Biol. Cybern. 88, 335–351 (2003)
DOI 10.1007/s00422-002-0386-2

Biological
Cybernetics

Effect of cross-trial nonstationarity on joint-spike events

S. Grün[1,*], **A. Riehle**[2], **M. Diesmann**[3]

[1] Max-Planck-Institute for Brain Research, Department of Neurophysiology, 60528 Frankfurt, Germany
[2] Equipe "Perception and Cognition," CNRS-INPC, Marseille, France
[3] Department of Nonlinear Dynamics, Max-Planck-Institut für Strömungsforschung, Göttingen, Germany

Received: 14 February 2002 / Accepted: 26 November 2002 /

Abstract. Common to most correlation analysis techniques for neuronal spiking activity are assumptions of stationarity with respect to various parameters. However, experimental data may fail to be compatible with these assumptions. This failure can lead to falsely assigned significant outcomes. Here we study the effect of nonstationarity of spike rate across trials in a model-based approach. Using a two-rate-state model, where rates are drawn independently for trials and neurons, we show in detail that nonstationarity across trials induces apparent covariation of spike rates identified as the generator of false positives. This finding has specific implications for the "shuffle predictor." Within the framework developed for our model, covariation of spike rates and the mechanism by which the shuffle predictor leads to wrong interpretation of the data can be discussed. Corrections for the influence of nonstationarity across trials by improvements of the predictor are presented.

1 Introduction

The analysis of the spatial structure, task dependence, and dynamics of spike time correlation is a central tool in the investigation of neuronal interaction and cortical processing (Perkel et al. 1967; Abeles 1982; Gerstein et al. 1989; Singer 1993). Assumptions about various aspects of the temporal statistics of the spike-generating processes underly most crosscorrelation techniques. Typical assumptions are that spike trains are generated by Poisson processes and that the rate functions driving the processes are stationary. Presumably the simplest example of a correlation technique is the peristimulus time histogram (PSTH) obtained from many realizations ("trials") of a single Poisson process. Here spike times are correlated with the occurrence of a "stimulus," a point in time used to align the individual trials. The histogram is only a meaningful measure of the potentially time-dependent rate if the spike-generating process is in each trial driven by the same (time-dependent) parameter.

Figure 1 illustrates a categorization of ways in which assumptions used in the analysis of spike data can be violated. While Fig. 1a is concerned with the nature of the spike-generating process, the remaining three panels distinguish basic types of nonstationarity of spike rate. An arbitrary experimental data set will contain contributions of all these aspects, and in general they need to be considered simultaneously when analyzing experimental data (e.g., Vaadia et al. 1988 provides instructive dot displays). Deviations of neuronal spiking from the statistics of a Poisson process have been demonstrated (see, e.g., Reich et al. 2000 for a recent study) and are incorporated in theoretical studies (e.g., Pauluis and Baker 2000; Baker and Lemon 2000; Baker and Gerstein 2000). A time dependence of the spike rate (Fig. 1b) naturally occurs if neurons exhibit a rate "response" to the event to which the trials are aligned. A standard method of coping with this type of nonstationarity is to perform the analysis in a sliding window of limited width (e.g., MacLeod and Laurent 1996; Roelfsema et al. 1997; Grün et al. 2002b). Within this window stationarity is assumed. However, the onset of a change in spike rate may vary across trials. This type of nonstationarity is called "latency variability" (Vaadia et al. 1988; Miller et al. 1992; Brody 1999a). Note this is not the stimulus-dependent difference in the onset of the observed response (e.g., Richmond et al. 1987; Coburn et al. 1990; Gawne et al. 1996) sometimes also called latency variability. Corrections for latency variability can in some cases be carried out by aligning the trials to a different event (Grün et al. 2002b) or by realignment of the trials exploiting similarity of the rate profiles of individual trials (Nawrot et al. 2002). Methods have been developed to include the effects of latency variability in the predictor for the correlation measure (Pauluis and Baker 2000; Baker and Gerstein 2001). Rate level differences across trials ("excitability," Brody 1999b; Fig. 1d, e.g., Gur et al.

Correspondence to: S. Grün
(e-mail: gruen@neurobiologie.fu-berlin.de,
Tel.: +49-30-83856635, Fax: +49-30-83855455)

**Present address*: Free University Berlin, Institute for Biology, Neurobiology, Berlin, Germany

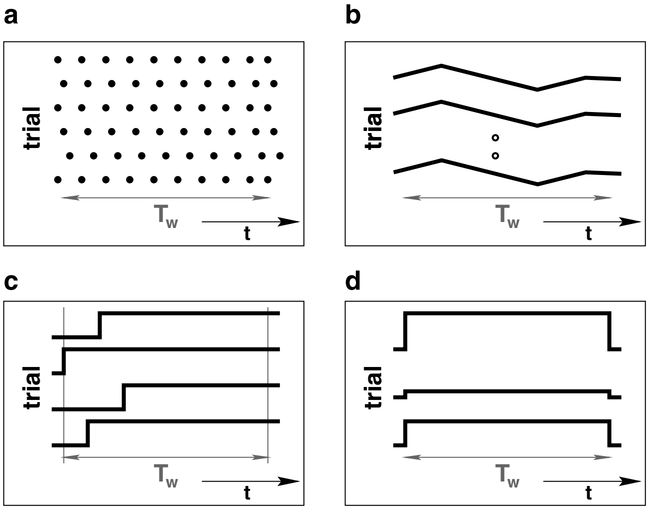


Fig. 1. Categorization of possible violations of assumptions conventionally used in the analysis of spiking activity. **a** Deviations from Poisson statistics. Spiking activity of a single neuron observed over some time interval T_w (spikes marked by *dots*, time t progresses horizontally) for a number of repetitions (trials, vertically arranged) under the same experimental conditions. Spike trains exhibit structure (here regularly spaced spike intervals) in conflict with the properties of a Poisson process (spike probability at t independent of position of last spike). **b** Nonstationarity of spike rate over time. Curves represent the time course of the rate function driving the spike-generating process in individual trials (indicated by arbitrary vertical offset). The spike rate is stationary across trials (*identical curves*); however, it varies in the course of time. **c** Latency variability. While the spike rate exhibits an identical time course in all trials, the rate profiles are not aligned with respect to the onset of the trial (curves show step at different locations in time). **d** Nonstationarity across trials. Although the spike rate is stationary during T_w in each trial, the spike rate varies across trials (curves show steps of different amplitudes)

1997; Nawrot et al. 2001) may be due to external or internal events not under the control of the experiment (“ongoing activity,” Arieli et al. 1996), variability of attention, or drifts in anesthesia level. This is what we call “nonstationarity across trials.” The above categorization considers nonstationarity in the activity of a single neuron. It is the presence of nonstationarity in single neurons that permits an additional relevant phenomenon: covariation of spike rate (e.g., Friston 1995; Vaadia et al. 1995a; Brody 1999b). Again, covariation of spike rate is an effect that can occur in time and across trials and does occur in neuronal data (e.g., Fries et al. 2001). Recent studies explored corrections for covariation of spike rates (Pauluis and Baker 2000; Baker and Gerstein 2001; Ben-Shaul et al. 2001).

In the ongoing debate in neuroscience on the role of temporal coordination of neuronal spike response for coding principles in the cortex, the occurrence of nonstationarity across trials and covariation of spike rate has gained importance. Spike rate is the classical measure for neuronal response, and its selectivity for specific stimuli and/or behavior has been contrasted to the information contained in spike timing (e.g., Oram et al. 2001, 1999; Shadlen and Newsome 1998; Baker and Lemon 2000; Richmond et al. 1999). Considering ensembles of neurons, the covariation of spike rate (or spike count) was suggested as a potential code (Oram et al. 2001). On the other hand, there is evidence for the involvement of the (fine) temporal coordination of spikes in neuronal processing (e.g., Singer and Gray 1995; Vaadia et al. 1995b; Riehle et al. 1997; Prut et al. 1998; Grün et al. 2002b). Thus, there is the possibility that both types of dynamics are present in a single data set. In order to disentangle the two coding strategies, a thorough understanding of the effects of nonstationarity on the predictors for correlation measures is required. In the present work, we use a model-based approach to investigate the effects of nonstationarity across trials. Simplicity of the model permits an analytical treatment and insight into the key mechanisms. The spike generators of the model neurons are assumed to be completely independent with nonstationary (stochastic) spike rates across trials (no covariation of spike rate). The system is analyzed in a realistic setting, where an experimental outcome consists of a finite number of trials. Surprisingly, although covariation of spike rate is not built into the model, it generates coherent rate changes in particular experimental outcomes and enables a characterization of the coherence in rate fluctuations. Nonstationarity across trials and covariation of spike rate can be analyzed in a common framework. Statements about the ensemble average as well as individual experimental outcomes can be made.

We begin our investigation by introducing a model for nonstationarity across trials (Sect. 2). Then (Sect. 3), the notion of “false positives” used throughout the study to quantify the effect of nonstationarity is presented. Basic properties of the model are explored using simulations. The next section (Sect. 4) develops a framework for the detailed analysis of the mechanisms by which nonstationarity enters correlation measures. Consideration of the mean coincidence count in a system of two trials already provides qualitative insight. The generalization of the approach to M trials explains model results with high precision. It is demonstrated that false positives in the two-rate-state model are fundamentally related to the effects of covariation of spike count. Consequences for the interpretation of individual experimental outcomes and possible corrections for nonstationarity across trials are discussed in Sect. 5 and illustrated using an experimental data set. Our framework enables a rigorous analysis of a widely used tool in correlation studies – the shuffle predictor – under conditions of nonstationarity across trials.

2 Two-rate-state model

Consider a system of two neurons, where in each trial a neuron can only be in one of two possible rate states (λ_1, λ_2). The difference of the two-rate levels $\Delta\lambda = \lambda_1 - \lambda_2$ defines the “degree” of nonstationarity across trials; rates are uniquely described with the additional specification of the mean rate $(\lambda_1 - \lambda_2)/2$. For each trial and neuron (stationary) spike rates are independently chosen according to occupation probabilities $P_\lambda(\lambda_1)$ and $P_\lambda(\lambda_2)$. The normalization condition

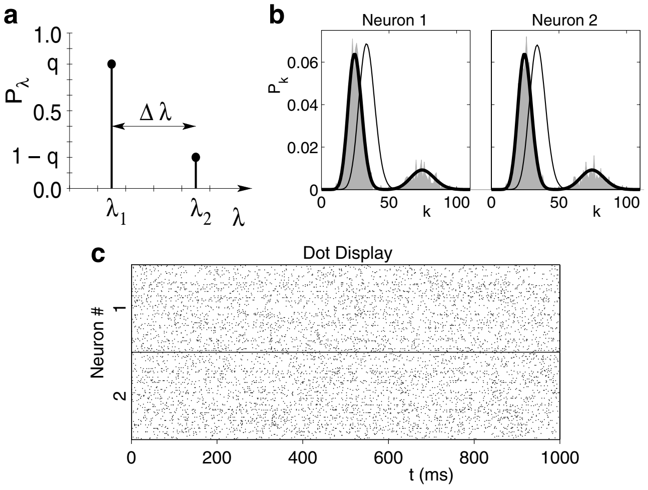


Fig. 2. Two-rate-state model of nonstationarity across trials (cf. Fig. 1d). **a** Distribution of spike rate P_λ across trials. During a trial spike rate is stationary at λ_i. Only two rate states, λ_1 and λ_2 separated by $\Delta\lambda = \lambda_2 - \lambda_1$, are allowed. For each trial and neuron the spike rate is drawn independently from the two possible rate levels and used as the parameter of a Poisson process. The probability of being in the lower rate state λ_1 is denoted by q and, by normalization, $1 - q$ for the higher rate λ_2. **b** Spike count distribution of the model defined in **a**. Two neurons are simulated for $M = 1000$ trials of duration $\mathcal{T} = 1000$ ms, with parameters $\lambda_1 = 25\,\mathrm{s}^{-1}$, $\lambda_2 = 75\ \mathrm{s}^{-1}$, and $q = 0.8$. Apart from statistical fluctuation the spike count histograms (*gray bars*, spike count k horizontal, probability of occurrence P_k vertical) of neuron 1 (*left panel*) and 2 (*right panel*) are identical. The *thick black curve* indicates the theoretical distribution given by Eq. 1. The *thin black curve* represents the distribution of spike counts expected on the basis of the average spike rate $q\lambda_1 + (1 - q)\lambda_2$ assuming stationarity. **c** Visual appearance of cross-trial nonstationarity in the dot display. The two boxes show the simultaneous spiking activity of the simulation described in **b** (*upper box*: neuron 1, lower: 2). *Dots* represent the spike times of the first 100 trials. In each box one vertical position is reserved for every trial; time t is advancing along the horizontal axis

$P_\lambda(\lambda_1) + P_\lambda(\lambda_2) = 1$ allows us to capture the asymmetry in the distribution of rates by a single parameter $q = P_\lambda(\lambda_1)$. Independent Poisson processes with individual parameters $\lambda \in \{\lambda_1, \lambda_2\}$, as specified above, generate the spike trains for each trial and neuron. A sketch of our model of nonstationarity across trials is shown in Fig. 2a.

A typical example of the raw data generated by the two-rate-state model is illustrated in Fig. 2c. The horizontally stripy appearance of the dot displays with individual "lines" or "bands" of dots popping out from a background of lower spike density is often observed in experimental data. Figure 2b shows the spike count histograms corresponding to the data set in Fig. 2c. Apart from fluctuations, the histograms for the two neurons are identical. The distribution of spike counts is bimodal. For each rate state, the spike counts k are distributed according to a Poisson distribution $\mathcal{P}$, the parameters being $\lambda_1\mathcal{T}$ and $\lambda_2\mathcal{T}$, respectively. $T = \mathcal{T}/h$ specifies the length of a trial, i.e., the number of time steps with bin size $h = 1$ ms. The bimodality stems from the relative contributions of the rate states to the ensemble of trials.

$$P_k(k) = \sum_i P_\lambda(\lambda_i)\mathcal{P}(k, \lambda_i\mathcal{T})$$
$$= q\mathcal{P}(k, \lambda_1\mathcal{T}) + (1 - q)\mathcal{P}(k, \lambda_2\mathcal{T}) \ . \tag{1}$$

This distribution clearly deviates from the unimodal distribution of spike counts obtained on the basis of the average spike rate $q\lambda_1 + (1 - q)\lambda_2$. A measure for the variability of neuronal spiking is the variance of the spike count divided by the mean spike count, known as the Fano factor (see Rieke et al. 1997 and references therein). As long as spikes are generated by a Poisson process, a Fano factor exceeding unity indicates nonstationarity across trials. Conversely, if the system is stationary across trials, conclusions about the nature of the process can be drawn.

3 Simulation results

Let us now utilize the model introduced in the preceding section to investigate the effect of nonstationarity across trials on measures of spike time correlation (e.g., Perkel et al. 1967; Abeles and Gerstein 1988; Aertsen et al. 1989; Martignon et al. 1995). We concentrate on correlations at zero time delay: spike coincidences. The unitary event method evaluates the statistical significance of the number of coincidences occurring in a given experimental data set (see Grün et al. 2002a for a detailed description and a discussion of the assumptions). At a predefined temporal precision h (discrete binning) the empirical coincidence count n^{emp} is compared to the expected coincidence count $\bar{n}^{\mathrm{pred}}$. Assuming stationarity across time and trials the expected value is obtained from the spike rates as the total number of time bins TM multiplied by the probability of observing a coincidence in a single bin:

$$\bar{n}^{\mathrm{pred}} = TM \cdot \frac{1}{M}\left(\sum_{i=1}^{M} p_{1,i}\right) \cdot \frac{1}{M}\left(\sum_{i=1}^{M} p_{2,i}\right) \ . \tag{2}$$

$p_{j,i}$ is the probability of neuron j emitting a spike in a time bin of trial i and M the number of trials. In experimental data, the underlying spike rates are unknown. Therefore, the probability of spike emission needs to be estimated from the average of the spike counts $k_{j,i}$ of individual trials. Consequently, with respect to our model, $\bar{n}^{\mathrm{pred}}$ is expressed in terms of the average $p_{j,i}$. While in model calculations we use $p_{j,i} = \lambda_{j,i}h$, for experimental data sets and in simulations the $p_{j,i}$ are replaced by the corresponding estimators $\hat{p}_{j,i} = k_{j,i}/T$. The significance of the number of coincidences in the data set is evaluated by the probability ("joint-p-value") of observing n^{emp} or more coincidences, assuming that for independent neurons the number of coincidences can be described by a Poisson distribution with parameter $\bar{n}^{\mathrm{pred}}$. If the significance measure is smaller than a requested significance level α (usually 0.01 or 0.05), the number of coincidences is said to deviate significantly from expectation.

The significance test implies that at a given significance level α, data sets from independent neurons should

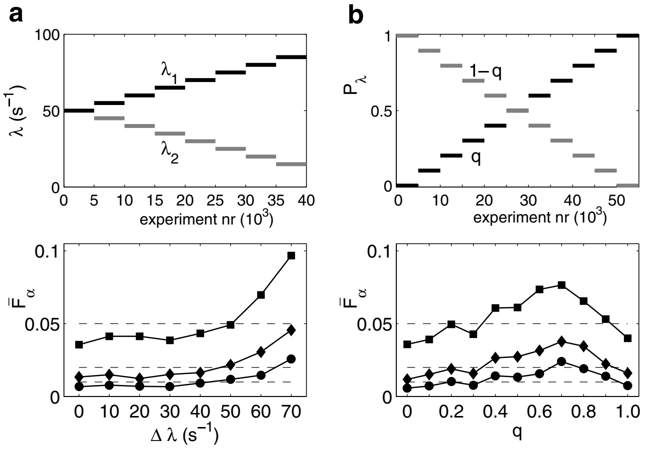


Fig. 3. False positives in the evaluation of spike coincidences under conditions of cross-trial nonstationarity. Data are generated according to the model described in Fig. 2. A set of $M = 100$ trials of duration $\mathcal{T} = 1000$ ms defines an experiment. Five thousand experiments are performed for a given parameter set $(\lambda_1, \lambda_2, q)$. In each experiment, the number of spike coincidences is compared to the predicted count, based on the estimated average spike rate across trials, and checked against a significance level α. The number of experiments with a significant outcome defines the fraction of false positives $\bar{F}_\alpha$. **a** Effect of the rate level difference $\Delta\lambda$. At $q = 0.5$ and mean spike rate $(\lambda_1 + \lambda_2)/2 = 50\,\mathrm{s}^{-1}$, $\Delta\lambda$ is varied from $0\ \mathrm{s}^{-1}$ to $70\ \mathrm{s}^{-1}$ in steps of $10\ \mathrm{s}^{-1}$. The *upper graph* illustrates the structure of the data set. *Curves* visualize the rate levels (vertical) λ_1 (*black*), λ_2 (*gray*) used in different experiments (horizontal) organized by increasing $\Delta\lambda$. Staircase-like appearance indicates the number of experiments performed for a particular parameter set. The *lower graph* shows the fraction of false positives $\bar{F}_\alpha$ (vertical) as a function of $\Delta\lambda$ (horizontal). Each data point is computed from the corresponding ensemble of experiments in the *upper graph*. Analysis is performed for three significance levels α: 0.01 (disks), 0.02 (diamonds), and 0.05 (squares); data points for corresponding α are connected by *straight lines*. *Dashed lines* indicate the theoretical fraction of false positives $f_\alpha = \alpha$. At small $\Delta\lambda$, $\bar{F}_\alpha$ falls below α because of discreteness of the distribution of coincidence counts and its approximation by a Poisson distribution. At large $\Delta\lambda$, nonstationarity causes an excess in $\bar{F}_\alpha$. **b** Effect of the occupation probability q. Same organization of graphs as in **a**. *Curves* in *upper graph* visualize occupation probability P_λ (vertical) of the two rate states, experiments (horizontal) organized by increasing q varied from 0.1 to 0.9 in steps of 0.1 with $\lambda_1 = 20\,\mathrm{s}^{-1}$ (*black*) and $\lambda_2 = 80\,\mathrm{s}^{-1}$ (*gray*). The *lower graph* shows $\bar{F}_\alpha$ (vertical) as a function of q (horizontal). $\bar{F}_\alpha$ exhibits a maximum at about $q = 0.7$

contain a significant number of coincidences in a fraction α of all cases. This ratio is called the fraction of false positives. False positive levels larger than α indicate that certain assumptions underlying the analysis are violated. Thus, ignorance of the presence of nonstationarity across trials may lead to a wrong interpretation of the experimental data.

In our model, nonstationarity across trials for a fixed mean spike rate $(\lambda_1 + \lambda_2)/2$ is characterized by two parameters: the difference in spike rate $\Delta\lambda$ and the relative occurrence of the two rate levels q. A simulated data set of M trials of duration $\mathcal{T}$ with a given nonstationarity $(\Delta\lambda, q)$ represents the outcome of one *experiment*. Figure 3 shows the fraction of experiments $\bar{F}_\alpha$ in which a significant number of coincidences is reported as a function of $\Delta\lambda$ and q. With increasing $\Delta\lambda$ the percentage of false positives increases monotonically. Above a rate level difference of $40\ \mathrm{s}^{-1}$, $\bar{F}_\alpha$ strongly deviates from α (Fig. 3a). For smaller $\Delta\lambda$, $\bar{F}_\alpha$ is somewhat smaller than α due to the discrete nature of coincidence counts (Roy et al. 2000; Grün et al. 2002a) and the approximation of the distribution of coincidence counts by a Poisson distribution (Gütig et al. 2002). The dependence of $\bar{F}_\alpha$ on q, however, is nonmonotonic. False positives increase with increasing q, reach a maximum at $q \approx 0.7$, and drop back to the expected level for q approaching 1.

Clearly both parameters of nonstationarity control the fraction of false positives provided the other parameter does not force the system to remain stationary. By definition the system is independent of q if $\Delta\lambda = 0$ and independent of $\Delta\lambda$ if $q = 0$ or $q = 1$. The numerical values of $\Delta\lambda$ and q at which large deviations of $\bar{F}_\alpha$ are observed suggest that without correction for nonstationarity across trials, unitary event analysis (Grün et al. 2002a) exhibits a certain robustness.

In the next section, we develop an analytical description of the effects observed in the two-rate-state model. This framework allows us to attain an understanding of the dependence of $\bar{F}_\alpha$ on $(\Delta\lambda, q)$ and, more importantly, insight into the meaning of $\bar{F}_\alpha$ for the single data set with which the experimenter is confronted.

4 Analysis of the two-rate-state model

Let us start the investigation with experiments composed of only two trials ($M = 2$). The limited combinatorics of the setup allows us to visualize all possible rate constellations and derive conclusions in a straightforward manner. At the same time the system is rich enough to show most of the effects observable in a system of more realistic size, say, $M \approx 30$. Equipped with basic insight into the properties of the model, we introduce, in the second step, a compact notation permitting an analytical treatment of systems with $M > 2$.

4.1 Two trials

In order to arrive at interpretable expressions, we pursue a strategy introduced earlier (Grün et al. 1999; Grün et al. 2002b) and develop the theory for the mean coincident count. Results are verified afterwards by simulations incorporating the full variability of coincidence counts originating from the variability of spike counts and random spike timing. The first task is to compute the expected number of coincidences n^{pred} for the constellation of rate states in a particular experiment:

$$n^{\mathrm{pred}} = T \cdot \left(\sum_{i=1}^{M} p_{1,i} \cdot p_{2,i} \right) . \tag{3}$$

The expected number of coincidences in a single trial is the product of the spike probabilities in the trial multiplied by the number of time steps T. The total number of coincidences n^{pred} is the sum of the predictors for individual trials (see also Pauluis and Baker 2000; Grün et al. 2002b). The expected number of coincidences based on averaging over trials is given by Eq. 2.

In single trials we can find one of four possible rate constellations:

$$s_1 = \begin{bmatrix} p_1 \\ p_1 \end{bmatrix}, \quad s_2 = \begin{bmatrix} p_2 \\ p_2 \end{bmatrix}, \quad s_3 = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \quad s_4 = \begin{bmatrix} p_2 \\ p_1 \end{bmatrix}, \tag{4}$$

where the rate state λ_k is indicated by the corresponding probability p_k of spike emission in time bin h. Row l of vectors s_i represents the rate state of neuron l. Possible experimental outcomes are described by the set of combinations (s_i, s_j). Thus, we have a natural arrangement of the 4^2 different experimental outcomes in a 4×4 matrix. For each of the 16 constellations we can compute n^{pred} as well as $\bar{n}^{\mathrm{pred}}$ and display the resulting expressions in matrix form (Fig. 4b,c; matrix elements show coincidence probability, and T is omitted for clarity). The two matrices differ only in their antidiagonal elements. In Fig. 4b pairs of elements on the antidiagonal are identical because of symmetry. In Fig. 4c all elements on the antidiagonal are identical because both neurons have an identical number of contributions from p_1 and p_2. It can be shown that all elements off the antidiagonal are identical for n^{pred} and $\bar{n}^{\mathrm{pred}}$ because at least one of the neurons is stationary across trials. Independent of $\Delta\lambda$ (assuming $\Delta\lambda > 0$), the relationship between n^{pred} and $\bar{n}^{\mathrm{pred}}$ for the different experimental outcomes is

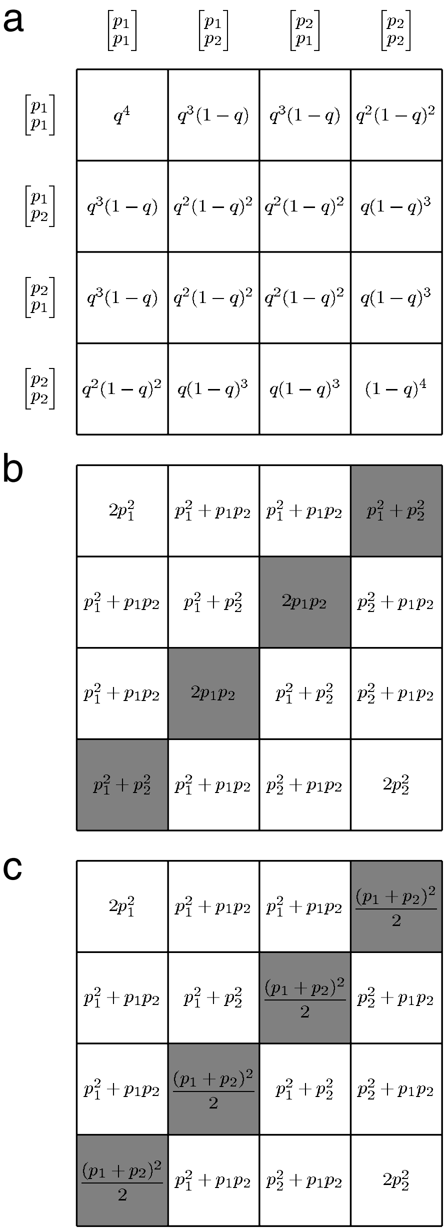
$$\begin{aligned} \bar{n}^{\mathrm{pred}} &< n^{\mathrm{pred}} \quad \text{for} \quad (s_1, s_2), (s_2, s_1) \\ \bar{n}^{\mathrm{pred}} &> n^{\mathrm{pred}} \quad \text{for} \quad (s_3, s_4), (s_4, s_3) \\ \bar{n}^{\mathrm{pred}} &= n^{\mathrm{pred}} \quad \text{else} . \end{aligned} \tag{5}$$

The effect of the second parameter q describing nonstationarity is expressed by the fact that different experimental outcomes generally do not occur with equal probability. The probability of a specific constellation (s_i, s_j) occurring is

$$Q(s_i, s_j) = q^{r_1} \cdot (1 - q)^{r_2} , \tag{6}$$

where r_k denotes the total number of data segments (irrespective of neuron and trial) in (s_i, s_j) subject to spike rate λ_k. Using the same layout as for n^{pred} and $\bar{n}^{\mathrm{pred}}$, Q can be arranged in a matrix (Fig. 4a). Obviously, the sum over all matrix elements equals unity and for $q = 0.5$, all constellations are equally likely to occur (0.5^4) because of $r_1 + r_2 = M$. In general, there are five different occupation probabilities. Due to the permutability of the trials, Q is symmetrical with respect to the diagonal. Elements on the antidiagonal are identical.

The sum over all matrix elements n^{pred}, weighted by the corresponding probability of occurrence Q, represents the coincidence count obtained in the average taken over the ensemble of possible experimental outcomes. Interestingly, the ensemble average of n^{pred} is identical to the ensemble average of $\bar{n}^{\mathrm{pred}}$. Thus, the ensemble average of the mean coincidence count does not provide us with any knowledge about nonstationarity across trials.



$$\begin{array}{c|cccc} & \begin{bmatrix} p_1 \\ p_1 \end{bmatrix} & \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} & \begin{bmatrix} p_2 \\ p_1 \end{bmatrix} & \begin{bmatrix} p_2 \\ p_2 \end{bmatrix} \\ \hline \begin{bmatrix} p_1 \\ p_1 \end{bmatrix} & q^4 & q^3(1-q) & q^3(1-q) & q^2(1-q)^2 \\ \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} & q^3(1-q) & q^2(1-q)^2 & q^2(1-q)^2 & q(1-q)^3 \\ \begin{bmatrix} p_2 \\ p_1 \end{bmatrix} & q^3(1-q) & q^2(1-q)^2 & q^2(1-q)^2 & q(1-q)^3 \\ \begin{bmatrix} p_2 \\ p_2 \end{bmatrix} & q^2(1-q)^2 & q(1-q)^3 & q(1-q)^3 & (1-q)^4 \end{array}$$

$$\begin{array}{|c|c|c|c|} \hline 2p_1^2 & p_1^2+p_1p_2 & p_1^2+p_1p_2 & p_1^2+p_2^2 \\ \hline p_1^2+p_1p_2 & p_1^2+p_2^2 & 2p_1p_2 & p_2^2+p_1p_2 \\ \hline p_1^2+p_1p_2 & 2p_1p_2 & p_1^2+p_2^2 & p_2^2+p_1p_2 \\ \hline p_1^2+p_2^2 & p_1^2+p_1p_2 & p_2^2+p_1p_2 & 2p_2^2 \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|} \hline 2p_1^2 & p_1^2+p_1p_2 & p_1^2+p_1p_2 & \frac{(p_1+p_2)^2}{2} \\ \hline p_1^2+p_1p_2 & p_1^2+p_2^2 & \frac{(p_1+p_2)^2}{2} & p_2^2+p_1p_2 \\ \hline p_1^2+p_1p_2 & \frac{(p_1+p_2)^2}{2} & p_1^2+p_2^2 & p_2^2+p_1p_2 \\ \hline \frac{(p_1+p_2)^2}{2} & p_1^2+p_1p_2 & p_2^2+p_1p_2 & 2p_2^2 \\ \hline \end{array}$$

Fig. 4. Occupation probability and coincidence probability for a system of two neurons and two trials. The system has 16 joint-rate states defined by the possible combinations of the two spike rates λ_1 and λ_2 at four independent positions. **a** Probability of finding the system in a specific rate state, organized as a matrix. Rate λ_k is expressed by the probability p_k of observing a spike in a time bin of duration h. Matrix elements of a row correspond to identical rates in trial 1, elements of a column to identical rates in trial 2. Component l of the vector to the left of row i and above column j specifies the probability p_k for neuron l in the respective trial. The matrix elements represent the probability that the system is in the specified state, where p_1 occurs with probability q and p_2 with $1 - q$ (Fig. 2a). **b** Expected number of coincidences per time bin, summed over the two trials (same layout as in **a**). **c** Expected number of coincidences per time bin based on rate average across trials summed over all trials (same layout as in **a**)

Note that for $M > 2$ a different matrix arrangement is more suitable: the two neurons (instead of the trials) span the dimensions of the matrix, and the 2^M rate combinations of the trials make up the rows and columns. This way the object remains two-dimensional. In this arrangement it is easy to see that the number of rate constellations where at least one of the neurons is stationary is $4 \cdot 2^M - 4$.

4.2 False positives

In the last section, we saw that the relationship between n^{pred} and $\bar{n}^{\text{pred}}$ is different for different experimental outcomes, while the comparison of the ensemble average of the two measures does not uncover the effect of nonstationarity. Therefore we need to investigate the origin of false positives in individual rate constellations (s_i, s_j). For a specific rate constellation, n^{pred} is the parameter of a Poisson distribution describing the distribution of coincidence counts occurring in realizations of this constellation of rates. Thus, the fraction of false positives f_α is the probability of obtaining a coincidence count equal to or larger than the significance threshold n_α:

$$f_\alpha = \sum_{n=n_\alpha}^{\infty} \frac{\left(n^{\text{pred}}\right)^n}{n!} \cdot \exp\left(-n^{\text{pred}}\right) \ , \tag{7}$$

approximating the adjusted α from below. However, given an experimental data set, n^{pred} is unknown and $\bar{n}^{\text{pred}}$ represents our estimate of the expected coincidence count. For some rate constellations n^{pred} and $\bar{n}^{\text{pred}}$ are not identical. Thus, the distribution of coincidence counts describing a realization of (s_i, s_j) (with mean n^{pred}) is shifted with respect to the distribution (with mean $\bar{n}^{\text{pred}}$) assumed for the significance test.

This situation is depicted in Fig. 5. f_α is still given by Eq. 7, but now the significance threshold n_α is adjusted for a distribution (with mean $\bar{n}^{\text{pred}}$) incompatible with the distribution underlying the data. For $\bar{n}^{\text{pred}} < n^{\text{pred}}$ (constellations (s_1, s_2) and (s_2, s_1)) the probability of observing coincidence counts larger than n_α exceeds α, and therefore a surplus in false positives is generated. When the estimate of the mean coincidence count $\bar{n}^{\text{pred}}$ is larger than the real mean of the distribution n^{pred} (constellations (s_3, s_4) and (s_4, s_3)), the fraction of false positives f_α falls below α.

Figure 6 shows the dependence of f_α on the parameters of nonstationarity $(\Delta\lambda, q)$ for the individual rate constellations (s_i, s_j). The dependence on $\Delta\lambda$ is shown in Fig. 6a. As expected from our analysis above, f_α is independent of $\Delta\lambda$ for matrix elements off the antidiagonal ($\bar{n}^{\text{pred}} = n^{\text{pred}}$). For the two constellations where both neurons change their rate in corresponding trials, (s_1, s_2) and (s_2, s_1), f_α increases rapidly with increasing $\Delta\lambda$. In contrast, f_α vanishes for the two constellations, (s_3, s_4) and (s_4, s_3), where both neurons have opposing rates in corresponding trials.

The dependence of f_α on q for constant $\Delta\lambda$ is shown in Fig. 6b. Variation of q does not lead to a variation in the fraction of false positives for individual constellations since q does not enter Eq. 7. However, assuming homogeneous sampling of the possible experimental outcomes, we need to consider that individual constellations do not occur with equal probability. The contribution of a rate constellation to the false positives observed in the ensemble is given by f_α, weighted by the probability of occurrence of the rate constellation

$$F_\alpha = f_\alpha \cdot Q \ . \tag{8}$$

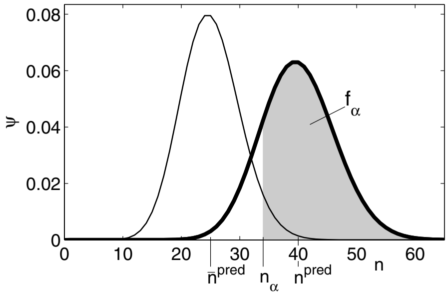


Fig. 5. Definition of false positives in the detection of significant spike coincidences. The significance of an empirical coincidence count n is evaluated on the basis of the distribution of coincidence counts ψ. Assuming a Poisson distribution (*thin curve*) parameterized by the expected coincidence count $\bar{n}^{\text{pred}}$ and a significance level α, n_α or more coincidences are required for significance. $\bar{n}^{\text{pred}}$ is estimated by averaging over trials. Under conditions of nonstationarity across trials, $\bar{n}^{\text{pred}}$ may be lower than the real expectation value n^{pred} (e.g., matrix elements (1,4) and (4,1) in Fig. 4). The distribution of coincidence counts of the system (*thick curve*) does not correspond to the distribution used for testing significance (*thin curve*). The probability f_α of obtaining a coincidence count $\geq n_\alpha$ (*gray area*) may exceed the adjusted fraction α (*gray area below thin curve*). f_α is the probability of obtaining a significant outcome in a data set of independent neurons: the fraction of false positives

The dependence of F_α on q is shown in Fig. 6b superimposed on the results for f_α. Different rate constellations exhibit a maximum of F_α at different values of q. The two homogeneous rate constellations (s_1, s_1) and (s_2, s_2) reach their maximum at $q = 0$ and $q = 1$, respectively. Here the probability of occurrence is unity and $F_\alpha = f_\alpha$. Rate constellations with the same contributions from λ_1 and λ_2 reach the maximum at $q = 0.5$. At this value all matrix elements occur with equal probability 1/16. In particular, at $q = 0.5$ the two rate constellations (s_1, s_2) and (s_2, s_1), with an f_α of practically 1 (Fig. 6a, at large $\Delta\lambda$), exhibit the maximal $F_\alpha = 1/16$ a matrix element can obtain. For comparison Fig. 6a shows the $\Delta\lambda$ dependence of F_α for a fixed q of 0.5. The contribution of (s_1, s_2) and (s_2, s_1) to the ensemble average limited by a maximal occurrence probability of 1/16 is visualized by the saturation of F_α at large $\Delta\lambda$.

We are now in a position to construct the ensemble average of false positives $\bar{F}_\alpha$ and compare our results to the simulations shown in Fig. 3. Independently of $\Delta\lambda$, the different rate constellations can be arranged into three classes (Eq. 5). Elements in the class with $\bar{n}^{\text{pred}} = n^{\text{pred}}$ generate a fraction of false positives $f_\alpha = \alpha$,

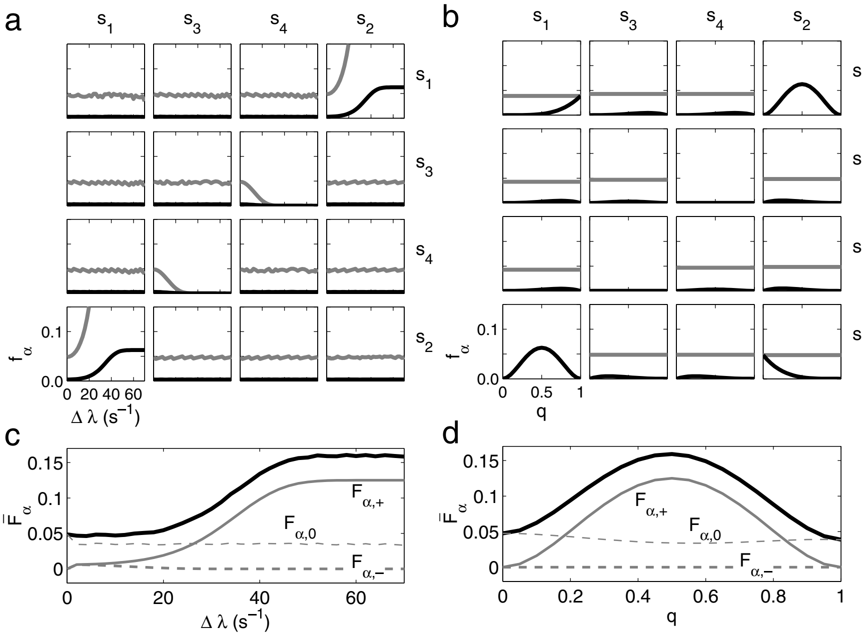


Fig. 6. False positives for possible experimental outcomes (*upper row*) and contributions to the ensemble average (*lower row*). Nonstationarity across trials parameterized by difference in rate levels $\Delta\lambda$ (*left column*) and occupation probability of rate states q (*right column*). **a** False positives f_α (*gray curves*, $\alpha = 0.05$) for individual states as a function of $\Delta\lambda$. Same system and layout of matrix as in Fig. 4, the four possible rate states of a single trial are specified by s_i (trial 1 vertical, trial 2 horizontal). The *black curve* shows the contribution to the ensemble average $\bar{F}_\alpha$: f_α weighted by the $\Delta\lambda$-independent probability of finding the system in the particular state (Fig. 4a). Graphs are computed for $(\lambda_1 + \lambda_2)/2 = 50\ \mathrm{s}^{-1}$, $q = 0.5$, trial duration $\mathcal{T} = 50\,\mathrm{s}$, and bin size $h = 1$ ms. Only the antidiagonal exhibits a dependence on $\Delta\lambda$. Matrix elements representing correlated rate changes, $(1, 4)$ and $(4, 1)$, show a dramatic increase in f_α. For anticorrelated rate changes, elements $(2, 3)$ and $(3, 2)$, f_α decreases with increasing $\Delta\lambda$. **b** False positives for individual states as a function of q. Same system, organization, and parameters as in **a**. Graphs are computed for $\Delta\lambda = 60\,\mathrm{s}^{-1}$; here certain states exhibit large f_α in **a** (at $q = 0.5$). *Gray curves* represent the q-independent value of f_α taken from **a** at the specified $\Delta\lambda$ (not shown: $f_\alpha \approx 0.5$ in $(1, 4)$,$(4, 1)$ and ≈ 0 in $(2, 3), (3, 2)$). Also elements off the antidiagonal show a q-dependence of $\bar{F}_\alpha$. The two states contributing most of the false positives exhibit a maximum of the weighted f_α at $q = 0.5$. **c** Contributions to the ensemble average $\bar{F}_\alpha$ as a function of $\Delta\lambda$. $\bar{F}_\alpha$ (*black curve*) is the sum of the f_α in individual states weighted by the probability of occurrence (sum of *black curves* in **a**). $\bar{F}_\alpha$ can be decomposed into a sum over contributions from three classes of states: $F_{\alpha,+}$ (*gray solid*) is the sum over states where the expected number of coincidences n^{pred} is larger than the cross-trial estimate $\bar{n}^{\mathrm{pred}}$ (elements $(1, 4)$ and $(4, 1)$ in **A**), $F_{\alpha,0}$ (*thin dashed*) collects the states with $n^{\mathrm{pred}} = \bar{n}^{\mathrm{pred}}$ (off antidiagonal elements in **a**), and $F_{\alpha,-}$ (*thick dashed*) the remaining ones with $n^{\mathrm{pred}} < \bar{n}^{\mathrm{pred}}$ (elements $(2, 3)$ and $(3, 2)$ in **a**). With $F_{\alpha,0}$ close to α (up to discreteness of the distribution) and $F_{\alpha,-}$ close to 0, $\bar{F}_\alpha$ is dominated by $F_{\alpha,+}$. **d** Contributions to the ensemble average as a function of q. Same representation as in **c**. The classification introduced in **c** also decomposes $\bar{F}_\alpha$ with respect to q

and we term their collective contribution to the ensemble average $F_{\alpha,0}$. The elements in the class with $\bar{n}^{\mathrm{pred}} < n^{\mathrm{pred}}$ have identical values for $\bar{n}^{\mathrm{pred}}$ and identical values for n^{pred}. Therefore, f_α is identical for all elements of this class and the total contribution to the ensemble average is called $F_{\alpha,+}$. The same statement is valid for the remaining class with $\bar{n}^{\mathrm{pred}} > n^{\mathrm{pred}}$ and contribution $F_{\alpha,-}$. Thus,

$$\bar{F}_\alpha = F_{\alpha,0} + F_{\alpha,+} + F_{\alpha,-} \tag{9}$$

and with factorized q-dependence

$$\begin{aligned}\bar{F}_\alpha &= (1 - 2w) \cdot f_\alpha(\bar{n}^{\mathrm{pred}} = n^{\mathrm{pred}}) \\ &\quad + w \cdot f_\alpha(\bar{n}^{\mathrm{pred}} < n^{\mathrm{pred}}) \\ &\quad + w \cdot f_\alpha(\bar{n}^{\mathrm{pred}} > n^{\mathrm{pred}})\end{aligned} \tag{10}$$

where $w = 2q^2(1 - q)^2$.

Fig. 6c shows $\bar{F}_\alpha$ and its constituents as functions of $\Delta\lambda$. Clearly, the ensemble average $\bar{F}_\alpha$ is dominated by $F_{\alpha,+}$, summarizing the effect of the corresponding rate constellations (s_1, s_2) and (s_2, s_1). The offset $\alpha \cdot 12/16$ is explained by the contribution of $F_{\alpha,0}$. The dominance of $F_{\alpha,+}$ is preserved in the q-dependence of $\bar{F}_\alpha$ (Fig. 6d). The bell-shaped q-dependence of (s_1, s_2) and (s_2, s_1) observed in Fig. 6b shapes the ensemble average. The drop of the contribution by $F_{\alpha,0}$ from α at $q = 0$ and $q = 1$ to $\alpha \cdot 12/16$ at $q = 0.5$ is irrelevant.

Comparison of Fig. 6c,d with the simulation results in Fig. 3 already shows a good qualitative correspondence. The monotonic dependence of $\bar{F}_\alpha$ on $\Delta\lambda$ and the non-monotonic dependence on q are captured. However, there are important differences. In the simulation, $\bar{F}_\alpha$ starts deviating from α at larger $\Delta\lambda$ and does not reach saturation (Fig. 3a). Furthermore, the q-dependence of $\bar{F}_\alpha$ observed in Fig. 3b is asymmetrical and has a maximum at ≈ 0.7, while Fig. 6d predicts a bell-shaped curve, the maximum located at $q = 0.5$.

The differences between the analytical treatment and the simulation are the Poisson approximation for the distribution of coincidence counts and the restriction to only two trials ($M = 100$ trials in the simulation). In the next section, we extend the analysis to $M > 2$ trials to check whether deviations from the simulations are due to our arbitrary restriction of the theory to two trials. In the analytical treatment, we have identified specific rate constellations dominating the fraction of false positives. This idea guides the analysis in the next section and is extended to provide a compact classification of rate constellations interpretable in the light of experimental data. It is demonstrated that the peak location of $\bar{F}_\alpha$ at $q = 0.7$ is typical for a larger, and more realistic, number of trials.

4.3 Framework for M trials

The number of possible rate constellations in M trials is given by 4^M and thus increases rapidly with the number of trials. In electrophysiological experiments, the number of trials is typically in the range of 20–100. Therefore, we must leave the direct approach used in Sect. 4.1, where we evaluated each individual rate constellation, and find a more compact description that is immune to the combinatorial explosion of possible constellations. In the study of a system with two trials, we have seen that the sequence of trials does not enter the calculation of false positives (and it should not do so by the definition of a "trial"). This leads us to a description where only the possible rate constellations in a single trial and the number of their occurrences in a data set of M trials are specified. We have already introduced a notation (Eq. 4) for the four possible rate constellations a trial can exhibit: s_1, s_2, s_3, s_4. Ignoring the sequence of rate constellations $s_{i(1)}, \ldots, s_{i(M)}$, an experiment is described by a vector $[n_1, n_2, n_3, n_4]$ specifying the number of trials n_i with rate constellation s_i. A valid experiment trivially fulfills $M = n_1 + n_2 + n_3 + n_4$. Thus, we have now changed our point of view (see illustration in Fig. 7). Instead of dealing with two neurons, we describe the system in terms of the M indistinguishable trials, each occupying one of four possible states. Thus, the total number of system states $[n_1, n_2, n_3, n_4]$ is given by the number of ways M indistinguishable "particles" can be distributed over four states. Equivalently, this is the number of "compositions" of M into four integers $n_i \geq 0$ (Petkovsek et al. 1996; Nijnhuis and Wilf 1978):

$$K_M = \begin{pmatrix} M+4-1 \\ 4-1 \end{pmatrix} = \begin{pmatrix} M+4-1 \\ M \end{pmatrix} \quad . \tag{11}$$

The two forms found in the literature are connected by the symmetry identity. We call $[n_1, n_2, n_3, n_4]$ a macrostate of the system because the "energies" (variables of interest) n^{pred} and $\bar{n}^{\text{pred}}$ depend only on the quadruplet. The corresponding number of different M-trial sequences (microstates) is given by the multinomial coefficient

$$K_{[n_1,n_2,n_3,n_4]} = \begin{pmatrix} M \\ n_1, n_2, n_3, n_4 \end{pmatrix} = \frac{M!}{n_1! \cdot n_2! \cdot n_3! \cdot n_4!} \quad . \tag{12}$$

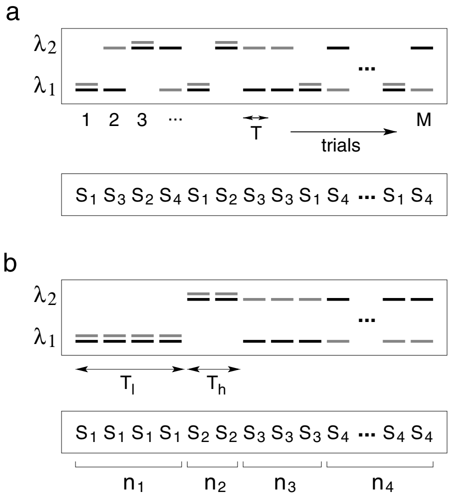


Fig. 7. Trials as particles. For $M > 2$ trials the viewpoint is changed from two particles (the neurons) being in any of 2^M combined rates states to a system of M particles (the trials) each being in one of four possible states. **a** Sketch of a particular M-trial rate constellation (microstate) of the system. Each trial (horizontal $1 \ldots M$, duration $\mathcal{T}$ indicated by length of bars) is in one of the states $\{s_1, s_2, s_3, s_4\}$ (cf. Fig. 4). States correspond to possible rate level combinations (λ_1, λ_2 vertical) of neuron 1 (*black bars*) and neuron 2 (*gray bars*). The microstate of the system is described by the sequence $s_{i(1)}, \ldots, s_{i(M)}$ (shown for example constellation). **b** Corresponding macrostate of the system. Same representation as in **a**, now trials are sorted with respect to s_i. The canonical arrangement of trials (a different microstate) represents a coherent rate step from low rate (s_1) to high rate (s_2) followed by anticorrelated rates (s_3 and s_4). The macrostate is described by $[n_1, n_2, n_3, n_4]$ specifying the number of trials n_i in s_i

The expressions for the coincidence counts in a macrostate (Eqs. 3 and 2) can now be stated in explicit form. The expected number of coincidences in a macrostate is

$$n^{\text{pred}} = T \cdot (n_1 \cdot p_1 p_1 + n_2 \cdot p_2 p_2 + n_3 \cdot p_1 p_2 + n_4 \cdot p_2 p_1) \tag{13}$$

and the number of coincidences predicted on the basis of rate averages reads

$$\begin{aligned} \bar{n}^{\text{pred}} &= T \cdot \frac{1}{M}(n_1 p_1 + n_2 p_2 + n_3 p_1 + n_4 p_2) \\ &\quad \times (n_1 p_1 + n_2 p_2 + n_3 p_2 + n_4 p_1) \quad . \end{aligned} \tag{14}$$

Note that each macrostate is assigned a unique value for n^{pred} and $\bar{n}^{\text{pred}}$; however, different macrostates may have identical values. The probability of observing a microstate (cf. Eq. 6) corresponding to $[n_1, n_2, n_3, n_4]$ is given by

$$\begin{aligned} Q_{[n_1,n_2,n_3,n_4]} &= (qq)^{n_1} \cdot ((1-q)(1-q))^{n_2} \\ &\quad \times (q(1-q))^{n_3} \cdot ((1-q)q)^{n_4} \quad , \end{aligned} \tag{15}$$

and consequently the occupation probability of the macrostate is

$$P_{[n_1,n_2,n_3,n_4]} = K_{[n_1,n_2,n_3,n_4]} \cdot Q_{[n_1,n_2,n_3,n_4]} \quad . \tag{16}$$

Trivially, the sum over the occupation probabilities of all macrostates equals unity.

The fraction of false positives f_α in a given macrostate can be calculated as shown for the case of $M = 2$ (Eq. 7). Hence, the contribution of the macrostate to the ensemble average is

$$F_{\alpha[n_1,n_2,n_3,n_4]} = P_{[n_1,n_2,n_3,n_4]} \cdot f_{\alpha[n_1,n_2,n_3,n_4]} \ . \tag{17}$$

The expression for the fraction of false positives observed in the ensemble average $\bar{F}_\alpha$ results from a generalization of Eq. 8:

$$\bar{F}_\alpha = \sum_{[n_1,n_2,n_3,n_4]} P_{[n_1,n_2,n_3,n_4]} \cdot f_{\alpha[n_1,n_2,n_3,n_4]} \ , \tag{18}$$

where the sum extends over all macrostates (compositions of M). Algorithms for the evaluation of Eq. 18 are available (e.g., Nijnhuis and Wilf 1978).

4.4 Evaluation of the M-trial system

The tools developed in the last section allow us to investigate the fraction of false positives in the ensemble average in dependence on the number of trials M. Figure 8 shows $\bar{F}_\alpha$ as a function of the parameters of nonstationarity $(\Delta\lambda, q)$ for different M. Interestingly, $\bar{F}_\alpha$ exhibits a nonmonotonic dependence on M. Starting at $M = 2$, $\bar{F}_\alpha$ increases with an increasing number of trials, reaches a maximum, and declines again (Fig. 8a). At $M = 100$, $\bar{F}_\alpha$ reaches values well below the curve for $M = 2$. This dependence on M is also expressed in the peak amplitude of $\bar{F}_\alpha$ in Fig. 8b. In addition, the location of the peak shifts from $q = 0.5$ at $M = 2$ to $q \approx 0.7$ for large M. Being defined on $q \in [0, 1]$, the bell-shaped curve becomes less symmetrical with increasing M. When $\bar{F}_\alpha$ in Fig. 8a for $M = 2$ has reached saturation (cf. Fig. 6c), it still increases for larger M.

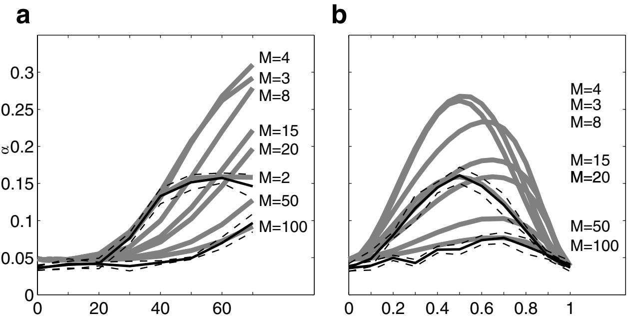


Fig. 8. False positives in systems with more than two trials. Same graphs and parameters as in Fig. 6c,d for different numbers of trials (*gray M* as labeled, $\alpha = 0.05$, $M\mathcal{T} = 100\,\mathrm{s}$). **a** False positives as a function of $\Delta\lambda$. The ensemble average $\bar{F}_\alpha$ increases monotonically with $\Delta\lambda$ for all M ($\Delta\lambda$ in steps of $10\,\mathrm{s}^{-1}$). The curve for $M = 2$ saturates for $\Delta\lambda > 40\,\mathrm{s}^{-1}$ at about $2/16$, for larger $\Delta\lambda$ curves for moderate M exceed this level. With further increasing M, $\bar{F}_\alpha$ passes a maximum and drops back below the value for $M = 2$. For a given $\Delta\lambda$, $\bar{F}_\alpha$ is nonmonotonic in M. Superimposed are simulation results (*black*) for $M = 2$ and $M = 100$ (cf. Fig. 3a). The *black solid curves* represent the mean of 5 times 1000 experiments, the *dashed curves* indicate mean ± standard deviation estimated from the five repetitions. Data for $M = 100$ copied from Fig. 3a. **b** False positives as a function of q. $\bar{F}_\alpha$ generally shows the nonmonotonic dependence on q already observed for $M = 2$ (Fig. 6), and the nonmonotonic dependence on M seen in **a** (q in steps of 0.1). With increasing M the peak of $\bar{F}_\alpha$ shifts from $q = 0.5$ to $q \approx 0.7$. Superimposed are simulation results as described in **a**. Data for $M = 100$ copied from Fig. 3b

Comparison with simulations of the full system for $M = 2$ and $M = 100$ (Fig. 8, black curves) shows that our analysis based on the mean coincidence counts (Eq. 18) accurately describes the dependence of $\bar{F}_\alpha$ on $(\Delta\lambda, q)$. In particular, the differences observed between the simulation results of Fig. 3 and the two-trial model (see Sect. 4.2, Fig. 6c,d) are explained by the restriction of the number of trials in the latter.

The surprising dependence of the peak amplitude of $\bar{F}_\alpha$ on M with a maximum at an intermediate value of M and low values at large M may already indicate that concentrating solely on the ensemble average is a dangerous procedure. Consider the following scenario: there are rate constellations that are much more effective in generating false positives than others and require a minimal number of trials for maximal performance. These specific experimental outcomes dominate the ensemble average at moderate M but become overwhelmed by the exponentially growing number of microstates (4^M) at large M. Obviously, for the experimenter confronted with a single data set better indicators than the ensemble average would be required in the above scenario. To gain insight into the properties of problematic rate constellations, further structuring of the set of macrostates is required. The second problem left unsolved by the analysis of the ensemble average is the question of why the peak of F_α shifts to $q \approx 0.7$ for large M. The next section approaches both of these problems and demonstrates their relationship.

4.5 Origin of false positives

In an earlier work (Grün et al. 2002b), we discussed the effect of temporal nonstationarity of rate on the fraction of false positives. In this situation, a coherent stepwise change of rate in the two neurons represents a worst-case scenario because the predictor for the coincidence count based on the temporal rate averages strongly underestimates the mean coincidence count. In the present study, spike rates are by definition stationary in time. However, we can think of the consecutive trials $1, \ldots, M$ as a temporal sequence of time segments. The ordering of simultaneous rates s_i in a microstate corresponding to macrostate $[n_1, n_2, n_3, n_4]$ does not enter the expressions for n^{pred} and $\bar{n}^{\mathrm{pred}}$. Therefore, we are free to choose any microstates as a canonical representation of the macrostate. We decide on the ordering whereby both neurons initially have low rates (n_1 times s_1), followed by a region where both neurons have high rates (n_2 times s_2), and two regions with opposing rates (n_3 times s_3, n_4 times s_4). This representation justifies our choice of indices for the different s_i (see Fig. 7b for illustration). Viewed as a temporal sequence, a coherent rate step is followed by anticorrelated rate levels.

Let us now parameterize the first two occupation numbers n_1 and n_2 by two measures more directly describing the coherent rate step. The relative length of the rate step is defined as

$$r = \frac{n_1 + n_2}{M} \ , \tag{19}$$

and the relative duration of the low rate and the high rate regime in the rate step is

$$q_r = \frac{n_1}{n_1 + n_2} \ . \tag{20}$$

In Fig. 9a, we have used the new variables r and q_r to structure the set of macrostates of a system with $M = 10$ trials in terms of false positives. We conclude that macrostates containing a rate step generate a high fraction of false positives. The longer the rate step (large r), the larger the fraction of false positives. For constant r, f_α reaches a maximum at the largest possible q_r smaller than 0.7. With increasing r the number of the allowed fractions q_r increases (Eq. 20). Hence, the observed maximum of f_α approaches 0.7 with increasing r due to the discrete nature of q_r.

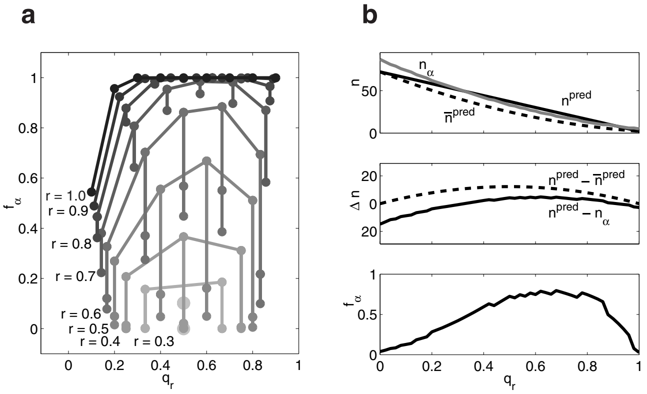


Fig. 9. Coherent rate steps explain the origin of false positives in M-trial systems. The first two occupation numbers of a macrostate $[n_1, n_2, n_3, n_4]$ can be parameterized by the relative length of the coherent rate step $r = (n_1 + n_2)/M$ and the relative duration of the two regimes in the rate step $q_r = n_1/(n_1 + n_2)$ (cf. Fig. 7b). **a** False positives f_α (vertical) as a function of rate step duration r and rate step proportion q_r (horizontal) for the individual macrostates of an example system ($\Delta\lambda = 70\,\mathrm{s}^{-1}$, $q = 0.5$, $M = 10$, $\mathcal{T} = 1000$ ms, $\alpha = 0.05$). Curves connect data (*dots*) for macrostates with identical r (as labeled, *dark gray* encodes large r, *large dots* for $r = 0.2$). Generally, the longer the rate step, the higher f_α. f_α is nonmonotonic in q_r. For large r, f_α reaches a maximum at $q_r \approx 0.7$. M restricts the allowed values of q_r to fractions of two integers $n_1/(rM)$. For a given r, the maximum occurs at the largest possible $q_r \leq 0.7$. Multiple values of f_α at identical (r, q_r) reflect macrostates distinguished by n_3 and n_4 (*vertically connected dots*). **b** Analysis of a coherent rate step in a continuous recording, corresponding to $r = 1$ and $M \to \infty$ with $M\mathcal{T} = 10 \cdot 1000$ ms (no discretization of q_r). The upper graph shows the dependence of n^{pred} (*black solid*), $\bar{n}^{\mathrm{pred}}$ (*dashed*), and n_α (*gray*, cf. Fig. 5) on q_r. Coincidence counts decrease with increasing fraction of time spent in the low-rate regime. At intermediate values of q_r the expected number of coincidences n^{pred} exceeds the significance threshold n_α. The middle graph shows the difference between expectation value and its estimate $n^{\mathrm{pred}} - \bar{n}^{\mathrm{pred}}$ (*dashed*) and the difference between expectation and significance threshold $n^{\mathrm{pred}} - n_\alpha$ (*solid*). The maximum of $n^{\mathrm{pred}} - \bar{n}^{\mathrm{pred}}$ occurs at $q_r = 0.5$, while the maximum of $n^{\mathrm{pred}} - n_\alpha$ occurs at $q_r \approx 0.7$. The lower graph shows the fraction of false positives f_α as a function of q_r. f_α is small when one of the rate regimes dominates. The position of the maximum can be explained by the maximum of $n^{\mathrm{pred}} - n_\alpha$ (middle graph). Steps are due to the discrete nature of n_α

Coherent rate steps are the most efficient generators of false positives in the ensemble of macrostates. Thus, structuring of the macrostates by (r, q_r) has reduced the problem to the analysis of false positives in a rate step. Fig. 9b shows n^{pred} and $\bar{n}^{\mathrm{pred}}$ as a function of q_r. With the expansion of the low rate regime (increasing q_r) the number of coincidences declines. However, the difference in the coincidence counts exhibits a maximum at $q_r = 0.5$ (Grün et al. 2002b). The error in the estimation of n^{pred} is largest when the low and the high rate regimes occur in equal proportion. The error vanishes if one regime dominates the average. The fraction of false positives, however, does not directly depend on the difference in coincidence counts. As f_α is the part of the distribution of coincidence counts to the right of the significance threshold n_α, the difference of n^{pred} and n_α is more important. This difference exhibits a maximum at $q_r \approx 0.7$ when the low rate regime already covers the larger part of the rate step (determines $\bar{n}^{\mathrm{pred}}$), and the coincidences created by the narrow high rate regime occur as a "surprise." The location of the maximum of f_α in the parameterization of the macrostates in Fig. 9a is now explained.

It remains to be answered at which parameters of nonstationarity the system becomes susceptible to the occurrence of macrostates contributing large f_α. Using Eq. 15 the occupation probability of a macrostate representing a pure rate step $[n_1, n_2, 0, 0]$, $r = 1$ is

$$Q_{[n_1,n_2,0,0]} = q^{2^{n_1}} \cdot (1-q)^{2^{n_2}} \tag{21}$$

$$= q^{2q_r M} \cdot (1-q)^{2(1-q_r)M} \ . \tag{22}$$

Thus, $Q_{[n_1,n_2,0,0]}$ has a maximum at $q = q_r$. In other words, pure rate steps with a specific proportion of the low and the high rate regime have the largest probability of occurring in the ensemble when the distribution of rates in the nonstationarity matches this proportion.

We have now seen that the fraction of false positives in the ensemble average $\bar{F}_\alpha$ reaches a maximum at the nonstationarity q where the probability of occurrence for pure rate steps with the most effective proportion q_r is maximal. Systems with a low number of trials can only realize certain q_r. Therefore, the location of the peak in $\bar{F}_\alpha$ (Fig. 8b) depends on the number of trials.

5 Discussion

Before we present an experimental data set to exemplify the results of our model analysis, let us first discuss solutions to the problem of nonstationarity across trials and consequences of nonstationarity for individual data sets.

5.1 Correcting for nonstationarity

In experimental data, covariation of spike rates across trials may occur for various reasons (Sect. 1). Such covariation of spike rate favors rate constellations corresponding to the coherent rate steps defined in Sect. 4.5. Figure 10 illustrates two strategies for coping with nonstationarity across trials. The first generates the distribution of coincidence counts on the basis of the spike counts in individual trials using a Monte-Carlo approach (Fig. 10a). The procedure still assumes that data are generated by a Poisson process (independent spike times). In the different surrogate data sets the spike counts are conserved, and the distribution assumes the correct mean n^{pred} and shape. In the second strategy, the expected coincidence count is computed for each trial i as $p_{1,i}p_{2,i} \cdot T$. Assuming that the coincidence count in each trial is a Poisson distributed random variable, we can use the sum of the expected coincidence counts as the parameter of a Poisson distribution (see Appendix B). This is why a Poisson distribution can be used to approximate the distribution of coincidence counts in Pauluis and Baker (2000) and Grün et al. (1999). The approach is motivated by the success in describing the fraction of false positives in M-trial systems considering the mean coincidence counts only (Fig. 8).

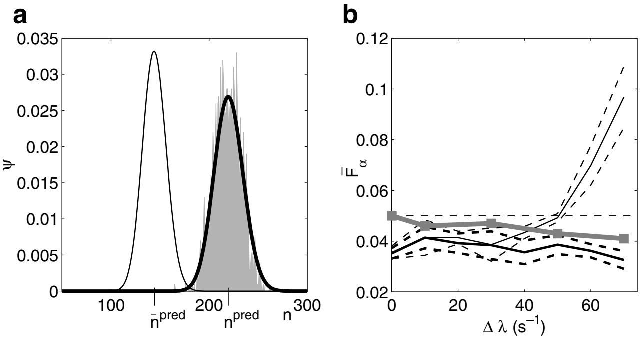


Fig. 10. Procedures to avoid false positives under conditions of nonstationarity across trials. **a** Distribution of coincidence counts ψ underlying the significance test for a simulated example experiment ($\Delta\lambda = 60\,\mathrm{s}^{-1}$, $q = 0.7$, $M = 100$, $\mathcal{T} = 1000\,\mathrm{ms}$, $(\lambda_1 + \lambda_2)/2 = 50\,\mathrm{s}^{-1}$, see Fig. 2). The experimental outcome is characterized by the M pairs of spike counts of the two neurons. The *gray histogram* is the distribution of coincidence counts generated from 1000 surrogate data sets with spike counts identical to the original experiment and homogeneously distributed random spike times (Monte-Carlo). *Black curves* are Poisson distributions with parameters (means) n^{pred} (*thick*, based on spike counts in individual trials) and $\bar{n}^{\mathrm{pred}}$ (*thin*, spike counts averaged across trials). **b** Ensemble average of false positives $\bar{F}_\alpha$ as a function of $\Delta\lambda$ for experiments and distributions as specified in **a**. For the Poisson distribution with parameter $\bar{n}^{\mathrm{pred}}$, $\bar{F}_\alpha$ (*thin black*) exceeds the adjusted level $\alpha = 0.05$ for $\Delta\lambda > 40\,\mathrm{s}^{-1}$ ($\Delta\lambda$ in steps of $10\,\mathrm{s}^{-1}$). The *solid curve* represents the mean of 5 times 1000 experiments ± standard deviation (*dashed*). The *gray curve* shows the results for the distribution generated by Monte-Carlo simulations (1000 experiments at $\Delta\lambda = 10, 30, 50, 70\,\mathrm{s}^{-1}$). The *thick black curve* shows the result for the Poisson distribution with parameter n^{pred} (same representation as for $\bar{n}^{\mathrm{pred}}$). Both curves account for nonstationarity; $\bar{F}_\alpha$ does not exceed α. *Gray curve* runs below α because of discretization (n_α) in macrostates with low expected coincidence count; *thick black curve* additionally lowered by Poisson approximation

An application of the two methods to simulated nonstationary data is shown in Fig. 10b. As we have done at the outset of our investigation (Fig. 3a), the fraction of false positives in the ensemble average is studied as a function of $\Delta\lambda$. While $\bar{F}_\alpha$ resulting from the across trial average of $\bar{n}^{\mathrm{pred}}$ exhibits the strong dependence on $\Delta\lambda$, the $\bar{F}_\alpha$ obtained using the alternative methods stay tightly below the adjusted value of α. $\bar{F}_\alpha$ generated for fixed spike counts (Monte-Carlo approach) falls below α because of the discrete nature of n_α. The approximation of the distribution of coincidence counts by a Poisson distribution leads to an additional lowering.

5.2 Individual experiments

In the following two sections, we demonstrate how nonstationarity across trials is reflected in statements about individual experimental data sets. To this end we discuss a standard tool, the shuffle predictor, and the likelihood of obtaining data with problematic rate constellations.

5.2.1 Consequences for the shuffle predictor. A commonly used predictor in crosscorrelation analysis is the "shuffle predictor" (Perkel et al. 1967; Glaser and Ruchkin 1976; Palm et al. 1988; Aertsen et al. 1989; Eggermont 1990). The idea is to generate a prediction on the basis of the original spike trains by destroying the intertrial correspondence of the simultaneously recorded spike trains but otherwise keeping the properties of the original data set (e.g., spike rate, temporal structure; see also Pipa and Grün 2002) intact. If the predictor is based on all possible trial combinations of the neurons, the predictor is commonly called a "full shuffle predictor." If trial combinations are generated by shifting the trials with respect to each other in wrap-around fashion (covering all possible shifts of one) it is called a "shift predictor." Sometimes a random subset of all possible trial combinations is used instead of the full shuffle. For our purposes the mean coincidence count in the ensemble of shuffles would serve as the predictor for the number of coincidences found in the original experiment. Clearly, application of trial shuffling implies that the processes generating the data are stationary across trials (i.e., rates are required to be stationary across trials).

Using the tools developed for our model of nonstationarity across trials, we can now investigate the properties of the shuffle predictor under conditions of nonstationarity. In our language, a particular experimental data set represents a microstate of the system. Application of trial shuffling destroys the original rate constellation, and another rate constellation (microstate) is created. Under shuffling, the system may change from the macrostate $[n_1, n_2, n_3, n_4]$ describing the original data set to another macrostate $[n'_1, n'_2, n'_3, n'_4]$. However, not all possible macrostates can be reached since the number of trials neuron 1 and neuron 2 are in one of the two rate levels λ_1, and λ_2 is predefined by the original data set. We have the following constraints connecting the set of reachable macrostates:

$$\begin{aligned} n_1 + n_3 &= n'_1 + n'_3 \\ n_1 + n_4 &= n'_1 + n'_4 \ , \end{aligned} \tag{23}$$

where the first equation conserves the number of trials neuron 1 is subject to rate λ_1 and the second equation does the same for neuron 2. The complementary conditions for λ_2 are automatically fulfilled. In the cases where at least one neuron is stationary (e.g., $n_1 + n_3 = M$), all shuffles remain in the same macrostate. In each reachable macrostate all possible microstates can be realized. Thus, for the full shuffle the probability distribution for reaching other macrostates can be constructed by computing the number of microstates for the individual reachable macrostates (Eq. 12), and normalization by the total number of reachable microstates (Fig. 11a). The expected coincidence count (Eq. 13) in a macrostate $[n'_1, n'_2, n'_3, n'_4]$ is generally different from the one for the original macrostate. However, due to Eq. 24, the expected coincidence count based on rate averages $\bar{n}^{\text{pred}}$ (Fig. 11a) is the same for all shuffles

$$\begin{aligned} \bar{n}^{\text{pred}} &= \mathcal{T}M \cdot (p_1 \cdot (n_1 + n_3) + p_2 \cdot (n_2 + n_4)) \\ &\quad \times (p_1 \cdot (n_1 + n_4) + p_2 \cdot (n_2 + n_3)) \\ &= \mathcal{T}M \cdot ((p_1 \cdot (n'_1 + n'_3) + p_2 \cdot (n'_2 + n'_4)) \\ &\quad \times (p_1 \cdot (n'_1 + n'_4) + p_2 \cdot (n'_2 + n'_3)) \ . \end{aligned} \tag{24}$$

It is a general result that in the case of full shuffling, the average n^{pred} equals $\bar{n}^{\text{pred}}$ (see Palm et al. 1988; Aertsen et al. 1989). Figure 11a illustrates the situation for a simulated experiment. The original data set may represent a microstate in any of the reachable macrostates. The distribution of n^{pred} is independent of the microstate from which the shuffling starts. Thus, if the covariation of spike rate in the experimental data set is high (Sect. 4.5), shuffling tends to destroy this coherence. In the example this corresponds to the case where shuffling starts in the macrostate with the largest n^{pred}, and shuffling predominantly transfers the system into the remaining three macrostates with considerably lower coincidence counts. Compared to the coincidence counts in the shuffle, the count of the experimental data set appears to be exceptionally large, which may lead to wrong interpretation of the data. Conversely, if the experimental data represent a microstate with small covariation, shuffling predominantly causes transitions into macrostates with larger rate coherence and larger coincidence counts. In the example, this corresponds to the situation where shuffling starts in the macrostate with lowest n^{pred}. Here, it appears that compared to the n^{pred} of the shuffles, the experimental data set has a lack of coincidences (see Grün et al. 2002a for a discussion of "false negatives").

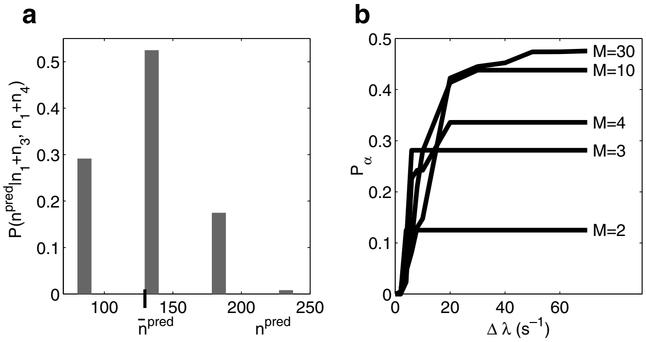


Fig. 11. Effect of nonstationarity across trials on the interpretation of individual experimental data sets. **a** Effects on the shuffle predictor for a particular experimental outcome (microstate) of a system with parameters ($\Delta\lambda = 70\,\text{s}^{-1}$, $M = 10$, $\mathcal{T} = 1000\,\text{ms}$, $(\lambda_1 + \lambda_2)/2 = 50\,\text{s}^{-1}$). In trial shuffling, the number of trials where neuron 1 is driven by rate λ_1, $n_1 + n_3$ ($n_1 + n_4$ for neuron 2) remains invariant (here 7); correspondingly, for λ_2 the invariant is $n_2 + n_4$ ($n_2 + n_3$ neuron 2, here $M - 7 = 3$). The constraints reduce the macrostates accessible by trial shuffling to 4 (bars) out of 286. The four macrostates (from left to right) are populated by 4200, 7560, 2520, and 120 microstates with respective n^{pred} 85.5, 134.5, 183.5, and 232.5 (horizontal). Normalization yields the probability (vertical) to reach a specific macrostate in a random shuffle. The average over all possible shuffles (full shuffling) leads to an expected number of coincidences of 129.6 (shuffle predictor corresponding to $\bar{n}^{\text{pred}}$), associating the four macrostates with false positive levels f_α of 0.0000, 0.1506, 0.9976, and 1.0000 (left to right). **b** Probability P_α (vertical) to perform an experiment for which the analysis reports false positives in a fraction of $f_\alpha > \alpha$ of the cases as a function of $\Delta\lambda$ (horizontal, steps of $5\,\text{s}^{-1}$, $(\lambda_1 + \lambda_2)/2 = 50\,\text{s}^{-1}$, $q = 0.5$, $M\mathcal{T} = 10\,\text{s}$). Curves represent results for a constant number of trials (M, as labeled). P_α is the fraction of microstates with $f_\alpha > \alpha$ at a given $\Delta\lambda$. Curves saturate because above a certain $\Delta\lambda$ macrostates differentiate into a class with $f_\alpha \leq \alpha$ and a class with $f_\alpha > \alpha$. For $M = 2$, P_α saturates at 2/16 (cf. Fig. 4)

5.2.2 Probability of false positives. We have analyzed the dependence of the fraction of false positives in the ensemble average $\bar{F}_\alpha$ (Fig. 8) on the parameters characterizing the nonstationarity $(\Delta\lambda, q)$. More importantly, for a given experimental data set a class of rate constellations was identified (Fig. 9) generating a large fraction of false positives f_α. However, it can be argued that specific rate constellations generally have only a low probability of occurrence. Thus, for the interpretation of an individual experiment the experimenter may be interested in aspects of the distribution of f_α. An example for such a measure is the probability P_α of obtaining a "problematic" experimental outcome, where f_α exceeds the adjusted level α. A second, related, measure is the expected f_α for experiments where α is exceeded. Having the distribution of macrostates $P_{[n_1,n_2,n_3,n_4]}$ (Eq. 16) available, such questions can be answered. f_α is determined by the macrostate $[n_1, n_2, n_3, n_4]$, therefore $P_{[n_1,n_2,n_3,n_4]}$ directly specifies the distribution of $f_{\alpha[n_1,n_2,n_3,n_4]}$, and we can write

$$P_\alpha = \sum_{\substack{[n_1,n_2,n_3,n_4] \\ f_\alpha > \alpha}} P_{[n_1,n_2,n_3,n_4]} \ . \tag{25}$$

Fig. 11b shows the dependence of P_α on $\Delta\lambda$ for experiments with different numbers of trials. The probability of obtaining false positives increases with $\Delta\lambda$. P_α saturates for large $\Delta\lambda$, with the saturation level increasing with the number of trials. The increase in P_α occurs because in macrostates where a rate step $[n_1, n_2]$ is combined with anticorrelated rates $[n_3, n_4]$, a minimal $\Delta\lambda$ is required to achieve dominance of the rate step. $\Delta\lambda$ differentiates the macrostates. At a certain $\Delta\lambda$ of an M-trial system, all macrostates that potentially create $f_\alpha > \alpha$ have separated and saturation is reached. With decreasing M, the number of macrostates where rate steps are combined with anticorrelated rates declines, and saturation is reached in fewer steps and at lower $\Delta\lambda$ (e.g., for $M = 2$ saturation is reached immediately for $\Delta\lambda > 0$). At small M, the $4 \cdot 2^M - 4$ microstates where at least one of neurons is stationary constitute a considerable fraction of the total number of microstates, and control P_α. For large M, the probability P_α of performing an experiment in which the probability of observing false positives is larger than the adjusted level α approaches 0.5.

5.3 Application to experimental data

Figure 12 illustrates the results of the analysis of an experimental data set from the primary motor cortex of the awake behaving monkey involved in a motor preparation task (Riehle et al. 2000). Data are analyzed for excess spike coincidences in a time-resolved manner (Grün et al. 2002b; Grün et al. 1999), correcting for the temporal nonstationarity of spike rate shown in Fig. 12a. The coincidence count in the data modulates in time and in certain periods deviates from the predicted coincidence count (Fig. 12c). Based on the cross-trial average of spike rate, the coincidence count reaches the significance threshold (Fig. 12d) in some of the instances where the empirical coincidence count exceeds the predicted one. Figure 12b shows the correlation coefficient R of the spike counts across trials as a function of the position of the analysis window. In terms of our model, the correlation coefficient of spike counts is a measure for the covariation (coherence) of the spike rates of the two neurons. Clearly, the correlation coefficient is time dependent. During the waiting period, R initially increases, reaches a maximum, subsequently decreases assuming negative values, and, after passing a minimum, returns to positive values. Thus the question arises of whether the observed deviations of the coincidence count from expectation are caused by the nonstationarity of spike count across trials. In order to check for this option the two methods for coping with nonstationarity across trials discussed in Sect. 5.1 are applied. The resulting time courses of the significance measure differ from the original analysis in detail. However, the overall time course is preserved (Fig. 12d,e). At $t \approx 250$ ms both methods report minor reduction in significance. This is the location where the correlation coefficient of spike counts (Fig. 12c) reaches a maximum. Based on these additional tests, we conclude that the excess coincidences classified as significant are probably due to the fine temporal coordination of spikes and are not an artifact of the covariation of spike counts.

The concept of the shuffle predictor suggests a useful reverse approach. Instead of comparing the experimental outcome to a distribution generated using many

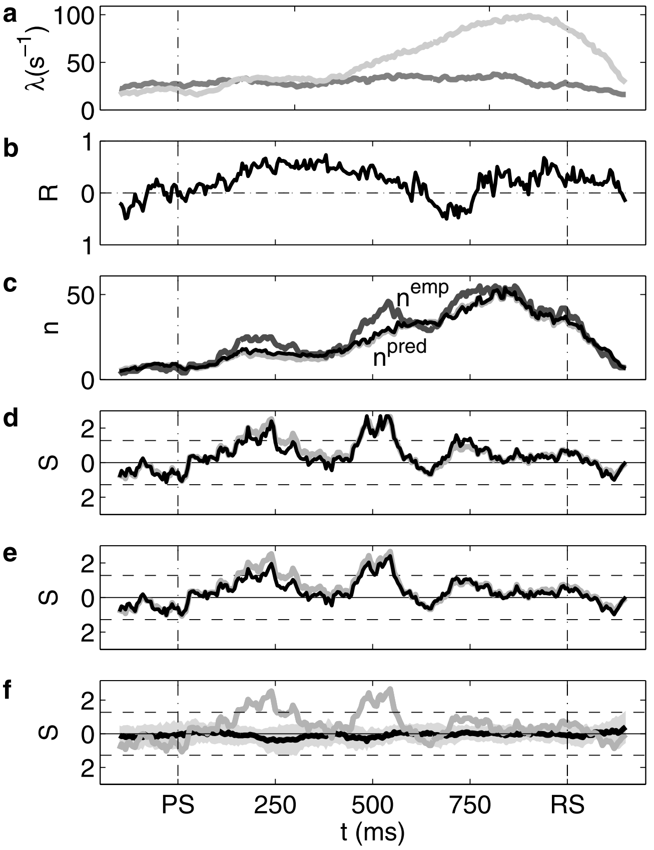


Fig. 12. Example of experimental data analyzed for unitary events considering covariation of spike count. Analysis of two simultaneously recorded neurons (21 trials) from primary motor cortex of the awake behaving monkey. At time (horizontal) PS the monkey received a preparatory signal indicating it was to wait until a go-signal (RS) instructed it to perform an arm movement to a predefined direction (detailed description of task and preparation in Riehle et al. 2000). Calculations are independently performed in a window of 100 ms shifted along the temporal axis in steps of 5 ms (Grün et al. 2002b). Results are displayed at the temporal center of window. **a** Spike rate averaged over trials (PSTH, neuron 1: *dark gray*, neuron 2: *light gray*), computation in sliding window equivalent to smoothing by a boxcar of 100 ms. **b** Time-resolved covariation R of spike count of neurons 1 and 2. The correlation coefficient of the spike counts of neurons 1 and 2 is taken as a measure of the covariation of spike counts over trials. **c** Number of coincidences. The *dark gray curve* shows the time course of the empirical number of coincidences n^{emp} (allowed temporal jitter 5 ms, Grün et al. 1999). The *light gray curve* is the predicted number of coincidences based on the rate averages (**a**) across trials ($\bar{n}^{\text{pred}}$). The *black thin curve* is the predictor based on trial-by-trial rate estimates (n^{pred}). **d-f** Different tests for the significance of the empirical coincidence count (levels $\alpha = 0.05$ and $1 - \alpha$, *dashed*). Significance is expressed in terms of the joint-surprise S, a logarithmic transformation of the cumulative probability of finding the observed number of coincidences or more (Grün et al. 2002a). S based on the comparison of n^{emp} and $\bar{n}^{\text{pred}}$ ("standard" unitary event analysis) is shown in *medium gray* in all three panels (copied). **d** S (*Black thin line*) based on the distribution of coincidence counts constructed from surrogate data with spike count in each trial and for both neurons identical to the experimental data. **e** S (*Black thin line*) based on the Poisson distribution with parameter n^{pred}. **f** Significance of shuffled experimental data. Spike trains of noncorresponding trials (repeated wrap-around shifting of trials) are analyzed for the significance of the coincidence count based on a Poisson distribution with parameter $\bar{n}^{\text{pred}}$ (invariant). The average over 21 possible combinations (not full shuffling) is shown in *black*; *light gray band* depicts ± standard deviation

shuffles, one can also apply the standard significance test to the individual shuffles. If trial shuffling successfully destroys the effects, causing a surplus in coincidences, the mean of the significance measure should be 0 with the fraction of false positives corresponding to α. In this approach, the rate profiles of individual trials and other features locked to trial onset are conserved. Figure 12f illustrates the result of such an analysis. Clearly, it appears unlikely that features in the original data set are caused by a random alignment of spikes.

One should be reminded that trial shuffling assumes stationarity across trials. Shuffling destroys, as intended, possible fine temporal coordination of spiking activity. Simultaneously, however, the shuffling procedure destroys the rate constellations of the original trials. Thus, in the presence of rate covariation across trials, shuffling decorrelates the rate states (Sect. 5.2.1). Solely due to this fact, the value of the significance measure for shuffled data can fall below the value obtained for the original arrangement (Fig. 11a). The dip in the mean of the significance measure at $t = 250$ ms may indicate effects of nonstationarity across trials.

In summary, although we found indications for nonstationarity across trials and for covariation of rate, in this particular data set the results of an analysis assuming stationarity are hardly affected. Nonstationarity does not seem to be strong enough to induce false positives. Obviously, this statement is not true for arbitrary experimental data sets. The particular data set was classified as “stationary across trials” by visual inspection before unitary event analysis was applied. However, we suggest that controls for nonstationarity across trials and covariance of spike rate be performed on a routine basis to avoid misinterpretation of the data.

6 Conclusions

The effects of nonstationarity across trials on crosscorrelation-type measures of spiking activity are analyzed using a model-based approach. The model describes nonstationarity across trials by assuming that in each trial neuronal spiking activity results from one of two spike rates. The difference of the two spike rates and their relative probability of occurrence constitute the parameters of the model. The significance of the number of coincident spike times found in simultaneous recordings from two neurons is used as an example for a measure potentially perturbed by nonstationarity. The effects of nonstationarity across trials are evaluated in terms of the occurrence of false positives, the fraction of experiments in which a significant outcome is falsely reported. Comparison of the analytical treatment with computer simulations shows that the system’s properties can be understood on the basis of the mean coincident count. A framework that assigns to each possible experimental outcome a collective “system state” enables the analysis of situations with a realistic number of trials. This scheme permits a classification of potential experimental outcomes uncovering efficient generators of false positives. Experiments showing coherent rate variations of the two neurons with the larger proportion of the trials spent in the low-rate regime rather than in the high-rate regime are most effective. Note that nonstationarity across trials is a prerequisite for the occurrence of covariation of spike rate but not sufficient. The model is minimal in the sense that nonstationarity is caused by transitions between just two rate states. In more general, multiple-rate-state models (the number of rate states is not restricted by M), coherent rate steps from low to high rates are still the effective generators of false positives. However, because of the increased total number of system states, the probability of occurrence of specific states is generally decreased. Thus, the two-rate-state model is extreme in the sense that it exhibits with high probability the relevant system states of more general models.

It is common knowledge that the application of the shift predictor requires stationarity across trials. The framework developed for our model connects failure of the shuffle predictor with the notion of covariation of spike rate. The mechanism by which the shuffle predictor fails is the following: shuffling reaches a certain subset of all possible rate constellations. If the original data set exhibits high covariation of spike rate, shuffling decorrelates and the prediction falls below the correct one.

The model allows an analysis of the ensemble average over many experiments but also predictions for individual experimental outcomes. While the ensemble average exhibits a certain robustness with respect to nonstationarity, in an individual experimental outcome false positives can be generated with high probability. The probability that in an individual experimental outcome false positives are observed with a probability larger than the adjusted level α approaches 0.5.

The analysis based on surrogate data reduces the number of false positives to its expected level. In the simplest case, surrogates are constructed by randomizing the spike times in individual trials while keeping the number of spikes constant. The distribution of coincidence counts on the basis of the spike counts in individual trials is known, however hard to evaluate by direct methods. Using the Poisson approximation, the distribution of coincidence is obtained by calculating the expected coincidence count on a trial-by-trial basis, using the spike counts or estimations of the instantaneous rates (Pauluis and Baker 2000; Grün et al. 2002b). Thus, when analyzing experimental data for spike coincidences, possible nonstationarity across trials combined with rate (spike count) covariations should be monitored. The main effects are compensated by improving the estimate of the mean coincidence count. More general approaches are currently being explored (Grün et al. 2001a; Pipa and Grün 2002).

Acknowledgements. We thank George Gerstein, Günther Palm, Gordon Pipa, Stefan Rotter, and Gaby Schneider for stimulating discussions and Denny Fliegner for advice on computational methods. Supported in part by the Volkswagen Foundation (SG). Part of the computations were performed using the parallel computing facilities at the Max-Planck-Institut für Strömungsforschung.

Appendices

A List of symbols used

$\mathcal{T}$:	temporal duration of the observation interval, $[\mathcal{T}] =$ unit of time
h:	time resolution, $[h] =$ unit of time
T:	temporal duration of observation interval in units of h, $[T] = 1$
M:	number of trials
λ_1, λ_2:	rate levels, $[\lambda_1] = [\lambda_2] = 1/$unit of time
$\Delta\lambda$:	rate level difference, $[\Delta\lambda] = 1/$unit of time
p_1, p_2:	spike probability in rate states λ_1, λ_2, respectively
$q, (1-q)$:	occupation probabilities of rate state λ_1, λ_2, respectively
$\bar{n}^{\mathrm{pred}}$:	expected number of coincidences based on averaged rates
n^{pred}:	expected number of coincidences based on the sum of the single trial expectancies
s_i:	vector representing single trial rate constellation i
n_i:	number of trials with s_i
$[n_1, n_2, n_3, n_4]$:	number of occurrences of rate states $[s_1, s_2, s_3, s_4]$ defining a macrostate
$K_{[n_1,n_2,n_3,n_4]}$:	number of microstates in macrostate $[n_1, n_2, n_3, n_4]$
$Q_{[n_1,n_2,n_3,n_4]}$:	occupation probability of macrostate $[n_1, n_2, n_3, n_4]$
$P_{[n_1,n_2,n_3,n_4]}$:	occurrence probability of macrostate $[n_1, n_2, n_3, n_4]$
f_α:	fraction false positives of a particular microstate
$F_{\alpha[n_1,n_2,n_3,n_4]}$:	fraction false positives of macrostate $[n_1, n_2, n_3, n_4]$
$\bar{F}_\alpha$:	fraction false positives in the ensemble average
S:	joint-surprise
α:	significance level

B Distribution of coincidence counts in multiple trials

Given the spike counts k_1^i, k_2^i of neuron 1 and 2 in a single trial i, the distribution of coincidence counts follows a hypergeometrical distribution (Palm et al. 1988; Gütig et al. 2002)

$$\mathcal{H}(n^i|k_1^i, k_2^i) = \frac{\binom{k_1^i}{n^i} \cdot \binom{T-k_1^i}{k_2^i - n^i}}{\binom{T}{k_2^i}} \quad . \tag{26}$$

Thus, for the distribution of coincidence counts in M trials we have

$$P(n) = \sum_{\substack{[n^1,\dots,n^M] \\ \sum n^i = n}} \prod_{i=1}^{M} \mathcal{H}(n^i|k_1^i, k_2^i) \quad , \tag{27}$$

where the sum extends over all compositions $[n^1, \dots, n^M]$ of n with M positions (Nijnhuis and Wilf 1978). The hypergeometrical distribution (Eq. 26) can be approximated by a Poisson distribution $\mathcal{P}$ with parameter $T \cdot \frac{k_1^i}{T} \cdot \frac{k_2^i}{T} = \frac{k_1^i k_2^i}{T}$:

$$\mathcal{P}\left(n^i \middle| \frac{k_1^i k_2^i}{T}\right) = \frac{\left(\frac{k_1^i k_2^i}{T}\right)^{n^i}}{n^i!} \cdot \exp - \frac{k_1^i k_2^i}{T} \quad . \tag{28}$$

Note that Eqs. 26 and 28 have the same mean $\frac{k_1^i k_2^i}{T}$. Even if the spike rates of the generating processes were known, one would use the spike counts to determine the appropriate parameter of Eq. 28. Whether $k_j^i/\mathcal{T}$ is a good estimator for $\lambda_{j,i}$ is of no importance in this view.

For the Poisson distribution we can exploit the fact that the sum of two Poisson distributed random variables is again Poisson distributed. Thus, for the Poisson approximation of Eq. 27 we have

$$P(n) \approx \sum_{\substack{[n^1,\dots,n^M] \\ \sum n^i = n}} \prod_{i=1}^{M} \mathcal{P}\left(n^i, \frac{k_1^i k_2^i}{T}\right) = \mathcal{P}\left(n \middle| \frac{1}{T} \sum_{i=1}^{M} k_1^i k_2^i\right) \quad . \tag{29}$$

Because of the importance of the above property of the Poisson distribution, we restate the textbook proof (e.g., Krengel 2000). The expected number of coincidences in a single trial i is abbreviated by $m_i = k_1^i k_2^i / T$, and it is sufficient to show the property for two random variables (here, $i = 1, 2$).

$$P(n) = \sum_{j=0}^{n} \mathcal{P}(j|m_1) \cdot \mathcal{P}(n-j|m_2) \tag{30}$$

$$= \sum_{j=0}^{n} \frac{m_1^j}{j!} \exp(-m_1) \cdot \frac{m_2^{n-j}}{(n-j)!} \exp(-m_2) \tag{31}$$

multiplying by unity

$$= \exp -(m_1 + m_2) \cdot \sum_{j=0}^{n} m_1^j m_2^{n-j} \cdot \frac{1}{(n-j)!j!} \cdot \frac{n!}{n!} \tag{32}$$

$$= \exp -(m_1 + m_2) \cdot \frac{1}{n!} \sum_{j=0}^{n} \binom{n}{j} m_1^j m_2^{n-j} \tag{33}$$

using the binomial theorem

$$= \exp -(m_1 + m_2) \cdot \frac{1}{n!} (m_1 + m_2)^n \tag{34}$$

$$= \mathcal{P}(n|m_1 + m_2) \quad \square \tag{35}$$

References

Abeles M (1982) Local Cortical Circuits: an electrophysiological study. Studies of brain function. Springer, Berlin Heidelberg New York

Abeles M, Gerstein GL (1988) Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. J Neurophysiol 60: 909–924

Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of 'effective connectivity.' J Neurophysiol 61: 900–917

Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science 273: 1868–1871

Baker SN, Gerstein GL (2000) Improvements to the sensitivity of gravitational clustering for multiple neuron recordings. Neural Comput 12: 2597–2620

Baker SN, Gerstein GL (2001) Determination of response latency and its application to normalization of cross-correlation measures. Neural Comput 13: 1351–1377

Baker SN, Lemon R (2000) Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance level. J Neurophysiol 84: 1770–1780

Ben-Shaul Y, Bergman H, Ritov Y, Abeles M (2001) Trial to trial variability in either stimulus or action causes apparent correlation and synchrony in neuronal activity. J Neurosci Meth 111: 99–110

Brody CD (1999a) Correlations without synchrony. Neural Comput 11: 1537–1551

Brody CD (1999b) Disambiguating different covariation types. Neural Comp 11: 1527–1535

Coburn KL, Ashford JW, Fuster JM (1990) Visual response latencies in temporal lobe structures as a function of stimulus information load. Behav Neurosci 104: 62–73

Eggermont JJ (1990) The correlative brain, vol 16. Studies of brain function. Springer, Berlin Heidelberg New York

Fries P, Neuenschwander S, Engel AK, Goebel R, Singer W (2001) Rapid feature selective neuronal synchronization through correlated latency shifting. Nat Neurosci 4: 194–200

Friston KJ (1995) Neuronal transients. Proc R Soc Lond B 261: 401–405

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in the striate cortex. J Neurophysiol 76: 1356–1360

Gerstein GL, Bedenbaugh P, Aertsen A (1989) Neuronal assemblies. IEEE Trans Biomed Eng 36: 4–14

Glaser EM, Ruchkin DS (1976) Principles of neurobiological signal analysis. Academic, New York

Grün S, Diesmann M, Aertsen A (2002a) 'Unitary events' in multiple single-neuron activity. I. Detection and significance. Neural Comput 14: 43–80

Grün S, Diesmann M, Aertsen A (2002b) 'Unitary events' in multiple single-neuron activity. II. Non-stationary data. Neural Comput 14: 81–119

Grün S, Diesmann M, Grammont F, Riehle A, Aertsen A (1999) Detecting unitary events without discretization of time. J Neurosci Meth 94: 67–79

Grün S, Galuske R, Singer W, Schmidt K (2001a) Analysis of parallel recordings from cat visual cortex for higher order coincidences. Abstr Soc Neurosci 27(2): 82–159

Grün S, Riehle A, Diesmann M (2001b) Unitary events analysis on the basis of surrogates. In: Elsner N, Kreutzberg G (eds) Proceedings of the 28th Göttingen Neurobiology Conference, Thieme, Stuttgart, p 251

Gur M, Beylin A, Snodderly M (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. J Neurosci 17: 2914–2920

Gütig R, Aertsen A, Rotter S (2002) Statistical significance of coincident spikes: count-based versus rate-based statistics. Neural Comput 14: 121–253

Krengel U (2000) Einführung in die Wahrscheinlichkeitstheorie und Statistik. 5th edn, Vieweg, Braunschweig

MacLeod K, Laurent G (1996) Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. Science 274: 976–979

Martignon L, von Hasseln H, Grün S, Aertsen A, Palm G (1995) Detecting higher-order interactions among the spiking events in a group of neurons. Biol Cybern 73: 69–81

Miller J, Riehle A, Requin J (1992) Effects of preliminary perceptual output on neuronal acitvity of the primary motor cortex. J Exp Psych 18: 1121–1138

Nawrot MP, Aertsen A, Rotter S (2002) Elimination of variability due to response latencies in neuronal spike trains. Biol Cybern (this issue)

Nawrot MP, Rodriguez V, Heck D, Riehle A, Aertsen A, Rotter S (2001) Trial-by-trial variability of spike trains in vivo and in vitro. Abstr Soc Neurosci 27: 6–49

Nijnhuis A, Wilf HS (1978) Combinatorical algorithms: for computers and Calculators, 2nd edn. Academic, New York

Oram MW, Hatsopoulos NG, Richmond BJ, Donoghue JP (2001) Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures. J Neurophysiol 86: 1700–1716

Oram MW, Wiener MC, Lestienne R, Richmond BJ (1999) Stochastic nature of precisely timed spike patterns in visual system neuronal responses. J Neurophysiol 81: 3021–3033

Palm G, Aertsen A, Gerstein GL (1988) On the significance of correlations among neuronal spike trains. Biol Cybern 59: 1–11

Pauluis Q, Baker SN (2000) An accurate measure of the instantaneous discharge probability, with application to unitary joint-event analysis. Neural Comput 12: 647–669

Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. Biophys J 7: 419–440

Petkovsek M, Wilf HS, Zeilberger D (1996) A = B. Peters, Wellesley, MA

Pipa G, Grün S (2003) Non-parametric significance estimation of joint-spike events by shuffling and resampling. Neurocomputing (in press)

Prut Y, Vaadia E, Bergman H, Haalman I, Slovin H, Abeles M (1998) Spatiotemporal structure of cortical activity: properties and behavioral relevance. J Neurophysiol 79: 2857–2874

Reich DS, Mechler F, Purpura KP, Victor JD (2000) Interspike intervals, receptive fields, and information encoding in primary visual cortex. J Neurosci 20: 1964–1974

Richmond BJ, Hertz JA, Gawne TJ (1999) The relation between v1 neuronal responses and eye movement-like stimulus representations. Neurocomputing 26: 247–254

Richmond BJ, Optican LM, Podell M, Spitzer H (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. J Neurophysiol 57: 132–146

Riehle A, Grammont F, Diesmann M, Grün S (2000) Dynamical changes and temporal precision of synchronized spiking activity in motor cortex during movement preparation. J Physiol Paris 94: 569–582

Riehle A, Grün S, Diesmann M, Aertsen A (1997) Spike synchronization and rate modulation differentially involved in motor cortical function. Science 278: 1950–1953

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) Spikes: exploring the neural code. MIT Press, Cambridge, MA

Roelfsema PR, Engel AK, König P, Singer W (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature 385: 157–161

Roy A, Steinmetz PN, Niebur E (2000) Rate limitations of unitary event analysis. Neural Comput 12: 2063–2082

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18: 3870–3896

Singer W (1993) Synchronization of cortical activity and its putative role in information processing and learning. Annu Rev Physiol 55: 349–374

Singer W, Gray C (1995) Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18: 555–586

Vaadia E, Aertsen A, Nelken I (1995a) 'Dynamics of neuronal interactions' cannot be explained by 'neuronal transients'. Proc R Soc Lond B 261: 407–410

Vaadia E, Haalman I, Abeles M, Bergman H, Prut Y, Slovin H, Aertsen A (1995b) Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. Nature 373: 515–518

Vaadia E, Kurata K, Wise SP (1988) Neuronal activity preceding directional and nondirectional cues in the premotor cortex of rhesus monkeys. Somatosens Mot Res 6: 207–230