

# Methods from neuroscience for studying networks using incomplete data

*in response to the A exam question posed by Veit Elser*

Bryan C. Daniels

*Department of Physics, Cornell University*

(Dated: January 13, 2009)

## Abstract

Studies of biological neural networks are limited by the amount of simultaneous data they can record. This restriction has led to techniques for characterizing networks using incomplete information, including parameter search algorithms, prior beliefs about the network structure, and effective models that can still make useful predictions. Similar techniques have also been used in studies of gene regulatory networks.

## I. INTRODUCTION

In biology, we often find large collections of relatively simple interacting elements that combine to create complicated structures with complex behavior. Neural networks and gene interaction networks are two such examples.

The communities studying these systems share a common problem: the available measurements are not able to resolve all the individual components or interactions (nor do we expect them to be able to do so soon, with the large numbers of components typically involved). We are left in a situation in which we are unable to fill in the middle-scale details. We know how a single neuron fires, and how a single gene transcription factor can alter the rate of the production of proteins. We also have a feeling for the broad behavior of information transfer in the brain, and the switching abilities of entire gene networks. But lacking is an understanding of exactly how the components fit into the whole.

The problem, then, is taking incomplete information about the behavior of a network and inferring something about how it works. The goal is to be able to at least match the output of a network with a model, and at most to have a full understanding of the individual players and interactions among them that produce the behavior in the actual system. In this short review, we will look at how neuroscience has tackled the challenges in studying a complex network when the available measurements cannot resolve every individual player. We will also briefly discuss how this relates to similar problems in the inference of gene networks.

### A. A quick introduction to neuroscience

Neurons are the subject of neuroscience, and a major goal is to understand how their collective behavior produces the information transport and processing abilities of biological neural networks. The dynamics of individual neurons have been explored in detail, and indeed have been a major focus of neuroscience [2]. Flows of ions through gated channels in the neuron's cell membrane lead to reproducible spikes in the potential across the membrane, and these spikes act as signals that can induce other neurons to spike [3]. The top trace in Fig. 1 shows an example recording of a neuron's membrane potential. Due to the complexity of the dynamics at the neuronal level, this has been a rich area of research.<sup>1</sup>

---

<sup>1</sup> Another way in which the idea of incomplete information comes up in neuroscience is in fitting single neuron models to data. In at least one study, fitting such a dynamical model led to interesting issues with ill-constrained parameters [4].

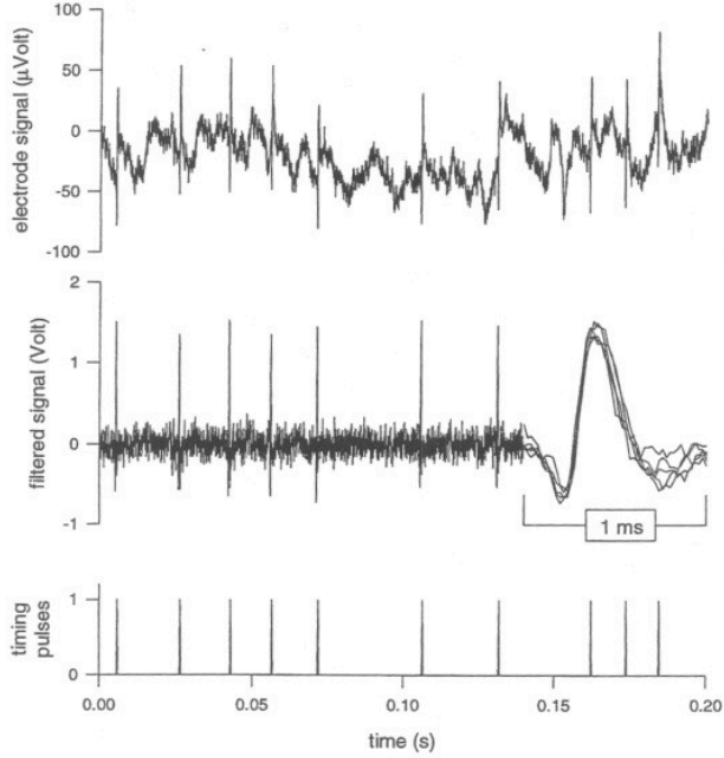


FIG. 1: Example potential recording from a single neuron, from Ref. [1]. The top trace shows the measured signal, the middle trace has filtered out low frequency noise, and the bottom trace shows the recorded spike times. The right section of the middle trace shows an expanded view with all the action potentials (spikes) overlaid to show their similarity to one another.

Brown *et al.* describe computational neuroscience as split into two groups: those who start with detailed biophysical neural models and build large artificial networks to see what behavior emerges, and those who try to directly analyze the growing amount of data coming from neuroscience experiments [5]. In dealing with large amounts of data, this latter group has espoused a less fine-grained modeling style, dealing more with signal processing and statistical methods than detailed dynamical models. They assert that neuronal-level dynamics are not needed to explain higher-order functions (cognitive behavior, information processing, etc.)—instead, spikes and their timing are assumed to carry all the relevant information, and all the other complexity of ion channels is abstracted away [1].<sup>2</sup>

Then, as in the bottom trace of Fig. 1, the output of a neuron is specified by a set of spike times,

<sup>2</sup> This standpoint has been reinforced by studies that have created decoding filters that can rather successfully reconstruct the stimulus given to a set of sensory neurons using only the output spike timing information; see, e.g., [6]. Still, proponents of a more dynamical and fine-grained perspective (such as Izhikevich [2]) might disagree with throwing away all this information.

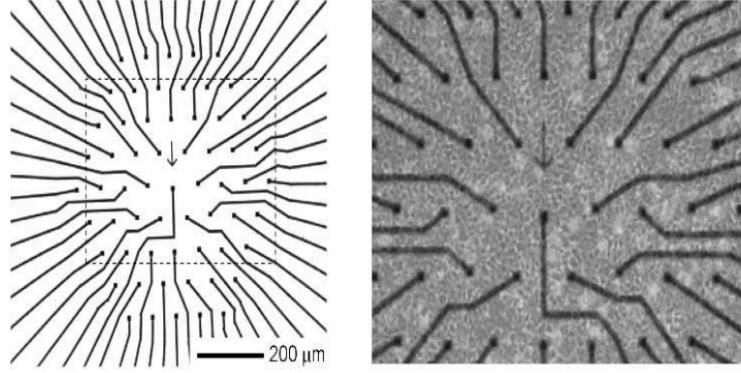


FIG. 2: Multielectrode array containing 61 electrodes in a hexagonal arrangement; figure from Ref. [8]. The photograph on the right shows neurons cultured directly on the array. Other types of multielectrode arrays can be used to take data from living subjects.

commonly known as a “spike train.” For fitting spike train data, neurons are usually assumed to fire spikes at random times in a Poisson process. And since a “point process” is defined in mathematics as a series of discrete events separated by random intervals, this is the name given to these neural models that output stochastic spike trains.

The typical methods of measuring the electrical state of a neuron use electrodes that either attach to a single neuron (intracellular electrodes) or are placed near a neuron (extracellular electrodes). Intracellular electrodes provide a better signal, but are generally too delicate to be used for *in vivo* experiments. Extracellular electrodes are easier to use, but they can pick up signals from multiple neurons [3].

Beginning in the 1960s, multiple-electrode devices began to be fabricated that could measure from several nearby points simultaneously [7]. A typical multielectrode array used today (see Fig. 2) may have about 50 extracellular electrodes regularly arranged over a distance of several hundred microns. Neurons can be either grown directly on the plate containing the multielectrode array [8], or the arrays can be implanted into living subjects [9].

## II. STUDYING COLLECTIONS OF NEURONS WITH DATA FROM A FEW

Ever since the activity of neurons has been measured, properties of groups of connected neurons have been of paramount interest to neuroscientists. Older methods focused on characterizing correlations between pairs of neurons at a time, and newer methods have attempted to fit data to large network models.

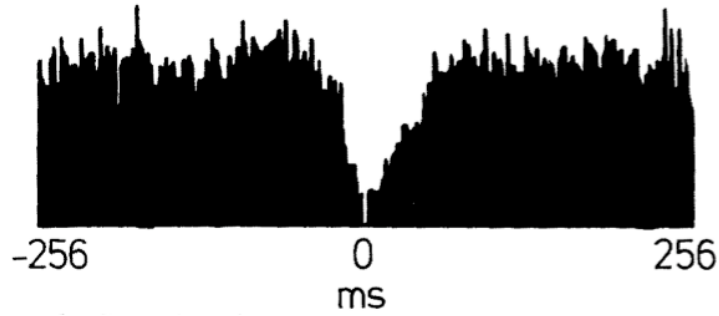


FIG. 3: A “cross-correlogram” indicating the functional connectivity between two neurons based on the relative timing of their spike trains; figure from Ref. [7]. The dip on both sides of zero indicates effective mutual inhibition: each neuron is less likely to fire soon after the other has fired.

#### A. Cross-correlation and pairwise methods

Early ventures into understanding neural network behavior had a strategy of working up from interactions between individual pairs of neurons, with the hope of piecing together a whole network from many such pairs.

A straightforward method that is often used to measure correlations between spike trains is the cross-correlogram (see Fig. 3), which bins spike train data for two neurons based on the delays between spikes of one neuron and spikes of the other [3, 7, 10]. This gives a qualitative picture of the “functional” connectivity between the two neurons: a dip in the cross-correlogram implies an effective inhibitory connection, and a peak implies an effective excitatory connection.

This type of pairwise method has been used to successfully characterize overall connectivity patterns between different brain regions and some local interactions in visual and auditory regions [11]. It has long been recognized, however, that measuring a correlation between the firing times of two neurons does not imply that they have a direct physiological connection [10, 11]. First, and most simply, intermediate unmeasured neurons could be causing an effective connection between the two measured neurons, while the neurons themselves are not directly connected. Second, as in Fig. 4, a shared source of input could be incorrectly interpreted as a connection. Thus measuring only pairwise correlations leaves ambiguity in the true connectivity. While clever techniques can pull out some additional information from cross-correlation data (for example, looking at the exact shape of the cross-correlations), these usually require human eyes and experience [11]. This motivates, then, methods that can describe more than two neurons at a time.

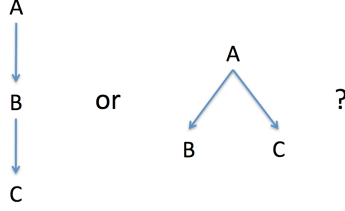


FIG. 4: Measuring only pairwise correlations between the output of neurons leaves ambiguity in the true connectivity. As a simple but typical example, it is difficult to distinguish a direct connection from a case of shared input.

### B. Model-based many-neuron models

Methods that attempt to describe many experimentally-measured neurons at once are relatively new, and consist mostly of so-called “model-based” methods.<sup>3</sup> The idea behind model-based methods is to fit experimental data to a predefined model, inferring the best parameters for the model using ideas from statistics such as maximum likelihood and Bayesian inference. The advantage to these methods is that they at least have the opportunity to make use of higher-order correlations (not just pairwise), since the data from every neuron is used at once.

The prototypical model for reproducing spike train data involves a stochastic simulation in which the probability density of the number of spikes emitted by neuron  $i$  at time  $t$  (or, equivalently, its instantaneous rate of firing) is given by its “stochastic intensity” or “conditional intensity”  $\lambda_i(t)$ . Each neuron has an intrinsic firing rate which it would have in the absence of the effects of other neurons, and its own firing affects  $\lambda_j$  for some or all of the other neurons, based on an interaction kernel that varies over time. For example, a typical model (see Fig. 5) would be defined by the conditional intensity [9, 11]

$$\lambda_i(t \mid \alpha_i) = \exp \left( \alpha_{i,0} + \sum_c \sum_m \alpha_{i,c,m} \mathcal{I}_{c,m}(t) \right), \quad (1)$$

where  $\alpha_{i,0}$  contains the intrinsic rate of firing of neuron  $i$  and  $\alpha_{i,c,m}$  contains its time-dependent effect on every other neuron (the  $c$  sum is over all other neurons, the  $m$  sum is over all past time windows, and  $\mathcal{I}$  is zero except if neuron  $c$  spiked  $m$  timesteps before the current time  $t$ , in which case it is one).

<sup>3</sup> There are also other methods that attempt to use the spike timing relations of many neurons to extract information about the network, but they are less easily interpreted as providing quantitative estimates of connectivity. These include gravitational clustering [12, 13], which represents each neuron as a particle and invents dynamics for pulling together neurons that fire in synchrony; and spike pattern classification methods [14, 15], which attempt to detect patterns of firing that are unlikely to happen by chance.

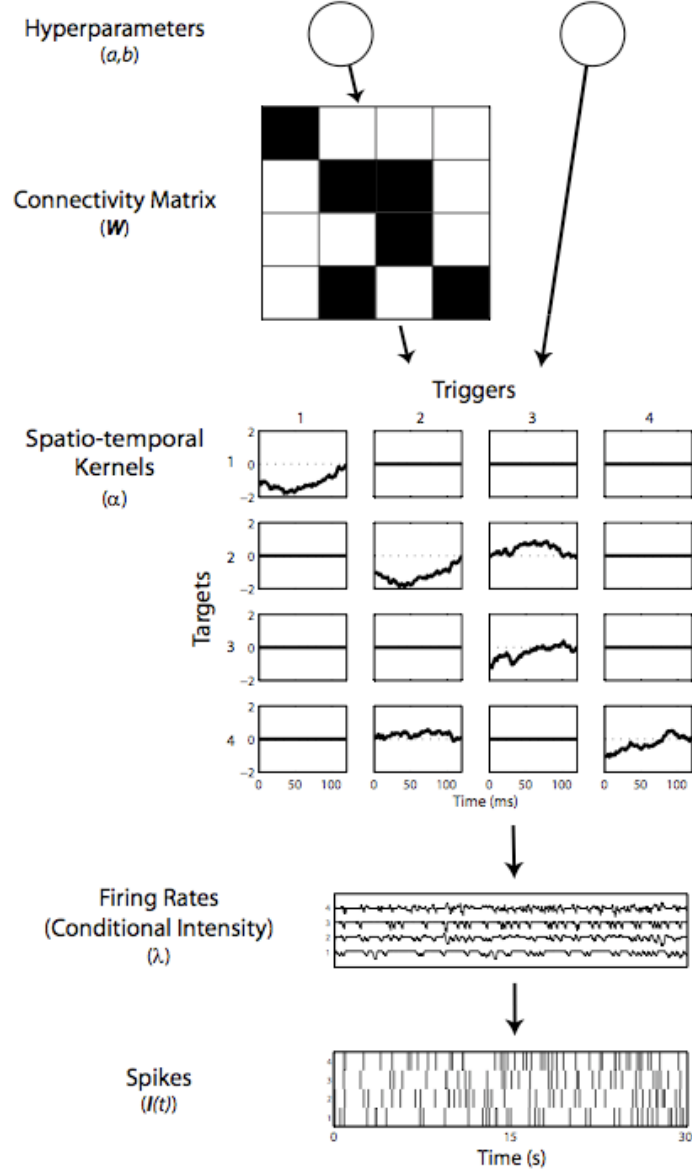


FIG. 5: Figure from Ref. [11] depicting their model. A connectivity matrix along with spatio-temporal kernels ( $\alpha$ ) describe the neural interactions. A simulation run would involve calculating the firing rates of each neuron ( $\lambda$ ) from its intrinsic properties and the effects of other neurons, which is then translated into a stochastic set of spike times. The inverse problem is finding the interaction kernels from a set of experimental spike data. The hyperparameters ( $a, b$ ) specify the strength of the priors used in the reconstruction [see Eq. (4)].

Using terminology from the statistics community, these neural models are usually assessed in terms of a goodness-of-fit measure known as the likelihood, the probability of obtaining the measured data given the model and a set of parameters. In contrast to fits to continuous-time

data, where the likelihood is usually written in terms of sums of squared residuals,<sup>4</sup> the likelihood of a stochastic point process is written as a function of the conditional intensity  $\lambda$  and the times of observed spiking  $N_i(t)$  for every neuron  $i$ . The log of the likelihood (or log-likelihood) turns out to be [11]

$$\log p(N(t) \mid \alpha) = \sum_i \left( \int_0^T \log \lambda_i(t \mid \alpha_i) dN_i(t) - \int_0^T \lambda_i(t \mid \alpha_i) dt \right). \quad (2)$$

The goal is then to find the set of parameters  $\hat{\alpha}$  that give the maximum likelihood. These maximum likelihood parameters then represent our best guess for the underlying connectivity of the network (or at least the best model for predicting the network behavior), and are not limited to using only pairwise correlations. Even though the log-likelihood is easy to compute using the above formula, searching for the maximum in such a high-dimensional space (which may have tens of thousands of dimensions) is a non-trivial task.

Other techniques from statistics are used as well; for instance, confidence intervals are found for the inferred parameters by calculating the Hessian of the log-likelihood function. In statistics language, this is the Fisher Information [5]:

$$I(\hat{\alpha}) = -\nabla^2 \log p(\hat{\alpha} \mid N(t)). \quad (3)$$

In the large-sample limit, the distribution of the maximum likelihood estimate  $\hat{\alpha}$  is Gaussian with covariance matrix  $I(\hat{\alpha})^{-1}$ ; thus  $I(\alpha)$  provides a measure of the range of allowable parameters.

Two main problems arise in such fits to large models. First, the high dimensionality of the parameter space makes the search for a solution difficult. Second, given that an adequate solution can be found with a reasonable amount of computing time, these problems are typically ill-constrained enough that many solutions are compatible with the data — can we find the “correct” solution?

### 1. *The problem of complexity and dimensionality in the search for solutions*

A major limitation to complex network models is that the number of parameters to be specified is typically large, such that, even if the model itself is easy to evaluate, finding the solution that best matches the input by exhaustively searching parameter space is impossible. This familiar problem is tackled in much the same way as inverse problems in other fields, with algorithms that attempt to search parameter space in an intelligent way.

---

<sup>4</sup> This is a consequence of the assumption of Gaussian error bars.



In the literature for inferring neural network parameters, methods from statistics are most prevalent, and most involve taking advantage of the structure of the problem by noting that calculating conditional probabilities is easier in a network by holding some nodes' parameters constant. The simplest such method, known as coordinate ascent, varies one or a few parameters while holding the others fixed [11, 16]. The Expectation-Maximization (EM) method is often used in maximum likelihood searches (e.g., Refs. [9, 17]), an iterative technique that uses intermediate values for some parameters to find closed-form expressions for maximizing the likelihood expectation with respect to other parameters. Another method used for neural networks is the "Gibbs sampler," a version of the Metropolis-Hastings algorithm that takes random steps in parameter space, also holding some parameters fixed in order to simplify the calculation [8].

A method championed by Mézard makes use of the locally tree-like structure of many networks, iteratively updating local "beliefs" of the true network structure in a way that would give an exact result for a tree graph. This method is shown to be more efficient than others, but can still run into trouble if the network is, as the authors say, "loopy" [18]. This work demonstrates a connection with other constraint satisfaction problems, providing interesting possible connections with ideas such as satisfiability (In what cases is network inference exponentially hard?) and spin glasses (Are neural networks "glassy," with many metastable states?).

## 2. *The problem of incomplete information: How close can we get to the "true" solution?*

Besides the computational complexity of searching for a solution, there is a problem that is perhaps more substantial: while an apparent goal would be to learn the actual topology and parameters describing the physiological network, we cannot expect to be able to fully characterize any neural network with current experimental methods. Can we make an inferred model that at least approximates "reality?"

There are at least two ways of tackling this problem: (1) use other information or beliefs to narrow down the possibilities and produce a full network that is as "realistic" as possible, or (2) accept the incompleteness of the model information and try for an intermediate sort of description.

Under the first approach, the model is ill-constrained: since there are many more parameters than measurements to constrain them, more information is needed to narrow down the possibilities. Although a search method may be able to find a solution, a common symptom of having insufficient data is over-fitting, where a model can fit well to a given dataset, but will poorly predict data that it has not seen.

In a Bayesian framework, the process of introducing additional constraints on a solution is achieved by adding a prior, representing expectations about solutions that are known beforehand to be more likely. This idea has been used in recent studies in the form of sparsity and smoothness priors. A sparsity prior represents the belief that any given neuron should interact with only a relatively small number of other neurons. This belief comes from other studies that show that small ensembles of neurons do have sparse functional connectivity [11]. A smoothness prior represents the belief that the degree of interaction between two neurons should vary smoothly in time; this is also backed up by other neural studies. Priors are incorporated into the Bayesian framework via a probability distribution that represents the best-guess likelihood of parameters before the information from the experiment is taken into account. In the above example, these priors could be written as [11]

$$p(\alpha_{i,c}) = \frac{1}{Z_{i,c}} \exp \left( -a \sum_m (\alpha_{i,c,m} - \alpha_{i,c,m-1})^2 - b \sum_m |\alpha_{i,c,m}| \right). \quad (4)$$

The first term gives a higher weight to smooth interaction kernels, and the second term favors smaller interactions (ideally zero) to encourage sparsity. The parameters  $a$  and  $b$  give the weighting of each prior with respect to the experimental likelihood, and must be chosen somewhat arbitrarily.<sup>5</sup> The results can be tested by comparing the fit model with spike train data that was not included in the fit. This approach has been shown to produce models that better fit unseen data, and it has successfully reproduced the clustering of neural connections inside anatomically distinct regions [11].

Even with priors to help constrain parameters, larger models become unwieldy so quickly that essentially all neural network models that attempt to fit real neural data do not simulate any more neurons than the number that were measured.<sup>6</sup> In this case, modelers are not searching for the connectivity of the full network, but rather for the “functional connectivity” [11, 21, 22] or “effective” interactions [23] between the measured neurons. These effective models reproduce the output of the measured neurons without trying to simulate every neuron in the system. This simplification is not overlooked in the neuroscience community, and the literature is full of concessions that their models do not reflect the actual underlying connectivity:

---

<sup>5</sup> The robustness of results to different prior weightings is often calculated in order to reduce anxiety about this arbitrariness (e.g., [11]). One recent study of a gene network model featured a particularly striking robustness of network parameters to prior weights [19].

<sup>6</sup> In fact, I was able to find only one researcher who seems to be interested in directly tackling the “unmeasured neuron” problem. Nykamp, a mathematician, describes our knowledge of the neural connectivity as being only up to “subpopulation ambiguity,” in that we simply cannot distinguish different neurons in small subpopulations. He points out: “We are unaware of others who explicitly address the effects of unmeasured neurons” [20].

Although it provides an accurate functional description of correlated spike responses, the generalized linear model does not reveal the biophysical mechanisms underlying the statistical dependencies between neurons: coupling does not necessarily imply anatomical connections between cells... — Pillow *et al.* [24]

Aertsen *et al.* realized that it is impossible to uniquely determine the underlying, “true” connectivity of a neural circuit without “exhaustive enumeration of all states of all elements.” What we infer is an “abbreviated description of an equivalent class of neural circuits,” ... a reconstruction of the circuit that best reproduces the observed spikes. — Stevenson *et al.* [11], quoting Aertsen *et al.* [25]

Moving even further in the direction of effective models, a recent stream of work has neglected time-dependence, directionality of interactions, and the expected sparsity of the network, using a model with simple pairwise interactions between every pair of neurons. Somewhat surprisingly, this type of model has been shown to reproduce observed spike trains quite well [16, 23, 26].

In this line of research, the models attempt to predict the overall distribution of spike firing patterns (see Fig. 6), and do not include any time dependence. The simplest such model is one in which each neuron has an independent mean rate of firing, and there is no correlated behavior. This can be shown to poorly predict the distribution of firing patterns (Fig. 6, right column). The next more complicated model would match not just the mean rates of firing, but also the pairwise correlations between every pair of neurons (Fig. 6, middle column). One could imagine next matching triplet correlations, then quartic correlations, and so on, eventually up to a model that explicitly includes the probability of every possible firing pattern (with combinatorially more terms to keep track of for each more complex model). The surprising fact is that multiple studies fitting data from retinal neurons have shown that the pairwise model accounts for over 90% of the information content of the full distribution, measured by comparing the entropy of the pairwise distribution with that of the full distribution [23, 26]. One of the studies further limited the interactions between neurons to pairwise *and* physically adjacent, and measured nearly all the neurons of a certain type in a small region of the retina; they found that they could account for 98% of “departures from statistical independence” [26]. In other words, in order to predict the firing patterns of a neural network, pairwise interactions are enough to get it mostly right.

Requiring that the model distribution fit mean rates and pairwise correlations is not enough to fully specify a model, so the standard trick is to use the distribution with the maximum entropy,

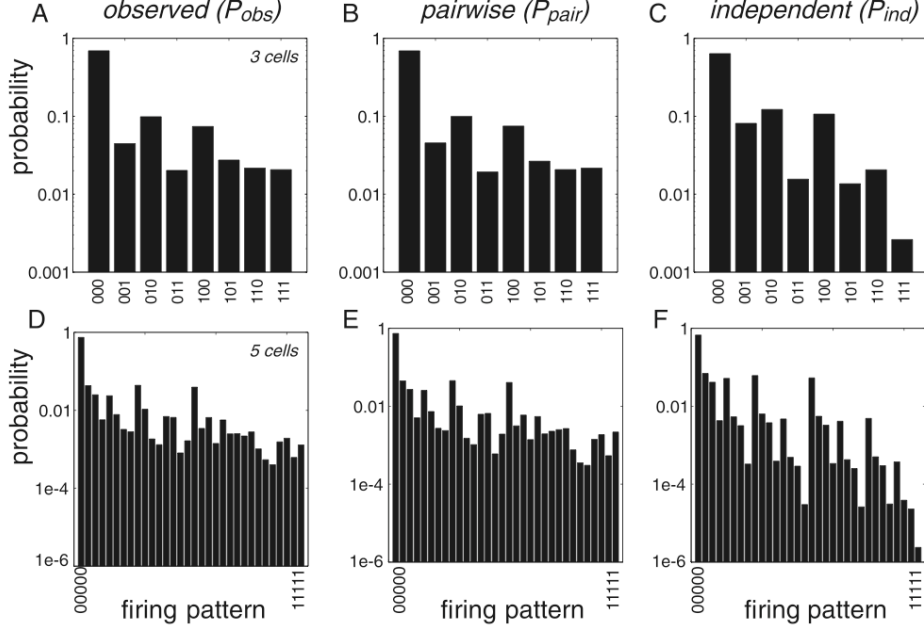


FIG. 6: Maximum entropy pairwise models explain the data well; figure from Ref. [26]. Shown are distributions of firing patterns from retinal neuron data: spike trains are first binned into time slices, and a firing pattern is defined by which neurons fire in each time slice. The left column shows the observed distributions, and the right two columns show predicted distributions from maximum entropy pairwise and independent models. The models are fit to spike train data from the same neurons, but distinct from the data shown in the left column. The top row shows firing patterns for 3 cells; the bottom row, 5 cells. Note the impressive match of the pairwise model to the data compared to the independent model.

a way of ensuring that the model does not produce any unwanted higher-order correlations.<sup>7</sup> It is interesting to notice that the maximum entropy second order (pairwise) theory corresponds to the Ising model [18, 23]; treating a spiking neuron as a spin with  $\sigma = +1$  and a non-spiking neuron as a spin with  $\sigma = -1$ , the maximum entropy second order distribution is [23]

$$p(\sigma_1, \sigma_2, \dots, \sigma_N) = \frac{1}{Z} \exp \left( \sum_i h_i \sigma_i + \frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j \right), \quad (5)$$

where the “local fields”  $h_i$  and “exchange interactions”  $J_{ij}$  are Lagrange multipliers that must be set to match the observed mean rates and pairwise correlations. (Note the similarity to Eq. (1); the only real difference is the time dependence of the interaction kernel).

Proponents of this method claim that, while it is further from direct estimation of the physiological basis of neural behavior, it promises to greatly simplify the task of modeling neural networks,

<sup>7</sup> This could be thought of as a prior favoring less informative models.

at least from a pragmatic predictive standpoint. It is important to point out that, besides being able to predict the firing pattern distributions, the approach has produced other interesting predictions. For instance, the magnitude of the observed interactions is small for small networks ( $J_{ij}$  is typically small compared to  $h_i$ ), but if the same interaction strengths are found in larger networks, an extrapolation predicts that groups of cells larger than about 200 will have very correlated behavior. This may correspond to a kind of error-correcting redundancy in the sensory system [23].

With a focus on models that describe correlations among the spike trains of only the measured neurons, it seems that the neural network community has dealt with incomplete information by aiming for effective models that can still produce interesting predictions.

### C. Challenges and limitations of current models

Compounding the problem of limited information, there are other uncertainties in current methods that limit the knowledge that can be gleaned from neural recordings; these have been recognized as possible problems, but have not yet been fully resolved.

One practical example is known as the problem of “spike sorting.” Since extracellular electrodes often pick up spikes from multiple nearby neurons, the usual first step in analysis of spike train data is determining how many neurons are being recorded, and then sorting out which neuron fired which spike (using, for example, the strength and shape of each spike). Many algorithms have been developed for this process, but it is unclear exactly how well they work; the ambiguity in spike sorting is often listed as a caveat in these studies [9, 14].

In addition, it is unclear whether the point process models used may be overly simplified. The relevant “state” of a neuron may be more complicated than simply a propensity for firing. Also, to facilitate statistical analysis, stationarity is usually assumed, such that the network and its dynamics do not change over time, even though adaptation is known to be important on the timescales of many neural recordings [14]. Finally, it may be important to go beyond functional connectivity and simulate all neurons in the network.

## III. COMPARISON WITH METHODS USED IN GENE NETWORK STUDIES

The study of neural networks has close similarities to the study of other biological networks, in particular gene transcription networks. Although the intrinsic stochasticity of neural models

makes the analysis somewhat different, modelers of gene networks must deal with the same issues that arise due to the complexity of the networks involved, and the fact that it is impossible to directly measure every interaction. First, gene networks present the same sort of inverse problem of fitting a large model to data, necessitating the use of intelligent searching schemes. Also, similar statistical techniques, such as Fisher Information, are used to more rigorously assess the significance of inferred parameters (e.g., [27]).

Network inference is a highly developed topic in the gene interaction literature. It is generally difficult to get dynamic measurements from standard measurement techniques (e.g., microarrays), making explicit dynamical modeling and fitting difficult. Instead, correlations between the variations in gene expression levels in different cellular conditions can be used to estimate connections between genes in much the same way as correlations between spike trains are used in neural networks. An example of a prominent analysis scheme for such network inference in gene networks is the ARACNE project [28]. Here, the problem of ambiguity of network structure produced by pairwise methods is recognized as important, and ideas of mutual information are used to remove as many connections as possible (resembling a sparsity prior) in an attempt to avoid false-positive connection inferences. The same Ising model formalism has been recognized as being useful for describing statistical correlations in this context as well [18, 28]. Approaches that explicitly use the idea of maximum entropy are also beginning to be used, and it is seen here also that pairwise interactions can explain most of the important genetic interactions [29].

A fundamental difference between analysis of gene networks and neural networks is in the questions that are being asked. In gene networks, the implicit goal has always been to characterize the set of genes and gene interactions that actually produce the observed behavior in the living cell, which are assumed to be the same for every such cell. In neuroscience, on the other hand, it is understood that not every brain will be wired in the same way. Thus the goal here is less focused. Although it would be beneficial to know every interaction between every neuron, this seems an unreachable goal (while, if most gene interactions are the same in every cell, it may be feasible to get enough information to eventually pin them all down). Then it seems reasonable to instead find a framework that can at least predict the output of a network, even if the model does not include every neuron, but instead aims to be an effective model. In short, while the discovery of specific gene interactions is most important in gene network studies, it is unknown what the most important aspects of neural networks are, and effective models may be good enough.

## IV. CONCLUSION

The characterization of biological neural networks presents an open and complex challenge. Early attempts looked only at the small scale correlations between pairs of neurons. The development of multielectrode arrays led to many-neuron models; these have presented a framework that revealed the challenges inherent in having an incomplete view of the underlying interacting elements. First, methods for intelligent searching have had to be developed for finding solutions to the inverse problem of fitting many-parameter models to data. More fundamentally, the lack of full information has forced the development of methods to either guess at probable realistic solutions using priors, or to redefine the goal and look for functional interactions and effective models. The analysis of gene transcription networks presents some of the same challenges, and similar methods have been applied. Ultimately, whether or not they can identify every individual interaction, these techniques will be tested by their ability to make useful predictions.

- 
- [1] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, 1999).
  - [2] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (MIT Press, 2007).
  - [3] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2001).
  - [4] P. Achard and E. de Schutter, PLoS Computational Biology **2**, e94 (2006).
  - [5] E. N. Brown, R. Barbieri, U. T. Eden, and L. M. Frank, *Likelihood Methods for Neural Spike Train Data Analysis* (CRC Press, Boca Raton, 2003), chap. 9, pp. 253–286.
  - [6] W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland, Science **252**, 1854 (1991).
  - [7] J. Krüger, Rev. Physiol. Biochem. Pharmacol. **98**, 177 (1983).
  - [8] F. Rigat, M. de Gunst, and J. van Pelt, Bayesian Analysis **1**, 733 (2006).
  - [9] M. Okatan, M. A. Wilson, and E. N. Brown, Neural Computation **17**, 1927 (2005).
  - [10] D. H. Perkel, G. L. Gerstein, and G. P. Moore, Biophys. J. **7**, 419 (1967).
  - [11] I. H. Stevenson, J. M. Rebesco, N. G. Hatsopoulos, Z. Haga, L. E. Miller, and K. P. Körding, IEEE Trans. on Neural Systems and Rehab. Eng. (in press) (2008).
  - [12] G. L. Gerstein and A. M. H. J. Aertsen, J. Neurophys. **54**, 1513 (1985).
  - [13] S. N. Baker and G. L. Gerstein, Neural Computation **12**, 2597 (2000).
  - [14] E. N. Brown, R. E. Kass, and P. P. Mitra, Nature Neuroscience **7**, 456 (2004).
  - [15] M. Abeles and G. L. Gerstein, J. Neurophys. **60**, 909 (1988).

- [16] T. Broderick, M. Dudík, G. Tkačik, R. E. Schapire, and W. Bialek, arXiv:0712.2437v2 (2007).
- [17] E. S. Chornoboy, L. P. Schramm, and A. F. Karr, *Biol. Cybern.* **59**, 265 (1988).
- [18] M. Mézard and T. Mora, arXiv:0803.3061v1 (2008).
- [19] S. Mukherjee and T. P. Speed, *PNAS* **105**, 14313 (2008).
- [20] D. Q. Nykamp, *Math. Biosciences* **205**, 204 (2007).
- [21] I. H. Stevenson, J. M. Rebecsco, L. E. Miller, and K. P. Körding, *Current Opinion in Neurobiology* (in press) (2008).
- [22] T. Gürel, U. Egert, S. Kandler, L. de Raedt, and S. Rotter, *Proceedings of International Joint Conference on Neural Networks* (2007).
- [23] E. Schneidman, M. J. B. II, R. Segev, and W. Bialek, *Nature* **440**, 1007 (2006).
- [24] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, *Nature* **454**, 995 (2008).
- [25] A. M. H. J. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm, *J. Neurophys.* **61**, 900 (1989).
- [26] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, *J. Neuroscience* **26**, 8254 (2006).
- [27] F. Casey, D. Baird, Q. Feng, R. N. Gutenkunst, J. J. Waterfall, C. R. Myers, K. S. Brown, R. A. Cerione, and J. P. Sethna, *IET Syst. Biol.* **1**, 190 (2007).
- [28] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano, *BMC Bioinformatics* **7**, S7 (2006).
- [29] T. R. Lezon, J. R. Banavar, M. Cieplak, A. Maritan, and N. V. Fedoroff, *PNAS* **103**, 19033 (2006).