

How Far can Neural Correlations Reduce Uncertainty? Comparison of Information Transmission Rates for Markov and Bernoulli Processes

Agnieszka Pregowska

*Institute of Fundamental Technological Research
Polish Academy of Sciences, ul. Pawinskiego 5B, 02-106 Warsaw, Poland
aprego@ippt.pan.pl*

Ehud Kaplan

*Icahn School of Medicine at Mount Sinai
One Gustave Levy Place, New York, NY 10029, USA
Department of Philosophy and History of Science, Faculty of Science
Charles University, Albertov 6, 128 43 Praha 2, Czech Republic
The National Institute of Mental Health
Topolová 748, 250 67 Klecany, Czech Republic
ehud.kaplan@gmail.com*

Janusz Szczepanski*

*Institute of Fundamental Technological Research
Polish Academy of Sciences, ul. Pawinskiego 5B, 02-106 Warsaw, Poland
jszczepa@ippt.pan.pl*

Accepted 9 January 2019
Published Online 7 March 2019

The nature of neural codes is central to neuroscience. Do neurons encode information through relatively slow changes in the firing rates of individual spikes (rate code) or by the precise timing of every spike (temporal code)? Here we compare the loss of information due to correlations for these two possible neural codes. The essence of Shannon's definition of information is to combine information with uncertainty: the higher the uncertainty of a given event, the more information is conveyed by that event. Correlations can reduce uncertainty or the amount of information, but by how much? In this paper we address this question by a direct comparison of the information per symbol conveyed by the words coming from a binary Markov source (temporal code) with the information per symbol coming from the corresponding Bernoulli source (uncorrelated, rate code). In a previous paper we found that a crucial role in the relation between information transmission rates (ITRs) and firing rates is played by a parameter s , which is the sum of transition probabilities from the no-spike state to the spike state and *vice versa*. We found that in this case too a crucial role is played by the same parameter s . We calculated the maximal and minimal bounds of the quotient of ITRs for these sources. Next, making use of the entropy grouping axiom, we determined the loss of information in a Markov source compared with the information in the corresponding Bernoulli source for a given word length. Our results show that in the case of correlated signals the loss of information is relatively small, and thus temporal codes, which are more energetically efficient, can replace rate codes effectively. These results were confirmed by experiments.

Keywords: Shannon information theory; information source; information transmission rate; firing rate; neural coding.

*Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

1. Introduction

Since the seminal work of Adrian and Zotterman¹ it has been recognized that in the nervous system information is transmitted among spiking neurons by trains of discrete electrical pulses, called action potentials or spikes. A second, nondigital, mode of communication, through gap junctions, is also common (see, for example, Refs. 2 and 3) but will not be considered here. It has been shown that firing rates of spikes change in a consistent manner with inputs. That has given rise to the notion that information is encoded in the neuronal firing rate. Recently, Wang *et al.*⁴ presented evidence that in the cerebral cortex various types of neural coding are controlled by the energy field and energy flow. Other work has shown that firing rate is directly related to energy cost,⁵ namely, energy consumption increases linearly with spiking frequency.^{6,7} On the other hand, many reports show (for example, Ref. 8) that the total number of spikes varies substantially from trial to trial during repeated presentations of the same stimulus. This variability has given rise to an alternative hypothesis, which states that additional information is contained in the precise timing of the spikes within the spike train. These two (not mutually exclusive) views of neural encoding and decoding are broadly called “rate-based” and “spike-based”.⁹

The question arises which neural mechanism(s) could be used to switch from rate to temporal coding and *vice versa*. It seems that the most natural is an application of certain learning mechanisms. This question can be considered on two levels. First, on the level of a single neuron and various biological processes, including ion channel mechanisms governing spike generation, and second, on the level of spiking neural networks (SNNs) architecture,¹⁰ with mechanisms of learning that allow to optimize information transmission processes. During the past few years artificial neural networks were developed, inspired by the structure and function of the human brain, with applications to learning and effective classification, pattern recognition, approximation problems, time series forecasting and optimization. Recent advances and the availability of computational power have increased the development of new, increasingly realistic biological models, and made it possible to develop more suitable algorithms. Therefore, the development of spiking

neural networks is the most natural and promising idea to get a new generation of much more efficient networks.¹⁰ This includes analysis of the role of models and types (e.g. excitatory, inhibitory) of neurons being adapted, network architecture and progress in learning methods, which are used to enhance the effectiveness of information processing. Efficient single-spiking SNN models were presented in Ref. 11, and applied to complicated pattern recognition problems, such as epileptic seizure detection. An extensive parametric analysis with application of new training algorithms was performed, to identify heuristic rules and optimum parameter values that yielded high accuracy (92.5%) in the EEG classification problem. In Ref. 12 a new multi-spiking neural network and a very effective training algorithm were presented. A novel network architecture and back-propagation training algorithm with learning mechanism, which allows communication via spike trains, was proposed. This idea demonstrates the training of a neural network, which exploits an increased degree of biological feasibility. The classification abilities of this brain-inspired algorithm were evaluated using three complex problems: the XOR problem, the Fisher iris problem and the problem of detecting epilepsy and seizures in EEG records. The simulations have shown their effectiveness compared to traditional artificial neuronal networks or simplistic single-spiking SNN neurons. In Ref. 13 a novel idea was developed, based on the integration of three different computing technologies, namely, neural networks, wavelets and chaos theory. These powerful computational algorithms were applied to epilepsy diagnosis and seizure detection, being modeled as effective classification of EEG signals. The spiking neural networks with their inherent property of brain dynamics constitute, in fact, a new generation of neural networks, which could provide explanations for dynamics in the cerebral cortex. This approach benefited from understanding cerebellar pathologies. In Ref. 14 a novel learning algorithm was proposed, based on an artificial system embedding the salient neuronal and plastic properties of the cerebellum, and operating in a closed loop. These closed-loop simulations reproduced several aspects of cerebellar pathologies revealed in human and animal experiments. In Ref. 15 a spatiotemporal-based learning algorithm named cross-correlated delay shift was proposed as a supervised learning method to learn

the association between precise patterns considering axonal and synapse delays with weight modulation. The efficacy of the proposed learning rules was validated through a real-world example, involving the detection of interictal spikes in EEG data from patients with epilepsy. An interesting weight-free learning algorithm that relies solely on adjusting the spatiotemporal delivery of neuron firings was proposed in Ref. 16. The idea of training neural networks by controlling cell activity, rather than synaptic weights, was motivated by techniques such as optogenetics that are having a significant impact in neuroscience. The advantage of the algorithm is that plasticity is altered by controlling cell activity, and thus it is potentially realizable in living biological neuronal networks via intracellular stimulation. Another mechanism of learning in which changes in the synaptic strength depend on the temporal correlations between the timing of the presynaptic and postsynaptic action potentials was proposed and exploited in Ref. 17. A biophysical model is used, in which the microcircuits of two neocortical neurons were coupled through synapses with spike-timing-dependent plasticity that produced the mirror function. This result indicates that hyperpolarized activated cationic current channelopathy may affect the mirror function in autism spectrum disorder (ASD), through a mechanism mediated by spike-timing-dependent plasticity. Therefore, the present model is useful in the study of the physiological mechanisms behind ASD. Recently, authors have proposed and developed network models named synfire chains¹⁸ and spiking neural P systems.¹⁹ Synfire chains address multi-stage signal transmission in the brain as a connectivity scheme that connects a sequence of neuron groups of roughly the same size. Spiking neural P systems are characterized by the fact that each neuron mimics a biological cell, and the communication between neurons is based on spikes. In Ref. 20 it was shown that with spike-timing-dependent plasticity, modulated by global population activity, long Synfire chains emerge in sparse random networks. The sparse network architecture prevents the chains from being short and cyclic, and shows that the formation of specific synapses is not essential for chain formation. A novel communication strategy on request among neurons in spiking neural P systems was proposed in Ref. 19. It was proven that such systems

are computationally universal, i.e. they are equivalent to Turing machines, as long as two types of spikes are used. Moreover, in Ref. 21 it was shown that even simplified communication systems (with a single type of neuron) on request still have powerful computation capabilities. Promising is the recent application of spiking neural networks for modeling cerebellar circuit,²² which is the most plastic structure of the brain, and plays a critical role in adaptive motor control by implementing three fundamental operations: prediction, timing and learning. An accurate analysis of model parameters, during different phases of the eye blink classical conditioning learning paradigm, showed that the main effect of transcranial magnetic stimulation was to alter plasticity at cortical synapses.

In general, learning mechanisms in the field of artificial neural networks refer to changes of synaptic connections, which are driven by the statistics of the input stimuli (unsupervised learning), or are based on reward-based learning, where the network parameters are optimized to achieve, for each stimulus, a desired behavior (supervised learning). Our results show that in studying the type of information transmission (rate coding *versus* temporal coding) in spiking neural P systems, an important role is played by the parameter s , which we called the “jumping” parameter. Taking into account that for experimentally observed typical firing frequencies the parameter s is close to 1, our analysis indicates that the amounts of information transmitted by both correlated and uncorrelated signals are comparable. For nontypical firing frequencies, as in the case of bursting, which is observed experimentally,²³ the value of s can be far from 1, and in this case the level of correlation of spikes can affect in a visible way the encoding method and the transmission rate being achieved.

In Ref. 24 it was shown that the temporal correlation between spikes determines the information content in finite-memory neural codes, both within a spike train and across spike trains produced in response to repeated presentations of the same stimulus. In Ref. 25, a specific transformation of spike trains into analog signals was applied to account for the fact that a spiking neuron is able to learn. It has been argued that associations of arbitrary spike trains in a supervised fashion allow the processing of spatiotemporal information encoded in the precise timing of spikes.

In our previous papers we compared directly information transmission rates (ITRs) with their corresponding firing rates^{26,27} in the case of binary information sources. Our results show that a parameter s (which we called a “jumping parameter”) played a crucial role in the characterization of neural coding. We showed out that depending on this parameter s , temporal coding can be more effective than rate coding.²⁷

In this paper, we compare transmission rates for two types of binary information sources: correlated sources and their corresponding independent sources. Making use of the entropy grouping axiom,²⁸ we analyze the relation of information transmitted by sources described as Markov processes and by those described as Bernoulli processes. Our results show that in this case too a crucial role is played by the parameter s . We found bounds for the quotients of ITRs for these sources, and also their quotients’ minimal and maximal values. We also determined the loss of Shannon information in Markov sources *versus* the corresponding Bernoulli sources for a given word length.

The paper is organized as follows. In Sec. 2, we briefly review the basic concepts of Information Theory, Bernoulli and Markov processes. In Sec. 3, we present the comparison of ITRs of spike trains coming from a Markov source and from the corresponding Bernoulli source. Section 4 contains the final remarks.

2. Entropy and Information

In Shannon’s theory a communication system is represented by: an input information source (stimuli source), a communication channel (neuronal network) and an output information source (output signal). In mathematical language sources of information are modeled as stationary discrete stochastic processes.²⁹ Discrete communication channels are defined by a system of conditional probabilities linking input and output symbols.^{28,30} In this paper, we study two types of output information sources, i.e. sources represented by Markov processes and by corresponding Bernoulli processes.^{28,31} First, we briefly recall the basic notation.²⁷

2.1. Entropy

Let Z^L be a set of all words (i.e. blocks) of length L , built of symbols (letters) from some finite

alphabet Z . Each word w can be treated as a message sent by information source \mathbf{Z} being a stationary stochastic process. If $P(w)$ denotes the probability the word $w \in Z^L$ occurs, then the information in the Shannon sense carried by this word is

$$I(w) := -\log_2 P(w). \quad (1)$$

Thus, the average information of the random variable W^L corresponding to the words of length L is called the Shannon block entropy, and is given by

$$H(W^L) := - \sum_{w \in Z^L} P(w) \log_2 P(w). \quad (2)$$

Since the word length L can be arbitrary, the block entropy does not perfectly describe the information source.^{28,30}

In the special case of a two-letter alphabet $Z = \{0, 1\}$ and the length of words $L = 1$ we introduce the following notation:

$$H_2(p) := H(W^1) = -p \log_2 p - (1-p) \log_2 (1-p), \quad (3)$$

where $P(1) = p$, $P(0) = 1-p$ are the associated probabilities.

2.2. Information transmission and firing rates

The appropriate measure for estimation of transmission efficiency of an information source \mathbf{Z} is the information transmitted on average by a single symbol, i.e. ITR.^{28,30}

Let us introduce the notation

$$\text{ITR}^{(L)}(\mathbf{Z}) := \frac{H(W^L)}{L}, \quad (4)$$

and in the limiting case

$$\text{ITR}(\mathbf{Z}) := \lim_{L \rightarrow \infty} \frac{H(W^L)}{L}. \quad (5)$$

This limit exists if and only if the stochastic process \mathbf{Z} is stationary.³⁰

The most commonly used definition of firing rate refers to the temporal average³²⁻³⁴ and is defined as

$$F_R = \frac{n_T}{T}, \quad (6)$$

where n_T denotes spike count in a given time window of length T (typically a few seconds). In practice, in order to get sensible averages, some reasonable number of spikes should occur within the time window.³⁴ Since the messages are treated as trajectories

of locally stationary stochastic process, the firing rate as defined by (6) is specific for a given information source provided T is large enough. Thus, $F_R \cdot \Delta\tau$ is related to the probability p of spike appearance, where $\Delta\tau$ is the time resolution or bin size.

2.3. Information sources

An information source must produce sequences of symbols, which from a mathematical point of view can be treated as trajectories of a stationary stochastic process $\mathbf{Z} = (Z_i)$, $i = 1, 2, \dots$, where Z_i are random variables³¹ taking the values from a finite alphabet.

Spike-train statistics are often modeled as stochastic point processes.^{35,36} Specifically, there is an extensive literature in which the spike interval distribution is modeled by Poisson point processes, see e.g. Refs. 37 and 38. Poisson point processes provide a good approximation to experimental data, especially when the refractory time scale, or more generally, any memory time scale in the spike generating mechanism, is short compared to the time scales such as the mean interspike interval. The application of the Poisson processes for modeling spike trains was proposed since the earliest descriptions,^{39,40} due to the proportional relationship between the mean and variance of many neural responses. The Poisson property was observed in a number of experimental data.⁴¹ On the other hand, it is known that Poisson point processes exhibit Markov properties.^{42,43} Moreover, there are three features of neural responses that have been shown to contribute to decreasing variability, including the refractory period,⁴⁴ bursting⁴⁵ and temporal correlations in neural responses on longer time scales.⁴⁶ All of these features can be modeled by including directly the history (i.e. applying the Markov property) dependence in spike-train prediction. Only in the past three decades have other models started to come into use to describe spike trains, such as the inhomogeneous Markov interval models^{47,48} and hidden Markov models (HMMs).^{49,50} In these processes, both the current experimental time and the time from the last spike are taken into account. Thus, the question of how spike train correlations affect information transmission is of high importance. Generalized additive models, based on the Markov interval models approach, was successfully proposed in

Ref. 47. The idea is based on direct estimation of the part of the model describing the history-dependent properties of spike generation, under the assumption of a constant rate, followed by an estimation of the modulatory part describing the response properties. Radons *et al.*⁴⁹ employed HMMs for decoding the identity of visual stimuli from recorded neural responses. Using this procedure, they were able to identify the presented visual stimulus with high accuracy. In Ref. 50, using hidden Markov model, the neural state transitions were detected for motor cortical prostheses. A technique was developed to automatically differentiate between baseline, plan and perimovement epochs of neural activity. A four-state modulated Markov process was used to model spontaneous neuronal firings of individual auditory cortical neurons.⁴⁸ Analytical expressions for the probability density functions that describe the interspike interval distribution and autocorrelation function were derived. The good agreement between experimental and theoretical interspike interval histograms and the autocorrelation functions allows interpretation of the system's parameters of the individual neurons in terms of slow and delta waves, and high-frequency oscillations observed in cortical networks.

The most commonly used method of digitalization spike trains was proposed in Refs. 38, 51 and 52. It is physically justified since spike trains are recorded with a finite time resolution $\Delta\tau$, so that in each time slice (bin) a spike is either present or absent. If the presence of spike is denoted by "1" and no spike by "0", then if we look at some time interval of length T , each possible spike train is equivalent to $\frac{T}{\Delta\tau}$ binary sequence which can be treated as trajectory of the stochastic process.

In Refs. 53 and 54 it was assumed that random variables which describe the generation of consecutive bits in the sequence representing spike train are independent. This means that these random variables are uncorrelated, i.e. their Pearson correlation coefficient (PCC) is equal to 0. Thus, assuming that 1 is generated with probability p (a spike is found in the bin), 0 is generated with probability $1 - p$ (a spike is not found), what we have is a Bernoulli process.^{30,31} Clearly, in the case of a Bernoulli process the distribution of k "ones" between the sequence of bits of length n does not influence the probability of such sequences. This probability is simply equal, for all such sequences, to $p^k(1 - p)^{n-k}$, and depends

only on the firing rate $\frac{k}{n}$. Consequently, since the Shannon information depends only on the probability, all such sequences transmit the same amount of information and we are in the *rate code* regime.

Following the entropy definition (2), the information transmission rates (4) and (5) for the Bernoulli process \mathbf{B} with the probability of bit 1 equal to p are determined as

$$\text{ITR}(\mathbf{B}(p)) = -p \log_2 p - (1-p) \log_2 (1-p). \quad (7)$$

Note that $\text{ITR}(\mathbf{B}(p))$ is equal to $H_2(p)$.

Now, let us assume that the generation of bits of the output signal from an information source is described by correlated random variables (in the sense of *PCC*), and this generation is governed by a Markov process \mathbf{M} . In general, a discrete Markov process is defined by a set of conditional probabilities $p_{j|i}$ describing changes from state i to the state j , where $i, j = 0, 1$, and by initial distribution probabilities. These changes are called transitions, and the probabilities associated with them are called transition probabilities. These probabilities can be put together into a matrix \mathbf{P} , called the transition matrix, which for the two-states process is of the form

$$\mathbf{P} := \begin{bmatrix} p_{0|0} & p_{0|1} \\ p_{1|0} & p_{1|1} \end{bmatrix} = \begin{bmatrix} 1 - p_{1|0} & p_{0|1} \\ p_{1|0} & 1 - p_{0|1} \end{bmatrix}. \quad (8)$$

This is a stochastic matrix, i.e. each of its columns sums to 1. Here, we assumed that the process is homogeneous in time. The probability evolution is governed by the master equation⁵⁵

$$\begin{bmatrix} p_{n+1}(0) \\ p_{n+1}(1) \end{bmatrix} = \begin{bmatrix} 1 - p_{1|0} & p_{0|1} \\ p_{1|0} & 1 - p_{0|1} \end{bmatrix} \cdot \begin{bmatrix} p_n(0) \\ p_n(1) \end{bmatrix}, \quad (9)$$

where n stands for the discrete time, $p_n(0)$ and $p_n(1)$ are probabilities of finding states “0” and “1” at time n , respectively. In the case of Markov processes the distribution of k “ones” between the sequence of bits of length n does influence the probability of such sequences. Consequently, since the Shannon information depends only on the probabilities, in general such sequences transmit different amounts of information. Here the patterns in the sequences of bits do play a role and we are in the *temporal code* regime.

The stationary solution of (9) is given by

$$\begin{bmatrix} p_{\text{eq}}(0) \\ p_{\text{eq}}(1) \end{bmatrix} = \begin{bmatrix} \frac{p_{0|1}}{(p_{0|1} + p_{1|0})} \\ \frac{p_{1|0}}{(p_{0|1} + p_{1|0})} \end{bmatrix}. \quad (10)$$

The entropy rate (6) of the Markov source with transition matrix defined by (8) reads³⁰ by definition

$$\begin{aligned} \text{ITR}(\mathbf{M}) &= P_{\text{eq}}(0)[-p_{1|0} \log_2 p_{1|0} \\ &\quad - (1 - p_{1|0}) \log_2 (1 - p_{1|0})] \\ &\quad + P_{\text{eq}}(1)[-p_{0|1} \log_2 p_{0|1} \\ &\quad - (1 - p_{0|1}) \log_2 (1 - p_{0|1})], \end{aligned} \quad (11)$$

or, making use of notation (3), it can be written in a compact form

$$\text{ITR}(\mathbf{M}) = p_{\text{eq}}(0)H_2(p_{1|0}) + p_{\text{eq}}(1)H_2(p_{0|1}). \quad (12)$$

For the latter, the use of the probability of state “1” is, in fact, understood as the firing rate, and is denoted by p ,

$$p := p_{\text{eq}}(1) = \frac{p_{1|0}}{(p_{0|1} + p_{1|0})}. \quad (13)$$

For the special case when $p_{0|1} + p_{1|0} = 1$, the Markov process \mathbf{M} becomes uncorrelated, and reduces to a Bernoulli process with $p = p_{1|0}$.

Under the above notation, we introduced²⁷ the “jumping” parameter s , which can be interpreted as the tendency of a transition from one state to the other state

$$s := p_{0|1} + p_{1|0}. \quad (14)$$

Note that $0 \leq s \leq 2$.

Using this notation, in the case of the Markov processes \mathbf{M} , we have

$$p = \frac{p_{1|0}}{s}. \quad (15)$$

For $0 \leq s \leq 1$, the firing frequency p can take on values from the interval $[0, 1]$, while for $1 \leq s \leq 2$ the values of p are limited to the interval $1 - \frac{1}{s} \leq p \leq \frac{1}{s}$. These limits of the range of p follow from (14), (15) and the inequality $s - 1 \leq p_{1|0} \leq 1$.

3. Results

In this section we compare directly the information transmission rates conveyed by spike trains coming from a Markov information source, as defined by (12), with ITR of spike trains coming from the

corresponding Bernoulli information source (7). It is natural to assume that the Bernoulli process corresponding to a given Markov process is defined by the stationary probabilities (10) of this Markov process.

Under the notations (7) and (12), we introduce the following Markov–Bernoulli information quotient Q_s , which is a function of p and depends on parameter s :

$$Q_s(p) := \frac{\text{ITR}(\mathbf{M}_s(p))}{\text{ITR}(\mathbf{B}_s(p))}. \quad (16)$$

Applying (12) and (15) we have

$$Q_s(p) := \frac{(1-p)H_2(ps) + pH_2((1-p)s)}{H_2(p)}. \quad (17)$$

Taking into account the range of p , we consider Q_s in the following two cases:

$$(a) \quad 0 \leq s \leq 1, \quad \text{here } 0 \leq p \leq 1, \quad (18)$$

$$(b) \quad 1 \leq s \leq 2, \quad \text{here } 1 - \frac{1}{s} \leq p \leq \frac{1}{s}. \quad (19)$$

In Fig. 1 we present, for some arbitrary values of s , typical traces of Q_s as a function of p both for lower values of the jumping parameter $0 \leq s \leq 1$ [panel (a)] and for higher values of the jumping parameter $1 \leq s \leq 2$ [panel (b)]. Observe that $Q_1(p) = 1$ for each p , due to the fact that for $s = 1$, the Markov process reduces to the Bernoulli process. In general, for s close to 1 the amounts of information carried

by correlated and corresponding uncorrelated signals are comparable, i.e. the loss of information by correlated signals is relatively small. Note that Q_s for every $s \neq 1$ exhibits one maximum only. One can check the symmetry property, i.e.

$$Q_s\left(\frac{1}{2} - r\right) = Q_s\left(\frac{1}{2} + r\right), \quad (20)$$

$$\text{for } 0 \leq r \leq \frac{1}{2}, \quad \text{in the case of } 0 \leq s \leq 1; \quad (21)$$

and

$$\text{for } 0 \leq r \leq \frac{1}{s} - \frac{1}{2}, \quad \text{in the case of } 1 \leq s \leq 2. \quad (22)$$

Thus, in both cases, the maximum Q_s^{\max} over p of the quotient Q_s for all values of parameter s is achieved for $p = \frac{1}{2}$ and by (17) it is equal to

$$Q_s^{\max} = Q_s\left(\frac{1}{2}\right) = H_2\left(\frac{s}{2}\right). \quad (23)$$

In Fig. 2 we show Q_s^{\max} as a function of s for $0 \leq s \leq 2$. Here one can see that the minimal values of the quotient $Q_s(p)$ for each s are reached at the end points of the intervals (18) and (19).

For $0 \leq s \leq 1$, making use of (17), the bounds are as follows:

$$\lim_{p \rightarrow 0} Q_s(p) = s \quad \text{and} \quad \lim_{p \rightarrow 1} Q_s(p) = s, \quad (24)$$

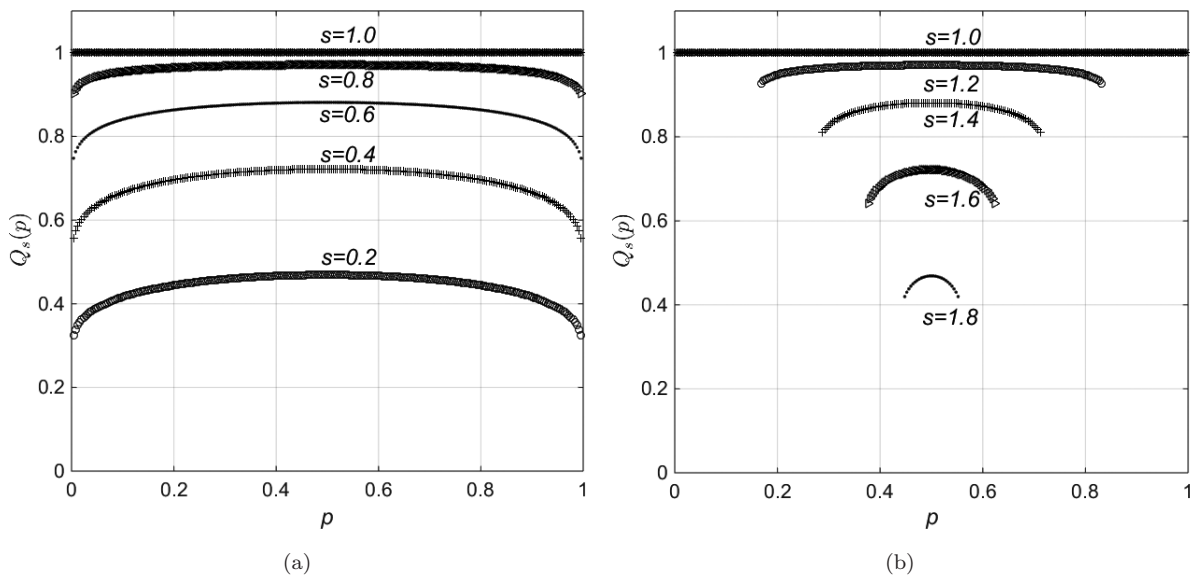


Fig. 1. The Markov–Bernoulli information quotient Q_s as a function of firing rate p for the chosen values of the jumping parameter s : (a) For parameters $0 \leq s \leq 1$ according to (18) the range of p is $[0, 1]$ and (b) for parameters $1 \leq s \leq 2$ due to (19) the range of p is $1 - \frac{1}{s} \leq p \leq \frac{1}{s}$.

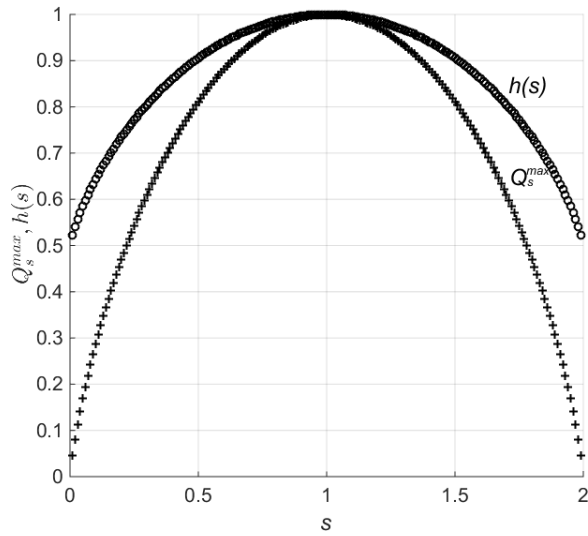


Fig. 2. The upper bound (23) of Q_s , i.e. Q_s^{\max} , and the limiting factor $h(s) = \frac{H_2(\frac{s}{2})+1}{2}$ in (33), (34) as functions of the jumping parameter s .

while for $1 \leq s \leq 2$ the bounds read

$$\lim_{p \rightarrow 1 - \frac{1}{s}} Q_s(p) = \frac{H_2(s-1)}{H_2\left(\frac{s-1}{s}\right)},$$

and by (20)

$$\lim_{p \rightarrow \frac{1}{s}} Q_s(p) = \frac{H_2(s-1)}{H_2\left(\frac{s-1}{s}\right)}. \quad (25)$$

Observe that for $s \rightarrow 1^+$ we have $\frac{H_2(s-1)}{H_2\left(\frac{s-1}{s}\right)} \rightarrow 1$.

Also, $g(s) := \frac{H_2(s-1)}{H_2\left(\frac{s-1}{s}\right)}$ as a function of s for $1 \leq s \leq 2$ is shown in Fig. 3.

Notice that for each s

$$f(s) := 2 - s \leq g(s). \quad (26)$$

The basic idea of Shannon Information Theory is to combine information with uncertainty. The higher the uncertainty of a given event, the more information is transmitted by such an event. The concept of entropy already addresses this idea.

To determine how far correlation reduces uncertainty, i.e. the amount of Shannon information, we compare the information per symbol transmitted by the words coming from a binary Markov source with the information per symbol coming from the corresponding Bernoulli source.

First we consider words of length 2. We make use of the grouping axiom of entropy.²⁸ It is known

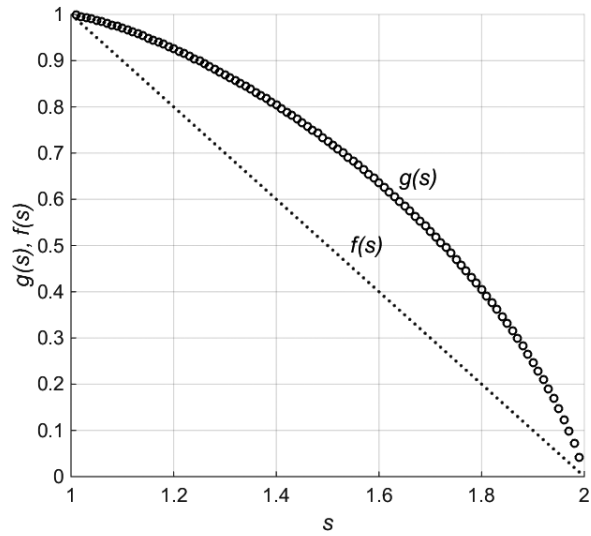


Fig. 3. The lower bound of Q_s namely $g(s) = \frac{H_2(s-1)}{H_2\left(\frac{s-1}{s}\right)}$ (24) and $f(s) = 2 - s$ (26) as functions of the jumping parameter $1 \leq s \leq 2$.

that the entropy function $H(Y)$ for a discrete random variable Y , under the assumptions of continuity, monotonicity, uncertainty of joint experiment and grouping axiom, is interpreted as the average uncertainty associated with the events $Y = z_i$, and is derived as $H(Y) = H(p_1, \dots, p_K) = -\sum_{j=1}^K p_j \log_2 p_j$, where p_i is probability of the event $Y = z_i$, $i = 1, 2, \dots, K$. Let us consider the quotient Q_s of the entropy of Markov source (11), (12) and a corresponding Bernoulli process (7). With the above notation and by (3) and (8) we have

$$\begin{aligned} Q_s &= \frac{p_{\text{eq}}(1)H_2(p_{0|1}) + p_{\text{eq}}(0)H_2(p_{1|0})}{H_2(p_{\text{eq}}(1))} \\ &= \frac{p_{\text{eq}}(1)H(p_{0|1}, 1 - p_{0|1}) + p_{\text{eq}}(0)H(p_{1|0}, 1 - p_{1|0})}{H(p_{\text{eq}}(1), 1 - p_{\text{eq}}(1))} \\ &= \frac{p_{\text{eq}}(1)H(p_{0|1}, p_{1|1}) + p_{\text{eq}}(0)H(p_{1|0}, p_{0|0})}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))}. \end{aligned} \quad (27)$$

Now, we express Q_s in the form

$$\begin{aligned} Q_s &= \frac{p_{\text{eq}}(1)H\left(\frac{p_{\text{eq}}(1)p_{0|1}}{p_{\text{eq}}(1)}, \frac{p_{\text{eq}}(1)p_{1|1}}{p_{\text{eq}}(1)}\right)}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\ &\quad + \frac{p_{\text{eq}}(0)H\left(\frac{p_{\text{eq}}(0)p_{1|0}}{p_{\text{eq}}(0)}, \frac{p_{\text{eq}}(0)p_{0|0}}{p_{\text{eq}}(0)}\right)}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))}. \end{aligned} \quad (28)$$

By adding and subtracting $H(p_{\text{eq}}(1), p_{\text{eq}}(0))$ in the numerator and making use of the grouping axiom,

$$\begin{aligned}
 Q_s &= \frac{p_{\text{eq}}(1)H\left(\frac{p_{\text{eq}}(1)p_{0|1}}{p_{\text{eq}}(1)}, \frac{p_{\text{eq}}(1)p_{1|1}}{p_{\text{eq}}(1)}\right)}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\
 &+ \frac{p_{\text{eq}}(0)H\left(\frac{p_{\text{eq}}(0)p_{1|0}}{p_{\text{eq}}(0)}, \frac{p_{\text{eq}}(0)p_{0|0}}{p_{\text{eq}}(0)}\right)}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\
 &+ \frac{H(p_{\text{eq}}(1), p_{\text{eq}}(0)) - H(p_{\text{eq}}(1), p_{\text{eq}}(0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\
 &= \frac{H(p_{\text{eq}}(1)p_{0|1}, p_{\text{eq}}(1)p_{1|1}),}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\
 &- \frac{H(p_{\text{eq}}(1), p_{\text{eq}}(0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} \\
 &= \frac{H(p(1, 0), p(1, 1), p(0, 1), p(0, 0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} - 1, \tag{29}
 \end{aligned}$$

where $p(i, j)$ denotes the probability of the words (i, j) , $i, j = 0, 1$.

For $0 \leq s \leq 1$ applying (23) and (24) to (29) we have

$$s \leq \frac{H(p(1, 0), p(1, 1), p(0, 1), p(0, 0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} - 1 \leq H_2\left(\frac{s}{2}\right). \tag{30}$$

For $1 \leq s \leq 2$, applying (23) and (25) to (29) we have

$$\begin{aligned}
 \frac{\frac{H_2(s-1)}{s}}{H_2\left(\frac{s-1}{2}\right)} &\leq \frac{H(p(1, 0), p(1, 1), p(0, 1), p(0, 0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} - 1 \\
 &\leq H_2\left(\frac{s}{2}\right), \tag{31}
 \end{aligned}$$

and applying (23) and (26) to (29) we have

$$\begin{aligned}
 2 - s &\leq \frac{H(p(1, 0), p(1, 1), p(0, 1), p(0, 0))}{H(p_{\text{eq}}(1), p_{\text{eq}}(0))} - 1 \\
 &\leq H_2\left(\frac{s}{2}\right). \tag{32}
 \end{aligned}$$

Let us consider the ITRs for the Markov process against the corresponding Bernoulli process for words of length 2. Making use of (30) and (32), and

using the notation (4), we obtain the following relations between these information sources:

$$\begin{aligned}
 \frac{s+1}{2} \text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(2)}(\mathbf{M}_s(p)) \\
 &\leq \frac{H_2\left(\frac{s}{2}\right) + 1}{2} \text{ITR}(\mathbf{B}_s(p)) \\
 &\text{for } 0 \leq s \leq 1; \tag{33}
 \end{aligned}$$

and

$$\begin{aligned}
 \frac{3-s}{2} \text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(2)}(\mathbf{M}_s(p)) \\
 &\leq \frac{H_2\left(\frac{s}{2}\right) + 1}{2} \text{ITR}(\mathbf{B}_s(p)) \\
 &\text{for } 1 \leq s \leq 2, \tag{34}
 \end{aligned}$$

where $\text{ITR}^{(2)}(\mathbf{M}_s(p))$ denotes the information transmission rate (4) of the Markov process for words of length 2, and $\text{ITR}(\mathbf{B}_s(p)) = \frac{1}{2} \text{ITR}^{(2)}(\mathbf{B}_s(p))$ is the information transmission rate of the corresponding Bernoulli process. Note that the correlation can reduce the ITR by as much as half, i.e.

$$\begin{aligned}
 \frac{1}{2} \text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(2)}(\mathbf{M}_s(p)) \\
 &\leq \text{ITR}(\mathbf{B}_s(p)) \text{ for } 0 \leq s \leq 1, \tag{35}
 \end{aligned}$$

$$\begin{aligned}
 \frac{1}{2} \text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(2)}(\mathbf{M}_s(p)) \\
 &\leq \text{ITR}(\mathbf{B}_s(p)) \text{ for } 1 \leq s \leq 2. \tag{36}
 \end{aligned}$$

Similar considerations for words of length $n (n \geq 2)$ led to the more general formulas:

$$\begin{aligned}
 \left[s \left(1 - \frac{1}{n} \right) + \frac{1}{n} \right] \text{ITR}(\mathbf{B}_s(p)) \\
 &\leq \text{ITR}^{(n)}(\mathbf{M}_s(p)) \\
 &\leq \left[H_2\left(\frac{s}{2}\right) \left(1 - \frac{1}{n} \right) + \frac{1}{n} \right] \text{ITR}(\mathbf{B}_s(p)) \\
 &\text{for } 0 \leq s \leq 1, \tag{37}
 \end{aligned}$$

$$\begin{aligned}
 \left[(2-s) \left(1 - \frac{1}{n} \right) + \frac{1}{n} \right] \text{ITR}(\mathbf{B}_s(p)) \\
 &\leq \text{ITR}^{(n)}(\mathbf{M}_s(p)) \\
 &\leq \left[H_2\left(\frac{s}{2}\right) \left(1 - \frac{1}{n} \right) + \frac{1}{n} \right] \text{ITR}(\mathbf{B}_s(p)) \\
 &\text{for } 1 \leq s \leq 2. \tag{38}
 \end{aligned}$$

Note that from (37) and (38) the following upper and lower bounds follow:

$$\begin{aligned}
 s\text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(n)}(\mathbf{M}_s(p)) \\
 &\leq \frac{H_2\left(\frac{s}{2}\right) + 1}{2} \text{ITR}(\mathbf{B}_s(p)) \quad \text{for } 0 \leq s \leq 1,
 \end{aligned}
 \tag{39}$$

$$\begin{aligned}
 (2-s)\text{ITR}(\mathbf{B}_s(p)) &\leq \text{ITR}^{(n)}(\mathbf{M}_s(p)) \\
 &\leq \frac{H_2\left(\frac{s}{2}\right) + 1}{2} \text{ITR}(\mathbf{B}_s(p)) \quad \text{for } 1 \leq s \leq 2,
 \end{aligned}
 \tag{40}$$

where in Ref. 30 $\text{ITR}(\mathbf{B}_s(p)) = H_2(p)$ and $n \geq 2$. Note that the bounds s and $2-s$ can be interpreted as $1 - \det \mathbf{P}$ and $\text{tr} \mathbf{P}$, respectively.

These results show that for the Markov processes for any word length, the reduction of information due to correlations is limited by the factor s or $2-s$. This finding supports the hypothesis that under certain conditions neurons can use temporal codes which are more energetically efficient compared to the more reliable rate code. It is interesting that the factors (bounds) in the above inequalities depend only on the jumping parameter s . This parameter is simply the sum of the conditional probabilities of transition from state to state. On the other hand, experiments show⁵⁶ that spiking frequency is in practice limited typically to 40 spikes within a time period of a few seconds, and the time resolution of the spikes being detected is typically in the range of 3 ms. Thus, it is justified to assume that, after digitalization (binning), the transition probability from the state in which there is a spike to the state where there is no spike is large (i.e. close to 1), while the transition probability from the state of no spike to the state where there is a spike is small (i.e. close to 0), and consequently the values of s are around 1. However, our results show that for s close to 1, the amounts of information carried by correlated (like temporal codes) and the corresponding uncorrelated (like rate code) signals are comparable. This suggests that when a neuronal system decides to use a temporal code, some trade-off between energetic cost and transmission reliability must be taken into account. Experiments confirm that such situations can occur

in the primary auditory cortex,⁵⁷⁻⁵⁹ the visual cortex⁶⁰ and also in the olfactory⁶¹ and the gustatory⁶² information processing systems.

4. Conclusions

Spiking neurons communicate with each other by means of small electric currents, transferring information via sequences of action potentials called spike trains, which can be viewed as a string of binary signals.^{53,63,64} It is still an open question whether the information contained in these binary signals is conveyed by the firing frequency or by the precise timing of the spikes. The nature of the code used by spike trains is closely related to whether the digitalized (binned) representation of messages is governed by *uncorrelated* stochastic processes (Bernoulli processes) or by *correlated* ones, such as some Markov processes.

We point out that the correlations we have considered in this paper refer *only* to correlations within a given spike train, and are thus distinct from correlations among spike trains emitted by several different neurons, a topic that has received a great deal of experimental and theoretical attention (see, for example, Refs. 65-67).

In this paper we have shown that, when information conveyed by spike trains coming from such different sources are compared, a crucial role is played by the same jumping parameter s , as we found in Ref. 27. We have found that the correlation-related loss of information for signals governed by Markov processes, when compared with the corresponding uncorrelated processes, is determined only by this parameter s . Experiments confirm that taking into account the frequency of neuronal signals and spike detection resolution, this parameter oscillates around 1. Our results show that for s close to 1, the amounts of information transmitted by correlated and corresponding uncorrelated signals are comparable. Thus temporal codes, which are more energetically efficient, can be used instead of rate codes. This was observed in a number of *in vivo* recordings of neuronal activity⁶⁸ and in the studies mentioned in that reference.

Acknowledgments

This paper has been partly conceived through participation in the EMBO Installation Grant of

Dr. Michal Komorowski (MK). Thus, we thank MK for valuable discussions.

References

1. E. D. Adrian and Y. Zotterman, The impulses produced by sensory nerve endings, *J. Physiol.* **61** (1926) 49–72.
2. K. Roy, S. Kumar and S. A. Bloomfield, Gap junctional coupling between retinal amacrine and ganglion cells underlies coherent activity integral to global object perception, *Proc. Natl. Acad. Sci. USA* **114**(38) (2017) E10484–E10493.
3. J. Song, K. Ampatzis, R. Bjornfors and A. El Manira, Motor neurons control locomotor circuit function retrogradely via gap junctions, *Nature* **529** (2016) 399–402.
4. R. Wang, I. Tsuda and Z. Zhang, A new work mechanism on neuronal activity, *Int. J. Neural Syst.* **25**(2) (2015) 1450037.
5. D. Attwell and S. Laughlin, An energy budget for signaling in the grey matter of the brain, *J. Cereb. Blood Flow Metab.* **21** (2001) 1133–1145.
6. A. Hasenstaub, S. Otte, E. Callaway and T. Sejnowski, Metabolic cost as a unifying principle governing neuronal biophysics, *Proc. Natl. Acad. Sci. USA* **345**(107) (2010) 12329–12334.
7. L. Kostal, P. Lansky and M. D. McDonnell, Metabolic cost of neuronal information in an empirical stimulus-response model, *Biol. Cybern.* **107** (2013) 355–365.
8. B. J. Richmond and L. M. Optican, Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex: II: Information transmission, *J. Neurophysiol.* **64**(2) (1990) 370–380.
9. R. Brette, Philosophy of the spike: Rate-based versus spike-based theories of the brain, *Front. Syst. Neurosci. Rev.* **9** (2015) 151.
10. S. Ghosh-Dastidar and H. Adeli, Spiking neural networks, *Int. J. Neural Syst.* **19**(4) (2009) 295–308.
11. S. Ghosh-Dastidar and H. Adeli, Improved spiking neural networks for EEG classification and epilepsy and seizure detection, *Integr. Comput.-Aided Eng.* **14**(3) (2007) 187–212.
12. S. Ghosh-Dastidar and H. Adeli, A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection, *Neural Netw.* **22**(10) (2009) 1419–1431.
13. H. Adeli and S. Ghosh-Dastidar, *Automated EEG-Based Diagnosis of Neurological Disorders: Inventing the Future of Neurology* (CRC Press, Boca Raton, 2010).
14. A. Geminiani, C. Casellato, A. Antonietti, E. D’Angelo and A. Pedrocchi, A multiple-plasticity spiking neural network embedded in a closed-loop control system to model cerebellar pathologies, *Int. J. Neural Syst.* **28**(5) (2018) 1750017.
15. L. Guo, Z. Wang, M. Cabrerizo and M. Adjouadi, A cross-correlated delay shift supervised learning method for spiking neurons with application to interictal spike detection in epilepsy, *Int. J. Neural Syst.* **27**(3) (2017) 1750002.
16. X. Zhang, G. Foderaro, C. Henriquez and S. Ferrari, A scalable weight-free learning algorithm for regulatory control of cell activity in spiking neuronal networks, *Int. J. Neural Syst.* **28**(2) (2018) 1750015.
17. G. Antunes, S. F. Faria da Silva and F. M. Simoes de Souza, Mirror neurons modeled through spike-timing-dependent plasticity are affected by channelopathies associated with autism spectrum disorder, *Int. J. Neural Syst.* **28**(5) (2018) 1750058.
18. M. Abeles, *Local Cortical Circuits* (Springer, Berlin, 1982).
19. L. Pan, G. Paun, G. Zhang and F. Neri, Spiking neural P systems with communication on request, *Int. J. Neural Syst.* **27**(8) (2017) 1750042.
20. F. Weissenberger, F. Meier, J. Lengler, H. Einarsson and A. Steger, Long synfire chains emerge by spike-timing dependent plasticity modulated by population activity, *Int. J. Neural Syst.* **27**(8) (2017) 1750044.
21. T. Wu, F. D. Bilbie, A. Paun, L. Pan and F. Neri, Simplified and yet Turing universal spiking neural P systems with communication on request, *Int. J. Neural Syst.* **28**(8) (2018) 1850013.
22. A. Antonietti, J. Monaco, E. D’Angelo, A. Pedrocchi and C. Casellato, Dynamic redistribution of plasticity in a cerebellar spiking neural network reproducing an associative learning task perturbed by TMS, *Int. J. Neural Syst.* **28**(9) (2018) 1850020.
23. C. M. Gray and D. A. McCormick, Chattering cells: Superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex, *Science* **274**(5284) (1996) 109–113.
24. A. Dettner, S. Munzberg and T. Tchumatchenko, Temporal pairwise spike correlations fully capture single-neuron information, *Nat. Commun.* **7** (2016) 13805.
25. A. Mohemmed, S. Schliebs, S. Matsuda and N. Kasabov, Span: Spike pattern association neuron for learning spatio-temporal spike patterns, *Int. J. Neural Syst.* **22**(4) (2012) 1250012.
26. A. Pregowska, J. Szczepanski and E. Wajnryb, Mutual information against correlations in binary communication channel, *BMC Neurosci.* **16** (2015) 32.
27. A. Pregowska, J. Szczepanski and E. Wajnryb, Temporal code versus rate code for binary information sources, *Neurocomputing* **216** (2016) 756–762.
28. R. B. Ash, *Information Theory* (John Wiley and Sons, New York, 1965).
29. C. E. Shannon, A mathematical theory of communication, *Bell Labs Tech. J.* **27**(3) (1948) 379–423.
30. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

31. W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley Series Probability and Statistics (John Wiley & Sons, New York, 1958).
32. J. L. van Hemmen and T. Sejnowski, *23 Problems in Systems Neurosciences* (Oxford University Press, Oxford, 2006).
33. M. Crumiller, B. Knight and E. Kaplan, The measurement of information transmitted by a neural population: Promises and challenges, *Entropy* **15** (2013) 3507–3527.
34. W. Gerstner, W. M. Kistler, R. Naud and L. Paninski, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press, Cambridge, 2014).
35. D. H. Perkel, G. L. Gerstein and G. P. Moore, Neuronal spike trains and stochastic point processes I: The single spike train, *Biophys. J.* **7**(4) (1967) 391–418.
36. D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods* (Springer, Berlin, 2003).
37. M. C. Teich and S. M. Khanna, Pulse-number distribution for the neural spike train in the cat's auditory nerve, *J. Acoust. Soc. Am.* **77**(3) (1985) 1110–1128.
38. F. Rieke, D. D. Warland, R. R. de Ruyter van Steveninck and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, 1997).
39. G. Werner and V. B. Mountcastle, Neural activity in mechanoreceptive cutaneous afferents: Stimulus-response relations, weber functions, and information transmission, *J. Neurophysiol.* **28** (1965) 359–397.
40. D. J. Tolhurst, J. A. Movshon and I. D. Thompson, The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast, *Exp. Brain Res.* **41** (1981) 414–419.
41. R. R. de Ruyter van Steveninck, G. D. Lewen, S. P. Strong, R. Koberle and W. Bialek, Reproducibility and variability in neural spike trains, *Science* **275** (1997) 1805–1808.
42. S. M. Ross, *Stochastic Processes* (Wiley-Interscience, New York, 1996).
43. A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes* (Tata McGraw-Hill Education, 2002).
44. P. Kara, P. Reinagel and R. C. Reid, Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons, *Neuron* **27**(3) (2000) 635–646.
45. R. Barbieri, M. C. Quirk, L. M. Frank, M. A. Wilson and E. N. Brown, Construction and analysis of non-Poisson stimulus-response models of neural spiking activity, *J. Neurosci. Methods* **105** (2001) 25–37.
46. B. B. Averbeck, Poisson or not Poisson: Differences in spike train statistics between parietal cortical areas, *Neuron* **62**(3) (2009) 310–311.
47. R. E. Kass and V. Ventura, A spike-train probability model, *Neural Comput.* **13**(8) (2001) 1713–1720.
48. T. Britvina and J. J. Eggermont, A Markov model for interspike interval distributions of auditory cortical neurons that do not show periodic firings, *Biol. Cybern.* **96** (2006) 245–264.
49. G. Radons, J. D. Becket, B. Dillfer and J. Kruger, Analysis, classification, and coding of multielectrode spike trains with hidden Markov models, *Biol. Cybern.* **71**(4) (1994) 359–373.
50. C. Kemere, G. Santhanam, B. M. Yu, A. Afshar, S. I. Ryu, T. H. Meng and K. V. Shenoy, Detecting neural-state transitions using hidden Markov models for motor cortical prostheses, *J. Neurophysiol.* **100**(4) (2008) 2441–2452.
51. E. T. Rolls, N. C. Aggelopoulos, L. Franco and A. Treves, Information encoding in the inferior temporal visual cortex: Contributions of the firing rates and the correlations between the firing of neurons, *Biol. Cybern.* **90** (2004) 19–32.
52. J. M. Amigo, J. Szczepanski, E. Wajnryb and M. V. Sanchez-Vives, Estimating the entropy rate of spike trains via Lempel-Ziv complexity, *Neural Comput.* **1386** (2004) 717–736.
53. D. MacKay and W. S. McCulloch, The limiting information capacity of a neuronal link, *Bull. Math. Biol.* **14** (1952) 127–135.
54. W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck and D. D. Warland, Reading a neural code, *Science* **252**(5014) (1991) 1854–1857.
55. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, 2007).
56. S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck and W. Bialek, Entropy and information in neural spike trains, *Phys. Rev. Lett.* **80**(1) (1998) 197–200.
57. R. C. deCharms and M. M. Merzenich, Primary cortical representation of sounds by the coordination of action-potential timing, *Nature* **381**(6583) (1996) 610–613.
58. D. Oertel, The role of timing in the brainstem auditory nuclei, *Annu. Rev. Physiol.* **61** (1999) 497–519.
59. X. Wang, T. Lu, D. Bendor and E. Bartlett, Neural coding of temporal information in auditory thalamus and cortex, *Neuroscience* **154** (2008) 294–303.
60. J. D. Victor and K. Purpura, Nature and precision of temporal coding in visual cortex: A metric-space analysis, *J. Neurophysiol.* **76** (1996) 1310–1326.
61. G. Laurent, Olfactory processing: Maps, time and codes, *Curr. Opin. Neurobiol.* **7** (1997) 547–553.
62. P. M. Di Lorenzo, J. Y. Chen and J. D. Victor, Quality time: Representation of a multidimensional sensory domain through temporal coding, *J. Neurosci.* **29** (2009) 9227–9238.
63. M. DeWeese, Optimization principles for the neural code, *Network, Comput. Neural Syst.* **7** (1996) 325–331.
64. J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky and E. P. Simoncelli,

- Spatio-temporal correlations and visual signaling in a complete neuronal population, *Nature* **454** (2008) 995–999.
65. B. B. Averbeck and D. Lee, Effects of noise correlations on information encoding and decoding, *J. Neurophysiol.* **95**(6) (2006) 3633–3644.
66. M. R. Cohen and A. Kohn, Measuring and interpreting neuronal correlations, *Nat. Neurosci.* **14**(7) (2011) 811–819.
67. H. G. Eyherabide and I. Samengo, When and why noise correlations are important in neural decoding, *J. Neurosci.* **33**(45) (2013) 17921–17936.
68. Ch. A. Baker, T. Kohashi, A. M. Lyons-Warren, X. Ma and B. A. Carlson, Multiplexed temporal coding of electric communication signals in mormyrid fishes, *J. Exp. Biol.* **216**(13) (2013) 2365–2379.