

This is the accepted manuscript made available via CHORUS. The article has been published as:

Ensemble inhibition and excitation in the human cortex: An Ising-model analysis with uncertainties

Cristian Zanoci, Nima Dehghani, and Max Tegmark

Phys. Rev. E **99**, 032408 — Published 7 March 2019

DOI: [10.1103/PhysRevE.99.032408](https://doi.org/10.1103/PhysRevE.99.032408)

Ensemble Inhibition and Excitation in the Human Cortex: an Ising Model Analysis with Uncertainties

Cristian Zanoci,^{*} Nima Dehghani,[†] and Max Tegmark[‡]

*Department of Physics, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA and*

*Center for Brains, Minds and Machines, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA*

(Dated: January 31, 2019)

The pairwise maximum entropy model, also known as the Ising model, has been widely used to analyze the collective activity of neurons. However, controversy persists in the literature about seemingly inconsistent findings, whose significance is unclear due to lack of reliable error estimates. We therefore develop a method for accurately estimating parameter uncertainty based on random walks in parameter space using adaptive Markov Chain Monte Carlo after the convergence of the main optimization algorithm. We apply our method to the activity patterns of excitatory and inhibitory neurons recorded with multielectrode arrays in the human temporal cortex during the wake-sleep cycle. Our analysis shows that the Ising model captures neuronal collective behavior much better than the independent model during wakefulness, light sleep, and deep sleep when both excitatory (E) and inhibitory (I) neurons are modeled; ignoring the inhibitory effects of I-neurons dramatically overestimates synchrony among E-neurons. Furthermore, information-theoretic measures reveal that the Ising model explains about 80% – 95% of the correlations, depending on sleep state and neuron type. Thermodynamic measures show signatures of criticality, although we take this with a grain of salt as it may be merely a reflection of long-range neural correlations.

I. INTRODUCTION

One of the main questions in neuroscience is how to accurately model the dynamics of networks of firing neurons. This question involves controversies not only about the correct dynamics, but also about the most relevant degrees of freedom in the network. One important example is the issue of temporal versus rate coding, i.e., whether the precise time of spiking or only the overall spiking rates matter in the description of network dynamics. A number of theoretical [1–3] and experimental [4–8] studies provide evidence for the importance of the exact timing of spikes, but this view has been challenged by alternative perspectives advocating for less time-constrained and more probabilistic models [9–13]. Attempts at answering this central question have been hampered both by experimental difficulties in acquiring adequate data and computational challenges related to the exponential (in system size) growth of the number of dependencies that a network model has to capture. Recent advances in experimental techniques using multielectrode arrays [14–17] enable us to simultaneously record the activity of large populations of neurons, further amplifying the need to formulate an effective theory describing the macroscopic characteristics of large neuronal networks given their numerous degrees of freedom. Statistical mechanics provides many examples of such theories that relate the macroscopic properties of matter to the interactions between its microscopic degrees of freedom.

One such family of models, known as maximum entropy models, consists of the least structured probability distributions that are consistent with a set of empirical statistics on finite data. In a landmark study, it was shown that pairwise maximum entropy models, also known as Ising models, based on average spiking probability and correlations between pairs of neurons, give a good description of the activity patterns in retinal ganglion cells [11]. Since then, these models have been widely used to describe the activity of ensembles of neurons in a variety of systems, both *in vitro* and *in vivo* [7, 18–25].

Despite Ising models’ success in describing the statistics of activity patterns, they also have certain limitations. First, it has been argued that higher-order neuron couplings could play an important role in population coding, so that pairwise couplings fail to capture the full dynamics [26–34], especially if exact spike timing is important [35, 36]. Second, the model’s reliability may be distance-dependent, leading to successful predictions for neurons separated by large distances, but poor fits to the activity of local clusters of neurons that might be dominated by high-order correlations due to distance dependent connectivity profiles [29, 37, 38]. Finally, Ising models may not be scalable to the full size of real neuronal networks [39–41].

Although pairwise maximum entropy models have known limitations and have been extensively studied before, they remain one of the few simple models that can explain the main characteristics of collective behavior. As new data sets for increasingly larger neuronal populations become available, it is imperative to rigorously test the applicability and predictive power of Ising models on these data sets. Unfortunately, without any approximations, the computational cost of making predictions us-

^{*} czanoci@mit.edu

[†] nima.dehghani@mit.edu

[‡] tegmark@mit.edu

ing these models grows exponentially with the number of neurons, thus rendering them intractable [24]. Moreover, in part due to these computational challenges, the existing literature on neural Ising models is largely devoid of any quantifications of uncertainties on their parameters, which makes it harder to resolve controversies about whether Ising models fit experimental data well [7, 11] or not [29, 39, 42].

It is therefore timely to develop an improved method for neural Ising modeling that can be applied to modern experimental data sets and can quantify parameter uncertainties while remaining computationally tractable. We will introduce such a method in this paper, and then use it to study the collective behavior of cortical excitatory and inhibitory neurons during the wake-sleep cycle (wakefulness, light sleep, and deep sleep) at multiple timescales. We seek to identify differences between excitatory and inhibitory neurons, as well as their distinctive behavior during wakefulness and sleep. Finally, we will also study the thermodynamic properties of the learned models.

II. METHODS

A. Data

We used data obtained from a set of multielectrode recordings in layers II/III of the human temporal cortex. Data was initially sampled at 30 kHz, then filtered and thresholded during the spike detection step. After spike sorting, a combination of morphological features of the spike waveforms along with the cross-correlogram of spike times were used to classify the cells as either excitatory (E) or inhibitory (I). This procedure produced a time-series of spikes for each of the N neurons (Panel (a) of Figure 1). The 12 hour recordings spanning overnight sleep were staged, yielding multiple state labels: awake, light sleep (stages II-III), deep sleep (SWS; slow-wave sleep), and REM (rapid eye movement). In Sec. III, we will use our method to analyze the awake, light sleep, and deep sleep states. Note that the data used in our analyses was devoid of any seizures. Additional details about the recordings and neuron classification procedure are presented in appendices A and B.

B. Maximum-entropy models

The maximum-entropy concept originates from statistical mechanics, and the connection between maximum-entropy models and classical thermodynamic distributions has been long known [43]. The main objective of maximum-entropy models is to construct a probability distribution that matches a set of empirical observations, but is otherwise as unconstrained and unstructured as possible. In this paper, we will focus on independent and

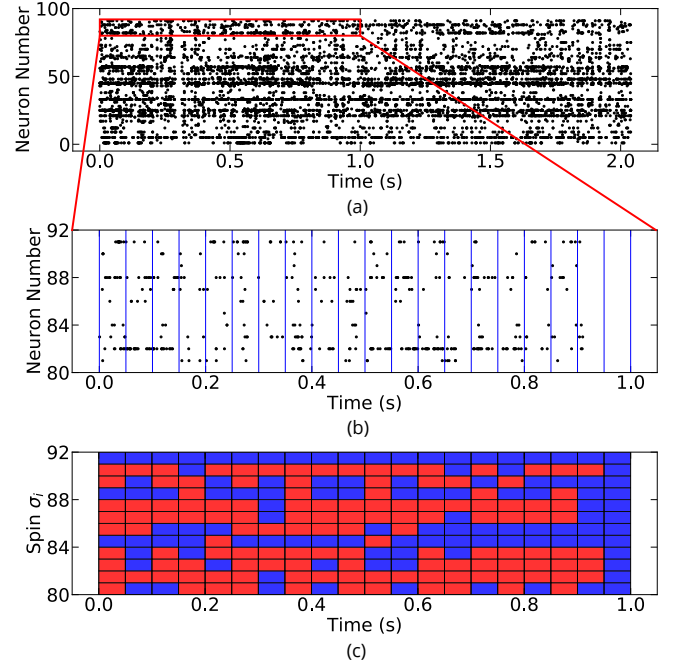


FIG. 1. Schematic representation of experimental data. (a) The spike train response of a set of 92 neurons in the human temporal cortex. (b) Discretization of the region delimited by the red rectangle into time bins of width $\Delta t = 50$ ms. (c) The raster corresponding to the discretization in panel (b), where red ($\sigma_i = +1$) denotes spiking of neuron i , and blue ($\sigma_i = -1$) represents silence.

pairwise models, which respectively match the first and second moments of the probability distribution of data.

Consider a network of N neurons for which we discretize each neuron's spike time-series into small windows of size Δt (Panel (b) of Figure 1). We performed this analysis for windows of size 20, 50, and 100 ms. For each neuron i , we assign $\sigma_i = +1$ if it generated an action potential (spiked) within this time window, or $\sigma_i = -1$ otherwise. Therefore, at each time step, our system can be described by a configuration vector $\boldsymbol{\sigma} \in \{-1, 1\}^N$, visualized as a column of N bits in Panel (c) of Figure 1.

Suppose our data consists of a set of activity patterns $\mathbf{X} = \{\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}, \dots, \boldsymbol{\sigma}^{(M)}\} \in \mathbb{R}^{N \times M}$, where each $\boldsymbol{\sigma}^{(k)}$ is a configuration vector. Then we can compute the mean spiking activity of each neuron

$$m_i = \langle \sigma_i \rangle_{\mathbf{X}} = \frac{1}{M} \sum_{k=1}^M \sigma_i^{(k)}, \quad (\text{II.1})$$

and the two-point function between pairs of neurons

$$Q_{ij} = \langle \sigma_i \sigma_j \rangle_{\mathbf{X}} = \frac{1}{M} \sum_{k=1}^M \sigma_i^{(k)} \sigma_j^{(k)}, \quad (\text{II.2})$$

where $\langle \cdot \rangle_{\mathbf{X}}$ denotes the empirical average with respect to our data \mathbf{X} . This average depends on the choice of

window size Δt . We also define the covariance matrix

$$C_{ij} = \langle \sigma_i \sigma_j \rangle_{\mathbf{X}} - \langle \sigma_i \rangle_{\mathbf{X}} \langle \sigma_j \rangle_{\mathbf{X}} = Q_{ij} - m_i m_j. \quad (\text{II.3})$$

Our goal is to build a model for the observed probability distribution $P^{(N)}(\boldsymbol{\sigma})$ which is consistent with the one- and two-point functions \mathbf{m} and \mathbf{Q} of the empirical data set \mathbf{X} . More formally, we derive the probability distribution by maximizing its entropy, subject to the constraints that enforce agreement with \mathbf{m} and \mathbf{Q} . Using the method of Lagrange multipliers to impose each constraint, the solution to the optimization problem is given by the Boltzmann distribution [43]:

$$P^{(2)}(\boldsymbol{\sigma}, \boldsymbol{\theta}) = \frac{e^{-\beta H(\boldsymbol{\sigma}, \boldsymbol{\theta})}}{Z(\boldsymbol{\theta})}, \quad (\text{II.4})$$

where

$$H(\boldsymbol{\sigma}, \boldsymbol{\theta}) = - \sum_{i=1}^N h_i \sigma_i - \sum_{i,j=1}^N J_{ij} \sigma_i \sigma_j = -\mathbf{h}^T \boldsymbol{\sigma} - \boldsymbol{\sigma}^T \mathbf{J} \boldsymbol{\sigma}, \quad (\text{II.5})$$

is the Hamiltonian of the system and $\boldsymbol{\theta} = (\mathbf{h}, \mathbf{J})$ is the parameter vector of our model. $P^{(2)}(\boldsymbol{\sigma}, \boldsymbol{\theta})$ is the probability that the network of neurons described by parameters $\boldsymbol{\theta}$ is in a state $\boldsymbol{\sigma}$. The partition function $Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma}, \boldsymbol{\theta})}$ normalizes the probability distribution. In physics applications, β is interpreted as the inverse temperature, but in the context of neuroscience it is simply a parameter that scales \mathbf{h} and \mathbf{J} , since the probability distribution depends on these parameters only through the combination $\beta \mathbf{h}$ and $\beta \mathbf{J}$. Therefore, without loss of generality, we can set $\beta = 1$ when fitting the model and absorb the scaling into our parameters \mathbf{h} and \mathbf{J} . The dependence of the system on this scaling parameter β will be explored in Section II C.

The Hamiltonian in Eq. (II.5) represents an energy function that assigns a weight to each configuration of spikes and silences. This Hamiltonian is identical to that of an Ising model for a system of interacting spins [44]. Therefore, we will refer to the pairwise model as the *Ising model*. In contrast to the usual Ising models in physics, where couplings typically reflect translational symmetry in some number of dimensions, these neural Ising models allow arbitrary all-to-all couplings. The parameter vector $\mathbf{h} \in \mathbb{R}^N$ can be interpreted as the intrinsic tendency of each neuron to fire and $\mathbf{J} \in \mathbb{R}^{N \times N}$ as the strength of pairwise interactions between the neurons. A positive J_{ij} favors the neurons firing together, while a negative J_{ij} does the opposite. We require that all the diagonal entries of \mathbf{J} (i.e. self-interactions) are zeros ($J_{ii} = 0$), since $\sigma_i^2 = 1$ implies that J_{ii} only contribute an irrelevant overall constant to our Hamiltonian. Without loss of generality, we take \mathbf{J} to be symmetric ($J_{ij} = J_{ji}$), since $\sum_{i,j} J_{ij} \sigma_i \sigma_j = \sum_{i,j} J_{ij} \sigma_j \sigma_i = \sum_{i,j} J_{ji} \sigma_i \sigma_j$.

Note that $\boldsymbol{\theta}$ has $N(N+1)/2$ independent components, which for a system with $N = 92$ neurons yields a parameter space of size 4,278. This is a reasonable number of

parameters to describe our data set with 234,848 entries (based on 50 ms binning). However, if we include higher-order interactions, then the model would have at least $O(N^3)$ parameters, which will require significantly more data points to avoid overfitting.

A simplified version of the pairwise model is obtained by assuming that $\mathbf{J} = 0$, i.e. that each neuron spikes independently of all the others. This defines what we will refer to as the *independent model* $P^{(1)}(\boldsymbol{\sigma}, \mathbf{h})$, which only constrains the mean spiking activity. Although the independent model is obviously not a realistic description of actual biological neural networks, it nonetheless provides a useful baseline comparison for the pairwise model. The main advantage of the independent model is that it is exactly solvable, since the partition function factorizes. One can easily show that its parameters must satisfy

$$h_i = \text{arctanh}(m_i). \quad (\text{II.6})$$

Unfortunately, such a closed-form correspondence between model parameters and data does not exist for the Ising model [44].

If we know the parameters $\boldsymbol{\theta}$ of the Ising model, then we can compute any thermodynamic quantity, although it may take an exponential amount of time to evaluate the partition function $Z(\boldsymbol{\theta})$. In particular, we can compute the mean spiking activity and the two-point correlation function:

$$m_i(\boldsymbol{\theta}) = \langle \sigma_i \rangle_{\boldsymbol{\theta}} = \sum_{\boldsymbol{\sigma}} P^{(2)}(\boldsymbol{\sigma}, \boldsymbol{\theta}) \sigma_i, \quad (\text{II.7})$$

$$Q_{ij}(\boldsymbol{\theta}) = \langle \sigma_i \sigma_j \rangle_{\boldsymbol{\theta}} = \sum_{\boldsymbol{\sigma}} P^{(2)}(\boldsymbol{\sigma}, \boldsymbol{\theta}) \sigma_i \sigma_j, \quad (\text{II.8})$$

where $\langle \cdot \rangle_{\boldsymbol{\theta}}$ denotes the expectation value with respect to our model. This is known as the forward Ising problem.

Our objective is to solve the inverse Ising problem - namely finding the best Ising model parameters $\boldsymbol{\theta}$ that describe our data \mathbf{X} . Although this inference problem is complicated for large networks, efficient algorithms for solving it are an active area of research [45], and a variety of methods have been used over the years to learn maximum entropy models. These methods include histogram Monte Carlo [46], minimum probability flow [47], adaptive cluster expansions [48, 49], and pseudo-likelihood [50]. In this work, we use a combination of Markov Chain Monte Carlo (MCMC) [51, 52] and gradient descent to iteratively estimate model averages of observables and update the parameters $\boldsymbol{\theta}$. A detailed description of our algorithm is presented in Appendix C.

Once we have learned the parameters of our model, a natural next step is to estimate the uncertainties associated with these parameters. As mentioned in Section I, this question has not been rigorously addressed in previous works on maximum entropy models. Our approach is to estimate the uncertainties using adaptive MCMC on

the space of parameter vectors θ . We perform a random walk using MCMC with a Gaussian proposal distribution centered at the solution of the inverse Ising problem, as found by our optimization procedure described in Appendix C. In order to determine the best value for the variance of the proposal distribution, we use adaptive MCMC to learn this parameter from previously generated Monte Carlo samples. A complete description of our method is given in Appendix D. One advantage of performing this random walk in parameter space is that we can fine-tune the solution previously obtained from the optimization algorithm.

C. Thermodynamic and information-theoretic quantities

The model parameters θ that we found can be interpreted as describing a system in thermal equilibrium at temperature $T = 1$, since as mentioned above, we set without loss of generality $\beta = 1/k_B T = 1/T = 1$. By analogy with a statistical mechanics system described by a Boltzmann distribution, we can introduce the temperature T as a scaling parameter of our Hamiltonian, which defines a one-parameter family of models whose thermodynamic properties can be explored. By varying T , we change the weights assigned to different activity patterns. It is important to emphasize that T is simply a model parameter, just like \mathbf{h} and \mathbf{J} , and we have no actual neural network that corresponds to this model at $T \neq 1$. Further, we estimate the heat capacity $C(T)$ and entropy $S(T)$ of our network. The heat capacity of a network of neurons can be interpreted as the variance of the surprise, where the surprise $-\log P(\sigma)$ determines how unexpected a particular activity pattern σ is for the network [22, 53, 54]. A small heat capacity indicates that all activity patterns appear with roughly the same probability, whereas a large heat capacity suggests that there is a balance between a few frequent patterns and multiple rare patterns [54, 55]. Moreover, divergences in the heat capacity can be used to determine the presence of a critical point.

The entropy S , on the other hand, can be used to compute the effective number 2^S of activity patterns of our system, which is an indicator of the size of the neural vocabulary. It also provides a bound on the network's capacity to encode and transmit information [22]. Furthermore, we can use the information-theoretic interpretation of entropy to assess the accuracy of our model. For this, we define a hierarchy of models, consisting of the independent $P^{(1)}(\sigma, \mathbf{h})$, pairwise $P^{(2)}(\sigma, \theta)$, and observed $P^{(N)}(\sigma)$ models, in this order. Each subsequent model captures more correlations among the data, with the empirical model capturing all the correlations. Given how the models are sorted from least to most structured, their respective entropies should satisfy $S_1 \geq S_2 \geq S_N$. The amount of correlation in the network is quantified by the “multi-information”

$$I_N \equiv S_1 - S_N, \quad (\text{II.9})$$

i.e., the decrease in entropy relative to the independent model [56, 57]. Similarly, $I_2 \equiv S_1 - S_2$ measures the decrease in entropy that is solely due to pairwise correlations. Therefore, the multi-information ratio I_2/I_N can be used to quantify the fraction of correlations captured by the pairwise model [11, 55].

To compute entropies and heat capacities, we take the following steps. For the empirical entropy S_N we use a low-bias estimator, specifically the Bayesian estimator with a centered Dirichlet mixture as its prior [58]. The entropy S_1 of the independent model can be computed analytically [44]. As for the entropy of the pairwise model, it is not feasible to compute the probability associated with every activity pattern, and even Monte Carlo sampling would lead to a poor approximation of the probability distribution due to the exponential growth of the phase space. A standard technique for approximating the entropy relies on integrating the heat capacity [22, 53]. However, this method requires generating Monte Carlo samples at many intermediate temperatures in order to get an accurate estimate of the integral, which can get computationally expensive. For our purposes, we therefore choose to use the Wang-Landau algorithm [59], which is better suited for this task and yields an estimate for both the entropy and the heat capacity. A summary of the algorithm is given in Appendix E.

III. RESULTS

In this section, we apply our methods to the above-mentioned data from *in vivo* multielectrode array recordings of neurons in the human temporal cortex. We construct maximum-entropy models of both inhibitory and excitatory neurons across multiple sleep stages. The figures presented in this section are based on the temporal bin size $\Delta t = 50$ ms, thus striking a good balance between capturing correlations among neurons and providing enough data for analysis. This value of Δt is slightly larger than the conventional 20 ms window used for retinal neurons [11, 22, 24, 60], reflecting the sparsity of activity patterns in the temporal cortex [61]. We repeated our analysis for both smaller ($\Delta t = 20$ ms) and larger ($\Delta t = 100$ ms) time bins, and confirmed that our conclusions also hold on these time scales and across all four data sets.

Figure 2 shows a sample fit of the pairwise model to the data recorded from 92 neurons in the human temporal cortex during wakefulness, with the neurons sorted in order of decreasing activity m_i . Panels (a) and (c) display the measured statistical properties \mathbf{m} and \mathbf{C} of the network, while panels (b) and (d) show the inferred model parameters \mathbf{h} and \mathbf{J} . This analysis shows that the majority of neurons have a very low probability of spiking, with the last 65 neurons firing in less than 3% of all

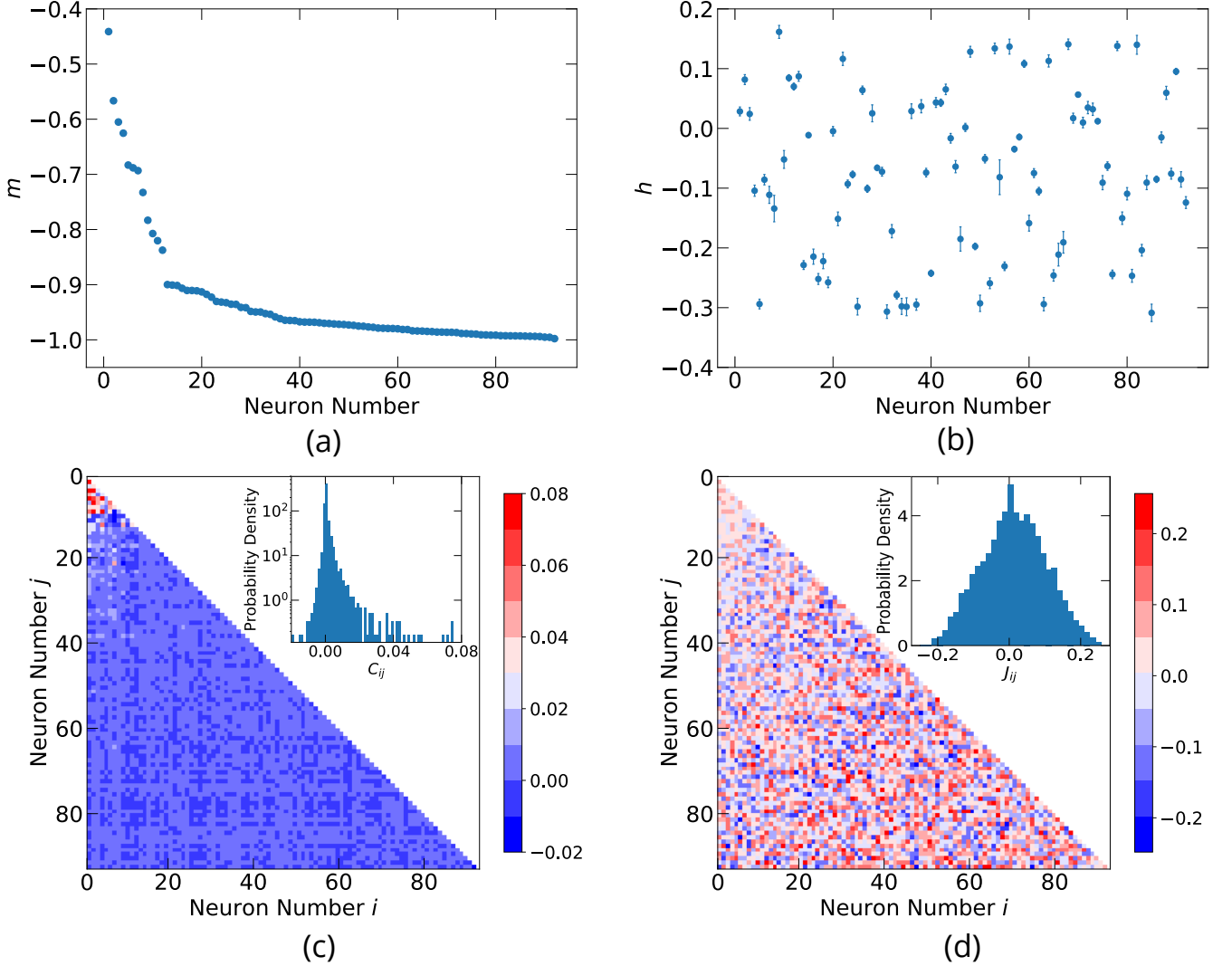


FIG. 2. Maximum entropy model for the population of 92 neurons in the awake state. (a) The mean spiking activity $m_i = \langle \sigma_i \rangle$. (b) The bias terms h_i of the inferred model, with a mean error of 0.01. Note that in the presence of interactions, the relationship between h_i and m_i no longer follows the closed form given in Eq. (II.6). (c) The correlation coefficients between pairs of neurons $C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$. The inset shows the population distribution of the correlation coefficients. (d) The pairwise coupling terms J_{ij} of the inferred model, with a mean error of 0.004. The inset shows the population distribution of the pairwise couplings. Note that panels (c) and (d) only show the lower triangle of the symmetric matrices C and J . In all panels, neurons are ordered by decreasing spiking frequency.

time bins. The values of the covariance C_{ij} for these less active neurons are close to zero, reflecting the fact that pairs of neurons are likely to be simultaneously silent. The majority of bias terms h_i take on negative values, thus showing the neurons' intrinsic tendency to remain silent.

The model uncertainties on \mathbf{h} are slightly larger than those on \mathbf{J} , but still below 5%. The couplings J_{ij} between neurons are widespread and can have either sign. The distribution of J_{ij} is seen to be roughly symmetric and centered around zero. This behavior is reminiscent of spin glasses [62], where competing interactions lead to frustration in the system. Therefore, we would expect the high-dimensional energy landscape to become

increasingly uneven and develop many local minima as the system size increases [22, 62]. This is consistent with the fact that multiple microscopic realizations of a system can lead to very similar macroscopic behaviors. We find that qualitatively similar observations and conclusions apply to both light and deep sleep states.

A. Reliable neurons

We confirm numerically that the optimization problem does indeed have multiple local minima. By running the algorithm described in Appendix C with different initializations, we observe that the optimization subroutine

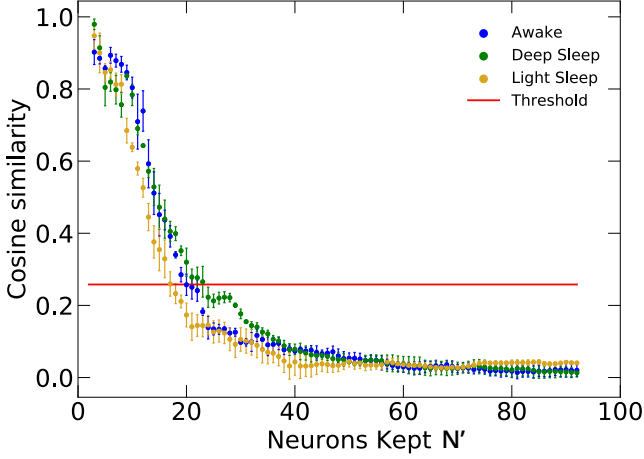


FIG. 3. Cosine similarity between all pairs of 10 parameter vectors θ estimated from 10 independent runs of our algorithm, when keeping only the N' most active neurons. Higher values of the cosine similarity indicate a smaller variance in parameters from run to run, so we see that the parameters of more active neurons can be more reliably measured. Our threshold for reliable neurons (horizontal line) is set at 0.25.

converges to a slightly different set of parameters each time. Since we want to be able to meaningfully interpret the model parameters θ , we will now investigate this issue.

To do so, we introduce cosine similarity as a measure of how reliably the parameters θ are replicated from run to run. The cosine similarity is a commonly used measure of the similarity of two vectors, defined simply as the cosine of the angle between them, computed via their dot product

$$\cos(\theta^{(i)}, \theta^{(j)}) \equiv \frac{\theta^{(i)} \cdot \theta^{(j)}}{|\theta^{(i)}| |\theta^{(j)}|}, \quad (\text{III.1})$$

where $|\theta^{(i)}| \equiv (\theta^{(i)} \cdot \theta^{(i)})^{1/2}$ denotes vector length. If two parameter vectors differ only slightly, then they are almost aligned in the parameter space and their cosine similarity is close to 1. If instead the difference between parameters is large, the two vectors will be roughly perpendicular in their high-dimensional parameter space and their cosine similarity will be near zero.

There are two candidate explanations for why our algorithm might produce different parameters θ for different initializations:

1. The hypothetical true values of \mathbf{m} and \mathbf{C} that we would measure if we had access to infinite data unfortunately lead to multiple local optima when fitting for θ .
2. These true \mathbf{m} and \mathbf{C} would give a unique local and global optimum θ , but the empirical \mathbf{m} and \mathbf{C} that we estimate from our finite data \mathbf{X} are sufficiently far from the true values, so that local optima arise.

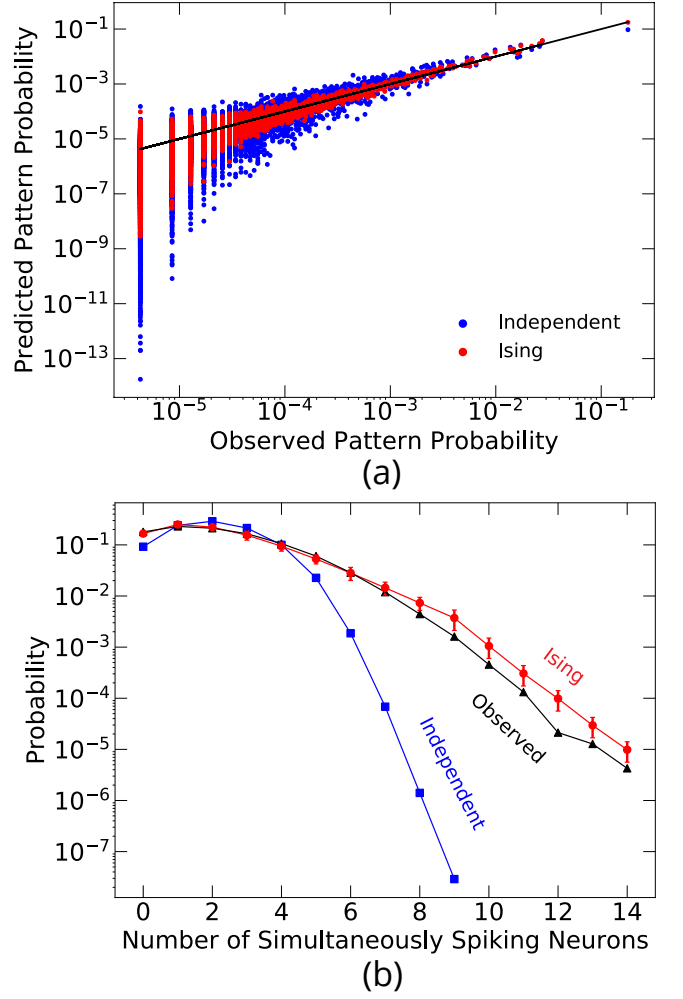


FIG. 4. Model predictions for the network of 19 reliable neurons during the awake state. (a) The probability of occurrence of each activity pattern predicted by the maximum entropy model ($P^{(1)}(\sigma, \mathbf{h})$ or $P^{(2)}(\sigma, \theta)$) is plotted against the observed pattern frequency from measured data ($P^{(N)}(\sigma)$). The black line corresponds to prediction matching observation. The apparent stratification on the lower end of the spectrum is due to the finite size of the data set. (b) The predicted and observed distribution of the number of simultaneously spiking neurons in each time bin. Error bars are asymmetrical because of the logarithmic scale. Similar results hold across all states, time scales, and data sets.

The more two neurons i and j spike, the more accurately and reliably we can estimate m_i and J_{ij} , whose uncertainties scale roughly as the square root of the number of spikes. To distinguish between explanations 1 and 2, we therefore perform 10 independent estimations of θ after discarding all but the N' most active neurons, and plot the average cosine similarity between pairs of parameter vectors as a function of N' in Figure 3. The results support the second hypothesis: parameters inferred from only the most active neurons are quite reliably recovered multiple times with different initializations, whereas

Neuron type	State	N	KL divergence per neuron		S_1/N	S_2/N	S_N/N	I_2/I_N
			Independent	Ising				
I	Awake	23	0.0146(1)	0.0044(1)	0.3326(1)	0.3227(2)	0.3204(2)	0.81(2)
I	Light Sleep	23	0.0325(1)	0.0102(1)	0.3578(1)	0.3337(2)	0.3282(2)	0.81(1)
I	Deep Sleep	23	0.0376(1)	0.0129(1)	0.3611(3)	0.3347(3)	0.3292(3)	0.83(2)
E	Awake	6	0.0008(2)	0.00010(3)	0.3975(3)	0.3893(3)	0.3890(3)	0.96(7)
E	Light Sleep	6	0.0010(2)	0.00030(3)	0.3205(3)	0.3102(2)	0.3092(3)	0.91(5)
E	Deep Sleep	14	0.0054(1)	0.0016(1)	0.2885(1)	0.2858(2)	0.2851(2)	0.80(8)
I and E	Awake	19	0.0185(1)	0.0056(1)	0.4826(2)	0.4698(2)	0.4673(2)	0.84(2)
I and E	Light Sleep	16	0.0323(1)	0.0116(1)	0.3976(1)	0.3811(2)	0.3792(1)	0.90(1)
I and E	Deep Sleep	23	0.0444(1)	0.0189(1)	0.4374(2)	0.4055(3)	0.4033(3)	0.94(1)

TABLE I. Information-theoretic quantities for populations of reliable inhibitory and excitatory neurons across different sleep states for data set 1. Similar results hold across all time scales and data sets. The numbers in parenthesis represent the uncertainty in the last digit.

those involving less active neurons are not.

We therefore define *reliable neurons* to be those for which the cosine similarity is above a given threshold, set here to 0.25. Intuitively, reliable neurons are those for which we have enough data to confidently infer their model parameters. For our data, this heuristic choice of threshold corresponds to neurons firing in at least 5% of the time windows, and selects approximately 20 neurons as reliable for each sleep state, roughly equally split between inhibitory and excitatory neurons. In the remainder of this section, we will only consider reliable neurons.

In a related approach based on reliable interactions [23], during the model fitting, only the frequent activity patterns of the network were taken into account and all the configurations whose occurrence rate was below a certain threshold were discarded. Our approach is similar, except that instead of discarding time segments, we discard the least active neurons.

B. Collective behavior

The first success of the pairwise model applied to reliable neurons can be seen when looking at the collective behavior in the network. Panel (a) of Figure 4 shows the probabilities that the independent and Ising models assign to each activity pattern. The Ising model is seen to significantly outperform the independent model, accurately predicting the observed pattern frequencies, except for the most rarely encountered patterns. Panel (b) of Figure 4 shows the spike synchrony, defined as the probability that a given number of neurons spike within the same time window, revealing that the independent model strongly underpredicts events with many synchronous spikes. The Ising model is seen to perform dramatically better.

C. Information-theoretic quantities

In addition to Figure 4, we can quantify the success of the Ising model by computing the information-theoretic quantities introduced in Section II C for different sleep states and subsets of neuron types. The results are summarized in Table I.

Recall that each activity pattern can be viewed as a vector of N bits that specifies which of the N neurons fired during a given time interval. Since the entropy S_i can be interpreted as the number of bits required to describe a typical pattern drawn from the probability distribution $P^{(i)}$, we expect $S_i = N$ if all neurons randomly fired or remained silent with equal probability. However, since neurons are mostly inactive, the entropy is seen to be significantly smaller ($S_1/N \sim 0.3 - 0.4$) for the independent model that accounts for these firing rates, and even smaller ($S_2 < S_1$) for the Ising model that also incorporates predictable regularities due to pairwise interactions. The entropy is the smallest ($S_N < S_2$) for the observed distribution, which includes regularities not captured by the Ising model.

The multi-information ratio $I_2/I_N = (S_1 - S_2)/(S_1 - S_N)$ (see Eq. II.9) is listed in the last column of Table I. It reveals that the Ising model captures the majority of the correlative structure in the data: in transitioning from the independent model to the observed probability distribution for neuron activity patterns, the Ising model accounts for 80% – 95% of the correlations, depending on sleep state and neuron type.

Lastly, for a given model $P^{(i)}$, the Kullback-Leibler (KL) divergence $D_{\text{KL}}(P^{(i)} \parallel P^{(N)})$ can be interpreted as the number of extra bits required to describe the observed neuron activity patterns simply because the model does not fit the data perfectly. Thus taking the KL divergence as a measure of how poorly a model fits the observed probability distribution of activity patterns, Table I shows that the Ising model outperforms the independent model for all the nine cases listed, typically by

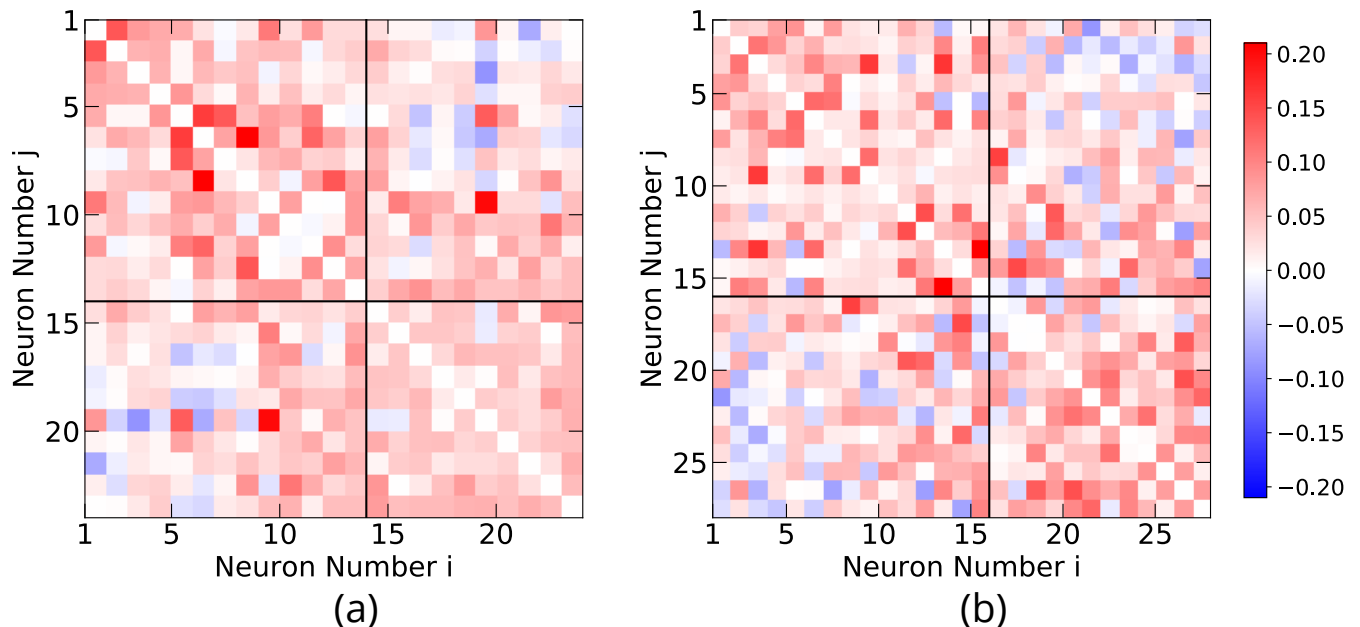


FIG. 5. The pairwise coupling matrix J inferred from the population of reliable neurons in the deep sleep state. Neurons are sorted by type, with (a) inhibitory neurons at positions 1 – 13 and excitatory neurons at positions 14 – 23 for data set 1, (b) inhibitory neurons at positions 1 – 15 and excitatory neurons at positions 16 – 27 for data set 2. The couplings J_{ij} among inhibitory neurons, as well as the couplings among excitatory neurons (diagonal blocks) are almost entirely positive. The couplings J_{ij} between inhibitory and excitatory neurons (off-diagonal blocks) display a mix of both positive and negative values. Similar results hold across all states, time scales, and data sets.

a factor greater than three. We also see that excitatory neurons are the easiest to predict, which is at least partly because we run our algorithm on a smaller subset of them. Furthermore, we observe that the Ising model provides a slightly better match to the observed activity pattern in the awake state. A worthwhile question for future theoretical work is to investigate whether more desynchronized states (such as the awake state) can in general be better modeled using purely pairwise interactions, without considering strong (synchronizing) external drives.

D. Excitatory and inhibitory neurons

We now turn to a central question of this paper: identifying differences between inhibitory and excitatory neurons. We begin by investigating the interactions between these two neuron types. Figure 5 shows the couplings J_{ij} between neurons of both types, revealing an interesting pattern: the couplings among I-neurons and among E-neurons are almost all positive, whereas the couplings between I- and E-neurons are a mix of positive and negative values, slightly skewed toward the positive. This observation suggests a non-trivial interaction between the I- and E-neurons, reflecting the ability of I-neurons to inhibit E-neurons in response to their excitation by E-neurons. The exact causal relationship and timescale of such interactions is hard to interpret given that we are

significantly subsampling (by orders of magnitude) from tens of thousands of neurons in these local circuits. In addition, the observed positive correlation among the I-neurons may hint to the existence of a strong common input or higher order correlations, both of which are inherently not addressable due to the size of the neuron pool and the nature of the pairwise model used in this study.

To further uncover the differences between inhibitory and excitatory neurons, we fit Ising models separately for each neuron type. Figure 6 shows that the Ising model works fairly well for the I-neurons alone, but fails completely for the E-neurons alone, dramatically overpredicting how often neurons spike together.

The success is arguably as noteworthy as the failure: it is striking that the Ising model works as well as it does even though the couplings of our observed neurons to the other roughly 10^{11} neurons in the brain are completely ignored. One interpretation of this success is the well-known hypothesis that collective neural dynamics occurs on a low-dimensional submanifold [63], such that observing a modest number of neurons suffices for determining a state's location in this submanifold.

Within the context of this interpretation, Figure 4 illustrates how a subsystem of merely 19 reliable neurons suffice to capture enough information about their surrounding dynamics to be accurately modeled as an isolated system, as long as both excitatory and inhibitory neurons are included. On the other hand, if the in-

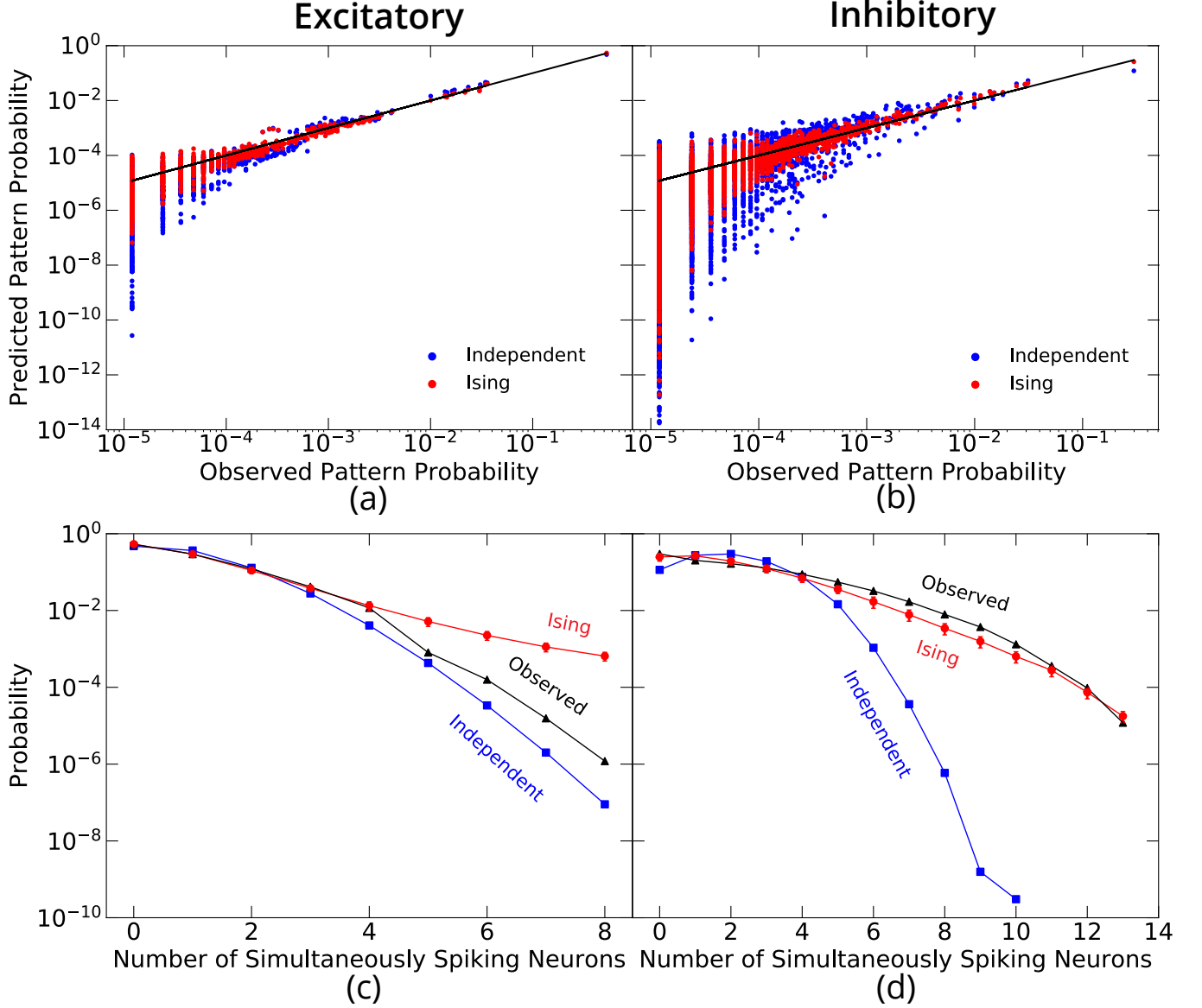


FIG. 6. Difference between E- and I-neurons. The probability of occurrence of each activity pattern predicted by the maximum entropy model ($P^{(1)}(\sigma, \mathbf{h})$ or $P^{(2)}(\sigma, \theta)$) is plotted against the observed pattern frequency from measured data ($P^{(N)}(\sigma)$) for a network of (a) 14 E-neurons, (b) 23 I-neurons. The black line corresponds to prediction matching observation. The predicted and observed distribution of the number of simultaneously spiking neurons in each time bin for a network of (c) 14 E-neurons, (d) 23 I-neurons. Error bars are not symmetrical because of the logarithmic scale. All plots are for the deep sleep state, but similar results hold across all states, time scales, and data sets.

hibitory effects of I-neurons are not modeled, then the synchrony among E-neurons is dramatically overestimated (see Panel (c) of Figure 6). Conversely, there is a slight hint that if the excitatory effects of E-neurons are not modeled, then the synchrony among I-neurons is slightly, but consistently, underestimated (see Panel (d) of Figure 6). In other words, accurate modeling of the joint system of E- and I-neurons requires sampling from both of its two distinct sub-populations.

E. Sleep states

Now that we have described the behavior of inhibitory and excitatory neurons that is common to all sleep states, we proceed to investigate the distinctions that appear between E and I neurons in different states. It is well-known that the spiking frequencies of many neurons vary strongly between sleep states and our data confirms this fact [64–68]. Additionally, we show that such a dependence on sleep states applies not only to the first moments \mathbf{m} (corresponding to firing frequencies), but also

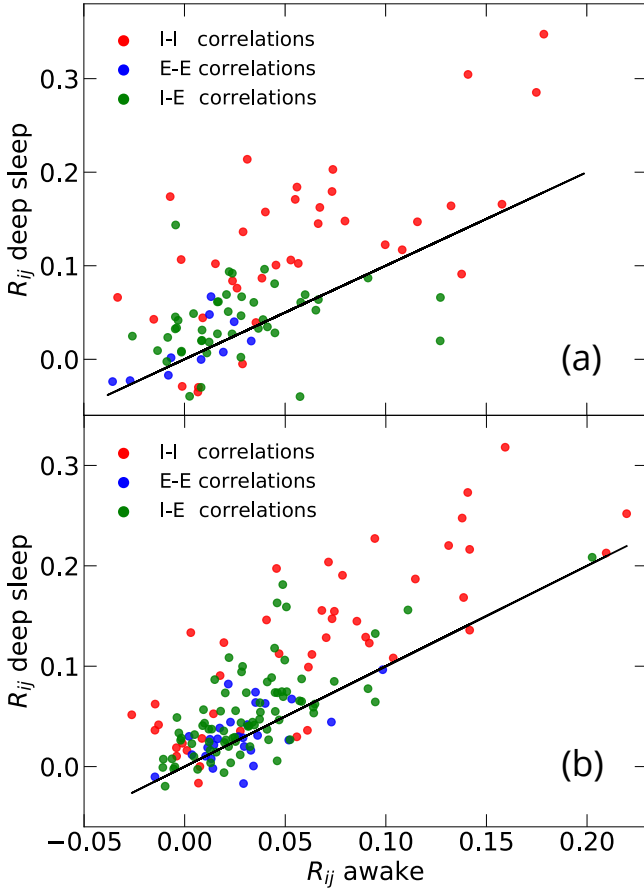


FIG. 7. Pearson correlation coefficients R_{ij} between the same set of (a) 14 (data set 1) (b) 18 (data set 2) reliable neurons during awake and deep sleep states. The coefficients R_{ij} are labeled based on the type of neurons i and j . The correlations are different during the two states, with I-I correlations being predominantly larger during deep sleep.

to second moments \mathbf{C} . Figure 7 compares the Pearson correlation coefficients $R_{ij} \equiv C_{ij}/[C_{ii}C_{jj}]^{1/2}$ between all neuron pairs. Note that we only study the neurons that are reliable in both states and split the correlation coefficients according to neuron types. Noticeably, most dots lie far from the diagonal line, meaning that the corresponding correlations are different during wakefulness and sleep. For both data sets, we observe that I-I correlations are slightly higher during deep sleep, which is indicated by their points lying predominantly above the diagonal line. This means that the network of inhibitory neurons is more internally correlated during sleep, perhaps reflecting the familiar slow-wave coherent oscillations associated with deep sleep.

F. Thermodynamic quantities

Finally, we study the thermodynamic properties of our models. Figure 8 shows the heat capacity and entropy

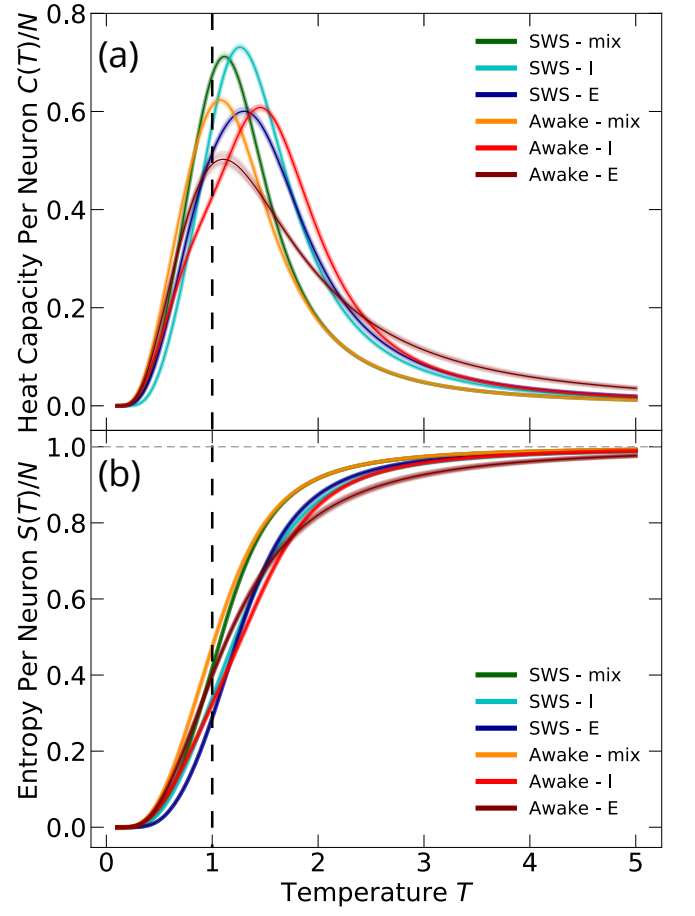


FIG. 8. Heat capacity and entropy in maximum entropy models. (a) Heat capacity per neuron $C(T)/N$ as a function of temperature for different types of neurons in the awake and SWS states. The heat capacity peaks around $T = 1$ (vertical dashed line). (b) Entropy per neuron $S(T)/N$ as a function of temperature for different types of neurons in the awake and SWS states. The entropy experiences a jump around $T = 1$ and saturates to its maximum value $S(T)/N = 1$ (horizontal dotted line) at high temperatures. Both panels indicate signatures of criticality. Errors are displayed as shaded regions around the main lines. Similar results hold across all time bins and data sets.

as a function of temperature for inhibitory, excitatory, and mixed networks of neurons in different sleep states. It is noteworthy that these thermodynamic curves look qualitatively similar across neuron types and sleep states despite the strong dependence on sleep state that we saw above at the level of individual neurons. We see that for all networks, the heat capacity peaks around the operating temperature $T = 1$, while the entropy exhibits a jump near that same temperature. This behavior is reminiscent of a phase transition, in which the system goes from an ordered, low-entropy state, to a disordered, high-entropy state [44]. At high temperatures, where all configurations have the same Boltzmann weight, the entropy approaches its maximum $S(T) = N$. The peak in

the heat capacity curves becomes sharper for larger system size N , which is consistent with finite-size scaling. In all the networks studied, the heat capacity peak is found to be at a slightly higher temperature than the operating point $T = 1$. Moreover, these signatures of criticality are robust to changes in parameters θ , as indicated by the small size of our error bars. The question of interpreting these criticality hints is subtle and we will return to it in Section IV B.

IV. DISCUSSION

In this paper, we have introduced a statistically rigorous and computationally efficient method for inferring an Ising model of the spiking activity of neural networks, and have applied it to excitatory and inhibitory human cortical neurons during the wake-sleep cycle. Our method provides accurate uncertainty estimates for all model parameters and derived quantities, as described in Appendix D, and remains tractable for large ($N \approx 100$) neuronal networks. This further improves previous work where parameters uncertainty was estimated either by repeatedly running the inference algorithm for different neuron subsets [11, 20, 22–24, 53, 54] or not at all [7, 18, 61]; the former approach may suffer from the population of neurons being inhomogeneous, consisting of different neuron types, being exposed to variable stimuli, or having drastically varying firing rates.

A. Modeling spiking behavior

We found that the Ising model describes neuronal collective behavior much better than the independent model throughout the sleep cycle as long as both excitatory (E) and inhibitory (I) neurons are modeled (Figure 4). These observations are consistent across sleep states and data sets. We found that the pairwise correlations in the Ising model accounted for 80% – 95% of all correlations in the data and that the Ising model yielded a KL divergence more than three times smaller than that of the independent model (Table I).

By modeling inhibitory and excitatory networks separately, we found that accurately predicting neuron synchrony requires sampling both E- and I-neurons. If the inhibitory effects of I-neurons were ignored, then synchrony among E-neurons was dramatically overestimated (Panel (c) of Figure 6). In contrast, the I-neurons could be fairly accurately modeled on their own, although ignoring the excitatory effects of the E-neurons caused a slight underestimation of their synchronous activity (Panel (d) of Figure 6). The inhibitory effect of I-neurons on E-neurons was also reflected by negative pairwise couplings J_{ij} between the two neuron types (Figure 5).

A recent study [61] also examined excitatory and inhibitory neurons in the human cortex during wakefulness and deep sleep, reporting that the spiking activity

is dominated by pairwise interactions during wakefulness but is population-wide during deep sleep [61], to an extent not captured by the Ising model for inhibitory neurons. While we found I-neurons to be accurately fit by the Ising model, we reproduced the conclusion of [61] that inhibitory neurons have higher average correlation during sleep, by analyzing the intra-class (I-I, E-E) and inter-class (E-I) interactions (Figure 7). Moreover, we found this higher I-I correlation to be more pronounced in deep sleep than in light sleep.

These findings further complement prior observed differences between excitatory and inhibitory neurons in the wake-sleep cycle. It has been shown that the cortical neurons manifest an overall multiscale balance, i.e. ensemble excitation and inhibition co-fluctuate, a property that is observed across multiple timescale and involves transient deviations from the absolute balance that are more prominent during deep sleep [69]. In addition, excitatory neurons show a tendency to be active more focally, manifesting a distance-dependent decay in their correlation, while in contrast, inhibitory neurons show a more robust correlation within the span of a cortical column [68]. These collective features are well matched with the envisioned significant role of the inhibitory neurons in gating information [70, 71] and in regulating oscillations including slow-wave sleep [68, 72].

B. Is the cortex critical?

Whether the collective neural activity is optimized to operate at a critical point [53, 73, 74] or reflects an “asynchronous irregular” (AI) regime [75, 76] (as a result of irregular firing with weak mean correlations despite substantial shared input) remains a controversial topic. A number of studies have used thermodynamic-based measures of population activity, such as the divergence of heat capacity and the power-law scaling of neural activity [53–55, 60, 77], as evidence in support of criticality. Each of these hypotheses has important implications for neural coding. On one hand, the critical regime provides optimal information retention [78], information integration [79], and maximum sensitivity to input variations [80]. On the other hand, the AI regime reflects a dynamic balance state of excitatory and inhibitory fluctuations [81, 82], providing non-redundant fast network responses [82].

We tested for criticality by measuring the heat capacity and entropy as a function of temperature. We found that although the activity and correlation between individual neurons varied dramatically with sleep state (Figure 7), the thermodynamic quantities remained qualitatively unchanged, with the heat capacity $C(T)$ peaking just after $T = 1$ (Panel (a) of Figure 8) and entropy $S(T)$ experiencing a significant jump at $T \approx 1$ (Panel (b) of Figure 8). These characteristics may suggest long-range correlation at the critical temperature, allowing the neurons to coordinate across extended distances. As a

consequence, the divergence of response functions would maximize the sensitivity of the system to the stimuli. In addition, the jump in entropy near the critical point implies a significant increase in the number of effective activity patterns available to the system, thus allowing it to exhibit a larger spectrum of possible responses.

However, we wish to draw attention to experimental and theoretical evidence for why we should take these criticality indications with a grain of salt. Experimentally, previous analysis of our data set have shown that across the wake-sleep cycle, human (as well as monkey and cat) cortical neurons do not show power-law scaling and are better fit with a multi-exponential model, suggesting that the underlying dynamics mirror the interaction of excitation and inhibition at multiple timescales [69, 83].

Turning to theoretical evidence, shared input to the network [84, 85], higher-order couplings [30], and randomly sub-sampled data [86, 87] can all masquerade as signatures of criticality. In fact, networks in self-sustained irregular regimes away from criticality can still manifest universal scaling functions [88]. Our data, and essentially any *in vivo* multielectrode measurements, dramatically subsample the neural activity and only provide partial measurements to estimate correlations, and leave us with latent variables that can only be inferred indirectly.

Finally, it has been suggested that the alleged criticality of the maximum entropy models can be a consequence of the inference procedure [89–92]. The basic argument is that the Ising model only supports long-range correlations when it is near-critical, such that if the data contains long-range correlations, then the Ising model that best fits these correlations is likely to be near-critical. The fact that we observe signatures of criticality independent of sleep state (awake, light sleep, and deep sleep), neuron type (E, I, and mixed), and timescale (20, 50, and 100 ms) thus suggests long-range correlations in all states that may or may not be due to critical behavior. A more detailed analysis of spatial and temporal correlations is required in order to settle the criticality controversy.

C. Outlook

To shed further light on the workings of biological neural networks, there are many opportunities to improve the method we have presented. The Ising model is simply the maximum-entropy model that matches all firing rates and equal-time two-point correlations. Therefore, straightforward generalizations involve including correlations between different times and including three-point functions and higher-order correlations.

The addition of higher-order moments as constraints has the potential to improve the model [28, 31–34, 93], especially since input nonlinearities have been argued to affect beyond-pairwise correlations [13, 27]. However, even the inclusion of three-point functions already re-

sults in a model with $O(N^3)$ parameters, which requires significantly more data points to avoid overfitting. Moreover, adding higher-order correlations renders the optimization problem even harder since the energy landscape will likely have even more local minima. Recently, several methods have been introduced for capturing beyond-pairwise correlations that may provide computationally tractable solutions [30, 94–96], but building accurate models of neuronal networks whose parameters and associated uncertainties can be inferred efficiently remains an active area of research.

Spatiotemporal extensions of the maximum entropy model to include correlations between different times have also been pursued in the past [97–99]. However, there are still plenty of interesting questions to study using these models and many valuable opportunities for improving their reliability, uncertainty estimation, and computational efficiency. For example, a spatiotemporal extension of our analysis should be able to explicitly quantify which neurons are exciting or inhibiting others at later times.

There are also ample opportunities to tackle the above-mentioned subsampling limitation, i.e., that current neuron data tends to be recorded from merely a minuscule fraction of all neurons. If it is correct that the relevant dynamics occur on a low-dimensional submanifold, then there is great value in further experimental and theoretical work to determine its dimensionality (and how many neurons suffice for capturing said dynamics).

Fortunately, rapid technological progress is enabling simultaneous high-quality recordings from ever-larger numbers of neurons. This will produce a gold mine of data that can be tested using our method and further improvements thereof to deepen our understanding of biological neural network dynamics.

ACKNOWLEDGMENTS

The authors would like to thank James Crutchfield, Cina Aghamohammadi, and Joshua Deutsch for helpful discussions, and the Center for Brains, Minds, and Machines (CBMM) for hospitality. This work was supported by NSF grant 1734870, the Foundational Questions Institute, and the Rothberg Family Fund for Cognitive Science. CZ was supported by the Whiteman Fellowship.

Appendix A: Patients and Recordings

For this study, we used 4 multielectrode temporal cortex recording sessions, each lasting 12 hours, including overnight sleep. The recordings were done using silicon-based NeuroProbe (from BlackRock Microsystems Inc.), composed of a 10×10 2D array of micro-electrodes, each 1 mm thick, separated by a $400 \mu\text{m}$ spacing [14, 15]. Four corner electrodes were used for grounding the electronics, resulting in 96 functional electrodes sampling the data at

30 kHz. All patients had focal epilepsy as confirmed by postoperative histology. Based on the post-excision histological exams, electrode tips reached the layers II/III of the middle temporal gyrus. The array implantation and surgical excision were performed for medical purposes according to IRB approved experiments. Patient multielectrode recording experiment approval was granted by the Institutional Review Boards of Massachusetts General Hospital, and Brigham & Women's Hospital. Sleep staging was performed based on video monitoring and a combination of scalp EEG and intracranial EEG recordings simultaneously acquired with the multielectrode array system. Sleep-wake categories were assigned as awake, light sleep, deep sleep (SWS), and rapid eye movement (REM). Due to the short duration of REM sleep, we did not use the REM sleep data in our study. For more details on patients and recordings, see the methods and supplementary material of prior publications [68, 72].

Appendix B: Spike Sorting and Cell Categorization

After thresholding the raw data for spike detection, the selected spikes were sorted offline using the first three principal components of spike waveforms from each electrode. This step was then followed by an automated clustering using expectation-maximization (EM). The overall morpho-functional characteristics of the spike waveform and putative mono-synaptic connections provided the final two cell classes of excitatory (E) and inhibitory (I). First, each cell's average spike waveforms was used to measure a variety of features, such as half-width of the positive peak, half-width of the negative peak, interval between negative and positive peaks (valley-to-peak), and the ratio of the negative to positive peak amplitude. Based on a K-means algorithm, these features were used to categorize the cells based on the morphology of the spike waveform of Fast-Spiking (FS, putative inhibitory) and Regular-Spiking (RS, putative excitatory) [100, 101]. The FS and RS functional labels were later tested and verified by the cross-correlograms interactions indicative of putative monosynaptic connections [102]. Only the verified categorized cells were used in the analyses. This resulted in 4 data sets, each with 92, 80, 36, and 30 neurons respectively. More details about spike sorting and cell classification are provided in the methods and supplementary material of [68, 69].

Appendix C: Learning the parameters of the Ising model

To solve the inverse Ising problem, we must find the parameters θ such that the Boltzmann distribution $P^{(2)}(\sigma, \theta)$ is as close as possible to the observed distribution $P^{(N)}(\sigma)$ over the data set $\mathbf{X} = \{\sigma^{(1)}, \sigma^{(2)}, \dots, \sigma^{(M)}\}$. More formally, this implies max-

imizing the likelihood

$$L(\mathbf{X}, \theta) = \prod_{k=1}^M \frac{e^{-H(\sigma^{(k)}, \theta)}}{Z(\theta)} = \frac{e^{-\sum_{k=1}^M H(\sigma^{(k)}, \theta)}}{Z^M(\theta)}. \quad (\text{C.1})$$

that the data is produced by our model. In practice, it is often more convenient to consider the log-likelihood

$$\begin{aligned} \log L(\mathbf{X}, \theta) &= -\sum_{k=1}^M H(\sigma^{(k)}, \theta) - M \log Z(\theta) \\ &= \sum_{k=1}^M \left(\mathbf{h}^T \sigma^{(k)} + \sigma^{(k)T} \mathbf{J} \sigma^{(k)} \right) - M \log Z(\theta) \\ &= M \sum_{i=1}^N h_i m_i + M \sum_{i,j=1}^N J_{ij} Q_{ij} - M \log Z(\theta) \\ &= M [\mathbf{h}^T \mathbf{m} + \text{Tr}(\mathbf{J} \mathbf{Q}) - \log Z(\theta)]. \quad (\text{C.2}) \end{aligned}$$

Notice that the log-likelihood depends only on one- and two-point correlation functions, and not on the entire data set. This is consistent with our expectation that \mathbf{m} and \mathbf{Q} provide sufficient information to learn the pairwise model.

The log-likelihood in Eq. (C.2) can be maximized by using optimization algorithms, such as gradient descent. The gradient of the log-likelihood is straightforward to compute [45] and is given by

$$\frac{\partial \log L(\mathbf{X}, \theta)}{\partial h_i} = M(m_i - m_i(\theta)), \quad (\text{C.3})$$

$$\frac{\partial \log L(\mathbf{X}, \theta)}{\partial J_{ij}} = M(Q_{ij} - Q_{ij}(\theta)), \quad (\text{C.4})$$

where $\mathbf{m}(\theta)$ and $\mathbf{Q}(\theta)$ denote the first and second moments predicted by the Ising model using Eqs. (II.7) and (II.8). The gradient descent update rule at each iteration is simply

$$\mathbf{h} \leftarrow \mathbf{h} + \eta[\mathbf{m} - \mathbf{m}(\theta)], \quad (\text{C.5})$$

$$\mathbf{J} \leftarrow \mathbf{J} + \eta[\mathbf{Q} - \mathbf{Q}(\theta)], \quad (\text{C.6})$$

where we have absorbed M into the learning rate η . Once we reach the maximum log-likelihood, the gradient and hence these updates will vanish, implying that our model will match the first and second moments of the observed distribution, as expected for a pairwise maximum entropy model.

In order to compute the average values $\mathbf{m}(\theta)$ and $\mathbf{Q}(\theta)$ on the right-hand side of the equations above, we need to sum over all 2^N possible configurations of the system. For large system sizes ($N > 30$), this computation

becomes intractable. We therefore use Markov Chain Monte Carlo (MCMC) [51, 52] with the usual Metropolis-Hastings sampling procedure to generate representative samples from $P^{(2)}(\boldsymbol{\sigma}, \boldsymbol{\theta})$ and estimate $\mathbf{m}(\boldsymbol{\theta})$ and $\mathbf{Q}(\boldsymbol{\theta})$ using Eqs. (II.7) and (II.8). In practice, we draw 10^6 samples and set $\eta = 0.01$.

At each iteration, we need to assess how closely our model matches the data. Ideally, we would compute the log-likelihood, but estimating the partition function takes an exponential amount of time. Therefore, we use the root-mean-squared-error (RMSE) between the predicted and measured one- and two-point functions as a proxy for monitoring the convergence of the optimization algorithm

$$\ell = \sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - m_i(\boldsymbol{\theta}))^2 + \frac{1}{N^2} \sum_{i,j=1}^N (Q_{ij} - Q_{ij}(\boldsymbol{\theta}))^2}, \quad (\text{C.7})$$

which is simply the likelihood gradient magnitude $|\nabla_{\boldsymbol{\theta}} L|$ except for two normalization factors. Our success criterion for stopping the optimization procedure is $\ell < 0.001$.

Appendix D: Estimating uncertainties on the parameters of the Ising model

1. MCMC on parameter space

Suppose that our optimization routine, described in Appendix C, converged to $\boldsymbol{\theta}_* = (\mathbf{h}_*, \mathbf{J}_*)$, which is the maximum likelihood estimate of our model parameters given a data set \mathbf{X} . We now start at $\boldsymbol{\theta}_*$ and perform a random walk in the space of parameters $\boldsymbol{\theta}$ using MCMC with a Gaussian proposal distribution and assuming a uniform prior. The algorithm is as follows:

1. Initiate the MCMC chain with $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_*$.
2. At each iteration t :
 - Draw a candidate step $\mathbf{s}_t \in \mathbb{R}^{N(N+1)/2}$ according to the multivariate Gaussian distribution $\mathcal{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the step covariance matrix. The proposed parameter vector is $\boldsymbol{\theta}' = \boldsymbol{\theta}_{t-1} + \mathbf{s}_t$.
 - Accept the proposed move with probability

$$p_{\text{accept}} = \min \left[1, \frac{L(\mathbf{X}, \boldsymbol{\theta}')}{L(\mathbf{X}, \boldsymbol{\theta}_{t-1})} \right] \\ = \min \left[1, e^{M((\mathbf{h}' - \mathbf{h}_{t-1})^T \mathbf{m} + \text{Tr}((\mathbf{J}' - \mathbf{J}_{t-1}) \mathbf{Q}) + \log \left(\frac{Z(\boldsymbol{\theta}_{t-1})}{Z(\boldsymbol{\theta}')} \right))} \right]. \quad (\text{D.1})$$

3. After generating k representative samples $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\}$, we can use them to compute the mean and standard deviation on $\boldsymbol{\theta}$, thus quantifying the uncertainty for the Ising model parameters.

2. Adaptive MCMC

One important technicality is choosing the step covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is too small, then most of our proposals will be accepted but we will not get to explore much of the parameter space. If $\boldsymbol{\Sigma}$ is too large, our steps will be very big and most of our proposals will be rejected. Moreover, we do not know how large $\boldsymbol{\Sigma}$ has to be along each dimension.

A standard way to deal with these problems is to employ adaptive MCMC techniques, which use the history of previous samples to update the proposal distribution dynamically at each step [103]. We can update $\boldsymbol{\Sigma}$ using either a fixed number of previous states, as is the case for Adaptive Proposal (AP) MCMC [104], or the whole chain generated so far, as in Adaptive Metropolis (AM) MCMC [105]. Updating the proposal distribution during our algorithm gives us a better chance at convergence. However, the adaptive algorithms lead to a stochastic process that is clearly no longer Markovian. Therefore, adaptive schemes may converge to incorrect distributions, since the standard ergodicity result no longer applies to non-Markovian processes. This is indeed the case for the AP algorithm, whereas AM has proven to be ergodic [105]. Therefore, we use the AM algorithm in this work.

The covariance matrix $\boldsymbol{\Sigma}_t$ at each iteration t is chosen according to

$$\boldsymbol{\Sigma}_t = \begin{cases} \boldsymbol{\Sigma}_0 & \text{if } t \leq t_0 \\ \lambda_d \text{Cov}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{t-1}) & \text{if } t > t_0 \end{cases} \quad (\text{D.2})$$

where t_0 is the initial period after which adaptation begins. Typically, we set $t_0 = 500$ and run the simulation for 10,000 steps. The scaling parameter $\lambda_d = 2.4^2/d$ depends only on the dimension of the parameter space $d = N(N-1)/2$ and is chosen so that it optimizes the mixing properties of the random walk in the case of Gaussian proposal and target distributions [106]. As an initial guess for the covariance matrix we choose the identity $\boldsymbol{\Sigma}_0 = \alpha \mathbf{I}_d$, scaled down by a factor $\alpha = 10^{-5}$ such that approximately half of the proposed steps get accepted during MCMC.

3. Approximating ratios of partition functions

The algorithm presented above relies on computing the ratio of partition functions in its second step. This can be performed exactly for small systems, where we can directly compute the partition function. However, it becomes unfeasible for larger systems, including those with $N \approx 100$ which we are interested in. Therefore we need a way to approximate the ratio of partition functions.

Although there are a few methods for directly estimating the partition function, such as the Wang-Landau algorithm (see Appendix E), they can be computationally expensive to run at each MCMC step and for now we are

more interested in approximating the ratio of partition functions $Z(\theta')/Z(\theta)$, rather than the partition functions themselves. We can re-write the ratio as follows:

$$\begin{aligned} \frac{Z(\theta')}{Z(\theta)} &= \frac{1}{Z(\theta)} \sum_{\sigma} e^{-H(\sigma, \theta')} \\ &= \sum_{\sigma} \frac{e^{-H(\sigma, \theta')}}{e^{-H(\sigma, \theta)}} P^{(2)}(\sigma, \theta) = \left\langle e^{-H(\sigma, \theta' - \theta)} \right\rangle_{\theta}. \end{aligned} \quad (\text{D.3})$$

Now we can approximate the right-hand side by drawing Monte Carlo samples from the model with parameters θ . If we draw M such samples $\sigma^{(k)}$, then

$$\frac{Z(\theta')}{Z(\theta)} \approx \frac{1}{M} \sum_{k=1}^M e^{-H(\sigma^{(k)}, \theta' - \theta)}, \quad (\text{D.4})$$

where the approximation sign was used to denote the fact that the two equations are equal only in the limit of infinite samples M . In practice, the right-hand side of Eq.(D.4) is a good approximation only when the distribution $P^{(2)}(\sigma, \theta)$ is close to $P^{(2)}(\sigma, \theta')$. If the two probability distributions are not close to each other, then most samples drawn from the θ -model will have a very low probability under the θ' -model and therefore make a negligible contribution to the sum in Eq. (D.4).

Fortunately, since we will be computing ratios of partition functions for model parameters drawn on consecutive iterations, θ_{t-1} and θ_t , and since our covariance matrix Σ tends to be very small, it is reasonable to assume that θ_{t-1} and θ_t are close enough to each other, such that Eq. (D.4) is applicable with relatively few ($M \approx 10^5$) samples. Therefore, the runtime at each MCMC iteration t will mostly consist of drawing M Monte Carlo samples from our previous model θ_{t-1} .

Appendix E: The Wang-Landau algorithm

The main idea behind the Wang-Landau algorithm [59] is to directly estimate the density of states $g(E)$, i.e., the number of states (configuration vectors) with a given energy E . Once we know the density of states, we can compute the partition function by grouping the terms in the sum by energy

$$Z(\theta) = \sum_{\sigma'} e^{-\beta H(\sigma', \theta)} = \sum_E g(E) e^{-E/T}. \quad (\text{E.1})$$

Other important thermodynamic quantities, such as the average energy, heat capacity, and entropy, follow from the partition function:

$$\langle E \rangle_{\theta} = \frac{1}{Z(\theta)} \sum_E g(E) E e^{-E/T}, \quad (\text{E.2})$$

$$C(T) = \frac{\langle E^2 \rangle_{\theta} - \langle E \rangle_{\theta}^2}{T^2}, \quad (\text{E.3})$$

$$S(T) = \frac{\langle E \rangle_{\theta} - F}{T} = \frac{\langle E \rangle_{\theta}}{T} + \ln Z(\theta), \quad (\text{E.4})$$

where $F = -T \ln Z(\theta)$ is the free energy. It is worth mentioning that $g(E)$ is independent of temperature. Hence we can compute the above quantities at any temperature T without rerunning the algorithm.

In order to estimate the density of states, the Wang-Landau algorithm performs a random walk in energy space and accepts the energy E associated with each configuration with a certain probability, designed to encourage the exploration of states with different energies [107]. For systems with continuous spectra, or systems with a lot of accessible energies (as is the case here), we begin by discretizing the spectrum into energy levels that are Δ apart. We assume that the energy spectrum is bounded, such that there is a finite number of energy levels. During the random walk, we keep a histogram $\rho(E)$ which is incremented each time we visit a state with energy E . The histogram has support only on the discretized, bounded spectrum. The random walk continues until the energy histogram $\rho(E)$ becomes flat, i.e. all its entries are within 20% of the mean value $\langle \rho(E) \rangle$. We check the flatness of the histogram every 10,000 steps.

The procedure during the random walk is as follows [107]:

1. Initialize $g(E) = 1$ and $\rho(E) = 0$ for all energies E .
2. At each iteration t :
 - Generate a new configuration by randomly flipping a state σ_i .
 - Let E_{t-1} and E' be the energies of the previous and current configurations respectively. Accept the new configuration and set $E_t = E'$ with probability

$$p_{\text{accept}} = \min \left(1, \frac{g(E_{t-1})}{g(E')} \right). \quad (\text{E.5})$$
- Otherwise, $E_t = E_{t-1}$.
- Update the density of states by a modification factor f , i.e. $g(E_t) \rightarrow f g(E_t)$.
3. If the histogram $\rho(E)$ is flat and $\ln f > \epsilon$, reduce the modification factor $f \rightarrow \sqrt{f}$, reset the histogram $\rho(E) = 0$, and proceed to step 2.

The modification factor f controls how well we approximate $g(E)$. Therefore, periodically decreasing f leads to a finer approximation of the density of states. Typical values for the parameters of the algorithm are $\Delta = 0.005$, $f = e = 2.7182$, and $\epsilon = 10^{-11}$.

- [1] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, 1991).
- [2] M. Diesmann, M.-O. Gewaltig, and A. Aertsen, *Nature* **402**, 529 EP (1999).
- [3] G. Kreiman, *Physics of Life Reviews* **1**, 71 (2004).
- [4] C. M. Gray and W. Singer, *Proceedings of the National Academy of Sciences* **86**, 1698 (1989).
- [5] M. Abeles, H. Bergman, E. Margalit, and E. Vaadia, *Journal of Neurophysiology* **70**, 1629 (1993).
- [6] A. Riehle, S. Grün, M. Diesmann, and A. Aertsen, *Science* **278**, 1950 (1997).
- [7] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, *Journal of Neuroscience* **26**, 8254 (2006).
- [8] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, *Nature* **454**, 995 EP (2008).
- [9] M. N. Shadlen and W. T. Newsome, *Journal of Neuroscience* **18**, 3870 (1998).
- [10] L. F. Abbott and P. Dayan, *Neural Computation* **11**, 91 (1999).
- [11] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, *Nature* **440**, 1007 (2006).
- [12] B. B. Averbeck, P. E. Latham, and A. Pouget, *Nature Reviews Neuroscience* **7**, 358 EP (2006).
- [13] K. Josić, E. Shea-Brown, B. Doiron, and J. de la Rocha, *Neural Computation* **21**, 2774 (2009).
- [14] P. K. Campbell, K. E. Jones, R. J. Huber, K. W. Horch, and R. A. Normann, *IEEE Transactions on Biomedical Engineering* **38**, 758 (1991).
- [15] K. E. Jones, P. K. Campbell, and R. A. Normann, *Annals of Biomedical Engineering* **20**, 423 (1992).
- [16] A. M. Litke, N. Bezayiff, E. J. Chichilnisky, W. Cunningham, W. Dabrowski, A. A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R. S. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, and A. Sher, *IEEE Transactions on Nuclear Science* **51**, 1434 (2004).
- [17] B. D. Allen, C. Moore-Kochlacs, J. G. Bernstein, J. Kinney, J. Scholvin, L. Seoane, C. Chronopoulos, C. Lamantia, S. B. Kodandaramaiah, M. Tegmark, and E. S. Boyden, *Journal of Neurophysiology* (2018), 10.1152/jn.00650.2017.
- [18] J. Shlens, G. D. Field, J. L. Gauthier, M. Greschner, A. Sher, A. M. Litke, and E. J. Chichilnisky, *Journal of Neuroscience* **29**, 5022 (2009).
- [19] S. Yu, D. Huang, W. Singer, and D. Nikolić, *Cerebral Cortex* **18**, 2891 (2008).
- [20] G. Tkačik, E. Schneidman, I. Berry, J. Michael, and W. Bialek, “Spin glass models for a network of real neurons,” (2009), [arXiv:0912.5409](https://arxiv.org/abs/0912.5409).
- [21] G. Tkačik, O. Marre, T. Mora, D. Amodei, M. J. B. II, and W. Bialek, *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P03011 (2013).
- [22] G. Tkačik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, and M. J. Berry, II, *PLOS Computational Biology* **10**, 1 (2014).
- [23] E. Ganmor, R. Segev, and E. Schneidman, *Proceedings of the National Academy of Sciences* **108**, 9679 (2011).
- [24] E. Ganmor, R. Segev, and E. Schneidman, *Journal of Neuroscience* **31**, 3044 (2011).
- [25] G. Tavoni, U. Ferrari, F. P. Battaglia, S. Cocco, and R. Monasson, *Network Neuroscience* **1**, 275 (2017).
- [26] B. Staude, S. Grün, and S. Rotter, *Frontiers in Computational Neuroscience* **4**, 16 (2010).
- [27] J. Zylberberg and E. Shea-Brown, *Phys. Rev. E* **92**, 062707 (2015).
- [28] F. Montani, R. A. A. Ince, R. Senatore, E. Arabzadeh, M. E. Diamond, and S. Panzeri, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **367**, 3297 (2009).
- [29] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, *Nature* **466**, 617 (2010).
- [30] J. H. Macke, M. Opper, and M. Bethge, *Phys. Rev. Lett.* **106**, 208102 (2011).
- [31] F. Montani, E. Phoka, M. Portesi, and S. R. Schultz, *Physica A: Statistical Mechanics and its Applications* **392**, 3066 (2013).
- [32] U. Koster, J. Sohl-Dickstein, C. M. Gray, and B. A. Olshausen, *PLOS Computational Biology* **10**, 1 (2014).
- [33] H. Shimazaki, K. Sadeghi, T. Ishikawa, Y. Ikegaya, and T. Toyozumi, *Scientific Reports* **5**, 9821 (2015).
- [34] D. A. Leen and E. Shea-Brown, *The Journal of Mathematical Neuroscience (JMN)* **5**, 17 (2015).
- [35] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia, *Neural Computation* **12**, 2621 (2000).
- [36] E. N. Brown, R. E. Kass, and P. P. Mitra, *Nature Neuroscience* **7**, 456 EP (2004).
- [37] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter, *PLOS Computational Biology* **7**, 1 (2011).
- [38] S. Yu, H. Yang, H. Nakahara, G. S. Santos, D. Nikolić, and D. Plenz, *Journal of Neuroscience* **31**, 17514 (2011).
- [39] Y. Roudi, S. Nirenberg, and P. E. Latham, *PLOS Computational Biology* **5**, 1 (2009).
- [40] Y. Roudi, E. Aurell, and J. A. Hertz, *Front Comput Neurosci* **3**, 22 (2009).
- [41] Y. Roudi, J. Tyrcha, and J. Hertz, *Phys. Rev. E* **79**, 051915 (2009).
- [42] V. Rostami, P. Porta Mana, S. Grün, and M. Helias, *PLOS Computational Biology* **13**, 1 (2017).
- [43] E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- [44] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, 2007).
- [45] H. C. Nguyen, R. Zecchina, and J. Berg, *Advances in Physics* **66**, 197 (2017).
- [46] T. Broderick, M. Dudik, G. Tkačik, R. E. Schapire, and W. Bialek, “Faster solutions of the inverse pairwise ising problem,” (2007), [arXiv:0712.2437](https://arxiv.org/abs/0712.2437).
- [47] J. Sohl-Dickstein, P. B. Battaglino, and M. R. DeWeese, *Phys. Rev. Lett.* **107**, 220601 (2011).
- [48] S. Cocco, S. Leibler, and R. Monasson, *Proceedings of the National Academy of Sciences* **106**, 14058 (2009).
- [49] S. Cocco and R. Monasson, *Phys. Rev. Lett.* **106**, 090601 (2011).
- [50] E. Aurell and M. Ekeberg, *Phys. Rev. Lett.* **108**, 090201 (2012).
- [51] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *The Journal of Chemical Physics* **21**, 1087 (1953).
- [52] W. K. Hastings, *Biometrika*, Vol. 57, No. 1, p. 97-109, 1970 **57**, 97 (1970).

- [53] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, *Proceedings of the National Academy of Sciences* **112**, 11508 (2015).
- [54] T. Mora, S. Deny, and O. Marre, *Phys. Rev. Lett.* **114**, 078105 (2015).
- [55] T. Mora and W. Bialek, *Journal of Statistical Physics* **144**, 268 (2011).
- [56] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
- [57] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, *Phys. Rev. Lett.* **91**, 238701 (2003).
- [58] E. W. Archer, I. M. Park, and J. W. Pillow, in *Advances in Neural Information Processing Systems 26* (Curran Associates, Inc., 2013) pp. 1700–1708.
- [59] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [60] M. L. Ioffe and M. J. Berry, II, *PLOS Computational Biology* **13**, 1 (2017).
- [61] T.-A. Nghiem, B. Telenczuk, O. Marre, A. Destexhe, and U. Ferrari, *Phys. Rev. E* **98**, 012402 (2018).
- [62] S. F. Edwards and P. W. Anderson, *Journal of Physics F: Metal Physics* **5**, 965 (1975).
- [63] R. A. Stefanescu and V. K. Jirsa, *PLOS Computational Biology* **4**, 1 (2008).
- [64] F. Amzica and M. Steriade, *Electroencephalography and Clinical Neurophysiology* **107**, 69 (1998).
- [65] A. Destexhe, D. Contreras, and M. Steriade, *Journal of Neuroscience* **19**, 4595 (1999).
- [66] M. Steriade, I. Timofeev, and F. Grenier, *Journal of Neurophysiology* **85**, 1969 (2001).
- [67] V. V. Vyazovskiy, U. Olcese, Y. M. Lazimy, U. Faraguna, S. K. Esser, J. C. Williams, C. Cirelli, and G. Tononi, *Neuron* **63**, 865 (2009).
- [68] A. Peyrache, N. Dehghani, E. N. Eskandar, J. R. Madsen, W. S. Anderson, J. A. Donoghue, L. R. Hochberg, E. Halgren, S. S. Cash, and A. Destexhe, *Proceedings of the National Academy of Sciences* **109**, 1731 (2012).
- [69] N. Dehghani, A. Peyrache, B. Telenczuk, M. Le Van Quyen, E. Halgren, S. S. Cash, N. G. Hatsopoulos, and A. Destexhe, *Scientific Reports* **6**, 23176 EP (2016).
- [70] M. Steriade, A. Nunez, and F. Amzica, *Journal of Neuroscience* **13**, 3266 (1993).
- [71] M. Steriade, *Neuronal substrates of sleep and epilepsy* (Cambridge University Press, 2003).
- [72] M. Le Van Quyen, L. E. Muller, B. Telenczuk, E. Halgren, S. Cash, N. G. Hatsopoulos, N. Dehghani, and A. Destexhe, *Proceedings of the National Academy of Sciences* **113**, 9363 (2016).
- [73] J. M. Beggs and D. Plenz, *Journal of Neuroscience* **23**, 11167 (2003).
- [74] D. R. Chialvo, *Nature Physics* **6**, 744 EP (2010).
- [75] A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, and A. S. Tolias, *Science* **327**, 584 (2010).
- [76] W. Softky and C. Koch, *Journal of Neuroscience* **13**, 334 (1993).
- [77] S. Yu, H. Yang, O. Shriki, and D. Plenz, *Frontiers in Systems Neuroscience* **7**, 42 (2013).
- [78] C. Haldeman and J. M. Beggs, *Phys. Rev. Lett.* **94**, 058101 (2005).
- [79] M. Tegmark, *Chaos, Solitons & Fractals* **76**, 238 (2015).
- [80] O. Kinouchi and M. Copelli, *Nature Physics* **2**, 348 EP (2006).
- [81] A. Renart, J. de la Rocha, P. Barthó, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, *Science* **327**, 587 (2010).
- [82] C. van Vreeswijk and H. Sompolinsky, *Science* **274**, 1724 (1996).
- [83] N. Dehghani, N. Hatsopoulos, Z. Haga, R. Parker, B. Greger, E. Halgren, S. Cash, and A. Destexhe, *Frontiers in Physiology* **3**, 302 (2012).
- [84] D. J. Schwab, I. Nemenman, and P. Mehta, *Phys. Rev. Lett.* **113**, 068102 (2014).
- [85] L. Aitchison, N. Corradi, and P. E. Latham, *PLOS Computational Biology* **12**, 1 (2016).
- [86] M. Nonnenmacher, C. Behrens, P. Berens, M. Bethge, and J. H. Macke, *PLOS Computational Biology* **13**, 1 (2017).
- [87] V. Priesemann, M. Wibral, M. Valderrama, R. Pröpper, M. Le Van Quyen, T. Geisel, J. Triesch, D. Nikolić, and M. H. J. Munk, *Frontiers in Systems Neuroscience* **8**, 108 (2014).
- [88] J. Touboul and A. Destexhe, *Phys. Rev. E* **95**, 012413 (2017).
- [89] I. Mastromatteo and M. Marsili, *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P10012 (2011).
- [90] M. Marsili, I. Mastromatteo, and Y. Roudi, *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P09003 (2013).
- [91] J. Tyrcha, Y. Roudi, M. Marsili, and J. Hertz, *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P03005 (2013).
- [92] A. Haimovici and M. Marsili, *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P10013 (2015).
- [93] S.-i. Amari, H. Nakahara, S. Wu, and Y. Sakai, *Neural Computation* **15**, 127 (2003).
- [94] C. Gardella, O. Marre, and T. Mora, *eNeuro* **3** (2016), 10.1523/ENEURO.0160-15.2016.
- [95] C. ODonnell, J. T. Goncalves, N. Whiteley, C. Portera-Cailliau, and T. J. Sejnowski, *Neural Computation* **29**, 50 (2017).
- [96] H. Huang, *Journal of Statistical Mechanics: Theory and Experiment* **2017**, 033501 (2017).
- [97] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs, *Journal of Neuroscience* **28**, 505 (2008).
- [98] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe, *Phys. Rev. Lett.* **102**, 138101 (2009).
- [99] J. Vasquez, O. Marre, A. Palacios, M. Berry, and B. Cessac, *Journal of Physiology-Paris* **106**, 120 (2012).
- [100] D. A. McCormick, B. W. Connors, J. W. Lighthall, and D. A. Prince, *Journal of Neurophysiology* **54**, 782 (1985).
- [101] P. Barthó, H. Hirase, L. Monconduit, M. Zugaro, K. D. Harris, and G. Buzsáki, *Journal of Neurophysiology* **92**, 600 (2004).
- [102] S. Fujisawa, A. Amarasingham, M. T. Harrison, and G. Buzsáki, *Nature Neuroscience* **11**, 823 EP (2008).
- [103] C. Andrieu and J. Thoms, *Statistics and computing* **18**, 343 (2008).
- [104] H. Haario, E. Saksman, and J. Tamminen, *Computational Statistics* **14**, 375 (1999).
- [105] H. Haario, E. Saksman, and J. Tamminen, *Bernoulli* **7**, 223 (2001).
- [106] A. Gelman, G. O. Roberts, W. R. Gilks, *et al.*, *Bayesian statistics* **5**, 42 (1996).

- [107] D. P. Landau, S.-H. Tsai, and M. Exler, *American Journal of Physics* **72**, 1294 (2004).