariant responses to stimulus features such as orientation and spatial frequency.

The approaches previously used to estimate spatial nonlinearities such as these from extracellular recordings are either inefficient or strongly constrained. In one approach, a subspace of the high dimensional stimulus space is identified that captures as much as possible of the nonlinear light response, and this subspace may then be associated with the action of subunits (Paninski 2003; Sharpee et al., 2004; Rust et al., 2005; Schwartz et al., 2006; Pillow et al., 2006; Rajan & Bialek, 2013; Park et al, 2013; Theis et al,2013; Liu et al., 2017). This approach involves relatively few assumptions, but requires large data sets, and the axes of the identified subspaces generally do not correspond to underlying biological mechanisms. A second major approach relies on convolutional assumptions – that all subunits are of identical form and organized on a spatial grid – in order to reduce data requirements (Vintch et al., 2015; Wu et al., 2015; McIntosh et al, 2016). This approach fails to capture biological diversity in subunits arising from variation in underlying circuit connectivity (Freeman et al., 2015). More recently, nonlinear subunits have been estimated by fitting flexible, high-dimensional cascaded response prediction models to neural data (Maheswaranathan et al., 2018; Shi et al., 2018), but such models have not been used to examine the spatial structure of multiple subunits that compose the receptive field. Finally, a spike-triggered matrix factorization approach has recently been developed (Liu et al., 2017), but this approach bears an uncertain relationship to models of neural response and may not be the most efficient or unbiased way to uncover the underlying biological mechanisms.

Here we present a robust, efficient approach based on *spike-triggered clustering*, a generalization of efficient spike-triggered methods that have been used for fitting linear and LN models. The method reveals subunit structure directly, makes few assumptions about subunit structure or organization, and has less demanding data requirements than more general methods. We apply the approach to multi-electrode recordings from parasol RGCs in macaque retina, where it reveals a gradual partitioning of the receptive field with a hierarchical organization of spatially localized and regularly spaced subunits, consistent with the input from a mosaic of bipolar cells. Motivated by this observation, we develop a novel prior for estimating such spatially localized subunits when data are limited. We extend the approach to provide a parsimonious model for RGC populations with shared subunits. A novel closed-loop stimulus, designed to produce no response in neurons with linear response properties, generated strong RGC responses that were explained substantially more accurately with the subunit model, and revealed more nonlinear responses in OFF compared to ON cells. The subunit model also explained structure in RGC responses to natural images not predicted by linear models, and when applied to data from primary visual cortex, revealed subunits with expected spatiotemporal structure, suggesting that the approach will generalize to other contexts and neural circuits.

Results

Our goal is to develop a general model for subunit computations in neural responses along with a methodology for parameter estimation, to test whether this model can predict responses over a wide range of stimuli, and to investigate the spatial properties of the estimated subunits.

Subunit response model and parameter estimation

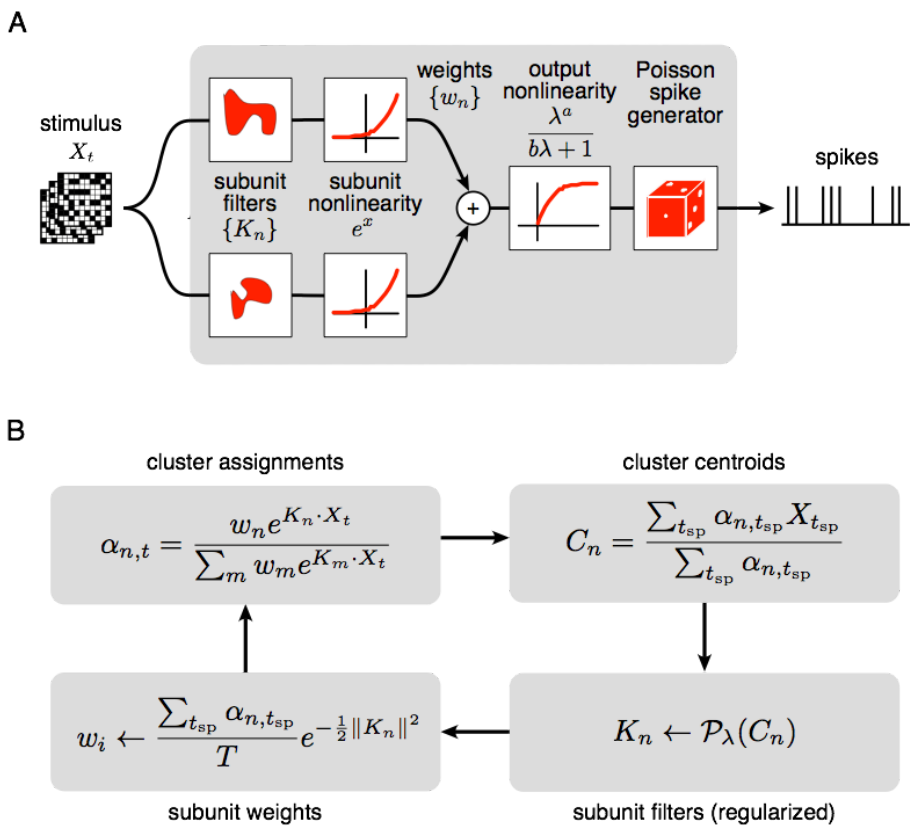

Figure 1: Spiking response model, and estimation through spike-triggered stimulus clustering. (A) The model is constructed as a cascade of two linear-nonlinear (LN) stages. In the first stage, subunit activations are computed by linearly filtering the stimulus (X_t) with kernels (K_n) followed by an exponential nonlinearity. In the second stage, a sum of subunit activations, weighted by (w_n), is passed through a saturating output nonlinearity (g), yielding a firing rate that drives an inhomogeneous Poisson spike generator. (B) Iterative fitting algorithm, partitioned into four steps. The subunit kernels (K_n) and weights (w_n) are randomly initialized, and used to compute soft cluster assignments ($\alpha_{n,t}$ - upper left), followed by cluster centroid computation (C_n - upper right), estimation of subunit kernels (K_n - lower right), and subunit weights (w_n - lower left). The summations are only over times when the cell generated a spike.

In the subunit model, RGC light responses are described as an alternating cascade of two

linear-nonlinear (LN) stages (Figure 1A). The visual stimulus (X_t) is defined as the image intensities over space and time prior to time t. In the first LN stage, the inner product of the stimulus with each of the linear subunit filters (K_n) is computed, followed by an exponential nonlinearity. In the second LN stage, the subunit outputs are weighted by non-negative scalars (w_n), summed, and passed through a final output nonlinearity with a flexible form: $g(x) = x^a / (bx+1)$. This yields the neuron's firing rate $R = g(\Sigma w_n \exp(K_n \bullet X_t))$, which drives a Poisson spike generator.

To predict neural responses using the model, the parameters (K_n, w_n, a, b) must be estimated from recorded data. The likelihood is not a concave function of the parameters, making the estimation problem difficult. However, the parameters can be efficiently estimated by iteratively alternating between two operations (see Methods): (1) estimate the subunits by clustering the spike-triggered stimuli, and (2) estimate the parameters of the output nonlinearity. In the first step, subunit filters are estimated by iteratively updating the soft-max weights of different subunits for each spike-triggered stimulus and then updating the cluster centers by a weighted average of spike-triggered stimuli (Figure 1B). In the second step, the parameters for output nonlinearity g(•) are estimated by maximizing the likelihood of the data.

Simulations show that this procedure yields accurate estimates of model parameters when sufficient data are available (Supplementary Figure S1). When experimental data are limited, reliable subunit estimates can be attained by regularizing the likelihood objective to capture prior knowledge about subunits. This leads to an augmentation of the fitting algorithm, incorporating a 'proximal' projection step that biases the subunit estimates to be more consistent with the prior information. For example, a sparsity inducing Laplacian prior adds an L1 regularization term to the objective function, which augments the fitting algorithm with a step in which a soft-thresholding operation is applied to the cluster centroids (Figure 1B). As discussed below (Figure 3), incorporating such priors can substantially improve the accuracy of estimated subunits.

Note that the model includes two hyperparameters - N and γ - that control the number of subunits and regularization strength, respectively. These can be explored systematically, or set by optimizing model predictions on held-out data (i.e., cross-validation).

Estimated subunits are spatially localized and non-overlapping

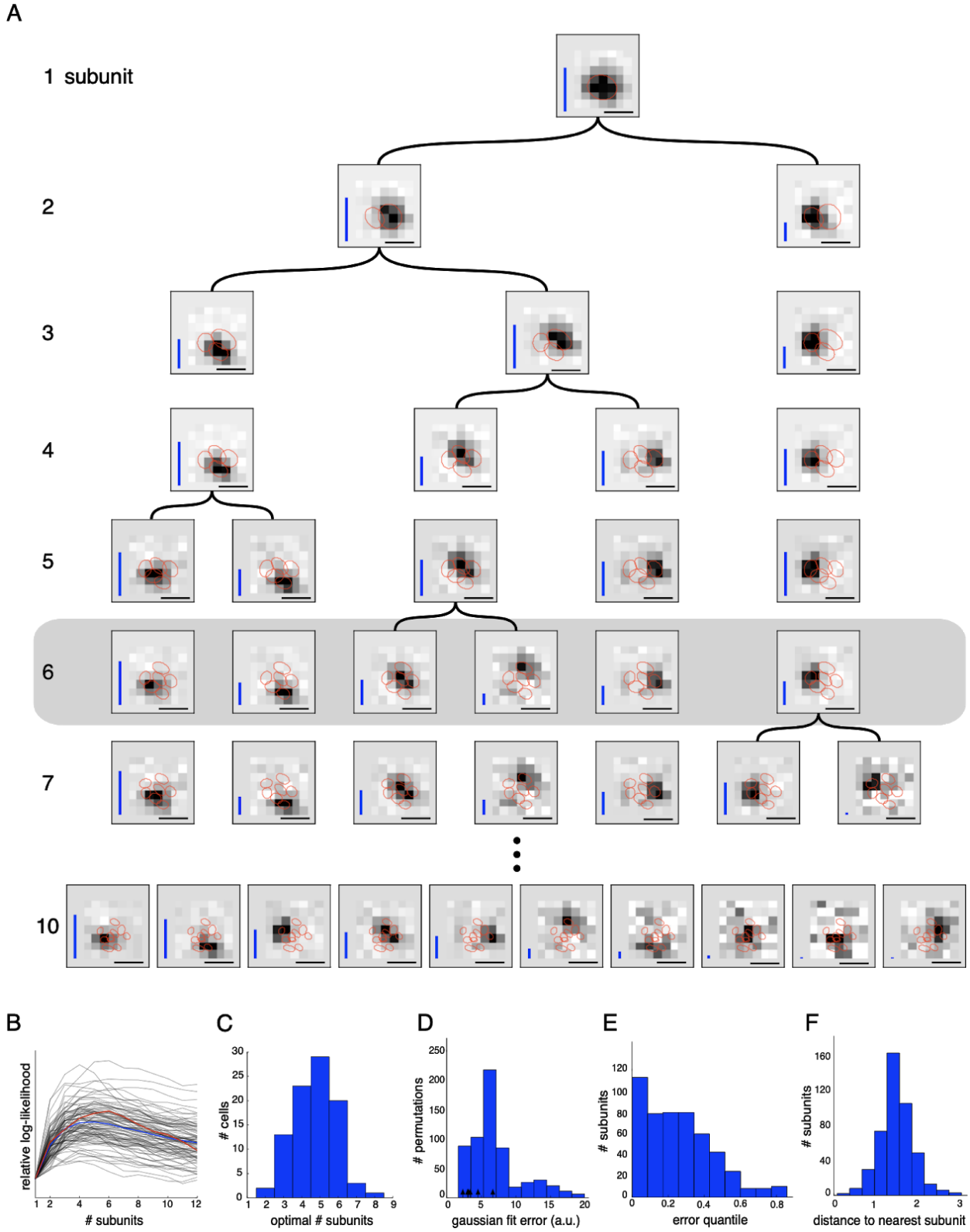

Figure 2: Estimated subunit properties. (A) Subunits, shown as grayscale images, estimated from OFF parasol cell responses to 24 min of white noise. Each pixel was temporally prefiltered with a kernel derived from the spike-triggered average (STA). Rows show estimated spatial subunits for successive values of N. Subunit locations are indicated with red ellipses, corresponding to the contour of a fitted two-dimensional Gaussian with standard deviation equal to the average nearest neighbor separation between subunits. As N increases, each successive set of subunits may be (approximately) described as resulting from splitting one subunit into two (indicated by arrows). Large N (e.g.,, last row) yields some subunits that are noisy or overlap substantially with each other. Height of vertical blue bars indicate the relative strength (average contribution to response over stimulus ensemble) of each subunit. Horizontal black bars indicate spatial scale (150μm). (B) Log-likelihood as a function of number of subunits (relative to single subunit model) for 91 OFF parasol cells (black) on 3 min of held-out test data, averaged across 10 random initializations of the model, from a distinct randomly sampled training data (24 min from remaining 27 min of data). Population average is shown in blue and the example cell from (A) is shown in red. (C) Distribution of optimal number of subunits across different cells, as determined by cross-validated log-likelihood on a held-out test set for OFF parasol cells. (D, E) Spatial locality of OFF parasol subunits, as measured by mean-squared error of 2D gaussian fits to subunits after normalizing with the maximum weight over space. Control subunits are generated by randomly permuting pixel weights for different subunits within a cell. For this analysis, the optimal number of subunits was chosen for each cell. (D) Distribution of MSE values for randomly permuted subunits for the cell shown in (A). MSE of 6 (optimal N) estimated subunits indicated with black arrows. (E) Distribution of quantiles of estimated OFF parasol subunits, relative to the distribution of MSE values for permuted subunits, across all cells and subunits. Null hypothesis has uniform distribution between 0-1. (F) Distribution of distances to nearest neighboring subunit within each OFF parasol cell. Distances are normalized by geometric mean of standard deviation of the gaussian fits along the line joining center of subunits. For this analysis, each cell is fit with 5 subunits (most frequent optimum from (C)).

To test the subunit model and estimation procedure on primate RGCs, light responses were obtained using large-scale multielectrode recordings from isolated macaque retina (Litke et al., 2004, Frechette et al., 2005). Spiking responses of hundreds of RGCs to a spatiotemporal white noise stimulus were used to classify distinct cell types, by examining properties of the spike-triggered average stimulus (STA), which provides a linear summary of light response (Frechette et al, 2005; Chichilnisky et al., 2002; Field et al; 2007). Complete populations of ON and OFF parasol RGCs covering the recorded region were examined further.

Fitting a subunit model reliably to these data was aided by decoupling the spatial and temporal properties of subunits. Specifically, although the fitting approach in principle permits estimation of full spatiotemporal subunit filters, the high dimensionality of these filters requires substantial data (and thus, long recordings). The parameter space was reduced by assuming that subunit responses are spatio-temporally separable, and that all subunits have the same temporal filter (consistent with known properties of retinal bipolar cells that are thought to form the RGC subunits). Given these assumptions, the temporal filter was estimated from the STA (see Methods). The stimulus was then convolved with this time course to produce an instantaneous spatial stimulus associated with each spike time. This temporally prefiltered spatial stimulus was then used to fit a model with purely spatial subunits.

The number of subunits, N, is a discrete parameter, and we examined its effect by fitting the model for different N, each with an independent initialization. Setting N=1 yielded a single subunit whose receptive field corresponds to a LN model. Typically, a model with two subunits partitioned this receptive field into spatially localized regions (Figure 2A, second row). Fitting the model with 3 or more subunits (Figure 2A, subsequent rows) typically caused one of the subunits from the preceding row to be partitioned further, while other subunits were largely unchanged. In principle this procedure could be repeated many times to capture all subunits in the receptive field. However, the number of model parameters increases with N, and the estimation accuracy decreases, with estimated subunits becoming noisier with larger overlap (Figure 2A, last row). This observation suggests an estimation approach in which the hierarchical introduction of subunits is built into the procedure, with the potential for greater efficiency (see Methods and Figure S2).

An optimal number of subunits was chosen for each cell as the N that maximized the cross-validated likelihood (i.e., the likelihood measured on a held out test set - Figure 2B). The typical optimum for OFF cells was 4-6 subunits (Figure 2C), a value that is governed by the true number of subunits, the resolution of the stimulus (which governs the dimensionality of the parameter space, see Figure 3), and the amount of data (number of spikes). Since the ON parasol cells had smaller optimum number of subunits (1-2) and much smaller increases in likelihood of the subunit model compared to an LN model (not shown), we focus subsequent analysis only on the OFF parasol cells.

The estimated subunits were larger and fewer in number than expected from the size and number of bipolar cells providing input to parasol RGCs at the eccentricities recorded (Jacoby et al., 2000; Schwartz & Rieke 2011; Tsukamoto & Omi 2015). However, two other features of estimated subunits suggested a relationship to the collection of bipolar cells contributing to the receptive field. First, estimated subunits were compact in space as shown in Figure 2A. This was quantified by comparing each subunit with the collection of subunits derived by randomly permuting the filter values at each pixel location across different subunits within a cell. Spatial locality for each subunit was evaluated by the mean-squared error (MSE) of 2D Gaussian fits. Compared to the permuted subunits, the estimated subunits had substantially lower error (Figure 2D, E). Second, the subunits “tiled” the RGC receptive field, in that they covered it with only small amounts of overlap, while leaving no holes. This was quantified by noting that on average, the neighboring subunits for a given cell were separated by ~1.5 times their width, with relatively little variation (Figure 2F).

Given these observations, a natural hypothesis for subunits observed with coarse stimulation is that they correspond to aggregates of neighboring bipolar cells. To test this intuition, the algorithm was applied to white noise responses generated from a simulated RGC with a realistic size and number of subunits in the receptive field center (Jacoby et al., 2000; Schwartz & Rieke 2011). For data simulated using coarse pixelated stimuli matched to those used in the experiments, the estimated subunits were fewer and larger than the bipolar cells, with each

subunit representing an aggregate of nearby underlying bipolar cells (Figure S1). The recoveredd subunits also exhibited spatial locality and tiling. Thus, the simulation supports the hypothesis that the subunits estimated from recorded data (Figure 2A) reflect aggregates of adjacent underlying bipolar cells.

Regularization for spatially localized subunit estimation

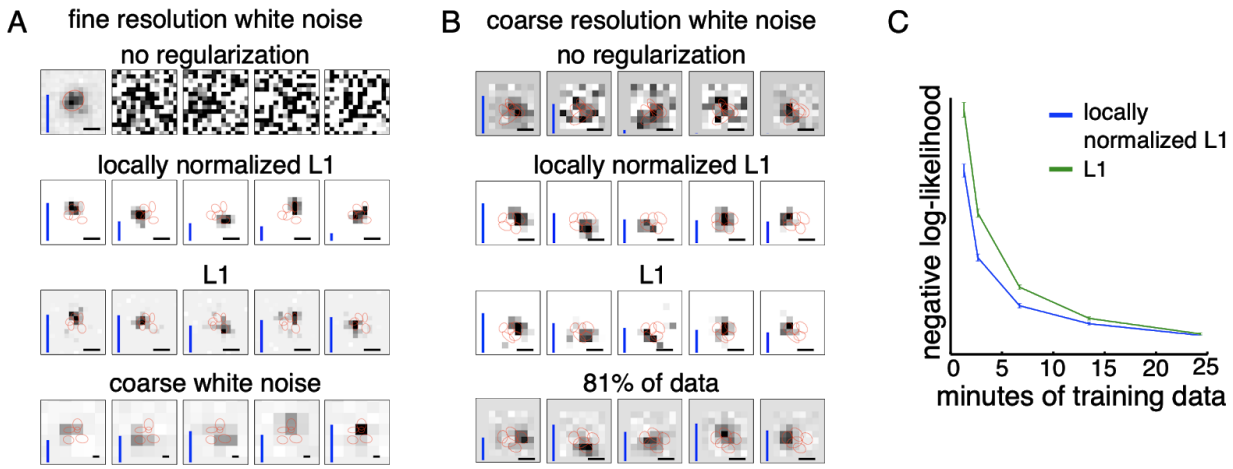

Figure 3: Spatially localized subunit estimation Comparison of different priors for estimating subunits using limited data. Examples of OFF parasol cells are shown. (A) Five subunits estimated using all data (1 hour 37 min) for fine resolution white noise without regularization (top row). The first estimated subunit is identical to the full receptive field and the others are dominated by noise. Locally normalized L1 (second row) and L1 (third row) regularization both give spatially localized subunits, with L1 regularization leaving more noisy pixels in background. In both cases, optimal regularization strength was chosen (from 0-1.8, steps of 0.1) based on performance on held-out validation data. The contours reveal interdigitation of subunits (red lines). Subunits estimated using white noise with 2.5x coarser resolution (24 min) and no regularization are larger, but at similar locations as subunits with fine resolution (bottom row). Scale bar: 75μm (B) For the cell in Figure 1A, 5 subunits estimated using the 3 min (10% of recorded data) of coarse resolution white noise responses are noisy and non-localized (top row). Similar to the fine case, using locally normalized L1 (second row) and L1 (third row) regularization both give spatially localized subunits, with L1 regularization subunits having noisier background pixels. The regularization strength (between 0 - 2.1, steps of 0.1) was chosen to maximize log-likelihood on held out data (last 5 min of data). Subunits estimated using 24 min (81% of data) of data are spatially localized and partition the receptive field (bottom row). Scale bars: 150μm (C) Held out log-likelihood for a 5 subunit model estimated from varying durations of training data with L1 (green) and locally normalized L1 (blue) regularization. Results averaged over 91 OFF parasol cells from Figure 1A.

To estimate subunits with a resolution approaching that of the underlying biology would require higher resolution stimuli. But higher resolution stimuli typically require longer duration of data. Specifically, to obtain a constant quality estimate of subunits as the stimulus resolution/dimensionality is increased requires a proportional increase in the number of spikes

(e.g. see Dayan & Abbott, 2001). Furthermore, higher resolution stimuli typically produce lower firing rates, because the effective stimulus contrast driving the photoreceptors is reduced. To illustrate the problem, the impact of increased resolution and smaller duration of data were examined separately. First, higher resolution stimuli led to poor subunit estimates, even when all the recorded data (97 minutes) were used (Figure 3A, top row): of the five estimated subunits, one resembled the receptive field, and the others were apparently dominated by noise and had low weights; thus, the algorithm effectively estimated a LN model. Second, for coarse resolution stimuli, using only 10% of the recorded data led to noisy subunit estimates (Figure 3B, top row) compared to the estimates obtained with 81% of data (24 minutes, Figure 3B, last row). These observations suggest that it would be useful to incorporate prior knowledge about subunit structure, and thus potentially obtain more accurate estimation with limited data.

To accomplish this, a prior on subunit filters was introduced, based on the observed spatial locality of estimated subunits (Figure 2). Previous studies (Liu et al., 2017; Maheswaranathan et al., 2018; MacFarland et al., 2013), have used L1 regularization (a common means of inducing sparsity) but this is agnostic to spatial locality (and indeed is invariant to spatial rearrangement). Instead of L1 regularization, we developed a novel locally normalized L1 (LNL1) regularizer that penalizes large weights, but only if all of the neighboring weights are small: $L(w_i) = \frac{|w_i|}{\varepsilon + \Sigma_{j \in neighbor(i)} |w_j|}$ for $\varepsilon = 0.01$. This penalty is designed to have a relatively modest influence on subunit size (compared to L1, which induces a strong preference for smaller subunits). The LNL1 penalty was incorporated using the projection operator in the clustering loop (Figure 1B). In this case, the appropriate projection operator is an extension of the soft-thresholding operator for the L1 regularizer, where the threshold for each pixel depends on the strength for neighboring pixels. The regularization strength is chosen to maximize cross-validated likelihood (see Methods).

The proposed prior improved subunit estimates for both of the limited data settings (high resolution, or small duration) presented above. In the case of the coarse resolution data, both L1 and LNL1 priors produced spatially localized subunits (Figure 3B, middle rows) whose locations matched the location of the subunits estimated using more data without regularization (Figure 3B, last row). This suggests that the proposed prior leads to efficient estimates without introducing new biases. Relative to L1 regularization, LNL1 regularization yielded fewer apparent noise pixels in the background (Figure 3B, middle rows). This improvement in spatial locality is reflected in more accurate response prediction with LNL1 regularization (Figure 3C).

In the case of high dimensional stimuli (Figure 3A, top row), both L1 and LNL1 priors were successful in estimating compact subunits (Figure 3A, middle rows) and matched the location of subunits obtained with a coarser stimulus (Figure 3A, bottom row). The estimated subunits tiled the receptive field, as would be expected from an underlying mosaic of bipolar cells. The LNL1 prior also led to subunit estimates with fewer spurious pixels compared to the L1 prior. Note that although these spurious pixels can also be suppressed by using a larger weight on the L1 prior, this approach led to smaller subunit size, which produced less accurate response predictions,

while LNL1 showed a much smaller dependence of subunit size (not shown). Hence, in both the conditions of limited data examined, the novel LNL1 prior yielded spatially localized subunits, fewer spurious pixels, and a net improvement in response prediction accuracy.

Parsimonious modeling of population responses using shared subunits

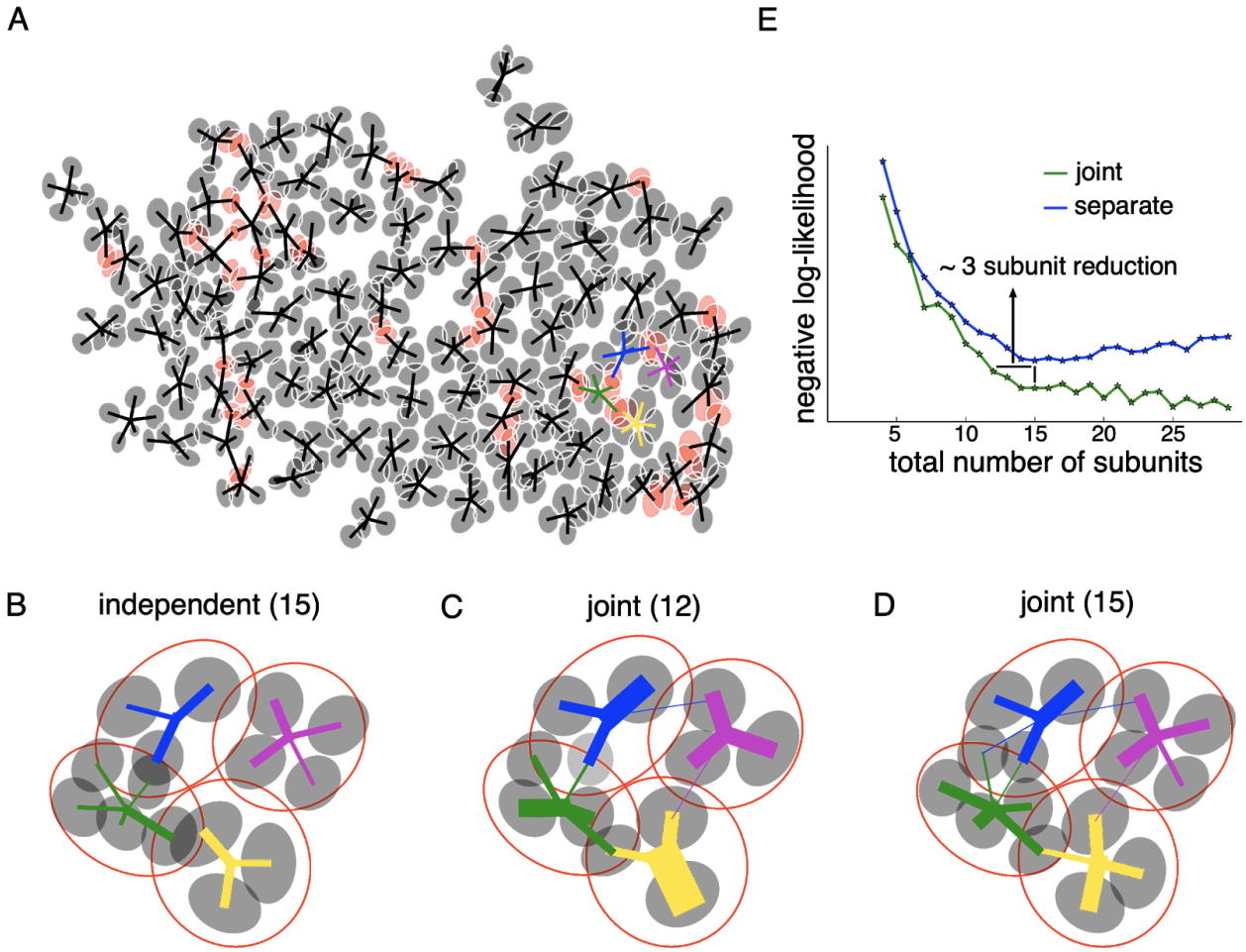

Figure 4: Joint estimation of subunits across multiple nearby cells. (A) Gaussian fits to the subunits estimated for an entire OFF parasol cell population (five subunits per cell, with poorly estimated subunits removed). Lines connect center of each cell to its subunits. Subunits which are closer to its neighbor than average (below 15 percentile) are indicated in red. (B) Different number of subunits for each cell is chosen (total 15 subunits) to give the highest summation of log-likelihood for four nearby cells. Gaussian fits to receptive field of the cells (red) and their subunits (gray). Connection strength from a cell (distinct color) to its subunits is indicated by thickness of line. (C) A common bank of 12 subunits estimated by jointly fitting responses for all four cells gives similar accuracy as (B). Sharing is indicated by subunits connected with lines of different colors (different cells) (D) A model with a common bank of 15 subunits (same number as B) gives better performance than than estimating the subunits for each cell separately. (E) The total negative log-likelihood on test data for four cells (y-axis) v/s total number of subunits (x-axis) for the two population models with subunits estimated jointly across cells (green) and best combination of separately estimated subunits (blue). Horizontal shift between curves indicates the reduction in total

number of subunits by jointly estimating subunits for nearby cells for similar prediction accuracy. The vertical shift indicates better performance by sharing subunits, for a fixed total number of subunits. In each case, the best value for locally normalized L1 regularization (in 0-3, step size 0.2) for a fixed number of subunits was chosen based on the performance on a separate validation dataset (see Methods).

To fully understand the visual processing in retina it is necessary to understand how a population of RGCs coordinate to encode the visual stimulus in their joint activity. The subunit estimation method extends naturally to joint identification of subunits in populations of neurons. Since neighboring OFF parasol cells have partially overlapping receptive fields (Gauthier et al., 2009), and sample a mosaic of bipolar cells, neighboring cells would be expected to receive input from common subunits. Indeed, in recorded OFF parasol mosaics, estimated subunits were frequently closer together than expected from the distribution of nearest-neighbor subunit distances obtained from single-cell data. For example, in one data set, the fraction of nearby subunits from different cells (spaced by less than 1SD of a Gaussian fit) was 14%, substantially higher than the predicted value of 9% based on the observed separation of subunits within a cell (Figure 2F). Examination of the mosaic suggests that these pairs of closely-spaced subunits may in fact correspond to single subunits shared between neighboring RGCs (Figure 4A). This raises the possibility that the estimated subunits of neighboring cells often reflect the same underlying biological substrate, and that joint estimation of subunits from the pooled data of neighboring RGCs may more efficiently reveal their collective spatial structure.

This hypothesis was tested by estimating a common “bank” of subunits to simultaneously predict the responses for multiple nearby cells. Specifically, the firing rate for the ith neuron was modeled as: $R_i = g_i(\Sigma w_{n,i} exp(K_n \bullet X_t))$. Here X_t is the stimulus, K_n are the filters for the common population of subunits, g_i is the output nonlinearity for the ith neuron,and $w_{n,i}$ is the non-negative weight of subunit n to the ith neuron, which is 0 if they are not connected. The model parameters were estimated by maximizing the summation of log-likelihood across cells, again in two steps. For the first step, the output nonlinearities were ignored for each cell and a common set of subunit filters were found, that clustered the spike-triggered stimuli for all the cells simultaneously. Specifically on each iteration, a) the relative weights for different subunits were computed for each spike-triggered stimulus and b) the cluster centers were updated using the weighted sum of spike-triggered stimuli across different cells (see Methods). Since the subunit filters could span the receptive fields of all the cells, leading to higher dimensionality, the spatial locality (LNL1) prior was used to efficiently estimate the subunits (Figure 3). In the second step of the algorithm, the non-linearities were then estimated for each cell independently (similar to Figure 1).

To quantify the advantages of joint subunit estimation, the method was applied to a group of 4 nearby OFF parasol cells. First, models with different number of subunits were estimated for each cell separately, and the combination of 15 subunits that maximized the cross-validated log-likelihood when summed over cells was chosen (Figure 4B). However, similar test log-likelihood was observed by using a joint model with only 12 subunits, effectively reducing model complexity by three subunits (Figure 4C). Examination of the subunit locations revealed

that pairs of overlapping subunits from neighboring cells in separate fits were replaced with one shared subunit (Figure 4B, C). Also, a joint model with 15 subunits showed higher prediction accuracy than a model in which 15 subunits were estimated separately (Figure 4D).

This trend was studied for different number of subunits in the filter bank, by comparing the total negative log-likelihood for these four cells associated with the joint model (Figure 4E, green) and a model in which subunits were estimated for each cell independently (Figure 4E, blue). Joint fitting provided more accurate predictions than separate fits, when using a large (>12) total number of subunits. For a smaller number of subunits, a horizontal shift in the observed likelihood curves revealed the reduction in the number of subunits that was obtained with joint estimation, without a reduction in prediction accuracy. Overall, jointly estimating a common collection of subunits for nearby cells resulted in a more parsimonious explanation of population responses.

Subunit model explains spatial nonlinearities revealed by null stimulus

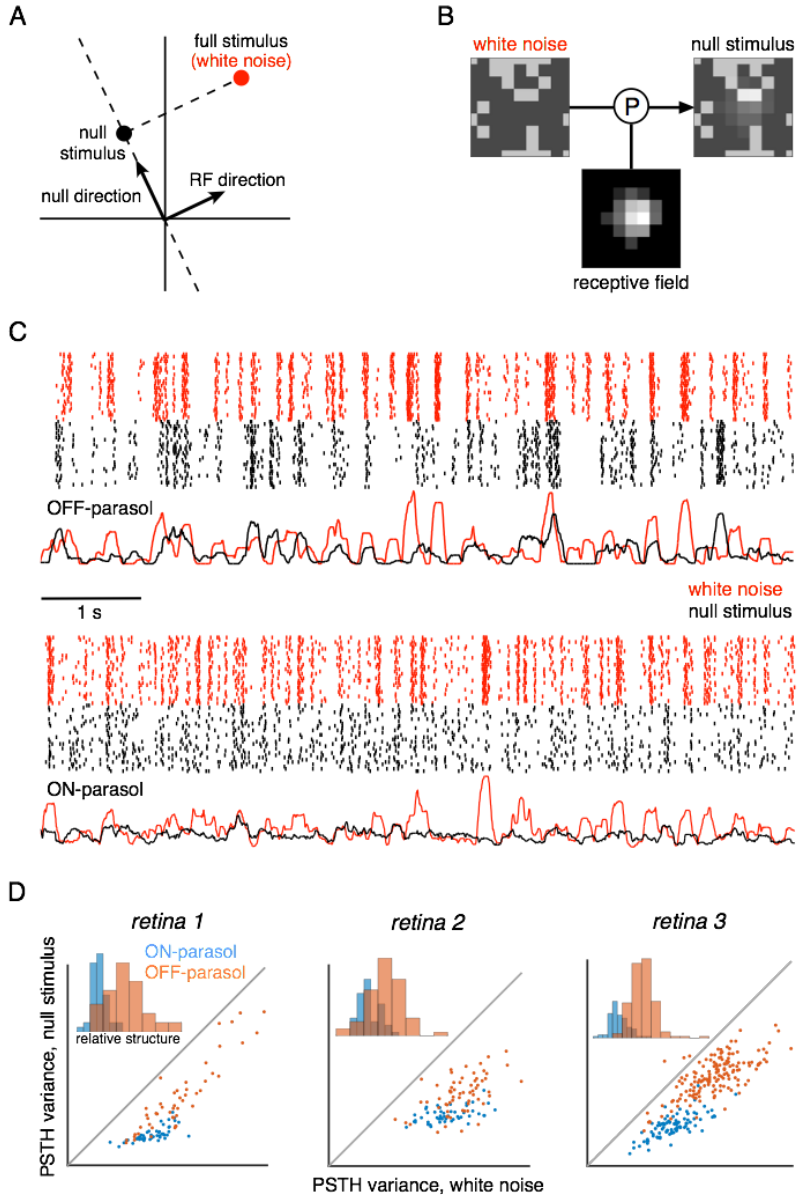

Figure 5: Cells respond to stimulus in null space of receptive field. (A) Construction of null stimulus, depicted in a 2-dimensional stimulus space. Each dimension of stimulus space consists of intensity along a particular pixel. A stimulus frame is represented as a point in this space. Each stimulus frame can be represented geometrically as the sum of the component along the receptive field (RF) direction (the response component of a linear nonlinear model) and the component orthogonal to the RF direction (the null component). (B) The null stimulus is constructed by projecting out the RF contribution from each frame of white noise stimulus. (C) Rasters representing responses for an OFF parasol (top) and ON parasol (bottom) cell to 30 repeats of 10s long white noise (red) and the corresponding null stimulus (black). (D) Response structure for ON parasol (blue) and OFF parasol (orange) populations for white noise (x-axis) and null stimulus (y-axis) across three preparations (different plots). The response structure was measured as variance of PSTH over time. Insets: Histogram of relative structure in white noise

responses that is preserved in null stimulus for ON parasol (blue) and OFF parasol (orange). Relative structure is measured by ratio of response structure in the null stimulus and response structure in white noise stimulus.

To more directly expose the advantage of a subunit model over commonly used models of RGC light response, experiments were performed using a visual stimulus that elicits no response in a LN model. Contrast-reversing spatial grating stimuli centered on the RF is an example of one such visual stimulus (Hochstein & Shapley 1976; Demb et al., 1999). However, to avoid recruiting additional nonlinearities, we wanted to use a stimulus with spatio-temporal statistics very similar to the diverse (high-entropy) white noise stimuli used for model characterization. Therefore, a stimulus was obtained by manipulating white noise so as to cancel the contribution of the linear RF.

Specifically, the computation performed by a hypothetical LN neuron can be represented in a stimulus space in which each axis represents stimulus contrast at a single pixel location (Figure 5A). In this representation, the LN response is determined by the projection of the stimulus vector onto the direction of the neuron's spatial RF. Thus, a "null stimulus" orthogonal to the RF can be computed by subtracting this projection from the original stimulus (Figure 5B). This null stimulus should yield zero response in a LN neuron. Note that although this approach is described above purely in the spatial domain, it generalizes to the full spatiotemporal RF, and the procedure based on the spatial RF alone will produce a null stimulus for a spatio-temporal linear filter that is space-time separable.

We tested whether this null stimulus could silence the response of RGCs, as follows. A 10-sec movie of white noise stimulus frames was projected into the intersection of the spatial null spaces corresponding to all cells of a single type. Responses were recorded from 30 repeated presentations of both the white noise and the null stimulus. RGC firing rates showed a modulation for both white noise and the null stimulus that was highly reproducible across repeats (Figure 6C, D), indicating that the visual inputs are not integrated linearly across the RF. Note that the null stimulus modulated OFF parasol cells more strongly than ON parasol cells, consistent with previous results obtained with white noise and natural scene stimuli (Chichilnisky & Kalmar 2002, Demb et al., 2001, Turner & Rieke 2016); henceforth, analysis is focused on OFF parasol cells.

The subunit model captured a substantial component of response nonlinearities revealed by the null stimulus. Subunit filters were estimated from responses to white noise (step 1 of model estimation, Figure 1), and response nonlinearities were estimated from responses to null stimulus (step 2 of model estimation, Figure 1). As expected, the single subunit (LN) model failed to capture responses to the held-out null stimulus (Figure 6A, second row). As the number of subunits was increased, the model's performance progressively increased (Figure 6A, last two rows). Prediction accuracy, defined as the correlation between the recorded and predicted firing rate, was evaluated across entire populations of OFF parasol cells in three recordings (Figure 6B). The single subunit model failed, with prediction accuracy near 0 (mean accuracy :

0.02, 0.12, 0.03 for three retinas). The prediction accuracy gradually increased with addition of more subunits, saturating at roughly 4-6 subunits (Figure 6B, red). In contrast, the addition of subunits showed marginal improvements for white noise stimuli, with signs of overfitting for large numbers of subunits (Figure 6B, black). Even though the log-likelihood increased for longer durations of a white noise stimulus (Figure 2B), the increase in correlation for repeats of shorter duration white noise was not significant. Thus, the response nonlinearities captured by the subunit model are more clearly exposed using the null stimulus.

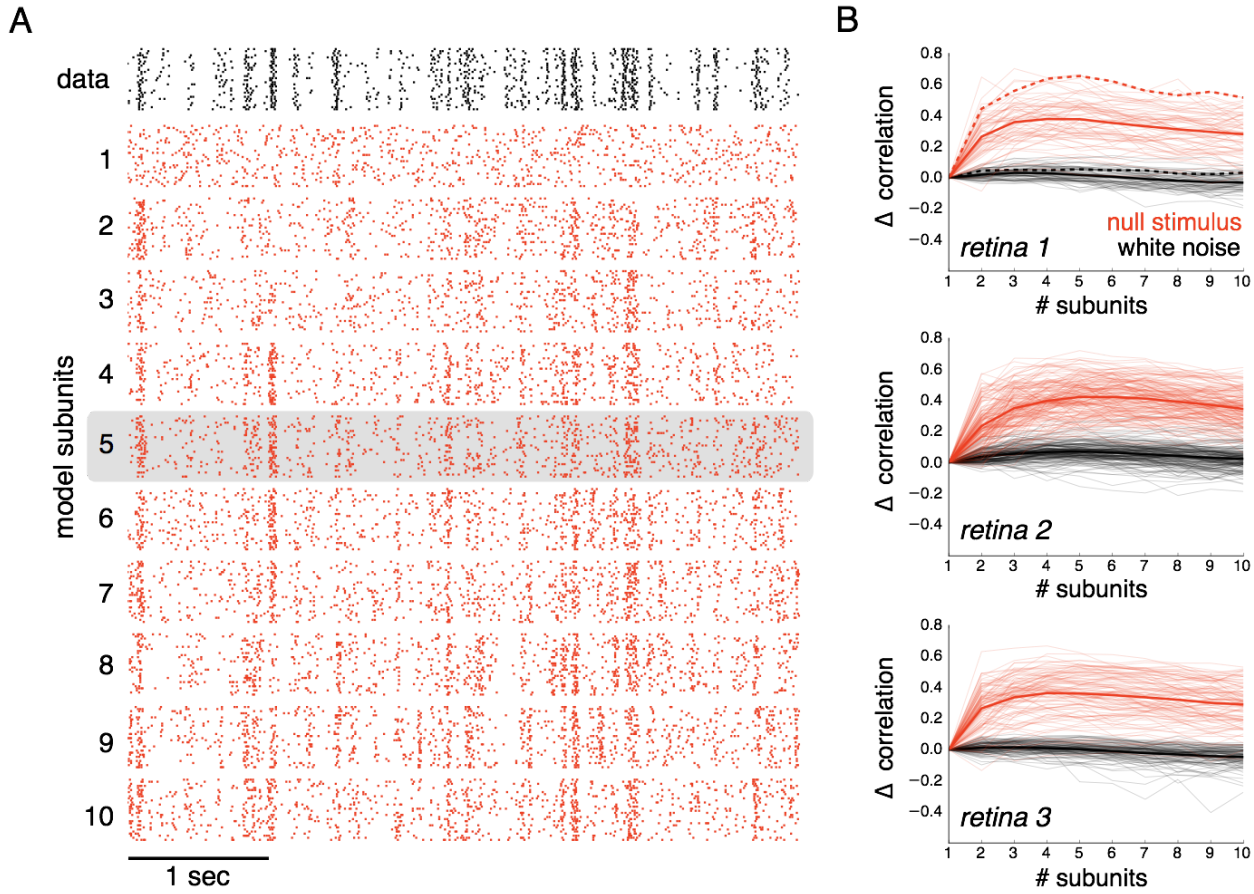

Figure 6: Subunits improve prediction of responses to null stimuli. (A) Rasters for recorded responses of an OFF parasol cell to 30 presentations of a 5 sec long null stimulus (top row, black). Predictions of models with increasing (1 to 10) number of subunits (subsequent rows, red). (B) The change in correlation between PSTH for recorded and predicted responses with different numbers of subunits across three preparations. Spatial kernels were estimated from 24 min of non-repeated white noise stimulus, with scales and output nonlinearity estimated from the first 5 sec of the repeated stimulus. Performance on the last 5 sec of the repeated stimulus was averaged over 10 fits, each with a random subsample of the non-repeated white noise stimulus. Individual OFF parasol cells (thin lines) and population average (thick lines) for the null stimulus (red) and white noise (black). The cell is (A) is indicated with dotted lines.

Subunits enhance response prediction for naturalistic stimuli

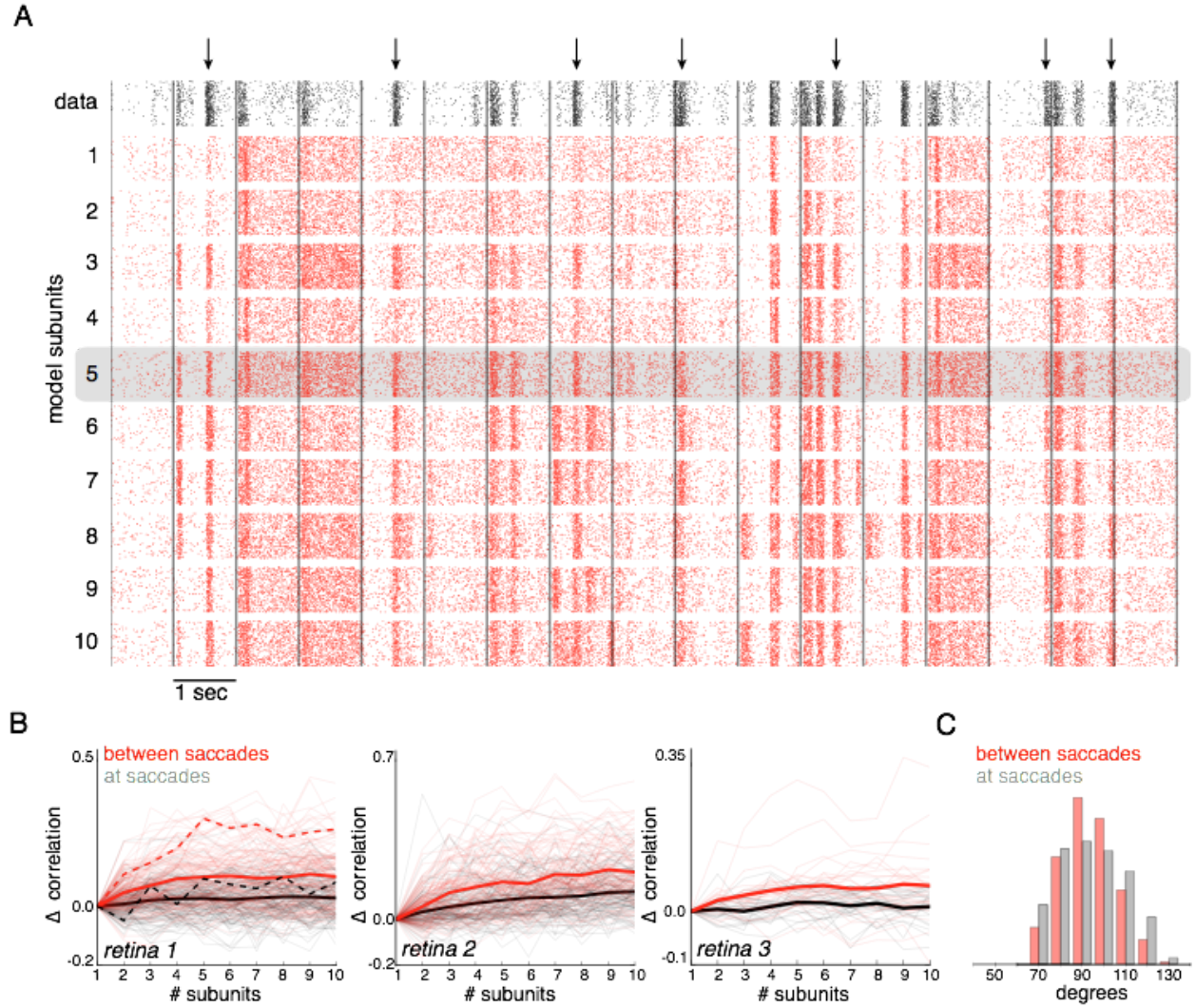

Figure 7: Subunits improve response prediction accuracy for NSEM. (A) Top row: Rasters of responses for an OFF parasol cell from 40 presentations for 30s long natural stimuli. Saccades to a new image every 1 sec (black lines). Subsequent rows indicate model predictions using different number of subunits (1 to 10) respectively. (B) Change in correlation between PSTH for recorded and predicted responses with different number of subunits across 3 preparations. Individual OFF parasol cells (thin lines) and population average (thick lines) for two conditions: 250 ms immediately following saccades (black) and inter-saccadic stimuli (red). Cell from (A) shown with dotted lines. (C) Distribution of angle between vectors representing the spatial receptive field and natural stimuli at saccades (black) and between saccades (red). Inter-saccadic stimulus shows more prevalence around 90 degrees (blue line).

A fuller understanding of the biological role of spatial nonlinearities requires examining how they influence responses elicited by natural stimuli. Images from the Van Hateren database

were shown for one second each, jittered according to the eye movement trajectories recorded during fixation by awake macaque monkeys (Z.M. Hafed and R.J. Krauzlis, personal communication; Heitmann et al, 2016, Van Hateren & van der Schaaf, 1998). The visual stimuli consisted of training data interleaved with identical repeats of testing data (see Methods). Subunit filters and time courses were first estimated using white noise stimuli (1st step, Figure 1), then the non-linearities and subunit magnitudes were fitted using natural stimuli (2nd step, Figure 1) (see Methods). Subsequently, the ability of the subunit model to capture responses to natural stimuli was measured.

Responses to natural stimuli showed a distinct structure, with strong periods of firing or silence immediately following saccades and occasional responses elicited by inter-saccadic eye-movements (Heitmann et al, 2016, Figure 7A black raster). This structure was progressively better captured by the model as the number of subunits increased, as shown by the rasters for an OFF parasol cell in Figure 7A. This finding was replicated in OFF parasol populations in three retinas, and quantified by the correlation between the recorded and predicted PSTH (Figure 7B). However, some of the firing events could not be captured even with a large number of subunits, suggesting additional mechanisms affecting natural scene responses.

Interestingly, further examination revealed that subunits led to greater improvement for responses elicited by inter-saccadic stimuli (small eye movements), compared to saccades (large eye movements) (Figure 7B). This could be explained by the fact that changes in luminance are more spatially uniform at saccades, resulting in more linear signaling by OFF parasol cells. Indeed, there were fewer stimuli in the null space of the receptive field (90 degrees) during saccades compared to between saccades (Figure 7C) and, the subunit model provides a more accurate explanation of responses to stimuli in the null space (Figure 6). Thus, these results indicate that spatial non-linearities explain the responses to natural stimuli, especially those elicited by jittering natural images.

Application to simple and complex cells in primary visual cortex

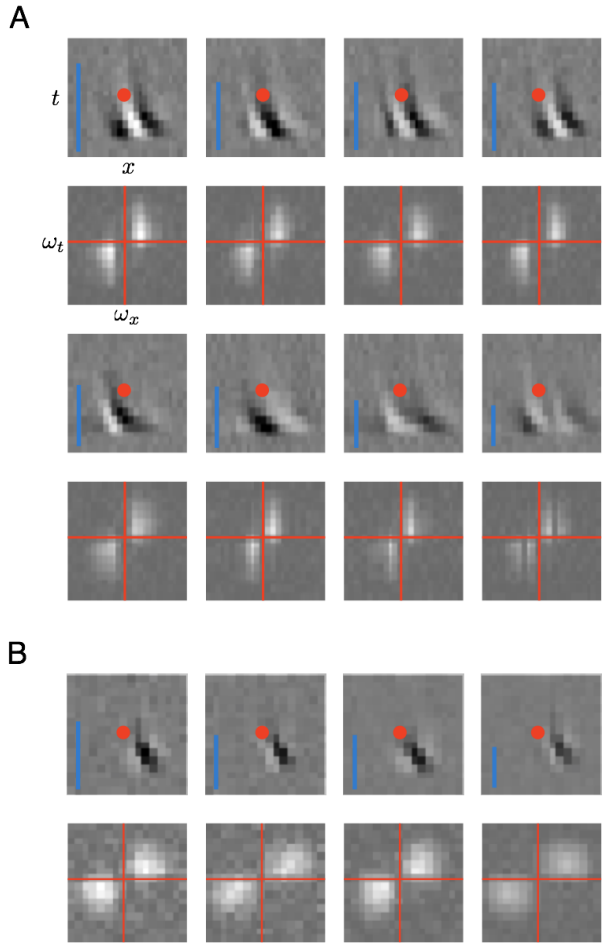

Figure 8: Application of subunit model to V1 (A) Estimated subunits using responses to flickering bar stimuli for the complex cell featured in Rust et al, 2005. Optimal number of subunits was found by cross-validation. Spatio-temporal filters (top row) and the fourier magnitude spectrum (bottom row) show translational invariance of estimated subunits. Red dots indicate center location for the spatio-temporal filters. Inset : Relative contribution to cell response for each subunit indicated by height of blue bar. (B) Same as (A) for the simple cell from Rust et al, 2005.

Identification of subunits using spike-triggered stimulus clustering should generalize to other neural systems with similar cascade of linear and non linear operations. We applied the unregularized method to data obtained from V1 simple and complex cells, responding to spatio-temporalyl oriented flickering bars (Rust et al, 2005). The number of cross validated subunits for the complex cell was greater than the simple cell (Figure 8), as reported previously (Rust et al., 2005). All subunits for a given cell had similar structure in space and time, but were

translated relative to each other (Figure 8 A, B, top rows), as confirmed by the very similar frequency spectrum (Figure 8 A,B, bottom rows). For the complex cell, this is consistent with the hypothesis that it receives inputs from multiple simple cells at different locations (Hubel & Wiesel, 1962; Adelson & Bergen 1985, Vintch et al., 2015). Thus, the method presented here extends to V1 data, producing subunit estimates that are not constrained by assumptions of orthogonality (Rust et al., 2005) or convolutional structure (Vintch et al., 2015), but broadly resemble the results of previous work.

Discussion

We developed an efficient modeling and estimation approach to study a ubiquitous motif of neural computation in sensory circuits: summation of respnoses of rectified subunits. The approach allowed efficient subunit estimation by clustering of spike-triggered stimuli, and the inclusion of priors for spatial localization to additionally constrain the fitting under conditions of limited data. The model and fitting procedure extended naturally to populations of cells, allowing joint fitting of shared subunits. We demonstrated the effectiveness of the method in capturing subunit computations in macaque parasol RGCs, specifically in stimulus conditions in which linear models completely fail, as well in V1 neurons. Below, we discuss the properties of the approach compared to others, and the implications of the application to retinal data.

Modeling linear-nonlinear cascades

The subunit model is a natural generalization of linear-nonlinear (LN) cascade models, and the estimation procedure a natural generalization of the spike-triggered estimation procedures that enabled their widespread use in modeling neural computations. In the present model framework, a maximum-likelihood estimate of model parameters leads to a spike-triggered clustering algorithm for obtaining the early linear filters that represent model subunits. These linear components are the key, high-dimensional entities that define stimulus selectivity, and, in the absence of a robust computational framework, would be difficult to estimate.

Related prior work (Liu et al., 2017) developed a spike-triggered non-negative matrix factorization (SNMF) method for estimating subunits, and applied it to recordings from salamander RGCs. The estimated subunits were shown to be well-matched to independent physiological measurements of bipolar cell receptive fields, an impressive validation of the method. In comparison to the method presented here, there are several technical issues worth noting. First, the assumptions of the SNMF algorithm are not directly matched to the structure of the subunit response model: the present model assumes nonnegative subunit outputs, while the SNMF factorization method assumes that the subunit filters themselves are nonnegative. The assumption of nonnegative filters is inconsistent with suppressive surrounds previously reported in retinal bipolar cells (Dacey et al., 2000; Fahey & Burkhardt 2003; Turner et al., 2018), and limits the application of the method to other systems such as V1 (Figure 8). However, a recent

modification of the SNMF method applies non-negativity constraints on subunit weights instead of the subunit filters (Jia et al., 2018), making it more suitable for other applications. Second, the SNMF method takes a sequence of additional steps to relate the estimated matrix factors to the parameters of an LNLN model. However in the present work, clustering of spike-triggered stimulus is shown to be equivalent to optimizing the prediction accuracy of a LNLN model, eliminating extra steps in relating model parameters to biology. Third, the comparison to bipolar cell receptive fields in the above studies relied on a regularization parameter that trades off sparseness and reconstruction error. This regularization can influence the size of the estimated subunits, and thus the comparison to bipolar receptive fields. In contrast, the present method includes a regularizer that encourages spatial contiguity of receptive fields, while exerting minimal influence on the size of estimated subunits, and is chosen using a cross-validation procedure.

Another study (Freeman et al., 2015) presented an approach to estimating subunits from OFF midget RGCs with visual inputs presented at single cone resolution, allowing a cellular resolution dissection of subunits. However, this approach assumed subunits that receive inputs from non-overlapping subsets of cones, which may not be an accurate assumption for some RGC types, including parasol cells. Moreover, the cone to subunit assignments were learned using a greedy procedure, which was effective for subunits composed of 1-2 cones, but is more likely to get stuck in local minima in more general conditions.

Recent studies fitted a large class of LNLN models to simulated RGC responses (McFarland et al., 2013), to recorded ON-OFF mouse RGC responses (Shi et al., 2018), and to flickering bar responses in salamander RGCs (Maheswaranathan et al., 2018) using standard gradient descent procedures. These approaches have the advantage of flexibility and simplicity. In contrast, we assume a specific subunit nonlinearity that leads to a specialized efficient optimization algorithm, and apply this to recorded data from primate RGCs. The choice of an exponential subunit non-linearity enables likelihood maximization for Gaussian stimuli, which leads to more accurate estimates with limited data, and reduces the computational cost of each iteration during fitting because the loss depends only on the collection of stimuli that elicit a spike (Ramirez et al., 2014). Additionally, the soft-clustering algorithm requires far fewer iterations than stochastic gradient descent algorithms used in fitting deep neural networks (Kingma et al, 2014; Duchi et al., 2011). Depending on the experimental conditions (e.g. type of stimuli, duration of recording, number of subunits to be estimated), the present approach may provide more accurate estimates of subunits, with less data, and more quickly.

Other recent studies applied convolutional and recurrent neural network models, fitted to natural scene responses, to mimic a striking variety of retinal response properties (McIntosh et al., 2016, Batty et al., 2016). The general structure of the model – convolutions and rectifying nonlinearities – resembles the subunit model used here. However, the number of layers and the interactions between channels do not have a direct correspondence to the known retinal architecture, and to date the specific properties of the inferred subunits, and their contributions

to the behaviors of the model, have not been elucidated. The complexity of the underlying models suggest that a comparison to the biology could be difficult.

In the context of V1 neurons, the present approach has advantages and disadvantages compared to previous methods. Spike-triggered covariance (STC) methods (Rust et al., 2005) estimate an orthogonal basis for the stimulus subspace captured by the subunits. However, the underlying subunits may not be orthogonal, and even if they are, are not guaranteed to align with the basis elements. The method also requires substantial data to obtain accurate estimates of the basis. To reduce the data requirements, other methods (Vintch et al, 2015; Wu et al., 2015) imposed an assumption of convolutional structure on the subunits. In comparison, the method presented here does not assume orthogonal or convolutional subunits. Instead, approximate translational invariance emerges from data, consistent with known physiological properties of V1 neurons.

A generalization of the STC approach (Sharpee et al., 2004; Paninski 2003) involves finding a subspace of stimulus space which is most informative about a neuron's response. This approach also yields a basis for the space spanned by subunits, but the particular axes of the space may or may not correspond to the subunits. Unlike the method presented here, this approach requires minimal assumptions about stimulus statistics, making it very flexible. However, methods based on information theoretic measures typically require much more data than are available in experimental recordings (Paninski 2003).

Although constraints on the structure of the subunit model (e.g. convolutional), and the estimation algorithm (e.g. spike-triggered covariance or clustering), both influence the efficiency of subunit estimation, the choice of priors used to estimate the subunits efficiently with limited data (i.e. regularization) is also important. In the present work, a regularizer was developed that imposes spatial contiguity, with minimal impact on the size of estimated subunits. A variety of other regularization schemes have been previously proposed for spatial contiguity for the estimation of V1 subunits (Park and Pillow, 2011). These could be incorporated in the estimation of the presented model using a corresponding projection operator.

As suggested by the results, another prior that could improve the subunit estimation is the hierarchical variant of the clustering approach (Figure S2). By splitting one of the subunits into two subunits at each step, models with different number of subunits can be estimated with greater efficiency, and preliminary exploration indicates that this may offer improvements in data efficiency and speed. However, potential tradeoffs emerging from the additional assumptions will require further study.

Revealing the impact of non-linear computations using null stimuli

The null stimulus methodology revealed the failure of LN models, and isolated the nonlinear behaviors arising from receptive field subunits. Previous studies have developed specialized stimuli to examine different aspects of the cascaded linear-nonlinear model. Contrast reversing gratings (CRGs) have been the most commonly used stimuli for identifying the existence of a spatial nonlinearity. Schwartz & Rieke 2012 extend this logic to more naturalistic stimuli, with changes in responses to translations and rotations of texture stimuli revealing the degree of spatial nonlinearity. Bölinger et al., 2012 developed a closed-loop experimental method to measure subunit nonlinearities, by tracing the intensities of oppositely signed stimuli on two halves of the receptive field to measure iso-response curves. Freeman et al., 2015 developed stimuli in closed loop that targeted pairs of distinct cones in the receptive field, and used these to identify linear and nonlinear summation in RGCs depending on whether the cones provided input to the same or different subunits.

The null stimulus approach introduced here offers a generalization of the preceding methods. Null stimuli can be generated starting from any desired stimulus ensemble (e.g., natural images), and the method does not assume any particular receptive field structure (e.g. radial symmetry). Construction of the null stimuli does, however, require fitting a linear (or LN) model to the data in closed loop. Projecting a diverse set of stimuli (e.g. white noise) onto the null space yields a stimulus ensemble with rich spatio-temporal structure that generates diverse neural responses. By construction, the null stimulus is the closest stimulus to the original stimulus that is orthogonal to the receptive field, and thus minimally alters other statistical properties of the stimulus that affect neural response. This property is useful for studying the effect of spatial nonlinearities in the context of specific stimuli, such as natural scenes. Whereas contrast reversing gratings do not reveal stronger nonlinearities in OFF parasols compared to ON parasols in responses to natural stimuli (Turner and Rieke 2017), the null stimulus captures these differences (Figure 6), highlighting the advantages of tailored stimulation with rich spatio-temporal responses.

Further applications and extensions

The assumptions behind the proposed subunit model and the application to primate RGC data enable efficient fitting and interpretability, but also lead to certain limitations that could potentially be overcome. First, the choice of exponential subunit nonlinearity improves efficiency by allowing the maximum expected likelihood approximation. Other nonlinearities that allow this approximation, such as rectified quadratic (Bölinger et al., 2012), could be explored. Second, the assumption of space-time separable RGC subunits significantly reduces the number of parameters that need to be estimated, but could be relaxed with more data. For example, a rank 2 approximation that allows for separate space-time separable center and surround filters for

each subunit may be useful (Schweitzer-Tong, Enroth-Cugell & Pinto, 1970), given the center-surround structure of bipolar cell receptive fields (Dacey et al., 2000, Turner et al., 2018).

The methods developed here may also prove useful in tackling additional challenging problems in neural circuitry and nonlinear coding, in the retina and beyond. First, applying the model to high-resolution visual stimulation of the retina could be used to reveal how individual cones connect to bipolar cells, the likely cellular substrate of subunits, in multiple RGC types (see Freeman et al., 2015). Second, the incorporation of noise in the model, along with estimation of shared subunits (Figure 4), could help to probe the origin of noise correlations in the firing of nearby RGCs. Third, extending the subunit model to include gain control, another ubiquitous nonlinear computation in retina and throughout the visual system, could provide a more complete description of the responses in diverse stimulus conditions (Heeger, 1992; Carandini & Heeger 2011). In addition to scientific advances, a more accurate functional model is critical for the development of artificial retinas for treating blindness. Finally, the successful application of the method to V1 data (Figure 8) suggests that the method could be effective in capturing computations in other neural circuits that share the nonlinear subunit motif.

Methods

Recordings

Detailed preparation and recording methods are described elsewhere (Litke et al., 2004; Frechette et al., 2005; Chichilnisky and Kalmar, 2002). Briefly, eyes were enucleated from 7 terminally anesthetized macaque monkeys (Macaca sp.) used by other experimenters in accordance with institutional guidelines for the care and use of animals. Immediately after enucleation, the anterior portion of the eye and the vitreous were removed in room light. The eye was stored in darkness in oxygenated Ames' solution (Sigma, St. Louis, MO) at 33°C, pH 7.4. Segments of isolated or RPE-attached peripheral retina (6-15mm temporal equivalent eccentricity (Chichilnisky and Kalmar, 2002), approximately 3mm x 3mm) were placed flat, RGC side down, on a planar array of extracellular microelectrodes. The array consisted of 512 electrodes in an isosceles triangular latticewith 60 µm inter-electrode spacing in each row and covering a rectangular region measuring 1800 µm x 900 µm. While recording, the retina was perfused with Ames' solution (35°C for isolated recordings and 33°C for RPE-attached recordings), bubbled with 95% O2 and 5% CO_2, pH 7.4. Voltage signals on each electrode were bandpass filtered, amplified, and digitized at 20 kHz.

A custom spike-sorting algorithm was used to identify spikes from different cells (Litke et al., 2004). Briefly, candidate spike events were detected using a threshold on each electrode, and voltage waveforms on the electrode and nearby electrodes in the 5ms period surrounding the time of the spike were extracted. Candidate neurons were identified by clustering the waveforms using a Gaussian mixture model. Candidate neurons were retained only if the assigned spikes exhibited a 1 ms refractory period and totaled more than 100 in 30 min of recording. Duplicate spike trains were identified by temporal cross-correlation and removed. Manual analysis was used to further select cells with a stable firing rate over the course of the experiment and with a spatially localized receptive field.

Visual Stimuli

Visual stimuli were delivered using the optically reduced image of a CRT monitor refreshing at 120 Hz and focused on the photoreceptor outer segments. The optical path passed through the mostly transparent electrode array and the retina. The relative emission spectrum of each display primary was measured with a spectroradiometer (PR-701, PhotoResearch) after passing through the optical elements between the display and the retina. The total power of each display primary was measured with a calibrated photodiode (UDT Instruments). The mean photoisomerization rates for the L, M, and S cones were estimated by computing the inner product of the primary power spectra with the spectral sensitivity of each cone type, and multiplying by the effective collecting area of primate cones (0.6 µm2 (Angueyra and Rieke, 2013; Schnapf et al., 1990). During white noise and null stimulus, the mean background illumination level resulted in photoisomerization rates of (4100,3900,1600) for the (L,M,S) cones. The pixel size was either 41.6 microns (8 monitor pixels on a side) or 20.8 microns (4 monitor pixels on a

side). A new white noise frame was drawn at refresh rates of 60 Hz (figure 2) or 30 Hz (Figure 5, 6). The pixel contrast (difference between the maximum and minimum intensities divided by the sum) was either 48% or 96% for each display primary, with mean contrast of 50%.

The details of natural stimuli are presented in Heitmann et al., 2016. Briefly, the stimuli consisted of images from the Van Hateren database, shown for one second each, jittered according to the eye movement trajectories recorded during fixation by awake macaque monkeys (Z.M. Hafed and R.J. Krauzlis, personal communication), (Heitmann et al, 2016, van Hateren and van der Schaaf, 1998). Image intensities produced mean photoisomerization rates of 1900, 1800, 700 for the L, M, S cones, respectively. The pixel width was 10.4 microns (2 monitor pixels on a side), and image frames refreshed at 120Hz. The training data consisting of 59 groups of 60 distinct natural scenes was interleaved with identical repeats of testing data consisting of 30 distinct natural scenes.

Subunit Estimation

The light responses of RGCs were modeled using a cascade of two linear-nonlinear (LN) stages, followed by a Poisson spike generator. Specifically, the instantaneous stimulus-conditional firing rate was modeled as $\lambda_t|X_t = g(\sum_n w_n \exp(K_n^T.X_t))$ where X_t is the visual stimulus at time t, K_n are the spatial subunit filters, w_n are non-negative weights on different subunits and $g(x) = \frac{x^a}{bx+1}$, $(b \geq 0)$ is the output nonlinearity.

The parameters $\{K_n, w_n, a, b\}$ are estimated by minimizing their negative log-likelihood given observed spiking responses Y_t to visual stimuli X_t, in two steps. In the first step, the parameters $\{K_n, w_n\}$ are estimated using a clustering procedure while ignoring the nonlinearity (g). In the second step, $\{g, w_n\}$ are estimated using gradient descent, with the K_n held fixed (up to a scale factor).

The negative log-likelihood for the first step can be simplified as follows:

$$\mathcal{L}(w_n, K_n) = \frac{1}{T}\sum_{t=1}^{t=T}\lambda_t - \frac{1}{T}\sum_{t=1}^{t=T} Y_t \log(\lambda_t) \tag{1}$$

$$\approx \mathbb{E}(\lambda) - \frac{1}{T}\sum_{t=1}^{t=T} Y_t \log(\lambda_t) \tag{2}$$

$$= \sum_{n=1}^{n=N} w_n e^{K_n^T K_n/2} - \frac{1}{T}\sum_{t=1}^{t=T} Y_t \log(\sum_{n=1}^{n=N} w_i e^{K_n^T X_t}) \tag{3}$$

$$\leq \sum_{n=1}^{n=N} w_i e^{K_n^T K_n/2} - \frac{1}{T}\sum_{t}\sum_{n=1}^{n=N} Y_t\ \alpha_{t,n}^0[(K_n - K_n^0)^T X_t + (\log(w_n) - \log(w_n^0))] + c(K_n^0, w_n^0)] \tag{4}$$

$$= \sum_{n=1}^{n=N}\left(w_i e^{K_n^T K_n/2} - \frac{1}{T}\sum_{t} Y_t\ \alpha_{t,n}^0[(K_n - K_n^0)^T X_t + (\log(w_n) - \log(w_n^0))] + c(K_n^0, w_n^0)\right) \tag{5}$$

where $\alpha_{t,n}^0 = \frac{w_n^0 e^{K_n^0 \cdot X_t}}{\sum_{m=1}^{m=N} w_j^0 e^{K_m^0 \cdot X_t}}$. Since the first term in the log-likelihood (Eq. 1) does not depend on recorded responses (Y_t), it is replaced with its expected value across stimuli (Eq. 2). Assuming the projections of the input distribution onto filters K_n are approximately Gaussian, this expectation can be computed using the moment generating function of a Gaussian distribution (Eq. 3). The second term only depends on the stimuli for which the cell generated spikes

($Y_t > 0$), reducing computational cost. Finally, to optimize the resulting non-convex function, a convex upper bound was minimized at each step. Specifically, the second term is replaced by a first-order Taylor approximation (Eq. 4) (the first term is already convex). Since the upper bound is tight at current parameter estimates, the successive convex optimization reduces the approximation of negative log-likelihood monotonically, leading to fewer iterations for convergence. This upper bound is separable across parameters of different subunits, as can be seen by re-arranging the summation over time and subunits (Eq. 5).

The minimization of the convex upper bound (Eq. 5) can accomplished using a clustering procedure composed of three sub-steps:

1. Update subunit activations (cluster weights) for each stimulus : $\alpha_{t,n} \leftarrow \frac{w_n e^{K_n . X_t}}{\sum_{m=1}^{m=N} w_j e^{K_m . X_t}}$.
2. Update linear subunit filters (cluster centers) : $K_n \leftarrow \frac{\sum_t Y_t \alpha_{t,n} X_t}{\sum_t Y_t \alpha_{t,n}}$.
3. Update relative weights for different subunits : $w_n \leftarrow \frac{\sum_t Y_t \alpha_{t,n}}{T} e^{-\frac{1}{2} K_n^T K_n}$.

The resulting subunits can be interpreted as a decomposition of the spike triggered average (STA), since the sum of estimated subunits (K_n), weighted by their expected contribution to the response ($w_n e^{\frac{1}{2} K_n^T K_n}$), is proportional to the spike-triggered average:

$$\sum_{n \in [N]} (w_n e^{\frac{1}{2} K_n^T K_n}) K_n = \sum_{n \in [N]} \left(\frac{\sum_t Y_t \alpha_{t,n}}{T} \right) K_n \tag{6}$$

$$= \sum_t \sum_{n \in [N]} \frac{Y_t \alpha_{t,n} X_t}{T} \tag{7}$$

$$= \frac{\sum_t Y_t X_t}{T} \tag{8}$$

where (6) comes form (3) (at convergence), (7) comes from (2) (at convergence), and (8) arises because the $\alpha_{t,n}$ sum to one. Step (6) may be interpreted as replacing the average spike rate, multiplied by contribution of each subunit. Step (7) may be interpreted as a weighted average of spike-triggered stimuli, where the weighting assigns responsibility for each spike to the subunits.

Incorporating priors

The log-likelihood can be augmented with a regularization term, to incorporate prior knowledge about the structure of subunits, thereby improving estimation in the case of limited data. Regularization is incorporated into the algorithm by projecting the subunit kernels (K_n) with a proximal operator (P_λ), derived from the regularization function, after each update of the cluster centers (step 2, figure 1). For a regularization term of the form $\lambda f(K)$, the proximal operator is defined as $\mathcal{P}_{\lambda,f}(K^0) = \text{argmin}_K \big(f(K) + (1/2\lambda) \|K - K^0\|_2^2 \big)$. For each result (Figures 3, 4) the regularization strength (λ) was chosen through cross-validation (see below).

For L_1-norm regularization (section 3), $f(K) = \sum_i |K^i|$, where K^i is the value of ith pixel, and the proximal operator is a soft-thresholding operator $\mathcal{P}_{\lambda,|.|}(K) = \text{max}(K - \lambda, 0) - \text{max}(-K - \lambda, 0)$. For the locally normalized L_1-norm regularization, the proximal operator of the Taylor approximation was used at each step. The Taylor approximation is a weighted L_1-norm with the weight for each pixel determined by the parameter value of the neighboring pixels. Hence, the proximal operator for ith pixel is $\mathcal{P}_{\lambda,|.|}(K^i) = \text{max}(K^i - a^i \lambda, 0) - \text{max}(-K^i - a^i \lambda, 0)$, where

$a^i = 1/(\epsilon + \sum_{j \in \text{Neighbors}(i)} |K^j|)$ for small $\epsilon = 0.01$. Simulations were used to verify that this regularizer induces a preference for spatially contiguous subunits, while being relatively insensitive to subunit area (compared with L_1 regularization).

Subunit estimation using hierarchical clustering

As shown in Figure 2, a hierarchical organization of subunits was observed in fits to RGC data, with one subunit broken down into two subunits each time the total number of subunits was increased by one. This hierarchical structure can be enforced explicitly to make the estimation procedure more efficient.

Since the softmax weights α_m (between 0 and 1) can be interpreted as the 'activation probability' of subunit m, hierarchical clustering is performed by estimating two children subunits m_1, m_2 with factorized activation probability : $\alpha_{m_1} = \alpha_m \alpha_{m_1|m}$, where $\alpha_{m_1|m}$ is the conditional probability of child m_1 given the activation of parent subunit with $\alpha_{m_1|m} = \frac{e^{K_{m_1}^T X + b_{m_1}}}{e^{K_{m_1}^T X + b_{m_1}} + e^{K_{m_2}^T + b_{m_2}}}$. The factorization is accurate if the activation of the parent subunit is the sum of activation of the children subunits ($e^{K_m . X + b_m} = e^{K_{m_1} . X + b_{m_1}} + e^{K_{m_2} . X + b_{m_2}}$).

The children subunits were initialized by adding independent random noise to parent subunits ($K_{m_i} = K_m + \epsilon_i$) and equally dividing the subunit weight into two ($e^{b_{m_i}} = \frac{e^{b_m}}{2}$). Hierarchical clustering is performed by iteratively updating the cluster assignments (α_{m_i}) and cluster centers ($K_{m_i},\ b_{m_i}$) for the children subunits using the factored α as outlined above.

The parent subunit for splitting is chosen greedily, to give maximum improvement in log-likelihood at each step. Hierarchical subunit estimation provides a computationally efficient way to estimate the entire solution path with different N. The estimated subunits obtained with this procedure are shown in Figure S2.

Population model

In Section 4, a method to estimate a common bank of subunits jointly using multiple nearby cells is described. For each cell c, the firing rate is given by $\lambda_{c,t} = g_c\ (\sum_{n=1}^{n=N} w_{n,c}\ e^{K_n . X_t})$ where $w_{n,c}$ are the cell specific subunit weights. The model parameters are estimated by maximizing the summation of log-likelihood across cells.

Similar to the single cell case, the estimation is performed in two steps: 1) Ignoring the output nonlinearity g_c, estimate $\{K_n, w_{n,c}\}$ by clustering, and 2) Estimate g_c, $w_{n,c}$ and the magnitude of K_n for each cell independently, using gradient descent with vector direction of $\{K_n\}$ fixed up to a scalar. Clustering in the first step can be interpreted as finding a common set of centroids to cluster the spike triggered stimuli for multiple cells simultaneously. The following steps are repeated iteratively:

1. Update subunit activations (cluster weights) for each stimulus and cell :

$$\alpha_{t,n,c} \leftarrow \frac{w_{n,c} e^{K_n . X_t}}{\sum_{m=1}^{m=N} w_{j,c} e^{K_m . X_t}}$$

2. Update linear subunit filters (cluster centers) using population responses :

$$K_n \leftarrow \frac{\sum_t \sum_c Y_{t,c} \alpha_{t,n,c} X_t}{\sum_t \sum_c Y_{t,c} \alpha_{t,n,c}}$$

3. Update subunit weights $(w_{n,c})$ for each cell :

$$w_{n,c} \leftarrow \frac{\sum_t \alpha_{t,n,c}}{T} e^{-\frac{1}{2}K_n^T K_n}$$

Application to neural responses

The subunit estimation algorithm was applied to a spatial segment of stimulus around the receptive field (5.2 micron boundary) and to the RGC spike counts over time (8.33ms bin size). The data were partitioned into three groups: testing data consisted of contiguous segment of data (last 10%) and rest of the data was randomly partitioned for training (90%) and validation (10%). Models were fit with different numbers of subunits N (1-12) and regularization values λ. Hyperparameter selection (N and λ) was performed by averaging the performance of multiple fits with different training/validation partitions and random initializations (see Figure captions for details). For a given number of subunits, the regularization value was chosen by optimizing the performance on validation data. Similarly, the most accurate model was chosen by optimizing over the number of subunits as well as regularization strength. When data with repeated presentations of the same stimulus were available (e.g. Figure 6, 7), the prediction accuracy was evaluated by measuring the correlation between peri-stimulus time histogram (PSTH) of recorded data and the predicted spikes for the same number of repetitions. The PSTH was computed by averaging the discretized responses over repeats and gaussian smoothing with standard deviation 16.66ms.

For null stimulus, responses from 30 repeats of a 10 sec long null stimulus were divided into training (first 5 sec) and testing (last 5 sec). Since the training data were limited, the subunit filters and weights (phase 1) were first estimated from long white noise stimulus and only the subunit scales and output nonlinearity (phase 2) were estimated using the null stimulus.

For natural scenes stimuli, the spatial kernel for subunits was first estimated from responses to the white noise stimulus (phase 1), and the scales of subunits and output nonlinearity were then fit using responses to a natural stimulus (phase 2). This procedure was used because subunits estimated directly from natural scenes did not show any improvement in prediction accuracy, and all the subunits were essentially identical to the RF, presumably because of the strong spatial correlations in natural images.

Previously published V1 responses to flickering bars were used for fitting subunits (see Rust et al., 2015 for details). In contrast to the retinal data, the two dimensions of the stimulus were time and one dimension of space (orthogonal to the preferred orientation of the cell). The number of subunits was chosen by cross validation as described above.

Simulated RGC model

To relate estimated subunits to bipolar cells, the model was fitted to responses generated from a simulated RGC (Appendix I). In the simulation, each RGC received exponentiated inputs from bipolar cells, which in turn linearly summed inputs from cones. The spatial dimensions of the simulation and the number of cones and bipolar cells were approximately matched to parasol RF center at the eccentricities of the recorded data (Schwartz and Rieke, 2011; Jacoby et al., 2000). Below, all the measurements are presented in grid units (g.u.), with 1 g.u. = 1 micron. The first layer consisted of 64 cones arranged on a jittered hexagonal lattice with nearest-neighbor distance 180 g.u., oriented at 60 degrees with respect to the stimulus grid. The location of each cone was independently jittered with a radially symmetric Gaussian with

standard deviation 12.6 g.u. Stimuli were pooled by individual cones with a spatio-temporally separable filter, with time course derived from a typical parasol cell and a Gaussian spatial filter (standard deviation = 12 g.u.). The photoreceptor weights were chosen to give a mean firing rate of ≈ 19 spikes/sec.

The second layer of cascade consisted of 12 bipolar cells, each summing the inputs from a spatially localized set of photoreceptors, followed by an exponential nonlinearity. The photoreceptors were assigned to individual bipolars by k-means clustering (12 clusters) of their locations. Finally, the RGC firing rate is computed by a positively weighted summation of bipolar activations. The bipolar weights were varied within 10% of the mean weight.

The subunit model was fitted on 1.5 hour long white noise stimulus such that the receptive field was roughly covered by 6 x 6 stimulus pixels, as in the recorded experimental data (Figure 2).

Null stimulus

We constructed null stimuli to experimentally validate the estimated subunits (Figures 5, 6). For a spatio-temporal stimulus movie, the goal is to find the closest stimulus such that the convolution with a spatio-temporal linear filter gives 0. For simplicity, we describe the algorithm only for a single cell, and the extension to multiple cells is straightforward. Let $a(x, y, t)$ represent the spatio-temporal linear filter, estimated by computing STA on white noise responses. The orthogonality constraint for a null stimulus $S(x, y, t)$ is written as:

$$\sum_{x,y} \sum_{\tau} a(x, y, \tau) S(x, y, t - \tau) = 0, \quad \forall t \in \{0 \cdots T\}$$

. The constraints for successive frames are not independent, but they become so when transformed to the temporal frequency domain. Writing $a(., ., \omega)$ and $S(., ., \omega)$ for the Fourier transform of $a(., ., t)$ and $S(., ., t)$ respectively, the constraints may be rewritten as:

$$\sum_{x,y} a(x, y, \omega) S(x, y, \omega) = 0, \quad \forall \omega \ \in \{\frac{2\pi i}{T}; i \in \mathbb{Z}\}$$

The spatial content can be projected onto the space orthogonal to the estimated spatial receptive field, for each temporal frequency. Specifically, given a vectorized stimulus frame S_w and a matrix A with columns corresponding to vectorized receptive fields of a set of cells, the null frame is obtained by solving the following optimization problem:

$$S_n = \text{argmin}_S \ \frac{1}{2} \|S - S_w\|_2^2, \ \text{ such that} A^T S = 0.$$

The solution may be written in closed form:

$$S_n = (I - A(AA^T)^{-1} A^T) S_w.$$

The receptive field for each cell was estimated by computing a space-time separable approximation of the spike-triggered average (STA), with the support limited to the pixels with value significantly different from background noise (absolute value of pixel value > 2.5 σ).

In addition to the orthogonality constraint, two additional constraints are imposed: a) each pixel must be in the range of monitor intensities (0-225) and b) the variance of each pixel over time must be preserved, to avoid contrast adaption in photoreceptors. These two constrains were incorporated using Dykstra's algorithm (Boyle and Dykstra, 1986), which iteratively projects

the estimated null stimulus into the subspace corresponding to individual constraints until convergence. Setting the contrast of initial white noise to 50% was necessary to satisfy these additional constraints. Finally, the pixel values of the resulting null stimuli were discretized to 8-bit integers, for display on a monitor with limited dynamic range.

For space-time separable (rank-1) STA (as in the recorded RGC data), spatio-temporal nulling gives the same solution as spatial-only nulling. To see this, assume $a(x, y, t) = a_1(x, y)a_2(t)$, a space-time separable linear filter. Hence, the null constraint can be re-written as

$$\sum_{x,y} \sum_{\tau} a_1(x, y)a_2(\tau)S(x, y, t - \tau) = \sum_{\tau} a_2(\tau)l(t - \tau) = 0 \ \ \forall t$$

where $l(t - \tau) = \sum_{x,y} a_1(x, y)S(x, y, t - \tau)$. In the frequency domain this is written as

$$a_2(\omega)l(\omega) = 0 \ \ \forall \ \omega$$

Since $a_2(t)$ has limited support in time, it has infinite support in frequency. Hence, $l(\omega) = 0 \ \ \forall \omega$ implying spatial-only nulling $l(t) = 0 \ \ \forall t$.

Acknowledgements
This work was supported by NSF IGERT Grant 0801700 (NB), NIH NEI F31EY027166 (CR), NSF GRFP DGE-114747 (CR, NB), Wu Tsai Neurosciences Institute Interdisciplinary Scholar Award (GG), Pew Charitable Trust Scholarship in the Biomedical Sciences (AS), Howard Hughes Medical Institute (ES), NIH NEI R01-EY021271 (EJC), NIH NEI P30-EY019005 (EJC). We thank Liam Paninski for helpful suggestions on the manuscript, Jill Desnoyer and Ryan Samarakoon for technical assistance, and H. Fox, M. Taffe, T. Albright, R. Krausliz, R. Siegel, K. Bankiewicz, C. Darian-Smith, J. Carmena, J. Wallis, E. Callaway, T. Moore, S. Morairty and the UC Davis Primate Center for providing access to retinas.

Author contributions
NS, NB, CR, AK, GG collected the MEA data, NS, EJC, ES conceived and designed the experiments, NS implemented the subunit model and analyzed data, AS and AL developed and supported the MEA hardware and software, NS, EJC and ES wrote the paper, all authors edited the paper, EJC and ES supervised the project.

Declaration of Interests
No competing interests.

References
Adelson, Edward H., and James R. Bergen. "Spatiotemporal energy models for the perception of motion." *Josa a* 2, no. 2 (1985): 284-299.

J. M. Angueyra and F. Rieke. Origin and effect of phototransduction noise in primate cone photoreceptors. Nature neuroscience, 16(11):1692, 2013.

Arcas, Blaise Agüera Y., and Adrienne L. Fairhall. "What causes a neuron to spike?." *Neural Computation* 15, no. 8 (2003): 1789-1807.

Baccus, Stephen A., and Markus Meister. "Fast and slow contrast adaptation in retinal circuitry." *Neuron* 36, no. 5 (2002): 909-919.

Batty, Eleanor, Josh Merel, Nora Brackbill, Alexander Heitman, Alexander Sher, Alan Litke, E. J. Chichilnisky, and Liam Paninski. "Multilayer recurrent network models of primate retinal ganglion cell responses." (2016).

Bölinger, Daniel, and Tim Gollisch. "Closed-loop measurements of iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina." *Neuron* 73, no. 2 (2012): 333-346

J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. In Advances in order restricted statistical inference, pages 28–47. Springer, 1986.

Carandini, Matteo, and David J. Heeger. "Normalization as a canonical neural computation." *Nature Reviews Neuroscience* 13, no. 1 (2012): 51.

Chander, Divya, and E. J. Chichilnisky. "Adaptation to temporal contrast in primate and salamander retina." *Journal of Neuroscience* 21, no. 24 (2001): 9904-9916.

Chichilnisky, E. J. "A simple white noise analysis of neuronal light responses." *Network: Computation in Neural Systems* 12, no. 2 (2001): 199-213.

Chichilnisky, E. J., and Rachel S. Kalmar. "Functional asymmetries in ON and OFF ganglion cells of primate retina." *Journal of Neuroscience* 22, no. 7 (2002): 2737-2747.

Dayan, Peter, Laurence F. Abbott, and L. Abbott. "Theoretical neuroscience: computational and mathematical modeling of neural systems." (2001).

Dacey, Dennis, Orin S. Packer, Lisa Diller, David Brainard, Beth Peterson, and Barry Lee. "Center surround receptive field structure of cone bipolar cells in primate retina." *Vision research* 40, no. 14 (2000): 1801-1811.

Demb, Jonathan B., Loren Haarsma, Michael A. Freed, and Peter Sterling. "Functional circuitry of the retinal ganglion cell's nonlinear receptive field." *Journal of Neuroscience* 19, no. 22 (1999): 9756-9767.

Demb, Jonathan B., Kareem Zaghloul, Loren Haarsma, and Peter Sterling. "Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina." *Journal of Neuroscience* 21, no. 19 (2001): 7447-7454.

Duchi, John, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization." *Journal of Machine Learning Research* 12, no. Jul (2011): 2121-2159.

Enroth-Cugell, Christina, and John G. Robson. "The contrast sensitivity of retinal ganglion cells of the cat." *The Journal of physiology* 187, no. 3 (1966): 517-552.

Enroth-Cugell, Christina, and Lawrence Pinto. "Algebraic summation of centre and surround inputs to retinal ganglion cells of the cat." *Nature* 226, no. 5244 (1970): 458.

Fahey, Patrick K., and Dwight A. Burkhardt. "Center-surround organization in bipolar cells: symmetry for opposing contrasts." *Visual neuroscience* 20, no. 1 (2003): 1-10.

Field, Greg D., Alexander Sher, Jeffrey L. Gauthier, Martin Greschner, Jonathon Shlens, Alan M. Litke, and E. J. Chichilnisky. "Spatial properties and functional organization of small bistratified ganglion cells in primate retina." *Journal of Neuroscience* 27, no. 48 (2007): 13261-13272.

Field, Greg D., Jeffrey L. Gauthier, Alexander Sher, Martin Greschner, Timothy A. Machado, Lauren H. Jepson, Jonathon Shlens et al. "Functional connectivity in the retina at the resolution of photoreceptors." *Nature* 467, no. 7316 (2010): 673.

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. and Chichilnisky, E.J., 2015. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *Elife*, *4*, p.e05241.

Frechette, Eric S., Alexander Sher, Matthew I. Grivich, Dumitru Petrusca, Alan M. Litke, and E. J. Chichilnisky. "Fidelity of the ensemble code for visual motion in primate retina." *Journal of neurophysiology* (2005).

Gauthier, Jeffrey L., Greg D. Field, Alexander Sher, Martin Greschner, Jonathon Shlens, Alan M. Litke, and E. J. Chichilnisky. "Receptive fields in primate retina are coordinated to sample visual space more uniformly." *PLoS biology* 7, no. 4 (2009): e1000063.

Gollisch, Tim, and Markus Meister. "Rapid neural coding in the retina with relative spike latencies." *science* 319, no. 5866 (2008): 1108-1111.

Heeger, David J. "Normalization of cell responses in cat striate cortex." *Visual neuroscience* 9, no. 2 (1992): 181-197.

Heitman, Alexander, Nora Brackbill, Martin Greschner, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. "Testing pseudo-linear models of responses to natural scenes in primate retina." *bioRxiv* (2016): 045336.

Hochstein, S., and R. M. Shapley. "Linear and nonlinear spatial subunits in Y cat retinal ganglion cells." *The Journal of Physiology* 262, no. 2 (1976): 265-284.

Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." *The Journal of physiology* 148, no. 3 (1959): 574-591.

Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *The Journal of physiology* 160, no. 1 (1962): 106-154.

Hunter, Ian W., and Michael J. Korenberg. "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models." *Biological cybernetics* 55, no. 2-3 (1986): 135-144.

Jacoby, Roy A., Allan F. Wiechmann, Susan G. Amara, Barbara H. Leighton, and David W. Marshak. "Diffuse bipolar cells provide input to OFF parasol ganglion cells in the macaque retina." *Journal of Comparative Neurology* 416, no. 1 (2000): 6-18.

Jia, Shanshan, Zhaofei Yu, Arno Onken, Yonghong Tian, Tiejun Huang, and Jian K. Liu. "Characterizing neuronal circuits with spike-triggered non-negative matrix factorization." *arXiv preprint arXiv:1808.03958* (2018).

Jones, Judson P., and Larry A. Palmer. "The two-dimensional spatial structure of simple receptive fields in cat striate cortex." *Journal of neurophysiology* 58, no. 6 (1987): 1187-1211.

Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*(2014).

Keat, Justin, Pamela Reinagel, R. Clay Reid, and Markus Meister. "Predicting every spike: a model for the responses of visual neurons." *Neuron* 30, no. 3 (2001): 803-817.

Korenberg, Michael J., and Ian W. Hunter. "The identification of nonlinear biological systems: LNL cascade models." *Biological cybernetics* 55, no. 2-3 (1986): 125-134.

Korenberg, M.J., Sakai, H.M. and Naka, K.L., 1989. Dissection of the neuron network in the catfish inner retina. III. Interpretation of spike kernels. *Journal of Neurophysiology*, *61*(6), pp.1110-1120.

Kuo, Sidney P., Gregory W. Schwartz, and Fred Rieke. "Nonlinear spatiotemporal integration by electrical and chemical synapses in the retina." *Neuron* 90, no. 2 (2016): 320-332.

Liu, Jian K., Helene M. Schreyer, Arno Onken, Fernando Rozenblit, Mohammad H. Khani, Vidhyasankar Krishnamoorthy, Stefano Panzeri, and Tim Gollisch. "Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization." *Nature communications* 8, no. 1 (2017): 149.

Litke, A.M., Bezayiff, N., Chichilnisky, E.J., Cunningham, W., Dabrowski, W., Grillo, A.A., Grivich, M., Grybos, P., Hottowy, P., Kachiguine, S. and Kalmar, R.S., 2004. What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Transactions on Nuclear Science*, *51*(4), pp.1434-1440.

Marmarelis, Panos Z., and Ken-Ichi Naka. "White-noise analysis of a neuron chain: an application of the Wiener theory." *Science* 175, no. 4027 (1972): 1276-1278.

McFarland, James M., Yuwei Cui, and Daniel A. Butts. "Inferring nonlinear neuronal computation based on physiologically plausible inputs." *PLoS computational biology*9, no. 7 (2013): e1003143.

McIntosh, Lane, Niru Maheswaranathan, Aran Nayebi, Surya Ganguli, and Stephen Baccus. "Deep learning models of the retinal response to natural scenes." In *Advances in neural information processing systems*, pp. 1369-1377. 2016.

Movshon, J. Anthony, Ian D. Thompson, and David J. Tolhurst. "Spatial summation in the receptive fields of simple cells in the cat's striate cortex." *The Journal of physiology* 283, no. 1 (1978): 53-77.

Nirenberg, Sheila, and Markus Meister. "The light response of retinal ganglion cells is truncated by a displaced amacrine circuit." *Neuron* 18, no. 4 (1997): 637-650.

Ölveczky, Bence P., Stephen A. Baccus, and Markus Meister. "Segregation of object and background motion in the retina." *Nature* 423, no. 6938 (2003): 401.

Paninski, Liam. "Convergence properties of some spike-triggered analysis techniques." In *Advances in neural information processing systems*, pp. 189-196. 2003.

Park, Il Memming, Evan W. Archer, Nicholas Priebe, and Jonathan W. Pillow. "Spectral methods for neural characterization using generalized quadratic models." In *Advances in neural information processing systems*, pp. 2454-2462. 2013.

Pillow, Jonathan W., and Eero P. Simoncelli. "Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis." *Journal of vision* 6, no. 4 (2006): 9-9.

Pillow, Jonathan W., Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E. J. Chichilnisky, and Eero P. Simoncelli. "Spatio-temporal correlations and visual signalling in a complete neuronal population." *Nature* 454, no. 7207 (2008): 995.

Rajan, Kanaka, and William Bialek. "Maximally informative “stimulus energies” in the analysis of neural responses to natural signals." *PloS one* 8, no. 11 (2013): e71959.

Real, Esteban, Hiroki Asari, Tim Gollisch, and Markus Meister. "Neural circuit inference from function to structure." *Current Biology* 27, no. 2 (2017): 189-198.

Ramirez, Alexandro D., and Liam Paninski. "Fast inference in generalized linear models via expected log-likelihoods." *Journal of computational neuroscience* 36, no. 2 (2014): 215-234.

Rust, Nicole C., Odelia Schwartz, J. Anthony Movshon, and Eero P. Simoncelli. "Spatiotemporal elements of macaque v1 receptive fields." *Neuron* 46, no. 6 (2005): 945-956.

Sakai, HIROKO M. "White-noise analysis in neurophysiology." *Physiological Reviews* 72, no. 2 (1992): 491-505.

J. Schnapf, B. Nunn, M. Meister, and D. Baylor. Visual transduction in cones of the monkey macaca fascicularis. The Journal of physiology, 427(1):681–713, 1990.

Schwartz, Greg, and Fred Rieke. "Nonlinear spatial encoding by retinal ganglion cells: when 1+ 1≠ 2." *The Journal of general physiology* 138, no. 3 (2011): 283-290.

Schwartz, Gregory W., Haruhisa Okawa, Felice A. Dunn, Josh L. Morgan, Daniel Kerschensteiner, Rachel O. Wong, and Fred Rieke. "The spatial structure of a nonlinear receptive field." *Nature neuroscience* 15, no. 11 (2012): 1572.

Schwartz, Odelia, Jonathan W. Pillow, Nicole C. Rust, and Eero P. Simoncelli. "Spike-triggered neural characterization." *Journal of vision* 6, no. 4 (2006): 13-13.

Sharpee, Tatyana, Nicole C. Rust, and William Bialek. "Analyzing neural responses to natural signals: maximally informative dimensions." *Neural computation* 16, no. 2 (2004): 223-250.

Turner, Maxwell H., and Fred Rieke. "Synaptic rectification controls nonlinear spatial integration of natural visual inputs." *Neuron* 90, no. 6 (2016): 1257-1271.

Turner, Maxwell H., Gregory W. Schwartz, and Fred Rieke. "Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina." *bioRxiv*(2018): 252148.

Theis, Lucas, Andrè Maia Chagas, Daniel Arnstein, Cornelius Schwarz, and Matthias Bethge. "Beyond GLMs: a generative mixture modeling approach to neural system identification." *PLoS computational biology* 9, no. 11 (2013): e1003356

Tsukamoto, Yoshihiko, and Naoko Omi. "OFF bipolar cells in macaque retina: type-specific connectivity in the outer and inner synaptic layers." *Frontiers in neuroanatomy* 9 (2015): 122.

Van Hateren, J. Hans, and Arjen van der Schaaf. "Independent component filters of natural images compared with simple cells in primary visual cortex." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265, no. 1394 (1998): 359-366.

Vintch, Brett, J. Anthony Movshon, and Eero P. Simoncelli. "A convolutional subunit model for neuronal responses in macaque V1." *Journal of Neuroscience* 35, no. 44 (2015): 14829-14841.

Wu, Anqi, Il Memming Park, and Jonathan W. Pillow. "Convolutional spike-triggered covariance analysis for neural subunit models." In *Advances in neural information processing systems*, pp. 793-801. 2015.

Supplementary

Validation of subunit estimation of simulated RGC responses

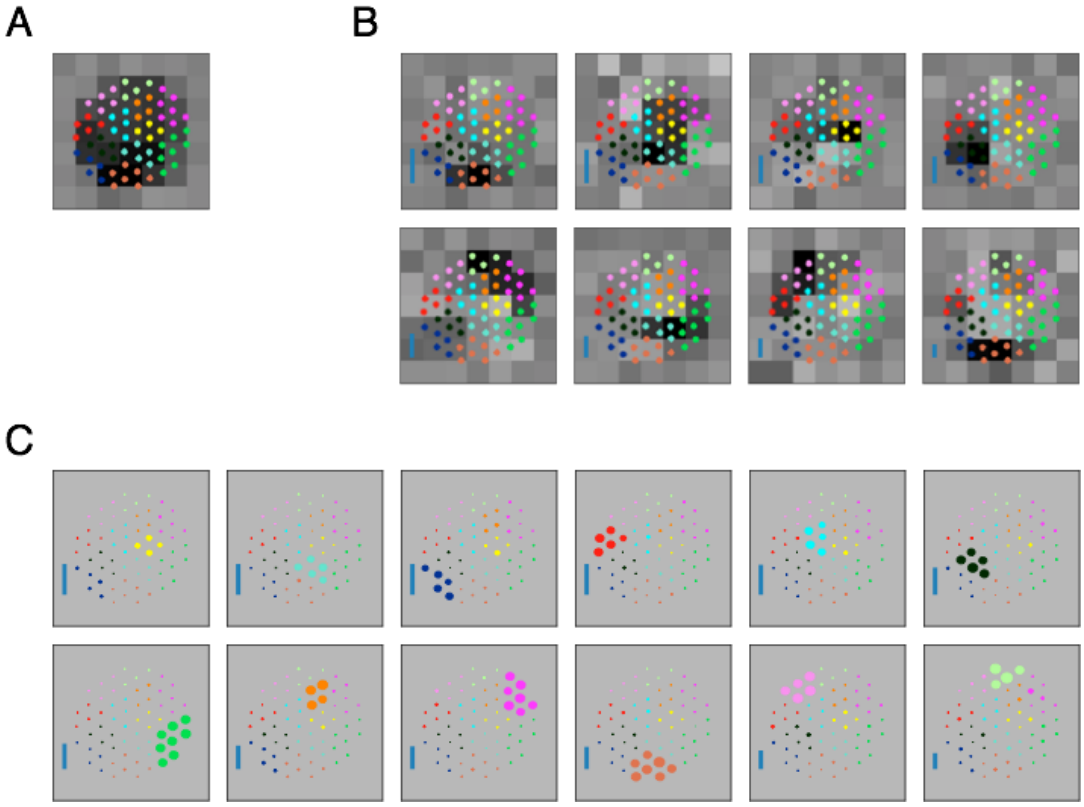

Figure S1: Validation of the subunit fitting algorithm on simulated RGC data. (A) Receptive field (spike-triggered average) of the simulated RGC with cascaded linear nonlinear units. Stimulus is temporally filtered by 64 photoreceptors, organized on a jittered hexagonal grid. Photoreceptor activations are summed by 12 bipolar cells, each connecting to 4-8 photoreceptors. Bipolar activations are exponentiated and added to give the poisson firing rate of the cell. Dots indicate photoreceptor locations, photoreceptors with same color connect to a single bipolar cell. See Methods for further details. (B) Subunits estimated from simulated RGC responses to 24 min of coarse white noise reveals fewer (8) subunits compared to the actual number of bipolar cells (12). The spatially localized subunits are aggregates of multiple, nearby underlying bipolar cells. The subunits are partially overlapping and tile to cover the receptive field. Inset: Blue bar height indicates the relative strength (average contribution to response over stimulus ensemble) of subunit. (C) Subunits estimated from simulated RGC responses to 6 hours of fine resolution white noise stimulus recovers the underlying bipolar cells. Size of dots indicate relative weights of different photoreceptors to the estimated subunits (12 total). Each subunit has strong inputs from photoreceptors connected to same bipolar cell. Inset: Same as B.

Estimating subunits by hierarchical clustering

Figure S2: Gradual partitioning of the receptive field into subunits by hierarchical clustering. Different number of subunits (rows) estimated by splitting one parent subunit into two subunits at each step. Children subunits estimated by soft-clustering the simulated spikes of the parent subunit, with the simulated spikes for parent subunit computed as the spiking activity of the cell, weighed by its soft-max subunit activation (see Methods). The parent subunit that gives maximum decrease in log-likelihood on training data is chosen for splitting. The choice of training data, preprocessing and figure details same as Figure 2. The achieved splitting of subunits is similar to the pattern of splitting in Figure 2 for small number of subunits (1- 5 subunits), but differs for larger number of subunits (6, 7 subunits). This suggests that enforcing the hierarchical constraint could lead to a more efficient estimation procedure.