Stimulus invariant aspects of the retinal code drive discriminability of natural scenes

Benjamin D. Hoshal*
Committee on Computational Neuroscience, University of Chicago, Chicago, Illinois 60637, USA

Caroline M. Holmes*
Department of Physics, Princeton University, Princeton, New Jersey, 08540

Kyle Bojanek and Jared Salisbury
Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA

Michael J. Berry, II
Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544

Olivier Marre
Institut de la Vision, Sorbonne Université, INSERM, CNRS, Paris, France

Stephanie E. Palmer
Department of Organismal Biology and Anatomy and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA Center for the Physics of Biological Function, Princeton University, Princeton, New Jersey 08544, USA

(Dated:)

Everything that the brain sees must first be encoded by the retina, which maintains a reliable representation of the visual world in many different, complex natural scenes while also adapting to stimulus changes. Decomposing the population code into independent and cell-cell interactions reveals how broad scene structure is encoded in the adapted retinal output. By recording from the same retina while presenting many different natural movies, we see that the population structure, characterized by strong interactions, is consistent across both natural and synthetic stimuli. We show that these interactions contribute to encoding scene identity. We also demonstrate that this structure likely arises in part from shared bipolar cell input as well as from gap junctions between retinal ganglion cells and amacrine cells.

While single cells individually encode specific stimulus features[1–3], it is their aggregate response that drives our perception[4–7]. For this reason, it is important to understand not only how individual cells respond to stimuli, but also how cells influence each other within a population [8–11]. Significant theoretical work has been devoted to understanding population responses [12–16], in tandem with experimental innovations in recording from a large number of cells simultaneously [17–20]. This creates an opportunity to move to more complex, dynamic stimuli and analyze the population code in terms of the readout goals of the downstream networks.

The natural environment has many complex spatio-temporal features that make neural encoding in the wild difficult to quantify and assess. Natural scenes vary in luminance over many orders of magnitude [21] and variance [22] [23], and have complicated temporal and spatial structure [24, 25]. Visual systems adapt to these changes on many scales in time and space. Neural systems show near-perfect adaptation to these changes [26], so a question remains about how brains recover scene-specific information once in an adapted state. These complexities and open questions have led many studies to investigate animal behavior in natural settings [27–30]. In this work, we quantify the structure of the neural code at the input end, and how it might support downstream readout that ultimately drives behavior in complex environments.

While natural scenes contain a multitude of higher order statistics, not all features are equally important. Even at the earliest stages of visual processing, the retina performs nonlinear computations to encode essential aspects of the visual scene. Retinal networks are flexible enough to encode a wide variety of complex stimulus features, such as object motion [31, 32], motion reversals [33–35], and omitted stimuli [36]. These early computations support efficient downstream readout by throwing away redundant information and preserving features that facilitate perception.

We use the responses of a salamander retina to natural stimuli to infer a minimal model of the population response structure of its output retinal ganglion cells. This structure is conserved across scenes, and it has functional consequences; it helps the population carry information about large-scale scene statistics. Finally, we show that this functional role requires only a sparse set of connections, and that these sparse couplings appear to arise from both shared input (bipolar and amacrine cells) and direct connections (gap junctions).

Results

Probing multiple naturalistic, dynamic inputs to the retina We make dense extracellular recordings from retinal ganglion cells (RGCs) in the larval tiger salamander (Fig. 1a) while presenting the retina with 20-second clips from five different movies (see Methods for details). Salamanders undergo metamorphosis, exposing them to both underwater and terrestrial environments while their retinal structure remains largely the same [37, 38]. Further, while salamanders are traditionally ambush predators, they still navigate through their environment, generating self-motion. The movies chosen represent a sampling of the wide variety of scenes that occur in the organism's ecological niche (Fig. 1b).

Because of the diversity of subjects and locations, the movies exhibit significantly different temporal and spatial correlation (Fig. 1d). Further, luminance correlations alone fail to capture behaviorally relevant features like motion, which arise from higher order structure. For example, object tracking in natural scenes reveal different shapes and timescales of the velocity autocorrelations for different scenes [39]. As animals navigate through different environments, they rapidly adapt to these changes in stimulus statistics [26, 40, 41]. Thus, playing a variety of scenes in a single experimental session allows for investigation into changes in the population structure across scenes.

Pairwise couplings are consistent across movies To fully describe a dynamic population code, we must enumerate 2^N possible states at each time point in the response. Even for modest N, a fully expressive code is both experimentally inaccessible and potentially unreadable by downstream networks. To summarize the population code, we use maximum entropy modeling, which has a history of success using $O(N^2)$ parameters to capture the structure in the data, even higher-order features not explicitly constrained by the model [42–49].

In applications of maximum entropy techniques in neuroscience, these models are constrained by the average firing rate of each cell and the correlation of each pair of cells [42], $\langle \sigma_i \rangle$ and $\langle \sigma_i \sigma_j \rangle$. We use a time-dependent maximum entropy model [50] that is also constrained by the time-varying firing rates averaged across repeated stimuli (see Methods for details). Our model takes the form

$$P(\vec{\sigma}^t) = \frac{1}{Z} e^{-\sum_i^N h_i^t \sigma_i^t - \sum_{i<j}^N J_{ij} \sigma_i^t \sigma_j^t}, \qquad (1)$$

and our constraints are on $\langle \sigma_i^t \rangle_k$, which captures each cell's individual response to the stimulus at time t averaged over trials, k, as well as $\langle \sigma_i^t \sigma_j^t \rangle_{t,k}$, the correlations between cells. These two constraints map to two sets of parameters, the time-dependent fields h_i^t and the static couplings J_{ij}, respectively. Interactions between the time-dependent fields h_i^t absorb any stimulus-dependent correlations, leaving the couplings J_{ij} to capture the noise correlations. Given that our model accurately predicts population activity (Fig. 2a), and that

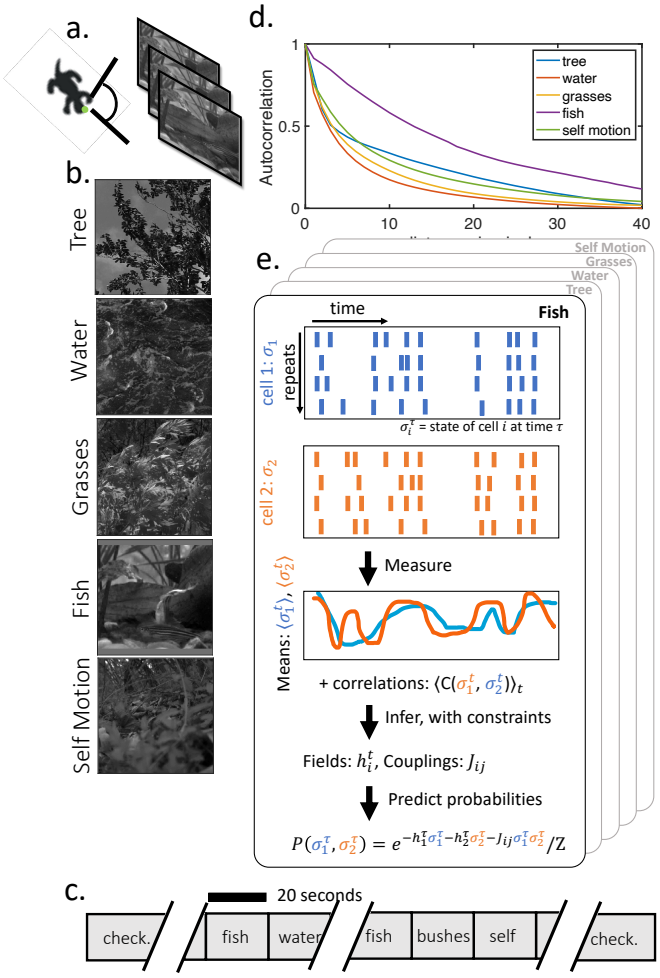

FIG. 1. **Measuring retinal ganglion cell responses to natural scenes.** (a) Voltage responses were recorded from the retinal ganglion cell layer of a salamander retina stimulated by natural movies. (b) Example frames from each of five natural scenes, which show, respectively, trees blowing in the wind; flowing water; ferns and grasses in a breeze; fish swimming; and woodland underbrush as viewed by a moving camera. The bottom image shows the aggregation of the receptive fields of the recorded population of neurons. (c) In order to probe the statistics of responses, natural scenes were repeated a minimum of 80 times in pseudorandom order. A checkerboard stimulus lasting 25 minutes was shown before and after the natural scenes. (d) The five natural movies used have different statistics, including (shown here) different spatial autocorrelation functions. (e) To model responses to each of the movies, time-dependent maximum entropy models are fit to each of the five natural scenes.

the fields are explicitly constrained by stimulus-induced single-cell statistics, we consider the matrices J_{ij} to carry the essence of the intrinsic, non-independent population structure.

In all movies, the cells significantly increase their firing rates in the first 200 ms (Fig. 2b) following the switch to a new stimulus. This is followed by a rapid decay back to a baseline firing rate. This is likely due to a strong pop-

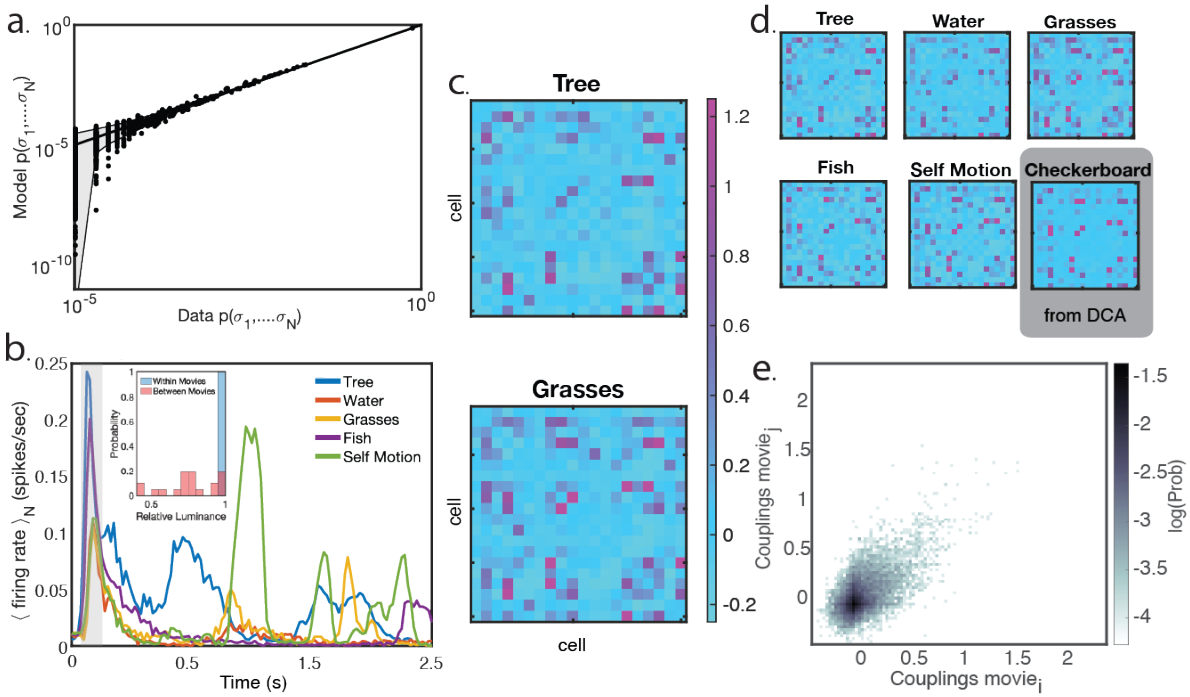

FIG. 2. **Retinal ganglion cell population maintains consistent couplings across a variety of natural stimuli** (a) Probability of population states, as measured from data and compared to the model. Gray shading indicates expected sampling noise in the estimates of probability from data. (b) Average population firing rate as a function of time for the first two seconds of each natural movie. In the first 200ms (grey bar), the population firing rate peaks. This may be due to a change in luminance within the aggregate population receptive field between stimuli (subplot). (c) Couplings for an example 20 cell group from two of the movies, Tree and Grasses. The structure of the coupling matrices are consistent across scenes. (d) Couplings across all five natural scenes and the checkerboard white noise stimulus. In all cases, couplings are consistent across stimuli. Even for the checkerboard, where another modeling procedure (DCA) had to be used, similar couplings are found. (e) Couplings are quantitatively similar across movies ($R^2 = 0.74$), and for any choice of group of cells; here, we show the values of the couplings J_{ij} for all pairs of movies, for ten different groups of cells.

ulation response to abrupt changes in luminance within their receptive fields (Fig. 2b inset) [51–53]. Despite this substantial adaptation to each movie, we find consistent J_{ij} matrices across movies, indicating that the noise correlations are conserved (Fig. 2c,d). Previous work has similarly found consistent couplings across visual inputs [50, 54, 55], but this is the first time this has been demonstrated across a range of naturalistic stimuli. These are entirely independently trained models, which separately learned the same couplings despite significantly varying scene statistics and population responses (Fig. 1). This strongly implies that this structure arises from the retina itself, rather than being inherited from the stimulus.

These consistent couplings are not unique to the particular 20-cell group analyzed in Fig. 2c and d. For a selection of randomly chosen groups, we plot the coupling J^{α}_{ij} between cells i and j in movie α against the couplings J^{β}_{ij}. We observe a strong correlation between the cell-cell interactions across movies (Fig. 2e). These couplings could arise in response to the shared long-timescale and length-scale correlations in natural scenes [24], or they could be an anatomical property of the retina. In order to investigate this, we need to compare the conserved couplings we observed in response to the natural movies to those found in response to an entirely different, non-naturalistic stimulus.

Finally, we investigate the population structure in response to a white-noise checkerboard stimulus. Due to a lack of repeats, we used a different method, Direct Coupling Analysis (DCA) [56], to infer these couplings (see Methods for details). Despite the many differences between this model and those above (stimulus, model details), we extract the same strong couplings as those found in response to the natural scenes. This strengthens the argument that the observed couplings are indicative of real biological interactions, not correlations inherited from the input.

Couplings allow for better decoding of scene identity What could these static, sparse couplings be used for downstream of the retina? While single cells adapt to switches between scenes after about one second and then fluctuate in response to ongoing dynamics

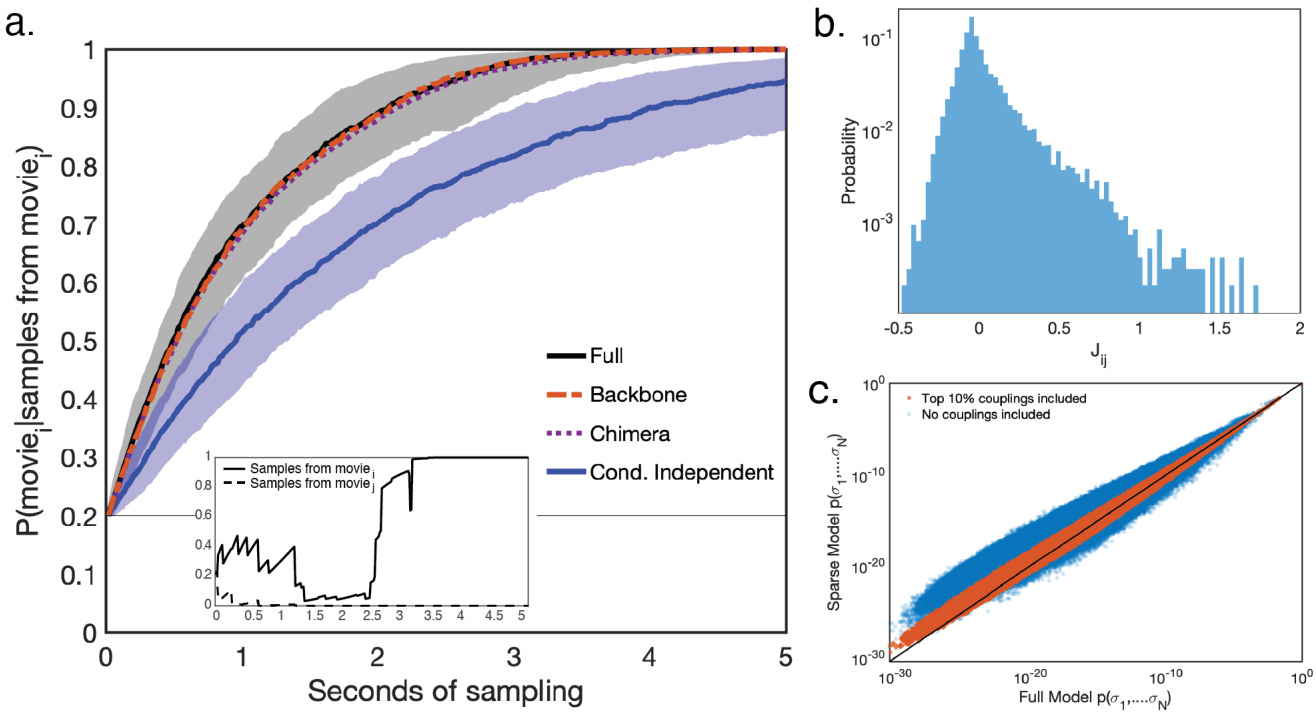

FIG. 3. **The conserved coupling structure aids in decoding scene identity.** (a) Probability of a given scene, given a particular spike train. The coupling population structure gives dramatically faster scene decoding than an independent model. Additionally, both the chimera model, which swaps J_{ij} matrices, and the sparse backbone model, which preserves only 10% of couplings, carry nearly as much scene information as the full model. Inset: probability of scene given spikes for a single trial of generated samples. (b) The distribution of J_{ij}s across many 20-cell models; we find that this distribution is heavy tailed. (c) Probabilities of population states, for our full model, a model with all J_{ij}s set to zero, and a model with sparsified J_{ij}s. The backbone model captures the predictions of the full model, while the independent model fails spectacularly.

within a scene, the population structure remains constant. Surprisingly, this scene-invariant, static backbone of interactions supports faster and more reliable readout of scene identity after this adaptation has occurred.

Interactions between cells in the retinal population have been observed in many other input contexts, and shape individual cell stimulus encoding [49]. The sparse pairwise structure that we observe may combine with individual cell fluctuations to change the overall encoding map between scenes. This interaction may support decoding scene identity. The movies contain a suite of higher order features that make each one readily discriminable to the human eye, but may also impact the local, correlated retinal population code in a decodable way.

To quantitatively test whether the couplings affect discrimination between scenes, we take advantage of the fact that each of our movies comes from a significantly different environment. This means that the information about scene statistics can be approximated by information about scene identity. We quantify the ability of an ideal observer to correctly identify a scene based solely on access to retinal output, given by the posterior $P(\text{scene identity}|\text{spikes})$, as a function of number of samples of the retinal response. This decoding task is similar to the real problem solved by downstream brain areas when an organism moves between scenes as it navigates its natural environment, and must trigger different behaviors and priors in different niches.

The quantity $P(\text{scene identity}|\text{spikes})$ describes how likely any particular scene is given a particular series of spikes. We measure this quantity by generating independent samples from our time-averaged models for each movie, $P(\vec{\sigma}) = \frac{1}{T}\sum_t^T P(\vec{\sigma}^t|t)$, and calculating $P(\text{scene identity}|\text{spikes})$ as a function of the number of generated samples using Bayes' law,

$$P(\text{scene identity}|\text{spikes}) = \frac{P(\text{spikes}|\text{scene identity})P(\text{scene identity})}{P(\text{spikes})}.$$

The terms on the right hand side of this equation can all be generated from our model. Because each sample has a defined length of 1/60 s, we can then convert the number of samples to a number of seconds of sampling.

By performing this analysis, we find that samples generated from just a 20-cell group carry enough information to correctly identify a scene within a few seconds (Fig. 3a). By contrast, spikes generated from a conditionally independent model, which is fit while constraining all J_{ij} to zero, take nearly twice as long to achieve the same

scene discriminability. This is surprising, as it was not *a priori* obvious that couplings would contribute to decoding [57].

While the population structure remains scene invariant, there are subtle changes in J_{ij} values across movies which might influence the neural code. Conversely, subtle changes in coupling strength might have minimal functional effect so long as the overall population structure remains consistent. To test whether small changes in coupling strength between movies affect scene discriminbability, we implement 'chimera' models. A chimera model for movie α uses the fields from that movie, $h_i^{t,\alpha}$, but replaces the couplings with those learned from a different movie, J_{ij}^{β}. This generates models that maintain the scene invariant coupling structure observed across movies while allowing individual coupling strengths to fluctuate. We find that spike trains generated from these models lead to similarly fast decoding as from the full model. This implies that fluctuations in the coupling values between scenes have little functional impact on scene readout. Instead, the scene-invariant population structure alone drives the improvement in scene discriminability.

We can then investigate whether the entire interacting population structure is important for decoding, or whether the J_{ij} interactions can be sparsified without sacrificing discriminability of scenes. The coupling distribution is heavy-tailed, with a sparse set of strong couplings (Fig. 3b) alongside many weak interactions. Previous work has conflicting reports on the relevance of weak couplings, where some show that weak couplings combine to have a large effect on population activity [42] while others suggest that ignoring weak interactions has minimal effect on population responses [45]. To investigate the role of strong couplings, we sparsify the J_{ij} matrix, leaving only the top 10% of couplings to shape population activity and re-train our model. We fit these 'backbone models' while constraining the weaker 90% of couplings from the full model to be zero. The backbone model is nearly as fast at identifying scenes as the full model, suggesting large-scale scene information is specifically preserved through the sparse, strong couplings rather than the combination of many weak couplings.

Additionally, upon comparing state probabilities, we find that the backbone model makes predictions that are very close to those of the full model, while the independent model fails significantly (Fig. 3c). These results suggest that the conserved population structure is dominated by a backbone of sparse strong couplings, and that these couplings play a functional role in preserving scene-level information.

Couplings arise from both gap junctions and shared bipolar cell input We have found consistent population structure that is dominated by sparse couplings. This structure is hard-wired into the retina code and could arise from many different circuit properties. Pairwise retinal couplings could be correlated with shared upstream input from a bipolar cell (Fig. 4a, left) , direct gap junctions between retinal ganglion cells (Fig. 4a top), or gap junctions between RGCs and a third neuron such as an amacrine cell (Fig. 4a right).

Nonlinear summation of bipolar cell (BC) inputs has been shown to be an integral component of retinal computation [58–62]. The convergence of BCs onto RGCs is modeled using nonlinear summation in so-called cascade models. These models explain a wide array of complex retinal computations (e.g. motion onset [58], omitted stimulus response [36] , background vs object motion [32], reversal response [33–35]). As a complement to that, BCs have diverging projections onto multiple RGCs on the retina [63]. While bipolar cells sample a smaller portion of the visual scene than RGCs, gap junctions networks between RGCs can expand the bipolar projective field out as far as ∼1mm [64] This BC divergence could play a role in shaping the population code in the retina and needs to be further explored with naturalistic, dynamic inputs.

In order to detect putative bipolar cell inputs onto the RGCs in our dataset, we use spike triggered independent component analysis (ST-ICA) [65] on the white noise checkerboard stimulus. ST-ICA models each RGC as the output of a temporal filter and spatial subunits. In a similar method, these subunits have been experimentally shown to map to bipolar cell inputs [66]. We show an example of the spatial subunits and temporal filters for three fast-OFF RGCs in Fig. 4b and Fig. 4c, two pairs of which (blue, pink; green, pink) demonstrate strong couplings. One coupling pair (blue,pink) exhibits a highly overlapped spatial subunit that we classify as a shared upstream input (see Methods). In Fig. 4c, temporal filters for two cells (blue, green) show near identical characteristics. This may arise from a gap junction connecting them and might explain their strong coupling [50].

Across the population, many RGCs share spatial subunits. These subunits closely align with the observed coupling matrix from the stimulus dependent maximum entropy models (Fig. 4d), demonstrating that strong couplings may arise in part from shared upstream input. However, not all cells with strong couplings share upstream input, and the presence of a shared subunit alone does not guarantee the existence of a strong coupling. Thus, strong couplings might arise from multiple sources within the retinal cell population.

Previous work has suggested that gap junctions may underpin couplings between RGCs [8, 50]. Here, we find evidence of gap junctions by inspecting the cross-covariance of responses after subtracting the trial-average (at zero lag this is the usual noise correlation). Pairwise noise correlations due to gap junctions can generally be split into two classes, direct RGC-RGC couplings and shared gap junctions with a third upstream neuron. The symmetric, medium width correlation we observe between some highly coupled cells without a shared subunit (Fig. 4d, red) likely arises from this second class, as direct RGC-RGC gap junctions lead to transient noise correlations at a sub-millsecond timescale [8]. Further-

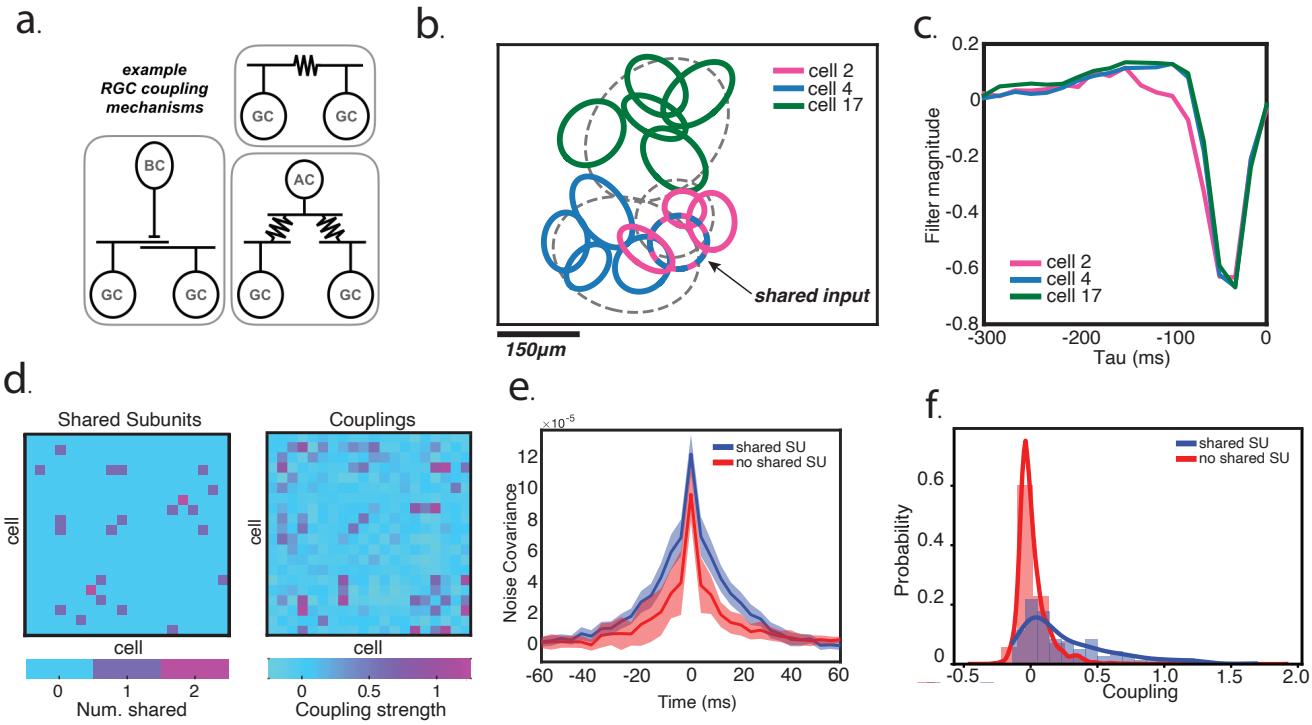

FIG. 4. **Strong couplings result from both shared inputs and gap junctions.** (a) Three potential pairwise coupling sources. Left, shared input from an upstream bipolar cell. Middle, a direct gap junction between two retinal ganglion cells. Right, gap junction connections to a shared third neuron an amacrine cell. (b) ST-ICA modeled spatial subunits for three cells, two of whom share a subunit. (c) Time filters for the same cells shown in b. (d) The average noise covariance for highly coupled cells. Highly coupled cells that share a subunit have a broader noise covariance in time than coupled cells without a shared subunit. (e) A comparison between shared subunits and couplings indicates many highly coupled cells shared upstream input. (f) Distribution of coupling strengths, both for cells sharing subunits and those that do not.

more, the broader noise correlations between RGCs with shared subunits demonstrates a timescale longer than can be explained from gap junctions alone and may indicate a coupling arising from shared upstream input (Fig. 4a, blue).

Shared input between RGCs, whatever the source, greatly increase the likelihood of a a strong coupling compared to RGC pairs without shared input (Fig. 4e). Of course, shared input does not *guarantee* a strong coupling between RGCs, but the long tail of strong coupling for pairs with shared input suggest that these mechanisms underpin our sparse network of strong interactions. We find that shared bipolar input and gap junctions work in consort to generate a sparse set of intrinsic correlations between RGCs.

Discussion

This work demonstrates that couplings between cells in a neural population are an important component of downstream readout of scene identity. While some studies show that independent models of the retinal code retain upwards of 90% of the response structure [67], this is typically agnostic to the downstream readout goals of the organism. Without a defined goal, it is impossible to determine whether aspects of the response structure lost by an independent encoding scheme relay information meaningful to the organism. It is possible that an independent readout preserves the majority of information available in the retinal population while failing to effectively convey critical features of the visual scene. Natural scenes probe a behaviorally relevant context to assess the impact of noise correlations on neural coding. These movies, like other natural inputs, drive a richer and more reliable code in the brain [68–70]. Comparing across movies reveals what the more subtle features in the neural code might be used for.

Our finding that sparse interactions sufficiently capture the functional impact of noise correlations on neural encoding elaborates on previous work that argued for a dense network of weak couplings in the retinal code [42, 46]. In these early models, both the fields and interactions between cells were static. The fluctuating fields we included (following the time-dependent maximum entropy model work [50]) absorb much of that structure, and capture the independent component of stimulus-driven changes in the neural population response. The

remaining sparse couplings are the key factor for efficient scene identification. A sparse backbone may be easier to implement and read out downstream. On the flip side, sparse codes might hamstring error correction [71, 72], so future work should explore how these costs and benefits trade-off for behaviorally relevant inputs and tasks.

We find that the noise correlations have a large effect on scene decoding, which may arise from small effects aggregated over time. It is not clear from the analysis performed here what precisely gives rise to the beneficial impacts of noise correlations on decoding. One possible answer is that the noise correlations may reflect changes in scene correlation structure. This may help recover scene specific information that is otherwise lost to single-cell-level adaptation.

Unraveling how this sparse but strong structure in the code is mechanistically supported is an important next step in this work. In some ways, the circuit structure in the eye differs from that found in the cortex. The retina is not a recurrent neural network; RGCs do not have direct synaptic coupling, and the photoreceptor-to-RGC circuit is largely feed-forward. To create a population code with sparse interactions, the retina needs to be wired around these structural constraints. These sparse interactions seem to be the result of common bipolar inputs and gap junction coupling between RGCs. What we have observed is sparse, strong, functionally important, exclusively non-synaptic RGC-RGC couplings. Both gap junctions and common bipolar inputs lead to stronger coupling between cells, but our analysis is not sensitive enough to tease apart whether these two types of coupling sources are mutually exclusive. Exclusivity would be an efficient way to implement a sparse backbone of specific cell-cell interactions. Future work to disentangle the circuit mechanisms giving rise to the sparse backbone might ultimately inform studies in cortex where gap junction coupling is also present [73–75].

Methods

Neural data Voltage traces from the RGC layer of a larval tiger salamander retina were recorded following the methods outlined in [17]. In brief, retina from a larval tiger salamander was isolated in darkness and pressed against a 252 channel multielectrode array. Recordings were taken during stimulus presentation and spike sorted using a mostly automated spike sorting algorithm. This technique captured a highly overlapping neural population of 93 cells that fully tiled a region of visual space. Spikes were binned at 60Hz for all analyses presented.

The binned data for both the checkerboard and movie stimuli can be found here: [LINK].

Stimuli during recording A 30Hz white noise checkerboard stimulus was played for 30 minutes prior to and after the natural scene stimuli.

Five different natural scenes lasting 20s were played in a pseudorandom order, and each were shown a minimum of 80 times. Specifically, they were shown (in order, for the tree, water, grasses, fish, and self motion movies) 83, 80, 84, 91, and 85 times. All natural scenes except for the tree stimulus were displayed at 60Hz. The tree stimulus was shown at 30 Hz and is repeated twice during each 20s epoch.

These movies were collected [insert info on sources], and are distributed alongside the data.

Maximum entropy modeling We followed the data-driven algorithm introduced in [76] for our maximum entropy modeling. This data-driven algorithm is a quasi-Newton method that allows for inference of model parameters $\mathbf{X}$, in our case time dependent fields h_i^t and couplings J_{ij}, without needing to compute the inverse model susceptibility matrix $\chi^{-1}[\mathbf{X}]$ at each time point during the learning dynamics.

As in [50] we learn a maximum entropy model with time-varying fields. Specifically, we learn a model of form

$$P(\sigma_1^t ... \sigma_N^t) \propto \exp\left(-\sum_i^N h_i^t \sigma_i^t - \sum_{i<j}^N J_{ij}\sigma_i^t \sigma_j^t\right). \quad (2)$$

The time dependent fields h_i^t capture the time-varying firing rate of cells σ_i in response to the stimulus. This means that stimulus dependent correlations between cells are absorbed into the fields, leaving the couplings J_{ij} to encapsulate the noise correlations between cells.

For several 20 cell groups, we validated model fits by comparing couplings in the first half and second half of each stimulus and ensuring couplings remained stable. To do this, we separately trained models on each half of the data.

All fits were done on groups of 20 cells, which were all subsets of the full population of 93 cells. These subsets were chosen at random

We additionally use sparse models to generate 3. For the independent model, we fit to the same fitting target, but constrained all couplings to be zero. For the backbone model, we first fit a full model with all couplings and fields. Then, we re-fit, constraining all but the top 10% of couplings from the first fitting to be zero.

DCA models When we were looking for a model that would give us access to the couplings in the white noise stimulus, we were faced with the lack of stimulus repeats. The stimulus-dependant maximum entropy model fundamentally relies on measurements of noise correlations, and as such does not work without stimulus repeats. However, we also wished for a method that would allow us to infer couplings that are only representative of the noise correlations, not couplings that included both shared stimulus inputs and noise correlations. In general, this is not possible, and standard maximum entropy model are not up to the task.

Here, however, we took advantage of a feature of this data: many of our cells have highly overlapping receptive fields. This means that two cells with similar RFs can be correlated two ways: because of a noise correlation, or because of shared stimulus drive. In particular, if several

cells are driven in the same way by the stimulus, they will all have correlations with each other, and therefore many methods would infer 'loops' of couplings.

However, in the protein community, DCA (Direct Coupling Analysis) was developed as a maximum-entropy technique with an emphasis on ignoring indirect couplings (i.e., breaking loops), and a prior asking for a sparse coupling matrix. This means that in our case with several cells all driven in the same way by the stimulus, many of those correlations will be dropped in favor of a sparser explanation for population activity, and in particular one should expect the remaining couplings to be the strongest correlations - those where there is both a biological coupling between the cells and a shared receptive field. This means that we do not expect quantitative agreement between this method and stimulus-dependant maximum entropy, but that we can hope to find the same backbone of strong couplings.

We do not expect this method to be perfect or to be generally applicable. Here, however, the fact that we obtained highly similar couplings to those found from stimulus dependant maximum entropy (a task at which stimulus independent maximum entropy models fail entirely) is proof in and of itself that this method was reasonably successful in our context.

In order to fit the DCA mdoel, we followed methods discussed in [56]. We chose a gauge where neural silence/activity are described by {0,1} to more easily relate the DCA network to couplings fit from the time dependent models.

Decoding scene identity

For the scene identity decoding introduced in Fig 3, we used a Bayesian approach to measure $P(\text{movie}|\text{spikes})$,

$$P(\text{movie}|\text{spikes}) = \frac{P(\text{spikes}|\text{movie})P(\text{movie})}{P(\text{spikes})}. \quad (3)$$

We generate sample spike trains from the learned probability distributions for each of the five movies to test these probabilities. As our models of spike probabilities are equilibrium models, with no relationships between consecutive states, we do this by sampling from the distribution $P(\vec{\sigma}) = \frac{1}{T}\sum_t^T P(\vec{\sigma}^t|t)$ independently for each state in our simulated spike train.

For each sampled spike train, we can calculate $P(\text{spikes}|\text{movie}_\alpha)$, where α indexes movies, simply by plugging into the probability distributions that define our models. We set a uniform prior, $P(\text{movie}) = 1/5$. Finally, $P(\text{spikes}) = \sum_\alpha P(\text{movie}_\alpha)P(\text{spikes}|\text{movie}_\alpha)$.

For each combination of movie and model choice, we generate 1000 spike trains. In Fig 3a, we show the median performance for each model at decoding, and the quartiles for the full and independent models.

Spike Triggered ICA To compute the Spike Triggered ICA (ST-ICA) we follow methods developed in [65].

We first compute the spike triggered average for each cell in a natural cubic spline basis. This is a common method to reduce the number of parameters needed for the model and ensure that the resulting receptive fields are smooth in space and time. We choose the number of splines such that the log-likelihood on held out white noise data is maximized.

To further reduce the number of parameters, we assume our receptive field is rank 1, it can be separated into a spatial filter and a temporal filter. Following these assumptions we use SVD to find this rank 1 approximation. This provably minimizes reconstruction error under the Frobenius norm. We then crop the spatial dimensions for each cell to the regions containing the receptive fields and convolve the stimulus with the temporal filters, which leaves only spatial degrees of freedom.

For each cell, we then have a matrix of size N spikes by M features, where each feature is a spatial pixel convolved with time filter. We use Preconditioned ICA [77], an algorithm for ICA that uses preconditioned L-BFGS, a low memory quasi newton optimization algorithm, for optimization to estimate 20 independent components. Resulting components were considered proper subunit candidates based on the presence of significant spatial autocorrelations, following methods in [66].

With a list of candidate subunits for each cell we then computed the activation of that subunit by projecting the time convolved stimulus onto each filter identified by ST-ICA. Two units were considered the same following methods developed in [78].

I. Acknowledgements

This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030); and by the National Institutes of Health BRAIN initiative (R01EB026943).

[1] Horace B Barlow. Summation and inhibition in the frog's retina. *The Journal of physiology*, 119(1):69, 1953.

[2] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959.

[3] Haldan K Hartline. The receptive fields of optic nerve fibers. *American Journal of Physiology-Legacy Content*, 130(4):690–699, 1940.

[4] David K Warland, Pamela Reinagel, and Markus Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of neurophysiology*, 78(5):2336–2350, 1997.

[5] William Bialek, Fred Rieke, Robert van Steveninck, and David Warland. Reading a neural code. *Advances in neural information processing systems*, 2, 1989.

[6] Tim Gollisch and Markus Meister. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–164, 2010.

[7] Stephen A Baccus, Bence P Ölveczky, Mihai Manu, and Markus Meister. A retinal circuit that computes object motion. *Journal of Neuroscience*, 28(27):6807–6817, 2008.

[8] Iman H Brivanlou, David K Warland, and Markus Meister. Mechanisms of concerted firing among retinal ganglion cells. *Neuron*, 20(3):527–539, 1998.

[9] Jonathon Shlens, Fred Rieke, and EJ Chichilnisky. Synchronized firing in the retina. *Current opinion in neurobiology*, 18(4):396–402, 2008.

[10] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[11] Luisa Ramirez and William Bialek. Compression as a path to simplification: Models of collective neural activity. *arXiv preprint arXiv:2112.14334*, 2021.

[12] David B Kastner, Stephen A Baccus, and Tatyana O Sharpee. Critical and maximally informative encoding between neural populations in the retina. *Proceedings of the National Academy of Sciences*, 112(8):2533–2538, 2015.

[13] Niru Maheswaranathan, Lane T McIntosh, David B Kastner, Josh B Melander, Luke Brezovec, Aran Nayebi, Julia Wang, Surya Ganguli, Stephen A Baccus, and Stanford University. Deep learning models reveal internal structure and diverse computations in the retina under natural scenes. *BioRxiv*, page 340943, 2018.

[14] Vicente Botella-Soler, Stéphane Deny, Georg Martius, Olivier Marre, and Gašper Tkačik. Nonlinear decoding of a complex movie from the mammalian retina. *PLoS computational biology*, 14(5):e1006057, 2018.

[15] Manuel Molano-Mazon, Arno Onken, Eugenio Piasini, and Stefano Panzeri. Synthesizing realistic neural population activity patterns using generative adversarial networks. *arXiv preprint arXiv:1803.00338*, 2018.

[16] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, 2019.

[17] Olivier Marre, Dario Amodei, Nikhil Deshmukh, Kolia Sadeghi, Frederick Soo, Timothy E Holy, and Michael J Berry. Mapping a complete neural population in the retina. *Journal of Neuroscience*, 32(43):14859–14873, 2012.

[18] Antal Berényi, Zoltán Somogyvári, Anett J Nagy, Lisa Roux, John D Long, Shigeyoshi Fujisawa, Eran Stark, Anthony Leonardo, Timothy D Harris, and György Buzsáki. Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *Journal of neurophysiology*, 111(5):1132–1149, 2014.

[19] Carolina Mora Lopez, Srinjoy Mitra, Jan Putzeys, Bogdan Raducanu, Marco Ballini, Alexandru Andrei, Simone Severi, Marleen Welkenhuysen, Chris Van Hoof, Silke Musa, et al. 22.7 a 966-electrode neural probe with 384 configurable channels in 0.13 μm soi cmos. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 392–393. IEEE, 2016.

[20] Nicholas A Steinmetz, Cagatay Aydin, Anna Lebedeva, Michael Okun, Marius Pachitariu, Marius Bauza, Maxime Beau, Jai Bhagat, Claudia Böhm, Martijn Broux, et al. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, 372(6539):eabf4588, 2021.

[21] Robert W Rodieck. *The first steps in seeing.* Sinauer Associates, 1998.

[22] Daniel L Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. *Physical review letters*, 73(6):814, 1994.

[23] Odelia Schwartz and Eero P Simoncelli. Natural signal statistics and sensory gain control. *Nature neuroscience*, 4(8):819–825, 2001.

[24] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345, 1995.

[25] J Hans van Hateren and Dan L Ruderman. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1412):2315–2320, 1998.

[26] Adrienne L Fairhall, Geoffrey D Lewen, William Bialek, and Robert R de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412(6849):787–792, 2001.

[27] Joseph J Atick and A Norman Redlich. What does the retina know about natural scenes? *Neural computation*, 4(2):196–210, 1992.

[28] Ilya Nemenman, Geoffrey D Lewen, William Bialek, and Rob R de Ruyter van Steveninck. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS computational biology*, 4(3):e1000025, 2008.

[29] Jelena Jovancevic-Misic and Mary Hayhoe. Adaptive gaze control in natural environments. *Journal of Neuroscience*, 29(19):6234–6238, 2009.

[30] Maxime JY Zimmermann, Noora E Nevala, Takeshi Yoshimatsu, Daniel Osorio, Dan-Eric Nilsson, Philipp Berens, and Tom Baden. Zebrafish differentially process color across visual space to match natural scenes. *Current Biology*, 28(13), 2018.

[31] Jerome Y Lettvin, Humberto R Maturana, Warren S McCulloch, and Walter H Pitts. What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47(11):1940–1951,

1959.

[32] Bence P Ölveczky, Stephen A Baccus, and Markus Meister. Segregation of object and background motion in the retina. *Nature*, 423(6938):401–408, 2003.

[33] Greg Schwartz, Sam Taylor, Clark Fisher, Rob Harris, and Michael J Berry. Synchronized firing among retinal ganglion cells signals motion reversal. *Neuron*, 55(6):958–969, 2007.

[34] Nishal P Shah, Nora Brackbill, Colleen Rhoades, Alexandra Kling, Georges Goetz, Alan M Litke, Alexander Sher, Eero P Simoncelli, and EJ Chichilnisky. Inference of nonlinear receptive field subunits with spike-triggered clustering. *eLife*, 9:e45743, mar 2020.

[35] Eric Y Chen, Janice Chou, Jeongsook Park, Greg Schwartz, and Michael J Berry. The neural circuit mechanisms underlying the retinal response to motion reversal. *Journal of Neuroscience*, 34(47):15557–15575, 2014.

[36] Greg Schwartz, Rob Harris, David Shrom, and Michael J Berry. Detection and prediction of periodic patterns by the retina. *Nature neuroscience*, 10(5):552–554, 2007.

[37] MTT Wong-Riley. Synaptic organization of the inner plexiform layer in the retina of the tiger salamander. *Journal of neurocytology*, 3(1):1–33, 1974.

[38] Dwight A Burkhardt, Patrick K Fahey, and Michael A Sikora. Natural images and contrast encoding in bipolar cells in the retina of the land-and aquatic-phase tiger salamander. *Visual neuroscience*, 23(1):35–47, 2006.

[39] Jared Salibury and Stephanie E. Palmer. A dynamic scale-mixture model of motion in natural scenes.

[40] Kerry J Kim and Fred Rieke. Slow na+ inactivation and variance adaptation in salamander retinal ganglion cells. *Journal of Neuroscience*, 23(4):1506–1516, 2003.

[41] Fred Rieke. Temporal contrast adaptation in salamander bipolar cells. *Journal of neuroscience*, 21(23):9445–9454, 2001.

[42] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[43] Jonathan W Pillow and Eero P Simoncelli. Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of vision*, 6(4):9–9, 2006.

[44] Einat Granot-Atedgi, Gašper Tkačik, Ronen Segev, and Elad Schneidman. Stimulus-dependent maximum entropy models of neural population codes. *PLoS computational biology*, 9(3):e1002922, 2013.

[45] Elad Ganmor, Ronen Segev, and Elad Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of sciences*, 108(23):9679–9684, 2011.

[46] Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, and Michael J Berry. Searching for collective behavior in a large network of sensory neurons. *PLoS computational biology*, 10(1):e1003408, 2014.

[47] Yasser Roudi, Sheila Nirenberg, and Peter E Latham. Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS computational biology*, 5(5):e1000380, 2009.

[48] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[49] Gašper Tkačik, Jason S Prentice, Vijay Balasubramanian, and Elad Schneidman. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424, 2010.

[50] Ulisse Ferrari, Stéphane Deny, Matthew Chalk, Gašper Tkačik, Olivier Marre, and Thierry Mora. Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons. *Phys. Rev. E*, 98:042410, Oct 2018.

[51] Tim Jarsky, Mark Cembrowski, Stephen M Logan, William L Kath, Hermann Riecke, Jonathan B Demb, and Joshua H Singer. A synaptic mechanism for retinal adaptation to luminance and contrast. *Journal of Neuroscience*, 31(30):11003–11015, 2011.

[52] Jonathan B Demb. Functional circuitry of visual adaptation in the retina. *The Journal of physiology*, 586(18):4377–4384, 2008.

[53] K Nakatani and K-W Yau. Calcium and light adaptation in retinal rods and cones. *Nature*, 334(6177):69–71, 1988.

[54] Oleksandr Sorochynskyi, Stéphane Deny, Olivier Marre, and Ulisse Ferrari. Predicting synchronous firing of large neural populations from sequential recordings. *PLoS computational biology*, 17(1):e1008501, 2021.

[55] Kristina D Simmons, Jason S Prentice, Gašper Tkačik, Jan Homann, Heather K Yee, Stephanie E Palmer, Philip C Nelson, and Vijay Balasubramanian. Transformation of stimulus correlations by the retina. *PLoS computational biology*, 9(12):e1003344, 2013.

[56] Martin Weigt, Robert A White, Hendrik Szurmant, James A Hoch, and Terence Hwa. Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72, 2009.

[57] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366, 2006.

[58] Eric Y Chen, Olivier Marre, Clark Fisher, Greg Schwartz, Joshua Levy, Rava Azeredo da Silveira, and Michael J Berry. Alert response to motion onset in the retina. *Journal of Neuroscience*, 33(1):120–132, 2013.

[59] Matías A Goldin, Baptiste Lefebvre, Samuele Virgili, Mathieu Kim Pham Van Cang, Alexander Ecker, Thierry Mora, Ulisse Ferrari, and Olivier Marre. Context-dependent selectivity to natural images in the retina. *Nature Communications*, 13(1):5556, 2022.

[60] Tim Gollisch. Features and functions of nonlinear spatial integration by retinal ganglion cells. *Journal of Physiology-Paris*, 107(5):338–348, 2013.

[61] Greg Schwartz and Fred Rieke. Nonlinear spatial encoding by retinal ganglion cells: when $1+1 \neq 2$. *Journal of General Physiology*, 138(3):283–290, 2011.

[62] Adrienne L Fairhall, C Andrew Burlingame, Ramesh Narasimhan, Robert A Harris, Jason L Puchalla, and Michael J Berry. Selectivity for multiple stimulus features in retinal ganglion cells. *Journal of neurophysiology*, 96(5):2724–2738, 2006.

[63] Hiroki Asari and Markus Meister. Divergence of visual channels in the inner retina. *Nature neuroscience*, 15(11):1581–1589, 2012.

[64] Hiroki Asari and Markus Meister. The projective field of retinal bipolar cells and its modulation by visual context. *Neuron*, 81(3):641–652, 2014.

[65] Holger; Schultz Simon R. Saleem, Aman; Krapp. Receptive field characterization by spike-triggered independent component analysis. *Journal of Vision*, 8(13), 2008.

[66] Jian K Liu, 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(1):1–14, 2017.

[67] Sheila Nirenberg, Steve M Carcieri, Adam L Jacobs, and Peter E Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411(6838):698–701, 2001.

[68] Rajeev V Rikhye and Mriganka Sur. Spatial correlations in natural scenes modulate response reliability in mouse visual cortex. *Journal of Neuroscience*, 35(43):14661–14680, 2015.

[69] Emmanouil Froudarakis, Philipp Berens, Alexander S Ecker, R James Cotton, Fabian H Sinz, Dimitri Yatsenko, Peter Saggau, Matthias Bethge, and Andreas S Tolias. Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. *Nature neuroscience*, 17(6):851–857, 2014.

[70] Uri Hasson, Rafael Malach, and David J Heeger. Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, 14(1):40–48, 2010.

[71] Jason L Puchalla, Elad Schneidman, Robert A Harris, and Michael J Berry. Redundancy in the population code of the retina. *Neuron*, 46(3):493–504, 2005.

[72] Elad Ganmor, Ronen Segev, and Elad Schneidman. A thesaurus for a neural population code. *Elife*, 4:e06134, 2015.

[73] Daniel S Friend and Norton B Gilula. Variations in tight and gap junctions in mammalian tissues. *The Journal of cell biology*, 53(3):758–776, 1972.

[74] Alejandro Peinado, Rafael Yuste, and Lawrence C Katz. Gap junctional communication and the development of local circuits in neocortex. *Cerebral Cortex*, 3(5):488–498, 1993.

[75] Ye Li, Hui Lu, Pei-lin Cheng, Shaoyu Ge, Huatai Xu, Song-Hai Shi, and Yang Dan. Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature*, 486(7401):118–121, 2012.

[76] Ulisse Ferrari. Learning maximum entropy models from finite-size data sets: A fast data-driven algorithm allows sampling from the posterior distribution. *Physical Review E*, 94(2):023301, 2016.

[77] Jean-Francois Cardoso Pierre Ablin and Alexandre Gramfort. Faster independent component analysis by preconditioning with hessian approximations. *IEEE Transactions on Signal Processing*, 66(15):4040–4049, 2018.

[78] Shanshan Jia, Zhaofei Yu, Arno Onken, Yonghong Tian, Tiejun Huang, and Jian K Liu. Neural system identification with spike-triggered non-negative matrix factorization. *IEEE Transactions on Cybernetics*, 2021.