Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The simplest maximum entropy model for collective behavior in a neural network

Gašper Tkačik[1], Olivier Marre[2,3], Thierry Mora[4], Dario Amodei[5,6], Michael J Berry II[3,6] and William Bialek[5,7]

[1] Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria
[2] Institut de la Vision, UMRS 968 UPMC, INSERM, CNRS U7210, CHNO Quinze-Vingts, F-75012 Paris, France
[3] Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[4] Laboratoire de Physique Statistique de l'École Normale Superieure, CNRS and Universites Paris VI and Paris VII, 24 rue Lhomond, 75231 Paris Cedex 05, France
[5] Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544, USA
[6] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
[7] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
E-mail: gtkacik@ist.ac.at, olivier.marre@gmail.com, tmora@lps.ens.fr, damodei@gmail.com, berry@princeton.edu and wbialek@princeton.edu

Received 22 August 2012
Accepted 30 October 2012
Published 12 March 2013

Online at stacks.iop.org/JSTAT/2013/P03011
doi:10.1088/1742-5468/2013/03/P03011

Abstract. Recent work emphasizes that the maximum entropy principle provides a bridge between statistical mechanics models for collective behavior in neural networks and experiments on networks of real neurons. Most of this work has focused on capturing the measured correlations among pairs of neurons. Here we suggest an alternative, constructing models that are consistent with the distribution of global network activity, i.e. the probability that K out of N cells in the network generate action potentials in the same small time bin. The inverse problem that we need to solve in constructing the model is analytically tractable, and provides a natural 'thermodynamics' for the network in the limit of large N.

1742-5468/13/P03011+10$33.00

J. Stat. Mech. (2013) P03011

We analyze the responses of neurons in a small patch of the retina to naturalistic stimuli, and find that the implied thermodynamics is very close to an unusual critical point, in which the entropy (in proper units) is exactly equal to the energy.

Keywords: phase diagrams (theory), neuronal networks (experiment), neuronal networks (theory)

ArXiv ePrint: 1207.6319

Many of the most interesting phenomena of life are collective, emerging from interactions among many elements, and physicists have long hoped that these collective biological phenomena could be described within the framework of statistical mechanics. One approach to a statistical mechanics of biological systems is exemplified by Hopfield's discussion of neural networks, in which simplifying assumptions about the underlying dynamics led to an effective 'energy landscape' on the space of network states [1]–[3]. In a similar spirit, Toner and Tu showed that simple stochastic dynamical models for coordinating the motion of moving organisms, as in flocks of birds or schools of fish, can be mapped to an effective field theory in the hydrodynamic limit [4, 5].

A very different way of constructing a statistical mechanics for real biological systems is through the maximum entropy principle [6]. Rather than making specific assumptions about the underlying dynamics, we take a relatively small set of measurements on the system as given, and build a model for the distribution over system states that is consistent with these experimental results but otherwise has as little structure as possible. This automatically generates a Boltzmann-like distribution, defining an energy landscape over the states of the system; importantly, this energy function has no free parameters, but is completely determined by the experimental measurements. As an example, if we look in small windows of time where each neuron in a network either generates an action potential (spike) or remains silent, then the maximum entropy distribution consistent with the mean probability of spiking in each neuron and the correlations among spikes in pairs of neurons is exactly an Ising spin glass [7]. Similarly, the maximum entropy model consistent with the average correlations between the flight direction of a single bird and its immediate neighbors in a flock is a Heisenberg model [8]. Starting with the initial work on the use of pairwise maximum entropy models to describe small ($N = 10$–15) networks of neurons in the retina, this approach has been used to describe the activity in a variety of neural networks [9]–[16], the structure and activity of biochemical and genetic networks [17, 18], the statistics of amino acid substitutions in protein families [19]–[25], and the rules of spelling in English words [26]. Here we return to the retina, taking advantage of new electrode arrays that make it possible to record from a large fraction of the ∼200 output neurons within a small, highly interconnected patch of the circuitry[8]. Our goal is not to

[8] A full account of the experiments will be given elsewhere. Briefly, experiments were performed on the larval tiger salamander, *Ambystoma tigrinum tigrinum*, in accordance with institutional animal care standards. Retinae were isolated from the eye in darkness [30], and the retina was pressed, ganglion cells down, against a custom fabricated array of 252 electrodes (size 8 μm, spacing 30 μm) [31]. The retina was superfused with oxygenated Ringer's medium (95% O_2, 5% CO_2) at 22 °C. Electrode voltage signals were acquired and digitized at 10 kHz by a 252 channel preamplifier (Multi-Channel Systems, Germany). The sorting of these signals into action potentials from individual neurons was done using the methods of [32, 33]. The stimulus was a 19 s grayscale movie clip of a swimming fish and water plants in a fish tank, which was repeated 297 times. It was presented using a CRT display (refresh rate 60 Hz), and focused on the photoreceptor layer of the retina using standard optics.

J. Stat. Mech. (2013) P03011

give a precise model, but rather to construct the simplest model that gives us a glimpse of the collective behavior in this system. For a different approach to simplification, see [27].

The maximum entropy approach is much more general than the construction of models based on pairwise correlations. To be concrete, we consider small slices of time during which each neuron in our network either generates an action potential or remains silent. Then the states of individual neurons are defined by $\sigma_i = 1$ when neuron i generates a spike, and $\sigma_i = -1$ when neuron i is silent. States of the entire network are defined by $\vec{\sigma} \equiv \{\sigma_i\}$, and we are interested in the probability distribution of these states, $P(\vec{\sigma})$. If we know the average values of some functions $f_\mu(\vec{\sigma})$, then the maximum entropy distribution consistent with this knowledge is

$$P(\vec{\sigma}) = \frac{1}{Z(\{g_\mu\})} \exp\left[-\sum_\mu g_\mu f_\mu(\vec{\sigma})\right], \tag{1}$$

where the couplings g_μ have to be adjusted to match the measured expectation values $\langle f_\mu(\vec{\sigma})\rangle$.

In any given slice of time, we will find that K out of the N neurons generate spikes, where

$$K = \tfrac{1}{2}\sum_{i=1}^{N}(\sigma_i + 1). \tag{2}$$

One of the basic characteristics of a network is the distribution of this global activity, $P_N(K)$. As an example, in figure 1 we show experimental results on $P_N(K)$ for groups of $N = 40$ neurons in the retina as it views a naturalistic movie. In these experiments (see footnote 8), we use a dense array of electrodes that samples 160 out of the ∼200 ganglion cells in a small patch of the salamander retina, and we divide time into bins of $\Delta\tau = 20$ ms. The figure shows the average behavior in groups of $N = 40$ cells chosen out of this network, under conditions where a naturalistic movie is projected onto the retina. The correlations between pairs of cells are small, but $P_N(K)$ departs dramatically from what would be expected if the neurons generated spikes independently.

How do we construct the maximum entropy model consistent with the measured $P_N(K)$? Knowing the distribution $P_N(K)$ is equivalent to knowing all its moments, so the functions $f_\mu(\vec{\sigma})$ whose expectation values we have measured are $f_1(\vec{\sigma}) = K$, $f_2(\vec{\sigma}) = K^2$, and so on. Thus we can write

$$P_N(\vec{\sigma}) = \frac{1}{Z(\{g_\mu\})} \exp\left[-\sum_{n=1}^{N} g_n K^n\right] = \frac{1}{Z_N} \mathrm{e}^{-V_N(K)}, \tag{3}$$

where $V_N(K)$ is some effective potential that we need to choose so that $P_N(K)$ comes out equal to the experimentally measured $P_N^{\mathrm{exp}}(K)$.

Usually the inverse problem for these maximum entropy distributions is hard. Here it is much easier. We note that

$$P_N(K) \equiv \sum_{\vec{\sigma}} \delta\left[K, \tfrac{1}{2}\sum_{i=1}^{N}(\sigma_i + 1)\right] P(\vec{\sigma}) \tag{4}$$

$$= \frac{1}{Z_N} \mathcal{N}(K, N)\, \mathrm{e}^{-V_N(K)}, \tag{5}$$

J. Stat. Mech. (2013) P03011

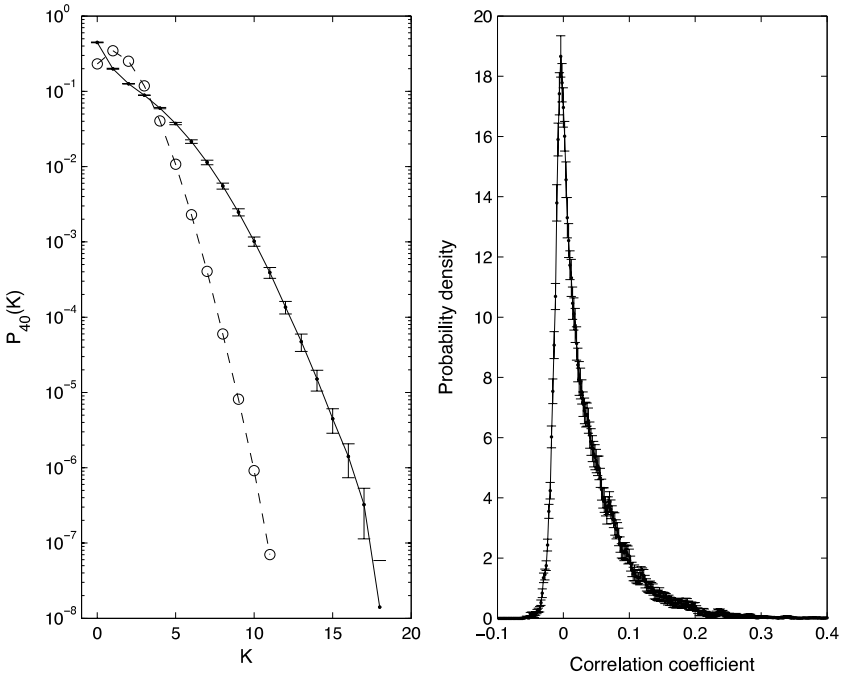

Figure 1. Experimental results for $P_N(K)$, in groups of $N = 40$ neurons. On the left, solid points show the distribution estimated by averaging over many randomly chosen groups of $N = 40$ cells out of the $N = 160$ in our data set; error bars are standard deviations across random halves of the duration of the experiment. Open circles are the expectation if cells are independent. On the right, the distribution of correlation coefficients among pairs of neurons in our sample. Because the experiment is long, the threshold for statistical significance of the correlations is very low, $|C_{\rm thresh}| \leq 0.01$. Almost all pairs of cells thus have significant correlations, but these correlations are weak.

where

$$\mathcal{N}(K, N) = \frac{N!}{(N-K)!K!}. \tag{6}$$

The log of this number is an entropy at fixed K, $S_N(K) \equiv \ln \mathcal{N}(K, N)$, so we can write

$$P_N(K) = \frac{1}{Z_N} \exp\left[S_N(K) - V_N(K)\right]. \tag{7}$$

Finally, to match the distribution $P_N(K)$ to the experimental measurement $P_N^{\rm exp}(K)$, we must have

$$V_N(K) = -\ln P_N^{\rm exp}(K) + S_N(K) - \ln Z_N. \tag{8}$$

In figure 2 we show the average results for $V_N(K)$ in networks of size $N = 40$.

We expect that both energy and entropy will be extensive quantities. For the entropy $S_N(K)$ this is guaranteed by equation (6), which tells us that as N becomes large, $S_N(K) \to Ns(K/N)$. It is an experimental question whether, in the networks we are studying, there is something analogous to a thermodynamic limit in which, for large N, we have $V_N(K) \to N\epsilon(K/N)$. This is illustrated on the right in figure 2, where for

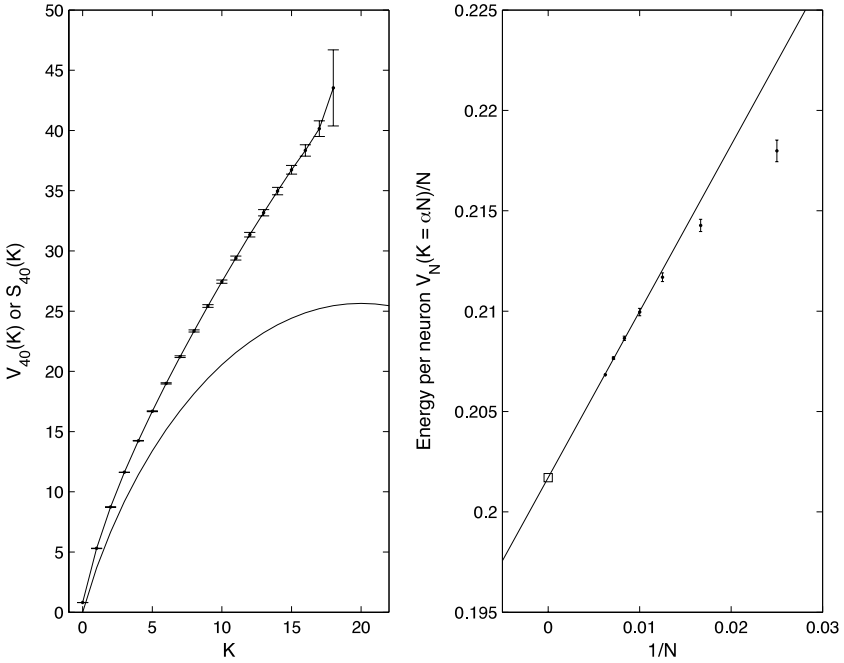

Figure 2. The effective potential and its dependence on system size. On the left, results for $N = 40$ neurons, showing both the potential $V_N(K)$ (points with error bars) and the entropy $S_N(K)$ (smooth curve); error bars are as in figure 1. On the right, the behavior of $V_N(K = \alpha N)/N$, for $\alpha = 0.05$, showing the dependence on N (points with error bars) and the extrapolation $N \to \infty$ (square).

$K/N = 0.05$ we study the dependence of the energy per neuron on $1/N$. There is a natural extrapolation to large N, and this is true for all the ratios of K/N that we tested.

In the $N \to \infty$ limit, the natural quantities are the energy and entropy per neuron, ϵ and s, respectively, and these are shown in figure 3. One clear result is that, as we look at more and more neurons in the same patch of the retina, we do see the emergence of a well defined, smooth relationship between entropy and energy $s(\epsilon)$. While most neural network models are constructed so that this thermodynamic limit exists, it is not so obvious that this should happen in real data. In particular, if we consider a family of models with varying N in which all pairs of neurons are coupled, the standard way of arriving at a thermodynamic limit is to scale the coupling strengths with N, and correspondingly the pairwise correlations are expected to vary with N. In constructing maximum entropy models, we cannot follow this path, since the correlations are measured and thus by definition do not vary as we include more and more neurons. Here we focus not on correlations but on the distribution $P_N(K)$, and thus the emergence of a thermodynamic limit depends on the evolution of this distribution with N.

In the thermodynamic limit we have

$$P_N(K) \to \frac{1}{Z_N} \exp\left[N\left(s(K/N) - \epsilon(K/N)\right)\right], \tag{9}$$

which means that the distribution of K/N should become sharply peaked around its most likely value. Put another way, as we look at larger and larger groups of cells, we expect to see the mean spike count $\langle K \rangle$ grow linearly with N, but also the variance $\langle (\delta K)^2 \rangle$

doi:10.1088/1742-5468/2013/03/P03011

J. Stat. Mech. (2013) P03011

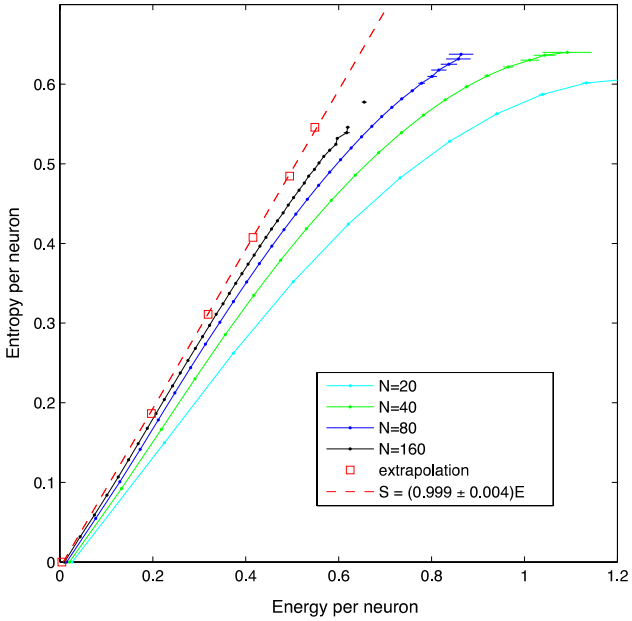

Figure 3. Entropy versus energy. We compute the effective energy per neuron, $\epsilon = V_N(K)/N$, averaged over multiple groups of N neurons chosen out of the 160 we have access to in the experiment, and then compare this with the entropy per neuron, $s = S_N(K)/N$. The extrapolation is as in figure 2, and the error bars in energy (visible only when larger than symbols) are as in figure 2.

should grow linearly with N, and hence the variance should grow linearly with the mean; the fractional variance in K, $\langle(\delta K)^2\rangle/\langle K\rangle^2$, should vanish as N becomes large. In fact, sampling 900 groups of cells from $N = 20$ to 160, we see that the variance has a very precise relation to the mean, but this is quadratic, not linear (figure 4). Further, the fractional variance does seem to approach a limit at large N, but this limit is one, not zero. Thus, the finite networks that we can observe seem typical of samples out of some idealized $N \to \infty$ network with well defined properties, and in this sense we have a thermodynamic limit, but these properties are quite unusual. As an aside, it is interesting that the relationship between mean and variance across different subnetworks is so tight, despite the fact that the system is very inhomogeneous, with the spiking probabilities of individual neurons varying by more than an order of magnitude.

As we look at increasing numbers of neurons, the mean spike count does grow linearly with N, on average across different choices of the N neurons. The fact that the variance of the spike count grows as $\langle(\delta K)^2\rangle \propto \langle K\rangle^2$ (figure 4) thus means that at large N the ratio $\langle(\delta K)^2\rangle/N$ will diverge. But, from figure 2, the energy is a smooth, monotonic function of K, with a derivative that never vanishes, at least over the range where we can make reliable estimates. The divergence of $\langle(\delta K)^2\rangle/N$ at large N thus suggests that the variance of the energy will diverge, even when normalized by the number of neurons. However, this normalized energy variance is the specific heat, and a diverging specific heat is a sign of a critical point. Let us see if we can make this more explicit.

We recall that the plot of entropy versus energy tells us everything about the thermodynamics of the system. In our maximum entropy construction, there is no real temperature—k_BT just provides units for the effective energy $V_N(K)$. However, if we

J. Stat. Mech. (2013) P03011

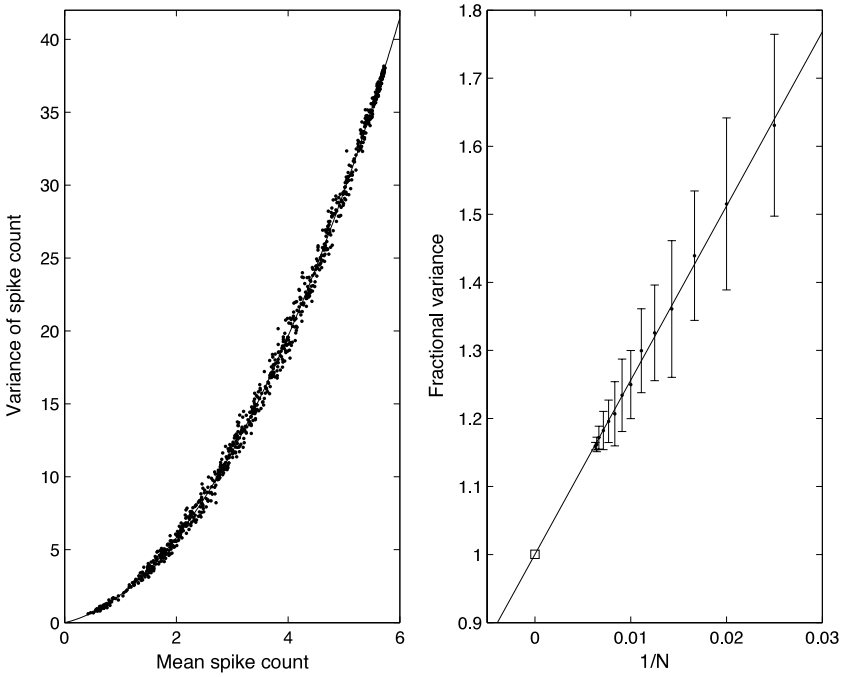

Figure 4. Variance and mean of spike counts. On the left, we select 900 groups of cells, from $N = 20$ to 160, and compute the mean and variance of spike counts K. On the right, we look at the fractional variance in spike counts; error bars are the standard deviation across multiple groups with the same N, and the square marks the extrapolation $N \to \infty$.

have a model for the energy as a function of the microscopic state of the system, then we can take this seriously as a statistical mechanics problem and imagine varying the temperature. More precisely, we can generalize equation (3) to consider

$$P_N(\vec{\sigma};\beta) = \frac{1}{Z_N(\beta)} \mathrm{e}^{-\beta V_N(K)}, \tag{10}$$

where the real system is at $\beta = 1$. Then in the thermodynamic limit we have the usual identities: the temperature is defined by $\partial s/\partial\epsilon = \beta$, the specific heat is $C = k_B\beta^2(-\partial^2 s/\partial\epsilon^2)^{-1}$, and so on. In particular, the vanishing of the second derivative of the entropy implies a diverging specific heat, a signature of a critical point.

In our case, since the real system is at $\beta = 1$, the behavior of the network will be dominated by states with an energy per neuron that solves the equation $\partial s/\partial\epsilon = 1$, but figure 3 shows us that, as we consider more and more neurons, the function $s(\epsilon)$ seems to be approaching $s = \beta_0\epsilon$, where $\beta_0 = 0.999 \pm 0.004$ is one within errors. If we had exactly $s = \epsilon$, then all energies would be solutions of the condition $\partial s/\partial\epsilon = 1$. Correspondingly, the specific heat C would diverge, signaling that the operating point $\beta = 1$ is a critical point. This is a very unusual critical point, since all higher derivatives of the entropy vanish [29].

More generally, when we try to describe the probability distribution over states $\vec{\sigma}$ using ideas from statistical mechanics, we are free to choose the zero of the (effective) energy as we wish. A convenient choice is that the unique state of zero spikes—complete

J. Stat. Mech. (2013) P03011

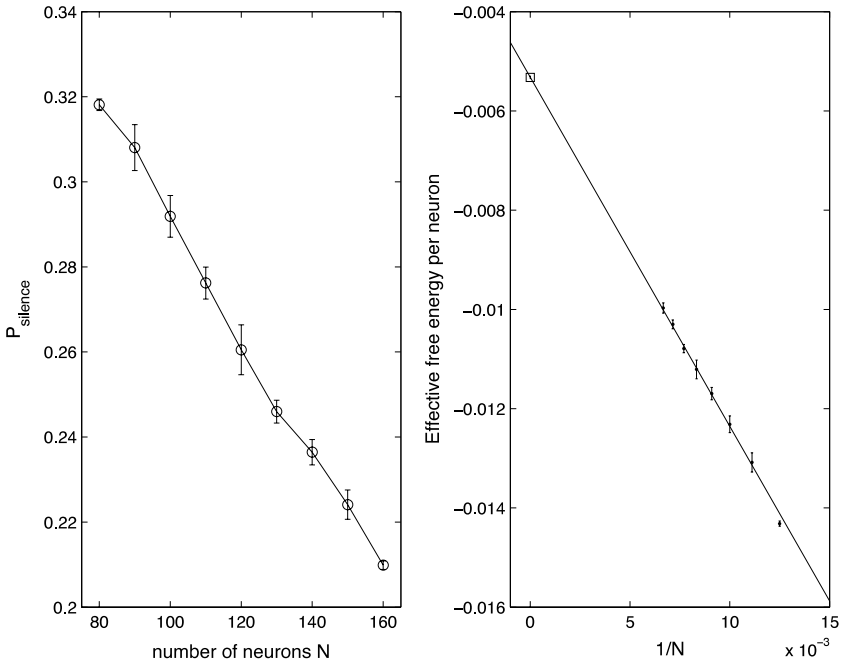

Figure 5. The probability of silence, and the effective free energy. On the left, the probability that a network of N neurons is in the silent state, where none of the cells generate a spike within a window $\Delta\tau$; error bars as in figure 1. Note that this probability declines very slowly with the number of neurons N. On the right, we translate the probability of silence into an effective free energy per neuron, and see that this varies linearly with $1/N$, yielding the extrapolation $N \to \infty$ (square).

silence in the network—should have zero energy. Unless there are exponentially many states with probability equal to the silent state (which seems unlikely), in the large N limit the entropy per neuron will also be zero at zero energy, but with this choice for the zero of energy, the probability of the silent state is given by $P_{\rm silence} = 1/Z$, and $Z = e^{-F}$, where F is the effective free energy, since we are at $\beta = 1$. Thus if we can measure this probability reliably, we can 'measure' the free energy, without any further assumptions. We see in figure 5 that the probability of silence falls as we look at more and more neurons, which makes sense since the free energy should grow with system size, but the decline in the probability of silence is surprisingly slow. We can make this more precise by computing the effective free energy per neuron, $f = F/N$, also shown. This is a very small number indeed, $f \sim -0.01$ at the largest values of $N = 160$ for which we have data.

We recall that, with $k_{\rm B}T = 1$, the free energy per neuron is $f = \langle\epsilon\rangle - s_{\rm total}$, where $\langle\epsilon\rangle$ denotes the average energy and $s_{\rm total}$ is the total entropy of the system, again normalized per neuron. Our best estimate of the entropy of the states taken on by the network is $s_{\rm total} \sim 0.2$ per neuron, which means that the free energy reflects a cancellation between energy and entropy with a precision of at least $\sim$5%. If we extrapolate to the thermodynamic limit the cancellation becomes even more precise, so that the extensive component of the free energy is $f_\infty = -0.0051 \pm 0.000\,03$ (figure 5). Notice that the small value of the free energy means that the silent state occurs frequently, and hence we can measure its probability very accurately, so the error bars are small. If we had a critical

J. Stat. Mech. (2013) P03011

system in which $s(\epsilon) = \epsilon$, the extensive component of the free energy would be exactly zero.

In a normal thermodynamic limit (and $\beta = 1$), $f_\infty = \epsilon^* - s(\epsilon^*)$, where ϵ^* is the energy at which $\partial s/\partial\epsilon = 1$. Geometrically, f_∞ is the intercept along the energy axis of a line with unit slope that is tangent to the curve $s(\epsilon)$ at the point ϵ^*. From above we have $s(0) = 0$, and then if $s(\epsilon)$ is concave ($\partial^2 s(\epsilon)/\partial\epsilon^2 < 0$, so that the specific heat is everywhere positive) we are guaranteed that f_∞ is negative, but to have $f_\infty \to 0$ then requires that $\partial s(\epsilon)/\partial\epsilon \leq 1$ at $\epsilon = 0$. In this scenario, pushing f_∞ toward zero requires both ϵ^* and $s(\epsilon^*)$ to approach zero, so that the network is in a (near) zero entropy state despite the finite temperature. This state would be similar to the critical point in the random energy model [28], but this seems inconsistent with the evidence for a nonzero entropy per neuron.

To have near zero free energy with nonzero entropy seems to require something very special. One possibility is to allow $\partial^2 s(\epsilon)/\partial\epsilon^2 > 0$, allowing phase coexistence between the $\epsilon = 0$ silent state and some other $\epsilon \neq 0$ state. The other possibility is to have $s(\epsilon) = \epsilon$, as suggested by figure 3. Thus, while the observation of a nearly zero free energy per neuron does not prove that the entropy is equal to the energy for all energies, it does tell us that the network is in or near one of a handful of unusual collective states.

The model we have considered here of course throws away many things: we are not keeping track of the identities of the cells, but rather trying to capture the global activity of the network. On the other hand, because we are considering a maximum entropy model, we know that what we are constructing is the least structured model that is consistent with $P_N(K)$. It thus is surprising that this minimal model is so singular. As we have emphasized, even without appealing to a model, we know that there is something special about these networks of neurons because they exhibit an almost perfect cancellation of energy and entropy. The more detailed maximum entropy analysis suggests that cancellation is not just true on average, but rather that the entropy is almost precisely equal to the energy as a function. This is consistent with hints of criticality in previous analyses, which extrapolated from much smaller groups of neurons [10, 13, 29], although much more remains to be done.

Acknowledgments

We thank A Cavagna, I Giardina, E Schneidman, G J Stephens, T Taillefumier, and A Walczak for helpful discussions. This work was supported in part by NSF Grants IIS-0613435 and PHY-0957573, by NIH Grants R01 EY14196 and P50 GM071508, by the Fannie and John Hertz Foundation, by the Human Frontiers Science Program, by the Swartz Foundation, and by the WM Keck Foundation.

References

[1] Hopfield J J, *Neural networks and physical systems with emergent collective computational abilities*, 1982 *Proc. Nat. Acad. Sci.* **79** 2554

[2] Amit D J, 1989 *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge: Cambridge University Press)

[3] Hertz J, Krogh A and Palmer R G, 1991 *Introduction to the Theory of Neural Computation* (Redwood City: Addison Wesley)

[4] Toner J and Tu Y, *Long-range order in a two-dimensional XY model: how birds fly together*, 1995 *Phys. Rev. Lett.* **75** 4326

J. Stat. Mech. (2013) P03011

[5] Toner J and Tu Y, *Flocks, herds, and schools: a quantitative theory of flocking*, 1998 *Phys. Rev.* E **58** 4828
[6] Jaynes E T, *Information theory and statistical mechanics*, 1957 *Phys. Rev.* **106** 620
[7] Schneidman E, Berry M J II, Segev R and Bialek W, *Weak pairwise correlations imply strongly correlated network states in a neural population*, 2006 *Nature* **440** 1007
[8] Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M and Walczak A, *Statistical mechanics for natural flocks of birds*, 2012 *Proc. Nat. Acad. Sci.* **109** 4786
Bialek W, Cavagna A, Giardina I, Mora T, Silvestri E, Viale M and Walczak A, 2011 arXiv:1107.0604 [physics.bio-ph]
[9] Shlens J, Field G D, Gaulthier J L, Grivich M I, Petrusca D, Sher A, Litke A M and Chichilnisky E J, *The structure of multi-neuron firing patterns in primate retina*, 2006 *J. Neurosci.* **26** 8254
[10] Tkačik G, Schneidman E, Berry M J II and Bialek W, *Ising models for networks of real neurons*, 2006 arXiv:q-bio/0611072
[11] Yu S, Huang D, Singer W and Nikolic D, *A small world of neuronal synchrony*, 2008 *Cerebral Cortex* **18** 2891
[12] Tang A *et al, A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro*, 2008 *J. Neurosci.* **28** 505
[13] Tkačik G, Schneidman E, Berry M J II and Bialek W, *Spin-glass models for a network of real neurons*, 2009 arXiv:0912.5409
[14] Shlens J, Field G D, Gaulthier J L, Greschner M, Sher A, Litke A M and Chichilnisky E J, *The structure of large-scale synchronized firing in primate retina*, 2009 *J. Neurosci.* **29** 5022
[15] Ohiorhenuan I E, Mechler F, Purpura K P, Schmid A M, Hu Q and Victor J D, *Sparse coding and higher-order correlations in fine-scale cortical networks*, 2010 *Nature* **466** 617
[16] Ganmor E, Segev R and Schniedman E, *Sparse low-order interaction network underlies a highly correlated and learnable neural population code*, 2011 *Proc. Nat. Acad. Sci.* **108** 9679
[17] Lezon T R, Banavar J R, Cieplak M, Maritan A and Federoff N V, *Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns*, 2006 *Proc. Nat. Acad. Sci.* **103** 19033
[18] Tkačik G, *Information flow in biological networks*, 2007 *Dissertation* Princeton University
[19] Bialek W and Ranganathan R, *Rediscovering the power of pairwise interactions*, 2007 arXiv:0712.4397 [q-bio.QM]
[20] Seno F, Trovato A, Banavar J R and Maritan A, *Maximum entropy approach for deducing amino acid interactions in proteins*, 2008 *Phys. Rev. Lett.* **100** 078102
[21] Weigt M, White R A, Szurmant H, Hoch J A and Hwa T, *Identification of direct residue contacts in protein–protein interaction by message passing*, 2009 *Proc. Nat. Acad. Sci.* **106** 67
[22] Halabi N, Rivoire O, Leibler S and Ranganathan R, *Protein sectors: evolutionary units of three-dimensional structure*, 2009 *Cell* **138** 774
[23] Mora T, Walczak A M, Bialek W and Callan C G, *Maximum entropy models for antibody diversity*, 2010 *Proc. Nat. Acad. Sci.* **107** 5405
Mora T, Walczak A M, Bialek W and Callan C G, 2009 arXiv:0912.5175 [q-bio.GN]
[24] Marks D S, Colwell L J, Sheridan R, Hopf T A, Pagnani A, Zecchina R and Sander C, *Protein 3D structure computed from evolutionary sequence variation*, 2011 *PLoS One* **6** e28766
[25] Sulkowska J I, Morocos F, Weigt M, Hwa T and Onuchic J N, *Genomics-aided structure prediction*, 2012 *Proc. Nat. Acad. Sci.* **109** 10340
[26] Stephens G J and Bialek W, *Statistical mechanics of letters in words*, 2010 *Phys. Rev.* E **81** 066119
Stephens G J and Bialek W, 2008 arXiv:0801.0253 [q-bio.NC]
[27] Macke J H, Opper M and Bethge M, *Common input explains higher-order correlations and entropy in a simple model of neural population activity*, 2011 *Phys. Rev. Lett.* **106** 208102
[28] Derrida B, *Random-energy model: an exactly solvable model of disordered systems*, 1981 *Phys. Rev.* B **24** 2613
[29] Mora T and Bialek W, *Are biological systems poised at criticality?*, 2011 *J. Stat. Phys.* **144** 268
Mora T and Bialek W, 2010 arXiv:1012.2242 [q-bio.QM]
[30] Puchalla J L, Schneidman E, Harris R A and Berry M J II, *Redundancy in the population code of the retina*, 2005 *Neuron* **46** 493
[31] Amodei D, *Network-scale electrophysiology: measuring and understanding the collective behavior of neural circuits*, 2011 *Dissertation* Princeton University
[32] Segev R, Goodhouse J, Puchalla J and Berry M J II, *Recording spikes from a large fraction of the ganglion cells in a retinal patch*, 2004 *Nature Neurosci.* **7** 1155
[33] Marre O, Amodei D, Deshmukh N, Sadeghi K, Soo F, Holy T E and Berry M J II, *Recording from a neural complete population in the retina*, 2012 *J. Neurosci.* **32** 14859

J. Stat. Mech. (2013) P03011