

## Common inputs in subthreshold membrane potential: The role of quiescent states in neuronal activity

Lisandro Montangie and Fernando Montani\*

*Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB), Universidad Nacional de La Plata,  
CONICET CCT-La Plata, Calle 59-789 (1900) La Plata, Argentina*



(Received 6 February 2018; published 6 June 2018)

Experiments in certain regions of the cerebral cortex suggest that the spiking activity of neuronal populations is regulated by common non-Gaussian inputs across neurons. We model these deviations from random-walk processes with  $q$ -Gaussian distributions into simple threshold neurons, and investigate the scaling properties in large neural populations. We show that deviations from the Gaussian statistics provide a natural framework to regulate population statistics such as sparsity, entropy, and specific heat. This type of description allows us to provide an adequate strategy to explain the information encoding in the case of low neuronal activity and its possible implications on information transmission.

DOI: [10.1103/PhysRevE.97.060302](https://doi.org/10.1103/PhysRevE.97.060302)

As experimental methods in neuroscience become capable of recording simultaneous activity of large populations of neurons, the complexity of the neurophysiological data highlights the necessity to develop a reliable framework accounting for the statistical features of synchronous activity patterns. Particularly, pairwise maximum entropy models (i.e., the Ising model) have been proven useful in characterizing synchronous neural activity [1–3]. In spite of their early success, further studies on the subject have found that interactions among pairs are not sufficient to fully capture statistics in different data [4–7]. Nevertheless, what are the functional implications and the source of these higher-order interactions are still open questions.

For instance, it has been shown that these complex interactions are dynamically modulated by the stimulus [8] and that they can impact coding properties [4]. Higher-order interactions are also necessary for accounting dynamics in ongoing neural avalanches [9] and it has been suggested that simultaneous silence, the coinactivation of neurons, is a ubiquitous feature of beyond-pairwise interactions and concisely summarizes them [10]. The latter finding is of paramount importance since the activity of neural populations is known to be sparse and this sparsity has numerous consequences [8].

A recent model that has shed some light on the matter is the dichotomized Gaussian (DG) model [11]. This model is capable of generating spike trains by thresholding a multivariate Gaussian random variable [12]. Despite being substantially simple, its spiking mechanism is similar to the one assumed in balanced networks, where the mean input to each neuron is typically smaller than the threshold and thus spikes are induced by fluctuations in the input [13]. It showed that higher-order correlations can arise as a consequence of this threshold operation of common input to a population of neurons and that these resulting interactions have a strong impact on the entropy, sparsity, and statistical heat capacity of the ensemble

[14]. Thus, the DG model has been reported to reproduce neural activity patterns better than the pairwise model [9].

In spite of the broad usefulness of this rather simple approach, one particular question has been eluded: what happens when the statistics of the sum of the inputs (or rather the membrane potential) is not Gaussian distributed? In vivo recordings of subthreshold membrane potential fluctuations in rat primary auditory cortex in both anesthetized and awake preparations [15,16] showed dynamics consisting of excursions much larger than the small fluctuations predicted by most random-walk models. Also, the local field potential activity distribution measured in [9] was found to be more heavy-tailed than expected by the Gaussian fit used. Thus, an extension of the DG model was proposed in [17–19] to take into account such deviations, showing how non-Gaussian distributed membrane potentials affected input-output statistics. In this Rapid Communication, we focus on the kurtosis of the input distribution as a measure of non-Gaussianity.

*Modeling non-Gaussian membrane potentials.* A commonly used theoretical approach is to consider that the input to the neuron consists of many small uncorrelated postsynaptic potentials summed together. In this case, according to the central limit theorem, the membrane potential is Gaussian distributed (assuming finite variance) and its fluctuations follow a random walk [11,12,14]. But the central limit theorem does not hold if correlations between far-ranging random variables are not negligible. This would be the case in which the presynaptic population is highly correlated: neurons might be silent most of the time, except during brief moments when large groups of them fire in a synchronized manner. One cannot distinguish between these two cases (or even among intermediate ones) by studying the output population firing pattern alone, since either model could account for essentially any observed spike train. That is assuming either uncorrelated or highly correlated presynaptic variables can lead to the same spiking pattern.

A key feature of this non-Gaussian dynamics is given by the more frequent occurrence of “outliers,” i.e., the all-or-none firing. Hence, a heavy-tailed distribution would be required

\*fmontani@gmail.com

one to capture this phenomena. This singular representation in the data would also require one to carefully account for the sparsity of the neural code. Also, one would like to represent the input of the presynaptic population using as few parameters as possible and at the same time providing flexibility to the representation.

Nonextensive statistical mechanics deals with strongly correlated random variables, sometimes referred to as global correlation [20,21]. The generalization of the classic entropy as the basis for generalizing the Boltzmann-Gibbs theory, reaches its maximum at the distributions usually named as  $q$  Gaussians. This fact, and a number of conjectures and numerical indications [22–24] suggest the existence of a  $q$  analog of the central limit theorem as well.  $q$  Gaussians are the probability density functions given by

$$f(x; q, \xi) = \frac{\sqrt{\xi}}{C_q} \exp_q(-\xi x^2), \quad (1)$$

where  $\exp_q(x) = [1 + (1 - q)x]^{1/(1-q)}$  is defined as the  $q$  exponential [21]. The parameters  $q$  and  $\xi$  are positive real numbers and  $C_q$  is a normalization constant. For  $1 < q < 3$ , these distributions exhibit heavier tails than a Gaussian distribution and in the limit of  $q = 1$  the latter is recovered. Thus, they are natural candidates to study the issue raised here, and the  $q$  analog of the central limit theorem justifies the use of this distribution in the presence of a large number of correlated inputs. The dichotomized  $q$  Gaussian ( $q$ -DG) model is proposed as an extension of the DG model that uses this family of functions [17–19]. Since finite variances are required for this model, we further restrict the values of  $q$  in the range  $[1, 5/3]$  [18,19].

Thus, we model a multivariate  $N$ -dimensional binary random variable  $X \in \{0, 1\}^N$  with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ . A sample from the  $q$ -DG distribution is obtained by first drawing a sample from an  $N$ -dimensional  $q$ -Gaussian random variable  $U$  and then thresholding it into 0 and 1. This thresholding operation will change the moments of the output, so  $X$ , in general, will not have the same mean and covariance as  $U$ . However, we can choose the mean  $h$  and covariance  $\alpha$  of  $U$  such that after truncation,  $X$  has the desired moments  $\boldsymbol{\mu}$  and  $\Sigma$  [12]. Input-output statistics were previously explored in [19], putting special emphasis on how higher-order correlations in the latent process influence population firing patterns.

Given the  $q$ -DG model, how can we best define a “distance” that quantifies the “tailedness” of the distribution of the input to the population of cells? For instance, DeWeese and Zador [15] used excess kurtosis, defined as the ratio of the fourth central moment divided by the fourth power of the standard deviation minus three (which is the kurtosis value expected for a Gaussian distribution); they found large values for their whole-cell patch-clamp recordings of subthreshold membrane potentials in many cases. The standard measure of kurtosis is related to the tails of the distribution, and higher kurtosis is the result of infrequent extreme deviations (or outliers) as opposed to frequent modestly sized deviations. Distributions with kurtosis greater than 3 (or positive excess kurtosis) are said to be leptokurtic. Predominantly positive kurtosis values for traces recorded in awake auditory cortex reflected the unimodal nature of membrane voltage distributions with long

tails. This measure, however, is static and cannot distinguish time durations among excursions. Let us emphasize that we are not focusing on the temporal statistics.

For  $q$ -Gaussian distributions, excess kurtosis is finite and defined as  $k(q) = 6(q - 1)/(7 - 5q)$  for  $1 < q < 7/5$ , which shows that they are leptokurtic in this range. Then, the relationship between the deformation parameter and kurtosis is not proportional: at first,  $k_q$  increases slowly as  $q$  becomes larger and then diverges since it approaches an asymptote. Thus, it establishes two possible regimes, one with low, close to zero kurtosis (“down state”) and another one with large, positive kurtosis (which corresponds to brief bumps or “Up state” [16]). Here, the terminology of “Up” and “Down” refer to cases of second- and higher-order interactions in the inputs that are different from the definition of “active” and “silent” states. The limit where the first ends and the second one begins depends on the criteria used: (i) we can set  $q = 1.25$  as the limit since it corresponds to the maximum tolerable value for excess kurtosis to be considered Gaussian-like ( $k = 2$ ) or (ii) evaluating the first and second derivative one can establish that that point is where kurtosis rapidly changes. In the following, an excess kurtosis value of  $k = 22.8$  ( $q = 1.38$ ), a value close to the average found by [15], is used as reference for the Up state.

*Heavier tails allow for diverse interaction structures.* First, we want to determine how kurtosis has an effect on the higher-order moments of the population activity distribution. If we consider a homogeneous population of  $N$  neurons (for simplicity), we can calculate the probability of exactly  $m$  cells firing within a given time window in this model as

$$P(m) = E_\varepsilon \left[ \binom{N}{m} [P(u > 0|\varepsilon)]^m [P(u \leq 0|\varepsilon)]^{N-m} \right]. \quad (2)$$

For each neuron in the  $q$ -DG model,

$$P(u_i > 0|\varepsilon) = \Phi_q \left( \frac{\sqrt{\alpha}\varepsilon + h}{\sqrt{1 - \alpha}} \right), \quad (3)$$

where  $\Phi_q(\cdot)$  is the cumulative distribution function of a  $q$  Gaussian and  $\varepsilon$  corresponds to a Gaussian noise random variable [19].  $h$  and  $\alpha$  are the input mean and covariance, respectively, equal for each neuron and pair of neurons. By symmetry, all activity patterns with the same number of spikes are equally likely. Such a model is fully characterized by the number of neurons that spike synchronously, i.e., the population spike count  $m = \sum_i X_i$ .

In general we can represent neuronal population activity by a binary vector  $\mathbf{x} = (x_1, \dots, x_N)$  in the space  $X$  of all binary vectors of length  $N$ , where  $x_i = 0$  if neuron  $i$  is silent in some time window and  $x_i = 1$  if it is firing one or more spikes. The probability distribution  $P(\mathbf{x})$  of observing a particular population response can be expressed using different coordinate systems. A useful way of characterizing the population activity distribution is by specifying the  $2^N - 1$  individual probability values; these are called the  $p$  coordinates. The probability can also be determined by the  $2^N - 1$  marginal probability values; these are called the  $\eta$  coordinates [25]. Provided  $P(\mathbf{x}) \neq 0$  for any  $\mathbf{x}$ , any such distribution can be expanded in the so-called

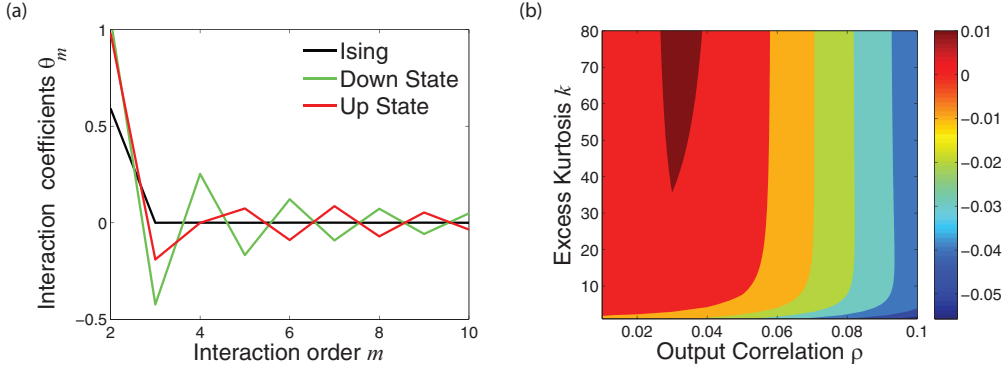


FIG. 1. (a) Interaction coefficients for a homogeneous population with mean firing rate  $\mu = 0.05$  and pairwise correlation coefficient  $\rho = 0.1$ , for three different models. Down and Up states correspond to excess kurtosis of  $k = 0$  and  $k = 22.8$ , respectively. (b) Strain of the population entropy rate for a homogeneous population generated with the  $q$ -DG model with  $\mu = 0.05$ .

log-linear model, or  $\theta$ -coordinates system [25,26]:

$$P(\mathbf{x}) = \exp \left( \sum x_i \theta_i + \sum_{i < j} x_i x_j \theta_{ij} + \sum_{i < j < k} x_i x_j x_k \theta_{ijk} + \dots + \sum_{i < \dots < N} x_i \dots x_N \theta_{i \dots N} - \psi \right), \quad (4)$$

where the  $2^N - 1$  different  $\theta$  coefficients uniquely determine the distribution. The use of this coordinate system to study probabilities and interactions was developed by Amari and co-workers [25,26].

Let us assume that the neural population is a fully homogeneous pool; all the parameters characterizing single neuron properties and interactions between any group of neurons do not depend on the precise identity of the considered neurons, but only on the number of neurons considered. With this assumption the probability distribution is now characterized by only  $N$  parameters. Due to the symmetry of the population all the  $\theta$  coordinates of a given order  $i$  are equal and can be represented by  $\theta_i$  (e.g., all interaction coefficients at order 3,  $\theta_{ijk}$ , are equal to a single parameter that we indicate by  $\theta_3$ ).

Within this information geometry framework [26], neuronal firing in a homogeneous population of size  $N$  can then be represented by the  $N$ th - order maximum entropy model in standard interaction coordinates  $\theta$  [4,10]. Using this representation, the probability of having exactly  $m$  neurons active at a certain time bin is

$$P(m) \sim \binom{N}{m} \exp \left[ \sum_{i=1}^N \binom{m}{i} \theta_i - \psi \right]. \quad (5)$$

We assume all the parameters characterizing single neuron properties and interactions between any group of neurons do not depend on the precise identity of the considered neurons, but rather on the number of neurons. Thus, the probability distribution is now characterized by only  $N$  parameters: all the  $\theta$  coordinates of a given order  $m$  are equal and can be represented by  $\theta_m$ .

Let us compare the equivalent equations (5) and (2) to extract information of how these correlation coefficients depend on input statistics. By solving linear equations, we can obtain the parameters  $\theta_m$  ( $m = 1, \dots, N$ ), and the normalization

factor  $\psi$ . This approach gives a good sense of how structured neural response is.

We also calculate these coefficients for the binary infinite-range Ising model and the DG model with the same mean firing rate  $\mu$  and pairwise correlation  $\rho$  [14]. The Ising model is a trivial case: all coefficients beyond second order are zero. The DG model exhibits higher-order correlations and presents an alternation in the signs of its coefficients with respect to the successive orders of interaction, as previously reported in [10]. Figure 1(a) shows that a sufficient departure from the expected kurtosis of a Gaussian membrane potential breaks the alternation in signs expected for the case  $q = 1$ . By the break of the alternation of signs we mean that in the Down state, not all even interaction coefficients are positive and odd ones are negative. Furthermore, the absolute values of the coefficients decrease for low-order interaction coefficients (i.e.,  $m = 3, 4, 5, 6$ ), suggesting a lesser higher-order correlated structure. Even so, this effect depends on the value of the pairwise correlation  $\rho$  and output firing rate  $\mu$  (which is in turn related to the neuron firing threshold).

Particularly, the third-order maximum entropy component (also called ‘‘strain’’ [6]) has been broadly used as a measure of departure from pairwise maximum entropy models. Triplet firing patterns provide insight into network behavior: if all three-neuron subsets have a positive strain, synchronous activity across the entire population is facilitated, whereas if all three-neuron subsets have a negative strain, global synchronization is suppressed. Using this measure, it was found that local firing patterns are distinctive: while multineuronal firing patterns at larger distances can be predicted by pairwise interactions, patterns within local clusters often show evidence of higher-order correlations [8].

The strain of a distribution over three binary random variables is the third-order interaction term in a maximum entropy model of third degree and can be calculated as

$$\theta_3 = \log \frac{P_{111} P_{100} P_{010} P_{001}}{P_{110} P_{101} P_{011} P_{000}}. \quad (6)$$

Each one of these probabilities can be calculated numerically for the  $q$ -DG, e.g.,  $P_{111} = P(m = 3)$ , since there is no closed-form analytic expression in general.

The triplet interaction coordinate in the Down state was found to be always negative and to decrease monotonically

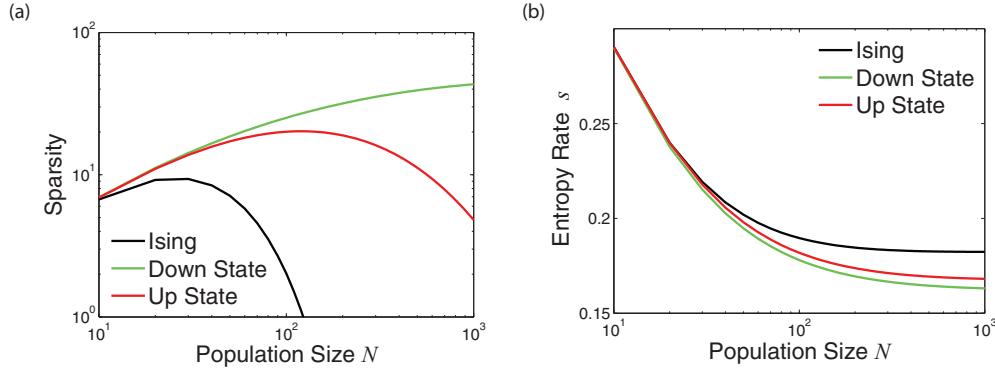


FIG. 2. (a) Sparsity and (b) entropy rate for a homogeneous population with  $\mu = 0.05$  and  $\rho = 0.1$ , considering the three different models. In both figures, Down and Up states correspond to  $k = 0$  and  $k = 22.8$ , respectively.

with increasing correlation coefficients: assuming random-walk dynamics makes it impossible to reach non-negative values of strain [14]. Figure 1(b) shows that common input can lead to zero or even positive strain if the subthreshold membrane potential distribution is sufficiently non-Gaussian. This result implies that higher-order correlations in the common input distribution are needed to have a higher occurrence of spike triplets. This condition also depends on firing rate and correlation coefficient: as spike probability becomes smaller and pairwise correlation decreases, these values become achievable. Nevertheless, this analysis is valid on a three-neuron model and does not necessarily hold for larger ensembles.

*Up states have different global statistics.* The number of parameters in Eq. (5) depends on the size of the population and thus, to study larger populations, further minimal representations of these higher-order features are needed. It was concluded that these beyond-pairwise interaction patterns are concisely summarized by simultaneous silence [10], constraining neural activity. Simultaneous silence refers to neuronal coinactivation: the occurrence of brief silent periods during which all neurons in the local network do not fire. Sparsity is considered to be an important component of population coding that relies on network inhibition [8,10]. Let us evaluate whether heavy-tailedness impacts output activity patterns in the  $q$ -DG (as a hallmark of sparsity in the input signal). Figure 2(a) shows sparsity (estimated as the probability of the population being quiet, i.e.,  $P(m=0)N$ , versus  $N$ .

Notice that in the Down state, Fig. 2(a) depicts that correlations increase sparsity, and in large populations is still dominated by higher-order interactions (in agreement with [14]). For larger values of excess kurtosis (i.e., the Up state), and thus larger departures from the random-walk dynamics, the distribution is less sparse than in the Down state for population sizes bigger than  $10^2$ . Still, activity remains much sparser than the Ising model. This implies that while higher-order interactions are still present, representation capacity in the neural population is more restricted due to the underlying beyond-pairwise correlations in the latent process. Hence, it defines an optimal ensemble size for coding.

Furthermore, we calculate entropy rate directly from the probability distribution as

$$s(N) = \frac{S(N)}{N} = \frac{1}{N} \sum_m P(m) \log \binom{N}{m}, \quad (7)$$

where  $S$  is the entropy of the system and  $N$  is the population size. In Fig. 2(b), it can be seen that a higher kurtosis value in the membrane potential distribution increases entropy rate. This difference does not change further for sizes larger than  $10^2$ , when entropy reaches its asymptotic value. Regardless, this higher entropy value is lower than the upper bound established by the pairwise maximum entropy model (as expected).

Finally, we study the impact of kurtosis on the heat capacity of the population. In terms of the spike train statistics, it is the normalized variance of the surprise of neural spike trains [27]. As a function of population size, we calculate it as

$$c(N) = \frac{1}{N} \sum_m P(m) \left[ \log P(m) - \log \binom{N}{m} - S \right]^2. \quad (8)$$

Introducing an artificial temperature parameter  $\beta = 1/T$ , we can rescale the probability as  $P_\beta(x) = P(x)^\beta / Z_\beta$  [2]. As previously shown in [14], the Down state (described by the DG model) elicits a sharply peaked and diverging specific heat. This is evidence for a physical system being at a critical point and implies that fluctuations around the entropy are nonvanishing in large populations. This was proven valid for common input models, being able to capture this behavior. As shown in Fig. 3(a), the  $q$ -DG model also confirms this for the Up state.

Specific heat scales linearly as size increases, regardless of kurtosis value in the input distribution. This is also the case for the difference in specific heat between the Down and Up states,  $\Delta c$  [see Fig. 3(b)]. Figure 3(c) depicts the power of criticality for the two states. Furthermore, it has been suggested that the rate at which the specific heat diverges provides a means of quantifying the strength of criticality [28,29]. We found that it increases rapidly at first as excess kurtosis becomes larger and then saturates in a maximum value that defines the Up state.

*Discussion.* Common input drives higher-order statistics in population firing patterns [14]. Nevertheless, previous studies have only focused on Gaussian-distributed subthreshold membrane potentials. In this Rapid Communication, we showed that the interaction structure and global statistics of the output of a population of threshold neurons is affected by nonrandom walk dynamics in the sum of its inputs.

Local correlations defined by interaction parameters imply a hierarchical organization [8]. They modestly reduce the amount of information that a cluster conveys while also modi-

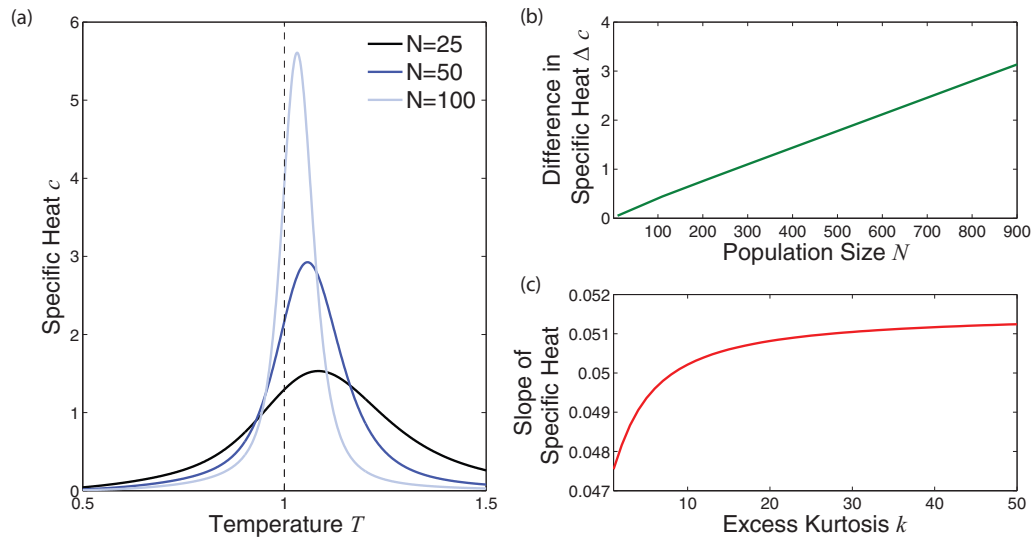


FIG. 3. (a) Specific heat of the Up state versus temperature for  $\mu = 0.05$ ,  $\rho = 0.1$ , and  $k = 22.8$ . (b) Difference in specific heat for the Up and Down states scales linearly with respect to population size. (c) Power of criticality (slope of  $c$  vs  $N$ ) is defined for the two states.

fyng the format of this information, creating sparser codes by increasing the periods of total quiescence and concentrating information into briefer periods of common activity. Hence, kurtosis acts as a modulation of the coupling throughout columns in the cortex. This is proposed as a mechanism to regulate local sensibility to stimuli of certain populations across the transmission of information. Systems near critical states are maximally responsive to certain external signals and this sensitivity may be useful to the auditory cortex in particular. It has been proposed that non-Gaussian membrane potentials in awake auditory cortex may reflect an adaptation to the rapid processing of auditory stimuli. Most of the systems that exhibit criticality in the thermodynamic sense also exhibit a wide range of time scales in their dynamics, so that it may provide a general strategy for neural systems to bridge the gap between the microscopic time scale of spikes and the

macroscopic time scales of behavior. It still seems unclear why subthreshold dynamics in the auditory cortex would differ from dynamics in other sensory cortices, although it is an intriguing possibility that this difference is related to the ability of the auditory cortex to exploit fine timing of brief sensory stimuli [30].

Furthermore, the variation of kurtosis through time would imply a change between two states of the membrane potential, if the output firing rate remains constant. This would in turn produce a bimodal distribution when switching from Down to Up state and vice versa.

*Acknowledgments.* We gratefully acknowledge funding from PIP 11220130100327CO (2014/2016) CONICET, Argentina (F.M.) and Universidad Nacional de La Plata, Argentina (project 11/X812). We also thank L. Guisandez for discussions.

- 
- [1] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, *Nature (London)* **440**, 1007 (2006).
- [2] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek, [arXiv:q-bio/0611072](https://arxiv.org/abs/q-bio/0611072).
- [3] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, *J. Neurosci.* **26**, 8254 (2006).
- [4] F. Montani, R. Ince, R. Senatore, E. Arabzadeh, M. Diamond, and S. Panzeri, *Philos. Trans. R. Soc. A* **367**, 3297 (2009).
- [5] Y. Roudi, S. Nirenberg, and P. Latham, *PLoS Comput. Biol.* **5**, e1000380 (2009).
- [6] I. Ohiorhenuan and J. Victor, *J. Comput. Neurosci.* **30**, 125 (2010).
- [7] E. Ganmor, R. Segev, and E. Schneidman, *Proc. Natl. Acad. Sci. USA* **108**, 9679 (2011).
- [8] I. Ohiorhenuan, F. Mechler, K. Purpura, A. Schmid, Q. Hu, and J. Victor, *Nature (London)* **466**, 617 (2010).
- [9] S. Yu, H. Yang, H. Nakahara, G. S. Santos, D. Nikolic, and D. Plenz, *J. Neurosci.* **31**, 17514 (2011).
- [10] H. Shimazaki, K. Sadeghi, T. Ishikawa, Y. Ikegaya, and T. Toyozumi, *Sci. Rep.* **5**, 9821 (2015).
- [11] S. Amari, H. Nakahara, S. Wu, and Y. Sakai, *Neural Comput.* **15**, 127 (2003).
- [12] J. H. Macke, P. Berens, A. S. Ecker, A. S. Tolias, and M. Bethge, *Neural Comput.* **21**, 397 (2009).
- [13] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker, and K. Josić, *Nat. Neurosci.* **19**, 383 (2016).
- [14] J. H. Macke, M. Opper, and M. Bethge, *Phys. Rev. Lett.* **106**, 208102 (2011).
- [15] M. R. DeWeese and A. M. Zador, *J. Neurosci.* **26**, 12206 (2006).
- [16] T. Hromádka, Anthony M. Zador, and Michael R. DeWeese, *J. Neurophysiol.* **109**, 1989 (2013).
- [17] F. Montani, E. Phoka, M. Portesi, and S. R. Schultz, *Physica A* **392**, 3066 (2013).

- [18] L. Montangie and F. Montani, *Physica A* **421**, 388 (2015).
- [19] L. Montangie and F. Montani, *Physica A* **471**, 845 (2017).
- [20] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).
- [21] C. Tsallis, in *New Trends in Magnetism, Magnetic Materials and Their Applications*, edited by J. L. Morán-Lopez and J. M. Sánchez (Plenum, New York, 1994), p. 451.
- [22] C. Vignat and A. Plastino, *J. Phys. A: Math. Theor.* **40**, F969 (2007).
- [23] S. Umarov, C. Tsallis, and S. Steinberg, *Milan J. Math.* **76**, 307 (2008).
- [24] S. Umarov and C. Tsallis, *J. Phys. A: Math. Theor.* **49**, 415204 (2016).
- [25] S. Amari and H. Nagaoka, *Methods of Information Geometry* (Oxford University Press, Oxford, UK, 2000).
- [26] S. Nakahara and S. Amari, *Neural Comput.* **14**, 2269 (2002).
- [27] T. Mora, S. Deny, and O. Marre, *Phys. Rev. Lett.* **114**, 078105 (2015).
- [28] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, *Proc. Natl. Acad. Sci. USA* **112**, 11508 (2015).
- [29] M. Nonnenmacher, C. Behrens, P. Berens, M. Bethge, and J. H. Macke, *PLoS Comput. Biol.* **13**, e1005718 (2017).
- [30] Y. Yang, M. R. DeWeese, G. H. Otazu, and A. M. Zador, *Nat. Neurosci.* **11**, 1262 (2008).