Information Transmission Rates of Cat Retinal Ganglion Cells

Christopher L. Passaglia1,2 and John B. Troy1
1Biomedical Engineering Department and the Neuroscience Institute, Northwestern University, Evanston, Illinois 60208; and 2Biomedical Engineering Department, Boston University, Boston, Massachusetts 02215

Submitted 14 August 2003; accepted in final form 29 October 2003

Passaglia, Christopher L., and John B. Troy. Information transmission rates of cat retinal ganglion cells. J. Neurophysiol. 91: 1217–1229, 2004. First published November 5, 2003; 10.1152/jn.00796.2003. To assess the information encoded in retinal spike trains and how it might be decoded by recipient neurons in the brain, we recorded from individual cat X and Y ganglion cells and visually stimulated them with randomly modulated patterns of various contrast and spatial configuration. For each pattern, we estimated the information rate of the cells using linear or nonlinear algorithms and for some patterns by directly measuring response probability distributions. We show that ganglion cell spike trains contain information from the receptive field center and surround, that the center and surround have similar signaling capacity, that antagonism between the mechanisms reduces information transmission, and that the total information rate is limited. We also show that a linear decoding algorithm can capture all of the information available in retinal spike trains about weak inputs, but it misses a substantial amount about strong inputs. For the strongest stimulus we used, the information rate of the best linear decoder averaged 40–70 bits/s across ganglion cell types, while the directly measured rate was around 20–40 bits/s greater. This implies that under certain stimulus conditions, visual information is encoded in the temporal structure of retinal spike trains and that a nonlinear decoding algorithm is needed to extract the temporally coded information. Using simulated spike trains, we demonstrate that much of the temporal structure may be explained by the threshold for spike generation and is not necessarily indicative of a complex coding scheme.

INTRODUCTION

Retinal ganglion cells are the sole messengers of visual information to the brain. They encode their messages with trains of optic nerve impulses, which target neurons in the brain must decipher in a reliable and timely manner. The task of deciphering the retinal code is complicated by spontaneous fluctuations in spike discharge that persist under steady uniform illumination and even in darkness (Kuffler et al. 1957; Troy and Robson 1992). The discharge noise renders the conversion of visual images into retinal spike trains imprecise. The same stimulus will evoke different responses, and the same response can result from different stimuli.

Our present understanding of how the brain might extract a message from variable neural responses is based largely on studies of the effect a stimulus has on the rate of spike discharge. In these studies, the discharge rate is usually specified by counting the spikes evoked in successive intervals of time (i.e., time average rate) or by measuring the intervals of time between successive spikes (i.e., instantaneous rate). This long-standing approach to neural coding has yielded many insights into the neural representation of visual information. Such insights include the discovery of a diversity of receptive field profiles among visual neurons (Enroth-Cugell and Robson 1966; Hubel and Wiesel 1962; Kuffler 1953) and their functional organization into parallel processing channels (Dacey 2000; Lennie 1980; Schiller 1992). Despite these valuable insights, the neural code for vision remains a subject of continued debate. Fueling the debate is the finding that some stimuli can modulate the fine structure of retinal, geniculate, and cortical spike trains (Bair and Koch 1996; Berry et al. 1997; Buračus et al. 1998; Liu et al. 2001; Reich et al. 1997; Reinagel and Reid 2000), raising the possibility that the precise patterning of spikes conveys information that is ignored by a rate-based coding scheme but perhaps not by the brain.

To clarify the importance of spike rates and spike patterns to neural coding, investigators have turned to information theoretic methods. These methods provide a means of quantifying the number of messages that a neuron can transmit about a set of sensory inputs. The methods have been applied to neural spike trains in two ways. Some studies have taken the approach of directly estimating information rate from the probability distribution of neural responses (de Ruyter van Steveninck et al. 1997; Strong et al. 1998). The advantage of this “direct approach” is that the information measures are free of experimental biases regarding the nature of the neural code. The disadvantage is that they provide little information about what messages are transmitted or how they are encoded. Moreover, many responses are needed to fully characterize the probability distribution, which greatly limits the variety of stimulus conditions that can be investigated in an experiment. In many cases, only a spatially uniform field modulated instantaneously over a wide contrast range has been examined (Berry and Meister 1998; Liu et al. 2001; Reinagel and Reid 2000), and the relevance of such a stimulus to natural neuronal function is questionable. Other studies have taken the approach of estimating information rate from the ratio of signal and noise power spectral density, on the assumption that the encoded signal sums with noise to produce a response probability distribution that is Gaussian (Borst and Theunissen 1999; Rieke et al. 1997). To specify the signal-to-noise ratio, various algorithms are used to optimally reconstruct the stimulus from the response or the response from the stimulus (Bialek et al. 1991; Haag and Borst 1998; van Hateren and Snippe 2001; Warland et al. 1997). This “reconstruction approach” is experimentally efficient and can provide insight about the coding mechanism. However, because of the simplifying assumption, it can only bound the information content of neural spike trains to a certain range, and this range might not be terribly confining.

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Here we use the reconstruction method to place bounds on the information transmission rate of cat retinal ganglion cells for stimulus patterns of various contrast and spatial configuration and use the direct method to evaluate how informative the bounds are. We find that a linear (rate-based) decoder captures all the information in retinal spike trains about weak visual inputs and a sizeable amount about strong inputs, but to extract the maximum information about strong inputs, a nonlinear decoder is required. A simple model of the retinal coding mechanism is considered that could account for the information embedded in the patterning of ganglion cell spikes.

**METHODS**

**Physiological preparation**

Anesthesia was induced in adult male cats (3.5–5 kg) with thioental sodium (20 mg/kg, iv, supplemented as needed) or, on a few occasions, with ketamine HCl (25 mg/kg, im) mixed with acepromazine (1 mg/kg, im). Following induction, a tracheotomy was performed, and the left femoral artery and both femoral veins were catheterized. The arterial catheter was connected to a blood pressure transducer. The venous catheters were connected to pumps that continuously infused ethyl carbamate (15–50 mg/kg/h after an initial loading dose of 200 mg/kg) to maintain anesthesia and pancuronium bromide (0.2 mg/kg/h) to achieve paralysis. Paralyzed animals were mechanically ventilated, and their blood pressure and heart rate were monitored to titrate the rate of anesthetic infusion. Body temperature and end-tidal CO$_2$ were also tracked and kept at normal levels. In addition to anesthetic and paralytic agents, animals were administered dexamethasone acetate (4 mg, im), atropine sulfate (0.3 mg, im), and cefazolin sodium (100 mg every 12 h, im). Ophthalmic solutions of 1% atropine and 2.5% phenylephrine hydrochloride were periodically added to the eye to dilate the pupils and retract the nictitating membranes. Contact lenses with artificial pupils (4 mm diam) were fitted bilaterally, and spectacle lenses were added to the optical path as needed to focus the stimulus onto the retina. All experimental procedures were approved by the Northwestern University Animal Care and Use Committee and were in accordance with the National Institutes of Health guidelines.

**Recording and visual stimulation**

A craniotomy was performed over the left or right optic tract, and a tungsten-in-glass microelectrode was advanced downward through a protective guide tube into the brain. After isolating the discharges of a single optic tract fiber, the retinal eccentricity of the recorded ganglion cell was determined by mapping its receptive field center onto a tangent screen on which the optic disk and major blood vessels surrounding the area centralis of each eye were drawn. The receptive field was then projected via an adjustable mirror onto a video display (Multiscan 17se, Sony) running at a frame rate of 150 Hz. The mean luminance of the display was 30 cd/m$^2$. For a 4-mm pupil, this amounts to a retinal illuminance of approximately 500 cat trolands, which lies in the low photopic range of the animal (Troy et al. 1999). Custom software controlled the display output and recorded spike times with 0.1-ms precision via stimulus generation (VSG2/2, Cambridge Research Systems) and data acquisition cards (AS1, Cambridge Research Systems). On-center X (on-X) cells, off-center X (off-X) cells, on-center Y (on-Y) cells, and off-center Y (off-Y) cells were identified by performing the modified null test with contrast-reversing gratings (Hochstein and Shapley 1976). After cell identification, the receptive field was precisely centered on the display (width: 30 cm, height: 22.5 cm, distance: 60 cm) by rotating the mirror until the response to a contrast-reversing bipartite field, oriented horizontally and then vertically, contained no component at the frequency of reversal.

Data were collected in epochs of 90 s during which cells were stimulated with spots and annuli of various diameter and time-varying luminance. Time variation was accomplished by setting the luminance of the spot and/or annulus to one of two values for each video frame according to a sequence of random binary numbers. The two values were chosen to modulate stimulus luminance without changing its mean. Hence, a spot or annulus having a contrast of 80%, by our usage of the term, fluctuated randomly between 6 and 54 cd/m$^2$ over time. To permit estimation of information rate, two types of random binary sequence were used for each stimulus condition. One repeated every 512 frames and the other did not. Each 512-frame segment (approximately 3.4 s) of the sequences was considered a trial, and the first trial was discarded to eliminate response transients. The rest of the trials exhibited response stationarity (e.g., Fig. 1C). For both sequences, a...
INFORMATION RATES OF X AND Y CELLS

Data analysis

From the set of responses to repeated and nonrepeated sequences, the information content of retinal spike trains was calculated via the reconstruction method and the direct method. Details of the calculations are provided in RESULTS as the methods are applied. Some of the calculations required a spectral description of the stimulus and response on each trial. These were obtained by binning the random binary sequence and spike train at 300 Hz (twice the frame rate) and Fourier transforming the resultant histograms. Since a trial lasted 512 frames, the histograms and their Fourier transforms contained 1,024 bins. Power spectra were computed from the stimulus and response transforms by squaring and summing the real and imaginary amplitudes of each Fourier component and averaging across trials. It was observed that information estimates for each of the four cell types appeared normally distributed when expressed in units of bps. χ² tests were performed, and a normality hypothesis could not be rejected for any cell type. Parametric tests were therefore used for all statistical comparisons of mean information rates.

RESULTS

Reported are data from 30 ON-X cells, 24 OFF-X cells, 30 ON-Y cells, and 31 OFF-Y cells. The cells ranged from 2 to 45° in retinal eccentricity, with a median eccentricity of 16°. Other types of cat ganglion cell were infrequently recorded and not studied.

Variability and reproducibility of retinal spike trains

Retinal ganglion cells can transmit information about a visual stimulus only when their pattern of spike discharge differs significantly from their maintained discharge, which is highly irregular (Fig. 1A). Several studies have shown that the distribution of interspike intervals at photopic light levels is well described by a renewal process with gamma-distributed intervals (e.g., Kuffler et al. 1957; Troy and Robson 1992). Such a process behaves as though it fires a spike on receiving n > 1 randomly timed inputs from a Poisson process. Because of the “integrate-and-fire” behavior, the variance in spike rate of a gamma process depends on the time interval of measurement. This can be seen from the power spectrum of the maintained discharge (Fig. 1B), which is not flat like that of a Poisson process (dashed line) but sigmoidal in shape. That such a shape is not the result of negative serial correlations in the spike train may be shown by scrambling the order of interspike intervals (Troy and Robson 1992).

Visual stimulation causes variations in ganglion cell spike rate to increase mainly at low temporal frequencies, as evidenced by the additional power below 30–40 Hz in response spectra to randomly modulated patterns of various contrast and spatial configuration (Fig. 1D). The increase in response power reflects a clustering of spikes during excitatory phases of the stimulus and a removal of spikes during inhibitory phases (Fig. 1C). For certain stimulus settings (e.g., a high contrast annulus or 1° spot), the bursts and pauses in ganglion cell activity can be highly reproducible from trial to trial. Such reliable patterning of spikes translated into response power spectral densities that were one to two orders of magnitude greater than the maintained discharge noise at many temporal frequencies. The marked differences between the maintained and driven discharge spectra and the changes in the driven discharge spectrum with stimulus contrast and spatial configuration indicate that cat ganglion cells can transmit large amounts of information and that the amount transmitted depends on the spatiotemporal characteristics of the visual input.

Estimating the information rates of retinal ganglion cells

We sought to quantify the information transmission rates of ganglion cells for a variety of visual inputs. Since the direct method of estimating information rate would not be practical for more than one input condition, we used the reconstruction method (Fig. 2A). This method requires less data because it assumes that the probability distribution of responses to a stimulus is well described by a Gaussian function whose variance equals the sum of signal and noise variances. As a result only signal-to-noise ratios need be measured. For such a communication channel, the information transmission rate I is

\[ I = \delta f \sum \log_2 [1 + SNR(f)] \]  

where \( SNR(f) \) is the ratio of signal-to-noise power spectral density at temporal frequency \( f \) and \( \delta f \) is the spectral bin width (approximately 0.3 Hz in our study). With neural spike trains, it is not known what constitutes signal and noise, let alone whether they are additive and Gaussian. Equation 1 may therefore underestimate or overestimate the information rate of ganglion cells depending on how signal and noise are defined. By exploiting the fact that Gaussian distributions have the greatest entropy for a given variance, the reconstruction method can, however, provide bounds on the information rate.

To place a lower bound, one presents a stimulus that is Gaussian over the frequency band to which a neuron can respond and searches for the linear transformation that maximizes the mutual information between the stimulus and response. Because the transformation is considered linear, the signal extracted from the response would also be Gaussian and have maximal entropy. The residual noise, on the other hand, could be non-Gaussian, but such noise would have even less entropy than Gaussian noise of the equivalent variance. Hence, the neuron must transmit information at or above this minimum rate. Whether a target neuron actually makes use of the information is irrelevant. It could decode the spike train in an inefficient manner. Nevertheless the information would still have been sent.

In placing lower bounds on the information rate of cat X and Y cells, we temporally modulated the visual stimulus on a set of trials according to different random binary sequences. Although the stimulus could alternate between only two values, it was effectively Gaussian over the frequency range of the cells because the rate of alternation was rapid and random. Evoked spike trains were linearly filtered to produce an estimate of the stimulus sequence on each trial having the least mean square

J Neurophysiol • VOL 91 • MARCH 2004 • www.jn.org

total of 2–10 epochs were collected, yielding 50–250 repeated and nonrepeated trials per stimulus condition. The entropy of the stimulus on each trial was 150 bits/s (bps) owing to the video frame rate. Between epochs of data collection, the location of the receptive field was regularly checked to ensure that it remained centered on the visual display. Only cells with stable and well-isolated spikes for the duration of data collection were analyzed.
error (reverse reconstruction technique). In some cases, the stimulus sequences were also filtered to produce a linear estimate of the evoked spike trains (forward reconstruction technique). Following the frequency domain formulation of Borst and Theunissen (1999), the best estimate of the stimulus $S_k(f)$ given by the linear decoding (i.e., reverse) filter $H_S(f)$ is

$$S_k(f) = H_S(f)R_k(f),$$

where $H_S(f) = \frac{\sum S_k(f)R_k(f)}{\sum S_k(f)S_k^*(f)}$ (2)

And, the best estimate of the response $\hat{R}_k(f)$ given by the linear encoding (i.e., forward) filter $H_R(f)$ is

$$\hat{R}_k(f) = H_R(f)S_k(f),$$

where $H_R(f) = \frac{\sum R_k(f)S_k^*(f)}{\sum R_k(f)R_k(f)}$ (3)

Asterisks denote the complex conjugate, and $S_k(f)$ and $R_k(f)$ are the Fourier transforms of the stimulus and response on trial $k$ computed as described in METHODS. The best linear estimate

---

**FIG. 2.** Estimating information transmission rate. The reconstruction method and the direct method were used to quantify the information transmission rates of cat retinal ganglion cells. Both methods share the same experimental paradigm. The cell is presented a stimulus pattern that is temporally modulated by different random waveforms on one set of trials and by identical random waveforms on another. The methods differ, however, in their analysis of neural responses. The reconstruction method estimates information rate by bounding it from below and above. A lower bound is placed using the 1st set of responses and an optimal linear filter $H_S$ to define the signal $\hat{S}$ encoded in the response $R$ about the stimulus $S$ (reverse reconstruction technique). And, an upper bound is placed using the average response $\bar{R}$ of the other response set to define the encoded signal. From the ratio of signal-to-noise power spectral density, lower and upper bound rates are calculated. The direct method estimates information rate by measuring, for each set of responses, the frequency of occurrence of all spike count patterns, or “words” $w$, of bin duration $\delta t$ and length $L$. For purpose of illustration, 3 spike trains that were recorded on successive trials from an ON-X cell are shown in B$_2$, along with the probability distribution of 4-bin words that was measured for the 3rd time bin ($t = 3$). From this and the word probability distributions for all other time bins, total response entropy and noise entropy are calculated. The difference between the 2 entropies is the cell’s information rate for a given $\delta t$ and $L$. 

---

**RECONSTRUCTION METHOD**

- $A_1$: Present Different Random Stimuli
- $A_2$: Present Identical Random Stimuli
- $A_3$: Fourier Transform Response
- $A_4$: Estimate Signal by Linear Filtering
- $A_5$: Calculate Signal-to-Noise Ratio

**DIRECT METHOD**

- $B_1$: Present Different Random Stimuli
- $B_2$: Present Identical Random Stimuli
- $B_3$: Measure Response Probabilities
- $B_4$: Calculate Total Response Entropy
was defined as the signal and the error in the estimate as noise $N_k(f)$. From Eq. 1, the lower bound on information rate $I_{LB}$ is

$$I_{LB} = \delta f \sum \log \left( 1 + \frac{\sum S_k(f) \hat{R}_k(f)}{\sum N_k(f) N^*_k(f)} \right)$$

where $N_k(f) = S_k(f) - \hat{S}_k(f)$.

$$I_{LB} = \delta f \sum \log \left( 1 + \frac{\sum R_k(f) \hat{R}_k^*(f)}{\sum N_k(f) N^*_k(f)} \right)$$

where $N_k(f) = R_k(f) - \hat{R}(f)$.

The lower bound rate is the same whether the forward or backward technique is used because the reconstruction is linear.

The information rate of a neuron can exceed the lower bound if it transforms the stimulus nonlinearly. Since the form of the nonlinearity is generally unknown, the encoded signal cannot be determined directly from the stimulus. However, in the presence of additive Gaussian noise, it can be estimated from the average neural response, and by presenting a stimulus that is Gaussian over the frequency range of the neuron, an upper bound may be placed. Unless the noise is non-Gaussian, the signal-to-noise adjustment was found to have a small effect on $I_{UB}$, reducing it by approximately 5 bits/s or less.

Figure 3A plots the average power spectrum of the random binary sequences used for lower and upper bound measurements. The lower bound stimulus spectrum is smooth because it is the average of many different random sequences, whereas the upper bound stimulus spectrum is for a repeated random sequence. Due to the limited frame rate of the display (150 Hz), both spectra show a sharp decline in power at high temporal frequencies. The empty symbols in Fig. 3B are the upper bound stimulus spectrum is for a repeated random sequence. Due to the limited frame rate of the display (150 Hz), both spectra show a sharp decline in power at high temporal frequencies. The empty symbols in Fig. 3B are.

The filled symbols in Fig. 3B plot the average power spectrum of noise in ganglion cell responses as determined by the linear reconstruction error (lower bound) and by deviations of individual responses from the ensemble average (upper bound). Notice that for each cell type, the response and re-

![Figure 3](Image)

**Figure 3.** Power spectral analysis of the stimulus and ganglion cell spike trains. A: average power spectrum of the ensemble of random binary sequences used for lower bound measurements and the standard random binary sequence used for upper bound measurements. B: average power spectrum of the response (empty symbols) and of noise in the response (filled symbols) of representative ON-X, ON-Y, OFF-X, and OFF-Y cells for lower and upper bound stimuli. Noise spectra were determined from the error in the best linear reconstruction of the stimulus given the response (top) or from deviations of individual responses from the average response (bottom) as outlined in Fig. 2. The stimulus was a 100% contrast spot of 0.5° diam for X cells and 2° diam for Y cells. The receptive field center of the cells was 0.5, 1.8, 0.5, and 1.4°, respectively, in Gaussian diameter. Their eccentricities were 9, 24, 16, and 9°. fps, frames per second.
response noise spectra superimpose above 30–40 Hz, which indicates that response power at high temporal frequencies was due almost entirely to spike discharge noise. That cat ganglion cells poorly encoded such stimulus frequencies is consistent with prior measurements at this photopic light level (Frischman et al. 1987). Notice also that the low-frequency plateau of the upper bound noise spectrum is 5- to 10-fold less than that of the lower bound noise spectrum. This finding was true irrespective of the random binary sequence used for upper bound measurement (data for other sequences not shown). It reflects the contributions of nonlinear mechanisms to retinal coding (Fig. 4), which appear as "noise" in lower bound measurements because a linear decoder cannot reconstruct them.

From the ratio of signal and noise power spectral density the bounds on information rate are computed. Figure 5 shows the value of this ratio as a function of temporal frequency for the cells in Fig. 3. That lower bound estimates of signal-to-noise ratio (filled symbols) exceeded upper bound estimates (empty symbols) at several frequencies should not be cause for concern. It happened because the distribution of power in a single random sequence varies above and below the mean of an ensemble of different sequences (Fig. 3A). Both stimuli, however, have the same variance (i.e., integral of the power spectrum). For the strong stimulus used here, lower bound estimates of signal-to-noise ratio fell around 3 over the frequency range of the cells, although a drop-off at low frequencies was noted with ON cells (48/60) and some OFF cells (8/55). In terms of information (right axis), this ratio translates to approximately 2 bits per frequency component. Information densities of >7 bits might be attainable from ganglion cell spike trains based on upper bound estimates of signal-to-noise ratio.

Summing information over temporal frequency and dividing by trial duration yields the lower and upper bounds on infor-

FIG. 4. Signal and noise contributions to ganglion cell responses. The thick line depicts the response spectrum of an arbitrary cat ganglion cell to a set of random binary sequences. The area shaded in black is the contribution of spike discharge noise to the response spectrum. It is given by the upper bound noise spectrum. The area shaded in gray is the signal contribution to the response spectrum. It generally contains a linear and nonlinear component. The linear component comprises the area bounded by the thick and thin lines, where the thin line is given by the lower bound (linear) noise spectrum. This component can be extracted from the cell’s response using a simple linear decoder. The nonlinear component comprises the gray area below the thin line. It can only be extracted using a nonlinear decoder. If the lower and upper bound noise spectra are the same, there is no nonlinear component and a linear decoder is optimal.

FIG. 5. Signal-to-noise ratio of ganglion cell responses. Ratio of signal-to-noise power spectral density given by the lower bound (filled symbols) and upper bound (empty symbols) measurements in Fig. 3. Right axis of plots expresses signal-to-noise ratio in terms of information density.

FIG. 6. Lower and upper bound information rates. Cumulative sum of lower bound (lower line) and upper bound (upper line) information densities in Fig. 5. Lower and upper bound rate for a given visual input is the sum of information over all temporal frequencies. The saturation of lower and upper bound rates of X cells indicates that nearly all information in their spike trains resided at temporal frequencies below 30–40 Hz at this photopic light level. Y cell spike trains may contain additional information at higher frequencies. bps, bits per second.
Information rates of the receptive field center and surround

Using the reconstruction method, we estimated the information rates of cat ganglion cells for the kinds of visual input to which the cells are most receptive. That is, since X and Y cells are known to have center-surround organization, we first examined what information they can transmit about a 100% contrast (black-white) spot of variable diameter centered on their receptive field (Fig. 7). To combine results from cells with different center sizes, spot diameter was divided by the smallest spot that gave a near maximum lower bound rate, which means that spots smaller than the center had a normalized size <1 and those larger than the center had a normalized size >1. For all four cell types, lower and upper bounds on information rate were the same on average for spots much smaller than the receptive field center. For this to occur, the signal encoded by the cells must have been linearly related to the visual input and to the average spike rate. The bounds on information rate differed greatly, however, for spots as large or larger than the center (≤30 bps for X cells and 100 bps for Y cells, P < 0.01), implying that the spike trains contained information about the stimulus that only a nonlinear decoding algorithm can access. It was also found that increasing spot size beyond a certain diameter actually reduced the information bounds of X cells (P < 0.01), presumably because of antagonism from the receptive field surround. The inhibitory effect was exerted mainly at temporal frequencies below approximately 10 Hz (e.g., compare 1 and 16° spot response spectra in Fig. 1D). It was less pronounced with Y cells, owing in part to their large surrounds. It could also be counterbalanced by a growth in signal-to-noise ratio at high frequencies. As a result, the information rate of Y cells typically held steady for medium- to large-size spots.

We further explored the contributions of the receptive field surround to retinal coding by altering the spatial configuration of the stimulus. Three patterns were specifically tested: a spot, an annulus, and a spot plus an annulus. Each pattern was presented at 100% contrast. The spot was optimized for center stimulation by adjusting its diameter to maximize information rate. The annulus was in turn optimized for surround stimulation by setting its inner diameter equal to the spot and its outer diameter to 16°, which well exceeds the characteristic diameter of X cell surrounds and covers much of Y cell surrounds (Linsenmeier et al. 1982; Troy et al. 1993). The spot and annulus were modulated by binary sequences drawn independently from the same set of random sequences so that both had the same statistics across trials, yet the combined pattern on any given trial differed visibly from a large spot. Since ganglion cells might encode features of both stimulus elements in their spike trains, lower bounds were separately calculated for the spot and annulus and added to get the rate for the combined pattern. For all four types of ganglion cell the bounds on information rate for the annulus were indistinguishable from those for the spot (Fig. 8), with one possible exception (P < 0.01 for the OFF-Y lower bound). Hence, not only did the cells transmit information about visual stimuli confined mainly to their receptive field surround, but the surround also appears to have a similar signaling capacity as the center. Driving both receptive field components asynchronously did not, however, double the information bounds. It did not lower the information bounds either, as synchronously driving the center and surround does (e.g., large spots in Fig. 7). Instead, less information was transmitted about the spot and annulus (light- and dark-gray bars, respectively), while the bounds for the combined pattern remained the same as those for the spot and annulus alone or perhaps changed slightly (P < 0.05 for both ON-X cell bounds). Coding the annulus thus came at the expense of coding the spot and vice versa. It would be interesting to see whether stimulating the center and surround synchronously but in antiphase (as with a modulated bull’s-eye pattern) would yield a similar result. It might be expected that such a stimulus would give higher information rates since the surround should have a facilitatory effect rather than an antagonistic one.

Since cat ganglion cells are sensitive to small fluctuations in luminance as well as large fluctuations, we next examined what information the cells can transmit about a spot of variable contrast matched in size to the receptive field center (Fig. 9).

To vary spot contrast from one set of trials to the next, the
maximum and minimum luminance of the random binary sequence was altered from its black-white setting in preceding figures. For all four cell types, the bounds on information rate increased linearly below approximately 25% contrast and were statistically indistinguishable on average (t-test not significant except for OFF-Y cells at 20% contrast). For higher contrast settings, the lower bound rate saturated at around 50 bps while the upper bound rate continued to grow, reaching a maximum of \( \sim 100 \) bps for X cells and \( \sim 150 \) bps for Y cells. Like the spot size measurements, the contrast measurements indicate that a linear decoding strategy is optimal for weak visual inputs. Exactly how much is uncertain because the reconstruction method makes no assumptions about the signal and noise in neural spike trains and estimates information rate directly from response probability distributions, much data were required. We thereby restricted the analysis to the 25 ganglion cells from which responses to a center-matched spot of 100% contrast were obtained for \( \geq 125 \) repeated and nonrepeated random binary sequences. Following Reinagel and Reid (2000), each spike train in a set of trials was represented as a string of \( T \) numbers corresponding to the spike count in successive time bins of duration \( \delta t \) (Fig. 2B, inset). Depending on the choice of \( \delta t \), the time bins could contain 0, 1, or more spikes. Once created, the strings were parsed bin by bin into patterns of numbers, or words, of length \( L \) and the frequency of occurrence of each word across the set of strings was measured (Fig. 2B\(_2\)). This produced an assortment of word probability distributions, one for every bin, whose average entropy \( E(L, \delta t) \) was computed as

\[
E(L, \delta t) = -\frac{1}{L(T-L)\delta t} \sum_{t=1}^{T-1} \sum_{w=1}^{L} P_w(t) \log P_w(t)
\]

where \( w \) is a specific word, \( W(L, \delta t) \) is the set of all possible words comprised of \( L \) bins of duration \( \delta t \), and \( P_w(t) \) is the probability of a specific word at time bin \( t \). Separate calculations were performed for repeated and nonrepeated trials. The former gave an estimate of the noise entropy \( E_N \) in the response, since a noiseless neuron would always generate the same output to a repeated input, and the latter gave an estimate of the total response entropy \( E_R \). The difference between the two is the information rate \( I \) of a cell for spike patterns of length \( L \) and temporal resolution \( \delta t \)

\[
I = E_R - E_N
\]
be multiples of one-eighth the frame interval (i.e., \(1/1,200\) s) so that the response strings contained timing information down to \(<1\) ms and an integer number of bins. Both the response entropy and noise entropy increase with temporal resolution because they are expressed in terms of rates and thus head toward infinity as \(\delta t\) approaches zero (Eq. 6). They both decrease with word length because retinal ganglion cells, like most neurons, do not fire spikes in a Poisson manner, not even when they are stimulated randomly (Fig. 1, C and D). Consequently, as word length increases, certain spike patterns occur more frequently than other patterns, reducing the entropy of the word probability distribution. Eventually, the exploding set of possible words and the finite data with which to estimate their frequencies of occurrence lead to inaccuracies in the word probability distribution and both entropy rates fall catastrophically. Since infinitely long words would be needed to reach the maximum entropy rate, we estimated it by linear extrapolation (dotted lines) as previously described (Strong et al. 1998). Extrapolated (arrows) and measured noise entropies were then subtracted from the total response entropy to give the average information rate of the cell for each bin size (Fig. 10B). The average information rate was always non-negative because a nonrepeated stimulus elicits a greater variety of spike patterns than a repeated stimulus. For small \(\delta t\), the information rate often increased with word length (solid line), which indicates that spike patterns \((L > 1)\) conveyed information beyond that of individual spikes \((L = 1;\) Reinagel and Reid 2000). That is to say, not all of the spikes fired by the cell were independent. Based on the maximum measured rate \((L = 4, \delta t = 0.8–1.7\) ms) and its proximity to the extrapolated rate (arrow), most of the pattern information was contained in words under 7 ms in duration. This is an even finer time scale than the frame interval of the display, suggesting that the spike patterns resulted from a nonlinearity in the retinal coding mechanism. Knowledge about the nonlinearity would be needed for a decoder to extract this pattern information.

The circles and triangles in Fig. 11A plot the directly measured information rates of the 25 ganglion cells for the 100% contrast spot. The directly measured rate of a cell was defined as the maximum extrapolated rate for all bin sizes. It should closely approximate the actual rate of the cells because the extrapolated rate was similar for the smallest two bin sizes. Among the group, the directly measured rate ranged from 60 to 148 bps and was greater for Y cells compared with X cells \((P < 0.01)\), consistent with the relative contrast sensitivities of the two cell types (Troy 1987), and was greater for ON cells compared with OFF cells \((P < 0.01)\). The bars in Fig. 11A mark the lower and upper bounds on information rate determined via the reconstruction method. For every cell, the directly measured rate fell within the bounds, even when the bounds were relatively restrictive (e.g., cells 4–6). It exceeded the lower bound by an average of 20 bps in X cells and 43 bps in Y cells, which amounted to 22 and 40% of the total bit rate of their respective spike trains. The contribution of nonlinear mechanisms to retinal coding was substantial.

While ON cells may have slightly higher information rates, OFF cells were more efficient at encoding information (Fig. 11B). The average number of bits per spike (bpk) was 1.7 ± 0.8 ms, the smallest bin size.
is little affected by spike threshold (illustrates that, for low gain settings, the average model response likely to be important. However, ON cells discharge spikes at a high and irregular rate (Fig. IA), and it might be expected that when the cells are visually stimulated the maintained discharge noise would render many of their spikes meaningless. Apparently retinal information transmission is not markedly degraded by spike discharge noise.

**Model simulations of retinal information transmission**

To gain insight into how target neurons in the brain might extract information that is encoded nonlinearly in ganglion cell spike trains, we considered what retinal mechanisms might cause a linear decoder to perform inadequately for strong inputs. Figure 12A plots the average response \( r_i(t) \) of an ON-Y cell to a center-matched spot repeatedly modulated by a 100% contrast random binary sequence, together with the best linear estimate of the response \( r_s(t) \). It is apparent on inspection of the waveforms that a major source of error in the linear estimate comes from the spiking mechanism, which does not permit negative spike rates. This suggests that rectification noise might explain, at least in part, the discrepancy between the lower bound and directly measured information rate of cat ganglion cells.

We tested the idea by creating artificial spike trains using a simple “integrate-and-fire” model of the spiking mechanism. The first step of the simulation was to convolve the best linear encoding filter \( h(l) \) with the set of random binary sequences \( s(t) \) presented to the cell. This step can be viewed as the transformation of the visual stimulus by the retinal network into the generator potential of the ganglion cell. The next step was to add Gaussian noise of variance \( \gamma \) to the spike generator input and integrate the result. Whenever the integral reached a threshold level \( \theta \), a spike was fired, and the integration was begun anew. The output \( r_{s,l}(t) \) of this linear-nonlinear model can be described mathematically as

\[
r_{s,l}(t) = \begin{cases} 
0, & \text{if } \sum_0^{L} s_i(t) + \alpha \gamma(t) + \beta h(t) < \theta \\
1, & \text{otherwise}
\end{cases}
\]

where \( \alpha \) determines the mean spike rate of the model neuron in the absence of a stimulus, \( \beta \) sets the gain of the model, and \( n(t) \) is Gaussian noise of unit variance. For the simulations \( \Delta t \) was 1/600 s. To make model spike trains semi-realistic, we fixed \( \alpha \) to the maintained discharge rate of the cell being simulated and adjusted \( \beta, \gamma, \) and \( \theta \) to give a driven discharge resembling that for the 100% contrast randomly modulated spot. Figure 12B shows that the model qualitatively reproduced the response power and low-frequency noise power of ganglion cell spike trains for this stimulus. We then varied \( \beta \) while holding the other three parameters constant and computed the information rate of model spike trains at each gain setting via both the reconstruction method and the direct method. Figure 12C illustrates that, for low gain settings, the average model response is little affected by spike threshold (inset \( c_1 \)), and model information rate is the same regardless of the method used to compute it. As with ganglion cell spike trains for weak visual inputs, a linear decoding strategy can capture all the information in model spike trains. For high gain settings, on the other hand, the generator signal frequently falls below spike threshold and the average model response gets clipped at zero spikes (inset \( c_2 \)). As with ganglion cell spike trains for strong visual inputs, the lower and upper bounds separate, and the directly measured information rate exceeds the lower bound rate. Differences in the lower bound and directly measured rates of X and Y cells can thus be partially, if not wholly, attributed to “spike patterns” of zeros that result from the stimulus-driven removal of spikes. An elaborate temporal coding scheme is not necessarily implied.
DISCUSSION

Our results show that cat retinal ganglion cells have the capacity to transmit an immense amount of information. This amount was generally largest for ON-Y cells and smallest for OFF-X cells. For the strongest stimulus with which we could drive the cells, the information rate was on the order of 100 bps or 1 bpk. Similarly, high information rates have been reported for cat lateral geniculate cells (Eckhorn and Pöpel 1975; Liu et al. 2001; Reingang and Reid 2000) and for visual neurons in other animals (de Ruyter van Steveninck et al. 1997) using the direct method. It implies that a single ganglion cell can encode 2^{100}, or approximately 10^{30}, different stimulus waveforms of 1-s duration. Since the receptive fields of each cell type tend to tile the retina with minimal center overlap (Wässle and Boycott 1991), the number of visual images that the brain could theoretically discriminate is astronomical.

To elicit high information rates, visual neurons must be driven with stimuli that are rich in spatiotemporal contrast. Such stimuli likely push the cells to their limits of operation or beyond. We have measured, for the first time, the information rates of different types of ganglion cells for a variety of visual inputs. To do so, a more efficient method, known as the reconstruction method, was applied. This method reduces data requirements at the cost of specificity: neural information rates are not measured directly but instead bounded to some range. We found that the information bounds for cat X and Y cells were the same whether the visual input maximally stimulated their receptive field center or their surround. Moreover, stimulating both receptive field components independently did not alter the bounds, while stimulating them synchronously generally lowered the bounds. Together these results indicate that ganglion cell spike trains contain information about visual signals from both the center and the surround, the center and surround mechanisms have similar signaling capacity, antagonism between the two mechanisms reduces the information conveyed by each, and the total information rate is limited. We also found that, for visual inputs of low-to-moderate strength, the bounds on information rate were essentially identical and thereby gave the true information rate of X and Y cells. Since the lower bound rate was determined by convolving ganglion cell spike trains with a linear filter, it is expected that target neurons in the brain could perform the same operation and decode all the information in the retinal output about such inputs. The bounds on information rate were found to differ greatly, however, for visual inputs of moderate-to-high strength. The lower bound rate saturated sharply at around 50–60 bps, while the upper bound rate increased monotonically with stimulus strength and always exceeded the directly measured rate. That the lower bound rate fell short of the directly measured rate indicates that within retinal spike trains there remained a substantial amount of information that a linear decoder cannot access. A similar conclusion was reached by others who compared the information conveyed by spike counts and spike patterns (Berry et al. 1997; de Ruyter van Steveninck et al. 1997; Reingang and Reid 2000). Hence, for strong visual inputs, a nonlinear decoder would be necessary to fully exploit the signaling capacity of retinal ganglion cells. Such a change in information transmission, from spike counts being important at low contrasts to spike patterns playing a greater role at high contrasts, has been seen in visual cortical neurons as well (Reich et al. 2001).

Whether brain neurons actually extract the maximal information from retinal spike trains is a separate and unanswered question. While the optimal linear decoder may miss information about strong inputs, we show that it would still discriminate up to approximately 2^{50} waveforms of 1-s duration per ganglion cell. Moreover, a given point in visual space is viewed by more than one type of ganglion cell. Overlap in the temporal properties of the various cell types makes their responses partially redundant. Between the information available to a linear decoder and redundancies in the retinal output, brain neurons might not need to extract all the information from every ganglion cell spike train to reliably construct the visual image.

Decoding the retinal output

If brain neurons do seek to extract maximum information from ganglion cell spike trains, they must undo the effects of retinal nonlinearities on the visual input. One nonlinearity that is likely to affect information transmission is the threshold for spike generation, so we incorporated it into a static linear-nonlinear model of the retinal coding mechanism. Model spike trains were found to exhibit many of the same properties as ganglion cell spike trains, including a difference between the lower bound and directly measured information rate for strong inputs. Since the only nonlinear element in the model was spike threshold, we conclude that it contributes to the so-called pattern information in retinal spike trains that a linear (rate-based) decoder cannot access.

Because it is difficult to see how ganglion cells could communicate anything about a visual stimulus when their generator potential is driven below spike threshold, it may seem strange that appending such a hard nonlinearity to the output of an otherwise linear encoder could increase information transmission to the directly measured rate. From an information theoretic standpoint, it happens because the threshold removes the cost of linear estimation errors during periods of inactivity, which allows the optimal linear encoder to better describe the average response during active periods. Since the information rate of a linear system is the same whether computed in a forward (stimulus-to-spikes) or reverse (stimulus-from-spikes) direction, we presume the spike threshold would have a related impact on the optimal linear decoder. From the standpoint of neurons in the brain, which might have spike trains from multiple types of ganglion cell at their disposal, a general strategy for decoding the retinal output may therefore be to process ganglion cell spike trains linearly and combine the signals decoded from ON and OFF cells when each cell type is active. This nonlinear decoding strategy is fairly simple in that it amounts to ignoring the cells during inactive periods when the linearly decoded signal would be erroneous.

The spike threshold is certainly not the only nonlinearity that affects retinal information transmission. It is well known, for example, that various gain control mechanisms operate in the retina as well. These nonlinear mechanisms adapt the retina to the mean light level (Shapley and Enroth-Cugell 1984) and to spatial and temporal contrast (Kim and Rieke 2001; Shapley and Victor 1978; Smirnakis et al. 1997). The stimuli that we presented were all at a single (photopic) light level, and in a given set of trials, at fixed contrast. This should have reduced the impact of light
adaptation and contrast adaptation on our results. Indeed, we found that the output of cat ganglion cells was generally stable by the end of the first trial, suggesting that adaptation to changes in the stimulus between trial sets was largely complete in a few seconds. We cannot, however, exclude the possibility that gain control mechanisms modulated the retinal output during each trial. Fast forms of adaptation have been demonstrated in cat and other animals (Kim and Rieke 2001; Shapley and Victor 1978). Such a dynamic nonlinearity could be responsible, in addition to the spike threshold, for the difference in lower bound and directly measured rates of cat ganglion cells at high stimulus contrasts. It could be incorporated into a more detailed model of the retinal coding mechanism as a stimulus-dependent change in the retinal filter (van Hateren et al. 2002; Victor 1987) or spike threshold (Keat et al. 2001).

**Rate and temporal coding**

Much research has focused on the relative importance of rate codes and temporal codes to neural information transmission, although the difference between the two coding schemes is not always clear. To make the distinction evident, it is usually said that rate codes place importance on the number of spikes fired per unit of time, whereas temporal codes ascribe meaning to the patterning of spikes within each time unit (deCharms and Zador 2000; Ferster and Spruston 1995; Rieke et al. 1997; Victor 1999). This description of neural coding works only when the time scale on which the brain operates is known. In practice, the time unit is often chosen arbitrarily, and so what may appear like temporal coding on one scale could be viewed as rate coding on another (deCharms and Zador 2000). To circumvent the problem it has been suggested that the high-frequency cutoff of neural responses is the time unit of physiological relevance (Theunissen and Miller 1995), meaning that messages would be rate encoded at coarser scales and temporally encoded at finer scales. Based on this definition and the merging of ganglion cell response and noise spectra in Fig. 3, the importance of the two schemes to retinal coding should be discernable at a time scale of 16 ms or less (i.e., one-half the period of the 30-Hz cutoff frequency given sampling criteria). From Fig. 6, it would then seem that X cells may have rate encoded all their messages, while Y cells may have temporally encoded a portion of theirs, as integrating beyond the cutoff frequency had little effect on the information bounds of the former cells but increased the upper bound rate of the latter.

Still, the possibility of pure rate coding for Y cells cannot be completely eliminated because the upper bound rate exceeded their directly measured rate even at 30 Hz.

An analysis of the linear and nonlinear information content of neural spike trains paints a perhaps more clear picture of what is rate and temporal coding. A spiking neuron anywhere in the nervous system can be categorized into one of four types of information channels (Fig. 13). The simplest type is one in which the spiking neuron, or the chain of neurons preceding it, produces a signal that is linearly related to the stimulus and the spike generator of the neuron encodes the signal linearly in its pattern of discharges (channel 1). All cat X and Y cells were found to behave like this channel for low-to-moderate strength inputs. It clearly communicates information using a rate code, since the stimulus can be recovered from the spike train without assuming a dependence between spikes. A second type of channel is one in which the stimulus transformation is linear but the signal encoder is nonlinear (channel 2). We found that this channel applied to cat ganglion cells for moderate-to-high strength inputs (i.e., the linear-nonlinear model). Unlike the first channel, it transmits some information temporally. This is because the spike rate cannot track the neural signal due to distortions in spike timing caused by rectification, refractoriness, or adaptation of the spiking mechanism. Only by compensating for the effect of these nonlinearities on the precise patterning of spikes can the temporally coded information be recovered (e.g., Berry and Meister 1998). The remaining two types of channels have a stimulus transformation that is nonlinear and a signal encoder that is linear (channel 3) or nonlinear (channel 4). Because of light adaptation and other nonlinear properties of the retinal network that were not invoked by the stimuli used here, a complete description of cat ganglion cell behavior ultimately involves these types of channels.

It is worth noting that the third channel conveys information using a rate code, since the encoder is linear, but may appear to use a temporal code if the stimulus transformation is very nonlinear. For example, suppose an apple and pear typically evoke from a neuron the spike trains “10100010” and “01001001” in a 32-ms interval. Without knowing the encoding properties of the neuron, one might be tempted to conclude that the fruit were temporally encoded because the number of spikes in the two trains is the same. Only the pattern of spikes is different. However, since the neural signal (i.e., generator potential) that gave rise to these spike trains is linearly recoverable in this hypothetical example, such a conclusion would be incorrect. That the temporal patterning of spikes differs in the two trains has nothing to do with the spiking mechanism using a “pattern code” to represent apples and pears and all to do with the dynamics of the generator potential elicited by the fruit. In contrast, the fourth channel involves temporal coding to some degree. In the extreme, this could mean that there are no rules governing the relationship between generator potentials and spike trains. The encoder basically uses the generator potential to look up in a random table of spike patterns the one that corresponds to an apple, pear, etc. Although such a scheme is highly nonphysiological, it serves to illustrate the realm of temporal coding.

Discussions of neural coding usually confound the conver-
sion of the stimulus to a neural signal with the subsequent conversion of that neural signal to a spike train. The picture of rate and temporal coding painted in Fig. 13 boils the debate down to the functional characteristics of the spiking mechanism since the terms have meaning only when neural messages are transmitted with spikes. Emphasis is thereby placed on the relationship between the stimulus and the neural signal, since this relationship is generally thought to determine what computations a neuron or neural network performs and thus what information there is for the spike train to encode. This is not to say the spiking mechanism is only involved in information transmission. While signal rectification caused by spike threshold or spike refractoriness could be regarded as a biophysically unavoidable consequence of firing spikes at abnormally low or high rates, other nonlinear behaviors like bursting (Mukherjee and Kaplan 1995; Reich et al. 2000; Reinagel et al. 1999) and spike frequency adaptation (Ahmed et al. 1998; Sanchez-Vives et al. 2000) are difficult to dismiss because they can act on time scales of the neural signal. For neurons that exhibit such behaviors the spiking mechanism could play a role in information processing as well as transmission.

ACKNOWLEDGMENTS

We thank Dr. Christina Enroth-Cugell for helpful comments on the manuscript.

GRANTS

This research was supported by National Eye Institute Grants R01-EY-06669 and F32-EY-06908.

REFERENCES


