

# On the Application of Information Theory to Neural Spike Trains

S. P. Strong

*Institute for Advanced Study, Princeton, New Jersey 08540*

Rob R. de Ruyter van Steveninck, William Bialek

*NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540*

Roland Koberle

*Instituto de Física de São Carlos, Universidade de São Paulo  
C. Postal 369, 13560-250 São Carlos SP Brasil*

The nervous system represents time-dependent signals in sequences of discrete action potentials or spikes; all spikes are identical so that information is carried only in the spike arrival times. We show how to quantify this information, in bits, free from any assumptions about which features of the spike train or input waveform are most important. We apply this approach to the analysis of experiments on a variety of systems, including some where we confront severe sampling problems, and discuss some of the results obtained and hopes for future extensions.

## 1 Introduction

As you read this text, optical signals reaching your retina are encoded into sequences of identical pulses, termed action potentials or spikes, that propagate along the  $\sim 10^6$  fibers of the optic nerve from eye to brain. This spike encoding appears almost universal, occurring in animals as diverse as worms and man, and spanning all the sensory modalities<sup>1</sup>. The molecular mechanisms for the generation and propagation of action potentials are well understood<sup>2</sup>, as are the mathematical reasons for the selection of stereotyped pulses by the dynamics of the nerve cell membrane<sup>3</sup>. Less well understood is the function of these spikes as a code<sup>4</sup>: How do the sequences of spikes represent the sensory world, and how much information is conveyed in this representation?

Experiments on spiking neurons are usually analyzed either by making assumptions about which features of the spike train are significant in the code, or by making assumptions about which features of the sensory input are being encoded. These assumptions become less compelling in deeper layers of the brain, so one would like a model independent approach.

Imagine that we present a long duration sensory stimulus drawn from some particular ensemble of stimuli, perhaps the natural one. During this long period the spike train varies, and we can quantify this variability by the entropy per unit time of the spike train, or entropy rate. We can sample the spike with fixed precision  $\Delta\tau$ , recording only the number of spike in each bin

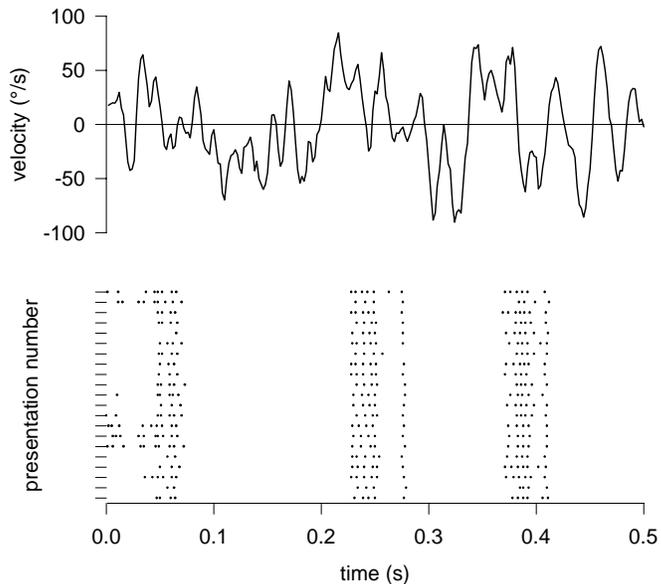


Figure 1: Angular velocity of a pattern moving across the fly's visual field produces a sequence of spikes in H1, indicated by dots. Repeated presentations produce slightly different spike sequences. For experimental methods see Ref. [13].

of this size, not their exact placement. The spike train is then converted to a sequence of integers whose entropy rate we are interested in estimating. If there is a deterministic mapping from input signals to output spike trains, then the entropy rate of this sequence is exactly the information rate, as defined by Shannon<sup>5</sup>, that the spike train, sampled at this precision, provides about the sensory stimulus. In fact repeated presentations of the same stimulus lead to different spike trains (Fig. 1). This variability can also be quantified by an entropy rate, which we call the conditional or noise entropy rate. The information that the spike train (recorded with precision  $\Delta\tau$ ) provides about the stimulus is the difference between the total spike train entropy rate and this conditional entropy rate. While both rates diverge logarithmically in the limit of  $\Delta\tau \rightarrow 0$ , the difference converges to the true, finite information rate.

The problem of measuring information transmission by spiking neurons is thus exactly the problem of estimating the entropy of integer sequences from examples. This analogy is model independent, making no assumptions about the structure of the input signal or the encoding process. These ideas allow us to quantify the response of neurons in a very general way, even when they are

responding to complex, time dependent stimuli.

## 2 The Potential Sampling Disaster

The most obvious way to estimate the entropy from examples of the configuration is to accumulate a large number of these configurations into a histogram, assume that the (normalized) count for each configuration approximates the probability of that configuration, and evaluate the entropy directly from the definition  $S(N, \Delta\tau) = -\sum_i p_i \log_2 p_i$  bits, where  $i$  indexes all possible configurations of integer counts in the  $N$  bins of size  $\Delta\tau$ .

This procedure works only if each configuration occurs many times. Imagine that we have a spike train with mean spike rate  $\bar{r} \sim 40$  spikes/s and we sample with a time resolution  $\Delta\tau = 3$  ms. The maximum entropy distribution consistent with this mean rate and a refractory period of 3 ms or more is a sequence of 0's and 1's, with the probability of a 1 being  $\bar{r}\Delta\tau = 0.12$ . The entropy of the configurations in a window of size  $T = N\Delta\tau = 99$  ms is

$$\begin{aligned} S_{\max}(N, \Delta\tau) &= N[-(\bar{r}\Delta\tau) \log_2(\bar{r}\Delta\tau) - (1 - \bar{r}\Delta\tau) \log_2(1 - \bar{r}\Delta\tau)] \quad (1) \\ &\sim 17.8 \text{ bits.} \end{aligned}$$

Naively, then, we need to see more than  $2^S \sim 2 \times 10^5$  examples of the 99 ms windows. If these windows have to be non-overlapping, then we need more than three hours of data, and one might think that we need *much* more data than this to insure that the probability of occupying each bin is estimated with reasonable accuracy. Such large quantities of data are generally inaccessible for experiments on real neurons.

Here we report that it is possible to make progress despite these pessimistic estimates. There are several ingredients: First, we examine explicitly the dependence of our entropy estimates on the size of the data set and find regular behaviors that can be extrapolated to the infinite data limit. Second, generalizing the suggestion by Ma<sup>8</sup>, we evaluate lower bounds to the entropy that are much less prone to sampling errors. Third, we are interested in the extensive component of the entropy in large time windows, and we find that a clean approach to extensivity is often visible before sampling problems set in. Finally, we have developed bounding procedures which are practical even in the limit of extreme undersampling of the conditional entropy (as few as two repetitions of only a few tens of seconds in length may suffice).

### 3 Direct Estimation

We begin our analysis with data taken from the H1 neuron of the common blowfly<sup>12</sup> for the case where spike times have been recorded with a precision of  $\Delta\tau = 3$  ms. For a window of  $T = 30$  ms—corresponding roughly to the behavioral response time of the fly<sup>9</sup>—we can estimate the entropy rather accurately by the naive procedure described above. The resulting entropies as a function of the amount of data included in the analysis vary only slightly (the finite data set corrections are less than a part in  $10^{-3}$ ). The dependence on system size is well fit by  $S_{\text{naive}} = S_0 + S_1/\text{size} + S_2/\text{size}^2$  and, under these conditions, we feel confident that  $S_0$  is the correct entropy for this combination of  $T$  and  $\Delta\tau$ . This procedure is applicable to any preparation where sufficient data can be gathered to allow for direct estimation of the entropies out beyond the time scale of significant correlations in the spike trains. We have used it successfully in other systems including the cricket cercal system and even the MT visual area of the macaque monkey (in a reanalysis of the data of Ref. <sup>7</sup>). However, even in the case of the fly where large amounts of data are available, if we extend the naive procedure to windows of  $T = 200$  ms, finite size corrections become large, the contribution of the second correction is significant and the extrapolation to infinite size is unstable. Fortunately, there are ways of dealing with this difficulty.

### 4 A Useful Upper Bound

First, the entropies estimated using windows of lengths short enough that direct estimation is possible already provide upper bounds to the total and conditional entropy rates. The simplest upper bound to the entropy rate can be obtained from dividing the entropy of the distribution of integer strings of size  $N$  by  $N\Delta\tau$ . Since this neglects all correlations between the different windows, it always overestimates the entropy and provides a crude upper bound. It can be substantially improved by utilizing an argument of Shannon's about compressing symbol sequences using the optimal predictor which utilizes only the previous  $M$  symbols. Imagine accumulating the probabilities of all the integer sequences seen in size  $M + 1$  bin windows. Now, given the contents of the previous  $M$  bins, one can say what the most likely integer for the next bin is, as well as the next most likely and so on. Following Shannon<sup>5</sup>, consider a compression scheme where the original series of integers is replaced by a series of integers that lists the rank of the actual integer in each bin, e.g. a 0 if the most probable integer occurred, a 1 is the next most probable, etc. It is clear that this procedure is invertible (modulo a finite initial

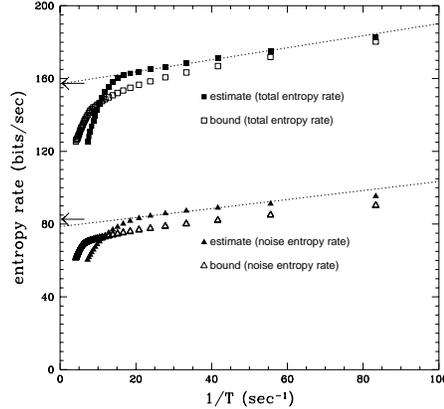


Figure 2: The total and noise entropies per unit time (in bits per second) are plotted versus the reciprocal of the window size (in  $\text{s}^{-1}$ ), with the time resolution held fixed at  $\Delta\tau = 3$  ms. Results are given both for the direct estimate and for the bounding procedure described in the text, and for each data point we apply the extrapolation procedures of Fig. 2. As discussed in the text, the plot is linear for window sizes that are not too long, where undersampling is severe, and not too short, a regime where correlations spoil extensivity. The linearity allows us to extrapolate to the entropy per unit time for words of infinite length. Arrows indicate upper bounds obtained by differentiating  $S(T)$ , as explained in the text. The information transmitted per unit time about the stimulus is the extrapolated total entropy per unit time minus the extrapolated noise entropy per unit time.

specification of the first  $M$  bins) so the new series has the same entropy rate as the original. We can now find an upper bound to its entropy rate, and thus to the true entropy rate, by neglecting all correlations between the integers in the new stream and estimating the entropy as:  $-(1/\Delta\tau) \sum_i p_i \log_2(p_i)$ , where  $p_i$  is the probability that a randomly chosen bin contained the integer  $i$ . The resulting upper bound is nothing other than  $1/(\Delta\tau)$  times the difference between the entropy estimate formed using windows of size  $M\Delta\tau$  and windows of size  $(M+1)\Delta\tau$ . This estimate of the entropy includes all correlations out to the time scale  $T = M\Delta\tau$ , and would be exact if there were no longer range correlations; it provides an upper bound to the entropy rate in any case and the same bounding procedure can be applied with a trivial generalization to the noise entropy.

For the case of the total entropy for H1, the bound becomes progressively tighter at larger  $M$ , until sampling problems set in. There is a broad plateau ( $\pm 2.7\%$ ) in the range  $18 < M\Delta\tau < 60$  ms, leading to the arrow shown in Fig. 2 at 157 bits/s, in excellent agreement with the extrapolation

## 5 A Useful Lower Bound

Ma<sup>8</sup> discussed the problem of entropy estimation in the undersampled limit. For probability distributions that are uniform on a set of  $N$  bins (as in the microcanonical ensemble), the entropy is  $\ln N$  and the problem is to estimate  $N$ . Ma noted that this could be done by counting the number of times that two randomly chosen observations yield the same configuration, since the probability of such a coincidence is  $1/N$ . In the undersampled limit, Ma's procedure is actually the optimal estimator of the entropy in the least-squares sense if all values of  $N$  are equally likely *a priori*. More generally, for non-uniform distributions the probability of a coincidence is  $P_c = \sum p_i^2$ , and it is clear that

$$\begin{aligned} S(N, \Delta\tau) &= -\sum p_i \log_2 p_i = -\langle \log_2 p_i \rangle \\ &\geq -\log_2 (\langle p_i \rangle) = -\log_2 P_c, \end{aligned} \quad (2)$$

so we can compute a lower bound to the the entropy by counting coincidences. This is the minimum entropy consistent with a given  $P_c$ , and it is one of the Renyi entropies<sup>10</sup>. It is also at the heart of algorithms for the analysis of attractors in dynamical systems<sup>11</sup>.

The bound in Eq. (2) is tightest for distributions that are close to uniform. The distributions of spike sequences cannot be uniform because the spikes are sparse. But the distribution of sequences with fixed spike count  $N_{\text{sp}}$  could be more nearly uniform, and so we apply the Ma bounding procedure independently in each  $N_{\text{sp}}$  sector. Thus we obtain a lower bound to the entropy,

$$\begin{aligned} S_{\text{Ma}} &= -\sum_{N_{\text{sp}}} P(N_{\text{sp}}) \\ &\quad \times \log_2 \left[ P(N_{\text{sp}}) \frac{2n_c(N_{\text{sp}})}{N_{\text{obs}}(N_{\text{sp}})[N_{\text{obs}}(N_{\text{sp}}) - 1]} \right], \end{aligned} \quad (3)$$

where  $n_c(N_{\text{sp}})$  is the number of coincidences observed among the words with  $N_{\text{sp}}$  spikes,  $N_{\text{obs}}(N_{\text{sp}})$  is the total number of occurrences of words with  $N_{\text{sp}}$  spikes, and  $P(N_{\text{sp}})$  is the fraction of words with  $N_{\text{sp}}$  spikes. Here and hereafter we drop the reference to the implicit dependence on  $N$  and  $\Delta\tau$  for all entropies, except when necessary for clarity.

The bound  $S_{\text{Ma}}$  suffers from two types of sampling errors. First, there are those associated with the estimate of  $P(N_{\text{sp}})$ . A slight generalization of the analysis of Ref. <sup>6</sup> shows that the leading error in the entropy computation for each  $N_{\text{sp}}$  is simply  $-[N_{\text{obs}}(N_{\text{sp}}) \ln 2]^{-1}$ . The second type of errors are associated with estimating  $\sum p_i^2$  from the observed number of coincidences

$n_c(N_{\text{sp}})$ . This produces an error in the contribution to the entropy at each  $N_{\text{sp}}$ ,

$$\Delta S_{\text{Ma}} = \frac{2}{N_{\text{obs}}(N_{\text{sp}})} \frac{\sum' p_i^3 - (\sum' p_i^2)^2}{(\ln 2)(\sum' p_i^2)^2} + \dots, \quad (4)$$

where the sum is taken only over words with  $N_{\text{sp}}$  spikes. This is again  $O(N_{\text{obs}}^{-1})$  and exponentially smaller than the sampling errors expected for the naive estimator. For the case of uniform probability distributions at fixed spike number, even this leading term vanishes.

In accord with these arguments, we found a very weak dependence of the Ma bound on the size of the data set. Furthermore, the error bars which we estimate by subdividing the data are also extremely small. These results give us confidence that the procedure in Eq. (3) gives a reliable lower bound to the entropy. In Fig. 2 we finally plot the total entropy as a function of the window size  $T$  for the H1 experiment, with results from both the naive procedure and from the Ma bound. For sufficiently large windows the naive procedure gives an answer smaller than the Ma bound, and hence the naive estimate must be experiencing sampling problems. Before this sampling disaster the lower bound and the naive estimate are never more than 10–15% apart. As with the upper bound, the same procedure can be applied with a trivial generalization to the noise entropy, and again the bound is always within 10% of the direct estimate, prior to the onset of sampling problems.

The point at which the naive estimate crashes into the Ma bound is also the point at which the second correction,  $S_2$ , becomes significant and we lose control over the extrapolation to the infinite data limit. This point occurs at a window size of roughly 100 ms. The sudden transition from accurate estimation to a crude lower bound occurs because the number of possible spike trains is growing exponentially with window size, and hence we are either very well sampled or disastrously undersampled. Conversely, the critical window size is only logarithmically dependent on the size of the data set, so we do not sacrifice much if we have more limited duration recordings.

## 6 Extrapolation to Infinite Window Size

If the correlations in the spike train have finite range, then the leading sub-extensive contribution to the entropy will be a constant. Thus, if we plot the entropy per unit time as a function of inverse window size, we should see a linear behavior which can be extrapolated to the infinite time limit. This is seen clearly in Fig. 2, and emerges before the sampling disaster. Indeed the break away from linear behavior is close to the point where the naive estimate

crashes into the Ma bound and  $S_2$  becomes large, confirming our identification of this point. Given the clean linear behavior in a well sampled region of the plot, we trust the extrapolation and arrive at an estimate of the entropy per unit time as seen with 3 ms resolution,  $157 \pm 3$  bits/s. This agrees well with the estimate based on the Shannon predictions scheme, which we argued should give a tight bound on the total entropy. Similar extrapolation is possible for the noise entropy, as shown in Figure 2. The difference between the two entropies is the information which the cell transmits,  $R_{\text{info}} = 78 \pm 5$  bits/s, or  $1.8 \pm 0.1$  bits/spike at a resolution  $\Delta\tau = 3$  ms.

## 7 The Severly Undersampled Limit for the Noise Entropy

In many preparations, the hundreds of repetitions of the stimulus possible for  $H1$  or even the tens of repetitions possible in the macaque visual cortex, may simply be out of reach. For example, we have analyzed data taken from human skin stretch receptors<sup>13</sup> where, frequently, only two repetitions of a dynamic stimulus of length 30 seconds were possible before the extracellular electrode shifted. In this case, the techniques outlined previously are wholly inadequate for estimating the conditional entropy, and hence the information. Fortunately, useful bounds may still be obtained in these circumstances.

First, we discuss a useful lower bound on the information from a small number of presentations of the stimulus. Imagine that one has recorded the spike trains (as before with precision  $\Delta\tau$ ) evoked by  $n$  repetitions of some relatively long dynamical stimulus. The spike trains should have no correlations with each other, other than those induced by the stimulus, since they are from different trials, well separated in time. In this case, the information,  $I_{n-1}$ , that  $n - 1$  spike trains supply about another spike train, must be *less than or equal to the information that one spike train supplies about the stimulus*,  $I_{\text{true}}$ :  $I_{n-1} < I_{\text{true}}$ . The intra-train information rate, given by the difference in the entropy rates of the  $n$  trains taken together ( $S_n$ ) and the sum of the entropy rate for  $n - 1$  trains ( $S_{n-1}$ ) and a single train ( $S_1$ ), is therefore a lower bound on the information rate of the spike train about the stimulus:

$$I_{\text{true}} > I_{n-1} = S_1 + S_{n-1} - S_n \quad (5)$$

Further, in principle, this bound converges to the true mutual information for large  $n$ :

$$\lim_{n \rightarrow \infty} I_{n-1} = I_{\text{true}} \quad (6)$$

It also provides the optimal lower bound to the information rate that is obtainable from  $n$  repetitions *without any assumptions about the encoding of the*

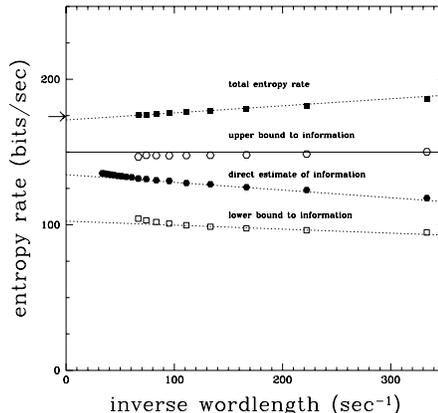


Figure 3: Results obtained from recordings in human skin stretch receptors. Top points (filled squares) are naive estimates of the total entropy rate extrapolated to infinite data. The arrow denotes the Shannon upper bound on the total entropy rate inferred from these data. Open hexagons are upper bound to information transmitted obtained from total entropy as above and lower bound to conditional entropy discussed in the text (only two repetitions were used). Solid line lies just above any of the estimated upper bounds and we therefore take it (150 bits/sec) as a true upper bound of the transmitted information. Filled hexagons are direct estimates of the transmitted information from the total and conditional entropies (extrapolated to infinite data). This was possible for this data set because 12 repetitions were recorded and some estimate of the conditional entropy could be made. Extrapolated estimate of the true information was 134.5 bits/sec. Open squares are the mutual information between the spike trains of repetitions 1 and 2, estimated for various window sizes. This gives a lower bound to the information the spike trains carry about the stimulus (see text), based on only two repetitions, of 102.5 bits/sec. Dashed lines are guides to the eye for extrapolating to the limit of infinite window size, assuming the approach to extensivity of the entropy discussed in the text (see also Fig. 3).

*information.* Practically, sampling problems occur when we try to estimate the total entropy rate for more than a handful of spike trains, even using all of the foregoing strategies. This limits utility of the intra-spike train information to small  $n$ , and for larger  $n$  the techniques described previously are superior. However, when few repetitions are available, the intra-spike train information can be quite useful. An example of results obtained for only two repetitions is shown in Fig. 3:

The naive estimates of entropy rates for both a single spike train,  $n$  spike trains and  $n - 1$  spike trains are all expected to have the form of true entropy rate plus constant times the inverse window size. Therefore, the naive estimate of the intra-spike train information,  $I_{n-1}^{\text{naive}}(N, \Delta\tau)$ , defined as:

$$I_{n-1}^{\text{naive}}(N, \Delta\tau) = \frac{1}{N\Delta\tau} (S_1(N, \Delta\tau) + S_{n-1}(N, \Delta\tau) - S_n(N, \Delta\tau)) \quad (7)$$

should have the form for large window sizes:  $I_{\text{true}}^{\text{intra}} + C_1/N + \text{exponentially small}$ , where  $I_{\text{true}}^{\text{intra}}$  is the true, intra-spike train information rate. Therefore, extrapolation to a true lower bound on the rate of transmitted information is, in general, practical, as shown in Fig. 4.

Also, shown in Figure 4 are upper bounds to the true information rate for a range window sizes. These are obtained from the naive estimate of the total entropy rate and a lower bound on the conditional noise entropy obtained at each window size. The lower bound is obtained from only two repetition of the stimulus by noting that:

$$\begin{aligned}
 S_{\text{noise}}(N, \Delta\tau) &= \langle -\sum_i p_i \log_2(p_i) \rangle_{\text{position in stimulus}} & (8) \\
 &= -\langle \log_2(p_i) \rangle_{\text{configurations}} \rangle_{\text{position in stimulus}} \\
 &> -\langle \log_2(\langle p_i \rangle_{\text{configurations}}) \rangle_{\text{position in stimulus}} \\
 &> -\langle \log_2(\langle \langle p_i \rangle_{\text{configurations}} \rangle_{\text{position within stimulus subset}}) \rangle_{\text{subsets of stimulus}} \\
 &> -\langle \log_2 P_c^{\text{subset}} \rangle_{\text{subsets}}
 \end{aligned}$$

where  $P_c$  is the probability that identical configurations were observed for the two spike trains on the two trials, given that the time window looked at fell within the given subset of the stimulus. Thus we need only specify a way of dividing the spike train up into subsets and then estimate  $P_c$  for each subset. The bound will be tighter the more for more uniform  $\langle \langle p_i \rangle_{\text{configurations}} \rangle$  within a given subset, and we therefore divided the stimulus into subsets where the windows in a subset all had the same number of spikes in the first trial. More restricted subsets yield a tighter bound, but cause more sampling problems, until, in the limiting case, each window is its own subset and the above is equivalent to the lower bound on the conditional entropy discussed previously.

## 8 Discussion

The temporal sequence of spikes provides a large capacity for transmitting information, as emphasized by MacKay and McCulloch nearly 45 years ago<sup>14</sup>. One central question in studies of the brain is whether this large capacity is used, or whether the variations in spike timing represent noise which must be averaged away<sup>15</sup>. We suggest that the proper quantitative formulation of this question is the comparison of information transmission rates (in bits/s) with the total entropy of the spike train (also in bits/s) as a function of the time resolution  $\Delta\tau$ <sup>16</sup>. Further, much of the current debate is focused on cells in the central nervous system<sup>15</sup>, where assumptions about what is being encoded should be viewed with caution. It therefore exceedingly important

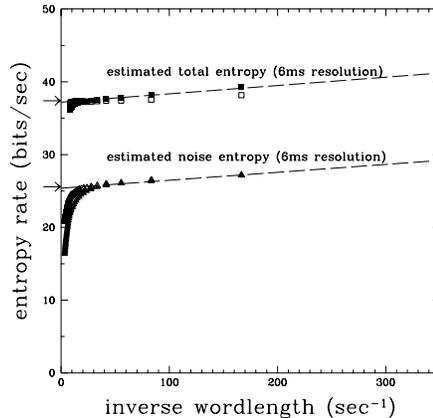


Figure 4: As for Fig. 3, but here analyzed data are those from MT area of macaque visual cortex (Ref. [7]) with timing precisions of 6 msec.

to have techniques for estimating quantities of fundamental importance, such as the Shannon information in the spike trains, which are applicable without assumptions. We believe that the analysis techniques discussed here are of nearly universal applicability, and in support of this we display in Fig. 4 results obtained for the information rate of a neuron in the motion sensitive region MT of the macaque monkey visual cortex.

At least for the H1 neuron, we believe that our results are sufficient to demonstrate the efficiency of the neural code for temporal resolutions approaching a millisecond. In support of this, we note that we have repeated the analysis of entropy and information at several different values of the time resolution  $\Delta\tau$ , from 0.7 ms to 800 ms. The information rate  $R_{\text{info}}$  has a very linear dependence on the log of the time resolution throughout much of this range; loss of one bit of precision in the specification of spike times ( $\Delta\tau \rightarrow 2\Delta\tau$ ) causes a  $\sim 10$  bit/s drop in the information rate.

The methods here are thus sufficient to demonstrate the importance of spike timing in this system down to the millisecond level. However, more importantly, information theory provides a general paradigm for establishing which features of the spike train are important for information transmission. To test that some feature of real neural spike trains is important for information transmission, one need only compare the information carried by the actual spike train with that carried by a modified version of it where the feature in

question has been degraded. Studying the information loss associated with coarsening timing precision is a specific example of this which demonstrates the importance of millisecond timing precision in H1. We have also undertaken analysis using information theory to demonstrate the importance of the super-Poisson reproducibility of spike counts in the H1 neuron<sup>17</sup>, and the road is open to a huge range of further applications of information theory in the study of neural spike data.

1. E. D. Adrian, *The Basis of Sensation* (W. W. Norton, New York, 1928).
2. D. J. Aidley, *The Physiology of Excitable Cells, 3d ed.* (Cambridge University Press, Cambridge, 1989)
3. A. L. Hodgkin and A. F. Huxley, *J. Physiol.* **117**, 500 (1952); D. G. Aronson and H. F. Weinberger, *Adv. Math.* **30**, 33 (1978).
4. For a review see F. Rieke, D. Warland, R. R. de Ruyter van Steveninck and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge MA, in press).
5. C. E. Shannon and W. Weaver (1949). *The Mathematical Theory of Communication* (University of Illinois Press, Urbana).
6. A. Treves and S. Panzeri, *Neural Comp.* **7**, 399 (1995).
7. K. H. Britten, M. N. Shadlen, W. T. Newsome and J. A. Movshon, *Visual Neuroscience* **10**, 1157 (1993).
8. S. K. Ma, *J. Stat. Phys.* **26**, 221 (1981).
9. M. F. Land and T. S. Collett, *J. Comp. Physiol.* **89**, 331 (1974).
10. A. Renyi, *Probability Theory* (North Holland, Amsterdam, 1970).
11. D. Ruelle, *Chaotic Evolution and Strange Attractors* (Cambridge University Press, Cambridge UK, 1989).
12. R. R. de Ruyter van Steveninck and W. Bialek, *Proc. R. Soc. Lond. Ser. B.* **234**, 379 (1988).
13. Benoni Eden, personal communication.
14. D. MacKay and W. S. McCulloch, *Bull. Math. Biophys.* **14**, 127 (1952).
15. M. N. Shadlen and W. T. Newsome, *Curr. Opin. Neurobiol.* **5**, 248 (1995) and W. R. Softky, *ibid.* **5**, 239 (1995); T. J. Sejnowski, *Nature* **376**, 21 (1995); D. Ferster and N. Spruston, *Science* **270**, 756 (1995); C. F. Stevens and A. Zador, *Curr. Biol.* **5**, 1370 (1995).
16. F. Rieke, D. Warland, and W. Bialek, *Europhys. Lett.* **22**, 151 (1993).
17. Rob R. de Ruyter van Steveninck, Geoffrey D. Lewen, S. P. Strong, R. Koberle and William Bialek, *Science* **275**, 1805 (1997).