You might find this additional information useful...

This article cites 43 articles, 26 of which you can access free at:
http://jn.physiology.org/cgi/content/full/96/5/2724#BIBL

This article has been cited by 2 other HighWire hosted articles:

Single neuron computation: from dynamical system to feature detector.
S. Hong, B. Aguera y Arcas and A. L. Fairhall
Neural Comput., December 1, 2007; 19 (12): 3133-3172.
[Abstract] [Full Text] [PDF]

Limits of Linear Rate Coding of Dynamic Stimuli by Electroreceptor Afferents
D. Gussin, J. Benda and L. Maler
J Neurophysiol, April 1, 2007; 97 (4): 2917-2929.
[Abstract] [Full Text] [PDF]

Updated information and services including high-resolution figures, can be found at:
http://jn.physiology.org/cgi/content/full/96/5/2724

Additional material and information about *Journal of Neurophysiology* can be found at:
http://www.the-aps.org/publications/jn

This information is current as of March 27, 2008 .

Selectivity for Multiple Stimulus Features in Retinal Ganglion Cells

Adrienne L. Fairhall,[1] C. Andrew Burlingame,[2] Ramesh Narasimhan,[1] Robert A. Harris,[3] Jason L. Puchalla,[2] and Michael J. Berry II[2]

[1]*Department of Physiology and Biophysics, University of Washington, Seattle, Washington;* [2]*Department of Molecular Biology, Princeton University, Princeton, New Jersey;* [3]*Department of Life Sciences, University of Sussex, United Kingdom*

Fairhall, Adrienne L., C. Andrew Burlingame, Ramesh Narasimhan, Robert A. Harris, Jason L. Puchalla, and Michael J. Berry II. Selectivity for multiple stimulus features in retinal ganglion cells. *J Neurophysiol* 96: 2724–2738, 2006. First published August 16, 2006; doi:10.1152/jn.00995.2005. Under normal viewing conditions, retinal ganglion cells transmit to the brain an encoded version of the visual world. The retina parcels the visual scene into an array of spatiotemporal features, and each ganglion cell conveys information about a small set of these features. We study the temporal features represented by salamander retinal ganglion cells by stimulating with dynamic spatially uniform flicker and recording responses using a multielectrode array. While standard reverse correlation methods determine a single stimulus feature—the spike-triggered average—multiple features can be relevant to spike generation. We apply covariance analysis to determine the set of features to which each ganglion cell is sensitive. Using this approach, we found that salamander ganglion cells represent a rich vocabulary of different features of a temporally modulated visual stimulus. Individual ganglion cells were sensitive to at least two and sometimes as many as six features in the stimulus. While a fraction of the cells can be described by a filter-and-fire cascade model, many cells have feature selectivity that has not previously been reported. These reverse models were able to account for 80–100% of the information encoded by ganglion cells.

INTRODUCTION

Recent technical advances allow recordings to be made from essentially every ganglion cell in a patch of retina during naturalistic stimulation (Frechette et al. 2004; Segev et al. 2004). This provides the opportunity for unprecedented understanding of the coding properties of ganglion cells, individually and as a population. Retinal image processing is sophisticated with many complex visual patterns modulating the firing of individual ganglion cells, such as direction of motion (Barlow and Levick 1965), object motion against a background (Lettvin et al. 1959; Olveczky et al. 2003), looming objects (Ishikane et al. 1999), static edges (Caldwell and Daw 1978), spatial generalization over local subunits (Victor and Shapley 1979), and deviations from temporal periodicity (Bullock et al. 1990). Systematic studies of the ganglion cell population using multielectrode arrays have shown great functional diversity (DeVries and Baylor 1997; Schneidman et al. 2002) and extensive receptive field overlap (Segev et al. 2004), allowing high-order visual features to be represented by multineuronal firing patterns (Puchalla et al. 2005; Schnitzer and Meister 2003).

Many studies have focused on measurement of the spatiotemporal receptive field, an approximation of ganglion cell function that implicitly assumes that the cell represents only one visual feature. However, retinal ganglion cell spike trains can represent multiple stimulus features, so more complete models are needed. The computation of Wiener kernels provides an estimate of the firing rate in an expansion in successive orders of the input (Wiener 1958). In practice, however, data limitations restrict the computation of the kernels to second order (Naka and Sakai 1991). Cascade models, where the stimulus is filtered through a linear stage then nonlinearly thresholded, provide a general framework for *forward* models, aimed at predicting a spike train given the stimulus. Elaborations of the cascade model include a further linear feedback stage that modulates the firing output depending on recent spike activity, thus including the effects of refractoriness (Gerstner and Kistler 2002; Keat et al. 2001; Paninski et al. 2004; Shapley and Victor 1979; Victor 1987, 1988). The linear filtering stage models the receptive field, whereas the threshold and the second linear filter are associated with the mechanism of spike generation.

Here, we apply second-order reverse correlation methods to discover new, detailed information about receptive field structure, in particular, the representation of multiple visual features by each spike (Aguera y Arcas et al. 2003; Brenner et al. 2000a; de Ruyter van Steveninck and Bialek 1988; Schwartz et al. 2002). While considering the units of representation in the spike train to consist of sequences of spikes has been shown to be fruitful (Berry and Meister 1998; Metzner et al. 1998; Reinagel and Reid 1998; Reinagel et al. 1999; Strong et al. 1998), here we will limit our discussion to the representation of the stimulus by single spikes only. *Reverse* models, describing what a spike reveals about the stimulus, offer a simple interpretation of the neural code of a ganglion cell: the visual features they identify directly illustrate the meaning of the neuron's spikes (Rieke et al. 1997). At the same time, the reverse model can be inverted using Bayes' rule to make a prediction of the neuron's output for arbitrary visual inputs (Aguera y Arcas et al. 2003; Brenner et al. 2000a). Furthermore, reverse models allow the use of information theory as an error measure that is well-matched to the aspects of a ganglion cell spike train that are important for conveying visual information and makes few other assumptions about the nature of the neural code (de Ruyter van Steveninck et al. 1997; Gawne and Richmond 1993; Panzeri and Schultz 2001; Panzeri et al. 1999; Reich et al. 2001; Reinagel and Reid 1998; Schneidman et al. 2003).

Address for reprint requests and other correspondence: M. J. Berry II, Dept. of Molecular Biology, Princeton University, Princeton, NJ 08544 (E-mail: berry@princeton.edu).

M E T H O D S

Physiological recording

Eyes were dissected from larval tiger salamanders (*Ambystoma tigrinum*) that had been light adapted for 1–3 h. The cornea was removed, and the remaining eyecup was cut into sections of quarter size or less. This preparation left the pigment epithelium and sclera intact for better stability during electrical recordings. Sections were placed with the ganglion cells facing a multi-electrode array (MultiChannel Systems MEA 60) and were perfused with oxygenated Ringer medium at room temperature (Balasubramanian and Berry 2002). Stable recordings of >12 h were achieved under these conditions. For a given stimulus condition, between 2 and 5 h of data were acquired.

Extracellular voltages were amplified (1,200-fold), filtered (200–5,000 Hz), and streamed to disk for off-line analysis at 10 KSamples · s^{-1} · channel^{-1}. Spike waveforms were sorted using the spike peak and width in a 2.5-ms window. Only well-isolated spike waveforms were used: we required that spike clusters were clearly visible above the noise and that the sorted spike trains had <0.3% and more typically between 0 and 0.1% of all interspike intervals <2 ms. This study is based on measurements of 59 cells recorded from seven retinas.

Visual stimulation

Retinas were stimulated with spatially uniform flicker presented using a white light-emitting diode (LED) diffused over the area of the array. Light intensities were chosen randomly at 120 or 180 Hz from a Gaussian distribution with a SD equal to 24–36% of the mean. Mean light levels ranged from 12 to 24 mW/m^2, corresponding to photopic vision. The relative absorption ratios for salamander red rods:red cones:blue cones are 1:0.96:0.61.

We stimulated the retina with white-noise segments of length 30 s repeated 10 times (5 min total), interspersed with random sequences of length 20 min. This basic pattern was repeated, using different random sequences but the same repeat sequence, for ≤6 h. The analysis of a given cell was restricted to the portion of the experiment where the cell's mean firing rate was approximately constant.

Model simulation

We implemented the filter-and-fire model described in Keat et al. (2001). There are two noise processes, one of which is additive and is summed with the filtered stimulus. This noise was Gaussian with a mean of zero and a SD of 0.15 (in units of stimulus SD). There is also a multiplicative noise which independently multiplies the amplitude of the afterhyperpolarization following each spike; this was a Gaussian variable with mean of one and SD of 0.085. The afterhyperpolarization used in the simulations shown was given by $A(t) = -0.6$ $\exp(-t/0.44s)$, and the filters were constructed as a sum of two alpha functions with timescales of order several hundred milliseconds. The threshold Θ took values of 1.5, 2, and 2.5. For each analysis, 30,000 spikes were used.

Covariance analysis

To estimate the temporal response properties of ganglion cells with a high degree of generality, we stimulated the retina with the wide variety of light patterns found in spatially uniform flicker and used covariance analysis to determine to which visual features each ganglion cell was sensitive. During this stimulus, denoted by $s(t)$, we recorded the occurrence time of every spike elicited from a ganglion cell, $\{t_i\}$, and selected the stimulus histories preceding every spike, $s_i(\tau) = s(t_i - \tau)$, where τ denotes the time index. We used 100 stimulus values preceding a spike. Thus for stimuli presented at a frame rate of 120 Hz, τ extended back to 833 ms before a spike,

whereas for 180-Hz stimuli, τ extended back to 556 ms before a spike. The mean, $\bar{s}_i(\tau)$, is the spike-triggered average. The covariance matrix is formed from the outer product of the stimulus histories averaged across all spikes

$$C_{jk} = \langle [s_i(\tau_j) - \bar{s}_i(\tau_j)][s_i(\tau_k) - \bar{s}_i(\tau_k)] \rangle_i$$

where the average $\langle \ldots \rangle$ is over all spike occurrences $\{t_i\}$.

We were interested in learning what features in the stimulus are relevant to spiking. These are the features for which the variance is altered from that of the prior distribution of stimuli, which is Gaussian along all stimulus dimensions, independent of the response. Thus we subtracted the covariance matrix of the prior itself, C^{prior}, to obtain a matrix representing the covariance *differences* (Brenner et al. 2000a)

$$\hat{C}_{jk} = C_{jk} - C_{jk}^{\text{prior}}$$

To find the relevant features, we diagonalized $\hat{C}$ to find its eigenmodes and corresponding eigenvalues. Each eigenmode can be thought of as a visual "feature" that may drive ganglion cell spiking, and its eigenvalue is equal to the variance of stimuli along this feature axis when a spike occurs. Because the prior stimulus distribution was subtracted from the covariance matrix, an eigenvalue of zero corresponds to the prior variance and indicates that the corresponding visual feature was not relevant to ganglion cell spiking. The number of significant eigenvalues gives an estimate for the dimensionality of the relevant stimulus space, and the corresponding eigenmodes span that space. Therefore the response function of the neuron may be completely determined by sampling in this subspace.

The significance of eigenvalues depends on sampling. We assessed eigenvalue significance by computing the eigenvalue spectrum as a function of the number of spikes N for a range of data fractions (Aguera y Arcas et al. 2003). Eigenvalues which were stable with respect to N were judged to be significant.

Information conveyed by the spike train

The information in the spike train can be evaluated directly using methods introduced in Strong et al. (1998) and Brenner et al. (2000b). The method of Strong et al. (1998) uses the distribution of spike words created from discretizing the spike train in time bins and concatenating the spike counts in these bins into a spike word; by varying the length of the spike words and the size of the time bin, this method can evaluate the information encoded by a neuron in the limit of infinite word length or zero time bin.

As we are evaluating how much information the observation of each spike gives about the stimulus ensemble, we used instead the method of Brenner et al. (2000b) and measured the information that the time of occurrence of a *single* spike conveys about the visual stimulus. To do this, we randomly generated a stimulus segment $s(t)$ of length $T = 30$ s from the ensemble of spatially uniform flicker, S. We repeated the same pseudo-random sequence many times to find $r(t)$, the time-varying firing rate of a ganglion cell responding to $s(t)$. These repetitions were interspersed between changing random sequences that were used to determine the overall mean firing rate $\bar{r}$ in response to the entire stimulus ensemble S. The information per second is given by the entropy difference between the response, $\bar{r}$, where the stimulus is an unknown sequence drawn from the ensemble S, and the response, $r(t)$, when the particular stimulus sequence $s(t)$ of duration T is known (Brenner et al. 2000b). We then computed

$$I_{\text{one spike}} = \frac{1}{r_{\text{mean}}T} \int_0^T dt\, r(t) \log_2 \frac{r(t)}{\bar{r}} \qquad (1)$$

where r_{mean} is the mean firing rate in response to the stimulus segment $r(t)$. Notice that *Eq. 1* involves two different average firing rates, r_{mean} and $\bar{r}$. The rate $\bar{r}$ appears in the logarithm because this is the best

estimate of the firing rate averaged over the entire stimulus ensemble *S*. We divide by a factor of r_{mean}, because the time integral in *Eq. 1* is taken over the repeated stimulus segment, and the average firing rate over this segment is r_{mean}. This puts the information in units of bits per spike rather than bits per second. See APPENDIX for a derivation of *Eq. 1*.

To compare this information benchmark against the information captured by models of a ganglion cell's response function, we needed to match r_{mean} to $\bar{r}$. To this end, we randomly selected stimulus segments of duration slightly shorter than *T* (ranging from 0.75 *T* to 0.95 *T*), each having a different mean firing rate, and recalculated the information for all such segments. We plotted $I_{one\ spike}$ versus r_{mean} and linearly interpolated to find the information for $r_{mean} \rightarrow \bar{r}$. We took the uncertainty in this interpolation to be our random error in measuring $I_{one\ spike}$; this random error was quite small, typically ~1–2%.

Values were corrected for bias due to the finite number of stimulus repeats by dividing the data into fractions with respect to the number of repeats, repeating the calculation, and computing the information as a function of 1/(sample size). We obtained the best quadratic fit to this dependence and extrapolated to infinite sample size (Golomb et al. 1997; Strong et al. 1998). The quadratic dependence gave a good fit for all cells that we have included.

This information measure makes no assumptions about the relationship between stimulus and response, and is an upper bound for the information captured by any particular low-dimensional model of this response function (Adelman et al. 2003; Aguera y Arcas et al. 2003). Because ganglion cell spike trains were sparse and precise, each spike carried a great deal of information: 4.9 ± 1.0 (SD) bits.

Information captured by the model

To compare the model with the directly computed information content of the spike train, we follow the argument of (Aguera y Arcas et al. 2003): starting with *Eq. 1*, we note that

$$\frac{r(t)}{\bar{r}} = \frac{P(\text{spike at } t | \vec{s})}{P(\text{spike at } t)} = \frac{P(\vec{s} | \text{spike at } t)}{P(\vec{s})} \quad (2)$$

The first step follows from the definition of $r(t)$, and the second step follows from Bayes' rule (Cover and Thomas 1991). In forming a *K*-dimensional model defined by the filters $\{f_1, \ldots, f_K\}$, we replace the stimulus $\vec{s}$ with its reduced, lower dimensional description in terms of the *K* projections of the stimulus, denoted by s_i, onto the filters f_i

$$\frac{P(\vec{s} | \text{spike at } t)}{P(\vec{s})} \rightarrow \frac{P(s_1, \ldots, s_K | \text{spike at } t)}{P(s_1, \ldots, s_K)}$$

The corresponding information in the *K*-dimensional model is therefore given by

$$I_{one\ spike}^{(K)}(f_1, f_2, \ldots, f_K)$$
$$= \int d^K s P(s_1, \ldots, s_K | \text{spike at } t) \log_2\left[\frac{P(s_1, \ldots, s_K | \text{spike at } t)}{P(s_1, \ldots, s_K)}\right] \quad (3)$$

where the time integral in *Eq. 1* has been replaced by an ensemble integral in *Eq. 3*. By the data-processing inequality (Cover and Thomas 1991), the information in the reduced description must be bounded by that in the full description

$$I_{one\ spike}^{(K)} \leq I_{one\ spike}$$

For reverse models with a single visual feature (1D), we computed this information for all of the significant eigenmodes as well as for the spike-triggered average (STA). For models with two visual features (2D), we computed the information for all combinations of two significant eigenmodes as well as for the joint distribution of the STA with each eigenmode orthogonalized with respect to the STA. The
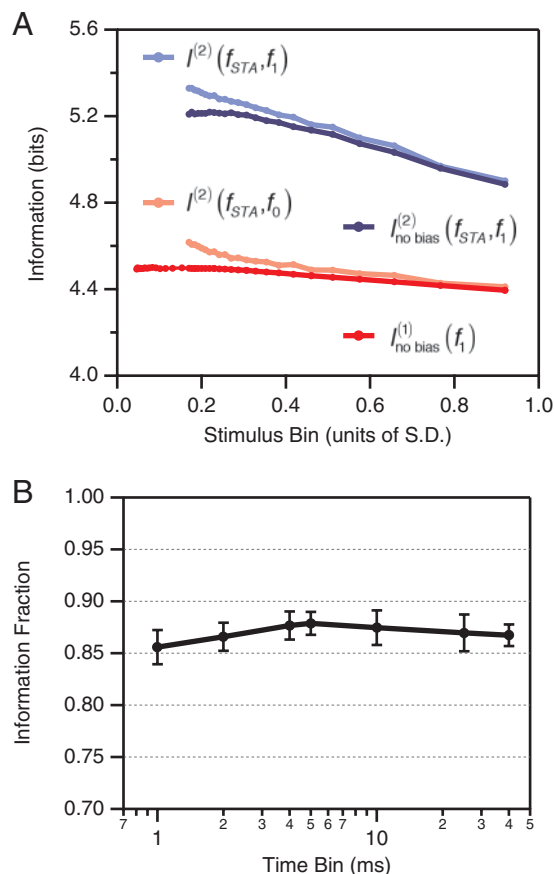

FIG. 1. Information calculations. *A*: we compared the information in a model with two visual features (2D) with the spike-triggered average (STA) and the most insignificant eigenmode $I^{(2)}(f_{STA}, f_0)$ (pink) vs. the bias-corrected information in a 1D model given by the STA, $I_{no\ bias}^{(1)}(f_{STA})$ (red) to determine the sampling bias in a 2D model. This is subtracted from raw 2D models, $I^{(2)}(f_{STA}, f_1)$ (light blue), to achieve a bias-corrected 2D model, $I_{no\ bias}^{(2)}(f_{STA}, f_1)$ (blue). *B*: information fraction plotted as a function of the time bin used in the covariance analysis and information calculations; error bars are SE over *n* = 5 cells.

data shown in Fig. 6 is from the latter case. Treating the filters as vectors, they were normalized such that the squared sum over the vector components equaled unity, and the stimulus was sampled in time bins equal to the frame time so that there were no temporal correlations. Because the prior stimulus distribution was independent in all stimulus directions and had a variance of unity, we used the analytically computed Gaussian prior for both $P(s_1)$ and $P(s_1, s_2)$.

To sample the spike-conditional distributions, we used a range of bin sizes and evaluated the resulting information as a function of bin size, δs. For 1D distributions, the value was very stable as a function of δs, but the 2D models depended more strongly on the bin size. We corrected our information calculation for finite sampling bias by considering the information carried by the eigenmode with an eigenvalue closest to zero; this visual feature should be most completely irrelevant to the firing of the neuron. By itself, this eigenmode should carry no information. Instead, its information was measured to be between 0.002 and 0.014 bits, depending on the number of spikes recorded for that ganglion cell. Therefore this information, as a function of δs, was subtracted from the values for all other information calculations for 1D models. Although this correction typically was small (<1%), it produced information values that were nearly constant down to a bin size of 0.05 times the prior SD (Fig. 1*A*).

To determine the bias for 2D models, we performed our calculation for the STA combined with the most insignificant eigenmode. Similarly, the information in this 2D model should be the same as the

information in the STA alone but was larger due to finite sampling. We took the sampling bias to be equal to the difference between the information in this 2D model and the information about the STA alone, as a function of δs. This bias was subtracted from all of our calculated information in 2D models, giving values that were nearly constant as a function of the bin size, δs (Fig. 1A). For the final value, we averaged over bin sizes in the range of 0.15–0.4 (normalized units); the SD was taken as an estimate of the random error.

We define the *synergy* in the 2D model as the difference in information between the 2D model and the summed information from the corresponding 1D models

$$I_{\text{syn}} = I^{(2)}_{\text{one spike}}(f_1, f_2) - [I^{(1)}_{\text{one spike}}(f_1) + I^{(1)}_{\text{one spike}}(f_2)] \qquad (4)$$

If the projections onto the two single stimulus dimensions were independent, the 2D distribution would simply be the product of the two 1D distributions, and the synergy would be zero. Recall that while the visual features determined by our analysis are necessarily *orthogonal*, the distribution $P(s_1, \ldots, s_K | \text{spike at } t)$ need not be statistically independent.

In all of our information calculations, we used a time bin that was matched to the frame time of the stimulus, either 5.56 or 8.33 ms. The same time bin was used for both full and model information as well as for the covariance analysis. To explore the limit of finer time scales, we ran an experiment in which the stimulus frame rate was 1,000 Hz, but where the temporal sequence of light intensities was low-pass filtered with a cut-off frequency of 120 Hz. This allowed us to extend our covariance analysis down to a time bin of 1 ms. For time bins in the range of 1 ms up to 40 ms, the information fraction captured by our analysis was roughly constant (Fig. 1B).

R E S U L T S

Simple model of ganglion cell function

Before describing the results of our covariance analysis for real ganglion cells, it is helpful to develop intuition using a simple model, where all elements of the spike generation are known and understood (Aguera y Arcas and Fairhall 2003). Keat et al. introduced an instantiation of the cascade model described in the preceding text—a spike response model (Gerstner and Kistler 2002)—which had considerable success in predicting the timing of spikes from a variety of ganglion cells (Keat et al. 2001). Therefore we first performed covariance analysis on this model of ganglion cell function. In the Keat model, the stimulus $s(t)$ is convolved with a single linear filter, $f^+(t)$. We denote this filter with the superscript $+$ to indicate that it has nonzero values only at positive times (due to causality). The output of this filtering stage, $g(t)$, is passed through a static nonlinearity which is simply a threshold, Θ. Each spike adds an afterhyperpolarization, $A(t) = -A_o \exp(-t/\chi)$, to the filtered stimulus $g(t)$. Noise is added both to the filtered stimulus prior to thresholding and to the amplitude of the afterhyperpolarization to account for the timing jitter and spike count variation in experimentally measured spike output.

Covariance analysis, illustrated in Fig. 2, finds the directions in stimulus space along which the variance is altered with respect to the prior stimulus distribution (see METHODS). Each direction (or eigenmode) has a corresponding eigenvalue, which describes how the spike-triggered stimulus variance differs from the prior: negative eigenvalues indicate that the spike-triggered stimulus distribution is narrower than the prior, positive values indicate that it is wider, and zero indicates no difference. The directions with eigenvalue significantly different from zero constitute the visual "features" about which the

occurrence of a spike conveys information. The set of all visual features revealed by covariance analysis constitute what we call a "reverse model" of the ganglion cell light response. These features describe what the brain can infer about the visual stimulus given the occurrence of a spike. This terminology is in contrast to a "forward model," such as that described by Keat et al. that predicts the probability of a spike given the visual stimulus. As we will determine the visual features using reverse correlation, they are indexed in the same sense as the stimulus, and have nonzero values only at negative times (again due to causality). We will therefore denote the visual features by $\{f_i^-(t)\}$ to indicate this sense of the index t.

The neuron fires whenever a decision variable, consisting of the stimulus convolved with the filter $f^+(t)$, plus an afterhyperpolarizing current convolved with the previous spike train, crosses threshold from below (see METHODS). Let us initially neglect the afterhyperpolarization and consider the case that the system simply fires a spike on an upward threshold crossing of $g(t) = \int_0^\infty f^+(\tau) s(t - \tau) d t$. Clearly, because spiking occurs where the filtered stimulus attains or exceeds the threshold value Θ, $f^-(t)$ is itself a relevant feature for spiking: any spike-triggering stimulus projected onto $f^-(t)$ will be close to Θ and will therefore have variance less than the prior and so should appear as a direction with negative eigenvalue.

A second feature is expected due to the threshold crossing: as the threshold is usually crossed from below, the time derivative $g'(t)$ of the filtered stimulus at the time of the spike should be positive (DeWeese 1995). Equivalently, the projection of the stimulus $s(t)$ onto the filter $f^{+\prime}(t)$ should be positive when there is a spike: if we take the time derivative of the filtered stimulus, we get $g'(t) = \int_0^\infty f^+(\tau) s'(t - \tau) d\tau$, where in practice the filter is taken to be zero at some finite time in positive t. Next, we can integrate by parts to get $g'(t) = -f^+(\tau)s(t - \tau)|_{\tau=0}^{\tau=\infty} + \int_0^\infty f^{+\prime}(\tau) s(t - \tau) d\tau$. The boundary term drops out, because the filter $f^+(t)$ is zero both at time 0 and for large positive t. Thus the time derivative of the filtered stimulus $g'(t)$ is seen to be a projection of the stimulus along a direction given by the time derivative of the filter, $f^{+\prime}(t)$. We note that $f^{+\prime}(t) = -f^{-\prime}(-t)]$ because the time derivative operator is anti-symmetric under $t \to -t$. It is important to remember that the eigenmodes identified by covariance analysis are only determined up to an arbitrary sign. Therefore we choose a sign convention such that the second visual feature is the negative time derivative of the first feature. With this choice, the projection of the stimulus onto the second visual feature tends to have positive values at the time of threshold crossing. And as a result, the distribution of these projections given a spike will have less variance than the a priori stimulus distribution and hence appear with negative eigenvalue.

With the afterhyperpolarization, the condition that spikes are only fired on an upward threshold crossing of $g(t)$ is weakened; the *sum* of the filtered stimulus and the combined AHPs from previous spikes, $h(t) = g(t) + \int_0^\infty A(\tau)\rho(t - \tau)d\tau$, must cross threshold, where $\rho(t) = \Sigma_i \delta(t - t_i)$ is the spike train. Thus subsequent spikes can occur either when the filtered stimulus is sufficiently large such that $h(t)$ still exceeds Θ or if $g(t)$ is still above threshold when the AHP decays away. The AHP does not affect the form of the stimulus filtering so $f(t)$ should still be a relevant mode. The dependence of spiking on the time derivative $f'(t)$ will, however, be weakened, as $h(t)$ may now cross threshold while $g(t)$ itself is decreasing.
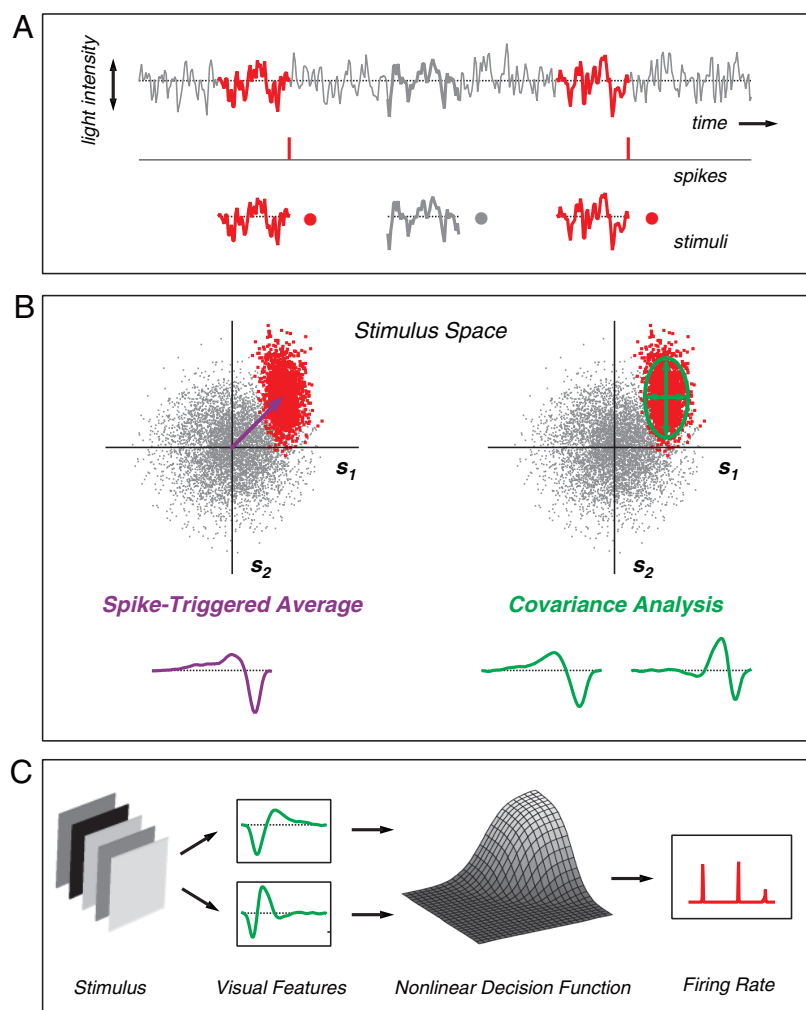
FIG. 2. Covariance analysis. *A*: ganglion cell spikes produced in response to a dynamically varying stimulus, here Gaussian white noise. Stimulus histories preceding spike occurrences are collected (red traces), along with prior stimulus samples, not correlated with the times of spikes (gray traces). Example taken from a real neuron. *B*: stimulus samples plotted in 2 arbitrary directions in stimulus space (gray dots). Viewed in the appropriate coordinate system, the spike-triggered stimuli forms a distinct cloud (red dots). The goal of our analysis is to find this appropriate coordinate system. The STA is the mean of the spike-triggered stimuli (purple); the covariance of the spike-triggered stimuli captures the coordinates of variation of the cloud, giving 2 visual features (green). *C*: equivalent forward model of spike generation filters the stimulus with multiple visual features and passes these values through a nonlinear decision function to produce a predicted firing rate as a function of time.

The filtered stimulus $g(t)$ typically remains above threshold for a time on the order of the correlation time of $g(t)$, ξ. If the time scale χ of the AHP is very long, following an initial spike the AHP can keep the decision variable below threshold for a time longer than ξ, preventing further spikes. For shorter χ, $A(t)$ decays away and $h(t)$ can recover to allow several spikes within time ξ. Thus burst-like events appear as a consequence of the AHP recovery; the burst duration is established by the correlation time of the filtered stimulus, ξ, and the typical burst interspike interval is governed by the AHP time scale, χ. For the first spike in a burst, the original criterion on $f'(t)$ holds. For all subsequent spikes, $h(t)$ will be at threshold while $g(t)$ is larger than threshold, but $g'(t)$ will be first positive and then negative within this time. Thus the relative importance of the derivative feature should depend on the time scale of $h(t)$: for large χ, the AHP case is close to the pure upward-threshold crossing case, and the derivative will be highly significant; for smaller χ, the derivative condition will be weakened.

Indeed, covariance analysis does recover both of these visual features (Fig. 3). The eigenvalues corresponding to these features depend weakly on the threshold, but the features themselves are unaffected by the value of the threshold (Fig. 3, *A* and *B*). The spike-triggered average is precisely a linear combination of the filter and its time derivative. Note that these visual features are the time-reversed versions of the respective filters in a forward model, so that the second feature is actually the negative time derivative of the first feature. To understand the structure of the spike-triggered stimuli in terms of these two features, we plot each spike-triggering stimulus history projected onto features 1 and 2 (see METHODS). This shows clearly the thresholding: the projection onto the first feature, by construction, is always near or greater than threshold, and the projection onto the second feature is mostly positive (Fig. 3*C*).

The interactions between spikes have been shown, for single neurons, to have a profound influence on the output of a white noise analysis (Aguera y Arcas and Fairhall 2003; Aguera y Arcas et al. 2003; Pillow and Simoncelli 2003; Powers et al. 2005). This effect is particularly dramatic for an integrate-and-fire-like neuron, where the integrator is reset after a spike, losing memory of the filtered stimulus. The Keat model does not make this assumption. Here we can control the strength of the interspike interaction with two parameters, the amplitude A_0 and time scale χ of the afterhyperpolarization. Varying the amplitude had no effect on the results of covariance analysis. As expected, when the time scale was reduced sufficiently, the second, derivative-like feature was no longer significant (data not shown). We have found no combinations of spike afterhyperpolarization parameters that give rise to more than two significant visual features.
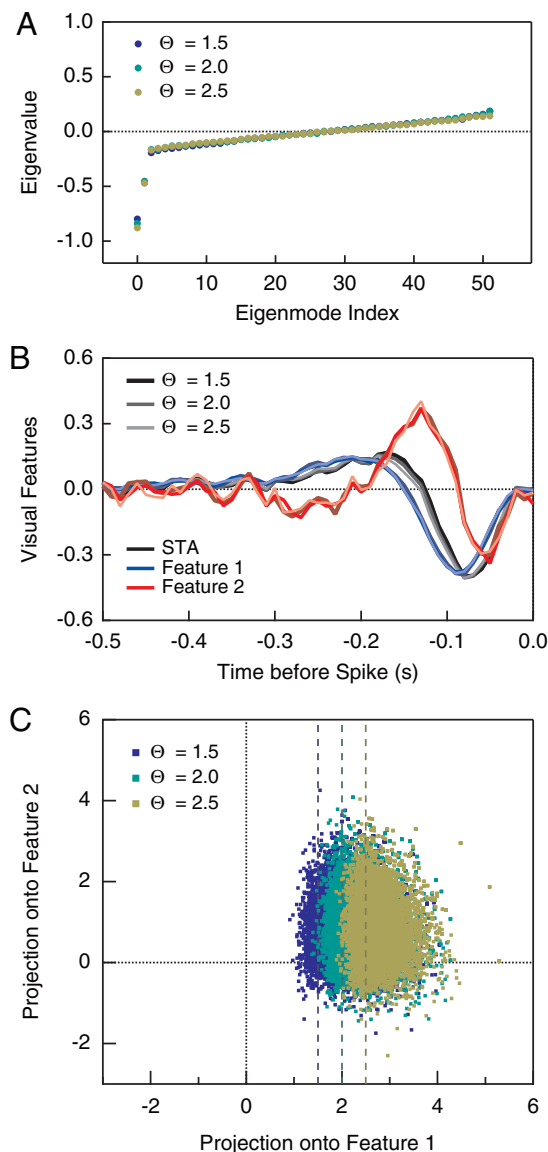
FIG. 3. Covariance analysis of the Keat model. *A*: eigenvalue spectrum in rank order for 3 different values of the threshold parameter, Θ. All other parameters of the model are held fixed. Equal numbers of spikes were used in the 3 cases to ensure comparable noise levels. *B*: STA (black) and 1st 2 modes (blue and red, respectively) are insensitive to the threshold (shown by shades of each color). *C*: projections of spike-triggering stimulus histories onto the 1st and 2nd modes for all 3 values of the threshold (shown by different color).

There are two additional model parameters, governing the amplitude of both additive and multiplicative sources of noise. The noise adds variability in the spike times. As long as the temporal jitter is smaller than the time scale of the filter itself, jitter in spike timing results only in a blurring of the threshold and weak dependence on the derivative mode. Thus there is some dependence of the eigenvalues on the amplitude of the noise, although noise levels appropriate for matching retinal timing jitter do not change the overall structure of the results (see DISCUSSION).

Feature selectivity of ganglion cells

We recorded spike trains from ganglion cells in the salamander retina while stimulating with spatially uniform

flicker, a stimulus ensemble that contains a rich variety of temporal patterns of light (see METHODS). As shown in previous studies, ganglion cell spike trains were found to consist of sparse and reliable firing events under these stimulus conditions (Berry et al. 1997). Covariance analysis revealed a great diversity of temporal features encoded by retinal ganglion cells (Fig. 4). We found at least two and as many as six significant visual features for individual cells. (Because the resolution of eigenvalues was limited by sampling noise, we only included ganglion cells from which we recorded at least ~2,000 spikes.) Within the ganglion cell population, a wide variety of feature selectivities was observed. Although the properties of individual cells did not fall into distinct classes, it is useful to divide the population into five qualitatively different types to describe this diversity. Figure 4 shows an example cell from each such response type. This shorthand description should not be taken as a claim that the ganglion cells could be clustered into five distinct functional classes. Instead, their response functions seemed to be distributed more continuously, with individual cells often having properties intermediate between these five types and the examples shown in Fig. 4 representing the range of possibilities that we observed.

The first type is the filter-and-fire cell (Fig. 4*A*; $n = 11$ cells), where the covariance analysis closely resembled that of the Keat model. These cells had two strongly negative eigenvalues (*left*). The corresponding visual features included one that resembled the STA and a second feature that was proportional to the time derivative of the first feature (*middle*). As discussed in the preceding text, we chose a sign convention in which this second feature was the negative time derivative of the first. Finally, projections of the stimulus onto these visual features at the occurrence time of a spike showed a threshold-like nonlinearity along both feature directions (*right*). Similar to the Keat model, the first feature had a slower time course than the STA. As for the Keat model, the STA is a linear combination of these two visual features, and it is sharpened in time by the addition of the derivative-like feature. All but one of these cells had a spike-triggered stimulus average that was either biphasic OFF or standard OFF-type (Puchalla et al. 2005). These cells have also been described as fast OFF in previous studies (Schnitzer and Meister 2003).

Bimodal cells (Fig. 4*E*; $n = 7$ cells) displayed two separate clouds in the plane formed by the projections onto two visual features. Such cells were strongly ON-OFF in their response to steps of light. If the STA was calculated separately for the two clouds seen in the projection space, one cloud had an ON-type STA and the other cloud was OFF-type (see Fig. 9). For these cells, the leading eigenvalue was positive. As this was the direction along which the variance change from the prior was most increased, this is the direction which maximized the bimodality of the spike-triggered distribution. One can think of this as the direction that best differentiates between the individual responses making up the two clouds. Although this feature is not the same as the STA, the STA was also always effective in separating the two clouds. For the example cell shown in Fig. 4*E*, there were many additional features that show repeated oscillations over a range of frequencies, like a Fourier basis. This structure was often seen for ganglion cells with higher average firing rates. These cells all had a fast OFF-type spike-triggered average, like the filter-and-fire cells.
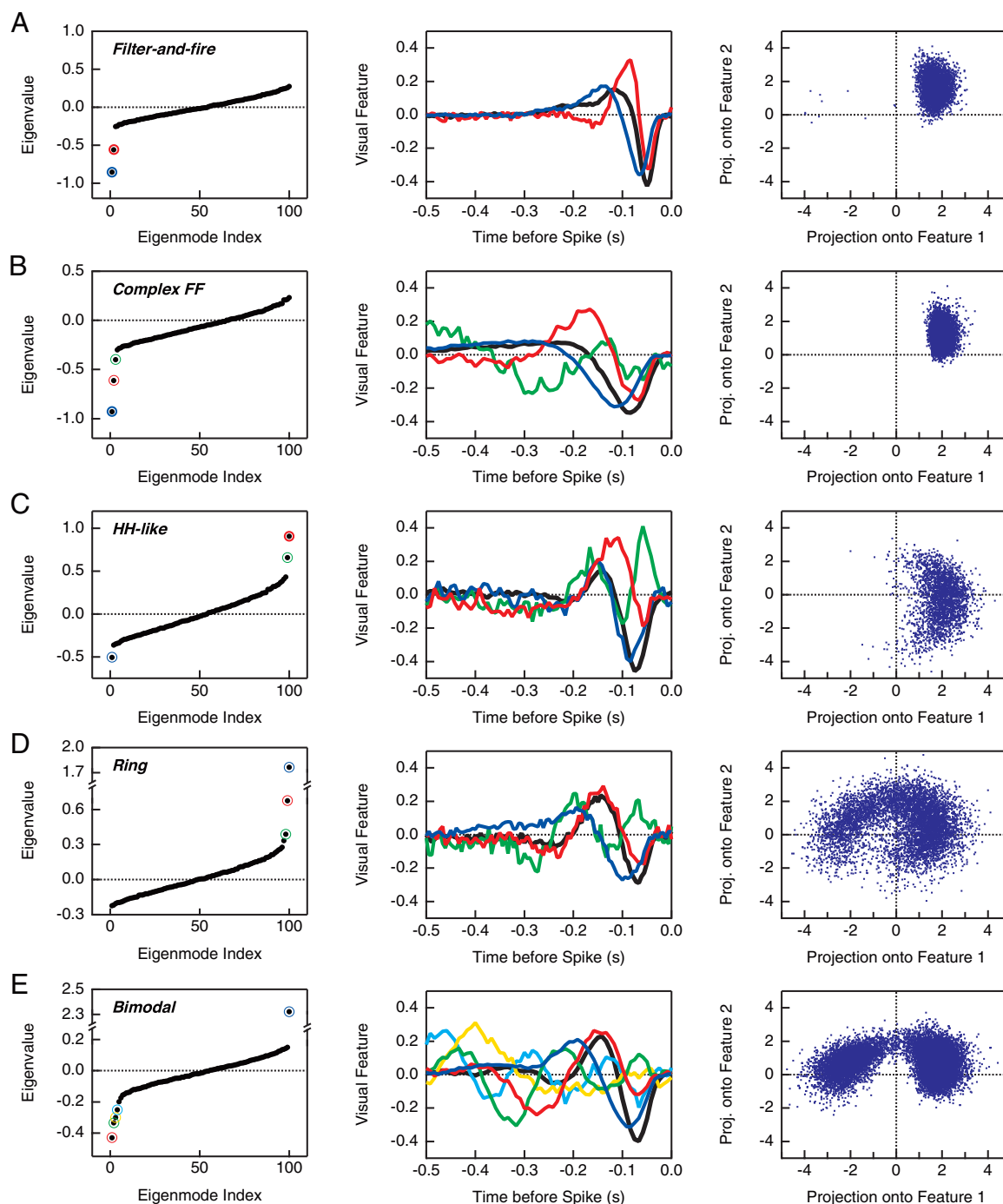
FIG. 4. Covariance analyses for examples of ganglion cells. *Left*: eigenspectrum, with significant modes marked by colored circles. *Middle*: STA (black) and the leading eigenmodes, colored as for their corresponding eigenvalues. *Right*: scatter plot of projections onto the 1st and 2nd eigenmodes; each point represents one spike. *A*: filter-and-fire cell. These cells have 2 eigenmodes, both negative. Their projections in the space of the 1st and 2nd modes resemble those of the model simulations. *B*: complex filter-and-fire cell. These are similar to the filter-and-fire cells, but with some slight differences: extra modes emerge, either with positive or negative sign. These extra modes are significant, in that they convey significant (if small) information. *C*: Hodgkin-Huxley-like cell. The derivative-like mode has a positive eigenvalue and correspondingly an increased variance relative to the prior. *D*: ring cell. The eigenspectrum contains several positive modes and the projections form a ring. *E*: bimodal cell. Two clear clouds emerge in the 2-dimensional plane. Many features are significant for this cell.

A number of cells had one strongly negative eigenvalue along with a positive second eigenvalue (Fig. 4*C; n = 8* cells). The visual feature corresponding to this second eigenvalue usually resembled a time derivative of the first visual feature, but the cloud of projections had a characteristically curved shape that is quite different from a filter-and-fire cell. This cloud extended equally through positive and negative projec-

tions of the second feature, so sensitivity to the derivative feature in this case clearly does not correspond to positive threshold crossing (DeWeese 1995). Instead, this structure is more typical of a stimulus feature that causes a spike timing shift as seen in simulations of a model neuron with Hodgkin and Huxley currents (Aguera y Arcas et al. 2003). Consequently, we call these *Hodgkin-Huxley-like* (HH-like) cells.

Most of these cells had a medium OFF-type spike-triggered average (Puchalla et al. 2005) with a roughly monophasic time course and a longer latency to peak than for filter-and-fire cells.

The two other response types had more complex structure. Ring cells (Fig. 4D; n = 6 cells) had several positive eigenvalues, and in the space of the first and second features, they showed a projection cloud resembling a ring. All slow OFF-type cells in our data set (n = 3) were ring cells; the other ring cells were fast OFF-type. Another group of cells was similar to filter-and-fire cells but had additional features or nontrivial structure in their multi-dimensional projections (complex FF; n = 14 cells). In the example shown (Fig. 4B), the third visual feature had oscillations well prior to the main lobe of the STA. In many other cases, complex-FF cells were weakly bimodal, where the projection cloud closely resembled the cell shown in Fig. 4B but also had a small fraction of the spikes elicited by an ON-type stimulus feature. Such weakly bimodal cells typically had a mixture of positive and negative eigenvalues.

Information conveyed about visual features

To determine how completely our reverse models can describe ganglion cell function, we compared the information captured by a given model to the total information available in the cell's spike train (Adelman et al. 2003; Aguera y Arcas et al. 2003). The available information was estimated directly by alternating between many nonrepeated segments of random flicker to sample the variety of ganglion cell responses and many repeats of a single, 30-s stimulus segment to sample the noise in that response (see METHODS, Eq. 1). The information captured by a given model of ganglion cell feature selectivity was calculated by comparing the distribution of stimuli given a spike to the prior stimulus distribution (see METHODS, Eq. 2). Because the models we considered ignored most of the possible features in the stimulus, the information they capture must be less than the information available in the spike train. However, if the ganglion cell is only sensitive to a small number of visual features, the model may be nearly complete.

First, we consider how much information we obtain from models that include only a single visual feature. Figure 5 plots the fraction of the available information captured by a single visual feature versus the corresponding eigenvalue from the covariance analysis. All of the significant visual features for the 59 cells in our study are displayed. It is worth noting a general feature of this plot: the distribution of eigenvalues has to be asymmetric because eigenvalues are bounded from below by −1, which can occur if the variance along the spike-triggered distribution is reduced to 0. In contrast, there is no a priori limit to the increase in variance shown by the spike-triggered distribution. We see that features with negative eigenvalues conveyed a large amount of visual information, in some cases up to almost 90% of the information encoded by the ganglion cell. Features with positive eigenvalues also made significant contributions, ranging up to >60% of a cell's available information. In general, the information conveyed by a visual feature increased with the absolute value of the eigenvalue, and visual features with negative eigenvalues tended to convey more information than those with positive eigenvalues of the same magnitude. Most of the points lying in the center of the plot between the positive and negative branches came from bimodal cells.
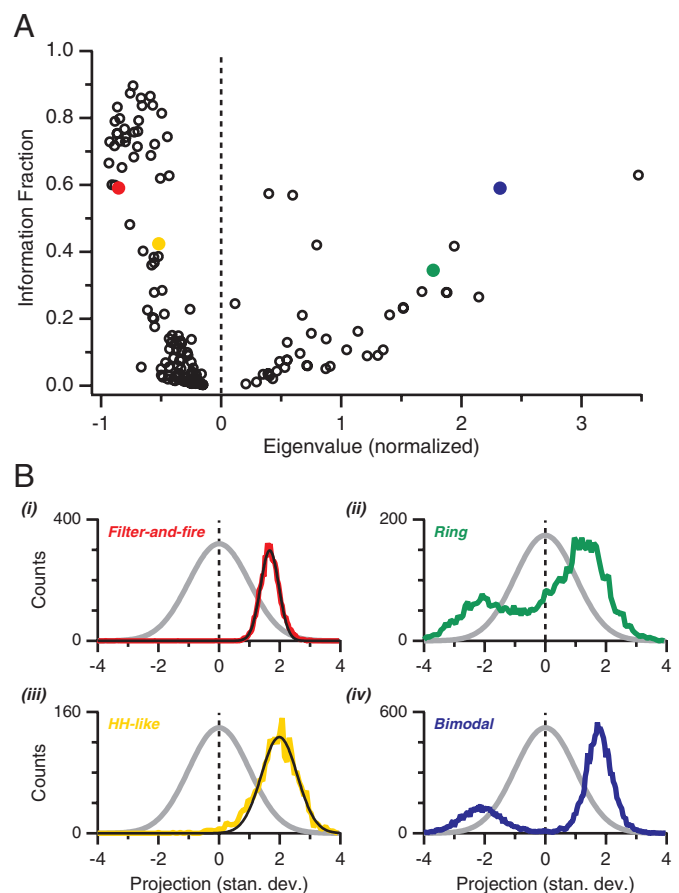

FIG. 5. One-dimensional models of ganglion cell function. A: information captured by individual eigenmodes plotted as a function of their corresponding eigenvalue. Four example eigenmodes are denoted by color. B: distribution of projections onto eigenmodes corresponding to particular eigenvalues, highlighted in the same color. i: typical distribution for a negative eigenvalue from a filter-and-fire cell (red). The distribution is well modeled as Gaussian (thin black line). The prior stimulus distribution is shown in gray. ii: positive eigenvalue with significant information from a ring cell (green). Although the distribution is broadened it is concentrated in values that have low probability in the prior distribution. iii: leading eigenvector from a Hodgkin-Huxley (HH)-like cell (yellow). The distribution of projections is less Gaussian than for the filter-and-fire (FF) cell (thin black line). iv: positive eigenvalue with large information from a bimodal cell (blue). Here the single dimension has separated the 2 modes of the bimodal distribution very well.

We have highlighted visual features from four example cells (Fig. 5B). The filter-and-fire cell (red) had a primary visual feature with an eigenvalue of −0.86 that captured 60% of the information in its spike train. The projection of stimuli along this direction at the time of a spike formed a narrow Gaussian distribution displaced far from the mean of the prior stimulus distribution (Fig. 5B, i). The Hodgkin-Huxley-like cell (yellow) also had a negative eigenvalue, but it captured less information; its distribution of projections was not as narrow or as nearly Gaussian as the FF cell (Fig. 5B, iii). Both the ring (green) and the bimodal (blue) cells had a visual feature with a positive eigenvalue, indicating that the variance along this direction of stimulus space was greater than for the prior. Their projections had bimodal distributions, indicating that both ON- and OFF-type visual features could elicit a spike from these cells (Fig. 5B, ii and iv). The bimodal cell can be seen to have cleaner
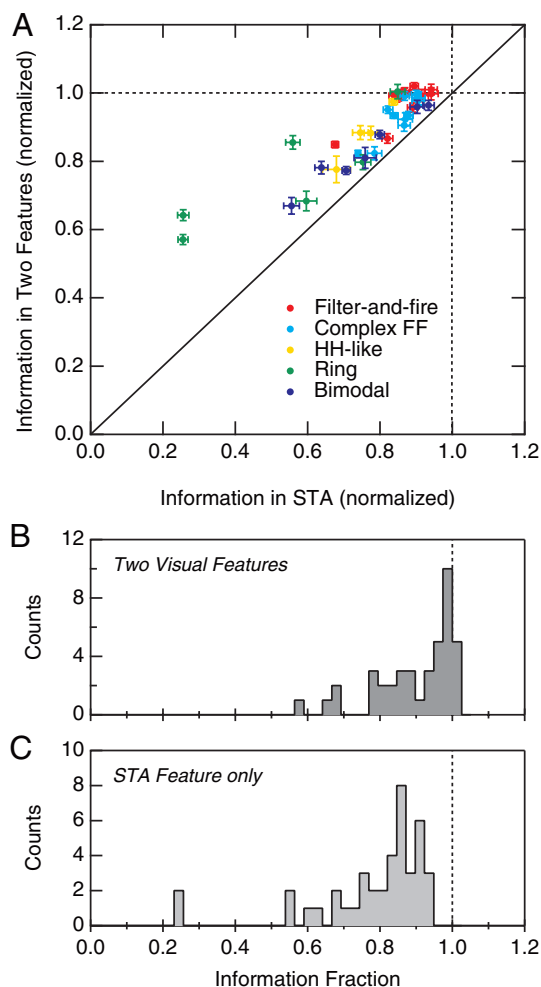
FIG. 6. Information tests. *A*: information captured by the best reverse model using 2 visual features plotted vs. the information in the model using only the STA as a visual feature. The black dotted line shows equality. *B*: histogram of the information captured by the best 2-feature models for *n* = 45 cells. For some cells, the model captures 100% of the information available in the ganglion cell's spike train. *C*: corresponding histogram of the information in the model using the STA as the visual feature.

the two-feature model was significantly better than the STA alone (Fig. 6*A*). For some cells, the two-feature model captured more than twice as much information as the STA alone. For instance, for the ring cell shown in Figs. 4 and 5, the ratio of information captured by the two-feature model to that captured by the STA alone is a factor of 2.5.

How do the two most important features contribute to produce the neuron's output? In the most complete model, the firing probability depends jointly on the projection of the stimulus onto two visual features, P(spike at $t|s_1, s_2$) (see METHODS). However, sampling such a two-dimensional distribution is necessarily more difficult than sampling in one dimension. Can one instead assume that the two features contribute independently? We answered this question by calculating the synergy in the two-feature description constructed as described in the preceding text from the STA and the leading eigenmodes orthogonalized with respect to the STA. The synergy is the difference between the information in the full 2D description and the sum of information in the two 1D descriptions P(spike at $t|s_1$) and P(spike at $t|s_2$) (see METHODS, *Eq. 4*). If the neuron's firing rate depended independently on the two visual features, the information would sum and the synergy would be zero. In Fig. 7, we plotted the synergy versus the information contained in the second visual feature alone. The comparison indicates that most cells had significant synergy, ≤0.5 bits or more, even when the second visual feature contributed very little information by itself. This indicates that the ganglion cell spikes provide significantly more information about the visual scene if the receptive field is measured over the joint space of two relevant features. Synergy typically was found for cells that had a curved or tilted shape to the geometry of their spike-triggered stimulus space.

Predicting the spiking rate

We can invert our reverse model using Bayes' rule to form a forward model that can predict the output of the ganglion cell for any arbitrary visual stimulus contained in the ensemble. Because the reverse model describes the stimulus features that lead to the occurrence of a single spike, the corresponding forward model predicts the probability that a spike will occur

separation between the two peaks in its projections; it also had a more positive eigenvalue, which captured more visual information.

Next we evaluated the completeness of reverse models that include two visual features. We formed such two-dimensional models either from all pairs of relevant visual features defined by covariance analysis or from the STA combined with the component of a relevant visual feature orthogonal to the STA. We found that the best model in almost all cases was the STA combined with the component of the visual feature that individually captured the most information, orthogonalized with respect to the STA. The two-feature models so constructed were highly successful descriptions of ganglion cell light responses: over the population of cells, they captured 89 ± 11% of the information; for some cells, the model captured 100% of the available information (Fig. 6*B*). We compared these two-feature models to one-feature models formed using the STA as this was often the most informative single stimulus feature. The STA alone was also very successful, capturing 78 ± 14% of the information (Fig. 6*C*). But in almost all cases,
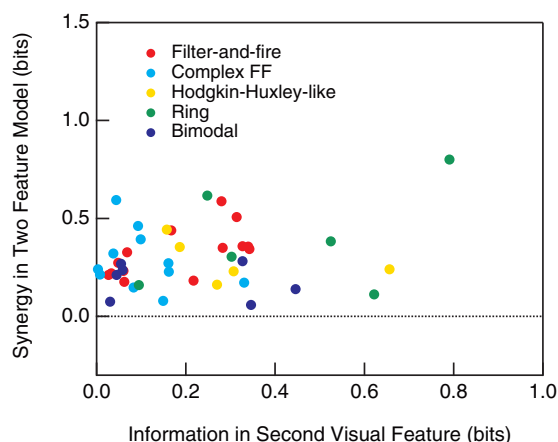

FIG. 7. Synergy between visual features. The synergy resulting from forming a joint model of the ganglion cell's response function using the STA feature and the orthogonal component of the most significant visual feature identified by covariance analysis. Dots are individual ganglion cells; color denotes response type.
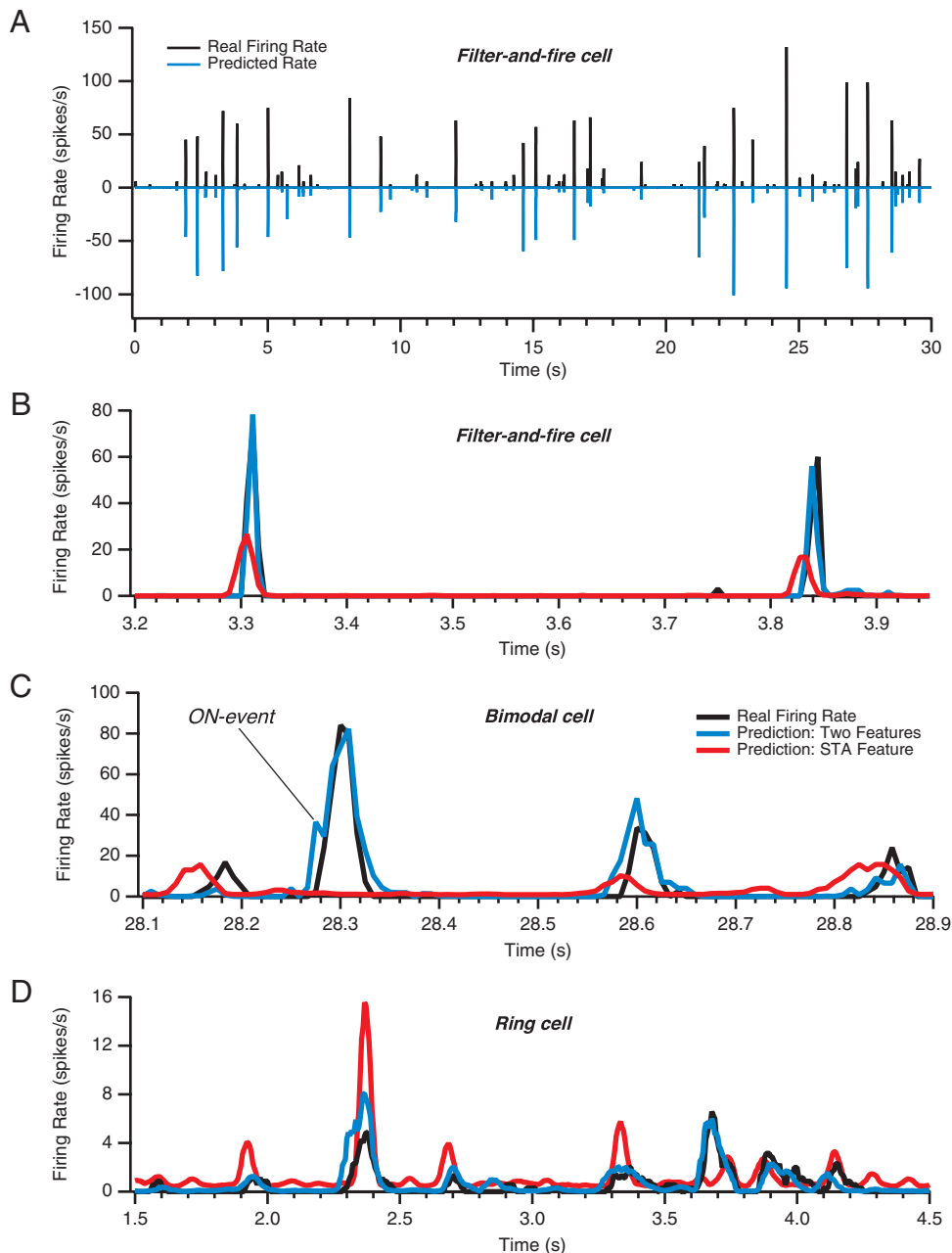
FIG. 8. Predictions of the spiking rate. *A*: real firing rate during a 30-s stimulus segment (black) plotted along with the predicted rate (blue; inverted scale) from a model using 2 visual features. This was the same filter-and-fire cell shown in Fig. 3. *B–D*: real firing rate of 3 different ganglion cells plotted on an expanded time scale (black) along with predicted rates for models using 2 visual features (blue) and for the model using only the STA as a visual feature (red). *B*: same filter-and-fire cell as in *A*. The 2D model captured 99% of the cell's information; the STA model captured 86%. *C*: bimodal cell (same as in Fig. 3). The 2D model captured 74% of the cell's visual information; the STA model captured 37%. The event at $t = 28.3$ s was driven by an increase in light intensity. *D*: ring cell (same as in Fig. 3). Note the low firing rate as well as the much larger time scale for this ganglion cell. The 2D model captured 63% of the cell's information; the STA model captured only 25%, in part because it erroneously predicted a low level of spontaneous firing.

given any stimulus (see METHODS, *Eq. 2*). Our reverse model was formed using the nonrepeated portion of the visual stimulus, but we tested the forward model on a different 30-s stimulus segment that was repeated many times. Therefore we are assessing the ability of the reverse model to generalize to arbitrary stimuli.

Using the projection onto two visual features as the reverse model captured ~100% of the encoded visual information for some cells. In this case, the corresponding forward model gave a very accurate prediction of the firing rate over the entire 30-s stimulus segment (Fig. 8*A*; filter-and-fire cell). Essentially every firing event produced by the ganglion cell was also predicted by the model: predicted events matched the occurrence of real events to within the event's temporal width for >99% of all of the cell's spikes. All of the periods of reliable ganglion cell silence were also correctly predicted: falsely

predicted events accounted for <1% of the predicted spikes. On an expanded scale, we can see that the reverse model with two visual features gave an excellent prediction of the detailed firing rate (Fig. 8*B*). Remaining errors were small deviations that were of the same magnitude as the trial-to-trial variability of the cell's response, which is precisely why the model captures all of the cell's visual information.

We can compare the performance of a model formed using only the STA as a visual feature. For this neuron, the STA model was very successful, capturing 86% of the encoded information. The corresponding prediction also was quite good, with predicted firing events for essentially all real events and silence elsewhere. However, a detailed comparison revealed systematic errors: the peak firing rate was generally too low, firing events were too wide, and in some cases, there was a systematic temporal offset (Fig. 8*B*).

For bimodal cells, the two-feature model gave accurate predictions, including both ON- and OFF-type stimulus events (Fig. 8C). However, the STA model could not predict ON-type events because the STA was always OFF type for these cells. For strongly bimodal cells, the prediction of the STA model often had a low peak firing rate and included some predicted firing events where there was no real event (Fig. 8C). This occurred in part because the STA could be quite unlike either of the visual features in the full model (Fig. 4E).

Ring cells exhibited slower modulations in their firing rate with considerably lower peak rate than other cell types (Fig. 8D). Although the model with two visual features typically captured much less of these cell's visual information, the prediction matched most events in the real firing rate with some errors in the peak rate. In contrast, the STA alone predicted maintained firing from the cell as well as many events where the real cell was silent. For this cell, the STA only captured 25% of the total visual information. These errors result from the ring-like structure of its stimulus projections at the time of a spike: stimuli with zero projection onto both features do not lead to spikes (Fig. 4D), but stimuli with zero projection onto a single feature do lead to spikes (Fig. 5B, ii). Thus the STA alone can give a quite adequate description of the light responses of some ganglion cells, particularly filter-and-fire cells, but for other cells with more exotic response properties, the prediction is qualitatively wrong.

DISCUSSION

We have demonstrated that the spiking behavior of many retinal ganglion cells is influenced by the presence of multiple temporal features in the stimulus. We found that models that included either one or two visual features captured a remarkably high fraction of the total information encoded by ganglion cell spike trains. Furthermore, the reverse model could be inverted to give an accurate prediction of a cell's firing for a novel stimulus pattern. Together, these results indicate that covariance analysis was highly successful at describing how retinal ganglion cells encode temporal patterns of light intensity. Across the population of ganglion cells, a great diversity of different feature selectivities was observed, constituting a surprisingly rich vocabulary of temporal features sent from eye to brain. This extensive vocabulary may allow neurons in subsequent brain circuits to form representations of many possible visual features from combinations of single ganglion cell spikes.

A substantial fraction of the cells displayed the computational structure of a filter-and-fire neuron. An important aspect of such a model is the explicit separation between the filtering properties of the retinal circuit and spike generation in the ganglion cell (often implemented in models by a simple threshold-crossing). For filter-and-fire cells, the spike-triggered average alone generally gave a very good description of the cell's spiking. However, the two visual features revealed by covariance analysis, along with the nonlinear decision function formed by the projection of the stimulus onto these features, show us this structure directly with no fitting required. We can interpret the features immediately as a filter and its derivative, and the plot of projections provides us with an estimate of the threshold, or the probability of spiking as a function of the projection. In contrast, Keat et al. had to perform nonlinear function optimization with 26 parameters, a daunting task (Keat et al. 2001). However, with a slight modification of the form of the cascade model, one can find globally optimal parameters, perhaps with an equally good fit to data and also recover the form of the spike afterhyperpolarization (Paninski et al. 2004).

Selectivity for more complex features

Many cells, however, show additional features, demonstrating that the filter-and-fire model does not capture all the relevant dynamics for these cells. Additional features may have two sources: the subset of the stimulus space that lead to a spike may be more complex than a single filter can describe, or the effects of the interaction between spikes—the effect of silence as well as spiking—may add complexity to the meaning of individual spikes, as has been seen for single neuron models (Aguera y Arcas and Fairhall 2003; Aguera y Arcas et al. 2001, 2003; Paninski et al. 2004). The complex FF cells show several deviations from the filter-and-fire picture, such as sensitivity to an additional visual feature having a negative eigenvalue. In the example shown in Fig. 4B, this extra visual feature involves a slow change in the light level at times prior to when the STA is significantly different from the mean light level. This additional feature can be thought of as modulating the firing rate depending on the recent visual context and may describe a form of local contrast adaptation. In other cases, these cells were weakly bimodal or had only a single significant feature with a positive eigenvalue.

Another subset of ganglion cells illustrate that the derivative-like visual feature need not be associated with a reduction in variance, as is the case for filter-and-fire models with threshold-crossing. For these cells, the additional visual feature had a positive eigenvalue, corresponding to an increase in the variance relative to the prior (Fig. 4C). This leads to an alternative interpretation of how a derivative-like visual feature arises: namely, from spike timing jitter. Imagine that a ganglion cell is performing template matching to one stimulus feature but that there is either noise or an additional stimulus feature that causes its spike to be shifted in time. This temporal jitter induces a time translation of the spike relative to the template, and the operation of time translation is equivalent to a time derivative, for small translations. This was analyzed in the case of a noisy temporal jitter in (Aldworth et al. 2005). In such a case, the distribution of projections onto the second, derivative-like feature is determined by the distribution of jitters, which may have a large variance. This behavior was observed in the Hodgkin-Huxley neuron (Aguera y Arcas et al. 2003), which leads us to speculate that for these cells the additional features that we observe may arise from the ion channels that generate spikes in the ganglion cell rather than from the filtering accomplished by the cell's presynaptic circuitry.

Two additional classes of cells emerged, ring and bimodal cells. For both of these types, the covariance analysis revealed many visual features. For the ring cells, the projections of the stimulus onto their two most informative features takes the form of a ring around the origin. The resulting bimodality along any one visual feature leads to positive eigenvalues or increased variance compared with the original stimulus. This structure is reminiscent of results for neurons in rat barrel
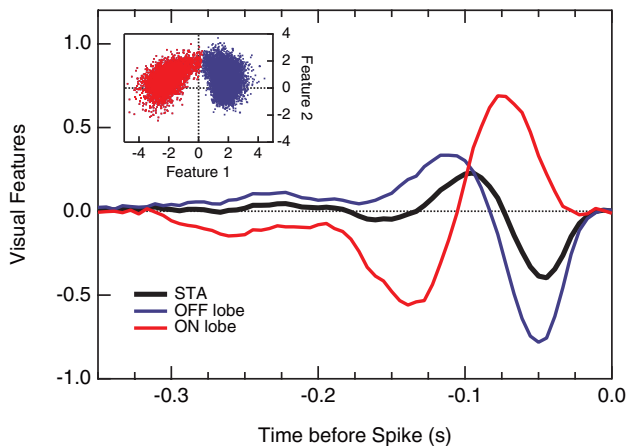
FIG. 9. ON and OFF responses for bimodal cells. STA (black) along with the average stimulus preceding OFF-type spikes (blue) and ON-type spikes (red), as defined by choosing positive or negative projections, respectively, onto the 1st eigenmode (*inset*).

cortex (Petersen and Diamond 2003), which are frequency sensitive but phase invariant and encode the power in a particular frequency band, as well as complex cells in V1 (Rust et al. 2004, 2005; Touryan et al. 2002), which exhibit invariance to the precise location of an oriented bar. It appears that a similar kind of phase invariance or sensitivity to stimulus power is being computed by ring cells in the retina.

Most strikingly, we observed a number of cells that showed a strong bimodality, with projections falling into two distinct clouds. Such cells are ON-OFF-type, with transient responses to both increases and decreases in illumination. The OFF-type responses were always seen to dominate, so that the STA was also OFF type. However, the multidimensional representation allows a clean separation of the two responses. If we then calculate the spike-triggered stimulus average separately for the two clouds, we see that one cloud of points represents an OFF-type stimulus feature, whereas the other cloud is ON type (Fig. 9). This analysis shows that there is a lack of exact symmetry between the ON and the OFF features. The generation of these two features may be the outcome of filtering through separate ON and OFF bipolar pathways, which have been shown to be asymmetric (Chichilnisky and Kalmar 2002; Rieke 2001). In other cascade models, these ON-OFF cells would require two separate channels of filtering with many more parameters and a corresponding increase in the complexity of finding optimal values for those parameters. In particular, the class of cascade models shown to have parameters that can be globally optimized cannot model bimodal cells (Paninski et al. 2004). On the contrary, this structure naturally emerges from covariance analysis with no fit parameters.

Diversity in the ganglion cell population

Although we have divided the ganglion cell population into several functional types for the purpose of discussion, we emphasize again that we saw no evidence that the population actually can be rigorously clustered into these five types. Based on the spike-triggered average alone, salamander ganglion cells can be divided into six broad functional types (Puchalla et al. 2005; Segev et al. 2006). We found a correspondence between these broad types and the full feature selectivity: for

instance, ring-like cells tended to be slow OFF-type and Hodgkin-Huxley-like cells tended to be medium OFF-type. However, this correspondence was not exact, indicating that ganglion cell selectivity to multiple stimulus features does not resolve functional subtypes. In a similar fashion, combining responses to other stimuli, such as spatially uniform steps of light, with the spike-triggered average also served to blur broad functional types rather than delineate subtypes (Segev et al. 2006).

One can also assign ganglion cells to functional classes using an information theoretic method that forms clusters by maximizing the information that spike trains convey about cell identity (Schneidman et al. 2002). Although this method is highly sensitive to the detailed spiking pattern produced by each ganglion cell, it is important to note that this technique assigns cells to functional classes without providing any direct insight into each cell's response function. After functional classes were defined, we examined the spike-triggered averages of all the cells. We found that this method distinguished all of the broad functional types that could be resolved by clustering the spatiotemporal receptive field (Puchalla et al. 2005; Segev et al. 2006) as well as some additional subtypes. However, these subtypes did not correspond well from one preparation to the next. Overall, the impression formed by previous studies of functional classification in the salamander retina is that ganglion cells of the same response polarity (e.g., ON or OFF) form a functional continuum. The present study only reinforces this conclusion.

Different models of complex feature selectivity

When ganglion cell spikes represent two or more visual features, one should bear in mind that those features define a stimulus *subspace* that is relevant for ganglion cell spiking. Any change of basis that remains in that same subspace provides an equivalent description of the neuron's response function, but the new basis will be spanned by different visual features. Illustrating this point, Schwartz et al. performed a slightly different kind of feature analysis on salamander ganglion cells (Schwartz et al. 2002). Where here we have simply compared the variance of all spike-triggered stimuli to the prior to identify features that influence the probability of firing, Schwartz et al. first removed the spike-triggered average from the stimulus by orthogonalizing the stimulus with respect to the STA. Then covariance analysis was performed between spikes and the altered stimulus. In this case, the STA feature was, by definition, excitatory. All other visual features appeared with negative eigenvalues in this analysis, and these were interpreted as suppressive stimulus axes.

As we have discussed with reference to the Keat model analysis, the STA is in general a linear combination of relevant visual features and time derivatives. Projecting the STA out of the spike-triggered stimulus distribution removes a component of the true excitatory feature as well as a component of the time derivative; it also preserves a component of both these features. The remaining components along the directions $f(t)$ and $f'(t)$ will in general have reduced variance, and so will appear with negative eigenvalues and thus be classified as "inhibitory." It is thus possible that the STA-centered analysis may confound some of the geometrical structure observed in our approach. This approach also may not properly capture the synergy

observed between the STA and other visual features (Fig. 7). Finally, if the STA-centered analysis was performed on ring cells, some of the additional features should still have a positive eigenvalue, which was not previously observed (Schwartz et al. 2002). Thus it is possible that the previous study analyzed a less complete sampling of ganglion cell response types than has been described here.

As we have discussed, noise-induced spike timing jitter can introduce a derivative component into covariance analysis (Aldworth et al. 2005). With respect to our simulations of the Keat model, the negative eigenvalue corresponding to the derivative mode can be made to become positive with the addition of sufficiently large jitter. We found that this occurs only when jitter is larger than the time scale of the filter itself. At that extreme, additional modes also start to appear. To examine the possibility that any of the eigen-mode structure we observed arose as a consequence of noise-induced jitter, we computed the jitter in the spike trains using a shuffled auto-correlation function (Berry et al. 1997) for all cells and compared it to the eigenvalue corresponding to the mode most similar to the derivative of the primary filter, when one existed. Jitter values were almost all between 2 and 10 ms, far less than the time scale of the filters. For the ring and bimodal cells, no single mode can generally be identified as derivative-like. Although one cell type, the ring cells, had unusually large jitter (~40 ms), we found no other relationship between the size of the jitter and the cell's eigenvalues or the eigenmode structure.

Future directions

An outstanding challenge for white-noise methods is to identify how the feature selectivity that they discover arises from neural circuitry or the biophysics of a spiking neuron. For example, a clear hypothesis for the generation of the bimodal cell response properties is that they reflect the ON and OFF nature of their bipolar cell inputs. How are the response properties of ring or complex-FF cells built up from their inputs? To what extent do the diverse response types observed correspond to a diversity in the responses of ganglion cells themselves, or rather of the connections in earlier layers? The response types observed in the salamander retina are reminiscent of more complex processing that takes place in mammalian cortex (Petersen and Diamond 2003; Rust et al. 2005; Touryan et al. 2002). It is tempting to speculate that the salamander retina might perform some of the complex feature representation that is accomplished at higher levels in mammals. If so, the relative ease of exploring the circuitry of feature selection in salamander retina may lead to new insights into cortical processing.

APPENDIX

Here we derive the expression for the information in a single spike. We can define the mutual information between the time of spikes and the stimulus by discretizing the spike train in a small time bin Δt and taking the limit of $\Delta t \to 0$. In this limit, each bin has only zero or one spike. We define the following probabilities

$$P(\text{spike}) \equiv P(1) = \bar{r}\Delta t$$

$$P(\text{silence}) \equiv P(0) = 1 - \bar{r}\Delta t$$

$$P(1|t) = r(t)\Delta t$$

$$P(0|t) = 1 - r(t)\Delta t$$

then the mutual information between the neural response and stimulus is

$$I(R; S) = H(R) - \langle H(R|t)\rangle_t$$

$$H(R) = -P(1) \log_2(P(1)) - P(0) \log_2(P(0))$$

$$\langle H(R|t)\rangle_t = -\langle P(1|t) \log_2(P(1|t))\rangle_t - \langle P(0|t) \log_2(P(0|t))\rangle_t$$

where we use the notation $\langle \ldots \rangle_t$ for an average over all time bins during the repeated stimulus segment. In the limit of small time bins, we can replace this average with an integral over time $\frac{1}{T}\int_0^T dt. \ldots$ Rearranging terms, we can write

$$I(R; S) = \tilde{I}_{\text{one spike}} + \tilde{I}_{\text{one silence}}$$

$$\tilde{I}_{\text{one spike}} = \langle r(t)\Delta t \log_2(r(t)\Delta t)\rangle_t - \bar{r}\Delta t \log_2(\bar{r}\Delta t)$$

$$\tilde{I}_{\text{one silence}} = \langle (1 - r(t)\Delta t) \log_2(1 - r(t)\Delta t)\rangle_t - (1 - \bar{r}\Delta t) \log_2(1 - \bar{r}\Delta t)$$

where $\tilde{I}_{\text{one\_spike}}$ is the information conveyed by the occurrence of a spike and $\tilde{I}_{\text{one silence}}$ is the information conveyed by the absence of a spike. Both quantities are measured in units of bits per time bin. To convert to units of bits per spike, we divide by the mean firing rate and by the time bin, remembering that the mean firing rate during the repeated stimulus segment is $\langle r(t)\rangle_t = r_{\text{mean}}$

$$I_{\text{one spike}} = \tilde{I}_{\text{one spike}}/r_{\text{mean}}\Delta t$$

Next we can substitute

$$I_{\text{one spike}} = \left\langle \frac{r(t)}{r_{\text{mean}}} \log_2\left(\frac{r(t)}{\bar{r}}\right) + \frac{r(t)}{r_{\text{mean}}} \log_2(\bar{r}\Delta t)\right\rangle_t - \frac{\bar{r}}{r_{\text{mean}}} \log_2(\bar{r}\Delta t)$$

$$= \left\langle \frac{r(t)}{r_{\text{mean}}} \log_2\left(\frac{r(t)}{\bar{r}}\right)\right\rangle_t + \left(1 - \frac{\bar{r}}{r_{\text{mean}}}\right)\log_2(\bar{r}\Delta t)$$

The first term is equal to *Eq. 1* of the text, in the limit of small time bins. The second term is zero if $r_{\text{mean}} = \bar{r}$. However, if the average firing rate during the repeated segment does not match the average firing rate during the entire stimulus ensemble, then the second term is not zero. Furthermore, this term is not well-behaved in the limit of small time bins.

However, if r_{mean} is close to $\bar{r}$, we will be able to find subsets of the repeated segment for which the average firing rate does match $\bar{r}$. We can use such segments to estimate the value of the first term over ranges of time for which $r_{\text{mean}} \to \bar{r}$. Using these stimulus segments (as described in METHODS), the information reduces to

$$I_{\text{one spike}} = \frac{1}{r_{\text{mean}}T}\int_0^T dt\, r(t) \log_2 \frac{r(t)}{\bar{r}}$$

Note that this expression neglects the information from silence, $I_{\text{one silence}}$. We do this because we are not evaluating the information from silence in the reverse model. Therefore a proper comparison of information captured by the reverse model versus information encoded in the spike train would not include a contribution due to information from silence. The information conveyed by silence makes a very small contribution, typically ~0.3% of the information conveyed by a spike (but this can be a bit larger for sloppy firing cells at a time bin of 8.33 ms).

Note also that unlike in Brenner et al. (2000b), we do not take the limit $\Delta t \rightarrow 0$. This is because we are comparing against the information of a reverse model evaluated at finite time resolution. So instead, we should match the time bin in *Eq. 1* to the temporal resolution of our model. In addition, we note that an extrapolation to zero time bin changes the encoded information by a very small amount, $\leq 1\%$.

REFERENCES

Adelman TL, Bialek W, and Olberg RM. The information content of receptive fields. *Neuron* 40: 823–833, 2003.

Aguera y Arcas B and Fairhall AL. What causes a neuron to spike? *Neural Comput* 15: 1789–1807, 2003.

Aguera y Arcas B, Fairhall AL, and Bialek W. What can a single neuron compute? In: *Advances in Information Processing Systems,* edited by Leen TK, Dietterich TG, and Tresp V. Cambridge, MA: MIT Press, 2001, p. 75–81.

Aguera y Arcas B, Fairhall AL, and Bialek W. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comput* 15: 1715–1749, 2003.

Aldworth ZN, Miller JP, Gedeon T, Cummins GI, and Dimitrov AG. Dejittered spike-conditioned stimulus waveforms yield improved estimates of neuronal feature selectivity and spike-timing precision of sensory interneurons. *J Neurosci* 25: 5323–5332, 2005.

Balasubramanian V and Berry MJ 2nd. A test of metabolically efficient coding in the retina. *Network* 13: 531–552, 2002.

Barlow HB and Levick WR. The mechanism of directionally selective units in the rabbit's retina. *J Physiol* 178: 477–504, 1965.

Berry MJ and Meister M. Refractoriness and neural precision. *J Neurosci* 18: 2200–2211, 1998.

Berry MJ, Warland DK, and Meister M. The structure and precision of retinal spike trains. *Proc Nat Acad Sci USA* 94: 5411–5416, 1997.

Brenner N, Bialek W, and de Ruyter van Steveninck R. Adaptive rescaling maximizes information transmission. *Neuron* 26: 695–702, 2000a.

Brenner N, Strong SP, Koberle R, Bialek W, and de Ruyter van Steveninck RR. Synergy in a neural code. *Neural Comput* 12: 1531–1552, 2000b.

Bullock TH, Hofmann MH, Nahm FK, New JG, and Prechtl JC. Event-related potentials in the retina and optic tectum of fish. *J Neurophysiol* 64: 903–914, 1990.

Caldwell JH and Daw NW. New properties of rabbit retinal ganglion cells. *J Physiol* 276: 257–276, 1978.

Chichilnisky EJ and Kalmar RS. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci* 22: 2737–2747, 2002.

Cover TM and Thomas JA. *Elements of Information Theory.* New York: Wiley, 1991.

de Ruyter van Steveninck R and Bialek W. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc Royal Soc Lond B Biol Sci* 234: 379–414, 1988.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, and Bialek W. Reproducibility and variability in neural spike trains. *Science* 275: 1805–1808, 1997.

DeVries SH and Baylor DA. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J Neurophysiol* 78: 2048–2060, 1997.

DeWeese M. *Optimization Principles for the Neural Code* (PhD thesis). Princeton, NJ: Princeton University, 1995.

Frechette ES, Sher A, Grivich MI, Petrusca D, Litke AM, and Chichilnisky EJ Fidelity of the ensemble code for visual motion in primate retina. *J Neurophysiol* 94: 119–135, 2004.

Gawne TJ and Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13: 2758–2771, 1993.

Gerstner W and Kistler W. *Spiking Neuron Models: Single Neurons, Populations, Plasticity.* Cambridge, UK: Cambridge Univ. Press, 2002.

Golomb D, Hertz J, Panzeri S, Treves A, and Richmond B. How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput* 9: 649–665, 1997.

Ishikane H, Kawana A, and Tachibana M. Short- and long-range synchronous activities in dimming detectors of the frog retina. *Vis Neurosci* 16: 1001–1014, 1999.

Keat J, Reinagel P, Reid RC, and Meister M. Predicting every spike: a model for the responses of visual neurons. *Neuron* 30: 803–817, 2001.

Lettvin JY, Maturana HR, McCulloch WS, and Pitts WH. What the frog's eye tells the frog's brain. *Proc I R E* 47: 1940–1951, 1959.

Metzner W, Koch C, Wessel R, and Gabbiani F. Feature extraction by burst-like spike patterns in multiple sensory maps. *J Neurosci* 18: 2283–2300, 1998.

Naka K and Sakai HM. The messages in optic nerve fibers and their interpretation. *Brain Res Brain Res Rev* 16: 135–149, 1991.

Olveczky BP, Baccus SA, and Meister M. Segregation of object and background motion in the retina. *Nature* 423: 401–408, 2003.

Paninski L, Pillow JW, and Simoncelli EP. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput* 16: 2533–2561, 2004.

Panzeri S and Schultz SR. A unified approach to the study of temporal, correlational, and rate coding. *Neural Comput* 13: 1311–1349, 2001.

Panzeri S, Schultz SR, Treves A, and Rolls ET. Correlations and the encoding of information in the nervous system. *Proc Biol Sci* 266: 1001–1012, 1999.

Petersen RS and Diamond M. Coding of dynamical whisker stimulation in the rat somatosensory cortex. *Soc Neurosci Abst* 57.1 2003.

Pillow JW and Simoncelli EP. Biases in white noise analysis due to non-Poisson spike generation. In: *Computational Neuroscience* (CNS*02). Chicago: Neurocomputing, 2003.

Powers RK, Dai Y, Bell BM, Percival DB, and Binder MD. Contributions of the input signal and prior activation history to the discharge behaviour of rat motoneurones. *J Physiol* 562: 707–724, 2005.

Puchalla JP, Schneidman E, Harris RA, and Berry MJ II. Redundancy in the population code of the retina. *Neuron* 46: 493–504, 2005.

Reich DS, Mechler F, and Victor JD. Independent and redundant information in nearby cortical neurons. *Science* 294: 2566–2568, 2001.

Reinagel P, Godwin D, Sherman SM, and Koch C. Encoding of visual information by LGN bursts. *J Neurophysiol* 81: 2558–2569, 1999.

Reinagel P and Reid RC. Visual stimulus statistics and the reliability of spike timing in the LGN. *Soc Neurosci Abstr* 24: 139, 1998.

Rieke F. Temporal contrast adaptation in salamander bipolar cells. *J Neurosci* 21: 9445–9454, 2001.

Rieke F, Warland D, de Ruyter van Steveninck RR, and Bialek W. *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press, 1997.

Rust NC, Schwartz O, Movshon JA, and Simoncelli EP. Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1 directionally selective neurons. *Neurocomputing* 58-60: 793–799, 2004.

Rust NC, Schwartz O, Movshon JA, and Simoncelli EP. Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 46: 945–956, 2005.

Schneidman E, Bialek W, and Berry MJ II. *An Information Theoretic Approach to the Functional Classification of Neurons.* Cambridge, MA: MIT Press, 2002.

Schneidman E, Bialek W, and Berry MJ 2nd. Synergy, redundancy, and independence in population codes. *J Neurosci* 23: 11539–11553, 2003.

Schnitzer MJ and Meister M. Multineuronal firing patterns in the signal from eye to brain. *Neuron* 37: 499–511, 2003.

Schwartz O, Chichilnisky EJ, and Simoncelli EP. Characterizing neural gain control using spike-triggered covariance. In: *Advances in Information Processing Systems,* edited by Dietterich TG, Becker S, and Ghahramani Z. Cambridge, MA: MIT Press, 2002, p. 269–276.

Segev R, Goodhouse J, Puchalla JP, and Berry MJ II. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nat Neurosci* 7: 1155–1162, 2004.

Segev R, Puchalla JP, and Berry MJ II. Oversampling of visual space by retinal ganglion cells. *J Neuropysiol* 95: 2277–2292, 2006.

Shapley R and Victor JD. The contrast gain control of the cat retina. *Vision Res* 19: 431–434, 1979.

Strong SP, Koberle R, Steveninck RRdRv, and Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett* 80: 197–200, 1998.

Touryan J, Lau B, and Dan Y. Isolation of relevant visual features from random stimuli for cortical complex cells. *J Neurosci* 22: 10811–10818, 2002.

Victor JD. The dynamics of the cat retinal X cell center. *J Physiol* 386: 219–246, 1987.

Victor JD. The dynamics of the cat retinal Y cell subunit. *J Physiol* 405: 289–320, 1988.

Victor JD and Shapley RM. The nonlinear pathway of Y ganglion cells in the cat retina. *J Gen Physiol* 74: 671–689, 1979.

Wiener N. *Nonlinear Problems in Random Theory.* Cambridge, MA: MIT Press, 1958.