Pinpointing connectivity despite hidden nodes within stimulus-driven networks

Duane Q. Nykamp
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455, USA
(Received 20 July 2007; revised manuscript received 11 June 2008; published 6 August 2008)

The effects of hidden nodes can lead to erroneous identification of connections among measured nodes in a network. For example, common input from a hidden node may cause correlations among a pair of measured nodes that could be misinterpreted as arising from a direct connection between the measured nodes. We present an approach to control for effects of hidden nodes in networks driven by a repeated stimulus. We demonstrate the promise of this approach via simulations of small networks of neurons driven by a visual stimulus.

DOI: 10.1103/PhysRevE.78.021902 PACS number(s): 87.19.L−, 87.18.Sn

I. INTRODUCTION

Determination of the connectivity structure of complex networks is hindered by one's inability to simultaneously measure the activity of all nodes. In experiments probing networks such as gene regulatory networks, computer networks, or neural networks, many hidden nodes could be interacting with the small set of measured nodes and corrupting estimates of connectivity in unknown ways. For example, if a hidden node had connections to two measured nodes, this common input could introduce correlations among the measured nodes, which might lead one to erroneously infer a connection between the measured nodes. Since, in many applications, one cannot simultaneously measure more than a tiny fraction of nodes, determining network structure is a challenge.

We present a promising approach to control for effects of hidden nodes and estimate connectivity among measured nodes of stimulus-driven networks. In particular, we can distinguish between causal connections[1] and common input originating from hidden nodes. Earlier versions [1–3] relied on models of the relationship between nodal activity and the stimulus. Our result here exploits a repeated stimulus to eliminate the need for such a model, greatly broadening the applicability of the analysis.

II. HISTOGRAM AND HISTORY MODEL

With a repeated stimulus, one can sample a node's activity to estimate its probability distribution conditioned on each stimulus time point. For simplicity, we assume a probability distribution determined by its mean[2] so that one needs only the average activity at each stimulus time. This average is analogous to the peristimulus time histogram (PSTH) commonly used to capture neurons' spiking activity as a function of stimulus time, and we will use the term PSTH as a shorthand for any such average activity. An important point is that one can estimate the PSTH without understanding how nodal activity is related to the stimulus.

We use the framework of Ref. [3] to build a model of the probability distribution of a node's activity conditioned not only on the stimulus, but also on the node's history and other nodes' activity. Let $R_s^{k,i}$ be the activity of node s at time i during stimulus repeat k. Ignoring for a moment the influence of other nodes, we allow the probability distribution of $R_s^{k,i}$ to depend on time i (but not explicitly on stimulus repeat k) and $\mathbf{R}_s^{k,<i}$, which represents all activity $R_s^{k,j}$ of node s at times $j<i$. For example, if we assume $R_s^{k,i}$ is a Poisson random variable (approximating, for example a point process [4]) that depends on its history $\mathbf{R}_s^{k,<i}$ linearly, we could write

$$\Pr(R_s^{k,i}|\mathbf{R}_s^{k,<i}) = \Gamma\left(R_s^{k,i}, g_s\left(P_s^i + \sum_{j>0} h_s^j R_s^{k,i-j}\right)\right), \tag{1}$$

where h_s^j is a kernel specifying the node's history dependence, $g_s(\cdot)$ is some (typically monotonic) function of its scalar argument, and

$$\Gamma(n,\lambda) = \frac{1}{n!}\lambda^n e^{-\lambda} \tag{2}$$

is the Poisson distribution. Since we do not assume a particular relationship between the stimulus and nodal activity, the function P_s^i depends arbitrarily on time point i but not on stimulus repeat k. The dependence of P_s^i on i is chosen so that the average of $R_s^{k,i}$ over all stimulus repeats k matches the PSTH.

We use the term histogram and history (HAH) model to refer to models that incorporate a dependence of a node's activity on its own history and also allow arbitrary dependence on a stimulus to match the PSTH. Model (1) is an example of a HAH model, specialized to the case where the expected value of $R_s^{k,i}$ is a linear function of the model parameters, except for the nonlinearity g_s. Such a nearly linear model is called a generalized linear model (GLM).

In a network, a node's activity could depend on the previous activity $\mathbf{R}^{k,<i}$ of all nodes. We define the directed graph of the network by letting $\bar{W}_{\tilde{s},s}^j$ be the (possibly zero) coupling strength from node $\tilde{s}$ onto node s at delay j. Adding linear coupling to the linear history in our example (1), we form the network HAH model,

[1] By causal connection, we refer to a directed path between measured nodes, possibly via one or more hidden nodes.

[2] Depending on the amount of data, one could also estimate higher moments or the full conditional probability.

$$\Pr(R_s^{k,i}|\mathbf{R}^{k,<i}) = \Gamma\left(R_s^{k,i}, \bar{g}_s\left(\bar{P}_s^i + \sum_{j>0} \bar{h}_s^j R_s^{k,i-j} + \sum_{\tilde{s}\neq s}\sum_{j>0} \bar{W}_{\tilde{s},s}^j R_{\tilde{s}}^{k,i-j}\right)\right), \quad (3)$$

where we have added bars over quantities to indicate they will be different from those in (1).

This analysis is not limited to GLMs such as (1) and (3), but can be applied to a more general class of HAH network models of the form

$$\Pr(R_s^{k,i}|\mathbf{R}^{k,<i}) = \bar{P}_s\left(R_s^{k,i}; i, \mathbf{R}_s^{k,<i}, \sum_{\tilde{s}\neq s}\sum_{j>0} \bar{W}_{\tilde{s},s}^j R_{\tilde{s}}^{k,i-j}\right), \quad (4)$$

where $\bar{P}_s$ is a probability distribution in its first argument that depends on time point i, the node's own history $\mathbf{R}_s^{k,<i}$, and the total coupling from other nodes summed linearly. (Since we will assume that the coupling $\bar{W}_{\tilde{s},s}^j$ is a small parameter, linear coupling will suffice.)

In the following description of the network analysis, we will leave the formulation of the HAH model generic in (4). However, to apply the analysis, one must choose a model that specifies the dependence of $R_s^{k,i}$ on history $\mathbf{R}_s^{k,<i}$, such as the linear form of (1) and (3). Since every possible history vector will not occur multiple times, a model-independent description of history dependence is unavailable. Nonetheless, history dependence is a single-node property, so development of suitable models is more tractable than for models of stimulus dependence (which could depend on the entire network).

III. DESCRIPTION OF NETWORK ANALYSIS

Our goal is to estimate the connectivity among measured nodes in the network despite the presence of hidden nodes. We assume the activity of each node can be represented by a model of the form (4), though the probability distribution may vary among nodes. If we let $\mathbf{R}$ denote the activity of all nodes s at all stimulus repeats k and all time points i, we can use Bayes theorem repeatedly to write the probability distribution of $\mathbf{R}$ in terms of model (4) as

$$\Pr(\mathbf{R}) = \prod_{s,k,i} \Pr(R_s^{k,i}|\mathbf{R}^{k,<i}) = \prod_{s,k,i} \bar{P}_s\left(R_s^{k,i}; i, \mathbf{R}_s^{k,<i}, \sum_{\tilde{s}\neq s}\sum_{j>0} \bar{W}_{\tilde{s},s}^j R_{\tilde{s}}^{k,i-j}\right). \quad (5)$$

We assumed that for a given time point i and stimulus repeat k, all nodes were independent conditioned on network history (i.e., we assumed that all network interactions occur at a delay of at least one time step).

Since model (5) contains many hidden nodes, we cannot determine its parameters. But we can fit a model to all the activity $\mathbf{R}_s$ of a single measured node, such as model (1). Using the same formalism as (5), we can write a generic HAH model for a single node's activity $\mathbf{R}_s$ as

$$\Pr(\mathbf{R}_s) = \prod_{k,i} \Pr(R_s^{k,i}|\mathbf{R}^{k,<i}) = \prod_{k,i} P_s(R_s^{k,i}; i, \mathbf{R}_s^{k,<i}, 0). \quad (6)$$

The single-node HAH model (6) is similar to (5), except that we have ignored the activity of all nodes except for one (setting $R_{\tilde{s}}^{k,i}=0$ for $\tilde{s}\neq s$). Since, of course, the single-node model will have different parameters, we denoted the change by removing the bar over the probability distribution.

We assume that the HAH model chosen for (6) is an identifiable model, meaning that we can identify all of its parameters from measurements of the activity of a single node in response to the stimulus. Since the dependence on time point i is arbitrary (presumably subject to some smoothness constraints), we imagine the parameters will be chosen so that the average of the activity $R_s^{k,i}$ over all stimulus repeats k will closely approximate the PSTH, hence motivating the use of the term HAH model. Regardless, the key assumption is that we can regard all the parameters of (6) as being known for each measured node, as these parameters can be obtained by fitting the model to activity of each measured node.

Relying on the previously detailed general procedure [3], we just give a brief sketch of the network analysis here. To proceed with the analysis, we want to link the single node model (6) to the network model (5). The left-hand side of (6) is simply the marginalization of the left-hand side of (5) over the activity of all other nodes $\tilde{s}\neq s$. For this reason, we refer to model (6) as the average model. For consistency, we need the product of the P_s to approximate the marginalization of the product of the $\bar{P}_s$.

We can analytically compute the marginalization of (5) if we assume that the coupling $\bar{W}_{\tilde{s},s}^j$ is weak.[3] We expand (5) to second order in $\bar{W}$, and explicitly sum over all possible combinations of the values of $R_{\tilde{s}}^{k,i}$ for $\tilde{s}\neq s$. As detailed in Ref. [3], each resulting term depends on the activity $R_{\tilde{s}}^{k,i}$ in terms of a polynomial in $R_{\tilde{s}}^{k,i}$ multiplied by a probability distribution $P_{\tilde{s}}$ or its derivative. The sum over all possible values of $R_{\tilde{s}}^{k,i}$ can be turned into a combination of moments of the probability distribution and related quantities. By equating the result of this analytic (though approximate) marginalization of (5) to the average model (6), we obtain an expression relating the unknown parameters of (5) to the known parameters of (6).

Although the parameters of the averaged model are not known for hidden nodes, we perform the above marginalization for each node and then rewrite the network model (5) in terms of the parameters from the averaged model (6). Then, we perform yet another marginalization of (5), this time marginalizing over the activity of just the hidden nodes. The result is an expression for the probability distribution of just the measured nodes, which we write as $\mathbf{R}_{\mathcal{Q}}$, where $\mathcal{Q}$ denotes

[3] We assume $\bar{W}_{\tilde{s},s}^j$ is a small parameter scaling with network size N as $1/N$ [2].

the set of indices q corresponding to measured nodes. The resulting probability distribution for measured node activity $\mathbf{R}_{\mathcal{Q}}$ is written in terms of the parameters of the averaged model (6). Of course, all the parameters corresponding to the arbitrary number of hidden nodes remain unknown. Since all the coupling parameters $\bar{W}^j_{\tilde{s},s}$ are also unknown, the parameters for the probability distribution of $\mathbf{R}_{\mathcal{Q}}$ are still underdetermined.

To close the system, we make one more approximation. Because we averaged over hidden node activity, the marginal probability distribution of $\mathbf{R}_{\mathcal{Q}}$ contains hidden node quantities only in terms of averages at different stimulus time points. These hidden node quantities vary with stimulus time point in unknown ways, as we do not know the structure of hidden node PSTHs. To simplify the equations, we simply ignore this unknown PSTH structure. Mathematically, the approximation is assuming all the hidden node PSTHs are flat, though one can justify this approximation with a weaker assumption that hidden node PSTH structure differs significantly from the PSTH structure of the measured nodes [2]. By ignoring the PSTH structure of hidden nodes, the averaged quantities from the hidden nodes no longer depend on the stimulus time point. This reduces the number of unknown quantities to a tractable number because it turns out that we can group all remaining hidden node quantities into two effective quantities.

The first hidden node quantity represents common input from hidden nodes onto a pair of measured nodes. Although common input could arise from an unspecified number of hidden nodes, the above approximation allows us to lump all of such effects into the common input measure that we denote by $U^j_{\tilde{q},q}$. The measure $U^j_{\tilde{q},q}$ is a sum over the average effect of common input connections $\bar{W}^{\tilde{j}-j}_{p\tilde{q}}$ and $\bar{W}^{\tilde{j}}_{pq}$ from hidden nodes p. These common input connections create a correlation between the activity $R^{k,i}_q$ of node q and the activity $R^{k,i-j}_{\tilde{q}}$ of node $\tilde{q}$ with a delay of j.

The second way that hidden nodes can create correlation among measured nodes is through an indirect chain of connections from one measured node $\tilde{q}$ to hidden node p (i.e., $\bar{W}^{\tilde{j}_1}_{\tilde{q}p}$) and then from hidden node p onto a second measured node q (i.e., $\bar{W}^{\tilde{j}_2}_{pq}$). We refer to such a chain of connections as consisting of a causal connections from node $\tilde{q}$ to node q. When we ignore the PSTH structure of hidden nodes, such indirect causal connections via hidden nodes appear in the probability distribution of $\mathbf{R}_{\mathcal{Q}}$ in exactly the same way as do the direct causal connections from $\tilde{q}$ to node q (i.e., $\bar{W}^j_{\tilde{q}q}$). Therefore, we cannot distinguish direct and indirect causal connections, and the second lumped quantity involving hidden nodes includes the direct causal connection along with the sum of all indirect causal connections via hidden nodes. (Since we computed only up to quadratic terms in the $\bar{W}$, the approximation captures just causal chains of two links.) We denote this total causal connection quantity by $W^j_{\tilde{q}q}$, where the absence of the overbar indicates that it is an effective parameter.

We end up with the following expression for the marginal probability distribution of the activity $\mathbf{R}_{\mathcal{Q}}$ of measured nodes:[4]

$$\Pr(\mathbf{R}_{\mathcal{Q}}) = \prod_{q,k,i} \Pr(R^{k,i}_q | \mathbf{R}^{k,<i}_{\mathcal{Q}}) \approx \prod_{q,k,i} P_q(R^{k,i}_q; i, \mathbf{R}^{k,<i}_q, \tilde{W}^{k,i}_q), \tag{7a}$$

where

$$\tilde{W}^{k,i}_q = \sum_{\tilde{q}\neq q} \sum_{j>0} W^j_{\tilde{q},q}[R^{k,i-j}_{\tilde{q}} - \mathrm{PSTH}^{i-j}_{\tilde{q}}] + \sum_{\tilde{q}\neq q} \sum_{j\geq 0} U^j_{\tilde{q},q} \frac{\partial P^{k,i-j}_{\tilde{q}}}{\partial w} \frac{1}{P^{k,i-j}_{\tilde{q}}}. \tag{7b}$$

We let $\mathbf{R}^{k,<i}_{\mathcal{Q}}$ represent previous activity of just the measured nodes and use the notation convention that indices q and $\tilde{q}$ correspond only to measured nodes. $\mathrm{PSTH}^i_{\tilde{q}}$ is the PSTH of node $\tilde{q}$ at time i (i.e., the expected value of $R^{k,i}_{\tilde{q}}$, which is independent of stimulus repeat k). We also used the shorthand notation for quantities from the averaged model (6)

$$P^{k,i}_q = P_q(R^{k,i}_q; i, \mathbf{R}^{k,<i}_q, 0),$$

$$\frac{\partial P^{k,i}_q}{\partial w} = \frac{\partial}{\partial w} P_q(R^{k,i}_q; i, \mathbf{R}^{k,<i}_q, w)|_{w=0}.$$

Note that (7) employs the averaged model (6) so that average effects of connections are included even when $\tilde{W}=0$. The factor $\tilde{W}^i_q$ represents deviations from the average that induce correlations between node q and the other measured nodes $\tilde{q}$. Given a determination of the P^i_q from fitting the averaged model (6) for each measured node q, the only unknowns are the W and U. We use the logarithm of (7) to find maximum likelihood estimates of W and U.

More intuition behind the difference between W and U can be obtained by specializing to the case where the probability distribution P_q is a Poisson distribution (2). If $P_q(R^{k,i}_q; i, \mathbf{R}^{k,<i}_q, w) = \Gamma(R^{k,i}_q, \lambda^i_q(\mathbf{R}^{k,<i}_q, w))$, then we can rewrite the expression for $\tilde{W}$ as[5]

$$\tilde{W}^i_q = \sum_{\tilde{q}\neq q} \sum_{j>0} W^j_{\tilde{q},q}[R^{k,i-j}_{\tilde{q}} - \mathrm{PSTH}^{i-j}_{\tilde{q}}] + \sum_{\tilde{q}\neq q} \sum_{j\geq 0} U^j_{\tilde{q},q}[R^{k,i-j}_{\tilde{q}} - \lambda^{i-j}_{\tilde{q}}(\mathbf{R}^{k,<i-j}_{\tilde{q}}, 0)] \frac{(\partial/\partial w)\lambda^{i-j}_{\tilde{q}}(\mathbf{R}^{k,<i-j}_{\tilde{q}}, w)|_{w=0}}{\lambda^{i-j}_{\tilde{q}}(\mathbf{R}^{k,<i-j}_{\tilde{q}}, 0)}. \tag{8}$$

For a Poisson random variable, one can read out two differences between the coefficients of the causal connection factor W and the common input factor U in (8) that enable distinction between W and U. The first difference appears in the square brackets and captures how history dependence has a different effect on causal connections than on common input. If there is a causal connection, any deviation of the activity of node $\tilde{q}$ from its average (the PSTH) is felt by node

[4] Compare Eq. (3.15) of Ref. [3]. We ignore terms quadratic in W to facilitate computations. We set $U^0_{\tilde{q},q}=0$ for $\tilde{q}>q$.

[5] Compare Eq. (3.17b) of Ref. [3].

q. On the other hand, if there is only common input from a hidden node, the activity of node $\tilde{q}$ is not transmitted to node q. Activity of node $\tilde{q}$ that can be predicted by its history dependence does not reflect activity of the common input node and has no effect on node q. Such activity is canceled out by $\lambda_{\tilde{q}}^{i-j}(\mathbf{R}_{\tilde{q}}^{k,<i-j},0)$. See Ref. [3] for details.

The second difference is the final factor of (8). The HAH model captures how the stimulus modulates with time i both the average activity $\lambda_{\tilde{q}}^{i}(\mathbf{R}_{\tilde{q}}^{k,<i},0)$ and the influence of inputs $(\partial/\partial w)\lambda_{\tilde{q}}^{i}(\mathbf{R}_{\tilde{q}}^{k,<i},w)|_{w=0}$. The final factor of (8) reflects how a causal connection has a different temporal modulation than common input (which depends on node $\tilde{q}$'s response to inputs). This difference exists as long as the common input node's PSTH differs from the measured nodes' PSTHs. See Ref. [2] for details in a general, model-dependent context.

IV. DEMONSTRATION OF RESULTS

To demonstrate the capabilities and limitations of our approach, we simulate small networks of neurons and attempt to distinguish between a direct connection and common input from an unmeasured neuron. In contrast to previous implementations [1–3], the HAH-based analysis allows us to simulate neurons with relatively complicated stimulus-dependence without concern for a method to reconstruct the model. We let $\mathbf{X}^i$ represent a two-dimensional visual stimulus at time i, and use a subunit model of a visual neuron analogous to the energy model along with divisive normalization [5]. We use a binary random variable where $R_s^{k,i}=1$ corresponds to a spike of neuron s in time bin i and stimulus repeat k. The simulated model is

$$\Pr(R_s^{k,i}=1|\mathbf{R}^{k,<i})=A_s\left(\frac{\sum_{j=1}^{4} I_{s,j}^i}{1+\sum_{j=5}^{8} I_{s,j}^i}+\sum_{\tilde{s},j}\bar{W}_{\tilde{s}s}^j R_{\tilde{s}}^{k,i-j}\right)_+^2$$

where $I_{s,j}^i=B_j[\mathbf{h}_{s,j}*\mathbf{X}^i]_+$, where $*$ represents convolution, and $[y]_+=\max\{y,0\}$. We used Gabor functions for the stimulus kernels $\mathbf{h}_{s,j}$ where the phases of each set of four kernels were in quadrature and the angle of the suppressive kernels in the denominator was orthogonal to that of the numerator kernels.[6] In this way, the neurons had a fundamentally nonlinear response to the stimulus. Note that this model has a fairly linear dependence on spike history which is exactly the same form as the interneuronal coupling (the final sum includes self-coupling $\bar{W}_{ss}$ terms).[7]

We reconstruct the circuitry with HAH models of GLM form (1), using binary random variables with mean $\lambda_q^i(\mathbf{R}_q^{k,<i},w)=g_q(P_q^i+\Sigma_j h_q^j R_q^{k,i-j}+w)$ where $g_q(y)=C_q\ln(1+e^{y+d_q})$. For each measured neuron, we compute maximum likelihood fits of the average model (1) to determine C_q, d_q, P_q^i, and h_q^i.[8]

In the first set of simulations, we let the stimulus be 100 repetitions of a 5 s sequence of random sinusoidal gratings [6] where a new grating was randomly selected every 50 ms.[9] (We view each time point as 1 ms.) We set the A_s so that each neuron fired 15 to 20 spikes/s, obtaining 7000–9000 spikes per neuron.

We simulated two networks under these conditions. The first contained two neurons where neuron 2 had a direct connection onto neuron 1. The second network contained three neurons where neuron 3 had common input connections onto the other two. Note that neuron 3 is unmeasured; we discard its spikes and analyze the spikes of just the first two neurons. The correlation between neurons 1 and 2 appears identical for both networks (Fig. 1). However, our new causal connection factor W picks out the direct connection [Fig. 1(a)] and the common input factor U picks out the common input [Fig. 1(b)]. We successfully controlled for common input from a neuron that remained unmeasured.

The rich random grating stimulus creates PSTHs with significant structure [see Fig. 3(a)], which is ideal for this analysis. It also reduces the chance that the PSTH of an unmeasured neuron matches the measured neurons. Since the neurons had strong history dependence, the HAH model could exploit both stimulus-dependent and history-dependent effects to distinguish the circuitry.

As a further test, we stimulated the same networks for 2 min with a simple stimulus: a grating drifting at 5 Hz.[10] We adjusted parameters so that all neurons fired at about 30 to 40 spikes per second, obtaining 4000–5000 spikes per neuron. The model neurons responded with little temporal modulation [see PSTH of Fig. 3(b)], so that our analysis could not exploit significant stimulus dependence to distinguish circuitry. Nonetheless, strong history-dependent effects

[6] At each position $\mathbf{z}=(z_1,z_2)$ and delay of t time steps, the kernel $\mathbf{h}_{s,j}$ was proportional to $t\exp(-t/\tau^h-|\mathbf{z}|^2/2\sigma_{s,j}^2)\cos(\mathbf{k}_{s,j}\cdot\mathbf{z}+\phi_{s,j})$ with $\mathbf{k}_{s,j}=2\pi f_{s,q}(\cos\psi_{s,j},\sin\psi_{s,j})$. The kernels were normalized so that $\mathbf{h}\cdot\mathbf{X}$ had unit variance (random gratings) or were unit vectors (drifting gratings). We set $\tau^h=40$, $\sigma_{1,j}=15$, $\sigma_{2,j}=20$, $\sigma_{3,j}=10$, $f_{1,j}=0.4/\sigma_{1,j}$, $f_{2,j}=0.3/\sigma_{2,j}$, $f_{3,j}=0.5/\sigma_{3,j}$, for $j=1,\ldots,8$; $\phi_{s,1}=\phi_{s,5}=0$, $\phi_{s,2}=\phi_{s,6}=\pi/2$, $\phi_{s,3}=\phi_{s,7}=\pi$, $\phi_{s,4}=\phi_{s,8}=3\pi/2$, $\psi_{1,j}=0$, $\psi_{2,j}=\pi/4$, $\psi_{3,j}=\pi/2$, $\psi_{s,j+4}=\psi_{s,j}+\pi/2$, $B_j=1$ for $s=1,2,3$ and $j=1,\ldots,4$; $B_j=0.2$ for $j=5,\ldots,8$.

[7] We set the self-coupling terms $\bar{W}_{ss}^j=-\infty$ for an absolute refractory period of τ_s^{ref}, after which we set $\bar{W}_{ss}^j=-a_s^{\text{hist}}e^{-j/\tau_s^{\text{hist}}}$. We let $\tau_1^{\text{ref}}=2$, $\tau_2^{\text{ref}}=3$, $\tau_3^{\text{ref}}=1$, $\tau_1^{\text{hist}}=10$, $\tau_2^{\text{hist}}=12$, $\tau_3^{\text{hist}}=6$, $a_1^{\text{hist}}=5$, $a_2^{\text{hist}}=3$, and $a_3^{\text{hist}}=2$. In some simulations, we divided all a_s^{hist} by 5. The interneuronal coupling was of the form $\bar{W}_{\tilde{s}s}^j=a_{\tilde{s}s}[(j-d_{\tilde{s}s})^2/\tau_w^3]e^{-(j-d_{\tilde{s}s})/\tau_w}$ for $j>d_{\tilde{s}s}$ and zero otherwise. We set $\tau_w=2$, $d_{21}=d_{32}=0$, and $d_{31}=4$. We set $a_{21}=2$ for direct connection networks. For common input networks, we set $a_{31}=a_{32}=6$ (random gratings) or $a_{31}=a_{32}=7$ (drifting gratings). Other $a_{\tilde{s}s}=0$.

[8] Although we do not exactly fit to the PSTH, the average of $g_q(\cdot)$ over stimulus repeats does converge to the PSTH.

[9] Each grating was on a 100×100 grid and was of the form $\pm(1/100)\text{cas}(2\pi\mathbf{k}\cdot\mathbf{z}/100)$ where $\text{cas}\,y=\cos y+\sin y$, $\mathbf{k}=(k_1,k_2)$, and $k_j\in\{-5,\ldots,5\}$.

[10] We carefully selected the spatial frequency and orientation of the grating so that each neuron responded, setting $\omega=0.005$, $\mathbf{k}=(k_1,k_2)$, $k_1=k_0\cos(\phi)$, $k_0\sin(\phi)$, $\phi=\pi/4$, and $k_0=0.033$.

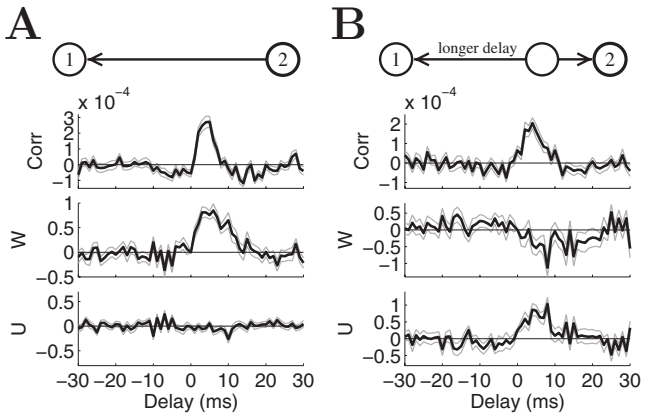

FIG. 1. Successful determination of circuitry for networks driven by random gratings. Results are shown from networks containing (a) a direct connection between two measured neurons and (b) common input from an unmeasured neuron onto two measured neurons, as schematized at top. In both cases, neuron 1's spikes are correlated with a delayed version of neuron 2's spikes, and the shuffle-corrected correlogram [7] (top panel) cannot distinguish the circuitry. The direct connection of (a) is correctly identified, as seen by the corresponding peak in the causal connection factor W. The hidden common input of (b) is correctly identified by the peak in U. (The opposite-sign reflection in W does not indicate a negative causal connection but is rather an artifact of the weak coupling assumption being violated [2]. It does not lead to ambiguity given the sign of the correlation.) Delay is the spike time of neuron 1 minus spike time of neuron 2, so we define $W^j = W_{12}^{-j}$ for $j \leq 0$, $W^j = W_{21}^j$ for $j > 0$, and equivalent for U^j. Thin gray lines indicate a bootstrap estimate of one standard error, calculated from 50 resamples.

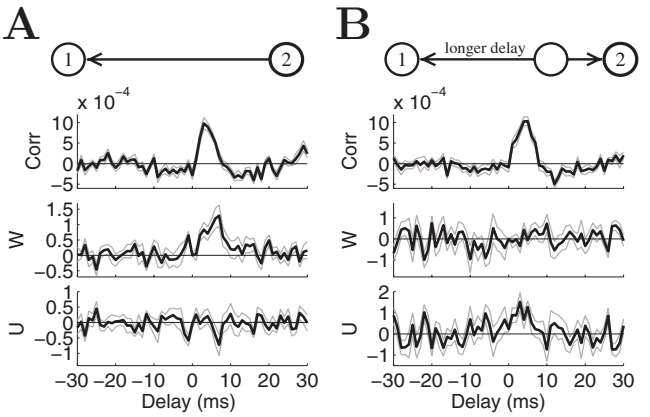

FIG. 2. Successful determination of circuitry for networks driven by drifting gratings. A direct connection network (a) and a common input network (b) are shown as in Fig. 1. Although the correlation was similar in both networks, the peak in the causal connection factor W identified the direct connection network (a) and the peak in the common input factor U identified the common input network (b). Since in this case the neurons' PSTHs showed little structure, the analysis primarily exploited the strong history dependence of the neurons' spike trains to distinguish the circuitry. Panels as in Fig. 1.

allowed the analysis to still distinguish the direct connection [Fig. 2(a)] as well as hidden common input [Fig. 2(b)].

However, if we reduced the history-dependence of the model neurons by a factor of 5 and stimulated with the simple drifting grating, the analysis could not correctly distinguish the hidden common input [Fig. 3(c)]. All neurons' PSTHs were similar with little structure, and the analysis could exploit neither stimulus dependence nor history dependence well enough to correctly determine network connectivity. We conclude that one should not apply this analysis to networks driven by a simple stimulus unless one has strong history dependence (for a point process, this means a strong deviation from a Poisson process). If one reduces the history dependence but uses the rich random grating stimulus, the analysis can still distinguish the common input network (not shown).

V. CONCLUSIONS

We have demonstrated that, if one drives a network so that the nodes have strong stimulus dependence and/or history dependence, one can control for common input arising from hidden nodes. In this way, one can pinpoint the causal connections among the measured nodes. Many methods have been used for determining causal connection in networks, including Granger causality [8], partial coherence [9], partial directed coherence [10], transfer entropy [11], mutual information [12], and other model regressions. Although some of these methods are designed to address common input, they can only control for common input arising from measured nodes. In networks with vast numbers of hidden nodes, the chances of measuring from all common input nodes is slim so that controlling for common input from measured nodes may have limited value. An alternative approach to investigating the effects of hidden nodes may be to fit a network

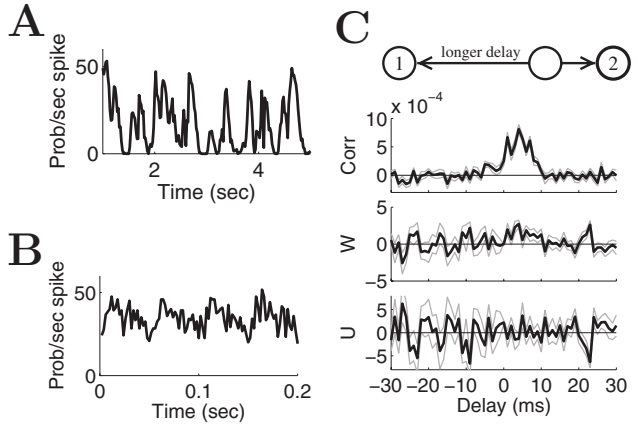

FIG. 3. (a) Sample PSTH from a neuron stimulated by random gratings. The rich structure in the PSTH allowed the analysis to exploit the neurons' stimulus dependence. (b) Sample PSTH from a neuron stimulated by drifting gratings. As the PSTH had little structure (note different temporal scale), the neurons' stimulus dependence provided little information about the circuitry. (c) The analysis was unable to determine the circuitry when neurons with little history dependence were driven by drifting gratings. W has a peak even though the correlation was due to common input from an unmeasured neuron. Panels as in Fig. 1.

model that contains a latent noise source [13].

Our analysis exploits a model-independent description of the relationship between nodal activity and the stimulus: the PSTH, or more generally, the probability distribution of nodal activity conditioned on stimulus time point. Nonetheless, making the subtle distinction between causal connections and hidden common input does require imposing models of the influence of history and other nodes. In particular, the key to success is an accurate reflection of how other nodes could influence the activity,[11] though the method is robust to small deviations (as demonstrated by our use of a different nonlinearity than used in the simulations). The approach also relies on observing sufficiently strong correlations so that they may be further decomposed into causal connection and common input components.

In the present work, we have overcome a major hurdle in the application of this approach by sidestepping the need to model how nodes respond to the stimulus (which could be a complex emergent property of the network). The modeling framework is sufficiently general to allow one to "plug in" a parametric class of models specifying how a given measured node could respond to its history and other nodes' activity. The initial successes demonstrate the promise of this approach toward pinning down network circuitry among measured nodes even in the presence of large numbers of hidden nodes.

[11]One challenge is accurately estimating the derivative with respect to coupling [last factor of (8)], needed to exploit stimulus modulation. Accurate results required carefully fitting even the factor C_s in $g_s(\cdot)$ despite it leading to a nonconvex optimization [14]. Our algorithm for determining C_s relies on history dependence.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation Grants No. DMS-0415409 and No. DMS-0719724.

[1] D. Q. Nykamp, SIAM J. Appl. Math. **65**, 2005 (2005).
[2] D. Q. Nykamp, Math. Biosci. **205**, 204 (2007).
[3] D. Q. Nykamp, SIAM J. Appl. Math. **68**, 354 (2007).
[4] D. R. Cox and V. Isham, *Point Processes* (Chapman and Hall, New York, 1980); D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer, New York, 1988).
[5] E. H. Adelson and J. R. Bergen, J. Opt. Soc. Am. A **2**, 284 (1985); E. P. Simoncelli and D. J. Heeger, Vision Res. **38**, 743 (1998).
[6] D. L. Ringach, G. Sapiro, and R. Shapley, Vision Res. **37**, 2455 (1997).
[7] D. H. Perkel, G. L. Gerstein, and G. P. Moore, Biophys. J. **7**, 419 (1967); A. M. H. J. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm, J. Neurophysiol. **61**, 900 (1989); G. Palm, A. M. H. J. Aertsen, and G. L. Gerstein, Biol. Cybern. **59**, 1 (1988).
[8] C. W. J. Granger, Econometrica **37**, 424 (1969).
[9] D. R. Brillinger, *Time Series: Data Analysis and Theory* (Holden Day, San Francisco, 1981).
[10] K. Sameshima and L. A. Baccalá, J. Neurosci. Methods **94**, 93 (1999); L. A. Baccalá and K. Sameshima, Biol. Cybern. **84**, 463 (2001).
[11] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[12] C. E. Shannon and W. Weaver, *The Mathematical Theory of Information* (University of Illinois Press, Urbana, IL, 1949).
[13] J. E. Kulkarni and L. Paninski, Network Comput. Neural Syst. **18**, 375 (2007).
[14] L. Paninski, Network Comput. Neural Syst. **15**, 243 (2004).