

RESEARCH ARTICLE

Bistability, non-ergodicity, and inhibition in pairwise maximum-entropy models

Vahid Rostami^{1*}, PierGianLuca Porta Mana¹, Sonja Grün^{1,2}, Moritz Helias^{1,3}

1 Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA BRAIN Institute I, Jülich Research Centre, Jülich, Germany, **2** Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany, **3** Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany

* v.rostami@fz-juelich.de



OPEN ACCESS

Citation: Rostami V, Porta Mana P, Grün S, Helias M (2017) Bistability, non-ergodicity, and inhibition in pairwise maximum-entropy models. *PLoS Comput Biol* 13(10): e1005762. <https://doi.org/10.1371/journal.pcbi.1005762>

Editor: Yasser Roudi, Det Medisinske Fakultet, NTNU, NORWAY

Received: December 6, 2016

Accepted: September 5, 2017

Published: October 2, 2017

Copyright: © 2017 Rostami et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and script processing the data to generate the figures of the manuscript are available from the DRYAD database (<http://datadryad.org/>) under the doi:[10.5061/dryad.n9f77](https://doi.org/10.5061/dryad.n9f77).

Funding: The work was carried out in the framework of the joint International Associated Laboratory (LIA) of INT (CNRS, AMU), Marseilles and INM-6, Jülich. Partially supported by HGF young investigator's group VH-NG-1028, Helmholtz portfolio theme SMHB, DFG Grant GR 1753/4-2 Priority Program (SPP 1665), and EU

Abstract

Pairwise maximum-entropy models have been used in neuroscience to predict the activity of neuronal populations, given only the time-averaged correlations of the neuron activities. This paper provides evidence that the pairwise model, applied to experimental recordings, would produce a bimodal distribution for the population-averaged activity, and for some population sizes the second mode would peak at high activities, that experimentally would be equivalent to 90% of the neuron population active within time-windows of few milliseconds. Several problems are connected with this bimodality: 1. The presence of the high-activity mode is unrealistic in view of observed neuronal activity and on neurobiological grounds. 2. Boltzmann learning becomes non-ergodic, hence the pairwise maximum-entropy distribution cannot be found: in fact, Boltzmann learning would produce an incorrect distribution; similarly, common variants of mean-field approximations also produce an incorrect distribution. 3. The Glauber dynamics associated with the model is unrealistically bistable and cannot be used to generate realistic surrogate data. This bimodality problem is first demonstrated for an experimental dataset from 159 neurons in the motor cortex of macaque monkey. Evidence is then provided that this problem affects typical neural recordings of population sizes of a couple of hundreds or more neurons. The cause of the bimodality problem is identified as the inability of standard maximum-entropy distributions with a uniform reference measure to model neuronal inhibition. To eliminate this problem a modified maximum-entropy model is presented, which reflects a basic effect of inhibition in the form of a simple but non-uniform reference measure. This model does not lead to unrealistic bimodalities, can be found with Boltzmann learning, and has an associated Glauber dynamics which incorporates a minimal asymmetric inhibition.

Author summary

Networks of interacting units are ubiquitous in various fields of biology; e.g. gene regulatory networks, neuronal networks, social structures. If a limited set of observables is accessible, maximum-entropy models provide a way to construct a statistical model for such

Grant 720270(HBP). All network simulations carried out with NEST (<http://www.nest-simulator.org>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

networks, under particular assumptions. The pairwise maximum-entropy model only uses the first two moments among those observables, and can be interpreted as a network with only pairwise interactions. If correlations are on average positive, we here show that the maximum entropy distribution tends to become bimodal. In the application to neuronal activity this is a problem, because the bimodality is an artefact of the statistical model and not observed in real data. This problem could also affect other fields in biology. We here explain under which conditions bimodality arises and present a solution to the problem by introducing a collective negative feedback, corresponding to a modified maximum-entropy model. This result may point to the existence of a homeostatic mechanism active in the system that is not part of our set of observable units.

Introduction

Correlated activity between pairs of cells was observed early on in the history of neuroscience [1, 2]. Immediately the question arose whether there is a functional interpretation of this observation [3], and this question is still with us. Hypotheses range from synchronous activation of neurons to bind representations of features into more complex percepts [4–7], to the involvement of correlations in efficiently gating information [8]. Direct experimental evidence for a functional role of correlated activity is the observation that the synchronous pairwise activation of neurons significantly deviates from the uncorrelated case in tight correspondence with behaviour. Such synchronous events have been observed in motor cortex [9, 10] at time points of expected, task-relevant information. In primary visual cortex they appear in relation to saccades (eye movements) [11, 12]. Another argument for the functional relevance of correlations is the robustness of signals represented by synchronous activity against noise [13]. Non-Gaussian distributions of membrane potentials of neurons indeed point towards the synchronized arrival of synaptic events [14, 15]. An opposite view regards correlated activity merely as an unavoidable epiphenomenon of neurons being connected and influencing one another [16]. In the worst case, both these views are partly true, prompting us to find ways to distinguish functionally relevant correlated events from the uninformative background.

In the context of experimental paradigms that perform repeated trials, the co-variability of neurons across trials has been termed “noise correlation”. Recurrent network models are able to reproduce and explain the weak magnitude and wide spread across pairs of second-order [17–23] and higher-order correlations [24, 25]. These simple dynamical models effectively map the statistics of the connectivity to the statistics of the activity. Even though they explain the uninformative part of correlated activity, it is unclear how to use them to distinguish this background from departures thereof.

The separation of the noise- or background correlation from functionally meaningful correlation is in addition hampered by the diverse dimensions of information processing’s not being completely orthogonal. Indeed, correlation transmission may be modulated by changes of firing rate [9]. Theory [26, 27] confirmed this entanglement in the regime of Gaussian fluctuating membrane potentials.

The dynamical-model approaches just outlined pivot on a more or less realistic physical description of the network, with some stochastic features. A complementary approach is also possible, fully pivoting on statistical models. The latter try to predict and characterize neuronal activity without relying on a definite physical network model. Statistical models have two convenient features. First, intuitive statistical working hypotheses usually translate into a *unique* statistical model [28, 29]; this fact streamlines the construction and selection of a such a model.

For example, the assumption that first- and second-order correlations recorded in an experiment are sufficient to predict the activity recorded in a new experiment, uniquely selects a truncated Gaussian model [29, 30]. Second, a successful statistical model implicitly restricts the set of possible dynamical physical models of the network: only those reflecting the well-modelled statistical properties are acceptable. Statistical models thus help in modelling the actual physical network structure.

A limit case of this kind of statistical models is obtained by choosing probability distributions having maximum entropy under the constraints of experimentally observed quantities [31, 32; in neuroscience see e.g. 33]. The suitability of such maximum-entropy distributions for neuronal activities has been tested in various experimental and simulated set-ups. For example, to explore the sufficiency of pairwise correlations or higher-order moments, or their predictive power for distribution tails [e.g. 34–48], and to characterize dynamical regimes [36, 49–51].

The probability distribution thus obtained, which includes the single-unit and pairwise statistics of the observation by construction, could help us to solve the background-correlation problem described above. In assigning to every observed activity pattern a probability, we obtain a measure of “surprise” for each such pattern; this surprise measure [e.g. 52, 53] is related to the logarithm of the probability and thus to Shannon’s entropy. Periods of activity with low probability correspond to large surprise: these patterns cannot be explained by the statistical properties that entered the construction of the probability distribution. In this way, we are able to effectively differentiate expected, less surprising events from those that are unexpected, surprising, and functionally meaningful.

Computing the maximum-entropy distribution from moment constraints—usually called the *inverse problem*—is simple in principle: it amounts to finding the maximum of a convex function. Hence optimization is straightforward [54, 55]. The maximum can be searched for with a variety of methods (downhill simplex, direction set, conjugate gradient, etc. [56, ch. 10]). The convex function, however, involves a sum over $\exp(N)$ terms, where N is the number of neurons. For 60 neurons, that is roughly twice the universe’s age in seconds, and modern technologies enable us to record *hundreds* of neurons simultaneously [57–60]. Owing to the combinatorial explosion for such large numbers of neurons, the convex function cannot be calculated, not even numerically. It is therefore “sampled”, usually via Markov-chain Monte Carlo techniques [61, 62]. In neuroscience the *Glauber dynamics*, also known as Gibbs sampling [61, 63, chap. 29], is usually chosen as the Markov chain whose stationary probability distribution is the maximum-entropy one. *Boltzmann learning* [64] is the iterative combination of sampling and search for the maximum, and is still considered the most precise method of computing a maximum-entropy distribution. Alternatively one may try to approximate the convex function by an analytic expression, as done with the mean-field [65, 66], Thouless-Anderson-Palmer [66, 67], and Sessak-Monasson [68, 69] approximations. The goodness of these approximations is usually checked against a Boltzmann-learning calculation [cf. 45].

Moment-constrained maximum-entropy models have also been used [70, 71] as generators of surrogate data, again via a Glauber dynamics. Such surrogates are used to implement a null hypothesis to estimate the statistical significance level of correlations between spike trains [70, 72–77].

The pairwise maximum-entropy model is applicable to *experimentally* recorded activities of populations of a couple hundreds neurons at most, so far; but its success, or lack thereof, cannot be automatically extrapolated to larger population sizes. Roudi et al. [78] gave evidence that the maximized Shannon entropy and other comparative entropies of such a model may present qualitatively different features above a particular population size. In the present paper we discuss a feature of the pairwise maximum-entropy model that may be problematic or undesirable: the marginal distribution for the population-averaged activity becomes *bimodal*,

and one of the modes may peak at high activities. In other words, maximum-entropy claims that the population should fluctuate between a regime with a small fraction of simultaneously active neurons, and another regime with a higher fraction of simultaneously active neurons; the fraction of the second regime can be as high as 90%. This feature of the maximum-entropy model has been observed before in several theoretical studies that assumed a homogeneous neuronal population [see e.g. 34, 41, 79, 80].

Our analysis has several points in common with Bohte & al.'s [34]. Bohte et al. wanted to see whether a maximum-entropy distribution can correctly predict the distribution of total activity, given only firing rates and pairwise correlations from a simulated network model as constraints. They found that both the simulation and the maximum-entropy model yield a bimodal distribution of total activity within particular ranges of firing rates and correlations. The fundamental difference from our work is that our experimental data do not show a bimodal distribution, but the maximum entropy model wrongly predicts such bimodality from the measured rates and correlations. More quantitatively, the pairwise correlation found in our data is much lower than that reported in Bohte et al.; in particular, it seems to belong to the range in which their simulation yielded a unimodal distribution [34, p. 169]. Their simulations therefore seem to corroborate that a second mode is biologically implausible in our correlation regime.

Amari & al. [79] notice the appearance of bimodal distributions for the averaged activity and analyse some of their features in the $N \rightarrow \infty$ limit. Their focus is on the correlations needed to obtain a “widespread” distribution in that limit. Our focus is on the bimodality appearing for large but finite N , and we find some mathematical results that might be at variance with Amari & al.'s. They seem to find [79, p. 135] that the Dirac-delta modes are at values 0 and 1; we find that they can appear also strictly within this range. They say [79, p. 138] that the “bigger peak” dominates as $N \rightarrow \infty$; we find that the height ratio between the peaks is finite and depends on the single and pairwise average activity, and for our data is about 2000 as $N \rightarrow \infty$ —an observable value for recording lengths achievable in present-day experiments.

We provide evidence that the bimodality of the pairwise model is bound to appear in applications to populations of more than a hundred neurons. It renders the pairwise maximum-entropy model problematic for several reasons. First, in neurobiological data the coexistence of two regimes appears unrealistic—especially if the second regime corresponds to 90% of all units being simultaneously active within few milliseconds. Second, two complementary problems appear with the Glauber dynamics and the Boltzmann-learning used to find the model's parameters. In the Glauber dynamics the activity alternately hovers about either regime for *sustained* periods, which is again unrealistic and rules out this method to generate meaningful surrogate data. In addition, the Glauber dynamics becomes practically *non-ergodic*, and the pairwise model *cannot be calculated at all* via Boltzmann learning or via the approximations previously mentioned [cf. 62, S 2.1.3; 61, chap. 29]. This case is particularly subtle because it can go undetected: *the non-ergodic Boltzmann learning yields a distribution that is not the maximum-entropy distribution one was looking for*.

Bohte & al. [34] remark that their neuronal-network simulation had to incorporate one inhibitory neuron, with the effect of “curtailing population bursts” [34, p. 175], because “the absence of inhibitory neurons makes a network very quickly prone to saturation” [34, p. 162]. This is something that a standard maximum-entropy distribution cannot do, hence a limitation in its predictive power. It is intuitively clear that lack of inhibition and bimodality are related problems: we show this in section “*Intuitive understanding of the bimodality: Mean-field picture*” using a simple mean-field analysis.

In the present work we propose a modified maximum-entropy model; more precisely, we propose a *reference probability measure* to be used with the method of *maximum relative*

entropy [e.g. 31, 81] (also called minimum discrimination information [82]; see [83] for a comparison of the two entropies). The principle and reference measure can be used with pairwise or higher-order constraints; standard maximum-entropy corresponds to a uniform measure. The proposed reference measure, presented in section “*Inhibited maximum-entropy model*”, solves three problems at once: (1) it leads to distributions without unrealistic modes and eliminates the bistability in the Glauber dynamics; (2) it leads to a maximum-entropy model that can be calculated via Boltzmann learning; (3) it can also “rescue” interesting distributions that otherwise would have to be discarded because incorrect. The reference measure we propose is neurobiologically motivated. It is a minimal representation of the statistical effects of inhibition naturally appearing in brain activity, and directly translates Bohte & al’s device of including one inhibitory neuron in the simulated network. Moreover, the reference measure has a simple analytic expression and the resulting maximum-entropy model *is still the stationary distribution of a particular Glauber dynamics*, so that it can also be used to generate surrogate data.

In the final “*Discussion*” we argue that the use of such a measure is not just an ad hoc solution, but a choice required by the underlying biology of neuronal networks: the necessity of non-uniform reference measures is similarly well-known in other statistical scientific fields, like radioastronomy and quantum mechanics.

The plan of this paper is the following: after some mathematical and methodological preliminaries, we show the appearance of the bimodality problem in the maximum-entropy model applied to an experimental dataset of the activity of 159 neurons recorded from macaque motor cortex. Then we use an analytically tractable homogeneous pairwise maximum-entropy model to give evidence that the bimodality problem will affect larger and larger ranges of datasets as the population size increases. We show that typical experimental datasets of neural activity are prone to this problem.

We then investigate the underlying biological causes of the bimodality problem and propose a way to eliminate it: using a minimal amount of inhibition in the network, represented in a modified Glauber dynamics that includes a minimal *asymmetric inhibition*. We show that this correction corresponds to using the method of maximum entropy with a different reference measure, as discussed above, and that the resulting maximum entropy distribution is the stationary distribution of a modified Glauber dynamics. We finally bring to a close with a summary, a justification and discussion of the maximum-entropy model with the modified reference measure, and a comparison with other statistical models used in the literature.

Results

Preliminaries: Maximum-entropy models and Glauber dynamics

Our study uses three main mathematical objects: the pairwise maximum-entropy distribution, a “reduced” pairwise maximum-entropy distribution, and the Glauber dynamics associated with them. We review them here; some remarks about their range of applicability are given in. Towards the end of the paper we will introduce an additional maximum-entropy distribution.

Pairwise maximum-entropy model. Neuronal activity is modelled as a set of sequences of spikes of N neurons during a finite time interval $[0, T]$. These spike sequences are discretized: we divide the time interval into n bins of identical length Δ equal to T/n , indexed by t in $1, \dots, n$. For each neuron i , the existence of one or more spikes in bin t is represented by $s_i(t) = 1$, and lack of spikes by $s_i(t) = 0$. With this binary representation, the activity of our population at time bin t is described by a vector: $s(t) := (s_i(t))$. We will switch freely between vector and component notation for this and other quantities.

Time averages are denoted by a circumflex: $\widehat{\cdot}$. The time-averaged activity of neuron i is denoted by m_i :

$$m_i := \widehat{s_i(t)} := \frac{1}{T} \sum_{t=1}^n s_i(t), \tag{1}$$

and the time average of the product of the activities of the neuron pair ij , called raw correlation or *coupled activity*, is denoted by g_{ij} :

$$g_{ij} := \widehat{s_i(t)s_j(t)} := \frac{1}{T} \sum_{t=1}^n s_i(t)s_j(t). \tag{2}$$

These time averages will be used as constraints for the maximum-entropy model.

The *pairwise* maximum-entropy statistical model [34, 35, 39, 40] assigns a time-independent probability distribution for the population activity $\mathbf{s}(t)$ of the form (time is therefore omitted in the notation):

$$P_p(\mathbf{s}|\mathbf{h}, \mathbf{J}) = \frac{1}{Z_p(\mathbf{h}, \mathbf{J})} \exp\left(\sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j\right), \tag{3}$$

$$Z_p(\mathbf{h}, \mathbf{J}) := \sum_{\mathbf{s}} \exp\left(\sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j\right);$$

the Lagrange multipliers $\mathbf{h}(\mathbf{m}, \mathbf{g})$ and $\mathbf{J}(\mathbf{m}, \mathbf{g})$ are determined by enforcing the equality of the time averages Eqs (1) and (2) with the single- and coupled-activity expectations, with their definitions

$$E_p(s_i) := \sum_{\mathbf{s}} s_i P_p(\mathbf{s}), \quad E_p(s_i s_j) := \sum_{\mathbf{s}} s_i s_j P_p(\mathbf{s}) \tag{4}$$

that is, enforcing

$$E_p(s_i) = m_i \quad \text{and} \quad E_p(s_i s_j) = g_{ij}. \tag{5}$$

By introducing the covariances \mathbf{c} and Pearson correlation coefficients $\boldsymbol{\rho}$,

$$c_{ij} := E(s_i s_j) - E(s_i)E(s_j),$$

$$\rho_{ij} := \frac{c_{ij}}{\sqrt{[E_p(s_i^2) - E_p(s_i)^2][E_p(s_j^2) - E_p(s_j)^2]}}, \tag{6}$$

the constraints above are jointly equivalent to

$$E_p(s_i) = m_i \quad \text{and} \quad c_{ij} = g_{ij} - m_i m_j \tag{7}$$

or

$$E_p(s_i) = m_i \quad \text{and} \quad \rho_{ij} = \frac{g_{ij} - m_i m_j}{\sqrt{(m_i - m_i^2)(m_j - m_j^2)}}. \tag{8}$$

The maximum-entropy distribution is unique if the constraints are convex. The covariance constraints $c_{ij} = g_{ij} - m_i m_j$ alone are not convex. In this case, uniqueness has to be checked separately [54, 55, 84]. On the other hand, the constraints $E_p(s_i) = m_i$ and $E_p(s_i s_j) = g_{ij}$ are separately convex, thus their conjunction $[E_p(s_i) = m_i] \wedge [E_p(s_i s_j) = g_{ij}]$ is convex too. The bijective

correspondence of the latter with $[E_p(s_i) = m_i] \wedge [c_{ij} = g_{ij} - m_i m_j]$ guarantees that the latter set of constraints is convex as well. What we have said about the covariances c also holds for the correlations ρ .

Reduced maximum-entropy model. If the time-averaged activities \mathbf{m} are homogeneous, i.e. equal to one another and to their population average \bar{m} , and the $N(N-1)/2$ time-averaged coupled activities \mathbf{g} are also homogeneous with population average \bar{g} , $\bar{g} := \frac{2}{N(N-1)} \sum_{i<j} g_{ij}$, then the pairwise maximum-entropy distribution has homogeneous Lagrange multipliers by symmetry: $h_i = h_r$ and $J_{ij} = J_r$. It reduces to the simpler and analytically tractable form

$$\begin{aligned}
 P_r(\mathbf{s} | h_r, J_r) &= \frac{1}{Z_r(h_r, J_r)} \exp \left[h_r N \bar{s} + \frac{1}{2} J_r N \bar{s} (N \bar{s} - 1) \right], \\
 Z_r(h_r, J_r) &:= \sum_{\mathbf{s}} \exp \left[h_r N \bar{s} + \frac{1}{2} J_r N \bar{s} (N \bar{s} - 1) \right],
 \end{aligned}
 \tag{9}$$

which assigns equal probabilities to all those activities \mathbf{s} that have the same *population-averaged activity* \bar{s} , defined as

$$\begin{aligned}
 \bar{s} &:= \frac{1}{N} \sum_{i=1}^N s_i, \\
 N \bar{s} &\in \{0, 1, 2, \dots, N\}.
 \end{aligned}$$

We denote *population averages* that are normalized to the number of neurons by an overbar: $\bar{\cdot}$. Then the quantity $N \bar{s}$ represents the sum of the activities of all neurons in the population at time bin t ; we call it *total activity*, and its plot is called “population time-histogram” in some works [cf 85, 86].

In this homogeneous case, the values of the multipliers appearing in Eq (9) are equal to their population averages: $h_i = h_r = \frac{1}{N} \sum_i h_i$ and $J_{ij} = J_r = \frac{2}{N(N-1)} \sum_{i<j} J_{ij}$. This distribution hence contains only information about how many neurons $N \bar{s}$ are active at any given point in time, but not the particular composition \mathbf{s} of active neurons. This simpler distribution can therefore also be interpreted as an approximation of the pairwise maximum-entropy one, achieved by disregarding population inhomogeneities of the constraints m_i and g_{ij} . But it is also an exact maximum-entropy distribution in its own right, obtained by only constraining the expectations for the population averages of the single and coupled activities,

$$\sum_i s_i = N \bar{s}, \quad \sum_{i<j} s_i s_j = N \bar{s} (N \bar{s} - 1) / 2,$$

to be equal to their measured time averages

$$E_r(N \bar{s}) = N \bar{m} \quad \text{and} \quad E_r(N \bar{s} (N \bar{s} - 1)) = \frac{N(N-1)}{2} \bar{g} := \sum_{i<j} g_{ij}.
 \tag{10}$$

For this reason we call the model Eq (9) a *reduced* pairwise maximum-entropy model. But in the inhomogeneous case the multipliers of the reduced model are *not* equal to the averages of the pairwise one: $h_i \neq \frac{1}{N} \sum_i h_i$, $J_r \neq \frac{2}{N(N-1)} \sum_{i<j} J_{ij}$.

It is straightforward to derive the probability distribution for the population average \bar{s} in this model, owing to its symmetry: the total number of active neurons in the population is $N \bar{s}$, and there are $\binom{N}{N \bar{s}}$ equally probable ways in which this is possible, each with probability by

Eq (9). Therefore,

$$\begin{aligned}
 P_r(\bar{s}|h_r, J_r) &= \frac{1}{Z_r(h_r, J_r)} \binom{N}{N\bar{s}} \exp \left[h_r N\bar{s} + \frac{1}{2} J_r N\bar{s} (N\bar{s} - 1) \right], \\
 Z_r(h_r, J_r) &:= \sum_{\bar{s}} \binom{N}{N\bar{s}} \exp \left[h_r N\bar{s} + \frac{1}{2} J_r N\bar{s} (N\bar{s} - 1) \right].
 \end{aligned}
 \tag{11}$$

This probability distribution $P_r(\bar{s})$ can, in turn, also be obtained applying a maximum-relative-entropy principle [31, 83], i.e. minimizing the relative entropy (or discrimination information)

$$H(P, P_0) := \sum_{\bar{s}} P(\bar{s}) \ln \frac{P(\bar{s})}{P_0(\bar{s})}
 \tag{12}$$

of $P(\bar{s})$ with respect to the reference distribution $P_0(\bar{s}) = 2^{-N} \binom{N}{N\bar{s}}$ while constraining its first two moments, or equivalently its first two factorial moments [87] ($E(N\bar{s}), E(N\bar{s}(N\bar{s} - 1)/2$)).

It is easy to see that in this model, by symmetry, we also have

$$E_r(s_i) = E_r(\bar{s}), \quad E_r(s_i s_j) = E_r \left(\frac{N\bar{s}(N\bar{s} - 1)}{N(N - 1)} \right),
 \tag{13}$$

$$c_{ij} = \bar{c} = E_r \left(\frac{N\bar{s}(N\bar{s} - 1)}{N(N - 1)} \right) - E_r(\bar{s})^2, \quad \rho_{ij} = \bar{\rho} = \frac{\bar{c}}{E_r(\bar{s}) - E_r(\bar{s})^2},
 \tag{14}$$

and $(E_r(\bar{s}), E_r(\frac{N\bar{s}(N\bar{s}-1)}{N(N-1)}))$, $(E_r(\bar{s}), \bar{c})$, $(E_r(\bar{s}), \bar{\rho})$ are equivalent sets of constraints (\bar{c} and $\bar{\rho}$ by themselves are not convex).

This reduced maximum-entropy model is mathematically very convenient because the Lagrange multipliers h_r, J_r can be easily found numerically (with standard convex-optimization methods like downhill simplex, direction set, conjugate gradient, etc. [56, ch. 10]) with high precision even for large (e.g. thousands) population sizes N .

The pairwise and reduced models are very similar to the *Gaussian ensemble* of statistical mechanics [88–90, and refs therein], in which the mean and variance of a system’s energy are constrained; it is intermediate in properties between the canonical and microcanonical ensembles.

The maximum-entropy models reviewed above use *time-averaged* data. Their probabilities are therefore time-invariant; they are stationary statistical models.

Glauber dynamics and Boltzmann learning. The normalization $Z_p(\mathbf{h}, \mathbf{J})$ appearing in the probability distribution Eq (3) requires the summation over 2^N states, typically in the range of $N \approx 100$. This calculation may require prohibitive amounts of time; we need a way to calculate the distribution that avoids the computation of the normalization. The probability $P_p(\mathbf{s}|\mathbf{h}, \mathbf{J})$ Eq (3) is identical in form to the stationary distribution of a Markov chain $\mathbf{s}(t) \mapsto \mathbf{s}(t + 1)$, the so-called asynchronous Glauber dynamics [63]. If this dynamics is ergodic, after an initial transient it generates states with a relative frequency distribution approximately equal to the stationary one. In this context the symmetric Lagrange multipliers \mathbf{J} are sometimes also referred to as “couplings”, in analogy to synaptic interactions between the units. The Lagrange multipliers \mathbf{h} are referred to as “biases” and may be interpreted as either a threshold or external input controlling the base activity of individual neurons. The temporal sequence of states produced by the Glauber dynamics is predominantly controlled by the time constant of the update rule and in general does not reflect the temporal evolution of the neuronal population activity.

If we assume that our uncertainty about the *evolution* of the population activity can be modelled by the Glauber dynamics of a binary network, we can employ the Glauber dynamics, with \mathbf{h}, \mathbf{J} parameters determined by the constraints Eq (5), to generate surrogate data that would allow us to implement a null-hypothesis for a statistical test, that includes the average activities and pairwise correlations as observed in the data. This procedure requires that the dynamics should be ergodic, sampling the entire state space. Otherwise, time averages obtained from the surrogate would not coincide with those of our experimentally observed data.

The Glauber dynamics is used as the sampling step in Boltzmann learning [64], as mentioned in the “Introduction”, to find the parameters $\mathbf{h}(\mathbf{m}, \mathbf{g})$ and $\mathbf{J}(\mathbf{m}, \mathbf{g})$ of the maximum entropy distribution Eq (3) having mean activities \mathbf{m} and raw pairwise correlations \mathbf{g} . Starting with some values of the multipliers $\hat{\mathbf{h}}$ and $\hat{\mathbf{J}}$, the distribution is sampled by the Glauber dynamics and the its averages of the single activity $\hat{\mathbf{m}}$ and the coupled activity $\hat{\mathbf{g}}$ are found for the current values of the multipliers. The latter are then adjusted in relation to the mismatch $|\hat{\mathbf{m}} - \mathbf{m}|$, $|\hat{\mathbf{g}} - \mathbf{g}|$ between the sampled averages and the required values from the experimental data. A new sampling is then performed, and so on until the mismatch lies below a prescribed accuracy.

The problem: Bimodality, bistability, non-ergodicity

We first show how the bimodality problem subtly appears with a set of experimental data, then explore its significance for larger population sizes and other samples of experimental data of brain activity.

Experimental data: Preliminary approximations. The data, provided by A. Riehle and T. Brochier (INT, CNRS-AMU Marseille, France), consist of the activity of a population of $N = 159$ single neurons recorded from motor cortex of macaque monkey for 15 minutes, using a 100-electrode “Utah” array as described in [60], but with a different behavioural design: here the monkey was awake and alert, but did not perform any task during the recording. This behavioural protocol is chosen for retrieving “resting” (or “ongoing”) state [91] data to characterize the “ground” state, in contrast to a task or functional state.

Fig 1A shows a raster plot (2s out of 15min for better visibility) of the activities $\mathbf{s}(t)$ of all recorded neurons. The population-averaged activity $\bar{s}(t)$ for this period is shown underneath. The distributions of the time-averaged single and coupled activities m_i, g_{ij} , and the corresponding empirical covariances c_{ij} measured in the full data set of 15 min are shown in panels B, C, D. The population averages of these time-averaged quantities are

$$\bar{m} = 0.0499, \quad \bar{g} = 0.00261, \quad \bar{c} = 0.000135, \quad \bar{\rho} = 0.00319. \quad (15)$$

As discussed in the previous section, the pairwise maximum-entropy model is a stationary statistical model. If we intended to analyse the dataset above with this model for a specific purpose—for example, characterizing a “ground state” of behavioural activity—then we would have to assess whether a stationary model would really suit these particular data and purpose. It would not be suitable, for example, to model transient aspects of neural activity. Our goal, however, is rather to analyse the general presence of bimodality in the model for data with ranges and orders of magnitude typical of recorded brain activity. In section “Bimodality of pairwise models for massively parallel data” we will see that our conclusions regarding bimodality are valid even if the population average \bar{m} is doubled or halved and if the Pearson correlation $\bar{\rho}$ becomes ten times smaller or larger; thus amply valid within any non-stationarity corrections [93]. Moreover, stationary maximum-entropy models have also been used with highly fluctuating data, e.g. from retinal cells [e.g. 35, 40, 43, 94], with the purpose of analysing

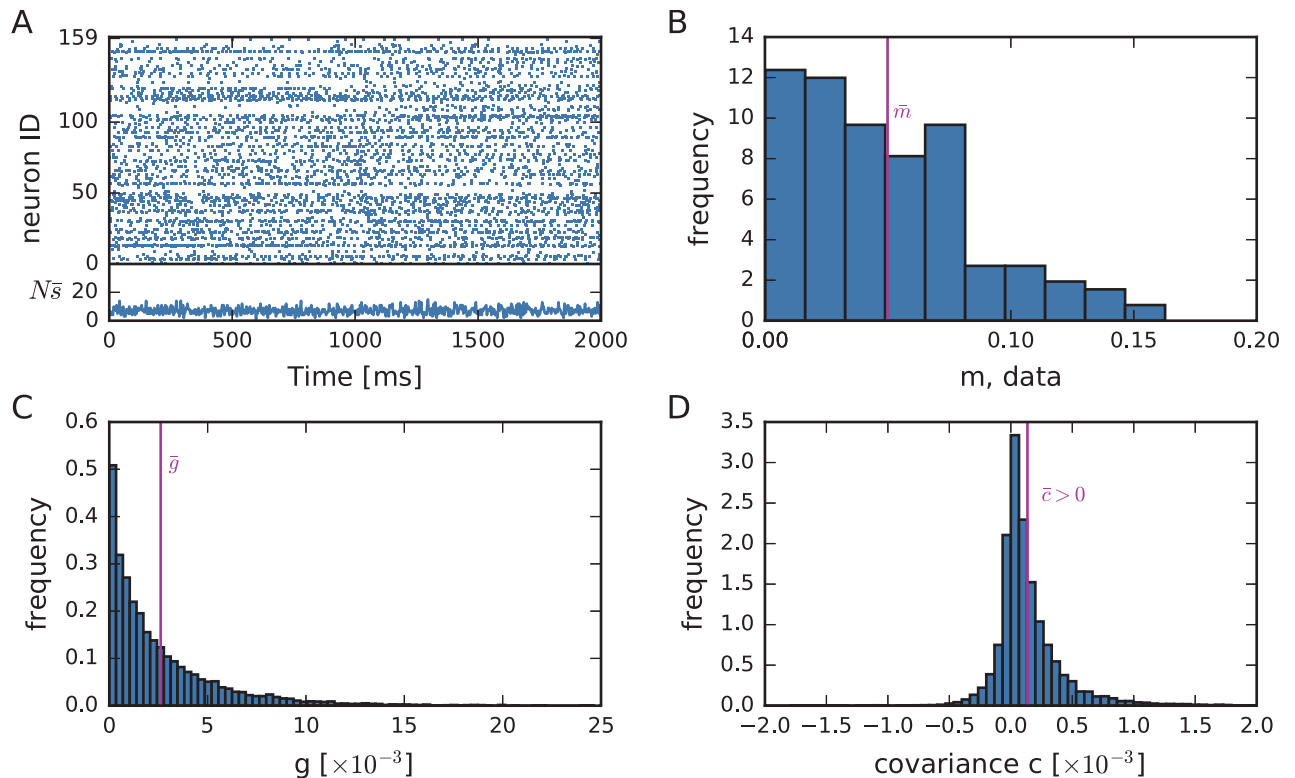


Fig 1. Experimental data and their empirical first- and second-order statistics. (A) Example raster display (snippet of 2s from the total data of 15 min) of $N = 159$ parallel spike recordings of macaque monkey during a state of “ongoing activity”. The experimental data are recorded with a 100-electrode “Utah” array (Blackrock Microsystems, Salt Lake City, UT, USA) with $400 \mu\text{m}$ interelectrode distance, covering an area of $4 \times 4 \text{ mm}^2$ (session: s131214-002). The total activity $N\bar{s}(t)$ shows the number of active neurons within each time bin t of width $\Delta = 3 \text{ ms}$. (B) Population distribution of the time-averaged activities m_i (in spikes/ Δ) of each of the neurons i , Eq (1). The vertical line marks the population average $\bar{m} = 0.0499$. (C) Population distribution of the time-averaged raw correlations (coupled activities) g_{ij} , Eq (2). The vertical line marks the population average, $\bar{g} := \frac{2}{N(N-1)} \sum_{i < j} g_{ij} = 0.00261$. (D) Population distribution of the covariances $c_{ij} = g_{ij} - m_i m_j$. The vertical line marks the slightly positive population average, $\bar{c} := \frac{2}{N(N-1)} \sum_{i < j} c_{ij} = 0.000135$. Histograms bins in B, C, D are computed with Knuth’s rule [92] and calculated over the full 15-minute long recording. Data courtesy by A. Riehle and T. Brochier.

<https://doi.org/10.1371/journal.pcbi.1005762.g001>

some of their information-theoretic properties rather than of modelling the data. For these reasons, and also for brevity, we do not address stationarity analyses and corrections in the present work.

We now need to find the parameters $\mathbf{h}(\mathbf{m}, \mathbf{g})$ and $\mathbf{J}(\mathbf{m}, \mathbf{g})$ of the maximum entropy distribution Eq (3) so that the mean activities \mathbf{m} and the raw pairwise correlations \mathbf{g} correspond to those measured in the data shown in Fig 1B and 1C. We try to find these parameters via Boltzmann learning with a Glauber dynamics, as explained in “Glauber dynamics and Boltzmann learning”.

We choose the sampling phase of the Boltzmann learning to have 10^6 timesteps; an example is shown in Fig 2A. This number of timesteps is large compared to Roudi et al. [45] (200 timesteps) or Broderick et al. [95] (400 timesteps). The preliminary approximations (\hat{h}_i, \hat{J}_{ij}) of the Lagrange multipliers obtained in this way are shown in Fig 2D. The final single and coupled activities are shown in Fig 2C, compared to the experimental data. The first and second moments are highly correlated with the experimental ones and seem to describe the data well. The preliminary approximation of the population-average probability distribution is shown in red in Fig 2B. Its tail disagrees with that of the empirical frequency distribution (dashed); but,

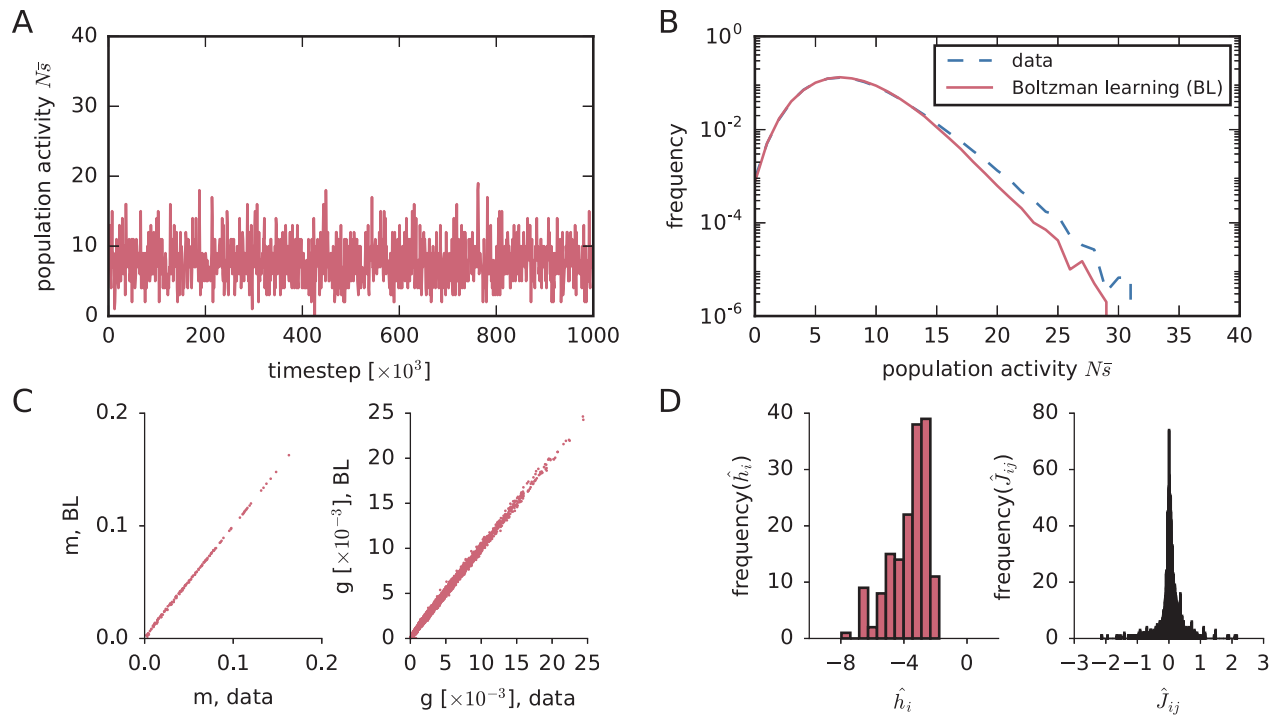


Fig 2. Preliminary results from Boltzmann learning. (A) Evolution of the total-population activity $N\bar{s}(t)$ of $N = 159$ neurons produced by the Glauber dynamics (implemented in NEST [96]; see section “Simulation of Glauber dynamics with NEST”) with 10^6 steps, during the sampling phase of the last Boltzmann-learning iteration. (B) Red, solid: Preliminary approximation of the probability distribution of the population-averaged activity, obtained via Boltzmann learning. Blue, dashed: empirical distribution of the population-averaged activity from the dataset shown in A. (C) Preliminary values of the time averages m_i and g_{ij} obtained from Boltzmann learning described in (A), versus the experimental ones shown in Fig 1. (D) Preliminary approximations of the population distributions of the Lagrange multipliers \hat{h}_i and \hat{J}_{ij} (associated with the averages m_i and g_{ij}) obtained via Boltzmann learning. Histogram bins in D are computed with Knuth’s rule [92].

<https://doi.org/10.1371/journal.pcbi.1005762.g002>

before discussing about curve-fitting properties, we want to make sure that our initial approximations are correct. In fact, we shall now see that these preliminary approximations are *not* correct in this case.

Appearance of bimodality. The preliminary results from Boltzmann learning do not show any inconsistency at this point. But now we sample the distribution for a much longer time: 5×10^7 steps, to verify whether these approximations have truly converged. The result is shown in Fig 3A. We find that after roughly 2×10^6 steps the whole population jumps to a high-activity regime and remains there until the end of the sampling. We have thus discovered that:

- our preliminary approximations of the Lagrange multipliers are wrong; their mismatch is shown in Fig 3C–3D;
- our preliminary approximation of the pairwise distribution, Fig 2B, is therefore wrong in two ways: it does not correspond to the (wrong) approximations of the Lagrange multipliers, and is not a pairwise maximum-entropy distribution;
- the Glauber dynamics has an additional metastable high-activity regime.

It is legitimate to wonder whether there are other metastable regimes. To test this possibility we start the dynamics with different numbers of initially active neurons. Two metastable regimes are observed (see Fig 3B): one at high activity and one at low activity. This means that

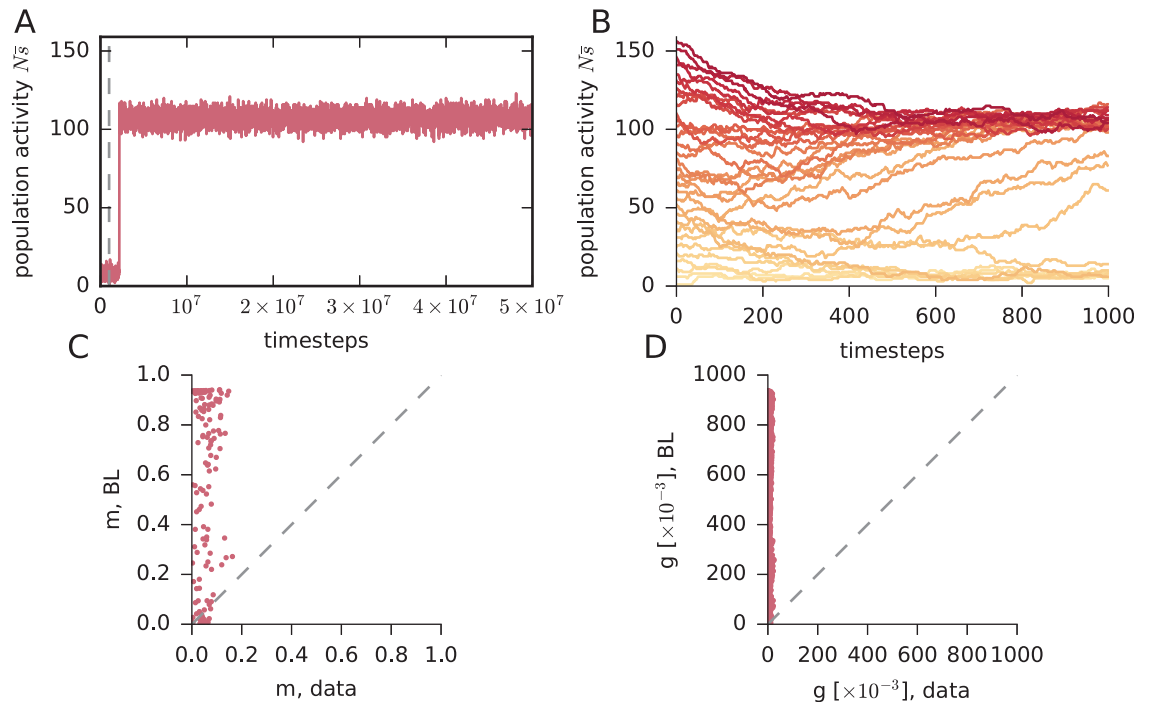


Fig 3. Longer sampling: Bistability. (A) Evolution of the total activity $N_s(t)$ produced by Glauber dynamics, as in Fig 2A, but with 5×10^7 steps. The dashed grey line marks the end of the previous shorter sampling of Fig 2A. (B) Evolutions of the total-population activity $N_s(t)$ obtained from several instances of Glauber dynamics. Each instance starts with a different value of the initial total activity $N_s(0)$, i.e. with $N_s(0)$ initially active neurons (chosen at random), and is represented by a different red shade, from $N_s(0) = 0$ (light red) to $N_s(0) = N$ (dark red). Note the two convergence basins, one at $N_s \approx 100$ and one at $N_s \approx 15$. (C, D) New values of the time averages m_i and g_{ij} versus the experimental ones. These new values are obtained from the longer Glauber dynamics described in (A) using the values of the Lagrange multipliers \hat{h}_i, \hat{J}_{ij} shown in Fig 2D, obtained from the previous Boltzmann learning. The plots clearly show that the values of \hat{h}_i, \hat{J}_{ij} found with the Boltzmann learning are not the ones yielding the constraints m_i, g_{ij} .

<https://doi.org/10.1371/journal.pcbi.1005762.g003>

the distribution associated with the initial, wrong approximations of the Lagrange multipliers of Fig 2D is *bimodal*, not unimodal as Fig 2B seemed to show.

Note that choosing as initial condition for the sampling procedure a state in the low-activity regime will not prevent the Glauber dynamics from entering the high activity state. As shown in Fig 3B, the system may still spontaneously transition into the high activity regime with a small but not negligible probability.

What do these facts imply? Let us recall that our primary goal is to model the data with an inhomogeneous pairwise maximum-entropy distribution. Boltzmann learning is just a procedure to find this distribution. This procedure explores the space of distributions in a particular way to find the correct one. What we just found says that this procedure entered a region of bimodal distributions in such space and got stuck there, without finding the correct distribution yet.

We must reflect on three main issues:

1. Boltzmann learning becomes impractically slow when it enters the bimodal region, because the Glauber dynamics that is part of this procedure becomes almost *non-ergodic*. Therefore it is an inefficient method to find the pairwise maximum-entropy distribution. Non-ergodicity is a known phenomenon in Monte Carlo methods; its solution requires longer sampling times or algorithms different from Glauber sampling [61, 62, 97, 98]. This problem also appears, for example, in the calculation of extensive parameters in finite-size statistical

mechanics in phase-transition regimes [99]. Note that bistability is not an *effect* of longer sampling; rather, longer sampling becomes a *necessity* because of bistability. This bistability is an inherent mathematical phenomenon caused by the positivity of the average correlations together with the large number N of units, as shown in sections “*Bimodality ranges and population size*” and “*Bimodality of the inhomogeneous model for large N* ”.

We could then try to use alternative procedures to find our desired distribution. The Thouless-Anderson-Palmer approximation [66, 67] and the Sessak-Monasson approximation [68, 69], for example, have been successfully used in the literature for this purpose. But unfortunately we find that these two approximations do not give the correct distribution either: properly sampled, the distributions they yield do not match the correlations and means of our data, just as in Fig 3C–3D. Evidently our data lie outside the domains of validity of these two approximations. Notably, the incorrect distributions given by these two approximations are also bimodal.

2. There may be one more problem ahead, though. Never mind that the procedures we know do not work; suppose we find a procedure that gives us the inhomogeneous pairwise distribution we are seeking. What should we do if this distribution turns out to be bimodal with a second mode at unrealistically high activities? Its Glauber dynamics would be bistable, yielding *sustained* periods of high activity, which would not be useful to generate meaningful, realistic surrogate data. Would we still be willing to use this maximum-entropy statistical model, or should we reject it altogether? And is the appearance of a bimodal distribution only peculiar to our data, or a more widespread feature of brain-activity data?
3. Besides, it is a pity that our initial approximation of the maximum-entropy distribution, Fig 2B, was incorrect. We had found *some* probability distribution, but it was *not* a pairwise maximum-entropy distribution; yet that distribution was modelling our data in an interesting way—and such modelling is our priority. This situation can be confusing, so let us explain it with an analogy. Imagine that we have some datapoints and we say “we want to fit the points with a *parabola*”. An incorrect fitting algorithm, however, gives us a curve that is not parabola. If this curve covers the datapoints in an interesting way, we may want to investigate what kind of curve it is. It could turn out to be a *hyperbole*, for example. We may then want to broaden our point of view and say “we want to fit the points with a *quadric*”, and use that curve. (We should still fix the fitting algorithm, though.) In our case, can we find out more about the probability distribution of Fig 2B? Could it also be a member of an extended maximum-entropy family for example?

We will shortly show that there is one solution that addresses these three issues all at once. We think, in fact, that it also addresses a fourth issue of maximum-entropy models, to be discussed later. We first analyse issues 2. and 3. in more detail; issue 1. above is subordinate to them.

Is the correct pairwise model bimodal? We would like to know whether the correct maximum-entropy distribution, which we have not found yet, is also bimodal like its incorrect approximation.

We make an educated guess by examining the analytically tractable reduced maximum-entropy model P_r , Eq (9). Using the population-averaged single and coupled activities as constraints, $E_r(\bar{s}_i) = \bar{m}$ and $E_r(\bar{s}_i \bar{s}_j) = \bar{g}$ from Eq (15), we numerically find the Lagrange multipliers of the reduced model:

$$h_r = -3.259, \quad J_r = 0.03859. \quad (16)$$

Note that in this case there is no sampling involved: the distribution can be calculated analytically, and the values Eq (16) are correct within the numerical precision of the maximization procedure (interior-point method [56, chap. 10]). The values of the expected single and couple activities, re-obtained by explicit summation (not sampling) from the corresponding reduced maximum-entropy distribution, agree with the values Eq (15) to seven significant figures.

The resulting reduced maximum-entropy distribution for the population-averaged activity, $P_r(\bar{s}|h_r, J_r)$, is shown in Fig 4A, together with the experimental frequency distribution of our data. Its corresponding Glauber dynamics with two metastable regimes is shown in Fig 4B. It shows a second maximum at roughly 90% activity.

A mathematically exact analysis of smaller subsets of our population with the inhomogeneous maximum-entropy model, and an analysis of the full population and large subsets of it with a reduced maximum-entropy model having higher-order constraints $E(N\bar{s})$, $E((N\bar{s})^2)$, $E\left(\frac{N\bar{s}(N\bar{s}-1)}{N(N-1)} - \left(\frac{N\bar{s}(N\bar{s}-1)}{N(N-1)}\right)^2\right)$ (the latter corresponding to the variance of the second moments), show that if a reduced maximum-entropy model is bimodal, the full inhomogeneous model is also bimodal, with a heightened second mode shifted towards lower activities with respect to the reduced model.

The bimodality encountered in the Boltzman learning, the bimodality of the reduced maximum-entropy model, the bimodality of the full maximum-entropy model for small populations, and finally the bimodality for the reduced model with higher-order constraints, together constitute strong evidence that the correct pairwise maximum-entropy distribution for our data is bimodal. In section “Bimodality of the inhomogeneous model for large N ” we prove that this must be true for large N , even if it were not true for our specific population size $N = 159$.

Bimodality ranges and population size. Next we want to address whether it is common or rare that the pairwise maximum-entropy method yields bimodal distributions for neuronal brain-activity data. For this purpose we first estimate the ranges of firing rates and correlations for which maximum-entropy yields a bimodal distribution; then we check whether typical experimental values fall within these ranges. We are particularly interested in how bimodality depends on the recorded population size N .

We again make an educated guess using the reduced maximum-entropy model, with distribution for the population average \bar{s} , $P_r(\bar{s}|h_r, J_r)$, Eq (11). The distribution has two maxima if it has one minimum for some value \bar{s}_m , $0 < \bar{s}_m < 1$. An elementary study of the convexity properties (second derivative) of this distribution shows that it has one minimum at $0 < \bar{s}_m < 1$ if

$$\frac{dP_r(\bar{s}|h_r, J_r)}{d\bar{s}} \Big|_{\bar{s}=\bar{s}_m} = 0, \quad \frac{d^2P_r(\bar{s}|h_r, J_r)}{d\bar{s}^2} \Big|_{\bar{s}=\bar{s}_m} > 0, \quad 0 < \bar{s}_m < 1, \tag{17}$$

These conditions can be solved analytically and give the ranges of the multipliers (h_r, J_r) for which bimodality occurs, parametrically in (\bar{s}_m, J_r) :

$$\left\{ \begin{array}{l} 0 < \bar{s}_m < 1, \\ J_r > \Psi'[1 + (1 - \bar{s}_m)N] + \Psi'(1 + \bar{s}_mN), \\ h_r(\bar{s}_m, J_r) = J_r/2 - \bar{s}_mN J_r - \Psi[1 + (1 - \bar{s}_m)N] + \Psi(1 + \bar{s}_mN), \end{array} \right. \tag{18}$$

where $\Psi(x) := d \ln \Gamma(x)/dx$, Γ being the Gamma function [100, chs 43–44]. We express the population-averaged activity $E_r(\bar{s})$ and the Pearson correlation $\bar{\rho}$, typically used in the literature, in terms of (h_r, J_r) using the definitions Eq (14) and the probability Eq (11). In this way we finally obtain the bimodality range for $(E_r(\bar{s}), \bar{\rho})$, parametrically in (\bar{s}_m, J_r) : the results are shown in Fig 5A for various values of the number of neurons N .

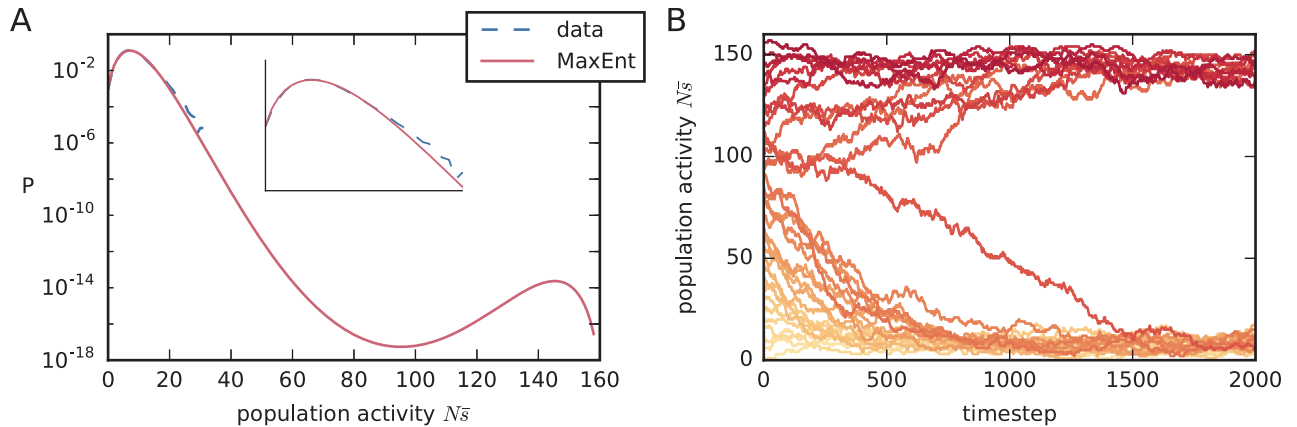


Fig 4. Reduced maximum-entropy model. (A) Red, solid: Probability distribution for the population-averaged activity, $P_r(\bar{s})$ given by the reduced model for our dataset Eq (15); note the two probability maxima. Blue, dashed: empirical frequency distribution of the population-averaged activity from our dataset. (B) Population-averaged activities $\bar{s}(t)$ obtained from several instances of the Glauber dynamics associated with the reduced model, with homogeneous couplings, $J_{ij} = J_r$, and biases, $h_i = h_r$, of Eq (16). As in Fig 3, each instance starts with an initial population activity $\mathbf{s}(0)$ having different values of population average $\bar{s}(0)$, and is represented by a different red shaded curve. The initial values range from $N\bar{s}(0) = 0$ (light red) to $N\bar{s}(0) = N$ (dark red).

<https://doi.org/10.1371/journal.pcbi.1005762.g004>

For each N we have a curve in the plane $E_r(\bar{s}), \bar{\rho}$. Values of $(E_r(\bar{s}), \bar{\rho})$ above such curves yield a bimodal pairwise maximum-entropy distribution in the homogeneous case, for the corresponding population size N . The plot notably shows that the range of constraints yielding bimodality increases with N . Fig 5B displays the probability distribution of the population-averaged activity for the constraints from our dataset Eq (15) but different values of N . When $N \lesssim 150$ the distribution has only one maximum at low activity, $\bar{s} \approx 0.0497$, and when $N \gtrsim 150$ a second probability maximum at high activity, $\bar{s} \approx 0.9502$, appears. The probability at this second maximum increases sharply until $N \approx 200$ and thereafter maintains an approximately stable value, roughly 6000 times smaller than the low-activity maximum. The minimum between the two modes becomes deeper and deeper as we increase N above 200.

As mentioned in the previous section, exact studies with small samples and studies with large samples and a reduced model with higher-order constraints indicate that the high-activity maximum in the inhomogeneous case is even larger (roughly 2000 times smaller than the low-activity one when $N = 1000$) and shifted towards lower activities ($\bar{s} \approx 0.25$ when $N = 1000$).

This can also be seen by adding a Gaussian jitter to the multipliers of the reduced case $h_i = h_r, J_{ij} = J_r$, thereby making the model inhomogeneous. The results for small and large jitter are shown in Fig 5C–5D, respectively. The basin of attraction of the second metastable regime is shifted to lower activities, and transitions between the two metastable regimes become more likely for larger jitters. This means that inhomogeneity makes the minimum in between the two modes shallower.

The population-averaged activity and Pearson correlation of our data (violet “3 ms” point in Fig 5A) fall within the bimodality range.

Bimodality of pairwise models for massively parallel data. Having found the bimodality ranges of firing rates and correlations in section “Bimodality ranges and population size”, we now ascertain whether our dataset is a typical representative leading to bimodal pairwise distributions, or an outlier. We take as reference the data summarized in Table 1 of Cohen & Kohn [37], which reports firing rates and spike-count correlations r_{SC} for several experimental recordings of brain activity. The reported firing rates correspond to population-averaged

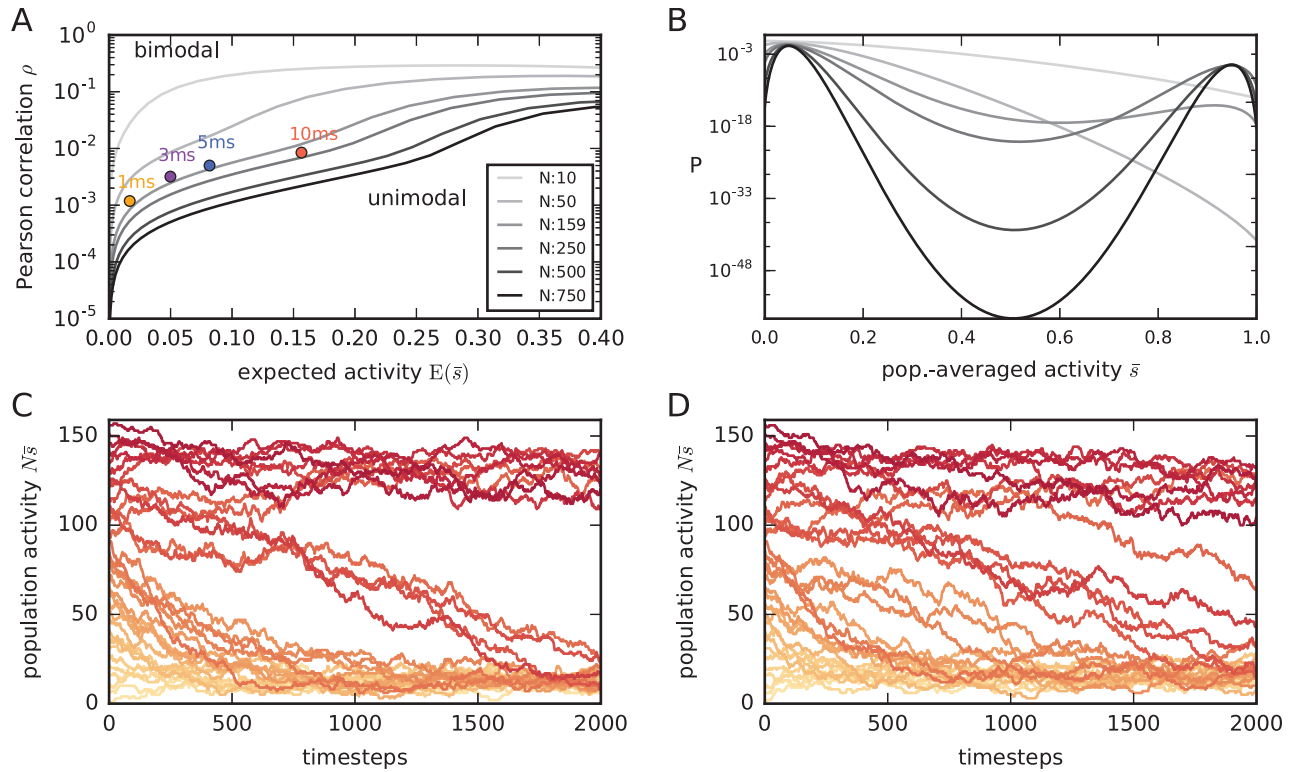


Fig 5. Bimodality ranges for the reduced model and effects of inhomogeneity. (A) The reduced maximum-entropy model Eq (9) yields a distribution $P_r(\bar{s})$ that is either unimodal or bimodal, depending on the number of neurons N and the values of the experimental constraints $(E_r(\bar{s}), \bar{\rho})$. Each curve in the plot corresponds to a particular N (see legend) and separates the values $(E_r(\bar{s}), \bar{\rho})$ yielding a unimodal distribution, below the curve, from those yielding a bimodal one, above the curve. The curves are symmetric with respect to $E_r(\bar{s}) = 0.5$ (ranges $E_r(\bar{s}) > 0.4$ not shown). Note how the range of constraints yielding bimodality increases with N . Coloured dots show the experimental constraints from our dataset for different time-binnings with widths $\Delta = 1$ ms, $\Delta = 3$ ms, $\Delta = 5$ ms, $\Delta = 10$ ms: all these binnings yield a bimodal distribution. (B) Probability distributions of the reduced model for the population-averaged activity, $P_r(\bar{s}|N)$, obtained using the constraints Eq (15) from our data set (3 ms purple dot in panel A) and different N (same colour legend as panel A). (C) Population-averaged activities $\bar{s}(t)$ from several instances of Glauber dynamics, all with the same normally-distributed couplings J_{ij} and biases h_i , with means as in Eq (16) and Fig 4B, and standard deviations $\sigma(J_{ij}) = 0.009$, $\sigma(h_i) = 0.8$. Each instance starts with an initial population activity $\mathbf{s}(0)$ having different values of the population average $\bar{s}(0)$, and is represented by a different red shade, from $N\bar{s}(0) = 0$ (light red) to $N\bar{s}(0) = N$ (dark red). Note how the basins of attraction of the two metastable regimes are wider than in the homogeneous case of Fig 4B. (D) The same as panel C, but with larger standard deviations $\sigma(J_{ij}) = 0.012$, $\sigma(h_i) = 1.08$; the jumps between the two metastable regimes become more frequent than in Fig 4B, indicating that the minimum between the modes becomes shallower as the inhomogeneity increases.

<https://doi.org/10.1371/journal.pcbi.1005762.g005>

activities \bar{m} ranging between 0.02 and 0.25 if we use 3 ms time bins; thus our data are well within this range.

The values reported for the correlations r_{SC} in [37] are given for the spike counts measured in large time intervals: several hundred of milliseconds. We therefore need a coarse estimate of the Pearson correlation coefficient ρ that would be measured on a fine temporal scale of 3 ms in the same experiment. Both r_{SC} and ρ are particular cases of the Pearson correlation coefficient r_{CCG} of spike counts in a window τ , as introduced by Bair et al. [101, App. A]:

$$r_{CCG_{ij}}(\tau) := \frac{E(v_i(\tau)v_j(\tau)) - E(v_i(\tau))E(v_j(\tau))}{\sqrt{[E(v_i(\tau)^2) - E(v_i(\tau))^2][E(v_j(\tau)^2) - E(v_j(\tau))^2]}} \quad (19)$$

$$\text{where } v_i(\tau) := \sum_{t=1}^{\tau} s_i(t),$$

i.e. $v_i(\tau)$ is the number of spikes of neuron i in the time window $\tau\Delta$. Here $s_i(t)$ are the binary representations of the spike trains binned with bin width Δ , in line with section “Pairwise maximum-entropy model”. This metric also equals the area between times $-\tau\Delta$ and $\tau\Delta$ under the cross-correlogram of neurons i and j (stationarity is assumed), calculated on the fine temporal time scale Δ . The spike count correlation r_{SC} corresponds to $\tau = n \equiv T/\Delta$, and our Pearson correlation ρ to $\tau = 1$. Several studies [we analysed: 101–105] report either measured values of $r_{CCG}(\tau)$ for different windows τ , or measured cross-correlograms. We studied, one by one, all the measures reported in the cited studies and numerically found that each of them satisfies $\rho \gtrsim r_{SC}/20$. We decide to take the lower bound $\bar{\rho} = \bar{r}_{SC}/20$ as a coarse estimate of ρ given r_{SC} , because it leads to points as far away from bimodality as possible, i.e. because it is biased *against* our conjecture.

Under these approximations—and notwithstanding the choice of estimates that keeps the data as far away from bimodality as possible—the largest part of the data summarized by Cohen & Kohn does fall in the bimodality region of Fig 5A for $N = 250$, and almost all data lies in the bimodality region for $N = 500$; see Fig 6. These data points have only an indicative value, but suggest that our dataset is not an outlier for the bimodality problem. If those data had been recorded from a population of 500 neurons, they would have yielded a bimodal pairwise maximum-entropy model because, as shown in section “Bimodality ranges and population size”, the more neurons we are able to record, the more likely the bimodality occurs. Thus the bimodality problem and its consequences need to be taken seriously. Our next question is then: Is there any way to eliminate the bimodality problem?

Analysis of the erroneous preliminary approximation of the distribution. In the last three sections we have partly addressed the second issue raised in section “Appearance of bimodality”: the distribution given by the pairwise maximum-entropy method is bound to be bimodal, not only for our data, but also for typical neuronal recordings of a couple of hundreds neurons or more.

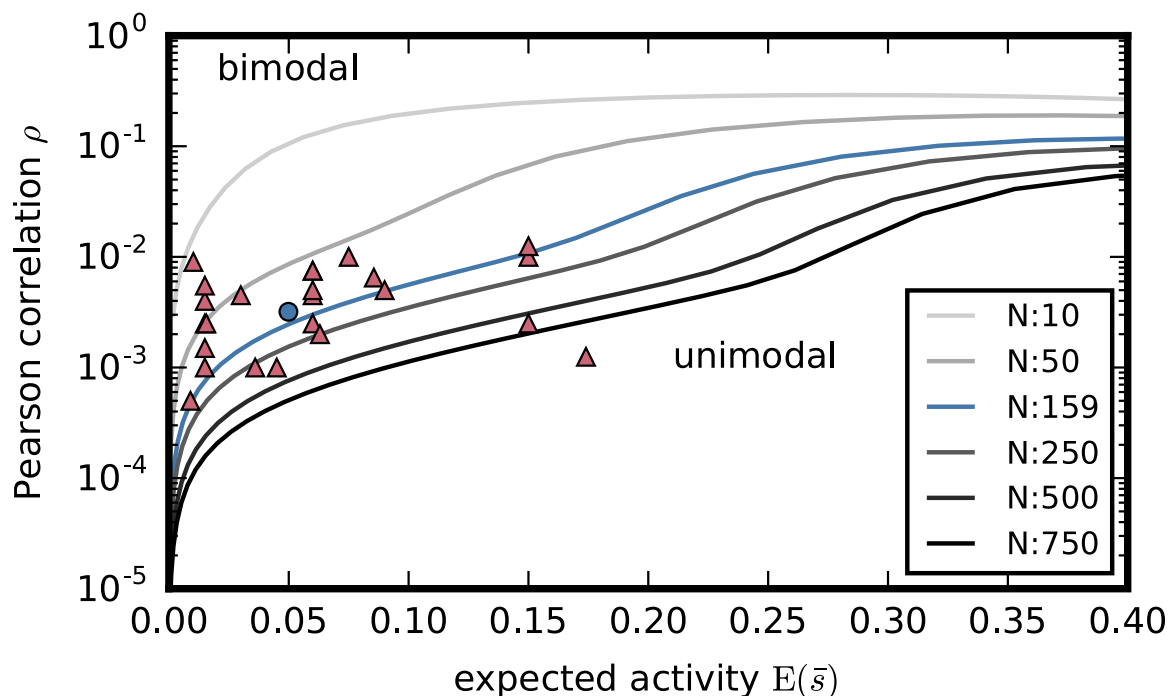


Fig 6. Bimodality for experimental data from neuroscientific literature. Mean activities and correlations ($E(\bar{s}), \bar{\rho}$) inferred from experimental data reported in Cohen & Kohn [37, Table 1], plotted upon the curves separating bimodal from unimodal maximum-entropy distributions of Fig 5A. The plot suggests that typical experimental neural recordings of 250 neurons and above are likely to lead to bimodal maximum-entropy pairwise distributions.

<https://doi.org/10.1371/journal.pcbi.1005762.g006>

We now address the third issue: to find out more about the first erroneous approximation of the probability distribution, shown in Fig 2B. It is important to remember that it is *not* the correct pairwise maximum-entropy distribution, and the initial erroneous approximations $\hat{\mathbf{h}}, \hat{\mathbf{J}}$ of the multipliers do *not* match the data. The approximation was erroneous because the sampling phase of the Boltzmann-learning algorithm was too brief. The time required to explore the full distribution is so long that the dynamics is non-ergodic for computational purposes. This non-ergodicity effectively truncates the sampling at states \mathbf{s} for which $\bar{s} \lesssim \theta$, where θ is the population-averaged activity at the trough between the two metastable regimes. This means that if the wrong multipliers $\hat{\mathbf{h}}, \hat{\mathbf{J}}$ are used in a “truncated” distribution

$$P_t(\mathbf{s}|\hat{\mathbf{h}}, \hat{\mathbf{J}}, \theta) \propto \begin{cases} \exp(\sum_i \hat{h}_i s_i + \sum_{i<j} \hat{J}_{ij} s_i s_j), & \bar{s} \leq \theta, \\ 0, & \bar{s} > \theta, \end{cases} \quad (20)$$

then the expectations of this distribution are close to the experimental time averages for the single and coupled activities: $E_t(s_i) = m_i$, $E_t(s_i s_j) = g_{ij}$. This truncated distribution, though, is obviously *not* a pairwise maximum-entropy distribution. It is an interesting distribution nevertheless, as Fig 2B shows. Could it be obtained or approximated with the maximum-entropy method in some other way?

Solution: An inhibited maximum-entropy model and Glauber dynamics

Let us briefly summarize our results so far and the reason why a maximum-entropy model yielding a bimodal distribution in the population-averaged activity is problematic:

- For commonly observed statistics of neuronal data, the pairwise maximum-entropy method yields a distribution with two distinct modes (bimodality), one of which at high activities—unrealistic in view of present neuroscientific data. The bimodality is bound to happen for large N as soon as the measured correlations are slightly positive. Moreover, the Glauber dynamics based on the pairwise model jumps between two metastable regimes and cannot be used to generate realistic surrogate data.
- The Boltzmann-learning procedure based on Glauber dynamics becomes practically non-ergodic and the Lagrange multipliers of the pairwise model are difficult or impossible to find. Standard analytic approximations fail as well.
- The initial erroneous approximation of the pairwise maximum-entropy distribution, obtained with a too short Boltzmann learning, shows an interesting fit with the data nevertheless. It would be interesting to know if it can be obtained or approximated with a generalized maximum-entropy method.

We will propose a solution that addresses all three issues at once. This solution pivots on the idea of *inhibition* and can be grasped with an intuitive explanation of how the bimodality arises.

Intuitive understanding of the bimodality: Mean-field picture. From the point of view of a network evolving with a Glauber dynamics with couplings \mathbf{J} and biases \mathbf{h} , the bimodality and bistability appear because the couplings \mathbf{J} are on average positive and make the network dominantly excitatory. The positivity of the couplings appears because the average correlation \bar{c} between the neurons, empirically measured, is positive (Figs 1D and 7A). This phenomenon is not unknown: bimodalities in the distribution of extensive quantities have a similar explanation in the statistical mechanics of finite-size systems [99, 106–108].

A naive mean-field analysis also confirms this. In such an approximation we imagine that each neuron is coupled to a field representing the mean activities of all other neurons [65; 109,

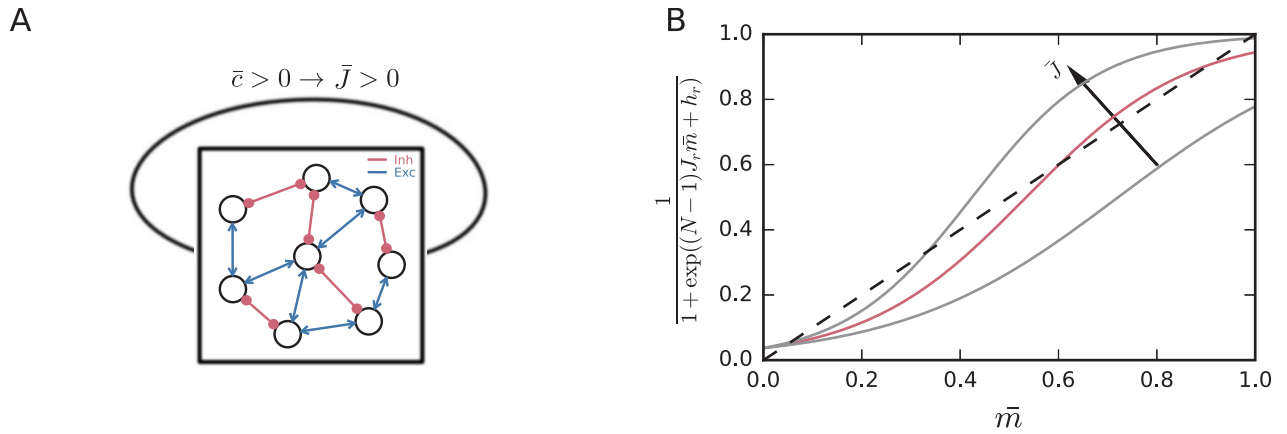


Fig 7. Mean-field picture. (A) Illustration of a self-coupled symmetric network that is self-excitatory on average. Arrow-headed blue lines (\rightarrow) represent excitatory couplings; circle-headed red lines (\rightarrow) represent inhibitory couplings. (B) Self-consistency solution of the naive mean-field equation, illustrated for different J_r . Larger J_r lead to two additional intersections, corresponding to an unstable and a stable solution. The red curve corresponds to the J_r calculated from our experimental data Eq (16).

<https://doi.org/10.1371/journal.pcbi.1005762.g007>

ch. 6]. From the point of view of entropy maximization, we are replacing the maximum-entropy distribution with one representing independent activities, having minimal Kullback-Leibler divergence from the original one [66, chs 2, 16, 17]. Given the couplings \mathbf{J} and biases \mathbf{h} , the mean activities \mathbf{m} must satisfy N self-consistency equations

$$\frac{1}{1 + \exp(\sum_{j \neq i} J_{ij} m_j + h_i)} = m_i. \tag{21}$$

In the homogeneous case they reduce to the equation $\{1 + \exp[(N - 1)J_r \bar{m} + h_r]\}^{-1} = \bar{m}$, corresponding to the intersection of two functions of \bar{m} : the diagonal line $\bar{m} \mapsto \bar{m}$, and the curve $\bar{m} \mapsto \{1 + \exp[(N - 1)J_r \bar{m} + h_r]\}^{-1}$ that depends parametrically on (h_r, J_r) ; see Fig 7B. For the Lagrange multipliers of our data, these curves intersect at two different values of \bar{m} , meaning that there are two solutions to the self-consistency equation, corresponding to two different mean activities. These approximately correspond to the maxima of the probability distribution for the population average in Fig 4A.

Importance of inhibition: Modified Glauber dynamics. In the neuronal network dynamics just analysed, with positive correlations on average, the second peak of high activity can be suppressed introducing an effective negative feedback loop. Such inhibitory mechanism can be represented by an additional neuronal unit I , having positive incoming couplings $J_{Ik} > 0$ and negative outgoing couplings $J_{kl} < 0$ with all other units k . This unit is therefore activated if the total activity of the other units is high, and once activated it provides negative input back to all other units. This situation is illustrated in Fig 8A. Such a stabilizing mechanism acts much in the same way as inhibition stabilizes the low-activity state in neuronal networks [110, 111]; it has also been used in the simulations by Bohte et al. [34]. The asymmetry of this mechanism is in contradiction with the symmetry of the couplings J_{ij} of the Glauber dynamics, however. We want to break this symmetry and add an asymmetric inhibitory feedback to the Glauber dynamics to avoid the bimodality of the probability distribution.

We preliminarily implement this idea in our Glauber dynamics, to observe its consequences, using first the incorrect approximations $(\hat{\mathbf{h}}, \hat{\mathbf{J}})$ of the multipliers that yielded a bimodal inhomogeneous pairwise distribution, and then the (correct) multipliers $h_i = h_r, J_{ij} = J_r$, Eq (16), of the reduced pairwise model. We connect all N neurons to a single inhibitory

neuron that instantaneously activates whenever their average activity exceeds a threshold $\theta \in \{0, 1/N, 2/N, \dots, (N-1)/N, 1\}$. Upon activation the inhibitory neuron inhibits the other N neurons (see Fig 8A) via N identical negative couplings $J_I < 0$. The results from the simulation of the inhibited Glauber dynamics are shown in Fig 8; in all cases the inhibitory coupling was $J_I = -24.7$ and the inhibition threshold $\theta = 0.3$.

The algorithm for sampling from this “inhibited” Glauber dynamics is explained in section “Inhibited Glauber dynamics”. It can be seen that the additional inhibitory neuron eliminates the bistability, leaving only the stable low-activity regime. The resulting homogeneous and inhomogeneous stationary distributions (in the inhomogeneous case J_{ij} and h_i are normally distributed as in Fig 5C) are either unimodal or have a second mode that is completely negligible, being tens or hundreds of orders of magnitude smaller than the first mode.

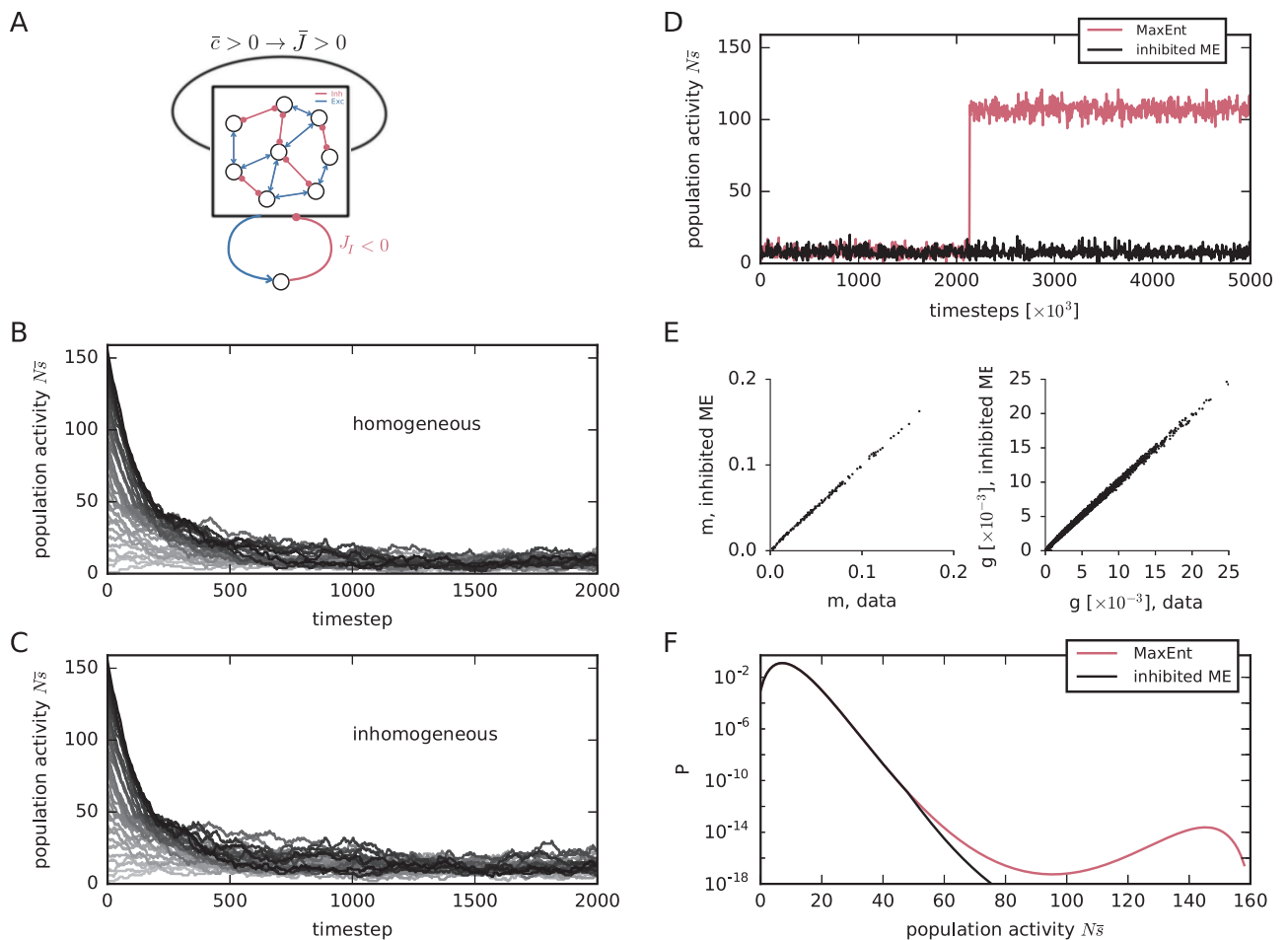


Fig 8. Asymmetric inhibition and elimination of bimodality and non-ergodicity. (A) Illustration of self-coupled network with additional asymmetric inhibitory feedback. Each neuron receives inhibitory input $J_I < 0$ from the additional neuron whenever the population-average \bar{s} becomes greater than the inhibition threshold θ . (B) Population-averaged activities $\bar{s}(t)$ from several instances of the inhibited Glauber dynamics, with $J_I = -24.7$, $\theta = 0.3$, and homogeneous $J_{ij} = J_i$, $h_i = h_i$ of Eq (16), as used for Fig 4B. Each instance starts with an initial population activity $\mathbf{s}(0)$ having different values of the population average $\bar{s}(0)$, and is represented by a different grey shade, from $N\bar{s}(0) = 0$ (light grey) to $N\bar{s}(0) = N$ (black). Note the disappearance, thanks to inhibition, of the bistability that was evident in the “uninhibited” case of Fig 4B. (C) Analogous to panel B, with $J_I = -24.7$, $\theta = 0.3$, but inhomogeneous normally distributed couplings and biases as in the uninhibited case of Fig 5C. The bistability again disappears thanks to inhibition. (D) Comparison of a longer (5×10^6 timesteps) Glauber sampling in the inhibited (black, $J_I = -24.7$, $\theta = 0.3$) and uninhibited (red) case, using the couplings and biases of Fig 2D obtained from our first Boltzmann learning. (E) Time averages m_i and g_{ij} obtained from Boltzmann learning for the inhibited model P_i , versus experimental ones. (F) Probability distribution of the population-averaged activity $P_i(\bar{s})$ given by the inhibited model Eq (22) for our dataset Eq (15), compared with the one previously given by the reduced model $P_r(\bar{s})$, Fig 4A.

<https://doi.org/10.1371/journal.pcbi.1005762.g008>

The inhibited Glauber dynamics can suppress the bistability for any network size N , with an appropriate choice of the inhibitory coupling $J_1 < 0$ and threshold θ .

Continuing our exploration, we check what happens if we use the inhibited Glauber dynamics in the sampling phase of Boltzmann learning for our initial problem, when we tried to find a stationary distribution having means and correlations shown in Fig 1B–1D. As shown in Fig 8E, the addition of the inhibitory neuron (again with $J_1 = -24.7$, $\theta = 0.3$) eliminates the second metastable state that appeared after 2×10^6 steps, cf. Fig 3. The resulting couplings and biases of the final stationary distribution are distributed as in Fig 2D—but note that this time the Boltzmann learning has converged, hence these are its correct final values.

However, it must be stressed that the stationary distribution thus found, using the inhibited Glauber dynamics, is *not* a pairwise maximum-entropy distribution, because the latter is the stationary distribution of the original Glauber dynamics, not of the modified one. If we use the inhibited dynamics in Boltzmann learning, we are abandoning the standard pairwise maximum-entropy model.

There is nevertheless a positive result: in the next section we show that the stationary distribution of the inhibited Glauber dynamics belongs to a generalized maximum-entropy family.

Inhibited maximum-entropy model. The pairwise maximum-entropy distribution Eq (3) is the stationary distribution of the Glauber dynamics with symmetric couplings. It is not the stationary distribution of the inhibited Glauber dynamics. But the following fact holds: *The stationary distribution of the inhibited Glauber dynamics (Fig 8) belongs to the maximum-entropy family. Its analytic expression is*

$$\begin{aligned}
 P_i(\mathbf{s}|\mathbf{h}, \mathbf{J}, J_1, \theta) &= \frac{1}{Z_i(\mathbf{h}, \mathbf{J}, J_1, \theta)} \times \\
 &\quad \exp \left[\sum_i h_i s_i + \sum_{i>j} J_{ij} s_i s_j + J_1 N G(\bar{s} - \theta) \right], \\
 Z_i(\mathbf{h}, \mathbf{J}, J_1, \theta) &:= \sum_{\mathbf{s}} \exp \left[\sum_i h_i s_i + \sum_{i>j} J_{ij} s_i s_j + J_1 N G(\bar{s} - \theta) \right], \\
 G(\bar{s} - \theta) &:= (\bar{s} - \theta) H(\bar{s} - \theta),
 \end{aligned}
 \tag{22}$$

where J_1 is the (negative, in our case) coupling strength from the inhibitory neuron to the other neurons, θ is the activation threshold of the inhibitory neuron, and H is the Heaviside step function. We call Eq (22) the *inhibited pairwise maximum-entropy model*. The proof that it is the stationary distribution of the inhibited Glauber dynamics is given in section “*Inhibited Glauber dynamics*”.

This maximum-entropy model is characterized by the new term $J_1 N G(\bar{s} - \theta)$ in the exponential, which we call “inhibition term”. The function $G(\bar{s} - \theta)$ is plotted in Fig 9 together with its exponential. We show in section “*Expansion of the inhibition term in terms of higher-order coupled activities*” that it can also be written as a linear combination of population-averaged K -tuple activities, $s_{i_1} s_{i_2} \cdots s_{i_K}$, for K equal to $N\theta + 1$ and larger:

$$N G(\bar{s} - \theta) = \sum_{K=N\theta+1}^N \binom{-N\theta}{K - N\theta - 1} \left(\sum_{i_1 < i_2 < \cdots < i_K} s_{i_1} s_{i_2} \cdots s_{i_K} \right),
 \tag{23}$$

the coefficients being generalized binomial coefficients [100, ch. 6; see also 112], which have alternating signs. For example, if $N = 5$ and $\theta = 3/5$,

$$NG(\bar{s} - \theta) = (s_2 s_3 s_4 s_5 + s_1 s_3 s_4 s_5 + s_1 s_2 s_4 s_5 + s_1 s_2 s_3 s_5 + s_1 s_2 s_3 s_4) - 3s_1 s_2 s_3 s_4 s_5.
 \tag{24}$$

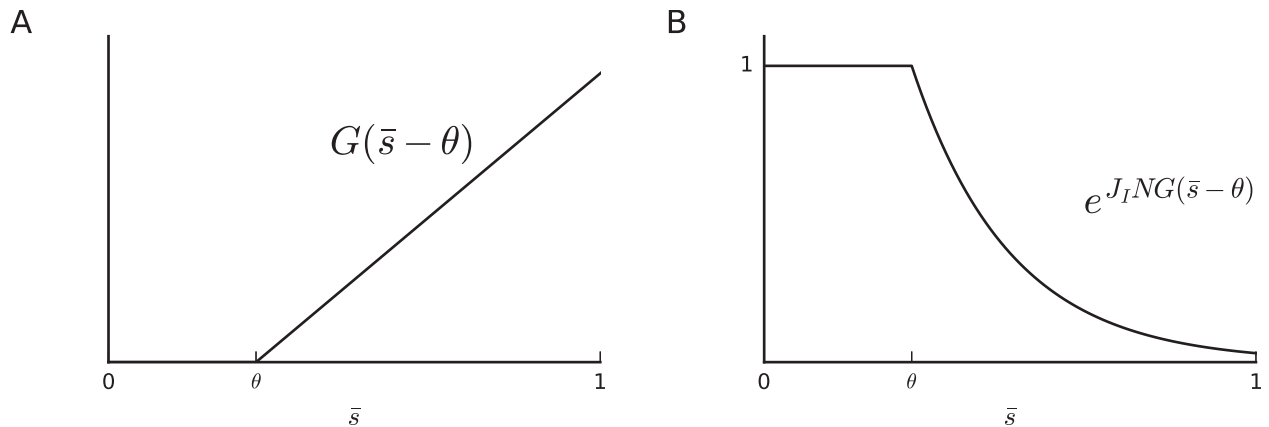


Fig 9. Reference measure. The function $G(\bar{s} - \theta)$ and its exponential for $J_I < 0$.

<https://doi.org/10.1371/journal.pcbi.1005762.g009>

This function differs from the additional function appearing in the maximum-entropy model by Tkačik et al. [50, 94, 113], which consists in $N + 1$ constraints enforcing the observed frequency distribution of the population average \bar{s} . For large data samples the constraints used in those works typically equal 0 for larger values of \bar{s} ; for reasons discussed in section “Range of applicability of maximum-entropy models”, the use of such extreme constraints may not be justified or meaningful.

The inhibited distribution $P_i(s)$ belongs to the maximum-entropy family in two different ways, the first preferable to the second:

1. It can be obtained by application of the maximum-relative-entropy (minimum-discrimination-information) principle [31, 83], with the pairwise constraints Eq (5), with respect to the reference distribution

$$P_0(s|J_I, \theta) \propto \exp [J_I N G(\bar{s} - \theta)], \quad N\theta \in \{0, 1, 2, \dots, N - 1, N\}, \quad (25)$$

also called “reference measure”, which assigns decreasing probabilities to states with average activities above θ ; see Fig 9B. This relative-maximum-entropy model can be interpreted as arising from a more detailed model in which we know that external inhibitory units make activities above the threshold θ increasingly improbable, like Bohte et al.’s model [34] for example. We discuss this in the “Discussion”. In this interpretation the parameters J_I and θ are chosen a priori.

2. Alternatively, it results from the application of the “bare” maximum-entropy principle given the pairwise constraints Eq (5) and an additional constraint for the expectation of $N G(\bar{s} - \theta)$:

$$\begin{aligned} E_i(N G(\bar{s} - \theta)) &= \sum_{N\bar{s}=N\theta}^N (N\bar{s} - N\theta) P_i(\bar{s}) \\ &= \sum_{K=N\theta}^N \binom{-N\theta}{-K+1} E_i \left(\sum_{i_1 < i_2 < \dots < i_K} s_{i_1} s_{i_2} \dots s_{i_K} \right). \end{aligned} \quad (26)$$

This is a constraint on what could be called the “tail first moment” of the distribution for the population-averaged activity $P_i(\bar{s})$: it determines whether the right tail of $P_i(\bar{s})$ has a small ($J_I < 0$) or heavy ($J_I > 0$) probability. It can also be seen as a constraint on the $N\theta$ -th

and higher moments, owing to Eq (23). In this interpretation the parameter J_I is the Lagrange multiplier associated with this constraint, hence it is determined by the data; the parameter θ is chosen a priori. Note, however, that experimental data are likely to give a vanishing time average of $N G(\bar{s} - \theta)$, so that $J_I = -\infty$. This interpretation has therefore to be used with care, for the reasons discussed in section “Range of applicability of maximum-entropy models”.

Two features of the inhibited maximum-entropy model Eq (22) are worth remarking upon:

1. The family of inhibited distributions P_i includes the pairwise family P_p , Eq (3), as the particular case $J_I = 0$. Note that if $J_I \neq 0$ the inhibited and uninhibited models with identical Lagrange multipliers (\mathbf{h}, \mathbf{J}) have *different* expectations for single and coupled activities:

$$\begin{aligned} E_i(s_i|\mathbf{h}, \mathbf{J}, J_I, \theta) &\neq E_p(s_i|\mathbf{h}, \mathbf{J}), \\ E_i(s_i s_j|\mathbf{h}, \mathbf{J}, J_I, \theta) &\neq E_p(s_i s_j|\mathbf{h}, \mathbf{J}), \end{aligned} \tag{27}$$

and therefore different covariances and correlations.

2. If the inhibitory parameter J_I is very large and negative, all activities \mathbf{s} such that $\bar{s} > \theta$ are assigned a negligible probability by the inhibited model. We therefore have the approximation (technically, pointwise convergence as $J_I \rightarrow -\infty$)

$$P_i(\mathbf{s}|\mathbf{h}, \mathbf{J}, J_I, \theta) \propto \begin{cases} \exp(\sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j), & \bar{s} \leq \theta, \\ 0, & \bar{s} > \theta, \end{cases} \quad J_I \ll 0, \tag{28}$$

where C is an appropriate normalization constant.

But the last expression is identical to Eq (20). Thus, *the inhibited maximum-entropy distribution P_i is approximately equal to the truncated distribution P_t* —the incorrect one Eq (20) obtained with our initial Boltzmann learning—having the same multipliers (\mathbf{h}, \mathbf{J}) and threshold θ :

$$P_i(\mathbf{s}|\mathbf{h}, \mathbf{J}, J_I, \theta) \approx P_t(\mathbf{s}|\mathbf{h}, \mathbf{J}, \theta), \quad J_I \ll 0, \tag{29}$$

and their expectations are also approximately equal. We have therefore addressed the third issue discussed in section “Appearance of bimodality”: the initial, incorrect but interesting approximation can actually be rescued, as we now explain.

Summary: Application of the inhibited maximum-relative-entropy model. Let us find a probability distribution for our data using the maximum-relative-entropy method, with reference measure Eq (25), $J_I = -24.7$, $\theta = 0.3$, given the single and pairwise constraints Eq (5) summarized in Fig 1C and 1D. To find the Lagrange multipliers (\mathbf{h}, \mathbf{J}) and the distribution Eq (22) we use the Boltzmann learning procedure with 5×10^6 timesteps. For the sampling phase we must use the inhibited Glauber dynamics, because its stationary distribution is Eq (22). As proven earlier (Fig 8) no bistability arises, so we do not need to worry about wrong approximations from undersampling; in fact, a much smaller number of timesteps would suffice.

The resulting multipliers and distribution are not shown because their plots are indistinguishable to the naked eye from the homologous ones in Fig 2B–2D. We have thus found that the interesting distribution shown in red in Fig 2B, although not a pairwise maximum-entropy distribution, still belongs to an enlarged maximum-entropy family: *it is an inhibited, inhomogeneous, pairwise maximum-relative-entropy distribution.*

The inhibited maximum-relative-entropy model therefore solves all these issues presented in section “*Appearance of bimodality*”:

1. The multipliers and distribution of this model can efficiently be found via Boltzmann learning with the typical number of timesteps used in the literature (200–400 [45, 95]), because its Glauber dynamics is not affected by bistability (Fig 8).
2. This model produces a distribution that has no unrealistic modes at high activities.
3. This distribution has interesting features and fitting properties (Fig 2B).

We can ask whether the solution of these three issues by the inhibited model is enough to warrant its use. In particular, does the use of the reference measure (Eq (25)) make sense from a neurobiological standpoint? In the following “*Discussion*” we argue that it does, and that it solves in fact a fourth issue of standard (i.e. uniform-measure) maximum-entropy models for neuronal networks.

Discussion

Summary

In this work we have shown that pairwise maximum-entropy models, widely used as reference distributions in the statistical description of the joint activity of hundreds of neurons, are poised to suffer from three interrelated problems when constrained with mean activities and pairwise correlations typically found in cortex:

1. Boltzmann-learning [64, 114] based on asynchronous Glauber dynamics [61, 63, chap. 29], used to find the Lagrange multipliers and distributions of these models, becomes practically non-ergodic (Fig 3), already for population sizes of roughly 50 neurons. The distribution is therefore difficult or impossible to find. Approximate methods like mean-field [65, 66], Thouless-Anderson-Palmer [66, 67], Sessak-Monasson [69, and refs therein] also break down in this case. This problem is known in the statistical mechanics of finite-size systems [99]. This non-ergodicity can go undetected; see “*Detection and further study of bimodality*” below.
2. Pairwise models are bound to give a *bimodal* probability distribution as soon as a critical number of units is exceeded. We have provided experimental evidence for this claim in section “*The problem: Bimodality, bistability, non-ergodicity*”. The first mode is observed in the data. But the model also predicts a second, *unobserved* mode at very high activities, with up to 90% of the population simultaneously active for long times. The probability of the second mode increases with population size. The Glauber dynamics based on this model jumps between two metastable regimes, remaining in each for long times (owing to its asynchronous update) and cannot be used to generate realistic surrogate data. As discussed in “*Is the correct pairwise model bimodal?*”, inclusion of third- or fourth-order correlations does not seem to cure this problem.
3. Interesting distributions that may be found as initial approximations of pairwise distributions (e.g. Fig 2B, red) may turn out to be incorrect owing to the first and second problems above. They have to be discarded for methodological reasons despite their interesting properties.

We have given an intuitive explanation of the common cause of these issues: positive pairwise correlations imply positive Lagrange multipliers between pairs of neuron, corresponding to a symmetric network that is excitatory on average. For typical values of correlations

observed in neuroscientific experiments, this network can therefore possess two metastable dynamic regimes, given sufficiently many units. The mechanism is identical to the ferromagnetic transition in the Ising model, as explained in “*Bimodality of the inhomogeneous model for large N* ”. An analogous bimodality appears in the statistical mechanics of finite-size systems [e.g. 108, 115, and refs therein]—but it is experimentally expected and verified there, unlike our neurobiological case.

Although we did not study maximum-entropy models typically used in other fields, like structural biology and genetic networks [116–118], social behavior in mammals [119, 120], natural image statistics [121, 122], and economics [123], the problems we have addressed are generic and emerge as soon as we study a large network with positive pairwise correlations on average; hence they might be of relevance to these fields.

In this work we have also suggested a remedy, based on the explanation above: the intuitive idea is to add a minimal asymmetric inhibition to the network, in the guise of an additional, asymmetrically coupled inhibitory neuron (Fig 8A) [cf. 34, p. 175]. This leads to an “inhibited” Glauber dynamics that is free from bistable regimes and has a unimodal stationary distribution $P_i(\mathbf{s})$, Eq (22). This dynamics depends on an inhibition-coupling parameter J_I and a threshold parameter θ .

Most important, we have shown that this new stationary distribution $P_i(\mathbf{s})$ belongs to the *maximum-entropy family*: it can be obtained with the maximum-relative-entropy method with respect to a reference measure, Eq (25) (Fig 9), that represents the neurobiologically natural presence of inhibition in the network. We call this model an “inhibited” pairwise maximum-entropy model.

The inhibited pairwise model solves all three problems above:

1. It can be found by Boltzmann learning with standard sampling times. In the present work we have not investigated whether analytic approximations like the Thouless-Anderson-Palmer or Sessak-Monasson ones can be adjusted to be applied to this model; but see point 3. below.
2. Its distribution does not have unrealistic modes at high average activities. The model allows us to decide how much any high-activity modes should be suppressed (parameter J_I) and the activity above which such modes are neurobiologically unrealistic (parameter θ).
3. It yields distributions similar to the interesting ones that should otherwise be abandoned on methodological grounds (wrong standard pairwise distributions). In this regard, if interesting distributions found in the literature turn out to be incorrect owing to undersampling, they could be “rescued” if reinterpreted as distributions of the inhibited model; see “*Detection and further study of bimodality*” below.

Detection and further study of bimodality

We wish to stress that the presence of bimodality and non-ergodicity can easily go unnoticed. Sampling from a bimodal distribution, the probability to switch to the second mode may be so small that it occurs over more sampling steps larger than those typically used in the literature, and the high mode is not visited during Boltzmann learning or surrogate generation. We then face a subtle situation: The obtained distribution is *not* a pairwise maximum-entropy distribution Eq (3)—the Lagrange multipliers are incorrect—yet a consistency check (also affected by undersampling) may wrongly seem to validate it, and also analytic approximations (outside of their convergence domain) may wrongly validate it.

The distribution found in this circumstance is not a standard pairwise distribution, but our *inhibited* maximum entropy distribution Eq (22), for appropriately chosen J_I and θ .

In this regard we urge researchers who have calculated pairwise (and even higher-order) maximum-entropy distributions for more than 50 neurons using short Boltzmann-learning procedures, to check for the possible presence of higher metastable regimes. The presence of bimodality and non-ergodicity can be checked, for example, by starting the sampling from different initial conditions, at low and high activities, looking out for bistable regimes [cf. 62, S 2.1.3]. Another way out of this problem is to use other sampling techniques or Markov chains different from the Glauber one [61, 62, 97, 98]. Alternatively, one may use the inhibited model Eq (22) with the standard approaches.

In the presence of inhomogeneous and randomly chosen parameters and large network sizes, the standard pairwise maximum-entropy distribution is mathematically identical with the Boltzmann distribution of the Sherrington & Kirkpatrick infinite-range spin glass [124, 125]. A more systematic analysis of the effect of inhomogeneity on the appearance of the second mode could therefore employ methods developed for spin glasses [126], which could produce approximate expressions for the inverse problem: the determination of Lagrange multipliers from the data. One may think of modifying the Thouless-Anderson-Palmer (TAP) mean-field approach [67, 127], generalizations of which exist for the asymmetric non-equilibrium case [93] appearing here due to the inhibitory unit. An appropriate modification of the ideas of Sessak and Monasson [68, 69] could also be an alternative. Another possibility is the use of cumulant expansions [17, 128], which unlike TAP-based approaches have the advantage of being valid also in regimes of strong coupling; recent extensions allow us to obtain the statistics at the level of individual units [129].

Bimodality in other models

In this work we have not investigated other models, like general linear models or kinetic Ising models for example. Considering the fundamental mechanism by which the bimodality arises, we expect similar problems in other models. The reasoning backing this hypothesis is this: Pairwise correlations in cortical areas are on average positive but very weak. In this limit we expect that these correlations require slightly positive “excitatory” couplings between units in most other models; an independent-pair approximation also suggests this [127]. As a result of this rough approximation determined at the level of individual pairs, we expect the couplings to be independent of the number of units of a dynamic or statistical model. With increasing number of units in the model the overall “excitatory feedback” $\sum_j J_{ij}$ will increase, and a simple mean-field analysis makes us expect the appearance of a second mode at a certain critical number, what in statistical mechanics is called a ferromagnetic transition; cf. Fig 7B. We expect similar ferromagnetic transitions to happen in a wide class of statistical models that only represent the observed, on average positively correlated units. Similar transitions are also reported in Bohte et al. [34] for a biological—as opposed to statistical—neuron model composed of excitatory neurons only. In fact, they had to introduce one inhibitory neuron in their model to avoid such transitions, which is also the idea behind our inhibitory term.

The bimodality problem could be cured by allowing for asymmetric connections, enabling the implementation of possibly hidden inhibitory units that stabilize the activity. For example, kinetic Ising models [130–132], which are maximum-entropy models over the possible histories of network activity [133–135], can have positive correlations among excitatory units in the asynchronous irregular regime, while their dynamics is stabilized by inhibitory feedback [see e.g. 136, Fig 3A]. Scaling of network properties with the number of units N is often studied in this context. In the asynchronous regime, mean pairwise correlations decrease as N^{-1} [18, 22, 110, 136]. This scaling is the result of a fictive experiment, typically used to derive a theoretical results in the $N \rightarrow \infty$ limit—any biological neuronal network has of course a certain fixed size

N . The mean correlation measured in a sample of size M , with $1 \ll M \leq N$, is by sampling theory expected to be roughly equal to the mean correlation of the full network, and does not vary much with M ; only the variance around this expectation declines to 0 as M approaches N .

Meaning and advantages of the inhibited model and its reference measure

The inhibited maximum-entropy model P_i , Eq (22), solves the problems discussed above; but we may ask if this is enough to motivate its use. We consider it an interesting model for at least two reasons. First, it actually is a class of models rather than a single specific model. In the present work we have focused on its use with pairwise constraints because these are still widely discussed in the literature. But the inhibition reference measure Eq (25) can be used with higher-order constraints or other kinds of constraints as well. We leave to future works the analysis of this possibility. Second, there are neurobiological reasons why the reference measure Eq (25) can be methodologically more appropriate than the uniform measure of the standard maximum-entropy method. Let us argue this point in more depth.

Standard (i.e. uniform reference measure) maximum-entropy distributions are often recommended as “maximally noncommittal” [137]. But this adjective needs qualification. Jaynes precised: “maximally noncommittal” by a certain criterion’—that the possible events or states be deemed to have a priori equal probabilities before any constraints are enforced [31]. When the initial probabilities are not deemed equal, for physical or biological reasons for example, reference measures appear. An important example of reference measure is the “density of states” that multiplies the Boltzmann factor $e^{-E/(kT)}$ in statistical mechanics [e.g. 138, ch. 16]: we cannot judge energy levels to be a priori equally probable because each one comprises a different amount of degrees of freedom. The proper choice of this reference measure is so essential as to be the first manifest difference between classical and quantum statistical mechanics, from “classical counting” to “quantum counting” of phase-space cells [138, ch. 16]. Owing to quantized energy exchanges, a quantum density of states is necessary in statistical mechanics; likewise we could say that owing to inhibitory feedback an inhibitory reference measure is necessary in the statistical mechanics of neuronal networks. The uniform reference measure of standard maximum-entropy expresses that network units have a priori equally probable $\{0, 1\}$ states. But these units are neurons, whose states are not a priori equally likely. The measure of the inhibited model P_i reflects this a priori asymmetry in a simplified way. There are surely other reference measures that reflect this asymmetry in a more elaborated way, but the one we have found is likely one of the simplest; cf. Bohte et al.’s [34] inhibitory solution.

The choice of an appropriate reference measure is critically important in neuroscientific inferences also for another reason. When maximum-entropy is used to generate an initial distribution *to be updated by Bayes’s theorem*, the choice of reference measure is not critical, because a poor choice gets anyway updated and corrected as new data accumulate. Not so when maximum-entropy is used to generate a sort of reference distribution that *will not be updated*, as is often done in neuroscience: an unnaturally chosen reference measure will then bias and taint all conclusions derived from comparisons with the maximum-entropy distribution.

The inhibited pairwise model can therefore be quite useful in all applications of the maximum-entropy model mentioned in “Introduction”. For example, it can serve as a realistic hypothesis against which to check or measure the prominence of correlations in simulated or recorded neural activities, to separate the low baseline level of correlation from the potentially behaviourally relevant departures thereof. The surprise measure to effect such separation would, according to the inhibited model, take into account the presence of inhibition and the

overall low level of activity that are natural in the cortex. The inhibited model can also be used for the generation of surrogate data which include the natural effect of inhibition besides the observed level of pairwise activity. It can also be useful in the study of the predictive sufficiency of pairwise correlations as opposed to higher-order moments, for example for distribution tails [e.g. 34–36, 38–44]; and in the characterization of dynamical regimes of neuronal activity [36, 49–51].

Choice of inhibition parameters

The inhibition reference measure Eq (25) contains the threshold θ and the inhibitory coupling J_I as parameters. The choice of their values depends on the point of view adopted about the measure. Three venues seem possible: (1) One might think of choosing (θ, J_I) to better fit the specific dataset under study, but this would counter the maximum-entropy spirit: the threshold cannot be a constraint, and the inhibitory coupling would acquire infinite values, as explained in section “*Inhibited maximum-entropy model*”. Moreover for our dataset this strategy would only give a worse fit (cf. Fig 2B) because the inhibition term flattens the distribution tails. (2) One might only want to get rid of the bistability of the Glauber dynamics and the bimodality of the distribution. In this case the precise choice of (θ, J_I) is not critical within certain bounds. The inhibition coupling $J_I < 0$ must be negative and sufficiently large to suppress activity once the population-averaged activity reaches θ . The self-consistency condition Eq (21) then gives $[1 + \exp(\sum_j^{j \neq i} J_{ij} m_j + h_i + J_I)]^{-1} \ll \theta$ for all i . The threshold θ can be safely set to any value between the highest observed population activity \bar{s} and the second fixed point of the self-consistency equation Eq (21), which is indicative of the second mode and is beyond $\bar{s} > 1/2$ (see Fig 7B) for the typically low mean activities observed in the cortex. (3) A methodologically sounder possibility, in view of the remarks about maximum-entropy measures given above, is to choose (θ, J_I) from general neurobiological arguments and observations. This was implicitly done in Bohte & al.’s neuron model [34] for example, but unfortunately they did not publish the values they chose. We leave the discussion of the neurobiological choice of these parameters to future investigations.

Relations to other work

Our inhibition term $J_I N G(\bar{s} - \theta)$, Eq (22), formally includes Shimazaki et al.’s “simultaneous silence” constraint [44] as the limit $J_I \rightarrow -\infty, \theta = 1/N$. Because of this limit their model has a sharp jump in probability at $\bar{s} = 1/N$: their constraint uniformly removes probability for $\bar{s} > 1/N$ and assigns it to the single point $\bar{s} = 0$. In contrast, our inhibited model P_i presents a kink but no jump for $\bar{s} = \theta$, with a discontinuity in the derivative proportional to J_I . But besides this mathematical relationship, our inhibition term and the “simultaneous silence” constraint have different motivations and uses. As discussed at length above and in section “*Inhibited maximum-entropy model*”, our term is best interpreted as a reference measure expressing the effects of inhibition, providing a biologically more suitable starting point [cf. 34] for maximum-entropy, rather than a constraint. Its goal is not to improve the goodness-of-fit for activities well below threshold, in contrast to earlier works [e.g. 35, 40, 50, 78, 80] and to the “simultaneous silence” constraint [44]. The goodness-of-fit is determined by the constraints alone. In this regard we do not present any improvement of the fit compared to a pure pairwise model. Future work could explore combinations of the here proposed reference measure and additional constraints that improve the fitness of the model.

Materials and methods

Range of applicability of maximum-entropy models

Maximum-entropy models are an approximate limit case of probability models by *exchangeability* [139–141], or *sufficiency* [141, 142, §§4.2–5]. This approximation holds if the constraints are empirical averages (e.g. time averages in our case) over enough many data compared with the number of points in the sample space. How much is “enough” depends on where the empirical averages lie within their physically allowed ranges: If they are well within their ranges, then a number of data values large but still smaller than the number of sample-space points may be enough. If the empirical averages are close or equal to their physically allowed extreme values, then the number of data values should be much larger than the number of sample-space points. If these conditions are not met the maximum-entropy method gives unreasonable or plainly wrong results, as can be ascertained by comparison with the non-approximated Bayesian model. Simple examples of these limitations are illustrated in [140, 141] together with the more reasonable predictions of the non-approximated Bayesian models [see also 61, p. 308].

A very large positive or negative Lagrange multiplier usually signals that the maximum-entropy method is inadequate, because the constraint corresponding to the multiplier is approaching its minimal or maximal allowed values. Consider our case, discussed in section “*The problem: Bimodality, bistability, non-ergodicity*”. The constraints are time-averages over roughly 300000 data points, and the sample space—the possible network states—has $2^N = 2^{159} \approx 7 \times 10^{47}$ points. Suppose that we want to use as constraints the $N + 1$ observed frequencies of the total activity $N\bar{s}$ [cf. 50, 94, 113]. Each frequency is bounded between 0 and 1. In our data the values $N\bar{s} = 24$ and $N\bar{s} = 28$ have non-zero frequencies, but the intermediate values $N\bar{s} \in \{25, 26, 27\}$ have zero frequencies—the minimum possible value. The Lagrange multipliers for the latter three frequencies would be $-\infty$. The maximum-entropy model would therefore predict that *it is possible for the network to have 24 or 28 simultaneously active neurons, but impossible for it to have 25, 26, or 27 active neurons*—not even in future recordings, if we interpret the model that way. Such a prediction is unreasonable, not to say a little silly. Under the assumption that each neuron is as likely as not to be active in each time bin, the probability that in 300000 time bins we observe all possible values of the total activity $N\bar{s} \in \{0, \dots, 159\}$ —each at least once—is of the order 10^{-1463} . This means that it is practically certain that some values of $N\bar{s}$ will not appear in our recording; not because of physical impossibility, but because of the exceedingly small number of observations compared with that of possible events. It is unreasonable to think that the three values 25, 26, 27 could not appear in a longer recording, yet the values 24 and 28 could. As signalled by the large value of the Lagrange multipliers, the conditions for the validity of the maximum-entropy limit are not satisfied in this case, and the method breaks down. The validity of the inhomogeneous pairwise model is similarly questionable if there are neuron pairs with zero coupled activity, $g_{ij} = 0$; some corrections to the method are necessary in that case.

The limitations of the maximum-entropy method are well-known [143] in the field of image reconstruction of astronomical sources, where this method has probably most successfully been applied for the longest time. In this field the maximum-entropy principle is today used differently: to generate a distribution on the space of prior distributions, rather than a prior itself [144, 145].

Glauber dynamics

Here we show that there is a temporal process that is able to sample from the the distribution $P_p(\mathbf{s}|\mathbf{h}, \mathbf{J})$ Eq (3). This temporal dynamics is called *Glauber dynamics*. It is an example of a

Markov chain on the space of binary neurons $\{0, 1\}^N$ [63]. At each time step a neuron s_i is chosen randomly and updated with the update rule

$$s_i \leftarrow 1 \text{ with probability } F_i(s) = g\left(\sum_{k \neq i} J_{ik}s_k + h_i\right) \text{ and } 0 \text{ else} \quad (30)$$

$$g(x) = \frac{1}{1 + \exp(-x)}, \quad (31)$$

where the coupling is assumed to be symmetric, $J_{ij} = J_{ji}$, and self-coupling is absent, $J_{ii} = 0$. The transition operator of the Markov chain, κ , only connects states that differ by at most one neuron, so for the transition of neuron i we can write, if $s^{i+} = (s_1, \dots, \underbrace{1}_{i\text{-th}}, \dots, s_N)$ and

$$s^{i-} = (s_1, \dots, \underbrace{0}_{i\text{-th}}, \dots, s_N),$$

$$\begin{aligned} \kappa(s^{i+}|s^{i-}) &= F_i(s^{i-}) \\ \kappa(s^{i-}|s^{i+}) &= 1 - F_i(s^{i+}). \end{aligned} \quad (32)$$

The pairwise maximum-entropy distribution $P_p(\mathbf{s}|\mathbf{h}, \mathbf{J})$ is stationary under the Markov dynamics above. The proof can be obtained as the $J_I = 0$ case of the proof, given below, for the inhibited pairwise maximum-entropy model.

Inhibited Glauber dynamics and its stationary maximum-entropy distribution

Inhibited Glauber dynamics. In the “inhibited” Glauber dynamics, the network of N neurons with states $s_i(t)$ has an additional neuron with state $s_I(t)$. The dynamics is determined by the following algorithm starting at time step t with states $\mathbf{s} = \mathbf{s}(t)$, $s_I = s_I(t)$:

1. One of the N units is chosen, each unit having probability $1/N$ of being the chosen one. Suppose i is the selected unit.
2. The chosen unit i is updated to the state $s'_i := s_i(t + 1)$ with probability

$$p(s'_i|\mathbf{s}, s_I) = (1 + \exp[(1 - 2s'_i)F_i(\mathbf{s}, s_I)])^{-1}$$

$$= \begin{cases} [1 + e^{F_i(\mathbf{s}, s_I)}]^{-1}, & \text{for } s'_i = 0, \\ [1 + e^{-F_i(\mathbf{s}, s_I)}]^{-1}, & \text{for } s'_i = 1, \end{cases}$$

$$\text{with } F_i(\mathbf{s}, s_I) := h_i + \sum_{k \neq i} J_{ik}s_k + J_I s_I.$$

Note the additional coupling from the neuron s_I , with strength J_I . This strength can have any sign, but we are interested in the $J_I \leq 0$ case; we therefore call s_I the “inhibitory neuron”.

3. The inhibitory neuron is deterministically updated to the state $s'_I := s_I(t + 1)$ given by

$$s'_I = H\left(\sum_k s_k/N - \theta\right), \quad (33)$$

corresponding to a Kronecker-delta conditional probability

$$p(s'_I|\mathbf{s}, s_I) = p(s'_I|\mathbf{s}) = \delta[s'_I - H(\sum_k s_k/N - \theta)]. \quad (34)$$

In other words, the inhibitory neuron becomes active if the population-averaged activity of the other neurons is equal to or exceeds the threshold θ .

4. The time is stepped forward, $t + 1 \rightarrow t$, and the process repeats from step 1.

The original Glauber dynamics, described in the previous section, is recovered when $J_I = 0$, which corresponds to decoupling the inhibitory neuron s_I .

The total transition probability can be written as

$$p(\mathbf{s}', s'_I | \mathbf{s}, s_I) = \frac{1}{N} \delta \left[s'_I - H \left(\sum_k s_k / N - \theta \right) \right] \times \sum_i [(1 + \exp[(1 - 2s'_i)F_i(\mathbf{s}, s_I)])^{-1} \prod_{k \neq i} \delta(s'_k - s_k)]; \quad (35)$$

the product of Kronecker deltas in the last term ensures that at most one of the N neurons changes state at each timestep.

The transition probabilities for the chosen neuron s_i and the inhibitory neuron s_I are independent, conditional on the state of the network at the previous timestep:

$$p(\mathbf{s}', s'_I | \mathbf{s}, s_I) = p(\mathbf{s}' | \mathbf{s}) p(s'_I | \mathbf{s}),$$

so the transition probability for the N neurons only can be written as

$$p(\mathbf{s}' | \mathbf{s}) = \frac{1}{N} \sum_i \left[(1 + \exp[(1 - 2s'_i)F_i(\mathbf{s})])^{-1} \prod_{k \neq i} \delta(s'_k - s_k) \right], \quad (36)$$

$$\text{with } F_i(\mathbf{s}) := h_i + \sum_{k \neq i} J_{ik} s_k + J_I H(\sum_k s_k / N - \theta). \quad (37)$$

This formula also shows that the transition probability for the network can alternatively be derived without explicitly introducing an inhibitory unit: starting from the modified activation function

$$p(s_i | s) = \{1 + \exp[(1 - 2s_i)F_i(s)]\}^{-1}$$

with F_i defined by Eq (39), the transition probability Eq (36) for the network follows from the additional requirement that only a single unit changes state within a single update.

Proof that the inhibited maximum-entropy model is the stationary distribution of the inhibited Glauber dynamics. The modified maximum-entropy distribution P_i , Eq (22), is the stationary distribution of a slightly modified version of the above dynamics, with the update rule

$$s'_I = H \left(\sum_{k \neq I} s_k / N - \theta \right), \quad (38)$$

and the use of N inhibitory neurons, one for each of the original N units. This dynamics has a slightly different transition probability, with activation function

$$F_i(\mathbf{s}) := h_i + \sum_{k \neq i} J_{ik} s_k + J_I H(\sum_{k \neq i} s_k / N - \theta) \quad (39)$$

instead of Eq (37). Note that the two dynamics are very similar for large enough N . To prove

the stationarity of inhibited maximum-entropy distribution P_i , we show that P_i satisfies the detailed-balance equality

$$p(\mathbf{s}'|\mathbf{s})P_i(\mathbf{s}) = p(\mathbf{s}|\mathbf{s}')P_i(\mathbf{s}') \quad \text{or} \quad \frac{p(\mathbf{s}'|\mathbf{s})}{p(\mathbf{s}|\mathbf{s}')} = \frac{P_i(\mathbf{s}')}{P_i(\mathbf{s})}, \quad \forall \mathbf{s}, \mathbf{s}', \tag{40}$$

which is a sufficient condition for stationarity [146–148].

First note that if \mathbf{s}' and \mathbf{s} differ in the state of more than one neuron, the transition probability $p(\mathbf{s}'|\mathbf{s})$ vanishes and the detailed-balance above is trivially satisfied. Also the case $\mathbf{s}' = \mathbf{s}$ is trivially satisfied. Only the case in which \mathbf{s}' and \mathbf{s} differ in the state of one unit, say s_i , remains to be proven. Assume then that

$$s'_i = 1, \quad s_i = 0, \quad \forall k \neq i, \quad s'_k = s_k; \tag{41}$$

by symmetry, if the detailed balance is satisfied in the case above it will also be satisfied with the values 0 and 1 interchanged.

Substituting the transition probability Eqs (36) and (39) in the left-hand side of the fraction form of the detailed balance Eq (40), and noting that $F_i(\mathbf{s}') = F_i(\mathbf{s})$, we have

$$\begin{aligned} \frac{p(\mathbf{s}'|\mathbf{s})}{p(\mathbf{s}|\mathbf{s}')} &= \exp[-F_i(\mathbf{s})]^{-1} \\ &= \exp \left[h_i + \sum_k^{k \neq i} J_{ik} s_k + J_i H \left(\sum_k^{k \neq i} s_k / N - \theta \right) \right]. \end{aligned} \tag{42}$$

Using the expression for the inhibited model P_i , Eq (22), in the right-hand side of the fraction form of the detailed balance Eq (40), we have

$$\begin{aligned} \frac{P(\mathbf{s}')}{P(\mathbf{s})} &= \exp \left\{ \left[\sum_{k \neq i} h_k s_k + h_i \cdot 1 \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \sum_{k, m \neq i}^{k \neq m} J_{mk} s_m s_k + \sum_{k \neq i} J_{mi} s_m \cdot 1 \right. \right. \\ &\quad \left. \left. + J_i N G \left(\sum_k^{k \neq i} \frac{s_k}{N} + \frac{1}{N} - \theta \right) \right] \right. \\ &\quad \left. - \left[\sum_{k \neq i} h_k s_k + h_i \cdot 0 \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \sum_{k, m \neq i}^{k \neq m} J_{mk} s_m s_k + \sum_{k \neq i} J_{mi} s_m \cdot 0 \right. \right. \\ &\quad \left. \left. + J_i N G \left(\sum_k^{k \neq i} \frac{s_k}{N} + \frac{0}{N} - \theta \right) \right] \right\} \\ &= \exp \left[h_i + \sum_k^{k \neq i} J_{ik} s_k + J_i H \left(\sum_k^{k \neq i} s_k / N - \theta \right) \right], \end{aligned} \tag{43}$$

where we have used the equality $NG(x + 1/N) - NG(x) = H(x)$, valid if $x = \sum_k^{k \neq i} s_k / N - \theta$ and

$N\theta \in \mathbf{Z}$. Comparison of formulae Eqs (42) and (43) finally proves that the detailed balance is satisfied also in the case Eq (41).

Simulation of Glauber dynamics with NEST

The neuron model `ginzburg_neuron` in NEST, a simulator for neural network models [96], implements the Glauber dynamics, if the parameters of the gain function are chosen appropriately. The gain function has the form

$$g_{\text{ginzburg}}(h) = c_1 h + \frac{c_2}{2} (1 + \tanh(c_3(h - \theta))). \tag{44}$$

With $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, setting $x = c_3(h - \theta)$, $c_1 = 0$, $c_2 = 1$, $c_3 = \frac{1}{2}$ it takes the form

$$\begin{aligned} g_{\text{ginzburg}}(h) &= \frac{1}{2} \frac{e^x + e^{-x} + e^x - e^{-x}}{e^x + e^{-x}}, \\ &= \frac{1}{1 + e^{-2x}} = \frac{1}{1 + e^{-(h-\theta)}}, \end{aligned} \tag{45}$$

which is identical to Eq (31).

Bimodality of the inhomogeneous model for large N

The large N limit for the inhomogeneous pairwise model can be studied employing results from spin glass theory [125]. The first point to realize is that for weak correlations the Lagrange multipliers J_{ij} are to dominant order determined only by the covariances between units i and j and by their respective mean activities. This follows from eq. (7) of Roudi et al. 2009 [127], which we expand in the limit of weak correlations (and hence only to linear order in J_{ij}) as

$$\begin{aligned} &\text{Roudi et al. eq. 7} \\ &\downarrow \\ -J_{ij} + \mathcal{O}(J_{ij}^2) &= [C^{-1}]_{ij} = [\{m_k(1 - m_k)\delta_{kl} + c_{k \neq l}\}]_{ij}^{-1} \\ &= \left[\{m_k(1 - m_k)\delta_{kl}\}^{-\frac{1}{2}} \left(\{\delta_{kl} + \frac{c_{k \neq l}}{\sqrt{m_k(1 - m_k)m_l(1 - m_l)}}\} \right)^{-1}_{ij} \{m_l(1 - m_l)\delta_{kl}\}^{-\frac{1}{2}} \right]_{i \neq j} \\ &\simeq -\frac{c_{ij}}{m_k(1 - m_k)m_l(1 - m_l)} + \mathcal{O}(c_{ij}^2), \end{aligned}$$

where we used the geometric series from the second to the third line. Since considering larger networks will not change the statistics of the c_{ij} (as long as we are within the local network of $N \simeq 10^3 - 10^4$ neurons), the Lagrange multipliers J_{ij} will, to leading order, follow the same distribution. In particular their population mean $\bar{J}_{ij} = \frac{1}{N(N-1)} \sum_{ij} J_{ij} \stackrel{N \gg 1}{\simeq} \mu$ and their variance $\delta J_{ij}^2 = \frac{1}{N(N-1)} \sum_{ij} (J_{ij} - \bar{J}_{ij})^2 \stackrel{N \gg 1}{\simeq} \sigma^2$ converge to values μ and σ^2 that are, to leading order, independent of N .

We now consider the “energy” associated with the maximum-entropy model

$$E(s) = -\frac{1}{2} \sum_{ij} J_{ij} s_i s_j - \sum_i h_i s_i.$$

For this expression to possess a well-defined $N \rightarrow \infty$ limit, we need (see [125], eqs. 1.3a and 1.3b) that $\mu = \tilde{J}_0/N$ and $\sigma^2 = \tilde{J}^2/N$, with N -independent quantities denoted by a tilde. We may therefore determine at which point we are in the phase diagram, shown in Fig 1 of

[125]. So we obtain the scaling relations

$$\tilde{J}_0 = N\mu.$$

$$\tilde{J} = \sqrt{N}\sigma.$$

We may now study what happens if we increase N . We therefore investigate how, for given and N -independent values of μ and σ , we move through the phase diagram of the model (see Fig 1 in [125]). The axes of this diagram are spanned by

$$\frac{\tilde{J}_0}{\tilde{J}} = \sqrt{N} \frac{\mu}{\sigma},$$

$$\frac{kT}{\tilde{J}} = \frac{1}{\sqrt{N}\sigma}.$$

So increasing N we will move to the lower right in the phase diagram, ultimately crossing the transition to ferromagnetic behaviour. This is the point at which the model becomes bistable. One may note that the position of this cross-over is not entirely correctly predicted by the replica-symmetric theory of [125]. The true solution, found by Parisi [149] is slightly displaced compared to the transition line in the diagram in Fig 1 of [125]. Still, as we are only interested in the limit $N \rightarrow \infty$, the result is the same and the model becomes bistable.

Expansion of the inhibition term in terms of higher-order coupled activities

Higher-order correlations are represented by products of K distinct activities, like e.g. $s_1 s_3 s_4 s_9$, with $K \in \{0 \dots, N\}$, whose expectations are the raw K -th moments of the distribution. There are $\binom{N}{K}$ such products for each given K . For a network activity $(s_1, \dots, s_N) \in \{0, 1\}^N$, each of those products amounts to either 0 or 1. More precisely, if the total activity is S , then $\binom{S}{K}$ of these products will equal 1 and the others will vanish; the binomial vanishes by definition if $K > S$, so it covers this case as well.

In the reduced, homogeneous case we can meaningfully sum together all products with K factors, because they have the same probability. Then, from what we said above, such sum equals $\binom{S}{K}$ when the total activity is S :

$$\sum_{i_1 < i_2 < \dots < i_K} s_{i_1} s_{i_2} \dots s_{i_K} = \binom{S}{K} \quad \text{if } \sum_i s_i = S. \tag{46}$$

We want to rewrite the logarithm of the inhibition term $NG(\bar{s} - \theta) := N(\bar{s} - \theta)H(\bar{s} - \theta)$ as a sum of such sums of K products, in order to interpret it as a combination of higher-order correlations:

$$NG(\bar{s} - \theta) = \sum_{K=0}^N f_K \left(\sum_{i_1 < i_2 < \dots < i_K} s_{i_1} s_{i_2} \dots s_{i_K} \right) = \sum_{K=0}^N f_K \binom{N\bar{s}}{K}, \tag{47}$$

with θ -dependent coefficients f_K . Let us find them.

Rewrite $NG(\bar{s} - \theta)$ as $G(S - \Theta)$, with $S := N\bar{s} \in \{0, \dots, N\}$ and $\Theta := N\theta$. The $(N + 1)$ -tuple v of numbers

$$v := (G(0 - \Theta), G(1 - \Theta), \dots, G(N - \Theta)) = (0, \dots, \underbrace{0}_{\Theta\text{th}}, 1, \dots, N - \Theta)$$

is a θ -dependent row-vector.

Expression Eq (47) can be interpreted as the matrix product fP of the row vector f —which we want to find—by the $(N + 1)$ -dimensional matrix P having element $\binom{S}{k}$ in its $(K + 1)$ th row and $(S + 1)$ th column. Such matrix is called a *Pascal* matrix [150, 151]; for example, for $N = 4$,

$$(P)_K^S = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Hence we have $v = fP$, and therefore $f = vP^{-1}$. The inverse P^{-1} of a Pascal matrix has elements $(P^{-1})_S^K = (-1)^{K-S} \binom{K}{S}$ [150, 151]; for example, for $N = 4$,

$$(P^{-1})_S^K = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 & -4 \\ 0 & 0 & 1 & -3 & 6 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

By multiplying the expressions of v and P^{-1} above we find that the row-vector $f = vP^{-1}$ is, explicating its dependence on Θ ,

$$f_k(\Theta) = \left\{ \begin{array}{cccccccc} (0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & \Theta = 0 \\ (0 & 0 & 1 & -1 & 1 & -1 & 1 & \dots & \Theta = 1 \\ (0 & 0 & 0 & 1 & -2 & 3 & -4 & \dots & \Theta = 2 \\ (0 & 0 & 0 & 0 & 1 & -3 & 6 & \dots & \dots \\ (0 & 0 & 0 & 0 & 0 & 1 & -4 & \dots & \\ \dots & & & & & & & & \\ K = 0 & K = 1 & K = 2 & \dots & & & & & \end{array} \right.$$

This solution has a convenient feature: if we increase N by 1, the matrix $(f_k(\Theta))$ of the N -dimensional solution acquires one new row and column, but the already existing entries remain unchanged.

We can thus write $G(S - \theta)$ as $\sum_{K=0}^N f_K(\theta) \binom{s}{K}$, with $f_K(\theta)$ given above. But $f_K(\theta) = 0$ if $K \leq \theta$, so we can also restrict the sum to $K > \theta$:

$$G(S - \theta) = \sum_{K=\theta+1}^N f_K(\theta) \binom{s}{K}. \tag{48}$$

Compare the matrix of values for $f_K(\theta)$ above with that for the generalized binomial coefficient [100, 112]:

$$\left(\begin{array}{c} -\theta \\ K - \theta - 1 \end{array} \right) = (-1)^{K-\theta-1} \binom{K-2}{K-\theta-1} = \left\{ \begin{array}{cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & \dots & \theta = 0 \\ & 1 & -1 & 1 & -1 & 1 & \dots & \theta = 1 \\ & & 1 & -2 & 3 & -4 & \dots & \theta = 2 \\ & & & 1 & -3 & 6 & \dots & \dots ; \\ & & & & 1 & -4 & \dots & \\ & & & & & & \dots & \\ K = 1 & K = 2 & \dots & & & & & \end{array} \right.$$

if $K > \theta$ we have $f_K(\theta) = \binom{-\theta}{K-\theta-1}$. We can therefore write Eq (48) more explicitly, recalling that $S \equiv N\bar{s}$, $\theta \equiv N\theta$, $G(S - \theta) \equiv NG(\bar{s} - \theta)$, and Eq (46), as:

$$\begin{aligned} NG(\bar{s} - \theta) &= \sum_{K=N\theta+1}^N \binom{-N\theta}{K - N\theta - 1} \binom{N\bar{s}}{K}, \\ &\equiv \sum_{K=N\theta+1}^N \binom{-N\theta}{K - N\theta - 1} \left(\sum_{i_1 < i_2 < \dots < i_K} s_{i_1} s_{i_2} \dots s_{i_K} \right), \end{aligned}$$

which is formula Eq (23).

Acknowledgments

We are grateful to Alexa Riehle and Thomas Brochier for providing the experimental data. PGLPM thanks the Forschungszentrum librarians for their always prompt help in finding arcane scientific works, Miri & Mari for encouragement and affection, Buster for filling life with awe and inspiration, and the developers and maintainers of LATEX, Emacs, AUCTEX, Python, Inkscape, Open Science Framework, bioRxiv, HAL, PhilSci, viXra, arXiv, Sci-Hub for making a free scientific exchange possible.

Author Contributions

Conceptualization: Vahid Rostami, PierGianLuca Porta Mana, Sonja Grün, Moritz Helias.

Data curation: Vahid Rostami.

Formal analysis: Vahid Rostami.

Funding acquisition: Sonja Grün, Moritz Helias.

Investigation: Vahid Rostami, PierGianLuca Porta Mana, Moritz Helias.

Methodology: Vahid Rostami, PierGianLuca Porta Mana.

Project administration: Sonja Grün, Moritz Helias.

Software: Vahid Rostami, PierGianLuca Porta Mana.

Supervision: PierGianLuca Porta Mana, Moritz Helias.

Validation: Vahid Rostami, PierGianLuca Porta Mana.

Visualization: Vahid Rostami.

Writing – original draft: Vahid Rostami, PierGianLuca Porta Mana, Moritz Helias.

Writing – review & editing: Vahid Rostami, PierGianLuca Porta Mana, Sonja Grün, Moritz Helias.

References

1. Perkel Donald H., Gerstein George L., and Moore George P. Neuronal spike trains and stochastic point processes: I. The single spike train. *Biophys. J.*, 7(4):391–418, 1967. <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.elsevier-2cce69bb-de3e-3481-95ea-9fa7927b1426/c/main.pdf>. See also [2]. [https://doi.org/10.1016/S0006-3495\(67\)86596-2](https://doi.org/10.1016/S0006-3495(67)86596-2)
2. Perkel Donald H., Gerstein George L., and Moore George P. Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophys. J.*, 7(4):419–440, 1967. <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.elsevier-3eda9334-4835-34f4-a2d4-8b17f5e8e281/c/main.pdf>. See also [1].
3. Gerstein George L. and Perkel Donald H. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science*, 881(164):828–830, may 16th 1969. <https://doi.org/10.1126/science.164.3881.828>
4. von der Malsburg Christoph The correlation theory of brain function. Internal report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, 1981.
5. Bienenstock Elie A model of neocortex. *Network: Comp. Neural Sys.*, 6:179–224, 1995. https://doi.org/10.1088/0954-898X_6_2_004
6. Singer W Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24(1):49–65, Sep 1999. [https://doi.org/10.1016/S0896-6273\(00\)80821-1](https://doi.org/10.1016/S0896-6273(00)80821-1) PMID: 10677026
7. Gray C. M. The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24:31–47, 1999. [https://doi.org/10.1016/S0896-6273\(00\)80820-X](https://doi.org/10.1016/S0896-6273(00)80820-X) PMID: 10677025
8. Salinas Emilio and Sejnowski Terrence J. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.*, 2(8):539–550, 2001. <https://doi.org/10.1038/35086012> PMID: 11483997
9. Riehle Alexa, Grün Sonja, Diesmann Markus, and Aertsen Ad Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, 278:1950–1953, 1997. <https://doi.org/10.1126/science.278.5345.1950> PMID: 9395398
10. Kilavik B. E., Roux S., Ponce-Alvarez A., Confais J., Gruen S., and Riehle A. Long-term modifications in motor cortical dynamics induced by intensive practice. *Journal of Neuroscience*, 29:12653–12663, 2009. <https://doi.org/10.1523/JNEUROSCI.1554-09.2009> PMID: 19812340
11. Maldonado Pedro, Babul Cecilia, Singer Wolf, Rodriguez Eugenio, Berger Denise, and Grün Sonja Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *Journal of Neurophysiology*, 100(3):1523–1532, 2008. <https://doi.org/10.1152/jn.00076.2008> PMID: 18562559
12. Ito J, Maldonado P, Singer W, and Grün S Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cereb Cortex*, 21(11):2482–2497, November 2011. <https://doi.org/10.1093/cercor/bhr020> PMID: 21459839
13. Schultze-Kraft Matthias, Diesmann Markus, Gruen Sonja, and Helias Moritz Noise suppression and surplus synchrony by coincidence detection. *PLoS Comput Biol*, 9(4):e1002904, 2013. <https://doi.org/10.1371/journal.pcbi.1002904> PMID: 23592953

14. DeWeese Michael R. and Zador Anthony M. Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *Journal of Neuroscience*, 26(47):12206–12218, 2006. <https://doi.org/10.1523/JNEUROSCI.2813-06.2006> PMID: 17122045
15. Teramae Jun-nosuke and Fukai Tomoki Complex evolution of spike patterns during burst propagation through feed-forward networks. *Biol. Cybern.*, 99(2):105–114, 2008. <https://doi.org/10.1007/s00422-008-0246-9> PMID: 18685860
16. Shadlen Michael N. and Newsome William T. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998. PMID: 9570816
17. Ginzburg Iris and Sompolinsky Haim Theory of correlations in stochastic neural networks. *Phys. Rev. E*, 50(4):3171–3191, 1994. <http://neurophysics.huji.ac.il/node/466>; <http://papers.cnl.salk.edu/PDFs/Theory%20of%20Correlations%20in%20Stochastic%20Neural%20Networks%201994-3835.pdf>. <https://doi.org/10.1103/PhysRevE.50.3171>
18. Renart Alfonso, De La Rocha Jaime, Bartho Peter, Hollender Liad, Parga N'estor, Reyes Alex, and Harris Kenneth D. The asynchronous state in cortical circuits. *Science*, 327:587–590, January 2010. <https://doi.org/10.1126/science.1179850> PMID: 20110507
19. Pernice Volker, Staude Benjamin, Cardanobile Stefano, and Rotter Stefan How structure determines correlations in neuronal networks. *PLoS Comput. Biol.*, 7(5):e1002059, May 2011. <https://doi.org/10.1371/journal.pcbi.1002059> PMID: 21625580
20. Pernice V and Rotter, S. Reconstruction of connectivity in sparse neural networks from spike train covariances. *Front. Comput. Neurosci. Conference Abstract: Bernstein Conference 2012*, 2012.
21. Trousdale J, Hu Y, Shea-Brown E, and Josic K Impact of network structure and cellular response on spike time correlations. *PLoS Comput. Biol.*, 8(3):e1002408, 2012. <https://doi.org/10.1371/journal.pcbi.1002408> PMID: 22457608
22. Tetzlaff Tom, Helias Moritz, Einevoll Gaute, and Diesmann Markus Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.*, 8(8):e1002596, 2012. <https://doi.org/10.1371/journal.pcbi.1002596> PMID: 23133368
23. Helias Moritz, Tetzlaff Tom, and Diesmann Markus Echoes in correlated neural systems. *New J Phys.*, 15:023002, 2013. <https://doi.org/10.1088/1367-2630/15/2/023002>
24. Jovanović Stojan, Hertz John, and Rotter Stefan Cumulants of Hawkes point processes. *Physical Review E*, 91(4):042802, apr 2015. ISSN 1539-3755. URL <http://link.aps.org/doi/10.1103/PhysRevE.91.042802>.
25. Jovanović Stojan and Rotter Stefan Interplay between Graph Topology and Correlations of Third Order in Spiking Neuronal Networks. *PLOS Computational Biology*, 12(6):e1004963, jun 2016. ISSN 1553-7358. URL <http://dx.plos.org/10.1371/journal.pcbi.1004963>. PMID: 27271768
26. De la Rocha Jaime, Doiron Brent, Shea-Brown Eric, Kresimir Josic, and Reyes Alex Correlation between neural spike trains increases with firing rate. *Nature*, 448(16):802–807, august 2007. <https://doi.org/10.1038/nature06028> PMID: 17700699
27. Shea-Brown Eric, Josic Kresimir, de la Rocha Jaime, and Doiron Brent Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.*, 100:108102, March 2008. <https://doi.org/10.1103/PhysRevLett.100.108102> PMID: 18352234
28. Diaconis Persi. Sufficiency as statistical symmetry. In: Felix E. Browder, editor. *Mathematics into the Twenty-first Century: 1988 Centennial Symposium August 8–12* (American Mathematical Society, Providence, USA, 1992), pages 15–26. 1992. First publ. 1991 as technical report <https://statistics.stanford.edu/research/sufficiency-statistical-symmetry>.
29. Dawid A. Philip Exchangeability and its ramifications. In: Damien Paul, Dellaportas Petros, Polson Nicholas G., and Stephens David A., editors. *Bayesian Theory and Applications* (Oxford University Press, Oxford, 2013), chapter 2, pages 19–29. 2013.
30. Freedman David A. Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Stat.*, 33(3):916–923, 1962. <https://doi.org/10.1214/aoms/1177704460>
31. Jaynes Edwin Thompson. Information theory and statistical mechanics. In: Ford K. W., editor. *Statistical Physics* (Benjamin, New York, 1963), pages 181–218. 1963. Repr. in: Edwin Thompson Jaynes. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics* (Kluwer, Dordrecht, reprint edition, 1989, first publ. 1983), Ed. by R. D. Rosenkrantz, ch. 4, 39–76. <http://bayes.wustl.edu/etj/node1.html>.
32. Sivia D. S. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford, 2 edition, 2006. Written with J. Skilling. First publ. 1996.

33. Latham Peter E. and Roudi Yasser Role of correlations in population coding. In: Quiroga Rodrigo Quian and Panzeri Stefano, editors. *Principles of Neural Coding* (CRC Press, Boca Raton, USA, 2013), chapter 7, pages 121–138. 2013. arXiv:1109.6524.
34. Bohte S. M., Spekreijse H., and Roelfsema P. R. The effects of pair-wise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neural Comp.*, 12(1):153–179, 2000. <https://doi.org/10.1162/089976600300015934>
35. Schneidman Elad, Berry Michael J. II, Segev Ronen, and Bialek William Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006. arXiv:q-bio/0512013, http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html. <https://doi.org/10.1038/nature04701> PMID: 16625187
36. Tkačik Gašper, Schneidman Elad, Berry, II, Michael J. and Bialek William. Ising models for networks of real neurons, 2006. arXiv:q-bio/0611072.
37. Cohen Marlene R. and Kohn Adam Measuring and interpreting neuronal correlations. *Nat. Neurosci.*, 14(7):811–819, 2011. <http://marlenecohen.com/pubs/CohenKohn2011.pdf>. <https://doi.org/10.1038/nn.2842> PMID: 21709677
38. Granot-Atedgi Einat, Tkačik Gašper, Segev Ronen, and Schneidman Elad Stimulus-dependent maximum entropy models of neural population codes. *PLoS Computational Biology*, 9(3):e1002922, 2013. arXiv:1205.6438. <https://doi.org/10.1371/journal.pcbi.1002922> PMID: 23516339
39. Martignon L., Von Hasein H., Grün Sonja, Aertsen Ad, and Palm Günther Detecting higher-order interactions among the spiking events in a group of neurons. *Biol. Cybern.*, 73(1):69–81, 1995. <https://doi.org/10.1007/BF00199057> PMID: 7654851
40. Shlens Jonathon, D. Field Greg, Gauthier Jeffrey L., Grivich Matthew I., Petrusca Dumitru, Sher Alexander, Litke Alan M., and Chichilnisky E. J. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.*, 26(32):8254–8266, 2006. See also correction: Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Matthew I. Grivich, Dumitru Petrusca, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. Correction, the structure of multi-neuron firing patterns in primate retina. *J. Neurosci.*, 28(5):1246, 2008. <https://doi.org/10.1523/JNEUROSCI.1282-06.2006> PMID: 16899720
41. Macke Jakob H., Opper Manfred, and Bethge Matthias The effect of pairwise neural correlations on global population statistics. Technical Report 183, Max-Planck-Institut für biologische Kybernetik, Tübingen, 2009a. http://www.kyb.tuebingen.mpg.de/publications/attachments/MPIK-TR-183_%5B0%5D.pdf.
42. Barreiro Andrea K., Gjorgjieva Julijana, Fred M. Rieke, and Shea-Brown Eric T. When are microcircuits well-modeled by maximum entropy methods? arXiv:1011.2797. See also: Andrea K. Barreiro, Eric T. Shea-Brown, Fred M. Rieke, and Julijana Gjorgjieva. When are microcircuits well-modeled by maximum entropy methods? *BMC Neurosci.*, 11(Suppl. 1):P65, 2010.
43. Ganmor Elad, Segev Ronen, and Schneidman Elad Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proc. Natl. Acad. Sci. (USA)*, 108(23):9679–9684, 2011. http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html. <https://doi.org/10.1073/pnas.1019641108>
44. Shimazaki Hideaki, Sadeghi Kolia, Ishikawa Tomoe, Ikegaya Yuji, and Toyozumi Taro Simultaneous silence organizes structured higher-order interactions in neural populations. *Sci. Rep.*, 5:9821, 2015. <https://doi.org/10.1038/srep09821> PMID: 25919985
45. Roudi Yasser, Tyrcha Joanna, and Hertz John Ising model for neural data: Model quality and approximate methods for extracting functional connectivity. *Phys. Rev. E*, 79(5):051915, 2009a. arXiv:0902.2885. <https://doi.org/10.1103/PhysRevE.79.051915>
46. Gerwin Sebastian, H. Macke Jakob, and Bethge Matthias Bayesian inference for generalized linear models for spiking neurons. *Front. Comput. Neurosci.*, 4:12, 2010. <https://doi.org/10.3389/fncom.2010.00012> PMID: 20577627
47. Macke Jakob H., Buesing Lars, Cunningham John P., Yu Byron M., Shenoy Krishna V., and Sahani Maneesh Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems (NIPS proceedings)*, 24:1350–1358, 2011a.
48. Macke Jakob H., Murray Iain, and Latham Peter E. Estimation bias in maximum entropy models. *Entropy*, 15(8):3109–3129, 2013. <http://www.gatsby.ucl.ac.uk/~pel/papers/maxentbias.pdf>. <https://doi.org/10.3390/e15083109>
49. Tkačik Gašper, Schneidman Elad, Berry, II, Michael J. and Bialek William. Spin glass models for a network of real neurons, 2009. arXiv:0912.5409.
50. Tkačik Gašper, Mora Thierry, Marre Olivier, Amodei Dario, Palmer Stephanie E., Berry Michael J. II, and Bialek William Thermodynamics and signatures of criticality in a network of neurons. *Proc. Natl. Acad. Sci. (USA)*, 112(37):11508–11513, 2014a. arXiv:1407.5946.

51. Mora Thierry, Deny Stéphane, and Marre Olivier Dynamical criticality in the collective activity of a population of retinal neurons. *Phys. Rev. Lett.*, 114(7):078105, 2015. arXiv:1410.6769. <https://doi.org/10.1103/PhysRevLett.114.078105> PMID: 25763977
52. Bartlett M. S. The statistical significance of odd bits of information. *Biometrika*, 39(3–4):228–237, 1952. <https://doi.org/10.1093/biomet/39.3-4.228>
53. Good Irving John The appropriate mathematical tools for describing and measuring uncertainty. In: Good Irving John *Good Thinking: The Foundations of Probability and Its Applications* (University of Minnesota Press, Minneapolis, USA, 1983), chapter 16, pages 173–177. 1957. First publ. 1957.
54. Mead Lawrence R. and Papanicolaou N. Maximum entropy in the problem of moments. *J. Math. Phys.*, 25(8):2404–2417, 1984. <http://bayes.wustl.edu/Manual/MeadPapanicolaou.pdf>. <https://doi.org/10.1063/1.526446>
55. Fang Shu-Cherng, Rajasekera J. R., and Tsao H.-S. J. *Entropy Optimization and Mathematical Programming*. Springer, New York, reprint edition, 1997.
56. Press William H., Teukolsky Saul A., T. Vetterling William, and Flannery Brian P. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 3 edition, 2007. First publ. 1988.
57. Nicolelis L M A, editor. *Methods for Neural Ensemble Recordings*. CRC Press, Boca Raton, Florida, 1998.
58. Buzsáki György Large-scale recording of neuronal ensembles. *Nat. Neurosci.*, 7(5):446–451, May 2004. <https://doi.org/10.1038/nn1233> PMID: 15114356
59. Berényi Antal, Somogyvári Zoltán, Nagy Anett J., Roux Lisa, Long John D., Fujisawa Shigeyoshi, Stark Eran, Leonardo Anthony, Harris Timothy D., and Buzsáki György Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J. Neurophysiol.*, 111(5):1132–1149, 2014. <http://www.buzsakilab.com/content/PDFs/Berenyi2013.pdf>. <https://doi.org/10.1152/jn.00785.2013> PMID: 24353300
60. Riehle Alexa, Wirtsohn Sarah, Grün Sonja, and Brochier Thomas Mapping the spatio-temporal structure of motor cortical lfp and spiking activities during reach-to-grasp movements. *Frontiers in Neural Circuits*, 7:48, 2013. <https://doi.org/10.3389/fncir.2013.00048> PMID: 23543888
61. MacKay David J. C. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, 2003. <http://www.inference.phy.cam.ac.uk/mackay/itila/>. First publ. 1995.
62. Landau David P. and Binder Kurt *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 4 edition, 2015. http://el.us.edu.pl/ekonofizyka/images/6/6b/A_guide_to_monte_carlo_simulations_in_statistical_physics.pdf, http://iop.vast.ac.vn/~nvthanh/cours/simulation/MC_book.pdf. First publ. 2000.
63. Glauber Roy J. Time-dependent statistics of the Ising model. *J. Math. Phys.*, 4(2):294–307, 1963. <https://doi.org/10.1063/1.1703954>
64. Ackley David H., Hinton Geoffrey E., and Sejnowski Terrence J. A learning algorithm for Boltzmann machines. *Cognit. Sci.*, 9(1):147–169, 1985. https://doi.org/10.1207/s15516709cog0901_7
65. Rayne Hartree Douglas The wave mechanics of an atom with a non-coulomb central field. Part I. Theory and methods. *Proc. Cambridge Philos. Soc.*, 24(1):89–110, 1928a. <http://sci-prew.inf.ua/index1.htm>. See also: Douglas Rayne Hartree. The wave mechanics of an atom with a non-coulomb central field. Part II. Some results and discussion. *Proc. Cambridge Philos. Soc.*, 24(1):111–132, 1928b. <http://sci-prew.inf.ua/index1.htm>. <https://doi.org/10.1017/S0305004100011919>
66. Opper Manfred and Saad David, editors. *Advanced Mean Field Methods: Theory and Practice*. MIT Press, Cambridge, USA, 2001.
67. Thouless D. J., W. Anderson P., and Palmer R. G. Solution of ‘Solvable model of a spin glass’. *Phil. Mag.*, 35(3):593–601, 1977. <https://doi.org/10.1080/14786437708235992>
68. Sessak Vitor and Monasson Rémi Small-correlation expansions for the inverse Ising problem. *J. Phys. A*, 42(5):055001, 2009. arXiv:0811.3574. <https://doi.org/10.1088/1751-8113/42/5/055001>
69. Sessak Vitor. *Inverse problems in spin models*. PhD thesis, École Normale Supérieure, Paris, 2010. arXiv:1010.1960.
70. Grün Sonja Data-driven significance estimation for precise spike correlation. *J. Neurophysiol.*, 101(3):1126–1140, 2009. <https://doi.org/10.1152/jn.00093.2008> PMID: 19129298
71. Macke Jakob H., Berens Philipp, Ecker Alexander S., Tolias Andreas S., and Bethge Matthias Generating spike trains with specified correlation coefficients. *Neural Comp.*, 21(2):397–423, 2009b. <https://doi.org/10.1162/neco.2008.02-08-713>
72. Fujisawa Shigeyoshi, Amarasingham Asohan, Matthew T. Harrison, and Buzsáki György Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.*, 11(7):823–833, 2008. <https://doi.org/10.1038/nn.2134> PMID: 18516033

73. Grün S., Riehle A., and Diesmann M. Effect of cross-trial nonstationarity on joint-spike events. *Biol. Cybern.*, 88(5):335–351, 2003. <https://doi.org/10.1007/s00422-002-0386-2> PMID: 12750896
74. Pipa G. and Grün S. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, 52–54:31–37, 2003. [https://doi.org/10.1016/S0925-2312\(02\)00823-8](https://doi.org/10.1016/S0925-2312(02)00823-8)
75. Pipa Gordon, Diesmann Markus, and Grün Sonja Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods. *Complexity*, 8(4):79–86, 2003. <https://doi.org/10.1002/cplx.10085>
76. Gerstein G. L. Searching for significance in spatio-temporal firing patterns. *Acta Neurobiol. Exp.*, 64:203–207, 2004.
77. Louis S., Gerstein G. L., Grün S., and Diesmann M. Surrogate spike train generation through dithering in operational time. *Front. Comput. Neurosci.*, 4(127), 2010. <https://doi.org/10.3389/fncom.2010.00127> PMID: 21060802
78. Roudi Yasser, Nirenberg Sheila, and Peter E. Latham Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't. *PLoS Computational Biology*, 5(5):e1000380, 2009b. arXiv:0811.0903. <https://doi.org/10.1371/journal.pcbi.1000380>
79. Amari Shun-ichi, Nakahara Hiroyuki, Wu Si, and Sakai Yutaka Synchronous firing and higher-order interactions in neuron pool. *Neural Comp.*, 15(1):127–142, 2003. <https://doi.org/10.1162/089976603321043720>
80. Macke Jakob H., Opper Manfred, and Bethge Matthias Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Phys. Rev. Lett.*, 106(20):208102, 2011b. arXiv:1009.2855. <https://doi.org/10.1103/PhysRevLett.106.208102>
81. Hobson Arthur A new theorem of information theory. *J. Stat. Phys.*, 1(3):383–391, 1969. <https://doi.org/10.1007/BF01106578>
82. Kullback Solomon The Kullback-Leibler distance. *American Statistician*, 41(4):340–341, 1987.
83. Hobson Arthur and Cheng Bin-Kang A comparison of the Shannon and Kullback information measures. *J. Stat. Phys.*, 7(4):301–310, 1973. <https://doi.org/10.1007/BF01014906>
84. Garrett Anthony J. M. Maximum entropy with nonlinear constraints: physical examples. In: Fougère Paul F., editor. *Maximum Entropy and Bayesian Methods: Dartmouth, U.S.A., 1989* (Kluwer, Dordrecht, 1990), pages 243–249. 1990.
85. Gerstein George L. Analysis of firing patterns in single neurons. *Science*, 131(3416):1811–1812, 1960. <https://doi.org/10.1126/science.131.3416.1811> PMID: 17753210
86. Gerstein George L. and Kiang Nelson Y.-S. An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophys. J.*, 1(1):15–28, 1960. http://research.meei.harvard.edu/eplhistory/Kiang_papers_CD_contents/1960_Gerstein_Kiang.pdf. [https://doi.org/10.1016/S0006-3495\(60\)86872-5](https://doi.org/10.1016/S0006-3495(60)86872-5) PMID: 13704760
87. Potts R. B. Note on the factorial moments of standard distributions. *Aust. J. Phys.*, 6(4):498–499, 1953. <https://doi.org/10.1071/PH530498>
88. Challa Murty S. S. and Hetherington J. H. Gaussian ensemble as an interpolating ensemble. *Phys. Rev. Lett.*, 60(2):77–80, 1988. <https://doi.org/10.1103/PhysRevLett.60.77> PMID: 10038203
89. Johal Ramandeep S., Planes Antoni, and Vives Eduard Statistical mechanics in the extended Gaussian ensemble. *Phys. Rev. E*, 68(5):056113, 2003. arXiv:cond-mat/0307646. <https://doi.org/10.1103/PhysRevE.68.056113>
90. Costeniuc M., Ellis R. S., Touchette H., and Turkington B. Generalized canonical ensembles and ensemble equivalence. *Phys. Rev. E*, 73(2):026105, 2006. arXiv:cond-mat/0505218. See also: M. Costeniuc, R. S. Ellis, Hugo Touchette, and B. Turkington. The generalized canonical ensemble and its universal equivalence with the microcanonical ensemble. *J. Stat. Phys.*, 119(5–6):1283–1329, 2005. arXiv:cond-mat/0408681. <https://doi.org/10.1103/PhysRevE.73.026105>
91. Arieli Amos, Sterkin Alexander, Grinvald Amiram, and Aertsen Ad Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871, 1996. http://material.brainworks.uni-freiburg.de/publications-brainworks/1996/journal%20papers/Arieli_Science96.pdf. <https://doi.org/10.1126/science.273.5283.1868> PMID: 8791593
92. Knuth Kevin H. Optimal data-based binning for histograms, 2013. arXiv:physics/0605197. First publ. 2006.
93. Roudi Yasser and Hertz John Mean field theory for nonequilibrium network reconstruction. *Phys. Rev. Lett.*, 106(4):048702, 2011. arXiv:1009.5946. <https://doi.org/10.1103/PhysRevLett.106.048702> PMID: 21405370
94. Tkačik Gašper, Marre Olivier, Amodei Dario, Schneidman Elad, Bialek William, and Berry Michael J. II. Searching for collective behavior in a large network of sensory neurons. *PLoS Computational Biology*, 10(1):e1003408, 2014b. arXiv:1306.3061. <https://doi.org/10.1371/journal.pcbi.1003408>

95. Tamara Broderick, Miroslav Dudik, Gašper Tkačik, Robert E. Schapire, and William Bialek. Faster solutions of the inverse pairwise Ising problem, 2007. arXiv:0712.2437.
96. Gewaltig Marc-Oliver and Diesmann Markus NEST (NEural Simulation Tool). *Scholarpedia*, 2(4):1430, 2007. <https://doi.org/10.4249/scholarpedia.1430>
97. Binder Kurt, editor. *Applications of the Monte Carlo Method in Statistical Physics*. Springer, Berlin, 2 edition, 1987. First publ. 1984.
98. Binder Kurt Applications of Monte Carlo methods to statistical physics. *Rep. Prog. Phys.*, 60(5):487–559, 1997. http://fisica.ciencias.uchile.cl/~gonzalo/cursos/SimulacionII/rpphys_binder97.pdf. <https://doi.org/10.1088/0034-4885/60/5/001>
99. Binder Kurt and Landau D. P. Finite-size scaling at first-order phase transitions. *Phys. Rev. B*, 30(3):1477–1485, 1984. <https://doi.org/10.1103/PhysRevB.30.1477>
100. Oldham Keit B., Myland Jan C., Jerome Spanier *An Atlas of Functions: with Equator, the Atlas Function Calculator*. Springer, New York, 2 edition, 2009. First publ. 1987.
101. Bair Wyeth, Zohary Ehud, and Newsome William T. Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J. Neurosci.*, 21(5):1676–1697, 2001. http://invice.net/biblio_database_dyva/woda/data/att/fc5e.file.pdf. PMID: 11222658
102. Mazurek Mark E. and Shadlen Michael N. Limits to the temporal fidelity of cortical spike rate signals. *Nat. Neurosci.*, 5(5):463–471, 2002. https://www.shadlenlab.columbia.edu/publications/publications/mike/mazurek_shadlen2002.pdf. PMID: 11976706
103. Kohn Adam and Smith Matthew A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.*, 25(14):3661–3673, 2005. <http://www.smithlab.net/publications.html>. <https://doi.org/10.1523/JNEUROSCI.5106-04.2005> PMID: 15814797
104. Smith Matthew A. and Kohn Adam Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.*, 28(48):12591–12603, 2008. <http://www.smithlab.net/publications.html>. <https://doi.org/10.1523/JNEUROSCI.2929-08.2008> PMID: 19036953
105. Bakhurin Konstantin I., Mac Victor, Peyman Golshani, and Sotiris C. Masmanidis Temporal correlations among functionally specialized striatal neural ensembles in reward-conditioned mice. *J. Neurophysiol.*, 115(3):1521–1532, 2016. <https://doi.org/10.1152/jn.01037.2015> PMID: 26763779
106. Labastie Pierre and Whetten Robert L. Statistical thermodynamics of the cluster solid-liquid transition. *Phys. Rev. Lett.*, 65(13):1567–1570, 1990. <http://cluster.physik.uni-freiburg.de/lehre/SS11/ueb11/Labastie1990.pdf>. <https://doi.org/10.1103/PhysRevLett.65.1567> PMID: 10042303
107. Chomaz Ph., Gulminelli Francesca, and Dufloy V. Topology of event distributions as a generalized definition of phase transitions in finite systems. *Phys. Rev. E*, 64(4):046114, 2001. arXiv:cond-mat/0010365. <https://doi.org/10.1103/PhysRevE.64.046114>
108. Gulminelli Francesca and Chomaz Ph. Failure of thermodynamics near a phase transition. *Phys. Rev. E*, 66(4):046108, 2002. <https://doi.org/10.1103/PhysRevE.66.046108>
109. Binney J. J., Dowrick N. J., Fischer A. J., and Newman M. E. J. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford University Press, Oxford, 2001. First publ. 1992.
110. van Vreeswijk Carl and Sompolinsky Haim Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726, December, 1996. <https://doi.org/10.1126/science.274.5293.1724> PMID: 8939866
111. Amit Daniel J. and Brunel Nicolas Dynamics of a recurrent network of spiking neurons before and following learning. *Network: Comput. Neural Syst.*, 8:373–404, 1997. https://doi.org/10.1088/0954-898X_8_4_003
112. Fowler David The binomial coefficient function. *Am. Math. Monthly*, 103(1):1–17, 1996. <https://doi.org/10.2307/2975209>
113. Tkačik Gašper, Marre Olivier, Mora Thierry, Amodei Dario, Berry Michael J. II, and Bialek William The simplest maximum entropy model for collective behavior in a neural network. *J. Stat. Mech.*, 2013: P03011, 2013. arXiv:1207.6319. <https://doi.org/10.1088/1742-5468/2013/03/P03011>
114. Hinton G. E. and Sejnowski T. J. Learning and relearning in Boltzmann machines. In: Rumelhart David E., McClelland James L., and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. (MIT Press, Cambridge, USA, 12th printing edition, 1999, first publ. 1986), chapter 7, pages 282–317, 507–516. 1999. First publ. 1986; <https://papers.cnl.salk.edu/PDFs/Learning%20and%20Relearning%20in%20Boltzmann%20Machines%201986-3239.pdf>.
115. Berry R. Stephen, Beck Thomas L., Davis Heidi L., and Jellinek Julius Solid-liquid phase behavior in microclusters. *Adv. Chem. Phys.*, 70(2):75–138, 1988.
116. Lezon Timothy R, Banavar Jayanth R, Cieplak Marek, Maritan Amos, and Fedoroff Nina V Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns.

- Proceedings of the National Academy of Sciences of the United States of America*, 103(50):19033–8, dec 2006. ISSN 0027-8424. URL <http://www.ncbi.nlm.nih.gov/pubmed/17138668> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1748172>. <https://doi.org/10.1073/pnas.0609152103> PMID: 17138668
117. Seno Flavio, Trovato Antonio, Banavar Jayanth R., and Maritan Amos Maximum entropy approach for deducing amino acid interactions in proteins. *Phys. Rev. Lett.*, 100:078102, Feb 2008. URL <https://link.aps.org/doi/10.1103/PhysRevLett.100.078102>. PMID: 18352600
 118. Mora Thierry, Walczak Aleksandra M, Bialek William, and Callan Curtis G Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(12):5405–10, mar 2010. ISSN 1091-6490. URL <http://www.ncbi.nlm.nih.gov/pubmed/20212159> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2851784>. <https://doi.org/10.1073/pnas.1001705107> PMID: 20212159
 119. Castellano Claudio, Fortunato Santo, and Loreto Vittorio Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, May 2009. URL <https://link.aps.org/doi/10.1103/RevModPhys.81.591>.
 120. Bialek William, Cavagna Andrea, Giardina Irene, Mora Thierry, Silvestri Edmondo, Viale Massimiliano, and Walczak Aleksandra M Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13):4786–4791, mar 2012. ISSN 1091-6490. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324025&tool=pmcentrez&rendertype=abstract>. <https://doi.org/10.1073/pnas.1118633109> PMID: 22427355
 121. Stephens Greg J., Mora Thierry, Tkačik Ga šper, and Bialek William Statistical thermodynamics of natural images. *Phys. Rev. Lett.*, 110:018701, Jan 2013. URL <https://link.aps.org/doi/10.1103/PhysRevLett.110.018701>. PMID: 23383852
 122. Saremi Saeed and Sejnowski Terrence J Hierarchical model of natural images and the origin of scale invariance. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8):3071–6, feb 2013. ISSN 1091-6490. URL <http://www.ncbi.nlm.nih.gov/pubmed/23382241> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3581899>. <https://doi.org/10.1073/pnas.1222618110> PMID: 23382241
 123. Sornette Didier Physics and financial economics (1776–2014): puzzles, Ising and agent-based models. *Reports on Progress in Physics*, 77(6):062001, jun 2014. ISSN 0034-4885. URL <http://stacks.iop.org/0034-4885/77/i=6/a=062001?key=crossref.a64c242d6daf6f9d12c71625a8e3d4d3>. <https://doi.org/10.1088/0034-4885/77/6/062001> PMID: 24875470
 124. Sherrington David and Kirkpatrick Scott Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35(26):1792–1796, 1975. <https://doi.org/10.1103/PhysRevLett.35.1792>
 125. Kirkpatrick Scott and Sherrington David Infinite-ranged models of spin-glasses. *Phys. Rev. B*, 17(11):4384–4403, 1978. <https://doi.org/10.1103/PhysRevB.17.4384>
 126. Fischer K. H. and Hertz J. A. *Spin glasses*. Cambridge University Press, Cambridge, reprint edition, 1993. First publ. 1991.
 127. Roudi Yasser, Aurell Erik, and Hertz John A. Statistical physics of pairwise probability models. *Front. Comput. Neurosci.*, 3:22, 2009c. arXiv:0905.1410. <https://doi.org/10.3389/neuro.10.022.2009>
 128. Buice Michael A., Cowan Jack D., and Chow Carson C. Systematic fluctuation expansion for neural network activity equations. *Neural Comp.*, 22(2):377–426, 2010. arXiv:0902.3925. <https://doi.org/10.1162/neco.2009.02-09-960>
 129. Dahmen David, Bos Hannah, and Helias Moritz Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium. *Physical Review X*, 6(3):031024, aug 2016. ISSN 2160-3308. URL <http://link.aps.org/doi/10.1103/PhysRevX.6.031024>.
 130. Mézard Marc and Sakellariou J. Exact mean-field inference in asymmetric kinetic Ising systems. *J. Stat. Mech.*, 2011:L07001, 2011. arXiv:1103.3433.
 131. Zeng Hong-Li, Aurell Erik, Alava Mikko, and Mahmoudi Hamed Network inference using asynchronously updated kinetic Ising model. *Phys. Rev. E*, 83(4):041135, 2011. arXiv:1011.6216. <https://doi.org/10.1103/PhysRevE.83.041135>
 132. Sakellariou Jason, Roudi Yasser, Mezard Marc, and Hertz John Effect of coupling asymmetry on mean-field solutions of the direct and inverse Sherrington-Kirkpatrick model. *Phil. Mag.*, 92(1–3):272–279, 2012. arXiv:1106.0452. <https://doi.org/10.1080/14786435.2011.634857>
 133. Jaynes Edwin Thompson. Macroscopic prediction. In: Haken Hermann, editor. *Complex Systems—Operational Approaches: in Neurobiology, Physics, and Computers* (Springer, Berlin, 1985), pages 254–269. 1985. Updated 1996 version at <http://bayes.wustl.edu/etj/node1.html>.
 134. Ge Hao, Pressé Steve, Ghosh Kingshuk, and Dill Ken A. Markov processes follow from the principle of maximum caliber. *J. Chem. Phys.*, 136(6):064108, 2012. arXiv:1106.4212. <https://doi.org/10.1063/1.3681941> PMID: 22360170

135. Lee Julian and Pressé Steve A derivation of the master equation from path entropy maximization. *J. Chem. Phys.*, 137(7):074103, 2012. <https://doi.org/10.1063/1.4743955> PMID: 22920099
136. Helias Moritz, Tetzlaff Tom, and Diesmann Markus The correlation structure of local cortical networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.*, 10(1):e1003428, 2014. <https://doi.org/10.1371/journal.pcbi.1003428> PMID: 24453955
137. Jaynes Edwin Thompson. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630, 1957a. <http://bayes.wustl.edu/etj/node1.html>, see also: Edwin Thompson Jaynes. Information theory and statistical mechanics. II. *Phys. Rev.*, 108(2):171–190, 1957b. <http://bayes.wustl.edu/etj/node1.html>. <https://doi.org/10.1103/PhysRev.106.620>
138. Callen Herbert B. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, New York, 2 edition, 1985. First publ. 1960.
139. Jaynes Edwin Thompson. Monkeys, kangaroos, and N , 1996. <http://bayes.wustl.edu/etj/node1.html>. First publ. 1986. (Errata: in equations (29)–(31), (33), (40), (44), (49) the commas should be replaced by gamma functions, and on p. 19 the value 0.915 should be replaced by 0.0915).
140. Porta Mana P.G.L. On the relation between plausibility logic and the maximum-entropy principle: a numerical study, 2009. arXiv:0911.2197. Presented as invited talk at the 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering “MaxEnt 2011”, Waterloo, Canada.
141. Porta Mana P.G.L. Maximum-entropy from the probability calculus: exchangeability, sufficiency, 2017. arXiv:1706.02561.
142. Bernardo José-Miguel and Smith Adrian F. *Bayesian Theory*. Wiley series in probability and mathematical statistics. Wiley, New York, reprint edition, 2000. First publ. 1994.
143. Weir Nicholas. Applications of maximum entropy techniques to HST data. In: P. J. Grosbøl and R. H. Warmels, editors. *Third ESO/ST-EFC Data Analysis Workshop*. (European Southern Observatory, Garching, 1991), pages 115–129. 1991.
144. Skilling John Classic maximum entropy. In: Skilling John, editor. *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. (Kluwer, Dordrecht, 1989), pages 45–52. 1989.
145. Carlos C. Rodríguez. Entropic priors, 1991. <http://omega.albany.edu:8008/>.
146. Kelly F. P. *Reversibility and Stochastic Networks*. Wiley, Chichester, 1979. http://www.statslab.cam.ac.uk/~frank/BOOKS/kelly_book.html.
147. van N. G.Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 3 edition, 2007. First publ. 1981.
148. Gardiner Crispin W. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer, Berlin, 3 edition, 2004. First publ. 1983.
149. Parisi Giorgio The order parameter for spin glasses: a function on the interval 0–1. *J. Phys. A*, 13(3):1101–1112, 1980. <https://doi.org/10.1088/0305-4470/13/3/042>
150. Call Gregory S. and Velleman Daniel J. Pascal’s matrices. *Am. Math. Monthly*, 100(4):372–376, 1993. <https://doi.org/10.2307/2324960>
151. Edelman Alan and Strang Gilbert Pascal matrices. *Am. Math. Monthly*, 111(3):189–197, 2004. <http://www-math.mit.edu/~gs/papers/pascal-work.pdf>. <https://doi.org/10.2307/4145127>