Statistical modelling of higher-order correlations in pools of neural activity

Fernando Montani\textsuperscript{a,b,*}, Elena Phoka\textsuperscript{e,d}, Mariela Portesi\textsuperscript{a,c}, Simon R. Schultz\textsuperscript{e}

\textsuperscript{a} Departamento de Física, Facultad de Ciencias Exactas, UNLP Calle 49 y 115. C.C. 67 (1900), La Plata, Argentina
\textsuperscript{b} IFLYSIB, CONICET & Universidad Nacional de La Plata, Calle 59-789, 1900 La Plata, Argentina
\textsuperscript{c} IFLP, CONICET & Universidad Nacional de La Plata, Calle 49 y 115, 1900 La Plata, Argentina
\textsuperscript{d} Department of Mathematics, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom
\textsuperscript{e} Department of Bioengineering, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom

HIGHLIGHTS

- We consider recordings from multiple neural units for a large number of neurons.
- We propose an approach within the framework of the extended central limit theorem.
- This allows us to characterise the extent of spike correlations higher than two.
- Infer dynamical properties of the network within pools of multiunit neural activity.
- Higher-order correlations needed to realise a widespread activity distribution.

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form 4 February 2013
Available online 22 March 2013

Keywords:
Extended central limit theorem
Large neural ensemble
Multiunit neural activity
Information geometry

ABSTRACT

Simultaneous recordings from multiple neural units allow us to investigate the activity of very large neural ensembles. To understand how large ensembles of neurons process sensory information, it is necessary to develop suitable statistical models to describe the response variability of the recorded spike trains. Using the information geometry framework, it is possible to estimate higher-order correlations by assigning one interaction parameter to each degree of correlation, leading to a \( (2^N - 1) \)-dimensional model for a population with \( N \) neurons. However, this model suffers greatly from a combinatorial explosion, and the number of parameters to be estimated from the available sample size constitutes the main intractability reason of this approach. To quantify the extent of higher than pairwise spike correlations in pools of multiunit activity, we use an information-geometric approach within the framework of the extended central limit theorem considering all possible contributions from higher-order spike correlations. The identification of a deformation parameter allows us to provide a statistical characterisation of the amount of higher-order correlations in the case of a very large neural ensemble, significantly reducing the number of parameters, avoiding the sampling problem, and inferring the underlying dynamical properties of the network within pools of multiunit neural activity.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

To understand how sensory information is processed in the brain, we need to investigate how information is represented collectively by the activity of a population of neurons. There is a large body of evidence suggesting that pairwise correlations
are important for information representation or processing in retina [1,2], thalamus [3], cerebral [4-6] and cerebellar cortices [7,8]. However, there is also evidence that in at least some circumstances, pairwise correlations do not by themselves account for multineuronal firing patterns [9,10]; in such circumstances, triplet and higher-order interactions are important. The role of such higher-order interactions in information processing is still to be determined, although they are often interpreted as a signature of the formation of Hebbian cell assemblies [11].

Higher-order correlations may be important for neural coding even if they arise from random fluctuations. Amari and colleagues [12] have suggested that a widespread distribution of neuronal activity can generate higher-order stochastic interactions. In this case, pairwise or even triplet-wise correlations do not uniquely determine synchronised spiking in a population of neurons, and higher-order interactions across neurons cannot be disregarded. Thus, to gain a better understanding of how neural information is processed, we need to study whether higher-order interactions arise from cell assemblies or stochastic fluctuations. Information-geometric measures can be used to analyse neural firing patterns including correlations of all orders across the population of neurons [13–18,9,10].

A straightforward way to investigate the neural activity of a large population of neurons is to use binary maximum entropy models incorporating pairwise correlations on short timescales [1,2]. To estimate this model, one has to consider a sufficient amount of data to measure the mean activity of individual neurons and correlations in pairs of neurons. This allows us to estimate the functional connectivity in a population of neurons at pairwise level. However, if higher-order correlations are present in the data, and as the number of possible binary patterns grows exponentially with the number of neurons, we would need to use an appropriate mathematical approach to go beyond a pairwise modelling. This is in general a difficult problem, as sampling even third-order interactions can be difficult in a real neurophysiological setting [9,10].

However, under particular constraints, this sampling difficulty can be substantially ameliorated. An example of such a constraint is pooling, in which the identity of which neuron fires a spike in each pool is disregarded. Such a pooling process reflects the behaviour of a simple integrate-and-fire neuron model in reading out the activity of an ensemble of neurons. It also reflects a common measurement made in systems neuroscience: the recording of multiunit neuronal activity, without spike sorting. Whether or not such constraints permit a complete description of neural information processing is still a matter of debate [6,10], but they may allow substantial insight to be gained into the mechanisms of information processing in neural circuits. A method for the statistical quantification of correlations higher than two, in the representation of information by a neuronal pool, would be extremely useful, as it would allow the degree of higher-order correlations to be estimated from recordings of multiunit activity (MUA)—which can be performed in a much broader range of circumstances than “spike-sortable” recordings can.

In this paper, we use a pooling assumption to investigate the limit of a very large neural ensemble, within the framework of information geometry [12–20]. In particular, we take advantage of the recent mathematical developments in q-geometry (and q-information geometry) and the extended central limit theorem [21–26] to provide a statistical quantification of higher-order spike correlations. This approach allows us to identify a deformation parameter to characterise the extent of spike correlations higher than two in the limit of a very large neuronal ensemble. Our method accounts for the different regimes of firing within the probability distribution and provides a phenomenological description of the data, inferring the underlying role of noise correlations within pools of multiunit neuronal activity.

2. Methodology

2.1. Information geometry and the pooled model

We represent neuronal firing in a population of size $N$ by a binary vector $X = (x_1, \ldots, x_N)$, where $x_i = 0$ if neuron $i$ is silent in some time window $\Delta T$ and $x_i = 1$ if it is firing (see Fig. 1(a)). Then, for a given time window, we consider the probability distribution of binary vectors, $(P(X))$. Any such probability distribution $(P(X))$ is made up of $2^N$ probabilities

$$P(X) = \text{Prob}(X_1 = i_1, \ldots, X_N = i_N) = P_{i_1 \ldots i_N}$$

subject to the normalisation

$$\sum_{i_1, \ldots, i_N = 0, 1} P_{i_1 \ldots i_N} = 1.$$  

As proposed by Amari and co-workers (see, for instance, Refs. [20,15]), the set of all the probability distributions $(P(X))$ forms a $(2^N - 1)$-dimensional manifold $S_N$. This approach uses the orthogonality of the natural and expectation parameters in the exponential family of distributions. It is also useful for analysing neural firing in a systematic manner based on information geometry. Any such probability distribution can be unequivocally determined using a coordinate system. One possible coordinate system is given by the set of $2^N - 1$ marginal probability values:

$$\bar{\eta}_{i} = E[x_i] = P(x_i = 1), \quad i = 1, \ldots, N$$

$$\bar{\eta}_{ij} = E[x_ix_j] = P(x_i = x_j = 1), \quad i < j$$

$$\vdots$$

$$\bar{\eta}_{123 \ldots N} = E[x_1 \ldots x_N] = P(x_1 = x_2 = \cdots = x_N = 1).$$

These are called the $\tilde{\eta}$-coordinates [20]. Moreover, provided $P(X) \neq 0$, any such distribution can be expanded as in Ref. [15]

$$P(X) = \exp \left\{ \sum_{i=1}^{N} x_i \tilde{\eta}_i + \sum_{i<j} x_i x_j \tilde{\eta}_i \tilde{\eta}_j + \sum_{i<j<k} x_i x_j x_k \tilde{\eta}_i \tilde{\eta}_j \tilde{\eta}_k + \cdots + x_1 \cdots x_N \tilde{\eta}_1 \cdots N - \psi \right\}, \quad (6)$$

where there are in total $2^N - 1$ different $\tilde{\eta}$ correlation coefficients that can be used to determine univocally the probability distribution. The $\tilde{\eta}$ forms a coordinate system, named $\tilde{\eta} - \text{coordinates}$, which correspond to e-flat structure in the $(2^N - 1)$-dimensional manifold $\mathcal{S}_N$. In Eq. (6), $\psi$ is a normalisation term. The $\tilde{\eta}$-coordinates and $\tilde{\theta}$-coordinates are dually orthogonal coordinates. The properties of the dual orthogonal coordinates allow the formulation of the generalised Pythagoras theorem that gives a decomposition of the Kullback–Leibler divergence to calculate contributions of different orders of interaction between two probability distributions.

These notions are rigorously developed in Refs. [12,14–20] within the framework of information geometry. Assigning one “interaction parameter” $\tilde{\theta}$ to each degree of correlation, we have a $(2^N - 1)$-dimensional model for a population with $N$ neurons. The basis of this formalism is shown in Fig. 1(b): once our coordinate $\tilde{\eta}$ is fixed we have a subset $E(\tilde{\eta})$ for each possible value of $\tilde{\eta}$ and an exponential family of distributions (with the same value of $\tilde{\eta}$). Notice that when higher-order correlations are considered, we need to construct an orthogonal multidimensional coordinate space to build the probability distribution, as is schematised in Fig. 1(b). Let us consider $M(\tilde{\eta}_1, \ldots, \tilde{\eta}_N)$ to be a submanifold of the marginals in $\mathcal{S}_N$. Then, the tangential direction of $M(\tilde{\eta}_1, \ldots, \tilde{\eta}_N)$ represents the direction in which only the pure correlation changes, while the tangential directions of $E(\tilde{\eta})$ span the directions in which only $\tilde{\eta}_1, \ldots, \tilde{\eta}_N$ change but $\tilde{\eta}$ is fixed. Directions of changes in the correlations and marginals are required to be mutually orthogonal.

It is important to point out that the estimation of all the parameters associated with higher-order correlations suffers greatly from a combinatorial explosion. Consider, for instance, a population of 50 neurons for which we would need therefore more than $10^{15}$ parameters. Thus, the number of parameters to be estimated from the available sample size constitutes the reason for the intractability of this approach.

However, if we assume that a neuron cannot process spikes from different neurons separately, then the labels of the neurons that fired each of the spikes are lost. In this case, the target neuron is only aware of the number of synchronous firings among its inputs. Similarly, a neurophysiological recording technique that disregards the identity of the neuron that fires each spike (i.e. it measures MUA) can only count the number of cells in the vicinity firing in a time window, rather than provide information about their pattern.

To investigate the effects of such processes, we consider a pooled model [13,12], where we assume (for mathematical convenience) a population of $N$ identical neurons. Rather than the full distribution of $X$ with probabilities $P(X)$ considered above, we now introduce a set of probabilities $P(k)$ where $k = 0, 1, \ldots, N$ represents the number of synchronous spiking neurons in the population during a time interval $\Delta T$.

This assumption greatly simplifies the analysis based on the coordinate systems described above. The probability distribution is now characterised by only $N$ parameters. Following Ref. [9], we refer to the probability values $P(k)$ as the $p$-coordinates (the pooled model is defined by $N$ parameters). Then, considering the number of neurons that are firing simultaneously (within $\Delta T$) in the pooled model, instead of the probability distribution equation (6), one has

$$P(1) = N \exp (\theta_1 - \psi(\tilde{\theta})), \quad (7)$$
and
\[ P(k) = \binom{N}{k} \exp \left( k \theta_1 + \sum_{i=2}^{k} \binom{k}{i} \theta_i - \psi(\theta) \right), \quad k = 2, \ldots, N \] (8)

while \( P(0) = \exp(-\psi(\theta)) \) corresponds to a completely silent response and is written in such a way so as to fulfill normalisation of the probabilities \( P(k) \). The marginals \( \eta_k \), indicating the probability of any \( k \) neurons firing within \( \Delta T \), are given by
\[ \eta_k = \sum_{i=k}^{N} \binom{N}{i} \frac{P(i)}{\binom{N}{i}}, \quad k = 0, 1, \ldots, N \ [13]. \]
In Eqs. (7) and (8), \( \theta_1 \) corresponds to the first order contribution to the log probabilities, while each \( \theta_i \) \((i = 2, \ldots, N)\) represents the effect on the log probabilities of interactions of order \( i \) in the neuronal pool. We will deal with the relative probability of simultaneous spiking. The ratio between \( P(k) \) and \( P(k-1) \) is given by the following expressions:
\[ \frac{P(1)}{P(0)} = N \exp(\theta_1), \] (9)
\[ \frac{P(2)}{P(1)} = \frac{N - 1}{2} \exp(\theta_1 + \theta_2), \] (10)
and if \( k > 2 \)
\[ \frac{P(k)}{P(k-1)} = \frac{N - k + 1}{k} \exp \left( \theta_1 + \sum_{i=2}^{k-1} \frac{(k-1)!}{(i-1)!(k-i)!} \theta_i + \theta_k \right). \] (11)

This approach has been applied in a previous publication [9] within the maximum entropy (ME) principle to evaluate whether higher-order interactions play a role in shaping the dynamics of neural networks. By fixing the correlation coordinate \( \theta_{k+1} = 0 \) in Amari’s formalism, the ME constraints are set up to order \( k \). This is, the ME principle is achieved by simple fixing \( \{\theta_i\}_{i=k} = 0 \) and the distribution is determined by \( \{\theta_i\}_{i<k} \). The constraints of the ME principle allow us to choose the most random distribution subject to these constraints [15,9]. Thus correlation structures that do not belong to the constrained features are simply removed. Let us consider for instance ME constraints at pairwise level that are set by fixing \( \{\theta_i\}_{i=2} = 0 \), therefore all correlation structures of order higher than two are removed.

In a recent paper, Yu et al. [27] have shown that higher-order correlation structures are quite crucial to get a better understanding of how information might be transmitted in the brain. They have reported the importance of higher-order interactions to characterise the cortical activity and the dynamics of neural avalanches, exhaustively analysing how the Ising like models [28,1,2,29] do not provide in general a fair description of the overall organisation of neural interactions when considering large neuronal systems.

### 2.2. A simple binomial approach

In order to get a better understanding of how to take the limit of a very large number of neurons, we first discuss a quite naive approach using Bernoulli trials and a binomial distribution. Despite its simplicity and limitations, it helps to provide some intuitive feedback.

Let us consider the activity of a neuronal population of \( N \) neurons in a specific time window \( \Delta T \). It should be noted that \( P(k) \) denotes the probability of \( k \) cells firing simultaneously and \( N - k \) being silent, with \( 0 \leq k \leq N \). Let us denote by \( \lambda \) the mean number of spiking cells, then
\[ \lambda = E_p[k] = \sum_{k=0}^{N} k P(k) = N p \]
where we introduced \( p = \frac{\lambda}{N} \).

If we assume that members of a population of neurons spike independently and with a fixed, constant probability, then the spiking activity of the population can be modelled as Bernoulli trials. For a population containing \( N \) neurons, each with a firing probability \( p \), the probability \( P(k) \) of having \( k \) neurons firing simultaneously is given by the binomial distribution:
\[ P(k) = b(k; N, p) = \binom{N}{k} p^k (1 - p)^{N-k}, \] with \( k = 0, 1, \ldots, N \). In the limit of large \( N \) and small \( p \), such that \( \lambda = N p \) is finite, one has the so-called Poisson regime [30]. In this case, the following relations hold for \( k \geq 1 \):
\[ \frac{b(k; N, p)}{b(k-1; N, p)} = \frac{\lambda}{k} + O \left( \frac{1}{N} \right), \] (12)
in the particular case \( k = 1 \) we have
\[ \frac{b(1; N, p)}{b(0; N, p)} = \lambda + O \left( \frac{1}{N} \right), \] (13)
while for \( k = 0 \) the distribution in the Poisson regime reads
\[ b(0; N, p) = \exp(-\lambda) + O \left( \frac{1}{N} \right). \] (14)
We now introduce these ansätze in Eqs. (9)–(11). The ratio between the distribution one would derive in the absence of knowledge of spike correlation, \( P(1) \), and the distribution one would obtain when no spike is being fired, \( P(0) \), is given by \( \frac{P(1)}{P(0)} = Ne^{\theta_1} \). Let us consider the limit \( N \to \infty \) with \( \lambda = Np \) finite (\( p \) is a small quantity). Then, comparing Eq. (13) with Eq. (9), we can write \( \lambda = Ne^{\theta_1} + O(\frac{1}{N}) \), which leads to

\[
\theta_1 \approx \ln \left( \frac{\lambda}{N} \right). \tag{15}
\]

Notice that in this regime, using Eq. (14), the probability distribution \( P(k) \) at \( k = 0 \) behaves as \( P(0) \equiv e^{-\psi} = e^{-\lambda} + O(\frac{1}{N}) \), which gives \( \psi \) approximately equal to \( \lambda \) for the normalising factor.

Introducing Eq. (15) into the probability of \( k \) neurons firing, Eq. (8), as \( \lim_{N \to \infty} \frac{N(N-1)...(N-(k+1))}{(N)^k} = 1 \), in the large \( N \) limit, we get \( P(k, \lambda) \approx \frac{\lambda^k}{k!} e^{-\lambda} e^{\sum_{i=1}^{k} \theta_i} \). Summing up, if a neuronal population behaves in the large \( N \) limit such that

\[
\psi(\theta) \approx \lambda, \tag{16}
\]

and accomplishes Eq. (15) then the probability of \( k \) cells firing simultaneously can be expressed as a \( \tilde{\delta} \)-deformed Poisson distribution such as:

\[
P_{\tilde{\delta}}(k, \lambda) \approx \frac{\lambda^k e^{-\lambda} e^{\tilde{\delta}(k)}}{k!}, \tag{17}
\]

where \( \tilde{\delta}(k) \equiv \sum_{i=2}^{k} \frac{k}{i} \theta_i \) is a deformation term that summarises correlations of all orders.

Eq. (17) corresponds to a Poisson distribution modulated by a multiple-correlation: \( e^{\tilde{\delta}(k)} \). Notice that in the analytical approach presented above we have taken the limit of \( N \to \infty \), but if only correlations up to order two dominated the neuronal spiking, the departure from a Poissonian behaviour in the probability of \( k \geq 2 \) firing neurons would be small. Indeed, applying Amari's prescription [12], we obtain \( \tilde{\delta}(2) = \theta_2 = O(\frac{1}{N}) \) and then \( e^{\tilde{\delta}(k)} \) is almost unity. But our derivation comprises all orders through the factor \( e^{\tilde{\delta}(k)} \).

It is important to note that we consider the mean number of spiking cells to be equal to the normalisation factor (see Eq. (16)). This assumption represents only a possible subset of the probability distributions \( \{P(X)\} \) presented in Eq. (6). Thus, in the limit of a very large number of neurons, and if higher-orders are taken into account, we cannot provide an analytical estimate of the entire set of all possible widespread distributions using the approach of Eq. (17).

More importantly, due to the high dimensionality involved in determining Eq. (6), or even in our approach of Eq. (17), we have to take just a few correlation terms to avoid having an intractable number of parameters [14,9].

But the more basic limitation of the above approach is due to the fact that taking a limit of a very large number of neurons or “thermodynamic limit” essentially means to account for the central limit theorem of statistics (CLT). The CLT articulates conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, can be considered as normally distributed [31]. In the next sections we discuss how to take the limit of a very large number of neurons accounting also for higher-order correlations in the probability distribution. Importantly, we consider the case of any probability distribution of firing, and do not just simply model the neural activity through a binomial distribution (or Bernoulli trials). We make use of an extension of the central limit theorem that also accounts for the case of correlated random variables [24]. This is particularly important since it allows us to extend analytical models accounting for the effect of higher-order correlations on the PDFs, making it possible to compute the emergent properties of the system in the “thermodynamic limit”.

2.3. q-information geometry

Recent studies on information geometry and complex systems have found a big number of distributions that in the asymptotic limit obey the power law rather than the Gibbs distribution [20,32,12,23,33]. These power law distributions seem to rule the asymptotic regime, and the q-exponential distributions used in non-extensive statistical mechanics are very useful for capturing such phenomena [32,24,22,23,25,26,20,21]. The geometrical structure of such probability distributions is termed q-geometry, and its mathematical foundations have been developed by Tsallis, Gell-Mann, Amari, and collaborators. They have in particular proved that the q-geometrical structure is very important to investigate systems with weakly and strongly correlated random variables [32,24,22,23,25,26,20,21].

In particular, recently Amari and Ohara [21] proved that it is possible to generalise the q-structure to any family of probability of distributions, and that the family of q-structure is ubiquitous since the family of all probability distributions can always be endowed with the structure of the q-exponential family for an arbitrary \( q \).

The q-exponential is defined as

\[
\exp_q(x) = \begin{cases} 
1 + (1 - q)(x) & \text{if } (1 + (1 - q)(x)) > 0, \\
0 & \text{otherwise}
\end{cases} \tag{18}
\]
where the limiting case of $q \to 1$ reduces to $\exp_1(x) = \exp(x)$. The $q$-exponential family is defined by generalising Eq. (6) (for a review see Ref. [21]) as

$$P_q(x, \Theta) = \exp_q\{\Theta \cdot x - \psi(\Theta)\},$$

(19)

where in particular the $q$-Gaussian distribution is defined as $p(x, \mu, \sigma) = \exp_q\left(\frac{(x-\mu)^2}{2\sigma^2} - \psi(\mu, \sigma)\right)$ (see Appendix B, for a connection with the correlation coordinates).

In analogy with the exponential families the $q$-geometry has a dually flat geometrical structure (as its coordinates are orthogonal) and accomplishes the $q$-Pythagorean theorem. The maximiser of the $q$-escort distribution is a Bayesian MAP (maximum $a$ posteriori probability) estimator as proved in Ref. [21]. Moreover, it is possible to generalise the $q$-structure to any family of probability distributions, because any parameterised family of probability distributions forms a submanifold embedded in the entire manifold. Altogether, we can introduce the $q$-geometrical structure to any arbitrary family of probability distributions and guarantee that the family of all the probability distributions belongs to the $q$-exponential family of distributions for any $q$ [21]. We refer to Fig. 2 for a schematic explanation of the $q$-exponential family.

Next, we take advantage of recent mathematical progress on $q$-geometry (and $q$-information geometry) to investigate the effect of high-order correlations on the probability distribution in the asymptotic limit.

3. Results

3.1. Beyond pairwise correlations

In statistical mechanics it is said that we reach the "thermodynamic limit" when the number of particles being considered reaches the limit $N \to \infty$. The thermodynamic limit is asymptotically approximated in statistical mechanics using the so-called central limit theorem (CLT) [31]. The CLT ensures that the probability distribution function of any measurable quantity is a normal Gaussian distribution, provided that a sufficiently large number of independent random variables with exactly the same mean and variance are being considered (see pp. 324–330 [31]). Thus, the CLT does not hold if correlations between random variables cannot be neglected.

Thus, the CLT articulates conditions under which a sufficiently large number of independent and identically distributed random variables, each with finite mean and variance, can be considered as normally distributed [31]. In particular, the CLT has been used by Amari and colleagues [12] to estimate the joint probability distribution of firing in a neuronal pool considering the limit of a very large number of neurons. Thus, in their approach pairwise correlations within the joint distribution of firing are quantified through the covariance $\langle U_iU_j \rangle$ of the weighted sum of inputs $U_i$ and $U_j$ of two given pairs of neurons ($i \neq j$, $i = 1 \ldots N$ and $j = 1 \ldots N$) [12]. This is $U_i = \sum_{j=1}^N w_{ij} - H$, where $w_{ij}$ is the connection weight from the $j$th input to the $i$th neuron ($H = E[U_i]$ denotes the mean). Importantly, these $U_i$ are being considered Gaussian due to the CLT [12], and thus the approach quantifies the amount of pairwise correlations through the covariance $\langle U_iU_j \rangle$.

In the presence of weak or strong correlations of any sort, the CLT has been generalised in recent publications by M. Gell-Mann, C. Tsallis, S. Umbarov, C. Vignat, and A. Plastino (see: Refs. [22–26]). They have proved that when a system with weakly or strongly correlated random variables is being considered, if we gather a sufficiently large number of such systems together, the probability distribution will converge to a $q$-Gaussian distribution [22–26]. This is in agreement with the theorems recently proved by Amari and Ohara [21], which permit the introduction of the $q$-geometrical structure to any arbitrary family of probability distributions, and guarantee that the family of all the probability distributions belongs to the $q$-exponential family of distributions.

We will use the "natural extension" of the central limit theorem (ECLT) proposed in Ref. [24], which accounts for cases in which correlations between random variables are non-negligible. This results in so-called $q$-Gaussians (instead of Gaussians) as the PDFs in the ECLT, as proved in Ref. [24]:

$$G_q(x) = \begin{cases} \left[1 + \frac{(1 - q)(-x^2)}{2}\right]^{\frac{1}{1-q}} & \text{if } \left(1 + \frac{(1 - q)(-x^2)}{2}\right) > 0, \\ 0 & \text{otherwise} \end{cases}$$

(20)

where $q$ is a (problem-dependent) positive real index. Notice that in the limit of $q = 1$ a normal Gaussian distribution is recovered as $\lim_{q \to 1} (1 + \frac{x^2}{N})^N = e$, which can be rephrased as $\lim_{q \to 1} (1 + (1 - q)x^2)^{\frac{1}{1-q}} = e$. In other words, the CLT is being recovered as $q \to 1$ [22–26].

Let us now consider the probability of exactly $k = N \cdot r$ (and thus $r = \frac{k}{N}$) neurons firing within a given time window $\Delta T$ across a population of $N$ neurons. In the framework of the pooled model we have: $P_r[r = \frac{k}{N}] = P_r[x_1 = x_2 = \ldots = x_k = 1, x_{k+1} = \ldots = x_N = 0]$, where the neuron $x_i$ is subject to a weighted sum of inputs $u_i$, thus $x_i = 1$ if and only if $u_i > 0$ and $x_i = 0$ if $u_i \leq 0$. Following [12], the neuronal pool receives common inputs $s_1, s_2, \ldots, s_M$ (as schematised in Fig. 3), and $u_i$ is weighted by the common inputs $u_i = \sum_{j=1}^M w_{ij}s_j - h$, where $w_{ij}$ are randomly assigned connection weights. These $u_i$ are $q$-Gaussian due to the ECLT [22–26]. Considering that the $u_i$ are subject to a $q$-Gaussian distribution $N_q(-h, 1)$, we define in analogy to Ref. [12] $u_i = \sqrt{-\alpha v_i} + \sqrt{\alpha} \varepsilon - h$, for $i = 1, \ldots, N$. We take $\alpha = E_q[u_iu_i]$ as a $q$-variance, $h = E_q[u_i]$. ...
as the q-mean, and two independent q-Gaussian random variables \( v_i \) and \( \epsilon \) subject to \( N_q(0, 1) \) (see Ref. [34] for a detailed description of q-Gaussian random variables).

In the following we will use what is commonly referred to as a factorisation approach in statistical mechanics [35,36], which is applicable in this case as we are considering weak correlations among neurons and the population of neurons is homogeneous. We name \( E_q \) as the expectation value taken with respect to a random variable \( \epsilon \), and \( P_i{\{u > 0|\epsilon}\} \) is the conditional probability for \( \epsilon \). This allows us to calculate the probability of having \( r = \frac{1}{N} \) neurons firing, separating the contribution of neurons that are firing \( [P_i{\{u > 0|\epsilon}\}]^k \) from those that are silent \( [P_i{\{u \leq 0|\epsilon}\}]^{N-k} \), as

\[
P_i\{x_1 = x_2 = \cdots = x_k = 1, x_{k+1} = \cdots = x_N = 0\} = E_q[P_i\{u_1; u_2; \ldots; u_k > 0, u_{k+1}; \ldots; u_N \leq 0|\epsilon\}]
\equiv E_q \left[ \left( \frac{N}{k} \right) [P_i(\{u > 0|\epsilon\})]^k[P_i(\{u \leq 0|\epsilon\})]^{N-k} \right].
\] (21)

In order to go beyond the pairwise estimation of [12] (performed within the CLT framework), we need to quantify the amount of correlations higher than two in the probability distributions. If we take the limit of \( N \to \infty \), in the framework of the q-Gaussian ECLT [22–26], instead of the Gaussian CLT as considered in Ref. [12], we can define:

\[
F_q(\epsilon) \equiv P_i\{u > 0|\epsilon\} = P_i \left( u_i > \frac{h - \sqrt{\alpha} \epsilon}{\sqrt{1-\alpha}} \right)
= \frac{1}{\sqrt{2\pi}} \int_{\frac{h - \sqrt{\alpha} \epsilon}{\sqrt{1-\alpha}}}^{\infty} \exp_{q} \left( -\frac{v^2}{2} \right) dv.
\] (22)

If we take \( q = 1 \) in the probability of having \( r = \frac{1}{N} \) neurons firing in Eq. (22), and the distributions within the integral of Eq. (22) corresponds to normal Gaussian distributions, and we are in the “CLT framework”. On the other hand, the “ECLT framework” corresponds to \( q > 1 \) and in this case the system has weakly or strongly correlated random variables. Thus the distributions within the integral of Eq. (22) are considered as q-Gaussian distributions, and correlations are quantified through \( q \).

Notice that if we consider the limit of the CLT framework (\( q = 1 \)), the previous expression reduces to

\[
F_{q=1}(\epsilon) = \frac{1}{2} \text{Erfc} \left( \frac{1}{\sqrt{2}} \frac{h - \sqrt{\alpha} \epsilon}{\sqrt{1-\alpha}} \right)
\]

(23)

where \( \text{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2)dt \) denotes the complementary error function. However, if the effect of correlations higher than two is not negligible then according to the ECLT: \( q > 1 \), thus

\[
\exp_{q} \left( -\frac{v^2}{2} \right) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \int_0^{\infty} dt t^{\frac{1}{q-1}-1} \exp \left( -t - t(q-1)\frac{v^2}{2} \right).
\] (24)
which is known as the Hilhorst transform [37], an integral representation widely used in generalised statistical mechanics [38]. Thus \( F_q(\varepsilon) \) reads

\[
F_q(\varepsilon) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \int_0^\infty \int_{b-\sqrt{\alpha \varepsilon}}^\infty dt \, dv \, t^{\frac{1}{q-1}-1} \exp\left(-t + t(q-1)\frac{v^2}{2}\right) .
\]

(25)

Using several non-trivial identities between Gauss hypergeometric functions and the incomplete beta function, we can exactly calculate the integrals expressed above (see Appendix A, for a detailed description of the math). And \( F_q(\varepsilon) \) in terms of a beta incomplete function reads

\[
F_q(\varepsilon) = \frac{1}{2\sqrt{2\pi}} \frac{B^{\frac{1}{2}}}{B^{\frac{1}{2}}} \left(\frac{1}{q-1} , \frac{1}{2}\right) ,
\]

(26)

where

\[
\xi_0(\varepsilon) = \frac{(q-1)(h-\sqrt{\alpha \varepsilon})^2}{2(1-\alpha)} .
\]

(27)

Eq. (26) allows us to calculate the probability (Eq. (21)) of exactly \( k \) neurons firing within a given time window \( \Delta T \) across a population of \( N \) neurons. Notice that the amount of correlations higher than two is quantified through \( q > 1 \), as when \( q \) is constrained to 1 it leads to \( F_{q=1}(\varepsilon) \), which is reduced to the estimation within the CLT (Eq. (23)) as in Ref. [12].

The joint distribution of firing can therefore be estimated as

\[
Q_q(r) \simeq NP \left[ \frac{r}{N} \right] = NE \left\{ \left(\frac{N}{k}\right) [F_q(\varepsilon)]^k [1 - F_q(\varepsilon)]^{N-k} \right\} .
\]

(28)

The expectation value \( E_\varepsilon \) can be estimated using the saddle point approximation [39,12]

\[
Q_q(r) = \sqrt{\frac{1}{r(1-r)}} \left[ z_q(\varepsilon_0) \right] \sqrt{\frac{1}{2\pi}} \exp\left[ Nz_q(\varepsilon_0) - \frac{\varepsilon_0^2}{2} \right] ,
\]

(29)

where \( z_q(\varepsilon_0) = r \log(F_q(\varepsilon_0)) + (1-r) \log(1-F_q(\varepsilon_0)) \). Within the saddle point approximation: \( \varepsilon_0 = \arg \max \varepsilon \in \mathbb{R} \left[ z_q(\varepsilon) \right] \) and \( \frac{dz_q(\varepsilon)}{d\varepsilon} = 0 \). The solution is \( \varepsilon_0 = F_q^{-1}(r) \), which implies \( r = F_q(\varepsilon_0) \), where \( r \) goes between \([0, 1]\) and \( \varepsilon_0 \) is defined for all real numbers. \( \xi_0(\varepsilon) \) (Eq. (27)) can take different values with \( \alpha \) and \( h \). Additionally, \( \xi_0(\varepsilon) \) depends on the degree of correlation of the network architecture, which is quantified by \( q \). Fig. 4 shows the behaviour of \( Q_q(r) \) for \( q = 1 \). Notice that a higher degree of correlation, \( q > 1 \), corresponds to a more widespread distribution. This is in agreement with the idea of Amari and Nakahara [12] that when taking a very large number of neurons higher-order correlations are needed to reproduce the behaviour of a widespread distribution.
Estimating Eq. (29) for \( q > 1 \) requires a non-trivial approach: finding \( \epsilon_0 = F_q^{-1}(r) \), and from Eq. (26) this means to estimate the inverse incomplete beta function of
\[
\frac{1}{B(p,q)} \left( \frac{1}{q-1}, \frac{1}{2} \right)
\]
(multiplied by the factor \( \frac{1}{2\sqrt{\pi q^2}} \)). The inverse beta functions are tabulated in exhaustive detail, thus we can perform the numerical estimation of Eq. (29). It is important to note that the incomplete beta function subroutines in Matlab are normalised as
\[
I_w(p,s) = \frac{\int_0^1 r^{p-1}(1-r)^{s-1}dr}{B(p,s)}
\]
and therefore we should also multiply by \( B_1(p,s) \) to estimate Eq. (26) numerically. Notice from Eq. (23) that if we take \( q = 1 \), then
\[
F_q^{-1}(r) \text{ is reduced to the estimation of an inverse complementary error function that is also a non-trivial mathematical operation.}
\]
When \( q = 1 \) we are in the CLT framework and just estimating pairwise correlations, and we are potentially missing higher-order correlations. In contrast, \( F_q^{-1}(r) \) allows us to quantify the amount of higher-order correlations as we are within the ECLT framework. It is important to point out that when \( q = 1 \), our previous findings reduce to the CLT limit estimations of Amari [12]. Thus, for real data, one can test for the presence of higher-order correlations by measuring the distribution of activity in multiunit recordings, and fitting \( q \), which represents the amount of higher-order correlations present in the distribution of firing. One can show by simple comparison how statistically different from the \( q = 1 \) case the measured distribution is.

In the next subsection (Section 3.2) we test the applicability of Eq. (29) by measuring the spiking rate of multiunit activity in all non-overlapping windows of length \( \Delta T \). We fit the parameter \( q \) to find the best-fitting function \( Q_q(r) \) in Eq. (29) for the experimental distribution. We then test the hypothesis of absence of higher-order correlation by comparison with a fit with \( q \) constrained to equal 1.

In Section 3.3 we consider a network simulation model in which the number of interconnected neurons is a parameter under control. We then evaluate the hypothesis that Eq. (29) permits characterising the internal dynamics of the network for a spatio-temporal neuronal dataset and quantifying the degree of higher-order correlations through \( q \).

A measure that is also particularly interesting in this context, since it was used in Ref. [40] to give an information theoretic proof of the CLT, is Fisher information
\[
I(Q_q) = \int_0^1 dr \left( \frac{dQ_q(r)}{dr} \right)^2 .
\]
This measure is useful for detecting dynamical changes in the PDFs (i.e. a sharper probability distribution function would tend to have higher Fisher information than a more widespread PDF). Discretising Eq. (30) as in Ref. [41], using a grid of size \( M \) to calculate the distribution, one obtains
\[
I(Q_q) = \frac{1}{4} \sum_{i=1}^{M-1} 2 \left( \frac{Q_{i+1} - Q_i}{Q_{i+1} + Q_i} \right)^2 .
\]
This provides us with an information theoretic measure to quantify the dynamical changes of the distribution. It is usually accepted when considering Fisher information and population codes of independent neurons that they encode information through bell-shaped tuning curves, and that the mean firing rate of a neuron is a Gaussian function of some variable (i.e. the external stimuli). In this case the slope of the distribution increases as the width of the tuning curve becomes smaller, and then a sharper distribution would have a higher slope and thus higher Fisher information. However, this would be the case if noise correlations are independent across neurons and independent of the tuning width, as sharpening in a realistic network does not guarantee a higher amount of Fisher information [42].

### 3.2 Experimental results

To provide an initial test of the applicability of this approach to the characterisation of higher-order correlations in pools of activity obtained from real neurophysiological data, we used a silicon microfabricated linear electrode array (NeuroNexus Inc., Michigan, USA) to record multiunit activity from the mouse barrel cortex (see Section 5). We used 5 min of spontaneous spiking cortical activity of three adult mice, in which 16 electrodes were placed to record the multiunit neural activity (three datasets with 16 different channels each). In the following, we apply our formalism to provide a statistical characterisation of higher-order correlations in pools of multiunit neural activity for each of the 16 different channels of the three datasets (thus in total 48 different recordings of 5 min in length), where we considered four non-overlapping window lengths of \( \Delta T = 25 \text{ ms}, \Delta T = 50 \text{ ms}, \Delta T = 100 \text{ ms}, \) and \( \Delta T = 200 \text{ ms} \). Our selection of these non-overlapping windows is related to the time windows typically used to investigate spike correlations and the firing rate distributions in Ref. [43–45]. The time windows of \( \Delta T = 25 \text{ ms} \) and \( \Delta T = 50 \text{ ms} \) are similar values to those used to investigate spike correlations in Ref. [43,44], and broader time windows of \( \Delta T = 100 \text{ ms} \) and \( \Delta T = 200 \text{ ms} \) have also been used to investigate the firing rate distributions [45], and the functional role of spike correlations [44]. The window length of \( \Delta T = 25 \text{ ms} \) is close to the values used in Ref. [2,1] to investigate the effect of spike correlations.

We estimated the normalised firing distribution \( Q_q(r) \), for experimental data, by measuring the spiking rate in all non-overlapping windows of length \( \Delta T \) (here 25, 50, 100, and 200 ms). We then fitted the parameter \( q \), representing the extent
of higher-order correlations present in the neuronal pool, to find the best-fitting function $Q_q(r)$ (as in Eq. (29)). Values of $q$ equal to unity imply the absence of higher-order correlations in the system. Fig. 5 shows the fitted $q$ extracted for 3 different sets of 16 simultaneously recorded channels of multiunit activity, of increasing depth in the cortex (in 50 µm steps). For each set, the recordings were taken using different time windows $\Delta T = 25$ ms (panels a–c), $\Delta T = 50$ ms (panels d–f), $\Delta T = 100$ ms (panels g–i) and $\Delta T = 200$ ms (j–l). Notice that the maximum and minimum values of $q$ correspond to channel 15 in panel i and channel 12 in panel l, respectively, for time windows of $\Delta T = 100$ ms and $\Delta T = 200$ ms.

In order to understand how well the model proposed in Eq. (29) fits the experimental data, we compared it with a fit with $q$ constrained to equal 1. We define the normalised firing rate as $r \equiv \langle k \rangle / N_{max}$, where $\langle k \rangle$ is the mean firing rate and $N_{max}$ is the maximum number of spikes. Figs. 6 and 7 show the experimental spontaneous distribution of firing $Q_q(r)$ as a function of $r$ (on logarithmic scale) for those channels with the maximum and the minimum values of $q$, respectively. That is, we use the estimated $q$ and the parameters $\alpha$ and $h$ that give the best fit for the distribution. The optimisation fitting criterion is the normalised mean squared error (NMSE), and the default error value is lower than 0.05 ($p$-value < 0.05). It is apparent that for all 25, 50, 100 and 200 ms time windows (Figs. 6 and 7(a), (b), (c) and (d), respectively), the theoretical curve is a remarkably good fit to the experimental distribution ($p$-value < 0.05). In contrast, the $q = 1$ curve does not come close to modelling the data satisfactorily over a wide range of firing rates.

To detect the dynamical changes in the probability distributions, we estimate Fisher information as in Eq. (31). On the one hand Fig. 8 shows Fisher Information versus $q$ for the entire data set. Fig. 9(a) shows Fisher information averaged over the 48 different channels, which becomes higher for smaller time windows. Notice from Fig. 9(b) that when considering the case $q = 1$, Fisher information takes higher values for 50, 100 and 200 ms time windows than for those in which $q > 1$ is considered (Fig. 9(a)). This is not the case for 25 ms time windows in which Fisher information takes a much higher value.
Fig. 6. Fit of $Q_r(r)$ as a function of the normalised firing rate $r$ (semi-log in the X axis). (a) $\Delta T = 25$ ms ($N_{\text{max}} = 130$), $q = 1.55$ and Fisher Info. equal to 0.1026 bits. (b) $\Delta T = 50$ ms ($N_{\text{max}} = 160$), $q = 1.5242$ and Fisher info. equal to 0.0489 bits. (c) $\Delta T = 100$ ms ($N_{\text{max}} = 200$), $q = 1.5483$ and Fisher info. equal to 0.0151 bits. (d) $\Delta T = 200$ ms ($N_{\text{max}} = 240$), $q = 1.3858$ and Fisher info. equal to 0.0204 bits, which corresponds to the minimum observed $q$ (Fig. 5 panel I: channel 12, diamond symbol). The goodness of the fit was evaluated by estimating the normalised mean squared error (NMSE), $p$-value $< 0.05$. The theoretical approach considering $q > 1$ (black dashed line) shows a remarkably good fit, in comparison to the case $q = 1$ (grey dashed line), to the experimental curves (grey circles joined by dotted lines).

for $q > 1$ than for those with $q$ constrained to be equal to 1. As expected, Fisher information is more substantial at shorter time windows, where the fine temporal precision at which the spikes may synchronise has a significant effect.

Overall our findings show that spontaneous distributions of firing with $q > 1$ should be considered to accurately reproduce the experimental dataset. The main advantage of this method is that through the inclusion of a deformation parameter $q$, which accounts for correlations higher than two within the probability distributions, we can quantify the degree of correlation and reproduce the experimental behaviour of the distribution of firing.

If we considered single neuron trial to trial fluctuations for a fixed stimulus in spike count fluctuations, the extent of noise correlations would depend on the width of the tuning curves as correlations come up from common inputs [42]. Thus, in this case Fisher information would depend on the sharpening of the distribution and also on the noise correlation, and therefore a sharper distribution would not necessarily imply a higher amount of Fisher information. This effect is shown, for instance, in Fig. 7(d) ($\Delta T = 200$ ms) and Fig. 7(b) ($\Delta T = 50$ ms); notice that (d) corresponds to a value of $q$ smaller than (b). The firing distribution of Fig. 7(d) has a higher slope than the one presented in Fig. 7(b), but this higher slope does not have a correspondence with a higher value of Fisher information (Fisher info. equal to 0.0150 bits in Fig. 7(d), and Fisher Info. equal to 0.0395 bits in Fig. 7(b)).

Summing up, we presented a formalism that provides us with an estimate of the degree of correlation for the distribution of firing within pools of multiunit neural activity. This method allows us to naturally distinguish how far this distribution of firing is from the one we would obtain if each neuron contributes within CLT ($q = 1$). It permits us, therefore, to quantify the amount of correlation, significantly reducing the number of parameters associated with the correlation coordinates.
Fig. 7. Fit of $Q_r(r)$ as a function of the normalized firing rate $r$ (semi-log in the X axis). (a) $\Delta T = 25$ ms ($N_{\text{max}} = 130$), $q = 1.604$ and Fisher info. equal to 0.0905 bits. (b) $\Delta T = 50$ ms ($N_{\text{max}} = 140$), $q = 1.6$ and Fisher info. equal to 0.0395 bits. (c) $\Delta T = 100$ ms ($N_{\text{max}} = 180$), $q = 1.6858$ and Fisher info. equal to 0.0142 bits, which corresponds to the maximum observed $q$ (Fig. 5 panel i: channel 15, diamond symbol). (d) $\Delta T = 200$ ms ($N_{\text{max}} = 240$), $q = 1.4478$ and Fisher info. equal to 0.0150 bits. Same curve labels as in Fig. 6.

3.3. A network simulation model

To further test our theoretical approach, we apply our formalism to a network simulation model in which the number for interconnected neurons is a parameter under control. We then analyse a simple network model in which neurons receive common overlapping inputs as in Fig. 3, considering that each neuron can be interconnected randomly with more than two neurons. The network simulation we use is the one developed in Ref. [46], which consists of cortical spiking neurons with axonal conduction delays and spike-timing-dependent plasticity (STDP). The magnitude of the synaptic weight between pre- and postsynaptic neurons depends on the timing of the spikes according to the STDP rules. Each neuron is described by a simple spiking model that allows us to reproduce the 20 most fundamental neurocomputational features of biological neurons.

The model exhibits several biologically realistic phenomena such as brain wave like oscillations and patterns of spontaneous cortical activity. Besides, it is quite simple and of easy implementation [46]. In the “spnet” model [46] each variable and parameter has a well defined biophysical meaning. Importantly, the interplay between conduction delays and STDP helps the spiking neurons to spontaneously self-organise into groups with patterns of time-locked activity. Thus, it produces stable firing patterns that would not be possible without these assumptions.

The main advantage of the proposed minimal network with conduction delays and STDP is that it helps to produce and maintain groups of neurons that have sustained time-locked activity. The number of neurons that fire simultaneously then would tend to increase with the number of interconnected neurons when considering time windows equal to the refractory period of a neuron. These time windows, typically used to estimate the joint distribution of firing, correspond to the time immediately following the transmission of a spike in which a neuron is incapable of responding to stimulation, until it regains its ability to transmit another spike (typically considered to be of about 20–25 ms). As our relevant parameter for this simulation is the number of interconnected neurons, $M$, a sustained activity is very important to test the hypothesis...
Fig. 8. Fisher information versus $q$ (theoretical approach considering $q > 1$), taking four different time windows and considering all the 48 available channels. Black up-triangles, $\Delta T = 25$ ms; dark down-triangles, $\Delta T = 50$ ms; dark grey squares, $\Delta T = 100$ ms; light grey circles, $\Delta T = 200$ ms.

Fig. 9. (a) Fisher information averaged over the 48 different channels (theoretical approach considering $q > 1$). Black bar, $\Delta T = 25$ ms; dark grey bar, $\Delta T = 50$ ms; grey bar, $\Delta T = 100$ ms; light grey bar, $\Delta T = 200$ ms. (b) Same as in 8 but considering the fit with $q$ constrained to be equal to 1.

of Amari and collaborators that weak higher-order interactions of almost all orders are required to realise a widespread activity distribution of a large population of neurons [12].

We run just a few minutes of the simulation [46] to capture synchronised firing activity of the neurons within a time window $\Delta T = 25$ ms. This allows us to prevent the algorithm to evolve to states in which the firing becomes more Poissonian and uncorrelated, also avoiding temporary shutdowns in neural activity. We run the simulation during 10 min, which is consistent with the typical recording time of multiunit activity in a real neurophysiological setting. This permits us to obtain sustained time locked activity for the set parameters we use to run the simulation, and to estimate the different joint probability distributions of firing changing the number of interconnected neurons.

Each neuron in the network is described by the simple spiking model [47]

$$v' = 0.04v^2 + 5v + 140 \cdot u + I$$  \hspace{1cm} (32)

$$u' = a(bv - u)$$  \hspace{1cm} (33)

with the auxiliary after-spike resetting

$$\text{if } v \geq +30 \text{ mV, then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases}$$  \hspace{1cm} (34)

where $v$ is the membrane potential of the neuron, and $u$ is a membrane recovery variable, which accounts for the activation of $K^+$ ionic currents and inactivation of $Na^+$ ionic currents, and gives negative feedback to $v$. After the spike reaches its apex at $+30$ mV (not to be confused with the firing threshold) the membrane voltage and the recovery variable are reset according to Eq. (34). The variable $I$ accounts for the inputs to the neurons [47,46].

Since we cannot simulate an infinite-dimensional system on a finite-dimensional lattice, we choose a network with a finite-dimensional approximation taking a time resolution of 1 ms. The network consists of $N = 1000$ neurons with the
Fig. 10. Distribution of firing $Q_q(r)$ versus the normalised firing rate $r$ (semi-log in the X axis), using 10 min of a simulated network that consists of $N = 1000$ neurons with the first $N_e = 650$ of excitatory RS type, and the remaining $N_i = 350$ of inhibitory FS ($\Delta T = 25$ ms). (a) Considering that each excitatory/inhibitory neuron is connected to $M = 2$ or $M = 3$ random neurons. We find no difference in the probability distribution of firing when $M = 2$ or $M = 3$ is considered, $q = 1$. (b) Considering that the number of randomly interconnected neurons is $M = 80$, $q = 1.4252$. The goodness of the fit was evaluated by estimating the normalised mean squared error (NMSE), $p$-value $< 0.05$.

first $N_e = 650$ being of excitatory RS type, and the remaining $N_i = 350$ of inhibitory FS type [47]. Each excitatory neuron is connected to $M = 2, 3, 20, 40, 60, 80, 100, 120, 140$ and $160$ random neurons, so that the probability of connection is $M/N = 0.002, 0.003, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14$ and $0.16$. Each inhibitory neuron is connected to $M = 2, 3, 20, 40, 60, 80, 100, 120, 140$ and $160$ excitatory neurons only. Synaptic connections among neurons have fixed conduction delays, which are random integers between 1 and 10 ms.

The main idea of this simulation is to investigate Amari’s hypothesis that correlations of almost all orders are needed to realise the widespread distribution of firing when a large number of neurons are considered and to show the behaviour of the parameter $q$ when the number of interconnected neurons is changed. In order to show that the probability distribution in the thermodynamic limit is not realised even when pairwise interactions, or third-order interactions, exist [12], we estimate the joint probability distribution of firing when each excitatory/inhibitory neuron is interconnected to $M = 2$ and $M = 3$ neurons. We run 10 min of simulated spiking activity, considering a window length of $\Delta T = 25$ ms, which is a time window close to the one used by [43,44,2,1] to investigate the effect of correlations. As shown in Fig. 10(a) $q = 1$ is a remarkably good fit to the simulated distribution of firing ($p$-value $< 0.05$) when each excitatory/inhibitory neuron is connected to $M = 2$, 3 random neurons. Notice that we find no difference in the probability distribution of firing when $M = 2$ and $M = 3$ are considered. However, when the number of interconnected neurons is higher, the distribution of firing becomes more widespread (see Fig. 10(b)) and $q$ increases as the number of interconnected neurons increases (Fig. 11). In the current simulation we took 1000 neurons; however, when we chose 100–200 neurons, we were also able to capture the effect of the higher-order correlation if the parameter $M$ was large enough to produce a widespread distribution (see Appendix C).

Summarising, by using the current formalism and measuring the joint distribution of firing as a function of the normalised firing rate in multiunit recordings, a neurophysiologist can then quantify the amount of correlation higher than two present in the neuronal data. The idea is that an experimenter, who is used to estimating the mean and the variance of a distribution, now will have to include a new parameter named $q$ that naturally distinguishes how far the distribution is from the one which only accounts for pairwise correlations. Estimating $q$ and the parameters $\alpha$ (q-variance) and $h$ (q-mean) that give the best fit for the joint distribution of firing, $Q_q(r)$ as in Eq. (29), provides the experimenter with a method to test the presence of higher-order correlations. This is a very important tool that can help neurophysiologists to gain new insights into the neuronal interdependence in the cortex and to investigate the underlying neural mechanisms that could generate higher-order correlation structures.

4. Discussion and conclusions

Approaches using binary maximum entropy models at a pairwise level have been developed considering a very large number of neurons on short time scales [1,2,29]. These models can capture essential structures of the neural population activity, however, due to their pairwise nature their generality has been subject to debate [48,49,10]. In particular, using an information geometrical approach, E. Ohiorhenuan and J. D. Victor have shown the importance of the triplets to characterise scale dependence in cortical networks. They introduced a measure called “strain” that quantifies how a pairwise-only model must be “forced” to accommodate the observed triplet firing patterns [10]. Thus, although models accounting for pairwise interactions have proved able to capture some of the most important features of population activity at the level of the retina [1,2], pairwise models are not enough to provide reliable descriptions of neural systems in general [48,49,10,50,51].
Fig. 11. Parameter $q$ estimated through the fit of the firing distribution $Q_q(r)$, versus the number of interconnected neurons $M$. The goodness of the fit was evaluated estimating the normalised mean squared error (NMSE), $p$-value $<0.05$.

Very little is known about how the information saturates as the number of neurons increases. It has been pointed out by Amari and colleagues [12] that as the number of neurons increases, pairwise or triplet-wise correlations are not enough to realise a widespread distribution of firing. Understanding how neural information saturates as the number of neurons increases would require the development of an appropriate mathematical framework to account for correlations higher than two in the thermodynamic limit.

It is important, therefore, to develop an appropriate mathematical approach to investigate systems with a large number of neurons, which could account for correlations of almost all orders within the distribution of firing.

In this paper we present a theoretical approach to quantify the extent of higher than pairwise spike correlation in pools of multiunit activity when taking the limit of a very large number of neurons. In order to do this, we take advantage of recent mathematical progress on $q$-geometry to investigate, in the asymptotic limit, the effect of higher-order correlations on the probability distributions within the ECLT framework [21–26]. The main basis of our formalism is that when taking the limit of a very large number of neurons within the framework of the CLT as in Ref. [12], we are losing information about higher-order correlations. Thus, in the new theoretical approach we take the limit of a very large number of neurons within the framework of the ECLT, instead of the CLT. The inclusion of a deformation parameter $q$ in the ECLT framework allows us to reproduce remarkably well the experimental distribution of firing and to avoid the sampling size problem of Eq. (6) due to the exponentially increasing number of parameters.

We estimated the normalised firing distribution $Q_q(r)$, from multiunit recordings, and fitted the parameter $q$ to find the best-fitting function $Q_q(r)$ (as in Eq. (29)). We showed by simple comparison how statistically different from the $q = 1$ case the measured distribution is (as it is reduced to the CLT pairwise estimations of [12]). Our theoretical predictions provided a remarkably good fit for the experimental distribution. We showed that the parameter $q > 1$ can capture higher-order correlations, which are salient features of the distribution of firing, when applied to our multiunit recording data obtained from the mouse barrel cortex. As higher-order correlations were present in the data, the distributions in the CLT framework do not fit the experimental data well.

Staude and collaborators [52–55] have introduced a quite powerful approach based on a continuous-time point process to investigate higher-order correlations in non-Poissonian spike trains. Higher-order interactions are very important to investigate the neuronal interdependence in the cortex and at population level of the neuronal avalanches [27]. More specifically, Plenz and collaborators [27] have applied a very powerful theoretical framework based on a Gaussian interaction model that takes into account the pairwise correlations and event rates and by choosing an intrinsic thresholding permit distinguishing higher-order interactions. In this approach the pattern probabilities for the so-called “DG model” [56,57] were estimated using the cumulative distribution of multivariate Gaussians and showed a high fitting precision of the experimental data. Our current theoretical formalism relies on a different basis: the recent progress made on the ECTL, and the main goal of our approach, is to provide a quantification of the amount of correlations higher than two when considering a large population of neurons. Thus using mathematical tools of non-extensive statistical mechanics [38,24–26,20,32,12,33,22,23], we introduced an approach that provides a quantification of the degree of higher-order correlation for a very large number of neurons.

It would be interesting to develop in the near future a paper showing a careful comparison of our current method with the one used by Plenz and collaborators. This would help us to investigate the possible link of the quantification through the $q$ parameter with the “DG model” [56,57,27] and to gain more insights for future analysis. Evaluating the degree of $q$ can help us to understand further the processing of information in the cortical network and to get more understanding of the non-linearities of information transmission within a neuronal ensemble. Although formally speaking the multiunit recordings are not in the “thermodynamic limit”, the current methodology presented in this paper is a completely new
5. Materials and methods

Recordings were made from adult female C56BL/6 mice, of 2–3 months of age. The animals were maintained in the Imperial College animal facility and used in accordance with UK Home Office guidelines. The experiments were approved by the UK Home Office, under Project License 70/6516 to S. R. Schultz. Mice were sedated by an initial intraperitoneal injection of urethane (1.1 g/kg, 10% w/v in saline) followed by an intraperitoneal injection of 1.5 ml/kg of Hypnorm/Hypnovel (a mix of Hypnorm in distilled water 1:1 and Hypnovel in distilled water 1:1; the resulting concentration being Hypnorm:Hypnovel:distilled water at 1:1:2 by volume), 20 min later. Atropine (1 ml/kg, 10% in distilled water) was injected subcutaneously. Further supplements of Hypnorm (1 ml/kg, 10% in distilled water) were administered intraperitoneally if required. Their body temperature was maintained at 37±0.5 °C with a heating pad. A tracheotomy was performed and an
endotracheal tube (Hallowell EMC) was inserted to maintain a clear airway as previously described [60]. After the animal was fixed on a stereotaxic frame, a craniotomy was performed, aimed at above barrel C2. A small window on the dura was exposed to allow insertion of the multi-electrode array. The exposed cortical surface was covered with artificial cerebrospinal fluid (in mM: 150 NaCl, 2.5 KCl, 10 HEPES, 2 CaCl₂, 1 MgCl₂; pH 7.3 adjusted with NaOH) to prevent drying. The linear probe (model: A1 × 16-mm-50-413, NeuroNexus Technologies) was lowered into the brain perpendicularly to the cortical surface, until all 16 electrode sites were indicating multiunit neural activity, and allowed to settle for 30 min before recording began.

Acknowledgements

This research was supported by PIP 0255/11 (CONICET), PIP 1177/09 (CONICET) and Pict 2007-806 (ANPCyT), Argentina (FM-MP), a BBSRC DTA studentship (EP) and EPSRC EP/E002331/1, UK (SRS).

Appendix A. Estimation of $F_q(\varepsilon)$ and $Q_q(r)$

We take the limit of $N \rightarrow \infty$, in the “ECLT framework” [22–26], instead of the Gaussian CLT as considered in Ref. [12], we can define:

$$F_q(\varepsilon) \equiv P_r(u > 0|\varepsilon) = P_r\left(u_i > \frac{h - \sqrt{\alpha \varepsilon}}{\sqrt{1 - \alpha}}\right) = \frac{1}{\sqrt{2\pi}} \int_{h - \sqrt{\alpha \varepsilon} \sqrt{1 - \alpha}}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv. \quad (A.1)$$

Notice that if we consider the limit of the CLT ($q = 1$), the previous expression reduces to

$$F_{q=1}(\varepsilon) = \frac{1}{2} \text{Erfc}\left(\frac{h - \sqrt{\alpha \varepsilon}}{\sqrt{1 - \alpha}}\right) \quad (A.2)$$

where \text{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} \exp(-t^2) dt denotes the complementary error function. However, if the effect of higher-order correlations than two are not negligible then according to the ECLT $q > 1$, thus

$$\exp_q\left(-\frac{v^2}{2}\right) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \int_{0}^{\infty} dt \, t^{\frac{1}{q-1}-1} \exp\left(-t - t(q - 1) \frac{v^2}{2}\right) \quad (A.3)$$

which is known as the Hilhorst transform [37], an integral representation widely used in generalised statistical mechanics [38]. Thus $F_q(\varepsilon)$ reads

$$F_q(\varepsilon) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \int_{0}^{\infty} \int_{h - \sqrt{\alpha \varepsilon} \sqrt{1 - \alpha}}^{\infty} dv \, dv^{q-1} \int_{0}^{\infty} dt \, t^{\frac{1}{q-1}-1} \exp\left(-t - t(q - 1) \frac{v^2}{2}\right) \quad (A.4)$$

and substituting $w = \sqrt{\frac{t(q-1)}{2}} v$, we can rewrite

$$\int_{h - \sqrt{\alpha \varepsilon} \sqrt{1 - \alpha}}^{\infty} dv \exp\left(-t - t(q - 1) \frac{v^2}{2}\right) = \frac{\exp(-t) \sqrt{\pi}}{\sqrt{\frac{t(q-1)}{2}}} \left(1 - \text{Erf}\left[\sqrt{\frac{t(q-1)}{2}} \frac{h - \sqrt{\alpha \varepsilon}}{\sqrt{1 - \alpha}}\right]\right) \quad (A.5)$$

where \text{Erf}(x) denotes the error function,

$$\text{Erf}(x) = 1 - \frac{\Gamma\left(\frac{1}{2}, x^2\right)}{\sqrt{\pi}} = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp(-t^2) dt \quad (A.6)$$

and

$$\Gamma\left(\frac{1}{2}, x^2\right) = 2 \int_{x}^{\infty} \exp(-t^2) dt. \quad (A.7)$$

Thus, using

$$\int_{0}^{\infty} x^{\mu-1} \exp(-\beta x) \Gamma(v, ax) dx = \frac{\alpha^v \Gamma(\mu + v)}{\mu(\alpha + \beta)^{\mu+v}} \, \text{E} \left( 1, \mu + v, 1 + \frac{\beta}{\alpha + \beta} \right)$$
in Eq. (A.1) where \( Re(\alpha + \beta) > 0, Re(\mu) > 0, Re(\mu + \nu) > 0 \) and \( _2F_1 \) denotes the Gauss hypergeometric function, we can derive a compact expression for \( F_q \) as

\[
F_q(\varepsilon) = \zeta(\varepsilon)_2F_1 \left( \begin{array}{c}
1, \frac{1}{q-1}; \frac{1}{q-1} + \frac{1}{2}; \frac{1}{1 + \xi_0(\varepsilon)}
\end{array} \right)
\]

where we named

\[
\zeta(\varepsilon) = \frac{\sqrt{\xi_0(\varepsilon)}}{2\sqrt{2\pi} \left( \frac{1}{q-1} - \frac{1}{2} \right) \sqrt{\frac{q-1}{2}(1 + \xi_0(\varepsilon))}^{\frac{q-1}{2}}}
\]

and

\[
\xi_0(\varepsilon) = \frac{(q-1)(h - \sqrt{ae})^2}{2(1-a)}.
\]

Making use of the fact that the hypergeometric functions accomplish the following identity

\[
_2F_1(a, b; c; z) = _2F_1(b, a; c; 1 - z) = (1 - z)^{c-a-b} _2F_1(c - a, c - b; c; z)
\]

(see Ref. [61]), we can name: \( a = 1; b = \frac{1}{q-1}; c = \frac{1}{q-1} + \frac{1}{2}; z = \frac{1}{1 + \xi_0(\varepsilon)} \) and therefore

\[
_2F_1 \left( \begin{array}{c}
1, \frac{1}{q-1}; \frac{1}{q-1} + \frac{1}{2}; \frac{1}{1 + \xi_0(\varepsilon)}
\end{array} \right) = \left( \frac{\xi_0(\varepsilon)}{1 + \xi_0(\varepsilon)} \right)^{-\frac{1}{2}} _2F_1 \left( \begin{array}{c}
\frac{1}{q-1} - \frac{1}{2}; \frac{1}{q-1} + \frac{1}{2}; \frac{1}{1 + \xi_0(\varepsilon)}
\end{array} \right).
\]

Then Eq. (A.8) reads as,

\[
F_q(\varepsilon) = \frac{1}{2\sqrt{2\pi}} \left( \frac{1}{q-1} - \frac{1}{2} \right) \sqrt{\frac{q-1}{2}(1 + \xi_0(\varepsilon))}^{\frac{q-1}{2}} \; _2F_1 \left( \begin{array}{c}
\frac{1}{q-1} - \frac{1}{2}; \frac{1}{q-1} + \frac{1}{2}; \frac{1}{1 + \xi_0(\varepsilon)}
\end{array} \right).
\]

Notice that the incomplete beta function [61] is defined as

\[
B_w(p, q) = \int_0^w t^{p-1}(1-t)^{q-1} dt = \frac{1}{p} \; _2F_1(p, 1 - q; p + 1, w)
\]

and therefore naming \( p = \frac{1}{q-1} - \frac{1}{2}, q = \frac{1}{2} \) and \( w = \frac{1}{1 + \xi_0(\varepsilon)} \) we can rewrite

\[
B \left( \frac{1}{q-1}, \frac{1}{2} \right) = \frac{1}{(\frac{1}{q-1} - \frac{1}{2}) (1 + \xi_0(\varepsilon))^{\frac{q-1}{2}}} \; _2F_1 \left( \begin{array}{c}
\frac{1}{q-1} - \frac{1}{2}; \frac{1}{q-1} + \frac{1}{2}; \frac{1}{1 + \xi_0(\varepsilon)}
\end{array} \right).
\]

This allows us to rewrite \( F_q(\varepsilon) \) in terms of a beta incomplete function with dependence on \( q \), as

\[
F_q(\varepsilon) = \frac{1}{2\sqrt{2\pi}} \sqrt{\frac{q-1}{2}} \; B \left( \frac{1}{q-1}, \frac{1}{2} \right) \left( \frac{1}{q-1} - \frac{1}{2} \right).
\]

The distribution of firing is approached as

\[
Q_q(r) \simeq NP_r \left\{ r = \frac{k}{N} \right\}
\]

\[
= NE_r \left\{ \left( \frac{N}{k} \right) [F_q(\varepsilon)]^k [1 - F_q(\varepsilon)]^{N-k} \right\}.
\]

Using that, \( \binom{N}{k} \simeq \exp(-Nr \log(r) - N(1-r) \log(1-r)) \sqrt{2\pi N r (1-r)} \) in the limit of large \( N \) we can write

\[
Q_q(r) = \sqrt{\frac{N}{(2\pi)^2(1-r)}} \int_0^\infty \text{d} \varepsilon \exp \left\{ N \left[ r \log \left( \frac{F_q(\varepsilon)}{r} \right) + (1-r) \log \left( \frac{1 - F_q(\varepsilon)}{1-r} \right) \right] - \frac{\varepsilon^2}{2} \right\},
\]

But notice that, in the definition of \( Q_q(r) \) the standard exponential is used since the correlation effects were previously included through the deformation parameter \( q \) within \( F_q(\varepsilon) \) (where we have used the ECLT [24]). The previous integral (Eq. (A.18)) is solved using the saddle point approximation [39,12], and then Eq. (29) is obtained. The goodness of the fit was evaluated estimating the normalised mean squared error (NMSE), \( p\text{-value} < 0.05 \). There is then fitted the \( q \) parameter to get the best fitting function \( Q_q(r) \) in Eq. (29) for the experimental distribution. We used the Matlab subroutine GFT2 to compute goodness of fit, for regression model, given matrix/vector of target and output values.
The goodness of the fit was evaluated estimating the normalised mean squared error (NMSE), $p$-value < 0.05.

Appendix B. The Gaussian distribution within the $q$-information geometry framework

The $q$-Gaussian distribution is defined as:

$$p(x, \mu, \sigma) = \exp_q \left( - \frac{(x - \mu)^2}{2\sigma^2} - \psi(\mu, \sigma) \right),$$

which can be rewritten as:

$$p(x, \mu, \sigma) = \exp_q \left( \frac{\mu}{\sigma^2} x - \frac{(x)^2}{2\sigma^2} - \frac{(\mu)^2}{2\sigma^2} - \log_q(\sqrt{2\pi}) \right)$$

where we can identify the correlation coordinates as

$$\theta_1 = \frac{\mu}{\sigma^2},$$

$$\theta_2 = \frac{-1}{2\sigma^2},$$

and the factor

$$\psi = \frac{(\mu)^2}{2\sigma^2} + \log_q(\sqrt{2\pi})$$

Appendix C. Simulation

We applied our formalism to a network simulation model in which the number of neurons is a parameter under control. In the following we consider a network of $N = 200$ neurons with the first $N_e = 80$ of excitatory RS type, and the remaining $N_i = 120$ of inhibitory FS type [47]. Each excitatory neuron is connected to $M = 80$ random neurons, and each inhibitory neuron is connected to $M = 80$ excitatory neurons only. Synaptic connections among neurons have fixed conduction delays of 10 ms. Notice that weak higher-order interactions of almost all orders are required for realising a widespread activity distribution of a large population of neurons [12]. This is in agreement with the hypothesis of Amari [12] that the widespread probability distribution in the thermodynamic limit is not realised even when pairwise interactions, or third-order interactions, exist (see Fig. C.1).

References


F. Montani et al./Physica A 392 (2013) 3066–3086