

Exactly solvable statistical physics models for large neuronal populations

Christopher W. Lynn^{1,2,3,*}, Qiwei Yu⁴, Rich Pang⁵, William Bialek^{3,4,6} and Stephanie E. Palmer^{7,8}¹Department of Physics, Quantitative Biology Institute, and Wu Tsai Institute, Yale University, New Haven, Connecticut 06520, USA²Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, New York, New York 10016, USA³Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA⁴Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA⁵Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA⁶Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10065, USA⁷Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA⁸Department of Physics, University of Chicago, Chicago, Illinois 60637, USA

(Received 18 December 2023; revised 27 June 2024; accepted 28 March 2025; published 19 May 2025)

Maximum-entropy methods provide a principled path connecting measurements of neural activity directly to statistical physics models, and this approach has been successful for populations of $N \sim 100$ neurons. As N increases in new experiments, we enter an undersampled regime where we have to choose which observables should be constrained in the maximum-entropy construction. The best choice is the one that provides the greatest reduction in entropy, defining a “minimax entropy” principle. This principle becomes tractable if we restrict attention to correlations among pairs of neurons that link together into a tree; we can find the best tree efficiently, and the underlying statistical physics models are exactly solved. We use this approach to analyze experiments on $N \sim 1500$ neurons in the mouse hippocampus, and we find that the resulting model captures key features of collective activity in the network.

DOI: [10.1103/PhysRevResearch.7.L022039](https://doi.org/10.1103/PhysRevResearch.7.L022039)

It has long been hoped that neural networks in the brain could be described using concepts from statistical physics [1–7]. More recently, our ability to explore the brain has been revolutionized by experimental techniques that record the activity from thousands of individual neurons, simultaneously [8–14]. Maximum-entropy methods connect these data to theory, starting with measured properties of the network and arriving at models that are mathematically equivalent to statistical physics problems [15,16]. In some cases, these models provide successful, parameter-free predictions for many detailed features of the neural activity patterns [7,17,18]. The same ideas have been used in contexts ranging from the evolution of protein families to ordering in flocks of birds [19–25]. But as experiments probe systems with more and more degrees of freedom, the number of samples that we can collect typically does not increase in proportion. Here we present a strategy for building maximum entropy models in this undersampled regime, and we apply this strategy to data from 1000+ neurons in the mouse hippocampus.

Consider a system of N variables $\mathbf{x} = \{x_i\}$, $i = 1, 2, \dots, N$. To describe the system, we would like to write down the probability distribution $P(\mathbf{x})$ over these microscopic degrees

of freedom, in the same way that we write the Boltzmann distribution for a system in equilibrium. In the example that we discuss below, all the x_i are observed at the same moment in time, but we could also include time in the index i , so that $P(\mathbf{x})$ becomes a probability distribution of trajectories.

From these N variables, we can construct a set of K operators or observables $\{f_v(\mathbf{x})\}$, $v = 1, 2, \dots, K$. The maximum-entropy approach takes this limited number of observables seriously, and insists that expectation values for these quantities predicted by the model match the values measured in experiment,

$$\langle f_v(\mathbf{x}) \rangle_P = \langle f_v(\mathbf{x}) \rangle_{\text{exp}}, \quad (1)$$

or more explicitly,

$$\sum_{\mathbf{x}} P(\mathbf{x}) f_v(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M f_v(\mathbf{x}^{(m)}), \quad (2)$$

where $\mathbf{x}^{(m)}$ is the m th sample out of M samples in total. Notice that to have control over errors in the measurement of all K expectation values, we need to have $K \ll MN$.

There are infinitely many distributions that obey these matching conditions, but the idea of maximum entropy is that we should choose the one that has the least possible structure, or equivalently generates samples that are as random as possible while obeying the constraints in Eq. (1). From Shannon we know that “as random as possible” translates uniquely to finding the distribution that has the maximum entropy consistent with the constraints [26,27]. The solution

*Contact author: christopher.lynn@yale.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

of this optimization problem has the form

$$P(\mathbf{x}) = \frac{1}{Z} \exp \left[- \sum_{v=1}^K \lambda_v f_v(\mathbf{x}) \right], \quad (3)$$

where the coupling constants λ_v must be chosen to satisfy Eq. (1). We emphasize that in using this approach, the model in Eq. (3) is something that needs to be tested—in systems such as networks of neurons, there is no H-theorem telling us that entropy will be maximized, nor is there a unique choice for the constraints that would correspond to the Hamiltonian of an equilibrium system.

Without a Hamiltonian, how should we choose the observables f_v ? Probability distributions define a code for the data [26] in which each state $\mathbf{x}$ is mapped to a code word of length

$$\ell(\mathbf{x}) = -\log P(\mathbf{x}), \quad (4)$$

so the mean code length for the data is

$$\langle \ell \rangle_{\text{exp}} \equiv -\frac{1}{M} \sum_{n=1}^M \log P(\mathbf{x}^{(n)}). \quad (5)$$

Combining Eqs. (1) and (3), $\langle \ell \rangle_{\text{exp}}$ is exactly the entropy of P . Thus, among all maximum entropy distributions, the one that gives the shortest description of the data is the one with minimum entropy. This “minimax entropy” principle was discussed 25 years ago [28], but has attracted relatively little attention.

Every time we add a constraint, the maximum possible entropy is reduced [29], and this entropy reduction is an information gain. Thus the minimax entropy principle tells us to choose observables whose expectation values provide as much information as possible about the microscopic variables. The problem is that finding these maximally informative observables is generally intractable: One must search over all possible sets of observables, and for each set one must solve the maximum-entropy problem for the parameters λ_v .

To make progress, we restrict the class of observables that we consider. With populations of $N \sim 100$ neurons, it can be very effective to constrain just the mean activity of each neuron and the correlations among pairs [16–18]. But this corresponds to $K \propto N^2$ constraints, and at large N we will violate the good sampling condition $K \ll NM$. To restore good sampling, we try constraining only N_c of the correlations, which define links between specific pairs of neurons that together form a graph $\mathcal{G}$. If we describe the individual neurons as being either active or silent, so that $x_i \in \{0, 1\}$, then the maximum entropy distribution is an Ising model on the graph $\mathcal{G}$,

$$P_{\mathcal{G}}(\mathbf{x}) = \frac{1}{Z} \exp \left[\sum_{(ij) \in \mathcal{G}} J_{ij} x_i x_j + \sum_i h_i x_i \right], \quad (6)$$

where $\{h_i, J_{ij}\}$ must be adjusted to match the measured expectation values $\langle x_i \rangle_{\text{exp}}$ and $\langle x_i x_j \rangle_{\text{exp}}$ for pairs $(ij) \in \mathcal{G}$. The minimax entropy principle tells us that we should find the graph $\mathcal{G}$ with a fixed number of links N_c such that $P_{\mathcal{G}}(\mathbf{x})$ has the smallest entropy while obeying these constraints. Even this problem remains intractable.

Calculations in statistical physics are hard due to feedback loops. Without loops, one can often make exact and efficient

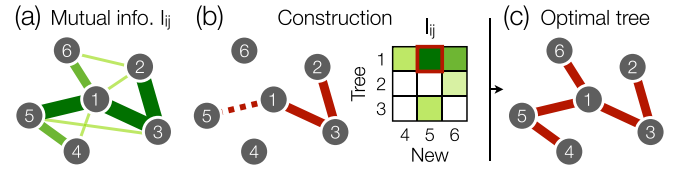


FIG. 1. Computing the optimal tree of pairwise correlations. (a) Mutual information between variables in a hypothetical system. (b) One can build the optimal tree one variable at a time, from an arbitrary starting variable, by iteratively adding the connection corresponding to the largest mutual information (dashed) from the tree (solid) to the remaining variables; this is Prim’s algorithm. (c) Optimal tree that minimizes the entropy $S_{\mathcal{T}}$ and maximizes the information $I_{\mathcal{T}}$.

calculations, as in one-dimensional systems or on Bethe lattices [30]. If the graph $\mathcal{G}$ has no loops, then it describes a tree $\mathcal{T}$, and we can analytically solve the maximum-entropy problem for the fields h_i and couplings J_{ij} as functions of the observables $\langle x_i \rangle$ and $\langle x_i x_j \rangle$,

$$J_{ij} = \ln \left[\frac{\langle x_i x_j \rangle (1 - \langle x_i \rangle - \langle x_j \rangle + \langle x_i x_j \rangle)}{(\langle x_i \rangle - \langle x_i x_j \rangle)(\langle x_j \rangle - \langle x_i x_j \rangle)} \right], \quad (7)$$

$$h_i = \ln \frac{\langle x_i \rangle}{1 - \langle x_i \rangle} + \sum_{j \in \mathcal{N}_i} \ln \left[\frac{(1 - \langle x_i \rangle)(\langle x_i \rangle - \langle x_i x_j \rangle)}{\langle x_i \rangle (1 - \langle x_i \rangle - \langle x_j \rangle + \langle x_i x_j \rangle)} \right], \quad (8)$$

where $\mathcal{N}_i$ are the neighbors of i on the tree [15,31]. Moreover, we can write the entropy of the model $P_{\mathcal{T}}$ as

$$S_{\mathcal{T}} = - \sum_{\mathbf{x}} P_{\mathcal{T}}(\mathbf{x}) \log P_{\mathcal{T}}(\mathbf{x}) = S_{\text{ind}} - \sum_{(ij) \in \mathcal{T}} I_{ij}, \quad (9)$$

where S_{ind} is the independent entropy of the individual variables, and I_{ij} is the mutual information between x_i and x_j [Fig. 1(a)]. Due to the maximum-entropy constraints, the mutual information I_{ij} in the model matches that in the data for $(ij) \in \mathcal{T}$. Thus, we can compute the entropy $S_{\mathcal{T}}$ of any maximum-entropy model on a tree without constructing the model itself.

To solve the minimax entropy problem, we want to find the tree $\mathcal{T}$ that minimizes the entropy $S_{\mathcal{T}}$ in Eq. (9), or, equivalently, maximizes the total mutual information

$$I_{\mathcal{T}} = \sum_{(ij) \in \mathcal{T}} I_{ij}. \quad (10)$$

We have therefore simplified the minimax entropy problem to a minimum spanning tree problem [15], which admits a number of efficient solutions. For example, rather than searching over all N^{N-2} possible trees, one can use Prim’s algorithm (Fig. 1) to find the tree $\mathcal{T}$ with maximum information $I_{\mathcal{T}}$ in $O(N^2)$ time [32]. This means that if the data are generated by an Ising model on a tree, then we are guaranteed to find the true interactions and our model is exact [33]. Thus, by restricting observables to pairwise correlations that form a tree, we can exactly solve not only the maximum-entropy problem [using Eqs. (7) and (8)], but also the minimax entropy problem for the optimal tree $\mathcal{T}$, even at very large N .

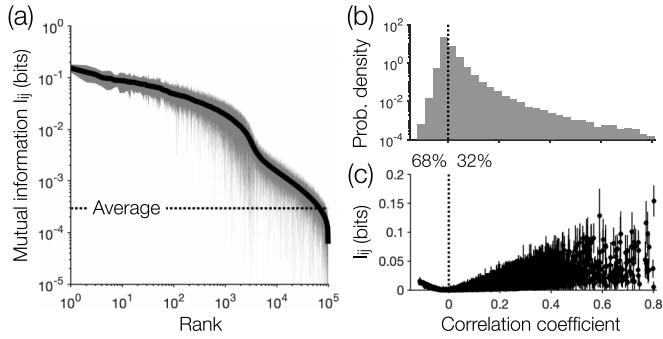


FIG. 2. Mutual information in a large population of neurons. (a) Ranked order of the significant mutual information values I_{ij} in a population of $N = 1485$ neurons in the mouse hippocampus [34]. Solid line and shaded region indicate estimates and errors (two standard deviations) of I_{ij} . (b) Distribution of correlation coefficients over neuron pairs, with percentages indicating the proportion of positively and negatively correlated pairs. (c) Mutual information I_{ij} vs correlation coefficient, where each point represents a distinct neuron pair. Estimates and errors are the same as in (a).

As emphasized above, our approach yields a model that needs to be tested. While maximum-entropy models traditionally constrain all $N(N - 1)/2$ pairwise correlations $\langle x_i x_j \rangle$ [16–18], here we only constrain the $N - 1$ corresponding to $(ij) \in \mathcal{T}$. At large N , this sparsity guarantees that we are well-sampled, but it also means that we only have access to a vanishingly small fraction $2/N$ of all pairwise correlations. Can such a sparse backbone of correlations contain enough information to capture features of the collective activity in the population?

To answer this question, we analyze calcium imaging data from CA1 in the mouse hippocampus as the animal runs along a virtual track [34]. Discussions of the hippocampus often focus on place cells, yet only about 30% of pyramidal cells in CA1 exhibit significant place fields in a single environment [35–38]. While individual place cells can be viewed as responding independently to external cues, to understand the collective activity of a mixed population one must consider interactions between neurons. This approach has been widely employed to study collective hippocampal functions including memory storage, pattern separation, and pattern completion [39–41]. Moreover, previous applications of the maximum-entropy principle in populations of $N \sim 100$ neurons have shown that the detailed patterns of activity can sometimes be explained entirely by the pairwise correlations between neurons [17,18].

Here, we consider calcium imaging from a population of $N = 1485$ neurons with images collected at 30 Hz for 39 min, yielding $M \approx 7 \times 10^4$ samples. The signal from each cell naturally binarizes into active/silent ($x_i = 1/0$) in each video frame [17], and we have sufficient data to estimate the mutual information I_{ij} with small errors [33]. Among all $\sim 10^6$ pairs of neurons, only 9% exhibit significant mutual information I_{ij} [Fig. 2(a); see [33] for details]. We see that the distribution of mutual information is heavy-tailed, which tells us that a small number of correlations contain orders of magnitude more information than average ($\bar{I} = 2.9 \times 10^{-4}$ bits). Additionally, while most pairs of neurons are negatively correlated

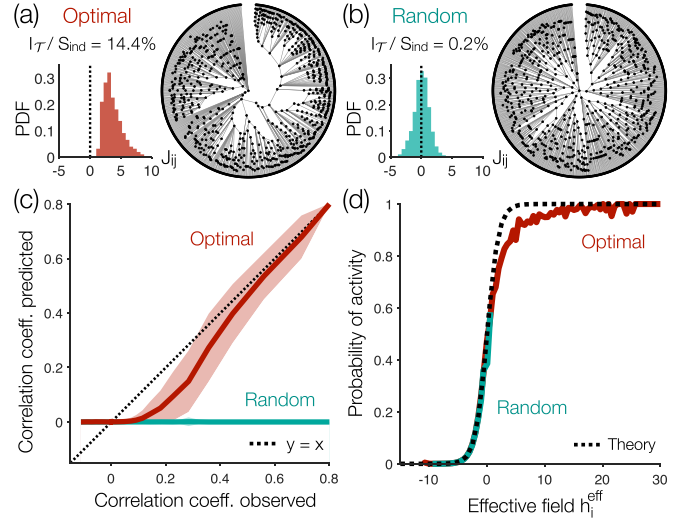


FIG. 3. Minimax entropy model of a large neuronal population. (a) Optimal tree for the neurons in Fig. 2 and distribution over interactions J_{ij} in the induced maximum-entropy model. The central neuron has the most connections, and those on the perimeter are leaves (with one connection) with distance from the center decreasing in the clockwise direction. (b) Same as (a) for a random tree. (c) Correlation coefficients predicted in the models vs those in the data, with the dashed line indicating equality. All pairs are divided evenly into bins along the x -axis, with solid lines and shaded regions reflecting means and standard deviations within bins. (d) Probability of the neuron to be active $x_i = 1$ vs effective field h_i^{eff} . Solid lines are computed over all neurons and all time points, and the dashed line indicates the theoretical prediction in Eq. (11).

[Fig. 2(b)], the strongest values of I_{ij} belong to pairs that are positively correlated [Fig. 2(c)], possibly reflecting the excitatory nature of pyramidal cells. Together, these observations suggest that a sparse network of positively correlated neurons may provide a large amount of information about the collective neural activity.

Applying our method, we identify the tree of maximally informative correlations [Fig. 3(a)], which captures $S_{\text{ind}} - S_{\mathcal{T}} = I_{\mathcal{T}} = 26.2$ bits of information; this is more than 50 times the average we find on a random tree [Fig. 3(b)]. Despite including only $2/N \approx 0.1\%$ of all pairwise correlations, we find that the optimal tree reduces the entropy of independent neurons by $I_{\mathcal{T}}/S_{\text{ind}} \approx 14.4\%$; this reduction in entropy is as strong as freezing the states of 214 randomly selected neurons.

In the model of Eq. (6), positive (negative) J_{ij} means that activity in neuron i leads to activity (silence) in neuron j . For random trees, the interactions J_{ij} are split almost evenly between positive and negative [Fig. 3(b)]; this is consistent with previous investigations of fully connected modes in smaller systems of $N \sim 100$ neurons [16,18,42]. But since the largest mutual information is associated with positive correlations [Fig. 2(c)], the optimal tree produces a model with strong interactions that are almost exclusively positive [Fig. 3(a)]. We have arrived at an Ising ferromagnet, with excitatory interactions similar to those between pyramidal neurons.

While the model $P_{\mathcal{T}}$ is defined to match a sparse subset of the observed correlations for $(ij) \in \mathcal{T}$, we can ask what $P_{\mathcal{T}}$ predicts for all pairs of neurons. The optimal tree

underestimates the strength of weak correlations, yet accurately predicts the strong correlations in the population [Fig. 3(c)]. Meanwhile, a typical set of correlations (that is, a random tree) captures almost none of the overall correlation structure in the data [Fig. 3(c)].

At the level of individual neurons, the model makes a clear prediction for the probability that a neuron spikes in response to the activity of the rest of the network:

$$P_T(x_i = 1 | \mathbf{x}_{-i}) = [1 + \exp(-h_i^{\text{eff}}(\mathbf{x}_{-i}))]^{-1}, \quad (11)$$

where $\mathbf{x}_{-i}$ represents the states of all neurons other than i , and $h_i^{\text{eff}}(\mathbf{x}_{-i}) = h_i + \sum_{j \in \mathcal{N}_i} J_{ij} x_j$ is an effective field. In a tree, each neuron only interacts with two other neurons in the population (on average). Although this sparsity restricts the range of predictions for random trees, the optimal tree still exhibits good agreement between theory and data across a wide range of spike probabilities [Fig. 3(d)]. These results suggest that a large amount of information about the neural activity is packed into a sparse network of strong correlations.

To interpret these results, we apply our framework to a canonical model of complex interacting systems: the Sherrington-Kirkpatrick (SK) model, which is equivalent to an Ising system with all-to-all interactions J_{ij} drawn from a zero-mean Gaussian [43]. We can investigate the impact of sparsity in the network by randomly removing interactions, resulting in a connection density d [Fig. 4(a)]. Simulating the original SK model ($d = 1$) with $N = 1485$ artificial neurons and $M \approx 7 \times 10^4$ samples to match the neural data, we find that the optimal tree captures $I_T/S_{\text{ind}} \approx 4.2\%$ of the independent entropy [Fig. 4(b)]; by contrast, a random tree only captures 0.5%. As we remove interactions from the system (decreasing d), the optimal tree uncovers a larger proportion of the interactions, thereby capturing more information about correlations in the system [Fig. 4(b)]; meanwhile, random trees become even less effective.

For the optimal tree to achieve the fraction $I_T/S_{\text{ind}} \approx 14.4\%$ observed in data [Fig. 3(a)], one must remove 99% of the interactions from the SK model ($d = 0.01$). For $N = 1485$, this corresponds to each neuron interacting with an average of only 15 other cells in the population. At $d = 0.01$, while the optimal tree fails to predict weak correlations, it captures both the strong positive and negative correlations in the system [Fig. 4(c)]. Moreover, it provides near-perfect predictions for the spiking probabilities of individual artificial neurons [Fig. 4(d)]. We therefore find that our minimax entropy framework becomes more accurate—resembling the neural data in Fig. 3—for sparse networks of interactions.

Thus far, we have focused on features of individual neurons and neuron pairs, but can our approach predict something about the activity of the population as a whole? One key measure of collective neural activity is the probability $P(k)$ that k out of the N neurons are simultaneously active within a window of time [16,18,25,42]. In the data, we see moments of both synchrony (large k) and silence (small k) that are much more likely than expected from independent neurons (Fig. 5). The optimal tree captures most of this collective behavior, correctly predicting ~ 100 -fold increases in the probability that $k \gtrsim 50$ neurons are active at once (Fig. 5, red). By contrast, a random selection of correlations yields predictions that are

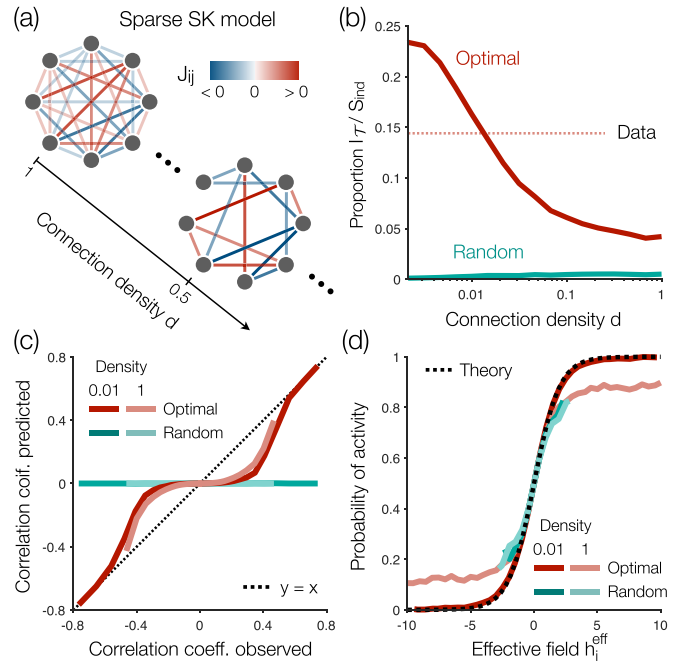


FIG. 4. Minimax entropy model of artificial neurons. (a) Sparse SK model defined by Eq. (6) with all-to-all interactions J_{ij} drawn from a zero-mean Gaussian with standard deviation $4/\sqrt{N}$ and fields $h_i = -\frac{1}{2} \sum_j J_{ij}$; this is equivalent to an Ising model with $x_i = \pm 1$, zero fields, and Gaussian interactions with standard deviation $1/\sqrt{N}$. We introduce sparsity by randomly removing $1 - d$ of the interactions and rescaling the strength of interactions to maintain a constant standard deviation. (b) Fractions of independent entropy I_T/S_{ind} captured by optimal and random trees as functions of the connection density d . Each point is averaged over 10 Monte Carlo simulations of the SK model with $N = 1485$ neurons and $M \approx 7 \times 10^4$ samples to match the data. Dashed line indicates the value of the optimal tree in neural data $I_T/S_{\text{ind}} \approx 14.4\%$. (c) Correlation coefficients predicted by optimal and random trees vs those observed in the SK model. (d) Probability of $x_i = 1$ vs effective field h_i^{eff} , with the dashed line indicating the theoretical prediction in Eq. (11). In (c) and (d), each line represents a single Monte Carlo simulation.

indistinguishable from independent neurons (Fig. 5, cyan). While the detailed patterns of neural activity are likely shaped by a web of weaker correlations, these results show that a sparse backbone of the strongest correlations can still produce the large-scale synchrony observed in neural systems.

In summary, it has been appreciated for nearly two decades that the maximum-entropy principle provides a link from data directly to statistical physics models, and this is useful in networks of neurons as well as other complex systems [15–25]. Less widely emphasized is that we do not have “the” maximum-entropy model, but rather a collection of possible models depending on what features of the system we choose to constrain. Quite generally, we should choose the features that are most informative, leading to the minimax entropy principle [28]. As we study larger and larger populations of neurons, we enter an undersampled regime in which selecting a limited number of maximally informative features is not only conceptually appealing but also a practical necessity.

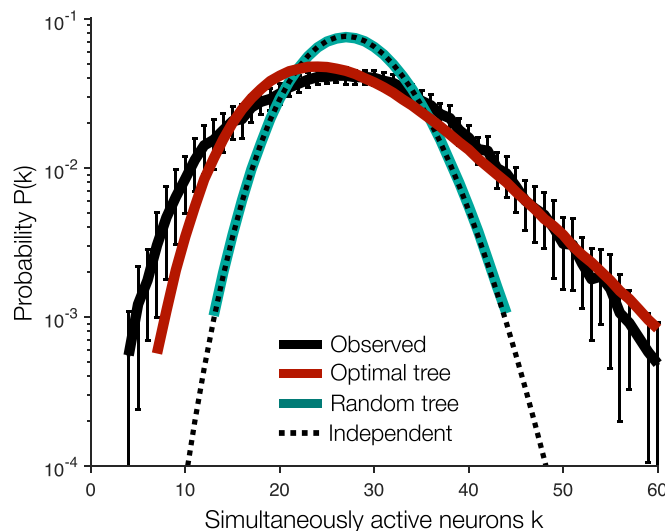


FIG. 5. Predicting synchronous activity. Distribution $P(k)$ of the number of simultaneously active neurons k in the data (black) and predicted by the optimal tree (red), a random tree (cyan), and independent neurons (dashed). To estimate probabilities $P(k)$ and error bars (two standard deviations), we first split the experiment into 1-min blocks to preserve dependencies between consecutive samples. We then randomly select one-third of these blocks and repeat 100 times. For each subsample of the data, we compute model predictions for $P(k)$ using Monte Carlo simulations with the same number of samples as the data.

The problem is that the minimax entropy principle is intractable in general. Here we have made progress in two steps. First, following previous successes, we focus on

constraining the mean activity and pairwise correlations. Second, we take the lesson of the Bethe lattice and select only pairs that define a tree. Once we do this, the relevant statistical mechanics problem is solved exactly, and the optimal tree can be found in quadratic time [15,31]. This means that there is a family of strongly interacting statistical physics models for thousands of neurons that we can construct in seconds on a personal computer. It is worth asking whether these models can capture any of the essential collective behavior in real networks. While our approach has important limitations (see an extended discussion in [33]), we find that the optimal tree correctly predicts strong correlations, individual spike probabilities, and the distribution of synchronous activity (Figs. 3 and 5). The key to this success is the heavy-tailed distribution of mutual information [Fig. 2(a)], and this in turn may be grounded in sparse networks of strong physical connections [44]. While these models cannot capture all aspects of collective behavior, our approach provides at least a starting point for scalable models of the much larger systems now becoming accessible to experiments.

We thank L. Meshulam and J. L. Gauthier for guiding us through the data of Ref. [34], and C. M. Holmes and D. J. Schwab for helpful discussions. This work was supported in part by the National Science Foundation through the Center for the Physics of Biological Function (PHY-1734030), by the National Institutes of Health through the BRAIN initiative (R01EB026943), by the James S. McDonnell Foundation through a Postdoctoral Fellowship Award (C.W.L.), and by Fellowships from the Simons Foundation and the John Simon Guggenheim Memorial Foundation (W.B.).

- [1] N. Wiener, *Nonlinear Problems in Random Theory* (MIT Press, Cambridge, MA, 1958).
- [2] L. N. Cooper, A possible organization of animal memory and learning, in *Collective Properties of Physical Systems: Proceedings of Nobel Symposium 24*, edited by B. Lundqvist and S. Lundqvist (Academic, New York, 1973), pp. 252–264.
- [3] W. A. Little, The existence of persistent states in the brain, *Math. Biosci.* **19**, 101 (1974).
- [4] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [5] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, 1989).
- [6] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).
- [7] L. Meshulam and W. Bialek, Statistical mechanics for networks of real neurons, [arXiv:2409.00412](https://arxiv.org/abs/2409.00412).
- [8] R. Segev, J. Goodhouse, J. Puchalla, and M. J. Berry II, Recording spikes from a large fraction of the ganglion cells in a retinal patch, *Nat. Neurosci.* **7**, 1155 (2004).
- [9] A. M. Litke, N. Bezayiff, E. J. Chichilnisky, W. Cunningham, W. Dabrowski, A. A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine *et al.*, What does the eye tell the brain? Development of a system for the large-scale recording of retinal output activity, *IEEE Trans. Nucl. Sci.* **51**, 1434 (2004).
- [10] J. E. Chung, H. R. Joo, J. L. Fan, D. F. Liu, A. H. Barnett, S. Chen, C. Geaghan-Breiner, M. P. Karlsson, M. Karlsson, K. Y. Lee *et al.*, High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays, *Neuron* **101**, 21 (2019).
- [11] D. A. Dombeck, C. D. Harvey, L. Tian, L. L. Looger, and D. W. Tank, Functional imaging of hippocampal place cells at cellular resolution during virtual navigation, *Nat. Neurosci.* **13**, 1433 (2010).
- [12] L. Tian, J. Akerboom, E. R. Schreier, and L. L. Looger, Neural activity imaging with genetically encoded calcium indicators, *Prog. Brain Res.* **196**, 79 (2012).
- [13] J. Demas, J. Manley, F. Tejera, K. Barber, H. Kim, F. M. Traub, B. Chen, and A. Vaziri, High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy, *Nat. Methods* **18**, 1103 (2021).
- [14] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux *et al.*, Neuropixels 2.0: A miniaturized high-density probe

- for stable, long-term brain recordings, *Science* **372**, eabf4588 (2021).
- [15] H. C. Nguyen, R. Zecchina, and J. Berg, Inverse statistical problems: From the inverse Ising problem to data science, *Adv. Phys.* **66**, 197 (2017).
- [16] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature (London)* **440**, 1007 (2006).
- [17] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Collective behavior of place and non-place neurons in the hippocampal network, *Neuron* **96**, 1178 (2017).
- [18] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Successes and failures of simplified models for a network of real neurons, *arXiv:2112.14735*.
- [19] T. R. Lezon, J. R. Banavar, M. Cieplak, A. Maritan, and N. V. Fedoroff, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns, *Proc. Natl. Acad. Sci. USA* **103**, 19033 (2006).
- [20] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing, *Proc. Natl. Acad. Sci. USA* **106**, 67 (2009).
- [21] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, Protein 3D structure computed from evolutionary sequence variation, *PLoS One* **6**, e28766 (2011).
- [22] A. Lapedes, B. Giraud, and C. Jarzynski, Using sequence alignments to predict protein structure and stability with high accuracy, *arXiv:1207.2484*.
- [23] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci. USA* **109**, 4786 (2012).
- [24] W. P. Russ, M. Figliuzzi, C. Stocker, P. Barrat-Charlaix, M. Socolich, P. Kast, D. Hilvert, R. Monasson, S. Cocco, M. Weigt *et al.*, An evolution-based model for designing chorismate mutase enzymes, *Science* **369**, 440 (2020).
- [25] C. W. Lynn, L. Papadopoulos, D. D. Lee, and D. S. Bassett, Surges of collective human activity emerge from simple pairwise correlations, *Phys. Rev. X* **9**, 011022 (2019).
- [26] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [27] E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**, 620 (1957).
- [28] S. C. Zhu, Y. N. Wu, and D. Mumford, Minimax entropy principle and its application to texture modeling, *Neural Comput.* **9**, 1627 (1997).
- [29] E. Schneidman, S. Still, M. J. Berry II, and W. Bialek, Network information and connected correlations, *Phys. Rev. Lett.* **91**, 238701 (2003).
- [30] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, USA, 2021), Vol. 14.
- [31] C. Chow and C. Liu, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Theor.* **14**, 462 (1968).
- [32] C. Moore and S. Mertens, *The Nature of Computation* (Oxford University Press, Oxford, 2011).
- [33] C. W. Lynn, Q. Yu, R. Pang, S. E. Palmer, and W. Bialek, companion paper, Exact minimax entropy models of large-scale neuronal activity, *Phys. Rev. E* **111**, 054411 (2025).
- [34] J. L. Gauthier and D. W. Tank, A dedicated population for reward coding in the hippocampus, *Neuron* **99**, 179 (2018).
- [35] J. O’Keefe and D. H. Conway, Hippocampal place units in the freely moving rat: Why they fire where they fire, *Exp. Brain Res.* **31**, 573 (1978).
- [36] J. O’Keefe and L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford University Press, 1978).
- [37] L. T. Thompson and P. J. Best, Place cells and silent cells in the hippocampus of freely-behaving rats, *J. Neurosci.* **9**, 2382 (1989).
- [38] S. Leutgeb, J. K. Leutgeb, A. Treves, M.-B. Moser, and E. I. Moser, Distinct ensemble codes in hippocampal areas CA3 and CA1, *Science* **305**, 1295 (2004).
- [39] M. C. Fuhs and D. S. Touretzky, A spin glass model of path integration in rat medial entorhinal cortex, *J. Neurosci.* **26**, 4266 (2006).
- [40] S. Káli and P. Dayan, The involvement of recurrent connections in area CA3 in establishing the properties of place fields: A model, *J. Neurosci.* **20**, 7463 (2000).
- [41] A. Samsonovich and B. L. McNaughton, Path integration and cognitive mapping in a continuous attractor neural network model, *J. Neurosci.* **17**, 5900 (1997).
- [42] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry II, and W. Bialek, Thermodynamics and signatures of criticality in a network of neurons, *Proc. Natl. Acad. Sci. USA* **112**, 11508 (2015).
- [43] D. Sherrington and S. Kirkpatrick, Solvable model of a spin-glass, *Phys. Rev. Lett.* **35**, 1792 (1975).
- [44] C. W. Lynn, C. M. Holmes, and S. E. Palmer, Heavy-tailed neuronal connectivity arises from Hebbian self-organization, *Nat. Phys.* **20**, 484 (2024).