

# **Classification of stationary neuronal activity according to its information rate**

Lubomir Kostal, Petr Lánský

*Institute of Physiology, Academy of Sciences of the Czech Republic, Videnska 1083,  
142 20 Prague 4, The Czech Republic*

## **Classification of stationary neuronal activity**

### **Address for correspondence:**

Lubomir Kostal

Institute of Physiology, Academy of Sciences of the Czech Republic,

Videnska 1083,

142 20 Prague 4,

The Czech Republic

tel: +420 2 4106 2276,

fax: +420 2 4106 2488,

e-mail: [kostal@biomed.cas.cz](mailto:kostal@biomed.cas.cz)

# Classification of stationary neuronal activity according to its information rate

## Abstract

We propose a measure of the information rate of a single stationary neuronal activity with respect to the state of null information. The measure is based on the Kullback-Leibler distance between two interspike interval distributions. The selected activity is compared with the Poisson model with the same mean firing frequency. We show that the approach is related to the notion of specific information and that the method allows us to judge the relative encoding efficiency. Two classes of neuronal activity models are classified according to their information rate: the renewal process models and the first-order Markov chain models. It has been proven that information can be transmitted changing neither the spike rate nor the coefficient of variation and that the increase in serial correlation does not necessarily increase the information gain. We employ the simple, but powerful, Vasicek's estimator of differential entropy to illustrate an application on the experimental data coming from olfactory sensory neurons of rats.

## 1 Introduction

It is generally accepted that the information in neuronal systems is transferred using the series of action potentials – the spike trains. There are two main hypotheses that attempt to classify possible ways in which the spike train may carry information: the frequency (rate) codes and the temporal spike codes (Gerstner and Kistler, 2002; Theunissen and Miller, 1995). Both hypotheses rely on the important assumption that single spikes are mutually indistinguishable. Therefore the spike trains can be

considered as a time series of point events. This is a widely accepted simplification that allows further analysis, especially from the information-theoretic point of view.

The idea of frequency coding was proposed by Adrian (1928). He showed that the number of spikes per some time window (spike frequency) non-linearly increases with the increase of the stimulus intensity. The idea of temporal spike coding (Perkel and Bullock, 1968; Theunissen and Miller, 1995), on the other hand, employs the timing of the spikes or their particular temporal pattern. This is a consequence of the fact that there are situations where time averaging is not possible. Also the high time variability in neuronal discharge, particularly within the cortex, see, e.g., Buracas and Albright (1999), may play some role and cannot be simply 'averaged out'. The extreme case, which gives an upper bound on the possible amount of information encoded (Zador, 1998), follows from the hypothesis that the precise value of each successive interspike interval (ISI) carries the information (in other words the information gain is maximized if there is no noise in the transmission). Though frequency codes and temporal spike codes are shown to be compatible in many cases (Gerstner and Kistler, 2002), it is clear that infinitely many different spike records may have the same frequency.

If the neuronal firing is stationary the mean spike frequency (the inverse of the mean ISI, Lánský et al. (2004)) carries the information from the frequency coding hypothesis point of view. However, as has been shown by Bialek et al. (1991) (see overview in Theunissen and Miller (1995)), the frequency codes in single neurons carry information about dynamic stimuli, which results in a non-stationary neuronal signal. The temporal coding, on the other hand, has been shown to occur almost exclusively under steady-state stimulus conditions (Fuller and Looft, 1984; Middlebrooks et al., 1994).

At the first approximation, neuronal firing under steady-state conditions is often described as a renewal process. This is justified by the observation that the

emitted spike resets the membrane potential of the neuron's body independently of the preceding synaptic processes (Abeles, 1982). In this case ISIs are described as independent realizations of a positive random variable  $T$ . The corresponding probability density function  $f(t)$  is thus the complete descriptor of such neuronal activity with the expected value of  $T$  (the mean ISI) denoted by  $E(T)$ .

Most of the time, however, though the neuronal firing is stationary there is a dependency structure between the observed ISIs (Chacron et al., 2001; Longtin and Racicot, 1996). The dependence may arise due to incomplete resetting of the membrane potential after the spike is emitted, which is experimentally observed especially in the distal parts of the neuron (Abeles, 1982). Such a type of neuronal firing is not a renewal process. The successive ISIs  $\{T_i\}$  are then statistically dependent with an expected value  $E(T) = E(T_i)$ . The activity is fully described by the joint probability density function  $f(t_1, t_2, \dots)$ , see, e.g., Cox and Lewis (1966) for details. The importance of the dependence in the ISI structure is also reflected in recent efforts to include the effect of serial correlation into neuronal models (Lindner, 2004; Lánský and Rodriguez, 1999; Sakai et al., 1999). Despite this effort we are not aware of any models where analytical results are available, except those vaguely mentioned in Lawrance (1972, p.215) for the model developed by Lampard (1968).

In this article we will try to quantify the information encoded by the temporal coding scheme, i.e., under the assumption of stationary stimuli conditions. The information theory, introduced by Shannon (1948) provides the mathematical basis for the task.

## 2 Theory and methods

Without loss of generality, information may be defined as a decrease in uncertainty (Shannon, 1948). For discrete random variables, the entropy,  $H$ , measures the

uncertainty and is thus closely related to the notion of information. The quantity that measures the uncertainty of a continuous random variable  $T$  with a probability density function  $f(t)$  is called differential entropy  $h(f)$ , also denoted  $h(T)$ , (Cover and Thomas, 1991):

$$h(f) = - \int_{-\infty}^{\infty} f(t) \ln f(t) dt. \quad (1)$$

Differential entropy  $h(f)$  does not share the same properties and intuitive interpretation as the entropy  $H$  (Shannon, 1948). Namely, it can be negative and its value changes with coordinate transforms. Thus it cannot be used as an absolute measure of the information content. Nevertheless, the most 'random' distribution is still the one that maximizes  $h$  under the given constraints. No discrete random variable appears in this article and for simplicity the term "entropy" is used for  $h(f)$ .

The measure of deviation of two probability density functions  $f(t)$  and  $g(t)$  that is related to the concept of entropy is the Kullback-Leibler (KL) distance (also relative entropy, Cover and Thomas (1991))

$$\text{KL}(f, g) = \int_{-\infty}^{\infty} f(t) \ln \frac{f(t)}{g(t)} dt. \quad (2)$$

The KL distance defined by formula (2) is not symmetric and does not satisfy the triangle inequality. Further, if there exists an interval such that  $g(t) = 0$  while  $f(t) \neq 0$  then  $\text{KL}(f, g) = \infty$ . On the other hand the KL distance is independent of coordinate transforms and  $\text{KL}(f, g) \geq 0$  with equality if, and only if,  $f(t) = g(t)$ . These properties make the KL distance suitable for measuring the information content described by the probability density  $f(t)$  relative to the reference probability density  $g(t)$ , which maximizes uncertainty. The interpretation of KL distance as a measure of the information content was pioneered by Tarantola and Valette (1982) in the general theory of inverse problems. The reference state  $g(t)$  is often described

as the state of null information or 'the state of total ignorance' (Tarantola, 1994). If the interpretation in terms of information is not needed, the KL distance can be still used as a tool to compare ISI distributions.

The above mentioned definitions can be extended into more dimensions, see Cover and Thomas (1991) for details. Let  $\{T_1, \dots, T_n\}$  be a set of  $n$  random variables described by the joint probability density function  $f(t_1, \dots, t_n)$ . Then the (joint) entropy  $h(T_1, \dots, T_n)$ , sometimes also denoted as  $h(f)$ , is defined by

$$h(T_1, \dots, T_n) = - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) \ln f(t_1, \dots, t_n) dt_1 \dots dt_n. \quad (3)$$

The joint entropy can be expressed as a sum of conditional entropies:

$$h(T_1, \dots, T_n) = \sum_{i=1}^n h(T_i | T_{i-1}, \dots, T_1), \quad (4)$$

where  $h(T_i | T_{i-1}, \dots, T_1) = - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_i) \ln f(t_i | t_{i-1}, \dots, t_1) dt_1 \dots dt_i$ . The independence bound for the joint entropy states that (Cover and Thomas, 1991, p. 232)

$$h(T_1, \dots, T_n) \leq \sum_{i=1}^n h(T_i) \quad (5)$$

with equality if, and only if, the variables  $T_i$  are independent, i.e., when the conditional densities are equal to the marginal ones  $f(t_i | t_{i-1}, \dots, t_1) = f(t_i)$  for each  $i$ . Furthermore, it is possible (Cover and Thomas, 1991, p. 273) to define the 'entropy per variable', or, the entropy rate  $\bar{h}$  of a stochastic process  $\{T_i\}_{i=1}^{\infty}$  as

$$\bar{h}(f) = \lim_{n \rightarrow \infty} \frac{1}{n} h(T_1, \dots, T_n). \quad (6)$$

The existence of the limit is guaranteed if the sequence  $\{T_i\}_{i=1}^{\infty}$  is stationary.

Similarly to formula (3), the KL distance of two joint probability density

functions  $f(t_1, \dots, t_n)$  and  $g(t_1, \dots, t_n)$  is

$$\text{KL}(f, g) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) \ln \frac{f(t_1, \dots, t_n)}{g(t_1, \dots, t_n)} dt_1 \dots dt_n. \quad (7)$$

In agreement with equation (6) we formally define the rate  $R$  of the KL distance as:

$$R = \lim_{n \rightarrow \infty} \frac{1}{n} \text{KL}(f(t_1, \dots, t_n), g(t_1, \dots, t_n)). \quad (8)$$

Expanding  $R$  employing formulas (3) and (6) gives

$$R = - \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) \ln g(t_1, \dots, t_n) dt_1 \dots dt_n \right] - \bar{h}(f) \quad (9)$$

In light of the KL distance as a measure of information we interpret  $R$  as the 'information rate' of a stochastic process described by the joint probability density function  $f(t_1, t_2, \dots)$  relative to the state of null information described by  $g(t_1, t_2, \dots)$ . This concept of information differs from the standard notion of mutual information introduced by Shannon (1948), though under special conditions a relation between the two can be found, as we demonstrate later. We will employ this definition to classify the available information content in several particular models of neuronal activity.

## 3 Results and discussion

### 3.1 The spike train as a renewal process

We apply formula (9) to find the information rate  $R$  of a renewal process describing the ISI generation. First, however, we have to determine uniquely the appropriate state of null information. The least informative state is in other words the most

random one, i.e. it bounds the entropy rate  $\bar{h}$  of all other possible processes from above. Formula (5) implies, that such sequence of random variables must consist of independent and identically distributed variables maximizing the entropy  $h$ . The maximum entropy distribution on  $(0, \infty)$  with fixed expected value is the exponential one described by the probability density function

$$g(t) = a \exp(-at), \quad (10)$$

where  $a > 0$  and the expected value equals to  $1/a$ . The entropy of the exponential distribution is  $h(g) = 1 - \ln a$ . The state of null information is thus described by the Poisson process which holds a prominent position in the neuronal modeling (Gerstner and Kistler, 2002). The Poisson-like firing has been experimentally observed in many situations, particularly in the cortical neurons, see Abeles (1982) for details. From formulas (4) and (6) it follows, that the entropy rate of a renewal process is equal to the entropy of the probability density. Thus the equation (9) reduces to

$$R = aE(T) - \ln a - h(f). \quad (11)$$

As we quantify the information gained beyond the hypothesis of frequency coding, we let distributions  $f(t)$  and  $g(t)$  have the same mean values,  $E(T) = 1/a$ . Formula (11) finally results in

$$R = 1 + \ln E(T) - h(f). \quad (12)$$

As mentioned, the entropy  $h(T)$  of a random variable  $T$  changes with the transformation  $T \rightarrow sT$  (with a 'scaling' constant  $s > 0$ ) as  $h(sT) = h(T) + \ln s$  (Cover and Thomas, 1991, p. 233). Due to this 'scaling property' the rate  $R$  in the formula (12) does not depend on the actual  $E(T)$ , which also follows from the fact that the rate is defined 'per ISI'. We see that the information rate of the renewal process gained

beyond frequency coding relative to the state of null information is reduced to the calculation (estimation) of  $h(f)$ .

In the following we compare the information rates  $R$  of several ISI distribution models described by probability density functions  $f(t)$  with two parameters. The rate  $R$  is parameterized by the coefficient of variation  $CV$  – the ratio of standard deviation to  $E(T)$ . While the advantage of  $CV$  is that for exponential distribution holds  $CV = 1$  (independently of  $E(T)$ ), the condition  $CV = 1$  does not in general imply exponentiality of the distribution. This feature was explored in greater detail with respect to experimental data in Kostal and Lansky (2006).

### 3.1.1 Gamma Model

The gamma distribution is one of the most frequent statistical descriptor of ISI (Levine, 1991; Mandl, 1992; Rieke et al., 1997). Its probability density function  $f(t)$  is defined by

$$f(t) = \frac{b^a t^{a-1} \exp(-bt)}{\Gamma(a)}, \quad (13)$$

with parameters  $a > 0, b > 0$  and  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$  the gamma function. It holds for equation (13):  $E(T) = a/b$  and  $CV = 1/\sqrt{a}$ . Using formula (12) we obtain the information rate  $R(CV)$

$$R(CV) = 1 - \ln CV^2 - \ln \Gamma(1/CV^2) + \frac{\Psi(1/CV^2) - 1}{CV^2} - \Psi(1/CV^2), \quad (14)$$

where  $\Psi(z) = \frac{d}{dz} \ln \Gamma(z)$  is the digamma function.

The rate  $R(CV)$  from equation (14) is plotted in Fig. 1A. It is useful to view the dependence  $R(CV)$  for every possible value of  $CV$  and  $R$ . To do this we employ a

conformal mapping  $(CV, R) \rightarrow (\hat{CV}, \hat{R})$  described by

$$\begin{aligned}\hat{CV} &= \arctan CV, \\ \hat{R} &= \arctan R.\end{aligned}\tag{15}$$

Due to this transformation, the whole quadrant  $(0, \infty) \times (0, \infty)$  is mapped onto section  $(0, \pi/2) \times (0, \pi/2)$  and the points  $\hat{CV} = \pi/2$ , resp.  $\hat{R} = \pi/2$  are identified with  $CV = \infty$ , resp.  $R = \infty$ . For convenience labels on Fig. 1A. correspond to the original scale  $(CV, R)$ .

The gamma density  $f(t)$  in equation (13) is exponential for  $a = 1$ , implying  $R(CV = 1) = 0$ . The information rate  $R$  tends to infinity for  $CV \rightarrow 0$ . This is a general property which can be seen directly from formula (12): for  $CV = 0$  the variable  $T$  is described by a  $\delta$ -function and the entropy is thus  $h(f) = -\infty$ , see Cover and Thomas (1991, p. 229) for details. On the other hand, the limit  $R(CV \rightarrow \infty) = \infty$  is true for the gamma distribution, but does not hold in general.

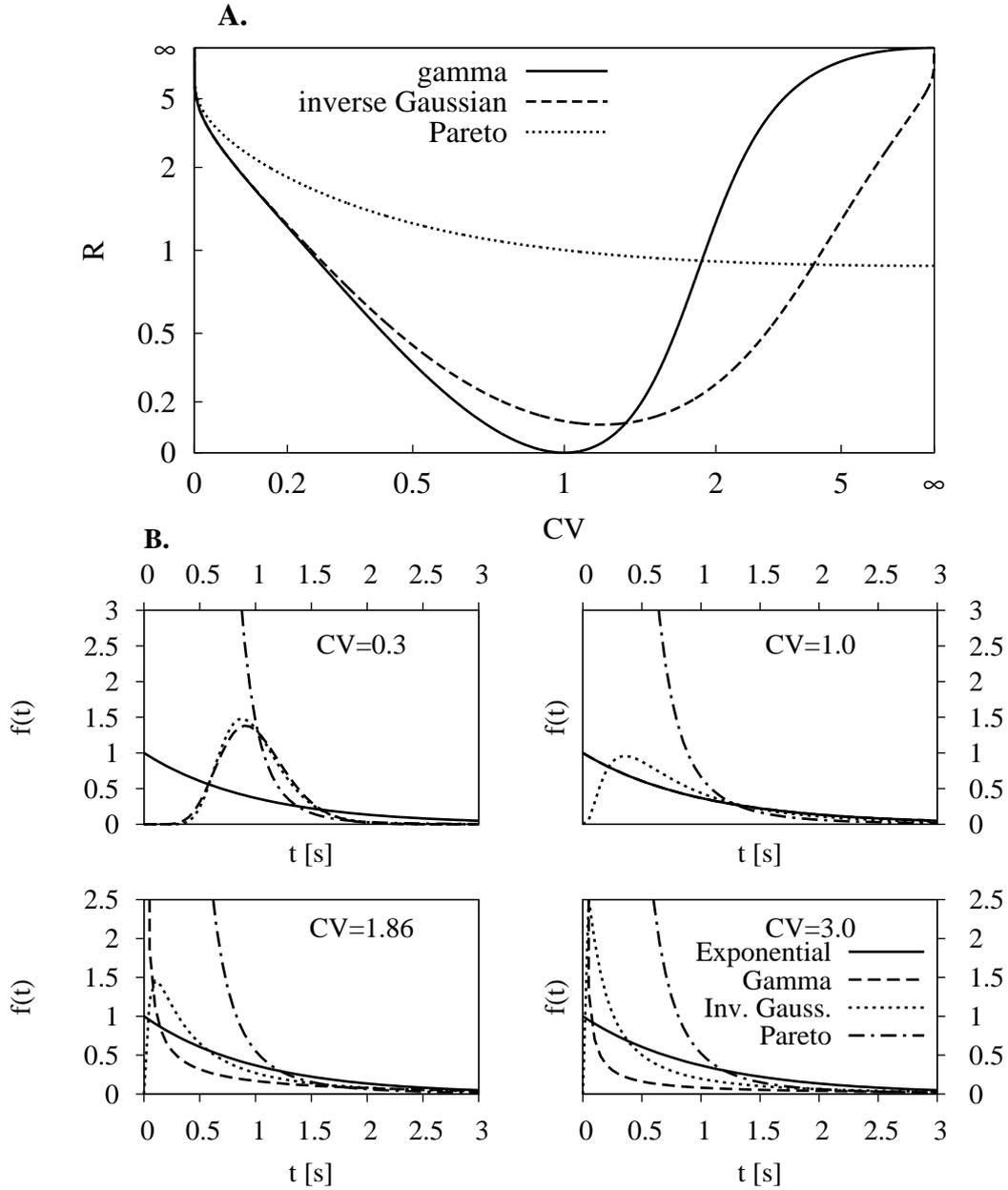
### 3.1.2 Inverse Gaussian model

The inverse Gaussian distribution of ISI can be obtained from the integrate-and-fire class of neuronal models and is also often fitted to experimental data (Gerstein and Mandelbrot, 1964; Levine, 1991). Its probability density is

$$f(t) = \sqrt{\frac{a}{2\pi bt^3}} \exp\left[-\frac{1}{2b} \frac{(t-a)^2}{at}\right],\tag{16}$$

with parameters  $a > 0$  and  $b > 0$ . In this case  $CV = \sqrt{b}$ ,  $E(T) = a$  and using the same technique as in the case of the gamma model the information rate  $R(CV)$  is obtained in the form

$$R(CV) = \frac{1}{2} \ln \frac{e}{2\pi} - \ln CV + \frac{3}{\sqrt{2\pi}} \frac{\exp(1/CV^2)}{CV} K_{1/2}^{(1,0)}(1/CV^2),\tag{17}$$



**Fig. 1.:** (Caption on the following page.)

where  $K_\nu^{(1,0)}(z)$  is the derivative of the modified Bessel function of the second kind  $K_\nu^{(1,0)}(z) = \frac{\partial}{\partial \nu} K_\nu(z)$  (Abramowitz and Stegun, 1972).

The resulting dependence  $R(CV)$  from equation (17) is plotted in Fig. 1A. using the transformation rules (15). The inverse Gaussian distribution is not exponential for  $CV = 1$  thus always  $R(CV) > 0$ . Furthermore the minimal rate (and thus the maximum similarity with the exponential model) does not occur at  $CV = 1$

**Fig. 1:** Information rates  $R$  of renewal processes in dependence on the  $CV$  (A). For  $CV = 1$  the rate of gamma model is zero, implying that at this point its distribution becomes exponential. The inverse Gaussian model is never exponential ( $R > 0$ ) and its minimal distance is at  $CV \approx 1.17$ . The Pareto model is also never exponential, but contrary to the gamma and inverse Gaussian cases its rate decreases with increasing  $CV$ , reaching the limit  $R(CV \rightarrow \infty) \approx 0.89$ . The information rate of inverse Gaussian grows slowly compared to that of the gamma distribution. The rates of gamma and inverse Gaussian are equivalent for  $CV \approx 1.31$ . For  $CV \ll 1$  the gamma and inverse Gaussian rates are hardly distinguishable, though this fact cannot be used to judge the degree of their mutual similarity, see (B), where the three corresponding probability density functions with  $E(T) = 1$  s are plotted for selected values of  $CV$ .

but at  $CV \approx 1.17$ . The rates of gamma and inverse Gaussian models become equivalent at  $CV \approx 1.31$ . In comparison with gamma model we see that though  $R(CV \rightarrow \infty) = \infty$  for both cases the gamma model approaches infinite information rate  $R(CV)$  much faster than the inverse Gaussian. On the other hand, for very small values of  $CV$  both rates are undistinguishable.

### 3.1.3 Pareto model

The Pareto distribution is not resulting from any theoretic neuronal model and we are not aware of any attempt to fit it to experimentally observed ISIs. We present it here to show a different kind of behavior than that of the above mentioned models.

The probability density function  $f(t)$  of the Pareto distribution is

$$f(t) = \begin{cases} 0, & t \in (0, b) \\ ab^a t^{-a-1}, & t \in [b, \infty) \end{cases} \quad (18)$$

with parameters  $a > 2$  and  $b > 0$ . In this case  $CV = 1/\sqrt{(a^2 - 2a)}$  and  $E(T) = ab/(a - 1)$ . Using formula (12), we arrive at the equation for the information rate  $R(CV)$  of the Pareto model in the form

$$R(CV) = CV^2 - CV\sqrt{1 + CV^2} + \ln\left(2 + \frac{1 + 2CV^2}{CV\sqrt{1 + CV^2}}\right), \quad (19)$$

for illustration see Fig. 1A. The Pareto model is never exponential and we can see that for increasing values of  $CV$  the rate  $R(CV)$  is slowly decreasing with the limit (which is also the minimal value of  $R(CV)$ ):  $R(CV \rightarrow \infty) = \ln 4 - 1/2 \approx 0.886$ . The information rates of Pareto and gamma models are equal for  $CV \approx 1.86$ , while for gamma and inverse Gaussian models it is equal for  $CV \approx 1.26$ . Nevertheless this fact cannot be used to judge the degree of similarity between the gamma and either the Pareto or inverse Gaussian models. The comparison of probability density functions  $f(t)$  of the inverse Gaussian, gamma and Pareto models can be seen in Fig. 1B for  $E(T) = 1$  s and  $CV \approx 1.86$ .

### 3.2 $R$ and specific information

Mutual information  $I(\mathcal{S}; \mathcal{R})$  (Cover and Thomas, 1991) determines the dependence between stimuli  $\mathcal{S}$  and responses  $\mathcal{R}$  (Borst and Theunissen, 1999). The information gained from a particular stimulus is known once the variability of responses across the whole set of stimuli is determined.  $I(\mathcal{S}; \mathcal{R})$  has no informative value if only one stimulus is presented. The coding efficiency of chosen stimulus can be judged

according to the deviation of the response from the spontaneous activity (Chacron et al., 2001), i.e., the most informative stimuli cause the largest difference. The information rate  $R$  provides a natural measure for this difference. Furthermore in the following we show that under certain conditions a link between  $R$  and  $I(\mathcal{S}; \mathcal{R})$  may be established.

The set of stimuli  $\mathcal{S} = \{s_i\}_{i=1}^n$  is discrete and the set of responses is realized by ISIs which can take any positive value. Mutual information can be formally expressed as  $I(\mathcal{S}; \mathcal{R}) = \sum_i p(s_i) i(\mathcal{R}|s_i)$ , where  $i(\mathcal{R}|s_i)$  is called the specific information due to the stimulus  $s_i$ . Analogously to DeWeese and Meister (1999) we express  $i(\mathcal{R}|s_i)$  as

$$i(\mathcal{R}|s_i) = h(\mathcal{R}) - h(\mathcal{R}|s_i). \quad (20)$$

From formula (20) it follows that the specific information is large for those stimuli that have only a few different responses associated with them because  $h(\mathcal{R}|s_i)$  is the uncertainty in response given stimulus  $s_i$ . If the stimulus  $s_i$  evokes only single possible response then it holds  $h(\mathcal{R}|s_i) = -\infty$ , because the probability density function of responses is realized by  $\delta$ -function.

We have restricted ourselves to the case in which the ISIs are described by a renewal process with probability density function  $f$  and the stimuli conditions are stationary in time. Under these two assumptions we can assign ISI distribution with density  $f$  to the stimulus  $s_i$  and the uncertainty in response becomes  $h(f) = h(\mathcal{R}|s_i)$ . The remaining term  $h(\mathcal{R})$  in formula (20) depends on the distribution of stimuli. It is possible to view  $h(\mathcal{R})$  as the entropy of the spontaneous neuronal activity  $h(\text{Spon})$ . The difference  $i(\text{Spon}|s_i) = h(\text{Spon}) - h(\mathcal{R}|s_i)$  does not add up to mutual information in the Shannon's sense. Nevertheless, the information rate computed employing the spontaneous activity differs only by a constant from the true mutual information rate and therefore the two measures behave similarly, see Chacron et

al. (2001) and Chacron et al. (2003) for details.

If the spontaneous activity is described by the Poisson process, then formula (20) corresponds to the expression for the information rate (12) and  $R$  coincides with specific information. If the spontaneous activity differs from the Poisson firing, then  $i(\text{Spon}|s_i)$  can be obtained from formula

$$i(\text{Spon}|s_i) = R(f) - R(\text{Spon}) = h(\text{Spon}) - h(f), \quad (21)$$

where  $R(\text{Spon})$  resp.  $R(f)$  are the information rates of the spontaneous activity and the activity in question. Note that we cannot directly employ a general (non-Poisson) spontaneous activity as the state of null information (which is required to maximize the entropy) because we would lose the interpretation in terms of information.

### 3.3 The spike train as a Markov chain

First we employ formula (9) to find the information rate  $R$  of a general stationary ISI model. Using the same reasoning as in the case of renewal process we find the state of null information to be the Poisson process. Simplifying equation (9) in this situation yields an expression similar to formula (11)

$$R = aE(T) - \ln a - \bar{h}(f) \quad (22)$$

and putting the mean ISI of the stochastic process described by  $f(t_1, t_2, \dots)$  and that of the state of null information equal, it yields

$$R = 1 + \ln E(T) - \bar{h}(f). \quad (23)$$

The available information rate of the general stationary stochastic process (gained beyond the idea of frequency coding) relative to the state of null information is again reduced to the calculation (estimation) of the entropy rate. (Obviously the relationship between  $R$  and the specific information holds also in the general stationary case.)

The computation of the entropy rate is however hardly possible in the general case (Cover and Thomas, 1991). Even though the limit is theoretically guaranteed to exist the convergence may be arbitrarily slow with increasing dimension of the joint probability density function. In the following we restrict ourselves to the class of stationary stochastic processes satisfying the first-order Markov property (Cox and Lewis, 1966)

$$\text{Prob}\{T_n \leq t_n | T_{n-1} = t_{n-1}, \dots, T_1 = t_1\} = \text{Prob}\{T_n \leq t_n | T_{n-1} = t_{n-1}\}. \quad (24)$$

In other words, each ISI depends only on the immediately preceding one and is conditionally independent of all other preceding ISIs. The Markov chain is therefore fully described by the joint probability density function  $f(t_1, t_2)$  of the two adjacent ISIs. Condition (24) simplifies the expression for the entropy rate and analogously to Cover and Thomas (1991, p. 66) we write

$$\bar{h}(f) = h(T_2|T_1) = - \int_0^\infty \int_0^\infty f(t_1, t_2) \ln f(t_2|t_1) dt_1 dt_2. \quad (25)$$

(To avoid indexing we denote  $T_1 \equiv X$  and  $T_2 \equiv Y$ .)

Formula (23) for the information rate  $R$  of the first-order Markov chain becomes

$$R = 1 + \ln E(Y) - h(Y|X). \quad (26)$$

Equation (26) can be rewritten using the mutual information  $I(X; Y)$ , the symmetric quantity that measures dependence between two random variables  $X, Y$  (Cover and Thomas, 1991)

$$I(X; Y) = h(Y) - h(Y|X). \quad (27)$$

Combining equations (26) and (27) separates the total information rate  $R$  into two parts

$$R = R_1 + I(X; Y), \quad (28)$$

where  $R_1 = h(g) - h(Y)$  is the information rate of the renewal process described by the marginal probability density function corresponding to the given Markov chain. The important property of mutual information is that  $I(X; Y) = 0$  if, and only if, the variables  $X$  and  $Y$  are independent, i.e., the Markov chain is reduced to the renewal process. Note that the above derived results can be extended to the  $k$ th-order class of Markov chains (conditional dependence on the  $k$  preceding states).

In the following we compare the information rates  $R$  of several Markov chain ISI models described by the probability density functions  $f(x, y)$ . We parameterize the rate  $R$  in dependence on the serial correlation  $\rho = [E(XY) - E(X)E(Y)]/[\sqrt{Var(X)}\sqrt{Var(Y)}]$ , which is frequently used to measure the dependence between variables  $X$  and  $Y$ . The resulting rate  $R$  is again independent of the expected value of the ISI.

### 3.3.1 Lawrance and Lewis model

The model introduced by Lawrance and Lewis (L-L) is described by the joint probability density function (Lawrance and Lewis, 1977)

$$f(x, y) = \frac{a^2 b}{1 - b + b^2} \left\{ U(bx - y) \frac{1 - b}{b^2} \exp \left[ -\frac{ab(x + y) - ay}{b^2} \right] + \exp \left[ -\frac{a(x + by)}{b} \right] + U(y - bx) \frac{(b - 1)^2}{b} \exp [-a(x - bx + y)] \right\} \quad (29)$$

with parameters  $a > 0$ ,  $b \in (0, 1)$  and  $U(x)$  the Heaviside unit step-function:  $U(x < 0) = 0$  and  $U(x \geq 0) = 1$ . The (first-order) serial correlation  $\varrho$  is

$$\varrho = b(1 - b), \quad (30)$$

from which the limitation of the L-L model follows: serial correlation is confined in interval  $(0, 1/4)$ . The marginal distribution of the L-L model is exponential with parameter  $a$ , see formula (10).

We investigate the information rate of the L-L model using formula (28). The term  $R_1$  is zero and thus the total rate  $R$  is equivalent to the mutual information,  $R = I(X; Y)$ . Two values of parameter  $b$  lead to the same value of the serial correlation in equation (30), thus we have to consider two solutions of  $R$  depending on which value of  $b$  was used

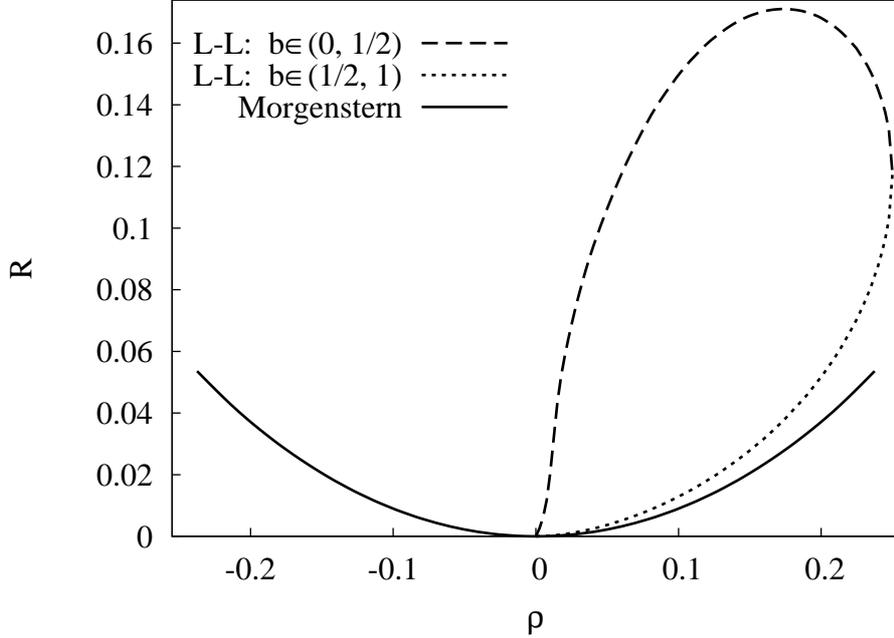
$$\begin{aligned} b \in (0, 1/2) : & \quad \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\varrho}, \\ b \in (1/2, 1) : & \quad \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\varrho}. \end{aligned}$$

The resulting  $R(\varrho)$  was carried out numerically and is plotted in Fig. 2. For  $b \in (1/2, 1)$  the rate increases monotonically from zero ( $R(\varrho = 0) = 0$ , the ISIs are independent) to its maximum value  $R(\varrho = 1/4) \approx 0.12$ . A more interesting result comes from examining the behavior of  $R$  for  $b \in (0, 1/2)$ . The maximum

value  $R \approx 0.17$  does not correspond to the maximum value of serial correlation, but is located at  $\rho \approx 0.17$ . From this value with increasing serial correlation we observe a decreasing rate, while the marginal distribution is still exponential. This seemingly paradoxical result comes from the fact that the value of serial correlation as a measure of dependency cannot be used to judge the degree of difference from the state of null information.

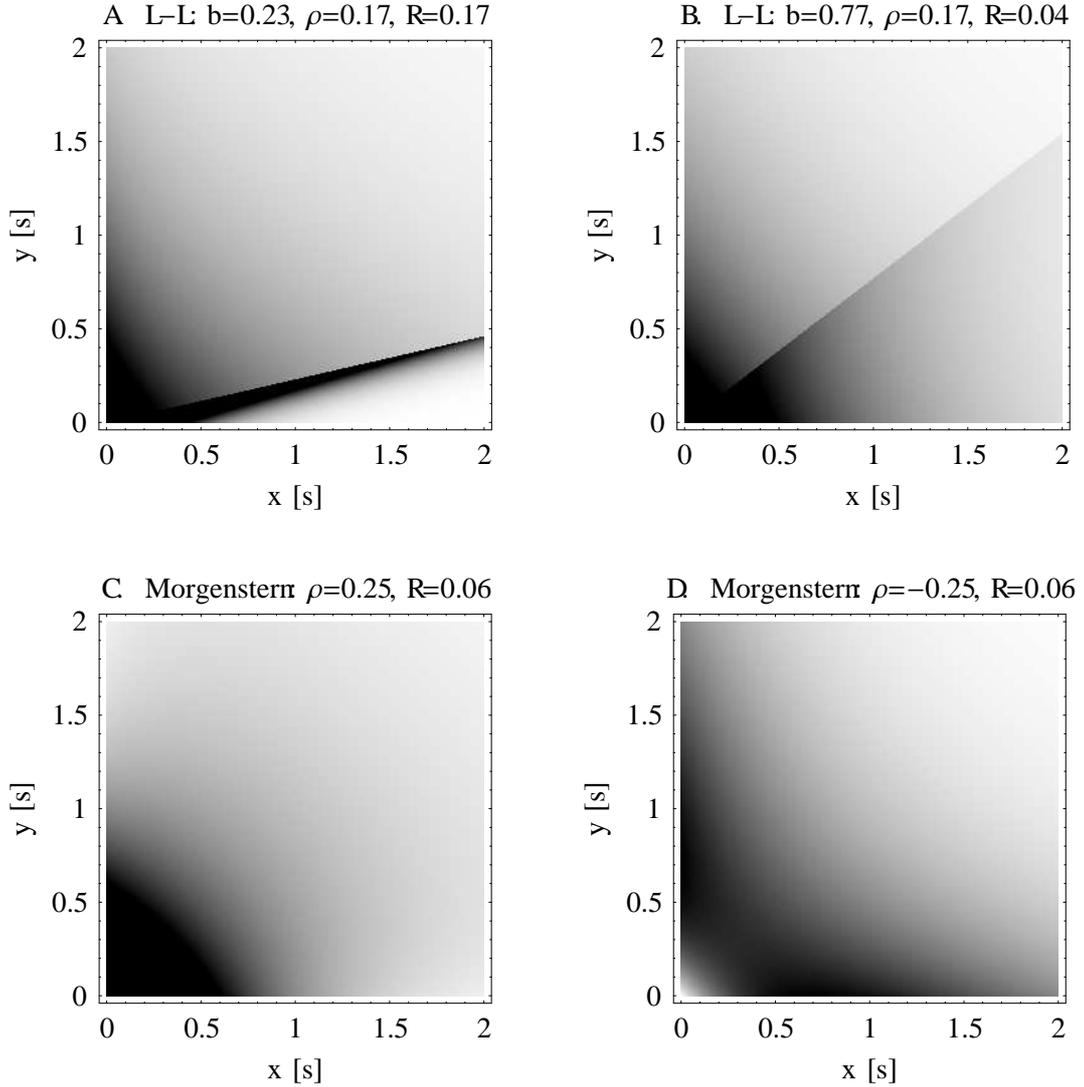
We can judge the qualitative behavior of the L-L model from the joint probability density plots shown in Fig. 3A and B. Though serial correlation is often found in experimental data, two-dimensional histograms corresponding to Fig. 3 are not presented. The length of the 'immediately preceding' ISI is given on the  $x$ -axis, the length of the current ISI is on the  $y$ -axis and the probability of their joint occurrence is indicated by the shade (a darker tone corresponds to a higher value, absolute numbers are not important). The positive serial correlation of the ISIs can be seen immediately for  $b = 0.23$  (A): short ISIs tend to be followed by short ones (the dark region for  $x < 0.5$  s and  $y < 0.5$  s), longer ISIs by comparatively long ones. Note especially the narrow dark band of highly probable  $(x, y)$  pairs in the lower part of the plot. This feature makes the existence of certain (short) sequences of ISIs more probable than others, i.e. it can be regarded as a simple mechanism of temporal pattern formation. Unfortunately, the serial correlation in this case is not large enough to make the effect visually pronounced (see also Fig. 4). Below the band of high probability one can see a triangle of nearly zero probability, e.g.,  $x > 1.5$  s can hardly be followed by  $y < 0.25$  s. The behavior of the L-L model for  $b = 0.77$  (B) differs in many aspects, though the value of serial correlation is the same as for  $b = 0.23$ . The sharp band of high probability is missing and the joint probability density is more diffused. Very short ISIs are often followed by even shorter ones than in the previous case (compare with Fig. 4D), but the preference for long ISIs following the longer ones ( $x > 1$  s) is not pronounced (note the darker

triangle below the diagonal). The difference between the two cases is captured in different values of the information rates  $R$ .

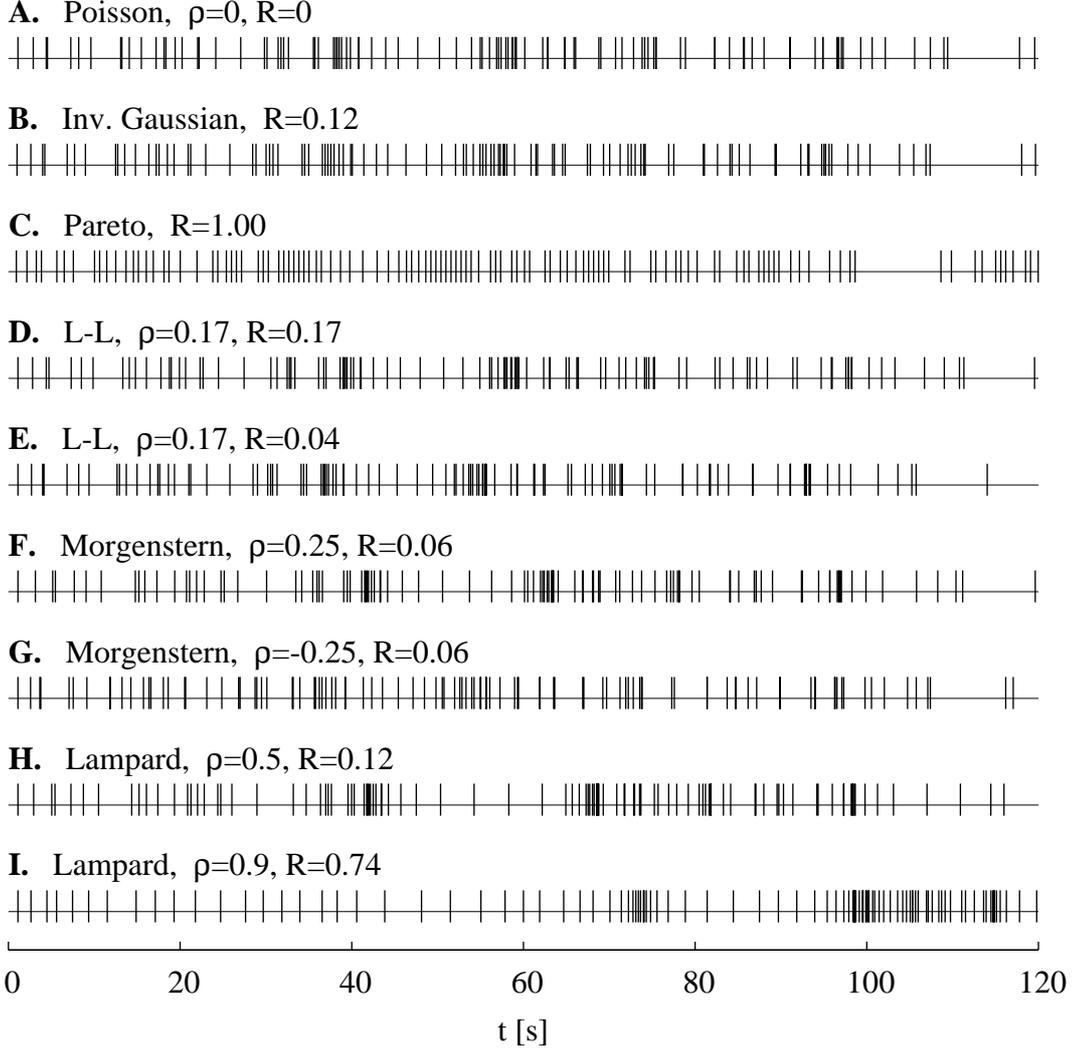


**Fig. 2:** The information rates  $R$  in dependence on the serial correlation  $\varrho$  of the Lawrance and Lewis (L-L) and Morgenstern ISI models. For  $b \in (1/2, 2)$  (the L-L model) the rate increases monotonically to its maximum  $R(\varrho = 1/4) \approx 0.12$ . For  $b \in (0, 1/2)$  the rate reaches its maximum at  $R(\varrho \approx 0.17) \approx 0.17$  and then decreases. The Morgenstern model exhibits simpler behavior: its rate is smaller compared to that of the L-L model at the corresponding value of  $\varrho$ .

The neuronal activity described by the L-L model is simulated in Fig. 4D for  $E(Y) = 1$  s and serial correlation  $\varrho = 0.17$  ( $b = 0.23$ ). Another realization corresponding to Fig. 3B is shown in Fig. 4E. Note that the simulations were always done with the same initial value of random seed. The relatively small differences from the Poisson case (Fig. 4A) are captured in comparatively small values of information rate  $R \approx 0.17$ .



**Fig. 3:** Density plots of the joint probability density functions  $f(x, y)$  of the Markov chain ISI models with  $E(Y) = 1$  s, darker tones correspond to higher values (absolute numbers are not important). The Lawrance and Lewis (L-L) model (A+B) is plotted for two different values of parameter  $b$ . Though the serial correlation  $\rho$  is the same in both situations the shapes of  $f(x, y)$  are different. This difference is captured by different values of the rate  $R$ . The Morgenstern model (C+D) is plotted for both extremal values of the serial correlation  $\rho = \pm 1/4$  with equal values of  $R \approx 0.6$ .



**Fig. 4.:** (*Caption on the following page.*)

### 3.3.2 Morgenstern model

The next model is constructed using the bivariate joint probability density function  $f(x, y)$  first described by Morgenstern (1956)

$$f(x, y) = a^2 e^{-2a(x+y)} [e^{a(x+y)} + 4\rho(e^{ax} - 2)(e^{ay} - 2)], \quad (31)$$

with parameters  $a > 0$  and  $\rho \in (-1/4, 1, 4)$ . The function  $f(x, y)$  is symmetric in its arguments  $f(x, y) = f(y, x)$ . The marginal distribution of this model is again

**Fig. 4:** Simulated spike trains for renewal models with  $CV = 1$  and Markov models with exponential marginal distributions in both cases  $E(T) = 1$ . The values of serial correlation  $\rho$  and information rate  $R$  are given. (A) The Poisson process. The inverse Gaussian model (B) is similar to (A) but lacking the extremely short ISIs (see the probability density plot in Fig. 1B). The Pareto model (C) differs strikingly from the Poisson process, with the main distinction being the 'dead time' (approx. 0.58s). The values of serial correlation  $\rho$  in the cases of the Lawrance and Lewis (D+E), Morgenstern (F+G) and Lampard models (H+I) are relatively too small to produce apparent change in comparison with the Poisson process – the effect is only slightly more pronounced for negative  $\rho$  (G) and for  $\rho = 0.5$  (H). The spike train generated according to Lampard model (I) with  $\rho = 0.9$  can be distinguished on the first sight.

exponential with parameter  $a$ . The serial correlation is equal to the parameter  $\rho$  in formula (31). The maximum serial correlation is again  $|\rho| = 1/4$ , but contrary to the previous case of L-L model it can also be negative.

We use the same approach in applying formula (26) as in the case of L-L model. The result was carried out numerically and is plotted for comparison with the L-L model in Fig. 2. The behavior of the Morgenstern model appears simple compared to that of the L-L model. The information rate is symmetric  $R(\rho) = R(-\rho)$  and  $R(\rho = 0) = 0$  implying that for zero serial correlation the ISIs are independent, as can be seen directly from formula (31). An interesting observation is that for the same value of  $\rho$  the rates  $R(\rho)$  of the Morgenstern and L-L models differ and that for any  $\rho \in (0, 1/4)$  the rates of the Morgenstern model are smaller.

Fig. 3C and D shows the  $f(x, y)$  for both extreme values of serial correlation. The behavior of the model is simpler compared to the L-L model, there are no

sharp regions of special interest and the probability density is smooth and symmetric around  $x = y$ . The positive correlation (C) is seen for ISIs smaller than 1 s, and then for large ISIs ( $x \approx 2$  s) which are usually not followed by very short ones. The reverse statements hold for the negative serial correlation (D). Very short ISIs are not followed by comparatively short ones (the light region for small  $x, y$ ) and very long ISIs ( $x > 2$  s) are preferably followed by shorter ones ( $y < 1$  s). Because the information rate is the same in both cases, we cannot distinguish between negative and positive serial correlation just by observing the value of  $R$ .

Neuronal firing that behaves according to the Morgenstern model is simulated in Fig. 4F and G. Due to the small values of serial correlation only small differences can be seen compared to the Poisson process (A). This is confirmed by a very small value of the information rate  $R \approx 0.06$ .

### 3.3.3 Lampard Model

The last example of the stationary Markov chain we employ as the ISI distribution model was first described by Lampard (1968). It describes a counter system whose inputs are a pair of independent Poisson processes. The advantage of this model over the previous two lies in its possible interpretation from the neurophysiological point of view (Lawrance, 1972).

The joint probability density function  $f(x, y)$  is defined by

$$f(x, y) = \frac{(1/\varrho - 1)^\xi}{xy\Gamma(\xi)} \left( \frac{a\xi^2 \sqrt{xy\varrho}}{|\varrho - 1|} \right)^{1+\xi} \exp \left[ \frac{a\xi(x+y)}{\varrho - 1} \right] I_{\xi-1} \left( \frac{2a\xi \sqrt{xy\varrho}}{1 - \varrho} \right), \quad (32)$$

where  $\varrho \in (0, 1)$  is the first-order serial correlation, parameters  $\xi > 0$ ,  $a > 0$  and  $I_\nu(z)$  is the modified Bessel function of the first kind (Abramowitz and Stegun, 1972). The marginal distribution of the model is the gamma distribution (compare with equation (13)):  $f(y) = (a\xi)^\xi y^{\xi-1} \exp(-a\xi y)/\Gamma(\xi)$ , from which follows that

parameter  $\xi$  is related to the *CV* by  $CV^2 = 1/\xi$ , and  $a$  describes the mean value of the ISI  $E(T) = 1/a$ . For  $\xi = 1$ , which implies  $CV = 1$ , the joint density  $f(x, y)$  is reduced to the Downton bivariate exponential density (Downton, 1970) with exponential marginal distributions:

$$f(x, y) = \frac{a^2}{1 - \varrho} \exp \left[ \frac{a(x + y)}{\varrho - 1} \right] I_0 \left( \frac{2a\sqrt{xy\varrho}}{1 - \varrho} \right) \quad (33)$$

To obtain the information rate  $R$  of the Lampard model we apply formula (28), because the term  $R_1$  is already given in the closed form by formula (14). Therefore the mutual information  $I = I(X; Y)$  remains to be calculated. We parameterize the mutual information  $I(\varrho)$  with the serial correlation (in agreement with the previous reasoning) and investigate its possible dependence on the *CV* through the remaining parameter  $\xi$ . The Bessel function  $I_\nu(z)$  can be expressed in a simple form for two particular values of parameter  $\nu$  (Abramowitz and Stegun, 1972)

$$\begin{aligned} I_{1/2}(z) &= \sqrt{\frac{2}{\pi}} \frac{\sinh z}{\sqrt{z}}, \\ I_{-1/2}(z) &= \sqrt{\frac{2}{\pi}} \frac{\cosh z}{\sqrt{z}}. \end{aligned}$$

The solution for the first condition  $\xi - 1 = 1/2$  is  $CV = \sqrt{2/3} \approx 0.816$  and for the second  $\xi - 1 = -1/2$  is  $CV = \sqrt{2} \approx 1.414$ . The joint density  $f(x, y)$  of the Lampard model then reduces to much simpler forms

$$CV = \sqrt{\frac{2}{3}} : f(x, y) = -\frac{9a^2}{2\pi\sqrt{(1-\varrho)\varrho}} \exp \left[ \frac{3}{2} \frac{a(x+y)}{\varrho-1} \right] \sinh \left( \frac{3a\sqrt{xy\varrho}}{\varrho-1} \right), \quad (34)$$

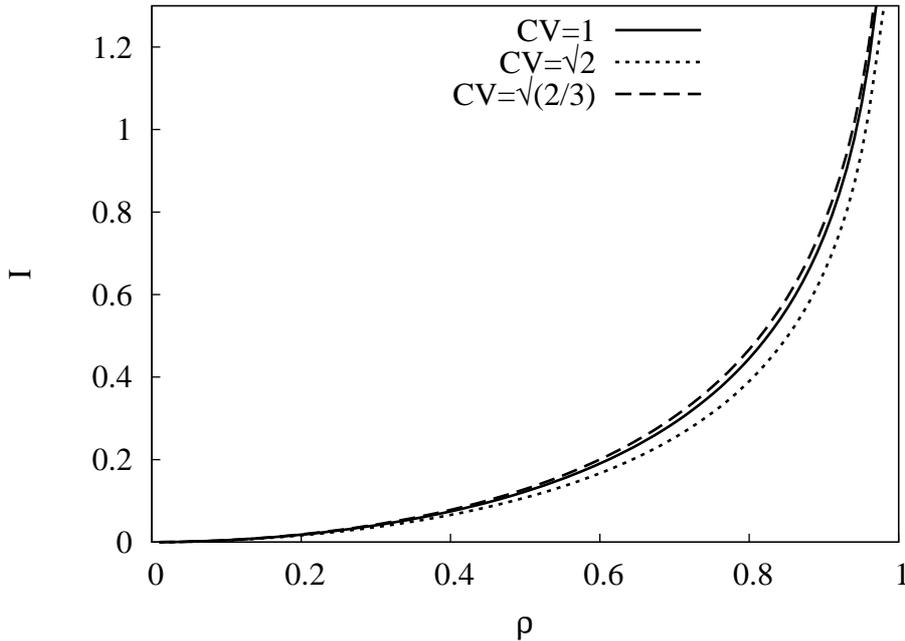
$$CV = \sqrt{2} : f(x, y) = \frac{a}{2\pi\sqrt{xy(1-\varrho)}} \exp \left[ \frac{1}{2} \frac{a(x+y)}{\varrho-1} \right] \cosh \left( \frac{a\sqrt{xy\varrho}}{\varrho-1} \right). \quad (35)$$

The reduced forms given by formulas (33) (for  $CV = 1$ ), (34) and (35) make the analytical integration of at least some parts in  $I(\varrho)$  possible. The remaining

integration can thus be carried out numerically with much better precision than using the general and rather complicated form (32).

The resulting mutual information  $I(\varrho)$  is plotted in Fig. 5 for the three above mentioned values of  $CV$ . The total information rate  $R$  can be computed using formula (28). For  $CV = 1$  and  $\varrho = 0$  the Lampard model reduces to the Poisson process, thus both  $R_1 = 0$  and  $I(X; Y) = 0$ . For  $CV = 1$  and  $\varrho > 0$  holds  $R_1 = 0$  and the mutual information corresponds to the total information rate  $R$  of the Lampard (=Downton) model. Similarly if  $CV \neq 1$  and  $\varrho = 0$ , then  $I(X; Y) = 0$  and the Lampard model reduces to the renewal process with gamma ISI distribution. Using formula (28) for the term  $R_1$  gives:  $R_1(CV = \sqrt{2/3}) \approx 0.044$  and  $R_1(CV = \sqrt{2}) \approx 0.216$ . The shapes of the curves  $I(\varrho)$  are very similar, starting from  $I(\varrho = 0) = 0$  and continuing with monotonous increase  $I(\varrho \rightarrow 1) \rightarrow \infty$ . Mutual information increases dramatically for  $\varrho > 0.8$ , where an arbitrarily small change in  $\varrho$  has a strong effect on the information rate, while the changes in the serial correlation value for  $\varrho < 0.5$  may be neglected. The dependence of  $I(\varrho)$  on the value of  $CV$  is relatively small, nevertheless, it cannot be attributed to numerical errors. For  $CV$  more deviated from one the effect is slightly more pronounced. Again we have a situation where the serial correlation  $\varrho$  is not a sufficient measure of the true dependence between two random variables, contrary to  $I(X; Y)$ .

A sample of neuronal firing behaving according to the Lampard model is shown in Fig. 4 for  $\varrho = 0.5$  (H) and  $\varrho = 0.9$  (I). At such a high value of serial correlation (especially in Fig. 4I) the pattern of spikes can be immediately distinguished from the other cases presented in the figure, though the marginal distribution (estimated, e.g., by histograms) of all the presented cases is exponential with equal parameters.



**Fig. 5:** The mutual information  $I = I(X; Y)$  of the Lampard model in dependence on serial correlation  $\rho$  for three values of  $CV$ . The shapes of the curves are very similar but not identical, which means that  $\rho$  does not measure the ISI dependence completely. A very sharp increase in  $I$  for  $\rho > 0.5$  suggests that low values of  $\rho$  have only marginal effect, while a small increase in  $\rho$  for  $\rho > 0.8$  can change  $R$  dramatically.

### 3.4 Experimental data

Formula (26) for the information rate  $R$  of the Markov chain can be expressed alternatively, substituting from the chain rule (4)

$$R = h(g) + h(X) - H(X, Y). \quad (36)$$

We see that the computation of  $R$  from experimental data for the renewal process and Markov chain is reduced to the estimation of entropy from one- and two-dimensional probability density functions. This makes equations (12) and (36)

applicable in experimental data analysis as the problem of entropy from data estimation is well exploited in literature, see, e.g., Beirlant et al. (1997); Tsybakov and Meulen (1996) for an overview of available techniques.

For one dimension (the renewal process) the simple and well researched Vasicek's estimator (Vasicek, 1976) gives reasonably good results on a wide range of data (Ebrahimi et al., 1992; Miller and Fisher III, 2003). Furthermore our own experience with simulated data shows that for sample sizes  $n = 500$  (the average size in the experimental data we used) the standard deviation is relatively small ( $\sigma < 0.07$ ) and a possible positive bias with respect to true values is negligible. It is also preferable to avoid estimations based purely on histograms because the choice of binwidth affects the results greatly. The support of ISI distributions is always positive which makes the application of kernel estimators problematic due to possible overlapping into negative values.

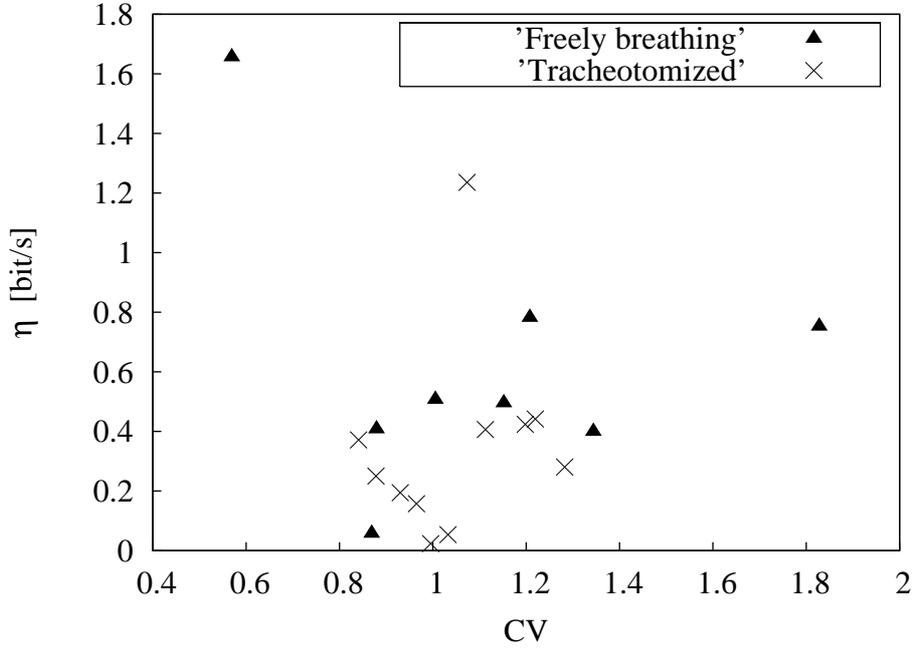
In this section we illustrate the use of information rate on experimental data in the case of renewal process. The data come from extracellular recordings made from olfactory receptor neurons of freely breathing and tracheotomized rats. Single-unit action potentials were recorded and more details on the data acquisition is described in Duchamp-Viret et al. (2003). The sample sizes range from (circa)  $n = 100$  to  $n = 2000$  ISIs and all records have been tested for stationarity and ISI independence (the Wald-Wolfowitz test, serial correlation, periodogram).

Given the  $n$  ranked ISIs  $\{t_{[1]} < t_{[2]} < \dots < t_{[n]}\}$  we used the entropy estimator proposed by Vasicek (1976)

$$h(\text{data}) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{n}{2m} (t_{[i+m]} - t_{[i-m]}) \right]. \quad (37)$$

The positive integer parameter  $m < n/2$  is set prior to computation and the two following conditions hold:  $t_{[i-m]} = t_{[1]}$  for  $(i-m) < 1$  and  $x_{[i+m]} = x_{[n]}$  for  $(i+m) > n$ .

The particular values of  $m$  corresponding to various values of  $n$  were determined by Ebrahimi et al. (1992).



**Fig. 6:** The information rate flow  $\eta$  in bits per second estimated from the experimental data in dependence on the  $CV$ . The stationary renewal activity of olfactory neurons in rats is compared for freely breathing (▲) (in the absence of any particular stimulus) and tracheotomized (×) animals. Except for two cases the  $\eta$  of tracheotomized animals is lower than in freely breathing ones. The results show that the activity in tracheotomized rats is closer to the state of null information (Poisson process) and that  $CV$  is not well related to the Poisson character of the process.

The information rate  $R$  represents the average information gained per ISI and does not depend on  $E(T)$ , i.e, the firing rate. To include the effect of faster vs. slower neuronal firing we examine the average distribution of information rate in

time. We define the information rate flow  $\eta$  by

$$\eta = \frac{1}{\ln 2} \frac{R}{E(T)}. \quad (38)$$

The factor  $1/\ln 2$  is used to change the logarithm base in equation (9) to 2, so the quantity  $\eta$  represents the average information gained (relative to the state of null information) in bits per second.

The estimated information rate flow  $\eta$  for the already mentioned two categories of data is plotted in Fig. 6. We see that the  $\eta$  in the tracheotomized case ( $\times$ ) is in most of the cases lower than that of the freely breathing ( $\blacktriangle$ ). This is verified independently by other methods in Duchamp-Viret et al. (2005), where also some further inferences from the data are made. Due to the properties of  $R$  low values of  $\eta$  indicate that the firing is close to the state of null information, therefore justifying the hypothesis that spontaneous firing is not informative.

## 4 Conclusions

The information rate  $R$  based on the Kullback-Leibler distance between two ISI models was proposed for a general case of stationary neuronal activity. If the reference state maximizes the entropy of the ISI probability distribution and if the mean values of both distributions are equal, then  $R$  measures the information rate per ISI due to the temporal coding scheme. Beside the introduction of  $R$  we proposed a related quantity, the information flow:  $\eta = R/E(T)$ , which measures the information gain per time unit. This quantity takes the firing rate of the neuron into account, thus even relatively small values of  $R$  must be taken into consideration when comparing the  $\eta$  of fast-firing neurons to the slower ones. The determination of  $R$  (and  $\eta$ ) requires the computation (estimation) of differential entropy which makes

the quantity applicable on suitable experimental or simulated data, as illustrated in the final section of this article. We showed that  $R$  is related to the mutual information and corresponds to the information due to specific stimulus.

We analyzed in detail several examples from two categories: the renewal process and the first-order Markov chain ISI models. The case of the renewal processes shows that even if neither spike frequency nor  $CV$  changes there still may be a gain of information. The chosen models behave differently for  $CV > 1$ , while for  $CV \ll 1$  their information rates  $R$  are very similar. On the other hand, the case of Markov chains indicates that if  $R$  is examined only due to the dependency among ISIs, small values of serial correlation  $\varrho$  imply small values of  $R$ . Moreover the increase in the serial correlation of ISIs does not necessarily increase the information rate. Though the Markov chain ISI models discussed here are not resulting from realistic neuronal models, it is nevertheless clear that the relation of serial correlation to the information rate is not simple. As the entropy rates  $\bar{h}$  were computed directly from the joint probability densities these Markov chain models serve well for testing purposes and further development of multi-dimensional entropy estimators. These estimators in turn may be used to estimate  $R$  in realistic models, where the joint probability density of ISIs is not available (including experimental data).

Theoretically the information rate  $R$  may tend to infinity. This is due to the fact that a continuous random variable generally carries an infinite amount of information (van der Lubbe, 1997, p. 171). Nevertheless this fact can be considered as merely formal and without consequences, in practice we are always working with finite precision on a finite time scale. Notably the cases  $CV \rightarrow 0$  or  $\varrho \rightarrow \pm 1$  would require an infinite timing precision of the neuronal firing.

## Acknowledgements

The authors wish to thank Patricia Duchamp-Viret and Michel Chaput for making their experimental data available. The authors are grateful to Nick Dorrel for linguistic correction. This work was supported by the Research project AV0Z 5011922, Center for Neuroscience LC554 and by the Academy of Sciences of the Czech Republic Grant (Information Society, 1ET400110401).

## References

- Abeles M 1982 *Local Cortical Circuits: An Electrophysiological Study*, Springer-Verlag, Berlin Heidelberg
- Abramowitz M, Stegun I A 1972 Handbook of mathematical functions, Dover, New York
- Adrian E D 1928 *The basis of sensation*, W. W. Norton, New York
- Beirlant J, Dudewicz E J, Györfi L, van der Meulen E C 1997 Nonparametric entropy estimation: an overview, *Int. J. Math. Stat. Sci.*, **6**, 17–39
- Bialek W, Rieke F, Van Steveninck R R, Warland D 1991 Reading a neural code, *Science*, **70**, 1629–1638
- Borst A, Theunissen F E 1999 Information theory and neural coding, *Nature neuroscience*, **2**, 947–957
- Buracas G T, Albright T D 1999 Gauging sensory representations in the brain, *Trends Neurosci.*, **22**, 303–309
- Chacron M J, Longtin. A, Maler L 2001 Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli, *J. of Neurosci*, **21**, 5328–5343
- Chacron M J, Longtin. A, Maler L 2003 The effects of spontaneous activity,

- background noise, and the stimulus ensemble on information transfer in neurons, *Network: Comput. Neural Syst.*, **14**, 803–824
- Cover T M, Thomas J A 1991 *Elements of information theory*, John Wiley & sons Inc., New York
- Cox D R, Lewis P A W 1966 *The Statistical Analysis of Series of Events*, John Wiley & Sons, Inc., New York
- DeWeese M R, Meister M 1999 How to measure the information gained from one symbol, *Network: Comput. Neural. Syst.*, **10**, 325–340
- Downton F 1970 Bivariate exponential distributions in reliability theory, *J. Roy. Statist. Soc. B*, **32**, 408–417.
- Duchamp-Viret P, Duchamp A, Chaput M A 2003 Single olfactory sensory neurons simultaneously integrate the components of an odour mixture, *Eur. J. Neurosci.*, **10**, 2690–2696
- Duchamp-Viret P, Chaput M A, Kostal L, Lánský P, Rospars J-P 2005 Patterns of spontaneous activity in single rat olfactory receptor neurons are different in normally breathing and tracheotomized animals, *Journal of Neurobiology*, **65**, 97–114
- Ebrahimi N, Habibullah M, Soofi E S 1992 Testing exponentiality based on Kullback-Leibler Information, *J. Statist. Soc. B*, **3**, 739–748
- Fuller M S, Looft F 1984 An information theoretic analysis of cutaneous receptor responses. *IEEE Trans. on Biomed. Eng.*, **314**, 377-383
- Gerstein G, Mandelbrot B 1964 Random walk models for the spike activity of a single neuron, *Biophys. J.*, **4**, 41–68
- Gerstner W, Kistler W 2002 *Spiking neuron models*, Cambridge, Cambridge University Press
- Kostal L, Lansky P 2006 Similarity of interspike interval distributions and information gain in a stationary neuronal firing, *Biol. Cybern.*, **94**, 157–167

- Lampard D G, 1968 A stochastic process whose successive intervals between events form a first order Markov chain *J. Appl. Probab.*, **5**, 648–668
- Lánský P, Rodriguez R 1999 Two-compartment stochastic model of a neuron, *Physica D*, **132**, 267–286
- Lánský P, Rodriguez R, Sacerdote L 2004 Mean instantaneous firing frequency is always higher than the firing rate, *Neural Comput.*, **16**, 477–489
- Lawrance A J 1972 Some models for stationary series of univariate events, *In: Stochastic point processes: statistical analysis, theory and applications*, ed. by P.A.W. Lewis, John Wiley & Sons, Inc., 199–256
- Lawrance A J, Lewis P A W 1977 An exponential moving-average sequence and point process (EMA1), *J. Appl. Probab.*, **14**, 98–113
- Levine M W 1991 The distribution of the intervals between neural impulses in the maintained discharges of retinal ganglion cells, *Biol. Cybern.*, **65**, 459–467
- Lindner B 2004 Interspike interval statistics of neurons driven by colored noise, *Phys. Rev. E*, **69**, 022901
- Longtin A, Racicot D M 1999 Assesment of linear and non-linear correlations between neural firing events, *In: Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, eds. C.D. Cutler and D.T. Kaplan, *Fields Institute Communications*, **11**, 223–239
- van der Lubbe J C A 1997 *Information theory*, Cambridge University Press, Cambridge
- Mandl G 1992 Coding for stimulus velocity by temporal patterning of spike discharges in visual cells of cat superior colliculus, *Vision Res.*, **33**, 1451–1475
- Middlebrooks J C, Clock A B, Xu L, Green D M 1994 A panoramic code for sound location by cortical neurons. *Science*, **264**, 842–844
- Miller E G, Fisher III J W 2003 ICA using space estimates of entropy, *J of Machine Learning Res*, **4**, 1271–1295

- Morgenstern D 1956 Einfache Beispiele Zweidimensionaler Verteilungen, *Mitt. Math. Stat.*, **8**, 234–235
- Perkel D H, Bullock T H 1968 Neural coding: a report based on a NRP work session, *Neurosci Res Program Bull*, **6**, 221-248
- Rieke F, Warland D, de Ruyter van Steveninck R R, Bialek W 1997 *Spikes: exploring the neural code*, Cambridge, MIT Press
- Sakai Y, Funanashi S, Shinomoto S 1999 Temporally correlated inputs to leaky integrate-and-fire models can reproduce spiking statistics of cortical neurons, *Neural Networks*, **12**, 1181–1190
- Shannon C 1948 A mathematical theory of communication, *AT&T Tech. J.*, **27**, 379–423
- Tarantola A 1994 *The Inverse Problem Theory*, Elsevier, Amsterdam
- Tarantola A, Valette B 1982 Inverse Problems = Quest for Information, *J. Geophys.*, **50**, 159–170
- Theunissen F, Miller J P 1995 Temporal encoding in nervous systems: a rigorous definition *J. Comput. Neurosci.* **2**, 149–162
- Tsybakov A B, van der Meulen E C 1996 Root- $n$  Consistent Estimators of entropy for densities with unbounded support, *Scand. J. Statist.*, **23**, 75–83
- Vasicek O 1976 A test for normality based on sample entropy, *J. Statist. Soc. B*, **38**, 54–59
- Zador A 1998 Impact of synaptic unreliability on the information transmitted by spiking neurons, *J. Neurophysiol.*, **79**, 1219–1229