

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

Gaspar Tkačik,^a Olivier Marre,^{b,e} Thierry Mora,^c Dario Amodei,^{d,f} Michael J Berry II,^{e,f} and William Bialek,^{d,g}

^a*Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria,*

^b*Institution de al Vision, UMRs 968 UPMC, INSERM,*

CNRS U7210, CHNO Quinze-Vingts, F-75012 Paris, France,

^c*Laboratoire de Physique Statistique de l'École Normale Supérieure,*

CNRS and Universites Paris VI and Paris VII,

24 rue Lhomond, 75231 Paris Cedex 05, France,

^d*Joseph Henry Laboratories of Physics, ^eDepartment of Molecular Biology,*

^f*Princeton Neuroscience Institute, and ^gLewis-Sigler Institute for Integrative Genomics*

Princeton University, Princeton, New Jersey 08544 USA

(Dated: July 27, 2012)

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 . 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:

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 lead 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 [27]. Our goal is not to 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 Ref [28].

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

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

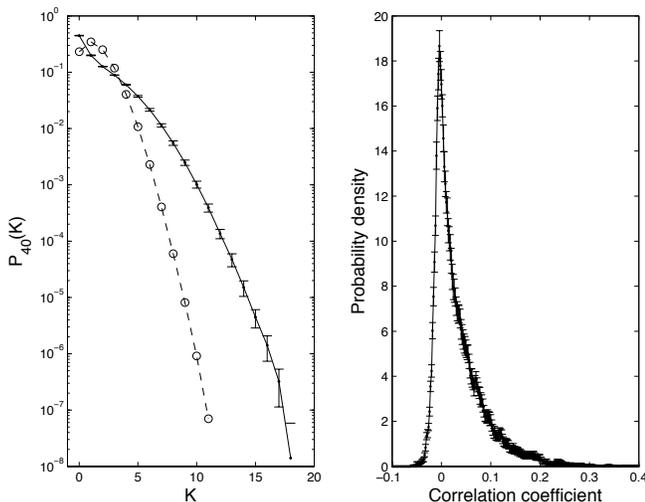


FIG. 1: Experimental results for $P_N(K)$, in groups of $N = 40$ neurons. At left, solid points show the distribution estimated by averaging over many randomly chosen groups of $N = 40$ cells out of the $N = 160$ in our data set; error bars are standard deviations across random halves of the duration of the experiment. Open circles are the expectation if cells are independent. At 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, $|C_{\text{thresh}}| \leq 0.01$. Almost all pairs of cells thus have significant correlations, but these correlations are weak.

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

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

One of the basic characteristics of a network is the distribution of this global activity, $P_N(K)$. As an example, in Fig 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 [27], 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

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

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 $P_N^{\text{exp}}(K)$.

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

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

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

where

$$\mathcal{N}(K, N) = N! / (N - K)! K!. \quad (6)$$

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

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

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

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

In Fig 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 Eq (6), which tells us that as N becomes large, $S_N(K) \rightarrow N s(K/N)$. 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 $V_N(K) \rightarrow N \epsilon(K/N)$. This is illustrated at right in Fig 2, where for $K/N = 0.05$ 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 $N \rightarrow \infty$ limit, the natural quantities are the energy and entropy per neuron, ϵ and s , respectively, and these are shown in Fig 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

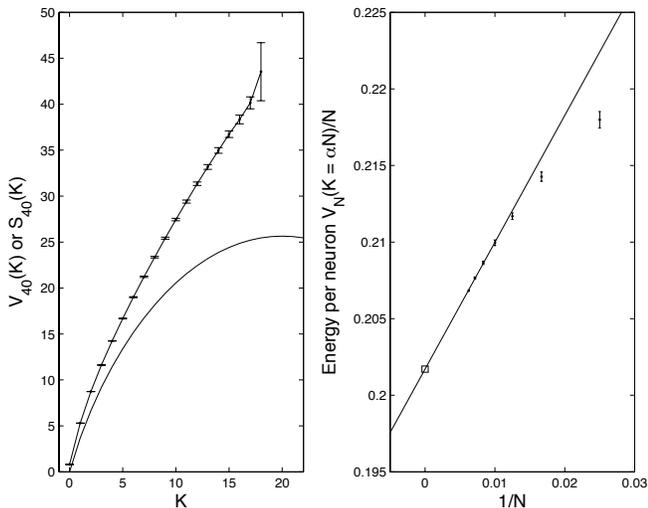


FIG. 2: The effective potential and its dependence on system size. At left, results for $N = 40$ neurons, showing both the potential $V_N(K)$ (points with error bars) and the entropy $S_N(K)$ (smooth curve); error bars are as in Fig 1. At right, the behavior of $V_N(K = \alpha N)/N$, for $\alpha = 0.05$,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

entropy models, we can't follow this path, since the correlations are measured and thus by definition don't 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 .

We recall that the plot of entropy vs. energy tells us everything about the thermodynamics of the system. In our maximum entropy construction, there is no real temperature— $k_B T$ just provides units for the effective energy $V_N(K)$. But if we 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 Eq (3) to consider

$$P_N(\vec{\sigma}; \beta) = \frac{1}{Z_N(\beta)} e^{-\beta V_N(K)}, \quad (9)$$

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 Fig 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

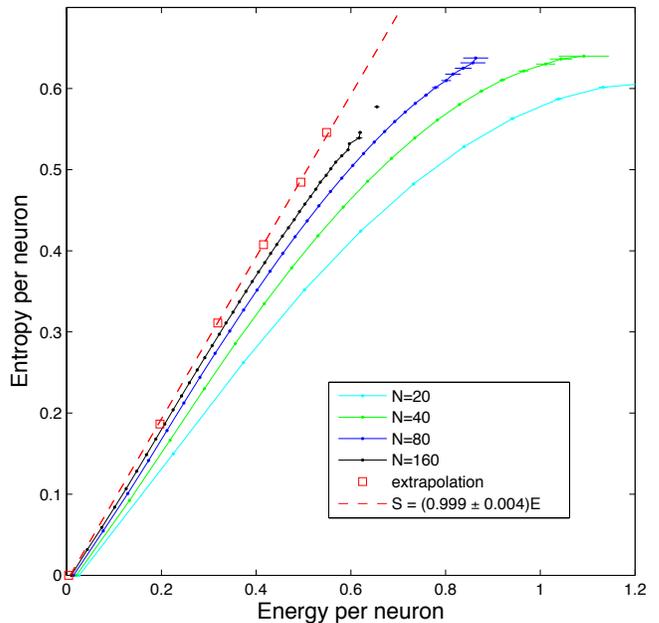


FIG. 3: Entropy vs. energy. We compute the effective energy per neuron, $\epsilon = V_N(K)/N$, averaged over multiple groups of N neurons chosen out of the 160 we have access to in the experiment, and then compare with the entropy per neuron, $s = S_N(K)/N$. The extrapolation is as in Fig 2, and error bars in energy (visible only when larger than symbols) are as in Fig. 1.

heat C would diverge, signaling that $\beta = 1$ is a critical point. This is a very unusual critical point, since all higher derivatives of the entropy vanish [30].

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 silence in the network—should have zero energy. Unless there are exponentially many states that 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 $P_{\text{silence}} = 1/Z$, 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 Fig 4 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 $k_B T = 1$, the free energy per neuron is $f = \langle \epsilon \rangle - s_{\text{total}}$, where $\langle \epsilon \rangle$ denotes the average energy and s_{total} is the total entropy of the system, again

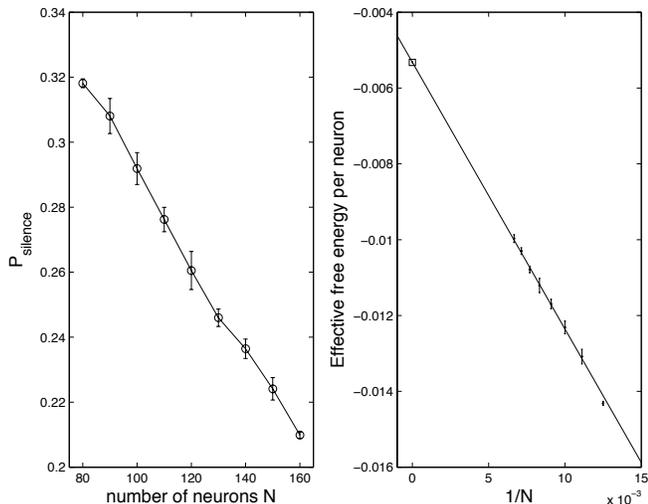


FIG. 4: The probability of silence, and the effective free energy. At left, the probability that a network of N neurons is in the silent state, where none of the cells generate a spike within a window $\Delta\tau$; error bars as in Fig 1. Note that this probability declines very slowly with the number of neurons N . At right, we translate the probability of silence into an effective free energy per neuron, and see that this varies linearly with $1/N$, the extrapolation $N \rightarrow \infty$ (square).

normalized per neuron. Our best estimate of the entropy of the states taken on by the network is $s_{\text{total}} \sim 0.2$ 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 $f_{\infty} = -0.0051 \pm 0.00003$ (Fig 4). 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 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 spe-

cific heat is everywhere positive) we are guaranteed that f_{∞} is negative. But to have $f_{\infty} \rightarrow 0$ 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 [29], 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 Fig 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, 30], although much more remains to be done.

We thank A Cavagna, I Giardina, E Schneidman, GJ 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.

-
- [1] JJ Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci (USA)* **79**, 2554–2558 (1982).
 - [2] DJ Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, 1989).
 - [3] J Hertz, A Krogh & RG Palmer, *Introduction to the Theory of Neural Computation* (Addison Wesley, Redwood City, 1991).
 - [4] J Toner & Y Tu, Long-range order in a two-dimensional XY model: How birds fly together. *Phys Rev Lett* **75**, 4326–4329 (1995).
 - [5] J Toner & Y Tu, Flocks, herds, and schools: A quantitative theory of flocking. *Phys Rev E* **58**, 4828–4858 (1998).
 - [6] ET Jaynes, Information theory and statistical mechanics. *Phys Rev* **106**, 620–630 (1957).
 - [7] E Schneidman, MJ Berry II, R Segev & W Bialek. Weak

- pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
- [8] W Bialek, A Cavagna, I Giardina, T Mora, E Silvestri, M Viale & A Walczak, Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci (USA)* **109**, 4786–4791 (2012); arXiv.org:1107.0604 [physics.bio-ph] (2011).
- [9] J Shlens, GD Field, JL Gauthier, MI Grivich, D Petrusca, A Sher, AM Litke & EJ Chichilnisky, The structure of multi-neuron firing patterns in primate retina. *J Neurosci* **26**, 8254–8266 (2006).
- [10] G Tkačik, E Schneidman, MJ Berry II & W Bialek, Ising models for networks of real neurons. arXiv:q-bio/0611072 (2006).
- [11] S Yu, D Huang, W Singer & D Nikolic, A small world of neuronal synchrony. *Cereb Cortex* **18**, 2891–2901 (2008).
- [12] A Tang, D Jackson, J Hobbs, W Chen, JL Smith, H Patel, A Prieto, D Petrusca MI Grivich, A Sher, P Hottowy, W Dabrowski, AM Litke & JM Beggs, A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* **28**, 505–518 (2008).
- [13] G Tkačik, E Schneidman, MJ Berry II & W Bialek, Spin-glass models for a network of real neurons. arXiv:0912.5409 (2009).
- [14] J Shlens, GD Field, JL Gauthier, M Greschner, A Sher, AM Litke & EJ Chichilnisky, The structure of large-scale synchronized firing in primate retina. *J Neurosci* **29**, 5022–5031 (2009).
- [15] IE Ohiorhenuan, F Mechler, KP Purpura, AM Schmid, Q Hu & JD Victor, Sparse coding and higher-order correlations in fine-scale cortical networks. *Nature* **466**, 617–621 (2010).
- [16] E Ganmor, R Segev & E Schneidman, Sparse low-order interaction network underlies a highly correlated and learnable neural population code, *Proc Natl Acad Sci (USA)* **108**, 9679–9684 (2011).
- [17] TR Lezon, JR Banavar, M Cieplak, A Maritan & NV Federoff, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci (USA)* **103**, 19033–19038 (2006).
- [18] G Tkačik, *Information Flow in Biological Networks* (Dissertation, Princeton University, 2007).
- [19] W Bialek & R Ranganathan, Rediscovering the power of pairwise interactions. arXiv:0712.4397 [q-bio.QM] (2007).
- [20] F Seno, A Trovato, JR Banavar & A Maritan, Maximum entropy approach for deducing amino acid interactions in proteins. *Phys Rev Lett* **100**, 078102 (2008).
- [21] M Weigt, RA White, H Szurmant, JA Hoch & T Hwa, Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci (USA)* **106**, 67–72 (2009).
- [22] N Halabi, O Rivoire, S Leibler & R Ranganathan, Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
- [23] T Mora, AM Walczak, W Bialek & CG Callan, Maximum entropy models for antibody diversity. *Proc Natl Acad Sci (USA)* **107**, 5405–5410 (2010); arXiv:0912.5175 [q-bio.GN] (2009).
- [24] DS Marks, LJ Colwell, R Sheridan, TA Hopf, A Pagnani, R Zecchina & C Sander, Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766 (2011).
- [25] JI Sulkowska, F Morocos, M Weigt, T Hwa & JN Onuchic, Genomics-aided structure prediction. *Proc Natl Acad Sci (USA)* **109**, 10340–10345 (2012).
- [26] GJ Stephens & W Bialek, Statistical mechanics of letters in words. *Phys Rev E* **81**, 066119 (2010); arXiv:0801.0253 [q-bio.NC] (2008).
- [27] 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 [31], and the retina was pressed, ganglion cell down, against a custom fabricated array of 252 electrodes (size 8 μm , spacing 30 μm) [32]. The retina was superfused with oxygenated Ringer’s medium (95% O₂, 5% CO₂) 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 Refs [33, 34]. 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.
- [28] JH Macke, M Opper & M Bethge, Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Phys Rev Lett* **106**, 208102 (2011).
- [29] B Derrida, Random-energy model: An exactly solvable model of disordered systems. *Phys Rev B* **24**, 2613–2626 (1981).
- [30] T Mora & W Bialek, Are biological systems poised at criticality? *J Stat Phys* **144**, 268–302 (2011); arXiv:1012.2242 [q-bio.QM] (2010).
- [31] JL Puchalla, E Schneidman, RA Harris & MJ Berry II, Redundancy in the population code of the retina. *Neuron* **46**, 493–504 (2005).
- [32] D Amodei, *Network-scale Electrophysiology: Measuring and Understanding the Collective Behavior of Neural Circuits* (Dissertation, Princeton University, 2011).
- [33] R Segev, J Goodhouse, J Puchalla & MJ Berry II, Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neurosci* **7**, 1155–1162 (2004).
- [34] O Marre, D Amodei, K Sadeghi, F Soo, TE Holy & MJ Berry II, Recording from a complete population in the retina. Submitted (2012).