IL NUOVO CIMENTO VOL. 17 D, N. 7-8 Luglio-Agosto 1995

Information Flow in Sensory Neurons (*).

M. DEWEESE (1)(2) and W. BIALEK (2)

(1) *Department of Physics, Princeton University - Princeton, NJ*
(2) *NEC Research Institute - 4 Independence Way, Princeton, NJ 08540*

(ricevuto il 24 Aprile 1995)

Summary. — Recent experiments show that the neural codes at work in a wide range of creatures share some common features. At first sight, these observations seem unrelated. However, we show that all of these features of the code arise naturally in a simple threshold crossing model when we choose the threshold to maximize the transmitted information. This maximization process requires neural adaptation to not only the d.c. signal level, as in conventional light and dark adaptation (for example), but also to the statistical structure of the signal and noise distributions. Interestingly, if we fix the threshold level, we can observe a peak in the transmitted information at a finite value of the input signal-to-noise ratio. However, when we allow the threshold to adapt to the statistical structure of the signal and noise, the transmitted information is always monotonically increasing with increasing input signal-to-noise ratio.

PACS 87.10 – General, theoretical, and mathematical biophysics (including logic of biosystems, quantum biology, and relevant aspects of thermodynamics, information theory, cybernetics, and bionics).
PACS 01.30.Cc – Conference proceedings.

1. – Introduction.

Most sensory receptor cells produce analog voltages and currents which are smoothly related to analog signals in the outside world. Before being transmitted to the brain, however, these signals are encoded in sequences of identical pulses called action potentials or spikes. As physicists, we would like to know if there is a universal principle at work in the choice of these coding strategies. By «universal» we mean: In (nearly) all modalities of (nearly) all creatures. By «principle» we mean something analogous to the least-action principle in classical dynamics, say. The existence of such a potentially powerful theoretical tool in biology is an appealing notion, but it

(*) Paper presented at the International Workshop «Fluctuations in Physics and Biology: Stochastic Resonance, Signal Processing and Related Phenomena», Elba, 5-10 June 1994.

may not turn out to be useful. Perhaps the function of biological systems is best seen as a co nplicated compromise among constraints imposed by the properties of biologica materials, the need to build the system according to a simple set of developn ental rules, and the fact that current systems must arise from their ancestor s by evolution through random change and selection. In this view, biology is history, nd the search for principles (except for evolution itself) is likely to be futile. Obvious l /, we hope that this view is wrong, and that at least some of biology is understa ndable in terms of the same sort of universal principles that have emerged in the phys ics of the inanimate world.

Adria n [1] noticed in the 1920's that every peripheral neuron he checked produced discrete, identical pulses no matter what input he administered. From the work of Hodgker and Huxley [2] we know that these pulses are stable non-linear waves which emerge from the non-linear dynamics describing the electrical properties of the nerve cell mer brane. These dynamics in turn derive from the molecular dynamics of specific i on channels in the cell membrane [3]. By analogy with other non-linear wave problems, we thus understand that when these signals have propagated over a long distance—*e.g.*, ≈ one meter from touch receptors in a finger to their targets in the spinal co rd—every spike has the same shape. This is an important observation since it implies t at all information carried by a spike train is encoded in the arrival times of the spik s. Since a creature's brain is connected to all of its sensory systems by such axons, a l the creature knows about the outside world must be encoded in spike arrival t mes.

Until recently, neural codes have been studied primarily by measuring changes in the rate of spike production by different input signals. Recently it has become possible to characterize the codes in information-theoretic terms, and this has led to the discovery of some potentially universal features of the code [4-10] (or see [11] for a brief sur mary). They are

1) *Very high information rates.* The record so far is 300 bits per second in a cricket r echanical sensor.

2) *High coding efficiency.* In cricket and frog vibration sensors, the information rate is within a factor of 2 of the entropy per unit time of the spike train.

3) *Linear decoding.* Despite evident non-linearities of the nervous system, spike trains can be *decoded* by simple linear filters. Thus we can write an estimate of the analog input signal $s(t)$ as $s_{\text{est}}(t) = \sum_i K_1(t - t_i)$, with K_1 chosen to minimize the mean-squared errors (χ^2) in the estimate. Adding non-linear $K_2(t - t_i, t - t_j)$ terms does not significantly reduce χ^2.

4) *Moderate signal-to-noise ratios.* All these examples of high information transmis sion rates have SNR of order unity over a broad bandwidth, rather than high SNR in a narrow band.

We s all try to tie all of these observations together by elevating the first to a principle The neural code is chosen to maximize information transmission where informat on is quantified following Shannon. We apply this principle in the context of a simple model neuron which converts analog signals into spike trains.

2. – Information theory.

In the 1940's, Shannon proposed a quantitative definition for «information» [12, 13]. He argued first that the average information available from observations of some event x_i is just the entropy of the distribution from which the x_i are chosen, and showed that this is the only definition consistent with several plausible requirements. This definition implies that the amount of information a signal can provide about some other signal is just the difference between the entropy of its *a priori* distribution and the entropy of its conditional distribution. The average of this quantity is called the mutual (or transmitted) information. Thus, we can write the amount of information that the spike train tells us about the signal as

$$(1)\quad I[\{t_i\} \to s(\tau)] = -\int \mathscr{D}t_i\, P[\{t_i\}] \log_2 P[\{t_i\}] - \int \mathscr{D}s\, P[s(\tau)] \Big(-\int \mathscr{D}t_i\, P[\{t_i\}|s(\tau)] \log_2 P[\{t_i\}|s(\tau)] \Big),$$

where $\int \mathscr{D}t_i$ is shorthand for integration over all arrival times $\{t_i\}$ and summation over all numbers of spikes N, and $\int \mathscr{D}s$ denotes integration over the space of functions $s(\tau)$. $P[\{t_i\}|s(\tau)]$ is the probability distribution for the spike train when the signal is given, and $P[s(\tau)]$ and $P[\{t_i\}]$ are the *a priori* distributions for the signal and spike train.

3. – Interlude: information and SNR.

The recent literature on «stochastic resonance» attempts to characterize the transmission of information in non-linear systems by measurements of the signal-to-noise ratio (SNR) response to sine wave inputs. In a *linear* system, such measurements at all frequencies provide, in principle, a complete characterization of the system from which the Shannon information can be calculated in any given ensemble of input signals. Generations of experimentalists have attempted such «complete characterization» experiments on neurons, and it seems fair to conclude that these highly non-linear adaptive systems are unlikely to yield to this approach.

The focus on SNR has an even more fundamental problem, namely that it is defined uniquely only in the limit of small signals where the noisy non-linear system exhibits a linear response. Furthermore, a single number for SNR can be meaningful only if there is a single number which characterizes the scale of the noise, and this effectively limits the analysis to near-Gaussian noise sources. As an example, if the nose $\eta(t)$ flips at random between $\pm\eta_{\max}$, the SNR is very small as $\eta_{\max} \to \infty$, but all signals $|s(t)| < \eta_{\max}$ are perfectly detectable.

For the case of a small signal in a Gaussian noise background, it is a theorem that the SNR at the output of a non-linear device must be less than or equal to the SNR at the input. Because this point has caused some confusion in the literature, we include here a brief proof. For small input signals the definition of output SNR which

corresponds to direct experimental measurements, *e.g.*, with a lock-in amplifier, is as follows:

$$\mathrm{SNR}_{\mathrm{out}}(\omega) = \frac{\langle |\langle \tilde{\mathscr{F}}(\omega)\rangle_\eta|^2\rangle_s}{\langle |\tilde{\mathscr{F}}_0(\omega)|^2\rangle_\eta}, \tag{2}$$

where $\mathscr{F}(\tau)$ is the non-linear functional of the input signal and noise that describes our device, tildes indicate Fourier transforms, and zero subscripts indicate that the input signal is set to zero. We expand $\mathscr{F}$ for small signal,

$$\langle \mathscr{F}(t)\rangle_\eta \approx \langle \mathscr{F}_0(t)\rangle_\eta + \int \mathrm{d}\tau\, s(\tau)\langle D(t-\tau)\rangle_\eta, \tag{3}$$

where we have defined

$$D(t-\tau) \equiv \frac{\delta \mathscr{F}_0(t)}{\delta \eta(\tau)}. \tag{4}$$

For Gaussian noise we may use the identity

$$\langle |\tilde{\mathscr{F}}_0(\omega)|^2\rangle_\eta = \langle |\eta(\omega)|^2\rangle_\eta \langle |\tilde{D}(\omega)|^2\rangle_\eta, \tag{5}$$

which holds no matter how large the noise is compared to the scale for non-linear response of the device. Finally we use the fact that $\langle x^2\rangle \geq \langle x\rangle^2$ to complete the proof:

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\langle |\tilde{s}(\omega)|^2\rangle_s |\langle \tilde{D}(\omega)\rangle_\eta|^2}{\langle |\tilde{\eta}(\omega)|^2\rangle_\eta \langle |\tilde{D}(\omega)|^2\rangle_\eta} \leqslant \frac{\langle |\tilde{s}(\omega)|^2\rangle_s}{\langle |\tilde{\eta}(\omega)|^2\rangle_\eta} = \mathrm{SNR}_{\mathrm{in}}. \tag{6}$$

4. – The threshold crossing model.

At the input, we have some signal $s(t)$ embedded in a noisy background $n(t)$. These continuous time series are drawn from their respective probability distributions. This is crucial: No information can be carried by the signal unless its entropy is an extensive quantity. In other words, if we choose to study a signal composed of a sine wave, the information carried by the signal will not grow linearly with the length of time we observe it, whether or not noise is present. In addition to this, we would like to compare our results to the performance of real neurons in as natural conditions as possible, so we should use ensembles of broad-band signals, not sine waves.

We study the following model for encoding the signal $s(t)$ in a spike train: First, the signal and noise are added and filtered to produce a new analog time series $y(t)$. We then produce a spike each time $y(t)$ crosses some threshold value θ with positive slope. If we allow $y(t)$ to be any functional of $s(t) + n(t)$, then our model can be used to study *any* possible digital encoding of an analog signal. For most of what we present here, we will only consider linear filtering and Gaussian signal and noise distributions, though we will have something to say about the general case. We will

find it useful to define a characteristic time for the signal as

$$\tau_s \equiv \sqrt{\frac{\langle s^2 \rangle_s}{\langle \dot{s}^2 \rangle_s}}\,, \tag{7}$$

and similarly for the noise.

Following the classic discussion by Rice [14, 15], we can express the sum of delta-functions resulting from the positive slope crossings of $y(t)$ through a threshold θ with the following definition:

$$\varrho(t) \equiv \sum_{i=1}^{N} \delta(t_i) = \delta(y(t) - \theta)\,\dot{y}(t)\,H(\dot{y}(t))\,, \tag{8}$$

where H is the step function and dots denote time derivatives. In this language the spike firing rate—the probability per unit time of generating a spike given a signal wave form $s(t)$—is just

$$r(t) = \langle \varrho(t) \rangle_\eta\,. \tag{9}$$

If the signal and noise are Gaussian and y involves only linear filtering, then the average firing rate is

$$\langle r \rangle_s = \frac{1}{2\pi} \sqrt{\frac{\langle \dot{y}^2 \rangle}{\langle y^2 \rangle}} \exp\left[\frac{-\theta^2}{2\langle y^2 \rangle}\right]. \tag{10}$$

In terms of ϱ, the conditional distribution that appears in eq. (1) can be written as

$$P[\{t_i\}\,|\,s(t)] = \frac{1}{N!} \left\langle \exp\left[-\int \mathrm{d}t\,\varrho(t)\right] \prod_i^N \varrho(t_i) \right\rangle_\eta\,. \tag{11}$$

Now all that is left is to define the signal and noise distributions and then we can write an explicit formulation for the mutual information in the threshold crossing model. Unfortunately, it is hopeless to evaluate this formula exactly, so we try to find a self-consistent solution for the information near the Poisson limit. In this limit the correlations between the spikes are small, so the timing of each spike gives nearly independent information. To find an approximation that is valid near this limit we expand the product of ϱ's in eq. (11) in a cluster expansion and expand the exponent about $\varrho = \langle \varrho \rangle_\eta$. The small parameter in this expansion is related to $\langle r \rangle_s \tau_c$, where τ_c is some average of the signal and noise correlation times.

The zeroth-order term is just what we would get if there were no correlations between the spikes. It is

$$R_{\text{info}} \equiv \frac{I[\{t_i\} \to s(\tau)]}{T} = \left\langle r \log_2 \left(\frac{r}{\langle r \rangle_s} \right) \right\rangle_s + \ldots\,. \tag{12}$$

This expression is valid for *any* analog to spike encoding process as long as the correlations between the spikes are small. For linear filtering and Gaussian signal and

noise, the zeroth-order term for the threshold crossing model can be calculated exactly,

$$(13)\quad R_{\text{info}} = \frac{\langle r\rangle}{\ln 2}\left(\frac{\theta^2 - \langle y^2\rangle}{2\langle y^2\rangle(1+\langle\eta^2\rangle/\langle s^2\rangle)} + \right.$$

$$\left. + \frac{1}{2}\ln\left(\frac{\langle\dot\eta^2\rangle\langle y^2\rangle}{\langle\eta^2\rangle\langle\dot y^2\rangle}\right) + \frac{2\langle\dot\eta^2\rangle}{\sqrt{\pi\langle\dot y^2\rangle\langle\dot s^2\rangle}} f\left(\frac{\langle\dot\eta^2\rangle}{\langle\dot s^2\rangle}\right) + \ln 2\right) + \dots,$$

where we have defined

$$(14)\quad f(a) \equiv \int_{-\infty}^{\infty} dz \exp[-az^2] \int_0^{\infty} dx\, x \exp[-(x-z)^2] \ln\left(\int_0^{\infty} dy\, y \exp[-(y-z)^2]\right),$$

which we can easily approximate for any value of a. We can also calculate the first correction to the zeroth-order term in this case. For the threshold θ set to maximize the first term, the first correction is indeed small as long as the time constants of the signal and noise are within a factor of 10 of each other. If we were studying a modulated Poisson process in which spikes are triggered according to a Poisson process with rate $r(t)$ dependent on the signal $s(t)$, all the correlations between spikes would be due to the time-dependent rate alone. In that case, the zeroth-order term would be an upper bound to the exact expression for the information. For the threshold crossing model, the corrections can be positive or negative. In particular, we find that the first correction to the zeroth-order term is negative when the correlation time of the noise is smaller than that of the signal. This is because the signal is not changing much from spike to spike, so the spikes are providing somewhat redundant information. Note that in the limit as $\tau_\eta \to 0$ our threshold crossing model is a modulated Poisson model. When the noise has a larger correlation time than the signal, the first correction is positive because the correlations in the spikes are due to the signal and add to the information we calculate from the first term.

5. – An aside about stochastic resonance.

If we fix the threshold and signal variance, and then vary the input SNR, we observe a peak in the information rate at a finite SNR value for some ratios of the noise and signal correlation times as is shown in fig. 1. This is true even if the signal is broad band so the «resonance» does not correspond to synchronization of the output to the signal and may not be obvious from the power spectrum of the spike train.

We get a similarly peaked picture if we plot the information rate *vs.* threshold for fixed SNR. If we adapt the threshold to maximize R_{info} at each SNR value, we find that the maximized information rate is always a monotonically increasing function of the input SNR everywhere that our expansion is valid (again see fig. 1). So it seems that if you can adapt your coding strategy, you discover that stochastic resonance effects disappear when you maximize R_{info}. In fact, if a model of a neuron exhibits stochastic resonance effects after it has been adapted to maximize R_{info}, then there is probably not enough adaptive freedom in the model to find the true optimum.

Stochastic resonance effects like the ones discussed here occur when information

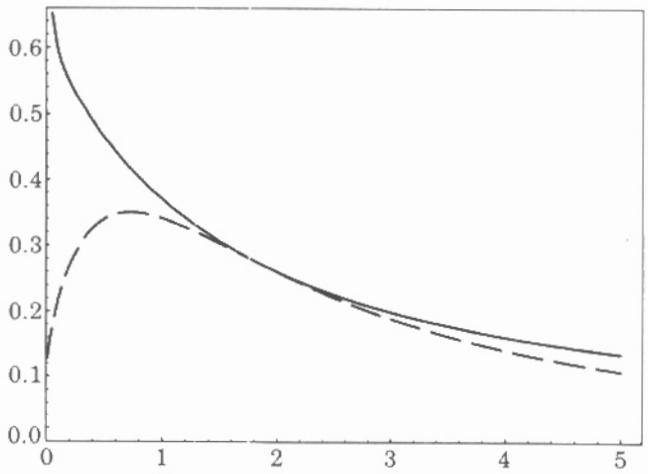

Fig. 1. – The dashed curve is the transmitted information rate *vs.* the input noise variance with the signal variance set to 1, $\tau_\eta/\tau_s = 0.1$, and the threshold fixed at $\sqrt{5}$. Notice that it goes through a maximum for a finite value for the SNR (≈ 1), which is very similar to the plots of $\mathrm{SNR}_{\mathrm{out}}$ *vs.* $\langle \eta^2 \rangle$ associated with stochastic resonance. The solid curve is the information rate for the same ratio of time constants, but with the threshold optimized at each value of the SNR. From the graph it is clear that the threshold setting for the dashed curve is the optimum for SNR near 1/2. For the entire region where our expansion is valid, we find that the information rate for the optimized threshold is monotonic with the input SNR, as it is here: Stochastic resonance effects evident when we fix the threshold disappear when it is optimized.

is selectively discarded in some encoding process. In our case, information about the analog signal is lost when we encode it in discrete spike times. Stochastic resonance has nothing to do with the dynamics of a system being studied. More generally, we can view the addition of noise to improve information transmission as a strategy for overcoming the *incorrect* setting of the threshold, θ. If θ is set to its optimum value, R_{info} is monotonic in SNR. This is analogous to the use of «dithering» to improve the performance of conventional analog-to-digital converters.

6. – Results from the linear filtered threshold crossing model.

Using the expression for the information rate shown in eq. (13) (which should be valid for τ_s and τ_η within a factor of 10 of each other), we can compare properties of the information rate in our model to those of experiments. We find that the information rate has a well-defined maximum near the Poisson limit. This is a non-trivial maximum since maximizing R_{info} does not correspond to maximizing $\langle r \rangle$ for a large region of parameter space including the physically relevant case where SNR is not too high and $\tau_s > \tau_\eta$.

At maximum information transmission in this region of parameter space, the (zeroth-order) information per spike has a strikingly simple form that depends only

on the SNR:

$$\frac{R_{\text{info}}}{\langle r \rangle} = \frac{1}{\ln 2}\left(\frac{1}{1 + 1/\text{SNR}}\right)\text{bits}. \tag{15}$$

When the SNR is small, the information per spike is linear with SNR, which we knew from low SNR expansions we did before attempting the general SNR case [11]. When the SNR is high, the information per spike approaches its highest value of $1/\ln 2$ bits, but clearly saturation occurs at rather modest SNR.

Having found an optimal setting of the threshold, we can study several properties of the resulting code [16]. To summarize, the information rate is robust to errors in the timing of individual spikes, with roughly 90% of the information retained when timing errors are $\approx 10\%$ of the mean interspike interval. This is in good agreement with experiment, and is another way of stating item number 2 of the introduction regarding the efficiency of codes. We can also analyze the optimal code using the same linear decoding methods used in analyzing real spike trains, except that here we know how much information is available in total. We find that reconstructing the signal with a linear filter captures 30% to 90% of the available information, in good agreement with the observation that the addition of non-linear terms to the reconstruction does not significantly improve the information rate. Indeed, the optimal code is in all respects similar to the real neural codes *except* that the real neurons reviewed above typically transmit two to three times as much information per spike, suggesting that we are missing something with our simple model. Perhaps real neurons are designed to maximize the information rate, but they are optimizing over a much larger space of coding strategies than we do with our linear filtering model.

It may also be that real neurons maximize something related to the information per spike rather than the information rate itself. If this is true, then there must be some penalty for high timing resolution in the spike train. Otherwise, the problem is not well defined since we can show that the information per spike diverges like the threshold squared as we let the threshold go to infinity. This produces very few spikes that each carry a lot of information, but this information can be recovered only if the spike times are «read» with arbitrarily high precision. This is an appealing candidate for a universal principle because both the production of spikes and the timing precision needed to maintain and read them can be thought of as metabolic costs to be minimized while trying to maximize the information transmitted. In addition to the heightened information per spike, the spikes would be more sparse on average, which would improve the linear decodability.

7. – Concluding remarks.

The four seemingly unrelated features that were common to several recent experiments on a variety of neurons are actually the natural consequences of maximizing the transmitted information. Specifically, they are all due to the relation between $\langle r \rangle_s$ and τ_c that is imposed by the optimization. We make a new prediction: Optimizing the code requires that the threshold adapt not only to cancel d.c. offsets, but it must adapt to the statistical structure of the signal and noise. Very recently, experimental hints at adaptation to statistical structure have been seen in the fly visual system [17] and in the salamander retina [18].

REFERENCES

[1] ADRIAN E. D., *The Basis of Sensation* (W. W. Norton, New York) 1928.
[2] KATZ B., *Nerve, Muscle, and Synapse* (McGraw-Hill, New York) 1966.
[3] SAKMANN B. and NEHER E. (Editors), *Single Channel Recording* (Plenum, New York) 1983.
[4] BIALEK W., RIEKE F., DE RUYTER VAN STEVENINCK R. R. and WARLAND D., *Science*, **252** (1991) 1854.
[5] RIEKE F., *Physical Principles Underlying Sensory Processing and Computation* (Dissertation, University of California at Berkeley) 1991.
[6] WARLAND D., *Reading Between the Spikes: Real-Time Processing in Neural Systems* (Dissertation, University of California at Berkeley) 1991.
[7] RIEKE F., BODNAR D. and BIALEK W., *Proceedings of the III International Congress of Neuroethology, Montreal* (1992).
[8] RIEKE F., YAMADA W., MOORTGAT K., LEWIS E. R. and BIALEK W., *Auditory Physiology and Perception: Proceedings of the IX International Symposium on Hearing*, edited by Y. CAZALS, L. DEMANY and K. HORNER (Elsevier, Amsterdam) 1992, p. 315.
[9] WARLAND D., LANDOLFA M., MILLER J. P. and BIALEK W., *Analysis and Modeling of Neural Systems*, edited by F. EECKMAN (Kluwer Academic) 1991, p. 327
[10] RIEKE F., WARLAND D. and BIALEK W., *Europhys. Lett.*, **22** (1993) 151.
[11] BIALEK W., DEWEESE M., RIEKE F. and WARLAND D., *Physica A*, **200** (1993) 581.
[12] SHANNON C. E., *Proc. I. R. E.*, **37** (1949) 10.
[13] BRILLOUIN L., *Science and Information Theory* (Academic Press, New York, N.Y.) 1962.
[14] RICE S. O., *Selected Papers on Noise and Stochastic Processes*, edited by N. WAX (Dover, New York) 1954, p. 133.
[15] KAC M., *Bull. Am. Math. Soc.*, **49** (1943) 314.
[16] DEWEESE M., *Optimization Principles for the Neural Code* (Dissertation, Princeton University) 1995.
[17] DE RUYTER VAN STEVENINCK R. R., BIALEK W., POTTERS M. and CARLSON R. H., *Statistical adaptation and optimal estimation in movement computation by the blowfly visual system*, in *IEEE International Conference on Systems, Man, and Cybernetics* (1994), p. 302.
[18] WARLAND D. and MEISTER M., unpublished.