LETTER Communicated by George Gerstein

Statistical Significance of Coincident Spikes: Count-Based Versus Rate-Based Statistics

Robert Gütig
guetig@biologie.uni-freiburg.de
Ad Aertsen
aertsen@biologie.uni-freiburg.de
Stefan Rotter
rotter@biologie.uni-freiburg.de
Neurobiology and Biophysics, Institute of Biology III, Albert–Ludwigs–University, 79104 Freiburg, Germany

Inspired by different conceptualizations of temporal neural coding schemes, there has been recent interest in the search for signs of precisely synchronized neural activity in the cortex. One method developed for this task is unitary-event analysis. This method tests multiple single-neuron recordings for short epochs with significantly more coincident spikes than expected from independent neurons. We reformulated the statistical test underlying this method using a coincidence count distribution based on empirical spike counts rather than on estimated spike probabilities. In the case of two neurons, the requirement of stationary firing rates, originally imposed on both neurons, can be relaxed; only the rate of one neuron needs to be stationary, while the other may follow an arbitrary time course. By analytical calculations of the test power curves of the original and the revised method, we demonstrate that the test power can be increased by a factor of two or more in physiologically realistic regimes. In addition, we analyze the effective significance levels of both methods for neural firing rates ranging between 0.2 Hz and 30 Hz.

1 Introduction

Thanks to advances in neurophysiological recording technology, it is now feasible to experimentally test biological hypotheses about cortical information processing and neuronal cooperativity on the basis of multiple single-neuron recordings (Aertsen, Bonhoeffer, & Krüger, 1987; Nicolelis, 1998). However, due to the stochastic appearance of neural response patterns in the cortex (Palm, Aertsen, & Gerstein, 1988), the neurobiological concepts in question must be translated into precise statistical hypotheses to be verified by specifically designed statistical tests.

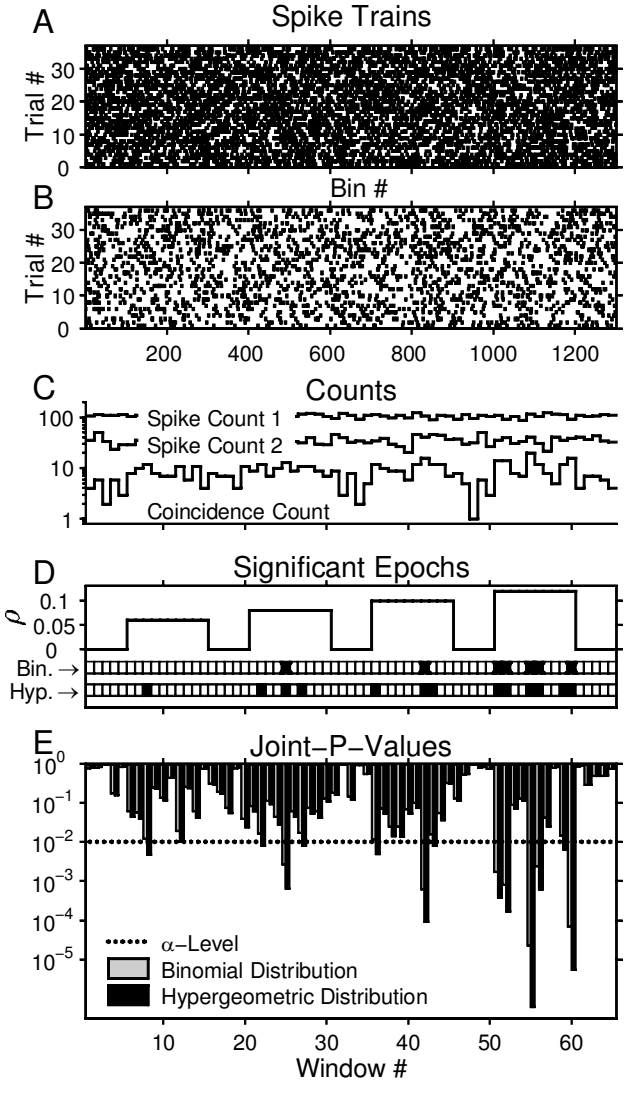
Inspired by a number of different conceptualizations of temporal neural coding schemes, such as correlational cell assemblies (Aertsen & Gerstein, 1991; Gerstein, Bedenbaugh, & Aertsen, 1989; von der Malsburg, 1981), coherent oscillations (Singer, 1993), and precise firing patterns (Abeles, 1982,

1991), there has been particular emphasis on the search for synchronized neural activity (Abeles & Gerstein, 1988; Abeles, Bergman, Margalit, & Vaadia, 1993; Kreiter & Singer, 1996; Prut et al., 1998; Singer, 1999). One of the methods developed for this task is unitary-event analysis (Grün, 1996; Grün, Diesmann, Grammont, Riehle, & Aertsen, 1999; Grün, Diesmann, & Aertsen, in press-a; Grün, Diesmann, & Aertsen, in press-b; Riehle, Grün, Diesmann, & Aertsen, 1997). This method searches recordings from multiple single neurons for epochs with distinctly more (near-)coincident spikes than expected from independent neurons obeying Poissonian spike statistics. The core of unitary-event analysis consists of computing the probabilities (joint-p-values) for the occurrence of a given minimum number of coincident spikes in short time segments, under the null hypothesis of independence. Segments with a joint-p-value below a fixed level of significance α are identified as significant epochs where the null hypothesis is rejected.

Here we demonstrate that by revising the original testing procedure, more specifically, by implementing a different coincidence count distribution, we can substantially increase the power of the method. Figure 1 compares the behavior of the original and the modified versions of the method on an empirically motivated, simulated data set of two neurons, stretching over epochs of correlated activity. Figure 1D depicts all epochs marked as significant by either of the two methods. The plot shows that the original version (Bin) misses epochs of synchronous neural activity that are detected by the modified version (Hyp).

The primary goal of this article is to investigate and compare two central properties—the power and the effective significance level—of the statistical test underlying the original and the revised versions of unitary-event analysis. Our investigation focuses on the identification of significant epochs; we will not treat statistical problems of interdependent testing, arising when

Figure 1: *Facing page*. Comparison of the original and the modified version of the statistical test underlying unitary-event analysis applied to simulated data. Matching the data of Riehle, Grün, Diesmann, and Aertsen (1997), we simulated 36 trials of two dependent neurons over 1300 bins (each of 5 ms duration) with the spike event probabilities $p_1 = 0.15$ and $p_2 = 0.05$, corresponding to the mean firing rates of 30 Hz and 10 Hz. (A, B) Raster plots of neurons 1 and 2. (C) The spike event and coincidence counts for 65 nonoverlapping analysis windows of 100 ms duration ($n = 36 \cdot \frac{100\,\text{ms}}{5\,\text{ms}} = 720$). (D) The upper part shows the correlation of spike counts ρ of the two neurons that was controlled through the stochastic model underlying the simulation of the data (see section 2 for details). The lower part depicts the epochs that were marked as significant by the original test (upper row) and the modified test (lower row) at a significance level of $\alpha = 0.01$ (dotted line in *E*). (E) The corresponding joint-p-values for both versions of the test, calculated in each of the analysis windows.



analysis windows overlap. Preliminary results have been presented in abstract form (Gütig, Rotter, Grün, & Aertsen, 2000).

2 Stochastic Model

First, we sketch the mathematical framework underlying our statistical assessment of the significance of coincident spiking of two neurons. Based on

this framework, in section 3 we summarize the statistical test underlying the original unitary-event analysis (Grün, 1996) and describe our revised version of the analysis method. One key element of our approach is to parameterize the stochastic dependence of the two neurons in terms of their spike correlation. This will be used in section 4 to calculate and compare the power of both statistical tests. The extension of this new approach to the case of three or more neurons is conceptually straightforward but numerically demanding. For clarity, most mathematical details are deferred to appendix A.

We assume the analysis to extend over a time window composed of n pairs of corresponding time bins (one for each neuron) of width Δt. Each bin can take either the value 1, denoting the observation of at least one spike in that time bin (we refer to this as a spike event), or the value 0, denoting the absence of spikes. Our treatment will refer only to joint observations of all n bins in the analysis window and disregard any information about the fine structure of the data in individual trials. Hence, an application of our results to pooled data from a multiple-trial design needs to assume stationarity across trials.

Two main assumptions will be made in the following:

A_1: The spike probabilities and the binwise spike correlations of the two neurons are the same for all n pairs of bins of the analysis window ("stationarity").

A_2: All bins, except for those in a pair, are stochastically independent ("serial independence").

Note that assumption A_1 is conceptually not essential to our framework, which is suited to treat a neural system with nonstationary spike probabilities. Doing so, however, would give rise to probability distributions with many parameters (cf. appendix C), greatly complicating the calculations. By contrast, assumption A_2 clearly imposes severe restrictions on the generality of the approach. Although it was not always explicitly stated, this assumption also underlies previous treatments of the topic (Grün, 1996; Roy, Steinmetz, & Niebur, 2000). The important task of overcoming the difficulties introduced by serially correlated spike trains is the subject of ongoing research.

As shown in appendix A, assumptions A_1 and A_2 allow us to completely characterize the probability space describing the realization of two neural spike trains by specifying the individual spike probabilities of the two neurons, p_1 and p_2, and their spike correlation ρ. We will abbreviate this parameter triplet throughout by $\xi := (p_1, p_2, \rho)$. The correlation ρ parameterizes the stochastic dependence of the two neurons, with the case of independently spiking neurons corresponding to $\rho = 0$. Note that for binary variables as used here, the notion of stochastic (in)dependence and the

Table 1: 2×2 Table of Counts for Analysis Windows Comprising n Bins.

	Spike from Neuron 1	No Spike from Neuron 1	
Spike from Neuron 2	k	$c_2 - k$	c_2
No Spike from Neuron 2	$c_1 - k$	$n - c_1 - c_2 + k$	$n - c_2$
	c_1	$n - c_1$	n

Note: The counts c_1 and c_2 denote the number of spike events from neurons 1 and 2, respectively, and k denotes the number of coincident spike events, in one particular observation

notion of correlation coincide. Thus, in our treatment of the test power in section 4, we will use nonvanishing values of ρ to quantify violations of the null hypothesis of independent firing. We will indicate the case of stochastic independence by writing ξ_0 instead of ξ.

The central statistics of the following treatment will be the individual spike counts, C_1 and C_2, and the coincidence count K, all derived from an observation of the two neurons in the analysis window. Since, according to our definition of spike events, each bin can hold at most one count, the total number of spike events observed from each neuron in a given analysis window is restricted to the integers between 0 and n. And the number of coincident spike events K cannot exceed either one of the corresponding spike counts C_1 and C_2. As shown in Table 1, the statistics C_1, C_2, and K give rise to a 2×2 table of counts for each realization of the analysis window. We note in passing that because of assumption A_1, the correlation ρ between spike events in coincident bins (see equation A.3) equals the correlation of the spike event counts of the two neurons, independently of the number of bins n.

Both tests for independence discussed in the following use the coincidence count K as their test statistic; they define statistical significance based on the number of coincident spikes observed within the analysis window. Thus, the probability distribution of this random variable, $P_\xi(K = k)$, directly enters the computation of the joint-p-values in each of the tests and therefore affects their statistical properties. The two methods, however, differ in the amount of information they draw from the data. To make this distinction clear, we discuss here both probability distributions for the case of independent neurons, that is, $P_{\xi_0}(K = k)$ where $\rho = 0$. The general forms of these distributions for arbitrary $\rho \in [-1, 1]$ and their derivations are given in appendix B. We emphasize that while the distributions for $\rho = 0$ suffice for the definition of the statistical tests in the next section, the calculation of test power in section 4 will rely on the general expressions from appendix B.

We first consider the situation where only the (constant) spike probabili-

ties p_1 and p_2 of both neurons are known. In particular, no further knowledge of the spike counts c_1 and c_2 is assumed. Then, based on assumptions A_1 and A_2, the probability distribution of K for independent neurons is given by the binomial distribution,

$$P_{\xi_0}(K = k) = \binom{n}{k} (p_1 p_2)^k (1 - p_1 p_2)^{(n-k)}. \tag{2.1}$$

Note that this distribution of coincidence counts was used in the original definition of unitary-event analysis (Grün, 1996), as well as in several recent contributions about the method (Grün et al., 1999; Roy et al., 2000).

In the second case, which is conceptually different, we consider the coincidence count distribution, making explicit use of the knowledge of the individual spike counts c_1 and c_2. In this case, we formulate the probability distribution conditional on the specific realization of the spike counts, that is, $P_\xi(K = k \mid c_1, c_2)$. Unlike the spike probabilities, these counts are readily accessible in any empirical spike train. It can be shown (cf. appendix B) that the conditional distribution of K for independent neurons is described by the hypergeometric distribution (cf. Palm et al., 1988):

$$P_{\xi_0}(K = k \mid C_1 = c_1, C_2 = c_2) = \frac{\binom{c_1}{k}\binom{n - c_1}{c_2 - k}}{\binom{n}{c_2}}. \tag{2.2}$$

Note that this distribution does not refer to the spike probabilities p_1 and p_2 anymore. Moreover, it can easily be verified that it is symmetric with respect to c_1 and c_2. The general form of the conditional distribution for $\rho \in [-1, 1]$ is much more complicated, though. Its explicit form is derived in appendix B.

3 Count-Based Versus Rate-Based Statistics

Typically, the test of a statistical hypothesis on the grounds of empirical data is based on a specially designed random variable, the so-called test statistic. In order to control the probability that the test fails, certain aspects of the distribution of the test statistic under the null hypothesis must be known. The probability that the null hypothesis is rejected, even if it is correct (type I error, or α-error), is usually fixed at some small value (e.g., 5 %). Essentially, this is achieved by adjusting the critical region of the test through the choice of an appropriate threshold of the test statistic (Mood, Graybill, & Boes, 1974). Similarly, the case where the null hypothesis is not rejected, even if it is false, is referred to as a type II error, or β-error. Given

a parametric model of deviations from the null hypothesis, which specifies the probability distribution of the test statistic under a given deviation, it is possible to calculate the β-error probability of the test. Based on this probability, one can directly assess the power of the test—the probability of correctly detecting a given violation of the null hypothesis.

Within the framework of the stochastic model defined in the previous section, including the assumptions A_1 and A_2, the null hypothesis underlying the detection of significant epochs comprises the following additional assumption for each analysis window:

H_0: The activity of the two neurons is stochastically independent, $\rho = 0$.

The original version of unitary-event analysis is based on the binomial coincidence count distribution (see equation 2.1). In this approach, empirical estimates for the parameters p_1 and p_2 on the basis of the spike counts c_1 and c_2,

$$\hat{p}_i = \frac{c_i}{n} \qquad (i = 1, 2), \tag{3.1}$$

are used to calculate the probability that $\tilde{k}$ or more coincident events are observed under the given conditions. This probability, here denoted by $J^{\mathrm{bin}}_{\xi_0}(\tilde{k}, c_1, c_2)$, is refered to as the joint-p-value (Grün, 1996)

$$J^{\mathrm{bin}}_{\xi_0}(\tilde{k}, c_1, c_2) := \sum_{k=\tilde{k}}^{n} \binom{n}{k} \left(\frac{c_1 c_2}{n^2}\right)^k \left(1 - \frac{c_1 c_2}{n^2}\right)^{(n-k)}. \tag{3.2}$$

We emphasize that this procedure uses the observed spike counts solely for the estimation of the spike probabilities. One effect of this is that the binomial distribution gives nonvanishing probabilities for the impossible outcome $k > \min(c_1, c_2)$. In addition, the binomial coincidence count distribution does not take into account the stochastic nature of the rate estimation procedure itself.

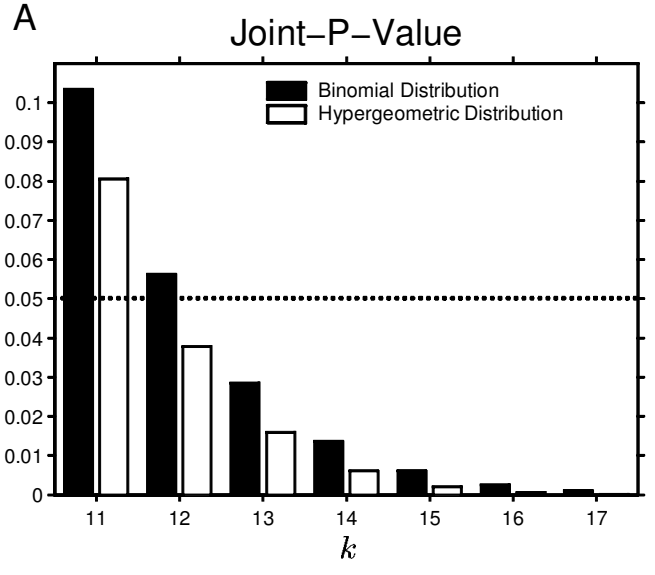
To make better use of the information contained in the spike counts, we propose to compute the joint-p-values from the conditional probabilities $P_{\xi_0}(k \mid c_1, c_2)$ instead. Given assumption H_0, together with the empirically accessible spike counts c_1 and c_2, these probabilities are determined by the hypergeometric distribution (see equation 2.2; Palm et al., 1988; Aertsen, Gerstein, Habib, & Palm, 1989; Lehmann, 1997). Thus, by using the empirical spike counts to specify the conditional distribution of coincidence counts (see equation 2.2), we can completely eliminate the rate estimation (see equation 3.1) from the testing procedure. Accordingly, the joint-p-value

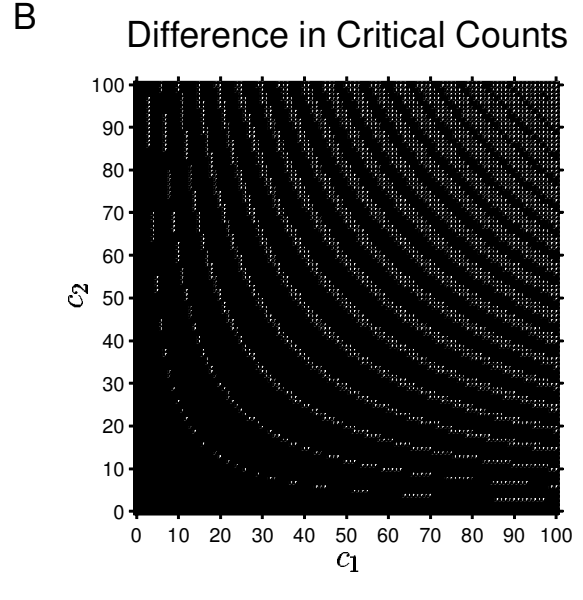
$J^{\mathrm{hyp}}_{\xi_0}(\tilde{k}, c_1, c_2)$ of an epoch with $\tilde{k}$ coincident spikes, based on the hypergeometric distribution, is given by

$$J^{\mathrm{hyp}}_{\xi_0}(\tilde{k}, c_1, c_2) := \sum_{k=\tilde{k}}^{\min(c_1,c_2)} \frac{\binom{c_1}{k}\binom{n-c_1}{c_2-k}}{\binom{n}{c_2}}. \tag{3.3}$$

To illustrate the difference between the two approaches, Figure 2A shows that for large enough values of the coincidence count k, the joint-p-values according to the binomial distribution exceed the corresponding probabilities based on the hypergeometric distribution. Thus, although the two distributions have equal mean, the binomial distribution overestimates the probability for the occurrence of large coincidence counts, for certain combinations of individual spike counts. The effect is that for the parameter values in Figure 2A, a statistical test using the hypergeometric distribution would classify an observation of $\tilde{k} \geq 12$ as significant ($\alpha = 0.05$; dotted line), whereas the corresponding test based on the binomial distribution would need at least $\tilde{k} = 13$ coincident spikes. Figure 2B shows the corresponding differences in these critical coincidence counts for all count combinations $c_1, c_2 \in \{0, \ldots, 100\}$. Light areas depict count combinations where the critical coincidence count for the binomial distribution exceeds the critical count of the modified test by one; for dark areas, the difference is zero. This example suggests that the use of the conditional coincidence count distribution may effectively lead to an increase in sensitivity of the test in certain parameter regimes. In fact, it is a result from mathematical statistics that the randomized version of the proposed test, called Fisher's exact test in the nonrandomized form presented here, is uniformly most powerful unbiased (Lehmann, 1997). A brief discussion of the randomized test is given in appendix E. Figure 2A also illustrates that in general, the joint-p-values of the critical coincidence counts do not equal the nominal significance threshold α. Moreover, for a given α-level, both tests will in general not operate on the same effective significance level.

Figure 2: *Facing page*. Comparison of binomial and hypergeometric distributions. (A) Joint-p-value: cumulative binomial and hypergeometric probability distributions for observing k or more coincidence counts for $n = 720$, $c_1 = 100$, and $c_2 = 51$. (B) Difference between the critical coincidence counts k_{crit} of both distributions for $n = 720$ and $\alpha = 0.05$. Light areas depict count combinations where the critical coincidence count of the binomial distribution exceeds the critical coincidence count of the hypergeometric distribution by one. For count combinations in the dark areas, the difference is zero.





Next to the issue of sensitivity of the two tests that will be addressed in the following section, we note one further important implication of using the hypergeometric coincidence count distribution. As shown in appendix C, the modified version of the test does not require that both neurons have con-

stant firing rates. Rather, it suffices if the spike probability of only one of the two neurons (either one) remains constant throughout the analysis window. Since the original method had to assume stationarity of both neural firing rates, this relaxation of the stationarity requirement A_1 implies an important extension of the class of empirical data that qualify for the statistical analysis described here. Note, however, that this relaxation concerns only the inferential statistical testing of H_0 as implemented by our revised method. The following investigation of the statistical properties of the method when applied to neurons that violate the hypothesis of independence still relies on stationarity in both neurons.

4 Test Power

To obtain a thorough understanding of the properties of the statistical significance tests defined above, we calculated the power function (Mood et al., 1974) of the tests for both coincidence count distributions. Specifically, this will allow us to quantify the advantage of using the hypergeometric count distribution concerning its test performance. For a given set of alternative hypotheses, the power of the statistical tests investigated here is the probability of obtaining a coincident count that yields a significant finding when tested against the null hypothesis. Loosely speaking, the power of the test measures the probability that the test will detect a given violation of the null hypothesis. According to the stochastic model defined in section 2, we use ρ to parameterize the set of alternative hypotheses. We emphasize that the power curves we will compute only characterize the sensitivity of the methods regarding violations of the null hypothesis of independent firing (H_0). Our treatment does not include violations of the other assumptions A_1 and A_2. We also note that the power curves will be calculated with respect to identical nominal significance levels α. Hence, the resulting values describe the more common nonrandomized application of the methods, as discussed in recent contributions (Grün, 1996; Grün et al., 1999; Pauluis & Baker, 2000; Roy et al., 2000). The effective significance levels of the tests, however, generally differ, and, hence, differences in test power will depend on differences in effective significance levels.

We introduce the auxiliary functions $k^{\mathrm{crit}}_{\mathrm{bin}}$ and $k^{\mathrm{crit}}_{\mathrm{hyp}}$:

$$k^{\mathrm{bin}}_{\mathrm{crit}}(c_1, c_2) := \min\{k \in \mathbb{N}\colon J^{\mathrm{bin}}_{\xi_0}(k, c_1, c_2) \le \alpha\} \tag{4.1}$$

$$k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) := \min\{k \in \mathbb{N}\colon J^{\mathrm{hyp}}_{\xi_0}(k, c_1, c_2) \le \alpha\}. \tag{4.2}$$

For any given combination of spike event counts c_1 and c_2, these k-values give the minimum number of coincident spikes that leads to a rejection of the null hypothesis at the significance level α, when tested against the null hypothesis underlying the corresponding coincidence count distribution. The probability of rejecting the null hypothesis for given counts c_1 and c_2

and ξ is then determined by

$$P_\xi(K \geq k_{\text{crit}}(c_1, c_2) \mid C_1 = c_1, C_2 = c_2), \tag{4.3}$$

which can be straightforwardly computed from equation B.7. It is important to note that this rejection probability is a function of the individual spike counts c_1 and c_2. This means that it is not an intrinsic property of the test itself, but rather a property of the test in combination with a specific empirical observation. Thus, to calculate the power of a test, we need to calculate the expectation value of the rejection probability with respect to the joint count distribution (see equation B.5), which is also specified through the parameter ξ of an alternative hypothesis. Thus, for a given ξ, the power π_ξ of the test is given by the expectation value of the rejection probability for each of the two underlying coincidence count distributions:

$$\begin{aligned}
\pi_\xi^{\text{bin}} &= \mathrm{E}_\xi\left[P_\xi(K \geq k_{\text{crit}}^{\text{bin}}(c_1, c_2) \mid C_1 = c_1, C_2 = c_2)\right] \\
&= \sum_{c_1, c_2 = 0}^{n} P_\xi(C_1 = c_1, C_2 = c_2) \\
&\quad \times P_\xi(K \geq k_{\text{crit}}^{\text{bin}}(c_1, c_2) \mid C_1 = c_1, C_2 = c_2) \qquad (4.4) \\
\pi_\xi^{\text{hyp}} &= \mathrm{E}_\xi\left[P_\xi(K \geq k_{\text{crit}}^{\text{hyp}}(c_1, c_2) \mid C_1 = c_1, C_2 = c_2)\right] \\
&= \sum_{c_1, c_2 = 0}^{n} P_\xi(C_1 = c_1, C_2 = c_2) \\
&\quad \times P_\xi(K \geq k_{\text{crit}}^{\text{hyp}}(c_1, c_2) \mid C_1 = c_1, C_2 = c_2). \qquad (4.5)
\end{aligned}$$

We refer to appendix D for an account on the numerical evaluation of these probabilities.

4.1 Difference in Test Power. In this section, we investigate the differences in power of the rate-based versus the count-based tests for significance of coincident spike activity. Specifically, we will discuss how this difference depends on the significance level α and the number of bins in the analysis window n and how this difference is affected by the spike probabilities p_1 and p_2. Apart from comparing the performance of the two methods, our analysis is also valuable for applications to empirical data. For a given set of parameters, knowledge of the power function allows us to assess whether a planned analysis is feasible or, vice versa, it can be used to guide the design of experiments. As mentioned before, we will use the spike correlation ρ to parameterize deviations from the null hypothesis of independent firing. The value of ρ will be varied over a physiologically realistic regime, as reviewed by Abeles (1991; (see also Aertsen & Gerstein, 1985).

We start the comparison by calculating the power curves of both analysis methods in a parameter regime that matches the empirical findings published by Riehle et al. (1997). Accordingly, we let the analysis window cover $n = 720$ bins of 5 ms each (originally collected from 36 trials) and set the spike probabilities of the two neurons to $p_1 = 0.15$ and $p_2 = 0.05$, corresponding to mean firing rates of 30 Hz and 10 Hz, respectively. We computed the power curves for the significance levels $\alpha = 0.01$, $\alpha = 0.05$ (this value was used by Riehle et al., 1997), and $\alpha = 0.1$.

Figure 3 shows that the tests based on the hypergeometric coincidence count distribution clearly outperform the original tests in terms of their power. Considering a spike correlation of $\rho = 0.1$, we see from Figure 3B that for $\alpha = 0.01$, the chance of rejecting the (false) null hypothesis of independent neurons is increased by over 0.1 compared to the original method. This corresponds to a relative increase in power of about 50% (see Figure 3).

From Figures 3A and 3B we also see that the difference between the tests increases when they are chosen to operate at more conservative significance levels (i.e., lower α). The inset of Figure 3A shows that for large analysis windows (here $n = 720$), the power decreases smoothly as α is decreased. It is interesting to note that the relative difference in test power (see Figure 3C) monotonically increases (up to 150%) as ρ approaches zero. Thus, other than the difference in power, which follows a bell-shaped curve (see Figure 3B), the relative difference in power becomes maximal for $\rho = 0$, that is, for independent neurons. This means that, as already suggested by Figure 2, the original method yields less significant findings in a low-correlation regime. In other words, for $\rho \to 0$, it effectively operates at a lower probability of yielding a false-positive finding than the count-based test. This is important, since it implies that the neglect of count information leads to a conservative behavior of the original test as applied in the past. However, as we will see, this is not the case for small values of $n \approx 20$ (i.e., for narrow analysis windows and single trial applications) in connection with specific values of α.

While both tests, as expected, gain power with increasing n (see Figure 4A), there is no qualitative change in the difference in power for larger analysis windows ($n \in \{100, \ldots, 700\}$, data not shown). The count-based test outperforms the rate-based version with differences in power qualitatively corresponding to the curves shown in Figure 3B. However, because of the overall increase in power, the peaks of the difference in power curves move toward smaller correlations as n grows. Therefore, the revised test is the method of choice when searching weak correlations in larger analysis windows ($n \gtrsim 100$).

Another potentially interesting regime is given by small analysis windows ($n \approx 20$), as would be needed for time-resolved single-trial analysis. Both tests suffer a considerable loss of power with decreasing n (see Figure 4A). Moreover, for small n, both methods become dominated by

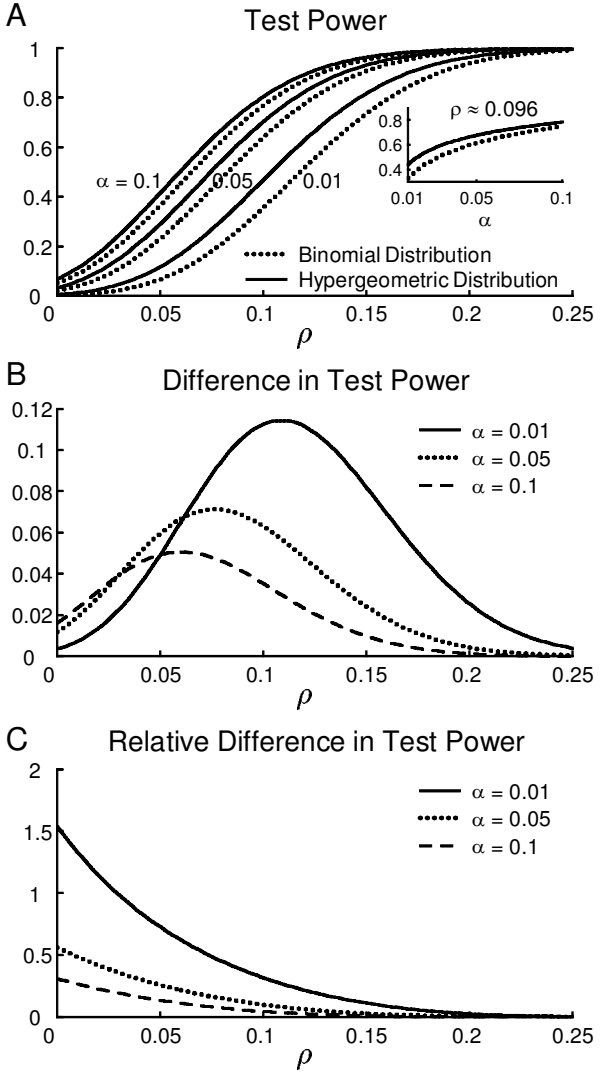


Figure 3: Dependence of both tests on nominal significance level. (A) Test power curves of the two methods for $n = 720$, $p_1 = 0.15$, $p_2 = 0.05$, and α-levels of 0.01, 0.05, and 0.1. The inset shows the test power for $\alpha \in \{0.01, 0.011, \ldots, 0.1\}$ at $\rho \approx 0.096$. (B) Differences in test power—the power values of the modified test minus the power values of the original version. (C) Difference in test power relative to the power of the original version of the test—the difference in power divided by the power of the original version. The analytical results shown in this figure were checked by computer simulations based on the Ran2 and the MT19937 random number generators (cf. Galassi et al., 1998).

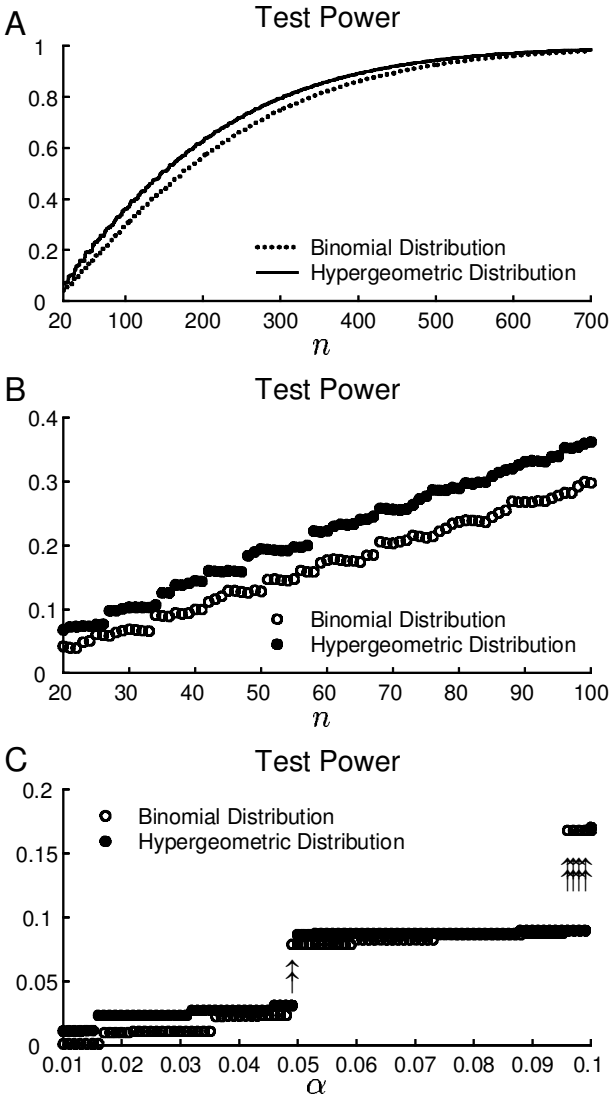


Figure 4: Dependence of both tests on analysis window size. (A, B) Test power curves for $\alpha = 0.05$, $p = p_1 = p_2 = 0.1$, $\rho \approx 0.19$, and $n \in \{20, 21, \ldots 700\}$. (C) Dependence of both tests on nominal significance level for $n = 20$, $p = p_1 = p_2 = 0.05$, $\rho \approx 0.26$, and $\alpha \in \{0.01, 0.011, \ldots, 0.1\}$. Arrows mark values of α where $\pi_\xi^{\text{bin}} > \pi_\xi^{\text{hyp}}$.

discreteness effects of the underlying spike statistics (see Figure 4B). Thus, the difference between the two methods becomes extremely sensitive to the nominal level of significance α, in a discontinuous way. For a small number of specific combinations of n and α (e.g. $n = 20$ and $\alpha = 0.049$), the power of the rate-based test can be even greater than the power of the count-based test (see Figure 4C). Thus, it would not, as was the case for large n, miss significant epochs but instead would indicate neural synchrony in cases that are not judged significant, once knowledge of the individual spike counts is incorporated into the testing procedure. Due to this behavior for small n, the original method should not be used in these parameter regimes.

Turning to the dependence of both tests on varying spike probabilities p_1 and p_2, we first let $p := p_1 = p_2$ and computed the power curves for p ranging from 0.01 to 0.15 (again, $n = 720$ and $\alpha = 0.05$). The results are shown in Figures 5A and 5B: as the power of both methods increases with growing spike probability p, the difference in their power also becomes larger. Thus, the advantage of the count-based method increases with higher neural firing rates. Also note that the power of both methods for low spike probabilities becomes small: the chance to detect a spike correlation of 0.1 is only around 30%. Finally, Figure 5C shows the difference in power for asymmetric spike probabilities $p_1 \neq p_2$ with constant product $p_1 p_2 = 0.0075$. Observe that the difference between the two tests grows as the asymmetry in the spike probabilities increases. This result indicates that the common assumption of equal spike probabilities underlying other investigations of the method (Grün et al., 1999; Roy et al., 2000) may shadow nontrivial properties of the tests, which are enhanced in regimes with different spike probabilities.

5 Effective Significance Level

Next to its power, the significance level of a statistical method is another important characteristic of its performance in practical applications. In this section, we analyze and compare the significance levels of the two tests. Generally, the significance level α of a statistical test denotes the probability that it leads to a rejection of its null hypothesis, even if it were correct (i.e., the probability of making a type I error). In principle, this probability is fixed before computing the test statistics. However, tests based on discrete random variables can effectively operate only at levels that correspond to p-values of actual realizations of the test statistic (cf. Figure 2; Mood et al., 1974). It is therefore necessary to differentiate between the nominal significance level (i.e., the value of α denoting the significance threshold) and the effective significance level (i.e., the largest possible p-value that still falls below the α threshold) (see also Roy et al., 2000). In principle, "randomized tests" provide the means to adjust the effective significance level of a test to its nominal level α by introducing an auxiliary random variable, not related to the data under testing. However, due to conceptual objections

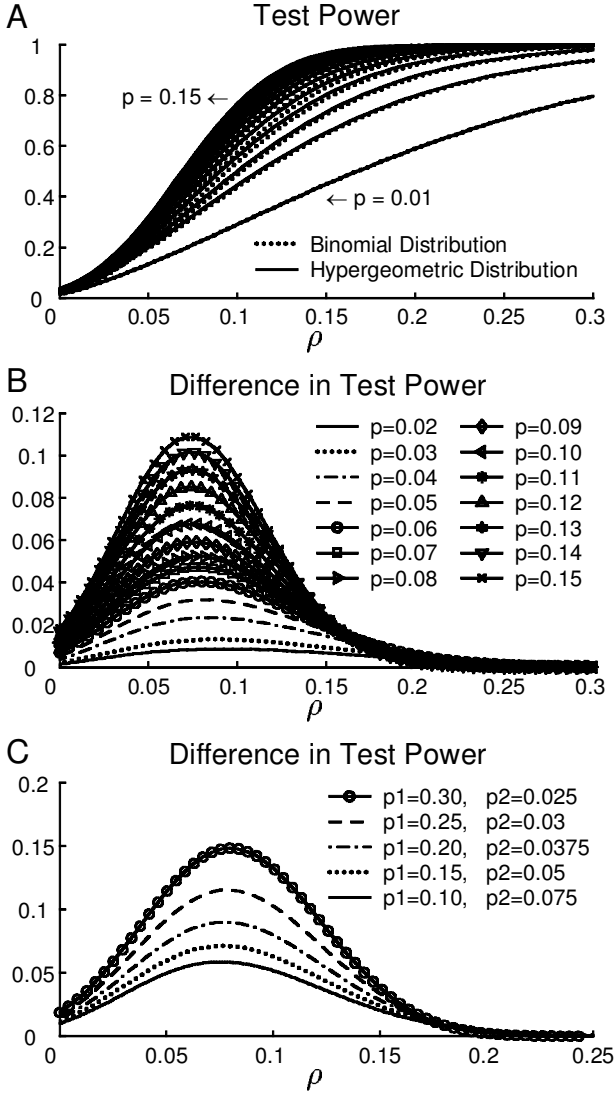


Figure 5: Dependence of both tests on spike probabilities ($n = 720$, $\alpha = 0.05$). (A) Test power curves of the two methods for "symmetrical" spike probabilities $p = p_1 = p_2 \in \{0.01, 0.02, \ldots, 0.15\}$. (B) Differences in test power corresponding to A. (C) Differences in test power for asymmetrical spike probabilities $(p_1, p_2) \in \{(0.1, 0.075), (0.15, 0.05), (0.2, 0.0375), (0.25, 0.03), (0.03, 0.025)\}$.

against this procedure, randomized tests are not commonly used in applied statistics (Mood et al., 1974), in spite of their attractive theoretical properties (Lehmann, 1997). A brief discussion of the randomization of the test investigated in this study is given in appendix E.

Given the critical coincidence counts $k^{\mathrm{bin}}_{\mathrm{crit}}(c_1, c_2)$ and $k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2)$ introduced in the previous section (see equations 4.1 and 4.2), the effective significance level of the tests investigated here is determined by the probability of obtaining a coincident count $k \geq k^{\mathrm{bin}}_{\mathrm{crit}}(c_1, c_2)$, evaluated with respect to the coincidence count distribution of the corresponding test, that is, by $J^{\mathrm{bin}}_{\xi_0}(k^{\mathrm{bin}}_{\mathrm{crit}}(c_1, c_2), c_1, c_2)$ and $J^{\mathrm{hyp}}_{\xi_0}(k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2), c_1, c_2)$, respectively (cf. equations 3.2 and 3.3). However, it is important to note that these effective significance levels can be calculated only if specific values of c_1 and c_2 are given. Only then is the coincident count distribution belonging to the corresponding test sufficiently specified. Due to this dependence on the specific realization of the counts C_1 and C_2, we will refer to these effective significance levels as count-dependent effective significance levels. Figure 6 shows the count-dependent effective significance levels for tests with a nominal significance level of $\alpha = 0.05$, for $n = 20$ and $n = 720$. The figure clearly demonstrates that the count-dependent effective significance levels of both tests fluctuate considerably between different realizations of the spike counts. In addition, it is interesting to note the effect of the neglect of count information by the original version of the test from the comparison of Figures 6A and 6C (see the figure caption for details).

We stress that the calculation of the count-dependent effective significance level does not take into account that the spike counts themselves are random variables. Thus, it is without doubt the appropriate measure to assess the effective significance level for the inferential statistical test of stretches of data with specific spike counts. However, one should be reluctant in interpreting this probability as the overall type I error probability of the method. This conceptual difference between the count-dependent effective significance level and the unconditional effective significance level (i.e., independent of the specific realization of the spike counts) gains crucial importance when interpreting the findings reported by Roy et al. (2000), although these authors do not seem to make this distinction consistently.

Instead of characterizing the method by its count-dependent effective significance level, we introduce the expected effective significance level α_{ξ_0}. The latter is defined as the expectation value of the count-dependent effective significance level with respect to the joint count distribution of independent neurons characterized by ξ_0. The interpretation of this measure as the expectation value of the count-dependent effective significance level rests on the additional assumption that the parameters $\xi_0 = (p_1, p_2, \rho = 0)$ of the joint count distribution are known. This assumption is not part of the null hypotheses of the tests investigated here. Thus, α_{ξ_0} has to be interpreted as the expectation value of the count-dependent effective significance level

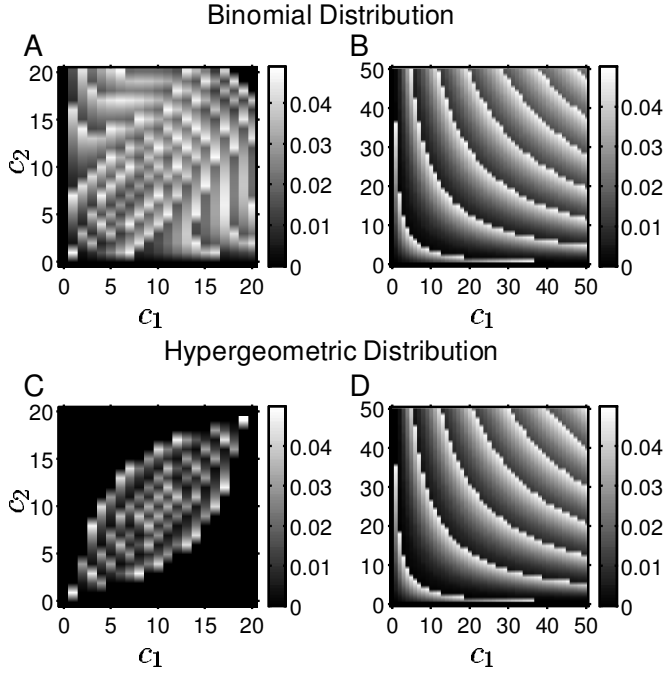


Figure 6: Count-dependent effective significance levels for $\alpha = 0.05$. (A, C) Small analysis window, $n = 20$. The black regions correspond to count combinations where the count-dependent effective significance level is zero; this occurs when either one of the counts is zero or when even the maximal possible number of coincidences—$k = 20$ for the binomial distribution and $k = \min(c_1, c_2)$ for the hypergeometric distribution—does not yield a significant p-value and, hence, the probability of rejecting the null hypothesis equals zero. (B, D) Large analysis window, $n = 720$. (A, B) Binomial coincidence count distribution. (C, D) Hypergeometric coincidence count distribution.

of the test when applied to neurons with spike probabilities p_1 and p_2. By contrast, the landscapes of count-dependent effective significance levels in Figure 6 do not depend on ξ_0, that is, on the values of p_1 and p_2. The parameter ξ_0 enters the computation of the expectation value α_{ξ_0} only through the joint count distribution underlying its definition.

Averaging over the joint count distribution for independent neurons (cf. appendix B), the expected effective significance level α_{ξ_0} is obtained in straightforward fashion:

$$\alpha^{\mathrm{bin}}_{\xi_0} = \mathrm{E}_{\xi_0}\left[J^{\mathrm{bin}}_{\xi_0}(k^{\mathrm{bin}}_{\mathrm{crit}}(c_1, c_2), c_1, c_2)\right]$$

$$= \sum_{c_1,c_2=0}^{n} P_{\xi_0}(c_1, c_2)\, J_{\xi_0}^{\mathrm{bin}}(k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2), c_1, c_2) \tag{5.1}$$

$$\alpha_{\xi_0}^{\mathrm{hyp}} = \mathrm{E}_{\xi_0}\left[J_{\xi_0}^{\mathrm{hyp}}(k_{\mathrm{crit}}^{\mathrm{hyp}}(c_1, c_2), c_1, c_2)\right]$$

$$= \sum_{c_1,c_2=0}^{n} P_{\xi_0}(c_1, c_2)\, J_{\xi_0}^{\mathrm{hyp}}(k_{\mathrm{crit}}^{\mathrm{hyp}}(c_1, c_2), c_1, c_2). \tag{5.2}$$

A discussion of a numerical evaluation of these expectation values is given in appendix D.

Since the test based on the hypergeometric coincidence count distribution makes no reference to the spike probabilities p_1 and p_2, it is independent of the "actual" spike probabilities of the investigated neurons. Thus, its expected count-dependent effective significance level $\alpha_{\xi_0}^{\mathrm{hyp}}$ reflects the probability that the test will reject its correct null hypothesis when applied to data from two independent neurons with spike probabilities ξ_0. Note that this is not the case for tests based on the binomial coincidence count distribution, where the testing procedure includes the estimation of the spike event probabilities p_i via $\hat{p}_i = c_i/n$ (see equation 3.1). In general, the estimated probabilities $\hat{p}_i$ will deviate from the actual parameters for most spike count combinations. Moreover, as pointed out in section 3, the binomial coincidence count distribution neglects the information contained in the actual spike counts. As a result, the count-dependent effective significance level based on these estimators will generally not correctly describe the probability of obtaining a coincident count $k \geq k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2)$. Therefore, the number $\alpha_{\xi_0}^{\mathrm{bin}}$ is of purely theoretical interest and does not bear any operational relevance.

Thus, to calculate the probability that a test based on the binomial coincidence count distribution will reject the null hypothesis when applied to independent neurons with spike probabilities p_1 and p_2, we need to compute the count-dependent probability to obtain a coincidence count $k \geq k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2)$ with respect to the hypergeometric distribution. Note, however, that $k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2)$ itself is calculated with respect to the binomial coincidence count distribution. Thus, again forming the expectation value with respect to the joint count distribution, we obtain

$$\epsilon_{\xi_0}^{\mathrm{bin}} = \mathrm{E}_{\xi_0}\left[J_{\xi_0}^{\mathrm{hyp}}(k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2), c_1, c_2)\right]$$

$$= \sum_{c_1,c_2=0}^{n} P_{\xi_0}(c_1, c_2)\, J_{\xi_0}^{\mathrm{hyp}}(k_{\mathrm{crit}}^{\mathrm{bin}}(c_1, c_2), c_1, c_2), \tag{5.3}$$

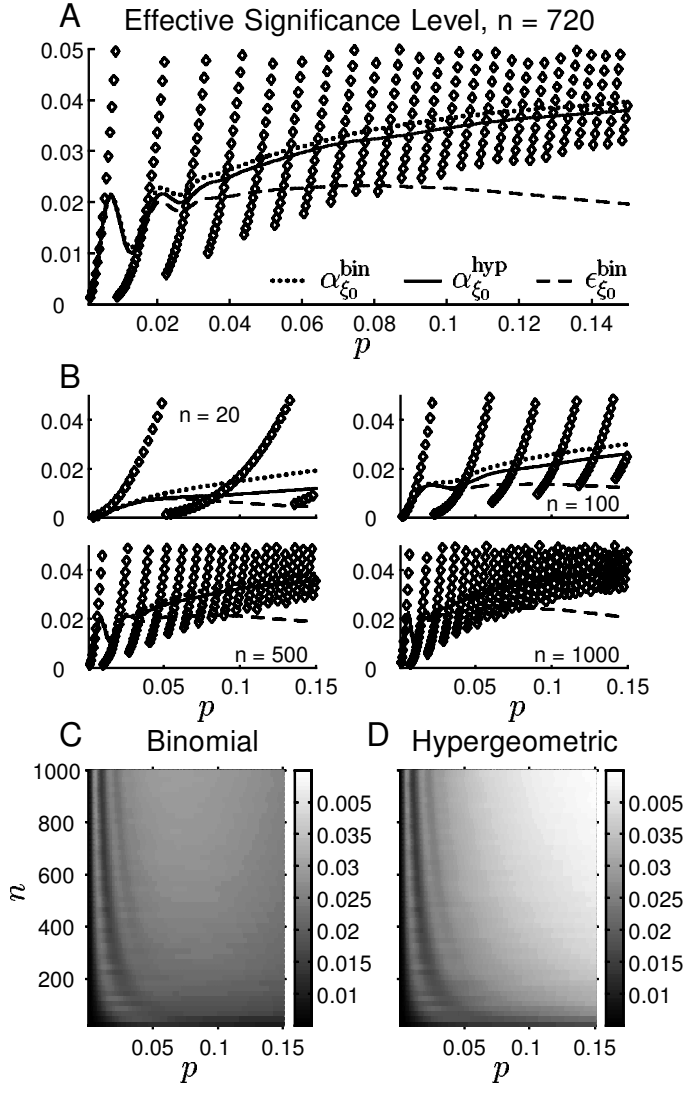
which we will refer to as α-error probability. It is clear from the above that for the corresponding α-error probability for the test based on the hypergeometric count distribution $\epsilon_{\xi_0}^{\mathrm{hyp}}$, we have $\epsilon_{\xi_0}^{\mathrm{hyp}} = \alpha_{\xi_0}^{\mathrm{hyp}}$.

5.1 Dependence of the Effective Significance Level on Spike Probabilities. In this section we investigate the dependence of the α-error probabilities on the spike probabilities p_1 and p_2, for both types of the test. Figure 7 shows the expected effective significance levels α_{ξ_0} and the α-error probability $\epsilon_{\xi_0}^{\mathrm{bin}}$ for $p = p_1 = p_2$ ranging from 0.001 to 0.15. The surface plots in Figures 7C and 7D display the dependence of these curves on the number of bins n. Overall, Figure 7 shows that the curves for the α-error probability of the modified method $\alpha_{\xi_0}^{\mathrm{hyp}}$ lie above the curves for the α-error probability of the original method $\epsilon_{\xi_0}^{\mathrm{bin}}$. Thus, especially for higher firing rates, where the difference between the curves amounts to a considerable fraction of the nominal significance level ($\alpha = 0.05$), the α-error probability of the modified method lies closer to the nominal significance level.

In Figures 7A and 7B, we also display the count-dependent effective significance levels that would arise from the binomial coincidence count distribution with coincidence probability p^2 (indicated by diamonds), as considered by Roy et al. (2000). Note that in order to obtain a continuous sample of these count-dependent effective significance levels, we did not restrict the values of p^2 to possible realizations of $c_1 c_2 / n^2$ but instead treated p as a continuous variable. In contrast to the view expressed by Roy et al. (2000), we emphasize that according to our formalism the interpretation of the ensuing sawtooth function (cf. Figure 7A) has to take into account that its independent variable p does not correspond to a neural spike probability, as it does in the calculation of α_{ξ_0} and $\epsilon_{\xi_0}^{\mathrm{bin}}$. Instead, this variable technically corresponds to an estimator of the spike probability, which would be based on realizations of the random variables C_1 and C_2 in applications of the method to experimental data.

In this context, it is important to recall that the count-dependent effective significance levels are different for different spike count combinations (cf. Figure 6). Thus, because of the stochastic nature of the spike counts, the tests will in general operate on different count-dependent effective significance levels for different realizations of the analysis window. Since this effect is due to the stochastic nature of the spike counts, it equally holds if the

Figure 7: *Facing page.* Dependence of expected effective significance level and α-error probability on analysis window size, for (A) $n = 720$ and (B) 20, 100, 500, and 1000, respectively. Diamonds depict corresponding count-dependent effective significance levels (cf. Roy et al., 2000; see text for details). (C, D) The α-error probabilities for the binomial and hypergeometric coincidence count distribution for $n \in \{20, 40, \ldots, 1000\}$. The oscillatory behavior of the curves for $n \gtrsim 100$ at low values of p reflects the periodic structure of the count-dependent effective significance level landscape (cf. Figure 6), which for small p dominates the expectation values because of the increasing localization of the joint count distributions. The analytical results shown in this figure were checked by computer simulations, based on the Ran2 random number generator (cf. Galassi et al., 1998).



probabilities p_1 and p_2 underlying the different realizations of the spike counts would remain perfectly constant. Therefore, contrary to the suggestion by Roy et al. (2000), the variation of count-dependent effective significance levels between different realizations of the analysis window is not an issue of neural firing rates. In fact, the count-dependent effective significance levels do not depend on the spike probabilities p_1 and p_2. Therefore, a change of neural firing rates would in no way influence the count-dependent effec-

tive significance levels—the shape of the sawtooth function in Figures 7A and 7B. Rather, it would lead to a change in the joint count distribution and, hence, affect the expected significance levels α_{ξ_0} and the α-error probability $\epsilon^{\mathrm{bin}}_{\xi_0}$. As shown in Figures 7A and 7B, computing these expectation values transforms the discrete sawtooth structure (diamonds) into smooth functions of p (curves). Thus, even for p as low as 0.02 (corresponding to a firing rate of 4 Hz for 5 ms bins) and n as in Figure 7, small changes in firing rate do not lead to large changes in the α-error probabilities of the two methods (as was claimed by Roy et al., 2000).

6 Discussion

We presented a modification of the statistical test underlying unitary-event analysis for the detection of neural synchrony. By incorporating the empirical spike counts into the calculation of the probability distribution of the test statistic (i.e., the coincidence count), we were able to remove the firing-rate estimation from the testing procedure. As a result, the distribution of the test statistic becomes independent of the a priori firing rates. The application of the modified method therefore avoids the problems of firing-rate estimation associated with statistical fluctuations in the spike counts.

To quantify the increase in sensitivity of the new method, we calculated and compared the test power of both tests with respect to violations of the null hypothesis of independent firing for various regimes of physiological parameters (firing rates; cf. Figure 5), degree of spike correlation (cf. Figures 3 and 5) and analysis parameters (size of analysis window; cf. Figure 4), and significance level (cf. Figure 3). The spike probabilities p_1 and p_2 were chosen such that the corresponding neural firing rates (for bin size $\Delta t = 5$ ms) lay between 2 Hz and 30 Hz. These results are of dual importance. First, they directly specify the probabilities of detecting given deviations from independent firing with the two methods. These probabilities are not only important quantities to characterize the performance of the statistical test, they are also critical in the context of experimental design: they allow one to choose appropriate values for analysis parameters such as the size of the time window necessary to verify a theoretically predicted dependence between two neurons. Second, the power curves allow us to quantify the effect of the suggested modification of the testing procedure on the performance of the test in comparison to the original version. This is important to reevaluate the results obtained by use of the original method and point out parameter regimes where the usage of the modified version of the test is especially crucial.

Overall, we found that for applications of the test to analysis windows comprising larger values of n ($n \gtrsim 100$), the modification proposed leads to an increase in test power of up to 0.12 (50% relative increase). This increase becomes especially pronounced for conservative nominal significance levels α, asymmetric firing regimes of the two neurons, high firing rates, and

moderate degrees of spike correlation. In general, for the firing rates analyzed, the peak values of the difference in test power were reached for values of the correlation between 0.05 and 0.15. This range corresponds to the empirical values of the asynchronous gain (ASG) found in the cortex (see reviews by Aertsen & Gerstein, 1985, and Abeles, 1991).

For short time windows ($n \approx 20$) we found that the test power of both methods falls below 0.2 (cf. Figure 4) for two neurons that operate below 30 Hz ($\Delta t = 5$ ms) with $\rho \approx 0.25$, corresponding to the maximal ASG for cortical neurons reported by Abeles (1991). Thus, when applying the test to single-trial data with short analysis windows comprising only a low number of bins, one has to face substantial reductions in test power. For these applications of the test, the increase in test power due to the modification of the method is vital. In addition, it is important to realize that for low values of n and for low firing rates, the discrete nature of the underlying test statistic dominates the properties of the test. Their dependence on n and α is complex in this regime, so special care should be taken with respect to the experimental design.

In addition, we have calculated the expected effective significance levels of the two tests and their α-error probabilities when applied to two neurons operating at equal rates, ranging from 0.2 Hz to 30 Hz ($\Delta t = 5$ ms). As discussed in detail in section 5, it is important to differentiate between the count-dependent effective significance level that does not depend on the neuronal spike probabilities and the expected effective significance level that depends on the spike probabilities through the joint count distribution. Contrasting the view expressed by Roy et al. (2000), who based their analysis on the count-dependent effective significance levels, our calculations show that the α-error probabilities of both methods vary only slowly as a function of firing rate above 4 Hz. Thus, in this regime, the probability of falsely indicating neural synchrony is only moderately sensitive to the firing-rate levels of the investigated neurons. While this result describes the behavior of the method when applied to neurons operating at certain firing rate levels, that is, independent of any specific empirical realization of spike counts (c_1, c_2), the count-dependent effective significance level (i.e., the effective significance level of the test when applied to a stretch of data with a specific combination of spike counts) does show considerable fluctuations depending on the joint counts in the analysis window (cf. Figure 6). This indeed implies fluctuations of the count-dependent effective significance level with respect to different realizations of the analysis window. We emphasize that these fluctuations are not the result of changes in neural firing rates, but the consequence of stochastic fluctuations in the counts themselves.

For firing rates below 4 Hz, the joint count distribution tends to concentrate on a small number of count combinations. This causes the changes of the α-error probabilities $\epsilon_{\xi_0}^{\mathrm{bin}}$ and $\epsilon_{\xi_0}^{\mathrm{hyp}}$ of the two versions of the test with

changes in neural firing rates to become more pronounced. For neurons operating at rate levels below 1 Hz, the joing count distribution becomes so narrow that the expectation value of the count-dependent effective significance level essentially behaves like the count-dependent effective significance level itself. Thus, the probability of producing a false positive when applying the method at firing rates below 1 Hz steeply decreases as the firing rate approaches lower values. As a consequence, in this firing-rate regime, the expected effective significance level of the test becomes sensitive to the firing rates of the investigated neurons. Since, by construction of the test, the effective significance level cannot surpass the nominal significance level α, the problem is not that the test could produce more false positives than expected. Rather, the decrease of the effective significance level for low firing rates implies that the test will effectively operate at a very conservative significance level. However, as can be seen from the power curves (cf. Figure 5), this decrease in effective significance level is accompanied by a decrease in test power. Thus, for very low firing rates, both versions of the test will fail to detect deviations from the null hypothesis of independent firing.

Finally, we could show that the use of the hypergeometric coincidence count distribution allows us to relax the stationarity requirement of the neuronal firing rates. In contrast to the original method, which had to assume the firing rates of both neurons to remain stationary over all bins of the analysis window, the modified test requires only one of the two neurons to have a stationary firing rate, while the other can follow an arbitrary time course. This generalization implies an important increase of applicability of the method to empirical data. We are currently investigating whether this "one-sided" stationarity criterion can be further relaxed, possibly by imposing joint (but weaker) requirements on both neuronal rate profiles. The robustness of the modified method with respect to the violation of the "one-sided" stationarity assumption will also be the subject of further research. A conceptually different approach to treat nonstationary neural data with count-based statistics could be based on the use of estimators for the instantaneous firing rate. Following recent work by Pauluis and Baker (2000), who implement a rate-based version of unitary-event analysis (Grün, 1996) in connection with an instantaneous rate estimation procedure, it might be interesting to use instantaneous rate estimation (Nawrot, Aertsen, & Rotter, 1999) together with the conditional coincidence count distribution used here for variable spike event probabilities. This approach seems capable of combining the advantages of count-based statistics with the improved applicability of instantaneous rate estimators to nonstationary data sets.

In conclusion, in view of the increase in test power, the increased interpretability of the significance measure, and the relaxation of the stationarity requirement, we clearly recommend implementation of the count-based rather than the rate-based version of this analysis method when testing the statistical significance of coincident spikes.

Appendix A: Stochastic Model

According to the definition of a spike event given in section 2, we define the probability space (Ω_i, P_{ξ_i}) for each pair of bins within the analysis window (indexed by $i \in \{1, \ldots, n\}$) on the basis of the sample space,

$$\Omega_i := \left\{ \omega_i = \left(\omega_{1,i}, \omega_{2,i}\right) : \omega_{1,i}, \omega_{2,i} \in \{0, 1\} \right\}, \tag{A.1}$$

and a probability P_{ξ_i} for each of the four possible outcomes (0,0), (0,1), (1,0), (1,1). A parameterization of these probabilities in terms of the individual spike event probabilities $p_{1,i}$ and $p_{2,i}$,

$$P_{\xi_i}\left(\omega_{1,i} = 1\right) = p_{1,i} \quad \text{and} \quad P_{\xi_i}\left(\omega_{2,i} = 1\right) = p_{2,i}$$
$$(p_{1,i}, p_{2,i} \in (0, 1)), \tag{A.2}$$

and the spike correlation ρ_i between the two neurons,

$$\mathrm{Corr}(\omega_{1,i}, \omega_{2,i}) = \rho_i \qquad (\rho_i \in (-1, 1)), \tag{A.3}$$

leads to the definition

$$P_{\xi_i}\left(\omega_i = (1, 1)\right) := p_{1,i}p_{2,i} + \rho_i R_i \tag{A.4}$$
$$P_{\xi_i}\left(\omega_i = (1, 0)\right) := p_{1,i}(1 - p_{2,i}) - \rho_i R_i \tag{A.5}$$
$$P_{\xi_i}\left(\omega_i = (0, 1)\right) := (1 - p_{1,i})p_{2,i} - \rho_i R_i \tag{A.6}$$
$$P_{\xi_i}\left(\omega_i = (0, 0)\right) := (1 - p_{1,i})(1 - p_{2,i}) + \rho_i R_i, \tag{A.7}$$

with $R_i = \sqrt{p_{1,i}(1 - p_{1,i})p_{2,i}(1 - p_{2,i})}$ and $\xi_i := (p_{1,i}, p_{2,i}, \rho_i)$.

Based on equations A.4 through A.7, the conditional probabilities to observe a spike event from neuron 2, given the behavior of neuron 1, are

$$\vartheta_i := P_{\xi_i}\left(\omega_{2,i} = 1 \mid \omega_{1,i} = 1\right) = p_{2,i} + \frac{\rho_i R_i}{p_{1,i}} \tag{A.8}$$

$$\varphi_i := P_{\xi_i}\left(\omega_{2,i} = 1 \mid \omega_{1,i} = 0\right) = p_{2,i} - \frac{\rho_i R_i}{1 - p_{1,i}}. \tag{A.9}$$

These conditional probabilities will be used in appendix B, yielding compact expressions of the probability distribution functions.

Starting from assumptions A_1 and A_2 as stated in section 2, we can define the probability space describing the entire analysis window by forming the product space (Ω, P_ξ) with

$$\Omega := \Omega_1 \times \cdots \times \Omega_n, \qquad P_\xi := \prod_{i=1}^{n} P_{\xi_i}, \tag{A.10}$$

and samples $\omega = (\omega_1, \ldots, \omega_n)$. Based on this product space, we define the discrete random variables,

$$C_m(\omega) := \sum_{i=1}^{n} \omega_{m,i} \qquad m = 1, 2, \tag{A.11}$$

denoting the total number of spike events from neuron m in the analysis window. Similarly,

$$K(\omega) := \sum_{i=1}^{n} \omega_{1,i} \cdot \omega_{2,i}, \tag{A.12}$$

denotes the number of coincident spike events from the two neurons.

Appendix B: Probability Distributions

For the derivations in this section, we will assume stationarity of all parameter triplets ξ_i in the analysis window, as formulated in assumption A_1. We let $\xi = \xi_i$ for all $i = 1, 2, \ldots, n$. Given the probabilities of the four possible spike event constellations of any two coinciding bins (see equations A.4–A.7) and using the conditional spike probabilities ϑ and φ (see equations A.8 and A.9), application of the multinomial distribution (cf. Feller, 1968) yields the probability of finding k coincident events and c_1, respectively c_2, spike events from the two neurons, that is, of making an observation ω with $K(\omega) = k$, $C_1(\omega) = c_1$, and $C_2(\omega) = c_2$:

$$\begin{aligned} &P_\xi(K = k, C_1 = c_1, C_2 = c_2) \\ &\quad = \frac{n!}{k!(c_1 - k)!(c_2 - k)!(n - c_1 - c_2 + k)!} \\ &\qquad \times \left[p_1\vartheta\right]^k \left[p_1(1 - \vartheta)\right]^{c_1 - k} \left[(1 - p_1)\varphi\right]^{c_2 - k} \\ &\qquad \times \left[(1 - p_1)(1 - \varphi)\right]^{n - c_1 - c_2 + k}. \end{aligned} \tag{B.1}$$

Using

$$P_\xi(C_1 = c_1) = \binom{n}{c_1} {p_1}^{c_1} (1 - p_1)^{(n - c_1)}, \tag{B.2}$$

it follows that

$$\begin{aligned} P_\xi(K = k, C_2 = c_2 \mid C_1 = c_1) &= \frac{P_\xi(K = k, C_1 = c_1, C_2 = c_2)}{P_\xi(C_1 = c_1)} \\ &= \binom{c_1}{k} \binom{n - c_1}{c_2 - k} \vartheta^k (1 - \vartheta)^{(c_1 - k)} \varphi^{(c_2 - k)} \\ &\quad \times (1 - \varphi)^{(n - c_1 - c_2 + k)}. \end{aligned} \tag{B.3}$$

Since the sample space Ω can be decomposed into disjoint subsets containing all elementary events with a specific coincidence count, we can write the joint count distribution as

$$\begin{aligned} P_\xi(C_1 = c_1, C_2 = c_2) &= P_\xi(C_1 = c_1) P_\xi(C_2 = c_2 \mid C_1 = c_1) \\ &= P_\xi(C_1 = c_1) \\ &\quad \times \sum_{k=0}^{\min(c_1,c_2)} P_\xi(K = k, C_2 = c_2 \mid C_1 = c_1), \end{aligned} \tag{B.4}$$

which by insertion of equation B.3 turns into

$$\begin{aligned} P_\xi(C_1 = c_1, C_2 = c_2) = &\binom{n}{c_1} p_1^{c_1} (1 - p_1)^{(n-c_1)} \\ &\times \sum_{k=0}^{\min(c_1,c_2)} \binom{c_1}{k} \vartheta^k (1 - \vartheta)^{(c_1-k)} \binom{n - c_1}{c_2 - k} \varphi^{(c_2-k)} \\ &\times (1 - \varphi)^{(n-c_1-c_2+k)}. \end{aligned} \tag{B.5}$$

Going back to equation B.1, it is now straightforward to derive the conditional coincidence count distribution. By inserting equations B.1 and B.5 into

$$P_\xi(K = k \mid C_1 = c_1, C_2 = c_2) = \frac{P_\xi(K = k, C_1 = c_1, C_2 = c_2)}{P_\xi(C_1 = c_1, C_2 = c_2)}, \tag{B.6}$$

we find

$$\begin{aligned} &P_\xi(K = k \mid C_1 = c_1, C_2 = c_2) \\ &= \frac{\binom{c_1}{k} \vartheta^k (1 - \vartheta)^{(c_1-k)} \binom{n-c_1}{c_2-k} \varphi^{(c_2-k)} (1 - \varphi)^{(n-c_1-c_2+k)}}{\sum_{k=0}^{\min(c_1,c_2)} \binom{c_1}{k} \vartheta^k (1 - \vartheta)^{(c_1-k)} \binom{n-c_1}{c_2-k} \varphi^{(c_2-k)} (1 - \varphi)^{(n-c_1-c_2+k)}}. \end{aligned} \tag{B.7}$$

For independent neurons ($\rho = 0$), the conditional probabilities ϑ and φ become equal to the spike probability of neuron 2 ($\vartheta = \varphi = p_2$). Therefore, it is a matter of straightforward substitution to derive the conditional coincidence count distribution $P_{\xi_0}(K = k \mid C_1 = c_1, C_2 = c_2)$ and the joint count distribution $P_{\xi_0}(C_1 = c_1, C_2 = c_2)$ for independent neurons from the general expressions in equations B.7 and B.5, respectively. Finally, the general form

of the unconditional coincidence count distribution (cf. section 2) is found by replacing p_2 of equation 2.1 with ϑ, so that the term $p_1\vartheta$ corresponds to the general form of the probability to observe a pair of coincident spikes within one bin (cf. equation A.4).

Appendix C: Nonstationary Rates

Relaxing the stationarity assumption A_1, we return to the general formulation of our stochastic model as developed in section 2 and appendix A. Thus, we replace the parameter triplet ξ with a vector of triplets $\vec{\xi}$. Its n components $\xi_i = (p_{1,i}, p_{2,i}, \rho_i)$ describe the probability space (Ω_i, P_{ξ_i}) of each individual pair of corresponding bins in the analysis window. Assuming stochastic independence of the two neurons (i.e., H_0), we can rewrite equation B.6 as

$$P_{\vec{\xi}_0}(K = k \mid C_1 = c_1, C_2 = c_2) = \frac{P_{\vec{\xi}_0}(K = k, C_1 = c_1, C_2 = c_2)}{P_{\vec{\xi}_0}(C_1 = c_1) P_{\vec{\xi}_0}(C_2 = c_2)}, \tag{C.1}$$

where we used $\vec{\xi}_0$ to indicate this parameter setting under the condition H_0.

Following assumption A_2, we can write the probability for a specific realization as the product of the probabilities of obtaining or not obtaining a spike event in each of the corresponding bins, respectively. Thus, the probability $P_{\vec{\xi}_0}(K = k, C_1 = c_1, C_2 = c_2)$ of making an observation with c_1 spike events from neuron 1, c_2 spike events from neuron 2, and k coincident events is given by the sum over the probabilities of all possible arrangements of this count configuration. For $M = \{1, 2, \ldots, n\}$, where n is the number of bins in the analysis window, we define the set M_{c_1} as the collection of all subsets of M with c_1 elements. Further, for any $\mu \in M_{c_1}$, we let the set $M_{c_2}^{\mu,k}$ denote the collection of all subsets of M with c_2 elements in total and k elements in common with μ. Using this notation, we have

$$\begin{aligned} &P_{\vec{\xi}_0}(K = k, C_1 = c_1, C_2 = c_2) \\ &\quad = \sum_{\mu \in M_{c_1}} \sum_{\lambda \in M_{c_2}^{\mu,k}} \prod_{i \in \mu} p_{1,i} \prod_{j \in M \setminus \mu} (1 - p_{1,j}) \prod_{l \in \lambda} p_{2,l} \prod_{m \in M \setminus \lambda} (1 - p_{2,m}). \end{aligned} \tag{C.2}$$

Using the same notation, the probability $P_{\vec{\xi}_0}(C_1 = c_1)$ is given by

$$P_{\vec{\xi}_0}(C_1 = c_1) = \sum_{\mu \in M_{c_1}} \prod_{i \in \mu} p_{1,i} \prod_{j \in M \setminus \mu} (1 - p_{1,j}). \tag{C.3}$$

The probability $P_{\vec{\xi}_0}(C_2 = c_2)$ can be expressed analogously. Thus, we can rewrite equation C.1 as

$$P_{\vec{\xi}_0}(K = k \mid C_1 = c_1, C_2 = c_2)$$

$$= \frac{\sum_{\mu \in M_{c_1}} \sum_{\lambda \in M_{c_2}^{\mu,k}} \prod_{i \in \mu} p_{1,i} \prod_{j \in M \setminus \mu} (1 - p_{1,j}) \prod_{l \in \lambda} p_{2,l} \prod_{m \in M \setminus \lambda} (1 - p_{2,m})}{\left[\sum_{\eta \in M_{c_1}} \prod_{i \in \eta} p_{1,i} \prod_{j \in M \setminus \eta} (1 - p_{1,j}) \right] \left[\sum_{\kappa \in M_{c_2}} \prod_{i \in \kappa} p_{2,i} \prod_{j \in M \setminus \kappa} (1 - p_{2,j}). \right]} \tag{C.4}$$

This equation describes the conditional probability that, given the individual spike event counts c_1 and c_2, one will observe k coincident events from two stochastically independent neurons with spike event probabilities according to $\vec{\xi}_0$.

Assuming a stationary rate for neuron 2 (i.e., $p_{2,i} = p_2$ for all i), equation C.4 reduces to

$$\begin{aligned} & P_{\vec{\xi}_0}(K = k \mid C_1 = c_1, C_2 = c_2) \\ &= \frac{\left[\sum_{\mu \in M_{c_1}} \prod_{i \in \mu} p_{1,i} \prod_{j \in M \setminus \mu} (1 - p_{1,j}) \right] \binom{c_1}{k} \binom{n - c_1}{c_2 - k} p_2^{c_2} (1 - p_2)^{n - c_2}}{\left[\sum_{\eta \in M_{c_1}} \prod_{i \in \eta} p_{1,i} \prod_{j \in M \setminus \eta} (1 - p_{1,j}) \right] \binom{n}{c_2} p_2^{c_2} (1 - p_2)^{n - c_2}} \\ &= \frac{\binom{c_1}{k} \binom{n - c_1}{c_2 - k}}{\binom{n}{c_2}}, \end{aligned} \tag{C.5}$$

which is the hypergeometric distribution as given in equation 2.2, independent of whether the firing rate of neuron 1 is stationary over the observation interval. Obviously, by interchanging the roles of the two neurons in the definitions of M_{c_1} and $M_{c_2}^{\mu,k}$, the same result can be obtained for a stationary rate in neuron 1—for $p_{1,i} = p_1$ for all i. In other words, stationarity of only one of the two neurons (either one) suffices to obtain the result in equation C.5.

Appendix D: Approximative Evaluation of the Expectation Values

The number of terms in the sums underlying the computation of the expectation values for the count-dependent test power and significance level with respect to the joint count distribution (cf. equations 4.4, 4.5, and 5.1–5.3) grows with n^2. Thus, the direct evaluation of the sums for realistic values of n, which can reach up to thousands, is rather unpractical. However, by using the fact that the mass of the joint count distribution is mainly concentrated at relatively few count combinations, it is straightforward to

calculate an approximate value of these quantities up to arbitrary precision δ, for example, the approximate test power $\pi_{\xi,\delta}^{\text{approx}}$. Selecting $\mathcal{B}_{\xi,\delta} \subset \{0, 1, \ldots, n\} \times \{0, 1, \ldots, n\}$ such that

$$\sum_{\mathcal{B}_{\xi,\delta}} P_\xi(C_1 = c_1, C_2 = c_2) \geq 1 - \delta, \tag{D.1}$$

we find the approximate test power $\pi_{\xi,\delta}^{\text{approx}}$ given by

$$\pi_{\xi,\delta}^{\text{approx}} = \sum_{(c_1,c_2)\in\mathcal{B}_{\xi,\delta}} P_\xi(C_1 = c_1, C_2 = c_2) P_\xi(K \geq k_{\text{crit}}(c_1, c_2) \mid c_1, c_2). \tag{D.2}$$

Since all probabilities are smaller than unity, it is clear from equation D.1 that $\pi_\xi - \pi_{\xi,\delta}^{\text{approx}} \leq \delta$. When computing the effective significance levels α_{ξ_0} and ϵ_{ξ_0}, the expectation value is formed over quantities smaller than α. Thus, the precision of the approximation of these significance levels improves to $\alpha\delta$.

Note that $\mathcal{B}_{\xi,\delta}$ is not uniquely defined through equation D.1. In our calculations, $\mathcal{B}_{\xi,\delta}$ was determined by successively adding up the mass of individual spike event count combinations until the cutoff value $1 - \delta$ was reached. To keep the number of individual spike count combinations entering $\mathcal{B}_{\xi,\delta}$ reasonably low, we started with the central term (cf. Feller, 1968) of the joint count distribution of independent neurons and iteratively added those count combinations from the surrounding of $\mathcal{B}_{\xi,\delta}$ that contributed most. Although not required for this procedure, this approach was motivated by the fact that the joint count distribution falls off monotonically with increasing distance from its central term.

Appendix E: Randomized Tests

The basic idea behind randomized tests (Mood et al., 1974) is to define a test through a critical function ψ that defines rejection probabilities for given empirical observations rather than through a fixed critical region of rejection $\mathcal{R}$. Through the incorporation of an additional independent random variable, it becomes possible to adjust the effective significance level of the test to match its nominal significance level α precisely, regardless of any discreteness of its test statistic.

Applying this concept to the test based on the hypergeometric coincidence count distribution, this means that the decision of the test will no longer be based on the critical region $\mathcal{R}_{c_1,c_2} = \{k\colon k \geq k_{\text{crit}}^{\text{hyp}}\}$ of overly critical coincident spike event counts k but rather on the critical function,

$$\psi_{c_1,c_2} := \begin{cases} 1 & k \geq k_{\text{crit}}^{\text{hyp}}(c_1, c_2) \\ \zeta_{c_1,c_2} & k = k_{\text{crit}}^{\text{hyp}}(c_1, c_2) - 1 \\ 0 & k < k_{\text{crit}}^{\text{hyp}}(c_1, c_2) - 1, \end{cases} \tag{E.1}$$

which for every element ω of the sample space Ω sets the probability of rejecting the null hypothesis. Thus, while the null hypothesis will always be rejected if $k \geq k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2)$ and never be rejected if $k < k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1$, the parameter ζ_{c_1,c_2} controls the rejection probability for all ω with $k = k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1$.

In order to adjust the effective significance level of the test to its nominal significance level α, we define ζ_{c_1,c_2} such that

$$\zeta_{c_1,c_2} := \frac{\alpha - P_{\xi_0}(K \geq k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2))}{P_{\xi_0}(K = k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1)}. \tag{E.2}$$

From here it is straightforward to see that the probability of falsely rejecting the null hypothesis of stochastically independent neurons reduces to

$$P_{\xi_0}(K \geq k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2)) + \zeta_{c_1,c_2} P_{\xi_0}(K = k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1) = \alpha, \tag{E.3}$$

and, thus, for all count combinations precisely corresponds to the nominal level of significance. Note that this procedure adjusts the probability of falsely rejecting the null hypothesis by introducing rejections of the null hypothesis with probability ζ_{c_1,c_2} for all count constellations with $k = k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1$. While this raises the probability of a false rejection to the nominal α-level of the test, and correspondingly increases its test power, the outcome of the test for a given set of data becomes a random variable. Thus, repeated applications of the test to the same data will in general lead to different findings. This indeterminacy of randomized tests is the reason for their restricted use in applied statistics (Mood et al., 1974).

By straightforward extension of equation 4.5, we find the power $\pi^{\mathrm{hyp}}_{\xi,\mathrm{Rnd}}$ of the randomized version of the test to be given by

$$\begin{aligned} \pi^{\mathrm{hyp}}_{\xi,\mathrm{Rnd}} = \sum_{c_1,c_2=0}^{n} & P_\xi(C_1 = c_1, C_2 = c_2) \\ & \times \Big[P_\xi(K \geq k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) \mid c_1, c_2) \\ & \quad + \zeta_{c_1,c_2} P_\xi(K = k^{\mathrm{hyp}}_{\mathrm{crit}}(c_1, c_2) - 1 \mid c_1, c_2) \Big]. \end{aligned} \tag{E.4}$$

Comparison of the power curves for the randomized versus the nonrandomized test in the same parameter regime as used in Figure 3 demonstrates that, as expected, the power curves of the randomized version of the test lie above the values reached by the nonrandomized version. The maximum increase in test power ranges from approximately 0.05 for $\alpha = 0.01$ ($\rho = 0.1$) to approximately 0.07 for $\alpha = 0.1$ ($\rho = 0.05$). Thus, the effect of randomization becomes larger for more permissive significance levels α. Finally, we note

that the randomized version of the test based on the hypergeometric coincidence count distribution, that is, of Fisher's exact test, is uniformly most powerful unbiased for testing independence in a 2×2 table (Lehmann, 1997).

Acknowledgments

We thank Hans Rudolf Lerche for valuable discussions and for calling our attention to Fisher's exact test. We thank Sonja Grün for helpful discussions and Arup Roy for kindly providing us with his manuscript prior to publication. We gratefully acknowledge Alexandre Kuhn for helpful comments on the manuscript. This work was supported in part by the Studienstiftung des deutschen Volkes, the German–Israeli Foundation for Scientific Research and Development (GIF), the Deutsche Forschungsgemeinschaft (DFG), and the Institut für Grenzgebiete der Psychologie, Freiburg.

References

Abeles, M. (1982). *Local cortical circuits: An electrophysiological study*. Berlin: Springer-Verlag.

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Abeles, M., Bergman, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol., 70*(4), 1629–1638.

Abeles, M., & Gerstein, G. L. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol., 60*(3), 909–924.

Aertsen, A., Bonhoeffer, T., & Krüger, J. (1987). Coherent activity in neuronal populations: Analysis and interpretation. In E. R. Canianello (Ed.), *Physics of cognitive processes* (pp. 1–34). Singapore: World Scientific Publishing.

Aertsen, A., & Gerstein, G. L. (1985). Evaluation of neuronal connectivity: Sensitivity of cross-correlation. *Brain Research, 340*, 341–354.

Aertsen, A., & Gerstein, G. L. (1991). Dynamic aspects of neuronal cooperativity: Fast stimulus-locked modulations of effective connectivity. In J. Krüger (Ed.), *Neuronal cooperativity* (pp. 52–67). Stuttgart: Springer-Verlag.

Aertsen, A., Gerstein, G., Habib, M., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of "effective connectivity." *J. Neurophysiol., 61*(5), 900–917.

Feller, W. (1968). *An introduction to probability theory and its application* (3rd ed.). New York: Wiley.

Galassi, M., Davies, J., Theiler, J., Gough, B., Priedhorsky, R., Jungman, G., & Booth, M. (1998). *GNU Scientific Library—Reference Manual* (0.4after ed.). Cambridge, MA: Free Software Foundation. Available online at: http://sources.redhat.com/gsl.

Gerstein, G. L., Bedenbaugh, P., & Aertsen, A. (1989). Neuronal assemblies. *IEEE Trans. Biomed. Eng., 36*, 4–14.

Grün, S. (1996). *Unitary joint-events in multiple–neuron spiking activity: Detection, significance, and interpretation*. Thun: Verlag Harri Deutsch.

Grün, S., Diesmann, M., & Aertsen, A. (in press-a). "Unitary events" in multiple single neuron spiking activity I. Detection and significance. *Neural Computation*.

Grün, S., Diesmann, M., & Aertsen, A. (in press-b). "Unitary events" in multiple single neuron spiking activity. II. Non-stationary data. *Neural Computation*.

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

Gütig, R., Rotter, S., Grün, S., & Aertsen, A. (2000). Significance of coincident spikes: Count-based versus rate-based statistics. In *Ninth Annual Computational Neuroscience Meeting CNS*2000* (p. 69). Brussels.

Kreiter, A. K., & Singer, W. (1996). On the role of neural synchrony in the primate visual cortex. In A. Aertsen & V. Braitenberg (Eds.), *Brain theory—Biological basis and computational principles* (pp. 201–227). Amsterdam: Elsevier.

Lehmann, E. (1997). *Testing statistical hypotheses* (2nd ed.). New York: Springer-Verlag.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). Tokyo: McGraw-Hill.

Nawrot, M., Aertsen, A., & Rotter, S. (1999). Single-trial estimation of neuronal firing rates. *J. Neurosci. Meth., 94*, 81–91.

Nicolelis, M. A. L. (Ed.). (1998). *Methods for neural ensemble recordings*. Boca Raton, FL: CRC Press.

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

Pauluis, Q., & Baker, S. N. (2000). An accurate measure of instantaneous discharge probability, with application to unitary joint-event analysis. *Neural. Comp., 12*, 647–669.

Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Hamutal, S., & Abeles, M. (1998). Spatiotemporal structure of cortical activity: Properties and behavioral relevance. *J. Neurophysiol., 79*(6), 2857–2874.

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

Roy, A., Steinmetz, P. N., & Niebur, E. (2000). Rate limitations of unitary event analysis. *Neural. Comp., 12*, 2063–2082.

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

Singer, W. (1999). Time as coding space. *Curr. Op. Neurobiol., 9*(2), 189–194.

von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Rep. No. 81-2). Göttingen: Max-Planck-Institute for Biophysical Chemistry.

Received August 1, 2000; accepted March 19, 2001.