Reliable biological communication with realistic constraints

Gonzalo G. de Polavieja*
Department of Theoretical Physics, Facultad de Ciencias C-XI, and Biophysics Group, Instituto "Nicolas Cabrera," Facultad de Ciencias, C-XVI, Universidad Autonoma de Madrid, Madrid 28049, Spain
(Received 1 June 2004; published 21 December 2004)

Communication in biological systems must deal with noise and metabolic or temporal constraints. We include these constraints into information theory to obtain the distributions of signal usage corresponding to a maximal rate of information transfer given any noise structure and any constraints. Generalized versions of the Boltzmann, Gaussian, or Poisson distributions are obtained for linear, quadratic and temporal constraints, respectively. These distributions are shown to imply that biological transformations must dedicate a larger output range to the more probable inputs and less to the outputs with higher noise and higher participation in the constraint. To show the general theory of reliable communication at work, we apply these results to biochemical and neuronal signaling. Noncooperative enzyme kinetics is shown to be suited for transfer of a high signal quality when the input distribution has a maximum at low concentrations while cooperative kinetics for near-Gaussian input statistics. Neuronal codes based on spike rates, spike times or bursts have to balance signal quality and cost-efficiency and at the network level imply sparseness and uncorrelation within the limits of noise, cost, and processing operations.

DOI: 10.1103/PhysRevE.70.061910 PACS number(s): 87.19.La, 87.17.Nn, 05.70.−a, 02.50.Tt

I. INTRODUCTION

Optimization theory has made an important impact in our understanding of biological structures like insect compound eyes or bones as well as in biological movement and behavior [1]. Optimality arguments complement mechanistic models by pointing to relevant constraints and common themes underlying different mechanisms. Here we are interested in optimal biological signaling. Arguments about the optimality of biological signaling have a long tradition, especially in neuroscience [2–6]. An important reason for the success of optimality theories of biological signaling is the applicability of information theory [7] and the links of this theory to statistical physics [2].

Recently, realistic constraints have been incorporated to the theory of optimal biological signaling for a better correspondence with experiments. Energy costs have been considered in the context of neuronal signaling [8] and theoretical results in this case have been shown to correspond to some cortical neurons [9,10]. We have generalized the optimal signaling theory by considering both costs and noise [11]. We found a much improved correspondence between theory and experiments in most cortical neurons when including both noise and cost constraints into the theory. Both signal quality and cost-efficiency were found to shape signaling. Similar discussions of the interplay of cost and noise have been given in the discussion of retinal signaling [12,13] and in other stochastic neural signaling problems [14].

Here we extend the general theory of optimal biological signaling to consider any type of constraints. A generalized Boltzmann distribution is obtained for a linear cost constraint, a generalized Poisson when the constraint is in the times and a generalized Gaussian when considering quadratic constraints. We also show optimal signaling at work in biochemical and neuronal signaling. The simple "rule of thumb" we obtain is that biological systems with a high information transfer must dedicate more output range to the more probable inputs and less to the noisier ones and to those paying a higher metabolic cost or time. We also show the links of the present theory to results in statistical physics, that are obtained here in the limit of no noise or no transitions.

The paper is organized as follows. Section II gives the statement of the maximization of the constrained information transfer. Section III discusses a particular case with analytical solution to gain insight into the general problem. Section IV gives an application to the case of enzymatic reactions. Section V gives the general solution to the optimal problem as a generalized Boltzmann form. Section VI discusses the case of temporal constraints, that results in a generalized Poisson distribution. Section VII gives the application to several cases of neuronal signaling. Section VIII discusses constraints for several components that reduce to results in statistical mechanics in the limit of no noise or no transitions. Finally, we discuss possible extensions of the results.

II. MAXIMUM INFORMATION TRANSFER WITHIN BIOLOGICAL CONSTRAINTS

Consider first the simple case of discrete states and a single constraint. The term information transfer is used informally in biology and its most studied quantitative counterpart is the information transfer proposed by Shannon [2,7,15]. Let $S=\{s_1,s_2,\ldots,s_k\}$ be the input states and $M=\{m_1,m_2,\ldots,m_l\}$ the output states. There is information transfer when there is a statistical dependency between input and output. This is expressed as the average *distance* between the joint distribution for input and output states $p(s_i,m_j)$ and the distribution corresponding to complete independency between them $p(s_i,m_j)_{ind}\equiv p(s_i)p(m_j)$ as

*Email address: gonzalo.polavieja@uam.es
URL: http://www.ft.uam.es/Neurociencia/gonzalo/Leech_page.htm

 ©2004 The American Physical Society

$$I(S;M) = \sum_{i,j} p(s_i,m_j)\log\left(\frac{p(s_i,m_j)}{p(s_i)p(m_j)}\right). \tag{1}$$

Our optimization problem consists in maximizing the information transfer I given a cost constraint and a noise structure. The cost constraint and the noise structure are included in the maximization in the following way. The cost constraint is given by fixing the average value E of a quantity ϵ, say the expected energy and the energy values of the systems states, respectively, as $E=\Sigma_i p(m_i)\epsilon_i$. This cost might be metabolic or, for example, an average time. Costs that are physically very different, like energy or time, can be treated identically from a mathematical point of view. The noise structure is fixed by an specified transition matrix $Q_{kj}=p(s_k|m_j)$ given by the probability that when the the output is in state m_j the input state was s_k. The transition matrix $\mathbf{Q}$ is introduced in the definition of the information transfer I in the following way. The information transfer I can be expressed as the difference between the output entropy and the noise entropy, $I(S;M)=H(M)-H(M|S)$, with $H(M)=-\Sigma_j p(m_j)\log p(m_j)$ the output entropy and $H(M|S)=-\Sigma_{j,k}p(s_j,m_k)\log p(m_j|s_k)$ the noise entropy. To express the noise entropy solely in terms of the output probabilities $\{p(m_i)\}$ and the matrix $\mathbf{Q}$, we use the following relations. We can write the noise entropy as $H(M|S)=\Sigma_j p(m_j)\xi_j$ with

$$\xi_j = -\sum_k Q_{kj}\log P_{jk}, \tag{2}$$

where the matrix $\mathbf{P}$ has elements $P_{jk}\equiv p(m_j|s_k)$ that can be expressed in terms of the transition matrix $\mathbf{Q}$ using Bayes' theorem [15] as

$$P_{jk} = \frac{p(m_j)Q_{kj}}{\sum_j p(m_j)Q_{kj}}. \tag{3}$$

The maximum information transfer problem consists then in finding the output probabilities $\{p(m_i)\}$ that maximize the information transfer I given the noise matrix $\mathbf{Q}$ and the cost constraint $E=\Sigma_i p(m_i)\epsilon_i$. Equivalently, using the method of the Lagrange multipliers, we can maximize

$$\mathcal{L} := I - \beta\left(\sum_j p(m_j)\epsilon_j - E\right) - \lambda\left(\sum_j p(m_j) - 1\right), \tag{4}$$

with I the information transfer and the second and third terms the cost constraint and the normalization, respectively.

III. AN ANALYTICAL EXAMPLE

To gain insight into the type of solutions we can obtain from the constrained maximum information problem, we first consider a simple analytical case with three states. The first two states have the same noise and cost and the third one is noiseless and has a higher cost, as depicted in Fig. 1. This means that $\epsilon_1=\epsilon_2$, $\epsilon_3>\epsilon_2$ and the noise matrix is given by $Q_{11}=Q_{22}=1-\rho$, $Q_{12}=Q_{21}=\rho$ and $Q_{33}=1$. The information transfer I for this case can be written as

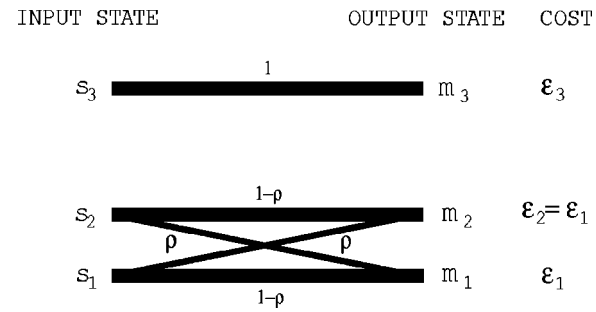

FIG. 1. Simple case of communication from three input states to three output states. Maximum information transfer with a constraint in the cost implies a penalization of the noisy states (the first two) and of the costly states (the third one). These abstract states can represent, for example, enzyme concentration values, the different firing patterns of a neuron, or the patterns of network activity.

$$I(S;M) = -2p(m_1)\log p(m_1) - p(m_3)\log p(m_3) - 2p(m_1)\xi, \tag{5}$$

with the ξ_j in Eq. (2) given by

$$\xi_1 = \xi_2 = -\rho\log\rho - (1-\rho)\log(1-\rho) \equiv \xi; \xi_3 = 0. \tag{6}$$

The output states that maximize $\mathcal{L}$ in Eq. (4) are

$$p(m_1) = p(m_2) = Z^{-1}\exp(-\beta\epsilon_1 - \xi), \tag{7}$$

$$p(m_3) = Z^{-1}\exp(-\beta\epsilon_3), \tag{8}$$

with $Z=2\exp(-\beta\epsilon_1-\xi)+\exp(-\beta\epsilon_3)$ the normalization constant and β given by the value of the average energy $2p(m_1)\epsilon_1+p(m_3)\epsilon_3=E$. The probabilities for the input states are obtained from $p(s_k)=\Sigma_i Q_{ki}p(m_i)$ to be $p(s_1)=p(s_2)=p(m_1)$ and $p(s_3)=p(m_3)$.

In the case of equal costs and equal noise for all states, the information transfer is maximized with equal output probabilities. In the case of all states having the same noise but different costs, the output states that maximize the information transfer have a Boltzmann distribution that penalizes more the output states with higher cost with a factor $\exp(-\beta\epsilon_n)$ for the nth state. In the case of all states having the same cost but different noise, output states are penalized more with higher noise with a factor $\exp(-\xi_n)$. In general, both cost and noise penalize output states in a form generalizing the Boltzmann distribution with an extra exponent penalizing the amount of noise.

From this example we can also obtain the relationship between the information transfer I and the average value of the constraint E substituting the output probabilities into the expression of I in Eq. (1) to obtain

$$I = \beta E + \log Z, \tag{9}$$

or in terms of the output and noise entropies as $H(S)=\beta E+\log Z+H(S|M)$ or $H(M)=\beta E+\log Z+H(M|S)$, respectively. In the limit of no transitions ($\rho\to 0$ in the example) the probabilities reduce to the Boltzmann form and the relation in Eq. (9) to the relationship familiar in statistical mechanics of the form $H=\beta E+\log Z$ with $H(S)=H(M)$.

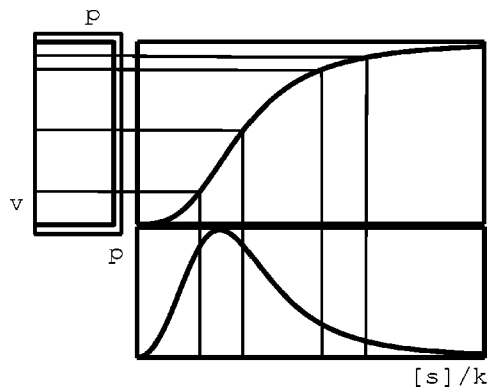

FIG. 2. Relationship between input and output distributions and the input to output transformation for maximum information transfer, here illustrated for the transformation from the dimensionless substrate concentration $[S]/k$ to the dimensionless reaction velocity v. In the simple case of no noise present and no cost constraint, the output distribution is a constant and the transformation is the integral of the input distribution. More output range is then dedicated to the more probable inputs, as illustrated by the two equal intervals of substrate concentration.

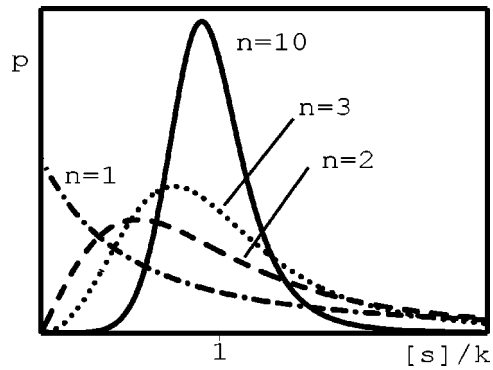

FIG. 3. Substrate distributions for simple reaction kinetics maximize the information transfer. For Michaelis-Menten kinetics ($n=1$) the distribution has a maximum at the lowest substrate concentration and a heavy tail. Increasing the cooperativity n of the kinetics corresponds to distributions with a maximum closer to the constant k in the reaction kinetics in Eq. (12).

IV. INFORMATION TRANSFER IN BIOCHEMICAL REACTIONS

We consider the transformation of input states given by the value of the concentration of substrate $[S]$ into the velocity of an enzymatic reaction v. We consider here the simple case of a single substrate. For generality, we consider the Hill equation, a phenomenological expression valid both for noncooperative and cooperative kinetics, of the form

$$v = \frac{[S]^n}{k^n + [S]^n}, \tag{10}$$

corresponding to Michaelis-Menten kinetics for $n=1$ and to cooperative kinetics for $n>1$, with the degree of cooperativity increasing with n. To understand the difference between noncooperative kinetics and cooperative kinetics for information transfer, we consider the simplest case of no noise present and no cost constraint. The maximization of the information transfer in this case gives $p(v)=\alpha$ with α a constant. In general the input and output probabilities are related by $p(s_k)=\Sigma_i Q_{ki} p(m_i)$. The simple case with no noise is a transformation of variables from $[S]$ to v so the probability densities are related as $p(v)dv=p([S])d[S]$. Using the optimal velocity density $p(v)=\alpha$, it follows that $p([S]) = \alpha(dv/d[S])$, that gives

$$p([S]) = \frac{k^n n[S]^{n-1}}{(k^n + [S]^n)^2}. \tag{11}$$

Figure 2 illustrates this relationship between the input to output transformation and the input and the output probability densities that maximize the information transfer. For maximum transfer with no noise (or same noise for all velocities) and no cost constraint the input density must be the derivative of the input to output transformation. This means that the more probable inputs have a larger output range. Two identical intervals of substrate values shown in Fig. 2 have very different output ranges, larger in the case of the more probable inputs.

Figure 3 shows the optimal substrate distribution for different values of cooperativity n. For the Michaelis-Menten case, $n=1$, the distribution has a maximum at $[S]=0$ and a heavy tail that makes the mean undefined. For cooperative kinetics the densities have their maximum at $\pi k n^{-1} \operatorname{cosec}(\pi/n)$ that increases with increasing n, and in the limit $n\to\infty$ tends to k. With increasing cooperativity the density is increasingly close to a Gaussian function centered at k and with decreasing width. The qualitative difference between the Michaelis-Menten and cooperative kinetics can be understood with the aid of Fig. 2. The largest slope of the reaction velocity in the Michaelis-Menten case is at low substrate concentrations, while for cooperative kinetics is close to k. For the most probable inputs to have a larger output representation, the substrate distribution has then to be maximal at low values for Michaelis-Menten and at values close to k for cooperative kinetics.

The match between substrate distribution and kinetics discussed in Fig. 2 is for a stationary substrate distribution. When this distribution varies with time, the kinetics must change accordingly to keep the match that assures maximal information transfer. The input to output transformation must then adapt to the changing statistics as

$$v([S];t) = \int_0^{[S]} d[S] P([S];t). \tag{12}$$

Figure 4 illustrates the case of a probability density of a substrate with a decreasing variance in time. The transformation that maximizes the information transfer then corresponds to a cooperativity increasing in time.

For the case of maximal information transfer with a cost constraint with costs of the form $\epsilon(v)$, the optimal probability density of substrate concentration is

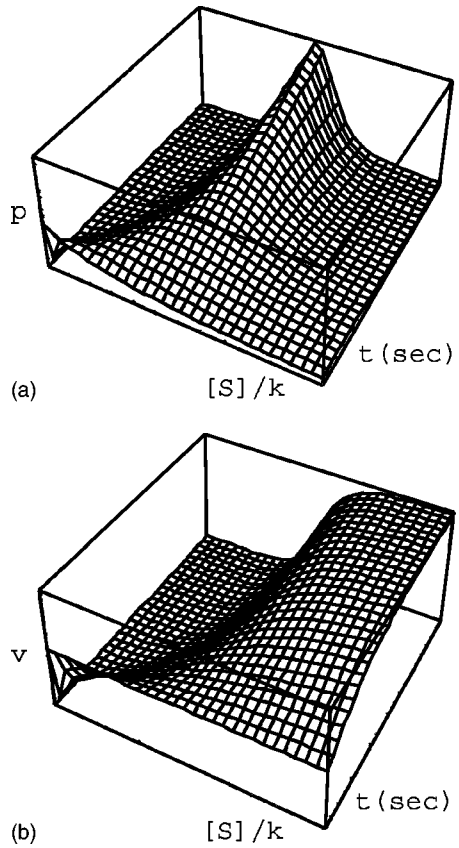

FIG. 4. Maximum information transfer for nonstationary substrate statistics. (a) Example of a substrate distribution that has a decreasing variance in time. (b) To keep the match between input statistics and kinetics that assures maximal transfer, the kinetics has to adapt when the change in substrate statistics. In this case the kinetics has to increase its degree of cooperativity.

$$P([S]) = Z^{-1}\frac{k^n n[S]^{n-1}}{(k^n+[S]^n)^2}\exp(-\beta\epsilon(v)). \tag{13}$$

The substrate values contributing to a larger output range are the most probable ones with the lower cost at the output. In the next section we will see that the noise acts as an extra cost as already seen in the example given in Sec. III.

V. GENERAL SOLUTIONS

To obtain the general solution we will use the alternating Blahut-Arimoto algorithm [15–17], used in communication theory to calculate the channel capacity and the rate distortion function [15]. The maximization of the constrained information information transfer,

$$\max_{p(m):\Sigma_j p(m_j)\epsilon_j=E}\sum_{j,k} p(m_j)Q_{kj}\log\frac{p(m_j)Q_{kj}}{p(m_j)\sum_i p(m_i)Q_{ki}}, \tag{14}$$

can be expressed as a double maximization [see Lemma 13.8.1 and Eq. 13.145 in [15]] as

$$\max_{P}\max_{p(m):\Sigma_j p(m_j)\epsilon_j=E}\sum_{j,k} p(m_j)Q_{kj}\log\frac{P_{jk}}{p(m_j)}. \tag{15}$$

The double maximization can be performed using the alternating Blahut-Arimoto algorithm [15–17]. This algorithm starts with a guess of the output distribution $p(m_i)$ (say, a random vector) and finds the P_{jk} from Eq. (3). A better guess for the output probabilities is then given by

$$p(m_j) = \frac{\exp-(\beta\epsilon_j - \sum_k Q_{kj}\log P_{jk})}{\sum_j \exp-(\beta\epsilon_j - \sum_k Q_{kj}\log P_{jk})}, \tag{16}$$

where β in Eq. (16) has to be evaluated at each step in the alternating algorithm from the energy constraint as

$$\frac{\sum_j \epsilon_j \exp-(\beta\epsilon_j - \sum_k Q_{jk}\log P_{jk})}{\sum_j \exp-(\beta\epsilon_j - \sum_k Q_{jk}\log P_{jk})} = E. \tag{17}$$

This output distribution $p(m)$ in Eq. (16) is then used for the next iteration and the operation is repeated until convergence. Csiszar and Tusnady [18] have shown that such an alternating algorithm converges to the maximum for this type of problems. For clarity, we give the recipe for calculation as follows.

Initialize $\mathbf{p^0}$ as a random vector.

FOR $t=0,1,2\ldots$ (until convergence),

$$P^t_{jk} = \frac{p^t(m_j)Q_{kj}}{\sum_j p^t(m_j)Q_{kj}}, \tag{18}$$

$$p^{t+1}(m_j) = \frac{\exp-(\beta^t\epsilon_j - \sum_k Q_{kj}\log P^t_{jk})}{\sum_j \exp-(\beta^t\epsilon_j - \sum_k Q_{kj}\log P^t_{jk})}, \tag{19}$$

where β^t in Eq. (19) has to be evaluated for each t from the energy constraint

$$\frac{\sum_j \epsilon_j \exp-(\beta^t\epsilon_j - \sum_k Q_{jk}\log P^t_{jk})}{\sum_j \exp-(\beta^t\epsilon_j - \sum_k Q_{jk}\log P^t_{jk})} = E. \tag{20}$$

ENDFOR

As in the simple analytical example in Sec. III, the probabilities have the form

$$p(m_j) = \hat{Z}^{-1}\exp(-\hat{\beta}\epsilon_j - \hat{\xi}_j), \tag{21}$$

where the hat symbol in $\hat{Z}$, $\hat{\beta}$ and $\hat{\xi}$ refers to the converged values. The distribution obtained has then the form of a generalized Boltzmann distribution with an exponential penalization of cost and noise.

VI. GENERALIZED POISSON DISTRIBUTIONS FOR TEMPORAL CONSTRAINTS

We have considered the maximization of the information transfer I given a cost constraint and obtained a generalized Boltzmann distribution. When considering that the output

states M take different times, these different times play the role of the costs and the Boltzmann distribution reduces to the Poisson distribution. In previous sections we have considered the maximization of the information transfer I. A maximization of the rate of information transfer I/T does not affect previous results, but it matters when the different costs are times. In particular, the maximization of the rate of information transfer I/T singles out a particular Poisson distribution, as we show in the following.

Let the input S be the times t and the output states the times τ. The information transfer has the form

$$I(t;\tau) = \int_0^\infty \int_0^\infty dt\, d\tau\, p(t,\tau) \log\left(\frac{p(t,\tau)}{p(t)p(\tau)}\right) \tag{22}$$

that can be written using entropies as $I(S;M)=H(M)-H(M|S)$, with $H(\tau)=-\int_0^\infty d\tau\, p(\tau)\log p(\tau)$ and $H(\tau|t)=\int_0^\infty d\tau \int_0^\infty p(\tau)p(t|\tau)\log p(\tau|t)$. The average time T is given by $T=\int_0^\infty d\tau\, p(\tau)\tau$. Maximizing information transfer I with an average time constraint T is equivalent to maximizing $\mathcal{L} := I-\beta T$. For the case in which all output states have the same noise, $H(\tau|t)=\alpha$, this maximization gives

$$p(\tau) = \frac{1}{T}\exp(-\tau/T), \tag{23}$$

that is a Poisson distribution corresponding to an information transfer of the form $I=\log T+1-\alpha$.

The maximization of the information rate I/T when all τ have the same noise is a particular Poisson distribution. Maximizing the function I/T respect to $p(\tau)$, gives $p(\tau)=(1/T^*)\exp(-\tau/T^*)$, with the specific average time $T^*=\exp(-\alpha)$ where α is the entropy of the noise when all output states have the same noise.

Including a refractory period τ_0 gives also a Poisson distribution but with an exponent that has a more complicated expression, $p(\tau)=\beta\exp(-\beta\tau)$, with β a solution of $\beta=(\tau_0+\beta^{-1})(-\log\beta-\alpha+1)$ that can be formally expressed as $\beta=\tau_0^{-1}\mathit{Product\ Log}(\tau_0\exp(-\alpha))$.

A more realistic case would consider that the different τ have different amounts of noise. In this case there are deviations from the Poisson distribution. The situation is analogous to that of Sec. V but in this case the output state costs are times and are expressed as

$$p(\tau_j) = \hat{Z}^{-1}\exp(-\hat{\beta}\tau_j - \hat{\xi}_j), \tag{24}$$

a generalized Poisson distribution that reduces to the Poisson distribution in the limit of no noise or when all output states have identical noise. For a maximization of the rate of information transfer I/T would select a particular β^* to be determined numerically.

VII. BALANCE BETWEEN SIGNAL QUALITY AND COST EFFICIENCY IN NEURONS

In this section we discuss maximum information transfer for neural codes based on spike rates (number of action potentials per second), spike times and bursts and a simple network configuration. For the spike rate code, we compare the theoretical result and recordings in awake monkeys from Ref. [9] (see also Ref. [10]). These recordings were obtained extracellularly from cortical neurons in awake macaques while they watch movies of natural scenes on a monitor.

Metabolic costs [19] and noise are known to influence neuronal signaling [20,21]. Measurements of cost and noise are however not detailed enough for our purposes so our strategy is to make simple assumptions. First we discuss the case of a neuronal code based on spike rates. We will make the following simple assumptions: (1) the output states are the neurons spiking rates, (2) the cost is linearly proportional to the spiking rate, and (3) the noise is here defined as deviations from the rate with a maximum probability obtained for the same stimulus and its underlying causes include extra spikes or their absence from network or intrinsic activity. In particular, we simply include noise as spikes produced when the noise-free state (the most probable state) corresponds to silence. These assumptions are formalized in the following way. Let $S=\{s_0=0, s_1=1, \ldots\}$ be the desired or noise-free rates and $M=\{m_0=0, m_1=1, \ldots\}$ the actual rates. According to assumption 2, the costs of the rates are $\epsilon(m_i)=\epsilon_0+\epsilon_1 i$ with ϵ_0 and ϵ_1 constants. In accord with assumption 3, $p(s_i|m_i)=1-A_i$ and $p(s_0|m_i)=A_i$, with A_i decreasing with the spike rate. We found that $A_i=a_0\exp(-\gamma i)$ is a simple function working well for all neurons tested. We have included these three assumptions into the algorithm in Sec. V and we found that

$$p(r) \approx \mathrm{Z}^{-1}\exp(-\beta r - \exp(-r/\alpha)) \tag{25}$$

approximated well the numerical and experimental results, where r is the rate. Note that different average costs E and different amounts of noise give different distributions but the functional form of the prediction is the same for all neurons within the above simple assumptions for the noise and cost constraint.

Figure 5 compares the theoretical prediction with recordings from visual cortex neurons. Similar fittings are obtained for all neurons presented in [9] and the two chosen illustrate the cases of small and large effect of the presence of noise. For high spiking rates the theoretical expression in Eq. (25) predicts an exponential decay that is clearly seen in Fig. 5(a). The exponent in this decay depends, according to the theory, on the mean cost. For low spike rates, the theoretical expression in Eq. (25) predicts a low usage to keep signal quality, as seen in Fig. 5(b). In general, there is a balance between signal quality and cost-efficiency that explains signal use, as seen in Fig. 5(c). Experimental data is then consistent with maximum information transfer with a linear cost constraint. According to the theory of optimal transfer and within the simple assumptions made above, neurons would then have the same functional form for the distribution as in Eq. (25) and the differences then arise due to different amounts of noise and different mean costs. Experiments using naturalistic stimulus could be designed to test whether signal usage varies according to the noise characteristics of the neuron, as predicted here. For the more general nonstationary case, the noise and cost might change dynamically. In this case the

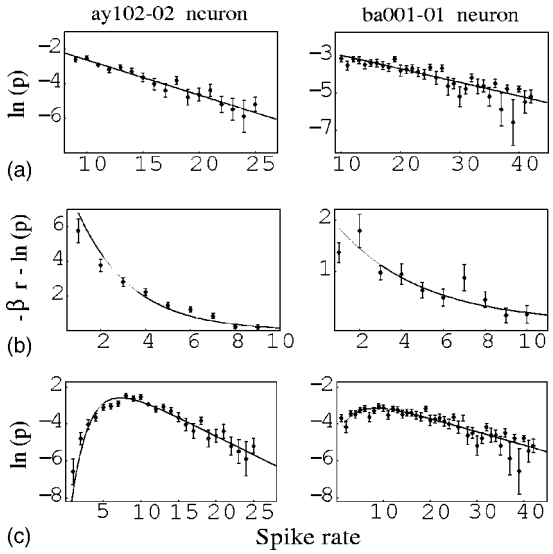

FIG. 5. Comparison of the theoretical prediction for maximum information transfer with a cost constraint in Eq. (25) (solid line) to the experimental spike rate distributions of visual cortex neurons. The spike rate (number of action potentials per second) was taken from. [9] Their recordings were obtained using extracellular electrodes in awake monkeys while they watch a monitor showing natural scenes. An exponential decay of the signal usage in (a) assures cost efficiency. Low signal usage at low spike rates in (b) assures signal quality. There is a good correspondence between theory and experiments for all spike rates, as shown in (c).

signal usage should always match this changes to maintain an expression like Eq. (25).

Faster signaling could be based on spike times instead of spike rates. As we saw in Sec. VI, the maximum rate of information corresponds in this case to a distribution of the form

$$p(\tau) = \hat{Z}^{-1} \exp(-\hat{\beta}^{*}\tau - \hat{\xi}(\tau)). \tag{26}$$

The noise is responsible for the deviation from a pure Poisson distribution. Near-Poisson distributions are routinely observed in neuronal recordings. Experimental measurements of noise, together with expression (26) would be able to distinguish if these experimental distributions emerge from optimal transmission or randomness.

The neuronal code could also be based on spike bursts. Maximal information transfer for bursting systems would be slightly different depending on the type of constraint. Pure maximal rate of information transfer would predict a generalized Poisson distribution with τ the duration of the burst (or $\tau_0+\tau$ if we include the silences between bursts). It is also possible that the relevant optimization is the information per unit energy I/E. This optimization gives a result analogous to the I/T optimization as the signaling cost associated to each burst is directly proportional to its duration. However, the total cost is $\epsilon=\epsilon_0+\epsilon_1 i$, with ϵ_0 a basal cost independent of the signaling. The value of ϵ_0 would not affect the maximization of information with a cost constraint, but it affects the maximization of I/E. To experimentally distinguish between between the I/T and I/E optimizations, one would need to compare the cases $\epsilon=\epsilon_0+\epsilon_1 i$ and $\tau=\tau_0+\tau_1\tau$.

The results from the single-neuron cases apply straightforwardly to the network case when the neurons are uncorrelated. In this case the output state is the pattern of network activity instead of the pattern of activity of a single neuron. Costly and noisy patterns of network activity should be avoided for high constrained communication. Let us consider the more general case of a given mean correlation. Concretely, we consider the transfer of an input S into the activity of N correlated neurons with activities $\{x_1, x_2, \ldots, x_N\}$ and mean pairwise correlations $\langle x_i x_j \rangle$ and for simplicity no other constraints. The probability maximizing the information transfer is then of the form

$$p(\mathbf{x}) = Z^{-1} \exp\left(-\sum_{i\neq j} \beta_{ij} x_i x_j - \xi(\mathbf{x})\right). \tag{27}$$

Maximum information transfer for this network case penalizes correlated states as well as noisy states. In general there should be a balance between these two factors as well as any other constraints.

Neurons are not only maximizing information transfer, but processing the information. Results from that processing must however be communicated reliably within constraints. There is a natural room in the present formulation for processing thanks to the matrix **Q** of transitions between input and output states. These transition states not only describe the noise (one to many transitions) but also any processing (many to one transitions).

VIII. OTHER CONSTRAINTS AND MORE LINKS TO STATISTICAL PHYSICS

Previous sections have considered a single constraint. In general there are several constraints E^k on the properties ϵ^k of the form $E_k=\Sigma_i p(m_i)\epsilon_i^k$. For this case, the maximum information transfer solutions have the form $p(m_i) = Z^{-1} \exp(-(\Sigma_k \beta^k \epsilon_i^k) - \xi_i)$.

We have also considered in previous sections that costs (metabolic or temporal) are linear. It is well known that the Gaussian distribution maximizes the entropy with a quadratic cost constraint [2]. Following the arguments in previous sections, maximizing the information transfer with a quadratic cost constraint will then give a generalized Gaussian distribution with the extra exponent penalizing noisy states.

More sophisticated constraints result in more elaborate distributions. In the following we give several complex constraints that do not have at present a direct application to biology, but that nevertheless illustrate the generality of the approach by obtaining other results from statistical physics. The reason for this success lies in the relationship of information theory and entropies [22]. Our approach contains previous results as particular cases when the noise is negligible or independent of the state.

Consider a system that has a constraint E, a constraint in the expected number of system elements and in the maximal number of elements in each state. Let $p(s_{ij})$ represent the probability that the the system state i is occupied by j ele-

ments. Following a procedure analogous to the maximization in Eq. (4) but now with the constraint expressed as $\Sigma_{i=1}^{K}\epsilon_i\Sigma_{j=0}^{M_i}jp(s_{ij})=E$, a constraint in the number of system elements as $\Sigma_{i=1}^{K}\Sigma_{j=0}^{M_i}jp(s_{ij})=N$ and the normalization constraint as $\Sigma_{j=0}^{M_i}p(s_{ij})=1$, with $\{M_i\}$ the maximum number of elements in each state i, we obtain the probability that maximizes the information transfer between the input (system) and output (measured) states as

$$p(s_{ij}) = \frac{\exp(-(\hat{\beta}\epsilon_i + \hat{\mu})j - \hat{\xi}_{ij})}{\sum_{j=0}^{M_i} \exp(-(\hat{\beta}\epsilon_i + \hat{\mu})j - \hat{\xi}_{ij})}, \tag{28}$$

where $\hat{\beta}$ and $\hat{\mu}$ are the values of the Lagrange multipliers for the constraint E and the total number of elements, respectively. To see the relationship with problems in statitical physics we take the limit of no transitions between states, $\xi\rightarrow 0$, and consider the case for which all states have same maximal value of elements, $M_i=M$ for all i. The proportion of elements in each state i reduces in this case to the Bose-Einstein distribution $n_i=\Sigma_{j=0}^{M}jp(s_{ij})=[\exp(-(\beta\epsilon+\mu)+1)]^{-1}$ or the Fermi-Dirac distribution $[\exp(-(\beta\epsilon+\mu)-1)]^{-1}$ for $M\rightarrow\infty$ and $M=1$, respectively. Viewing the input and output states as the system and measured states in statistical experiments, the common statistical distribution follow for maximum information transfer with simple constraints in the no transitions or low-noise limit.

IX. DISCUSSION

The theory of biological signaling benefits from the results of information theory, that provides the rate of information transfer as the functional to be optimized in communication systems. Instead of calculating global optima of the information transfer, an approach closer to biological systems is to include relevant constraints. Here we have included noise and costs constraints into information theory to include the limitations in signaling that biological systems are confronted with. We have asked which are the distributions that maximize the information transfer given any noise and any constraints. We have obtained generalized versions of the Boltzmann, Gaussian, Poisson and other distributions with an extra term measuring the amount of noise. The biological transformations that maximize the transfer have been shown to be those that dedicate more output range to the more probable inputs and least range to the noisy states and the states with a higher contribution to constraints. Noncooperative reactions are best suited for the transfer of substrates with a maximum at low concentrations while cooperative reactions for near-Gaussian distributions. The firing of neurons is penalized for low rates to minimize the effect of noise and for high rates for cost efficiency.

Some extensions of the results may proof particularly useful. The effect of the processing through the matrix **Q** should be studied in particular applications. The importance of chemical reactions as means of communication and their construction to adapt to input statistics and to reduce the effect of noise has been discussed here, but further theoretical and experimental work is needed to show its relevance.

The theoretical results point to new experimental venues. Both noise and cost measurements are needed to understand the statistics of signal usage. Experiments designed to understand mechanisms should also measure the statistics in naturalistic conditions as they may be matched for high information transfer.

ACKNOWLEDGMENTS

Discussions with Simon Laughlin, Mikko Juusola, William Bialek and Ilya Nemenman are gratefully acknowledged. I am especially indebted to Vijay Balasubramanian and Michael Berry for discussing their independent work on metabolically efficient signaling in the vertebrate retina. I am also grateful to Stefano Panzeri for sending the data from Ref. [9]. This work is partially supported with grants from MCyT and fBBVA.

[1] R. M. Alexander, *Optima for Animals* (Princeton University Press, Princeton, NJ, 1996).
[2] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).
[3] S. B. Laughlin, Z. Naturforsch. C **36c**, 910 (1981).
[4] J. H. van Hateren, Biol. Cybern. **68**, 23 (1992).
[5] J. H. van Hateren, J. Comp. Physiol., A **171**, 157 (1992).
[6] J. H. van Hateren, Nature (London) **360**, 68 (1992).
[7] C. E. Shannon, AT&T Bell Labs. Tech. J. **27**, 379 (1948).
[8] W. B. Levy and R. A. Baxter, Neural Comput. **8**, 531 (1996).
[9] A. Treves, S. Panzeri, E. T. Rolls, M. Booth, and E. A. Wakeman, Neural Comput. **11**, 601 (1999).
[10] R. Baddeley, L. F. Abbot, M. C. A. Booth, F. Sengpiel, T. Freeman, E. A. Wakeman, and E. T. Rolls, Proc. R. Soc. London, Ser. B **264**, 1775 (1997).
[11] G. G. de Polavieja, J. Theor. Biol. **657**, 657 (2002).
[12] V. Kimber Balasubramanian and M. Berry, Neural Comput. **13**, 799 (2001).
[13] V. Balasubramanian and M. Berry, Network **13**, 531 (2002).
[14] S. Schreiber, C. K. Machens, A. V. M. Herz, and S. B. Laughlin, Neural Comput. **14**, 1323 (2002).
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1998).
[16] R. E. Blahut, IEEE Trans. Inf. Theory **18**, 460 (1972).
[17] S. Arimoto, IEEE Trans. Inf. Theory **18**, 14 (1972).
[18] I. Csiszar and G. Tusnady, Stat. Decisions Suppl. **1**, 205 (1984).
[19] S. B. Laughlin, R. R. de Ruyter van Stevenick, and J. C. Anderson, Nat. Neurosci. **1**, 36 (1998).
[20] C. Koch, *Biophysics of Computation* (Oxford University Press, New York, 1999).
[21] J. A. White, J. T. Rubinstein, and A. R. Kay, Trends Neurosci. **23**, 131 (2000).
[22] E. T. Jaynes, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT Press, Cambridge, MA, 1978).