INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **13** (2002) 531–552 PII: S0954-898X(02)53153-5

A test of metabolically efficient coding in the retina

Vijay Balasubramanian[1,3] **and Michael J Berry II**[2]

[1] David Rittenhouse Laboratories, University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

E-mail: vijay@endive.hep.upenn.edu and berry@princeton.edu

Received 20 May 2002, in final form 5 September 2002
Published 25 September 2002
Online at stacks.iop.org/Network/13/531

Abstract

We tested the hypothesis that aspects of the neural code of retinal ganglion cells are optimized to transmit visual information at minimal metabolic cost. Under a broad ensemble of light patterns, ganglion cell spike trains consisted of sparse, precise bursts of spikes. These bursts were viewed as independent neural symbols. The noise in each burst was measured via repeated presentation of the visual stimulus, and the energy cost was estimated from the total charge flow during ganglion cell spiking. Given these costs and noise, the theory of efficient codes predicts an optimal distribution of symbol usage. Symbols that are either noisy or costly occur less frequently in this optimal code. We found good qualitative and quantitative agreement with the measured distribution of burst sizes for ganglion cells in the tiger salamander retina.

1. Introduction

Optimization principles constitute a fundamental approach to understanding the structure and function of the nervous system. By quantifying the performance of neurons, information theory provides very general tools for constructing such principles. An intuitive design, especially for the sensory periphery, is that a neuron strives to maximize the mutual information between its input and output, subject to realistic constraints. Noise, either in the input or added by the neuron, is a fundamental constraint. Several researchers have had success predicting how the receptive fields of retinal neurons should be organized to maximize the transmission of information about natural visual images (Barlow 1961, 2001, Atick 1992, van Hateren 1992). This success might lead one to imagine that many neurons strive to maximize their information transmission rate.

However, there is a difficulty: spiking neurons, which constitute the vast majority of neurons in the vertebrate brain, produce action potentials at average rates far below the rate that maximizes transmitted information. While axons can sustain precise firing at rates up to

[3] Both authors contributed equally to this work.

hundreds of spikes per second (Berry *et al* 1997, de Ruyter *et al* 1997, Buracas *et al* 1998, Reinagel and Reid 2000, Kara *et al* 2000), many neurons fire much less frequently. For instance, ganglion cells in the salamander retina fire 1–5 spikes s^{-1} under varied stimulation (Berry *et al* 1997), neurons in the LGN of the cat fire 5–15 spikes s^{-1} in response to natural movies (Stanley *et al* 1999), and pyramidal neurons in the rat cortex fire at 1–4 spikes s^{-1} (Schoenbaum *et al* 1999, Fanselow and Nicolelis 1999), to name just a few. Therefore, one must conclude either that these neurons are not well designed for information transmission or that other constraints are important.

In fact, neurons are some of the most energy intensive cells in the body (Wong-Riley 1989), and spikes comprise a significant portion of their energy usage (Siesjo 1978, Ames 2000, McIlwain and Bachelard 1985, Atwell and Laughlin 2001). This suggests that neural codes might be constrained by a requirement of energy efficiency. Previous work has developed a general theory of energy-efficient codes that predicts the optimal usage distribution of a set of discrete, independent neural symbols, each subject to an energy cost and corrupted by noise (Levy and Baxter 1996, Baddeley *et al* 1997, Balasubramanian *et al* 2001, de Polavieja 2002, Schreiber *et al* 2002). We used this prediction to make a quantitative test of the hypothesis that the distribution of symbols in the neural code of retinal ganglion cells is energy efficient. Under a wide variety of stimulus conditions, the output of ganglion cells is a sparse set of discrete firing events (Berry *et al* 1997). Each firing event is a tight burst that we view as a composite neural symbol characterized by its number of spikes. The noise of each firing event was measured by repeated presentations of the same stimulus, and the energy cost was estimated by simulations of the ionic current flow during spiking. The theoretically optimal distribution of burst sizes was found to be in both qualitative and quantitative agreement with experiment—both noisy and expensive symbols were used less often.

Our work provides support to the hypothesis that energy costs are an important constraint for the neural code implemented by retinal ganglion cells. However, as we discuss, further work is necessary to compare the single parameter appearing in our model to the actual biophysics of energy consumption. Perhaps the most interesting fact emerging from this work is the importance of accounting for noise. The reduced frequency of noisy firing events in our data is significant and correlates well with predictions from optimal coding theory.

2. Materials and methods

2.1. Experimental methods

Experiments were performed on isolated retinas from the larval tiger salamander (*Ambystoma tigrinum tigrinum*). The retina was perfused with Ringer's medium (110 mM NaCl, 22 mM $NaHCO_3$, 2.5 mM KCl, 1.6 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM D-glucose), which was oxygenated by continuously bubbling a 95/5% O_2/CO_2 mixture through it. Action potentials from retinal ganglion cells were recorded using a multi-electrode array (see Berry *et al* (1997) and Meister *et al* (1994) for further details) while the preparation was carried out at room temperature. This temperature is realistic for tiger salamanders in their natural habitat (Roth 1987). The retina was stimulated with illumination from a computer monitor that was focussed onto the photoreceptor layer.

Stimuli were drawn from a random flicker ensemble, in which the monitor screen was divided into square regions and intensity values for each square were chosen every 30 ms from a Gaussian distribution. The mean light level was in the photopic range for salamander (7 mW m^{-2}), and the temporal root-mean-squared fluctuation (contrast) was 35% of the mean. This level of contrast approximates that found under more natural viewing conditions (van der

Schaaf and van Hateren 1996). Experiments were performed using squares that ranged in size from 134 μm, which was roughly the size of a ganglion cell's receptive field centre, up to 1070 μm, as well as with spatially uniform stimuli.

Spike trains were recorded during 30 or 60 repeated presentations of stimulus sequences lasting from 120 to 540 s under spatially uniform flicker or during 60 repeats of six, 30 s stimulus segments having different spatial modulation. Total recording times ranged from 120 up to 270 min, with an average of 192 min. A total of 41 ganglion cells from four retinas were recorded for long enough times so that the error in measuring their burst size distribution was smaller than the probability itself (see section 2.6 for details). The majority (69%) were fast OFF-type cells, identified by their unique reverse correlation to random flicker stimulation (Smirnakis *et al* 1997). These cells resemble Y cells in the cat and M cells in the monkey. The rest of the ganglion cells (31%) were slower OFF-types. In the tiger salamander, as in most amphibians, the overwhelming majority (80–90%) of ganglion cells are OFF-type (Smirnakis *et al* 1997, Roth 1987).

2.2. Firing event identification

Under random flicker stimulation, salamander ganglion cells respond in sparse, highly precise firing 'events' (Berry *et al* 1997). Figure 1 shows typical spike trains recorded from a salamander ganglion cell under two different kinds of checkerboard random flicker. Most events were easily identified as single peaks in the firing rate flanked by periods of silence (figure 1(C)). In fact, 89% of all event boundaries were set at times when the ganglion cell was completely silent. However, in some instances, more ambiguous peaks in the firing rate were observed. In order to provide a consistent definition of firing events, we identified event boundaries as reproducible minima in the firing rate. For each minimum ν, the ratio of the firing rate at its adjacent peaks (p_1, p_2) to that at the minimum was required to exceed a threshold value, $\sqrt{p_1 p_2}/\nu > \phi$, with 95% confidence. The threshold was set to $\phi = 2$; event boundaries were rather insensitive to its value (see Berry *et al* (1997) for addition details). Once event boundaries were drawn, all spikes between successive boundaries were assigned to a single event. We denote the number of spikes at event number a on stimulus trial b as n_{ab}. If no spikes were observed in a given trial, n_{ab} was zero. The average number of spikes in that event is $N_a = \langle n_{ab} \rangle_b$, and the variance is $\mathrm{var}(N_a) = \langle n_{ab}^2 \rangle_b - \langle n_{ab} \rangle_b^2$.

2.3. Energy efficient neural codes

The properties of an optimally energy efficient neural code were defined in earlier work (Levy and Baxter 1996, Balasubramanian *et al* 2001, de Polavieja 2002, Schreiber *et al* 2002). Here, we imagine that a neural circuit maps stimuli deterministically onto a discrete set of desired neural symbols $\{Y\}$, and that noise is added before the output of actual neural symbols $\{Z\}$ (see figure 3(A)). The goal of an energy efficient code is to find the probability p_k of using each output symbol z_k that optimizes information transmission per unit metabolic cost. In the absence of noise, the solution is a Boltzmann distribution:

$$p_k = \mathrm{e}^{-\beta E_k} \tag{1}$$

where β is set such that $\sum_k \mathrm{e}^{-\beta E_k} = 1$. However, noise profoundly alters the optimal symbol distribution. One of the main contributions of Balasubramanian *et al* (2001) was to show how noise can be incorporated into the theory of energy efficient codes, and to demonstrate that small changes in noise levels can sometimes result in dramatic changes in the optimal probabilities of symbol use. For example, in an energy efficient code we would expect inexpensive symbols to be used more often. However, because of the competition with noise, it can happen that a

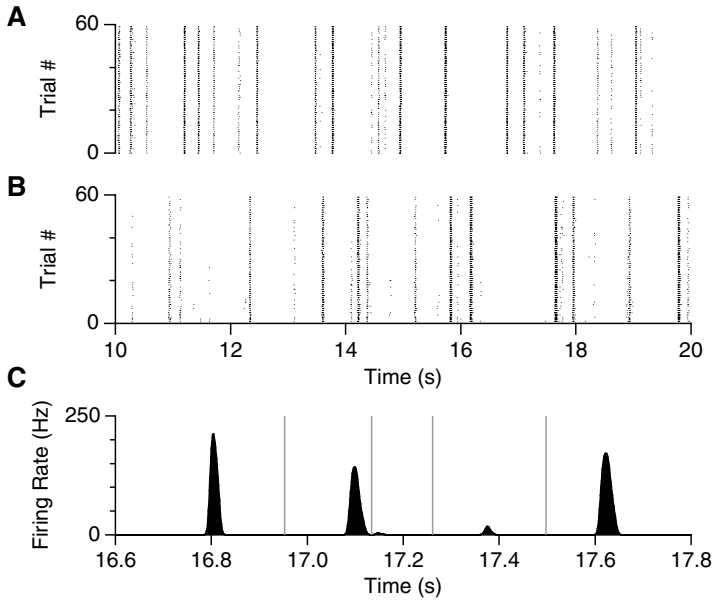

Figure 1. Spike trains recorded from a salamander ganglion cell under 60 repeated presentations of random flicker with (A) 534 μm checkers and (B) 134 μm checkers. (C) Firing rate as a function of time, measured as a histogram over spike trains using a 3 ms time bin, for the 534 μm checkers, exhibiting clear firing events. Event boundaries are shown as vertical, grey lines. Note that there is an event boundary near time = 17.13 s, where the cell is not unambiguously silent.

cheap but noisy symbol is heavily suppressed. Hence it is of critical importance to carefully measure and incorporate noise in biological systems. One of the primary purposes of this work is to show how this is performed in the analysis of retinal coding.

Noise is expressed as a transition matrix $Q_{k|j}$ that describes the likelihood that the intended message or *presymbol* y_j emerges from the channel as the transmitted symbol z_k. Given this 'noise matrix', an iterative algorithm generalizing classic work by Arimoto and Blahut (Arimoto 1972, Blahut 1972, 1987) computes the symbol distribution that optimizes information transmission at a fixed average cost E, essentially by treating the number of bits of information lost to noise as an additional incurred cost. By iteratively optimizing the transmitted information subject to the fixed average energy cost E, the algorithm chooses an optimal distribution q_j for the presymbols y_j entering the noisy information channel, leading to transmission of the symbols z_k with optimal probabilities $p_k = \sum_j Q_{k|j} q_j$.

Using this algorithm, we can always find the probability distribution of coding symbols that optimizes the information I they transmit about a sensory stimulus at a fixed average cost E. Numerical optimization of I/E then yields the code that maximizes information transmission per unit metabolic cost. The necessary ingredients of this model are the energy cost of each output symbol, E_k, and the noise in each presymbol, $Q_{k|j}$. A detailed description of the algorithm appears in Balasubramanian *et al* (2001). All numerical optimizations in this paper were carried out using the FindMinimum routine in the commercial software package Mathematica 4.0. This routine uses the steepest descent method and finds minima to six digits of precision, greater than necessary given the error bars on our data.

The energy efficient codes derived in this manner have three salient features that are relevant for this study. First, more energetic symbols are suppressed in the optimal distribution.

Second, noisier symbols are also suppressed. Both these facts are illustrated in panels B and C of figure 3. Finally, a useful property of metabolically efficient codes is that the distribution over symbols that maximizes bits per energy is invariant under rescalings of the energy costs $E_n \rightarrow \lambda E_n$, where λ is some fixed number (see (Balasubramanian *et al* 2001) for discussion). This fact allows us to eliminate one parameter in our model; the linear energy model $E_n = a_1 + a_2 n$ can be rescaled by a factor of $1/a_1$ to $E_n = 1 + bn$, while leaving the optimal distribution unchanged. The resulting parameter $b = a_2/a_1$ might be interpreted as the relative cost of a spike compared to the baseline cost of keeping the cell alive. However, it might also be thought of as effectively including energy costs from other neurons in the circuit.

2.4. Classifying presymbols

In order to provide a consistent and statistically significant classification of firing events into presymbols, we compiled the distribution of spike counts across repeated trials for each event and used the following algorithm:

(1) Compare the number of observations of the most common event size B_1 to the number of observations of the next most common B_2. If B_1 exceeds B_2 with 95% confidence (assuming Poisson counting statistics), then the event arises from an integer presymbol corresponding to the most common burst size. Most events (68%) were classified as arising from integer presymbols. An example of the distribution of spike counts for an event that was classified as arising from an integer presymbol is shown in figure 5(A).
(2) If B_1 does not exceed B_2 with statistical confidence, compare B_1 to the observations of the third most common event size B_3. If B_1 exceeds B_3 with 95% confidence, then the event is a half-integer presymbol corresponding to the two most common event sizes. The vast majority of these events were split between successive spike counts, such as 1 and 2 spike bursts, accounting for 28% of all events. An example is shown in figure 5(B). The remaining 0.2% of all events had bimodal burst size distributions, often between 0 and 2 spikes, shown in figure 5(C). Because the latter events were so rare, we simply treated them as integer presymbols corresponding to the most common burst size.
(3) If B_1 does not exceed B_3 with statistical confidence, check whether the three most common event sizes have successive values. If so, check if the most common burst size is flanked by the next two most common. In this case, assume that the event is a noisier integer event corresponding to the most common burst size (2% of all events fell in this category). Otherwise, assume that this is a noisier half-integer event characterized by the two most common burst sizes (2% of all events fell in this category).
(4) If the three most common burst sizes were not successive, we could not clearly characterize the presymbol. As only 0.05% of all events were in this final category, we again assumed that they were noisy integer events corresponding to their most common burst size.

The results section motivates the classification algorithm above in terms of the measured characteristics of burst size distributions and noise matrices for different events.

2.5. Burst size distributions and noise matrices

For each ganglion cell, we compiled a burst size distribution $P(z)$ from the spike counts for individual events and trials, n_{ab}. Increasingly large bursts were increasingly rare, and their probability consequently had large uncertainty due to random error. Therefore, in our analysis we only included bursts up to a maximum burst size, $z_{\max}$. We denote the number of observations of burst size z in a single stimulus trial by $N(z)$. The probability that a randomly

selected burst in the trial has size z is then $P(z) = N(z)/\sum_z N(z)$, where we only include z up to $z_{\max}$. Our criterion for choosing $z_{\max}$ was that ∼98% of all bursts were included. Over our population of ganglion cells, $z_{\max}$ ranged from 4 up to 8 spikes, with an average of 5.0.

We compiled the noise matrix $Q_{z|y}$ by estimating which presymbol y_a gave rise to each firing event (see classifying presymbols) and adding the histogram of n_{ab} for event a over repeated trials b to the row of Q corresponding to y_a. We denote the number of observations of burst size z given presymbol y as $L_{z|y}$, and the number of observations of presymbol y as M_y. Each row of the noise matrix was separately normalized by the number of observed presymbols, $Q_{z|y} = L_{z|y}/(M_y R)$, where R is the number of stimulus repeats. This normalization gives $\sum_z Q_{z|y} = 1$ as it should in order that $Q_{z|y}$ describe the probability that a given presymbol y produces the output z. As discussed above, the measured burst size distribution was truncated at a given $z_{\max}$ for each cell, which can create unwanted 'edge effects' in the theoretical analysis of the energy optimal burst size distribution. In order to avoid these edge effects, we kept burst sizes greater than $z_{\max}$ in the noise matrix. However, after the theoretical prediction of the optimal burst size distribution $P^*(z)$ was obtained, we truncated it at $z = z_{\max}$ and re-normalized to unity, $\sum_{z=1}^{z_{\max}} P^*(z) = 1$. This allowed us to compare the predictions of the theory properly with the measured distribution $P(z)$. To make this comparison quantitative, we measured the chi-squared error between theory and experiment, normalized by the number of degrees of freedom:

$$\chi^2 = \frac{1}{z_{max} - 2} \sum_{z=1}^{z_{\max}} \frac{(P^*(z) - P(z))^2}{\mathrm{var}[P(z)]}, \tag{2}$$

where $\mathrm{var}[P(z)]$ is the variance in the probability of measuring burst size z (defined below).

2.6. Error analysis

In making quantitative comparisons between optimal and actual neural codes, we must treat the uncertainty in our measurements properly. Assigning error bars to the burst size distribution is not trivial, because repeated presentations of the stimulus do not yield independent measurements of the burst size distribution: these measurements are related to each other by the noise matrix. However, we can use the fact that random flicker is completely uncorrelated from one stimulus frame to the next. Therefore, the stimulus patterns that cause each type of firing event occur independently and with a constant probability per unit time. The number of presymbols of each type elicited by a stimulus sequence of fixed length will vary with Poisson statistics:

$$\mathrm{var}[M_y] = M_y. \tag{3}$$

For each presymbol, variations in the number of output spikes observed on different stimulus trials are also independent and obey Poisson statistics. This implies that variations in $L_{z|y}$, the number of observations of z given presymbol y, are also Poisson:

$$\mathrm{var}[L_{z|y}] = L_{z|y}. \tag{4}$$

We can use equations (3) and (4) to find the error $\mathrm{var}(N(z))$. By definition of the noise matrix, $N(z) = \sum_y Q_{z|y} M_y = (1/R) \sum_y L_{z|y}$. Therefore

$$\mathrm{var}[N(z)] = (1/R) \sum_y \mathrm{var}(L_{z|y}) = (1/R) \sum_y L_{z|y} = N(z). \tag{5}$$

In order to find the error in the probability distribution over burst sizes $P(z)$, we use standard error propagation:

$$P(z) = \frac{N(z)}{\sum_{z'} N(z')} \tag{6}$$

$$\mathrm{var}[P(z)] = \left(\frac{1}{\sum_{z'} N(z')} + \frac{N(z)}{(\sum_{z'} N(z'))^2} \right) \mathrm{var}[N(z)] = \left(\frac{\sum_{z'} N(z') + N(z)}{(\sum_{z'} N(z'))^2} \right) N(z) \tag{7}$$

As we have discussed, a crucial input to the theory is the experimentally measured noise in each ganglion cell. Poisson variation in the number of presymbols y and the resulting observations of outputs z imply that the entries in the noise matrix $Q_{z|y} = L_{z|y}/M_y R$ have variances

$$\mathrm{var}[Q_{z|y}] = \frac{2L_{z|y}}{M_y R} = 2Q_{z|y} \tag{8}$$

where we have once again used standard error propagation. Since we use a numerical algorithm to derive the metabolically optimal code, it is not possible to derive an analytic expression for the error bars in the theoretical prediction. Furthermore, performing a complete numerical error propagation is prohibitively time consuming. Instead, we have estimated the effects of the measurement errors on the theory by propagating the most important uncertainties, namely those in the diagonal entries of the noise matrix, through the algorithm that determines the optimal metabolically efficient code. The resulting errors are reported in figure 7 by including the maximum and minimum deviations in the predicted probability for each symbol when the diagonal entries of the noise matrix were perturbed by plus or minus one standard deviation. In the limit of small errors, this min–max error is similar to the result of standard error propagation. Further discussion appears in the results.

2.7. Ion channel simulations

We studied the energy cost associated with electrical activity using a single-compartment model of tiger salamander ganglion cells due to Fohlmeister and Miller (FM) (1997). This model includes five membrane currents: voltage gated sodium (I_{Na}), voltage gated calcium (I_{Ca}), delayed rectifier potassium (I_{K}), inactivating potassium (I_{A}) and calcium-gated potassium current ($I_{\mathrm{K,(Ca)}}$). In addition to these ionic currents, the model has membrane leak and capacitance, and includes calcium sequestration and pumping with a relaxation mechanism having a single time constant.

All parameters were as given in table 1 and the text of Fohlmeister and Miller (1997) except as specified below. The leak reversal potential was chosen as $V_{\mathrm{leak}} = -65$ mV, so that after a long period with $I_{\mathrm{stim}} = 0$, the membrane approaches a resting potential of $V = V_{\mathrm{leak}}$. The leak conductance was chosen to be $\bar{g}_{\mathrm{leak}} = 0.05$ mS cm^{-2}. (This is the value given in the text of Fohlmeister and Miller (1997) and not in their table 1.) The time constant for the removal of calcium from the intracellular space either by pumping out of the cell or by sequestering in internal stores was chosen to be $\tau_{\mathrm{Ca}} = 50$ ms. Finally, we assumed a cell radius $r = 25$ μm and carried out all simulations at room temperature.

With these parameters fixed, we studied the energetics of spiking by stimulating the model with brief current pulses following long resting periods. The charge flow in each ionic channel was obtained by integrating the corresponding current as a function of time. The ATP consumption in ganglion cell soma required for transporting this charge will be linear in the amount of charge, allowing us to determine how the energy cost depends on the number of spikes and time intervals between spikes in a burst. The results of these simulations are described in the results.

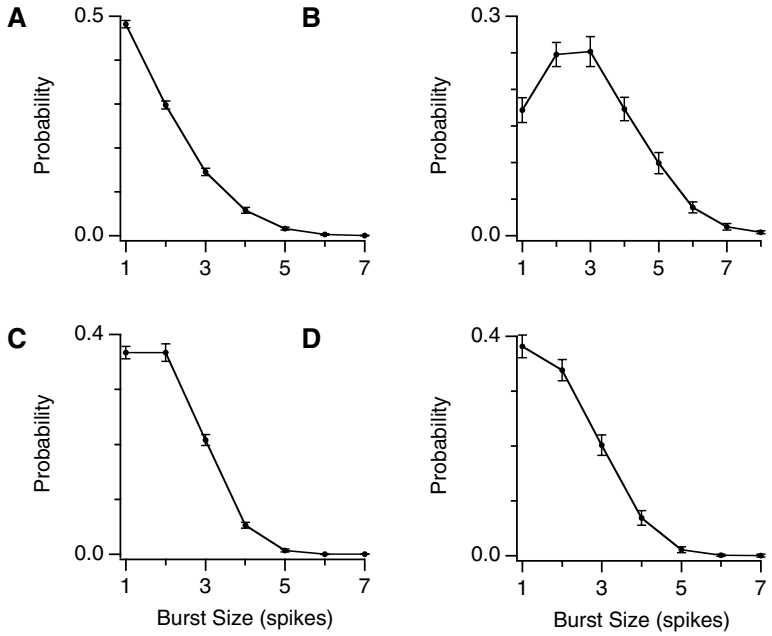

Figure 2. Burst size distributions for four representative salamander ganglion cells, showing (A) pure exponential shape, (B) significant suppression of small bursts, (C) and (D) flattening of the distribution for small bursts.

3. Results

3.1. Burst size distributions

Recordings were made from 41 ganglion cells in the larval tiger salamander retina while stimulated with the diverse spatial and temporal patterns of light found in random flicker (see the materials and methods section). Under random flicker stimulation, these ganglion cells respond in sparse, highly precise episodes of firing (see figure 1; Berry *et al* 1997, Berry and Meister 1998, Meister and Berry 1999). These firing 'events' were tight bursts containing one to eight spikes with inter-spike intervals of 5–10 ms. Their first spike was elicited with timing that typically jittered by ∼4 ms from one repeated stimulus presentation to the next. The total number of spikes in an event typically varied by ∼0.5 spike. In between events, the spontaneous firing rate was strictly zero. Events occurred at rates of 1–3 s^{-1}, with inter-event time intervals broadly distributed from 50 ms up to 3 s. Because of their high precision, each firing event can serve as an individual visual message of high fidelity. Because of the long inter-event time intervals and the consequently weak correlations between successive firing events, each event can serve as an independent visual message. This led us to consider each firing event as a composite coding symbol, characterized by its time of occurrence and total number of spikes.

Because events with different numbers of spikes can encode different visual messages, this paper focuses on the distribution of burst sizes used by retinal ganglion cells to represent a broad ensemble of visual stimuli. Figure 2 shows such distributions for four cells along with error bars that reflect the expected Poisson variation for stimuli drawn from the same ensemble (see the materials and methods section). A characteristic feature shared by all cells is the sharp

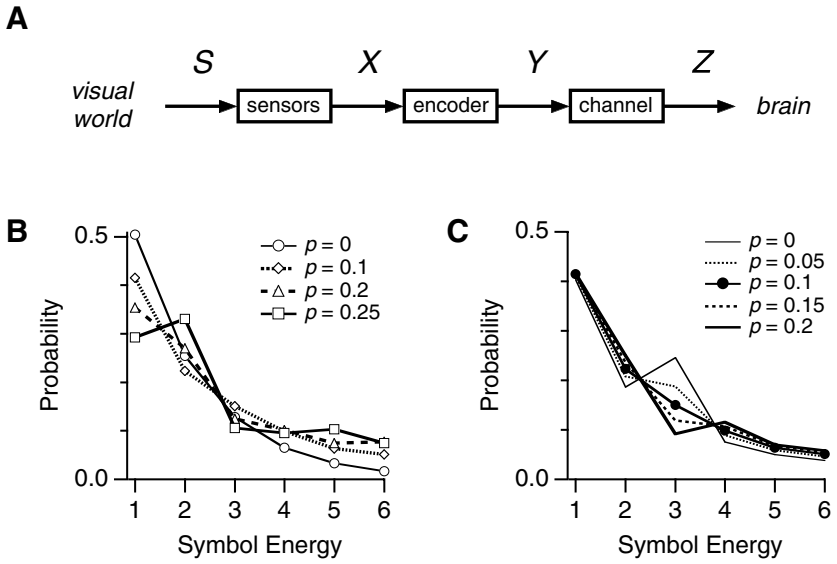

Figure 3. (A) Simplified schematic of our model of a sensory system. (B) and (C) Effects of noise in the model: the optimal usage probability of each symbol is plotted for different levels of noise. In both (B) and (C), there are six output symbols, and the energy cost of symbol k is $E_k = k$. In (B), the noise matrix has the following 'diagonal' form: $\Pr(z|y) = 0$, except for $\Pr(z = k|y = k) = 1 - 2p$ for $k = 1 \ldots 6$; $\Pr(z = k + 1|y = k) = \Pr(z = k - 1|y = k) = p$ for $k = 2 \ldots 5$; $\Pr(z = 2|y = 1) = \Pr(z = 5|y = 6) = 2p$. The optimal usage distribution is exponential when $p = 0$, showing suppression of energetically expensive symbols. As the noise increases, there is a marked deviation from an exponential symbol distribution. In (C), the noise matrix is of the same form as in (B), but with $p = 0.1$ for all k except $k = 3$, for which the noise probabilities are varied. This panel shows that noise in a single symbol strongly suppresses the use of that symbol in the optimal distribution. Results in (B) first appeared in Balasubramanian *et al* (2001).

falloff of bursts of large size. An energy efficient code would behave in this way since large bursts cost more energy. Some of the cells have the surprising feature of a suppression of small bursts (figure 2(B)), and most show a flattening at small burst sizes (figures 2(C) and (D)). Naively, this seems to point away from energy efficiency as an organizing principle for the retinal code, since smaller bursts cost less energy. However, efficient coding also requires that noisy symbols be used less frequently, and as we shall see, small bursts are suppressed in precisely those cells where these bursts are not reliably produced from trial to trial.

3.2. Structure of the model

In order to assess the energy efficiency of the retinal code, we developed a simple model of a sensory system in which the properties of an optimally energy efficient code could be solved (Balasubramanian *et al* 2001). In this model, the action of a neural circuit was idealized as occurring in three stages: stimuli in the external world $\{S\}$ are detected by an array of sensors and converted into neural signals $\{X\}$, these signals are encoded deterministically into a discrete set of composite symbols $\{Y\}$, and finally noise is added before the output $\{Z\}$ emerges (see figure 3(A)).

This model idealizes the action of the retina by combining all types of noise into a single stage, where the intended neural symbol Y is transformed into the actual symbol Z. In particular, photon noise is lumped together with less well understood internal sources of

variability in the retina. Because of this abstract representation of noise, the internal symbols Y cannot be directly measured, nor can the noise be broken down into its distinct components. While it would be interesting to have a model that allows for separate study of, say, the effects on an optimal code of photon noise as a function of light level, we shall see that our approach has the advantage that all of the noise parameters can be directly measured from data.

In the model, the output $\{Z\}$ is a string of discrete, independent neural symbols $\{z_0, z_1, \ldots, z_n\}$. Each symbol z_k has an energy cost E_k associated with its production. In the case of the retina, individual firing events are viewed as the output symbols. Firing events are distinguished only by their number of spikes: z_1 is a one-spike burst, z_2 is a two-spike burst, etc. We add a symbol z_0 representing every 50 ms period of silence. (Our results display very little sensitivity to the value of the time bin used to discretize periods of silence, as we have checked by picking a variety of time bins.) Because spiking represents a significant energy expenditure for a neuron, firing events with different numbers of spikes have significantly different energy costs. In what follows, we argue that the energy cost of a burst is linearly related to its number of spikes.

Given a set of costs $\{E_0, E_1, \ldots, E_n\}$ for the output symbols z_k, and a noise matrix $Q_{k|j} = \Pr(z_k|y_j)$ which gives the probability that a presymbol y_j yields an output symbol z_k, we can calculate the distribution of symbols z_k which optimizes the information transmitted per unit metabolic cost (see methods) (Balasubramanian *et al* 2001). The theory predicts that both noisy and energy expensive symbols are suppressed in the optimal code. For example, figures 3(B) and (C) show a model with six input and six output symbols with energies $E_k = k$. In the absence of noise, the symbol y_j is transmitted as $z_k = y_j$, and bits per energy is maximized when the usage distribution of outputs z_k is exponential ($p_k = \mathrm{e}^{-\beta E_k}$). Figure 3(B) shows that the addition of uniform noise distorts the optimal symbol distribution, with a particularly strong effect on the symbol y_1. Figure 3(C) shows that adding noise only to y_3 leads to marked suppression in its use, since this symbol is then not a reliable conveyor of information. Indeed, it is shown in Balasubramanian *et al* (2001) that small changes in the noise can result in large changes in the optimal distribution of output symbols. So, to study efficient coding by a ganglion cell, we must carefully characterize its noise.

3.3. Noise in ganglion cells

Repeated presentations of the same input to the retina lead to a distribution of output burst sizes for each firing event. Figure 4(A) plots the variance of the burst size distribution of a given event against its mean for one ganglion cell. Most values have $\mathrm{var}(N) < 0.5$, indicating that the number of spikes in a firing event is precisely determined by the stimulus. The solid line at the bottom of the graph shows the theoretical lower bound on the variance of events with a given mean, obtained when the event uses only outputs n and $n-1$ with the appropriate proportions. This theoretical lower bound drops to zero for integer values of the mean spike number and rises to 0.25 for half-integer values. A corresponding pattern is seen for the actual firing events: lower variance occurs near integer values of the mean spike number. This pattern is demonstrated by the thick line in figure 4(A), which is an average over the observed variance of all events as a function of mean burst size.

This difference between high and low variance events arises fundamentally from the nature of the mapping between the continuous patterns of light occurring in the environment and the discrete outputs of the retina. The set of all patterns of light constitutes a high-dimensional space of stimuli, and we can view the retinal mapping as dividing stimulus space into regions corresponding to presymbols $\{\mathrm{Y}\}$ (see figure 4(B) for a schematic). Noise has the effect of mapping a given stimulus onto a nearby location in stimulus space. When the input starts out

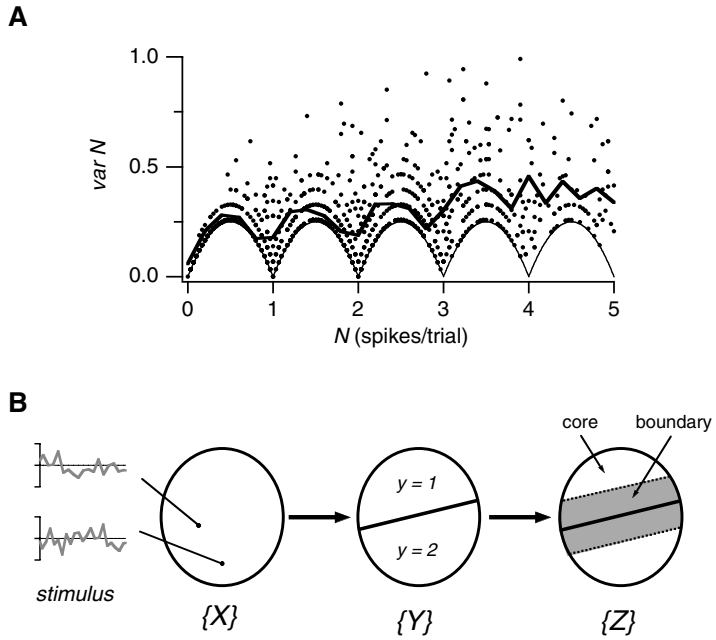

Figure 4. (A) Variance of the spike count in a firing event plotted as a function of average spike number for 1043 events recorded from one ganglion cell. (B) Deterministic mapping of stimulus to presymbol: noisy boundary versus reliable core. Visual stimuli occupy a large space, where each point is a pattern of light intensity as a function of time (shown on far left). This stimulus space $\{X\}$ is mapped deterministically into presymbols $\{Y\}$, which are in turn mapped into the output symbols $\{Z\}$ by the noise matrix. The figure shows two regions of presymbol space which should be mapped into one or two spikes, respectively, in the absence of noise. In the presence of noise, stimuli that are mapped into the core of these regions will be reliably encoded into the same number of spikes, while stimuli mapped into the boundary regions as marked will be readily corrupted, being transmitted sometimes as one spike and sometimes as two. See the text for discussion.

near the boundary between two presymbols, noise will be much more likely to change the resulting output symbol than if the stimulus is far from the boundary. For example, figure 4(B) shows regions of stimulus space corresponding to one- and two-spike outputs. Inputs that lie in the 'core' (unshaded) of each region will be mapped more reliably to output burst sizes, while inputs that lie in the 'boundary' (shaded) regions will be noisier. As long as the outputs of a neural circuit are discrete, there will exist boundary regions in the input space that are more susceptible to noise.

3.4. Defining the noise matrix

The noise matrix in the model, $Q_{k|j}$, describes the transition probability between presymbols y_j and observed burst sizes z_k. The presymbols y_j are idealized, noise-free versions of the circuit's desired output z_k. As noise in the retina originates already at the level of photon absorption by the photoreceptors, there is actually no place in the retinal circuit where one can attempt to measure y_j. However, the mapping from $X \to Y$ in figure 3 is deterministic, so we can associate a unique value of y_j with each firing event. Because every single observation of a firing event is corrupted by noise, we must observe many instances of every firing event to determine y. Our task is to use the distribution of burst sizes for each firing event to infer that event's presymbol.

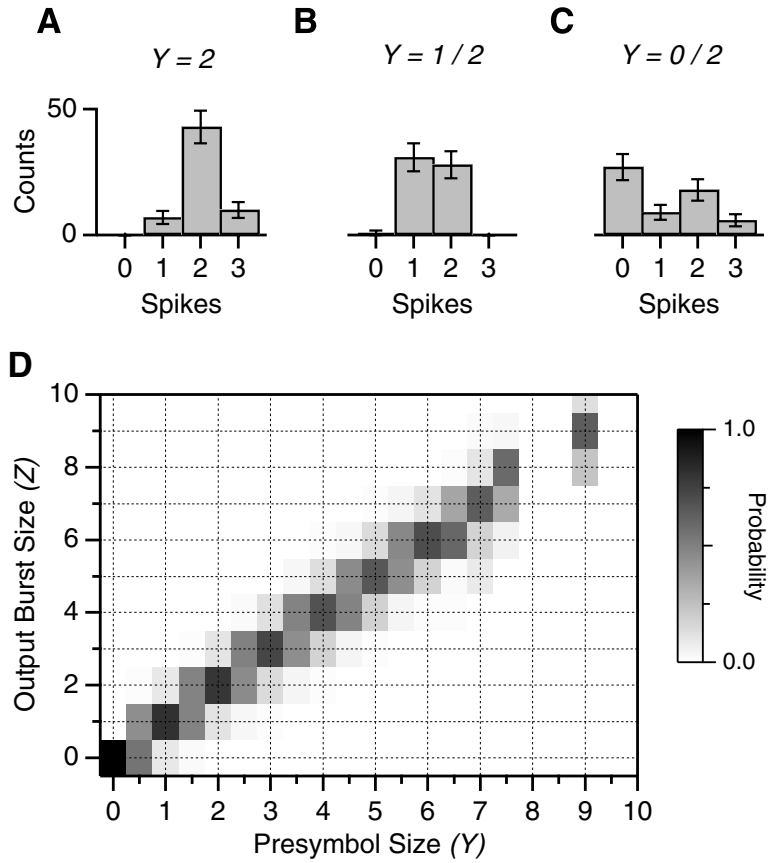

Figure 5. (A)–(C) Distribution of spike counts for three representative firing events of types 2, 1/2 and 0/2. (D) Depiction of the noise matrix for a ganglion cell with the transition probability $Q_{z|y}$ for a presymbol y to emerge as an output symbol z plotted in greyscale. For this cell, $z_{\max} = 5$.

As the variability in spike number was very low, most firing events had a narrow, unimodal distribution of burst sizes. Some of these distributions had their peak at a single spike count, while others had a peak shared between two successive burst sizes. The former case corresponds to a core or integer presymbol, with a value equivalent to the most common burst size (see figure 5(A)). The latter case corresponds to a boundary or half-integer presymbol, split between two successive burst sizes (see figure 5(B)). A small percentage of events had bimodal burst size distributions (see figure 5(C)) or had peaks spread among three or more burst sizes. These events were not well characterized by either integer or half-integer presymbols, but they occurred so infrequently ($<0.2\%$ of all events) that their awkward classification should not effect our results appreciably. The algorithm used to classify presymbols is described in the materials and methods section. Figure 5(D) shows the typical structure of a noise matrix. The block on the jth row and kth column represents the value of $Q_{k|j}$ in the shading depth—darker colours represent larger entries. Note that integer presymbols are mostly mapped onto single output burst sizes, while half-integer presymbols are mostly mapped onto two burst sizes. This structure is simply a reflection of the relatively low noise in our measured spike trains.

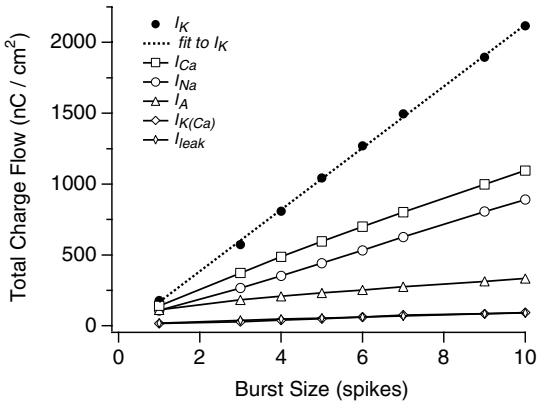

Figure 6. Total charge flow in the FM model plotted as a function of the number of spikes in a 50 ms period. For the delayed rectifier potassium current, a linear curve fit is shown as a dashed line.

3.5. Energetics of spiking

In addition to the noise matrix, the second ingredient we need in order to compare the theory of energy-efficient codes to experimental burst-size distributions is the energy cost of each neural symbol. A significant component of the total energy cost of a neuron results from its electrical signalling (Siesjo 1978, Laughlin *et al* 1998, Atwell and Laughlin 2001). In particular, action potentials are accompanied by large ionic currents that must then be reversed by pumps at a cost in energy. For instance, Na^+/K^+ ATP-ase, the most common pump in the nervous system, uses one molecule of ATP to pump 3 Na^+ out of the cell and 2 K^+ in to the cell. By estimating the total charge flow during a burst, we can consequently estimate the energy cost of that burst.

We studied a single compartment model of the tiger salamander ganglion cell developed by FM (see materials and methods). Their model used six ionic currents with conductance parameters set to closely match the cell membrane voltage measured during spikes and bursts (Fohlmeister and Miller 1997). We excited the FM model with current pulses of increasing duration and amplitude to produce bursts of action potentials. The total charge flow in each of the channel types was computed for a variety of burst sizes by integrating the ionic current. The results are displayed in figure 6, along with a linear fit to the data from the delayed-rectified K^+ current.

Clearly the charge flow, and hence the ATP consumption, is linear with burst size in the FM model and insensitive to the time intervals between spikes in a burst. What is more, the total charge flow associated with action potentials greatly exceeds the resting charge leakage through the cell membrane. These results held for all the amplitudes and durations of injected current that were studied. While there are certainly other costs and constraints relevant to biological information processing (see the discussion), this shows that the energy cost of producing bursts in real ganglion cells is likely to be both large and linear in burst size. Therefore, it is reasonable to assume such a linear energy cost model when assessing the energy efficiency of the retinal code.

3.6. Theory meets experiment

Accordingly, we assume that a burst of size n has a cost linear in n : $E_n = E_0 + E_s n$. Here, E_0 is interpreted as the cost of 50 ms of silence, and E_s is the added cost of a single spike.

Since codes that maximize bits per energy are invariant under rescalings of the cost model: $E_n \rightarrow \lambda E_n$ (Balasubramanian *et al* 2001), we can rescale the costs by $1/E_0$ and arrive at

$$E_n = 1 + bn \tag{9}$$

as a one-parameter model of the cost of a burst of size n. The energy slope $b = E_s/E_0$ is the ratio of the cost of producing a spike to the cost of 50 ms of silence. Given the typical $\sim$5 ms refractory period between action potentials, a slope of $b = 1$ would imply that an action potential is of order 10 times as expensive as an equal duration of silence.

With the measured noise and the above linear cost, we used the algorithm presented in the materials and methods section to compute the energy-efficient code for different energy slope parameters b. The number of zero symbols $N(z_0)$ depends on the time bin used to discretize silence. Because our choice of a 50 ms time bin is arbitrary, we left the zero symbol out of the burst size distribution that we attempted to fit with our theory. Furthermore, large bursts were too infrequent to use in our analysis, so we truncated the distribution at a largest burst size $z_{\max}$ (see materials and methods). Therefore, after computing the optimal distribution over all output symbols in our model $(0, 1, 2, \ldots)$, we dropped the 0 as well as bursts with sizes greater than $z_{\max}$. We then renormalized the theoretical optimal distribution over finite sized bursts $(1, 2, \ldots, z_{\max})$. This enabled direct comparison with the experimentally measured distribution of burst sizes. For each cell in our data set we performed fits of the theory to the experimentally measured distribution by varying the energy slope b and minimizing χ^2.

3.6.1. Qualitative agreement. Figure 7 shows representative fits for four cells stimulated by spatially uniform flicker and displays excellent qualitative agreement. For all cells, the sharp falloff of large bursts was accurately reproduced by the theory. Most cells had some suppression of one-spike bursts. Interestingly, these cells also had higher noise in the one-spike bursts, which were consequently suppressed in the theoretical efficient code. Some cells showed a marked dip in the use of small bursts (e.g. figure 7(B)). As shown, such dips were also reproduced by the theory, again because of the less frequent use of noisy symbols in an efficient code.

3.6.2. Single stimulus condition. Figure 8(A) shows a histogram of the goodness of fit, expressed as χ^2 per degree of freedom (see equation (2)), found for ganglion cells stimulated by spatially uniform flicker. Whenever we refer to χ^2 in the text it should be understood as χ^2 per degree of freedom. Low χ^2 values are obtained for all cells. The higher values typically occur when the fit fails to match at one point leading to a large χ^2 contribution. For instance, note in figure 7(C) that $\chi^2 = 7.2$ due to a large mismatch in the five-spike burst, even though the fit of theory to experiment looks quite good to the eye. Figure 8(B) shows a histogram of the energy slope b values obtained from fits to the same cells. Most values fall around $b = 0.5$–3.

3.6.3. Multiple stimulus conditions. We measured burst size distributions and noise under random flicker stimulation with different checker sizes, ranging from 134 μm (smaller than the typical receptive field centre) to 1070 μm. Previous work (Smirnakis *et al* 1997) has shown that the retina adapts its function when the spatial scale of the visual stimulus changes. We found that both the burst size distributions and their noise were different in each stimulus condition. However, we might expect that the ratio of the energy cost of spiking to silence should remain roughly constant. Therefore, as a further test of the hypothesis of metabolic efficiency we performed fits of the theory to the experimentally measured burst distribution under two or more stimulus conditions using a single value of the energy slope b.

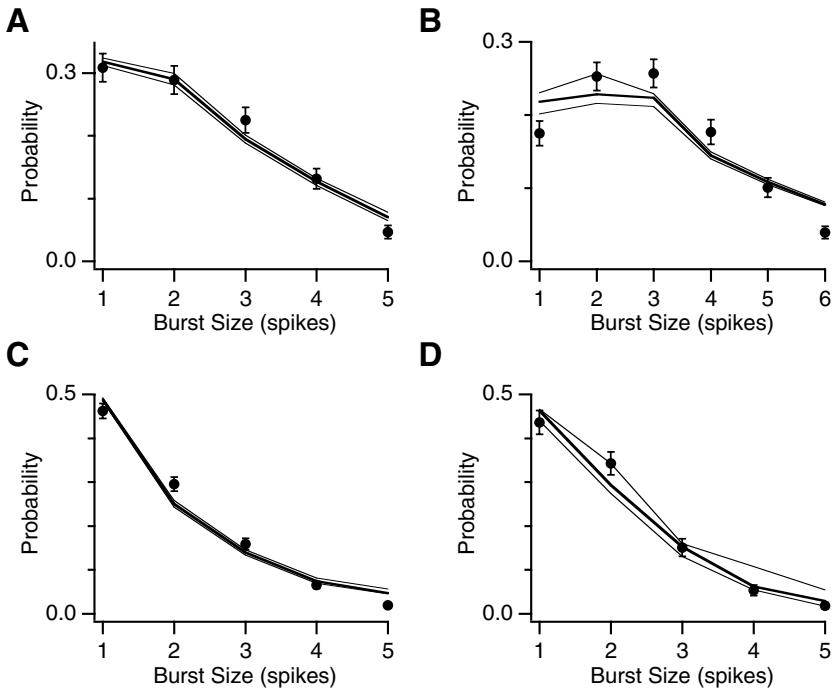

Figure 7. Curve fits of an optimally energy efficient code to the experimental burst size distribution for four cells made by finding the best value of the energy cost parameter b. Dots with error bars are the experimental data; the heavy central line is the fit. The thin upper and lower lines are numerically estimated upper and lower ranges on the theoretical fit arising from propagation of the error bars in the noise matrix through the theory (see the materials and methods section). Values of (χ^2, b) for cells in (A)–(D) are (1.6, 1.1), (5.1, 0.6), (7.2, 1.0) and (1.6, 2.6), respectively.

Figure 8(C) shows a histogram (black) of the goodness of fit for ganglion cells measured under two different stimulus conditions. Most values of χ^2 per degree of freedom fell near one, again indicating good quantitative agreement. Figure 8(D) shows a histogram (black) of energy slope values derived from these fits, with most values clustered around $b = 0.5$–1. We also measured five ganglion cells under more than two stimulus conditions. Fits to their burst size distributions yielded χ^2 values similar to the cells measured under only two stimulus conditions (figure 8(C), grey). Interestingly, their energy slope values were even more tightly clustered around $b = 0.7$ (figure 8(D), grey).

Of course, separate fits to each of the different stimulus conditions will lead to somewhat different optimal values of the energy slope b in each case. However, the error measure, χ^2, typically varies slowly with b, and as a result there is a single choice of b that gives good χ^2 for all the stimulus conditions. In this way the hypothesis of energy efficiency passes our multiple stimulus condition test.

The effects of noise and energy. The good quantitative agreement between an optimally energy efficient code and the actual retinal code requires that both the energy costs and noise of neural symbols be taken into account. To demonstrate this point, we compared our model to two others: one which neglects the noise of transmission through a channel, and another which includes the empirical noise, but assumes all neural symbols have the same energy

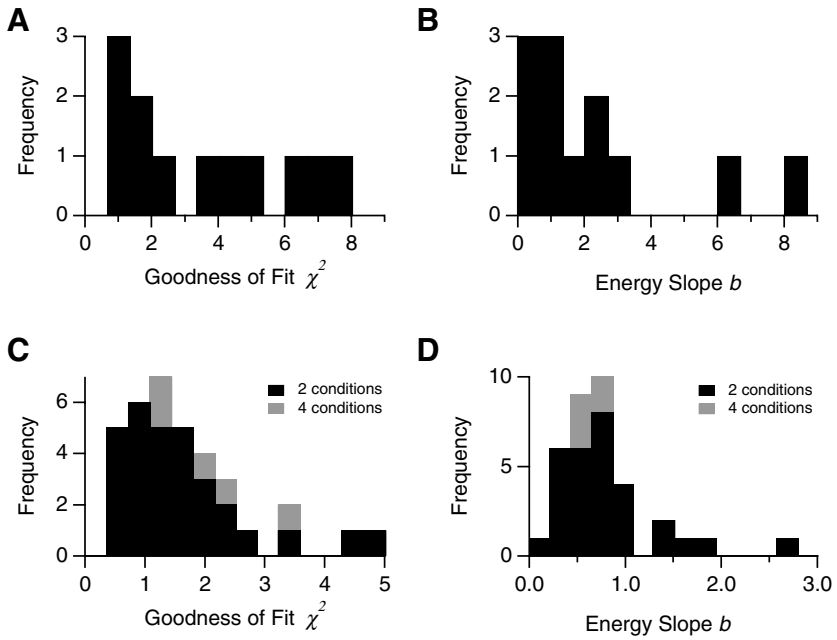

Figure 8. (A) Histogram of goodness-of-fit values, χ^2, and (B) best energy slope values, b, for ganglion cells stimulated by spatially uniform flicker. (C) Histogram of goodness-of-fit values, χ^2, and (D) best energy slope values, b, for ganglion cells stimulated by spatially modulated flicker. For spatially modulated flicker, curve fits used the same energy slope b for stimuli from two spatial scales (shown in black) or four spatial scales (shown in grey). The values derived from four spatial scales (grey) are appended on top of those from two spatial scales (black) for clarity.

cost. As shown in figure 9(A), models neglecting the cost of transmission (dashed curve) fail to capture the rarity of large bursts, while models neglecting noise (dotted curve) miss the suppression of noisy small bursts. In contrast, the full model (solid curve) captures both the suppression of large, expensive bursts and small noisy ones as seen in the data (dots with error bars). Figure 9(B) compares χ^2 of the fits to the data using the full and partial models. The dark and open circles represent models that neglect noise and energy costs, respectively. It is clear that a model incorporating both factors is a significant quantitative improvement.

To better demonstrate the connection between a symbol's noise and its frequency in the retinal code, we carried out the following procedure. First, we took the energy slope b for which the optimal energy-efficient code best matched the data, as described above. Then, we computed the distribution of symbols for an optimal code having the same energy slope b, but no noise (see equation (1)). To express whether a given symbol is used more or less frequently than expected by such an optimal, noiseless code, we defined the 'relative probability' as the ratio between the measured burst size probability and that for the optimal noiseless code. Figure 10(A) illustrates the calculation of relative probability for one ganglion cell. Now, recall that the probability that a presymbol remains uncorrupted by noise is given by the 'diagonal' noise matrix element $Q_{k|k}$. Then, $\eta_k \equiv 1 - Q_{k|k}$ is a measure of the noisiness of the neural symbol z_k. We calculated the ratio of each symbol's noise level η_k to the average of the noise levels of all the symbols for a given cell. This quantity, which we call the relative noise, expresses whether a given symbol is more or less noisy than the cell's other symbols. Figure 10(B) plots the relative probability versus the relative noise for a set of burst sizes from

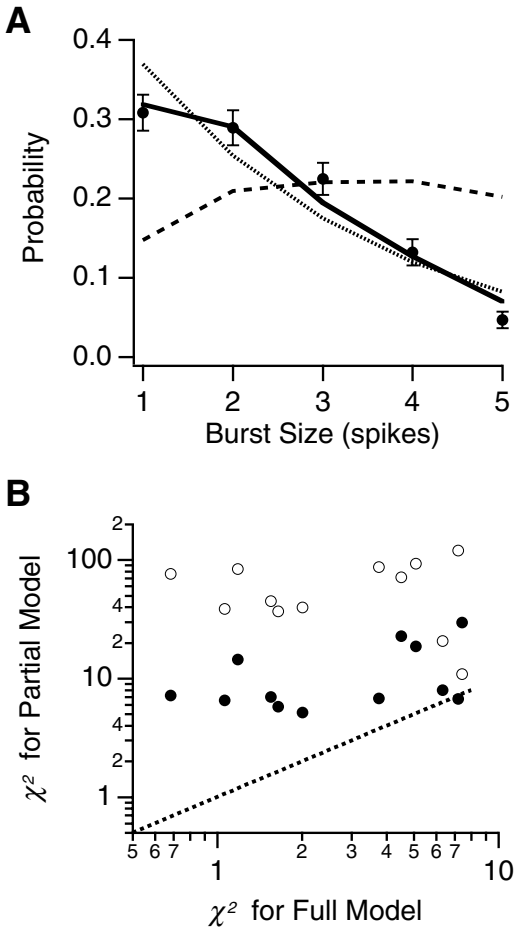

Figure 9. (A) Experimental burst size distribution (circles) for a ganglion cell shown against the full optimal model (dark curve) as well as models without noise (dotted curve) and without energy (dashed line). The full model is seen to match the data significantly better than either partial model. Respective χ^2 values for this cell are 1.6 (full model), 7.0 (no noise) and 45 (no energy). (B) The χ^2 deviation between the burst size distribution of model and data for partial models without noise (dark circles) and without energy (open circles) plotted against χ^2 for the full model. The dashed line shows the case where the partial and full models have the same error.

ganglion cells stimulated by spatially uniform flicker. A strong correlation between lower noise and higher probability is seen. This effect is much stronger than proportionality, i.e. a $\sim$10% decrease in noise results in a $\sim$40% decrease in its usage frequency for most cells (see dashed line). Thus, the decreased use of noisy symbols predicted by information theory is reproduced in the experimental data.

4. Discussion

In summary, we have shown that one aspect of the neural code of retinal ganglion cells, the distribution of burst sizes used to represent a broad stimulus ensemble, is consistent with the principle of metabolic efficiency. Bursts that are noisy or energetically costly are used less

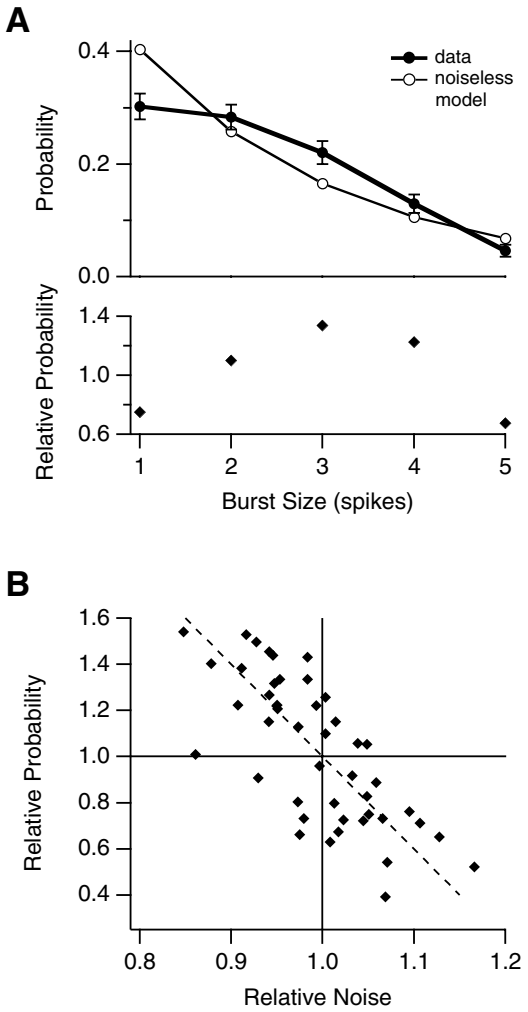

Figure 10. (A) Illustration of the effects of noise. The top panel plots the burst size distribution (solid circles) for a ganglion cell along with the burst size distribution for an optimal code with the same energy slope b but no noise (open circles). The bottom panel shows the relative probability (diamonds), which is the ratio between the probability of each burst size in the data to the probability for the optimal, noiseless model. (B) The relative probability of burst sizes plotted versus their relative noise for ganglion cells stimulated by spatially uniform flicker (diamonds). Each symbol represents one burst size for one cell. The dashed line with a slope of -4 is drawn as a guide to the eye.

frequently, as is found for an energy efficient code. By measuring the noise in each firing event and estimating the energy cost in the soma up to one free parameter, we compared the actual burst size distributions of ganglion cells to the distribution that optimizes the amount of visual information transmitted per unit energy. Good agreement was found, even across multiple stimulus conditions. Particularly interesting was the reduced frequency of noisy events relative to more reliable firing events, as such neural symbols convey less visual information and hence should be suppressed in an energy efficient code. These results lend weight to the hypothesis that the structure of the retinal code is determined in part by the need for energy efficiency.

Other kinds of models

Our results do not rule out explanations of a ganglion cell's burst distribution other than energy efficiency. For instance, some of the qualitative features of our data, such as the rarity of large bursts, are consistent with a simple threshold-crossing mechanism, where large bursts require a very strong stimulus that occurs infrequently in the random flicker ensemble. Recently, Keat *et al* have shown that a model consisting of a linear filter followed by threshold crossing can account for many features of the spike trains elicited by spatially uniform flicker (Keat *et al* 2001). However, a mechanistic description of the retinal code does not exclude an optimization principle: regardless of whether any given optimization principle holds, there must always exist *some* mechanistic explanation.

An optimization principle is a different *kind* of description. Its value lies in explaining data with relatively few arbitrary parameters. In this respect, our model, with only one free parameter, compares favourably with a detailed mechanistic model having 21 free parameters (Keat *et al* 2001), although only four to six of these parameters of this model significantly affect the burst size distribution. In addition, an optimization principle is not just an algorithm for predicting aspects of a neural code; such a principle also provides a reason for why the code is structured as it is. In other words, it seeks to explain the choice of operating parameter values in a mechanistic description of the same phenomenon. Our results support the hypothesis that retinal circuitry is configured to transmit more visual information per unit energy cost than other possible configurations.

Relevance of energy costs

Why should retinal energy efficiency matter to an animal, considering that the retina comprises only a small fraction of most vertebrates' bodies? In fact, neurons are some of the most energy-intensive cells in the body. For instance, the human brain accounts for 20% of daily energy intake but only 2% of body mass (Rolf and Brown 1997). Electrical and chemical signalling has been estimated to account for up to 85% of a neuron's energy budget (Atwell and Laughlin 2001). These data suggest that the use of energy efficient neural codes could be quite important for an animal. Furthermore, the energy saved in the retina by an optimal code may underestimate the total savings for the animal. Retinal spike trains lead to cascades of synaptic currents and action potentials in the central brain, the costs of which might be reduced by using an efficient retinal code. In addition, the retina employs much of the same molecular machinery—from ion channels to biochemical signalling cascades—found throughout the brain. As a result, mutations in this machinery that allow for greater energy efficiency in the retina may simultaneously enable other parts of the brain to save energy.

One might also wonder why energy efficiency is important for ganglion cell spike trains, because photoreceptors are known to be metabolically expensive, and variations in the neural code considered in this paper do not affect this initial stage of the visual pathway. Ames (1992) has made a detailed analysis of energy costs in the rabbit retina. He finds that while the photoreceptors do cost more ATP to run, the ganglion cell layer uses anaerobic respiration, so that its cost in terms of glucose consumption is comparable to the photoreceptor layer. Since the ganglion cell layer does not receive a direct blood supply in either the rabbit or salamander retinas, it is likely that the ganglion cell layer of salamander uses anaerobic respiration, too. In addition, a simple estimate of the total membrane area of the optic nerve in the salamander indicates that its energy cost may equal the entire retina for reasonable firing rates, and this cost will be reduced by an efficient code.

Previous work on energy efficiency

Energy efficiency has also been explored for neurons in the primary visual cortex (area V1) of the cat and inferotemporal cortex (IT) of the macaque monkey by Baddeley *et al* (1997). They found that the distribution of spike counts was very nearly exponential over a large range of window sizes. As an exponential spike count distribution maximizes the entropy of the spike train at fixed average firing rate (Levy and Baxter 1996, Baddeley 1996, Balasubramanian *et al* 2001), they interpreted this as evidence that the code of cortical neurons is optimized for energy efficiency. This result held for a variety of stimulus conditions including natural scenes, white noise and spontaneous firing. However, Baddeley *et al* (1997) did not include the effects of noise in their analysis. As shown in Balasubramanian *et al* (2001), an exponential distribution of spike counts only optimizes information at fixed energy when the noise is zero (see also figure 3(B)). Indeed, noise can dramatically modify the energy-efficient spike count distribution, for example by heavily suppressing the use of inexpensive but noisy signalling patterns.

A subsequent study of IT neurons in the monkey by the same group (Treves *et al* 1999) showed that only a minority of recorded cells (14% during natural viewing and 25% during a fixation task) had exponential spike count distributions. Many of the cells showed a strong suppression of the smallest spike counts, as is also seen for retinal ganglion cells (figure 2). De Polavieja (2002) analysed this data using the method of Balasubramanian *et al* (2001) to incorporate the effects of noise on the distribution of firing rates that optimized information at fixed energy. Here, noise was modelled phenomenologically with a single fit parameter. Noise suppressed the usage of low firing rates, giving good agreement with the experimental data.

Our analysis improves on previous efforts in two ways. First, we divided the spike train into firing events. This serves to minimize the correlations between successive coding symbols, so that their information content is nearly independent. In contrast, when the spike train is arbitrarily divided into fixed time windows, short-range correlations due to the refractory period and the tendency to burst are ignored. This typically leads to an overestimate of the information contained in each symbol as well as an overestimate of their noise. Second, we directly measured the neural noise, so that its effects can be incorporated without fit parameters. Direct measurement of the noise is important for retinal ganglion cells, because individual cells display a great variety of noise characteristics. This variety was found to be very well matched by the usage of burst sizes (see figure 10).

A drawback of measuring the noise is that it requires many repeats of the same stimulus. As a consequence, it is much harder to sample the ensemble of stimuli in the finite duration of an electrophysiological recording. This problem is particularly acute for natural stimuli, as the variety of this ensemble is huge and not well characterized. In this regard, the approach of focusing only on the entropy can make immediate progress for visual neurons under more natural stimulation (Baddeley *et al* 1997). For our analysis to be extended to the case of natural stimulation, it will be necessary to simplify the description of a neurons noise, such that fewer stimulus repeats are needed to measure it.

Future directions

Our analysis would benefit from several extensions. The first is developing a model that explicitly represents time. Because our model treats the spike train merely as an ordered list of firing events, it leaves out all of the information contained in the exact time at which each event occurs. In addition, we were forced to use an arbitrary time bin, 50 ms, to identify periods of silence as a zero-spike burst, which prevented us from predicting the overall firing rate of a ganglion cell. Other important extensions include testing energy efficiency for

different functional types of ganglion cells, different species, and more stimulus conditions, to assess the generality of our result. Finally, our observation of the reduced frequency of noisy events relative to more reliable firing events brings up an interesting question of mechanism: Can ganglion cells detect and adapt to their own noise dynamically, or alternatively are their operating conditions fixed in such a way as to be well matched to their noise characteristics?

It is worth remembering that the optimization principle used in this paper need not be interpreted as energy efficiency. Strictly speaking, we have only shown that the distribution of ganglion cell burst sizes maximizes information transmission subject to a cost that is proportional to its number of spikes. It is very intuitive to associate this cost function with metabolic cost. However, a similar distribution of burst sizes results from the constraint of maximizing information transmitted in a fixed window of time (de Polavieja 2002). Other intuitive constraints on neural signalling might also yield a similar cost function. Further exploration is necessary to distinguish between these possibilities. In particular, it is essential to tie the single parameter in our model to the detailed biophysics of retinal information transmission.

Acknowledgments

VB was supported initially by the Society of Fellows and the Milton Fund of Harvard University and later by DOE grant DOE-FG02-95ER40893. MB was supported by the Pew Charitable Trust and the E Mathilda Ziegler Foundation. VB is grateful to the ITP, Santa Barbara and the Aspen Centre for Physics for hospitality while this work was in progress. We are grateful to Don Kimber and Markus Meister for collaboration on previous work directly related to this article.

References

Ames A 1992 3rd energy requirements of CNS cells as related to their function and to their vulnerability to ischemia: a commentary based on retina *Can. J. Physiol. Pharmacol.* **70** S158–64

Ames A 2000 3rd CNS energy metabolism as related to function *Brain Res. Rev.* **34** 42–68

Arimoto S 1972 An algorithm for computing the capacity of an arbitrary discrete memoryless channel *IEEE Trans. Inf. Theory* **18** 14–20

Atick J J 1992 Could information theory provide an ecological theory of sensory processing? *Network: Comput. Neural Syst.* **3** 213–51

Atwell D and Laughlin S B 2001 An energy budget for signalling the grey matter of the brain *J. Cereb. Blood Flow Metab.* **21** 1133–45

Baddeley R J 1996 An efficient code in V1? *Nature* **381** 560–1

Baddeley R, Abbott L F, Booth M C A, Sengpiel F, Freeman T, Wakeman E A and Rolls E T 1997 Responses of neurons in primary and inferior temporal visual cortices to natural scenes *Proc. R. Soc.* B **264** 1775–83

Balasubramanian V, Kimber D and Berry M J 2001 Metabolically efficient information processing *Neural Comput.* **13** 799–816

Barlow H B 1961 Possible principles underlying the transformation of sensory messages *Sensory Communication* ed W A Rosenblith (Cambridge, MA: MIT Press) pp 217–34

Barlow H B 2001 Redundancy reduction revisited *Network: Comput. Neural Syst.* **12** 241–53

Berry M J II, Warland D W and Meister M 1997 The structure and precision of retinal spike trains *Proc. Natl Acad. Sci. USA* **94** 5411–16

Berry M J II and Meister M 1998 Refractoriness and neural precision *J. Neurosci.* **18** 2200–11

Blahut R E 1972 Computation of channel capacity and rate distortion functions *IEEE Trans. Inf. Theory* **18** 460–73

Blahut RE 1987 *Principles and Practice of Information Theory* (Reading, MA: Addison-Wesley)

Buracas G T, Zador A M, deWeese M R and Albright T D 1998 Efficient discrimination of temporal patterns by motion-sensitive neurons in the primate visual cortex *Neuron* **20** 959–69

de Polavieja G G 2002 Errors drive the evolution of biological signalling to costly codes *J. Theor. Biol.* **214** 657–64

de Ruyter van Steveninck R R, Lewen G D, Strong S P, Koberle R and Bialek W 1997 Reproducibility and variability in neural spike trains *Science* **275** 1805–8

Fanselow E E and Nicolelis M A L 1999 Behavioral modulation of tactile responses in the rat somatosensory system *J. Neurosci.* **19** 7603–16

Fohlmeister J F and Miller R F 1997 Impulse encoding mechanisms of ganglion cells in the tiger salamander retina *J. Neurophysiol.* **78** 1935–47

Gillespie J H 1998 *Population Genetics: A Concise Guide* (Baltimore, MD: Johns Hopkins University Press)

Kara P, Reinagel P and Reid R C 2000 Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons *Neuron* **27** 635–46

Keat J, Reinagel P, Reid R C and Meister M 2001 Predicting every spike: a model for the responses of visual neurons *Neuron* **30** 803–17

Laughlin S B, de Ruyter van Steveninck R and Anderson J C 1998 The metabolic cost of neural information *Nature Neurosci.* **1** 36–41

Levy W B and Baxter R A 1996 Energy-efficient neural codes *Neural Comput.* **8** 531–43

McIlwain H and Bachelard H S 1985 *Biochemistry and the Central Nervous System* 5th edn (London: Churchill Livingstone)

Meister M and Berry M J II 1999 The neural code of the retina *Neuron* **22** 435–50 (review)

Meister M, Pine J and Baylor D A 1994 Multi-neuronal signals from the retina: acquisition and analysis *J. Neurosci. Methods* **51** 95–106

Reinagel P and Reid R C 2000 Temporal coding of visual information in the thalamus *J. Neurosci.* **20** 5392–400

Rolf D F S and Brown G C 1997 Cellular energy utilization and molecular origin of standard metabolic rate in mammals *Physiol. Rev.* **77** 731–58

Roth G 1987 *Visual Behavior in Salamanders* (Berlin: Springer)

Schoenbaum G, Chiba A A and Gallagher M 1999 Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning *J. Neurosci.* **19** 1876–84

Schreiber S, Machens C K, Herz A V M and Laughlin S B 2002 Energy efficient coding with discrete stochstic events *Neural Comput.* **14** 1323–46

Siesjo B 1978 *Brain Energy Metabolism* (New York: Wiley)

Smirnakis S M, Berry M J, Warland D K, Bialek W and Meister M 1997 Adaptation of retinal processing to image contrast and spatial scale *Nature* **386** 79–83

Stanley G B, Li F F and Dan Y 1999 Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus *J. Neurosci.* **19** 8036–42

Strong S P, Koberle R, de Ruyter can Steveninck R R and Bialek W 1997 Entropy and information in neural spike trains *Phys. Rev. Lett.* **80** 197–200

Treves A, Panzeri S, Rolls E T, Booth M and Wakeman E A 1999 Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimulation *Neural Comput.* **11** 601–31

van der Schaaf A A and van Hateren J H 1996 Modelling the power spectra of natural images-statistics and information *Vis. Res.* **36** 2759–70

van Hateren J H 1992 Real and optimal neural images in early vision *Nature* **360** 68–70

Wong-Riley M T T 1989 Cytochrome oxidase: an endogenous metabolic marker for neuronal activity *Trends Neurosci.* **12** 94–101