Use of information theory in studies of efficiency of cortical coding of natural stimuli

— term paper for Ph/Bi 103 —

Daniel Wagenaar, March 2000

1 Introduction

How does the nervous system encode sensory information? This question has been at the core of neurscience ever since its conception in the early 1900's. It still hasn't been answered to any satisfactory extent, but some promising insights have been gained, not least due to the recent joined efforts between biologists and theorists in statistics and other fields. This paper relates some of these efforts.

Before getting into details, this first section provides a very brief overview of some of the techniques used traditionally to probe neuronal coding. A more in-depth treatment, with pointers to recent literature is given by Buračas et al. in [1].

1.1 Variance of spike counts

Perhaps the simplest method to examine what external stimuli particular neurons code for, is to measure their firing rates after presentation of various stimuli. One can then find which of the stimuli presented yields the largest spike count in a given time window. Various statistical tests are available[1] to gauge the significance of any difference in the rates obtained.

Simplicity comes at a price though. Firstly, by looking at first order statistics (counts) only, one immediately loses sight of any temporal structure. Secondly, since spike counting implicitly presupposes that the firing is a Poisson process, it is generally necessary to have long samples or to average over large number of trials in order to have acceptable standard errors on the counts.

A useful extension of this technique is to compute the covariance of spike counts from two neurons, rather than relating the spike count from one neuron to exernal stimuli. The same limitations apply, and choosing the width of time windos becomes a more tricky issue.

1.2 Temporal (cross-)correlation

The first step into gaining insight into the temporal structure of a process is usually obtaining temporal correlograms. One can plot a neuron's autocorrelation function, i.e. a histogram of its

---

[1] For an excellent introduction to biological applications of statistics, see [2].

inter-spike intervals, or one can plot the cross-correlogram of two neurons, i.e. the distribution of intervals between a spike on neuron *A* and a spike on neuron *B*.

This technique quickly tells you whether the behaviour of the two neurons is correlated, and in many cases whether there is a causal relationship between their firings. Again, the choice of time bin width is crucial, and the technique only works if the relationship consists of directly induced firing: if it is effectively a pattern from the first neuron that triggers a pattern in the second neuron, this will not show up clearly in the cross-correlogram.

1.3 Mutual information between spike trains

The mutual information between two data streams is a completely general measure of their relatedness, which doesn't rely on the coding mechanisms used by the two streams. The definition and some of the mathematical properties of mutual information will be detailed in the next section.

The coding scheme independence makes mutual information well suited for studying relations between pairs of neurons when one of the neurons is hypothesized to be far downstream from the other, or for studying relations between a neuron's firing pattern and a stream of external stimuli with complex structure.

One important limitation of the concept is that it does not deal with time delays between the two data streams. When it is suspected that one neuron may react to another neuron or to a stimulus after a fixed delay, it is possible to manually input this delay into the algorithm, or to determine the delay by maximizing the mutual information with respect to the delay. If the delay is variable, the method breaks down, except in simplifying cases.

2 Some information theory

This section will present a mathematical definition of mutual information and some related concepts necessary to understand the definition. This presentation aims for simplicity rather than full mathematical formality. Most of the following material is closely based upon an introductory text on information theory aimed at applications in artificial neuron network research [3].

Formally, information theory considers data streams consisting of long sequences of messages. Each message is a word x taken from a dictionary X, with a probability p_x. It is not required that subsequent messages are wholly independent, as long as in the long run the probabilities p_x are adhered to.

2.1 Entropy

The entropy for a data stream (or more strictly: for a dictionary) is defined as:

$$H(X) = -\sum_x p_x \log_2 p_x.$$

Apart from a numerical factor, this definition coincides with the definition of entropy in statistical physics.

In the limit of infinitely long streams, the entropy becomes equal to the average length (in bits) of the shortest possible decodable code that could encode the messages in the stream. Thus, entropy can be viewed as a measure of the uncompressibility of a data stream. Entropy is not a very useful measure of the information contents of a stream, since incompressibility, and hence entropy, is highest for a totally random stream.

The following are some key properties of the entropy:

1. For any dictionary X, the entropy is non-negative: $H(X) \geq 0$, with equality iff there there is only one word in X.
2. An upper bound on entropy is given by $H(X) \leq \log_2 N$, with equality iff all words have equal probability.
3. The entropy of a derived data stream Y, whose messages y are obtained from the data stream X as $y = f(x)$ for some (deterministic) function f cannot be higher than the original entropy: $H(f(X)) \leq H(X)$, with equality iff f is injective (invertible).

One very important issue when computing the entropy of a stream, is to decide what the length of the individual messages is[1]. For example, with spike trains as a data stream, one could pick the individual interspike intervals as messages, or multiplets of such intervals. The resulting entropy will be different. This is a strong limitation on the claim of generality made above. In effect, some external judgement has to be applied to decide what constitutes a message, although in the next section it will be shown that there may be algorithmic solutions to this problem too.

2.2 Mutual information

Much more interesting than the incompressibility of a data stream, is the amount of information this stream supplies about another stream. This quantity is measured by the mutual information,

---

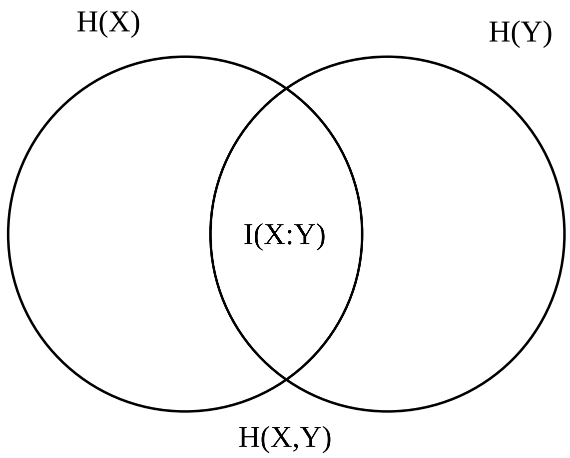
[1] Information theory is not an experimental science, warn Rieke et al. ([4] p.148): just as there is no direct method to measure the entropy of a gas, there is no direct method to measure the entropy of a real datastream — the only instance one can obtain the exact entropy is when the full probability distribution of messages in the stream is known, which obviously requires that one knows what constitutes a message. In physics, entropy change can be measured by observing heat flow, but there is no known analogue to heat flow in information theory.

defined for two streams X and Y by

$$I(X:Y) = H(X) + H(Y) - H(X,Y).$$

Here $H(X)$ and $H(Y)$ are the entropies of the individual stream, and $H(X,Y)$ is the entropy of the joint stream defined as $H(X,Y) = -\sum_{x,y} p_{x,y} \log_2 p_{x,y}$, where $p_{x,y}$ is the probability of randomly picking the pair (x,y) when picking x from X and y from Y. Explicitly, the information can be written as $I(X:Y) = \sum_{x,y} p_x p_y \log \frac{p_{x,y}}{p_x p_y}$.

If the above seems a strange definition, consider the following analogy: suppose two people standing close together look at the world through monoculars. Each person will see a small area of the environment: one sees a circle $H(X)$, the other sees a circle $H(Y)$. In general, these circles may overlap. If that is the case, then Xanthippe can predict a part $I(X:Y)$ of what Yo-anne will see.



Some important properties of the mutual information are:

1. Mutual information is non-negative: $I(X,Y) \geq 0$, with equality iff the distributions X and Y are independent.

2. The mutual information $I(X:Y)$ cannot exceed the entropy $H(X)$ or $H(Y)$.

3. The mutual information between a derived data stream $f_n(X)$ and Y cannot exceed the mutual information between X and Y:

   $$I(f_\eta(X):Y) \leq I(X:Y),$$

   where f_η is any deterministic or stochastic function.

This last property states the intuitive truth that if you tell me a story about a friend, and I relate that story to another friend, the story I tell will contain less (accurate) information about your friend then your story did. This property is very useful, because it states that by making simplified assumptions about what quantity actually carries the information in a spike train, you can never *over*estimate the mutual information, and similarly, if you make simplified assumptions as to what your stimulus is like, you will derive an underestimate of the mutual information.

3 The use of mutual information in neurobiology

3.1 Foundations

Before using information theory in neurobiology, we have to ascertain that it applies:

> **Supposition**: Organisms care about how much information passes through a channel.

The value of this ‘optimal use of resources’ argument can be hotly debated. To me it seems that one has to make some assumptions in order to get anywhere, and information flow seems a more relevant quantity (being a general property of data processing systems) than spike count (which is an accident of the particular hardware). I will return to this question in a later section.

3.2 Quantizing the neuronal data stream

Information theory grew out of an engineering problem: how to efficiently encode a given set of messages into a channel with limited capabilities. In neurobiology one does not generally know what messages are being transmitted by a neuron. Instead, one has to decide heuristally what amount of data to treat as a message for computing information theoretical quantities. This makes estimating mutual information a bit of an art. In practice researchers often associate words to spike trains of length on the order of 100 ms, and split these words into ‘letters’ corresponding to the number of spikes in subtrains of length around 20 ms. Both arguments of biological plausibility and requirements of statistics are used to arrive at these numbers. They can be varied at will, as long one makes sure that enough statistics remain to closely approximate the probability distribution in the reduced space of letters and words (as opposed to the full space of possible unquantized spike trains). If these constraints are satisfied, properties 3 of the entropy and the mutual information guarantee that the number obtained from these calculations is a lower bound, and one can safely adjust the parameters such as to obtain the highest possible lower bound[1].

---

[1] Even in the case of undersampled distributions some useful bounds can be obtained [5].

3.3 Components of entropy

The choices to be made for the information theoretical analysis of spike data do not end here. One has to choose what is going to be considered signal and what is noise. The entropy of a spike train is the sum of four components:

1. Mutual information with feature of stimulus under investigation
2. Mutual information with other features of stimulus
3. Mutual information with local events (not stimulus related)
4. Intrinsic noise entropy (e.g. limitations of synaptic efficacy)

Component 1 will always be viewed as a signal, and component 4 is always noise, but components 2 and 3 can be classified as either.

3.4 Estimators of mutual information

As there is no way to directly measure mutual information in experimental situations, several estimator algorithms have been devised. Here, I introduce two, which are used in experiments discussed in the next section.

3.4.1 Linear signal reconstruction à la Bialek

The first method [4] is based on the observation that the information between a signal and a spike train can never be less than the information between a signal and a reconstruction of that signal solely based on the spike train. From an observed spike train

$$x(t) = \sum_i \delta(t - t_i),$$

where t_i are the spike arrival times, the method tries to estimate the signal $s(t)$ via a Wiener kernel K_1:

$$s_{\text{est}}(t) = \sum_i K_1(t - t_i).$$

This kernel is found by minimizing some error functional such as:

$$E[s; s_{\text{est}}] = \left\langle \int dt \left(s(t) - s_{\text{est}}(t) \right)^2 \right\rangle,$$

where the average is over the set of all stimuli presented.

A considerable advantage of this method is that it solves the problem of time delay estimation naturally, and that it doesn't depend on choosing timewindows and bins at the level of the spike train. The disadvantage is that kernel estimation is rather involved.

SKRvSB98 fig 3 here

Figure 1: Total and noise entropy as estimated using different time windows (window size *increases* to the *left*). Filled squares are obtained from a method estimating finite data set corrections (which are found to be $< 10^{-3}$), open squares represent the Ma bound.

3.4.2 Direct method à la Strong et al.

An alternative method is pioneered in [6]. In this method the information is computed as the difference between an estimate of the total entropy in the spike train and an estimate of the 'noise' entropy in the spike train:

$$I(X:S) = H(X) - H(X|S) =: H_{\text{total}}(X) - H_{\text{noise}}(X),$$

where X represents the spike train codes at a particular point in time, and S represents the stimilus. Note that the rightmost member of this equation does not refer to the stimulus explicitly, so again time delay problems are resolved naturally. The solution is less clean though than for the reconstruction method: it requires that a stimulus of length (much) longer than the delay be presented multiple times. Strong et al. [6] describe a method to estimate the entropy and information per unit time in the limit of infinite time windows based on calculating total and noise entropies $H_{\text{total}}(X)$ and $H_{\text{noise}}(X)$ for a series of finite windows (Figure 1). The method assumes that there are no (detailed) long time correlations within spike trains.

The stimuli used in this study were pseudo random time series of vertical gratings displayed on an oscilloscope, varying at a time scale of 500 μs, much less than the typical integration time of the neural system under consideration. The stimulus pattern repeated every 33 s, and the total length of the experiment was 100 cycles [7].

4 Stimulus representation in primate visual cortex

In a famous series of articles [8, 9, 10], Richmond et al. investigated how visual stimuli are represented in neuronal firing patterns in monkey inferior temporal cortex. They used carefully selected stimuli with fixed overall intensity which are orthogonal in bitmap space. These stimuli were shown for 800 ms, during and after which neuronal response was recorded. They find that the response to a particular stimulus is highly consistent across repeated exposure. The time structure of the responses is categorized using principal component analysis [11]. For all neurons investigated most of the variance of the responses was accounted for by the first few principal components: the number of significant components was typically 2, 3 or 4.

Finally, the authors compute the information between the spike train and the stimulus in two ways: first by binning the spike counts, and secondly by binning PCA coefficients. They find that individual cells transmit a relatively low amount of information about each stimulus: $I_c = .426 \pm .040$ for count based calculations, $I_3 = .883 \pm .054$ for PCA (first 3 components only) based calculations. Since I_3 is much higher than I_c, the authors conclude that the temporal patterns of the responses carry information beyond what is encoded simply in the count. Furthermore, the coefficients of the PCs were uncorrelated, and considering more than one component made for a larger estimate of I: $I_1 = .403 \pm .039$; $I_2 = .642 \pm .045$. Thus the authors conclude that the neurons encode information about more than one aspect of the stimuli.

Note that the information is by no means high enough for single cells to identify the stimulus. Even with throughput as high as I_3, an ensemble of at least some 7 or 8 neurons would be needed to identify a stimulus. The authors do not however mention information rates obtained for such ensembles of cells.

The amount of per-stimulus information (Box 1) varied considerably across stimuli. This indicates that the response to some stimuli is more stable than for others. The authors do not elaborate this point. The conceptual difficulty associated with $I(R : s)$ is well illustrated by comparing this per-stimulus information with the average spike count $C(s)$ observed after presentation of the same stimulus (Figure 2). The authors claim that these numbers are essentially uncorrelated, but it is striking that the highest information appears to be transferred about the stimulus that causes the lowest firing rate. This can be explained as follows: when the average spike count is low in a set of spike trains R', the entropy in that set is also low if the trains obey anything resembling a modulated Poisson process. Since $I(R : s)$ is computed by subtracting the entropy in the set of spike trains observed after presenting stimulus s from the total entropy, $I(R : s)$ will be artificially high in this situation.

Box 1. Information per stimulus

An alternative way to write the mutual information between two data streams is

$$I(X:Y) = H(X) - H(X|Y),$$

the difference between the entropy in X and the entropy in X after Y has been fixed. ($H(X|Y)$ is simply defined as $H(X|Y) = H(X,Y) - H(Y)$.). In this form, some authors try to disentangle the amount of information in X about a given message from Y: we can write

$$I(X:Y) = H(X) - \sum_y p_y H(X|y) = \sum p_y [H(X) - H(X|y)] \equiv \sum_y p_y I(X:y),$$

where $H(X|y) = \sum_x p_{x|y} \log_2 p_{x|y}$. They then use $I(R:s)$ to define the stimulus s that optimally stimulates the neuron R. The meaning of $I(R:s)$ is not quite obvious though. It would be much safer conceptually to only consider optimally differentiable stimulus pairs, i.e. to use simple stimulus sets $P(S) = \{p(s_0) = .5, p(s_1) = .5\}$, and find the set S for which $I(R:S)$ is maximal.

OR4 and 5a here

Figure 2: Single stimulus information versus observed spike counts. **(a)** Information rate for particular stimuli (averaged over all presentations of that stimulus), and spike count associated with that stimulus. **(b)** Scatter plot of the same data, showing supposed lack of correlation.

4.1 Constant vs rapidly changing motion stimuli

Richmond et al. acknowledge that their method does not yield insight into the role of temporal modulation of the stimuli. This hiatus wasn't filled until a few years later. In [12], Buračas et al. present an awake monkey both with statically moving patterns and with patterns that are varied on timescales on the order of 30 to 300 ms and observe the response from neurons in middle temporal cortex.

They show that for static stimuli essentially all information is transmitted in the first few hundred milliseconds: even though the cells continues to fire during the entire stimulus presentation, BZDA fig 3 here

Figure 3: Mutual information between spike train in MT cortex and constant motion stimulus: **(a)** Cumulative, **(b)** Per unit time. The cell cannot distinguish fully between a set of eight stimuli (solid lines), but can essentially perfectly distinguish a set of two stimuli (dotted lines). For both stimulus sets, most information is transferred in the first 200 ms post stimulation.

considering the entire fire pattern barely yields more information about the stimulus (Figure 3). This clearly makes biological sense, as typical behavioral reaction times to such simple stimuli are on the order of a few hundred milliseconds. The amount of information transmitted about a typical stimilus is similar to the numbers found in [10]: little more than one bit per presentation. Thus it might look as if the information rate (in bits per second) of these cells is very small.

On presentation of rapidly changing stimuli, the situation is rather different. The cells can still recognize individual stimuli with the same precision as when those stimuli are presented for a much longer time. Therefore the information rate is much higher for these stimuli. Thus the authors argue that the standard practice of using prolonged stimuli may lead to significant underestimation of the capapabilities of neuronal systems[1].

4.2 Timing precision

Buračas et al. show that the high amount of information transferred about quickly varying stimuli is not due to an improved recognition percentage of stimuli, or in reduced inter-presentation variation: for both constant motion and rapidly changing stimuli, the resulting spike trains are well modeled by a modulated Poisson process[2]. Instead, the higher information transfer should be attributed to the high timing precision with which the stimuli are detected (Figure 4). Quantita-

---

[1] It should be noted that stationary stimuli are mainly used because they are relatively easy to generate and they yield spike trains that are easier to analyse by spike count statistics. Researchers generally agree that more rapidly varying stimuli would be preferable in principle, as they correspond more closely to natural system operation.

[2] The authors are not very specific about the meaning of a Poisson process modulated at a rate close to the mean firing rate.

BZDA fig 1 and 4b here

Figure 4: Response to a constant motion stimulus (left) and a rapidly changing stimulus (right). The top panels show firing rates, while the bottom panels show spike trains for individual trials. The second trace in the top right panel shows the stimulus: high indicates motion in direction of preference [highest induced firing rate], low in opposite direction.

tively, the timing precision of the first spike after a stimulus transition to preferred direction varied between 2 ms and 10 ms for different cells.

Obviously, showing constant stimuli hardly probes such precision. In fact, no time locking appears after the first 100 ms for constant motion stimulation.

4.3 Stimulus reconstruction vs Strong's direct method

Buračas et al. use both information estimators introduced in section 3.4. They find that the direct method yields a consistently higher number for the mutual information (Figure 5), a fact they attribute to the reconstruction method only measuring the information between the spike train and the characteristics of the stimulus under explicit control, while the direct method considers even the subtle unintentional differences. The authors argue that this shows that a neuron may encode more info about a stimulus than just the parameter you think you change. This conclusion is not at all obvious. Firstly, there is no a priori reason why the reconstruction method could not reconstruct such unintentional features, and secondly both methods are lower bound estimators, and it is dangerous to make inferences about the relative sizes of the quantities these estimators approximate from the relative sizes of the estimates. Additionally, inter cell variability of information rates was considerable.

BZDA fig 6abc here

Figure 5: Information rates as estimated by reconstruction and direct methods. **(a)** Raw data, lines connect data for individual cells. **(b)** Information per spike, showing that far more information is transmitted about precise timing than about exact simulus direction. **(c)** Total information transferred during stimuli lasting 1 s.

4.3.1 Earlier stages in visual processing

Clearly, high timing precision in MT cortex would be impossible without hardware support from earlier visual centres. Indeed, high timing precision has been reported in several such centres. For example, Berry et al. studied response in retinal ganglion to rapidly changing random noise patterns [13]. Their experiments were performed on dissociated retina with very similar results from rabbit and larval salamander retina, indicating a high degree of generality. The stimulus consisted of uniform intensity light modulated by gaussian noise at 30 ms. Typically, the same 20 s long stimulus segments was presented 100 times, although the authors also used much longer segments and even random checkerboard stimuli.

They find that ganglion cells in both species react specifically to certain features of the random noise input, such as sudden increases or decreases of illumination. For rapidly changing stimuli, most cells did not sustain a background firing rate. Instead, they fired with very high timing precision at particular moments in the stimulus segment (Figure 6).

The timing precision was typically better than 10 ms for the salamander cells, and better than 20 ms for the rabbit cells. Salamander cells exhibited quiet periods (firing rate less than 5% of maximum) that spanned over 90% of the recording time. Quiet periods in rabbit cells comprised about 75% of time. When the cells did spike, the variability in spike count was much less then expected for a Poission process (Figure 7). The authors calculate that upto six times as much information is carried by the timing than by the spike count of the responses.

Very interestingly, the timing precision deteriorated with the square root of stimulus contrast over a very wide range, as did the spike count precision[1], while the fraction of quiet periods barely changed except for very low contrast (4% or less). The authors thus hypothesize that the background firing rates observed by other experimenters when showing long constant luminosity stimuli, may be due to the retina adjusting to extreme low contrast situations. The background

---

[1] Poisson level variability was obtained at about 5% contrast.

BWM97 fig 1 here

Figure 6: Stimulus segment **(a)**, and responses **(b)** of several cells (S1, S2: salamander, R1, R2: rabbit). Spike trains from S1 and R2 for twenty trials each are shown on the right.

BWM97 fig 2 here

Figure 7: Timing variability **(a)** and spike count variability **(b)** for a typical salamander ganglion cell. The dotted line in (b) shows the Poisson relation, while the solid lines indicate limits imposed by counts being integers.

firing would then be caused by amplification of random events in upstream cells. They suggest that the high precision sector of the behaviour may be more natural, and that precisely timed events consisting of a few spikes only rather than modulated Poisson coding could be the basis of information processing in the retina and across much of the visual and other sensory systems. Suggestions of high timing precision have a long history in the auditory system, where source localization crucially depends on such precision. More recently, recurrence of fixed firing patterns has been studied in other brain areas ([14] and several papers by M. Abeles et al.).

5 Discussion

5.1 The relevance of information theory to neurobiology

A system where flow of information is optimized should be highly favoured by evolutionary pressure, as maintaining neurons is energetically very expensive and optimal use of resources for decision taking can often be a matter of life or death. This argument is elaborated in [15]. Strong et al. [6] cite a lecture by Atick [16] also advancing this point of view, but I haven't been able to verify that source.

5.2 What is 'optimal' stimulation?

Traditionally, optimal stimulation has been defined as the stimulus that generates the highest spike count[1]. In view of the fact that studies such as [13] show that spike count isn't always the only, or even the most important parameter in neuronal response, alternative definitions seem to be required. One alternative that has been suggested is to use the stimulus for which the transferred information is highest. These suggestions tend to propose $I(R:s)$ (Box 1) as the fundamental quantity. However, this quantity has uncertain conceptual foundations. A better approach would be to discard the notion of optimal stimuli in favour of optimal stimulus sets, defined as the set S of stimuli for which the information $I(R:S)$ is largest for a particular neuron. To my knowledge, this approach has not been used to date.

References

[1] G. T. Buračas and T. D. Albright, *Gauging sensory representations in the brain*, TINS **22** (1999) 303.

[2] R. R. Sokal and F. J. Rohlf, *Biometry*, 3rd ed., W. H. Freeman and company, 1995.

[3] A. C. C. Coolen, *Information theory in neural networks*, King's College London, The Strand, London WC1, UK, 1997, Lecture notes of a course in the dept. of mathematics, available from http://www.mth.kcl.ac.uk/∼tcoolen.

[4] F. Rieke, D. K. Warland, R. R. de Ruyter van Steveninck, and W. Bialek, *Spikes*, MIT Press, 1997.

[5] S. K. Ma, *(title unknown)*, J. Stat. Phys. **26** (1981) 221.

---

[1] It should be noted that the term 'rate code' has eroded quite a bit since the 1920s, since many people now compute rates in very short time windows using multi-trial averaging.

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

[7] R. R. de Ruyter van Steveninck and W. Bialek, *Real-time performance of a movement-sensitive neuron in the blowfly visual sstem: coding and information transfer in short spike sequences*, Proc.R.Soc. **B234** (1988) 379–414.

[8] B. J. Richmond, L. M. Optican, M. Podell, and H. Spitzer, *Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex – I. Resonse characteristics*, J. Neurophys. **57** (1987) 132–146.

[9] B. J. Richmond and L. M. Optican, *Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex – II. Quantification of response waveform*, J. Neurophys. **57** (1987) 147–161.

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

[11] C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, 1995.

[12] G. T. Buračas, A. M. Zador, M. R. DeWeese, and T. D. Albright, *Efficient discrimination of temporal pattern by motion-sensitive neurons in primate visual cortex*, Neuron **20** (1998) 959–969.

[13] M. J. Berry, D. K. Warland, and M. Meister, *The structure and precision of retinal spike trains*, PNAS **94** (1997) 5411–5416.

[14] Z. Nádasdy, H. Hirase, A. Czurkó, J. Csicsvari, and G. Buzsáki, *Replay and time compression of recurring spike sequences in the hippocampus*, J. Neuroscience **19** (1999) 9497–9507.

[15] M. R. DeWeese, *Optimization principles for the neural code*, Advances in Neural Processing Systems (D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds.), vol. 8, MIT Press, 1996.

[16] J. J. Atick, in *Princeton Lectures on Biophysics* (W. Bialek, ed.), World Scientific, 1992, pp. 223–289.