

# Impact of Higher-Order Correlations on Coincidence Distributions of Massively Parallel Data

Sonja Grün<sup>1,\*</sup>, Moshe Abeles<sup>2</sup>, Markus Diesmann<sup>1</sup>

<sup>1</sup>Theoretical Neuroscience Group, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-Shi, 351-0198 Saitama, Japan

<sup>2</sup>Gonda Brain Research Center, Bar Ilan University, Ramat Gan 52900, Israel  
\*corresponding author, gruen@brain.riken.jp

Aug 8th, 2008

## Abstract

The signature of neuronal assemblies is the higher-order correlation structure of the spiking activity of the participating neurons. Due to the rapid progress in recording technology the massively parallel data required to search for such signatures are now becoming available. However, existing statistical analysis tools are severely limited by the combinatorial explosion in the number of spike patterns to be considered. Therefore, population measures need to be constructed reducing the number of tests and the recording time required, potentially for the price of being able to answer only a restricted set of questions.

Here we investigate the population histogram of the time course of neuronal activity as the simplest example. The amplitude distribution of this histogram is called the complexity distribution. Independent of neuron identity it describes the probability to observe a particular number of synchronous spikes.

On the basis of two models we illustrate that in the presence of higher-order correlations already the complexity distribution exhibits characteristic deviations from expectation. The distribution reflects the presence of correlation of a given order in the data near the corresponding complexity. However, depending on the details of the model also the regime of low complexities may be perturbed.

In conclusion we propose that, for certain research questions, new statistical tools can overcome the problems caused by the combinatorial explosion in massively parallel recordings by evaluating features of the complexity distribution.

*Key words and phrases: spike synchronization, higher-order synchrony, massively, parallel spike trains*

# 1 Introduction

Following the hypothesis that assembly activity is expressed by temporal relations between the spiking activity of the participating neurons, neuronal responses need to be observed and analyzed with respect to temporal structure. With massively parallel recordings becoming available chances to observe the signature of assembly activity are increasing and indeed the availability of massively parallel spike data is escalating rapidly (e.g. Nicolelis et al., 1997; Csicsvari et al., 2003; Ikegaya et al., 2004). However, at this point in time we lack the corresponding analysis tools (Brown et al., 2004). Most of the existing methods are based on pairwise analysis (e.g. Aertsen et al., 1989; Nowak et al., 1995; Kohn & Smith, 2005; Shmiel et al., 2006), approaches to analyze correlations between more than two neurons do exist but typically work only for a small number of neurons (e.g. Abeles & Gerstein, 1988; Dayhoff & Gerstein, 1983; Grün et al., 2002a,b) or consider pair correlations only while analyzing the ensemble (e.g. Gerstein et al., 1985; Shlens et al., 2006; Schneidman et al., 2006). To extend existing methods designed to work on small number of neurons to massively parallel data is generally not feasible. One reason is that these methods typically assess individual spike patterns, e.g. coincidences with an identification of the participating neurons, or spatio-temporal spike pattern. An extension to many neurons would lead to a combinatorial explosion. This particularly holds for methods which include significance tests that do not only test against full independence, but detect higher-order correlations (e.g. Martignon et al., 1995; Nakahara & Amari, 2002; Schneider & Grün, 2003; Gütig et al., 2003; Ehm et al., 2007). Additional complications are the limited number of samples in experimental data, in particular if data are non-stationary. Only a few approaches exist that can handle and analyze massively parallel data for higher-order correlations. These approaches are based either on the model assumption of synfire chains (Schrader et al., 2008) or on compound Poisson processes (Staude et al., 2007, 2008).

Here we aim at a fast screening method that can detect correlation within massively parallel spike data. We base our approach on the distribution of the sum of spikes across

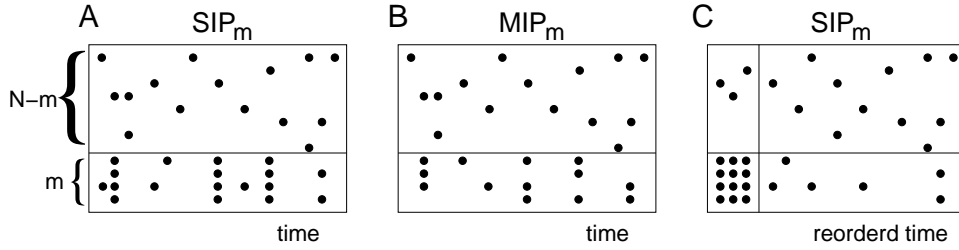


Figure 1: Sketches of the models used for generating parallel spike data with higher-order spike coincidences in  $m$  out of  $N$  processes (A:  $SIP_m$  model, B  $MIP_m$  model). Rate  $p$  of the neurons is the same in both models and also for the  $m$  correlated and  $N - m$  uncorrelated neurons. The bottom  $m$  neurons in the  $SIP_m$  model contain coincident events at rate  $\alpha$  involving all  $m$  neurons. These neurons also contain uncorrelated background spikes at rate  $p - \alpha$ . In the  $MIP_m$  model the  $m$  neurons typically contain coincidences of lower order than  $m$  depending on the copy probability  $\epsilon$ , and do not contain background spikes. The rate of insertion is  $\alpha = p/\epsilon$ . (C) shows the  $SIP_m$  model with reordered time bins. In the left part the bottom  $m$  neurons contain coincidences of order  $m$  in all time steps exhibiting the dependent fraction, the right part shows the independent fraction of the time bins.

neurons as reflected in the population histogram. In particular we explore how coincidence patterns of higher-order are reflected in this measure and if the order of the correlation can be identified. The dependence on parameters relevant for experimental data are studied using numerical and analytical methods.

Preliminary results have been presented in abstract form (Grün et al., 2003).

## 2 Correlation Model

We model massively parallel spike trains as  $N$  parallel stationary processes. For the generation of correlation we use slightly modified versions of two types of models, recently published by Kuhn et al. (2003). The basic idea underlying both models is to have a hidden Poisson ‘mother’ process of rate  $\alpha$ , from which spikes are copied into parallel child processes according to a given probability  $\epsilon$ . If  $\epsilon = 1$  the model is named ‘single interaction process’ (SIP). All spikes of the mother process are present in all  $N$  child processes, such that synchronous higher-order spike events across all neurons are induced for each spike in the mother process. In case  $\epsilon < 1$  the model is called ‘multiple interaction process’ (MIP). Here, not necessarily all neurons receive a copy of the mother process’ spikes, and the neurons receiving a spike are randomly chosen out of all  $N$  processes. As a result, the

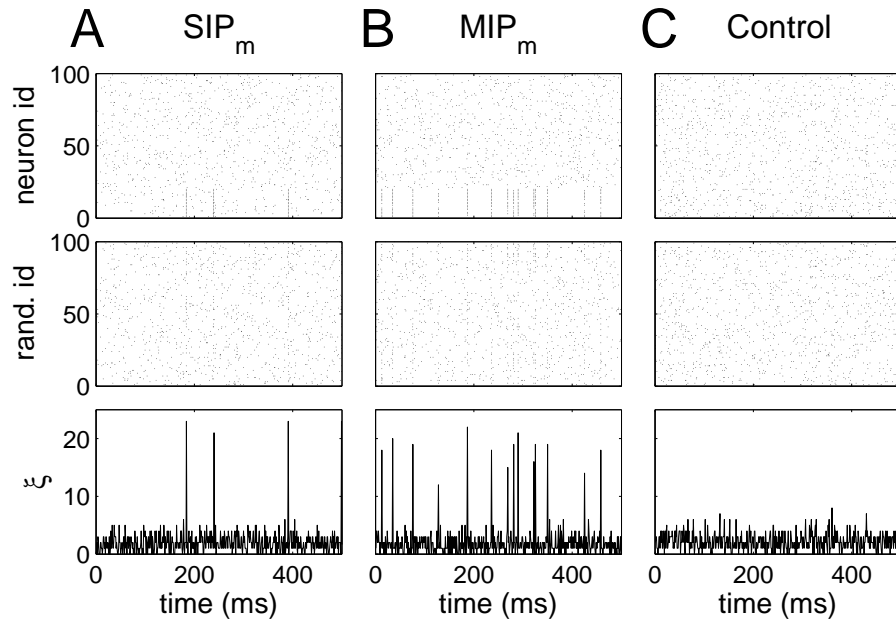


Figure 2: Dot displays of realizations of the (A)  $SIP_m$  model, (B)  $MIP_m$  model and (C) control data generated by spike time randomization of the data of the  $MIP_m$  model in (B). The parameters common to all panels are  $N = 100$ , rate  $p = 0.02$ , and  $h = 0.001s$ . The duration shown is  $T = 500ms$ . In (A) neurons  $1 \dots 20$  ( $m = 20$ ) are correlated with  $\alpha = 0.005$ , in (B) neurons  $1 \dots 20$  ( $m = 20$ ) are correlated with  $\epsilon = 0.8$ . The middle row shows the same data as in the top row with random ordering of the neuron identifiers. The bottom row shows the population histograms (bin size  $h = 0.001s$ ) identical for the top and the middle panel of each column.

generated synchronous patterns differ in composition of neurons and in their complexity, i.e. the number of spikes in the pattern.

Here, we modify the models as follows. First, instead of expressing the model as Poisson processes in continuous time we formulate it in discretized time as Bernoulli processes. The reason is that we anyway aim to detect coincident events via binning. Formulating the model in continuous time with subsequent binning would lead to additional effects not relevant for the present study (cmp. Staude et al., 2008). Second, in a realistic recording session with  $N$  electrodes we do not expect to record from the same assembly at all electrodes. Instead we expect to observe at most a subset of  $m < N$  neurons to participate in the same assembly. Therefore we copy the spikes of the mother process into  $m$  selected processes only. Third, correlated activity found in experimental data is typically embedded in uncorrelated 'background' firing activity (e.g. Riehle et al., 1997; Abeles et al., 1993; Prut et al., 1998). Therefore we 'dilute' the activity of the fully synchronized neurons with uncorrelated background spikes.

Thus, we define the following models of correlated activity.

**Single interaction process in  $m$  out of  $N$  neurons: SIP $_m$ .** From a Bernoulli process of rate  $\alpha$ , discretized in bins of width  $h$  (typically 1ms) we copy spikes with probability  $\epsilon = 1$  into  $m$  out of  $N$  single neuron channels. Into each of the  $m$  processes of the correlated neurons we in addition to the synchronized events inject uncorrelated spikes modeled by a Bernoulli process of rate  $p_b = p - \alpha$  (see Fig. 1A for a sketch of the model, and Fig. 2A for an example realization with ordered (top) and randomized (middle) neuron identifiers. Under the constraint that all  $N$  neurons have the same firing rate  $p$  we model the  $N - m$  uncorrelated processes as independent Bernoulli processes of rate  $p$ . The parameter ranges we consider are the typical firing rates of cortical neurons (from a few to  $\approx 100$ Hz) and coincidence rates up to a few Hz as extracted from experimental cortical data (Grün et al., 1999). Thus  $\alpha$  is typically small relative to the firing probability  $p$ .

**Multiple interaction process in  $m$  out of  $N$  neurons: MIP $_m$ .** From a Bernoulli process of rate  $\alpha$ , discretized in bins of width  $h$  (typically 1ms) we copy spikes with a probability  $\epsilon \leq 1$  into  $m$  out of  $N$  neurons (for  $\epsilon = 1$  this corresponds to SIP $_m$  without

background and  $\alpha = p$ ). Here we do not insert background spikes into the  $m$  processes since due to the reduced copy probability also isolated spikes are generated appearing as background. The firing rate of these neurons is  $p = \alpha \cdot \epsilon$ . To fulfill the constraint that all neurons have the same firing rates the  $N - m$  uncorrelated processes are modelled as a Bernoulli processes with firing probability  $p$  (see Fig. 1B for a sketch of the model, and Fig. 2B for an example realization with ordered (top) and randomized (middle) neuron identifiers).

**Control data.** For comparison we generate control data without any correlated component. To this end we randomize the spike times (bins occupied by a spike) of  $\text{SIP}_m$  and  $\text{MIP}_m$  realizations along the temporal axis. This conserves the spike counts of the original data thereby avoiding additional variance (see Fig. 2C for an example of control data for the  $\text{MIP}_m$  model realization in Fig. 2B).

### 3 Distribution of coincidence counts

Next we explore how correlation between groups of neurons is expressed in simple measures like the population histogram and the distribution of the corresponding counts per bin. Thus we ignore individual spike constellations across the neurons and restrict the description to the distribution of the number of spikes of all  $N$  neurons within a bin, i.e. the complexity  $\xi$  (shown in Fig. 2, bottom row). We derive analytical descriptions of the distributions for both model types and then compare these to simulations.

#### 3.1 Coincidence count distribution of $\text{SIP}_m$ model

The mathematical description of the coincidence distribution of correlated processes can be split into two contributions. Given that the time bins are independent, i.e. the processes have no memory, we can reorganize the bins in time (see Fig. 1C). Let us first collect the bins that contain injected coincidences (‘dependent part’), then consider the rest i.e. all the time bins containing uncorrelated activity (‘independent part’). The probability distribution of the dependent part is characterized by coincidences of order  $m$  with firing probability  $\alpha$  in the  $m$  out of  $N$  neurons. The  $N - m$  neurons fire independently with

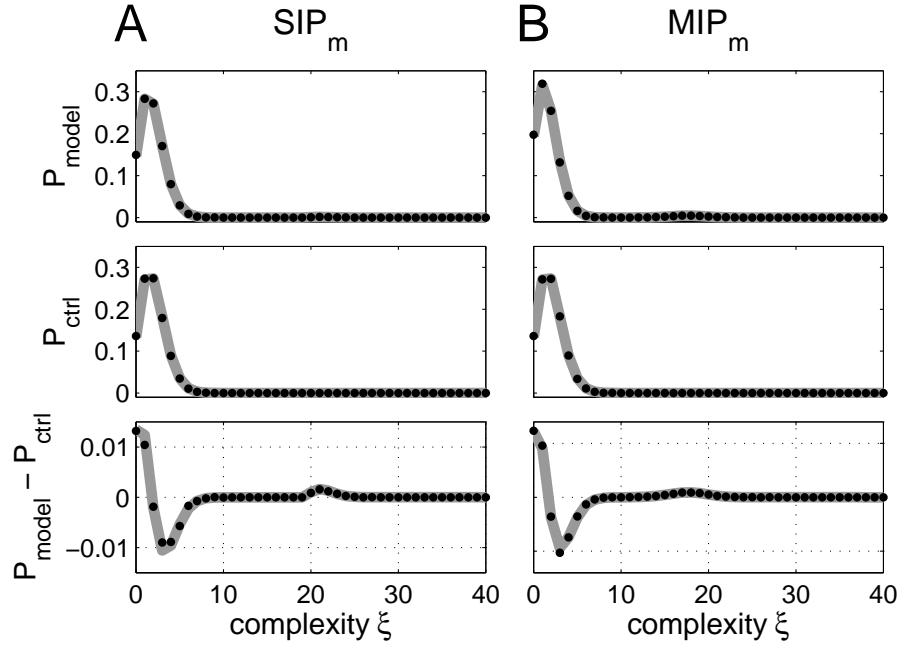


Figure 3: Coincidence probability distributions of the  $SIP_m$  model (A) and the  $MIP_m$  model (B). Distributions (horizontal axis: coincidence complexity  $\xi$  ranging for better visibility only from 0 to 40) are constructed from realizations of duration  $T = 100$ s at the resolution of the data  $h = 0.001$ s, with  $N = 100$  neurons and firing probability  $p = 0.02$  for all neurons.  $m = 20$  neurons contain correlated activity. In case of  $SIP_m$  the coincidence rate is  $\alpha = 0.005$ , in case of  $MIP_m$  the copy probability is  $\epsilon = 0.8$  resulting in an insertion probability of  $\alpha = 0.025$ . Top row: probability distribution of the raw coincidence counts; middle: results for the respective uncorrelated control data; bottom: difference distribution (model - control). The gray curves represent analytical results, black dots show results of simulations.

rate  $p$ . The coincidence distribution of only those  $N - m$  neurons follows a binomial distribution

$$B(i, N - m, p) = \binom{N - m}{i} p^i \cdot (1 - p)^{(N - m - i)} \quad . \quad (1)$$

These coincident events meet injected events of order  $m$ . Thus we obtain for the distribution of synchronized events of order  $\xi = i + m$  in the  $N$  neurons:

$$P_{SIP,dep}(\xi) = B(\xi - m, N - m, p) \quad . \quad (2)$$

Since injected coincidences are of order  $m$  only patterns of complexities with at least  $m$  spikes are found; the probability for synchronous events of complexity  $\xi < m$  is 0. The expectation value is

$$\langle \xi_{SIP,dep} \rangle = (N - m) \cdot p + m \quad . \quad (3)$$

The coincidence distribution of the remaining bins (independent part) is characterized by chance coincidences only. As before, the  $N - m$  neurons exhibit chance coincidences according to a binomial distribution with firing probability  $p$  as expressed in Eq. 1. However, the firing probability of the  $m$  neurons in the independent part  $p_b = p - \alpha$  is reduced by the injection probability of the synchronous events. The coincidence probability of these neurons is  $B(j, m, p_b) = \binom{m}{j} p_b^j \cdot (1 - p_b)^{(m - j)}$ . Consequently, we obtain the probability to observe a coincidence of a given complexity  $\xi$  by considering all possible combinations summing up to  $\xi$  of the number of coincidences among the group of  $N - m$  neurons and the number of coincidences among the group of  $m$  neurons. This is expressed by the convolution of the two binomial distributions:

$$P_{SIP,indep}(\xi) = \sum_i B(i, N - m, p) \cdot B(\xi - i, m, p_b) \quad (4)$$

abbreviated as

$$= B(i, N - m, p) * B(\xi - i, m, p_b)$$

where we define  $B(i, M, p) = 0$  for  $i < 0$  and  $i > M$ . Thus the expectation value of this part of the complexity distribution is

$$\langle \xi_{SIP,indep} \rangle = (N - m) \cdot p + m \cdot (p - \alpha) = Np - m\alpha \quad . \quad (5)$$

Finally, the total coincidence distribution is the sum of the distributions of the two parts Eq. 2 and Eq. 4 weighted by the relative number of bins containing injected coincidences and its complement respectively. For large values of  $T$  we just consider the mean fractions  $\alpha$  of the dependent part and  $(1 - \alpha)$  of the independent part:

$$P_{SIP_m}(\xi) = \alpha \cdot P_{SIP,dep}(\xi) + (1 - \alpha) \cdot P_{SIP,indep}(\xi) \quad . \quad (6)$$

For simplicity we ignore the loss of spikes induced by the injection of coincident events into background activity with subsequent clipping (Grün et al., 1999) since for the range of parameters studied the probability for such collisions is very small ( $\alpha \cdot p_b$ ).

### 3.2 Coincidence count distribution of MIP<sub>m</sub> model

In case of the MIP<sub>m</sub> model coincidences are inserted according to a given probability  $\epsilon$  which defines the probability to copy spikes from the mother process into  $m$  of the  $N$  neurons. Let us again consider first only the part of the bins in which the mother process contained a spike (dependent part). The probability distribution for coincidences within the  $m$  neurons only is  $B(i, m, \epsilon)$  (cmp. Kuhn et al., 2003). These events occur with the rate of the mother process, i.e. with probability  $\alpha$ . The spiking probability of the  $m$  neurons is  $p = \epsilon \cdot \alpha$ . For the remaining  $N - m$  neurons the same expression for the coincidence distribution holds as given above in Eq. 1. Thus, we yield for the dependent part:

$$P_{MIP,dep}(\xi) = B(i, m, \epsilon) * B(\xi - i, N - m, p) \quad . \quad (7)$$

The mean complexity is

$$\langle \xi_{MIP,dep} \rangle = (N - m) \cdot p + m \cdot \epsilon \quad . \quad (8)$$

Thus, MIP<sub>m</sub> differs from SIP<sub>m</sub> in the sense that not all  $m$  neurons are correlated with order  $m$ . Furthermore, SIP<sub>m</sub> contains background spikes in the correlated neurons, whereas MIP<sub>m</sub> does not. For the part of the bins without a spike in the mother process (independent part) the description of the distribution is similar to the one for the SIP<sub>m</sub> model (Eq.

4), however here the equation simplifies to chance coincidences from the  $N - m$  neurons only:

$$P_{MIP, indep}(\xi) = B(\xi, N - m, p) \quad . \quad (9)$$

The mean complexity is

$$\langle \xi_{MIP, indep} \rangle = (N - m) \cdot p \quad . \quad (9)$$

The total probability distribution then results from weighting by the relative contributions:

$$\begin{aligned} P_{MIP, dep}(\xi) &= \alpha \cdot P_{MIP, dep} + (1 - \alpha) \cdot P_{MIP, indep} \\ &= \alpha \cdot B(i, m, \epsilon) * B(\xi - i, N - m, p) + (1 - \alpha) \cdot B(\xi, N - m, p) \quad . \end{aligned}$$

Fig. 3 illustrates the coincidence count distributions for the  $SIP_m$  and the  $MIP_m$  model.

### 3.3 Control Data

In case of uncorrelated data (no insertion), the firing probability of all processes in Eq. 4 is  $p$  for both model types, and thus Eq. 4 reduces to:

$$P_{ctrl}(\xi) = B(i, m, p) * B(\xi - i, N - m, p) = B(\xi, N, p) \quad (8)$$

by using the addition formula for binomial coefficients. The same result holds for the control data for the  $MIP_m$  model. The expectation value for the mean complexity is  $\langle \xi_{ctrl} \rangle = Np$ . This distribution reflects the coincidence distribution assuming full independence of the processes subject to the constraint of identical firing rates.

## 4 Comparing Model and Control Data

In the following we compare model and control data based on our analytical derivations and simulations. In particular we are interested to know in how far model and control data differ and, in view of data analysis, how this knowledge can be used to detect correlation. As the population dot displays in Fig. 2B demonstrate, model and control data are visually not distinguishable if the neuron identifiers are randomly arranged. Also the coincidence

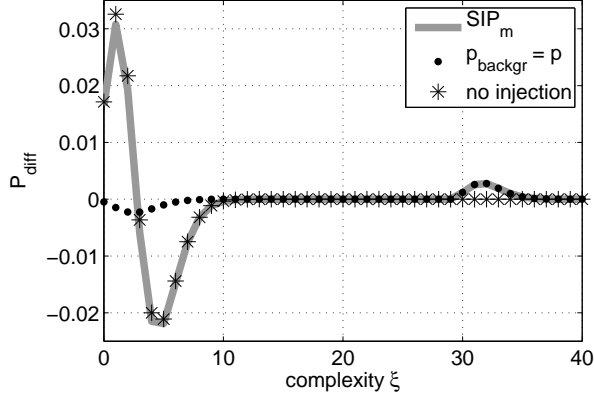


Figure 4: Constituents of the complexity distribution. The thick gray curve shows the difference between the probability distribution of the  $SIP_m$  model and the respective control ( $m = 30$ ,  $N = 100$ ,  $p = 0.03$ ,  $\alpha = 0.01$ ,  $T = 100s$ , and  $h = 0.001s$ ). The black dots show the difference distribution for a  $SIP_m$  model in which the background probability of the  $m$  neurons containing coincidences is elevated to match the rate  $p$  of the  $N - m$  independent ones. Both  $SIP_m$  models exhibit the same excess hump starting at  $\xi = 30$ , but for the latter model the pronounced biphasic feature at low complexities is reduced to a small dimple (cmp. Eq. 4). The asterisks show the difference distribution for the original  $SIP_m$  model without injected coincidences; the  $m$  neurons only have uncorrelated background spikes at rate  $p - \alpha$ . The excess of high complexities is absent but the biphasic feature is conserved.

count distributions are visually very similar Fig. 3 (top, middle). Therefore we subtract the control data from the model data to highlight potential net excess coincidences in the model data Fig. 3 (bottom). Here deviations become clearly visible: at low complexities the model data contain more coincidences than the control, at slightly higher complexities they contain less coincidences, and for complexities at about  $\xi = m$  there is again an excess of coincidences.

These features are characteristic for both models (Fig. 3, bottom row), however for  $SIP_m$  excess coincidences occur with a hump at a value of  $\xi$  slightly above  $m$ , for  $MIP_m$  the hump is located below  $\xi = m$ . Due to the chosen copy probability of  $\epsilon = 0.8$  for  $MIP_m$ , the probability for patterns of complexity  $\xi = m$  is low, and therefore the complexity of the hump is at  $\xi < m$ . In contrast, for  $SIP_m$  the hump is at values  $\xi > m$ , since patterns of complexity  $m$  are inserted, and by chance meet background spikes of the  $N - m$  uncorrelated neurons.

The origin of the differences of the model data and the control data at low complexities

is less obvious. Let us therefore restate the expression for the difference of the  $SIP_m$  model and the control data (subtract Eq. 3.3 from Eq. 6):

$$\begin{aligned} P_{SIP,diff}(\xi) &= P_{SIP_m}(\xi) - P_{ctrl}(\xi) \\ &= \alpha \cdot B(\xi - m, N - m, p) + (1 - \alpha) \cdot B(i, N - m, p) * B(\xi - i, m, p_b) \\ &\quad - B(\xi, N, p) \quad . \end{aligned}$$

For the sake of simplicity we assume that coincidences are inserted into all  $N$  processes. Thus Eq. 4 simplifies to:

$$P_{SIP,diff}(\xi) = \alpha P_{dep} + (1 - \alpha) B(\xi, N, p_b) - B(\xi, N, p)$$

with

$$P_{dep} = \begin{cases} 0 & \text{if } \xi < N \\ 1 & \text{if } \xi = N \end{cases} \quad (6)$$

Injected coincidences enter  $P_{SIP,diff}(\xi)$  only at  $\xi = N$  with probability  $\alpha$ . The remaining term expresses the contributions of the chance coincidences and their difference to the control data. Obviously, the distribution of the coincidences of the control data is shifted to a higher mean value ( $Np$ ) as compared to the chance coincidences of the independent part of the model data ( $Np_b$ ). The subtraction of the two distributions leaves a positive peak at small complexities, and a negative peak at somewhat higher complexities (Fig. 4). This difference can almost completely be compensated by increasing the rate of the independent part to  $p$ . In this case Eq. 4 reduces to:

$$\begin{aligned} P_{SIP,diff}(\xi) &= \alpha P_{dep} + (1 - \alpha) \cdot B(\xi, N, p) - B(\xi, N, p) \\ &= \alpha P_{dep} - \alpha \cdot B(\xi, N, p) \end{aligned}$$

or in the general case of  $m < N$

$$P_{SIP,diff}(\xi) = \alpha \cdot B(\xi - m, N - m, p) - \alpha \cdot B(\xi, N, p) \quad .$$

Whereas the entries at complexities  $\xi > m$  are not affected Fig. 4 (cmp. the gray curve and black dots at the hump around  $\xi = 32$ ), the biphasic feature at low complexities is reduced to a dimple reflecting the binomial distribution of chance coincidences scaled by

$-\alpha$ . The negative weight originates from the normalization constraint of the correlated data. The biphasic feature at low complexities can be replicated (Fig. 4, black asterisks) by independent data in which  $m$  neurons have a reduced firing probability  $p - \alpha$  as it is the case in the independent part of the  $\text{SIP}_m$  model.

In conclusion, the biphasic feature at low complexities in the difference distribution of model and control data is due to the constraint of all neurons having the same firing probability and not due to the injected coincidences. The feature can be eliminated by adjusting the background rate of the correlated neurons to a rate comparable to the rate of the independent neurons. As outlined in the discussion section, we may be able to exploit this observation to differentiate between candidate mechanisms for the generation of correlated spiking in the neuronal system.

## 5 Parameter Variations

After having understood how inserted higher-order coincident spike events influence the coincidence distribution, we now study parameters relevant for the analysis of neurophysiological data. Typical questions on an experimental data set are: Is there correlation in the data? What is the order of the correlation, i.e. how many and which neurons are involved? Furthermore, synchronous spike events may occur with a temporal jitter that does not correspond to the bin width chosen for analysis.

### 5.1 Variation of correlation order $m$

In the  $\text{SIP}_m$  model coincidences of synchronized spike events of order  $m$  are inserted into  $m$  neurons and in Fig. 2A we studied the coincidence distributions for  $m = 20$  in  $N = 100$  neurons. Now we are interested to see how the systematically varied order  $m$  of the injected coincidences affects the distribution. As before we study the distribution of the correlated data, the control data and their difference, however now visualized by a color code along the horizontal axis ( $\xi$ ) and for increasing  $m$  along the vertical axis (Fig. 5A). Again, we find the biphasic feature for low complexities, which hardly varies with increasing  $m$  since the insertion and background rates are not changed.

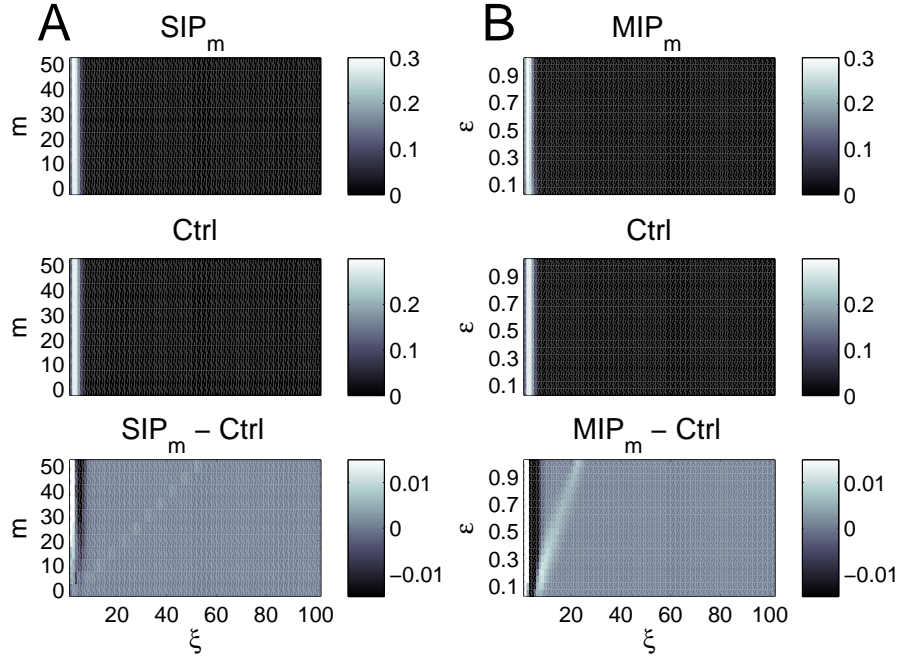


Figure 5: Dependence of the complexity distribution on the order of the correlation (top: model data, middle: control data, bottom: difference of model and control). (A) variation (vertical) of  $m$  for the  $\text{SIP}_m$  model ( $m = 5 \dots 50$  in steps of 5,  $\alpha = 0.005$ ) and (B) variation of  $\epsilon$  for the  $\text{MIP}_m$  model ( $\epsilon = 0.05 \dots 1$  in steps of 0.05). The gray code indicates the probability to observe a coincidence pattern of a certain complexity  $\xi$  (horizontal). The data in the top panels result from simulations of the respective models with parameters  $N = 100$ ,  $p = 0.02$ ,  $T = 100\text{s}$ , and  $h = 0.001\text{s}$ . In the  $\text{MIP}_m$  model  $\alpha$  is adjusted for each  $\epsilon$  value to account for  $p = \alpha \cdot \epsilon$ . The control data are generated by temporal randomization of the spikes of the model data.

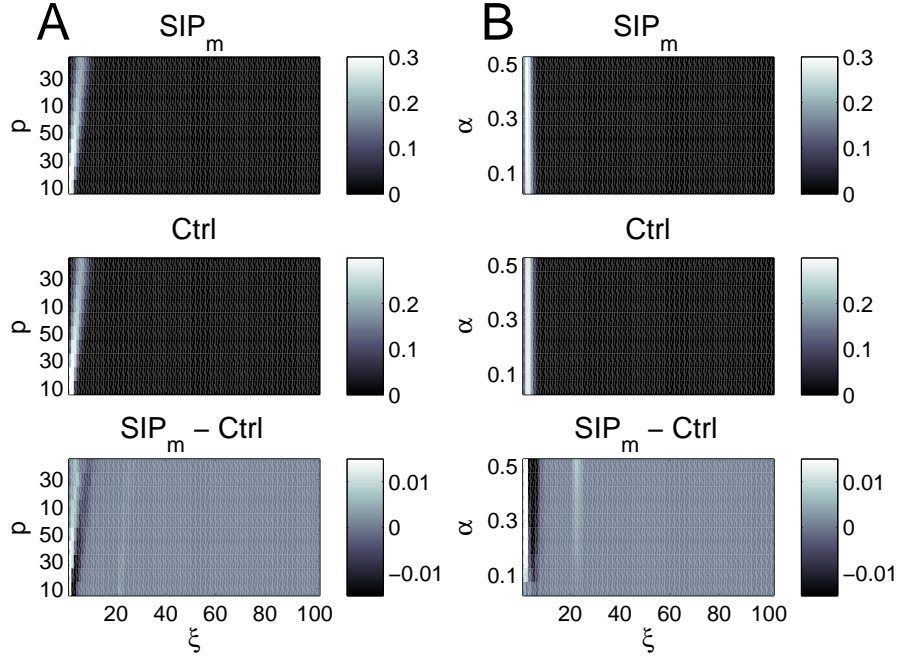


Figure 6: Dependence of the complexity distribution on firing and coincidence probability in the  $SIP_m$  model (top: model data, middle: control data, bottom: difference of model and control data). Common parameters of the simulations are:  $m = 20$ ,  $N = 100$ ,  $T = 100s$ ,  $h = 0.001s$ . The gray code indicates the probability to observe a coincidence pattern of a certain complexity  $\xi$  (horizontal). (A) The firing probability  $p = 0.005 \dots 0.05$  is varied in steps of 0.05 with a constant coincidence probability of  $\alpha = 0.05$ . (B) The coincidence probability  $\alpha = 0.002 \dots 0.02$  is varied in steps of 0.002 at a constant firing probability of  $p = 0.02$ .

The inserted coincidences, again, are not visible in the raw coincidence matrix. Only in the difference matrix (Fig. 5A,bottom) with increasing  $m$  excess coincidences appear at a complexity always somewhat higher than  $m$  since inserted coincidences of order  $m$  by chance meet background spikes which increases the complexity of the detected coincidence patterns.

For small  $m$ , the coincidences due to insertion overlap with the features due to the constraint on spike rate and the biphasic shape is disguised.

## 5.2 Variation of copy probability $\epsilon$

In physiological terms the copy probability  $\epsilon$  of the  $MIP_m$  model corresponds to the participation probability of the neurons in an assembly activation. In Fig. 5B we keep the total number of neurons  $N$  constant, as well as the number of neurons  $m$  (here

20) in which coincidences are inserted. With increasing  $\epsilon$  the mean complexity of the excess coincidences increase linearly according to Eq. 8. For  $\epsilon = 1$ , which corresponds to  $\text{SIP}_m$  without background, the mean complexity reaches a value slightly above  $m$  (Fig. 5B, bottom). The amplitude of the hump decreases with increasing  $\epsilon$  which is due the requirement of constant firing rate subject to the relation  $p = \epsilon \cdot \alpha$ . To fulfill this constraint the rate of the mother process  $\alpha$  has to decrease with increasing  $\epsilon$ . At the same time the variance of the complexity of the excess coincidences becomes larger according to  $m \cdot \epsilon \cdot (1 - \epsilon)$ , but due to the simultaneous decrease of  $\alpha$  the hump appears less wide in the difference plot.

Summarizing the variation of  $m$  and  $\epsilon$  leaves us with the insight that from the position of the excess hump we cannot directly conclude on the number of neurons involved in the correlation if the underlying model is not known. The hump position could either directly reflect the number of correlated neurons for  $\text{SIP}_m$ , or is indicating a smaller number of neurons than are actually involved in the correlation due a value of  $\epsilon < 1$  in the  $\text{MIP}_m$  model. As a consequence the underlying model must first be identified. One option is to extract the coincidence patterns which exhibit excess complexities and analyze them for their individual composition. If always the same set of neurons is active in a pattern we can conclude on  $\text{SIP}_m$ . However, if patterns are composed of subsets of a particular superset of neurons we can conclude on a  $\text{MIP}_m$  type model.

### 5.3 Variation of firing rates

Fig. 6A, top panel shows the dependence of the coincidence distribution on the total firing probability  $p$  for constant  $N$ ,  $m$ , and coincidence probability  $\alpha$  in case of  $\text{SIP}_m$ . By increasing  $p$  also the background spike probability  $p_b$  of the neurons with injected coincidences increases ( $p_b = p - \alpha$ ). Consequently, the mean complexity of the peak of the independent part of the coincidence distribution increases according to Eq. 5. The complexity at the hump of the excess coincidences (Fig. 6A, bottom) mainly corresponding to the dependent part also increases with the firing probability  $p$  but with a smaller slope, i.e. according to Eq. 3 with  $(N - m)p + m$ , since  $(N - m)p < Np$ .

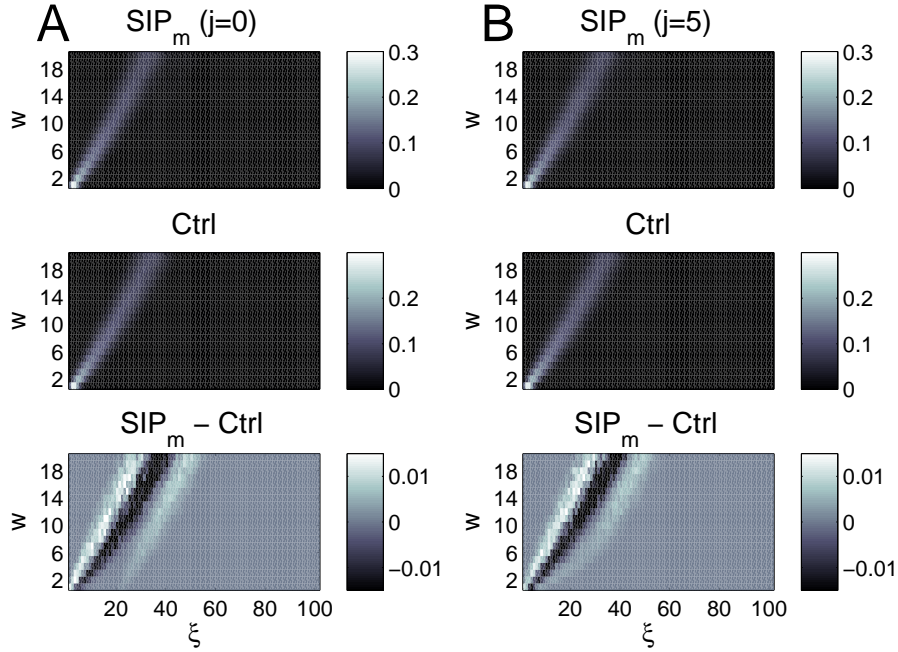


Figure 7: Coincidence probability distribution (top: model data, middle: control data, bottom: difference of model and control) of the  $SIP_m$  model under variation of the bin width  $w$  (varied from 1 to 20 in steps of 1 in units of  $h = 0.001s$ , vertical axis). The analyzed data sets differ in the temporal jitter of the inserted coincidences: in (A) the coincidences are exact without temporal jitter, in (B) the jitter is  $j = 5h$ . The jitter is generated by randomly displacing each spike of each neuron within a window of  $\pm 0.5j$  centered at its original position, thereby generating coincident spike events with a maximal distance of  $j$ . The gray code indicates the probability of occurrence of coincidence patterns as a function of complexity  $\xi$  (horizontal). The data in the top panels result from simulations of the  $SIP_m$  model with parameters  $N = 100$ ,  $T = 100s$ ,  $h = 0.001s$ ,  $m = 20$ ,  $\alpha = 0.05$  and  $p = 0.02$ .

Similar considerations hold for changes of the coincidence rate  $\alpha$  (Fig. 6B). Keeping all other parameters constant, a change of the coincidence injection probability  $\alpha$  only affects the hump height, but not its complexity. Only the complexity of the independent part at small complexities is affected, since an increase of the coincidence probability leads to a decrease of the background probability in the  $m$  neurons. Consequently, as discussed in section 4, the difference in chance coincidences of the model and the control data increases (cmp. Eq. 4).

#### 5.4 Variation of bin width vs. temporal jitter

Coincident spike events of pairs of cortical neurons typically have a temporal jitter of a few ms (see e.g. Grün et al., 1999; Pazienti et al., 2008). Such a jitter can be modeled by

copying the spikes of the mother processes not always into exactly the same bin across the neurons, but to allow copying into neighboring bins with some probability. For the sake of simplicity, here we decide for a rectangular distribution of spike times as described in the caption of Fig. 7.

One option to detect jittered coincidences, is to adjust the bin width (e.g. Grün et al., 1999, 2002a) to the precision of the spikes. Since in an electrophysiological experiment the appropriate bin width cannot be known in advance, we analyze the same data set with increasing bin width. Thus, instead of counting the simultaneously emitted spikes at the resolution of the data  $h$ , we now count the spikes within  $w$  neighboring time steps.  $w$  is called the bin width and we also call the sum of events complexity although a single neuron may contribute more than one spike.

Similar to the foregoing displays we then plot the coincidence probability distribution as a function of the complexity  $\xi$  horizontally, and along the vertical axis as a function of the bin width  $w$  (in units of  $h$ ). Fig. 7 shows in A the result for a non-jittered data set and in B the result for a data set where coincidences have a jitter of up to  $j = 5\text{ms}$ . Even in the case of no jitter the complexity of the peak of the chance coincidences increases with bin width. This holds for both, the model and the control data. The reason is that with increasing bin width the probability to detect spikes within a bin trivially increases due to background activity. For the independent part the coincidence probability distribution for a bin width  $w$  reads:

$$P_{SIP, indep}(\xi, w) = (1 - \alpha) \cdot B(i, (N - m) \cdot w, p) * B(\xi - i, m \cdot w, p_b)$$

Thus the mean of the independent part of the complexity distribution increases as  $w(N - m) \cdot p + wm \cdot p_b$ . The detected complexities are generally higher than in the original case where the width of the bins corresponds to the resolution  $h$  of the generating Bernoulli process.

For the dependent part we obtain:

$$P_{SIP, dep}(\xi, w) = \alpha \cdot B(\xi - m, (N - m)w, p) \quad . \quad (5)$$

The mean complexity is  $m + w(N - m)p$ .

If coincidences are jittered, only part of the coincidences are detected for  $wh < j$ . For small  $w$  the complexity of the detected excess coincidences is reduced to the number of spikes of an injected coincidence falling into one bin. Furthermore, the counts of these events is smaller than the injected number and therefore the hump of the excess coincidences is small. From  $wh = j$  on, all inserted coincidences are detected (neglecting debris due to exclusive binning; Grün et al., 1999). The complexity of the hump is now increasing faster with  $w$  as for the range of bin sizes for  $w < j$ . This change of slope at  $w = j$  may be used as an indicator for the precision of synchronization in the data.

## 6 Discussion

The goal of this study is to learn if a simple measure like the population histogram generated from parallel spike data can be used to detect correlation between the spike trains. To study such a situation we re-formulated two types of models for parallel point processes, which enables us to generate many parallel spike trains with a subgroup of them being correlated and exhibiting coincident spiking. As a measure of correlation between the spike trains we use the coincidence count distribution which is the amplitude distribution of the population histogram, i.e. the sum of spikes across the neurons within a predefined bin size as a function of time.

We illustrate that correlations are neither directly visible in the population dot display nor in the coincidence distribution, since background spikes act as a strong noise component. These spikes generate a large number of entries in the coincident count distribution at low complexities dominating the distribution. We choose to account for effects of background firing rate by comparing the model data to control data, which contain the same number of spikes per neuron but with the single neurons exhibiting fully independent firing. Still, the distribution of the original and the control data appear very similar by visual inspection, and only the difference of the two gives indication of correlation in the model data.

We find that correlation in a subset of the neurons can be clearly identified by entries in the difference matrix at the complexities close to the order of correlation contained

in the data. The exact complexity observed depends on the assumed underlying model. For  $SIP_m$  the minimum complexity of the excess coincidence entries reflects the number of correlated neurons, for  $MIP_m$  the maximum. The two models can be distinguished by extracting the time bins of entries of high complexity and analyzing the detailed composition of the neurons contributing to the patterns. In  $SIP_m$  a subset of  $m$  neurons would always fire together, for  $MIP_m$  a subset of  $m$  neurons would be involved but would contribute with different compositions of neurons. The latter seems to be a quite realistic assumption. Realizations of the synfire chain model would exhibit such activity in each group of the chain, since not all neurons in a group need to be active for stable propagation of a synfire run (Diesmann et al., 1999). A typical participation probability for each individual neuron, which corresponds to  $\epsilon$  in the  $MIP_m$  model, is about 0.8

If the number of correlated neurons, however, becomes small compared to the chance complexities the hump of the excess coincidences drowns in the noise. However, for  $m$  as small as 5 the hump of the excess coincidences already separates from the background. Larger background firing rate would indeed shift the peak of the chance coincidences to somewhat higher complexities, but also shifts the hump of the excess coincidences due to enhanced chances to meet background spikes.

An obvious biphasic feature in the difference coincidence distribution occurs at low complexities which we identified as being due to the constraint of all neurons having the same firing probability. As a consequence of the presence of correlated spikes the background firing of the neurons in the correlated data set is reduced. Therefore the distribution exhibits a decreased probability for small complexities as compared to the fully independent control data. Thus, if coincident events are injected into processes containing the same rate as the independent processes, the biphasic feature is absent and the difference coincidence distribution becomes almost flat in this regime.

This observation provides an interesting option to distinguish two potential mechanisms for the generation of synchronous spike events in the neuronal system. In one scenario additional synchronized events are generated by the system without affecting the ongoing activity, which would correspond to the case where the neurons receiving

additional coincidences have an increased rate (by  $p - p_b$ ) as compared to the neurons without coincidences. An example situation is where some of the observed neurons are part of an occasionally active feed-forward subnetwork (synfire chain, Schrader et al., 2008). Alternatively, spikes could be shifted such that they become synchronized with others. This would correspond to the case where firing rates are the same for all neurons. An example situation is where spikes become locked to global network oscillations.

The comparison of experimental data with data that implement control models is a standard approach in the correlation analysis of parallel spike trains (e.g. Pipa & Grün, 2003; Ikegaya et al., 2004; Shmiel et al., 2006). Such a comparison is mostly formulated as a significance test where the control data realize a specific null-hypothesis. The analysis presented here is rather thought of as a fast scanning procedure helping to decide if a data set contains interesting correlation which should further be analyzed in more detail. It provides a comparison of original data to control data realizing full statistical independence. The latter can be analytically described by a Bernoulli distribution. However, also a numerical realization can be achieved easily in practice by randomizing the bins of each neuron in time. Such a surrogate also conserves the spike count of each neuron. In contrast, realizations of Bernoulli processes with the firing probability estimated from the data would introduce additional variance.

A next step would be to add a statistical test to the discussed approach. One way of quantifying differences of the complexity distributions would be to perform a test on the full distribution (e.g. Pipa & Grün, 2003; Pipa et al., 2008) another to calculate and compare the moments or cumulants (Staudte et al., 2007, 2008). The latter also provides the option to compare the experimental data to models which include successively higher orders of correlation as implemented in Staudte et al. (2008), which then enable statements on the minimal order consistent with the data.

Other aspects that need to be considered in extensions of the current approach are typical features of experimental data: non-stationarity of the firing rate in time, inhomogeneous rates of the different neurons, and temporal modulation of the correlation (e.g. Riehle et al., 1997). Such situations can rapidly reach a level of complexity severely

impeding analytical descriptions but the flourishing idea of surrogates can come to the rescue (see Grün (2008) for a review).

**Acknowledgments.** This work was initiated during a scientific stay of SG and MD in 2002 at the Hadassah Medical School of the Hebrew University in Jerusalem financed by the NEURALCOMP-ICNC program. The project was supported in part by the Volkswagen Foundation, the Stifterverband für die deutsche Wissenschaft, the BCCN Berlin and Freiburg (BMBF grants 01GQ01413 and 01GQ0420), and DIP F1.2.

## References

- 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. M. H. J., Gerstein, G. L., Habib, M. K., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of ‘effective connectivity’. *J. Neurophysiol.* *61*(5), 900–917.
- Brown, E. N., Kaas, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* *7*(5), 456–461.
- Csicsvari, J., Henze, D. A., Jamieson, B., Harris, K. D., Sirota, A., Barth, P., Wise, K. D., & Buzsaki, G. (2003). Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J. Neurophysiol.* *90*, 1314–1323.
- Dayhoff, J. E., & Gerstein, G. L. (1983). Favored patterns in spike trains. I. detection. *J. Neurophysiol.* *49*(6), 1334–1348.
- Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* *402*(6761), 529–533.
- Ehm, W., Staude, B., & Rotter, S. (2007). Decomposition of neuronal assembly activity via empirical de-poissonization. *Electron. J. Statist.* *1*, 473–495.
- Gerstein, G. L., Perkel, D. H., & Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously recorded populations of neurons: Detection and measurement. *J. Neurosci.* *5*(4), 881–889.
- Grün, S. (2008). Data driven significance estimation for precise spike correlation (invited review). *J. Neurophysiol.*.. submitted.

- Grün, S., Abeles, M., & Diesmann, M. (2003). The impact of higher-order correlations on coincidence distributions of massively parallel data. In *Proc. 5th Meeting German Neuroscience Society*, pp. 650–651.
- Grün, S., Diesmann, M., & Aertsen, A. (2002a). ‘Unitary Events’ in multiple single-neuron spiking activity. I. Detection and significance. *Neural Comput.* *14*(1), 43–80.
- Grün, S., Diesmann, M., & Aertsen, A. (2002b). ‘Unitary Events’ in multiple single-neuron spiking activity. II. Non-Stationary data. *Neural Comput.* *14*(1), 81–119.
- Grün, S., Diesmann, M., Grammont, F., Riehle, A., & Aertsen, A. (1999). Detecting unitary events without discretization of time. *J. Neurosci. Methods* *94*(1), 67–79.
- Gütig, R., Aertsen, A., & Rotter, S. (2003). Analysis of higher-order neuronal interactions based on conditional inference. *Biol. Cybern.* *88*(5), 352–359.
- Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., & Yuste, R. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Science* *5670*(304), 559–564.
- Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlations in primary visual cortex of the Macaque. *J. Neurosci.* *25*(14), 3661–3673.
- Kuhn, A., Aertsen, A., & Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* *1*(15), 67–101.
- Martignon, L., von Hasseln, H., Grün, S., Aertsen, A., & Palm, G. (1995). Detecting higher-order interactions among the spiking events in a group of neurons. *Biol. Cybern.* *73*, 69–81.
- Nakahara, H., & Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Comput.* *14*, 2269–2316.
- Nicolelis, M., Ghazanfar, A., Faggin, B., Votaw, S., & Oliverira, L. (1997). Reconstructing the engram: simultaneous, multisite, many single neuron recordings. *Neuron* *18*(4), 529–537.

- Nowak, L. G., Munk, M. H., Nelson, J. I., James, A., & Bullier, J. (1995). Structural basis of cortical synchronization. I. Three types of interhemispheric coupling. *J. Neurophysiol.* *74*(6), 2379–2400.
- Pazienti, A., Diesmann, M., & Grün, S. (2008). The effectiveness of systematic spike dithering depends on the precision of cortical synchronization. *Brain Research.* in press.
- Pipa, G., & Grün, S. (2003). Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing* *52–54*, 31–37.
- Pipa, G., Wheeler, D., Singer, W., & Nikolic, D. (2008). Neurovidence: Reliable and efficient analysis of an excess or deficiency of joint-spike events. *J. Comput. Neurosci.* in press.
- 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*(5345), 1950–1953.
- Schneider, G., & Grün, S. (2003). Analysis of higher-order correlations in multiple parallel processes. *Neurocomputing* *52–54*, 771–777.
- Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* *440*, 1007–1012.
- Schrader, S., Grün, S., Diesmann, M., & Gerstein, G. (2008). Detecting synfire chain activity using massively parallel spike train recording. submitted.
- Shlens, J., Field, G. D., Gauthier, J. L., Grivich, Matthew I, P. D., Sher, A., Litke, A. M., & Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* *26*(32), 8254–8266.

- Shmiel, T., Drori, R., Shmiel, O., Ben-Shaul, Y., Nadasdy, Z., Shemesh, M., Teicher, M., & Abeles, M. (2006). Temporally precise cortical firing patterns are associated with distinct action segments. *J Neurophysiol* 96(5), 2645–2652.
- Staude, B., Rotter, S., & Grün, S. (2007). Detecting the existence of higher-order correlations in multiple single-unit spike trains. In *Society for Neuroscience*, Volume 103.9/AAA18 of *Abstract Viewer/Itinerary Planner*, Washington, DC.
- Staude, B., Rotter, S., & Grün, S. (2008). Inferring assembly-activity from population spike trains. to be submitted.