Data-Driven Significance Estimation for Precise Spike Correlation

Sonja Grün

Theoretical Neuroscience Group, Riken Brain Science Institute, Wako-Shi, Japan

Submitted 23 October 2008; accepted in final form 10 December 2008

The principles of neuronal information processing are still not well understood. In particular, the mechanisms underlying neuronal coding continue to be debated. There are two main views: rate coding emphasizes that firing rate is used as the information carrier (Shadlen and Movshon 1999), whereas temporal coding emphasizes the role of precisely timed spikes (Singer 1999). The notion behind the latter is that coordination of spike timing between neurons indicates interaction within neuronal groups (cell assemblies; Hebb 1949). Experimental studies provide support for both perspectives, and both coding schemes may well coexist. However, this discussion is often implicitly a discussion about the underlying analysis methods and being able to decide between the two options implies that we must understand how to differentiate the two (Staude et al. 2008). In particular, showing that there is temporal coordination that is beyond what is expected by chance is indeed a difficult issue, because false-positive results are prone to occur if certain properties of the spike trains are not properly considered. For example, if the firing rates of the neurons increase simultaneously, the number of coincident spike events measured across the neurons will trivially increase. Thus it is the task of the analysis method to show that there are even more coincident spike events than are expected by the increased firing rates.

Because of the stochastic nature of the firing of cortical neurons and the typically low firing rates, joint spike occurrences are relatively rare and thus require advanced methods to obtain reliable statistics. One solution that immediately comes to mind is to average over (long) stretches of data, which is possible, but only if firing rates are stationary. However, experimentalists make every effort to make neurons respond to the experimental manipulation, i.e., to induce temporary changes in their firing rates. Thus approaches must be found to improve statistics for relatively short time windows. For this reason, experiments are repeated under the same conditions to get enough samples by averaging across trials. This, however, requires that the statistics are stationary across the trials. Unfortunately, this assumption may be violated by, for example, effects of anesthetics or attention, in which case averaging of parameters is not valid. These and other features of experimental data, which are particularly prominent in data of awake, behaving animals, need to be considered in correlation analysis. This review is intended to provide an overview of potential obstacles and possible routes to overcome them. I illustrate the various strategies by using as an example analysis tool, the Unitary Events (UE) analysis method. The method was specifically designed to test the hypothesis that cortical neurons coordinate their spiking activity in brief volleys of synchronous spikes. It detects the presence of spike coincidences in simultaneously recorded multiple single-unit spike trains and evaluates their statistical significance. Because the method is well calibrated and thoroughly tested, it provides a suitable framework for this discussion. The conclusions, however, are more generally valid for other correlation analysis techniques, ranging from cross-correlation analysis to spatio-temporal pattern analysis.

The paper is structured as follows: after a brief review of the basic principles of the UE analysis, I introduce various issues that arise in the analysis of precise spike correlation in experimental data, discuss their influence on UEs, and present potential corrections. This includes how to properly adjust to the temporal scale of correlation, how to handle various types of nonstationarity, and how to treat deviations from Poisson statistics. Solutions based on analytical methods and methods based on surrogate data are discussed. An additional section focuses on the impact of data preprocessing such as spike sorting. In a general discussion, I compare various methods for generation of control data, discuss their costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Address for reprint requests and other correspondence: S. Grün, Theoretical Neuroscience Group, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-Shi, 351-0198, Japan (E-mail: gruen@brain.riken.jp).
properties, and relate them to other analysis tools. Finally, I summarize issues that need to be addressed in future studies.

**DETECTION AND ANALYSIS OF PRECISE SPIKE CORRELATION**

Detection of joint-spike events

The UE (Grün et al. 2002a) analysis method is designed to detect coincident spike patterns between two or more simultaneously recorded spike trains and to evaluate their significance. The specific questions addressed by this analysis are 1) do the simultaneously recorded neurons show correlations of their firing activity, 2) is any such correlation specific to subgroups of the neurons, and 3) do these correlations change dynamically as a function of stimuli or behavior?

In this method, the spiking activity of each neuron is represented as a sequence of zeros and ones after appropriate time discretization (e.g., $h = 1 \text{ ms}$), a one representing the existence of at least one spike (clipping), and a zero the absence of a spike (Fig. 1A). Under the assumption of stationary firing, the firing probability $p_i$ per neuron $i$ can be estimated by evaluating its frequency, i.e., the number of ones $c_i$ divided by the total number of bins $n = T/h$ of the observed time interval of duration $T$: $p_i = \frac{c_i}{n}$. A similar statistic can be obtained for the coincidence patterns composed of the zeros and ones of a single time bin across the neurons. There are at most $2^N$ different coincidence patterns in data of $N$ simultaneously observed neurons. The actual number of different coincidence patterns found is typically lower, but a unique index $k$ can be assigned to each pattern. For each pattern $k$, we determine the number of occurrences yielding its empirical count $n_{\text{emp}}^k$.

**Significance of joint spike events**

We are interested in detecting whether there are coincidence patterns that occur significantly more often than expected on the basis of their firing rates. To this end, we compare the empirical number of occurrences of pattern $k$ to the expected one by calculating the joint probability of occurrence in each of the neurons assuming statistical independence: $P_{\text{exp}} = \prod_{i=1}^{n} \varphi_i$, with $\varphi_i = p_i$ if neuron $i$ contributes a spike to the pattern, and $\varphi_i = 1 - p_i$ if neuron $i$ contributes no spike (where $p_i$ is the firing probability per bin of neuron $i$). The expected number of occurrences per pattern $k$ is simply given by $n_{\text{exp}}^k = P_{\text{exp}} \cdot T/M$.

To improve the statistics of our estimators, we can sum the counts across the trials, if the firing rate for each of the neurons is stationary across the different trial repetitions (cross-trial stationarity). The firing probabilities of the neurons are estimated by $p_i = \frac{1}{M} \sum_{j=1}^{M} c_{i,j}$, with $c_{i,j}$ as the number of spikes of neuron $i$ in trial $j$ and $M$ the number of trials. Correspondingly, for the number of coincidences we sum the number of occurrences across trials for each pattern: $n_{\text{emp}}^k = \sum_{j=1}^{M} n_{i,j}^k$. The expected number of occurrences is computed as above, but now with the firing probabilities retrieved from summing across trials and multiplied with the number of trials considered: $n_{\text{exp}}^k = P_{\text{exp}}^k \cdot T \cdot M / h$.

Next we evaluate whether the empirical number of coincidences significantly deviates from the expected number. Therefore we test if the number of empirical coincidences is consistent with the coincidence distribution resulting from independent processes. This distribution can be expressed analytically if the spike trains follow Poisson statistics. In this case, the distribution is described by a Poisson distribution (see Grün et al. 2002a for the derivation and Grün et al. 2003 for an

---

**Figure 1.** A: representation of simultaneous spike trains and significance estimation of coincident spike events. A: spike trains of $N$ simultaneously recorded spike trains represented in matrix form, each line corresponds to a different neuron, of total observation time $T$. Spikes are represented as 1 s in time bins of width $h$ occupied with 1 or more spikes (clipping, 0 otherwise). Right: firing probability ($p_1, \ldots, p_N$) of each of the neurons. The gray box highlights an example spike pattern across the neurons $v_{\text{det}}(t)$ detected. B: coincidence count distribution for one particular pattern $k$. The expected number of coincidences of this pattern ($n_{\text{exp}}^k$) defines the mean of this distribution. The significance value of the empirical coincidence count ($n_{\text{emp}}^k$) is expressed by the joint-$p$-value ($jp$) given by the area under the distribution from ($n_{\text{emp}}^k$) to infinity (black shaded). C: logarithmic transformation of the significance value $jp$ (horizontal axis) into the surprise measure $S$ (vertical axis). The arrow points to the surprise value corresponding to the shaded area in B ($S = 1.67, jp = 0.05$). The significance level of $1\%$ corresponds to $S = 2$. (Figure modified from Grün et al. 2002a.)
alternative derivation) with the mean defined by the expected number $n_{exp}^k$ for coincidence pattern $k$. Having the distribution available, we can calculate the significance of the empirical number of coincidences $n_{emp}^k$ as the $p$-value (here called joint-$p$-value, $j_p$), i.e., the probability of observing at least $n_{emp}^k$ coincidences (Fig. 1B). If $j_p$ is smaller than a predefined significance level $\alpha$, we can conclude excess synchrony. If $j_p$ is larger than $1 - \alpha$, we conclude significantly missing coincidences. If excess is detected for the respective coincidence pattern, we call the instantiations of the patterns UE. Because highly significant events are indicated by very small $j_p$ values, for better visualization (Fig. 1C), we logarithmically transform the $j_p$ value into the surprise measure: $S(j_p) = \log\left(\frac{1 - j_p}{j_p}\right)$ (Palm et al. 1988).

This measure is zero for no deviation from expectation, positive for more coincidences than expected, and negative for less than expected. Large values of $S$ indicate significance (e.g., significance at the 1% level results in a surprise measure of 2.0).

Temporal precision of joint spike events

One way to capture potential jitter of coincident spike events is to adjust the bin width accordingly. The choice of the bin width $w$ (in units of time steps $h$) for detecting coincidences is best if the width just catches the temporal jitter of the coincidences. Of course, this width cannot be known in advance, but predictions may be available based on the biophysical properties of the neuronal system under study. One way of optimally adjusting the bin width is by varying $w$ systematically. Using simulated data, generated by injection of coincident events in otherwise independent data (the same type of model is assumed in all sections on UE), it has been shown that the significance is largest at the optimal bin width (Fig. 2A) (Grün et al. 1999). The peak in the surprise $S$ can be understood as follows: Up to the optimal bin width, more and more of the existing coincidences are detected. At the optimal width, the maximal number is reached. For larger allowed coincidence widths, the number of coincidences only increases because of chance coincidences, thereby reducing the relative contribution of excess coincidences and consequently the significance.

However, binning has a considerable drawback: there is a high probability that coincidences are split by the bin borders such that contributing spikes fall into neighboring bins and thus are not detected as coincidences. The division of the time axis into disjoint bins (DB) can lead to a considerable loss of the originally existing coincidences of $\pm60\%$ for bin sizes equal to the temporal jitter of the coincidences (Grün et al. 1999). One way to avoid this is to leave the data on a rather fine time scale and shift the spike trains against each other up to the allowed coincidence width $b$ [multiple shift method (MS)] (Grün et al. 1999). In this case, the analytical expression for the expected number of coincidences has to be adjusted to account for the various shifts $l$, and for two parallel spike trains $i$ and $M$ trials reads: $n_{emp}^{M}\text{MS} = \sum_{i=1}^{M} \sum_{l=1}^{2b+1} p_{i=1,j=1} p_{i=2,j=l}$, where $p_{i,j,l}$ is the firing probability of neuron $i$ in trial $j$ at a shift $l$ (see Eq. 15 in Grün et al. 1999). This approach is more sensitive to excess
coincidences, and the significance at optimal coincidence width is considerably larger compared with disjoint binning. An extension of the method to coincidences in more than two parallel spike trains is implemented in NeuroXidence (Pipa et al. 2008a).

Furthermore, by systematic variation of the maximal shift, the MS method enables one to derive the typical temporal precision of excess synchronous events in experimental data (Fig. 2B) as that coincidence width that shows the largest significance. A time-resolved analysis of the latter showed that the temporal precision of coincident events can also change as a function of the behavioral demands (Riehle et al. 2000).

COPING WITH STATISTICAL FEATURES OF EXPERIMENTAL DATA

The basic approach of the outlined analysis of precise spike coincidences relies on a number of assumptions, which are typically not fulfilled in experimental neuronal data. Most obvious is the observation that firing rates change as a function of time (Fig. 3A). Another type of nonstationarity is the change of the base level of firing across trials (Fig. 3B) or latency variability of the onset of firing rate changes. The spike train interval statistics of experimental data often indicate that the data do not follow Poisson statistics (Fig. 3C). In the following, we discuss how to cope with such realistic properties of experimental data by incorporating them into the statistical test.

Nonstationarity in time

In the UE analysis, it is assumed that the firing rate of the neurons is stationary within the evaluated time interval of duration $T$. If rates change as a function of time, the average rate is not a good description of the rate profile. As a consequence, the expected number of coincidences may not be correct. The worst case scenario is the situation where the rate changes in stepwise fashion from one rate level to another. Estimating the firing rate by averaging over time leads to an underestimation of the rate in the high rate regime. This results in an expected number of coincidences lower than what is truly expected, and as a consequence, may lead to false positive (fp) results. The longer the time interval $T$ and the larger the rate step, the higher the probability for fps (see Grün et al. 2003).

The most intuitive way to treat such cases would be to cut the data into stationary pieces and perform the analysis separately in each of those. However, to find segments that are jointly stationary for all the neurons under consideration, we require methods for a reliable rate estimation and then the detection of joint-stationary regions (see also APPENDIX D in Grün et al. 2002b). Although this is an appealing idea, its premise is contingent on reliable rate estimation, which is not a trivial task (e.g., Nawrot et al. 1999; Shimazaki and Shinomoto 2007; Ventura et al. 2002). In addition, temporal segments corresponding to the joint-stationary regions would be analyzed independently from each other, and thus a smooth transition in time would not be given. For these reasons, this route has not been explored further.

An alternative idea is to slide a window of predefined width $T$ along the data and carry out the UE analysis separately in each of the windows (Grün et al. 2002b). The data from time segments of the different trials are concatenated and analyzed as one long data stretch. The result of the analysis of a window is represented at the center of the window, thereby providing a time resolved analysis tool. The empirical and expected number of coincidences per pattern $k$ may be obtained as a function of time, as well as the significance of their difference (Fig. 4).

The choice of the width of the sliding window requires a balance of the window being small enough to fulfill the requirement of quasi-stationarity, yet providing enough sample bins (number of bins in $T$ times the number of trials $M$) and number of spikes for reliable statistics (Roy et al. 2000). An extreme case would be to choose the window to be of the size of a bin, as in the joint-peristimulus time histogram (JPSTH) analysis method (Aertsen et al. 1989). This would result in a better time resolution. However, the statistics have to be performed solely on the data of a single bin from across the trials and therefore require a large number of trials (typically several hundred). Thus choosing a wider window reduces the temporal precision (because of the smoothing effect of the window) but enables one to perform the analysis with less trials.

This approach has a useful side effect. The time-resolved analysis shows potential modulation of synchrony. The time scales of the firing rates and the rates of the coincidence events can be different. Even with temporally stationary firing rates, synchrony may be modulated on a time scale of tens or hundreds of milliseconds. Thus such an approach enables one to directly relate the dynamics of synchrony to the specific behavior of the animal (Grammont and Riehle 1999, 2003; Maldonado et al. 2008; Riehle et al. 1997, 2000). The window size in that respect is optimal if the $jp$ value shows a triangle like shape under the assumption of homogeneous occurrence of synchrony within the time interval (Grün et al. 2002b). As

![Fig. 3](http://jn.physiology.org/)

**Fig. 3.** Sketch of characteristic features of experimental data (shown for 1 neuron only). A: nonstationarity of firing rates in time. Within the time interval $T$ selected for analysis, the rate changes (black curve). The time course is the same for each trial expressing cross-trial stationarity. B: cross-trial nonstationarity. The firing rates are stationary within a trial of duration $T$, but have different absolute levels in different trials (black lines). C: spike trains with non-Poisson interspike intervals (ISIs), expressed here by spike occurrences (dots) with regular time intervals. This feature holds for all trials, thereby fulfilling cross-trial stationarity with respect to rates and intervals distributions. (Modified from Grün et al. 2003.)
Nonstationarity across trials

The approaches discussed thus far are based on rate estimates obtained by averaging across trials. This is justified if firing rates are stationary across the trials. However, experimental data often violate this assumption as a consequence of, for example, a variation of the depth of the anesthesia or a change in the level of attention of the animal. In analyses that involve the evaluation of first moments only (e.g., the PSTH), this aspect is often neglected. However, for analysis approaches that involve higher statistical moments, neglecting effects of cross-trial nonstationarity may lead to false-positive results (Ben-Shaul et al. 2001; Brody 1999a,b; Grün et al. 2003; Pauluis and Baker 2000; Ventura et al. 2005b).

Grün et al. (2003) studied in detail the influence of cross-trial nonstationarity of firing rates on the significance estimation in the UE analysis. In this work, cross-trial nonstationarity is modeled by drawing one of two possible (stationary) rate levels (their distance being the degree of nonstationarity) independently for each trial and each neuron (Fig. 5A). The occupation probability of the two rate states is an additional parameter of the model. Thus in this model, the degree of nonstationarity across trials and the occupation probability of the rate levels can be varied systematically. The study shows that the rate of false positives increases with the degree of nonstationarity (Fig. 5B). However, the most effective generator of false positives turns out to be co-variation of firing rates across trials (Fig. 5C), i.e., where both neurons are in the same of the two rate states for each trial. The rate of false positives is maximal, if in ~70% of the trials the neurons are jointly in the low rate state and the rest in the high rate state. It should be noted that this result can also be interpreted in the context of nonstationarity in time by interpreting the trials as time seg-

FIG. 4. Time-resolved Unitary Events (UE) analysis of data from 2 neurons (motor cortex of behaving monkey, details in Riehle et al. 1997). A: comparison of measured, i.e., empirical (dark gray), and expected (light gray) coincidence rates calculated in sliding a windows of 100-ms duration (steps of 3 ms corresponding to bin width). The expected coincidence rate is calculated by the product of the individual firing rates obtained as averages across the trials (M = 36, Poisson model assuming quasi-stationarity of neurons). B: resulting time resolved significance level expressed by the surprise measure S. The dashed line corresponds to a significance level of 5% (top: for excess coincidences, bottom: for lacking coincidences). The deviation of the empirical from the expected is modulated in time. Pronounced deviations appear at 900, 1,200, and 1,500 ms (vertical lines; line at 0 indicates start of trial). These points in time correspond to behaviorally relevant instances in time but are absent of any stimulus input or arm movements. (Modified from Riehle et al. 1997.)

FIG. 5. Cross-trial nonstationarity of firing rates. A: two rate-state model. Top: distribution of spike rate. Two rate levels are allowed (λ1, λ2); their difference Δλ = |λ1 − λ2| defines the degree of cross-trial nonstationarity. The probability of being in λ1 is denoted by p and to be in λ2 by 1 − p. Bottom: visual appearance of cross-trial nonstationarity in the dot display (shown for 1 neuron only). Inset: distribution of coincidence counts underlying the significance test for a simulated example experiment (Δλ = 60 Hz, q = 0.7, M = 100, T = 1,000 ms, λ1 = 30 Hz, λ2 = 90 Hz). The gray histogram is the distribution of coincidence counts generated from 1,000 surrogate datasets with spike counts identical to the original experiment and homogeneously distributed random spike times. Black curves are Poisson distributions with parameters (means) nexp (thin, based on spike counts in individual trials) and nexp (thick, spike counts averaged across trials). Main panel: False positives (fp) as a function of Δλ for experiments and distributions as specified in the inset. The thick black line shows the results based on the Poisson distribution with parameter nexp. The fp rate exceeds the significance level of α = 5% for Δλ = 40 Hz (in steps of 10 Hz). The gray band indicates ±SD. The gray curve shows the results for the distribution generated by spike time randomization. The thin black line shows the result for Poisson distribution with parameter nexp (gray band: ±SD). The latter 2 distributions account for cross-trial nonstationarity; fp does not exceed α. C: coherent rate steps explain the origin of false positives. The rate composition of the 2 neurons in each trial can take 4 possible states, i.e., low-low, high-high, low-high, and high-low rate combinations. Their respective number of occurrences (n1, n2, n3, n4) in an M trial system defines the covariance of rates across trials. It can be parametrized by the relative length of the coherent rate step r = (n1 + n2)M and the relative duration of the 2 regimens in the rate step q = n1/n2. Low-high and high-low rate combinations are thereby neglected. The graphs show the fp rate (vertical) as a function of r and q, for an example system (Δλ = 70 Hz, q = 0.5, M = 10, T = 1,000 ms, α = 5%). Curves connect data (●) for M trial systems with identical r (labeled, dark gray encodes large r, large circles for r = 0.2). The larger the rate step r, the larger the fp rate. (Modified from Grün et al. 2003.)
ments. In this view, the most efficient generator of false positives is a step of coherent firing rates. The longer a coherent rate step, the more false positives are expressed. (For a related aspect, see Kass and Ventura 2006.)

The basic approach to solving the problem caused by cross-trial nonstationarity is to include the nonstationarity into the predictor. Grün et al. (2003) discussed two possible solutions. One is analytical and the other numeric, both with slightly different underlying assumptions. The analytical approach is to calculate the expected number of coincidences on a trial-by-trial basis. Therefore the expected number of coincidences is calculated on the basis of the firing probabilities \( p_{ij} \) of each of the neurons \( i \) in each of the trials \( j = 1 \ldots M \), by calculating the expected joint probability per trial \( P_j = p_{1j}p_{2j} \) and summing all these probabilities derived for all trials to obtain the total joint probability \( P_{\text{exp}} = \sum_j P_j \). The total expected number of coincidences is finally derived as \( n_{\text{exp}} = P_{\text{exp}} \cdot \frac{T}{h} \) and indicates the mean of the Poisson distribution as for the stationary case. The latter relies on the fact that the sum of Poisson distributed variables is again Poisson distributed. Hence, the assumption of neuronal Poisson spike trains is implied (see Grün et al. 2003 for a slight relaxation of this condition).

The alternative approach is to generate the coincidence distribution by surrogate data. To destroy potential spike correlation, spike times within each trial are randomized for each neuron. In this procedure, the exact spike counts per trial are additionally preserved. From such a surrogate dataset, the total number of coincidences is derived as the sum of the number of coincidences found in the various trials. Repeating this procedure many times yields the coincidence distribution that is used for estimating the significance of the empirical coincidences (significance by bootstrap) instead of the analytical Poisson distribution. However, it is important to keep in mind that this approach destroys the original spike train structure and therefore implicitly assumes Poisson spike train statistics. As shown in Grün et al. (2003), both of these procedures fully account for the nonstationarity of rates across trials and avoid false positives that may be induced by the cross-trial nonstationarity (Fig. 5B).

Alternatively, spike dithering can be used to generate surrogate data that also preserve spike counts trial-by-trial. Dithering (Gerstein 2004; Maldonado et al. 2008; Pazienti et al. 2007, 2008), in the literature also called jittering (Abeles and Gat 2001; Butts et al. 2007; Date et al. 1998; Hatsopoulos et al. 2003; Jones et al. 2004), or teetering (Shmel et al. 2006), randomly displaces each individual spike within a small time window around its original position. We prefer to use the term dithering for indicating the manipulation of the data in contrast to the temporal jitter of joint spike events inherent in the experimental data. Applying this procedure destroys the exact timing of the spikes and thus also temporal relations between spikes of simultaneously observed neurons. Pazienti et al. (2007) studied how effectively the dithering procedure destroys precise spike coincidences by comparing the number of observed coincidences before and after dithering. Two different methods in use for coincidence detection are compared: disjunct binning and multiple shift (Fig. 6A). As expected, the larger the dither width, the more coincidences are destroyed, and the fewer still detected. However, the decay rate of the destruction depends on the details of the coincidence detection method. Dithering in combination with disjunct binning is most effective in destroying coincidences and most effective if both neurons are dithered compared with only one. However, dithering in combination with the multiple shift detection method leads to a very slow destruction rate (Fig. 6B). As discussed in section Temporal precision of joint spike events

![FIG. 6. Effectiveness of dithering in destroying coincident spike events. A: spike dithering and 2 methods of coincidence detection. Filled bins indicate spike occurrence; the width of the bins indicates the time resolution \( h \) (typically 1 ms). Top: generation of surrogate data. Original simultaneous spike data of neurons 1 and 2 are shown as gray filled bins. Coincidences are assumed to be precise (i.e., no jitter, within the same bin). In surrogate spike trains (black filled bins) all original spikes are independently dithered with uniform probability in the range \( \pm s \) (in units of \( h \)). Middle: in the disjunct binning (DB) method, coincidences are detected in exclusive windows of width \( w \) to allow a temporal jitter of the spikes. Only spikes within the same window (between thick vertical lines) are counted as a coincidence. Bottom: in the multiple shift (MS) method, spike coincidences are detected if the distance between spikes is smaller than or equal to a maximal shift defined as an a priori parameter \( b \) here: \( b = \pm 5 \). A leftward shift of the spike train 2 by \( 5h \) and a rightward shift of \( 3h \) leads to detection of both coincidences. B: probability of coincidence detection with DB and MS as a function of dither range \( s \). DB is shown for 1-neuron dithering [DB(1n), dark gray] and for 2-neuron dithering [DB(2n), black], MS for 2n (MS, light gray). Three values of coincidence width: thin curve for \( b = 0, w = 1 \); thick curve with knobs for \( b = 5, w = 6 \), and thick curves for \( b = 10, w = 11 \). (Modified from Pazienti et al. 2007.)](http://jn.physiology.org/10.1152/jn.00339.2008)
multiple shift is very robust and therefore partly compensates for the consequences of the dithering. Trivially, the larger the coincidence width allowed in the analysis, the less steep is the decay. Interestingly, the intuition that a dither of the same size as the allowed coincidence width would lead to a destruction rate of 50% of the coincidences does not hold; it is less. The exact decay rate depends on the details of the coincidence detection method. Maldonado et al. (2008) showed existence of precise synchronization in data from V1 of a free viewing monkey by systematically increasing the dither width and by the observation of a steady decay of excess correlation. The decay rate corresponds fully to what is theoretically expected given the MS coincidence detection method (Pazienti et al. 2008).

Generating the coincidence distribution by trial shuffling is not appropriate in the case of cross-trial nonstationarity of firing rates. Because the order of the trials of the various neurons are randomized (trial shuffling), new trial combinations across the neurons are created. For example, originally co-varying firing rates of the neurons across trials are destroyed, and as a result, the mean of the coincidence distribution is smaller and is shape is narrower. As a consequence, even coincidences that occurred by chance may be indicated as significant and thus constitute false positives. Conversely, in data sets without rate co-variation, it can be generated by trial shuffling with the consequence that existing spike correlation may be overlooked (see Grün et al. 2003).

Let us mention that latency variability, i.e., the variability of the onset of rate changes across trials, can in some cases be considered as combinations of nonstationarity in time and across trials. Solutions are discussed and provided by combining the methods discussed for the two types of nonstationarities in this and previous reviews (Baker and Gerstein 2001; Nawrot et al. 2003; Pauluis and Baker 2000; Ventura 2004). In the case of slowly varying firing rates, dithering can handle cross-trial nonstationarity, including latency variability. Another approach is to align the data properly, i.e., at the onset of the rate changes, if the experimental context still permits a meaningful interpretation (Grün et al. 2002b; Nawrot et al. 2003).

Deviation from Poisson

Neuronal spike trains typically deviate from Poisson interval statistics. This poses the question to what extent correlation analyses are affected by the properties of the individual processes. For the framework of the UE analysis, this question has recently been studied for renewal processes (Pipa et al. 2008b). In particular, it has been studied whether an error in the assumed process type (here: Poisson) influences the significance estimation of spike coincidences of non-Poisson spike trains. Spike trains are modeled as renewal processes with interspike intervals (ISIs) drawn from gamma- or log-normal distributions, currently considered to be reasonable models for experimental spike trains (see references in Tuckwell 1988 and Nawrot et al. 2008). Figure 7A shows dot displays of example realizations of a gamma process for three parameter settings. The processes are parametrized by their CV ($C_v$), i.e., the SD of the ISI distribution divided by its mean. Processes with higher ISI variability than Poisson ($C_v = 1$) have a $C_v > 1$, and processes with more regular ISIs than Poisson have values $< 1$. The coincidence distribution for such processes can be generated by repeated simulations of such processes and counting the occurrence of coincident spike events. Figure 7B shows that the shape of the distribution changes dramatically with the $C_v$. As a consequence, the significance levels of coincident spike events are altered if these distributions are directly compared (Pipa et al. 2008b).

However, the UE analysis is in large parts unaffected if the spike trains have gamma- or log-normal renewal statistics. Only for very regular processes ($C_v \ll 1$), the UE analysis leads to an increased number of false positives, but the significance of coincident events of processes with moderate $C_v$ ($< 1$) or bursty processes ($C_v > 1$) tends to be underestimated (Fig. 7C). The reason that the UE analysis does not generate false positives for bursty processes, is two fold. First, it operates on binned and clipped spike trains, which considerably reduces the burstiness and leads to a Poisson like coincidence distribution. Second, the UE analysis adjusts the mean of the distribution used for the significance evaluation according to the spike counts in the data under evaluation.

Interestingly, most physiological data have a $C_v > 0.2$ and $\leq 1$ (Nawrot et al. 2008), leading to the conclusion that when Poisson processes are wrongly assumed this leads to a more conservative test but not to fps. A further study extended to non-renewal processes with first-order negative serial correlation, as found in experimental data (Nawrot et al. 2008), shows basically the same general results, but the fps rate as a function of the $C_v$ has a higher modulation depth (Grün et al. 2008b).

---

**FIG. 7.** Impact of non-Poisson spike trains on significance estimation. A: dot displays of 3 different examples of realizations of gamma processes (left to right: $C_v = 3$; $C_v = 1$ (Poisson); $C_v = 0.1$). B: comparison of coincidence distributions from 2 parallel spike trains (dashed: Poisson; gray: gamma process with $C_v = 0.1$; firing rates are 50 Hz; $M = 10^5$ trials of 5-s duration, bin width of $w = 4$ ms, no clipping). Both processes are parameterized by the product of bin width and firing rate, here 0.1. Probability of false positives (UE analysis, Poisson assumption; significance level: 1%) for coincidences detected in pairs of independent gamma- (solid, gray) or log-normal (black dashed) processes (rates were $\lambda = 50$ Hz, 10,000 realizations of 5 s for each $C_v$).

---

*J Neurophysiol* • VOL 101 • MARCH 2009 • www.jn.org
Taken together, these studies show that the structure of the spike trains influences the significance evaluation of coincident spike events. If Poisson statistics are assumed, significance may be under- or overestimated depending on the $C_v$ of the analyzed processes. Obviously, it is preferable to use the proper coincidence distribution for the significance test. One solution would be to model the simultaneous processes as independent stochastic processes and to realize numerically the coincidence distribution on this basis. However, this requires knowledge about the statistics of the measured data for a proper model selection, which is not a trivial task because the parameter estimation is often confounded with changes in the firing rates (see General Discussion). Another way to create the proper distribution without having to assume a certain point process model is to generate the distribution directly from the measured data by random trial shuffling (Ito 2007; Pipa and Grün 2003).

The shuffling procedure for the UE analysis is quite time consuming, because it has to be performed for each position of the sliding window and for all possible coincidence patterns therein. One efficient way to generate the distribution without the need to generate all possible shuffles is to resample counts resulting from a subset of trial shuffles (Pipa and Grün 2003). Further reduction of computation time is achieved by generating only the part of the distribution which is relevant for the significance estimation (the tail) instead of the complete distribution (Pipa et al. 2003).

However, as already mentioned, trial shuffling requires data to be stationary across trials, and a violation of this assumption may have serious consequences. A variant of the UE analysis, called NeuroXidence (Pipa et al. 2008a), suggests another solution. The method also follows the approach to use the original spike trains for generation of the null hypothesis. However, instead of trial shuffling, the trial order of the measured spike trains is preserved, but the spike trains are shifted against each other by a small amount of time. Repeated random shifts smaller than the time scale of the modulation of the firing rate generate a set of coincidence counts where correlated spike events are destroyed. Significance of the empirical number of coincidences is evaluated by comparison with the counts resulting from the shifted versions using a $t$-test. Analysis for significance of joint-spike events of experimental data from motor cortex of monkey with a moderate $C_v$ (0.74 and 0.95) confirmed the results derived using other predictors with slightly higher significance (Pipa et al. 2007), which may reflect the fact that the method is accounting for the ISI properties of the spike trains.

**Preprocessing of Data**

Identifying neuronal groups that are currently part of an active assembly assumes that the analyzed spike trains reflect single unit data. However, a number of processing steps are involved to extract such data. One important step is to extract the single unit activities that contribute to the measured extracellular signal by spike sorting. Although spike sorters are under continuous development for the last 20 years or so, sorting remains a difficult and error prone procedure (Harris et al. 2000; Lewicki 1999; see references in Pazienti et al. 2006). Sorting methods are typically not fully automated but still require human assistance for proper visual selection of spike shapes or for clustering similar shapes; i.e., individuals follow different strategies. One researcher may prefer to include all spikes in an effort not to loose any spikes and, thereby, occasionally include spikes from other neurons; others want to make certain to only group the spikes of nearly identical shapes and with a high probability of being from only one neuron and thereby sometimes miss spikes from a neuron. The former would lead to false-positive errors (FP) and the latter to false-negative errors (FN), both of which also occur in almost all automated sorting procedures (Harris et al. 2000).

We need to understand to what degree such misses or insertions of spikes influence the results of correlation analysis. More specific questions come to mind: do falsely included spikes generate false positive correlation? Did missed spikes lead to an increase or decrease of the estimated correlation? Pazienti and Grün (2006) studied these questions in the framework of the UE analysis; others have treated similar questions in respect to cross-correlation analysis (Bar-Gad et al. 2001; Bedenbaugh and Gerstein 1997; Gerstein 2000).

Pazienti et al. (2006) explored the impact of sorting errors on the results of the UE analysis method and found that FN and FP sorting errors always reduce significance in the context of pairwise analysis (Fig. 8). However, FN errors lead to a stronger reduction of significance than FP errors. This also holds for the cross-correlation analysis (Bedenbaugh and Gerstein 1997; Gerstein 2000). If more than two parallel processes are analyzed, the picture becomes more complicated (Pazienti and Grün 2007). Spikes deleted from original patterns because of FN errors result in a decrease of the complexity (number of spikes) of the pattern; FP errors may enhance pattern complexity. However, the probability of the latter is very small compared with the former. Thus spike sorting errors interfere with higher-order correlations and have the tendency to reduce pattern complexity and to enhance the significance of lower complex patterns. Consequently, we may assume that the size of an active assembly is typically larger than detected.

**General Discussion**

We showed for the example of the UE analysis that spike correlation analysis is sensitive to detailed features of the spike train statistics of the single neuron. Analysis approaches that include incorrect assumptions about the data may lead to false-positive or false-negative results. For example, failure to consider various types of nonstationarity can result in the detection of significant spike correlation, although these only occur at chance level. This not only holds for the UE analysis technique but has also been reported for other types of spike correlation analyses (Baker and Gerstein 2000, 2001; Brody 1999a,b; Gerstein 2004; Oram et al. 1999; Richmond et al. 1999; Staude et al. 2008; Tetzlaff et al. 2008; Ventura 2005a,b). Thus analysis of precise spike correlation requires careful consideration of the statistical features of the experimental data and consequently a proper choice of procedures for estimating the significance of the correlation.

The key to avoiding false-positive results lies in the proper formulation of the null hypothesis (McLelland et al. 2007; Richmond 1999). Therefore all features of the experimental data not of interest for the question at hand should
be considered and included in the null hypothesis. The complexity of the experimental data often prevents the use of analytical approaches for statistical tests or parametric testing. However, as shown for the UE method, analytical treatment is possible under strict assumptions regarding the processes. In cases where an analytical approach is not possible, numerical methods are nowadays used given the recent increase in available computer power (Fellous et al. 2004; Fujisawa et al. 2008; Stark and Abeles 2005). In this context, the generation of artificial, surrogate data is an established method. A number of examples were discussed in the context of the UE method. Surrogate data are also used for significance estimation of occurrence of other specific features, e.g., spike-pair coincidences (Fujisawa et al. 2008; Shmiel et al. 2006; Ventura et al. 2005a,b), coincidence patterns (Grün et al. 2003; Pipa and Grün 2003; Pipa et al. 2003), or spatio-temporal patterns (Abeles and Gat 2001; Baker and Gerstein 2000; Baker and Lemon 2000; Gerstein 2004). The general strategy is to design surrogate data such that the most prominent (best would be all) features of the experimental data are conserved, but the features to be studied specifically are destroyed. In the case of precise joint spike events, the interest is to keep the statistical features of the spike trains of the individual neurons as similar to the original as possible but to destroy the exact relative timing of spikes from different neurons.

Table 1 provides a list of approaches used for such a task divided into modeling-based approaches and data-based approaches. It also indicates the appropriateness of each approach in accounting for the features of the data discussed above. Solutions suggested in the framework of the UE analysis are marked in gray. Furthermore, Table 1 lists the assumptions of each of the methods, as well as the features they conserve and those they destroy. Methods differ in their ability to destroy features. One could try to characterize the degree of destruction by an objective measure, like the similarity measure for spike trains developed by Victor and Purpura (1996). However, such a characterization would not give us insight into the specific feature of the data that is problematic in the respective analysis. Therefore we characterize each by commonly used measures in neuroscience, divided into measures relevant for single spike trains and for parallel spike trains. For single neurons, we consider the average rate across time and trials: the time-dependent estimate of the firing rate as measured by the PSTH, the ISI distribution, the spike count per trial (SpC), and average spike count across trials (tot SpC). For neurons in parallel, we consider co-variation of firing rate (co-var rate) or spike counts (co-var SpC) across neurons within trials and the population histogram (POPH) as a measure of the number of spikes across neurons in small time bins (per trial) or averaged across trials (tot POPH). The common goal of all approaches is to detect precisely correlated spiking activity across the neurons; therefore they all destroy—in different ways—the exact spike timing relation across the neurons. Some are able to conserve the individual spike train structure (sp-tr
The methods’ abilities to account for prominent features of experimental data (section 3) are marked \([X]\): perfect conservation of measures; \((X)\) approximate conservation of measures, i.e., conservation in the statistical sense). Approaches that were illustrated in the framework of the UE method are marked in gray. Further are listed the assumptions of the approach, the measures they conserve, and those they destroy. These aspects are further divided into considerations in respect to single neurons or to parallel neuron. SpC, spike count; non-stat, non-stationary; ISI, interspike interval distribution; cr-tr, cross-trial; co-var, co-variation; sp-tr struc, spike train structure; entries commented by ‘tot’ means that the respective measure is averaged over all trials, otherwise the measures are meant to be estimated from individual trials; POPH, population histogram; PSTH, peristimulus time histogram. Note, the latter 2 are conceptually different measures: conservation of the POPH means that the exact spike sum per bin across neurons is preserved. In contrast conservation of the PSTH means that the time course of the firing rate is statistically preserved but not necessarily the exact spike sum per bin across trials.

The methods’ abilities to account for prominent features of experimental data (section 3) are marked \([X]\): perfect conservation of measures; \((X)\) approximate conservation of measures, i.e., conservation in the statistical sense). Approaches that were illustrated in the framework of the UE method are marked in gray. Further are listed the assumptions of the approach, the measures they conserve, and those they destroy. These aspects are further divided into considerations in respect to single neurons or to parallel neuron. SpC, spike count; non-stat, non-stationary; ISI, interspike interval distribution; cr-tr, cross-trial; co-var, co-variation; sp-tr struc, spike train structure; entries commented by ‘tot’ means that the respective measure is averaged over all trials, otherwise the measures are meant to be estimated from individual trials; POPH, population histogram; PSTH, peristimulus time histogram. Note, the latter 2 are conceptually different measures: conservation of the POPH means that the exact spike sum per bin across neurons is preserved. In contrast conservation of the PSTH means that the time course of the firing rate is statistically preserved but not necessarily the exact spike sum per bin across trials. The methods’ abilities to account for prominent features of experimental data (section 3) are marked \([X]\): perfect conservation of measures; \((X)\) approximate conservation of measures, i.e., conservation in the statistical sense). Approaches that were illustrated in the framework of the UE method are marked in gray. Further are listed the assumptions of the approach, the measures they conserve, and those they destroy. These aspects are further divided into considerations in respect to single neurons or to parallel neuron. SpC, spike count; non-stat, non-stationary; ISI, interspike interval distribution; cr-tr, cross-trial; co-var, co-variation; sp-tr struc, spike train structure; entries commented by ‘tot’ means that the respective measure is averaged over all trials, otherwise the measures are meant to be estimated from individual trials; POPH, population histogram; PSTH, peristimulus time histogram. Note, the latter 2 are conceptually different measures: conservation of the POPH means that the exact spike sum per bin across neurons is preserved. In contrast conservation of the PSTH means that the time course of the firing rate is statistically preserved but not necessarily the exact spike sum per bin across trials.

**Table 1. Methods for implementing the null-hypothesis, divided into model based and data based**

<table>
<thead>
<tr>
<th>Method</th>
<th>Problems accounted for</th>
<th>Assumptions</th>
<th>Conserves</th>
<th>Destroys</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Homogeneous Poisson process</td>
<td>(X)</td>
<td>Poisson stat</td>
<td>average rate</td>
</tr>
<tr>
<td>B</td>
<td>Homogeneous model process</td>
<td>(X)</td>
<td>Model process stat</td>
<td>average rate</td>
</tr>
<tr>
<td>C</td>
<td>Inhomogeneous Poisson process</td>
<td>(X)</td>
<td>Poisson Non-stat</td>
<td>PSTH</td>
</tr>
<tr>
<td>D</td>
<td>Inhomogeneous model process</td>
<td>(X)</td>
<td>Model process Non-stat</td>
<td>PSTH</td>
</tr>
<tr>
<td>E</td>
<td>Spike time randomization (within single trials)</td>
<td>X</td>
<td>Poisson stat</td>
<td>SpC</td>
</tr>
<tr>
<td>F</td>
<td>ISI shuffling (within single trials)</td>
<td>X</td>
<td>Renewal stat rate</td>
<td>SpC</td>
</tr>
<tr>
<td>G</td>
<td>ISI shuffling (across trials)</td>
<td>X</td>
<td>Renewal stat rate cr-tr stat</td>
<td>tot SpC</td>
</tr>
<tr>
<td>H</td>
<td>Trial shuffling</td>
<td>X</td>
<td>cr-tr stat</td>
<td>PSTH</td>
</tr>
<tr>
<td>I</td>
<td>Shift predictor</td>
<td>X</td>
<td>Short-term cr-tr stat</td>
<td>PSTH</td>
</tr>
<tr>
<td>J</td>
<td>Spike shuffling across neurons (within trials)</td>
<td>X</td>
<td>cr-tr stat</td>
<td>PSTH</td>
</tr>
<tr>
<td>K</td>
<td>Spike exchange across neurons (within trials)</td>
<td>X</td>
<td>st rate SpC co-var Poisson</td>
<td>PSTH</td>
</tr>
<tr>
<td>L</td>
<td>Dithering</td>
<td>X</td>
<td>sp-tr struc (approx)</td>
<td>PSTH (smoothed)</td>
</tr>
<tr>
<td>M</td>
<td>Soft dithering (according to ISI distribution)</td>
<td>X</td>
<td>sp-tr struc (approx)</td>
<td>PSTH (smoothed)</td>
</tr>
<tr>
<td>N</td>
<td>Shifting spike trains of neurons against each other</td>
<td>X</td>
<td>Sp-tr struc</td>
<td>PSTH</td>
</tr>
</tbody>
</table>

struc), others conserve the POPH or only the tot POPH, and others can only conserve the co-variation of the firing rates. In the following, we will briefly relate these approaches to various studies reported in the literature and their respective use and purposes.

Modeling spike trains as stochastic processes (model based) is a very elegant way of generating surrogate data. Because of their simplicity, homogeneous processes are often assumed (Table 1, A and B), which (applied in sliding windows) enables treatment of nonstationary data (Grün et al. 2002b; see section above for an analytical version of such a model based approach). Inhomogeneous Poisson processes that reproduce the dynamics of the firing rate according to the estimated firing rate profiles are also often used (Roxin et al. 2008) but rely on a proper estimate of the time-dependent firing rate, which is hard to derive (Nawrot et al. 1999; Shimazaki and Shinomoto 2007; Ventura et al. 2002). More complicated processes that reproduce the ISIs, e.g., gamma- or log-normal processes or processes that also reproduce higher order interval statistics (Table 1, C and D), are also used (Farkhooi et al. 2008). However, these methods rely on proper estimates of additional model parameters. This is not trivial because, for example, the interval statistics of experimental data are often confounded by changes of the firing rates (Baker and Lemon 2000; Johnson 1996). This problem may be solved by time-rescaling approaches that make the firing stationary before deriving the ISI distribution and estimating parameters (Brown et al. 2002; Nawrot et al. 1999, 2008). However, this relies on the assumption that the process parameters do not change in time. If they do, parameters have to be estimated in a time-dependent manner. Even more complex models, e.g., which change their internal structure as a function of time, are becoming to become available (Baker and Lemon 2000; Brown et al. 1998; Johnson 1996; for a recent review, see Itskov et al. 2008; Kass et al. 2005 and references therein; Truccolo et al. 2005). Although it is still not possible to estimate all parameters from the experimental data, a more complete characterization of neuronal spiking dynamics (stochastic model) is helpful be-
cause it enables a more precise description of the assumptions. Consequently, more specific statements about the experimental data are possible.

Other methods operate directly on the original data (data based) and modify the data according to specific rules (see Fig. 9 for illustrations of the respective approaches). Most of these methods belong to the class of resampling procedures, whereas others are rather heuristic approaches. Spike time randomization (Table 1E), i.e., random replacement of the spike times of a neuron within a given time interval, assumes stationary firing rate therein. If applied in a trial-by-trial manner, the method accounts for cross-trial nonstationarity (Grün et al. 2003). The procedure is closely related to the generation of a stationary Poisson process, but here the spike counts are additionally matched. However, the ISI structure of the original data are destroyed. On the contrary, other methods particularly focus on the conservation of the ISI distribution by shuffling the ISIs of a neuron (Table 1F), thereby preserving the spike count in each trial (Ikegaya et al. 2004; Masuda and Aihara 2003; Nadasdy et al. 1999). Potential co-variation of spike counts across neurons is therefore automatically preserved. In attempts to additionally account for nonstationarity in time, such procedures may be applied in shorter time windows, e.g., in a sliding window fashion, to also approximate time-dependent changes. For the case

![Diagrams](http://jn.physiology.org/)

of ISI shuffling, there may be not enough ISIs for shuffling within a time window and therefore shuffling may be across trials (Table 1G), which, however, assumes cross-trial stationarity.

Trial shuffling (Table 1H) is a well-established way of destroying the relationship between simultaneously recorded spike trains. For each neuron, the sequence of trials is randomly reordered, thereby destroying the trial relations across neurons and thus potential spike correlation (Gerstein and Perkel 1972). The internal structure of the spike trains, including time dependent variations of parameters, is left intact. Potential co-variation of spike counts across the neurons is destroyed. The drawback of the method is that nonstationarity across trials (with respect to rate or other properties of the spike trains) cannot be tolerated, because a recombination of these features, e.g., rate, may additionally change the chance occurrence of joint spike events. For the shift predictor (Table 1I), spike trains are not randomly assigned, but trials are shifted systematically against each other (Gerstein and Perkel 1972). The shift predictor is mainly of historical interest as an approximation of the full shuffle. However, in the presence of long-term nonstationarity across trials, the shift predictor may still be useful.

Some methods for the evaluation of spatio-temporal spike patterns follow a different approach, where surrogates are not generated by separately modifying the spike trains of individual neurons but by shuffling spikes across neurons within the same trial (Table 1J) (Ikegaya et al. 2004; Nadasdy et al. 1999) or by exchanging spikes across neurons (Table 1K) (Ikegaya et al. 2004). The latter preserves the population histogram, but neither of them preserves the PSTHs or the ISI distributions, and they assume co-stationary of the data. None of the properties of experimental data listed in Table 1 (Problems) is accounted for.

Spike dithering (Table 1L; also known as jittering or teetering) destroys the exact timing of individual spikes and therefore also the exact relative spike timing across neurons (Date et al. 1998; Nadasdy et al. 1999 (1-sided dither); Abeles and Gat 2001; Butts et al. 2007; Gerstein 2004; Hatsopoulos et al. 2003; Jones et al. 2004; Maldonado et al. 2008; Pazienti et al. 2007, 2008; Shmiel et al. 2006). This treatment effectively leads to a smoothing of the firing rates but also to a slight modification of the ISI distribution. Nevertheless, dithering is currently accepted as one of the best methods for the generation of surrogates because it simultaneously preserves many features of the spike trains with relatively high accuracy—the spike train structure, and thus the ISI distribution, the spike counts per trial, the firing rate profile, and trivially all of these features simultaneously for the neurons. Obviously, the effectiveness of the destruction of spike correlation depends not only on the chosen bounds for the dithering but also on the exact method used to detect correlated spike events (Maldonado et al. 2008; Pazienti et al. 2007, 2008). Another version of the dithering method that preserves the ISI distribution very well applies random displacement of spikes according to a two-dimensional ISI probability distribution estimated from the data (Table 1L) (Gerstein 2004).

Instead of modifying the spike trains, it was suggested (Harrison and Geman 2008; Pipa et al. 2008a) to shift the whole spike trains against the others by random amounts (Table 1N). Repeated random shifts generate a set of coincidence counts where correlated spike events are destroyed. Significance of the empirical number of coincidences is evaluated by comparison with the counts resulting from the shifted versions. Thereby—in contrast to the spike dither methods—the structure of the spike train is conserved, as well as co-varying features. The maximal shift is chosen larger than the temporal precision of the joint spike events to ensure effective destruction of the latter. This smooths the PSTHs.

In summary, surrogate methods can reproduce several aspects of the statistical features of the data, but none are able to cope with all features perfectly. A number of methods cannot intrinsically deal with nonstationarity of the firing rate in time (Table 1, A, B, and E–G) but can still be reasonably well applied in a sliding window approach (Grünewald et al. 2002b). However, dithering methods in their respective variants seem currently the most appropriate methods for the generation of proper surrogates, because of their abilities to account for all required properties almost perfectly.

The creation of surrogate data is an appealing way to generate control data and conceptually simple. However, for a proper choice of the surrogate, one has to be very clear about the implicit assumptions and about the features destroyed. Both may also have an impact on the test one wants to perform. This implies that the choice of surrogate data may depend on the data analysis method used. We can expect that such mutual influence also occurs with other analysis methods, e.g., those that focus on temporally delayed spike patterns (Prut et al. 1998). Thus surrogates have to be considered and tested in the context of the analysis method intended to be used. Therefore we urge one to thoroughly test and calibrate analysis techniques before applying them to experimental data. In contrast, testing a method by applying it to experimental data is only of limited help. For calibration, we need full control over the test data. Statistical models of parallel spike trains that contain a number of relevant features of spike trains but also the correlation structure to be tested are available (Baker and Gerstein 2000; Brette 2008; Grünewald et al. 1999, 2002ab, 2008; Kuhn et al. 2003; Macke et al. 2008; Niebur 2007; Staude et al. 2007).

We emphasize a similar concern with respect to the impact of data preprocessing. As discussed above, spike sorting can have a major impact on the results of the UE analysis and on the conclusions drawn. This was also shown in the context of cross-correlation analysis of data with a pronounced autostructure (Bar-Gad et al. 2001). These authors showed that, when combined with sorting failure in the detection of overlapping spikes, the shape of the cross-correlation function can be completely altered. In their case, zero- and near-zero delay coincidences were lacking, but delayed coincidences were emphasized. This may also be taken as evidence that the effects discussed here for coincident spike events have even more complex aspects to be considered if spike patterns with temporal delays are analyzed.

OUTLOOK AND CHALLENGES

Presently, control data implement the null hypothesis of (full) independence of the simultaneous spike trains. However, even in the infancy of theoretical neuroscience, researchers sought to detect correlated subgroups of neurons for the identification of neuronal assemblies (Abeles 1991; Abeles et al.
It is plausible that the mechanisms underlying neural processing are still not well understood because of systematic undersampling of the neuronal system. Therefore we have to considerably increase the number of simultaneously recorded single neurons. This constitutes not only a technical challenge for the experimentalists but also a challenge for developing reliable analysis of such data, in particular for the analysis of spike correlation (see Brown et al. 2004 for a recent review). Simply following along the lines of existing methods would imply a combinatorial explosion of parameters and conditions. Thus new perspectives are required (Berger et al. 2007; Eldawlatly et al. 2008; Grün et al. 2008a; Staude et al. 2007; Stuart et al. 2002). Despite the ever increasing computer power, it will always be hard to handle such analysis and, in particular, outcomes, because the results must be distilled and compiled into dimensions that the human brain is capable of understanding. New methods are expected to be most successful if their statistics are based on clear hypotheses based on low-parameter (network) models. Biophysical models of single neurons or simple network structures are already used to explain certain aspects of experimental data (Galan et al. 2008; Tetzlaff et al. 2008; Tiesinga et al. 2008). Furthermore, it is conceivable that, in the near future, neuronal network models will improve to the extent that they serve as fairly complete descriptions of the biological system. Hopefully, future analysis methods can be calibrated by data resulting from biologically realistic network models (Schrader et al. 2008) that naturally implement such hypotheses.

ACKNOWLEDGMENTS

I thank S. N. Baker for organizing the Engineering and Physical Sciences Research Council–funded Newcastle workshop on Spike Train Analysis 2006, which inspired this review; G. Pipa and S. Louis for support in the preparation of part of the figures; B. Staude and M. Denker for comments on the manuscript; and M. Diesmann for useful discussions and comments on an earlier version of the manuscript.

REFERENCES


Amari S. Measure of correlation orthogonal to change in firing rate. Neural Comput In press.


It is plausible that the mechanisms underlying neural processing are still not well understood because of systematic undersampling of the neuronal system. Therefore we have to considerably increase the number of simultaneously recorded single neurons. This constitutes not only a technical challenge for the experimentalists but also a challenge for developing reliable analysis of such data, in particular for the analysis of spike correlation (see Brown et al. 2004 for a recent review). Simply following along the lines of existing methods would imply a combinatorial explosion of parameters and conditions. Thus new perspectives are required (Berger et al. 2007; Eldawlatly et al. 2008; Grün et al. 2008a; Staude et al. 2007; Stuart et al. 2002). Despite the ever increasing computer power, it will always be hard to handle such analysis and, in particular, outcomes, because the results must be distilled and compiled into dimensions that the human brain is capable of understanding. New methods are expected to be most successful if their statistics are based on clear hypotheses based on low-parameter (network) models. Biophysical models of single neurons or simple network structures are already used to explain certain aspects of experimental data (Galan et al. 2008; Tetzlaff et al. 2008; Tiesinga et al. 2008). Furthermore, it is conceivable that, in the near future, neuronal network models will improve to the extent that they serve as fairly complete descriptions of the biological system. Hopefully, future analysis methods can be calibrated by data resulting from biologically realistic network models (Schrader et al. 2008) that naturally implement such hypotheses.
SIGNIFICANCE ESTIMATION FOR PRECISE SPIKE CORRELATION


