

Decoding Stimulus Variance from a Distributional Neural Code of Interspike Intervals

Brian Nils Lundstrom and Adrienne L. Fairhall

Department of Physiology and Biophysics, University of Washington, Seattle, Washington 98195

The spiking output of an individual neuron can represent information about the stimulus via mean rate, absolute spike time, and the time intervals between spikes. Here we discuss a distinct form of information representation, the local distribution of spike intervals, and show that the time-varying distribution of interspike intervals (ISIs) can represent parameters of the statistical context of stimuli. For many sensory neural systems the mapping between the stimulus input and spiking output is not fixed but, rather, depends on the statistical properties of the stimulus, potentially leading to ambiguity. We have shown previously that for the adaptive neural code of the fly H1, a motion-sensitive neuron in the fly visual system, information about the overall variance of the signal is obtainable from the ISI distribution. We now demonstrate the decoding of information about variance and show that a distributional code of ISIs can resolve ambiguities introduced by slow spike frequency adaptation. We examine the precision of this distributional code for the representation of stimulus variance in the H1 neuron as well as in the Hodgkin–Huxley model neuron. We find that the accuracy of the decoding depends on the shapes of the ISI distributions and the speed with which they adapt to new stimulus variances.

Key words: adaptation; information theory; neural coding; noise; invertebrate; spike patterns; computation

Introduction

What constitutes “the neural code”? Spike trains from single neurons may transmit information via the number of spikes in some time window, the absolute timing of spikes, and spike patterns. Spike rate and spike timing can be thought of as a continuum, with information represented by spike count at different time scales in different organisms and brain areas (Bialek et al., 1991; Mainen and Sejnowski, 1995; Reinagel and Reid, 2002). Information encoding via spike patterns and interspike intervals (ISIs), however, relies on relationships between spikes. In the identified neuron H1 of the fly visual system the meaning of particular sequences of spike intervals has been determined by using reverse correlation (de Ruyter van Steveninck and Bialek, 1988; Brenner et al., 2000). These multiple spike symbols permit stimulus reconstructions that increase in fidelity with the increasing number of spikes per symbol (Rieke et al., 1997). The meaning of spike patterns such as bursts has been examined in the retina (Meister and Berry, 1999), primate visual cortex (Reich et al., 2000), the lateral geniculate nucleus (Lesica and Stanley, 2004), and conductance-based models (Kepecs and Lisman, 2003). In the early auditory system (Cariani, 1999) single spikes are phase-locked to oscillations of the incoming sound wave. Although spe-

cific cycles of the wave may fail to evoke a spike, the distribution of intervals across the cell population shows clear peaks at the fundamental wave length and its harmonics and thus represents pitch. In this paper we take a different approach: we consider the distribution of intervals generated in time by a single neuron.

Stimulus decoding has been analyzed extensively in the neural code, usually when the stimulus distribution does not change with time (Bialek et al., 1991; Haag and Borst, 1997; Warland et al., 1997; Strong et al., 1998; Egelhaaf and Warzecha, 1999; Stanley et al., 1999; Brenner et al., 2000; Keat et al., 2001). Typically, however, statistics of the sensory environment change in time, and for many sensory systems the coding strategy of the system also changes. In the retina (Shapley and Victor, 1979; Kim and Rieke, 2001; Baccus and Meister, 2002), visual cortex (Carandini and Ferster, 1997), the auditory system (Kvale and Schreiner, 2004; Dean et al., 2005), and somatosensory cortex (M. Maravall, R. S. Petersen, A. L. Fairhall, E. Arabzadeh, and M. E. Diamond, unpublished observations) a changing stimulus variance results in contrast adaptation. During contrast adaptation the relationship between the input stimulus and the output firing probability may change to match the new dynamic range of the stimulus. In the case of H1 this takes a particularly elegant form: the input/output functions have the same shape, but the stimulus is scaled by its SD (Brenner et al., 2000; Fairhall et al., 2001). This rescaling potentially renders the code ambiguous with respect to stimulus variance. Although this ambiguity may be resolved by the spike rate, which depends on the variance envelope, spike rate also adapts, with spike frequency adaptation occurring over many seconds.

We claim that it is possible to decode stimulus statistics rapidly by monitoring the local statistics of spike intervals (Fairhall et al., 2001). Our goal here is to examine the speed, accuracy, and precision with which this decoding strategy can represent a time-

Received Jan. 17, 2006; revised June 29, 2006; accepted July 27, 2006.

This research was funded in part by Nippon Electric Company (NEC) Research Institute and by a Burroughs Wellcome Careers at the Scientific Interface Grant to A.L.F.; B.N.L. also was supported by the Medical Scientist Training Program, a fellowship from the National Institute of General Medical Sciences (T32 07266), as well as an Achievement Rewards for College Scientists Fellowship. We are grateful to Geoff Lewen and Robert de Ruyter van Steveninck for assistance with the experiments and to William Bialek, Michael Berry, Blaise Aguera y Arcas, Matt Higgins, and William Spain for discussions.

Correspondence should be addressed to Adrienne Fairhall, Department of Physiology and Biophysics, University of Washington, Health Sciences Building G424, Box 357290, Seattle, WA 98195. E-mail: fairhall@u.washington.edu.

DOI:10.1523/JNEUROSCI.0225-06.2006

Copyright © 2006 Society for Neuroscience 0270-6474/06/269030-08\$15.00/0

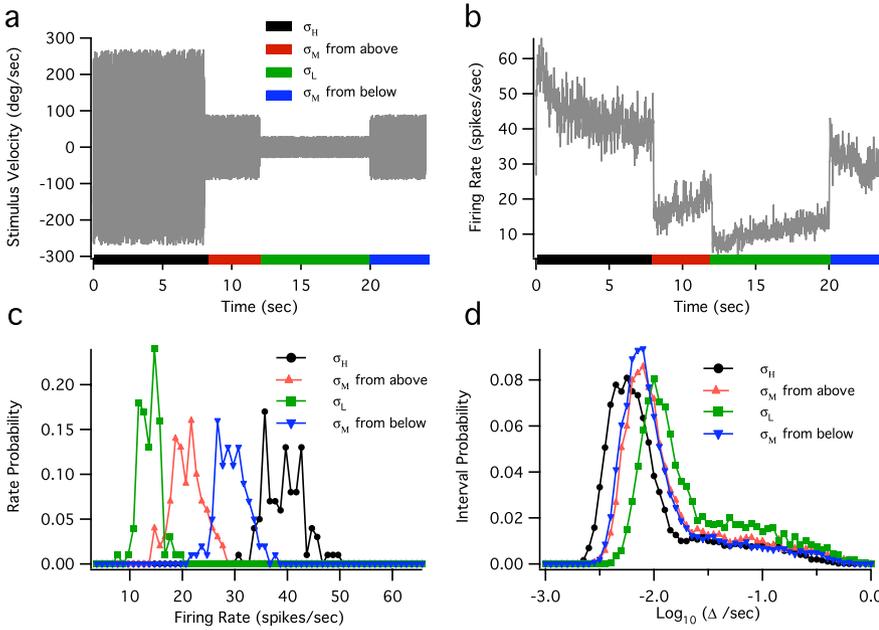


Figure 1. Ambiguity in the rate with respect to the stimulus SD is reduced by considering spike intervals. **a**, A schematic of the velocity stimulus to the fly. This three-state 24 s stimulus was repeated for 2 h. Stimulus SDs were $\sigma_H = 3\sigma_M = 9\sigma_L$. **b**, Mean spike rate for the three-state switching experiment with bin size of 10 ms. **c**, Rate distributions $P(r|\sigma_j)$ from the last second of each epoch. **d**, ISI distributions $P(\Delta|\sigma_j)$ from the last second of each epoch. Note that stimuli with σ_M lead to two distinct rate distributions (**c**, triangles) but very similar ISI distributions (**d**, triangles).

varying variance. We consider distributional coding of visual stimulus motion in the fly-identified neuron H1 and, to demonstrate its generality, in a Hodgkin–Huxley (HH) model neuron.

Materials and Methods

Fly data. Spike times were recorded from the H1 neuron (Franceschini et al., 1989) in the *Calliphora vicina* fly (Fairhall et al., 2001). Action potentials were recorded extracellularly as the fly viewed a horizontally moving vertical bar pattern. Bars moved across an oscilloscope screen at a standardized distance from the animal with random velocity, the velocities of which were chosen from a uniform distribution with unit variance ($\sigma^2 = \sigma = 1$) multiplied by some constant. Stimulus frames were drawn every 2 ms, with a new velocity chosen for each frame. In the three-state switching experiment the stimuli with three different SDs, $\sigma_H = 3\sigma_M = 9\sigma_L$, were presented in the sequence as follows: $\sigma_H \rightarrow \sigma_M \rightarrow \sigma_L \rightarrow \sigma_M$ (see Fig. 1a). Each SD was presented for the same duration per cycle; the sequences with SDs σ_H and σ_L were each presented in 8 s blocks, whereas the two presentations of σ_M were each 4 s. Data were recorded for 2 h or 300 consecutively presented 24 s trials. In steady-state experiments the constant variance stimuli, with SDs spanning a 60-fold range, were presented in random order to a single fly as described above for 15 min each over the course of a day.

Simulated data. A standard HH space-clamped model neuron was used to simulate changes in the neuronal membrane potential with respect to time (Hodgkin and Huxley, 1952). The equations were solved numerically by using fourth-order Runge–Kutta integration with a fixed time step of 0.05 ms. Injected current was simulated by a series of zero-mean normally distributed random numbers that were smoothed by a Gaussian filter (full-width half-maximum, 0.6 ms). Spike times were identified as the upward crossing of the voltage trace at 0 mV, with a resting potential of -65 mV. For steady-state data 10 1 h data runs were simulated, with normalized SDs spanning an eightfold range.

Estimating probability distributions. Normalized base 10 logarithmic distributions of ISIs were created for spike train data. Because spike intervals span a wide range, using a logarithmic instead of a linear scale allows for differences in short intervals to be noticeable; bin widths were $0.05 \log_{10} s$, with range $[-2.7 \text{ to } 0.2]$ for the fly and $[-2.1 \text{ to } 1.7]$ for the

model neuron. One was added systematically to all bins before the normalization of distributions that would be in the denominator of a ratio to avoid division by zero; when sample number $n \gg 1$, this is analogous to assuming a uniform previous distribution and then finding the mean of the posterior probability distribution by using Bayes' rule (cf. Gelman, 1995; Johnson et al., 2001). According to Bayes' rule, the previous distribution is multiplied by a likelihood and divided by the probability of all data to give the posterior distribution. If the previous distribution is a uniform distribution, the posterior distribution is then the two-parameter Beta distribution $\text{Beta}(\alpha, \beta)$, which in this case is $\text{Beta}(1 + x, 1 + n - x)$, where x is observed successes and n is the number of samples. The mean of any Beta distribution is $\alpha/(\alpha + \beta)$, thus yielding a mean of $(1 + x)/n$ when $n \gg 1$. Avoiding zero bins allows all log ratios of these distributions to be computed; in practice, the treatment of these values noticeably affects results when either the distributions are undersampled or the supports of the probability distributions are markedly different, such as is the case between the extreme curves of Figure 4b.

Discriminating between two distributions. Because our goal is to evaluate the information contained in the spike interval distributions, we use the Kullback–Leibler divergence, a statistical measure that is related to mutual information, to quantify the difference between two probability distributions as follows:

$$D_{KL}(P, Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}. \quad (1)$$

This measure is zero if and only if $P = Q$ and is otherwise > 0 . Unlike a true distance measure, the Kullback–Leibler divergence is not symmetric, $D_{KL}(P, Q) \neq D_{KL}(Q, P)$, and does not obey the triangle inequality. For our purposes this is advantageous because it allows us to quantify the potential asymmetry in observing changes in different directions, i.e., comparing low-to-high variance changes with high-to-low variance changes.

D_{KL} quantifies the intrinsic classification difficulty (Johnson et al., 2001). Although D_{KL} does not give the absolute probability of error for a specific decoder, it does limit the ultimate performance of any classifier; a one-bit increase in D_{KL} corresponds to a twofold decrease in error probability. Given a decoding model with some set of parameters, for small parameter changes the smallest achievable mean-squared estimation error is inversely proportional to D_{KL} ; this is the Cramer–Rao bound (Johnson et al., 2001).

Discrimination by using ISIs: binary choice. Given a sequence of ISIs generated in response to a statistically steady-state stimulus, we would like to quantify our ability to decode responses by classifying them according to whether they were caused by a stimulus with SD σ_A or σ_B .

We wish to compute the Kullback–Leibler divergence between the distributions of n -interval sequences $\Delta_1, \dots, \Delta_n$ generated by different stimulus SDs, $D_{cKL}^{(n)} = D_{KL}(P(\Delta_1, \dots, \Delta_n|\sigma_B) \| P(\Delta_1, \dots, \Delta_n|\sigma_A))$. Because in practice the multidimensional distributions $P(\Delta_1, \dots, \Delta_n|\sigma)$ are very difficult to sample for $n > 2$, we will assume that successive intervals are independent. In this case, $D_{cKL}^{(n)}$ is given by the cumulative D_{KL} as follows:

$$D_{cKL}^{(n)} = \sum_{i=1}^n D_{KL}(P(\Delta_i|\sigma_B) \| P(\Delta_i|\sigma_A)) \\ = \sum_{i=1}^n \sum_{\Delta_i} P(\Delta_i|\sigma_B) \log_2 \frac{P(\Delta_i|\sigma_B)}{P(\Delta_i|\sigma_A)}. \quad (2)$$

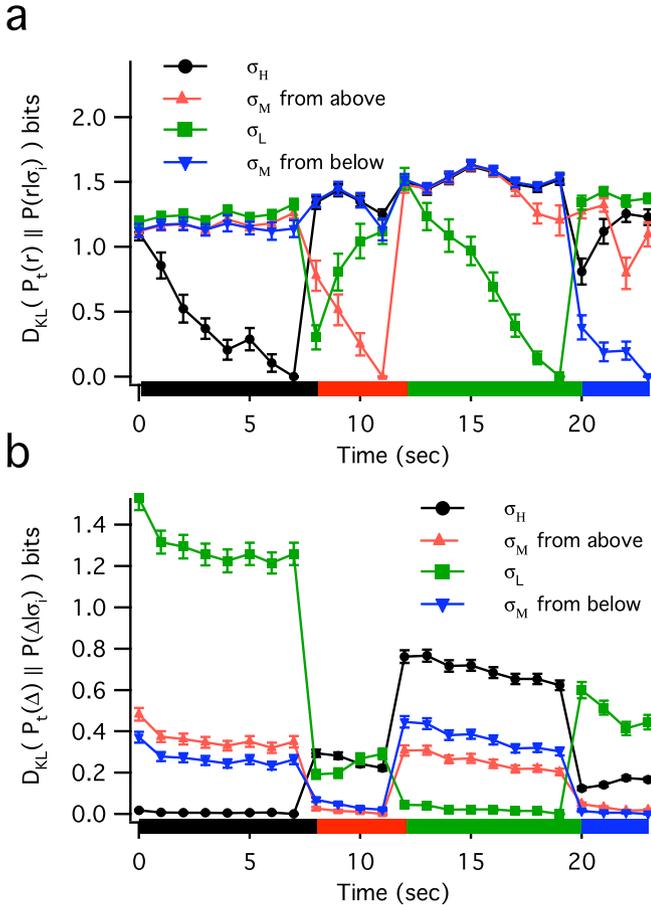


Figure 2. Kullback–Leibler divergence D_{KL} between the time-dependent distributions and the relevant reference distributions for rate $D_{KL}[P_t(r), P(r|\sigma)]$ (a) and intervals $D_{KL}[P_t(\Delta), P(\Delta|\sigma)]$ (b). Each distribution was created from 1 s (times 300 repeats) of this 24 s stimulus pattern. Reference distributions are sampled from the final second of each of the four epochs, hence $D_{KL} = 0$ at 7, 11, 19, and 23 s. Although rate distributions change continuously throughout the epoch (a), interval distributions quickly reach a steady state (b). Error bars represent the SD.

We estimated the distributions $P(\Delta_i|\sigma_A)$ and $P(\Delta_i|\sigma_B)$ from 10,000 unique random samples of interval sequences from the steady-state fly and model neuron data. To de-bias and estimate the variance of D_{CKL} , we applied the Bootstrap method (Press, 1992; Johnson et al., 2001) by drawing 10,000 sequences of n ISIs randomly with replacement from the initial set of 10,000 sequences. After repeating this procedure 200–300 times, we found the mean μ and the SD σ . Using the mean, we de-biased $D_{CKL}^{(n)}$ by subtracting $\mu - D_{CKL}^{(n)}$, and we took the SD as the error. Note that, for steady-state data, the following applies: $D_{CKL}^{(n)} = nD_{CKL}(P(\Delta|\sigma_B)||P(\Delta|\sigma_A))$. We note that, for well sampled steady-state data, calculating $D_{CKL}^{(n)}$ is equivalent to calculating the expectation of a decision variable D_n as follows:

$$D_n = \sum_{i=1}^n \log \frac{P(\Delta_i|\sigma_B)}{P(\Delta_i|\sigma_A)}. \quad (3)$$

The D_{CKL} values of Figure 4 are related to both the $D_{CKL}^{(n)}$ values, which assess the ability to decode from an n -interval sequence, as well as to a possible log-likelihood decision variable (Fairhall et al., 2001), which could be used in decoding. In Fairhall et al. (2001), the signal-to-noise ratio (SNR) of D_n was used to determine the reliability of decoding, where SNR is defined as the square of the mean over the variance of D_n .

Multiple choices: calculating the information that one interval gives about the stimulus variance. When there are more than two choices, we compute the mutual information between the set of intervals Δ and the stim-

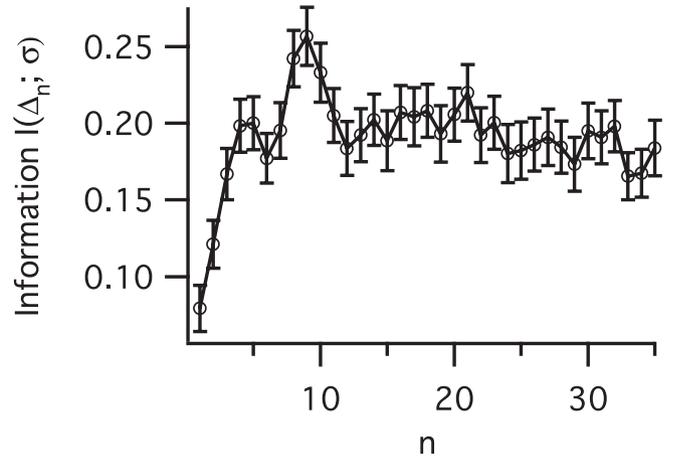


Figure 3. The information $I(\Delta_n; \sigma)$ gained from successive spikes after a switch between SDs as calculated by Equation 4. The probability $P(\Delta_n|\sigma_i)$ of a given interval value for the n th interval after an SD switch is calculated for each of the three σ values in Figure 1. After approximately three intervals, the information per spike reaches steady state, suggesting that interval coding quickly adapts to the new stimulus statistics. Error bars represent the SD.

ulus SDs σ_i . This quantifies the average reduction in uncertainty about the stimulus source that is gained by observing a single interval. The mutual information is defined as the difference between the total entropy $H[P(\Delta)]$ and the noise entropy H_s . In this application the noise entropy is the entropy of $P(\Delta|\sigma)$, the distribution of Δ conditioned on the SD σ , averaged over the set of possible values of σ as follows: $H_s = -\sum_{\Delta} P(\Delta|\sigma) \log_2 P(\Delta|\sigma)$ (Cover and Thomas, 1991; Rieke et al., 1997; Dayan and Abbott, 2001). The difference between the total entropy and the noise entropy can be rewritten as the following:

$$I(\Delta; \sigma) = \sum_{\Delta} \sum_{\sigma} P(\sigma) P(\Delta|\sigma) \log_2 \frac{P(\Delta|\sigma)}{P(\Delta)}. \quad (4)$$

This information can be calculated either for the steady-state distributions or for subsets of intervals with a particular temporal relationship to the stimulus envelope, such as for the n th interval after a change in stimulus variance as in Figure 3. SDs for each data point are found by using a Bootstrapping method as described above. All distributions are well sampled because $n \gg m$ ($n = [300\ 600\ 300]$, $m = 59$), where n is the sample number and m is the number of bins. Additionally, bias is negligible because $V/B^2 \approx [n(\log m)^2]/m^2$, where V and B are the variance and bias of our estimator, respectively (Paninski, 2003). Therefore, we evaluate Equation 4 directly.

Testing the assumption of independence. To test the assumption of interval independence, we calculate the mutual information between an interval Δ_1 and Δ_n , the n th interval after it in a data sequence, as follows:

$$I(\Delta_n; \Delta_1) = \sum_{\Delta_n} \sum_{\Delta_1} P(\Delta_n, \Delta_1) \log_2 \frac{P(\Delta_n, \Delta_1)}{P(\Delta_n)P(\Delta_1)}. \quad (5)$$

Note that Equation 5 is symmetric with respect to the two intervals. Because of the limited sampling of the joint distributions, our estimates of these probability distributions will tend to lead to an overestimate of information (Treves and Panzeri, 1995); however, because $n \gg m$ ($n = 30,000$, $m = 59^2 = 3481$), the bias-corrected estimators are effective (Paninski, 2003). We correct for calculating the information for all of our data, I_{full} , and for two halves of our data, $I_{\text{half } 1}$ and $I_{\text{half } 2}$; the corrected information is then $I_{\text{corr}} = 2I_{\text{full}} - 0.5(I_{\text{half } 1} + I_{\text{half } 2})$. This method derives from the fact that the first correction term in undersampled biased information is $C_1 = (2n \ln 2)^{-1}(m_s - 1)(m_r - 1)$, where n is the sample size and m_s and m_r are the sizes, or number of bins, of the stimulus and response sets, respectively (Treves and Panzeri, 1995; Panzeri and Treves, 1996), and allows us to calculate I_{corr} without explicitly finding C_1 . We verified that C_1 is indeed linear with regard to smaller subsets of

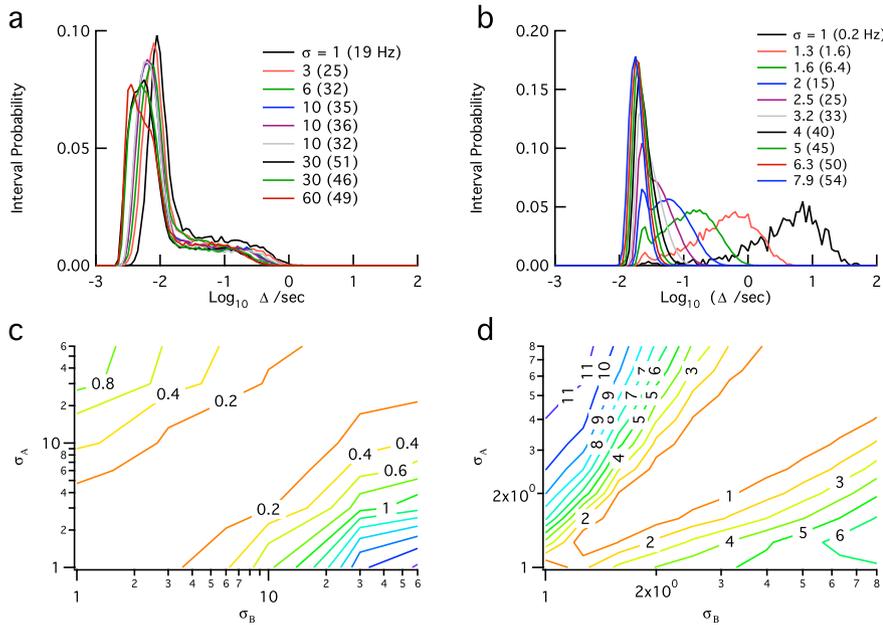


Figure 4. Calculating the difference between different steady-state ISI distributions. Displayed are log-|SI| distributions in response to zero-mean random stimuli from fly H1 neuron (**a**) and HH model neuron (**b**) in which stimulus SDs (normalized to the lowest SD) are shown in the figure key (right) and the mean firing rates (Hz) are shown in parentheses. Bin width is 0.05 \log_{10} units; the sum of all bins equals one. Shown also are contours of the Kullback–Leibler divergence D_{KL} between the log-|SI| distributions for the fly H1 neuron (**c**) and HH model neuron (**d**) as calculated from Equation 1. In the fly there is generally a larger difference between the distributions for the transition from a low-variance to high-variance condition, i.e., for $\sigma_A < \sigma_B$, rather than vice versa. The model neuron shows the opposite trend. The distributions corresponding to the three smallest stimulus SD values for the HH model neuron shown in **b** are omitted from additional analyses.

data as well; thus the first-order term is a good approximation of the error.

Estimating a time-varying stimulus variance. We demonstrate an explicit decoding of this interval code to estimate the time-varying SD of the stimulus by using the HH model. Probability distributions $P(\Delta|\sigma)$ are estimated in response to a range of steady-state stimulus SDs (see Fig. 4*b*). A time-varying stimulus variance then can be estimated from n successive output intervals of the HH model, using a maximum likelihood estimator as follows:

$$\sigma_{ML} = \arg \max_{\sigma} \left(\sum_n \log P(\Delta_n|\sigma) \right), \quad (6)$$

where the estimate is based on an implicit assumption of a locally unchanging variance. Alternatively, stimulus variance can be estimated from spike rate via a firing rate–current variance curve analogous to a firing rate–current mean ($f-I$) curve. Because the HH model does not display spike frequency adaptation, the responses of the model to statistical changes in the stimulus are nearly instantaneous, and estimates from intervals and spike rate should be similar.

Results

In the case of H1 it has been shown previously that the spike rate varies adaptively with the stimulus variance envelope (Fairhall et al., 2001). Here we present a case in which spike frequency adaptation leads to explicit ambiguity in the firing rate response to the local value of stimulus variance. In this case we show that the distribution of ISIs presents an alternative decoding variable to spike rate. We examine the speed, accuracy, and precision with which this distributional code conveys information about the stimulus variance.

Resolving ambiguity

We construct a three-state experiment in which rate responses show adaptation in different directions during presentation of

the same variance stimulus (Fig. 1). In this experiment we switch the SD of the white noise stimulus among three values, $\sigma_H = 3\sigma_M = 9\sigma_L$, presented in the sequence $\sigma_H \rightarrow \sigma_M \rightarrow \sigma_L \rightarrow \sigma_M$ (Fig. 1*a*). Thus the system experiences a stimulus SD of σ_M while adapting both to an upward transition $\sigma_L \rightarrow \sigma_M$ and to a downward transition $\sigma_H \rightarrow \sigma_M$. Spike rates corresponding to stimulus SD σ_M from the two transitions differ noticeably (Fig. 1*b*). To estimate the distribution of possible firing rates, we collect the firing rates produced in response to the random white noise stimuli at a given time with respect to the variance envelope. Figure 1*c* displays the distributions of spike counts in 10 ms bins obtained from the last second of each variance epoch. The rate distributions sharing the stimulus SD σ_M overlap equally with those resulting from σ_H and σ_L (Fig. 1*c*). However, the steady-state spike interval distributions for σ_M are nearly identical and clearly differ from those resulting from σ_H and σ_L (Fig. 1*d*).

Therefore, if one were to use the spike rate to estimate the stimulus variance, one might conclude erroneously that four different stimulus variances were used rather than three, although the spike interval distributions would identify correctly the three different stimuli variances. Because we are computing rate by counting spikes in a given time window, small changes in the tails of log-|SI| distributions can lead to large changes in the spike rate such that similar ISI distributions result in different rate distributions, as demonstrated in Figure 1, *c* and *d*. For example, adding a few long intervals has a negligible effect on the interval distribution but a dramatic effect on the estimated rate, because these few intervals span a long period of time.

To quantify how the locally sampled distributions of spike rate (Fig. 2*a*) and spike intervals (Fig. 2*b*) evolve with time, we use the Kullback–Leibler divergence D_{KL} , defined in Equation 1. We sample rate $P_t(r)$ and interval $P_t(\Delta)$ distributions (300 repeats) in a 1 s bin at each time point t . We then compare these with their respective reference distributions $P(r|\sigma_i)$ and $P(\Delta|\sigma_i)$, which we determine from the rates and intervals sampled in the last second of each variance epoch. In Figure 2*a* we see that the rate distributions continue to evolve with time throughout the epoch until they are, by construction, identical to the reference distribution in the final second. The evolution of the rate is evident through the continuous change in the D_{KL} measure. However, in Figure 2*b*, D_{KL} for the interval distributions approaches a steady state of close to zero within approximately the first second. Furthermore, when the stimulus has SD σ_M , the interval D_{KL} values from the two σ_M steady-state distributions are approximately equally close to zero.

Figures 1*d* and 2*b* suggest that information about stimulus variance can be obtained quickly from ISIs after a transition. Therefore, we calculate the time-dependent mutual information provided by each successive interval after a variance switch according to Equation 4 in Materials and Methods. For the n th interval after a switch, Δ_n , we calculate $I(\Delta_n; \sigma)$ and find that the amount of information provided by each interval increases rap-

idly to a steady state (Fig. 3). Given a three-state discrimination, there is a maximum of $\log_2(3) = 1.585$ bits of information to obtain. If the assumption of independent intervals is valid, perfect discrimination then can be achieved with ~ 8 –12 intervals.

Evaluating the accuracy and precision of the distributional code

How accurately can such an interval code convey stimulus information, and what are the speed/accuracy tradeoffs in estimating the variance from a sequence of intervals? One expects that the distributions are most discriminable in the steady state and that adaptation will slow the acquisition of information about a variance change. Therefore, we examined steady-state data from the *C. vicina* fly viewing uniformly distributed white noise motion stimuli with a constant variance and data from a HH model neuron (Hodgkin and Huxley, 1952) driven by a Gaussian-distributed white noise current input with constant variance.

ISI distributions of the fly and simulated data demonstrate a shift of the peak toward shorter intervals, with increasing stimulus variance that is more marked in the fly ISI distributions than in the simulated data (Fig. 4*a,b*). Although both distributions display tails at long time scales, the tails of the fly ISI distributions have a consistent slope, whereas the relative area under the HH tails increases dramatically with decreasing variance. The presence of these long time scale tails allows for significant changes in spike rate without a large change in the overall ISI distribution. Contours of the Kullback–Leibler divergences between distributions generated from stimuli with SD σ_A and σ_B are plotted in Figure 4, *c* and *d*, and the asymmetry in the D_{KL} is readily apparent because the greatest distance for the fly data is between responses resulting from low- to high-variance switches in contrast to the model neuron in which the opposite is true.

Asymmetry in the discrimination of variance transitions has been noted previously in a fly experiment of time-varying variance, in which steady-state information transmission was reached more quickly after an increase in stimulus variance than after a decrease in variance (Fairhall et al., 2001). In the detection of a variance change in a Gaussian distribution, this asymmetry follows directly from signal detection theory: it is easier to spot an unexpected large-variance outlier than to realize that too many values are falling in the center of the distribution (DeWeese and Zador, 1998; Fairhall et al., 2001; Snippe et al., 2004). Our results, which examine steady-state data, suggest that in decoding the change the opposite also can be true; the interval distributions are highly non-Gaussian, and the direction of the asymmetry in D_{KL} will depend on the particular shape of the two distributions. For decoding during adaptation at least two asymmetries may compete: different rates of adaptation either can reinforce or can counteract the asymmetry in reading out the non-Gaussian signal from the interval data. Even in steady state there may be two causes of asymmetry: again, the possible asymmetry of the D_{KL} divergence between distributions (Fig. 4*c,d*) and an asymmetry in the variance of a discriminator variable, such as the SNR discriminator based on Equation 3. Even when the D_{KL} is higher for a low-to-high variance transition, a very large variance can make this transition less reliably detected from the resulting intervals than a high-to-low transition (data not shown).

To quantify how well information about stimulus variance can be decoded from sequences of n intervals, we assume successive intervals are independent and use the cumulative $D_{cKL}^{(n)}$, Equation 2, to compare responses (see Materials and Methods). Figure 5 shows how the estimated $D_{cKL}^{(n)}$ varies as a function of stimulus SD ratio for several values of n . It is evident that discrimination is

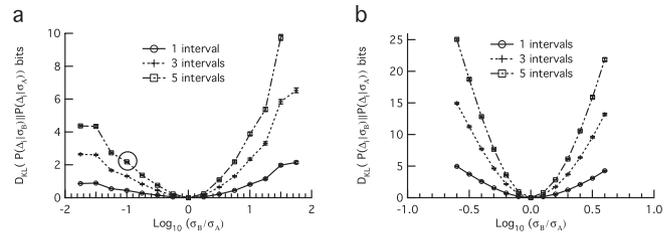


Figure 5. Cumulative Kullback–Leibler divergence ($D_{cKL}^{(n)}$) between neural responses, as defined in Equation 2, as a function of the log ratio of two stimulus SDs, σ_B and σ_A . This divergence is related to the probability of misclassification in which a distance of one bit corresponds to a twofold reduction in error and is a measure of the capacity to discriminate between two ISI distributions resulting from two stimulus variances, as in Figure 4, *a* and *b*. $D_{cKL}^{(n)}$ is for $n = 1, 3$, or 5 ISIs from fly data for a given SD ratio (*a*) and from the HH model neuron (*b*). For example, the circle represents the average discrimination for all 10-fold ratios of stimulus SD decrease, e.g., 10 to 1, 30 to 3, and 60 to 6. Error bars represent the SD.

relatively easier for the model neuron than for the fly H1 neuron. The asymmetries of Figure 4, *c* and *d*, also are reflected here such that discrimination is easier for low-to-high variance changes in the fly data, and high-to-low variance changes are discriminated more easily from model neuron data.

As a measure of discrimination we use a difference of one bit, which corresponds to a twofold reduction in the probability of classification error by a decoder (Johnson et al., 2001). Discriminating stimuli with SDs that differ by a factor of 10 in the fly (Fig. 6*a,c*) requires approximately three intervals and ~ 100 ms. For HH output the discrimination is relatively easier: discrimination of variances differing by a factor of 3 requires approximately three spikes and < 100 ms (Fig. 6*b,d*). Although the time to discrimination typically increases with the number of spikes required, this is not always true and depends on the shape of the ISI distributions. For example, when we compare distributions that differ significantly in the long time scale tails, few spikes may be needed to reach a given level of discrimination, but the average required time may be long if the discriminating spike intervals lie in the long time scale tails.

Because this measure of decoding ability assumes independence of successive ISIs, we test this assumption in two ways. First, we calculate D_{cKL} values for a two-interval discrimination, using both the joint distribution of the two ISIs, which takes into account interval correlations, and the independent approximation, Equation 3. For the fly data we find that, as expected, discrimination with the joint distribution is slightly better but on the whole very similar (Fig. 7*a*). Second, we calculate the mutual information $I(\Delta_1; \Delta_n)$ between the first and the n th interval in a sequence, according to Equation 5, for each of the distributions in Figure 4*a* and plot their mean values (Fig. 7*b*). This gives us a measure of the correlations between intervals and is equivalent to the D_{cKL} between the joint and independent distributions of Δ_1 and Δ_n . After a separation of more than one interval, the intervals, although slightly correlated, have very small mutual information and can be considered as approximately independent (Fig. 7*b*). For the HH neuron there is approximately zero correlation after a single interval.

Real-time decoding

Until now, we have assessed the ability with which one, in principle, could decode information about stimulus variance. However, we also would like to decode stimulus variance explicitly. Because we have unlimited data from the HH model, we use the simulation to test the plausibility of one-trial real-time decoding

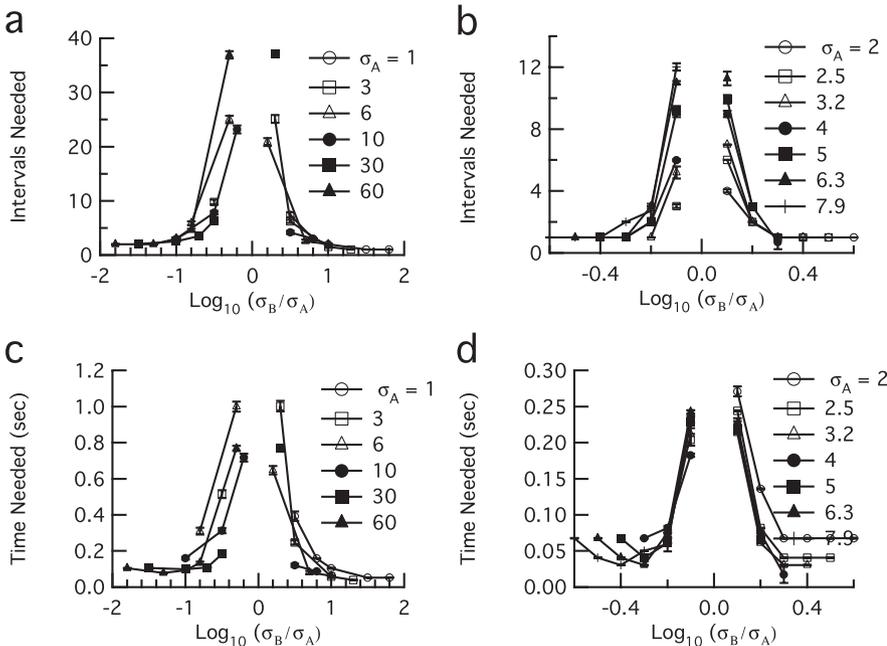


Figure 6. Number of spikes and corresponding mean time needed for discrimination, where discrimination is defined as a distance of one bit. For each data series, the stimulus SD of the comparison distribution (σ_A) is held constant while the SD of the source distribution varies (σ_B). **a**, Number of H1 spikes required to discriminate between two variances. **b**, HH spikes required to discriminate between two variances. **c**, Mean time (seconds) in fly H1 neuron required to discriminate between two variances. **d**, Mean time (seconds) in HH neuron required to discriminate between two variances. Discrimination requires considerably fewer spikes and less time in the model neuron as compared with the fly neuron. The figure key values (right) for σ_A and σ_B are in normalized units of SD, which correspond to those in Figure 4.

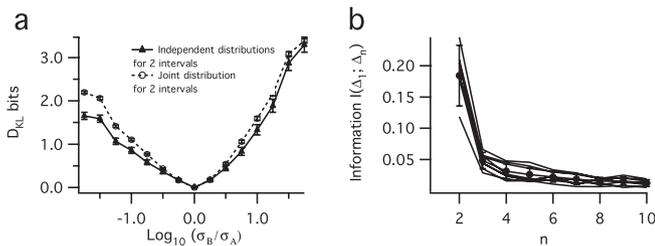


Figure 7. Testing the independence assumption. **a**, Kullback–Leibler divergences calculated by using independent distributions for $P(\Delta_1|\sigma)$ and $P(\Delta_2|\sigma)$ (triangles) and joint distributions $P(\Delta_1, \Delta_2|\sigma)$ (circles) as a function of the log ratio of two stimulus SDs, σ_B and σ_A , for independent distributions; error bars represent the SD. **b**, Mutual information, Equation 5, between the first spike interval and the n th interval of fly data for interval sequences that result from steady-state zero-mean stimuli with varying SDs. The sharp decrease between $n = 2$ and $n = 3$ suggests that dependence between intervals is confined mainly to the first and second interval; other intervals effectively are drawn randomly from the probability distribution. This information calculation was corrected for undersampling, as described in Materials and Methods. Gray lines represent the mutual information between intervals from different variance distributions, whereas the black line represents the mean; error bars represent the SD across different stimulus conditions.

of a continuously varying variance envelope. Therefore, we now estimate stimulus variance using one example of a 60 s spike train from the HH neuron (Fig. 8). We assume that we have access to a library of reference distributions $P(\Delta|\sigma)$ in which each distribution is gathered from a steady-state stimulus variance experiment (Fig. 4b). From this we can use a maximum likelihood estimator to estimate the value of the time-varying stimulus variance: the estimate is the SD that yields the highest probability of n successive spike intervals, as in Equation 6. In principle, given a suitable parametric family of distributions as a function of variance, one might compute these results analytically. However, here we use

sampled data with 10 values of the variance spaced such that successive variances are ~ 1.6 times the previous (given Fig. 5b, this is a reasonable discretization scale for decoding with the use of more than one interval). If we consider two successive intervals, an estimate requires on average 22 ms of spike train data (Fig. 8a), whereas an estimate based on 15 intervals requires a mean time of 314 ms (Fig. 8b). Using these mean times as the window width, we also estimate stimulus SD from the number of spikes in this window. As the number of intervals considered or the window width increases, variance estimates better match the true stimulus variance (Fig. 8c,d). Because the HH model does not display spike frequency adaptation, we do not expect any ambiguity, as exemplified in Figure 1, and estimates from intervals and spike rate should converge for long-enough windows (Fig. 8e). However, calculating spike rate in short windows severely discretizes the possible estimates of variance (Fig. 8a, red circles), and for short times the calculations based on intervals can give better variance estimates even when adaptation is not present (Fig. 8e).

Discussion

We have examined the ability of a code based on the local distribution of ISIs to convey information about stimulus variance in an identified fly visual neuron and in a simulated HH neuron. We present a specific example in the fly visual system in which adapting spike rates measured locally give insufficient information to determine stimulus variance, whereas an ISI distributional code allows this discrimination to be performed with reasonable accuracy. We find that this code can be decoded rapidly to convey accurate information about the local stimulus variance, that the precision of this code depends on the particular shape of the ISI distribution, and that this code is effective in the general and simple HH model neuron.

During adaptation in H1 the rate distributions do not provide a unique mapping to stimulus variance, whereas ISI distributions provide an approximately one-to-one mapping (Fig. 1). This results from the observation that rate distributions evolve slowly with time (on the order of seconds), whereas ISI distributions quickly (~ 200 ms) reach a steady state (Fig. 2). The ISI distributions permit variance discrimination as quantified by the Kullback–Leibler divergence, which depends on the shape of the ISI distribution (Figs. 4–6). This can lead to asymmetries with respect to variance increases or decreases in the time to decode the variance change. Decoding with ISI distributions can be used to estimate a time-varying stimulus variance (Fig. 8).

As demonstrated in Figures 1–3, this distributional code can yield stimulus variance information quickly that is unavailable from spike rate; qualitatively, this is based on the shapes of the ISI distributions, which reach steady state after a sudden variance change more quickly than does spike rate. That the ISI distributions may be very similar although the spike rate distributions are different and evolving may seem paradoxical, because the firing rate can be defined as the inverse of the mean ISI. However, here we have considered rate computed from spike counts in a given

time window. Short intervals are characteristic of a given variance, and variance adaptation leads to a slow change in the proportion of short and long intervals. In the switch to a new variance there is very rapid change in the typical pattern of short intervals, although the relative probability of long intervals adapts more slowly. These small changes in the long interval tails of the distributions lead to significant variations in the rate, as we have observed during adaptation. These features typically render the rate unsuitable to estimate accurately over the short times relevant for the encoding of natural stimuli (Gautrais and Thorpe, 1998). Furthermore, different parts of the interval distribution adapt at different speeds, and we accentuate this feature by plotting log-ISI distributions.

The robustness of this distributional code is related to the shape of the steady-state distributions. The ISI distributions of both the fly H1 neuron and the HH model vary systematically with the variance of the driving stimulus. These variations can be analyzed quantitatively by using the Kullback–Leibler divergences (Figs. 4*c,d*, 5, 6). It has been noted previously that it is easier, both theoretically and experimentally, to discriminate a variance increase than a variance decrease (DeWeese and Zador, 1998; Fairhall et al., 2001; Snippe et al., 2004). However, here we have considered the speed and accuracy of decoding from a non-Gaussian distribution, where this may not hold. As has been pointed out (DeWeese and Zador, 1998), the speed of recognizing a change in distribution depends on the occurrences of outliers after the change. In our results the distributions corresponding to different stimulus variances have distinct and non-Gaussian forms; Figure 4, *a* and *b*, includes cases in which the occurrence of distinctive outliers happens more frequently for the low-variance distribution than for the high-variance distribution, thus resulting in easier identification of the low-variance rather than the high-variance source.

In addition to coding via spike rate, spike timing, and specific spike patterns, distributional coding may provide another means for the neuron to transmit information. Although information in spike rate typically is thought to be transmitted rapidly via averaging across populations of neurons (Shadlen and Newsome, 1998), we have shown that this distributional code provides a means of conveying ensemble information rapidly (~10–100 ms) via the output of a single neuron. Other results also show a strong dependence of the interval statistics on the variance of the driving stimulus (Johannesma, 1969; Hunter et al., 1998; Wang and Wang, 2000; Hunter and Milton, 2003). Furthermore, results from the HH neuron demonstrate that this variance dependence is present for a simple conductance-based model neuron and is likely to be present for many neural types. The decoding scheme that we propose shows that, in principle, this information is available. A biophysical implementation of such a decoding may involve an array of tuned synapses showing short-term depression and facilitation (Tsodyks and Markram, 1997; Varela et al., 1997) to provide sensitivity to interval sequence. Thus a library of prob-

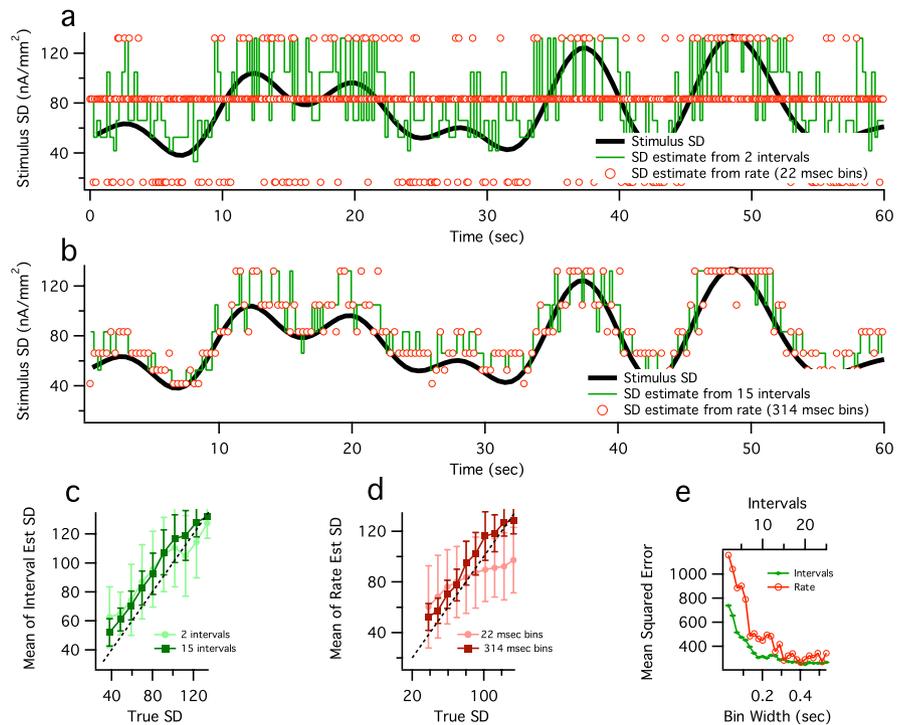


Figure 8. Decoding a time-varying stimulus SD from HH spike trains by using 2 ISIs (*a*) and 15 ISIs (*b*, green lines). For each data point we find σ , for which $P(\Delta|\sigma)$ is maximized for either 2 or 15 intervals, as in Equation 6. Stimulus SD is estimated from the rate calculated in a window width equal to the mean time of 2 or 15 intervals (22 and 314 ms, respectively; red circles). The instantaneous stimulus SD is shown in black. At very short times, the range of possible estimates is limited when calculated from rate, but not from intervals. For longer times, the estimates from rate and intervals are comparable, because the HH model does not show slow spike frequency and would not be expected to display the ambiguity between rate and intervals as in fly neuron (Fig. 1). Shown is a comparison of estimated versus true SD, using intervals (*c*) and spike rate (*d*), in which the vertical bars represent the SDs of the estimated SD, and mean-squared error of the estimated SD (*e*) as a function of interval number or spike rate bin width.

ability distributions $P(\Delta|\sigma)$ (as in Fig. 4*b*) might be “stored” at synapses in which a preferred ISI sequence leads to a maximal postsynaptic response.

Adaptation is functionally very important for sensory systems, allowing the system to tailor its responses to the local stimulus and thereby increase information transmission. However, adaptation leads to ambiguity, whereby the relationship between stimulus and response is no longer one-to-one but depends on the statistical context of the stimulus. There are three possible solutions to this problem. Information may be conveyed by other neurons in the network, or it may be discarded entirely, as may be the case for light/dark adaptation of photoreceptors. We suggest a third possibility: contextual information may be conveyed by the statistical properties of the spike train. Although the first two of these possibilities also may be true for H1, we have demonstrated that information about the statistical ensemble indeed may be extracted. The dependence of the interval statistics on the stimulus ensemble may be a generic property of neural response, because it is observed in neurons as simple as the HH model (Wang and Wang, 2000). It therefore is tempting to suggest that using this dependence to convey ensemble information may be a universal capability of neural computation.

References

- Baccus SA, Meister M (2002) Fast and slow contrast adaptation in retinal circuitry. *Neuron* 36:909–919.
- Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252:1854–1857.

- Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. *Neural Comput* 12:1531–1552.
- Carandini M, Ferster D (1997) A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* 276:949–952.
- Cariani P (1999) Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast* 6:147–172.
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley.
- Dayan P, Abbott LF (2001) *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge, MA: Massachusetts Institute of Technology.
- Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8:1684–1689.
- de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc R Soc London B Biol Sci* 234:379–414.
- DeWeese M, Zador A (1998) Asymmetric dynamics in optimal variance adaptation. *Neural Comput* 10:1179–1202.
- Egelhaaf M, Warzecha AK (1999) Encoding of motion in real time by the fly visual system. *Curr Opin Neurobiol* 9:454–460.
- Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412:787–792.
- Franceschini N, Riehle A, Le Nestour A (1989) Directionally selective motion detection by insect neurons. In: *Facets of vision: compound eyes from Exner to Autrum and beyond* (Stavenga DG, Hardie RC, eds), pp 360–390. Berlin: Springer.
- Gautrais J, Thorpe S (1998) Rate coding versus temporal order coding: a theoretical approach. *Biosystems* 48:57–65.
- Gelman A (1995) *Bayesian data analysis*, Ed 1. London: Chapman and Hall.
- Haag J, Borst A (1997) Encoding of visual motion information and reliability in spiking and graded potential neurons. *J Neurosci* 17:4809–4819.
- Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol (Lond)* 117:500–544.
- Hunter JD, Milton JG (2003) Amplitude and frequency dependence of spike timing: implications for dynamic regulation. *J Neurophysiol* 90:387–394.
- Hunter JD, Milton JG, Thomas PJ, Cowan JD (1998) Resonance effect for neural spike time reliability. *J Neurophysiol* 80:1427–1438.
- Johannesma PIM (1969) *Stochastic neural activity: a theoretical investigation*. PhD thesis, p 91. University of Nijmegen, The Netherlands.
- Johnson DH, Gruner CM, Baggerly K, Seshagiri C (2001) Information-theoretic analysis of neural coding. *J Comput Neurosci* 10:47–69.
- Keat J, Reinagel P, Reid RC, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. *Neuron* 30:803–817.
- Kepecs A, Lisman J (2003) Information encoding and computation with spikes and bursts. *Network Comput Neural Syst* 14:103–118.
- Kim KJ, Rieke F (2001) Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J Neurosci* 21:287–299.
- Kvale MN, Schreiner CE (2004) Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol* 91:604–612.
- Lesica NA, Stanley GB (2004) Encoding of natural scene movies by tonic and burst spikes in the lateral geniculate nucleus. *J Neurosci* 24:10731–10740.
- Mainen ZF, Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. *Science* 268:1503–1506.
- Meister M, Berry 2nd MJ (1999) The neural code of the retina. *Neuron* 22:435–450.
- Paninski L (2003) Estimation of entropy and mutual information. *Neural Comput* 15:1191–1253.
- Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. *Network Comput Neural Syst* 7:87–107.
- Press WH (1992) *Numerical recipes in C: the art of scientific computing*, Ed 2. Cambridge, UK: Cambridge UP.
- Reich DS, Mechler F, Purpura KP, Victor JD (2000) Interspike intervals, receptive fields, and information encoding in primary visual cortex. *J Neurosci* 20:1964–1974.
- Reinagel P, Reid RC (2002) Precise firing events are conserved across neurons. *J Neurosci* 22:6837–6841.
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) *Spikes: exploring the neural code*. Cambridge, MA: Massachusetts Institute of Technology.
- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18:3870–3896.
- Shapley R, Victor JD (1979) The contrast gain control of the cat retina. *Vision Res* 19:431–434.
- Snippe HP, Poot L, van Hateren JH (2004) Asymmetric dynamics of adaptation after onset and offset of flicker. *J Vis* 4:1–12.
- Stanley GB, Li FF, Dan Y (1999) Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J Neurosci* 19:8036–8042.
- Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and information in neural spike times. *Phys Rev Lett* 80:197–200.
- Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Comput* 7:399–407.
- Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci USA* 94:719–723.
- Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Nelson SB (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci* 17:7926–7940.
- Wang Y, Wang ZD (2000) Information coding via spontaneous oscillations in neural ensembles. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 62:1063–1068.
- Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol* 78:2336–2350.