Synergy in a Neural Code

Naama Brenner
NEC Research Institute, Princeton, NJ 08540, U.S.A.

Steven P. Strong
Institute for Advanced Study, Princeton, NJ 08544, and NEC Research Institute, Princeton, NJ 08540, U.S.A.

Roland Koberle
Instituto di Física de São Carlos, Universidade de São Paulo, 13560–970 São Carlos, SP Brasil, and NEC Research Institute, Princeton, NJ 08540, U.S.A.

William Bialek
Rob R. de Ruyter van Steveninck
NEC Research Institute, Princeton, NJ 08540, U.S.A.

We show that the information carried by compound events in neural spike trains—patterns of spikes across time or across a population of cells—can be measured, independent of assumptions about what these patterns might represent. By comparing the information carried by a compound pattern with the information carried independently by its parts, we directly measure the synergy among these parts. We illustrate the use of these methods by applying them to experiments on the motion-sensitive neuron H1 of the fly's visual system, where we confirm that two spikes close together in time carry far more than twice the information carried by a single spike. We analyze the sources of this synergy and provide evidence that pairs of spikes close together in time may be especially important patterns in the code of H1.

1 Introduction ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬

Throughout the nervous system, information is encoded in sequences of identical action potentials or spikes. The representation of sense data by these spike trains has been studied for 70 years (Adrian, 1928), but there remain many open questions about the structure of this code. A full understanding of the code requires that we identify its elementary symbols and characterize the messages that these symbols represent. Many different possible elementary symbols have been considered, implicitly or explicitly, in previous work on neural coding. These include the numbers of spikes in time windows of fixed size and individual spikes themselves. In cells that

produce bursts of action potentials, these bursts might be special symbols that convey information in addition to that carried by single spikes. Yet another possibility is that patterns of spikes—across time in one cell or across a population of cells—can have a special significance, a possibility that has received renewed attention as techniques emerge for recording the activity of many neurons simultaneously.

In many methods of analysis, questions about the symbolic structure of the code are mixed with questions about what the symbols represent. Thus, in trying to characterize the feature selectivity of neurons, one often makes the a priori choice to measure the neural response as the spike count or rate in a fixed window. Conversely, in trying to assess the significance of synchronous spikes from pairs of neurons, or bursts of spikes from a single neuron, one might search for a correlation between these events and some particular stimulus features. In each case, conclusions about one aspect of the code are limited by assumptions about another aspect. Here we show that questions about the symbolic structure of the neural code can be separated out and answered in an information-theoretic framework, using data from suitably designed experiments. This framework allows us to address directly the significance of spike patterns or other compound spiking events. How much information is carried by a compound event? Is there redundancy or synergy among the individual spikes? Are particular patterns of spikes especially informative?

Methods to assess the significance of spike patterns in the neural code share a common intuitive basis:

- Patterns of spikes can play a role in representing stimuli if and only if the occurrence of patterns is linked to stimulus variations.

- The patterns have a special role only if this correlation between sensory signals and patterns is not decomposable into separate correlations between the signals and the pieces of the pattern, such as the individual spikes.

We believe that these statements are not controversial. Difficulties arise when we try to quantify this intuitive picture: What is the correct measure of correlation? How much correlation is significant? Can we make statements independent of models and elaborate null hypotheses?

The central claim of this article is that many of these difficulties can be resolved using ideas from information theory. Shannon (1948) proved that entropy and information provide the only measures of variability and correlation that are consistent with simple and plausible requirements. Further, while it may be unclear how to interpret, for example, a 20% increase in correlation between spike trains, an extra bit of information carried by patterns of spikes means precisely that these patterns provide a factor-of-two increase in the ability of the system to distinguish among different sensory inputs. In this work, we show that there is a direct method

of measuring the information (in bits) carried by particular patterns of spikes under given stimulus conditions, independent of models for the stimulus features that these patterns might represent. In particular, we can compare the information conveyed by spike patterns with the information conveyed by the individual spikes that make up the pattern and determine quantitatively whether the whole is more or less than the sum of its parts.

While this method allows us to compute unambiguously how much information is conveyed by the patterns, it does not tell us what particular message these patterns convey. Making the distinction between two issues, the symbolic structure of the code and the transformation between inputs and outputs, we address only the first of these two. Constructing a quantitative measure for the significance of compound patterns is an essential first step in understanding anything beyond the single spike approximation, and it is especially crucial when complex multi-neuron codes are considered. Such a quantitative measure will be useful for the next stage of modeling the encoding algorithm, in particular as a control for the validity of models. This is a subject of ongoing work and will not be discussed here.

2 Information Conveyed by Compound Patterns

In the framework of information theory (Shannon, 1948), signals are generated by a source with a fixed probability distribution and encoded into messages by a channel. The coding is in general probabilistic, and the joint distribution of signals and coded messages determines all quantities of interest; in particular, the information transmitted by the channel about the source is an average over this joint distribution. In studying a sensory system, the signals generated by the source are the stimuli presented to the animal, and the messages in the communication channel are sequences of spikes in a neuron or a population of neurons. Both the stimuli and the spike trains are random variables, and they convey information mutually because they are correlated. The problem of quantifying this information has been discussed from several points of view (MacKay & McCulloch, 1952; Optican & Richmond, 1987; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997; Strong, Koberle, de Ruyter van Steveninck, & Bialek, 1998). Here we address the question of how much information is carried by particular "events" or combinations of action potentials.

2.1 Background. A discrete event E in the spike train is defined as a specific combination of spikes. Examples are a single spike, a pair of spikes separated by a given time, spikes from two neurons that occur in synchrony, and so on. Information is carried by the occurrence of events at particular times and not others, implying that they are correlated with some stimulus features and not with others. Our task is to express the information

conveyed, on average, by an event E in terms of quantities that are easily measured experimentally.[1]

In experiments, as in nature, the animal is exposed to stimuli at each instant of time. We can describe this sensory input by a function $s(t')$, which may have many components to parameterize the time dependence of multiple stimulus features. In general, the information gained about $s(t')$ by observing a set of neural responses is

$$I = \sum_{\text{responses}} \int Ds(t')P[s(t') \text{ \& response}] \log_2 \left(\frac{P[s(t') \text{ \& response}]}{P[s(t')]P[\text{response}]}\right), \quad (2.1)$$

where information is measured in bits. It is useful to note that this mutual information can be rewritten in two complementary forms:

$$\begin{aligned}
I &= -\int Ds(t')P[s(t')]\log_2 P[s(t')] \\
&\quad + \sum_{\text{responses}} P[\text{response}] \int Ds(t')P[s(t')|\text{response}]\log_2 P[s(t')|\text{response}] \\
&= S[P(s)] - \langle S[P(s|\text{response})]\rangle_{\text{response}}, \quad (2.2)
\end{aligned}$$

or

$$\begin{aligned}
I &= -\sum_{\text{responses}} P(\text{response})\log_2 P(\text{response}) \\
&\quad + \int Ds(t')P[s(t')] \sum_{\text{responses}} P[\text{response}|s(t')]\log_2 P[\text{response}|s(t')] \\
&= S[P(\text{response})] - \langle S[P(\text{response}|s)]\rangle_s, \quad (2.3)
\end{aligned}$$

where S denotes the entropy of a distribution; by $\langle \cdots \rangle_s$ we mean an average over all possible values of the sensory stimulus, weighted by their probabilities of occurrence, and similarly for $\langle \cdots \rangle_{\text{responses}}$. In the first form, equation 2.2, we focus on what the responses are telling us about the sensory stimulus (de Ruyter van Steveninck & Bialek, 1988): different responses point more or less reliably to different signals, and our average uncertainty about the sensory signal is reduced by observing the neural response. In the second form, equation 2.3, we focus on the variability and reproducibility of the neural response (de Ruyter van Steveninck, Lewen, Strong, Koberle,

---

[1] In our formulation, an event E is a random variable that can have several outcomes; for example, it can occur at different times in the experiment. The information conveyed by the event about the stimulus is an average over the joint distribution of stimuli and event outcomes. One can also associate an information measure with individual occurrences of events (DeWeese & Meister, 1998). The average of this measure over the possible outcomes is the mutual information, as defined by Shannon (1948) and used in this article.

& Bialek, 1997; Strong et al., 1998). The range of possible responses provides the system with a capacity to transmit information, and the variability that remains when the sensory stimulus is specified constitutes noise; the difference between the capacity and the noise is the information.

We would like to apply this second form to the case where the neural response is a particular type of event. When we observe an event E, information is carried by the fact that it occurs at some particular time t_E. The range of possible responses is then the range of times $0 < t_E < T$ in our observation window. Alternatively, when we observe the response in a particular small time bin of size Δt, information is carried by the fact that the event E either occurs or does not. The range of possible responses then includes just two possibilities. Both of these points of view have an arbitrary element: the choice of bin size Δt and the window size T. Characterizing the properties of the system, as opposed to our observation power, requires taking the limit of high time resolution ($\Delta t \to 0$) and long observation times ($T \to \infty$). As will be shown in the next section, in this limit, the two points of view give the same answer for the information carried by an event.

2.2 Information and Event Rates. A crucial role is played by the event rate $r_E(t)$, the probability per unit time that an event of type E occurs at time t, given the stimulus history $s(t')$. Empirical construction of the event rate $r_E(t)$ requires repetition of the same stimulus history many times, so that a histogram can be formed (see Figure 1). For the case where events are single spikes, this is the familiar time-dependent firing rate or poststimulus time histogram (see Figure 1c); the generalization to other types of events is illustrated by Figures 1d and 1e. Intuitively, a uniform event rate implies that no information is transmitted, whereas the presence of sharply defined features in the event rate implies that much information is transmitted by these events (see, for example, the discussion by Vaadia et al., 1995). We now formalize this intuition and show how the average information carried by a single event is related quantitatively to the time-dependent event rate.

Let us take the first point of view about the neural response variable, in which the range of responses is described by the possible arrival times t of the event. [2] What is the probability of finding the event at a particular time t? Before we know the stimulus history $s(t')$, all we can say is that the event can occur anywhere in our experimental window of size T, so that the probability is uniform $P(\text{response}) = P(t) = 1/T$, with an entropy of $S[P(t)] = \log_2 T$. Once we know the stimulus, we also know the event rate $r_E(t)$, and so our uncertainty about the occurrence time of the event is reduced. Events will occur preferentially at times where the event rate is large, so the probability distribution should be proportional to $r_E(t)$; with proper

---

[2] We assume that the experiment is long enough so that a typical event is certain to occur at some time.
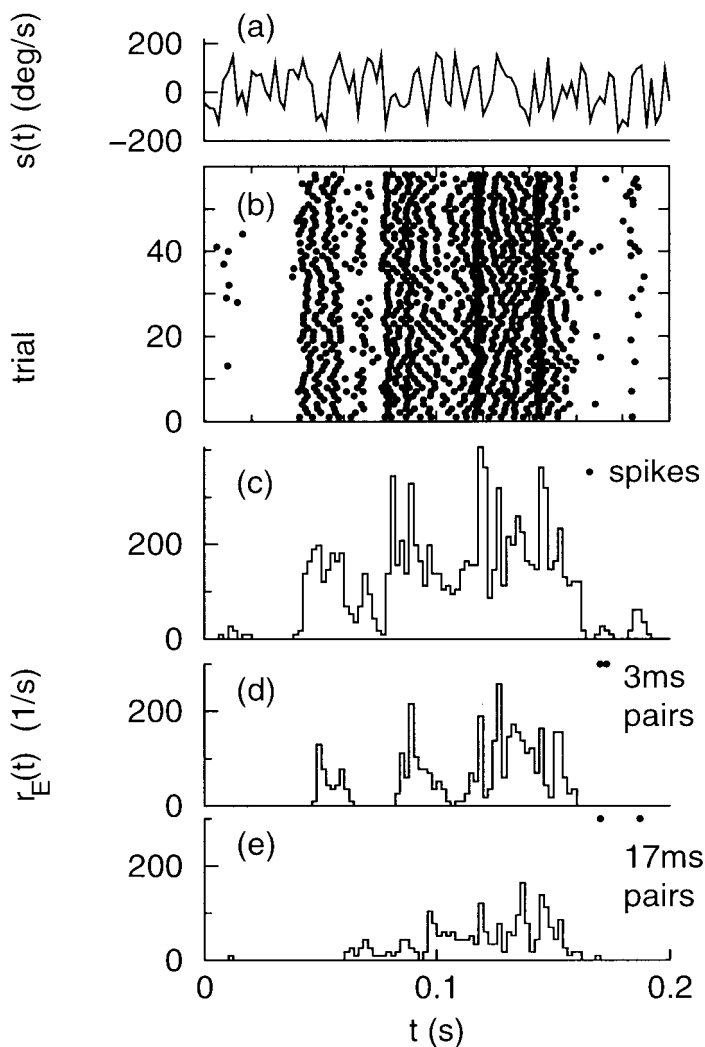
Figure 1: Generalized event rates in the stimulus-conditional response ensemble. A time-dependent visual stimulus is shown to the fly (a), with the time axis defined to be zero at the beginning of the stimulus. This stimulus runs for 10 s and is repeatedly presented 360 times. The responses of the H1 neuron to 60 repetitions are shown as a raster (b), in which each dot represents a single spike. From these responses, time-dependent event rates $r_E(t)$ are estimated: (c) the firing rate (poststimulus time histogram); (d) the rate for spike pairs with interspike time $\tau = 3 \pm 1$ ms and (e) for pairs with $\tau = 17 \pm 1$ ms (e). These rates allow us to compute directly the information transmitted by the events, using equation 2.5.

normalization $P(\text{response}|s) = P(t|s) = r_E(t)/(T\bar{r}_E)$. Then the conditional entropy is

$$S[P(t|s)] = -\int_0^T dt \, P(t|s) \log_2 P(t|s)$$

$$= -\frac{1}{T}\int_0^T dt \, \frac{r_E(t)}{\bar{r}_E} \log_2 \left(\frac{r_E(t)}{\bar{r}_E T}\right). \tag{2.4}$$

In principle one should average this quantity over the distribution of stimuli $s(t')$; however, if the time T is long enough and the stimulus history sufficiently rich, the ensemble average and the time average are equivalent. The validity of this assumption can be checked experimentally (see Figure 2c). The reduction in entropy is then the gain in information, so

$$I(E; s) = S[P(t)] - S[P(t|s)]$$

$$= \frac{1}{T}\int_0^T dt \, \left(\frac{r_E(t)}{\bar{r}_E}\right) \log_2 \left(\frac{r_E(t)}{\bar{r}_E}\right), \tag{2.5}$$

or equivalently

$$I(E; s) = \left\langle \left(\frac{r_E(t)}{\bar{r}_E}\right) \log_2 \left(\frac{r_E(t)}{\bar{r}_E}\right) \right\rangle_s, \tag{2.6}$$

where here the average is over all possible value of s weighted by their probabilities $P(s)$.

In the second view, the neural response is a binary random variable, $\sigma_E \in \{0, 1\}$, marking the occurrence or nonoccurrence of an event of type E in a small time bin of size Δt. Suppose, for simplicity, that the stimulus takes on a finite set of values s with probabilities $P(s)$. These in turn induce the event E with probability $p_E(s) = P(\sigma_E = 1|s) = r_E(s)\Delta t$, with an average probability for the occurrence of the event $\bar{p}_E = \sum_s P(s) r_E(s)\Delta t = \bar{r}_E\Delta t$. The information is the difference between the prior entropy and the conditional entropy: $I(E; s) = S(s) - \langle S(s|\sigma_E)\rangle$, where the conditional entropy is an average over the two possible values of σ_E. The conditional probabilities are found from Bayes' rule,

$$P(s|\sigma_E = 1) = \frac{P(s)p_E(s)}{\bar{p}_E}$$

$$P(s|\sigma_E = 0) = \frac{P(s)(1 - p_E(s))}{(1 - \bar{p}_E)}, \tag{2.7}$$

and with these one finds the information,

$$I(E; s) = -\sum_s P(s) \log_2 P(s) + \sum_{\sigma_E = 0,1} P(s|\sigma_E) \log_2 P(s|\sigma_E)$$

$$= \sum_s P(s) \left[p_E(s) \log_2 \left(\frac{p_E(s)}{\bar{p}_E}\right) + (1 - p_E(s)) \log_2 \left(\frac{1 - p_E(s)}{1 - \bar{p}_E}\right)\right]. \tag{2.8}$$
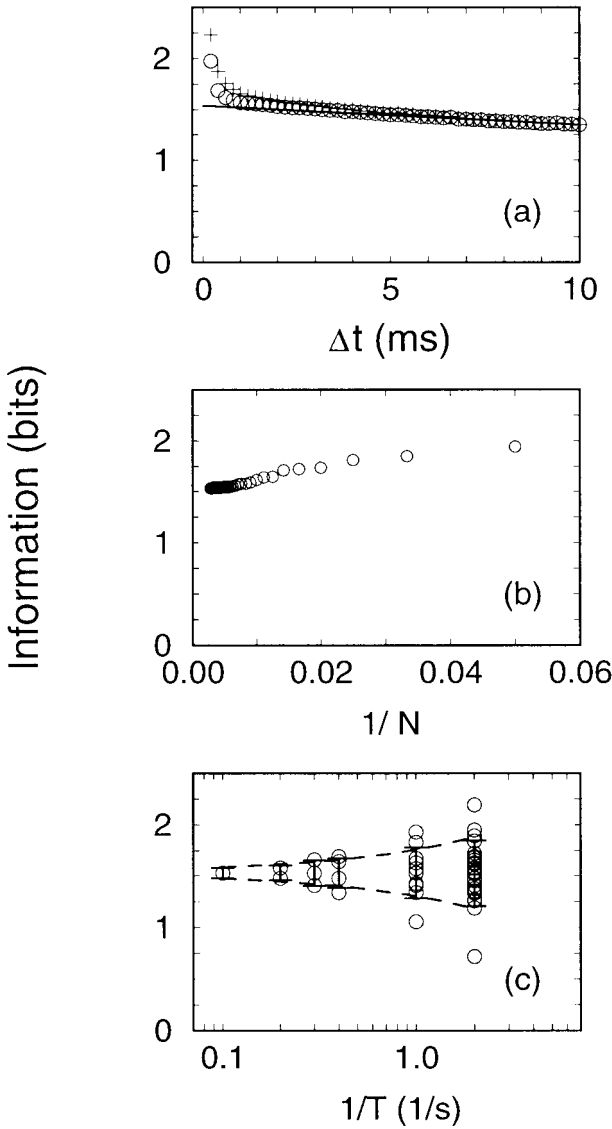
Taking the limit $\Delta t \to 0$, consistent with the requirement that the event can occur at most once, one finds the average information conveyed in a small time bin; dividing by the average probability of an event, one obtains equation 2.6 as the information per event.

Equation 2.6 and its time-averaged form, equation 2.5, are exact formulas that are valid in any situation where a rich stimulus history can be presented

repeatedly. It enables the evaluation of the information for arbitrarily com-
plex events, independent of assumptions about the encoding algorithm.
The information is an average over the joint distribution of stimuli and re-
sponses. But rather than constructing this joint distribution explicitly and
then averaging, equation 2.5 expresses the information as a direct empirical
average: by estimating the function $r_E(t)$ as a histogram, we are sampling the
distribution of the responses given a stimulus, whereas by integrating over
time, we are sampling the distribution of stimuli. This formula is general
and can be applied for different systems under different experimental con-
ditions. The numerical result (measured in bits) will, of course, depend on
the neural system as well as on the properties of the stimulus distribution.
The error bars on the measurement of the information are affected by the
finiteness of the data; for example, sampling must be sufficient to construct
the rates $r_E(t)$ reliably. These and other practical issues in using equation 2.5
are illustrated in detail in Figure 2.

2.3 The Special Case of Single Spikes. Let us consider in more detail
the simple case where events are single spikes. The average information
conveyed by a single spike becomes an integral over the time-dependent
spike rate $r(t)$,

$$I(1 \text{ spike; s}) = \frac{1}{T} \int_0^T dt \ \left(\frac{r(t)}{\bar{r}}\right) \log_2 \left(\frac{r(t)}{\bar{r}}\right). \tag{2.9}$$

It makes sense that the information carried by single spikes should be re-
lated to the spike rate, since this rate as a function of time gives a complete

---

Figure 2: *Facing page.* Finite size effects in the estimation of the information
conveyed by single spikes. (a) Information as a function of the bin size Δt used
for computing the time-dependent rate $r(t)$ from all 360 repetitions (circles) and
from 100 of the repetitions (crosses). A linear extrapolation to the limit $\Delta t \to 0$ is
shown for the case where all repetitions were used (solid line). (b) Information
as a function of the inverse number of repetitions N, for a fixed bin size $\Delta t =$
2 ms. (c) Systematic errors due to finite duration of the repeated stimulus $s(t)$.
The full 10-second length of the stimulus was subdivided into segments of
duration T. Using equation 2.9 the information was calculated for each segment
and plotted as a function of $1/T$ (circles). If the stimulus is sufficiently long that
an ensemble average is well approximated by a time average, the convergence
of the information to a stable value as $T \to \infty$ should be observable. Further,
the standard deviation of the information measured from different segments of
length T (error bars) should decrease as a square root law $\sigma \propto 1/\sqrt{T}$ (dashed
line). These data provide an empirical verification that for the distribution used
in this experiment, the repeated stimulus time is long enough to approximate
ergodic sampling.

description of the "one-body" statistics of the spike train, in the same way that the single particle density describes the one-body statistics of a gas or liquid. This is true independent of any models or assumptions, and we suggest the use of equations 2.9 and 2.5 to test directly in any given experimental situation whether many body effects increase or reduce information transmission.

We note that considerable attention has been given to the problem of making accurate statistical models for spike trains. In particular, the simplest model—the modulated Poisson process—is often used as a standard against which real neurons are compared. It is tempting to think that Poisson behavior is equivalent to independent transmission of information by successive spikes, but this is not the case (see below). Of course, no real neuron is precisely Poisson, and it is not clear which deviations from Poisson behavior are significant. Rather than trying to fit statistical models to the data and then to try to understand the significance of spike train statistics for neural coding, we see that it is possible to quantify directly the information carried by single spikes and compound patterns. Thus, without reference to statistical models, we can determine whether successive spikes carry independent, redundant, or synergistic information.

Several previous works have noted the relation between spike rate and information, and the formula in equation 2.9 has an interesting history. For a spike train generated by a modulated Poisson process, equation 2.9 provides an upper bound on information transmission (per spike) by the spike train as a whole (Bialek, 1990). Even in this simple case, spikes do not necessarily carry independent information: slow modulations in the firing rate can cause them to be redundant. In studying the coding of location by cells in the rat hippocampus, Skaggs, McNaughton, Gothard, and Markus (1993) assumed that successive spikes carried independent information, and that the spike rate was determined by the instantaneous location. They obtained equation 2.9 with the time average replaced by an average over locations. DeWeese (1996) showed that the rate of information transmission by a spike train could be expanded in a series of integrals over correlation functions, where successive terms would be small if the number of spikes per correlation time were small; the leading term, which would be exact if spikes were uncorrelated, is equation 2.9. Panzeri, Biella, Rolls, Skaggs, and Treves (1996) show that equation 2.9, multiplied by the mean spike rate to give an information rate (bits per second), is the correct information rate if we count spikes in very brief segments of the neural response, which is equivalent to asking for the information carried by single spikes. For further discussion of the relation to previous work, see appendix A.

A crucial point here is the generalization to equation 2.5, and this result applies to the information content of *any* point events in the neural response—pairs of spikes with a certain separation, coincident spikes from two cells, and so on—not just single spikes. In the analysis of experiments, we will emphasize the use of this formula as an exact result for the infor-

mation content of single events, rather than an approximate result for the spike train as a whole, and this approach will enable us to address questions concerning the structure of the code and the role played by various types of events.

3 Experiments in the Fly Visual System

In this section, we use our formalism to analyze experiments on the movement-sensitive cell H1 in the visual system of the blowfly *Calliphora vicina*. We address the issue of the information conveyed by pairs of spikes in this neuron, as compared to the information conveyed independently by single spikes. The quantitative results of this section—information content in bits, effects of synergy, and redundancy among spikes—are specific to this system and to the stimulus conditions used. The theoretical approach, however, is valid generally and can be applied similarly to other experimental systems, to find out the significance of various patterns in single cells or in a population.

3.1 Synergy Between Spikes. In the experiment (see appendix B for details), the horizontal motion across the visual field is the input sensory stimulus $s(t)$, drawn from a probability distribution $P[s(t)]$. The spike train recorded from H1 is the neural response. Figure 1a shows a segment of the stimulus presented to the fly, and Figure 1b illustrates the response to many repeated presentations of this segment. The histogram of spike times across the ensemble of repetitions provides an estimate of the spike rate $r(t)$ (see Figure 1c), and equation 2.5 gives the information carried by a single spike, $I(1 \text{ spike}; s) = 1.53 \pm 0.05$ bits. Figure 2 illustrates the details of how the formula was used, with an emphasis on the effects of finiteness of the data. In this experiment, a stimulus of length $T = 10$ sec was repeated 360 times. As seen from Figure 2, the calculation converges to a stable result within our finite data set.

If each spike were to convey information independently, then with the mean spike rate $\bar{r} = 37$ spikes per second, the total information rate would be $R_{\text{info}} = 56$ bits per second. We used the variability and reproducibility of continuous segments in the neural response (de Ruyter van Steveninck et al., 1997; Strong et al., 1998) to estimate the total information rate in the spike train in this experiment and found that $R_{\text{info}} = 75$ bits per second. Thus, the information conveyed by the spike train as a whole is larger than the sum of contributions from individual spikes, indicating cooperative information transmission by patterns of spikes in time. This synergy among spikes motivates the search for especially informative patterns in the spike train.

We consider compound events that consist of two spikes separated by a time τ, with no constraints on what happens between them. Figure 1 shows segments of the event rate $r_\tau(t)$ for $\tau = 3 \ (\pm 1)$ ms (see Figure 1d),
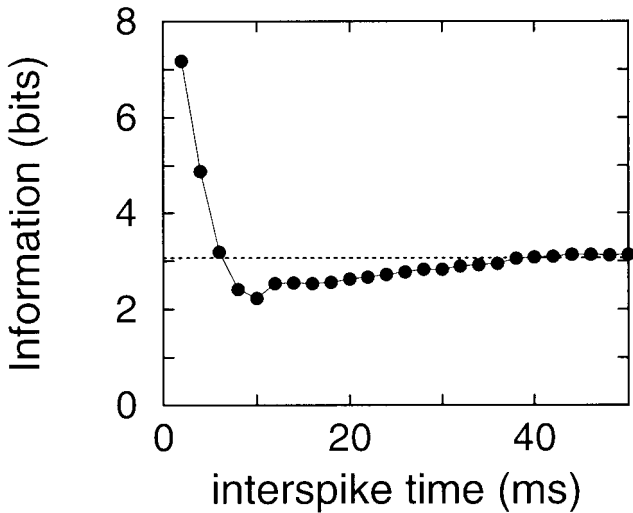
Figure 3: Information about the signal transmitted by pairs of spikes, computed from equation 2.5, as a function of the time separation between the two spikes. The dotted line shows the information that would be transmitted by the two spikes independently (twice the single-spike information).

and for $\tau = 17$ (± 1) ms (see Figure 1e). The information carried by spike pairs as a function of the interspike time τ, computed from equation 2.5, is shown in Figure 3. For large τ, spikes contribute independent information, as expected. This independence is established within approximately 30 to 40 ms, comparable to the behavioral response times of the fly (Land & Collett, 1974). There is a mild redundancy (about 10–20%) at intermediate separations and a very large synergy (up to about 130%) at small τ.

Related results were obtained using the correlation of spike patterns with stimulus features (de Ruyter van Steveninck & Bialek, 1988). There the information carried by spike patterns was estimated from the distribution of stimuli given each pattern, thus constructing a statistical model of what the patterns "stand for" (see the details in appendix A). Since the time-dependent stimulus is in general of high dimensionality, its distribution cannot be sampled directly and some approximations must be made. de Ruyter van Steveninck and Bialek (1988) made the approximation that patterns of a few spikes encode projections of the stimulus onto low-dimensional subspaces, and further that the conditional distributions $P[s(t)|E]$ are gaussian. The information obtained with these approximations provides a lower bound to the true information carried by the patterns, as estimated directly with the methods presented here.

3.2 Origins of Synergy. Synergy means, quite literally, that two spikes together tell us more than two spikes separately. Synergistic coding is often discussed for populations of cells, where extra information is conveyed by patterns of coincident spikes from several neurons (Abeles, Bergmann, Margalit, & Vaadia, 1993; Hopfield, 1995; Meister, 1995; Singer & Gray, 1995). Here we see direct evidence for extra information in pairs of spikes across time. The mathematical framework for describing these effects is the same, and a natural question is: What are the conditions for synergistic coding?

The average synergy $\text{Syn}[E_1, E_2; s]$ between two events E_1 and E_2 is the difference between the information about the stimulus s conveyed by the pair and the information conveyed by the two events independently:

$$\text{Syn}[E_1, E_2; s] = I[E_1, E_2; s] - (I[E_1; s] + I[E_2; s]). \tag{3.1}$$

We can rewrite the synergy as:

$$\text{Syn}[E_1, E_2; s] = I[E_1; E_2|s] - I[E_1; E_2]. \tag{3.2}$$

The first term is the mutual information between the events computed across an ensemble of repeated presentations of the same stimulus history. It describes the gain in information due to the locking of compound event (E_1, E_2) to particular stimulus features. If events E_1 and E_2 are correlated individually with the stimulus but not with one another, this term will be zero, and these events cannot be synergistic on average. The second term is the mutual information between events when the stimulus is not constrained or, equivalently, the predictability of event E_2 from E_1. This predictability limits the capacity of E_2 to carry information beyond that already conveyed by E_1. Synergistic coding (Syn > 0) thus requires that the mutual information among the spikes is increased by specifying the stimulus, which makes precise the intuitive idea of "stimulus-dependent correlations" (Abeles et al., 1993; Hopfield, 1995; Meister, 1995; Singer & Gray, 1995).

Returning to our experimental example, we identify the events E_1 and E_2 as the arrivals of two spikes, and consider the synergy as a function of the time τ between them. In terms of event rates, we compute the information carried by a pair of spikes separated by a time τ, equation 2.5, as well as the information carried by two individual spikes. The difference between these two quantities is the synergy between two spikes, which can be written as

$$\text{Syn}(\tau) = -\log_2\left(\frac{\bar{r}_\tau}{\bar{r}^2}\right) + \frac{1}{T}\int_0^T dt \frac{r_\tau(t)}{\bar{r}_\tau} \log_2\left[\frac{r_\tau(t)}{r(t)r(t-\tau)}\right]$$
$$+ \frac{1}{T}\int_0^T dt \left[\frac{r_\tau(t)}{\bar{r}_\tau} + \frac{r_\tau(t+\tau)}{\bar{r}_\tau} - 2\frac{r(t)}{\bar{r}}\right]\log_2[r(t)]. \tag{3.3}$$

The first term in this equation is the logarithm of the normalized correlation function, and hence measures the rarity of spike pairs with separation τ;

the average of this term over τ is the mutual information between events (the second term in equation 3.2). The second term is related to the local correlation function and measures the extent to which the stimulus modulates the likelihood of spike pairs. The average of this term over τ gives the mutual information conditional on knowledge of the stimulus (the first term in equation 3.2). The average of the third term over τ is zero, and numerical evaluation of this term from the data shows that it is negligible at most values of τ.

We thus find that the synergy between spikes is approximately a sum of two terms, whose averages over τ are the terms in equation 3.2. A spike pair with a separation τ then has two types of contributions to the extra information it carries: the two spikes can be correlated conditional on the stimulus, or the pair could be a rare and thus surprising event. The rarity of brief pairs is related to neural refractoriness, but this effect alone is insufficient to enhance information transmission; the rare events must also be related reliably to the stimulus. In fact, conditional on the stimulus, the spikes in rare pairs are strongly correlated with each other, and this is visible in Figure 1a: from trial to trial, adjacent spikes jitter together as if connected by a stiff spring. To quantify this effect, we find for each spike in one trial the closest spike in successive trials, and measure the variance of the arrival times of these spikes. Similarly, we measure the variance of the interspike times. Figure 4a shows the ratio of the interspike time variance to the sum of the arrival time variances of the spikes that make up the pair. For large separations, this ratio is unity, as expected if spikes are locked independently to the stimulus, but as the two spikes come closer, it falls below one-quarter.

Both the conditional correlation among the members of the pair (see Figure 4a) and the relative synergy (see Figure 4b) depend strongly on the interspike separation. This dependence is nearly invariant to changes in image contrast, although the spike rate and other statistical properties are strongly affected by such changes. Brief spike pairs seem to retain their identity as specially informative symbols over a range of input ensembles. If particular temporal patterns are especially informative, then we would lose information if we failed to distinguish among different patterns. Thus, there are two notions of time resolution for spike pairs: the time resolution with which the interspike time is defined and the absolute time resolution with which the event is marked. Figure 5 shows that for small interspike times, the information is much more sensitive to changes in the interspike time resolution (open symbols) than to the absolute time resolution (filled symbols). This is related to the slope in Figure 2: in regions where the slope is large, events should be finely distinguished in order to retain the information.

3.3 Implications of Synergy. The importance of spike timing in the neural code has been under debate for some time now. We believe that some issues in this debate can be clarified using a direct information-theoretic approach. Following MacKay and McCulloch (1952), we know that marking
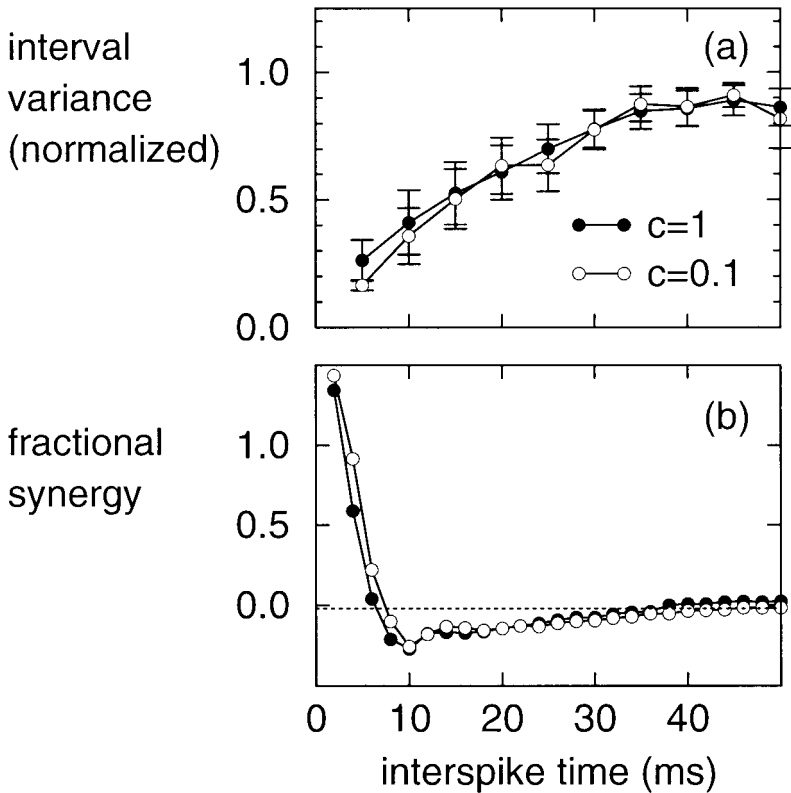
Figure 4: (a) Ratio between the variance of interspike time and the sum of variances of the two spike times. Variances are measured across repeated presentations of same stimulus, as explained in the text. This ratio is plotted as a function of the interspike time τ for two experiments with different image contrast. (b) Extra information conveyed cooperatively by pairs of spikes, expressed as a fraction of the information conveyed by the two spikes independently. While the single-spike information varies with contrast (1.5 bits per spike for $C = 0.1$ compared to 1.3 bits per spike for $C = 1$), the fractional synergy is almost contrast independent.

spike arrival times with higher resolution provides an increased capacity for information transmission. The work of Strong et al. (1998) shows that for the fly's H1 neuron, the increased capacity associated with spike timing indeed is used with nearly constant efficiency down to millisecond resolution. This efficiency can be the result of a tight locking of individual spikes to a rapidly varying stimulus, and it could also be the result of temporal patterns providing information beyond rapid rate modulations. The anal-
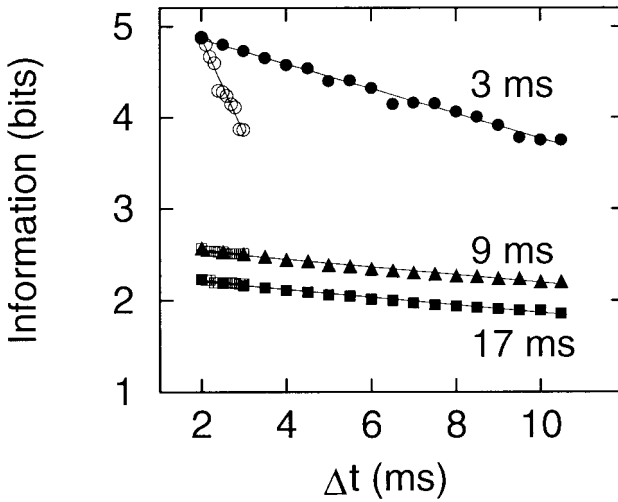
Figure 5: Information conveyed by spike pairs as a function of time resolution. An event—pair of spikes—can be described by two times: the separation between spikes (relative time) and the occurrence time of the event with respect to the stimulus (absolute time). The information carried by the pair depends on the time resolution in these two dimensions, both specified by the bin size Δt along the abscissa. Open symbols are measurements of the information for a fixed 2 ms resolution of absolute time and a variable resolution Δt of relative time. Closed symbols correspond to a fixed 2 ms resolution of relative time and a variable resolution Δt of absolute time. For short intervals, the sensitivity to coarsening of the relative time resolution is much greater than to coarsening of the absolute time resolution. In contrast, sensitivity to relative and absolute time resolutions is the same for the longer, nonsynergistic, interspike separations.

ysis given here shows that for H1, pairs of spikes can provide much more information than two individual spikes, and information transmission is much more sensitive to the relative timing of spikes than to their absolute timing. The synergy is an inherent property of particular spike pairs, as it persists when averaged over all occurrences of the pair in the experiment.

One may ask about the implication of synergy to behavior. Flies can respond to changes in a motion stimulus over a time scale of approximately 30 to 40 ms (Land & Collett, 1974). We have found that the spike train of H1 as a whole conveyed about 20 bits per second more than the information conveyed independently by single spikes (see Section 3.1). This implies that over a behavioral response time, synergistic effects provide about 0.8 bit of additional information, equivalent to almost a factor of two in resolving power for distinguishing different trajectories of motion across the visual field.

4 Summary

Information theory allows us to measure synergy and redundancy among spikes independent of the rules for translating between spikes and stimuli. In particular, this approach tests directly the idea that patterns of spikes are special events in the code, carrying more information than expected by adding the contributions from individual spikes. It is of practical importance that the formulas rely on low-order statistical measures of the neural response, and hence do not require enormous data sets to reach meaningful conclusions. The method is of general validity and is applicable to patterns of spikes across a population of neurons, as well as across time. Specific results (existence of synergy or redundancy, number of bits per spike, etc.) depend on the neural system as well as on the stimulus ensemble used in each experiment.

In our experiments on the fly visual system, we found that an event composed of a pair of spikes can carry far more than the information carried independently by its parts. Two spikes that occur in rapid succession appear to be special symbols that have an integrity beyond the locking of individual spikes to the stimulus. This is analogous to the encoding of sounds in written English: each of the compound symbols *th*, *sh*, and *ch* stands for sounds that are not decomposable into sounds represented by each of the constituent letters. For spike pairs to act effectively as special symbols, mechanisms for "reading" them must exist at subsequent levels of processing. Synaptic transmission is sensitive to interspike times in the 2–20 ms range (Magelby, 1987), and it is natural to suggest that synaptic mechanisms on this timescale play a role in such reading. Recent work on the mammalian visual system (Usrey, Reppas, & Reid, 1998) provides direct evidence that pairs of spikes close together in time can be especially efficient in driving postsynaptic neurons.

Appendix A: Relation to Previous Work

Patterns of spikes and their relation to sensory stimuli have been quantified in the past using correlation functions. The event rates that we have defined here, which are connected directly to the information carried by patterns of spikes through equation 2.5, are just properly normalized correlation functions. The event rate for pairs of spikes from two separate neurons is related to the joint poststimulus time histogram (PSTH), defined by Aertsen, Gerstein, Habib, and Palm (1989; Vaadia et al., 1995). Making this connection explicit is also an opportunity to see how the present formalism applies to events defined across two cells.

Consider two cells, A and B, generating spikes at times $\{t_i^A\}$ and $\{t_i^B\}$, respectively. It will be useful to think of the spike trains as sums of unit

impulses at the spike times,

$$\rho^A(t) = \sum_i \delta(t - t_i^A) \tag{A.1}$$

$$\rho^B(t) = \sum_i \delta(t - t_i^B). \tag{A.2}$$

Then the time-dependent spike rates for the two cells are

$$r^A(t) = \langle \rho^A(t) \rangle_{\text{trials}}, \tag{A.3}$$
$$r^B(t) = \langle \rho^B(t) \rangle_{\text{trials}}, \tag{A.4}$$

where $\langle \cdots \rangle_{\text{trials}}$ denotes an average over multiple trials in which the same time-dependent stimulus $s(t')$, or the same motor behavior observed. These spike rates are the probabilities per unit time for the occurrence of a single spike in either cell A or cell B, also called the PSTH. We can define the probability per unit time for a spike in cell A to occur at time t and a spike in cell B to occur at time t', and this will be the joint PSTH,

$$\text{JPSTH}^{AB}(t, t') = \langle \rho^A(t) \rho^B(t') \rangle_{\text{trials}}. \tag{A.5}$$

Alternatively, we can consider an event E defined by a spike in cell A at time t and a spike in cell B at time $t - \tau$, with the relative time τ measured with a precision of $\Delta \tau$. Then the rate of these events is

$$r_E(t) = \int_{-\Delta \tau/2}^{\Delta \tau/2} dt' \, \text{JPSTH}^{AB}(t, t - \tau + t') \tag{A.6}$$

$$\approx \Delta \tau \text{JPSTH}^{AB}(t, t - \tau), \tag{A.7}$$

where the last approximation is valid if our time resolution is sufficiently high. Applying our general formula for the information carried by single events, equation 2.5, the information carried by pairs of spikes from two cells can be written as an integral over diagonal "strips" of the JPSTH matrix,

$$I(E;s) = \frac{1}{T} \int_0^T dt \frac{\text{JPSTH}^{AB}(t, t - \tau)}{\langle \text{JPSTH}^{AB}(t, t - \tau) \rangle_t} \log_2 \left[\frac{\text{JPSTH}^{AB}(t, t - \tau)}{\langle \text{JPSTH}^{AB}(t, t - \tau) \rangle_t} \right], \tag{A.8}$$

where $\langle \text{JPSTH}^{AB}(t, t - \tau) \rangle_t$ is an average of the JPSTH over time; this average is equivalent to the standard correlation function between the two spike trains.

The discussion by Vaadia et al. (1995) emphasizes that modulations of the JPSTH along the diagonal strips allow correlated firing events to convey information about sensory signals or behavioral states, and this information

is quantified by equation A.8. The information carried by the individual cells is related to the corresponding integrals over spike rates, equation 2.9. The difference between the the information conveyed by the compound spiking events E and the information conveyed by spikes in the two cells independently is precisely the synergy between the two cells at the given time lag τ. For $\tau = 0$, it is the synergy—or extra information—conveyed by synchronous firing of the two cells.

We would like to connect this approach with previous work that focused on how events reduce our uncertainty about the stimulus (de Ruyter van Stevenick & Bialek, 1988). Before we observe the neural response, all we know is that stimuli are chosen from a distribution $P[s(t')]$. When we observe an event E at time t_E, this should tell us something about the stimulus in the neighborhood of this time, and this knowledge is described by the conditional distribution $P[s(t')|t_E]$. If we go back to the definition of the mutual information between responses and stimuli, we can write the average information conveyed by one event in terms of this conditional distribution,

$$
\begin{aligned}
I(E; s) &= \int Ds(t') \int dt_E P[s(t'), t_E] \log_2 \left(\frac{P[s(t'), t_E]}{P[s(t')]P[t_E]} \right) \\
&= \int dt_E P[t_E] \int Ds(t') P[s(t')|t_E] \log_2 \left(\frac{P[s(t')|t_E]}{P[s(t')]} \right).
\end{aligned}
\tag{A.9}
$$

If the system is stationary, then the coding should be invariant under time translations:

$$
P[s(t')|t_E] = P[s(t' + \Delta t')|t_E + \Delta t'].
\tag{A.10}
$$

This invariance means that the integral over stimuli in equation A.9 is independent of the event arrival time t_E, so we can simplify our expression for the information carried by a single event,

$$
\begin{aligned}
I(E; s) &= \int dt_E P[t_E] \int Ds(t') P[s(t')|t_E] \log_2 \left(\frac{P[s(t')|t_E]}{P[s(t')]} \right) \\
&= \int Ds(t') P[s(t')|t_E] \log_2 \left(\frac{P[s(t')|t_E]}{P[s(t')]} \right).
\end{aligned}
\tag{A.11}
$$

This formula was used by de Ruyter van Stevenick and Bialek (1988). To connect with our work here, we express the information in equation A.11 as an average over the stimulus,

$$
I(E; s) = \left\langle \left(\frac{P[s(t')|t_E]}{P[s(t')]} \right) \log_2 \left(\frac{P[s(t')|t_E]}{P[s(t')]} \right) \right\rangle_s.
\tag{A.12}
$$

Using Bayes' rule,

$$
\frac{P[s(t')|t_E]}{P[s(t')]} = \frac{P[t_E|s(t')]}{P[t_E]} = \frac{r_E(t_E)}{\bar{r}_E},
\tag{A.13}
$$

where the last term is a result of the distributions of event arrival times being proportional to the event rates, as defined above. Substituting this back to equation A.11, one finds equation 2.6.

Appendix B: Experimental Setup

In the experiment, we used a female blowfly, which was a first-generation offspring of a wild fly caught outside. The fly was put inside a plastic tube and immobilized with wax, with the head protruding out. The proboscis was left free so that the fly could be fed regularly with some sugar water. A small hole was cut in the back of the head, close to the midline on the right side. Through this hole, a tungsten electrode was advanced into the lobula plate. This area, which is several layers back from the compound eye, includes a group of large motion-detector neurons with wide receptive fields and strong direction selectivity. We recorded spikes extracellularly from one of these, the contralateral H1 neuron, Franceschini, Riehle & le Nestour, 1989; Hausen & Egelhaaf, 1989). The electrode was positioned such that spikes from H1 could be discriminated reliably and converted into pulses by a simple threshold discriminator. The pulses were fed into a CED 1401 interface (Cambridge Electronic Design), which digitized the spikes at 10 μs resolution. To keep exact synchrony over the duration of the experiment, the spike timing clock was derived from the same internal CED 1401 clock that defined the frame times of the visual stimulus.

The stimulus was a rigidly moving bar pattern, displayed on a Tektronix 608 high-brightness display. The radiance at average intensity $\bar{I}$ was about 180 mW/(m^2 × sr), which amounts to about 5×10^4 effectively transduced photons per photoreceptor per second (Dubs, Laughlin, & Srinivasan, 1984). The bars were oriented vertically, with intensities chosen at random to be $\bar{I}(1 \pm C)$, where C is the contrast. The distance between the fly and the screen was adjusted so that the angular subtense of a bar equaled the horizontal interommatidial angle in the stimulated part of the compound eye. This setting was found by determining the eye's spatial Nyquist frequency through the reverse reaction (Götz, 1964) in the response of the H1 neuron. For this fly, the horizontal interommatidial angle was 1.45 degrees, and the distance to the screen 105 mm. The fly viewed the display through a round 80 mm diameter diaphragm, showing approximately 30 bars. From this we estimate the number of stimulated ommatidia in the eye's hexagonal raster to be about 612.

Frames of the stimulus pattern were refreshed every 2 ms, and with each new frame, the pattern was displayed at a new position. This resulted in an apparent horizontal motion of the bar pattern, which is suitable to excite the H1 neuron. The pattern position was defined by a pseudorandom sequence, simulating independent random displacements at each time step, uniformly distributed between −0.47 degrees and +0.47 degrees (equivalent to −0.32 to +0.32 omm, horizontal ommatidial spacings). This corresponds to a dif-

fusion constant of $18.1(°)^2$ per second or $8.6\,\text{omm}^2$ per second. The sequence of pseudorandom numbers contained a repeating part and a nonrepeating part, each 10 seconds long, with the same statistical parameters. Thus, in each 20 second cycle, the fly saw a 10 second movie that it had seen 20 seconds before, followed by a 10 second movie that was generated independently.

Acknowledgments

References

Abeles, M., Bergmann, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol., 70,* 1629–1638.

Adrian, E., (1928). *The basis of sensation*. London: Christophers.

Aertsen, A., Gerstein, G., Habib, M., & Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity." *J. Neurophysiol., 61*(5), 900–917.

Bialek, W. (1990). Theoretical physics meets experimental neurobiology. In E. Jen (Ed.), *1989 Lectures in Complex Systems* (pp. 513–595). Menlo Park, CA: Addison-Wesley.

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. Lond. Ser. B, 234,* 379.

de Ruyter van Steveninck, R. R., Lewen, G., Strong, S., Koberle, R., & Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science, 275,* 275.

DeWeese, M. (1996). Optimization principles for the neural code. *Network, 7,* 325–331.

DeWeese, M., & Meister, M. (1998). How to measure information gained from one symbol. *APS Abstracts.*

Dubs, A., Laughlin, S., & Srinivasan, M. (1984). Single photon signals in fly photoreceptors and first order interneurones at behavioural threshold. *J. Physiol., 317,* 317–334.

Franceschini, N., Riehle, A., & le Nestour, A. (1989). Directionally selective motion detection by insect neurons. In D. Stavenga & R. Hardie (Eds.) *Facets of Vision*, page 360. Berlin: Springer-Verlag.

Götz, K. (1964). Optomotorische Untersuchung des Visuellen Systems einiger Augenmutanten der Fruchtfliege *Drosophila*. *Kybernetik, 2,* 77–92.

Hausen, K., & Egelhaaf, M. (1989). Neural mechanisms of visual course control in insects. In D. Stavenga & R. Hardie (Eds.), *Facets of Vision*, page 391. Springer-Verlag.

Hopfield, J. J. (1995). Pattern recognitin computation using action potential timing for stimulus representation. *Nature, 376*, 33–36.

Land, M. F., & Collett, T. S. (1974). Chasing behavior of houseflies (*Fannia caniculaus*): A description and analysis. *J. Comp. Physiol., 89*, 331–357.

MacKay, D., & McCulloch, W. S. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophys., 14*, 127–135.

Magelby, K. (1987). Short-term synaptic plasticity. In G. M. Edelman, V. E. Gall, & K. M. Cowan (Eds.), *Synaptic function* (pp. 21–56). New York: Wiley.

Meister, M. (1995). Multineuronal codes in retinal signaling. *Proc. Nat. Acad. Sci. (USA), 93*, 609–614.

Optican, L. M., & Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III: Information theoretic analysis. *J. Neurophys., 57*, 162–178.

Panzeri, S., Biella, G., Rolls, E. T., Skaggs, W. E., & Treves, A. (1996). Speed, noise, information and the graded nature of neuronal responses. *Network, 7*, 365–370.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J., 27*, 379–423, 623–656.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, 555–586.

Skaggs, W. E., McNaughton, B. L., Gothard, K. M., & Markus, E. J. (1993). An information-theoretic approach to deciphering the hippocampal code. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing, 5* (pp. 1030–1037). San Mateo, CA: Morgan Kaufmann.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., & Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett., 80*, 197–200.

Usrey, W. M., Reppas, J. B., & Reid, R. C. (1998). Paired-spike interactions and synaptic efficiency of retinal inputs to the thalamus. *Nature, 395*, 384.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neural interactions in monkey cortex in relation to behavioural events. *Nature, 373*, 515–518.