Available online at www.sciencedirect.com
ScienceDirect
Mathematical Biosciences 205 (2007) 204–251
Mathematical Biosciences
www.elsevier.com/locate/mbs

A mathematical framework for inferring connectivity in probabilistic neuronal networks

Duane Q. Nykamp *,[1]

School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church Street, Minneapolis, MN 55455, USA

Received 21 February 2006; received in revised form 10 August 2006; accepted 24 August 2006
Available online 5 September 2006

Abstract

We describe an approach for determining causal connections among nodes of a probabilistic network even when many nodes remain unobservable. The unobservable nodes introduce ambiguity into the estimate of the causal structure. However, in some experimental contexts, such as those commonly used in neuroscience, this ambiguity is present even without unobservable nodes. The analysis is presented in terms of a point process model of a neuronal network, though the approach can be generalized to other contexts. The analysis depends on the existence of a model that captures the relationship between nodal activity and a set of measurable external variables. The mathematical framework is sufficiently general to allow a large class of such models. The results are modestly robust to deviations from model assumptions, though additional validation methods are needed to assess the success of the results.

Keywords: Neural networks; Correlations; Causality; Maximum likelihood; Point process

* Tel.: +1 612 625 0338; fax: +1 612 626 2017.
E-mail address: nykamp@math.umn.edu
[1] This research was supported in part by the National Science Foundation Grant DMS-0415409.

0025-5564/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2006.08.020

1. Introduction

In many cases where one wishes to infer the connectivity among nodes in a network, only a small fraction of the nodes can be sampled simultaneously. For example, neuroscientists can measure the individual activity of only tens, or possibly hundreds, of neurons simultaneously, despite the fact that the neural networks of even small areas of the brain contain millions of neurons. Similar situations exist when studying, for example, gene networks, communication networks, or social networks. The recurring feature is the presence of large numbers of unmeasured nodes. The effects of connections involving these unmeasured nodes can confound attempts to infer the "subnetwork" of connections among the measured nodes. For example, connections between a single unmeasured node and multiple measured nodes could cause the measured nodes to appear connected even if no direct connection among the measured nodes exists. The focus of this work is a mathematical framework that can, under suitable assumptions, provide a method to control for the effects of unmeasured nodes.

We restrict our attention to networks that can be represented as directed graphs (digraphs), where the direction of a connection indicates a causal connection from one node to another. We view each node as having some measurable "activity" (we will be more specific below). A causal connection means the activity in one node affects the future activity of another node (but not the reverse, unless there also happens to exist a reciprocal connection). We can hence refine our goal to inferring the *causal subnetwork* of connections among the measured nodes (see Fig. 1). We wish to identify the causal influences among measured nodes, distinguishing such causal network connections from those connections involving unmeasured nodes where no such causal influence is present. For example, in the case where a single unmeasured node has connections onto multiple measured nodes (a "common input" configuration, illustrated with dashed thick lines in Fig. 1), there is no causal connection among the measured nodes. Hence, to reconstruct the causal subnetwork among measured nodes, we must control for the possibility of common input connections from unmeasured nodes.

If measured node 1 has a causal connection onto measured node 2 (as in Fig. 1), one would expect the connection to introduce correlations into the activity of the nodes. The activity of node 2 would be correlated with a delayed version of the activity of node 1, where the delay corresponds to the time required for the influence of node 1's activity to affect node 2. However, common input connections from an unmeasured node could induce similar correlation between the measured nodes' activity. Consider the example from Fig. 1 where an unmeasured node has a connection onto measured nodes 3 and 4. If the connection onto node 4 has a shorter delay than the connection onto node 3, then the activity of node 3 will be correlated with a delayed version of the activity of node 4. From observing this correlation in the activity of nodes 3 and 4, one might naively and incorrectly conclude that node 3 received a causal connection from node 4.

This paper describes a method that can distinguish the causal subnetwork from common input connections, subject to a certain form of ambiguity in the identity of nodes that we call "subpopulation" ambiguity. We argue that, at least in a large number of neuroscience experiments, this subpopulation ambiguity is already present in how nodes are identified. Hence, even without the presence of unmeasured nodes, a determination of connectivity will be subject to this subpopulation ambiguity. When our analysis can be applied, the presence of unmeasured nodes adds little additional ambiguity to the determination of connectivity.

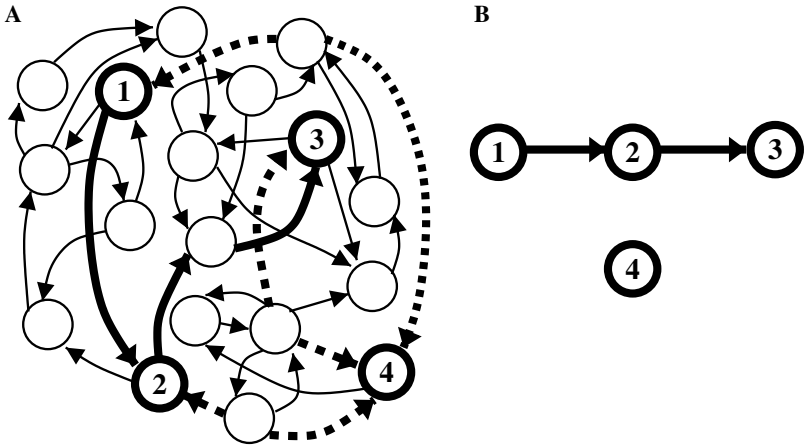

Fig. 1. Extracting the causal subnetwork among measured nodes from a network containing unmeasured nodes. (A) A sample network. Four measured nodes are illustrated by numbered circles with thick borders; unnumbered circles correspond to unmeasured nodes. Arrows indicate the direct causal connections of the network. Our goal is to identify the causal subnetwork among the measured nodes, illustrated with solid thick lines and depicted separately in panel (B). (Note that if one counts arbitrarily long chains in the network, every node might have a causal connection to almost every other node. As discussed below, since we assume the connection strength is weak, we include the effect of chains of at most two edges.) In this example, there are causal connections of two or fewer edges from 1 to 2 and from 2 to 3. We view longer arrows as indicating a longer delay. Hence, the common input connections from unmeasured nodes (illustrated with dashed thick lines) could lead to the appearance as though there were causal connections from 1 to 4, from 2 to 4 and from 4 to 3. We must control for such effects from common input to determine the causal subnetwork among measured neurons. (B) The causal subnetwork among the measured neurons extracted from panel (A). If we could successfully control for all common input, we would determine the absence of causal connections between neuron 4 and the other measured neurons.

This research is motivated by neuroscience applications, and this initial formulation is designed to be immediately applicable to neuroscience experiments. For clarity and to emphasize that some elements of the analysis are specialized to neuroscience, we will primarily use the language of neuroscience, referring to the network nodes as neurons. Nonetheless, we believe the basic approach can be generalized to have wider application.

In Section 2, we detail the subpopulation ambiguity that is present in our results. In Section 3, we describe the model framework and the assumptions of the analysis. We present the analysis of the model in Section 4 and derive our estimates of causal network structure. We demonstrate the results applied to simulations in Section 5 and discuss the results in Section 6.

2. The subpopulation ambiguity

The definition of a subpopulation of neurons is based on the relationship between neural activity and any measured external variables. (A group of neurons whose activity has a similar relationship to the external variables will be considered part of the same subpopulation.) We first describe the external variables before discussing the subpopulation ambiguity.

2.1. Measurable external variables

Denote the vector of measurable external variables by **X**. These external variables will depend on the experiment performed and the properties of the neurons being measured. The external variables can include any measurable quantity for which one can model the relationship to the activity of neurons.

For example, if the experiment includes an animal looking at images on a computer monitor, good representations of **X** would depend on the type of neurons being measured. For neurons involved in low level visual processing, **X** could include the sequence of pixel values shown on the monitor (Fig. 2, left). For neurons involved in higher levels of visual processing, **X** could be a more abstract vector somehow indexing the images or features of the images to which the neurons are sensitive. On the other hand, if the experiment involves animal movement, the vector **X** could include animal position (Fig. 2, middle) or limb position, depending on what factors modulate the neural response. **X** could also include food presentations (Fig. 2, right) or other rewards if neurons are sensitive to such rewards.

For the mathematical analysis of this paper, the specific form of the external variables is irrelevant. We will simply assume the relationship between neural activity and **X** can be modeled and make some basic assumptions about such a model. The main role the external variables play is in the interpretation of the results, namely understanding the subpopulation ambiguity.

For ease in exposition, we will most frequently refer to the external variable vector as being a representation of an image stimulus on a computer monitor (Fig. 2, left). Then, we can use the language that neurons are "responding to" the stimulus and be more concrete in our description of a subpopulation. However, we use this language solely to simplify the descriptions; all the results are valid for any type of external variable.

2.2. The idea of a subpopulation of neurons

The concept of a subpopulation is motivated by the type of experimental data available in many neuroscience experiments. In many experiments where one monitors the electrical activity of individual neurons in animals' brains, the precise identity of the measured neurons remains unknown (see Fig. 3). Researchers typically use electrodes that do not enter individual neurons. Instead, these electrodes are simply slowly pushed into the brain until they come sufficiently close to a single neuron so that its activity can be isolated from that of other neurons. Other than the neuron's

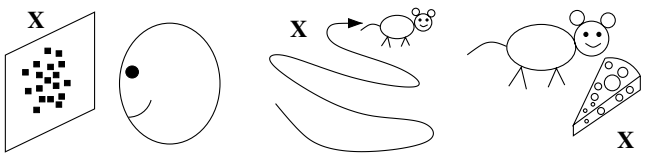

Fig. 2. Examples of measurable external variables. Left: in an experiment where an animal is looking at a computer monitor, the external variable vector **X** could be the sequence of pixel values on the monitor. Middle: in an experiment where the neural response is modulated by the position of an animal, **X** could include animal position. Right: if neurons respond to reward, **X** could include presentation of food, such as the cheese pictured.

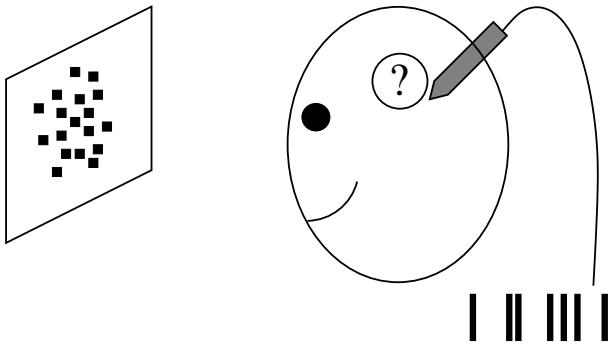

Fig. 3. Illustration of the subpopulation ambiguity already present in many neuroscience experiments. Here an animal is looking at a stimulus on a computer screen while the activity of a neuron is being recorded by an electrode (gray pentagon). The electrode was "blindly" lowered into the animal's brain and moved until the electrical activity of a single neuron (circle) can be distinguished. The electrode does not enter the neuron, but merely happens to be sufficiently close to the neuron so that its discrete output events, or spikes, can be detected. The information from the electrode is the sequence of times where that neuron spiked (illustrated by sequence of vertical black bars). As the electrode does not enter the neuron, the precise identity of the neuron is never known (illustrated by question mark). The neuron is characterized only by the relationship between its activity (spikes) and the stimulus (or other external variables). In other words, the neuron is identified only by its subpopulation, defined by those neurons whose activity is related to the stimulus in the manner observed.

general location in the brain, the only data from which to identify the neuron is the set of times in which it emitted its discrete output events, called spikes.

As a result, neurons are identified only by the relationship between their spiking activity and the external variables.[2] We define a subpopulation to capture the essence of this characterization. If we view the neurons as "responding to" the stimulus (external variables), a subpopulation can be most simply stated as a group of neurons that respond similarly to the stimulus. (More generally, it is a group of neurons whose activity has a similar relationship to the external variables.) The definition of a subpopulation can be made more precise in the context of a model.

Presumably, a subpopulation will contain more than one neuron (though mathematically that is not required). In fact, many neurons in the brain seem to respond similarly, so subpopulations are likely to be large. Hence, any conclusions about neurons that are drawn from this type of experimental data will inherently include this subpopulation ambiguity; any conclusion must be interpreted in light of the fact that only the subpopulation, not the precise identity of the neuron, can be determined.

Note that we are not introducing any assumptions about the properties of neurons when we claim conclusions contain subpopulation ambiguity. We are simply observing the fact that ambiguity is introduced when one measures only spike times, and we use the term subpopulation ambiguity to characterize this ambiguity.

[2] Moreover, different individual animals from the same species do not in general have the same set of neurons. Hence, if one wishes to generalize beyond the individual animal being studied, the precise identity of a given neuron is not even meaningful.

In particular, conclusions about connectivity based on spike time measurements will also be subject to subpopulation ambiguity. Even if we were to measure from two neurons and somehow knew that those exact two neurons were connected, our conclusion would still contain the subpopulation ambiguity. Our best conclusion would be, for example, that a neuron with the response properties of neuron 1 had a causal connection onto a neuron with the response properties of neuron 2. In other words, we could simply conclude that a neuron from the subpopulation of neuron 1 had a causal connection onto a neuron from the subpopulation of neuron 2.

We stress that the presence of subpopulation ambiguity in connectivity estimates does not indicate that one is measuring a connection from one group of neurons (subpopulation) onto another group of neurons (subpopulation). Nor does the ambiguity imply that all neurons in a given subpopulation have the same type of connections onto all neurons of another subpopulation. The statement is simply that an identification of a connection between two individual neurons is subject to subpopulation ambiguity in the identity of the neurons involved.

2.3. Determining connectivity in the presence of unmeasured neurons

We demonstrated above how, given the ambiguity in much experimental data, the best one can do is determine connectivity subject to subpopulation ambiguity, even if one did not have to control for the presence of unmeasured neurons. The essence of this paper is that we can determine connectivity subject to subpopulation ambiguity *even in the presence of unmeasured neurons*.

This determination of connectivity in the presence of unmeasured neurons is possible because we do not have to control for all types of common input connections. If the best we can do is determine connectivity subject to subpopulation ambiguity, the identification of connectivity is not sensitive to further ambiguity in the precise identity of neurons within a given subpopulation. In fact, we show below that certain forms of common input can be misconstrued as a causal connection without invalidating the determination of connectivity.

Consider the common input connection onto neuron 2 and neuron 4 in Fig. 4. Because we view longer arrows as representing longer delays, the common input from the unmeasured neuron will make the activity in neuron 4 be correlated with a delayed version of the activity in neuron 2. Hence, the relationship between the activity of neuron 2 and neuron 4 will mimic a causal connection from neuron 2 onto neuron 4. Note that in this case, the unmeasured common input neuron is from the subpopulation of neuron 2 (as indicated by gray shading). It is perfectly fine (given subpopulation ambiguity) to misconstrue this common input connection as a causal connection from neuron 2 onto neuron 4. Indeed, the correlation between the activity of neuron 2 and neuron 4 is caused (in part) by a connection from neuron 2's subpopulation onto neuron 4. Although neuron 2 itself does not have a causal connection onto neuron 4, the unmeasured neuron that does have a causal connection onto neuron 4 is indistinguishable from neuron 2 (in experiments where neurons are identified only by subpopulation). Little additional ambiguity[3] in the conclusion about connectivity is introduced by the fact that we cannot tell if neuron 2 itself is connected

[3] The only additional ambiguity is that, when connectivity must be interpreted strictly in terms of subpopulations, we cannot infer a causal connection from subpopulation 1 onto subpopulation 4 from the existence of both a connection from subpopulation 1 onto subpopulation 2 and a connection from subpopulation 2 onto subpopulation 4. Compare Fig. 4B with Fig. 4C.

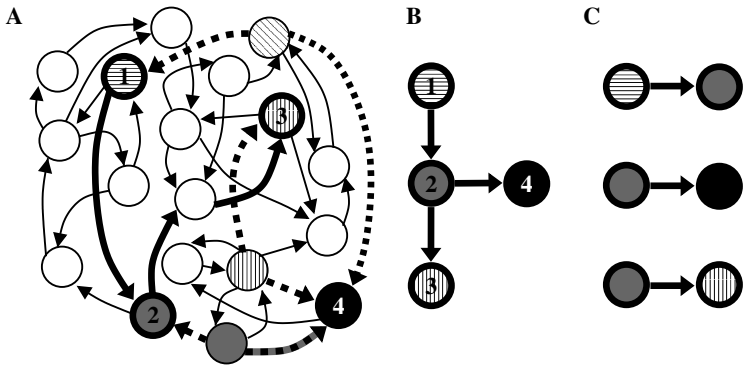

Fig. 4. Illustration of connectivity subject to subpopulation ambiguity for the network of Fig. 1. (A) The four measured neurons (numbered) and the three unmeasured neurons giving common input connections are filled to indicate membership in one of five subpopulations: black, gray, vertical hatched, horizontal hatched, or diagonal hatched. (The subpopulations of the remainder of the neurons are not indicated.) Since longer arrows indicate longer delays, the common input connections (thick dashed arrows) could be misidentified as causal connections from 1 to 4, from 2 to 4, and from 4 to 3. To determine connectivity correctly subject to subpopulation ambiguity, we do not have to control for all types of common input. Even if we misidentify the common input to neuron 2 and neuron 4 as a causal connection from 2 to 4, we will still have correctly determined the connectivity subject to subpopulation ambiguity; indeed, the common input includes a causal connection onto neuron 4 from a neuron within neuron 2's subpopulation (thick gray-black dashed arrow). (B) The resulting causal subnetwork in the case that the common input is misidentified as a connection from neuron 2 to neuron 4. Note that we must interpret each connection separately in terms of subpopulations (as in panel C) to avoid erroneous conclusions, such as the existence of a causal connection from neuron 1's subpopulation onto neuron 4's subpopulation. (C) The correct interpretation of the connectivity subject to subpopulation ambiguity identified in panel (B). We have identified a causal connection from a horizontal hatched neuron onto a gray neuron, from a gray neuron onto a black neuron and from a gray neuron onto a vertical hatched neuron.

to neuron 4 or if another neuron in neuron 2's subpopulation is connected to neuron 4. We can still infer the fact that there is an individual neuron within neuron 2's subpopulation that is connected to an individual neuron within neuron 4's subpopulation.

In contrast, the other two common input connections must be distinguished from causal connections. These common input connections induce correlations in the neurons' activity that mimic a causal connection from neuron 1 onto neuron 4 and from neuron 4 onto neuron 3. The correlations are not caused by any causal connection from neuron's 1 subpopulation onto neuron 4's subpopulation or any causal connection from neuron 4's subpopulation onto neuron 3, respectively.[4] To obtain an accurate estimate of the connectivity subject to subpopulation ambiguity, we must detect that the correlations between the activity of neuron 1 and neuron 4 and between the activity of neuron 3 and neuron 4 are due to common input connections from unmeasured neurons.

[4] The fact that, in the second case, the common input neuron is from neuron 3's subpopulation does not allow us to misidentify this connection. Although there is a causal connection from neuron 3's subpopulation onto neuron 4, the delays are set up so that the common input mimics a causal connection from neuron 4 onto neuron 3.

We turn now to the development of the analysis that can distinguish these latter common input configurations from causal connections.

3. Description of the model

3.1. The model network

Our results are based on a fairly generic class of probabilistic causal network models in discrete time. Since the measured activity of a neuron is the sequence of spike times, we model the activity as a point process. Rather than introduce standard point process notation (see, e.g., Refs. [1,2]), we jump immediately to the formulation in discrete time, which is all we need for our analysis.

Initially, imagine that we have discretized time sufficiently finely so that a neuron can have at most one spike in each time bin. Then, the response of neuron s in time bin i is simply a binary random variable R_s^i, where $R_s^i = 1$ if neuron s spiked in time bin i and is zero otherwise.[5] Let $\mathbf{R}$ be the vector of the responses of all neurons for all time bins. To specify our network model, we simply need to specify $\Pr(R_s^i = 1 | \mathbf{R}^{<i} = \mathbf{r}^{<i}, \mathbf{X} = \mathbf{x})$, i.e., the probability of a spike in neuron s at time i conditioned on the external variables $\mathbf{X}$ and the responses of all neurons for all time bins before i, which we denote by $\mathbf{R}^{<i}$.

Although we are primarily interested in fine temporal discretizations where at most one spike can occur per time bin, we will for simplicity in the form of the equations relax the assumption that R_s^i is a binary random variable and allow multiple spikes of a neuron in a single time bin. If we assume that neuronal responses in time bin i depend on the responses of neurons for time bins only before i, R_s^i becomes a Poisson random variable.[6] For small temporal discretizations, the probability that $R_s^i > 1$ becomes vanishingly small, and the difference between the binary and Poisson models is negligible.

We stress that we are not assuming that the responses of neurons are Poisson processes. The response of a neuron still depends on its spike history. For sufficiently small time bins, it approaches a generic history-dependent point process [1,2]. Hence, as long as one uses small time bins, one can regard the use of a Poisson distribution simply as a mathematical convenience.

We write our model equation as follows. Denote by $\Gamma(n, \lambda)$ the Poisson distribution with mean λ:

$$\Gamma(n, \lambda) = \frac{1}{n!} \lambda^n \mathrm{e}^{-\lambda}. \tag{3.1}$$

Then, for $r_s^i = 0, 1, 2, \ldots$, we will model the probability that R_s^i is r_s^i by $\Gamma(r_s^i, \lambda)$, where the mean number of spikes λ is a function of neurons' previous spikes and the external variables. If

[5] Throughout this paper, we will use the subscripts p, q, and s to denote neuronal indices and the superscripts i and j to denote temporal indices.

[6] This means, in particular, that multiple spikes of a neuron in a single time bin are independent, conditioned on the previous spikes of all neurons and the external variables. Because this is not realistic (a neuron's spike does strongly influence the probability that the neuron will immediately spike again), the model is a better approximation of biology with small temporal discretizations, where the probability of multiple spikes in a time bin is vanishingly small.

$\lambda \ll 1$, this distribution is approximately a binary distribution, as $\Gamma(0,\lambda) \approx 1-\lambda$ and $\Gamma(1,\lambda) \approx \lambda$ (and $\Gamma(n,\lambda) \approx 0$ for $n > 1$).

We assume that λ, the mean of R_s^i, depends on previous spikes in the following way. To represent a direct causal connection from neuron $\tilde{s}$ onto neuron s, let the factor $\bar{W}_{\tilde{s}s}^j$ denote the direct influence of the activity of neuron $\tilde{s}$ on the activity of neuron s after a delay of j times steps. If $\bar{W}_{\tilde{s}s}^j = 0$ for all j, then there is no direct causal connection from $\tilde{s}$ to s. If $\bar{W}_{\tilde{s}s}^j \neq 0$ for some $j > 0$, then there is a direct causal connection from $\tilde{s}$ to s (and we would draw an arrow from $\tilde{s}$ to s in the graph of Fig. 1). Note that $\bar{W}$ could be positive or negative, and so the model can represent both excitatory and inhibitory neurons. As we will assume linear coupling, $\bar{W}_{\tilde{s}s}^j$ is simply the temporal kernel of the connection from $\tilde{s}$ to s. We model the total "effect" of the previous spikes of all neurons on R_s^i by $\sum_{\tilde{s}}\sum_{j>0}\bar{W}_{\tilde{s}s}^j R_{\tilde{s}}^{i-j}$, where $\tilde{s}$ ranges over the index of all neurons. (We do not specify the total number of neurons in our notation; we assume it is some unknown finite integer.)

We write the mean of R_s^i simply as a function of the above connectivity term and the external variables

$$\lambda = f\left(\sum_{\tilde{s}}\sum_{j>0}\bar{W}_{\tilde{s}s}^j R_{\tilde{s}}^{i-j}, \mathbf{X}; \bar{\theta}_s^i\right).$$

Within restrictions discussed below, one can freely choose the function $f(w,\mathbf{x};\theta)$ to model key aspects of how the neural response depends on the external variables (and on the input due to network connectivity). The vector $\bar{\theta}_s^i$ denotes the parameters of the function f.

Putting this all together, the model for the probability distribution of the response of neuron s at time i is

$$\Pr(R_s^i = r_s^i | \mathbf{R}^{<i} = \mathbf{r}^{<i}, \mathbf{X} = \mathbf{x}) = \Gamma\left(r_s^i, f\left(\sum_{\tilde{s}}\sum_{j>0}\bar{W}_{\tilde{s}s}^j r_{\tilde{s}}^{i-j}, \mathbf{x}; \bar{\theta}_s^i\right)\right). \tag{3.2}$$

Using Bayes' rules and the fact that we assume the responses of all neurons at a given time step are independent, conditioned on the spiking history,[7] the probability distribution for any spike sequence $\mathbf{R}$ of the network, conditioned on the external variables, is

$$\begin{aligned}\Pr(\mathbf{R} = \mathbf{r} | \mathbf{X} = \mathbf{x}) &= \prod_i \prod_s \Pr(R_s^i = r_s^i | \mathbf{R}^{<i} = \mathbf{r}^{<i}, \mathbf{X} = \mathbf{x}) \\ &= \prod_i \prod_s \Gamma\left(r_s^i, f\left(\sum_{\tilde{s}}\sum_{j>0}\bar{W}_{\tilde{s}s}^j r_{\tilde{s}}^{i-j}, \mathbf{x}; \bar{\theta}_s^i\right)\right),\end{aligned} \tag{3.3}$$

where both s and $\tilde{s}$ range over all neuron indices and i ranges over all time points of an experiment.[8]

[7] Again, the model captures the biology better with small time bins, as the spikes of all neurons are modeled to be conditionally independent in a single time bin.

[8] For simplicity in notation, we do not explicitly make a special case of the beginning of the experiment. In practice, one identifies a memory period over which a neuron's response could depend on other neuron's spikes and analyzes the spikes obtained after that memory period, conditioned on spikes during the memory period.

A few additional comments are in order. First, the vector **R** includes the spikes of those neurons that remain unmeasured; hence, much of **R** will remain unobservable. (Addressing the fact that most of **R** is unknown is the entire focus of this paper.)

Second, the fact that we are modeling spiking probability conditioned on the external variables **X** does not necessarily mean the neurons are responding to **X**, as would be the case if **X** were a stimulus. **X** could represent a motor output that is a consequence of neural activity. Eq. (3.2) could represent how, once we observe the motor behavior, we would model the probability distribution of the neural activity that previously took place.

Third, the structure of the problem would be unchanged if we replaced the probability distribution $\Gamma(n,\lambda)$ with another (continuous or discrete) probability distribution. In this case, the model would not represent a point process of neural spikes but could represent the activity of nodes in another type of network. At least in principle, one should be able to perform the subsequent analysis with other probability distributions, allowing the results to be applied to other networks.

Fourth, note that all model parameters in Eq. (3.3) (meaning $\bar{W}$ and $\bar{\theta}$) have bars over them. In the subsequent model analysis, we assume that the actual probability distribution of the **R** is determined by these model parameters through Eq. (3.3). Since many neurons are unmeasured, we will be unable to recover these parameters exactly. We will reserve unbarred model parameters to denote the effective parameters that we will be able to calculate from the response of measured neurons. In fact, the ability to calculate an effective θ will be the key assumption restricting the choice of model function f, as discussed in the following section.

3.2. Restrictions of model choice

Our analysis approach is to completely separate the single neuron analysis from the network analysis (i.e., we employ a modular approach to attacking the problem of network connectivity). In the framework of model (3.2) (which is the essence of Eq. (3.3)), all the single neuron properties are embedded in the model function f. Hence, choosing an appropriate f is primarily a problem in modeling a single neuron (the only network aspect of f itself is the single scalar argument for the coupling term). For the network analysis that is the focus of this paper, the details of f will be irrelevant. We assume an appropriate f has been chosen that has the required properties outlined as follows.

The important restriction on the choice of model function f is that one must have an algorithm for determining the "effective parameters" of f from measured spikes of a single neuron when all coupling is ignored. In particular, we assume the existence of an algorithm that can determine the effective parameters, denoted by an unbarred θ, by fitting the equation

$$\Pr(R_s^i = r_s^i | \mathbf{X} = \mathbf{x}) = \Gamma(r_s^i, f(0, \mathbf{x}; \theta_s^i)). \tag{3.4}$$

Of course, for the results of the analysis to be valid, the function f and the algorithm for effective parameters θ must be chosen so that Eq. (3.4) captures critical elements of how the neural response depends on the external variables **X**. Devising a good form of f and concomitant fitting algorithm is a difficult problem and outside the scope of this paper. As we demonstrate with simulation results, the approach can be robust to some deviations from the model f but can also fail due to such deviations. One advantage of the modular approach is that once one develops a model and fitting algorithm, one can immediately use that model with this network approach.

We emphasize that this fitting procedure is a single neuron task, as this step uses just the spikes of neuron s, ignoring the spikes of all other neurons. By setting the first argument of f to zero before determining θ_s^i, we effectively pretend that neuron s is uncoupled from all other neurons. Since, by assumption, the spikes R_s^i were really generated by Eq. (3.2), the response of neuron s is actually determined by the responses of other neurons and the connectivity. The effective θ_s^i will consequently depend not only on the original model parameters $\bar{\theta}_s^i$, but also on the connectivity parameters $\bar{W}$ and the parameters of other neurons. We refer to Eq. (3.4) as the averaged model, as we are essentially averaging over all the responses of all other neurons, as discussed below.

The notation of the averaged model (3.4) allows an arbitrary dependence of the effective parameter θ_s^i on time i, leading to the appearance that the θ_s^i could be independent of each other. Presumably, to obtain good estimates of the θ_s^i from a single observation of each R_s^i, one must impose strong dependence among the θ_s^i for nearby i. As this paper focuses on network issues, we simply assume that any such single neuron model requirements have been satisfied. In our examples, we limit ourselves to models where θ_s^i for all i are functions of a single θ_s (see Section 5.2). In principle, one could use models that include adaptation and other factors that allow the model parameters to significantly change over time.

For the rest of the paper, we will assume that, for the measured neurons, the θ_s^i are known parameters. Note that this assumption means that we can calculate $f(w, \mathbf{x}; \theta_s^i)$ for any value of w and any time point i. This assumption restricts how one might add coupling to a model. The dependence on the coupling argument w must be chosen so that, from only the spikes of neuron s itself, one must be able to estimate any parameters required to calculate the effect of nonzero w on f. We also assume that $f(w, \mathbf{x}; \theta_s^i)$ is twice continuously differentiable in w. We let $'$ denote differentiation with respect to the first argument of f, so that the first and second derivatives of f with respect to w are $f'(w, \mathbf{x}; \theta_s^i)$ and $f''(w, \mathbf{x}; \theta_s^i)$. The above assumptions imply that we can calculate these derivatives after we determine θ_s^i and that the derivatives of $f(w, \mathbf{x}; \theta)$ with respect to w are equivalent to derivatives with respect to some function of $\mathbf{x}$ and θ.

3.3. The weak coupling assumption

Model (3.3) includes the effects of arbitrary combinations of indefinitely long chains of network connections. To reduce the lengths of chains we need to consider, we make a weak coupling assumption by treating the coupling $\bar{W}$ as a small parameter. In our analysis, we simply make a second-order approximation in $\bar{W}$, expanding Eq. (3.3) in linear and quadratic powers of the $\bar{W}_{\tilde{s}s}^j$. (Recall, we assume f is C^2 as a function of the argument where the $\bar{W}_{\tilde{s}s}^j$ appear, so the error in this approximation is $O(\bar{W}^3)$.)

However, such an argument does not fully elucidate the consequences of our weak coupling assumption. There are important consequences in the scaling of coupling strength with network size, the identity of the neurons that appear in Eq. (3.3), and possible network topologies that are considered in the analysis.

First, the analysis is designed to be applied to large networks, such as the brain's neural networks. For this reason, we briefly examine the consequences of the weak coupling in the large N limit, where N is the (unknown) number of neurons. Assuming sufficient derivatives of f, if we expanded Eq. (3.3) in higher-orders of $\bar{W}$, we would obtain terms that are kth order in $\bar{W}$ summed over all nodes of the network k times. To neglect such terms, we need the sum of $\bar{W}$ over

the network to be a small parameter. In other words, in the large N limit, we need the average value of $\bar{W}$ to scale like $1/N$. (As this scaling law pertains only to the average $\bar{W}$, all $\bar{W}^j_{\tilde{s}s}$ do not need to uniformly scale like $1/N$.)

Second, the weak coupling assumption implies that the individual neurons of Eq. (3.3) must represent lumped models that already incorporate effects of other neurons projecting to them. When we expand in $\bar{W}$, we view the terms with $\bar{W}$ as perturbations off the uncoupled probability with mean $f(0, \mathbf{X}; \bar{\theta}^i_s)$. Since this uncoupled probability is, by assumption, the largest term, either the neuron responds only weakly to the external variables (in which case the analysis would not work well) or the uncoupled model must have a strong dependence on $\mathbf{X}$. The relationship between the external variables and the response of most neurons is established through connections via other neurons. Hence, the uncoupled model already includes effects of connections from other neurons (i.e., is a lumped model of the effects of all these neurons). The weak connections that we include in Eq. (3.3) simply modulate the neural response around the baseline response that is captured by the lumped model.

This interpretation of the uncoupled models as lumped models mitigates some of the strength of the scaling assumption. Eq. (3.3) does not need to explicitly include all neurons in a particular neural network, in which case the large N limit is less important. In fact, our results can be viewed as simply fitting small networks to the measured neuron data. However, for ease of exposition, we adhere to the interpretation of a large network throughout this paper, as this is the way we typically motivate the analysis.

The third consequence of the weak coupling assumption is that it greatly simplifies the network topology. We no longer have to consider the directed graph (digraph) of the network in its entirety. We can consider separately each subgraph that is a (not necessarily connected) digraph of one or two edges. Because the following analysis is based on the weak coupling assumption, we can simply analyze the effects of all possible digraphs of one or two edges and sum them up. In this way, we view the large digraph of the entire network as being approximated by a superposition of such smaller digraphs (see Fig. 5). Moreover, we seek to derive an expression for the activity of only the measured neurons, and the activity of the measured neurons is affected only by those digraphs that project onto measured nodes (we will confirm this intuitive claim with analysis). Hence, it turns out that we need to concern ourselves only with those digraphs in which each node has a directed path toward a measured node (such as all the digraphs highlighted in Fig. 5).

Since we effectively replace the original digraph with a superposition of these small digraphs, we can ignore the effects of long network chains. (This is our justification for including only paths of two or fewer edges in our illustration of the causal subnetwork in Fig. 1.) One important consequence is that we can focus on the case where there are only two measured nodes. No digraph of two or fewer edges involving unmeasured nodes can introduce correlations among three or more measured node. (Any connected component of a graph with two edges has at most three nodes.) Hence, all the interesting effects of unmeasured nodes occur already with just two measured nodes. We do perform the analysis for an arbitrary number of measured nodes, but all our intuitive explanations and examples will involve only two measured nodes.

In principle, one can analyze higher-order expansions. Unfortunately, the number of digraphs one needs to examine increases exponentially with the number of edges (i.e., the order of the expansion).

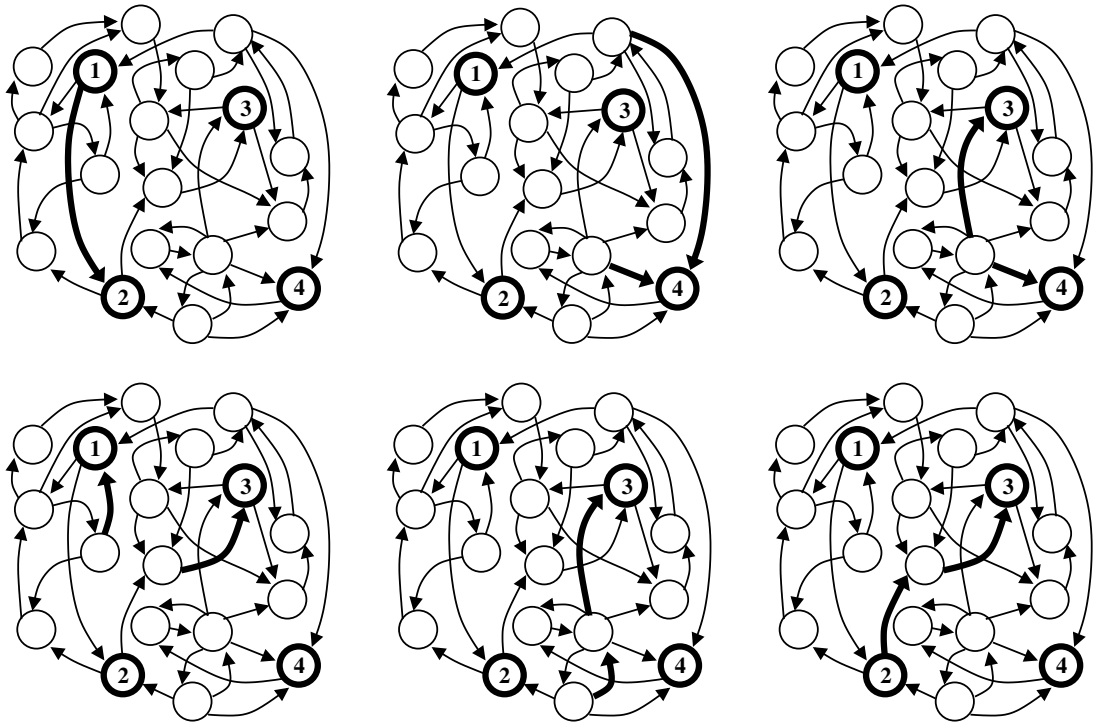

Fig. 5. An illustration of the decomposition of the directed graph (digraph) from Fig. 1 into a superposition of digraphs of two or fewer edges. This decomposition is a consequence of our weak coupling assumption. Each panel depicts a digraph (thick arrows) that corresponds to a single term in the expansion of model (3.3) in $\bar{W}$. Although not shown, the expansion of model (3.3) does include digraphs that do not project to the measured nodes (numbered circles). However, since the effects of such digraphs do not influence the measured nodes and drop from our analysis, we show examples only of those digraphs where each node has a directed connection (of one or two edges) to a measured node. To allow a direct mapping between the digraphs and the terms of the expansion in $\bar{W}$, we must include digraphs of a single edge counted twice (not shown). In our analysis, the full digraph is effectively replaced by a superposition of such smaller digraphs. Note that each causal connection or common input connection highlighted in Fig. 1 can be represented as one of these smaller digraphs.

4. Analysis of model network

We seek to develop a method to determine the connectivity among measured neurons under the assumption that the activity of all neurons was generated according to the network model (3.3). The presence of unmeasured neurons will prevent us from completely succeeding, as our connectivity estimates will be subject to the subpopulation ambiguity discussed in Section 2.

We divide the set of all neural indices into two non-overlapping sets: $\mathcal{Q}$ containing the indices of measured neurons and $\mathcal{P}$ containing indices of unmeasured neurons (clearly $\mathcal{Q} \cup \mathcal{P}$ is the set of all neural indices). Our task is to approximate the connectivity kernels $\bar{W}^j_{q_1 q_2}$ for $q_1, q_2 \in \mathcal{Q}$ from the measured spikes $\mathbf{R}_{\mathcal{Q}}$ (the subset of $\mathbf{R}$ containing the responses of measured neurons) and the external variables $\mathbf{X}$.

We proceed in three basic steps. First, we analyze the responses of an individual neuron to derive an expression for the effective parameters θ_s^i (of the averaged model (3.4)) in terms of the original model parameters (of the network model (3.3)). Second, starting with the network model (3.3), we derive an expression for the probability of any observed sequence of measured neuron spikes, writing everything in terms of the effective parameters. Third, by allowing subpopulation ambiguity (Section 2), we can isolate the effect of common input from unmeasured neurons and develop a maximum likelihood estimator of the connectivity kernels $\bar{W}^j_{q_1q_2}$.

4.1. Individual neuron analysis

When one measures from a single neuron of the brain and fits a model such as (3.4) (i.e., ignores the spikes of all other neurons), one is effectively averaging over the activity of all other neurons. In this section, we explicitly compute this average under the assumption that the neurons' spikes were actually generated by the network model (3.3). The result of this process will be an expression for the effective parameters θ defined in the averaged model (3.4) in terms of the original model parameters $\bar{\theta}$ and $\bar{W}$.

Recall from Section 3.2 that we assume the existence of an algorithm to determine the θ_s^i from the responses of neuron s and the external variables $\mathbf{X}$. Hence, for measured neuron $q \in \mathscr{Q}$, we can view $f(0,\mathbf{x};\theta_q^i)$ (and the derivatives $f'(0,\mathbf{x};\theta_q^i)$ and $f''(0,\mathbf{x};\theta_q^i)$) as known quantities. Once we have derived expressions for the original model parameters in terms of the effective parameters, we will be able to rewrite the network model (3.3) in terms of the effective parameters.

Although we do not know θ_p^i for unmeasured neuron $p \in \mathscr{P}$, our network analysis will turn out to be simpler if we write expressions involving the original model parameters $\bar{\theta}_p^i$ in terms of the (unknown) effective parameters θ_p^i. Hence, in this section we do not need to distinguish measured from unmeasured neurons and will analyze the response of neuron $s \in \mathscr{Q} \cup \mathscr{P}$.

The averaged model (3.4) is simply the network model (3.3) averaged over all time steps and all but one neural response:

$$\Pr(R_{s_1}^{i_1} = r_{s_1}^{i_1} | \mathbf{X} = \mathbf{x}) = \sum_{\substack{\mathbf{r} \\ r_{s_1}^{i_1} \text{ fixed}}} \Pr(\mathbf{R} = \mathbf{r} | \mathbf{X} = \mathbf{x}). \tag{4.1}$$

Here, we are summing over all (infinite) possibilities of spike combinations $\mathbf{R}$ of the entire network, subject only to the restriction that $R_{s_1}^{i_1}$ is fixed at the value $r_{s_1}^{i_1}$. We substitute in the structure of the model network (3.3):

$$\Pr(R_{s_1}^{i_1} = r_{s_1}^{i_1} | \mathbf{X} = \mathbf{x}) = \sum_{\substack{\mathbf{r} \\ r_{s_1}^{i_1} \text{ fixed}}} \prod_{i_2} \prod_{s_2} \Gamma\left(r_{s_2}^{i_2}, f\left(\sum_{\tilde{s}} \sum_{j>0} \bar{W}^j_{\tilde{s}s_2} r_{\tilde{s}}^{i_2-j}, \mathbf{x}; \bar{\theta}_{s_2}^{i_2} \right) \right). \tag{4.2}$$

Note that all products and sums involving forms of s are implicitly taken over all the indices in the network $\mathscr{Q} \cup \mathscr{P}$.

We define the following shorthand notation:

$$\begin{aligned}
\bar{\Gamma}^i_s &= \Gamma(r^i_s, f(0, \mathbf{x}; \bar{\theta}^i_s)),\\
\partial\bar{\Gamma}^i_s &= \frac{\partial}{\partial w}\Gamma(r^i_s, f(w, \mathbf{x}; \bar{\theta}^i_s))\bigg|_{w=0},\\
\partial^2\bar{\Gamma}^i_s &= \frac{\partial^2}{\partial w^2}\Gamma(r^i_s, f(w, \mathbf{x}; \bar{\theta}^i_s))\bigg|_{w=0}.
\end{aligned} \tag{4.3}$$

(Note the bars to indicate these expressions contain the original model parameters $\bar{\theta}^i_s$.) Also, to simplify notation, we define $\bar{W}^j_{\bar{s}s} = 0$ for $j \leqslant 0$ so that we do not have to restrict the sum over j to positive j. Because of this simplification, the causal structure of the network connections is no longer made explicit in the notation. Nonetheless, the causality of the connections will remain an important restriction.

As described in Appendix A, we take the second-order Taylor series of (4.2) in $\bar{W}$. Once we have done that, we can explicitly compute the average over all $\mathbf{r}$, as each r^i_s appears in the equation in a simple manner. This procedure requires that one enumerates all the possible digraphs that correspond to each term in the Taylor series. Most of these digraphs disappear in the averaging process. After averaging, one obtains the following expression for the activity of neuron s_1 at time i_1:

$$\begin{aligned}
\Pr(R^{i_1}_{s_1} = r^{i_1}_{s_1} | \mathbf{X} = \mathbf{x}) = \bar{\Gamma}^{i_1}_{s_1} &+ \sum_{s_2}\sum_{j_2} \bar{W}^{j_2}_{s_2 s_1} f(0, \mathbf{x}; \bar{\theta}^{i_1 - j_2}_{s_2}) \partial\bar{\Gamma}^{i_1}_{s_1}\\
&+ \frac{1}{2}\sum_{s_2}\sum_{j_2} (\bar{W}^{j_2}_{s_2 s_1})^2 f(0, \mathbf{x}; \bar{\theta}^{i_1 - j_2}_{s_2}) \partial^2\bar{\Gamma}^{i_1}_{s_1}\\
&+ \frac{1}{2}\sum_{s_2, s_3}\sum_{j_2, j_3} \bar{W}^{j_2}_{s_2 s_1} \bar{W}^{j_3}_{s_3 s_1} f(0, \mathbf{x}; \bar{\theta}^{i_1 - j_2}_{s_2}) f(0, \mathbf{x}; \bar{\theta}^{i_1 - j_3}_{s_3}) \partial^2\bar{\Gamma}^{i_1}_{s_1}\\
&+ \sum_{s_3, s_2}\sum_{j_2, j_3} \bar{W}^{j_2}_{s_2 s_3} \bar{W}^{j_3}_{s_3 s_1} f(0, \mathbf{x}; \bar{\theta}^{i_1 - j_3 - j_2}_{s_2}) f'(0, \mathbf{x}; \bar{\theta}^{i_1 - j_3}_{s_3}) \partial\bar{\Gamma}^{i_1}_{s_1}\\
&+ \mathrm{O}(\bar{W}^3).
\end{aligned} \tag{4.4}$$

The sum in the first line corresponds to a single connection onto neuron s_1. The second and third lines correspond to two connections onto neuron s_1. The last line corresponds to a path of two connections onto neuron s_1. Since there is no restriction on the identities of neurons s_3 and s_2, they could be identical to each other or neuron s_1, corresponding to loops in the network. Fig. 6 illustrates the digraphs included in Eq. (4.4).

According to the averaged model (3.4), this expression should be equal to $\Gamma(r^{i_1}_{s_1}, f(0, \mathbf{x}; \theta^{i_1}_{s_1}))$. However, the right-hand side of Eq. (4.4) is not quite a second-order approximation to a Poisson distribution.[9] Nonetheless, if we fit $\Pr(R^{i_1}_{s_1} = r^{i_1}_{s_1} | \mathbf{X} = \mathbf{x}) = \Gamma(r^{i_1}_{s_1}, f(0, \mathbf{x}; \theta^{i_1}_{s_1}))$, the right-hand side of Eq. (4.4) is our best approximation to $\Gamma(r^{i_1}_{s_1}, f(0, \mathbf{x}; \theta^{i_1}_{s_1}))$.

In parallel to the definition of the $\bar{\Gamma}^i_s$ in Eq. (4.3), we define corresponding notation for unbarred Γ^i_s:

[9] The right-hand side of Eq. (4.4) is almost a second-order approximation to a Poisson distribution. A brief calculation shows that the only deviation is the presence of the term with $\frac{\partial^2\Gamma(n,\lambda)}{\partial\lambda^2}$ that is hidden in the $(\bar{W}^{j_2}_{s_2 s_1})^2$ term.

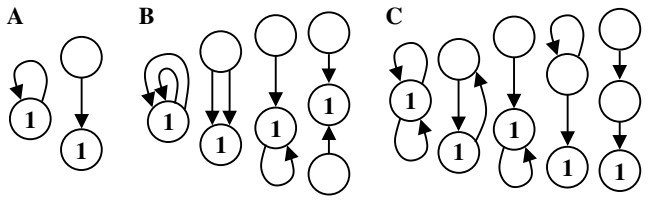

Fig. 6. Illustration of the digraphs of two or fewer edges that influence the response of a single node. This is an exhaustive list of digraphs of two or fewer edges where each node has a directed path toward node 1. In each panel, the left digraphs are special cases of the rightmost digraph with some nodes being identified with each other. These digraphs correspond to the lines in Eq. (4.4), where the neuron with index s_1 corresponds to node 1. Each line in (4.4) corresponds to multiple digraphs, as indices s_3 and s_2 could be equal to each other or to s_1. Also, a digraph could appear in multiple terms since these representations ignore the temporal component. (A) Digraphs corresponding to the sum from the first line of (4.4). The first digraph is the case $s_2 = s_1$. (B) Digraphs corresponding to the second and third line of (4.4). The second line contains only the first two digraphs, but these digraphs are repeated in the third line. (C) Digraphs corresponding to the fourth line of (4.4). Note that the first and third digraphs are equivalent to the first and third digraphs in B. The fact that these digraphs appear twice illustrates some of the subtlety in the correspondence between digraph structure and network behavior. Panel (B) represents the effect of simultaneous inputs reaching node 1. Panel (C) represents the effect of a chain of length two reaching node 1. For example, the first digraph in panel (B) corresponds to the second-order effects of inputs reaching node 1 and can be viewed as double-counting a single connection. The first digraph in panel (C) corresponds to the loop of the second digraph in panel (C) specialized to the case when the two nodes are identified with each other. The first digraph in panel (C) represents, for example, how a spike in neuron 1 can cause a second spike in neuron 1, which in turn influences the later spiking probability of neuron 1.

$$\begin{aligned}
\Gamma_s^i &= \Gamma(r_s^i, f(0, \mathbf{x}; \theta_s^i)),\\
\partial\Gamma_s^i &= \frac{\partial}{\partial w}\Gamma(r_s^i, f(w, \mathbf{x}; \theta_s^i))|_{w=0},\\
\partial^2\Gamma_s^i &= \frac{\partial^2}{\partial w^2}\Gamma(r_s^i, f(w, \mathbf{x}; \theta_s^i))|_{w=0}.
\end{aligned} \tag{4.5}$$

Then, as described in Appendix A, we can derive the following equations for expressions involving the original barred parameters written in terms of the effective parameters:

$$\begin{aligned}
\bar{\Gamma}_s^i = \Gamma_s^i &- \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0, \mathbf{x}; \theta_{s_2}^{i-j_2})\partial\Gamma_s^i\\
&- \frac{1}{2}\sum_{s_2}\sum_{j_2} (\bar{W}_{s_2 s}^{j_2})^2 f(0, \mathbf{x}; \theta_{s_2}^{i-j_2})\partial^2\Gamma_s^i\\
&+ \frac{1}{2}\sum_{s_2,s_3}\sum_{j_2,j_3} \bar{W}_{s_2 s}^{j_2}\bar{W}_{s_3 s}^{j_3} f(0, \mathbf{x}; \theta_{s_2}^{i-j_2}) f(0, \mathbf{x}; \theta_{s_3}^{i-j_3})\partial^2\Gamma_s^i + \mathrm{O}(\bar{W}^3),
\end{aligned} \tag{4.6a}$$

$$\partial\bar{\Gamma}_s^i = \partial\Gamma_s^i - \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0, \mathbf{x}; \theta_{s_2}^{i-j_2})\partial^2\Gamma_s^i + \mathrm{O}(\bar{W}^2), \tag{4.6b}$$

$$\partial^2\bar{\Gamma}_s^i = \partial^2\Gamma_s^i + \mathrm{O}(\bar{W}), \tag{4.6c}$$

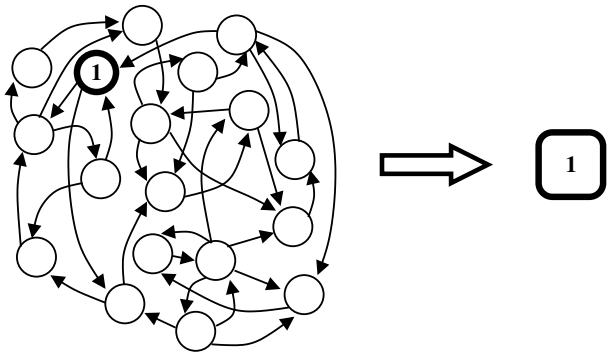

Fig. 7. Illustration of the averaged model of a single neuron. The averaged model (square on the right labeled by 1) includes the effect of the entire network pictured at the left on the response of a single neuron (circle labeled by 1). The averaged model is an average over the activity of the entire network, although, due to the weak coupling assumption, the effect of most neurons averages away.

$$f(0,\mathbf{x};\bar{\theta}^i_s) = f(0,\mathbf{x};\theta^i_s) - \sum_{s_2}\sum_{j_2} \bar{W}^{j_2}_{s_2 s} f(0,\mathbf{x};\theta^{i-j_2}_{s_2}) f'(0,\mathbf{x};\theta^i_s) + \mathrm{O}(\bar{W}^2) \tag{4.6d}$$

$$f'(0,\mathbf{x};\bar{\theta}^i_s) = f'(0,\mathbf{x};\theta^i_s) - \sum_{s_2}\sum_{j_2} \bar{W}^{j_2}_{s_2 s} f(0,\mathbf{x};\theta^{i-j_2}_{s_2}) f''(0,\mathbf{x};\theta^i_s) + \mathrm{O}(\bar{W}^2) \tag{4.6e}$$

$$f''(0,\mathbf{x};\bar{\theta}^i_s) = f''(0,\mathbf{x};\theta^i_s) + \mathrm{O}(\bar{W}) \tag{4.6f}$$

We refer to Eqs. (4.6) as the effective parameter equations (4.6). We will use them repeatedly to write expressions in terms of effective parameters.

To convert to effective parameters, we averaged over the activity of the entire network (subject to the weak coupling assumption) in order to represent the response of a single neuron using the averaged model (3.4). Graphically, this corresponds to reducing the effect of the entire network into a single averaged model, as illustrated in Fig. 7. The averaged model captures how the response of the neuron is related to the external variables and will be used to define subpopulations of neurons below.

As illustrated by Fig. 7, using effective parameters greatly simplifies the representation of the neural response. In the network analysis of the next section, connections onto just a single neuron will be absorbed into the effective parameters of that neuron, and we will be left with only those connections that introduce correlations among measured neurons. Such simplifications are one reason we convert everything to effective parameters, even for unmeasured neurons where the effective parameters are not known.[10]

4.2. Network analysis

In this second step of analysis, we derive an expression for the probability of any observed sequence of measured neuron spikes. We obtain this expression by averaging the network model

[10] This simplification due to effective parameters was already evident in the first effective parameter equation (4.6a) itself, where there is no direct representation of the digraphs containing a path with two edges leading to neuron s (Fig. 6C). The effect of such paths is included in the effective parameters of neuron s_2 in the first line of (4.6a).

(3.3) over the spikes of all unmeasured neurons, invoking the weak coupling assumption, and using the effective parameter equation (4.6) to write everything in terms of effective parameters.

We denote by $\mathbf{R}_{\mathscr{Q}}$ the responses of the measured neurons and by $\mathbf{R}_{\mathscr{P}}$ the responses of the unmeasured neurons. To find the probability of observing a sequence $\mathbf{r}_{\mathscr{Q}}$ of measured neuron spikes, we average the network model Eq. (3.3) over all possible values of $\mathbf{R}_{\mathscr{P}}$, denoted by a sum over $\mathbf{r}_{\mathscr{P}}$:

$$\begin{aligned}\Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x}) &= \sum_{\mathbf{r}_{\mathscr{P}}} \Pr(\mathbf{R} = \mathbf{r} | \mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{r}_{\mathscr{P}}} \prod_i \prod_s \Gamma\left(r_s^i, f\left(\sum_{\tilde{s}} \sum_{j>0} \bar{W}_{\tilde{s}s}^j r_{\tilde{s}}^{i-j}, \mathbf{x}; \bar{\theta}_s^i\right)\right).\end{aligned} \tag{4.7}$$

We proceed in a manner similar to the single neuron analysis. As described in Appendix B, we expand Eq. (4.7) in a Taylor series to second-order in $\bar{W}$. Then, we enumerate all the digraphs that correspond to each term in the Taylor series, separating out all possible combinations of effects from unmeasured neurons. Once we average over all possible unmeasured neuron activity and use the effective parameter equation (4.6) to write everything in terms of the effective parameters, the resulting equation simplifies to

$$\begin{aligned}\Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x}) &= \Gamma + \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{i_1,j_1} \bar{W}_{q_2 q_1}^{j_1} \left[r_{q_2}^{i_1 - j_1} - f(0, \mathbf{x}; \theta_{q_2}^{i_1 - j_1}) \right] \partial_{W_{q_1}^{i_1}} \Gamma \\ &- \frac{1}{2} \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{i_1,j_1} (\bar{W}_{q_2 q_1}^{j_1})^2 f(0, \mathbf{x}; \theta_{q_2}^{i_1 - j_1}) \partial_{W_{q_1}^{i_1}} \partial_{W_{q_1}^{i_1}} \Gamma \\ &+ \frac{1}{2} \sum_{q_1,q_2,q_3,q_4 \in \mathscr{Q}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}_{q_2 q_1}^{j_1} \bar{W}_{q_4 q_3}^{j_2} \left[r_{q_2}^{i_1 - j_1} - f(0, \mathbf{x}; \theta_{q_2}^{i_1 - j_1}) \right] \left[r_{q_4}^{i_1 - j_2} - f(0, \mathbf{x}; \theta_{q_4}^{i_1 - j_2}) \right] \partial_{W_{q_1}^{i_1}} \partial_{W_{q_3}^{i_2}} \Gamma \\ &+ \frac{1}{2} \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{p_1 \in \mathscr{P}} \sum_{\substack{i_1,j_1,j_2 \\ (q_2,j_2) \neq (q_1,j_1)}} \bar{W}_{p_1 q_1}^{j_1} \bar{W}_{p_1 q_2}^{j_2} f(0, \mathbf{x}; \theta_{p_1}^{i_1}) \partial_{W_{q_1}^{i_1 + j_1}} \partial_{W_{q_2}^{i_1 + j_2}} \Gamma \\ &+ \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{p_1 \in \mathscr{P}} \sum_{i_1,j_1,j_2} \bar{W}_{q_2 p_1}^{j_2} \bar{W}_{p_1 q_1}^{j_1} \left[r_{q_2}^{i_1 - j_1 - j_2} - f(0, \mathbf{x}; \theta_{q_2}^{i_1 - j_1 - j_2}) \right] f'(0, \mathbf{x}; \theta_{p_1}^{i_1 - j_1}) \partial_{W_{q_1}^{i_1}} \Gamma \\ &+ \mathrm{O}(\bar{W}^3),\end{aligned} \tag{4.8}$$

where we have defined the following notation

$$\begin{aligned}\Gamma &= \prod_{q \in \mathscr{Q}} \prod_i \Gamma(r_q^i, f(0, \mathbf{x}; \theta_q^i)), \\ \partial_{W_{q_1}^{i_1}} \Gamma &= \frac{\partial}{\partial W_{q_1}^{i_1}} \left[\prod_{q_2 \in \mathscr{Q}} \prod_{i_2} \Gamma(r_{q_2}^{i_2}, f(W_{q_2}^{i_2}, \mathbf{x}; \theta_{q_2}^{i_2})) \right]_{\mathbf{W}=\mathbf{O}}, \\ \partial_{W_{q_1}^{i_1}} \partial_{W_{q_2}^{i_2}} \Gamma &= \frac{\partial^2}{\partial W_{q_1}^{i_1} \partial W_{q_2}^{i_2}} \left[\prod_{q_3 \in \mathscr{Q}} \prod_{i_3} \Gamma(r_{q_3}^{i_3}, f(W_{q_3}^{i_3}, \mathbf{x}; \theta_{q_3}^{i_3})) \right]_{\mathbf{W}=\mathbf{O}}.\end{aligned} \tag{4.9}$$

As before, we do not explicitly denote the causality of the connections, but implicitly assume that all $\bar{W}^j_{s_1 s_2} = 0$ for $j \leqslant 0$.

Because we have written everything in terms of effective parameters, only those few digraphs that introduce correlations among measured neurons remain in Eq. (4.8) (see Fig. 8). The first three lines of (4.8) contain the effects of connections among the measured neurons themselves (Fig. 8A). The last two lines contain the two types of effects from unmeasured neurons that can be mediated through digraphs of two or fewer edges: common input from a single unmeasured neuron (Fig. 8B) and a connection between measured neurons via a single unmeasured neuron (Fig. 8C).

The transformation to effective parameters greatly simplified the network. Compared to the huge number of digraphs we had before converting to effective parameters (see Eq. (B.3) of Appendix B), we are left with just a few. As illustrated in Fig. 9, we have extracted out an effective network that contains the causal subnetwork among the measured neurons as well the common input connections that we highlighted in Fig. 1.

Despite the simplifications, it may still not be evident that we have actually made progress toward our ultimate goal of determining the causal subnetwork among the measured neurons. We still have not controlled for the effects of common input from unmeasured neurons. Only after we allow for subpopulation ambiguity we will be able to separate the effects due to common input from the effects due to the causal subnetwork.

4.3. Allowing subpopulation ambiguity

In this third step of the analysis, we will finally estimate the causal subnetwork, though it will be subject to subpopulation ambiguity discussed in Section 2. We accomplish this goal by analyzing separately two cases. First, we analyze the (unrealistic) case where all unmeasured neurons are from different subpopulations than the measured neurons. In this first case, we can determine the causal subnetwork without any subpopulation ambiguity. In the second (realistic) case, we

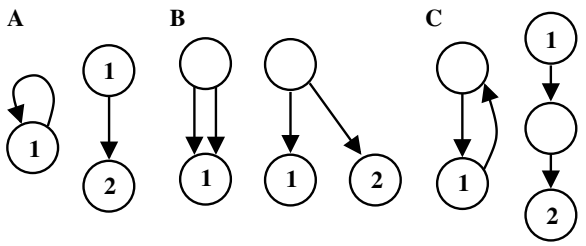

Fig. 8. Illustration of the digraphs of two or fewer edges that remain in our network analysis. Numbered circles correspond to measured nodes; unnumbered circles correspond to unmeasured nodes. (A) The only digraphs of one edge are connections between measured nodes, either a connection between a pair of distinct nodes or a connection from a node onto itself. These digraphs correspond to the sum in the first line of Eq. (4.8). We do not show the digraphs of two edges involving only measured nodes, which correspond to the second and third lines of Eq. (4.8). (B) Digraphs involving common input from a single unmeasured node onto one or two measured nodes, corresponding to the fourth line of Eq. (4.8). (C) Digraph involving a directed path of two edges between measured nodes via a single unmeasured node (from the last line of (4.8)). These digraphs are similar to those of panel (A), except for the presence of the intermediate unmeasured node.

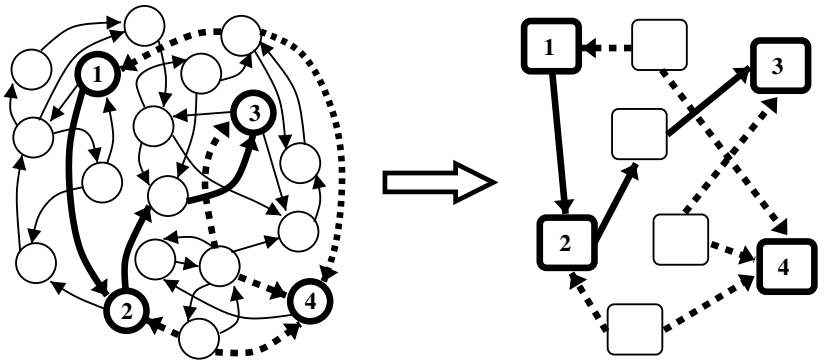

Fig. 9. Illustration of the effective network involving the measured nodes. By averaging over the activity of all unmeasured nodes and then using the averaged model (squares) represented in Fig. 7, we simplify the network into the effective network containing connections that introduce correlations among the measured nodes. The left panel is Fig. 1, and the right panel is the simplification of the network represented by Eq. (4.8). Numbered circles or squares represent measured nodes; unnumbered circles or squares represent unmeasured nodes. As in Fig. 1, thick solid lines indicate the causal subnetwork among measured nodes and thick dashed lines indicate common input from unmeasured nodes.

do not restrict the subpopulation of unmeasured neurons. In this second case, we can still determine the causal subnetwork, but only with the subpopulation ambiguity. But before looking at either case, we need to define a subpopulation in the context of our model.

4.3.1. Subpopulation definition

As described in Section 2.2, a subpopulation can be thought of as a group of neurons whose activity has a similar relationship to the external variables. The relationship between a neuron's activity and the external variables is captured by the averaged model (3.4). Roughly speaking, two neurons are in the same subpopulation if they have similar effective parameters.

We first outline an effective definition of a subpopulation. This effective definition will be a more relaxed definition than required for the analysis, as described below. We fix a delay j and look at the response of neuron s_1 ($R^i_{s_1}$) and the response of neuron s_2 from j time steps previous ($R^{i-j}_{s_2}$). According to the averaged model (3.4), the expressions $f(0,\mathbf{X};\theta^i_{s_1})$ and $f(0,\mathbf{X};\theta^{i-j}_{s_2})$ are the expected values of the random variables $R^i_{s_1}$ and $R^{i-j}_{s_2}$ conditioned on the external variables $\mathbf{X}$. The expected values themselves, though, vary as a function of time bin i. If $f(0,\mathbf{X};\theta^i_{s_1})$ and $f(0,\mathbf{X};\theta^{i-j}_{s_2})$ are highly positively correlated across time, then both neurons' responses have a similar relationship to the external variables and the neurons are from the same subpopulation (when viewed at delay j). If the correlation is small or negative, the neurons are from different subpopulations.

To write an expression for this correlation, we first define the average of f over time, conditioned on $\mathbf{X}=\mathbf{x}$, as

$$\mu_{s_1} = \langle f(0,\mathbf{x};\theta^i_{s_1})\rangle = \frac{1}{M}\sum_i f(0,\mathbf{x};\theta^i_{s_1}), \tag{4.10}$$

where the sum is over all time bins and M is the total number of time bins. Similarly, we define a variance and covariance

$$\begin{aligned}\operatorname{var}(s_1) &= \langle f(0,\mathbf{x};\theta^i_{s_1})^2\rangle - \mu^2_{s_1} = \frac{1}{M}\sum_i f(0,\mathbf{x};\theta^i_{s_1})^2 - \mu^2_{s_1},\\ \operatorname{cov}(s_1,s_2,j) &= \langle f(0,\mathbf{x};\theta^i_{s_1}) f(0,\mathbf{x};\theta^{i-j}_{s_2})\rangle - \mu_{s_1}\mu_{s_2}\\ &= \frac{1}{M}\sum_i f(0,\mathbf{x};\theta^i_{s_1}) f(0,\mathbf{x};\theta^{i-j}_{s_2}) - \mu_{s_1}\mu_{s_2}.\end{aligned} \tag{4.11}$$

We base our relaxed definition of subpopulation on the correlation coefficient

$$cc^j_{s_1s_2} = \frac{\operatorname{cov}(s_1,s_2,j)}{\sqrt{\operatorname{var}(s_1)\operatorname{var}(s_2)}}. \tag{4.12}$$

Note that $-1 \leqslant cc^j_{s_1s_2} \leqslant 1$. If $cc^j_{s_1s_2}$ is large (near 1), then neurons s_1 and s_2 are from the same subpopulation when viewed at the delay j. Otherwise, neurons s_1 and s_2 are from different subpopulations when viewed at the delay j. Note that one can view the correlation coefficient (4.12) as a measure of similarity between the effective parameters of neurons s_1 and s_2.

To work through the analysis, we need a more stringent definition of different subpopulations. If we assume neurons s_1 and s_2 are from different subpopulations when viewed at the delay j, we will assume that to $O(\bar{W})$, the averaged model $f(0,\mathbf{X};\theta^i_{s_1})$ and its derivatives are independent from $f(0,\mathbf{X};\theta^{i-j}_{s_2})$ and its derivatives. For functions (or derivatives) F and G, we will assume

$$\left\langle F\left(f\left(0,\mathbf{x};\theta^i_{s_1}\right)\right) G\left(f\left(0,\mathbf{x};\theta^{i-j}_{s_2}\right)\right)\right\rangle = \left\langle F\left(f\left(0,\mathbf{x};\theta^i_{s_1}\right)\right)\right\rangle \left\langle G\left(f\left(0,\mathbf{x};\theta^{i-j}_{s_2}\right)\right)\right\rangle \tag{4.13}$$

in terms that are second-order in $\bar{W}$.

Despite the fact that the analysis uses this stringent definition, the more relaxed definition based on correlation coefficients is easier to quantify and better reflects the effective grouping of neurons into subpopulations that we see in simulations. For example, we find that neurons with a negative correlation coefficient are still effectively from different subpopulations, despite not being independent in the sense required for the analysis. For simplicity of discussion, we use the language of subpopulations as being discrete entities, though in fact there is a gray area of intermediate values of $c^j_{s_1s_2}$ where results are mixed. See Ref. [3] for a further discussion on the precise identity of subpopulations in the context of a former approach. The general results hold for the new approach presented here.

We stress that, since subpopulations are based on the averaged model, their definition does include the average effect of connections from other neurons. A neuron's subpopulation is determined by its effective parameters. Connections from other neurons will change the relationship between a neuron's activity and the external variables, and hence will change that neuron's effective parameters. Nonetheless, even if neurons s_1 and s_2 are connected, the correlation coefficient (4.12) does not include the precise correlations in individual spikes that are due to that connection.

4.3.2. Different subpopulation case

We first look at the case when all unmeasured neurons are from different subpopulations. This will allow us to simplify the two lines of Eq. (4.8) involving unmeasured neurons.

We detail this process for the common input term (second-to-last line of (4.8)). Using the notation of the previous section, we write the sum over i_1 as M times the expected value $\langle\cdot\rangle$. Then, we invoke our subpopulation assumption by using Eq. (4.13):

$$\frac{1}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{\substack{i_1,j_1,j_2\\(q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{p_1q_1}\bar{W}^{j_2}_{p_1q_2}f(0,\mathbf{x};\theta^{i_1}_{p_1})\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma$$

$$=\frac{M}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{\substack{j_1,j_2\\(q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{p_1q_1}\bar{W}^{j_2}_{p_1q_2}\left\langle f(0,\mathbf{x};\theta^{i_1}_{p_1})\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma\right\rangle$$

$$=\frac{M}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{\substack{j_1,j_2\\(q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{p_1q_1}\bar{W}^{j_2}_{p_1q_2}\left\langle f(0,\mathbf{x};\theta^{i_1}_{p_1})\right\rangle\left\langle\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma\right\rangle$$

$$=\frac{1}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{\substack{i_1,j_1,j_2\\(q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{p_1q_1}\bar{W}^{j_2}_{p_1q_2}\langle f(0,\mathbf{x};\theta_{p_1})\rangle\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma.$$

In the last step, we removed the i_1 from $\theta^{i_1}_{p_1}$ to stress that the averaged quantity does not depend on i_1.

We perform the same manipulations on the last line of Eq. (4.8), allowing us to rewrite Eq. (4.8) as

$$\begin{aligned}\Pr(\mathbf{R}_{\mathcal{Q}}=\mathbf{r}_{\mathcal{Q}}|\mathbf{X}=\mathbf{x})&=\Gamma+\sum_{q_1,q_2\in\mathcal{Q}}\sum_{i_1,j_1}\bar{W}^{j_1}_{q_2q_1}[r^{i_1-j_1}_{q_2}-f(0,\mathbf{x};\theta^{i_1-j_1}_{q_2})]\partial_{W^{i_1}_{q_1}}\Gamma\\&-\frac{1}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{i_1,j_1}(\bar{W}^{j_1}_{q_2q_1})^2f(0,\mathbf{x};\theta^{i_1-j_1}_{q_2})\partial_{W^{i_1}_{q_1}}\partial_{W^{i_1}_{q_1}}\Gamma\\&+\frac{1}{2}\sum_{q_1,q_2,q_3,q_4\in\mathcal{Q}}\sum_{\substack{i_1,j_1\\i_2,j_2}}\bar{W}^{j_1}_{q_2q_1}\bar{W}^{j_2}_{q_4q_3}\left[r^{i_1-j_1}_{q_2}-f(0,\mathbf{x};\theta^{i_1-j_1}_{q_2})\right]\left[r^{i_1-j_2}_{q_4}-f(0,\mathbf{x};\theta^{i_1-j_2}_{q_4})\right]\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_3}}\Gamma\\&+\frac{1}{2}\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{\substack{i_1,j_1,j_2\\(q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{p_1q_1}\bar{W}^{j_2}_{p_1q_2}\langle f(0,\mathbf{x};\theta_{p_1})\rangle\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma\\&+\sum_{q_1,q_2\in\mathcal{Q}}\sum_{p_1\in\mathcal{P}}\sum_{i_1,j_1,j_2}\bar{W}^{j_2}_{q_2p_1}\bar{W}^{j_1}_{p_1q_1}[r^{i_1-j_1-j_2}_{q_2}-f(0,\mathbf{x};\theta^{i_1-j_1-j_2}_{q_2})]\langle f'(0,\mathbf{x};\theta_{p_1})\rangle\partial_{W^{i_1}_{q_1}}\Gamma+\mathrm{O}(\bar{W}^3).\end{aligned}\tag{4.14}$$

The key result is that, since we replaced the unmeasured neuron factors with their temporal averages, these factors no longer depend on i_1. This allows us to group all the effects due to common input from unmeasured neurons into a single factor. Furthermore, the indirect causal path from q_2 to q_1 via unmeasured p_1 (last line of (4.14)) is now in the same form as the direct causal connection from q_2 to q_1 (first line).

We define the effective causal connection factor W and effective common input factor U by[11]

[11] Note that this definition introduces arbitrary asymmetry into the definition of $U^0_{q_1q_2}$ since the ordering of the neuron indices q_1 and q_2 is arbitrary.

$$W^{j_1}_{q_2q_1} = \bar{W}^{j_1}_{q_2q_1} + \sum_{p_1 \in \mathscr{P}} \sum_{j_2} \bar{W}^{j_1-j_2}_{q_2p_1} \bar{W}^{j_2}_{p_1q_1} \langle f'(0, \mathbf{x}; \theta_{p_1}) \rangle,$$

$$U^{j_1}_{q_2q_1} = \begin{cases} \sum_{p_1 \in \mathscr{P}} \sum_{j_2} \bar{W}^{j_2}_{p_1q_2} \bar{W}^{j_1+j_2}_{p_1q_1} \langle f(0, \mathbf{x}; \theta_{p_1}) \rangle & \text{for } j_1 > 0, \\ \sum_{p_1 \in \mathscr{P}} \sum_{j_2} \bar{W}^{j_2}_{p_1q_2} \bar{W}^{j_2}_{p_1q_1} \langle f(0, \mathbf{x}; \theta_{p_1}) \rangle & \text{for } j_1 = 0 \text{ and } q_2 < q_1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

We substitute these expressions into Eq. (4.14). Combining the first and last sums is equivalent to replacing $\bar{W}$ in the first sum by the effective W of Eq. (4.15). The common input effects of the second-to-last line can be written in terms of U. (We wrote the sum for $j_1 \neq 0$ as twice the sum for $j_1 > 0$.) Since $W = \bar{W} + O(\bar{W}^2)$, we can simply erase the bars of the $\bar{W}$ in the remaining second-order terms. We arrive at the equation

$$\begin{aligned} \Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x}) = \Gamma &+ \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{i_1,j_1} W^{j_1}_{q_2q_1} [r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})] \partial_{W^{i_1}_{q_1}} \Gamma \\ &- \frac{1}{2} \sum_{q_1,q_2 \in \mathscr{Q}} \sum_{i_1,j_1} (W^{j_1}_{q_2q_1})^2 f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \Gamma \\ &+ \frac{1}{2} \sum_{\substack{q_1,q_2,q_3,q_4 \in \mathscr{Q} \\ i_1,j_1 \\ i_2,j_2}} W^{j_1}_{q_2q_1} W^{j_2}_{q_4q_3} \left[r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \right] \left[r^{i_1-j_2}_{q_4} - f(0, \mathbf{x}; \theta^{i_1-j_2}_{q_4}) \right] \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_3}} \Gamma \\ &+ \sum_{q_2,q_1 \in \mathscr{Q}} \sum_{i_1} \sum_{j_1>0} U^{j_1}_{q_2q_1} \partial_{W^{i_1-j_1}_{q_2}} \partial_{W^{i_1}_{q_1}} \Gamma \\ &+ \sum_{\substack{q_2,q_1 \in \mathscr{Q} \\ q_2<q_1}} \sum_{i_1} U^{0}_{q_2q_1} \partial_{W^{i_1}_{q_2}} \partial_{W^{i_1}_{q_1}} \Gamma + O(\bar{W}^3). \end{aligned} \tag{4.16}$$

Recall that $\mathbf{R}_{\mathscr{Q}}$ consists of the responses of all measured neurons, which are indexed by $q \in \mathscr{Q}$. For an experiment, our data will be the measured value $\mathbf{r}_{\mathscr{Q}}$ of the spikes $\mathbf{R}_{\mathscr{Q}}$ and the measured value $\mathbf{x}$ of the external variables $\mathbf{X}$. By the assumption of Section 3.2, we are able to determine the parameters θ^i_q from this data set by analyzing just the spikes of neuron q and $\mathbf{X}$. Hence, the only unknown parameters in Eq. (4.16) are the causal connection $W^j_{\tilde{q}q}$ and common input $U^j_{\tilde{q}q}$ factors.

Our goal is to find those values of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$ for which our model will best fit the data. In particular, given our measured values $\mathbf{r}_{\mathscr{Q}}$ and $\mathbf{x}$, we will seek values of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$ that will maximize $\Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x})$, the probability that our model predicts we should have observed the spikes $\mathbf{r}_{\mathscr{Q}}$ given that our external variables were $\mathbf{x}$. In other words, we will seek maximum likelihood estimators of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$.

Naively, one might expect that one could find the values of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$ that maximize the right-hand side of Eq. (4.16). However, the expansion in $\bar{W}$, though allowable in the presence of weak coupling, has resulted in an expression that is no longer a probability. The most obvious problem is that Eq. (4.16) is linear in U so that it is unbounded in U. Attempts to find parameters that maximize (4.16) will lead to estimates of U that diverge to plus or minus infinity. Before we can find a maximum likelihood estimator, we need to convert Eq. (4.16) back to a probability.

With one exception, Eq. (4.16) is a second-order approximation to a product of Poisson probability distributions. The exception is the same one we encountered with the single neuron anal-

ysis: the term that captures second-order effects from a single connection (the second line of (4.4) for the single neuron analysis, and the second line of (4.16) here). In the single neuron case, we could ignore this discrepancy because we were just performing algebraic manipulations. Here, however, we wish to build a maximum likelihood estimator from the form of the probability distribution $\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x})$, so we need to ensure it is a probability.

To convert the right-hand side of (4.16) to a probability, we modify our weak coupling assumption of Section 3.3 to insist that any single connection is not too large. This additional assumption does not affect the scaling argument or the assumptions on network topology because we are not changing our assumption on any quantities that appear in a sum over all neurons in the network. We simply assume that we can ignore the second-order effects of a term $(\bar{W}^j_{\tilde{q}q})^2$ that is not summed over the network. (The fact that $(\bar{W}^j_{\tilde{q}q})^2$ is summed over $q \in \mathcal{Q}$ is not problematic because the set of measured neurons is a small fixed subset of neurons.)

With this assumption, we can simply erase the second line of Eq. (4.16). To simplify the common input terms with $U^{j_1}_{q_2 q_1}$, we will use the identity derivable from the definitions (3.1) and (4.9) that for $(q_1, j_1) \neq (q_2, 0)$ and $f(0, \mathbf{x}; \theta^{i_1 - j_1}_{q_2}) > 0$,

$$\partial_{W^{i_1-j_1}_{q_2}} \partial_{W^{i_1}_{q_1}} \Gamma = \left[r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \right] \frac{f'(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})}{f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})} \partial_{W^{i_1}_{q_1}} \Gamma. \tag{4.17}$$

Since $U^0_{q_1 q_1} = 0$, we do not need an identity for $\partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \Gamma$. To address the case when $f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) = 0$, we observe that in the computation of the maximum likelihood estimators described below, we will need to evaluate $\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x})$ only when $\mathbf{r}_{\mathcal{Q}}$ is the spike sequence actually observed from the network. If $f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) = 0$, then the averaged model (3.4) states that the probability of observing $R^{i_1-j_1}_{q_2} = r^{i_1-j_1}_{q_2}$ when $r^{i_1-j_1}_{q_2} > 0$ is precisely zero. Hence, if the model is accurately capturing the data, one would expect the measured value $r^{i_1-j_1}_{q_2}$ to be zero.[12] If $r^{i_1-j_1}_{q_2} = 0$, then one can argue by continuity that $\partial_{W^{i_1-j_1}_{q_2}} \partial_{W^{i_1}_{q_1}} \Gamma = -f'(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \partial_{W^{i_1}_{q_1}} \Gamma$ even if $f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) = 0$. We implicitly assume one makes this substitution in the case $r^{i_1-j_1}_{q_2} = 0$.

With these substitutions, we write the probability of a measured spike sequence as

$$\begin{aligned}
\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x}) \approx \Gamma &+ \sum_{q_1, q_2 \in \mathcal{Q}} \sum_{i_1, j_1} W^{j_1}_{q_2 q_1} [r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})] \partial_{W^{i_1}_{q_1}} \Gamma \\
&+ \frac{1}{2} \sum_{q_1, q_2, q_3, q_4 \in \mathcal{Q}} \sum_{\substack{i_1, j_1 \\ i_2, j_2}} W^{j_1}_{q_2 q_1} W^{j_2}_{q_4 q_3} \left[r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \right] \left[r^{i_1-j_2}_{q_4} - f(0, \mathbf{x}; \theta^{i_1-j_2}_{q_4}) \right] \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_3}} \Gamma \\
&+ \sum_{q_2, q_1 \in \mathcal{Q}} \sum_{\substack{i_1, j_1 \\ j_1 > 0}} U^{j_1}_{q_2 q_1} \left[r^{i_1-j_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) \right] \frac{f'(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})}{f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})} \partial_{W^{i_1}_{q_1}} \Gamma \\
&+ \sum_{\substack{q_2, q_1 \in \mathcal{Q} \\ q_2 < q_1}} \sum_{i_1} U^0_{q_2 q_1} \left[r^{i_1}_{q_2} - f(0, \mathbf{x}; \theta^{i_1}_{q_2}) \right] \frac{f'(0, \mathbf{x}; \theta^{i_1}_{q_2})}{f(0, \mathbf{x}; \theta^{i_1}_{q_2})} \partial_{W^{i_1}_{q_1}} \Gamma,
\end{aligned} \tag{4.18}$$

[12] If a measured value $r^{i_1-j_1}_{q_2}$ is nonzero when $f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2}) = 0$, then, theoretically, the model must be a bad approximation of the network and one should use a different model. In practice, one may allow for the possibility of slight deviations from the model, and rescue the analysis by preventing $f(0, \mathbf{x}; \theta^{i_1-j_1}_{q_2})$ from dropping below some tiny number.

where the $\approx$ indicates that this is not a true second-order approximation in $\bar{W}$ since we removed the $(W^{j_1}_{q_2q_1})^2$ term. Now, we can pull all terms into Γ to write the result as a product of (dependent) Poisson distributions, yielding an equation in the same form as our original model network (3.3):

$$\begin{aligned}\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x}) \approx \prod_i \prod_{q\in\mathcal{Q}} \Gamma\Bigg(r^i_q, f\Bigg(&\sum_{\tilde{q}\in\mathcal{Q}} \sum_{j>0} W^j_{\tilde{q}q}[r^{i-j}_{\tilde{q}} - f(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})] \\ &+ \sum_{\tilde{q}\in\mathcal{Q}} \sum_{j>0} U^j_{\tilde{q}q}[r^{i-j}_{\tilde{q}} - f(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})] \frac{f'(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})}{f(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})} \\ &+ \sum_{\substack{\tilde{q}\in\mathcal{Q}\\ \tilde{q}<q}} U^0_{\tilde{q}q}[r^i_{\tilde{q}} - f(0,\mathbf{x};\theta^i_{\tilde{q}})] \frac{f'(0,\mathbf{x};\theta^i_{\tilde{q}})}{f(0,\mathbf{x};\theta^i_{\tilde{q}})}, \mathbf{x}; \theta^i_q\Bigg)\Bigg). \end{aligned} \tag{4.19}$$

Recall that U is $O(\bar{W}^2)$, so that Eq. (4.18) is a valid second-order expansion of Eq. (4.19).

We summarize Eq. (4.19) in Fig. 10. By assuming that all unmeasured neurons are from different subpopulations than the measured neurons, we were able to lump causal effects at a given delay into a single factor $W^j_{\tilde{q}q}$ and all common input effects at a given delay into a single factor $U^j_{\tilde{q}q}$.

Eq. (4.19) is indeed a probability, so we have rectified the problems with Eq. (4.16). Given the measured spikes $\mathbf{r}_{\mathcal{Q}}$ and measured external variables $\mathbf{x}$, we can first obtain the θ^i_q (according to Section 3.2), and then compute the values of the only unknown parameters, the causal connection $W^j_{\tilde{q}q}$ and common input $U^j_{\tilde{q}q}$ factors, by maximizing the probability of observing $\mathbf{r}_{\mathcal{Q}}$ as predicted by Eq. (4.19). To find such maximum likelihood estimators of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$, we need to numerically find the maximum of the log likelihood, the logarithm of Eq. (4.19) (the actual values being exceedingly small as to cause numerical underflow).

In general, one has to worry about non-global local maxima of the log likelihood, as algorithms to find maxima can get stuck in local maxima. Following Ref. [4], we can derive a

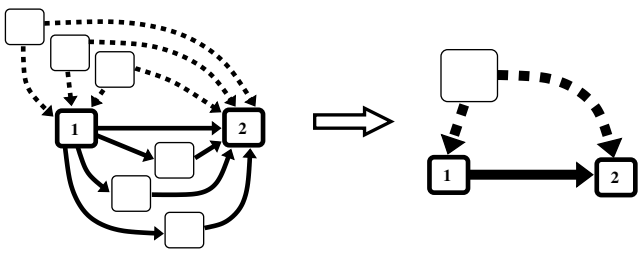

Fig. 10. Illustration summarizing the effect of assuming all unmeasured neurons are from different subpopulations than the measured neurons. For some fixed delay j, the left panel illustrates many common input connections (thick dashed lines) and causal connections (thick solid lines) that introduce correlations between neuron 2 and neuron 1 delayed by j time steps. If we assume that all the unmeasured neurons (unlabeled squares) are from different subpopulations than neuron 1 and neuron 2, the network simplifies to the effective network in the right panel. In this case, we have a single effective common input connection (thick dashed lines, corresponding to $U^j_{\tilde{q}q}$ in Eq. (4.19)) and a single effective causal connection (thick solid lines, corresponding to $W^j_{\tilde{q}q}$ in Eq. (4.19)).

condition on f under which the log likelihood will be concave, which will guarantee that no non-global local maxima exist. For fixed values of $\mathbf{r}_{\mathcal{Q}}$, $\mathbf{x}$ and θ, the log likelihood is of the form

$$\log \Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x}) \approx C + \sum_{i,q} r_q^i \log f(A_q^i \mathbf{y}, \mathbf{x}; \theta_q^i) - \sum_{i,q} f(A_q^i \mathbf{y}, \mathbf{x}; \theta_q^i), \tag{4.20}$$

where each A_q^i is a fixed matrix, C is a constant, and $\mathbf{y}$ is the vector of all the parameters $W_{\tilde{q}q}^j$ and $U_{\tilde{q}q}^j$. (Recall that $\Gamma(n, \lambda)$ is the Poisson distribution (3.1).)

Since concavity is preserved under addition and $r_q^i \geqslant 0$, the log likelihood will be concave in $\mathbf{y}$ if $f(w, \mathbf{x}; \theta)$ is convex in w and $\log f(w, \mathbf{x}, \theta)$ is concave in w. (Since the argument of f is linear in $\mathbf{y}$, the contours of f in $\mathbf{y}$ are linear, and the issue of concavity is reduced to the one-dimensional dependence of $f(w, \mathbf{x}; \theta)$ on w.) See Ref. [4] for more details on this argument in a more general setting and a description of the (quite restrictive) set of functions that are both convex and log-concave. The requirements to be both convex and log-concave include that f must be monotonically increasing in w and that it must grow at least linearly in w.

It turns out that we can get away with more general forms of f because we assume that the W and U are small. We can invoke our weak coupling assumption to make a quadratic approximation in W and U of the log likelihood.[13] This quadratic function can have no non-global local maxima. One can find the unique point where its Jacobian matrix is zero by taking a single Newton iteration starting from $W = 0$ and $U = 0$. One can take further iterations or use an alternative method to solve the fully nonlinear equation. But, if the resulting maximum differs greatly from the quadratic maximum and one is using a form of f that is not both convex and log-concave, one should investigate if the strong coupling has induced the presence of multiple local maxima.

In practice, we have achieved good results even with relatively strong coupling and a sigmoidal f by starting at $W = 0$ and $U = 0$ and iterating to a critical point using a modified version of Powell's Hybrid method as implemented in the GNU Scientific Library [5]. (As computing the Hessian matrix of the log likelihood is computationally intensive, this algorithm is faster than Newton's method.) Despite obtaining good results with strong coupling, one must keep in mind that the results with strong coupling go beyond the state of the analysis and must be used with caution.

4.3.3. Same subpopulation case

We now address the realistic case where unmeasured neurons can be from the same subpopulation as measured neurons. In this case, we can no longer exactly determine the causal subnetwork among measured neurons. Instead, we can determine the causal subnetwork only subject to the subpopulation ambiguity (Section 2).

To determine connectivity subject to the subpopulation ambiguity, we simply use the algorithm we derived above, and do not need to make any further modifications. In other words, given an

[13] This quadratic approximation will significantly differ from any of the second-order approximations presented above, not only because it is an approximation of the logarithm of the likelihood, but also because it will contain many terms not considered in our weak coupling assumption. In particular, it will contain terms that are quadratic in U as well as cross terms between W and U. The (complicated) quadratic approximation is of the form $c + J\mathbf{y} + \frac{1}{2}\mathbf{y}^t H \mathbf{y}$ where c is a constant and $\mathbf{y}$ is the vector of all Ws and Us. J and H are the Jacobian and Hessian matrices, respectively, of the log likelihood, evaluated when W and U are zero.

observed spike sequence $\mathbf{r}_{\mathscr{D}}$, we find values of $W^j_{\tilde{q}q}$ and $U^j_{\tilde{q}q}$ that maximize Eq. (4.19), even though we derived that equation subject to different assumptions.

This result is primarily a numerical observation, though we can obtain some insight by examining Eq. (4.8), the last equation before we made assumptions about subpopulations. We examine only the term corresponding to common input, looking at the effect of a common input connection from unmeasured neuron p_1 onto measured neurons q_1 and q_2, with delays j_1 and j_2, respectively,

$$\sum_{i_1} \bar{W}^{j_1}_{p_1q_1} \bar{W}^{j_2}_{p_1q_2} f(0,\mathbf{x};\theta^{i_1}_{p_1}) \partial_{W^{j_1+j_1}_{q_1}} \partial_{W^{i_1+j_2}_{q_2}} \Gamma$$

(extracted from the second-to-last line of Eq. (4.8)). We assume that $j_2 > j_1$, so this common input introduces correlations in the activity of neurons q_1 and q_2 that mimic a causal connection from q_1 onto q_2 with delay $j_3 = j_2 - j_1$. Using Eq. (4.17), we can rewrite the common input expression as

$$\sum_{i_2} \bar{W}^{j_1}_{p_1q_1} \bar{W}^{j_1+j_3}_{p_1q_2} f(0,\mathbf{x};\theta^{i_2-j_3-j_1}_{p_1}) \left[r^{i_2-j_3}_{q_1} - f(0,\mathbf{x};\theta^{i_2-j_3}_{q_1}) \right] \frac{f'(0,\mathbf{x};\theta^{i_2-j_3}_{q_1})}{f(0,\mathbf{x};\theta^{i_2-j_3}_{q_1})} \partial_{W^{i_2}_{q_2}} \Gamma, \tag{4.21}$$

where $i_2 = i_1 + j_2$.

First, we look at the case where neuron p_1 is from the same subpopulation as neuron q_1 when viewed at the delay j_1, analyzing the extreme case where $f(0,\mathbf{x};\theta^{i_2-j_3-j_1}_{p_1}) = f(0,\mathbf{x};\theta^{i_2-j_3}_{q_1}) + \mathrm{O}(\bar{W})$. Then, a second-order approximation to expression (4.21) is

$$\sum_{i_2} \bar{W}^{j_1}_{p_1q_1} \bar{W}^{j_1+j_3}_{p_1q_2} \left[r^{i_2-j_3}_{q_1} - f(0,\mathbf{x};\theta^{i_2-j_3}_{q_1}) \right] f'(0,\mathbf{x};\theta^{i_2-j_3}_{q_1}) \partial_{W^{i_2}_{q_2}} \Gamma. \tag{4.22}$$

By examining Eq. (4.18), we see only a hint that the common input of expression (4.22) might act like a causal connection. Compared to the connection of the first line of (4.18), expression (4.22) has an additional f' factor. On the other hand, the common input of the last line of (4.18) has an additional f'/f factor. Hence, the cancellation of the factor $1/f(0,\mathbf{x};\theta^{i_2-j_3}_{q_1})$ above has increased the similarity of (4.22) to the causal connection and decreased its similarity to the common input of (4.18).

We show via simulations that despite the additional f' factor, the effect of expression (4.22) is similar to that of a causal connection.[14] In particular, common input from the subpopulation of neuron q_1 will appear as a causal connection in our maximum likelihood estimators based on Eq. (4.19), as illustrated in Fig. 11A. Consequently, our results are subject to the subpopulation ambiguity of Section 2.

Second, we look at the case where neuron p_1 is from the same subpopulation as neuron q_2 when viewed at the delay $j_1 + j_3$, analyzing the extreme case where $f(0,\mathbf{x};\theta^{i_2-j_3-j_1}_{p_1}) = f(0,\mathbf{x};\theta^{i_2}_{q_2}) + \mathrm{O}(\bar{W})$. In this case, no terms cancel in the common input of (4.21), and the similarity to the causal connection of Eq. (4.18) is not increased. Simulations show that this common input from the subpopulation of neuron q_2 will still appear as a common input in our maximum likelihood estimators based

[14] Although it would seem in principle possible to distinguish the effect of (4.22) from both the causal connection and common input of Eq. (4.18), the differences among the three effects are too subtle, and attempts to make the three-way distinction have proven too sensitive to noise.

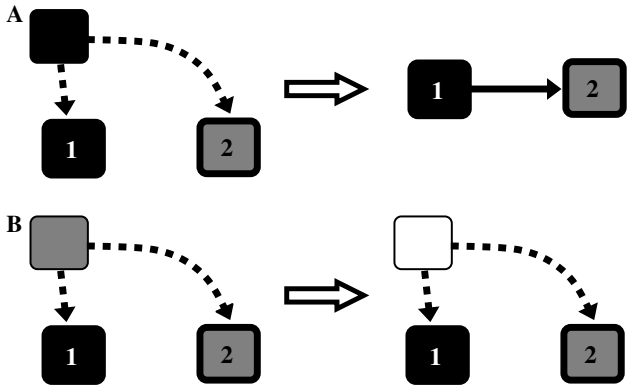

Fig. 11. The effect of common input from one of the measured neuron's subpopulations. In this illustration, the connection to measured neuron 2 has a longer delay than the connection to measured neuron 1, introducing correlations in the measured neurons' activity that is similar to a causal connection from neuron 1 onto neuron 2. Subpopulation is indicated by shading (black, gray, or white). (A) Common input from an unmeasured neuron in neuron 1's subpopulation (left panel) appears like a causal connection from neuron 1 onto neuron 2 (right panel). Since the common input configuration contains a causal connection from neuron 1's subpopulation, this misidentification still leads to a correct characterization of the connectivity subject to the subpopulation ambiguity. (B) Common input from an unmeasured neuron in neuron 2's subpopulation (left panel) is correctly identified as common input (right panel). Distinguishing this common input configuration from a casual connection is required even if we allow subpopulation ambiguity because there is no causal connection from neuron 2's subpopulation onto neuron 1's subpopulation.

on Eq. (4.19), as illustrated in Fig. 11B. This result is crucial to maintain an accurate estimate of connectivity even if we allow subpopulation ambiguity, as this common input configuration does not have a causal connection from neuron q_1's subpopulation onto neuron q_2's subpopulation.

5. Simulation results

To illustrate our approach, we simulated several small networks. We designated two neurons as measured neurons and recorded the spike times of only those two neurons. The remaining neurons were unmeasured, and we ignored their spike times in the analysis. From the spike times of the two measured neurons and the external variables, we attempted to determine the connectivity between the two measured neurons.

5.1. The simulated networks

In this section, we describe the external variables and models that we simulated to generate the data.

5.1.1. The simulated external variables

We viewed the simulated neurons as responding to a visual stimulus represented by the external variable **X**. At each time i, the visual stimulus was a gray-scale rectangular image (e.g., on a com-

puter screen) that is constant along vertical lines. We let X_j^i represent the value of the jth line of the image at time i, where X_j^i indicates the brightness of line j relative to a background gray (positive values of X_j^i indicate brighter lines and negative values indicate darker lines). Each image contained $N_0 = 100$ lines ($j = 0, 1, \ldots, N_0 - 1$). The entire visual stimulus was a movie consisting of a sequence of such images that spanned 500,000 time bins.

The movie was a sequence of sinusoidal grating images, where each image is presented for ten consecutive time bins (see Fig. 12). Let $\mathbf{I}^k$ be a sinusoidal image with wave number k defined so that the jth line I_j^k is given by $I_j^k = \mathrm{cas}(2\pi kj/N_0)$, where $\mathrm{cas}x = \cos x + \sin x$. The $\mathbf{I}^k$ are Hartley basis functions defined so that they form an orthonormal set. We generated our stimulus movie $\mathbf{X}$ by, every ten time bins, randomly selecting a new image, with replacement, from the set composed of the $\mathbf{I}^k$ and $-\mathbf{I}^k$, for $k = -14, -13, \ldots, 13, 14$.

5.1.2. The simulated models

We simulated two types of models to generate the test data for our analysis. The first model was a linear–nonlinear model, where the response probability of each neuron was a linear function of the stimulus composed with a static, memoryless nonlinear function. The probability distribution of the response R_s^i of neuron s at time bin i, conditioned on previous spikes $\mathbf{R}^{<i}$ of all neurons and the stimulus $\mathbf{X}$, was

$$\Pr(R_s^i = r_s^i | \mathbf{R}^{<i} = \mathbf{r}^{<i}, \mathbf{X} = \mathbf{x}) = \Gamma\left(r_s^i, g\left(\bar{\mathbf{h}}_s^i \cdot \mathbf{x} + \sum_{\tilde{s}} \sum_{j>0} \bar{W}_{\tilde{s}s}^j r_{\tilde{s}}^{i-j}; \bar{A}_s, \bar{y}_s, \bar{\epsilon}_s\right)\right), \tag{5.1}$$

where $\Gamma(n, \lambda)$ is the Poisson distribution (3.1), the first sum is over all neurons $\tilde{s}$ in the network, and $g_s(y; \bar{A}_s, \bar{y}_s, \bar{\epsilon}_s)$ is an error function nonlinearity with argument y and parameters $\bar{A}_s$, $\bar{y}_s$ and $\bar{\epsilon}_s$ defined by

$$g(y; \bar{A}_s, \bar{y}_s, \bar{\epsilon}_s) = \frac{\bar{A}_s}{2}\left[1 + \mathrm{erf}\left(\frac{y - \bar{y}_s}{\bar{\epsilon}_s\sqrt{2}}\right)\right]. \tag{5.2}$$

We let $\bar{\mathbf{h}}_s$ be a spatio-temporal linear kernel for neuron s. We view the kernel as sliding along the stimulus with time, and let $\bar{\mathbf{h}}_s^i$ be the kernel shifted for time point i. We write the convolution of the kernel with the stimulus as the dot product $\bar{\mathbf{h}}_s^i \cdot \mathbf{X}$ (implicitly viewing the temporal index of the stimulus as going backward in time).

In our simulations, we set the form of the kernels to capture some features of the response of neurons in visual cortex [6]. For vertical line $j = 0, 1, \ldots, N_0$ and temporal index t, we used the form

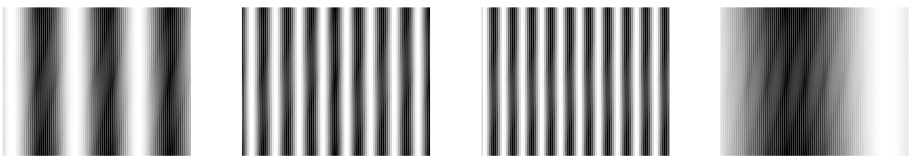

Fig. 12. Example of a sequence of the sinusoidal grating images from the movie used as the visual stimulus in the simulations. Each image was presented for ten units of time, after which a new image was randomly selected.

$$\bar{h}_s(j,t) = (t-b_s)\exp\left(-\frac{t-b_s}{\tau_s} - \frac{(j-c)^2}{2\sigma_s^2}\right)\cos(2\pi f_s(j-c)+\phi_s)$$

for $t > b_s$ and $\bar{h}_s(j,t) = 0$ otherwise. We set $c = (N_0 - 1)/2$ so that the kernels were centered on the image; we set the remaining kernel parameters, b_s, τ_s, σ_s, f_s, and ϕ_s, individually for each neuron. We formed the vector $\bar{\mathbf{h}}_s$ by sampling $\bar{h}_s(j,t)$ for $t = 1,2,\ldots,100$, and normalized the resulting vector so that the variance of $\bar{\mathbf{h}}_s^i \cdot \mathbf{X}$ would be one.

Our second model was similar to the linear–nonlinear model, except that the relationship between the neural activity and the stimulus was more fundamentally nonlinear. In this model, we convolved the stimulus separately with two linear kernels $\bar{\mathbf{h}}_{s,1}^i$ and $\bar{\mathbf{h}}_{s,2}^i$ as well as their opposites $-\bar{\mathbf{h}}_{s,1}^i$ and $-\bar{\mathbf{h}}_{s,2}^i$. We added the coupling effects to the convolution with $\bar{\mathbf{h}}_{s,1}^i$, took a nonlinear function of each convolution, and then added together the four results.

We included a "refractory period," so that after each time a neuron spiked, it could not spike for the subsequent time bin. Hence each R_s^i could be only zero or one, and R_s^i was set to zero if $R_s^{i-1} = 1$. If the neuron had not spiked in the previous two time bins, the probability of a spike was[15]

$$\Pr(R_s^i = 1 | \mathbf{R}^{<i} = \mathbf{r}^{<i}, R_s^{i-1} = 0, \mathbf{X} = \mathbf{x}) = 1 - \Gamma(0, \lambda_s^i) \tag{5.3}$$

where

$$\lambda_s^i = \bar{A}_{s,1}\left[\bar{\mathbf{h}}_{s,1}^i \cdot \mathbf{x} + \sum_{\tilde{s}}\sum_{j>0} \bar{W}_{\tilde{s}s}^j r_{\tilde{s}}^{i-j} - \bar{y}_{s,1}\right]_+^2 + \bar{A}_{s,2}\left[-\bar{\mathbf{h}}_{s,1}^i \cdot \mathbf{x} - \bar{y}_{s,2}\right]_+^2 + \bar{A}_{s,3}\left[\bar{\mathbf{h}}_{s,2}^i \cdot \mathbf{x} - \bar{y}_{s,3}\right]_+^2 + \bar{A}_{s,4}\left[-\bar{\mathbf{h}}_{s,2}^i \cdot \mathbf{x} - \bar{y}_{s,4}\right]_+^2. \tag{5.4}$$

We used the half-squaring nonlinearity ($[y]_+^2 = y^2$ if $y>0$ and is zero otherwise) with parameters $\bar{A}_{s,k}$ and $\bar{y}_{s,k}$ for $k = 1,2,3$, and 4. We chose the parameters of $\bar{\mathbf{h}}_{s,2}$ to be identical to those of $\bar{\mathbf{h}}_{s,1}$, except that we altered its spatial phase by one quarter period. (We set $\phi_{s,1} = \phi_s$ and $\phi_{s,2} = \phi_s + \pi/2$.) Hence, model (5.3) was a generalization of the energy model [7] commonly used to model the responses of visual neurons. We view each the four terms in (5.4) as "subunits" and refer to model (5.3) as a subunit model.

For all simulations, we modeled the effects of connections between neurons as alpha-functions of the form

$$\bar{W}_{s_1s_2}^j = B_{s_1s_2}\frac{j - d_{s_1s_2}}{\tau_w^2}\exp\left(-\frac{j-d_{s_1s_2}}{\tau_W}\right)$$

for $j > d_{s_1s_2}$ and $\bar{W}_{s_1s_2}^j = 0$ otherwise. The parameter $d_{s_1s_2}$ specified the delay in the connection from neuron s_1 onto neuron s_2, and $B_{s_1s_2}$ was the strength of the connection. For all connections, we set $\tau_W = 0.5$. For each simulation, we repeated the above movie 10 times. We set the parameters to obtain approximately 30,000–40,000 spikes per neuron.

[15] We used the probability $1 - \Gamma(0,\lambda)$ so that the probability of a spike was set to be the probability of one or more spikes under the Poisson probability distribution.

5.2. The models used for analysis

To perform our analysis, we need to choose a model and a method to determine its effective parameters, as discussed in Section 3.2. We use the linear–nonlinear model (5.1) and the linear–quadratic–nonlinear model [8], as we have developed an algorithm to determine their parameters from movies of orthogonal images, such as the stimulus described above. The algorithm for determining the parameters is described in Appendix C, which is a modification of that presented in Ref. [8].

To write our model Eqs. (3.3) in terms of the linear–nonlinear model (5.1), we define the model function f as

$$f(w, \mathbf{x}; \bar{\theta}_s^i) = g\left(\bar{\mathbf{h}}_s^i \cdot \mathbf{x} + w; \bar{A}_s, \bar{y}_s, \bar{\epsilon}_s\right). \tag{5.5}$$

Hence, the vector of model parameters $\bar{\theta}_s^i$ contains the linear kernel $\bar{\mathbf{h}}_s^i$ and the error function nonlinearity parameters $\bar{A}_s$, $\bar{y}_s$ and $\bar{\epsilon}_s$. For each measured neuron $q \in \mathcal{Q}$, the effective single neuron parameters θ_q^i (containing the effective linear kernel $\mathbf{h}_s^i$ and the effective error function nonlinearity parameters A_s, y_s, and ϵ_s) are determined by fitting Eq. (3.4) with $f(0, \mathbf{x}; \theta_q^i) = g(\mathbf{h}_q^i \cdot \mathbf{x}; A_q, y_q, \epsilon_q)$ from the spikes of neuron q, using the procedure described in Appendix C.

To obtain the linear–quadratic–nonlinear model, we add a quadratic kernel $\bar{\mathbf{h}}_{s,2}$ to the linear nonlinear model. The quadratic kernel is defined in a similar way as the linear kernel, except that we convolve the kernel with the vector denoted by $\mathbf{X}^2$ that has components $(X_j^i)^2$, the squares of the stimulus values. Exactly as we did for the linear kernel, we write this convolution as $\bar{\mathbf{h}}_{s,2}^i \cdot \mathbf{X}^2$, where the temporal index of the kernel indicates shifting in time, and normalize $\bar{\mathbf{h}}_{s,2}$ so that the variance of $\bar{\mathbf{h}}_{s,2}^i \cdot \mathbf{X}^2$ is one. Denoting the linear kernel now by $\bar{\mathbf{h}}_{s,1}^i$, the model function f for the linear–quadratic–nonlinear model is

$$f(w, \mathbf{x}; \bar{\theta}_s^i) = g\left(\sqrt{1-\bar{\alpha}_s}\,\bar{\mathbf{h}}_{s,1}^i \cdot \mathbf{x} + \sqrt{\bar{\alpha}_s}\,\bar{\mathbf{h}}_{s,2}^i \cdot \mathbf{x}^2 + w; \bar{A}_s, \bar{y}_s, \bar{\epsilon}_s\right), \tag{5.6}$$

where $\bar{\alpha}_s$ (restricted to $0 \leqslant \bar{\alpha}_s \leqslant 1$) indicates the relative weight of the quadratic kernel. In this case, the vector of model parameters $\bar{\theta}_s^i$ contains the linear and quadratic kernels, the parameter $\bar{\alpha}_s$, and the error function parameters $\bar{A}_s$, $\bar{y}_s$, and $\bar{\epsilon}_s$. The effective single neuron parameters θ_q^i of each measured neuron q are determined by fitting Eq. (3.4) with

$$f(0, \mathbf{x}; \theta_q^i) = g\left(\sqrt{1-\alpha_s}\mathbf{h}_{1,q}^i \cdot \mathbf{x} + \sqrt{\alpha_s}\mathbf{h}_{2,q}^i \cdot \mathbf{x}^2; A_q, y_q, \epsilon_q\right)$$

using the procedure described in Appendix C. As above, θ_q^i contains the unbarred versions of the parameters in $\bar{\theta}_q^i$.

Both the linear–nonlinear model and the linear–quadratic–nonlinear model satisfy the requirements of Section 3.2. Note that, since w is just added to the $\bar{\mathbf{h}} \cdot \mathbf{x}$ terms, we can determine $f(w, \mathbf{X}; \theta_q^i)$ and $f'(w, \mathbf{X}; \theta_q^i)$ once we have fit $f(0, \mathbf{X}; \theta_q^i)$ in the single neuron analysis. Hence, once we determine the θ_q^i from the single neuron analysis, we can compute all quantities in (4.19) except for the W and U.[16]

[16] We used cross-validation to estimate $f(w, \mathbf{X}; \theta_q^i)$ and $f'(w, \mathbf{X}; \theta_q^i)$. We divided the data into 10 segments. When computing these quantities for each time bin i from one of these segments, we calculated the parameters θ_q^i from the data in the other 9 segments.

Note that the discretization into temporal bins is artificial, at least for biological systems. For simplicity, we simulated models in discrete time, using time bins of size one. We performed all our analyses using time bins of size two.

5.3. Linear–nonlinear model simulations

For the first set of tests, we simulated networks of linear–nonlinear neurons (5.1). Each network contained two measured neurons that are indexed by $q = 1, 2$. All other neurons were unmeasured, and we ignored their spikes in all analyses.

We analyzed the results to calculate our causal connection factor W and common input factor U using the linear–nonlinear model (5.5). We also calculated a commonly used measure of the correlation between two spike trains, the shuffle-corrected correlogram or covariogram [9–11]. We can calculate this measure because we have repeated the stimulus multiple times. For each measured neuron $q = 1, 2$, we average its activity over all stimulus repeats, which we write as $\langle R_q^i | \mathbf{X} \rangle$. Then, the covariogram for measured neuron 1 and neuron 2 is

$$C^j = \langle R_1^i R_2^{i-j} \rangle - \left\langle \langle R_1^i | \mathbf{X} \rangle \langle R_2^{i-j} | \mathbf{X} \rangle \right\rangle, \tag{5.7}$$

where the remaining $\langle \cdot \rangle$ indicate an average over the entire experiment. We index the covariogram by the delay j, which corresponds to the spike time of neuron 1 minus the spike time of neuron 2. For display purposes, we plot the W and U factors using the same convention, defining

$$W^j = \begin{cases} W_{12}^{-j} & \text{for } j < 0, \\ 0 & \text{for } j = 0, \\ W_{21}^{j} & \text{for } j > 0, \end{cases}$$

$$U^j = \begin{cases} U_{12}^{-j} & \text{for } j \leqslant 0, \\ U_{21}^{j} & \text{for } j > 0. \end{cases}$$

We begin with two simple networks: a network of two measured neurons with a causal connection and a network of two uncoupled measured neurons that both receive a connection from a third, unmeasured neuron. The results are shown in Fig. 13. We chose parameters so that the covariogram (top plot) in both networks has a peak at a delay of four units of time. Note that one cannot distinguish the networks from the covariograms alone. The covariograms have virtually the same form regardless of whether the two measured neurons have a causal connection or are correlated simply due to a common input connection.

The bottom two panels of Fig. 13 demonstrate that our analysis can indeed distinguish the causal connection from the common input, at least for the case when the unmeasured neuron is from a distinct subpopulation. In Fig. 13A, the causal connection factor W is significantly positive at a delay of 4, while the common input factor U is virtually zero. In Fig. 13B, on the other hand, it is the common input factor U that is significantly positive at a delay of 4. (See below for an explanation why the causal connection factor W dips negative at a delay of 4.) Hence, we can conclude that the correlation is due to a causal connection in Fig. 13A and the correlation is due to common input in Fig. 13B.

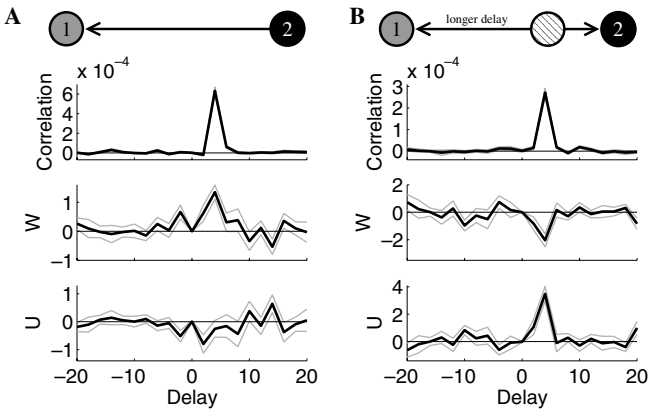

Fig. 13. Distinguishing a causal connection from common input in simulations of linear–nonlinear networks. (A) The results from analyzing the spikes of two neurons with a causal connection, as schematized at the top. Neuron 2 has a causal connection onto neuron 1 with a delay of four units of time. Shading or hatching indicates subpopulation. The correlation between the spikes of neuron 1 and neuron 2, as measured by the covariogram (5.7), is plotted in the top panel. The covariogram indicates that the spikes of neuron 1 are correlated with the spikes of neuron 2 delayed by four units of time. The causal connection factor W, but not the common input factor U, is positive at a delay of 4, indicating the presence of a causal connection from neuron 2 onto neuron 1. Delay is spike time of neuron 1 minus spike time of neuron 2. In all three plots, thin gray lines indicate a bootstrap estimate of one standard error, calculated by resampling from the set of stimulus repetitions 100× (when the standard error is small, gray lines are obscured by the black line). Simulation parameters: $\bar{A}_1 = \bar{A}_2 = 1$, $\bar{y}_1 = 5.4$, $\bar{y}_2 = 7.6$, $\bar{\epsilon}_1 = 2.0$, $\bar{\epsilon}_2 = 3.0$, $b_1 = b_2 = 0$, $\tau_1 = 40$, $\tau_2 = 50$, $\sigma_1 = 10$, $\sigma_2 = 15$, $f_1 = 0.08$, $f_2 = 0.04$, $\phi_1 = 0$, $\phi_2 = 2\pi/3$, $B_{21} = 3.0$, $d_{21} = 3$, $B_{11} = B_{12} = B_{22} = 0$. (B) The results from analyzing the spikes of two neurons that received common input connections from a third, unmeasured, neuron. Panels as in (A). As schematized at the top, the connection from the unmeasured neuron (unlabeled) onto neuron 1 had a longer delay (by four units) than the connection onto neuron 2. Hence, the covariogram reveals the correlation at a delay of 4. The plots of W and U reveal that this correlation is due to common input rather than a causal connection from neuron 2 onto neuron 1, as the common input factor U is positive at a delay of 4. Most parameters as in panel (A). Exceptions and additional parameters (the unmeasured neuron is indexed by 3): $\bar{A}_3 = 1.0$, $\bar{y}_1 = 5.7$, $\bar{y}_2 = 7.9$, $\bar{y}_3 = 10.0$, $\bar{\epsilon}_3 = 4.0$, $b_3 = 0$, $\tau_3 = 45$, $\sigma_3 = 20$, $f_3 = 0.06$, $\phi_3 = 4\pi/3$, $d_{31} = 5$, $d_{32} = 1$, $B_{31} = B_{32} = 7$, $B_{ij} = 0$ for all other i and j.

Such conclusions, however, must be interpreted with the subpopulation ambiguity described in Section 2. The next two tests demonstrate that we indeed obtain a description of connectivity subject to subpopulation ambiguity as claimed in Section 4.3.3. We simulate the same common input network of Fig. 13B, but change the parameters of the unmeasured neuron so that it is first in the subpopulation of neuron 2 and then in the subpopulation of neuron 1. The results are shown in Fig. 14. Common input from neuron 2's subpopulation (Fig. 14A), but not common input from neuron 1's subpopulation (Fig. 14B), appears like a causal connection from neuron 2 onto neuron 1. Since common input from neuron 2's subpopulation does contain a connection from neuron 2's subpopulation onto neuron 1, these results still accurately reflect the connectivity when interpreted with subpopulation ambiguity. (Compare the results of Fig. 11 to Fig. 14, where the roles of neuron 1 and neuron 2 are reversed.)

Note that in Figs. 13B and 14B, W dips negative at the delay of 4. This reciprocal behavior frequently happens when connections are too strong to strictly satisfy our weak coupling assumption (Section 3.3). Such behavior does not present ambiguity in the interpretation of the results. In

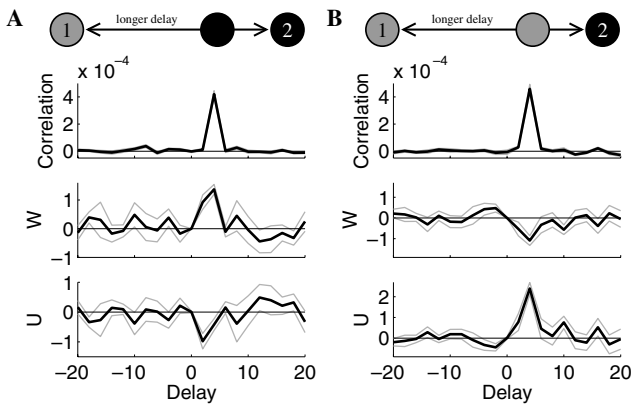

Fig. 14. The presence of subpopulation ambiguity in the identification of causal connections (cf. Fig. 11). Panels as in Fig. 13. In both examples, common input from an unmeasured neuron has a longer delay to neuron 1, creating correlations in the spikes of neuron 1 and neuron 2 that mimic a causal connection from neuron 2 onto neuron 1. (A) When the unmeasured common input neuron is from neuron 2's subpopulation (cc^j_{32} from Eq. (4.12) exceeded 0.9), the causal connection factor W is positive, incorrectly indicating a causal connection from neuron 2 onto neuron 1. Nonetheless, the determination of connectivity is correct if one allows subpopulation ambiguity, as the network does contain a connection from neuron 2's subpopulation onto neuron 1. Parameters as in Fig. 13B, except that $b_2 = 2.0$, $\tau_3 = 50$, $\sigma_3 = 15$, $f_3 = 0.035$, $\phi_3 = 2\pi/3$. (B) Even when the unmeasured common input neuron is from neuron 1's subpopulation (cc^j_{31} from Eq. (4.12) exceeded 0.9), the common input factor W is positive, correctly indicating that the correlation was due to common input. Parameters as in Fig. 13B, except that $y_2 = 7.8$, $b_1 = 6.0$, $\tau_3 = 40$, $\sigma_3 = 10$, $f_3 = 0.075$, $\phi_3 = 0$.

both cases, we know that the results indicate a positive correlation due to common input rather than a negative correlation due to a causal connection because the spikes are positively correlated, as demonstrated by the covariogram.

Because of the weak coupling assumption (Section 3.3), our analysis explicitly ignores the effects of chains of connections that are longer than two edges. However, if we include in our simulations such longer chains that introduce correlations between a pair of measured neurons, the analysis will still categorize the source of those connections as arising from a causal connection or common input (subject, of course, to subpopulation ambiguity). If neuron 2 has a long chain of causal connections onto neuron 1, and these connections are strong enough to induce significant correlations between neuron 1 and neuron 2, then the analysis classifies the correlations as resulting from a causal connection (Fig. 15A). On the other hand, if neuron 1 and neuron 2 receive common input from an unmeasured neuron via long chains of connections, and that common input neuron is from a distinct subpopulation, the analysis classifies the correlations as resulting from common input (Fig. 15B). Hence, we can obtain an estimate of the causal subnetwork even in the presence of such long chains of connections.

5.4. Subunit model simulations

For the second set of tests, we simulated networks of subunit model neurons (5.3). The subunit model was chosen so that it deviates from both the linear–nonlinear (5.5) and linear–quadratic–nonlinear (5.6) models for which we have implemented our analysis. The most fundamental

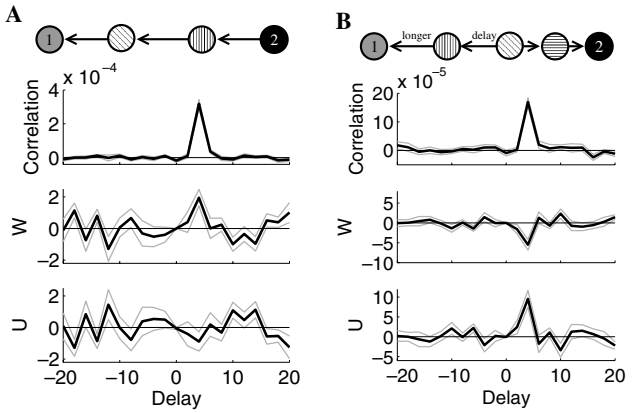

Fig. 15. Distinguish causal connections even with long chains of connections. The analysis can distinguish long chains of causal connections from long chains of common input. Panels as in Fig. 13. (A) A causal connection from neuron 2 to neuron 1 through a chain of unmeasured neurons is identified as a causal connection by W. Most parameters as in Fig. 13B. Exceptions and additional parameters (the unmeasured neurons are indexed by 3 and 4): $\bar{A}_4 = 1$, $\bar{y}_1 = 5.8$, $\bar{y}_1 = 5.9$, $\bar{y}_2 = 7.7$, $\bar{y}_3 = 10.3$, $\bar{y}_4 = 6.9$, $\epsilon_4 = 2.5$, $b_4 = 0$, $\tau_4 = 40$, $\sigma_4 = 15$, $f_4 = 0.07$, $\phi_4 = \pi$, $d_{24} = 0$, $d_{43} = 1$, $d_{31} = 0$, $B_{24} = B_{43} = B_{31} = 10$, $B_{ij} = 0$ for all other i and j. (B) A common input connection onto neuron 1 and neuron 2 via chains of unmeasured neurons is identified as a common input connection by U. Most parameters as in panel (A). Exceptions and additional parameters (the unmeasured neurons are indexed by 3, 4, and 5): $\bar{A}_5 = 1$, $\bar{y}_2 = 8.0$, $\bar{y}_3 = 10$, $\bar{y}_4 = 7.1$, $\bar{y}_5 = 9.2$, $\epsilon_5 = 3.5$, $b_5 = 0$, $\tau_5 = 50$, $\sigma_5 = 10$, $f_5 = 0.05$, $\phi_5 = 5\pi/3$, $d_{34} = 4$, $d_{41} = 2$, $d_{35} = d_{52} = 1$, $B_{34} = B_{41} = B_{35} = B_{52} = 11$, $B_{ij} = 0$ for all other i and j.

difference is the four separate subunits that create a relationship between the spikes and the stimulus that cannot be captured exactly by the linear–nonlinear (5.5) or linear–quadratic–nonlinear (5.6) models. Moreover, even if only one subunit were present, the subunit model (5.3) uses a half-squaring nonlinearity, a binary spike distribution, and a refractory period in contrast to the error function nonlinearity, Poisson spike distribution, and no refractory period assumed by the analysis. Hence, the subunit model simulations are designed to test the performance of our analysis under deviations from model assumptions.

We simulate the same type of network for all our subunit model simulations. The network contains a causal connection from measured neuron 2 onto measured neuron 1. In addition, an unmeasured neuron from a different subpopulation has common input connections onto both neuron 1 and neuron 2. This time, the common input connection has a longer delay onto neuron 2, so the common input leads to correlations that mimic a causal connection from neuron 1 onto neuron 2. Since we use the convention that delay equals the spike time of neuron 1 minus the spike time of neuron 2, the causal connection creates correlations at a positive delay and the common input creates correlations at a negative delay. Hence, our analysis will be successful if W is positive at a positive delay (set to be a delay of 4) and U is positive at a negative delay (set to be a delay of -4).

Results from our first subfield model simulation are shown in Fig. 16. When we analyzed the spikes of neuron 1 and neuron 2 using the linear–nonlinear model (5.5) (Fig. 16A), we failed to distinguish the causal connection from the common input. The estimates of W and U are large for most delays, yet both estimates stay within approximately one standard error from zero at the delays of 4 and -4. This negative result is not surprising because the linear–nonlinear model

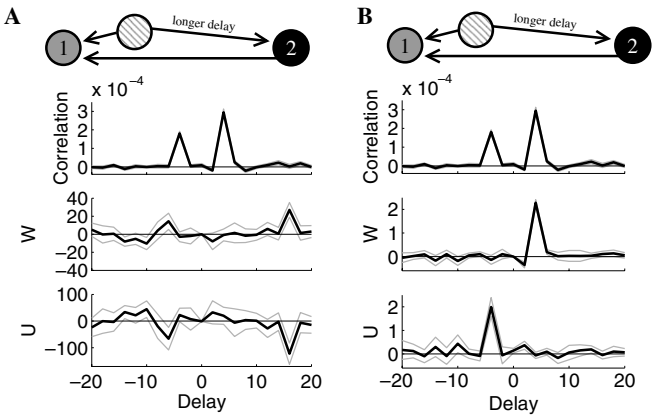

Fig. 16. Results from subfield model simulations with significant contributions from many subfields. Here we simulated a network that has a causal connection from neuron 2 onto neuron 1 (at a delay of 4) as well as common input from an unmeasured neuron with delay to neuron 2 that is four units of time longer (corresponding to a delay of -4 in our convention). Panels as in Fig. 13. (A) The analysis based on the linear–nonlinear model failed to determine the sources of the correlations. At the delays of 4 and -4, both W and U differed insignificantly from zero. Both W and U fluctuated between large values (note scales). The results were invalid because the linear–nonlinear model failed to capture the nonlinear nature of the neurons' response to the stimulus. (B) The analysis based on the linear–quadratic–nonlinear model did successfully distinguish the causal connection at a delay of 4 (W is significantly positive) and the common input at a delay of -4 (U is significantly positive). Although the linear–quadratic–nonlinear model only crudely approximates the nonlinear nature of the subfield model, it captures enough for the analysis to determine the connectivity pattern. Kernel parameters as in Fig. 13B except that $f_3 = 0.02$. Other parameter: $A_{1,1} = 0.006$, $A_{2,1} = 0.007$, $A_{3,1} = 0.009$, $A_{1,2} = 0.005$, $A_{2,2} = 0.006$, $A_{3,2} = 0.008$, $A_{1,3} = 0.004$, $A_{2,3} = 0.005$, $A_{3,3} = 0.007$, $A_{1,4} = A_{2,4} = A_{3,4} = 0.002$, $y_{1,k} = 0.1$, $y_{2,k} = 0.2$, $y_{3,k} = 0.3$ for $k = 1, 2, 3$, and 4, $d_{21} = 3$, $d_{31} = 1$, $d_{32} = 5$, $B_{21} = 3$, $B_{31} = B_{32} = 8$, $B_{i,j} = 0$ for all other i and j.

captures only the (rectified) linear component of the response, but the simulated neurons' response had very little linear component. (When fitting the linear–quadratic–nonlinear model, the α_s exceeded 0.9.) The presented approach requires the use of a model that better corresponds to the neural response.

On the other hand, the analysis based on the linear–quadratic–nonlinear model (5.6) does accurately identify both the causal connection and the common input (Fig. 16B). The analysis successfully distinguishes the common input (at a delay of -4) from the causal connection (at a delay of 4). This result is notable because the linear–quadratic–nonlinear model only crudely approximates some of the spatial nonlinearity of the subfield model (5.3). The fact that the quadratic terms captured enough of the nonlinear response properties to obtain accurate results provides evidence that the approach is at least somewhat robust to deviations from the model assumptions.

Next, we simulated the same network, except that we changed the subfield weights so that the response was dominated by a single subunit ($\bar{A}_{s,1} \gg \bar{A}_{s,2}, \bar{A}_{s,3}, \bar{A}_{s,4}$ for all neurons s). In this way, the differences between the subunit model (5.3) and the linear–nonlinear model (5.5) were minimized. Fig. 17A shows that the analysis based on the linear–nonlinear model successfully distinguishes between the causal connection and the common input. (Results using the linear–quadratic–nonlinear model (5.6) were similar.) In this case, there was a sufficient linear component to the response for the linear–nonlinear model to give good results. (When fitting the linear–quadratic–nonlinear model, the α_s values were between 0.15 and 0.4.)

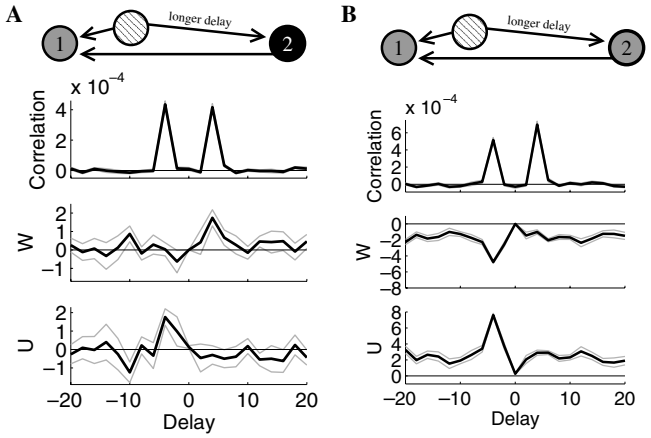

Fig. 17. Results from subfield model simulations dominated by one subfield. We simulated the same network as in Fig. 16, so that the effect of a causal connection appeared at a delay of 4 and common input appeared at a delay of −4. Panels as in Fig. 13. (A) Since one subfield dominates, the subfield model deviates only modestly from the linear–nonlinear model. In this case, analysis based on the linear–nonlinear model correctly identifies the causal connection at a delay of 4 and the common input at a delay of −4. Parameters as in Fig. 16 except that $A_{1,1} = 0.01$, $A_{2,1} = 0.013$, $A_{3,1} = 0.023$, $A_{s,2} = A_{s,3} = 0.002$, $A_{s,4} = 0.001$, $b_s = 0$ for $s = 1,2,3$. (B) The analysis is less robust to model deviations when the measured neurons are from the same subpopulation. (Here, cc^j_{21} of (4.12) exceeded 0.8.) These results are from a simulation identical to that of panel (A), except that the kernel parameters of neuron 2 were modified to closely match those of neuron 1. The causal connection at a delay of 4 was not correctly identified, as W has no peak there. Parameters as in panel (A) except that $f_2 = 0.075$, $\phi_2 = \pi/6$.

The approach is less robust to deviations from the model assumptions when the measured neurons are from the same subpopulation. For example, if we change the parameters of neuron 2 to be similar to those from neuron 1 and repeat the simulation of Fig. 17A, we fail to obtain correct estimates of the connectivity. As shown in Fig. 17B, the linear–nonlinear model fails to capture the neural response properties well enough to distinguish the connectivity. The common input factor U does have a large positive peak at a delay of −4, but W does not have a positive peak at a delay of 4. Also, W is significantly negative and U is significantly positive at most delays. (Analysis with the linear–quadratic–nonlinear model performs no better; in this case, U even shows a slight peak at a delay of 4.) If we set the weight of all subunits but one to zero, then the analysis does again correctly determine the connectivity (not shown). Hence, the failure shown in Fig. 17B indicates that small deviations from the model assumptions can lead to invalid results if the measured neurons are from the same subpopulation.

6. Discussion

We have developed a model-based modular approach to identifying causal connections in a neural network where many neurons remain unmeasured. The approach is modular because the analysis works with a large class of models (which determine the $f(w, \mathbf{X}; \theta)$ of Eq. (3.3)). The network analysis can use models, and algorithms to determine their parameters, that have been developed independently. When a model captures the neurons' behavior sufficiently well,

we can distinguish causal connections from common input that originates from unmeasured neurons (subject to the subpopulation ambiguity described in Section 2). The approach is modestly robust to deviations from the model when analyzing connectivity between two measured neurons that have dissimilar response properties.

We stress that though our identification of causal connections contains the subpopulation ambiguity described in Section 2, the identified connections are between individual neurons, not between groups of neurons. (Subpopulation ambiguity only plays a role in interpreting the precise identity of those individual neurons.) The activity of each node on the network model (3.3) is a spike train of a single neuron. The analysis is designed to be applied to simultaneous measurements of spikes trains of individual neurons.

6.1. Two heuristic explanations for the results

The distinction between the causal connections and the common input is captured in the final Eq. (4.19) for the likelihood of observing the measured neurons' spikes (which we obtained by averaging over unmeasured neurons' spikes). In this equation, the only difference between the causal connection W terms and the common input U terms is the factor of f'/f in the U term. A heuristic explanation for this factor is that in the common input case, unlike the causal connection case, the correlation is induced because both measured neurons are receiving a connection. When expanding the full network Eq. (4.7) in terms of $\bar{W}$, derivatives of f correspond to the effect of receiving a connection. Hence, in the U terms of the final Eq. (4.19), the presence of the connection onto neuron $\tilde{q}$ has replaced a $f(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})$ factor with the derivative $f'(0,\mathbf{x};\theta^{i-j}_{\tilde{q}})$. We exploit the difference between f and its derivative with respect to W to distinguish common input from a causal connection. (See below for a discussion on how to make the approach more robust in case this difference is small.)

We also reiterate a heuristic explanation that follows more naturally from the mathematics of Ref. [3], but still gives some intuition why one may be able to distinguish causal connections from common input even in the present context. In the causal connection network of Fig. 13A, if the external variables $\mathbf{X}$ happened to be optimal for neuron 2 followed by optimal for neuron 1, the effect of the connection would be enhanced. (Neuron 1 would be highly likely to receive an input from neuron 2 just as it was primed to fire a spike due to the external variables.) In contrast, in the common input network of Fig. 13B, there is no connection from neuron 2's subpopulation onto neuron 1's subpopulation. Hence, if the external variables happened to be optimal for neuron 2 followed by optimal for neuron 1, there is no connection whose effect would be enhanced. This heuristic argument suggests that the relationship between the external variables and the measured neurons' spike pairs should be different in the causal connection configuration as opposed to the common input configuration. Although we do not explicitly rely on such differences between causal connection and common input in the analysis, the fact that such differences exist enables our analysis to succeed in determining the causal connections.[17]

[17] Note that in the common input configuration of Fig. 14A (but not in the common input configuration of Fig. 14B), there is a connection from neuron 2's subpopulation onto neuron 1's subpopulation. The effect of this connection would also be enhanced when the external variables happened to be optimal for neuron 2 followed by optimal for neuron 1. Hence this heuristic argument can also explain the presence of the subpopulation ambiguity.

6.2. Relationship to other work

The present approach builds on earlier work [3,12] that addressed the same problem of determining connectivity in the presence of unmeasured neurons and introduced the concept of a subpopulation. The earlier work applied only to neurons responding to a white noise stimulus in a manner that could be well approximated by a linear–nonlinear model. In contrast, the present work allows the use of a large class of models that relate neuronal response to measurable external variables. Moreover, the mathematical approach has completely changed in the present work. The earlier work employed a combination of spike correlation analysis [9–11] and white noise analysis [13–15], while the present work is based on maximum likelihood estimators. Using maximum likelihood estimators has increased the statistical efficiency of the approach. However, as demonstrated by the long simulations required, the approach still requires large amounts of data to make a successful determination of causal connections.

There is a large literature about determining connections between neurons [9–11,16–18]. Methods such as partial coherence [16] are designed to account for common input from *measured* neurons. Recent examples [17,18] employ maximum likelihood estimates similar to what we have used here. However, we are unaware of others who explicitly address the effects of *unmeasured* neurons, which is the focus of this paper.

6.3. Validation methods

The demonstrated failures in Figs. 16A and 17B indicate the importance of validating the results before they can be trusted to be an accurate representation of the causal structure of the network. When a model f is used so that the effective single neuron model (3.4) does not sufficiently well capture the neurons' behavior, the network analysis may lead to inaccurate results. Developing methods to identify those invalid cases will be crucial for the viability of this approach.

Standard point process model analyses, such as those based on the time-rescaling theorem [2,19], will be one place to start the development of validation methods. Such analyses should be able to detect cases such as that of Fig. 16A, where the model fails to capture the relationship between the neural response and the external variables. However, these analyses also detect those deviations from the model to which the network analysis is robust.

Since any model of the form (3.4) will only approximately capture the actual neuron response, deviations from the model prediction will be present for any model. A useful validation method must be able to distinguish between critical deviations that invalidate the network analysis and harmless deviations to which the network analysis is robust. We have not yet developed validations that meet this requirement.

For example, model (3.4) for any f should fail validations based the time-rescaling theorem on all simulations of the subunit model (5.3) (Figs. 16 and 17). The subunit model includes a refractory period while the model (3.4) is an inhomogeneous Poisson process. The refractory period should induce a deviation from Poisson that will be detected after rescaling time. However, the presence of the refractory period did not appear to degrade the performance of our network analysis.

Moreover, such analyses treat the response of each neuron separately and hence cannot distinguish between the invalid case of Fig. 17B and the case of Fig. 17A in which we could properly

determine the causal connection. In both cases, the response of the individual neurons deviated from the linear–nonlinear model (using Eqs. (3.4) and (5.5)) by an equal amount. The difference between the two cases is that only in Fig. 17B were there stimulus-dependent correlations [20] between neuron 1 and neuron 2 (because both neurons responded to similar stimulus features). We believe the network analysis failed in this case because the model did not adequately capture those stimulus-dependent correlations. A validation method that could detect a poor representation of these stimulus-dependent correlations may be able to indicate when the results based on that model would be invalid.

6.4. Further extensions

Other than robustness to model deviations and validations to detect critical deviations, one further issue may need to be addressed before this approach can be confidently applied to analyzing experimental data. As mentioned above, the current implementation of this approach depends on exploiting the difference between the model function f and its derivative with respect to W. As a consequence, if $f(w, \mathbf{x}; \theta)$ had an exponential dependence on w, the difference between the W term and the U term would disappear, and the method would fail. This observation was one motivation for choosing an error function nonlinearity (5.2) in our models. This sigmoidal dependence on w created a significant difference between $f(0, \mathbf{x}; \theta^{i-j}_{\tilde{q}})$ and its derivative.

However, an exponential functional form of f is commonly used in models, in part because it allows one to prove nice properties of the likelihood surface [4]. Moreover, even if one employs a sigmoidal nonlinearity in the analysis, one may legitimately worry of the consequences if the actual data are generated by a process with a nearly exponential dependence on w.

A sigmoidal form of spiking probability can represent saturation of a neuron's response when it is driven strongly. Such an effect could be obtained even with an exponential form if one allowed a strong history dependence in the spiking probability. For example, one could model a refractory period by letting the $\bar{W}^j_{qq}$ factors be large and negative for small j. If one included such strong history dependence, it could not be expanded out in a Taylor series. The averaged model (3.4) would have to include the dependence of the probability of a spike R^i_p at time i on the history of spikes R^{i-j}_p at previous times $i - j$. Extending the approach to include strong history dependence may increase the robustness to the form of nonlinearity in f. Moreover, with such non-Poisson spiking behavior included in the averaged model (3.4), history dependence such as from a refractory period will no longer be a deviation from model predictions. Hence, validation approaches such as those involving the time-rescaling theorem might only detect other types of deviations such as those that actually invalidate the results of our network analysis.

Another direction to take this research is to investigate the effects of higher-order powers of $\bar{W}$ such as those that capture effects of three or more connections. Given the combinatorial explosion that occurs as one adds higher-order effects, one will be limited in how far one can push this type of analysis. Nonetheless, additional insights into the network structure may be attainable with judicious inclusion of some higher-order effects, especially if one merely desires to determine which of a small set of causal connectivity patterns is more likely to have produced the observed data.

In conclusion, we believe the presented approach to network analysis can evolve into a useful tool for analyzing the causal structure in networks, provided one can model the relationship between the response of a node and a set of measurable external variables. To create reliable tools, we must develop better approaches to validating the robustness of the approach to model deviations. Once such validation methods are available, this approach may provide a window into the causal structure of the brain's neural networks or other networks even when one can simultaneously measure the activity of only a relatively small number of nodes.

Appendix A. Derivation of single neuron results

We outline the calculation to obtain an expression for the original parameters of model (3.3) in terms of the effective parameters of the averaged model (3.4). We derive this relationship by averaging model (3.3) over the activity of all neurons except one.

We start with (4.2) for $R^{i_1}_{s_1}$, the activity of node s_1 at time i_1, expressed as an average over possible values of $\mathbf{r}$, which we rewrite here:

$$\Pr(R^{i_1}_{s_1} = r^{i_1}_{s_1} | \mathbf{X} = \mathbf{x}) = \sum_{\substack{\mathbf{r} \\ r^i_{s_1} \text{ fixed}}} \prod_{i_2} \prod_{s_2} \Gamma\left(r^{i_2}_{s_2}, f\left(\sum_{\tilde{s}} \sum_{j>0} \bar{W}^j_{\tilde{s}s_2} r^{i_2-j}_{\tilde{s}}, \mathbf{x}; \bar{\theta}^{i_2}_{s_2} \right) \right).$$

We take a second-order Taylor series with respect to $\bar{W}$. In terms of the shorthand notation (4.3), the activity of neuron s_1 at time i_1 is approximately

$$\begin{aligned}
\Pr(R^{i_1}_{s_1} = r^{i_1}_{s_1} | \mathbf{X} = \mathbf{x}) = & \sum_{\substack{\mathbf{r} \\ r^{i_1}_{s_1} \text{ fixed}}} \prod_{s_2, i_2} \bar{\Gamma}^{i_2}_{s_2} \\
& + \sum_{\substack{\mathbf{r} \\ r^{i_1}_{s_1} \text{ fixed}}} \sum_{s_2, \tilde{s}_2} \sum_{i_2, j_2} \bar{W}^{j_2}_{\tilde{s}_2 s_2} r^{i_2-j_2}_{\tilde{s}_2} \partial \bar{\Gamma}^{i_2}_{s_2} \prod_{\substack{s_3, i_3 \\ (s_3,i_3) \neq (s_2,i_2)}} \bar{\Gamma}^{i_3}_{s_3} \\
& + \frac{1}{2} \sum_{\substack{\mathbf{r} \\ r^{i_1}_{s_1} \text{ fixed}}} \sum_{s_2, \tilde{s}_2, \tilde{s}_3} \sum_{i_2, j_2, j_3} \bar{W}^{j_2}_{\tilde{s}_2 s_2} \bar{W}^{j_3}_{\tilde{s}_3 s_2} r^{i_2-j_2}_{\tilde{s}_2} r^{i_2-j_3}_{\tilde{s}_3} \partial^2 \bar{\Gamma}^{i_2}_{s_2} \prod_{\substack{s_3, i_3 \\ (s_3,i_3) \neq (s_2,i_2)}} \bar{\Gamma}^{i_3}_{s_3} \\
& + \frac{1}{2} \sum_{\substack{\mathbf{r} \\ r^{i_1}_{s_1} \text{ fixed}}} \sum_{\substack{s_2, \tilde{s}_2 \\ s_3, \tilde{s}_3}} \sum_{\substack{i_2, j_2, i_3, j_3 \\ (s_3,i_3) \neq (s_2,i_2)}} \bar{W}^{j_2}_{\tilde{s}_2 s_2} \bar{W}^{j_3}_{\tilde{s}_3 s_3} r^{i_2-j_2}_{\tilde{s}_2} r^{i_3-j_3}_{\tilde{s}_3} \partial \bar{\Gamma}^{i_2}_{s_2} \partial \bar{\Gamma}^{i_3}_{s_3} \prod_{\substack{s_4, i_4 \\ (s_4,i_4) \neq (s_2,i_2) \\ (s_4,i_4) \neq (s_3,i_3)}} \bar{\Gamma}^{i_4}_{s_4} + \mathrm{O}(\bar{W}^3).
\end{aligned} \tag{A.1}$$

Ignoring temporal indices, each term in Eq. (A.1) corresponds to a (not necessarily connected) digraph with one or two edges corresponding to the combination of the $\bar{W}^j_{\tilde{s}s}$ present in the term. For example, terms in the third line contain $\bar{W}^{j_2}_{\tilde{s}_2 s_2} \bar{W}^{j_3}_{\tilde{s}_3 s_2}$, which corresponds to the digraph where two neurons ($\tilde{s}_2$ and $\tilde{s}_3$) have a connection onto neuron s_2.

Note that all network effects have been removed from the $\bar{\Gamma}$ (and the derivatives $\partial\bar{\Gamma}$ and $\partial^2\bar{\Gamma}$). Since each $\bar{\Gamma}^i_s$ depends only on a single r^i_s, we can factor each term of Eq. (A.1) into factors that depend only on a single r^i_s and evaluate the sum over all r^i_s independently from all other factors. For example, since $\bar{\Gamma}^i_s$ is the probability distribution of a Poisson random variable, the sum of $\bar{\Gamma}^i_s$ times a power of r^i_s (over all possible values of r^i_s) is simply a moment of the Poisson distribution. We can also explicitly calculate the sums involving derivatives of $\bar{\Gamma}^i_s$ from the definition in Eq. (3.1). The results are summarized in the following:

$$\sum_{r_s^i} \bar{\Gamma}_s^i = 1 \qquad \sum_{r_s^i} \partial \bar{\Gamma}_s^i = \sum_{r_s^i} \partial^2 \bar{\Gamma}_s^i = 0$$
$$\sum_{r_s^i} r_s^i \bar{\Gamma}_s^i = f(0, \mathbf{x}; \bar{\theta}_s^i) \qquad \sum_{r_s^i} (r_s^i)^2 \bar{\Gamma}_s^i = f(0, \mathbf{x}; \bar{\theta}_s^i)(1 + f(0, \mathbf{x}; \bar{\theta}_s^i))$$
$$\sum_{r_s^i} r_s^i \partial \bar{\Gamma}_s^i = f'(0, \mathbf{x}; \bar{\theta}_s^i) \qquad \sum_{r_s^i} r_s^i \partial^2 \bar{\Gamma}_s^i = f''(0, \mathbf{x}; \bar{\theta}_s^i). \tag{A.2}$$

The first equation in (A.2) is simply the probability of having any type of activity. In each term of Eq. (A.1), most of the $\bar{\Gamma}_s^i$ are alone and disappear by summing to one in this way.

We know that the activity of neuron s_1 should not be affected by digraphs that contain nodes without a path to neuron s_1. These terms will contain a derivative of $\bar{\Gamma}_s^i$ that is not matched by a r_s^i. According to (A.2), these unmatched derivatives sum to zero, so the terms disappear as expected.

More generally, to simplify Eq. (A.1), we separate all the different types of digraphs hidden in its terms. To accomplish this combinatorial task, we find all the ways in which the neuron subscripts and temporal superscripts of Eq. (A.1) can be identified. (One must remember the causal restriction that $j > 0$.) Because the terms are not summed over $r_{s_1}^{i_1}$, the neuron subscript s_1 and the temporal superscript i_1 play a special role. For example, in the second line of (A.1), we identify $(s_2, i_2) = (s_1, i_1)$ so that the term corresponds to a connection onto neuron s_1 that affects its activity at time i_1. (These are the only terms from the second line that survive.) Once this identification is made, we calculate the sum over all r_s^i for $(s, i) \neq (s_1, i_1)$ using the identities of Eq. (A.2). In the end, we obtain Eq. (4.4), which we repeat here:

$$\begin{aligned}
\Pr(R_{s_1}^{i_1} = r_{s_1}^{i_1} | \mathbf{X} = \mathbf{x}) = \bar{\Gamma}_{s_1}^{i_1} &+ \sum_{s_2} \sum_{j_2} \bar{W}_{s_2 s_1}^{j_2} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) \partial \bar{\Gamma}_{s_1}^{i_1} \\
&+ \frac{1}{2} \sum_{s_2} \sum_{j_2} (\bar{W}_{s_2 s_1}^{j_2})^2 f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) \partial^2 \bar{\Gamma}_{s_1}^{i_1} \\
&+ \frac{1}{2} \sum_{s_2, s_3} \sum_{j_2, j_3} \bar{W}_{s_2 s_1}^{j_2} \bar{W}_{s_3 s_1}^{j_3} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) f(0, \mathbf{x}; \bar{\theta}_{s_3}^{i_1 - j_3}) \partial^2 \bar{\Gamma}_{s_1}^{i_1} \\
&+ \sum_{s_3, s_2} \sum_{j_2, j_3} \bar{W}_{s_2 s_3}^{j_2} \bar{W}_{s_3 s_1}^{j_3} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_3 - j_2}) f'(0, \mathbf{x}; \bar{\theta}_{s_3}^{i_1 - j_3}) \partial \bar{\Gamma}_{s_1}^{i_1} \\
&+ \mathrm{O}(\bar{W}^3).
\end{aligned}$$

Combining this equation with the averaged model (3.4), and using the unbarred shorthand notation (4.5), we obtain the following expression for Γ in terms of original model parameters:

$$\begin{aligned}
\Gamma_{s_1}^{i_1} = \bar{\Gamma}_{s_1}^{i_1} &+ \sum_{s_2} \sum_{j_2} \bar{W}_{s_2 s_1}^{j_2} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) \partial \bar{\Gamma}_{s_1}^{i_1} \\
&+ \frac{1}{2} \sum_{s_2} \sum_{j_2} (\bar{W}_{s_2 s_1}^{j_2})^2 f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) \partial^2 \bar{\Gamma}_{s_1}^{i_1} \\
&+ \frac{1}{2} \sum_{s_2, s_3} \sum_{j_2, j_3} \bar{W}_{s_2 s_1}^{j_2} \bar{W}_{s_3 s_1}^{j_3} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_2}) f(0, \mathbf{x}; \bar{\theta}_{s_3}^{i_1 - j_3}) \partial^2 \bar{\Gamma}_{s_1}^{i_1} \\
&+ \sum_{s_3, s_2} \sum_{j_2, j_3} \bar{W}_{s_2 s_3}^{j_2} \bar{W}_{s_3 s_1}^{j_3} f(0, \mathbf{x}; \bar{\theta}_{s_2}^{i_1 - j_3 - j_2}) f'(0, \mathbf{x}; \bar{\theta}_{s_3}^{i_1 - j_3}) \partial \bar{\Gamma}_{s_1}^{i_1} \\
&+ \mathrm{O}(\bar{W}^3).
\end{aligned} \tag{A.3}$$

In the rest of our analysis, we will want to replace expressions in terms of $\bar{\theta}$ with expressions in terms of θ. To accomplish this, we need expressions for Γ_s^i and $f(0,\mathbf{x};\bar{\theta}_s^i)$, as well as their derivatives, that include only effective parameters. To begin, we rearrange Eq. (A.3) and drop the subscripts off s_1 and i_1,

$$\begin{aligned}\bar{\Gamma}_s^i =& \Gamma_s^i - \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2})\partial\bar{\Gamma}_s^i \\ &- \frac{1}{2}\sum_{s_2}\sum_{j_2} (\bar{W}_{s_2 s}^{j_2})^2 f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2})\partial^2\bar{\Gamma}_s^i \\ &- \frac{1}{2}\sum_{s_2,s_3}\sum_{j_2,j_3} \bar{W}_{s_2 s}^{j_2}\bar{W}_{s_3 s}^{j_3} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2}) f(0,\mathbf{x};\bar{\theta}_{s_3}^{i-j_3})\partial^2\bar{\Gamma}_s^i \\ &- \sum_{s_3,s_2}\sum_{j_2,j_3} \bar{W}_{s_2 s_3}^{j_2}\bar{W}_{s_3 s}^{j_3} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_3-j_2}) f'(0,\mathbf{x};\bar{\theta}_{s_3}^{i-j_3})\partial\bar{\Gamma}_s^i \\ &+ \mathrm{O}(\bar{W}^3). \end{aligned} \tag{A.4}$$

Next, we invoke the assumption from Section 3.2 that the derivatives of $f(w,\mathbf{x};\theta)$ with respect to w are equivalent to derivatives with respect to some function of $\mathbf{x}$ and θ. Hence, if we differentiate the left-hand side of (A.4) with respect w, we need to differentiate those f on the right-hand side that contain θ_s^i. Since we will need only a first-order approximation to the first derivative, we just differentiate the first line of (A.4) to obtain

$$\partial\bar{\Gamma}_s^i = \partial\Gamma_s^i - \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2})\partial^2\bar{\Gamma}_s^i + \mathrm{O}(\bar{W}^2)$$

We repeat this process to conclude that a zeroth-order approximation of the second derivative is

$$\partial^2\bar{\Gamma}_s^i = \partial^2\Gamma_s^i + \mathrm{O}(\bar{W}).$$

Next, we take the expected value of the above three equations by multiplying by r_s^i, summing over all possible values of r_s^i, and using the identities of Eq. (A.2). Retaining only the order of approximation that we will need later, we write the results as

$$\begin{aligned} f(0,\mathbf{x};\bar{\theta}_s^i) =& f(0,\mathbf{x};\theta_s^i) - \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2}) f'(0,\mathbf{x};\bar{\theta}_s^i) + \mathrm{O}(\bar{W}^2) \\ f'(0,\mathbf{x};\bar{\theta}_s^i) =& f'(0,\mathbf{x};\theta_s^i) - \sum_{s_2}\sum_{j_2} \bar{W}_{s_2 s}^{j_2} f(0,\mathbf{x};\bar{\theta}_{s_2}^{i-j_2}) f''(0,\mathbf{x};\bar{\theta}_s^i) + \mathrm{O}(\bar{W}^2) \\ f''(0,\mathbf{x};\bar{\theta}_s^i) =& f''(0,\mathbf{x};\theta_s^i) + \mathrm{O}(\bar{W}). \end{aligned}$$

Note that the above equations imply that all the expressions involving $\bar{\theta}_s^i$ are a zeroth-order approximation to the same expressions involving θ_s^i. Therefore, in any approximation, we can simply erase the bars off the $\bar{\theta}$ in the highest-order terms without decreasing the order

of the approximation. To obtain a second-order approximation of $\bar{\Gamma}^i_s$, we also need to make a first-order approximation to the $\bar{W}^{j_2}_{s_2 s}$ term in the first line of Eq. (A.4) (i.e., substitute in the above first-order corrections for $\partial\bar{\Gamma}^i_s$ and $f(0,\mathbf{x};\bar{\theta}^i_s)$). We end up with the effective parameter equation (4.6) giving the original model parameters in terms of the effective parameters of the averaged model (3.4).

Appendix B. Derivation of network results

We outline the calculation by which we obtain Eq. (4.8) for the probability of observing a given sequence $\mathbf{r}_{\mathscr{Q}}$ of spikes from the measured neurons. We begin with Eq. (4.7) for the probability of observing $\mathbf{r}_{\mathscr{Q}}$, which we rewrite here:

$$\Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x}) = \sum_{\mathbf{r}_{\mathscr{P}}} \prod_i \prod_s \Gamma\left(r^i_s, f\left(\sum_{\tilde{s}} \sum_{j>0} \bar{W}^j_{\tilde{s}s} r^{i-j}_{\tilde{s}}, \mathbf{x}; \bar{\theta}^i_s\right)\right).$$

Recall that the sum is over all possible values of $\mathbf{r}_{\mathscr{P}}$, i.e., over all possible combinations of spikes from the unmeasured neurons. We expand in $\bar{W}$ to obtain

$$\begin{aligned}
\Pr(\mathbf{R}_{\mathscr{Q}} = \mathbf{r}_{\mathscr{Q}} | \mathbf{X} = \mathbf{x}) &= \sum_{\mathbf{r}_{\mathscr{P}}} \prod_{s_2, i_2} \bar{\Gamma}^{i_2}_{s_2} \\
&+ \sum_{\mathbf{r}_{\mathscr{P}}} \sum_{s_2, \tilde{s}_2} \sum_{i_2, j_2} \bar{W}^{j_2}_{\tilde{s}_2 s_2} r^{i_2 - j_2}_{\tilde{s}_2} \partial\bar{\Gamma}^{i_2}_{s_2} \prod_{\substack{s_3, i_3 \\ (s_3,i_3)\neq(s_2,i_2)}} \bar{\Gamma}^{i_3}_{s_3} \\
&+ \frac{1}{2} \sum_{\mathbf{r}_{\mathscr{P}}} \sum_{s_2, \tilde{s}_2, \tilde{s}_3} \sum_{i_2, j_2, j_3} \bar{W}^{j_2}_{\tilde{s}_2 s_2} \bar{W}^{j_3}_{\tilde{s}_3 s_2} r^{i_2 - j_2}_{\tilde{s}_2} r^{i_2 - j_3}_{\tilde{s}_3} \partial^2\bar{\Gamma}^{i_2}_{s_2} \prod_{\substack{s_3, i_3 \\ (s_3,i_3)\neq(s_2,i_2)}} \bar{\Gamma}^{i_3}_{s_3} \\
&+ \frac{1}{2} \sum_{\mathbf{r}_{\mathscr{P}}} \sum_{\substack{s_2, \tilde{s}_2 \\ s_3, \tilde{s}_3}} \sum_{\substack{i_2, j_2, i_3, j_3 \\ (s_3,i_3)\neq(s_2,i_2)}} \bar{W}^{j_2}_{\tilde{s}_2 s_2} \bar{W}^{j_3}_{\tilde{s}_3 s_3} r^{i_2 - j_2}_{\tilde{s}_2} r^{i_3 - j_3}_{\tilde{s}_3} \partial\bar{\Gamma}^{i_2}_{s_2} \partial\bar{\Gamma}^{i_3}_{s_3} \prod_{\substack{s_4, i_4 \\ (s_4,i_4)\neq(s_2,i_2) \\ (s_4,i_4)\neq(s_3,i_3)}} \bar{\Gamma}^{i_4}_{s_4} \\
&+ \mathrm{O}(\bar{W}^3),
\end{aligned} \tag{B.1}$$

which is analogous to Eq. (A.1) for the single neuron case. As before, we do not explicitly denote the causality of the connections, but implicitly assume all $\bar{W}^j_{s_1 s_2} = 0$ for $j \leqslant 0$.

Eq. (B.1) contains arbitrary digraphs of two or fewer edges hidden in its terms, depending on the identities of the neurons indexed by various forms of s, which could be measured or unmeasured neurons. Proceeding exactly as we did for the single neuron case, we separate out the different ways in which subsets of the indices could correspond to combinations of unmeasured neurons. Only the identities of unmeasured neurons matter, because we are averaging only over the activity of unmeasured neurons. For this step of the analysis, we do not modify the factors involving measured neurons.

For each identified combination of indices that correspond to the activity of a single unmeasured neuron at a single time bin, we calculate the sum over all possible number of spikes using

the appropriate identity in Eq. (A.2). To simplify the notation, we implicitly assume that all indices that are a form of q are measured indices ($q, \tilde{q} \in \mathcal{Q}$) and all indices that are a form of p are unmeasured indices ($p, \tilde{p} \in \mathcal{P}$). We also use the following shorthand combinations of the $\bar{\Gamma}^i_s$ of Eq. (4.3)

$$\begin{aligned}
\bar{\Gamma} &= \prod_{q\in\mathcal{Q}} \prod_i \bar{\Gamma}(r^i_q, f(0,\mathbf{x};\bar{\theta}^i_q)) \\
\partial_{W^{i_1}_{q_1}} \bar{\Gamma} &= \frac{\partial}{\partial W^{i_1}_{q_1}} \left[\prod_{q_2\in\mathcal{Q}} \prod_{i_2} \bar{\Gamma}(r^{i_2}_{q_2}, f(W^{i_2}_{q_2}, \mathbf{x}; \bar{\theta}^{i_2}_{q_2})) \right]_{\mathbf{W}=\mathbf{O}} \\
\partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \bar{\Gamma} &= \frac{\partial^2}{\partial W^{i_1}_{q_1} \partial W^{i_2}_{q_2}} \left[\prod_{q_3\in\mathcal{Q}} \prod_{i_3} \bar{\Gamma}(r^{i_3}_{q_3}, f(W^{i_3}_{q_3}, \mathbf{x}; \bar{\theta}^{i_3}_{q_3})) \right]_{\mathbf{W}=\mathbf{O}}.
\end{aligned} \tag{B.2}$$

The probability of observing a sequence of spikes in the measured neurons becomes

$$\begin{aligned}
\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x}) = \bar{\Gamma} &+ \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{q}_1 q_1} r^{i_1-j_1}_{\tilde{q}_1} \partial_{W^{i_1}_{q_1}} \bar{\Gamma} \\
&+ \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{p}_1 q_1} f(0,\mathbf{x};\bar{\theta}^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \bar{\Gamma} \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{q}_1 \\ q_2,\tilde{q}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{q}_1 q_1} \bar{W}^{j_2}_{\tilde{q}_2 q_2} r^{i_1-j_1}_{\tilde{q}_1} r^{i_2-j_2}_{\tilde{q}_2} \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \bar{\Gamma} \\
&+ \sum_{\substack{q_1,\tilde{p}_1 \\ q_2,\tilde{q}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{p}_1 q_1} \bar{W}^{j_2}_{\tilde{q}_2 q_2} f(0,\mathbf{x};\bar{\theta}^{i_1-j_1}_{\tilde{p}_1}) r^{i_2-j_2}_{\tilde{q}_2} \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \bar{\Gamma} \\
&+ \frac{1}{2} \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} (\bar{W}^{j_1}_{\tilde{p}_1 q_1})^2 f(0,\mathbf{x};\bar{\theta}^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \bar{\Gamma} \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{p}_1 \\ q_2,\tilde{p}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{p}_1 q_1} \bar{W}^{j_2}_{\tilde{p}_2 q_2} f(0,\mathbf{x};\bar{\theta}^{i_1-j_1}_{\tilde{p}_1}) f(0,\mathbf{x};\bar{\theta}^{i_2-j_2}_{\tilde{p}_2}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \bar{\Gamma} \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{p}_1 \\ q_2}} \sum_{\substack{i_1,j_1,j_2 \\ (q_2,j_2)\neq(q_1,j_1)}} \bar{W}^{j_1}_{\tilde{p}_1 q_1} \bar{W}^{j_2}_{\tilde{p}_1 q_2} f(0,\mathbf{x};\bar{\theta}^{i_1}_{\tilde{p}_1}) \partial_{W^{i_1+j_1}_{q_1}} \partial_{W^{i_1+j_2}_{q_2}} \bar{\Gamma} \\
&+ \sum_{\substack{p_1,\tilde{q}_1 \\ q_2}} \sum_{j_1,i_2,j_2} \bar{W}^{j_1}_{\tilde{q}_1 p_1} \bar{W}^{j_2}_{p_1 q_2} r^{i_2-j_2-j_1}_{\tilde{q}_1} f'(0,\mathbf{x};\bar{\theta}^{i_2-j_2}_{p_1}) \partial_{W^{i_2}_{q_2}} \bar{\Gamma} \\
&+ \sum_{\substack{p_1,\tilde{p}_1 \\ q_2}} \sum_{j_1,i_2,j_2} \bar{W}^{j_1}_{\tilde{p}_1 p_1} \bar{W}^{j_2}_{p_1 q_2} f(0,\mathbf{x};\bar{\theta}^{i_2-j_2-j_1}_{\tilde{p}_1}) f'(0,\mathbf{x};\bar{\theta}^{i_2-j_2}_{p_1}) \partial_{W^{i_2}_{q_2}} \bar{\Gamma} \\
&+ \mathrm{O}(\bar{W}^3).
\end{aligned} \tag{B.3}$$

The number of digraphs represented by Eq. (B.3) is enormous, and we do not attempt to characterize them here. Instead, we simplify the equation by writing everything in terms of effective parameters by using the effective parameter Eqs. (4.6) repeatedly.

The first step is to replace the $\bar{\Gamma}$ of the first term with the effective Γ defined by Eq. (4.9). We take the product of the first effective parameter Eq. (4.6) over all $q_1 \in \mathcal{Q}$ and all i, retaining only a second-order approximation in $\bar{W}$. We write the result in terms of Γ and its derivatives defined in (4.9). After distinguishing between all combinations of measured and unmeasured indices, the result is the following equation:

$$\begin{aligned}
\bar{\Gamma} =& \Gamma - \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{q}_1 q_1} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) \partial_{W^{i_1}_{q_1}} \Gamma \\
&- \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{p}_1 q_1} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \Gamma \\
&- \frac{1}{2} \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} (\bar{W}^{j_1}_{\tilde{q}_1 q_1})^2 f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \Gamma \\
&- \frac{1}{2} \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} (\bar{W}^{j_1}_{\tilde{p}_1 q_1})^2 f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \Gamma \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{q}_1 \\ q_2,\tilde{q}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{q}_1 q_1} \bar{W}^{j_2}_{\tilde{q}_2 q_2} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{q}_2}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \Gamma \\
&+ \sum_{\substack{q_1,\tilde{p}_1 \\ q_2,\tilde{q}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{p}_1 q_1} \bar{W}^{j_2}_{\tilde{q}_2 q_2} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1}) f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{q}_2}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \Gamma \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{p}_1 \\ q_2,\tilde{p}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{p}_1 q_1} \bar{W}^{j_2}_{\tilde{p}_2 q_2} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1}) f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{p}_2}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \Gamma + \mathrm{O}(\bar{W}^3).
\end{aligned}$$

We substitute this expression into Eq. (B.3) to remove the bars from the first term. We also remove the bars from all terms that are second-order in the $\bar{W}$, as we can do that without changing the order of the approximation. The probability becomes

$$\begin{aligned}
\Pr(\mathbf{R}_{\mathcal{Q}} = \mathbf{r}_{\mathcal{Q}} | \mathbf{X} = \mathbf{x}) =& \Gamma + \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{q}_1 q_1} r^{i_1-j_1}_{\tilde{q}_1} \partial_{W^{i_1}_{q_1}} \bar{\Gamma} \\
&- \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{q}_1 q_1} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) \partial_{W^{i_1}_{q_1}} \Gamma \\
&+ \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{p}_1 q_1} f(0,\mathbf{x};\bar{\theta}^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \bar{\Gamma} \\
&- \sum_{q_1,\tilde{p}_1} \sum_{i_1,j_1} \bar{W}^{j_1}_{\tilde{p}_1 q_1} f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1}) \partial_{W^{i_1}_{q_1}} \Gamma \\
&- \frac{1}{2} \sum_{q_1,\tilde{q}_1} \sum_{i_1,j_1} (\bar{W}^{j_1}_{\tilde{q}_1 q_1})^2 f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) \partial_{W^{i_1}_{q_1}} \partial_{W^{i_1}_{q_1}} \Gamma \\
&+ \frac{1}{2} \sum_{\substack{q_1,\tilde{q}_1 \\ q_2,\tilde{q}_2}} \sum_{\substack{i_1,j_1 \\ i_2,j_2}} \bar{W}^{j_1}_{\tilde{q}_1 q_1} \bar{W}^{j_2}_{\tilde{q}_2 q_2} \left[r^{i_1-j_1}_{\tilde{q}_1} r^{i_2-j_2}_{\tilde{q}_2} + f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{q}_1}) f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{q}_2}) \right] \partial_{W^{i_1}_{q_1}} \partial_{W^{i_2}_{q_2}} \Gamma
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{\substack{q_1,\tilde{p}_1\\ q_2,\tilde{q}_2}}\sum_{\substack{i_1,j_1\\ i_2,j_2}}\bar{W}^{j_1}_{\tilde{p}_1q_1}\bar{W}^{j_2}_{\tilde{q}_2q_2}f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1})\left[r^{i_2-j_2}_{\tilde{q}_2}+f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{q}_2})\right]\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_2}}\Gamma\\
&+\sum_{\substack{q_1,\tilde{p}_1\\ q_2,\tilde{p}_2}}\sum_{\substack{i_1,j_1\\ i_2,j_2}}\bar{W}^{j_1}_{\tilde{p}_1q_1}\bar{W}^{j_2}_{\tilde{p}_2q_2}f(0,\mathbf{x};\theta^{i_1-j_1}_{\tilde{p}_1})f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{p}_2})\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_2}}\Gamma\\
&+\frac{1}{2}\sum_{\substack{q_1,\tilde{p}_1\\ q_2}}\sum_{\substack{i_1,j_1,j_2\\ (q_2,j_2)\neq(q_1,j_1)}}\bar{W}^{j_1}_{\tilde{p}_1q_1}\bar{W}^{j_2}_{\tilde{p}_1q_2}f(0,\mathbf{x};\theta^{i_1}_{\tilde{p}_1})\partial_{W^{i_1+j_1}_{q_1}}\partial_{W^{i_1+j_2}_{q_2}}\Gamma\\
&+\sum_{\substack{p_1,\tilde{q}_1\\ q_2}}\sum_{j_1,i_2,j_2}\bar{W}^{j_1}_{\tilde{q}_1p_1}\bar{W}^{j_2}_{p_1q_2}r^{i_2-j_2-j_1}_{\tilde{q}_1}f'(0,\mathbf{x};\theta^{i_2-j_2}_{p_1})\partial_{W^{i_1+j_2}_{q_2}}\Gamma\\
&+\sum_{\substack{p_1,\tilde{p}_1\\ q_2}}\sum_{j_1,i_2,j_2}\bar{W}^{j_1}_{\tilde{p}_1p_1}\bar{W}^{j_2}_{p_1q_2}f(0,\mathbf{x};\theta^{i_2-j_2-j_1}_{\tilde{p}_1})f'(0,\mathbf{x};\theta^{i_2-j_2}_{p_1})\partial_{W^{i_1+j_2}_{q_2}}\Gamma+\mathrm{O}(\bar{W}^3).
\end{aligned}
\tag{B.4}
$$

The last step is to get rid the bars on the first-order terms. Combining the effective parameter Eqs. (4.6) with the various definitions of the $\bar{\Gamma}$ and Γ (Eqs. (4.3), (B.2), (4.5), and (4.9)), we obtain the expression

$$
\begin{aligned}
\partial_{W^{i_1}_{q_1}}\bar{\Gamma}=&\partial_{W^{i_1}_{q_1}}\Gamma-\sum_{q_2,\tilde{s}_2}\sum_{i_2,j_2}\bar{W}^{j_2}_{\tilde{s}_2q_2}f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{s}_2})\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_2}}\Gamma+\mathrm{O}(\bar{W}^2)\\
=&\partial_{W^{i_1}_{q_1}}\Gamma-\sum_{q_2,\tilde{q}_2}\sum_{i_2,j_2}\bar{W}^{j_2}_{\tilde{q}_2q_2}f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{q}_2})\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_2}}\Gamma\\
&-\sum_{q_2,\tilde{p}_2}\sum_{i_2,j_2}\bar{W}^{j_2}_{\tilde{p}_2q_2}f(0,\mathbf{x};\theta^{i_2-j_2}_{\tilde{p}_2})\partial_{W^{i_1}_{q_1}}\partial_{W^{i_2}_{q_2}}\Gamma+\mathrm{O}(\bar{W}^2).
\end{aligned}
$$

Applying this formula and the effective parameter Eq. (4.6d), we arrive at Eq. (4.8) for the probability of observing the measured spike sequence, written solely in terms of effective parameters.

Appendix C. Outline of model-fitting algorithms

The algorithm for determining the parameters of the linear–nonlinear model (5.5) and the linear–quadratic–nonlinear model (5.6) is a modification of that presented in Ref. [8]. The main difference was that we used an error function nonlinearity (5.2) rather than an exponential nonlinearity.

If we fix the parameter A_q, we followed the algorithm in Ref. [8] modified for an error function nonlinearity in order to find the kernels $h^i_{1,q}$, $h^i_{2,q}$, and the parameters y_q, ϵ_q, and α_q. We then fix the kernels at these estimates but refine the estimates of y_q, ϵ_q. and α_q, obtaining maximum likelihood estimates of these parameters, conditioned on the fixed values of $h^i_{1,q}$, $h^i_{2,q}$, and A_q. (As we cannot show the maximum likelihood surface has no non-global local maxima, it is important that we start with good estimates using the procedure modified from Ref. [8].) We end up with values of all the parameters as a function of A_q. (For the linear–nonlinear model, we set $\alpha_s = 0$ and fit $h^i_q = h^i_{1,q}$, y_q, and ϵ_q in an analogous manner.)

Our last step is to attempt to find a maximum likelihood estimate of A_q, where the other parameters were a function of A_q as described above. We looked for a maximum on the interval $[0, A^0]$, where A^0 was 10 times the time bin size.

References

[1] D.R. Cox, V. Isham, Point Processes, Chapman & Hall, New York, 1980.
[2] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, Springer-Verlag, New York, 1988.
[3] D.Q. Nykamp, Revealing pairwise coupling in linear–nonlinear networks, SIAM J. Appl. Math. 65 (2005) 2005.
[4] L. Paninski, Maximum likelihood estimation of cascade point-process neural encoding models, Network: Comput. Neural Syst. 15 (2004) 243.
[5] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, F. Rossi, Gnu Scientific Library Reference Manual, second ed., Network Theory Ltd., Bristol, United Kingdom, 2003. Available from: <http://www.gnu.org/software/gsl/>.
[6] S. Marcelja, Mathematical description of the responses of simple cortical cells, J. Opt. Soc. Am. 70 (1980) 1297.
[7] E.H. Adelson, J.R. Bergen, Spatiotemporal energy models for the perception of motion, J. Opt. Soc. Am. A 2 (1985) 284.
[8] D.Q. Nykamp, Measuring linear and quadratic contributions to neuronal response, Network: Comput. Neural Syst. 14 (2003) 673.
[9] D.H. Perkel, G.L. Gerstein, G.P. Moore, Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains, Biophys. J. 7 (1967) 419.
[10] A.M.H.J. Aertsen, G.L. Gerstein, M.K. Habib, G. Palm, Dynamics of neuronal firing correlation: modulation of "effective connectivity", J. Neurophysiol. 61 (1989) 900.
[11] G. Palm, A.M.H.J. Aertsen, G.L. Gerstein, On the significance of correlations among neuronal spike trains, Biol. Cybern. 59 (1988) 1.
[12] D.Q. Nykamp, Reconstructing stimulus-driven neural networks from spike times, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Advances in Neural Information Processing Systems 15, MIT Press, Cambridge, MA, 2003, p. 309.
[13] P.N. Marmarelis, V.Z. Marmarelis, Analysis of Physiological Systems: The White Noise approach, Plenum Press, New York, 1978.
[14] E. DeBoer, P. Kuyper, Triggered correlation, IEEE Trans. Biomed. Eng. 15 (1968) 169.
[15] Y. Dan, J.-M. Alonso, W.M. Usrey, R.C. Reid, Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus, Nat. Neurosci. 1 (1998) 501.
[16] J.R. Rosenberg, A.M. Amjad, P. Breeze, D.R. Brillinger, D.M. Halliday, The Fourier approach to the identification of functional coupling between neuronal spike trains, Prog. Biophys. Mol. Biol. 53 (1989) 1.
[17] K.D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, G. Buzsáki, Organization of cell assemblies in the hippocampus, Nature 424 (2003) 552.
[18] M. Okatan, M.A. Wilson, E.N. Brown, Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity, Neural Comput. 17 (2005) 1927.
[19] E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, L.M. Frank, The time-rescaling theorem and its application to neural spike train data analysis, Neural Comput. 14 (2001) 325.
[20] D.Q. Nykamp, Spike correlation measures that eliminate stimulus effects in response to white noise, J. Comput. Neurosci. 14 (2003) 193.