

CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains

Benjamin Staude · Stefan Rotter · Sonja Grün

Received: 20 December 2008 / Revised: 7 August 2009 / Accepted: 1 September 2009 / Published online: 28 October 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Recent developments in electrophysiological and optical recording techniques enable the simultaneous observation of large numbers of neurons. A meaningful interpretation of the resulting multivariate data, however, presents a serious challenge. In particular, the estimation of higher-order correlations that characterize the cooperative dynamics of groups of neurons is impeded by the combinatorial explosion of the parameter space. The resulting requirements with respect to sample size and recording time has rendered the detection of coordinated neuronal groups exceedingly difficult. Here we describe a novel approach to infer higher-order correlations in massively parallel spike trains that is less susceptible to these problems. Based on the superimposed activity of all recorded neurons, the cumulant-based inference of higher-order correlations (CuBIC) presented here exploits the fact that the absence of higher-order correlations imposes also strong constraints on correlations of lower order. Thus,

estimates of only few lower-order cumulants suffice to infer higher-order correlations in the population. As a consequence, CuBIC is much better compatible with the constraints of *in vivo* recordings than previous approaches, which is shown by a systematic analysis of its parameter dependence.

Keywords Multichannel recordings · Spike train analysis · Higher-order correlations · Cell assembly hypothesis

1 Introduction

More than 50 years after its first conception (Hebb 1949), the idea that entities of thought or perception are represented by the coordinated activity of (large) neuronal groups has lost nothing of its attraction (e.g., Wennekers et al. 2003; Harris 2005; Sakurai and Takahashi 2006). Now known as the “assembly hypothesis”, the concept of a single, unified principle underlying such diverse computational tasks as association and pattern completion (e.g., Palm 1982), binding (e.g., Singer and Gray 1995), language processing (e.g., Wennekers et al. 2006), or memory formation and retrieval (e.g., Pastalkova et al. 2008) has inspired numerous theoretical and experimental neuroscientists. However, whether or not the dynamic formation of cell assemblies constitutes a fundamental principle of cortical information processing remains a controversial issue of current research (e.g., Singer 1999; Shadlen and Movshon 1999; van Vreeswijk 2006). While initially mainly technical problems limited the experimental surge for support of the assembly hypothesis (state-of-the-art electrophysiological setups in the last century

Action Editor: Rob Kass

B. Staude · S. Grün
Unit of Statistical Neuroscience, RIKEN Brain Science
Institute, Wako-Shi, Japan

B. Staude (✉) · S. Rotter
Bernstein Center for Computational Neuroscience,
Freiburg & Faculty of Biology, Albert-Ludwig University,
Hansastr. 9a, 79104 Freiburg, Germany
e-mail: staude@bccn.uni-freiburg.de

S. Grün
Bernstein Center for Computational Neuroscience,
Berlin, Humboldt Universität zu, Berlin, Germany
e-mail: gruen@brain.riken.jp

allowed to record only few neurons simultaneously), the recent advent of multi-electrode array and optical imaging techniques reveals fundamental shortcomings of available analysis tools (Brown et al. 2004).

Based on the efficacy of synchronized presynaptic spikes to reliably generate output spikes (Abeles 1982; König et al. 1995; Diesmann et al. 1999), the temporal coordination of spike timing is a commonly accepted signature of assembly activity (e.g., Gerstein et al. 1989; Abeles 1991; Singer et al. 1997; Harris 2005). Consequently, approaches to detect assembly activity have focused on the detection of correlated spiking. Statistically, coordinated spiking of large neuronal groups has been associated to higher-order correlations among the corresponding spike trains (Martignon et al. 1995, 2000), where “genuine higher-order correlations” are assumed, if coincident spikes of a neuronal group cannot be explained/predicted by the firing rate and pairwise correlations alone. Mathematical frameworks for the estimation of such correlations exist (Nakahara and Amari 2002; Martignon et al. 1995; Schneider and Grün 2003; Gütig et al. 2003), however, run into combinatorial difficulties, as they assign one correlation parameter for each group of neurons and thus require in the order of 2^N parameters for a population of N simultaneously recorded neurons. Comparing the number of parameters to the available sample size of typical electrophysiological recordings (e.g. a population of $N = 100$ neurons implies $\sim 10^{30}$ parameters while 100 s of data sampled at 10 kHz provide $\sim 10^6$ samples) illustrates the principal infeasibility of this approach. In fact, the estimation of such higher-order correlations runs into severe practical problems even for populations of $N \sim 10$ neurons (Martignon et al. 1995, 2000; Del Prete et al. 2004).

Due to the limitations of multivariate approaches, most studies in favor of cortical cell-assemblies resort to pairwise interactions. And indeed, the existence and functional relevance of pairwise interactions has been demonstrated in various cortical systems and behavioral paradigms (e.g., Eggermont 1990; Vaadia et al. 1995; Kreiter and Singer 1996; Riehle et al. 1997; Kohn and Smith 2005; Sakurai and Takahashi 2006; Fujisawa et al. 2008). However, although pairwise analysis can indicate highly correlated groups of neurons (Berger et al. 2007; Fujisawa et al. 2008), knowledge of higher-order correlations is essential to conclude on large, synchronously firing assemblies (see examples in Fig. 4 for illustration). The importance to transcend pairwise descriptions of parallel spike trains is further underscored by the strong impact of higher-order correlations on the dynamics of cortical neurons (e.g., Bohte et al. 2000; Kuhn et al. 2003). Taken together, we believe that

the increasing number of simultaneously observable neurons can only be exploited by analysis techniques that go beyond mere pairwise correlations but are nevertheless applicable given the limited sample size of electrophysiological data (Brown et al. 2004).

In this study, we present a novel cumulant based inference procedure for higher-order correlations (CuBIC). The method is specifically designed to detect the presence of higher-order correlations under the constraints of the limited sample size typical of standard experimental paradigms. This is achieved by combining three basic ideas. First, CuBIC is based on the binned and summed spiking activity across all recorded neurons (population spike count). This transforms the multivariate problem to estimate all $\sim 2^N$ model parameters into a parsimoniously parametrized univariate problem, thereby dramatically reducing the required sample size. Furthermore, pooling the spiking activity avoids the need for sorting the multi-unit spike trains recorded on a single electrode into isolated single neuron spike trains. Correlations among the individual spike trains are measured by the cumulants of the population spike count. Importantly, the cumulant correlations of higher order used here do not conform to the higher-order parameters of the exponential family used by, e.g., Schneidman et al. (2006). Second, we use the compound Poisson process (e.g., Snyder and Miller 1991; Daley and Vere-Jones 2005) as a flexible, intuitive, and analytically tractable model of correlated spiking activity. Here, correlations among spike trains are modeled by the insertion of additional coincident events in continuous time (Holgate 1964; Ehm et al. 2007; Johnson and Goodman 2008; Staude et al. 2010). As a consequence, the conceptual difference between the discretely sampled data (population spike count) and the continuous time spike trains is an explicit feature of the present framework. And third, we formalize the observation that even small correlations of lower order can imply synchronized spiking of large neuronal groups (Amari et al. 2003; Benucci et al. 2004; Schneidman et al. 2006).

The first and second ideas are elaborated in Section 2. Based on the compound Poisson process, Section 3 thoroughly formalizes the third idea, where we analytically derive confidence intervals for a hierarchy of statistical hypothesis tests. The tests are then combined to compute a lower bound $\hat{\xi}$ for the order of correlations in a given data set. Importantly, the inferred lower bound $\hat{\xi}$ can considerably exceed the order of the estimated correlations. Thereby, CuBIC avoids the direct estimation of higher-order correlations, which is practically infeasible for orders of $m \gtrsim 5$, but nevertheless reveals their presence. As shown by

in this study does not assume binary variables. As a consequence, our term “complexity” as the “number of spikes per bin in the population” does not comply with the “number of active neurons per bin” used in binary frameworks.

To regard the values of $Z(s)$ in different bins as i.i.d. random variables, we here assume that $Z(s)$ and $Z(s+k)$ are independent for $k \neq 0$ (zero memory), and that the distribution of $Z(s)$ does not depend on the time bin s (stationarity). We name the resulting distribution of population spike counts

$$f_Z(l) := \Pr\{Z = l\}$$

the “complexity distribution” of the population. The validity of these assumptions with respect to real spike trains, and potential adaptations of CuBIC, are discussed in Section 5.3.

Despite ignoring the specific neuron IDs that contribute to the patterns $\mathbf{X}(s)$, the complexity distribution nevertheless contains information about the correlation structure of the population. For instance, if the counting variables $\{X_i\}_{i \in \{1, \dots, N\}}$ are independent Poisson counts, the corresponding population spike count, being the sum of the independent Poisson variables, is again a Poisson variable, and thus the complexity distribution f_Z is a Poisson distribution. Correlations change the relative probabilities for patterns of high and low complexities as compared to the independent case, as can be seen by comparing the complexity distribution of the correlated population and its independent Poisson fit in Fig. 1 (blue bars and dashed gray line, respectively; see also Grün et al. 2008a and Louis and Grün 2009). We quantify such deviations from independence by means of the cumulants $\kappa_m[Z]$ of the population spike count Z .

2.1.1 Correlations and cumulants

Like the more familiar (raw) moments $E[Z^m]$ of a random variable Z , the cumulants $\kappa_m[Z]$ characterize the shape of its distribution (see Appendix A and, e.g., Stratonovich 1967; Gardiner 2003). Most common are the first two cumulants, the expectation and the variance. These admit simple expressions in terms of the moments: $\kappa_1[Z] = E[Z]$ and $\kappa_2[Z] = E[Z^2] - E[Z]^2 = \text{Var}[Z]$. Similar expressions for higher cumulants, however, are exceedingly complicated (see Stuart and Ord 1987 for explicit expressions for $m \leq 10$).

The most important property of cumulants, and their advantage over the raw moments is that the cumulant of the sum of independent random variables is the sum of their cumulants. For $m = 2$ and $Z = \sum_{i=1}^N X_i$, for

instance, we have the well-known variance-covariance relationship

$$\begin{aligned} \kappa_2[Z] &= \text{Var}[Z] = \text{Var}\left[\sum_{i=1}^N X_i\right] \\ &= \sum_{i=1}^N \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j]. \end{aligned} \quad (1)$$

Equation (1) shows that $\text{Var}[\sum_{i=1}^N X_i] = \sum_{i=1}^N \text{Var}[X_i]$ if $\sum_{i \neq j} \text{Cov}[X_i, X_j] = 0$, which is the case if the X_i are (pairwise) uncorrelated. Cumulant correlations of higher order, the so-called “mixed” or “connected” cumulants, are a straightforward generalization of the covariance in exactly this sense: if the population has neither pairwise nor triple-wise cumulant correlations, then $\kappa_3[Z] = \sum_{i=1}^N \kappa_3[X_i]$ (see Appendix A for a concise definition, and Gardiner 2003, for a general introduction). Higher-order cumulant correlations generalize the covariance also with respect to the fact that $E[X_1 X_2] = E[X_1] E[X_2]$ implies $\text{Cov}[X_1, X_2] = 0$. That is, if the multivariate random vector \mathbf{X} decomposes into independent subgroups, then the cumulant correlations vanish (Streitberg 1990). For notational consistency, we will from now on stick to the cumulant notation, e.g., use “first/second cumulant” instead of the more familiar terms “mean/variance”.

A further consequence of Eq. (1) is that $\kappa_2[Z]$ is influenced only by the single process statistics (via $\sum_{i=1}^N \text{Var}[X_i]$) and pairwise correlations (via $\sum_{i \neq j} \text{Cov}[X_i, X_j]$). No higher-order correlations contribute to the second cumulant. This holds also for higher cumulants, i.e. $\kappa_m[Z]$ depends on correlations among the X_i of maximal order m . And finally, in the same way that $\kappa_2[Z]$ depends on pairwise correlations and the single process statistics, also $\kappa_m[Z]$ does not measure pure m th order correlations, but depends on correlations of all orders up to m . While a correction of the second cumulant for the influence of the single process statistics would be straightforward (subtracting $\sum_{i=1}^N \text{Var}[X_i]$ in Eq. (1), see Appendix B), correcting higher cumulants for the influence of correlations of lower order is exceedingly complicated. We therefore employ a parametric model for Z , the compound Poisson process (see next section), the parameters of which can be interpreted straightforwardly in terms of higher-order correlations among the X_i .

We would like to stress that the higher-order correlations defined by cumulants differ strongly from the higher-order parameters of the exponential family used by, e.g. Martignon et al. (1995), Nakahara and Amari (2002), Shlens et al. (2006). The relationship between

these two frameworks is discussed in more detail in Section 5.1 (see also Staude et al. 2010).

2.2 Model

As opposed to the *discretized*, i.e. binned, population spike count $Z(s)$ of the previous section, the proposed model operates in *continuous*, i.e. unbinned time. That is, we model the process $z(t) = \sum_{i=1}^N x_i(t)$, where $x_i(t) = \sum_j \delta(t - t_j^i)$ denotes the i th unbinned, continuous-time spike train ($i = 1, \dots, N$) with spike-event times t_j^i . The model we propose for $z(t)$ is that of a compound Poisson process (CPP)

$$z(t) = \sum_j \delta(t - t_j) a_j, \tag{2}$$

where the event times t_j constitute a Poisson process, and the marks a_j are i.i.d. integer-valued random variables, drawn independently for all t_j . The marks a_j determine the number of neurons that fire at time t_j , and will be referred to as the “amplitude” of the event at time t_j . The probability that an event has a specific amplitude is determined by the amplitude distribution f_A , i.e. $f_A(\xi) = \Pr\{a_j = \xi\}$ (see Fig. 1). The Poisson process that generates the events t_j is called the “carrier process” of the model and its rate ν is the “carrier rate”. Processes of this type are also referred to as generalized, or marked, Poisson processes (see e.g. Snyder and Miller 1991 for a general definition and Ehm et al. 2007 for an application to spike train analysis).

To interpret the CPP as the lumped process of a correlated population of N spike trains, and to utilize the CPP for the generation of artificial data, the events of $z(t)$ are assigned to individual processes $x_i(t)$ ($i = 1, \dots, N$). The simplest model draws the a_j process IDs that receive a spike at t_j as a random subset from $\{1, \dots, N\}$, independently for all event times t_j (Fig. 1). This results in a homogeneous population of correlated Poisson processes, where all processes have the same rate, all pairs of processes have identical correlations, and so on. The SIP/MIP models presented in Kuhn et al. (2003) are special cases of this more general model (see also Appendix B for examples). As the inference procedure CuBIC presented in the remainder of this study is based solely on the summed activity, however, it is not affected by the details of the copying procedure. In particular, CuBIC does not presuppose that the population is homogeneous.

2.3 Relating measurement and model

To interpret the continuous-time model parameters ν and f_A in terms of correlations among the counting

variables X_i , we now relate the former to the cumulants of the population spike count Z . First of all, note that as the carriers process $z(t)$ is a Poisson process, the population spike counts in different bins are i.i.d. For $l = 1, \dots, N$, define the processes $y_l(t)$ that determine all event times t_j in Eq. (2) with given amplitude $a_j = l$ (compare Appendix A). Then the CPP $z(t)$ admits the representation $z(t) = \sum_{l=1}^N l \cdot y_l(t)$. As a consequence, the discretized population spike count satisfies $Z = \sum_{l=1}^N l Y_l$, where the Y_l are the counting variables obtained from the processes $y_l(t)$ using a bin size h . As the event times t_j of $z(t)$ follow a Poisson process and the subsequent amplitudes a_j are independent, the $y_l(t)$ are independent Poisson processes. The rate ν_l of $y_l(t)$ is given by $\nu_l = f_A(l) \cdot \nu$ ($l = 1, 2, \dots$). Hence, the Y_l are Poisson variables with parameter $\nu_l h$. As a consequence, all cumulants of Y_l are identical and given by $\kappa_m[Y_l] = \nu_l h$ ($m = 1, 2, \dots$; Gardiner 2003). The scaling behavior of the cumulants $\kappa_m[l Y_l] = l^m \kappa_m[Y_l]$ for all m (Mattner 1999) yields

$$\begin{aligned} \kappa_m[Z] &= \kappa_m \left[\sum_l l \cdot Y_l \right] \stackrel{\{Y_l\} \text{ independent}}{=} \sum_l l^m \kappa_m[Y_l] \\ &= \sum_l l^m \nu_l h. \end{aligned}$$

Using $\nu_l = f_A(l) \cdot \nu$ and $\sum_{l=1}^N l^m \cdot f_A(l) = E[A^m] =: \mu_m[A]$, we finally have

$$\kappa_m[Z] = \mu_m[A] \nu h. \tag{3}$$

Equation (3) is the central equation of this study and requires a few more remarks. First, it implies that the first two cumulants of the population spike count, $\kappa_1[Z]$ and $\kappa_2[Z]$, are determined solely by the carrier rate ν and the first two moments of the amplitude distribution, $\mu_1[A]$ and $\mu_2[A]$. As $\kappa_1[Z]$ and $\kappa_2[Z]$ determine firing rates and pairwise correlations in the population (Appendix B, Eqs. (24) and (26)), we conclude that all CPP models with identical carrier rate ν and identical $\mu_1[A]$ and $\mu_2[A]$ generate populations with identical rates and pairwise correlations. Differences in higher-order correlations can thus be modeled by choosing different higher moments for f_A (see Section 2.4 for the precise relationship between the entries of the amplitude distribution and higher-order correlations in the population). This makes the CPP a very flexible and convenient model to generate artificial data sets with identical firing rates and pairwise correlations, yet different higher-order correlations (see Section 4 and Appendix B for examples, and Kuhn et al. 2003; Ehm et al. 2007; Staude et al. 2010 for alternative parametrizations).

Second, the moments of the strictly positive, integer-valued variable A increase with the order, i.e. satisfy $\mu_m[A] \leq \mu_{m+1}[A]$ for $m \geq 1$. By Eq. (3), this implies that also the cumulants of Z increase with the order, i.e. $\kappa_m[Z] \leq \kappa_{m+1}[Z]$. This is a reformulation of the fact that insertion of joint spikes as done in the CPP generates only positive correlations (e.g., Brette 2009; Johnson and Goodman 2008, Johnson and Goodman, unpublished manuscript), and it shows that certain combinations of cumulants cannot be realized by the CPP.

2.4 Terminology

The conceptual difference between the measured data (the counting variables X_i and the complexity distribution f_Z) and the parametric model that we assume to underlie this measurement (the CPP) is crucial for the remainder of this study. To illustrate this difference, note that f_Z depends on the bin size h used in the discretization. Larger bin widths increase the probabilities for high complexities, while smaller bin widths increase the probability for empty bins. The parameters of the CPP, on the other hand, do not change with the bin size. For instance, if all events of $z(t)$ have amplitude $a_j = 1$, i.e. the amplitude distribution has a single peak at $\xi = 1$, the population consists of independent Poisson processes, irrespective of the bin size used for the analysis, or even the number of recorded neurons. In other words: patterns of complexity $Z \geq 2$ can occur by chance, while events of amplitude $\xi \geq 2$ do not occur by chance but imply correlations in the population. The order of the correlation, in turn, is determined by the amplitudes of the events in the corresponding CPP (see Appendix A for the proof):

Theorem 1 *Let $z(t) = \sum_{i=1}^N x_i(t)$ be a compound Poisson process with amplitude distribution f_A , and let $\mathbf{X} = (X_1, \dots, X_N)^T$ be the vector of counting variables obtained from the $x_i(t)$ with a bin width h . Then the components of \mathbf{X} have correlations of order m if and only if f_A assigns non-zero probabilities to amplitudes $\geq m$.*

The above theorem confirms the intuitive conception that, within the framework of the CPP, correlations of a certain order m require injected coincidences into at least m processes (events of amplitudes $\geq m$). Also, a population with only pairwise but no higher-order correlations has events of maximal amplitude $m = 2$. Importantly, correlations in the above theorem are defined strictly on the basis of the discretized counting variables X_i . As a consequence, they do not resolve

(and do not depend on) the perfect temporal precision of the coincident events in the CPP: if the events of $z(t)$ were copied into the individual processes with a temporal jitter that is small with respect to the bin size h , the correlations among the counting variables will hardly be affected (see Section 5.3.3 for a more detailed discussion). Note also that events of amplitude ξ induce not only correlations of order ξ , but also of orders $< \xi$.

3 Cumulant based inference of higher-order correlations (CuBIC)

This section describes our cumulant based inference procedure for higher-order correlations (CuBIC). The outcome of CuBIC is a lower bound $\hat{\xi}$ on the order of correlation in the spiking activity of large groups of simultaneously recorded neurons. As will be shown in Section 4, this lower bound can exceed the order of the cumulants that were estimated for the inference. Thereby, CuBIC can provide statistical evidence for large correlated groups without the discouraging requirements on sample size that direct tests for higher-order correlations have to meet. This is achieved by exploiting constraining relations among correlations of different orders. For illustration, consider as an example the extreme situation of 4 simultaneously recorded neurons, where all neuron pairs have a correlation coefficient of $c = 1$. As $c = 1$ implies identity for all pairs of spike trains, all four spike trains of this example must be identical (Fig. 2, left). In other words, a spike in one neuron implies joint spike events in all the other neurons. Thus, the data must have correlation of order 4.

The key observation of this example is that the order of correlation we inferred ($\hat{\xi} = 4$) exceeds the order m of the measured correlations (here $m = 2$). CuBIC formalizes this example in the framework of the CPP, and generalizes it in two aspects:

1. Assume the correlation coefficients had been somewhat smaller than 1: do we still need correlation of order 4 to explain the measured pairwise correlation? Or would correlation of order 3 as in Fig. 2, middle, or even of order 2 as in Fig. 2, right, suffice to explain the measured correlations? In other words: What order of correlation is minimally required to explain the measured pairwise correlations?
2. Can we formulate a similar reasoning for measured correlations of higher order? What order of correlation is minimally required to explain the measured correlations of third (fourth,...) order?

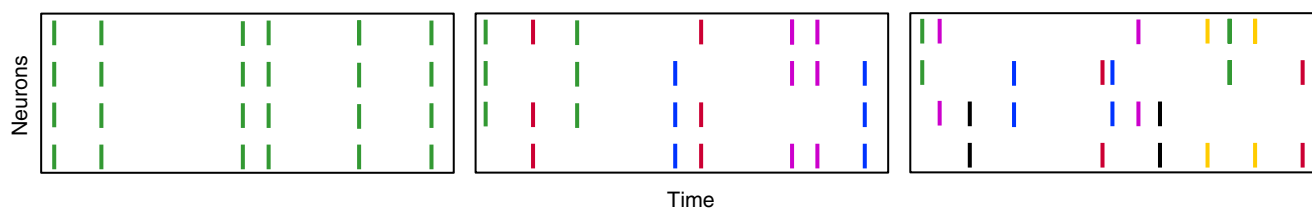


Fig. 2 Strong pairwise correlations imply higher-order correlations. Raster plots show populations that have a single event amplitude that differs in the three panels (*left*: all events have amplitude 4, *middle*: amplitude 3, *right*: amplitude 2). Note that an increase in pairwise correlations in the right population can

only be achieved by increasing the event amplitude and thereby the order of correlation. The plots thus illustrate the populations that are maximally correlated with correlations of order ≤ 4 (*left*), ≤ 3 (*middle*) and ≤ 2 (*right*)

The next section details our approach to answer the first question. The results are then generalized to answer the second question (Section 3.2). Finally, we explain how CuBIC combines the results to infer a lower bound for the order of correlation in a given data set (Section 3.3).

3.1 Testing pairwise correlations ($m = 2$)

We approach the first aspect with a hierarchy of statistical hypothesis tests, labeled by the integer ξ . For fixed ξ , the null hypothesis $H_0^{2,\xi}$ states that the pairwise correlations (thus the superscript “2”) in the population are compatible with the assumption that there is no correlation beyond order ξ ; the alternative $H_a^{2,\xi}$ states that correlation of order higher than ξ is necessary to explain the pairwise correlations. Rejection of e.g. $H_0^{2,4}$ in favor of $H_a^{2,4}$ thus implies the existence of correlation of at least order 5. The test statistics to decide between these alternatives is the second sample cumulant of the population spike count, and the compound Poisson process provides the framework to analytically derive the required confidence bounds.

3.1.1 The null hypothesis $H_0^{2,\xi}$

The formulation of a null hypothesis requires a careful distinction between the random variable that describes the experiment, and the data sample that is tested against the hypothesis. For a fixed value of the test parameter ξ , this section formulates the null hypothesis $H_0^{2,\xi}$, using the random variable Z' that describes the population spike count of a neuronal population. Section 3.1.2 explains how a given data sample $\{Z'_1, \dots, Z'_L\}$ is tested against $H_0^{2,\xi}$.

The null hypotheses $H_0^{2,\xi}$ is based on the CPP model whose population spike count has the maximal second cumulant under the constraints that (a) the first cumulant equals that of a given population spike count Z' , and (b) the maximal order of correlation in the model is ξ . Using the amplitude distribution f_A to parametrize

its correlation structure, this CPP model is the solution of the constrained maximization problem

$$\begin{aligned} \kappa_{2,\xi}^* &:= \max_{\nu, f_A} \{\kappa_2[Z]\} \\ \text{subject to } \kappa_1[Z] &= \kappa_1[Z'] \\ \text{and } f_A(k) &= 0 \text{ for } k > \xi, \end{aligned} \tag{4}$$

where ν and f_A are the CPP parameters that determine the population spike count Z . We rewrite Eq. (4) by observing that the second constraint ($f_A(k) = 0$ for $k > \xi$) together with Eq. (3) implies $\kappa_2[Z] = \mu_2[A]\nu h = \sum_{l=1}^{\xi} l^2 f_A(l)\nu h$. With $\nu_l = f_A(l) \cdot \nu$ (Section 2.3), we thus have

$$\kappa_2[Z] = \sum_{l=1}^{\xi} l^2 \nu_l h = \begin{pmatrix} 1^2 \\ \vdots \\ \xi^2 \end{pmatrix} \cdot \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_{\xi} \end{pmatrix} h, \tag{5}$$

where the dot denotes the standard scalar product. Using Eq. (5) and the vector notation $\vec{\xi}_m := (1^m, \dots, \xi^m)^T$ and $\vec{\nu}_{\xi} := (\nu_1, \dots, \nu_{\xi})^T$, the maximization problem of Eq. (4) becomes

$$\begin{aligned} \kappa_{2,\xi}^* &:= \max_{\vec{\nu}_{\xi}} \left\{ \vec{\xi}_2 \cdot \vec{\nu}_{\xi} h \right\} \\ \text{subject to } \vec{\xi}_1 \cdot \vec{\nu}_{\xi} h &= \kappa_1[Z']. \end{aligned} \tag{6}$$

In Eq. (6), both the function to maximize ($\vec{\xi}_2 \cdot \vec{\nu}_{\xi} h$) and the constraint ($\vec{\xi}_1 \cdot \vec{\nu}_{\xi} h = \kappa_1[Z']$) depend linearly on the parameters $\vec{\nu}_{\xi} = (\nu_1, \dots, \nu_{\xi})$. Problems of this type, so-called Linear Programming Problems, are uniquely solvable, e.g. using the Simplex Method (Press et al. 1992, Chapter 10.8). The solution yields the upper bound for the second cumulant $\kappa_{2,\xi}^*$ and the corresponding parameter vector $\vec{\nu}_{\xi}^*$. The carrier rate and amplitude distribution of the CPP that maximizes Eq. (6) are then given by $\nu^* = \sum_{l=1}^{\xi} \nu_l^*$ and $f_{A^*}(l) = \nu_l^*/\nu^*$.

Assume that the combination of firing rates and pairwise correlations in the population can be realized with correlations of order $\leq \xi$. Then, the second cumulant of

its population spike count Z' must be smaller than the upper bound $\kappa_{2,\xi}^*$ computed in the previous section. We thus formulate the null hypothesis

$$H_0^{2,\xi} : \kappa_2[Z'] \leq \kappa_{2,\xi}^*.$$

The alternative hypothesis

$$H_a^{2,\xi} : \kappa_2[Z'] > \kappa_{2,\xi}^*$$

states that, within the framework of the CPP model, the pairwise correlations in the population imply the presence of correlation of order $> \xi$.

3.1.2 Test statistics and their distributions

The derivations in the previous section require the cumulants $\kappa_1[Z']$ and $\kappa_2[Z']$ of the random variable Z' . Given only a data sample $\{Z'_1, \dots, Z'_L\}$ of size L , we estimate these quantities by the standard sample mean and (unbiased) sample variance (Stuart and Ord 1987)

$$k_1 := \hat{\kappa}_1[Z'] = \frac{1}{L} \sum_{l=1}^L Z'_l,$$

$$k_2 := \hat{\kappa}_2[Z'] = \frac{1}{L-1} \sum_{l=1}^L (Z'_l - k_1)^2.$$

The test of the data sample against $H_0^{2,\xi}$ requires the distribution of the test statistics k_2 under the null hypothesis. To derive this distribution, recall that $E[k_2] = \kappa_2[Z']$ and $\text{Var}[k_2] = \frac{\kappa_4[Z']}{L} - \frac{2\kappa_2[Z']^2}{L-1}$, where the $\kappa_i[Z']$ are the unknown cumulants of the variable Z' that underlies the sample (Stuart and Ord 1987). The trick of CuBIC is to assume in $H_0^{2,\xi}$ that Z' is the population spike count of the CPP that solves Eq. (6) after the unknown cumulant $\kappa_1[Z']$ has been substituted with its estimate k_1 . Under this assumption, all cumulants of Z' can be computed by inserting the model parameters v^* and f_{A^*} into Eq. (3), which yields

$$\text{Var}[k_2] = \left(\frac{\mu_4[A^*]}{L} - \frac{\mu_2[A^*]^2}{L-1} \right) v^* h. \tag{7}$$

Given sample sizes of $L > 10,000$ (roughly corresponding to 10–100 s of data, see Section 4.2 and Fig. 6(d)), the distribution of k_2 under $H_0^{2,\xi}$ is well approximated by a normal distribution. Taken together, under $H_0^{2,\xi}$ the test statistics k_2 is normal with mean $\kappa_{2,\xi}^*$ and variance given by Eq. (7), such that the p -value for the test against $H_0^{2,\xi}$ is

$$p_{2,\xi} = \int_{k_2}^{\infty} \frac{1}{\sqrt{2\pi \text{Var}[k_2]}} \exp\left(-\frac{(t - \kappa_{2,\xi}^*)^2}{2 \text{Var}[k_2]}\right) dt.$$

The rejection of the null hypothesis $H_0^{2,\xi}$ for a given ξ implies that the pairwise correlations in the data are not compatible with the assumption that there is no correlation beyond order ξ . Thus, rejecting $H_0^{2,\xi}$ implies that $\xi + 1$ is a lower bound for the order of correlation.

3.2 Testing higher-order correlations ($m > 2$)

We now generalize the tests with measured pairwise correlations ($m = 2$) to exploit also estimated correlations of higher order ($m > 2$). The difference to the tests with $m = 2$ is that the upper bound for the m th cumulant involves all cumulants up to order $m - 1$. The generalization of the maximization problem (Eq. (4)) reads

$$\kappa_{m,\xi}^* := \max_{v, f_A} \{\kappa_m[Z]\}$$

subject to $\kappa_i[Z] = \kappa_i[Z']$ for all $i < m$ (8)

and $f_A(k) = 0$ for $k > \xi$,

where, as before, the $\kappa_i[Z']$ are the cumulants of the random variable Z' . Using the v -parametrization of Eq. (5), we can rewrite Eq. (8) as the Linear Programming Problem

$$\kappa_{m,\xi}^* := \max_{\vec{v}_\xi} \left\{ \vec{\xi}_m \cdot \vec{v}_\xi h \right\} \tag{9}$$

subject to $\Xi^{m-1} \vec{v}_\xi h = \vec{\kappa}_{m-1}[Z']$,

where Ξ^{m-1} is a $(\xi \times m - 1)$ -dimensional coefficient matrix with entries $\Xi_{i,l}^{m-1} = l^i$, and $\vec{\kappa}_{m-1}[Z'] = (\kappa_1[Z'], \dots, \kappa_{m-1}[Z'])^T$ is the column-vector of the first $m - 1$ cumulants of Z' . Solving Eq. (9) yields both the maximal m th cumulant $\kappa_{m,\xi}^*$ and the rate vector \vec{v}_ξ^* , which again can be used to compute the corresponding model parameters v^* and f_{A^*} .

In direct analogy to the case $m = 2$, the (m, ξ) -null hypothesis $H_0^{m,\xi}$ states that correlation of orders $1, 2, \dots, m$ are compatible with the assumption that there is no correlation beyond order ξ . This implies that $\kappa_m[Z']$ falls below the upper bound $\kappa_{m,\xi}^*$, and we define

$$H_0^{m,\xi} : \kappa_m[Z'] \leq \kappa_{m,\xi}^*.$$

To test a sample $\{Z'_1, \dots, Z'_L\}$ against $H_0^{m,\xi}$, we estimate correlations of order m by the m th sample cumulant of the population spike count, the so-called the m th k -statistics k_m (Stuart and Ord 1987; the sample mean and unbiased sample variance defined above are the first two k -statistics). And again, we assume that Z' is the population spike count of the CPP model that solves Eq. (4) after the unknown $\kappa_i[Z']$ have been replaced by the estimates k_i ($i = 1, \dots, m - 1$). Then, the

test statistics k_m is normally distributed with mean $\kappa_{m,\xi}^*$ and variance $\text{Var}[k_m]$, where expressions for $\text{Var}[k_m]$ can be found in the literature (Stuart and Ord 1987). The p -value of $H_0^{m,\xi}$ is thus given by

$$p_{m,\xi} = \int_{k_m}^{\infty} \frac{1}{\sqrt{2\pi \text{Var}[k_m]}} \exp\left(-\frac{(t - \kappa_{m,\xi}^*)^2}{2 \text{Var}[k_m]}\right) dt. \quad (10)$$

Note that after substituting the true cumulant vector $\vec{k}_{m-1}[Z]$ by its estimate $\vec{k}_{m-1} := (k_1, \dots, k_{m-1})^T$ in Eq. (9), the resulting constraint $\Xi^{m-1} \vec{v}_\xi = \vec{k}_{m-1}$ consists of $m - 1$ equations for the ξ positive parameters v_1, \dots, v_ξ . Unfortunately, there can be combinations of estimated cumulants for which these equations cannot be solved. In this case, Eq. (9) does not possess a solution, and the corresponding null hypothesis cannot be tested (see also Section 3.3).

3.3 Computing the lower bound

In the preceding section, CuBIC was presented as a collection of hypothesis tests $H_0^{m,\xi}$, labeled by the indices m (the order of the estimated cumulant) and ξ (the maximal order of correlation in the null). We now combine these tests to infer a lower bound $\hat{\xi}$ for the order of correlation in a given data set. To do so, recall that rejecting $H_0^{m,\xi}$ for given m and ξ implies that the combination of the first m cumulants requires correlation of order $> \xi$. As a consequence, every rejected hypothesis $H_0^{m,\xi}$ implies that $\xi + 1$ is a lower bound for the highest order of correlation in the data. The question thus is: which of these bounds should we use?

Evidently, we aim to infer the highest order of correlation that is present in the data. Hence, the objective is to find the maximum of all the lower bounds that were obtained from the hierarchy of tests. We thus aim for the pair (m, ξ) with the highest value of ξ such that the corresponding null hypothesis $H_0^{m,\xi}$ is rejected.

A conceptual algorithm for this search is presented in Fig. 3. It consists of two nested loops: for a fixed order of the estimated cumulant m , the inner loop searches the highest ξ for which $H_0^{m,\xi}$ is rejected; the outer loop iterates over subsequent orders m . The free parameters of the algorithm are the test level α , and, to ensure termination of the loops, upper bounds for the cumulant (m_{max}) and the maximal order of correlation assumed in the null (ξ_{max} ; see Section 5 for their choice). After initializing the test variables ($m = 2, \xi = 1, \hat{\xi}_m = 1$), we estimate the first m cumulants from the data by computing the corresponding k -statistics. Next, we check if Eq. (9) is solvable for the current values of m and ξ (yellow box). This check consists of two steps. The first step checks if the first $m - 1$ estimated

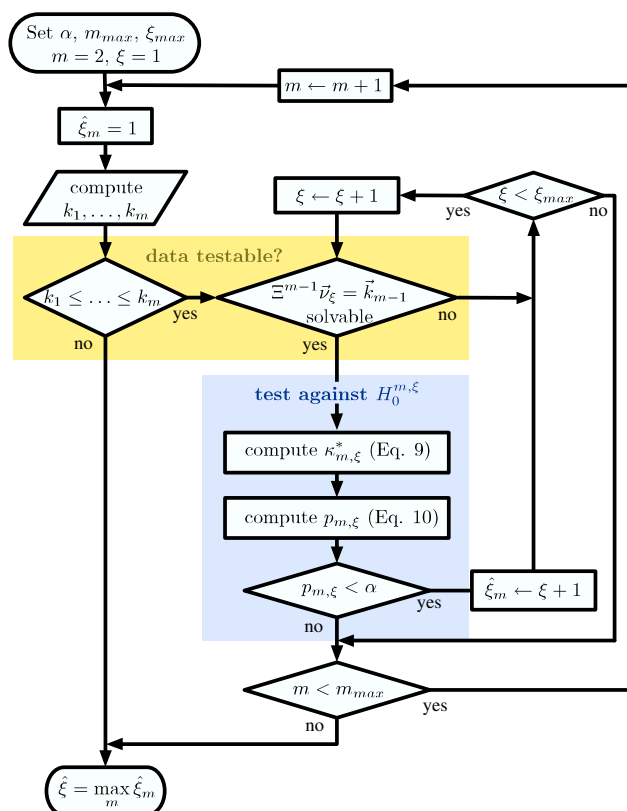


Fig. 3 Procedure to infer the highest lower bound $\hat{\xi}$ for the order of correlation in a given data set. The pair (m, ξ) with the highest ξ such that the corresponding null hypothesis $H_0^{m,\xi}$ is rejected is found by two nested loops. For any given m , the inner loop over ξ increases the lower bound $\hat{\xi}_m$, whenever Eq. (9) can be solved (right rhomb in upper yellow box) and the corresponding hypothesis is rejected ($p_{m,\xi} < \alpha$). This loop terminates ($p_{m,\xi} \geq \alpha$ or $\xi \leq \xi_{max}$) with $\hat{\xi}_m = \max\{\xi | H_0^{m,\xi} \text{ rejected}\} + 1$, which is the highest lower bound for the current value of m . The outer loop runs over the order of estimated cumulants m . The procedure terminates if the first $m - 1$ estimated cumulants violate the constraints of the CPP model (left rhomb in upper yellow box), or if the the order of the cumulant m reached the predefined upper bound m_{max} (rhomb below lower blue box). Finally, the bounds for different m are compared and their maximum $\hat{\xi} = \max_m \hat{\xi}_m$ is returned

cumulants increase with order m , a requirement of the CPP model assumed to underlie the data (left rhomb in yellow box; compare last paragraph of Section 2.3). If this is not the case, then $k_1 \leq k_2 \leq \dots \leq k_{m'-1}$ is false for all $m' \geq m$, which implies that Eq. (9) cannot be solved for $m' \geq m$. If the solution, i.e. the maximal cumulant $\kappa_{m',\xi}^*$ and the model parameters v^* and f_{A^*} , are not available, however, the corresponding p -values (Eq. (10)) cannot be computed. In this case, no hypothesis $H_0^{m',\xi}$ with $m' \geq m$ can be tested, and the procedure is terminated. If the first $m - 1$ estimated cumulants are in principle compatible with the CPP,

we check in the second step if the current value of ξ can solve Eq. (9) (right rhomb in yellow box). If this is not the case, ξ is increased by 1 until either Eq. (9) can be solved, or ξ reaches its upper bound ξ_{max} . In the former case, the data are tested against $H_0^{m,\xi}$ (blue box), in the latter case the test is skipped. At rejection of $H_0^{m,\xi}$ (rightward arrow at the box $p_{m,\xi} < \alpha$) we set $\hat{\xi}_m = \xi + 1$ and, unless ξ reached the upper bound ξ_{max} , repeat the inner loop with ξ set to $\xi + 1$. The loop terminates with $\hat{\xi}_m = \max\{\xi | H_0^{m,\xi} \text{ rejected}\} + 1$, which is the largest lower bound for the current value of m . As we reject the “zeroth” hypothesis $H_0^{m,0}$ by convention (every data set has correlations of order > 0), this holds even if the tests for the current value of m were skipped. The inner loop is repeated with new m set to $m + 1$, until m reaches the upper bound m_{max} . Finally, the

lower bounds obtained for the different values of m are compared, and their maximum is returned as the absolute lower bound $\hat{\xi}$. A MatLab-implementation of the proposed algorithm is available upon request.

4 Simulation results

4.1 Illustration

Before we present a systematic analysis of the proposed procedure in Section 4.2, we illustrate CuBIC’s sensitivity by a comparison of three sample data sets (Fig. 4). All three data sets are based on the same number of neurons ($N = 100$) with identical firing rates ($\lambda = 10$ Hz) and pairwise correlations ($c = 0.01$ in $N_C = 30$

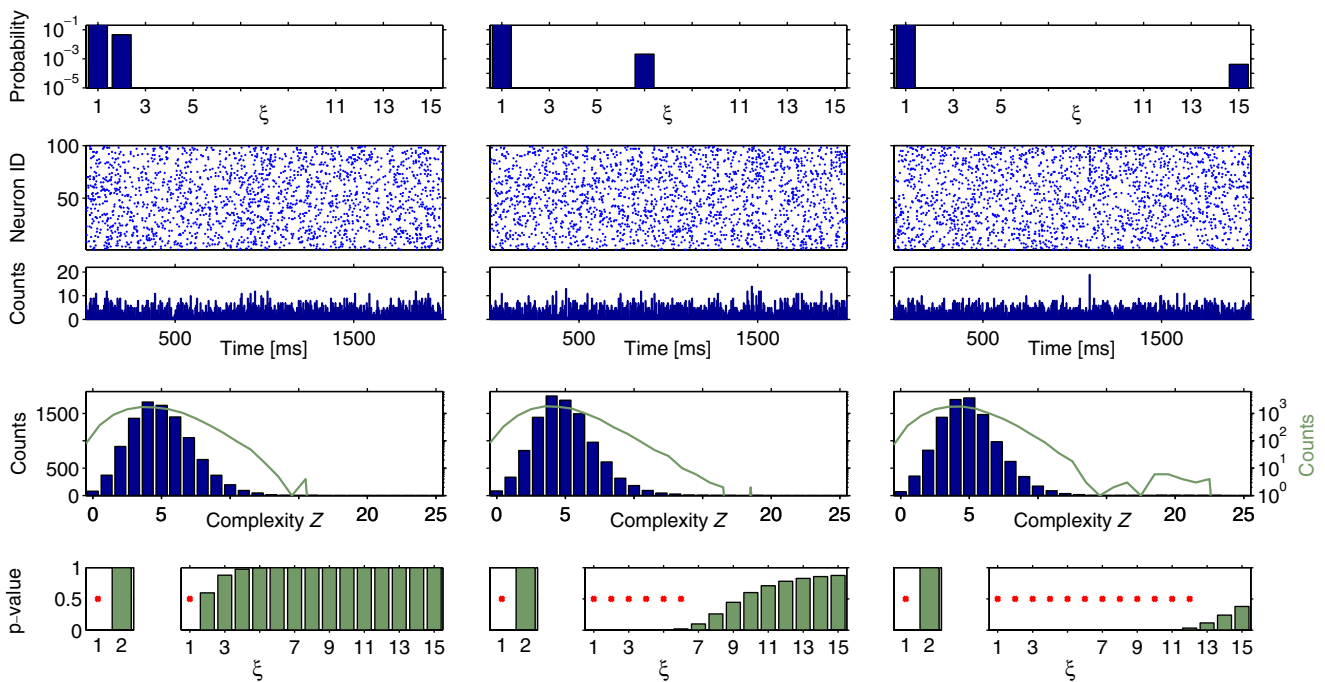


Fig. 4 CuBIC results for three example data sets with identical rates and pairwise correlations, but different higher-order correlations (each column shows one data set). Amplitude distributions (top, logarithmic y-scale) illustrate the correlation structure of the respective CPP. Below are the raster displays of the activities of $N = 100$ processes demonstrated for a period of the first 2 s of a total simulation time of $T = 100$ s. The third panels from top show the population spike counts of the data, computed with a bin width of $h = 5$ ms. Their histograms (of the total simulation time) are shown in the fourth panel as the complexity distribution Z (blue bars) and its logarithmically transformed version (green line graph, y-axis on the right). The bottom panels show the results of CuBIC, i.e. the p -values for subsequent hypothesis tests assuming increasingly higher ξ in the null-hypotheses $H_0^{m,\xi}$. The left graph shows the results for the second cumulant ($m = 2$), the

graph on the right uses the third cumulant ($m = 3$), respectively. Red filled circles denote p -values below a significance level of $\alpha = 0.05$. Correlation in the data sets was induced by events of amplitude $\xi_{syn} = 2$ (Set1, left), $\xi_{syn} = 7$ (Set2, middle), and $\xi_{syn} = 15$ (Set3, right), that were injected in addition to the background spikes at $\xi = 1$. In all data sets, the carrier rate ν and the relative probabilities of events of amplitude 1 and ξ_{syn} were chosen to mimic $N = 100$ neurons with identical firing rates (10 Hz). We let 70 of those neurons fire independently (neuron IDs 1–70) and $N_C = 30$ to form a homogeneously correlated subgroup with a count-correlation coefficient of $c = 0.01$ (neuron IDs 71–100), yielding higher-order event rates $\nu_{\xi_{syn}}$ of $\nu_2 = 43.5$ Hz in Set1, $\nu_7 = 2.07$ Hz in Set2 and $\nu_{15} = 0.41$ Hz in Set3. Test parameters: $\alpha = 0.05$, $\xi_{max} = 15$, $m_{max} = 4$

and $c = 0$ in the remaining 70 neurons). The difference between the data sets lies solely in their higher-order correlation structure, i.e. their maximal order of correlation (Fig. 4, top panels). All sets have amplitudes at $\xi = 1$ that generate independent “background spikes”. Correlations are induced by additional events of amplitudes $\xi_{syn} > 1$ that differ across data sets: $\xi_{syn} = 2$ in Set1 (left), $\xi_{syn} = 7$ in Set2 (middle), and $\xi_{syn} = 15$ in Set3 (right). Thus, Set1 has correlation of order 1 and 2, Set2 has correlations up to order 7, and Set3 has correlations up to order 15. The identical pairwise correlations in the three data sets are achieved by adjusting the rate $v_{\xi_{syn}} = v \cdot f_A(\xi_{syn})$ of the higher-order events (see Appendix B for the relationship between the model parameters v , f_A and the population statistics N , N_C , λ and c). Test parameters were set to $\xi_{max} = 15$ and $m_{max} = 4$.

The similarity of the raster displays and population spike counts (Fig. 4, 2nd and 3rd from top) of the three data sets may illustrate that such standard visualization methods cannot reveal the differences in the higher-order properties of the data sets. Larger probabilities for patterns of high complexity induced by the higher-order correlations present in Set2 and Set3 are visible in the complexity distributions (third row of panels) only when plotted on a logarithmic scale (green lines). The low rates of the higher-order events imposed by the small pairwise correlations ($v_7 = 2.07$ Hz in Set2, $v_{15} = 0.41$ Hz in Set3) makes their detection in the three distributions (bars) very difficult on a linear scale.

The identical pairwise correlations in all three data sets are reflected in identical results for tests with $m = 2$. For all data sets, $H_0^{2,\xi}$ is rejected ($p < 0.05$, indicated by red marks) only for $\xi = 1$, implying that the pairwise correlations are significant and events of amplitude 1 are not enough to explain the data. As $H_0^{2,\xi}$ is retained for $\xi > 1$ in all data sets ($p > 0.05$), the lower bound for $m = 2$ is $\hat{\xi}_2 = 2$ (see Fig. 3), i.e. the tests with $m = 2$ do not imply correlation of order ≥ 2 in any of the data sets.

Test results with $m = 3$ differ strongly for the three data sets (Fig. 4, lower panels, right graphs). For Set1, only $H_0^{3,1}$ is rejected and all $H_0^{3,\xi}$ for $\xi \geq 2$ are retained. Hence $\hat{\xi}_3 = 2$. For Set2, $H_0^{3,\xi}$ is rejected for $\xi = 1, \dots, 6$, yielding $\hat{\xi}_3 = 7$, while for Set3 the rejection of all null-hypotheses for $\xi \leq 12$ yields $\hat{\xi}_3 = 13$. In all tests with $m = 4$ (data not shown), the smallest values for ξ that solved Eq. (9) yielded non-significant p values (for Set1 $p_{4,2} = 0.5$, for Set2 $p_{4,9} = 0.8$, for Set3 $p_{4,16} = 0.67$). Hence, no hypothesis with $m = 4$ was rejected except for $H_0^{4,0}$ which is rejected by convention. Thus, $\hat{\xi}_4 = 1$ in all data sets. For the total lower bounds $\hat{\xi} =$

$\max_m \{\hat{\xi}_m\} = \max \{\hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4\}$, we thus obtain $\hat{\xi} = 2$ for Set1, $\hat{\xi} = 7$ for Set2, and $\hat{\xi} = 13$ for Set3.

The lower bound corresponds to the maximal order of correlation in Set1 ($\xi_{syn} = 2$) and Set2 ($\xi_{syn} = 7$). Only for Set3 the lower bound falls short of the true maximum by 2 ($\xi_{syn} = 15$), but nevertheless indicates correlations of higher order than in Set2. Taken together, the three examples illustrate that CuBIC reliably detects differences in higher-order statistics, even if firing rates (here $\lambda = 10$ Hz) and very small pairwise correlations (here $c = 0.01$ in 30 out of 100 neurons) are identical.

4.2 Parameter dependence of test performance

Several parameters are likely to influence CuBIC’s performance. For instance, the number of recorded neurons, N , the collection of firing rates, $\lambda_1, \dots, \lambda_N$, the duration of the experiment, T , and the bin size, h , influence the amount of available data and can thus be expected to effect test results. Also the correlation structure, i.e. the rate and the order of injected coincidences, parametrized by the $N - 1$ probabilities of the amplitude distribution f_A , is likely to affect the performance of the method. However, some of the parameters mentioned above are in fact redundant with respect to the test statistics used here, i.e. the k -statistics of the population spike count Z . For instance, Z depends only on the summed firing rate $\Lambda = \sum_{i=1}^N \lambda_i$, but not on the precise combination of the rates of individual neurons. To analyze the parameter dependence of CuBIC (Section 4.2.3), we focus on parameters that are relevant in experimental data and avoid the above-mentioned redundancies. Furthermore, we restrict our analysis to tests with the third cumulant, thus study the inner loop in Fig. 3 for $m = 3$. To keep notation simple, we drop the subscript 3 and write $\hat{\xi} = \hat{\xi}_3$ in the remainder of this section. The upper bound for ξ was $\xi_{max} = 30$ in all simulations.

4.2.1 Parameters

We investigate the performance of CuBIC with respect to five parameters: the duration of the experiment, T , the total firing rate of the population, $\Lambda = \sum_{i=1}^N \lambda_i$, the bin width used to discretize the data, h , and two parameters, ξ_{syn} and ρ , that characterize the correlation structure (amplitude distribution, f_A) of the data. We here consider amplitude distributions that consist of two isolated peaks only: a “background-peak” at $\xi = 1$ and a second peak at $\xi = \xi_{syn}$ that determines the maximal order of correlation in the data (see for examples

Fig. 4, top row). The relative height of the two peaks is parametrized by the moments ratio

$$\rho = \frac{\mu_2[A]}{\mu_1[A]}, \tag{11}$$

where $\mu_i[A] = E[A^i] = \sum_{\xi=1}^N \xi^i f_A(\xi)$ is the i th moment of A .

Substituting the moments $\mu_i[A]$ in Eq. (11) by the cumulants of Z via Eq. (3) shows that ρ equals the Fano Factor of the population spike count, i.e. $\rho = \frac{\kappa_2[Z]}{\kappa_1[Z]} = \frac{\text{Var}[Z]}{E[Z]}$ (see e.g., Kumar et al. 2008; Kriener et al. 2008). This interpretation is valid for arbitrary amplitude distributions, and provides a method to estimate ρ from data. For a more intuitive interpretation of ρ , consider a population of N neurons with identical firing rates, where a sub-population of N_C neurons forms a homogeneously correlated subgroup, while the remaining $N - N_C$ neurons fire independently (compare rasters displays in Fig. 4). In this case ρ relates to the (average) pairwise count correlation coefficient c of the correlated subgroup

$$c = \frac{N(\rho - 1)}{N_C(N_C - 1)}. \tag{12}$$

For a derivation of this relation see Appendix B, Eq. (29). In case of a completely homogeneous population with $N = N_C$, this simplifies to $c = \frac{\rho - 1}{N - 1}$ (compare e.g., Kumar et al. 2008; Kriener et al. 2008). In either case, the correlation parameter ρ determines the effect of higher-order events on the pairwise correlations in the population.

4.2.2 Quantification of test performance

We assess CuBIC’s performance by Monte-Carlo techniques. That is, we compute the lower bound for the order of correlation $\hat{\xi}$ (Figs. 3 and 5(a)) for each of 1000 data sets that are simulated from the CPP model with identical parameter combinations. The resulting (estimated) distribution $p(\hat{\xi})$ then indicates the range of lower bounds for this particular parameter combination. We thus study the parameter dependence of CuBIC’s performance by the parameter dependence of the distribution of $p(\hat{\xi})$ (Fig. 5(b)). For the specific parameter combination of Fig 5(a), for instance, lower bounds fall almost exclusively between $\hat{\xi} = 8$ and $\hat{\xi} = 14$ (Fig. 5(b), bottom panel). Increasing the correlation parameter ρ gradually sharpens and shifts $p(\hat{\xi})$ to higher values of $\hat{\xi}$ (Fig. 5(b)). CuBIC performs optimally if the lower bound corresponds to the maximal order of correlation in all Monte-Carlo simulations, i.e. if $\hat{\xi} = \xi_{syn}$ for all data sets. In this case, $p(\hat{\xi})$ is a delta-peak located at ξ_{syn} (Fig. 5(b), top panel).

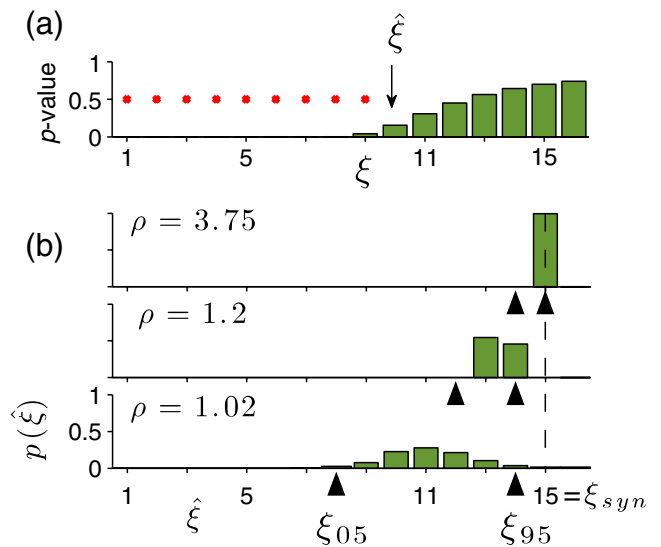


Fig. 5 Quantification of test performance. (a) p -values for tests against $H_0^{3,\hat{\xi}}$ and the resulting lower bound for the order of correlation $\hat{\xi} = \max\{\xi | H_0^{3,\xi} \text{ rejected}\} + 1$ (here $\hat{\xi} = 10$) for one data set (parameters: $\xi_{syn} = 15$, $\Lambda = 10^3$ Hz, $T = 100$ s, $\rho = 1.02$ and $h = 1$ ms). Red filled circles denote rejected hypotheses, i.e. p -values below 0.05. (b) Three estimated distributions of lower bounds $p(\hat{\xi})$, each obtained via Monte-Carlo simulations of 1000 data sets with identical parameter combinations. Parameters as in (a), only for different values of the correlation parameter (bottom: $\rho = 1.02$, middle: $\rho = 1.2$; top: $\rho = 3.75$). The two percentiles $\xi_{05} = \max\{\hat{\xi} | \sum_{k=\hat{\xi}+1}^{\infty} p(k) > 0.95\}$ and $\xi_{95} = \max\{\hat{\xi} | \sum_{k=\hat{\xi}+1}^{\infty} p(k) < 0.05\}$ are marked by arrowheads (see text for interpretation), also at the corresponding parameter values in Fig. 6(c). Observe the shift and sharpening of $p(\hat{\xi})$ towards the maximal order of correlation in the data $\xi_{syn} = 15$ (dashed line) for increasing ρ

To further reduce complexity, we will discuss the parameter dependence of $p(\hat{\xi})$ by means of the two percentiles (black triangles in Fig. 5(b))

$$\xi_{05} = \max \left\{ \hat{\xi} \mid \sum_{k=\hat{\xi}+1}^{\infty} p(k) > 0.95 \right\}$$

$$\xi_{95} = \max \left\{ \hat{\xi} \mid \sum_{k=\hat{\xi}+1}^{\infty} p(k) < 0.05 \right\}.$$

CuBIC performs optimally if $p(\hat{\xi})$ is a delta peak as in Fig. 5(b), top panel, i.e. if $\xi_{05} + 1 = \xi_{syn} = \xi_{95}$. More generally, a value of $\xi_{05} \gg 0$ indicates that CuBIC reliably infers the existence of higher-order correlations ($\hat{\xi} \gg 0$ in 95% of all cases), while a value of $\xi_{95} \leq \xi_{syn}$ guarantees that the actual order is not overestimated ($\hat{\xi} > \xi_{syn}$ in less than 5% of all cases, no false positives).

As mentioned before (Section 3.3), certain combinations of sample cumulants, e.g. $k_2 < k_1$, represent

untestable data sets and yield lower bounds of $\hat{\xi} = 1$. To avoid border effects for extremely small pairwise correlations ($\rho < 1.05$), we set $\hat{\xi} = 1$ not only if $k_2 < k_1$ but also if k_2 does not significantly exceed k_1 , i.e. if $H_0^{2,1}$ is retained (marked as dotted lines in Fig 6(c)).

4.2.3 Results

Figure 6 shows the dependence of CuBIC’s performance on the population firing rate Λ (Fig. 6(a)), the duration of the experiment T (Fig. 6(b)), the correlation strength (population Fano Factor) ρ (Fig. 6(c)) and the bin width h (Fig. 6(d)) for different values of ξ_{syn} (crosses in Fig. 6(a–d) denote the default parameters that are fixed while only one of them is varied).

Changes in the population firing rate Λ , the duration of the experiment T , and the correlation strength ρ have qualitatively very similar effects on the performance (Fig. 6(a), (b), and (c), respectively). For all analyzed values of ξ_{syn} (dark blue, green, and yellow lines) and small values of either of the parameters, i.e. $\Lambda = 10$ Hz (Fig. 6(a)), $T = 1$ s (Fig. 6(b)) or $\rho = 1.0087$ (Fig. 6(c)), the distributions of lower bounds $p(\hat{\xi})$ extend from $\xi_{05} = 0$ (lower lines) to $\xi_{95} < \xi_{syn}$ (upper lines below colored triangles). Hence, lower bounds might not exceed the trivial value of $\hat{\xi} = 1$ for such small parameter values. Increasing either of the parameters gradually shifts $p(\hat{\xi})$, i.e. ξ_{05} and ξ_{95} , to higher values of $\hat{\xi}$, thus indicating improved performance. The exact quality of the performance, captured by the width of $p(\hat{\xi})$, i.e. the distance between ξ_{05} and ξ_{95} , and in particular its difference from ξ_{syn} (colored triangles, compare also Fig. 5(b)), depends crucially on the order of correlation ξ_{syn} present in the data.

For $\xi_{syn} = 7$ (dark blue lines), CuBIC performs optimally ($\xi_{05} + 1 = 7 = \xi_{95}$, compare Fig. 5(b)) for experiment durations of $T > 100$ s (Fig. 6(b)) and a wide range population Fano Factors ($\rho \geq 1.17$, Fig. 6(c)). Such optimality is not guaranteed for the analyzed range of the population firing rate Λ , although population rates above $\Lambda = 400$ Hz yield a lower percentile of $\xi_{05} = 5$, indicating close-to-optimal performance ($\hat{\xi} \geq 6 = \xi_{syn} - 1$ in 95% of all cases, Fig. 6(a)). Smaller values of either of the parameters shift $p(\hat{\xi})$ to lower values, indicating impaired performance. Nevertheless, as the lower percentile ξ_{05} is considerably larger than zero even for population rates as low as $\Lambda = 20$ Hz, experiment durations of $T = 6$ s and population Fano Factors of $\rho = 1.02$, CuBIC reliably detects the presence of higher-order correlations for a wide range of parameter combinations.

In the other extreme, where the same pairwise correlations (identical value of ρ) are realized by correlations

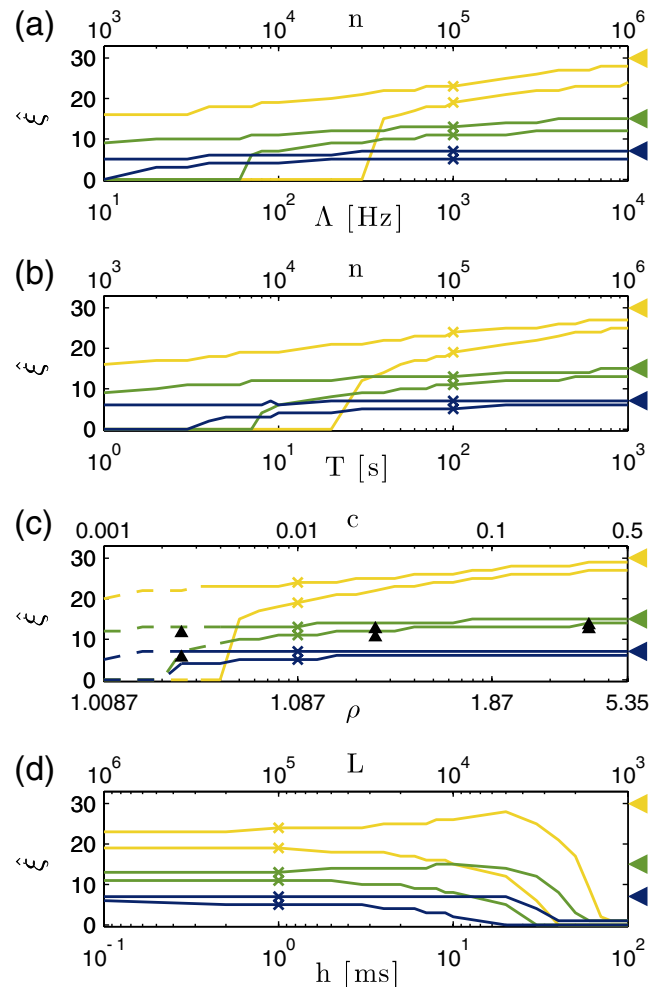


Fig. 6 Parameter dependence of CuBIC for $m = 3$. Shown are the percentiles ξ_{05} and ξ_{95} (lower and upper curves of same color, respectively; see Fig. 5(b)), as a function of (a) the population firing rate Λ , (b) the duration of the experiment T , (c) the correlation strength ρ , and (d) the bin width h for data sets with correlation of maximal order $\xi_{syn} = 7$ (dark blue), $\xi_{syn} = 15$ (green) and $\xi_{syn} = 30$ (yellow). Crosses denote default parameter combinations ($\Lambda = 10^3$ Hz, $T = 100$ s, $\rho = 1.087$ and $h = 1$ ms) one of which is varied in the different panels. Additional x axes show corresponding values of the total spike count $n = \Lambda T$ (a and b), the pairwise correlation coefficient c assuming a population of $N = 100$ neurons, $N_C = 30$ of which are homogeneously correlated (c), and the sample size $L = \frac{T \cdot 1000}{h}$ (d). Arrowheads on the right indicate ξ_{syn} , which is the optimal value for ξ_{95} . Data points for the dashed lines in (c) contain data sets where k_2 did not significantly exceed k_1 (compare Section 4.2.2). Upward arrowheads in (c) denote the percentiles for the parameter combinations of Fig. 5(b)

of order $\xi_{syn} = 30$ (yellow lines), $p(\hat{\xi})$ falls short of its optimum ($\xi_{05} + 1 = 30 = \xi_{95}$, see Fig. 5(b), top panel) even for high values of all investigated parameters, i.e. population rates of $\Lambda = 10^4$ Hz, experiment durations of $T = 10^3$ s and correlation strengths of $\rho = 5.35$ (Fig. 6(a), (b) and (c), respectively). Note, however,

that also here $\xi_{05} \gg 0$ even for moderate values of Λ , T and ρ , implying that the presence of higher-order correlations is reliably detected, albeit not their actual order. For the default values $\Lambda = 10^3$ Hz, $T = 100$ s and $\rho = 1.087$ (yellow crosses in Fig. 6(a–d)), for instance, lower bounds for $\xi_{syn} = 30$ fall between $\hat{\xi} = 20$ and $\hat{\xi} = 24$ in 90% of the analyzed data sets ($\xi_{05} = 19$, $\xi_{95} = 24$).

Test performance for $\xi_{syn} = 15$ (green lines) is somewhere in between the often optimal results for $\xi_{syn} = 7$ and the more indicative results for $\xi_{syn} = 30$. In short, optimal performance is achieved only for high population Fano Factors ($\rho \geq 3.75$, rightmost black triangle in Fig. 6(c)), while the presence of higher-order correlations is reliably detected ($\hat{\xi} \gg 0$) for a wide range of parameters ($\Lambda > 70$ Hz, Fig. 6(a), $T \geq 80$ s, Fig. 6(b), and $\rho \geq 1.02$ Hz, Fig. 6(c)).

The qualitative similarity between Fig. 6(a) and (b) result from the identical effect of the varied parameters, Λ and T , on the expected spike count $n = \Lambda T$ (Fig. 6(a), (b), top x axis). As n affects the quality of the estimators k_i , both Λ and T influence test results in a similar manner. The similar ρ -dependence results from the fact that increasing ρ increases the probability for high-amplitude events $f_A(\xi_{syn})$, which simplifies their detection (Fig. 6(c)).

Changes in the bin width h (Fig. 6(d)) affect CuBIC's performance for all values of ξ_{syn} in a very similar way. Increasing h decreases the lower percentile from $\xi_{05} \gg 0$ at $h = 0.1$ ms to $\xi_{05} = 0$ for $h = 100$ ms, irrespective of ξ_{syn} . The upper percentiles ξ_{95} (upper lines) can even show non-monotonic behavior: for $\xi_{syn} = 15$ (green) and $\xi_{syn} = 30$ (yellow), ξ_{95} increases for values between $1 \leq h \leq 60$ ms and decreases for $h > 60$ ms. The median of $p(\hat{\xi})$, which lies roughly in the middle of ξ_{05} and ξ_{95} , however, remains approximately constant for $h \leq 60$ ms, at least for $\xi_{syn} = 15$ and $\xi_{syn} = 30$ (green and yellow, respectively). Therefore, the average lower bound remains approximately constant as long as $h \leq 10$ ($L \geq 10^4$, top axis), but only the variability of $\hat{\xi}$, i.e. the distance between ξ_{05} and ξ_{95} , increases for this range of h .

We wish to stress that the larger variability of $\hat{\xi}$ for increased bin sizes is not so much due to the increased average spike count per bin $\mu = \Lambda h$, but rather the resulting decreased sample size $L = \frac{T}{h}$ (Fig. 6(d), top x -axis). The expected spike count $\mu = \Lambda h$ changes with the population rate Λ and the bin width h in the exact same manner. However, while increasing μ via Λ (Fig. 6(a)) from the default value $\Lambda = 10^3$ Hz (crosses) to $\Lambda = 10^4$ Hz does not alter test performance for $\xi_{syn} = 7$ (dark blue lines), increasing μ via the bin width from $h = 1$ ms to $h = 10$ ms (Fig. 6(d)) drastically reduces the lower percentile ξ_{05} (lower dark blue line

in Fig. 6(d)). The decrease in sample size increases the variance of our test statistic (compare Eq. (7)) and therefore decreases test power.

4.2.4 Summary of results

The simulation results of the previous section show that CuBIC reliably detects the presence of higher-order correlations for a wide range of parameter values. As CuBIC is not designed to directly estimate the order of correlation, the inferred lower bound might not always correspond to the maximal order of correlation present in a given data set. The most important parameter is the sample size L , which is required to be $\geq 10^4$ to reliably detect higher-order correlations (Fig. 6(d)), corresponding to e.g. an experiment of $T = 50$ s duration analyzed with a bin size of $h = 5$ ms. Furthermore, total spike counts of $n = 10^4$ (e.g. $N = 100$ neurons firing at 1 Hz for $T = 100$ s; Fig. 6(a) and (b)) and a correlation parameter (population Fano Factor) in the order of $\rho = 1.087$ suffice for the test to reliably detect the presence of higher-order correlations (lower percentiles $\xi_{05} > 0$, green and dark blue lower lines Fig. 6(a–c)) if correlations in the data are of moderate order ($\xi_{syn} = 7$ and $\xi_{syn} = 15$, green and dark blue lines).

5 Discussion

The identification of active cell-assemblies in simultaneously recorded spike trains has been the topic of extensive research (Gerstein et al. 1985, 1989; Sakurai 1998; Martignon et al. 2000; Grün et al. 2002a, b; Harris 2005; Schrader et al. 2008). Analysis tools that go beyond pairs of neurons, however, face severe limitations due to limited sample sizes of typical electrophysiological recordings (Martignon et al. 1995, 2000; Brown et al. 2004). The cumulant based inference of higher-order correlations (CuBIC) presented in this study avoids the need for extensive sample sizes. This is achieved by a combination of three ingredients: (1) pooling, i.e. using the population spike count and its cumulants to measure correlations in the population; (2) using the compound Poisson process (CPP) as a parametric model for correlated spiking populations; and (3) exploiting the constraining relations among correlations of lower and higher order in large neuronal pools.

Initially, we presented CuBIC as a hierarchy of hypothesis tests $H_0^{m,\xi}$, indexed by the order of the estimated cumulants m and the maximal order of correlation ξ allowed in the null. The algorithm of Fig. 3 combines these tests to infer a lower bound $\hat{\xi}$ on the order of correlation in a given data set (see also

Section 3.3). As free parameters, the algorithm requires upper bounds for m (m_{max}) and ξ (ξ_{max}) to be chosen by the experimenter. To choose m_{max} , recall that CuBIC is extremely sensitive for higher-order correlations already with $m = 3$ (Fig. 6). For the parameters analyzed here, tests with $m = 4$ increased $\hat{\xi}$ only occasionally, as the smallest ξ that solved Eq. (9) typically yielded non-significant p -values (data not shown). Furthermore, the reliable estimation of high cumulants is known to require vast sample sizes. Therefore, although tests with $m > 3$ can in principle yield additional information, we do not expect that choosing the bound for m larger $m_{max} = 4$ improves test results in situations with sample sizes compatible with electrophysiological experiments. To choose the upper bound for ξ , note that ξ_{max} is also an upper bound for $\hat{\xi}$. As the maximal order of correlation in the data has to be smaller than the number of recorded neurons N , we know that $\hat{\xi} \leq N$ must hold. This suggests to choose $\xi_{max} = N$, as higher values of ξ_{max} cannot increase $\hat{\xi}$. If the analysis was based on multi-unit activity and the number of recorded neurons is unknown, an a posteriori criterion on whether the chosen value for ξ_{max} was high enough can be obtained by inspecting tests with $\xi = \xi_{max}$. That is, if $H_0^{m, \xi_{max}}$ was rejected for some m , hypothesis with $\xi > \xi_{max}$ might also have been rejected, potentially resulting in higher $\hat{\xi}$. Thus, if tests against $H_0^{m, \xi_{max}}$ are retained for all m , the chosen ξ_{max} was high enough.

The procedure was calibrated with rather restricted model classes, namely amplitude distributions with two isolated peaks. Thus, the question arises how differences in the correlation structure would influence CuBIC's performance. In the numerical examples of this study, CuBIC used only the first three cumulants of the data, which depend on the first three moments of the amplitude distribution (compare Eq. (3)). As a consequence, if the data under consideration had a different amplitude distribution, however with similar first three moments, CuBIC will produce similar results. If, for instance, the additional isolated peak at ξ_{syn} is exchanged by a somewhat distributed peak centered at ξ_{syn} , the inferred lower bound will be similar.

Importantly, CuBIC is not designed to directly estimate the “true” maximal order of correlation in the data. Furthermore, the neuron IDs that realize the correlations are currently not resolved. It is important to keep in mind, however, that approaches that pose such specific questions to the data face severe limitations with respect to the size of the analyzed populations (e.g., Shlens et al. 2006; see also Section 5.1). Thus, we regard the less specific lower bound $\hat{\xi}$ obtained by CuBIC as the price to pay when analyzing large populations. Renouncing the specific questions is precisely

what allows CuBIC to detect higher-order correlations in data sets with previously unreachable population size ($N \sim 100$) and sample sizes compatible with physiological experiments (see Figs. 4 and 6).

5.1 Model dependence of higher-order correlations

The role of correlations for cortical information processing has been lively debated (Abeles 1991; Shadlen and Newsome 1994; Riehle et al. 1997; Singer 1999; Abbott and Dayan 1999; Kohn and Smith 2005; Schneidman et al. 2006; Shlens et al. 2006; Laurent 2006; Pillow et al. 2008). While in the meantime pairwise correlations are widely accepted as a relevant feature of parallel spike trains, the relevance of higher-order correlations is controversial. In fact, several recent studies raised doubts about the importance of higher-order interactions. In two studies that explore multi-neuron firing patterns in primate (Shlens et al. 2006) and salamander (Schneidman et al. 2006) retina, data were collected for fairly long times (~ 1 hour), and distributions of the binary patterns P_N for subpopulations with $N \leq 10$ were estimated. Based on this, a surrogate distribution was generated that reproduces the pairwise correlations in the population, but disregards all higher-order parameters: the “maximum entropy distribution” P_2 . A comparison of P_N with P_2 provides estimates on the amount of information contained in the higher-order parameters. The results of both studies indicate that only a negligible fraction of the total information is contained in the higher-order parameters (less than 2% in primate retina according to Shlens et al. 2006, and less than 10% in salamander retina according to Schneidman et al. 2006), and a recent study reports similar values for neo-cortical networks (Tang et al. 2008). In other words: almost the entire information available in the distribution of multi-neuron firing patterns can already be inferred from pairwise correlations. So why bother estimating higher-order parameters?

The consequences of the maximum entropy results for the relevance of CuBIC are difficult to judge. First of all, the maximum entropy approach does not appear to be very sensitive for higher-order correlations. As shown by Shlens et al. (2006), additional higher-order coincidences with a rate of ~ 0.3 Hz among the $N = 7$ analyzed neurons would not have altered their results. This “coincidence rate”, however, lies in the parameter range of Set3 of Fig. 4 (here $\nu_{15} = 0.41$ Hz). In this range CuBIC detected correlations of order ≥ 13 , illustrating the comparatively low sensitivity of the maximum entropy approach for existing higher-order correlations.

Second, the combinatorial explosion of the parameter space limits maximum entropy approaches to only relatively small populations ($N \lesssim 10$), unless additional assumptions are imposed on the data (Montani et al. 2009; Shlens et al. 2009). Theoretical evidence suggests, however, that the relevance of higher-order correlations increases with the size of the neuronal population (Roudi et al. 2009). In fact, using $N = 24$ unsorted multi-unit spike trains of the rodent whisker system, Montani et al. (2009) report a significant increase in information if additional to the pairwise interactions also triplet interactions are taken into account. Taken together, we firmly believe that the analysis of large groups with methods that are sensitive to higher-order correlations is inevitable to uncover their role in cortical computation.

Our third remark concerns the conceptual differences between cumulant correlation as measured here, and the higher-order parameters used in the maximum entropy framework. The latter is based on the prominent binary exponential family (Martignon et al. 1995; Nakahara and Amari 2002; Martignon et al. 2000; Gütig et al. 2003; Del Prete et al. 2004; Shimazaki et al. 2009). In this setting, spike trains are represented as binary sequences, such that $X_i(s) = 1$ if the i th neuron has one or more spikes in the s th bin, and $X_i(s) = 0$ otherwise. Assuming stationarity and no memory, the multivariate distribution function of the binary pattern vector \mathbf{X} admits the representation

$$p_{\mathbf{X}} = \exp \left[\sum_i \theta_i X_i + \sum_{i < j} \theta_{ij} X_i X_j + \sum_{i < j < k} \theta_{ijk} X_i X_j X_k + \dots + \theta_{1\dots N} X_1 \dots X_N - \psi \right].$$

In this framework, higher-order correlations are reflected in nonzero “higher-order parameters” θ . For example, θ_{ijk} quantifies third-order interactions of the triplet (X_i, X_j, X_k) . Importantly, these higher-order parameters are neither equivalent, nor related in any straightforward way to the cumulant correlations employed by CuBIC (see Section 2.1.1 and Appendix A). A major difference, for instance, results from the equivalence of cumulant-based correlations with the concept of coincident events addressed by the CPP (Theorem 1). Given that these coincidences are perfectly synchronized, changes in the bin width h do not affect cumulant-based correlations. In contrast, as changes in the bin width alter the shape of $p_{\mathbf{X}}$, the interaction parameters θ inevitably change with h . Observe that the bin size dependence of θ does not result from the fact that larger bins measure correlations on a

broader time scale as the time scale of correlation was assumed to be instantaneous, but is rather inherent to the exponential family. More generally, cumulant correlations determine to what extent expectation values of products factorize (e.g., Stratonovich 1967, and our Section 2.1.1), while the higher-order parameters in the exponential family measure to what extent the probability of certain binary patterns can be explained by the probabilities of its sub-patterns (e.g., Gütig et al. 2003). Taken together, the two frameworks are designed for different purposes, and can hence be expected to be sensitive to different features of the data. A better insight into their relationship presents an important path for future research (see also Staude et al. 2010, for a more detailed discussion).

Our main motivation to use cumulant correlations is that they enable us to analyze large populations of simultaneously recorded spike trains. As opposed to the θ parameters, cumulant correlations can be inferred from the population spike count Z . Thereby, the multivariate problem to estimate correlations for all $2^N - 1$ subgroups is transferred into a parsimoniously parametrized univariate problem with N parameters. The small sample size required by CuBIC and its ability to analyze populations of $N > 100$ neurons is a direct consequence of this property. Furthermore, the compound Poisson process presents a very intuitive interpretation of cumulant correlations (Theorem 1) that directly implements the concept of temporally coordinated spikes (see also Staude et al. 2008). And finally, the exponential family captures imprecise correlations only at the cost of ignoring spikes by clipping, as it requires the variables X_i to be binary (e.g., Del Prete et al. 2004), whereas cumulant correlations can be measured among arbitrary spike counts. We conclude that the adequate framework to analyze a given data set should be carefully chosen in view of the specific scientific question. For the detection of temporally coordinated spikes in large neuronal populations, we believe that cumulant-based correlations as exploited by CuBIC are well-suited.

5.2 Relation to other tools

While algorithms for the generation of correlated spike trains are becoming available (e.g., Kuhn et al. 2003; Niebur 2007; Staude et al. 2008; Brette 2009; Macke et al. 2008; Krumin and Shoham 2009), only few of these models can be exploited for data analysis. In an approach similar to ours, Ehm et al. (2007) derive estimators of the continuous time parameters (analogous to our carrier rate and amplitude distribution) from the Poissonized bin counts (here: “population spike

count”) via Fourier inversion of the empirical characteristic function. Instead of inferring lower bounds as in CuBIC, “empirical de-Poissonization” presented by Ehm et al. (2007) thus directly estimates the orders of correlation in the data. However, the more explicit parameters appear to impose constraints on the data, such that particular care is required if the expected number of spikes per bin exceeds a critical value. A comparison of CuBIC and the empirical de-Poissonization to resolve their respective advantages and disadvantages is currently under way (Reimer et al. 2009).

Besides the higher-order approaches described above, the majority of available analysis tools for massively parallel spike trains either exploit pairwise statistics to infer inter-correlated groups by using some sort of clustering (Gerstein and Aertsen 1985; Gerstein et al. 1985; Berger et al. 2007; Grün et al. 2008b; Fujisawa et al. 2008), or test the occurrence of specific patterns against the assumption of complete (Grün et al. 2002a; Pipa et al. 2008) or partial independence (Berger et al. 2010). The three examples depicted in Fig. (4), however, illustrate the limitations of both approaches. First, the three data sets have identical second order statistics, making it impossible for pairwise approaches to differentiate between them. Second, the combinatorial explosion prevents the analysis of populations sizes of $N \sim 100$ neurons with pattern-based approaches. And third, it is unclear whether either approach is sensitive enough to detect correlated groups when pairwise correlation are as low as discussed here.

A novel approach that is recently gaining attention starts from the analytically tractable generalized linear model (GLM), and estimates its parameters by maximum likelihood techniques (Paninski 2004; Truccolo et al. 2005; Okatan et al. 2005; Pillow et al. 2008). The power of this framework is that estimated parameters can be interpreted in terms of network properties like e.g. synaptic interactions between nerve cells, stimulus-response properties and the history dependence of individual spike trains. To understand how such anatomical and biophysical attributes of neuronal populations relate to higher-order correlations in the joint spiking of groups of neurons as detected by CuBIC is subject of current research.

Constraining relations among correlations of different order have also been studied by Amari et al. (2003), Johnson and Goodman (unpublished manuscript) and Benucci et al. (2004). The latter study defined a combinatorial algorithm to obtain upper bounds for certain observables. Closer to our approach is the study of Johnson and Goodman (unpublished manuscript), who derive explicit constraints among correlation coefficients of different orders within the CPP framework.

However, both studies do not provide confidence intervals, hence are not applicable to data analysis in a straightforward way. Amari et al. (2003) investigated the relation between population variance and higher-order correlations in the information geometric framework (Amari and Nagaoka 2000) for the limit of large population sizes ($N \rightarrow \infty$). The result is similar to our findings in the sense that a high population variance implies higher-order correlations. However, results from large N provide only qualitative suggestions for data analysis.

5.3 Consequences of underlying model

An important ingredient of the present study is to use the CPP as a model for correlated spike trains. A consequence of this model-driven approach is that the reliability of obtained test results depends crucially on how well the CPP is suited to describe populations of spiking neurons in general, and assembly activity in particular.

5.3.1 Stationarity

We presented the CPP as a model of a stationary population, i.e. where individual spike trains have constant firing rates. However, cortical spike trains typically undergo both stimulus-driven and internally generated changes in firing rate. To analyze also such non-stationary data sets, the relatively small required sample size enables us to use a “sliding-window” approach (see e.g., Grün et al. 2002b), where data are cut into quasi-stationary windows of e.g. 100 ms and the statistics is therein obtained over trials. For the case of 100 trials analyzed with a bin width of $h = 1$ ms, this would yield $L = 10^4$ samples, which is perfectly in the required range for CuBIC to yield lower bounds close to the actual order of correlation (compare Fig. 6(d)). An alternative could be to directly include the non-stationarities in the model (Stauder and Rotter 2009). Both approaches present important aspects of future research.

5.3.2 Single-process properties

Recent studies of cortical spike trains criticize the use of Poisson processes as models for single spike trains (Amarasingham et al. 2006; Nawrot et al. 2008). Without doubt, inclusion of process types deviating from Poisson in the generating model presents a major challenge for the presented procedure. However, criticism against the Poisson model for cortical spike trains is typically motivated by either non-exponential

interval statistics (e.g., Maimon and Assad 2009), or non-Poissonian spike count distributions for large bin sizes ($h > 50$ ms) and high firing rates under 'optimal' stimulation ($\lambda > 50$ Hz; e.g. Amarasingham et al. 2006). Both of these aspects are of little impact if the counting statistics is sampled with small ($h \leq 10$ ms) bin sizes. In such a scenario, and in particular for low firing rates (e.g. Lennie 2003; Lee et al. 2006; Olshausen and Field 2006; Maldonado et al. 2008), neurons will have at most one spike per bin, leading to almost identical counting statistics for Poisson processes and other process types (e.g. gamma processes; van Vreeswijk 2006; Nawrot et al. 2008). Note also that the CPP does not directly model individual processes as Poisson processes, but only assumes the carrier processes to be Poisson. Importantly, the latter can be a good approximation even if individual processes deviate from strict Poisson statistics. For instance, if background spikes are Poisson and high-amplitude events are very regular but have a very low rate (as in Fig. 4), the carrier process is still well approximated by a Poisson process. More generally, the sum of increasingly sparse point processes converge to a Poisson process (Daley and Vere-Jones 2005, but see Lindner 2006; Câteau and Reyes 2006). Although these aspects need to be considered and will be subject of further research, we therefore expect CuBIC to yield reliable results even if single processes deviate from the Poisson assumption (Staude et al. 2007).

5.3.3 Precision of coincidences

Cortical correlations are reported on a variety of time scales, from precise correlations within a few few ms (Grün et al. 1999; Kohn and Smith 2005; Pazienti et al. 2008; Desbordes et al. 2008) to slow covariations of firing rates on a time scale of > 50 ms (e.g., Kohn and Smith 2005; Smith and Kohn 2008). The perfect temporal precision of the joint spikes generated by the CPP may thus be questioned as a biologically plausible model of experimental spike trains. In the present study, the CPP was not intended to realistically model spike trains in continuous time, however. We presented the CPP strictly to interpret the correlated counting variables X_i , obtained with a given bin width h . Allowing a temporal jitter in the synchronous events, e.g. by jittering the coincident events during the inserting process (see e.g. Staude et al. 2008; Brette 2009), results in comparable counting variables X_i if this jitter is small compared to the applied bin size (Pazienti et al. 2008). Although its sensitivity for such imprecise coincidences remains to be assessed systematically, we expect CuBIC to tolerate imprecise coincidences with small but realistic jitter.

A somewhat different issue is the detection of correlations with systematic temporal offset, e.g. due to functional or "effective" connectivity (Aertsen et al. 1989; Fujisawa et al. 2008). As an extension of CuBIC to detect also non-zero-lag higher-order correlations, one may shift individual spike trains against each other and then analyze this shifted population (cmp. Grün et al. 1999). As tests for different shifts will generally not be independent, however, appropriate corrections that account for multiple testing have to be applied. Furthermore, the combinatorial explosion that emerges from shifting all spike train against each other with different delays requires smart algorithms to reduce the resulting computational cost.

5.3.4 Positivity of the correlations

A limitation of the CPP is that the insertion of synchronous spike events models spike coordination, not spike avoidance. As a consequence, correlations in the CPP can only be positive (Staude et al. 2008; Brette 2009). The CPP cannot model negatively correlated spike counts, that is, troughs in the cross-correlation function. Whether or not this is relevant for CuBICs applicability depends crucially on the properties of electrophysiological data. The rare reports of negative correlations, especially among cortical spike trains (e.g., Nevet et al. 2007), however, makes us confident that this is not an important limitation.

6 Conclusions

Based on Hebb's original ideas (Hebb 1949), the coordinated spiking of large neuronal groups is the conceptual foundation of various "brain theories" (e.g., Eggermont 1990; Abeles 1991). However, limitations of available analysis tools has rendered the exploitation of the increasing number of simultaneously observable neurons extremely difficult. The CuBIC method presented in this study is a novel procedure to detect higher-order correlations in massively parallel spike trains, which overcomes these limitations. In particular, the sample size required by our method lies in a range that is compatible with electrophysiological experiments. CuBIC assumes data to be stationary and single processes the sum of which is appropriately modeled by the CPP, and its extension to account for dynamically changing, non-Poissonian data is the most important aspect of future work. Taken together, we regard CuBIC as a unique opportunity to analyze large populations of neurons for coordinated spike timing.

Acknowledgements We would like to thank Abigail Morrison, Imke Reimer, and Stefano Cardanobile for helpful comments on earlier versions of this manuscript. Funded by NaFöG Berlin, the German Ministry for Education and Research (BMBF grants 01GQ01413 and 01GQ0420 to the BCCN Berlin and Freiburg, respectively), the IGPP Freiburg, the Stifterverband für die Deutsche Wissenschaft, the Initiative and Networking Fund of the Helmholtz Association within the Helmholtz Alliance on Systems Biology, and the German Research Foundation (DFG SFB 780).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

A Higher-order correlations and the CPP

A pair of random variables X_1 and X_2 is uncorrelated, if the expectation of their product factorizes

$$E[X_1 X_2] = E[X_1] E[X_2],$$

and the covariance $\text{Cov}[X_1, X_2] := E[X_1 X_2] - E[X_1] E[X_2]$ measures the degree of correlation between them. As the covariance is bounded by the square root of the individual variances, one quantifies pairwise correlations by the dimensionless correlation coefficient

$$c := \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1] \text{Var}[X_2]}}. \tag{13}$$

Cumulant-correlations are a generalization of the (co-)variance to measure correlations among more than two r.v.s (e.g., Stratonovich 1967; Stuart and Ord 1987; Streitberg 1990; Papoulis and Pillai 2002; Gardiner 2003; Staude et al. 2010). The cumulants $\kappa_m[Z]$ ($m \in \mathbb{N}$) of a r.v. Z are defined in terms of the coefficients of the power series expansion of the logarithm of its characteristic function

$$\Theta_Z(u) := \log E[e^{iuZ}] = \sum_{m=1}^{\infty} i^m \frac{u^m}{m!} \kappa_m[Z], \tag{14}$$

where $\Theta_Z(u)$ is called the “cumulant generating function” (c.g.f.) of Z . To relate the cumulants to the more familiar (raw) moments $\mu_m[Z] := E[Z^m]$, recall that the latter are defined in terms of the coefficients of the power series expansion of the characteristic function

$$\gamma_Z(u) := E[e^{iuZ}] = \sum_{m=1}^{\infty} i^m \frac{u^m}{m!} \mu_m[Z]. \tag{15}$$

Expressions relating cumulants to moments are obtained by writing the logarithm of the right hand side of Eq. (15) as a power series in u , and collecting the coefficients. For the first two cumulants, this yields

$$\kappa_1[Z] = E[Z], \text{ and } \kappa_2[Z] = E[Z^2] - E[Z]^2 = \text{Var}[Z].$$

Similar expressions for higher m are increasingly complicated, but algorithms for their computation are available (e.g. Stuart and Ord 1987; Di Nardo et al. 2008).

Given two independent r.v.s X_1 and X_2 , we have

$$\begin{aligned} \Theta_{X_1+X_2}(u) &:= \log E[e^{i(uX_1+uX_2)}] \\ &= \log E[e^{iuX_1} e^{iuX_2}] \\ &\stackrel{\{X_i\} \text{ independent}}{=} \log (E[e^{iuX_1}] E[e^{iuX_2}]) \\ &= \log E[e^{iuX_1}] + \log E[e^{iuX_2}] \\ &= \Theta_{X_1}(u) + \Theta_{X_2}(u) \end{aligned}$$

Writing the c.g.f.s as power series and comparing coefficients yields the central property of cumulants: the cumulant of the sum of independent random variables is the sum of the individual cumulants

$$\kappa_m[X_1 + X_2] = \kappa_m[X_1] + \kappa_m[X_2] \text{ for all } m$$

If X_1 and X_2 are not independent, the cumulants of $X_1 + X_2$ contain also “mixed”, or “connected” cumulants. For $m = 2$, we obtain the well-known equation

$$\begin{aligned} \kappa_2[X_1 + X_2] &= \kappa_2[X_1] + \kappa_2[X_2] + 2\kappa_{1,1}[X_1, X_2] \tag{16} \\ &= \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}[X_1, X_2] \tag{17} \end{aligned}$$

Equation (16) implies that the second cumulant of $X_1 + X_2$ is additive-linear, if and only if $\kappa_{1,1}[X_1, X_2] = \text{Cov}[X_1, X_2] = 0$, i.e. if X_1 and X_2 are uncorrelated. Higher-order cumulant correlations generalize the covariance in exactly this sense: they determine the “degree of linearity” of the higher cumulants (e.g., Stratonovich 1967; Gardiner 2003). Formally, higher-order cumulants arise as coefficients of the c.g.f. of vector-valued variables $\mathbf{X} = (X_1, \dots, X_N)$

$$\Theta_{\mathbf{X}}(\mathbf{u}) := \log E[e^{i \sum_{j=1}^N u_j X_j}] \tag{18}$$

$$= \sum_{m=1}^{\infty} i^m \sum_{\{J=(j_1, \dots, j_N) | \sum_j j_i = m\}} \kappa_J[\mathbf{X}] \frac{u_1^{j_1} \dots u_N^{j_N}}{j_1! \dots j_N!} \tag{19}$$

Higher-order cumulant correlations exist, if the corresponding mixed cumulant $\kappa_J[\mathbf{X}]$ is non-zero. To be precise,

Definition 1 Let \mathbf{X} be a random vector. Denote the number of non-zero elements of a multi-index $J = (j_1, \dots, j_N) \in \mathbb{N}^N$ by n_J . Then the components of \mathbf{X} are said to have correlations of order m if there exists $J \in \mathbb{N}^N$ with $n_J = m$ such that the corresponding mixed cumulant is non-zero, i.e. $\kappa_J[\mathbf{X}] \neq 0$.

For multi-indices J with only one non-zero element, e.g. $J = (2, 0, 0, \dots, 0)$, the mixed cumulants κ_J in Eq. (18) are simply the univariate cumulants of Eq. (14): $\kappa_{(2,0,\dots,0)}[\mathbf{X}] = \kappa_2[X_1] = \text{Var}[X_1]$. Furthermore, a generalization of $\text{Cov}[X, X] = \text{Var}[X]$ is that mixed cumulants with identical arguments are the univariate cumulants of the appropriate order

$$\kappa_J[X, \dots, X] = \kappa_{j_1+\dots+j_N}[X]. \tag{20}$$

Evidently, the c.g.f. of a sum of r.v.s $Z = \sum_{i=1}^N X_i$ is the multivariate c.g.f. with all arguments set to u

$$\begin{aligned} \Theta_Z(u) &:= \log E[e^{iu \sum_{j=1}^N X_j}] = \log E[e^{i \sum_{j=1}^N u X_j}] \\ &= \Theta_{\mathbf{X}}(u, \dots, u). \end{aligned}$$

Inserting into Eq. (18) and comparing coefficients with Eq. (14) yields

$$\kappa_m[Z] = \sum_{\{J=(j_1,\dots,j_N) \mid \sum_j j_i=m\}} \frac{m!}{j_1! \cdots j_N!} \kappa_J[\mathbf{X}],$$

showing in particular that the m th cumulant of Z depends on correlations of maximal order m .

Proof of Theorem 1 To prove Theorem 1 of Section 2.4, we present a multivariate version of the compound Poisson process (CPP) defined in Section 2.2 (compare Ehm et al. 2007; Staude et al. 2010). Consider N neurons $\{x_i(t)\}_{i \leq N}$ and let \mathcal{A} be the set of all nonempty subsets of $\{1, \dots, N\}$. For each $M \in \mathcal{A}$, we assign a stationary Poisson process $y'_M(t)$ with rate ν'_M that determines the points in time where the neurons $\{x_i(t)\}_{i \in M}$ have a joint spike event. Assuming all the $y'_M(t)$ s to be independent, it is straightforward to verify that $z(t) = \sum_{M \in \mathcal{A}} y'_M(t) |M|$ is a CPP as defined in Section 2.2, where $|M|$ denotes the number of elements of M . The processes $y_l(t)$ of Section 2.3 are given by $y_l(t) = \sum_{|M|=l} y'_M(t)$, and we obtain the parameters of $z(t)$ (carrier rate ν and amplitude distribution f_A) by $\nu = \sum_{M \in \mathcal{A}} \nu'_M$ and $f_A(k) = \frac{1}{\nu} \sum_{|M|=k} \nu'_M$. Evidently, f_A assigns non-zero probabilities to events of amplitude $\geq m$ if and

only if at least one of the rates ν'_M with $|M| \geq m$ is positive. Using the above definition, Theorem 1 is thus equivalent to the equivalence of the following two statements:

- (a) at least one m th order mixed cumulant is non-zero, i.e. there exists a multi-index $J \in \mathbb{N}^N$ with $n_J \geq m$ such that $\kappa_J[\mathbf{X}] \neq 0$
- (b) at least one of the processes $y'_M(t)$ that determines coincidences of at least m of the spike trains $x_i(t)$ has non-zero rate, i.e. there exists $M \in \mathcal{A}$ with $|M| \geq m$ and $\nu'_M > 0$

The proof is not difficult, yet notationally cumbersome for arbitrary N . The idea is to decompose Z into the counting variables Y'_M of the $y'_M(t)$ and show that their mixed cumulants $\kappa_J[Y'_M]$ vanish if J has more non-zero elements than the multi-index associated to M .

With the variables Y'_M , the argument of $\kappa_J[\mathbf{X}]$, the pattern vector $\mathbf{X} = (X_1, \dots, X_N)$, can be written as

$$\mathbf{X} = \sum_{t=1}^N \sum_{|M|=t} \delta_M \cdot Y'_M, \tag{21}$$

where $\delta_M \in \{0, 1\}^N$ is the binary vector whose i th component is 1 if $i \in M$ and 0 otherwise. For given $J \in \mathbb{N}^N$, we thus have

$$\kappa_J[\mathbf{X}] = \kappa_J \left[\sum_{t=1}^N \sum_{|M|=t} \delta_M \cdot Y'_M \right] = \sum_{t=1}^N \sum_{|M|=t} \kappa_J [\delta_M \cdot Y'_M], \tag{22}$$

where the second equality follows from the independence of the Y'_M s. Clearly $\kappa_J[\delta_M \cdot Y'_M] = 0$ for all M with $n_J > |M|$, as in this case the order of the measured correlation n_J exceeds the number of non-zero components of the argument. In fact, $\kappa_J[\delta_M \cdot Y'_M] \neq 0$ if and only if $\text{supp}(J) \subset M$, as these are the cases where we do not correlate with the variable 0 (the support $\text{supp}(J)$ is the set of non-zero elements of J). In this case, $\kappa_J[\delta_M \cdot Y'_M]$ is a cumulant of the Poisson variable Y'_M which satisfies $\kappa_J[\delta_M \cdot Y'_M] = \kappa_{j_1+\dots+j_N}[Y'_M] = \nu'_M h$ (Eq. (20)). Hence,

$$\kappa_J [\delta_M \cdot Y'_M] = \begin{cases} \nu'_M h, & \text{supp}(J) \subset M \\ 0, & \text{otherwise.} \end{cases}$$

This implies

$$\kappa_J[\mathbf{X}] = \sum_{\{M \mid \text{supp}(J) \subset M\}} \nu'_M h. \tag{23}$$

Now we can prove the equivalence of statements (a) and (b).

“(a)⇒(b)”: Let $J^+ \in \mathbb{N}^N$ be such that $n_{J^+} \geq m$ and $\kappa_{J^+}[\mathbf{X}] \neq 0$. Then by Eq. (23), $\sum_{\{M|\text{supp}(J^+) \subset M\}} v'_M h \neq 0$, and as all $v'_M \geq 0$ there must be at least one $M^+ \in \mathcal{A}$ with $\text{supp}(J) \subset M^+$ and $v'_{M^+} > 0$. As $n_{J^+} \geq m$, this M^+ must have at least m elements, hence $|M^+| \geq m$.

“(b)⇒(a)”: Let $M^+ \in \mathcal{A}$ with $|M^+| \geq m$ and $v'_{M^+} > 0$ and let J^+ be the associated multi-index. Then $n_{J^+} = |M^+| \geq m$ and

$$\kappa_{J^+}[\mathbf{X}] = \sum_{\{M|\text{supp}(J^+) \subset M\}} v'_M h \geq v'_{M^+} h > 0 \quad \square$$

B Model parameters and population statistics

This appendix relates the parameters of the compound Poisson process, i.e. the carrier rate ν and the amplitude distribution f_A , to parameters used in the context of physiological data, i.e. firing rates λ and pairwise count correlation coefficients c , for a population of N neurons.

Firing rates To compute the average rate λ , observe that $\lambda = \frac{1}{hN} \sum_{i=1}^N E[X_i] = \frac{1}{Nh} E[Z] = \frac{1}{Nh} \kappa_1[Z]$. With Eq. (3), we thus have

$$\lambda = \frac{E[A]}{N} \nu \tag{24}$$

Pairwise Correlations To compute the count correlation coefficient c for a fixed bin width h (Eq. (13)), recall that by definition of the variance

$$\kappa_2[Z] = \text{Var}[Z] = \sum_{i=1}^N \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j].$$

Denoting the population average variance and covariance by

$$v := \frac{1}{N} \sum_{i=1}^N \text{Var}[X_i], \quad r := \frac{1}{N(N-1)} \sum_{i \neq j} \text{Cov}[X_i, X_j],$$

the population Fano-Factor $\frac{\kappa_2[Z]}{\kappa_1[Z]}$ reads

$$\begin{aligned} \frac{\kappa_2[Z]}{\kappa_1[Z]} &= \frac{\sum_{i=1}^N \text{Var}[X_i]}{\sum_{i=1}^N E[X_i]} + \frac{\sum_{i \neq j} \text{Cov}[X_i, X_j]}{\sum_{i=1}^N E[X_i]} \\ &= \frac{Nv}{\sum_{i=1}^N E[X_i]} + \frac{N(N-1)r}{\sum_{i=1}^N E[X_i]} \end{aligned}$$

The Poisson property of the individual processes implies $E[X_i] = \text{Var}[X_i]$, which yields $\sum_{i=1}^N E[X_i] = Nv$.

Using the identity of the population Fano-Factor with the moments ratio $\frac{\kappa_2[Z]}{\kappa_1[Z]} = \frac{\mu_2[A]}{\mu_1[A]} = \rho$, we finally obtain

$$\rho = 1 + (N-1) \frac{r}{v} \tag{25}$$

Note that by Eq. (25), the pairwise correlations, contained in r , do not depend on the details of the amplitude distribution f_A , but are completely determined by the ratio of its first two moments ρ .

If the population is homogeneous (all neurons have the same firing rate, all pairs are equally correlated, and so on), the average quantities v and r are the actual variance and covariance of the processes. In this case, the count correlation coefficient is given by $c = \frac{r}{v}$, which implies

$$c = \frac{\rho - 1}{N - 1}. \tag{26}$$

Conversely, the moments ratio is given by

$$\rho = 1 + c(N - 1). \tag{27}$$

Correlated subgroups Now consider the case where N_C of the N neurons form a homogeneously correlated subgroup while the remaining $N_I = N - N_C$ neurons are independent (see rasters in Fig. 4 for examples). In this case, we can decompose Z into the contribution of the correlated, Z_C , and the uncorrelated, Z_I , subpopulations: $Z = Z_C + Z_I$. Now $E[Z_C] = \frac{N_C}{N} E[Z]$ and $E[Z_I] = \text{Var}[Z_I] = \frac{N_I}{N} E[Z]$. We can thus write

$$\begin{aligned} \rho E[Z] &= \text{Var}[Z] = \text{Var}[Z_C] + \text{Var}[Z_I] \\ &= \text{Var}[Z_C] + \frac{N_I}{N} E[Z]. \end{aligned}$$

Dividing by $E[Z_C] = \frac{N_C}{N} E[Z]$ yields

$$\frac{\rho N}{N_C} = \frac{\text{Var}[Z_C]}{E[Z_C]} + \frac{N_I}{N_C}. \tag{28}$$

For a homogeneous population as the correlated subgroup, the Fano Factor $\frac{\text{Var}[Z_C]}{E[Z_C]}$ is related to the pairwise correlation coefficient via Eq. (27), i.e. $\frac{\text{Var}[Z_C]}{E[Z_C]} = 1 + c(N_C - 1)$. Inserting into Eq. (28) and solving for c yields the correlation coefficient of the correlated subgroup

$$c = \frac{N(\rho - 1)}{N_C(N_C - 1)} \tag{29}$$

Single interaction process The amplitude distributions used in Section 4 consist of two isolated peaks

$$f_A(\xi) = f_A(1)\delta_{1,\xi} + f_A(\xi)\delta_{\xi_{syn},\xi} \tag{30}$$

where $\delta_{i,k}$ is the Kronecker-Delta. The “single interaction process” (Kuhn et al. 2003) is the special case

with $\xi_{in} = N$. Because $f_A(1) + f_A(\xi_{syn}) = 1$, this model has three free parameters ν , $f_A(1)$ and ξ_{syn} , which are related to N , λ and ρ (or c) by Eqs. (24) and (29). Straightforward computation yields that homogeneous populations with mean firing rate λ and correlation parameter ρ are achieved by setting

$$\nu = \frac{\lambda N[\rho - 1 + \xi_{syn}(\xi_{syn} - \rho)]}{\xi_{syn}(\xi_{syn} - 1)}$$

and

$$f_A(1) = \frac{\xi_{syn}(\xi_{syn} - \rho)}{(\rho - 1) + \xi_{syn}(\xi_{syn} - \rho)}$$

Given λ and ρ are constrained, the event amplitude ξ_{syn} that realizes the correlation can be freely chosen as long as $\nu > 0$ and $\eta \in [0, 1]$. This is guaranteed if

$$1 \leq \rho \leq \xi_{syn} \leq N.$$

References

- Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*, 91–101.
- Abeles, M. (1982). Role of cortical neuron: Integrator or coincidence detector? *Israel Journal of Medical Sciences*, *18*, 83–92.
- Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex* (1st Ed.). Cambridge: Cambridge University Press.
- Aertsen, A., Gerstein, G., Habib, M., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of “effective connectivity”. *Journal of Neurophysiology*, *61*(5), 900–917.
- Amarasingham, A., Chen, T.-L., Harrison, M. T., & Sheinberg, D. L. (2006). Spike count reliability and the poisson hypothesis. *Journal of Neuroscience*, *26*(4), 801–809.
- Amari, S.-i., & Nagaoka, H. (2000). *Information geometry*. New York: AMS and Oxford University Press.
- Amari, S.-i., Nakahara, H., Wu, S., & Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. *Neural Computation*, *15*, 127–142.
- Benucci, A., Verschure, P. F. M. J., & König, P. (2004). High-order events in cortical networks: A lower bound. *Physical Review E*, *70*(051909).
- Berger, D., Borgelt, C., Louis, S., Morrison, A., & Grün, S. (2010). Efficient identification of assembly neurons within massively parallel spike trains. *Computational Intelligence and Neuroscience*, *21*. doi:10.1155/2010/439648.
- Berger, D., Warren, D., Normann, R., Arieli, A., & Grün, S. (2007). Spatially organized spike correlation in cat visual cortex. *Neurocomputing*, *70*(10–12), 2112–2116.
- Bohte, S. M., Spekreijse, H., & Roelfsema, P. R. (2000). The effects of pair-wise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neural Computation*, *12*, 153–179.
- Brette, R. (2009). Generation of correlated spike trains. *Neural Computation*, *21*, 188–215.
- Brown, E. N., Kaas, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature Neuroscience*, *7*(5), 456–461.
- Câteau, H., & Reyes, A. (2006). Relation between single neuron and population spiking statistics and effects on network activity. *Physical Review Letters*, *96*, 058101.
- Daley, D. J., & Vere-Jones, D. (2005). *An introduction to the theory of point processes, Vol. 1: Elementary theory and methods* (2nd Ed.). New York: Springer.
- Del Prete, V., Martignon, L., & Villa, A. E. P. (2004). Detection of syntonies between multiple spike trains using a coarse-grain binarization of spike count distributions. *Network: Computation in Neural Systems*, *15*, 13–28.
- Desbordes, G., Jin, J., Weng, C. L. N. A., Stanley, G. B., & Alonso, J.-M. (2008). Timing precision in population coding of natural scenes in the early visual system. *PLoS Biology*, *6*, e324.
- Di Nardo, E., Guarino, G., & Senato, D. (2008). A unifying framework for k-statistics, polykays and their multivariate generalizations. *Bernoulli*, *14*(2), 440–468.
- Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, *402*(6761), 529–533.
- Eggermont, J. J. (1990). The correlative brain. In *Studies of brain function* (Vol. 16). Berlin: Springer.
- Ehm, W., Staude, B., & Rotter, S. (2007). Decomposition of neuronal assembly activity via empirical de-poissonization. *Electronic Journal of Statistics*, *1*, 473–495.
- Fujisawa, S., Amarasingham, A., Harrison, M., & Buzsaki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neuroscience*, *11*, 823–833.
- Gardiner, C. W. (2003). *Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer Series in Synergetics* (3 Ed., No. 13). New York: Springer.
- Gerstein, G. L., & Aertsen, A. M. H. J. (1985). Representation of cooperative firing activity among simultaneously recorded neurons. *Journal of Neurophysiology*, *54*(6), 1513–1528.
- Gerstein, G. L., Bedenbaugh, P., & Aertsen, A. (1989). Neuronal assemblies. *IEEE Transactions on Biomedical Engineering*, *36*, 4–14.
- Gerstein, G. L., Perkel, D. H., & Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously recorded populations of neurons: Detection and measurement. *Journal of Neuroscience*, *5*(4), 881–889.
- Grün, S., Abeles, M., & Diesmann, M. (2008a). Impact of higher-order correlations on coincidence distributions of massively parallel data. In *Lecture notes in computer science. ‘Dynamic brain—from neural spikes to behaviors’* (Vol. 5286, pp. 96–114).
- Grün, S., Abeles, M., & Diesmann, M. (2008b). Impact of higher-order correlations on coincidence distributions of massively parallel data. In *Lecture notes in computer science. ‘Dynamic brain—from neural spikes to behaviors’* (Vol. 5286). New York: Springer.
- Grün, S., Diesmann, M., & Aertsen, A. (2002a). ‘Unitary events’ in multiple single-neuron spiking activity. I. Detection and significance. *Neural Computation*, *14*(1), 43–80.
- Grün, S., Diesmann, M., & Aertsen, A. (2002b). ‘Unitary Events’ in multiple single-neuron spiking activity. II. Non-stationary data. *Neural Computation*, *14*(1), 81–119.
- Grün, S., Diesmann, M., Grammont, F., Riehle, A., & Aertsen, A. (1999). Detecting unitary events without discretization of time. *Journal of Neuroscience Methods*, *94*(1), 67–79.
- Gütig, R., Aertsen, A., & Rotter, S. (2003). Analysis of higher-order neuronal interactions based on conditional inference. *Biological Cybernetics*, *88*(5), 352–359.
- Harris, K. (2005). Neural signatures of cell assembly organization. *Nature Reviews. Neuroscience*, *5*(6), 339–407.

- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Holgate, P. (1964). Estimation for the bivariate poisson distribution. *Biometrika*, *51*, 241–245.
- Johnson, D., & Goodman, I. (2008). Inferring the capacity of the vector Poisson channel with a bernoulli model. *Network: Computation in Neural Systems*, *19*, 13–33.
- Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlations in primary visual cortex of the Macaque. *Journal of Neuroscience*, *25*(14), 3661–3673.
- König, P., Engel, A. K., Roelfsema, P. R., & Singer, W. (1995). How precise is neuronal synchronization. *Neural Computation*, *7*, 469–485.
- Kreiter, A. K., & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of awake macaque monkey. *Journal of Neuroscience*, *16*(7), 2381–2396.
- Kriener, B., Tetzlaff, T., Aertsen, A., Diesmann, M., & Rotter, S. (2008). Correlations and population dynamics in cortical networks. *Neural Computation*, *20*, 2185–2226.
- Krumin, M., & Shoham, S. (2009). Generation of spike trains with controlled auto-and cross-correlation functions. *Neural Computation*, *21*, 1642–1664.
- Kuhn, A., Aertsen, A., & Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, *1*(15), 67–101.
- Kumar, A., Rotter, S., & Aertsen, A. (2008). Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *Journal of Neuroscience*, *28*(20), 5268–5280.
- Laurent, G. (2006). Shall we even understand the fly's brain? In L. van Hemmen, & T. Sejnowski (Eds.), *23 problems in systems neuroscience* (Chap. 8, pp. 21). Oxford: Oxford University Press.
- Lee, A., Manns, I., Sakmann, B., & Brecht, M. (2006). Whole-cell recordings in freely moving rats. *Neuron*, *51*, 399–407.
- Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*, 493–497.
- Lindner, B. (2006). Superposition of many independent spike trains is generally not a Poisson process. *Physical Review E*, *73*, 022901.
- Louis, S., & Grün, S. (2009). Estimating the temporal precision and size of correlated groups of neurons from population activity. In *BMC Neuroscience* (Vol. 10, pp. P253).
- Macke, J., Berens, P., Ecker, A. S., Tolias, A. S., & Bethge, M. (2008). Generating spike trains with specified correlation coefficients. *Neural Computation*, *21*(2), 397–423.
- Maimon, G., & Assad, J. (2009). Beyond poisson: Increased spike-time regularity across primate parietal cortex. *Neuron*, *42*, 426–440.
- Maldonado, P., Babul, C., Singer, W., Rodriguez, E., Berger, D., & Grün, S. (2008). Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *Journal of Neurophysiology*, *100*, 1523–1532.
- Martignon, L., Deco, G., Laskey, K., Diamond, M., Freiwald, W., & Vaadia, E. (2000). Neural coding: Higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Computation*, *12*, 2621–2653.
- 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. *Biological Cybernetics*, *73*, 69–81.
- Mattner, L. (1999). What are cumulants? *Documenta Mathematica*, *4*, 601–622.
- Montani, F., Ince, R. A. A., Senatore, R., Arabzadeh, E., Diamond, M. E., & Panzeri, S. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1901), 3297–3310.
- Nakahara, H., & Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Computation*, *14*, 2269–2316.
- Nawrot, M. P., Boucsein, C., Rodriguez Molina, V., Riehle, A., Aertsen, A., & Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods*, *169*, 374–390.
- Nevet, A., Morris, G., Saban, G., Arkadir, D., & Bergman, H. (2007). Lack of spike-count and spike-time correlations in the substantia nigra reticulata despite overlap of neural responses. *Journal of Neurophysiology*, *98*, 2232–2243.
- Niebur, E. (2007). Generation of synthetic spike trains with defined pairwise correlations. *Neural Comput*, *19*, 1720–1738.
- Okatan, M., Wilson, M. A., & Brown, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation*, *17*(9), 1927–1961.
- Olshausen, B. A., & Field, D. J. (2006). What is the other 85 percent of V1 doing. In L. van Hemmen, & T. Sejnowski (Eds.), *23 problems in systems neuroscience* (Chap. 8, pp. 182–211). Oxford: Oxford University Press.
- Palm, G. (1982). Neural assemblies. An alternative approach to artificial intelligence. *Studies of Brain Function* (Vol. 7). Berlin: Springer.
- Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, *15*, 243–262.
- Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4th ed.). Boston: McGraw-Hill.
- Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*, 1322–1327.
- Pazienti, A., Maldonado, P. E., Diesmann, M., & Grün, S. (2008). Effectiveness of systematic spike dithering depends on the precision of cortical synchronization. *Brain Research*, *1225*, 39–46.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*, 995–999.
- Pipa, G., Wheeler, D., Singer, W., & Nikolić, D. (2008). Neuroxidance: A non-parametric test on excess or deficiency of joint-spike events. *Journal of Computational Neuroscience*, *25*, 64–88.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C* (2 Ed.). Cambridge: Cambridge University Press.
- Reimer, I. C. G., Staude, B., & Rotter, S. (2009). Detecting assembly activity in massively parallel spike trains. In H. Bähr, & I. Zerr (Eds.), *Proceedings of the 8th meeting of the German neuroscience society/30th Göttingen neurobiology conference Neuroforum, Supplement* (Vol. 1).
- 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.
- Roudi, Y., Nirenberg, S., & Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't. *PLoS Computational Biology*, *5*(5), e1000380+.
- Sakurai, Y. (1998). The search for cell assemblies in the working brain. *Behavioural Brain Research*, *91*, 1–13.

- Sakurai, Y., & Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *Journal of Neuroscience*, *6*(40), 10141–10153.
- 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. *Journal of Neurophysiology*, *100*, 2165–2176.
- Shadlen, M. N., & Movshon, A. J. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, *24*, 67–77.
- Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*(4), 569–579.
- Shimazaki, H., Amari, S., Brown, E. N., & Grün, S. (2009). State-space analysis on time-varying correlations in parallel spike sequences. *Proc. IEEE international conference on acoustics, speech, and signal processing (ICASSP)*, 3501–3504.
- Shlens, J., Field, G. D., Gauthier, J. L., Greschner, M., Sher, A., Litke, A. M., et al. (2009). The structure of large-scale synchronized firing in primate retina. *Journal of Neuroscience*, *29*(15), 5022–5031.
- Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., & Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, *26*(32), 8254–8266.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*(1), 49–65.
- Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: Necessity, signature and detectability. *Trends in Cognitive Sciences*, *1*(7), 252–261.
- Singer, W., & Gray, C. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.
- Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, *28*(48), 12591–12603.
- Snyder, D. L., & Miller, M. I. (1991). *Random point processes in time and space*. New York: Springer.
- Stauder, B., Grün, S., & Rotter, S. (2010). Higher order correlations. In S. Grün, & S. Rotter (Eds.), *Analysis of parallel spike trains*. New York: Springer.
- Stauder, B., & Rotter, S. (2009). Higher-order correlations in non-stationary parallel spike trains: Statistical modeling and inference. In *BMC Neuroscience* (Vol. 10, pp. P108).
- Stauder, B., Rotter, S., & Grün, S. (2007). Detecting the existence of higher-order correlations in multiple single-unit spike trains. In *Society for neuroscience abstract viewer/itinerary planner* (Vol. 103.9/AAA18). Washington, DC.
- Stauder, B., Rotter, S., & Grün, S. (2008). Can spike coordination be differentiated from rate covariation? *Neural Computation*, *20*, 1973–1999.
- Stratonovich, R. L. (1967). *Topics in the theory of random noise*. New York: Gordon & Breach Science.
- Streitberg, B. (1990). Lancaster interactions revisited. *The Annals of Statistics*, *18*(4), 1878–1885.
- Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics* (5 Ed.). London: Griffin.
- Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J. L., Patel, H., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro*. *Journal of Neuroscience*, *28*, 505–518.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, *93*, 1074–1089.
- Vaadia, E., Aertsen, A., & Nelken, I. (1995). 'dynamics of neuronal interactions' cannot be explained by 'neuronal transients'. *Proc Biol Sci*, *261*(1362), 407–410.
- van Vreeswijk, C. (2006). What is the neural code? In L. van Hemmen, & T. Sejnowski (Eds.), *23 problems in systems neuroscience* (Chap. 8, pp. 143–159). Oxford: Oxford University Press.
- Wennekers, T., Garagnani, M., & Pulvermüller, F. (2006). Language models based on hebbian cell assemblies. *Journal of Physiology (Paris)*, *100*(1–3), 16–30.
- Wennekers, T., Sommer, F., & Aertsen, A. (2003). Cell assemblies - editorial. *Theory in Biosciences*, *122*, 1–4.