

WHITE NOISE ANALYSIS OF COUPLED LINEAR-NONLINEAR SYSTEMS

DUANE Q. NYKAMP*

Abstract. We present an asymptotic analysis of two coupled linear-nonlinear systems. Through measuring first and second input-output statistics of the systems in response to white noise input, one can completely characterize the systems and their coupling. The proposed model is similar to a widely used phenomenological model of neurons in response to sensory stimulation and may be used to help characterize neural circuitry in sensory brain regions.

Key words. neural networks, correlations, Weiner analysis, white noise

AMS subject classification. 92C20

1. Introduction. Most electrophysiology data from intact mammalian brains is recorded using an extracellular electrode which remains outside neurons. When the electrode is positioned near a neuron, it can record the neuron's output events, called spikes, because spike magnitudes are sufficiently large. The internal state of a neuron, including small fluctuations in response to its inputs, cannot be measured.

When only output spikes are measurable, one cannot directly measure the effect of a connection from one neuron to another. If neuron 1 is connected to neuron 2, then an output spike of neuron 1 will perturb the internal state of neuron 2. If the internal state cannot be measured, this perturbation can be inferred only via its effect on the spike times of neuron 2. In general, the spike times of a neuron will be a function of many inputs coming from many other neurons. This complexity makes reliable inferences on the structure of neuron circuits from spike time data a formidable challenge.

Explicit mathematical models may lead to tools that can address this challenge. Through model analysis, one may develop methods to infer aspects of network structure from spike times, subject to the validity of the underlying model. In this paper, we derive a method to reconstruct the connectivity between two isolated neurons based on a simple linear-nonlinear model (see below) of neural response to white noise. Although this model greatly simplifies the reality of the brain's neural networks, the results from this analysis can be used to analyze neurophysiology data provided that they are interpreted within the limitations of the model [12].

Numerous researchers have used white noise analysis to describe the response of neurons to a stimulus. The most common use of white noise analysis has been to analyze the response properties of single neurons [11, 4, 5, 8, 9, 2, 3, 16, 18, 7, 6]. Recently, researchers have begun to apply the techniques of white noise analysis to simultaneous measurements of multiple neurons [15, 1, 19], although without explicitly modeling neural connectivity. In Ref. [12], we showed how, in white noise experiments, interpretation of spike time data is especially difficult because standard correlation measures confound stimulus and connectivity effects. We demonstrated correlation measures that remove the stimulus effects based on the linear-nonlinear model.

In this paper, we present the asymptotic analysis of the linear-nonlinear model that underlies the correlation measures of Ref. [12]. Subject to a first order approxi-

*Department of Mathematics, University of California, Los Angeles, CA 90095 (nykamp@math.ucla.edu). Supported by a NSF Mathematical Sciences Postdoctoral Research Fellowship.

mation in the coupling magnitude, we derive a method to completely reconstruct the coupled system from first and second input-output statistics. As a consequence of this reconstruction, we obtain a correlation measure, which we call \mathcal{W} , that approximates the neuronal coupling.

Although the analysis below can be used for any pair of coupled linear-nonlinear systems, we refer to the systems as *neurons* since neuroscience was the motivation of this analysis and this choice simplifies the description.

In section 2, we describe the linear-nonlinear model. In section 3, we derive expressions for the input-output statistics for the case when the neurons are uncoupled. We derive the corresponding expressions for the cases of unidirectional coupling in section 4 and generalize the results to mutual coupling in section 5. We demonstrate the method with simulations in section 6 and discuss the results in section 7.

2. The model. The standard model underlying most white noise analyses of neural function is the linear-nonlinear model of neural response to an input \mathbf{X} ,

$$\Pr(R^i = 1 | \mathbf{X} = \mathbf{x}) = g(\mathbf{h}^i \cdot \mathbf{x}), \quad (2.1)$$

where the response R^i at discrete time point i is one if the neuron spiked and zero otherwise. The neural response depends on the convolution of the kernel \mathbf{h} with the stimulus. The stimulus \mathbf{X} is a vector whose components represent the spatio-temporal sequence of stimulus values, such as the sequence of pixel values for each refresh of a computer monitor.

The neural response depends on the convolution of stimulus with a kernel \mathbf{h} , normalized so that $|\mathbf{h}| = 1$. The kernel can be viewed as sliding along stimulus with time; it represents the spatio-temporal stimulus features to which the neuron responds. We let \mathbf{h}^i denote the kernel shifted for time point i and write the convolution of the kernel with the stimulus as the dot product $\mathbf{h}^i \cdot \mathbf{X}$ (implicitly viewing the temporal index of the stimulus as going backward in time). The function $g(\cdot)$ is the neuron's output non-linearity (representing, for example, its spike generating mechanism). Although the linear-nonlinear system is only a phenomenological approximation of complex biology, it can be simply characterized by standard white noise analysis [13]. The ease of an explicit mathematical analysis is a prime motivation for choosing the linear-nonlinear model and white noise input.

We propose a model that augments the linear-nonlinear framework to include the effects of neural connections between two neurons. After neuron q spikes, the probability that neuron p spikes j time steps later is modified by the connectivity factor \bar{W}_{qp}^j . In a caricature of synaptic input acting at subthreshold levels (of the voltage, or internal state, of a neuron), the term \bar{W}_{qp}^j is added underneath the nonlinearity so that

$$\Pr(R_p^i = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_q = \mathbf{r}_q) = g_p\left(\mathbf{h}_p^i \cdot \mathbf{x} + \sum_{j \geq 0} \bar{W}_{qp}^j r_q^{i-j}\right) \quad (2.2)$$

where $p, q \in \{1, 2\}$ represent the index of the neurons, $q \neq p$, and $R_p^i \in \{0, 1\}$ is the response of neuron p at time i .¹

¹With the exceptions of W and T , we will use capital variables to denote random quantities. In addition, we will use subscripts to denote neuron index and superscripts to denote temporal indices.

We assume the output nonlinearity can be approximated as an error function

$$g_p(s) = \frac{\hat{r}_p}{2} \left[1 + \operatorname{erf} \left(\frac{s - \bar{T}_p}{\epsilon_p \sqrt{2}} \right) \right], \quad (2.3)$$

where \hat{r}_p is the maximum firing rate, \bar{T}_p is the threshold, ϵ_p defines the steepness of the nonlinearity, and $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$. Note that $\lim_{x \rightarrow \infty} g_p(x) = \hat{r}_p$ and $\lim_{x \rightarrow -\infty} g_p(x) = 0$. The error function nonlinearity is assumed so that we can derive analytic results. As demonstrated in section 6, the results apply to more general nonlinearities.

So that the input \mathbf{X} is a discrete approximation of temporal or spatio-temporal white noise, we let each of its n components be standard normal random variables. We do not explicitly distinguish spatial versus temporal components of the input in our notation because they are treated identically in the analysis. To keep the notation simple, time is represented only by the temporal index of the kernels \mathbf{h}_p^i and the spikes R_p^i . With this convention, the probability density function of \mathbf{X} is simply

$$\rho_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} e^{-|\mathbf{x}|^2/2}. \quad (2.4)$$

In the next sections, we consider special cases of the coupling \bar{W} . For each case, we calculate the expected values of the responses $E\{R_p^i\}$, the ‘‘correlation’’² between the stimulus and the spikes of each neuron $E\{\mathbf{X}R_p^i\}$, and the ‘‘correlation’’ between the spikes of the two neurons $E\{R_1^i R_2^{i-k}\}$. Since these statistics can be estimated when one can obtain only the spike times from the neurons, they are readily measurable in neurophysiology experiments. We base our reconstruction of the linear-nonlinear system of Eq. (2.2) on these input-output statistics. Most importantly, we will reconstruct the coupling terms \bar{W}_{pq}^j .

3. Uncoupled neurons. In this section, we assume that the neurons are uncoupled so that the responses of neurons are independent conditioned on the input.³ In this case, the response probabilities obey Eqs. (2.2) and (2.3) with $\bar{W}_{pq}^j = 0$.

The analysis of the single neuron statistics reduces to the case of individual neurons. As detailed in Ref. [13], the first two input-output statistics are given by

$$E\{R_p^i\} = \frac{\hat{r}_p}{2} \operatorname{erfc} \left(\frac{\delta_p \bar{T}_p}{\sqrt{2}} \right) \quad (3.1)$$

and

$$E\{\mathbf{X}R_p^i\} = \frac{\delta_p}{\sqrt{2\pi}} e^{-\frac{\delta_p^2 \bar{T}_p^2}{2}} \mathbf{h}_p^i, \quad (3.2)$$

where

$$\delta_p = \frac{1}{\sqrt{1 + \epsilon_p^2}}, \quad (3.3)$$

²We recognize that the statistics $E\{\mathbf{X}R_p^i\}$ and $E\{R_1^i R_2^{i-k}\}$ are not actually correlations. We use the term since these statistics are consistently called correlations in the neuroscience literature. We hope the reader will forgive our loose use of the term. The stimulus-spike correlation $E\{\mathbf{X}R_p^i\}$ can be thought of as the average stimulus that precedes each spike of neuron p .

³Meaning $\Pr(R_1^i = 1 \& R_2^j = 1 | \mathbf{X}) = \Pr(R_1^i = 1 | \mathbf{X}) \Pr(R_2^j = 1 | \mathbf{X})$.

and $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$. Note that, since both the input and the system are stationary, the results are independent of time index i (except for the temporal index of the linear kernel). Assuming one knew \hat{r}_p , the nonlinearities $g_p(\cdot)$ could be computed by estimating ϵ_p and \bar{T}_p from these statistics [13]. One could also obtain the unit vectors \mathbf{h}_p^i from $E\{\mathbf{X}R_p^i\}/|E\{\mathbf{X}R_p^i\}|$. In this simple case, one does not even need to measure $E\{R_1^i R_2^{i-k}\}$ to reconstruct the system.

Before we calculate an expression for $E\{R_1^i R_2^{i-k}\}$, we define the angles between the linear kernels, which turn out to be the only important geometry of the kernels for white noise input. Let $\bar{\theta}_{pq}^k$ be the angle between kernel q and kernel p shifted k units in time

$$\cos \bar{\theta}_{pq}^k = \mathbf{h}_p^{i-k} \cdot \mathbf{h}_q^i. \quad (3.4)$$

This angle is of course independent of time index i . (The inner product can be represented as a cosine because kernels were normalized to be unit vectors.) Note that $\bar{\theta}_{qp}^{-k} = \bar{\theta}_{pq}^k$. We always define the corresponding sine is by $\sin \theta = \sqrt{1 - \cos^2 \theta}$ so that $\sin \theta \geq 0$.

For a given time shift k , without loss of generality, assume \mathbf{h}_1^i is the first unit vector \mathbf{e}_1 (in stimulus space) and \mathbf{h}_2^{i-k} is a linear combination of the first two unit vectors:⁴

$$\begin{aligned} \mathbf{h}_1^i &= \mathbf{e}_1 \\ \mathbf{h}_2^{i-k} &= \mathbf{e}_1 \cos \bar{\theta}_{21}^k + \mathbf{e}_2 \sin \bar{\theta}_{21}^k. \end{aligned}$$

Assuming the nonlinearities satisfy

$$\lim_{x \rightarrow -\infty} g_p(x) = 0, \quad (3.5)$$

we compute the correlation between the spikes of neuron 1 and the spikes of neuron 2, by changing variables and integrating by parts twice. In each integration by parts, one boundary term disappears due to Eq. (3.5), and the other boundary term is incorporated into the complementary error functions:

$$\begin{aligned} E\{R_1^i R_2^{i-k}\} &= \frac{1}{2\pi} \int g_1(x_1) g_2(x_1 \cos \bar{\theta}_{21}^k + x_2 \sin \bar{\theta}_{21}^k) e^{-\frac{x_1^2 + x_2^2}{2}} dx_1 dx_2 \\ &= \frac{1}{2\pi} \int g_1(u) g_2(v) \exp\left(-\frac{u^2}{2} - \frac{(v - u \cos \bar{\theta}_{21}^k)^2}{2 \sin^2 \bar{\theta}_{21}^k}\right) \frac{du dv}{\sin \bar{\theta}_{21}^k} \\ &= \frac{1}{4} \int g_1'(u) g_2'(v) \operatorname{derfc}\left(\frac{u}{\sqrt{2}}, \frac{v}{\sqrt{2}}, \cos \bar{\theta}_{21}^k\right) du dv \end{aligned} \quad (3.6)$$

where we have defined a double complementary error function

$$\operatorname{derfc}(a, b, c) = \frac{4}{\pi} \int_a^\infty dy e^{-y^2} \int_{\frac{b-cy}{\sqrt{1-c^2}}}^\infty dz e^{-z^2}. \quad (3.7)$$

The function derfc is a two dimensional analogue of the complementary error function. The integral is taken over the intersection of the two half planes $\mathbf{x} \cdot \mathbf{u} > a$

⁴Since stimulus is rotationally invariant, we can rotate axis so that \mathbf{h}_1^i is parallel to the first axis and \mathbf{h}_2^{i-k} lies in the span of the first two axes. Recall that $|\mathbf{h}_p^i| = 1$.

and $\mathbf{x} \cdot \mathbf{v} > b$, where \mathbf{u} and \mathbf{v} are two unit vectors with inner product $\mathbf{u} \cdot \mathbf{v} = c$. (Here, \mathbf{x} represents a generic vector.) Note that $\text{derfc}(a, b, 0) = \text{erfc}(a)\text{erfc}(b)$ and $\text{derfc}(a, b, c) = \text{derfc}(b, a, c)$.

When the nonlinearity is an error function (Eq. (2.3)), we substitute into Eq. (3.6) and use formula (B.8) to obtain

$$E\{R_1^i R_2^{i-k}\} = \frac{\hat{r}_1 \hat{r}_2}{4} \text{derfc}\left(\frac{\delta_1 \bar{T}_1}{\sqrt{2}}, \frac{\delta_2 \bar{T}_2}{\sqrt{2}}, \delta_1 \delta_2 \cos \bar{\theta}_{21}^k\right). \quad (3.8)$$

Eqs. (3.1), (3.2), and (3.8) are the expressions for the input-output statistics for the simple case of uncoupled neurons.

4. Unidirectional coupling. Let the coupling from neuron 2 to neuron 1 (\bar{W}_{21}^j) be nonzero, but keep $\bar{W}_{12}^j = 0$. Then the probability of a spike of neuron 1 at time k is dependent not only on the input but also on the spikes of neuron 2 for times before k , as given by Eq. (2.2). The probability of neuron 2 spiking remains the same as in section 3 so that the input-output statistics $E\{R_2^i\}$ and $E\{\mathbf{X}R_2^i\}$ are unchanged.

In what follows, we calculate expressions for the remaining input-output statistics. We first show that effective parameters of the system can be calculated from $E\{R_p^i\}$ and $E\{\mathbf{X}R_p^i\}$. We next show how the coupling \bar{W}_{21}^j can be calculated from $E\{R_1^i R_2^{i-k}\}$.

We assume that \bar{W}_{21}^j is small, and compute a first order approximation by dropping terms that are second order or higher in \bar{W}_{21}^j . Since from now on all equalities will be within $O(\bar{W}^2)$, we will, for simplicity, use $=$ to mean equal within $O(\bar{W}^2)$.

4.1. Mean rate of neuron 1. In this section, we show that the mean rate of neuron 1 is nearly identical to the uncoupled case of Eq. (3.1), only with the original threshold \bar{T}_1 replaced with an effective threshold T_1 to be defined below.

The general expression for the mean rate of neuron 1, calculated in Appendix A.3, is

$$E\{R_1^i\} = \frac{1}{\sqrt{2\pi}} \int g_1(u) e^{-\frac{u^2}{2}} du + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2\sqrt{2\pi}} \int g_1(u) g_2(v) e^{-\frac{u^2}{2}} \text{erfc}\left(\frac{v - u \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du dv. \quad (4.1)$$

Note that the mean rate $E\{R_1^i\}$ is independent of i (as it must be).

When the nonlinearities are error functions (Eq. (2.3)), the first term is the uncoupled mean rate (Eq. (3.1)). We use formula (B.5), to simplify the \bar{W}_{21}^j term so that the mean rate of neuron 1 is

$$E\{R_1^i\} = \frac{\hat{r}_1}{2} \text{erfc}\left(\frac{\delta_1 \bar{T}_1}{\sqrt{2}}\right) + \frac{\hat{r}_1 \hat{r}_2 \delta_1}{2\sqrt{2\pi}} e^{-\frac{\delta_1^2 \bar{T}_1^2}{2}} \sum_{j \geq 0} \bar{W}_{21}^j \text{erfc}\left(\frac{\delta_2 \bar{T}_2 - \delta_1^2 \delta_2 \bar{T}_1 \cos \bar{\theta}_{21}^j}{\sqrt{1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j}}\right)$$

Using the Taylor series for $\text{erfc}\left(\frac{\delta_1 \bar{T}_1 + x}{\sqrt{2}}\right)$, we pull the second term into the error function (making only a $O(\bar{W}^2)$ error), obtaining

$$E\{R_1^i\} = \frac{\hat{r}_1}{2} \text{erfc}\left(\frac{\delta_1 T_1}{\sqrt{2}}\right), \quad (4.2)$$

where we let $T_2 = \bar{T}_2$ and have defined the effective threshold for neuron 1:

$$T_1 = \bar{T}_1 - \sum_{j \geq 0} \frac{\hat{r}_2 \bar{W}_{21}^j}{2} \operatorname{erfc} \left(\frac{\delta_2 T_2 - \delta_1^2 \delta_2 \bar{T}_1 \cos \bar{\theta}_{21}^j}{\sqrt{2(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}} \right). \quad (4.3)$$

The mean rate of neuron 1 is identical to that of an uncoupled neuron with the effective threshold T_1 .

4.2. Correlation of spikes of neuron 1 with stimulus. We calculate the general expression for the correlation between the spikes of neuron 1 with the stimulus in Appendix A.4, obtaining

$$\begin{aligned} E\{\mathbf{X}R_1^i\} &= \frac{1}{\sqrt{2\pi}} \left[\int g'_1(u) e^{-\frac{u^2}{2}} du \right. \\ &\quad \left. + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2} \int g'_1(u) g'_2(v) u e^{-\frac{u^2}{2}} \operatorname{erfc} \left(\frac{v - u \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k} \right) du dv \right] \mathbf{h}_1^i \\ &\quad + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2\pi} \int g'_1(u) g'_2(v) \exp \left(-\frac{u^2 - 2 \cos \bar{\theta}_{21}^j uv + v^2}{2 \sin^2 \bar{\theta}_{21}^j} \right) du dv \mathbf{h}_{21}^{\perp ji}, \end{aligned} \quad (4.4)$$

where we define $\mathbf{h}_{21}^{\perp ji}$ as the component of \mathbf{h}_2^{i-j} that is perpendicular to \mathbf{h}_1^i ,

$$\mathbf{h}_{21}^{\perp ji} = \frac{\mathbf{h}_2^{i-j} - \cos \bar{\theta}_{21}^j \mathbf{h}_1^i}{\sin \bar{\theta}_{21}^j}. \quad (4.5)$$

Because of the coupling, $E\{\mathbf{X}R_1^i\}$ is no longer parallel to the linear kernel \mathbf{h}_1^i . Each term in the last sum of Eq. (4.4) indicates how the coupling \bar{W}_{21}^j leads to a component of $E\{\mathbf{X}R_1^i\}$ that is perpendicular to \mathbf{h}_1^i .

When the nonlinearities are error functions (Eq. (2.3)), we use Eq. (4.5) and formulas (B.1), (B.6), and (B.2) to obtain the following expression for the correlation between the stimulus and the spikes of neuron 1

$$\begin{aligned} E\{\mathbf{X}R_1^i\} &= \mu_1^0 \left[1 - \sum_{j \geq 0} \frac{\hat{r}_2 \bar{W}_{21}^j \delta_1^2 \delta_2 \cos \bar{\theta}_{21}^j}{\sqrt{2\pi(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}} \exp \left(-\frac{[\delta_2 T_2 - \delta_1^2 \delta_2 T_1 \cos \bar{\theta}_{21}^j]^2}{2(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)} \right) \right] \mathbf{h}_1^i \\ &\quad + \mu_1^0 \sum_{j \geq 0} \frac{\hat{r}_2 \bar{W}_{21}^j \delta_2}{\sqrt{2\pi(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}} \exp \left(-\frac{[\delta_2 T_2 - \delta_1^2 \delta_2 T_1 \cos \bar{\theta}_{21}^j]^2}{2(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)} \right) \mathbf{h}_2^{i-j}, \end{aligned} \quad (4.6)$$

where

$$\mu_p^0 = \frac{\hat{r}_p \delta_p}{\sqrt{2\pi}} e^{-\frac{\delta_p^2 T_p^2}{2}}. \quad (4.7)$$

One key to obtaining Eq. (4.6) was using the exponential's Taylor series to bring the effective threshold T_1 (4.3) into the exponential of the first term. We let $T_2 = \bar{T}_2$ and simply replaced \bar{T}_1 with T_1 in all other terms (making an $O(\bar{W}^2)$ error).

4.3. Reconstruction from the mean rate and stimulus-spike correlations. Eqs. (4.2) and (4.6) give expressions for the first two input-output statistics of the linear-nonlinear system (2.2) with unidirectional coupling. These equations show

that the coupling has both changed the effective threshold and altered the direction of $E\{\mathbf{X}R_1^i\}$ so that it is no longer parallel to the kernel \mathbf{h}_1^i .

Because of these modifications, we can no longer recover \bar{T}_1 or \mathbf{h}_1^i (or $\cos\bar{\theta}_{21}^j$) from $E\{R_1^i\}$ and $E\{\mathbf{X}R_1^i\}$ as outlined in section 3. Nonetheless, subject to one more assumption, one can recover the effective threshold T_1 , the original δ_1 , and an effective angle between the kernels. As shown below, one simply views the neurons as uncoupled and reconstructs the neuron parameters as in section 3. This procedure does not use $E\{R_1^i R_2^{i-k}\}$. We will be able use this last statistic to determine the coupling \bar{W}_{21}^j .

4.3.1. Effective angle between kernels. When the neurons were uncoupled, the linear kernel \mathbf{h}_1^i could be determined by the normalized stimulus-spike correlation $E\{\mathbf{X}R_1^i\}/|E\{\mathbf{X}R_1^i\}|$. Although this measurement no longer yields the kernel, we can treat it as an effective kernel and define the effective angle between kernels by

$$\cos\theta_{pq}^k = \frac{E\{\mathbf{X}R_p^{i-k}\}}{|E\{\mathbf{X}R_p^{i-k}\}|} \cdot \frac{E\{\mathbf{X}R_q^i\}}{|E\{\mathbf{X}R_q^i\}|}. \quad (4.8)$$

In this case of unidirectional coupling, neuron 2 is unaffected, and the effective angle between neuron 1 and 2 is $\cos\theta_{21}^k = \mathbf{h}_2^{i-k} \cdot E\{\mathbf{X}R_1^i\}/|E\{\mathbf{X}R_1^i\}|$.

We rewrite Eqs. (4.2) and (4.6) in terms of the measurable effective angle as follows. The magnitude of the stimulus-spike correlation, within $O(\bar{W}^2)$, is

$$|E\{\mathbf{X}R_1^i\}| = \mu_1^0 \left[1 + \sum_{j \geq 0} \frac{\hat{r}_2 \bar{W}_{21}^j (1 - \delta_1^2) \delta_2 \cos\bar{\theta}_{21}^j}{\sqrt{2\pi(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}} \exp\left(-\frac{[\delta_2 T_2 - \delta_1^2 \delta_2 T_1 \cos\bar{\theta}_{21}^j]^2}{2(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}\right) \right] \quad (4.9)$$

so that

$$\cos\theta_{21}^k = \cos\bar{\theta}_{21}^k + \sum_{j \geq 0} \frac{\hat{r}_2 \bar{W}_{21}^j \delta_2 (\cos\theta_{22}^{k-j} - \cos\bar{\theta}_{21}^j \cos\bar{\theta}_{21}^k)}{\sqrt{2\pi(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}} \exp\left(-\frac{[\delta_2 T_2 - \delta_1^2 \delta_2 T_1 \cos\bar{\theta}_{21}^j]^2}{2(1 - \delta_1^2 \delta_2^2 \cos^2 \bar{\theta}_{21}^j)}\right).$$

Since $\cos\theta_{21}^k$ is within $O(\bar{W})$ of $\cos\bar{\theta}_{21}^k$, we can replace $\cos\bar{\theta}_{21}^k$ by $\cos\theta_{21}^k$ in the last terms (making only an $O(\bar{W}^2)$ error), and write $\cos\bar{\theta}_{21}^k$ in terms of $\cos\theta_{21}^k$:

$$\cos\bar{\theta}_{21}^k = \cos\theta_{21}^k - \sum_{j \geq 0} \bar{W}_{21}^j C_{21}^{jk}, \quad (4.10)$$

where

$$C_{pq}^{jk} = (\cos\theta_{pp}^{k-j} - \cos\theta_{pq}^j \cos\theta_{pq}^k) \mu_{pq}^j, \quad (4.11)$$

$$\mu_{pq}^k = \frac{\hat{r}_p \delta_p \exp(-\frac{1}{2}[\lambda_{pq}^k]^2)}{\sqrt{2\pi(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k)}}, \quad (4.12)$$

and

$$\lambda_{pq}^k = \frac{\delta_p T_p - \delta_p \delta_q^2 T_q \cos\theta_{pq}^k}{\sqrt{1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k}}. \quad (4.13)$$

Note that $\mu_p^0 \mu_{pq}^k = \mu_q^0 \mu_{qp}^{-k}$.

We now make an $O(\bar{W}^2)$ error by replacing $\cos \bar{\theta}_{21}^k$ with $\cos \theta_{21}^k$ in the stimulus-spike correlation,

$$E\{\mathbf{X}R_1^i\} = \mu_1^0 \left[\left(1 - \sum_{j \geq 0} \bar{W}_{21}^j \delta_1^2 \cos \theta_{21}^j \mu_{21}^j \right) \mathbf{h}_1^i + \sum_{j \geq 0} \bar{W}_{21}^j \mu_{21}^j \mathbf{h}_2^{i-j} \right]. \quad (4.14)$$

and in the expression for the effective threshold (4.3),

$$\bar{T}_1 = T_1 + \sum_{j \geq 0} \bar{W}_{21}^j \eta_{21}^j \quad (4.15)$$

where

$$\eta_{pq}^k = \frac{\hat{r}_p}{2} \operatorname{erfc} \left(\frac{\lambda_{pq}^k}{\sqrt{2}} \right). \quad (4.16)$$

4.3.2. Effective nonlinearity parameters. As shown above, $\cos \theta_{21}^k$, not the original $\cos \bar{\theta}_{21}^k$, is the measurable inner product between the kernels. We next show that, with one additional mild assumption, the parameters T_1 and δ_1 are the nonlinearity parameters measured from $E\{\mathbf{X}R_1^i\}$ and $E\{R_1^i\}$ when treating neuron 1 as an independent neuron as in section 3.

The magnitude of $E\{\mathbf{X}R_1^i\}$ is

$$|E\{\mathbf{X}R_1^i\}| = \mu_1^0 \left(1 + \sum_{j \geq 0} \bar{W}_{21}^j (1 - \delta_1^2) \cos \theta_{21}^j \mu_{21}^j \right). \quad (4.17)$$

This expression simplifies to μ_1^0 if we assume that $\bar{W}_{21}^j \delta_1 \delta_2 (1 - \delta_1^2) \cos \theta_{21}^j$ is small enough to ignore. Since we have already assumed that \bar{W}_{21}^j is small, we simply need to assume that $\delta_2 (1 - \delta_1^2) \cos \theta_{21}^j$ is small to have an expression that is second order in a small parameter. This expression is the product of three factors less than one. It will be small if the nonlinearities are not very sharp or if the kernels of the neurons are not nearly aligned.

With this approximation, the stimulus-spike correlation is

$$|E\{\mathbf{X}R_1^i\}| \approx \mu_1^0 = \frac{\hat{r}_1 \delta_1}{\sqrt{2\pi}} e^{-\frac{\delta_1^2 T_1^2}{2}}. \quad (4.18)$$

Recall that the mean rate of neuron 1 (Eq. (4.2)) is

$$E\{R_1^i\} = \frac{\hat{r}_1}{2} \operatorname{erfc} \left(\frac{\delta_1 T_1}{\sqrt{2}} \right).$$

These results are the same as (3.1) and (3.2) for an uncoupled neuron with nonlinearity parameters δ_1 and T_1 . One can determine δ_1 and T_1 from these equations (assuming \hat{r}_1 is known).

Using only $E\{R_p^i\}$ and $E\{\mathbf{X}R_p^i\}$ in this manner, one can calculate effective nonlinearity parameters of both neuron 1 and neuron 2, as well as the effective angle between the linear kernels. We next show how the connectivity \bar{W}_{21} can be determined from the remaining input-output statistic $E\{R_1^i R_2^{i-k}\}$.

4.4. Correlation between spikes of neuron 1 and 2. We calculate the general expression for the correlation between the spikes of neuron 1 and 2 in Appendix A.5, obtaining the complicated expression

$$\begin{aligned}
E\{R_1^i R_2^{i-k}\} &= \frac{1}{4} \int g'_1(u_1) g'_2(u_2) \operatorname{derfc}\left(\frac{u_1}{\sqrt{2}}, \frac{u_2}{\sqrt{2}}, \cos \bar{\theta}_{21}^k\right) du_1 du_2 \\
&+ \frac{\bar{W}_{21}^k}{2\sqrt{2\pi}} \int g'_1(u_1) g'_2(u_2) e^{-\frac{u_1^2}{2}} \operatorname{erfc}\left(\frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du_1 du_2 \\
&+ \sum_{j \geq 0, j \neq k} \frac{\bar{W}_{21}^j}{4\sqrt{2\pi}} \int du_1 du_2 du_3 g'_1(u_1) g'_2(u_2) g'_2(u_3) e^{-\frac{u_1^2}{2}} \\
&\times \operatorname{derfc}\left(\frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}, \frac{u_3 - u_1 \cos \bar{\theta}_{21}^j}{\sqrt{2} \sin \bar{\theta}_{21}^j}, \frac{\cos \theta_{22}^{k-j} - \cos \bar{\theta}_{21}^j \cos \bar{\theta}_{21}^k}{\sin \bar{\theta}_{21}^j \sin \bar{\theta}_{21}^k}\right).
\end{aligned} \tag{4.19}$$

When the nonlinearities are error functions (Eq. (2.3)), we simplify this expression using three formulas ((B.8), (B.9), and (B.5)) and Eqs. (4.10), (4.15), (4.7), (4.16), and (4.11). We use the following Taylor series expansions of $\operatorname{derfc}(a, b, c)$,

$$\begin{aligned}
\operatorname{derfc}(a+x, b, c) &= \operatorname{derfc}(a, b, c) - \frac{2x}{\sqrt{\pi}} e^{-a^2} \operatorname{erfc}\left(\frac{b-ca}{\sqrt{1-c^2}}\right) + O(x^2) \\
\operatorname{derfc}(a, b, c+x) &= \operatorname{derfc}(a, b, c) + \frac{2x}{\pi\sqrt{1-c^2}} e^{-\frac{a^2-2abc+b^2}{1-c^2}} + O(x^2),
\end{aligned}$$

to pull terms for the effective threshold T_1 and effective kernel inner product $\cos \theta_{21}^j$ into the first term. All other terms are $O(\bar{W})$, so we can simply drop the bars from \bar{T}_1 and $\cos \bar{\theta}_{21}^j$.

Defining

$$\nu_{pq}^k = \frac{\hat{r}_p \hat{r}_q}{4} \operatorname{derfc}\left(\frac{\delta_p T_p}{\sqrt{2}}, \frac{\delta_q T_q}{\sqrt{2}}, \delta_p \delta_q \cos \theta_{pq}^k\right), \tag{4.20}$$

$$\tilde{\nu}_{pq}^{kj} = \begin{cases} \eta_{pq}^k & \text{for } j = k, \\ \frac{(\hat{r}_p)^2}{4} \operatorname{derfc}\left(\frac{\lambda_{pq}^k}{\sqrt{2}}, \frac{\lambda_{pq}^j}{\sqrt{2}}, \xi_{pq}^{kj}\right) & \text{otherwise,} \end{cases} \tag{4.21}$$

and

$$\xi_{pq}^{kj} = \frac{\delta_p^2 \cos \theta_{pp}^{k-j} - \delta_p^2 \delta_q^2 \cos \theta_{pq}^j \cos \theta_{pq}^k}{\sqrt{(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^j)(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k)}}, \tag{4.22}$$

the correlation between the spikes of neuron 1 and 2 becomes

$$E\{R_1^i R_2^{i-k}\} = \nu_{21}^k + \sum_{j \geq 0} A_{21}^{kj} \bar{W}_{21}^j \tag{4.23}$$

where

$$A_{pq}^{kj} = \mu_q^0 [\tilde{\nu}_{pq}^{kj} - \eta_{pq}^k \eta_{pq}^j + (\cos \theta_{pq}^k \cos \theta_{pq}^j - \cos \theta_{pp}^{k-j}) \mu_{pq}^k \mu_{pq}^j], \tag{4.24}$$

and μ_p^0 , μ_{pq}^k , λ_{pq}^k , and η_{pq}^k are defined in Eqs. (4.7), (4.12), (4.13), and (4.16) respectively. The term ν_{21}^k in Eq. (4.23) is analogous to the correlation observed in the uncoupled case (Eq. (3.8)), and the sum represents additional correlations due to the coupling terms \bar{W}_{21}^j . A discussion of some properties of A_{pq}^{kj} is given in the next section.

The important fact to note about Eq. (4.23) is that, with the exception of the \bar{W}_{21}^j , every factor on the right hand side can be calculated from the mean rates $E\{R_p^i\}$ and stimulus-spike correlations $E\{\mathbf{X}R_p^i\}$. Eq. (4.23) can then be solved to determine the \bar{W}_{21}^j .

5. Mutual Coupling. Let both \bar{W}_{21}^j and \bar{W}_{12}^j be nonzero so that the neurons are mutually coupled. Then the probability of a spike of neuron p at time k depends not only on the input but also on the spikes of neuron q for times before k , as given by Eq. (2.2).

Since we assume that \bar{W}_{pq}^j is small and compute a first order approximation, the mutual interaction results are identical to the unidirectional results of section 4 applied in both directions. The effect of neuron p on neuron q is $O(\bar{W})$, so the effect of neuron p on itself through neuron q is $O(\bar{W}^2)$ and can be ignored. We can ignore second (and higher) order interactions.

The mutual coupling case involves no more work beyond the unidirectional case. The statistics for neuron 1 are unchanged, and the statistics for neuron 2 become analogous to those for neuron 1. The expression for $E\{R_1^i R_2^{i-k}\}$ simply adds a sum in terms of the \bar{W}_{12}^j coupling.

We summarize the model and resulting equations. We are given the system

$$\Pr(R_p^i = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_q = \mathbf{r}_q) = g_p\left(\mathbf{h}_p^i \cdot \mathbf{x} + \sum_{j \geq 0} \bar{W}_{qp}^j r_q^{i-j}\right) \quad (5.1)$$

for $p, q \in \{1, 2\}$, $q \neq p$, with

$$g_p(x) = \frac{\hat{r}_p}{2} \left[1 + \operatorname{erf}\left(\frac{x - \bar{T}_p}{\epsilon_p \sqrt{2}}\right) \right]. \quad (5.2)$$

We assume we know \hat{r}_p . We can reconstruct the system from the following input-output statistics: $E\{R_p^i\}$, $E\{\mathbf{X}R_p^i\}$, and $E\{R_1^i R_2^{i-k}\}$.

First, we calculate $\delta_p = 1/\sqrt{1 + \epsilon_p^2}$ and an effective threshold T_p from $E\{R_p^i\}$ and $|E\{\mathbf{X}R_p^i\}|$ using the equations⁵

$$E\{R_p^i\} = \frac{\hat{r}_p}{2} \operatorname{erfc}\left(\frac{\delta_p T_p}{\sqrt{2}}\right) \quad (5.3)$$

and

$$|E\{\mathbf{X}R_p^i\}| \approx \mu_p^0 = \frac{\hat{r}_p \delta_p}{\sqrt{2\pi}} \exp\left(-\frac{\delta_p^2 T_p^2}{2}\right). \quad (5.4)$$

Then, we calculate the effective angle between the kernels by

$$\cos \theta_{pq}^k = \frac{E\{\mathbf{X}R_p^{i-k}\} \cdot E\{\mathbf{X}R_q^i\}}{|E\{\mathbf{X}R_p^{i-k}\}| |E\{\mathbf{X}R_q^i\}|}. \quad (5.5)$$

⁵The fact that T_p and $E\{\mathbf{X}R_p^i\}$ are given by equations analogous to Eqs. (4.15) and (4.14) is not needed for the reconstruction.

The last step is to calculate the coupling \bar{W} from the spike correlations with delays $k = -N, \dots, N$,

$$E\{R_1^i R_2^{i-k}\} = \nu_{21}^k + \sum_{j \geq 0} A_{21}^{kj} \bar{W}_{21}^j + \sum_{j \geq 0} A_{12}^{-kj} \bar{W}_{12}^j \quad (5.6)$$

where

$$\begin{aligned} A_{pq}^{kj} &= \mu_q^0 [\tilde{\nu}_{pq}^{kj} - \eta_{pq}^k \eta_{pq}^j + (\cos \theta_{pq}^k \cos \theta_{pq}^j - \cos \theta_{pp}^{k-j}) \mu_{pq}^k \mu_{pq}^j], \\ \nu_{pq}^k &= \frac{\hat{r}_p \hat{r}_q}{4} \operatorname{derfc} \left(\frac{\delta_p T_p}{\sqrt{2}}, \frac{\delta_q T_q}{\sqrt{2}}, \delta_p \delta_q \cos \theta_{pq}^k \right), \\ \tilde{\nu}_{pq}^{kj} &= \begin{cases} \eta_{pq}^k & \text{for } j = k, \\ \frac{(\hat{r}_p)^2}{4} \operatorname{derfc} \left(\frac{\lambda_{pq}^k}{\sqrt{2}}, \frac{\lambda_{pq}^j}{\sqrt{2}}, \xi_{pq}^{kj} \right) & \text{otherwise,} \end{cases} \\ \eta_{pq}^k &= \frac{\hat{r}_p}{2} \operatorname{erfc}(\lambda_{pq}^k / \sqrt{2}), \\ \mu_{pq}^k &= \frac{\hat{r}_p \delta_p \exp(-\frac{1}{2}[\lambda_{pq}^k]^2)}{\sqrt{2\pi(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k)}}, \\ \lambda_{pq}^k &= \frac{\delta_p T_p - \delta_p \delta_q^2 T_q \cos \theta_{pq}^k}{\sqrt{1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k}}, \end{aligned}$$

and

$$\xi_{pq}^{kj} = \frac{\delta_p^2 \cos \theta_{pp}^{k-j} - \delta_p^2 \delta_q^2 \cos \theta_{pq}^j \cos \theta_{pq}^k}{\sqrt{(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^j)(1 - \delta_p^2 \delta_q^2 \cos^2 \theta_{pq}^k)}}.$$

We assume that we have chosen the number of delays (given by $k = -N, \dots, N$) so that \bar{W}_{21}^j and \bar{W}_{12}^j for $j = 0, \dots, N$ are all the nonzero connectivity terms of the system. Unfortunately, even though the \bar{W} are the only unknowns left in the system (5.6), we still have $2N + 2$ unknowns with only $2N + 1$ equations.

To reduce the number of unknowns, we simply do not attempt to distinguish \bar{W}_{21}^0 from \bar{W}_{12}^0 . Although there is no reason these should be identical, the best we can do is calculate their sum. To solve the equations, we define a new \bar{W}^j by

$$\bar{W}^j = \begin{cases} \bar{W}_{12}^{-j} & \text{for } j < 0, \\ \bar{W}_{12}^0 + \bar{W}_{21}^0 & \text{for } j = 0, \\ \bar{W}_{21}^j & \text{for } j > 0. \end{cases} \quad (5.7)$$

Our new equation for the \bar{W} is then

$$E\{R_1^i R_2^{i-k}\} = \nu_{21}^k + \sum_j \tilde{A}^{kj} \bar{W}^j \quad (5.8)$$

where

$$\tilde{A}^{kj} = \begin{cases} A_{12}^{-k, -j} & \text{for } j < 0, \\ \frac{1}{2}(A_{12}^{-k0} + A_{21}^{k0}) & \text{for } j = 0, \\ A_{21}^{kj} & \text{for } j > 0. \end{cases} \quad (5.9)$$

If we let $\mathcal{S}^k = E\{R_1^i R_2^{i-k}\} - \nu_{21}^k$, we can write the solution of Eq. (5.8) for \bar{W}^j in matrix-vector notation as $\bar{W} = \tilde{A}^{-1}\mathcal{S}$, where \tilde{A}^{-1} denotes the matrix inverse of \tilde{A} . This solution of Eq. (5.8) for \bar{W}^j modifies the correlations in $E\{R_1^i R_2^{i-k}\}$ in two ways. First, the subtraction of ν_{21}^k removes correlations due solely to the fact that neurons are responding to the same stimulus. (See Ref. [12] for a detailed discussion.) Second, inverting the matrix \tilde{A} eliminates the filtering of \bar{W} by the temporal structure of \mathbf{h}_1^i and \mathbf{h}_2^i .

The relevant temporal structure of the kernels is captured by $\cos \bar{\theta}_{pq}^k = \mathbf{h}_p^{i-k} \cdot \mathbf{h}_q^i$. Clearly, $\cos \bar{\theta}_{pp}^0 = |\mathbf{h}_p^i|^2 = 1$. If, with this exception, $\cos \bar{\theta}_{pq}^k = 0$, (so that the kernels are orthogonal to each other and temporal shifts of themselves), then the effects of \bar{W} are not filtered by the kernels. \tilde{A} is a diagonal matrix, and inverting \tilde{A} simply scales the measured correlations. (To see this fact, recall that $\text{derfc}(a, b, 0) = \text{erfc}(a)\text{erfc}(b)$ and that we can interchange $\cos \bar{\theta}_{pq}^k$ and $\cos \theta_{pq}^k$ in expressions defining A since it appears in $O(\bar{W})$ terms.)

As the inner products $\cos \bar{\theta}_{pq}^k$ increase, the off-diagonal elements of \tilde{A} grow. In fact, the inner products of the kernels with themselves ($\cos \bar{\theta}_{pp}^k$) will be close to 1 for k near 0 if the structure of the kernels changes slowly with time. Typically, the off-diagonal elements of \tilde{A} will still be substantially less than the diagonal elements even with large $\cos \bar{\theta}_{pp}^k$, and inversion of \tilde{A} will be stable. However, close examination of equations defining \tilde{A} reveals that off-diagonals could become equal to the diagonal in the extreme case of very sharp nonlinearities and other parameter limits. (Parameters needed are $\epsilon_1 = \epsilon_2 = 0$ so that $\delta_1 = \delta_2 = 1$, as well as $\hat{r}_1 = \hat{r}_2 = 1$, $\cos \bar{\theta}_{pp}^k = 1$ for $k \neq 0$ and $\cos \bar{\theta}_{pq}^j = 0$ for $p \neq q$.)⁶ In this case, the matrix \tilde{A} could become almost singular, and its inversion would not be stable.

Outside this extreme case, the matrix \tilde{A} is well-conditioned, and solving Eq. (5.8) for \bar{W} removes the filtering caused by the temporal structure of the kernels. Subject to the validity of the model (5.1), the result will faithfully reconstruct the underlying connectivity.

6. Results. To demonstrate the reconstruction procedure, we simulate a pair of coupled linear-nonlinear neurons (Eq. (5.1)) responding to white noise input and use the above method to estimate the parameters. We assume that the maximum output rates \hat{r}_p are known using alternative methods such as those described in Ref. [13]. Then, from the responses R_p^i and the discrete white noise input X , one can estimate $E\{R_p^i\}$, $E\{\mathbf{X}R_p^i\}$, and $E\{R_1^i R_2^{i-k}\}$ for $p = 1, 2$ and $k = -N, \dots, N$. The maximum delay parameter N must be chosen large enough so that the $E\{R_1^i R_2^{i-k}\}$ capture the effects of the \bar{W}^j . In the examples, we set $N = 30$.

The calculations depend on estimating the inner products $E\{\mathbf{X}R_p^i\} \cdot E\{\mathbf{X}R_q^{i-k}\}$. We estimate each correlation by $E\{\mathbf{X}R_p^i\} \approx \langle \mathbf{X}R_p^i \rangle$, where $\langle \cdot \rangle$ represents averaging over a data set. A naive estimate of the inner product by $E\{\mathbf{X}R_p^i\} \cdot E\{\mathbf{X}R_q^{i-k}\} \approx \langle \mathbf{X}R_p^i \rangle \cdot \langle \mathbf{X}R_q^{i-k} \rangle$ will be highly biased, especially when the dimension of the kernels \mathbf{h}_p^i and \mathbf{h}_q^i is large. To reduce the bias, we estimate the covariance between the factors of each term defining $\langle \mathbf{X}R_p^i \rangle \cdot \langle \mathbf{X}R_q^{i-k} \rangle$ and subtract it from the estimate.⁷

For our simulations, we used kernels \mathbf{h}_p^i that mimic linear kernels of neurons in

⁶Note that $\lim_{c \rightarrow 1} \text{derfc}(a, a, c) = 2\text{erfc}(a)$.

⁷This bias reduction is equivalent to estimating the product of expected values of two random variables Y and Z using the formula for covariance $E\{YZ\} - \text{cov}(Y, Z) = E\{Y\}E\{Z\}$. For more details on bias reduction of inner products, see Appendix B of Ref. [13].

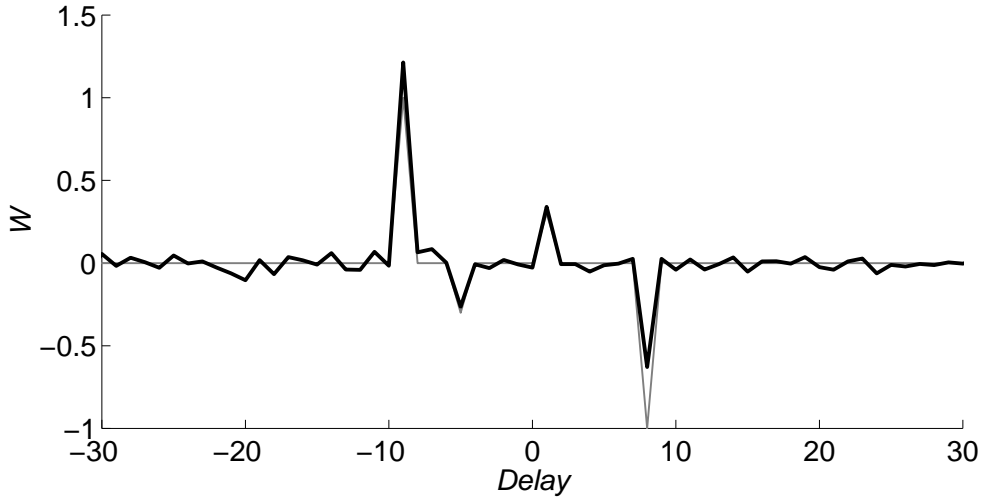


FIG. 1. Estimated connectivity \mathcal{W} (thick black line) when the nonlinearities are error functions. For comparison, the simulated connectivity \bar{W} is shown with a thin gray line. \mathcal{W} agrees quantitatively with \bar{W} , though the magnitudes of the large peaks differ. Delay is in units of time and is the spike time of neuron 1 minus the spike time of neuron 2.

visual cortex [10]. We used the spatio-temporal linear kernels of the form

$$h_p(\mathbf{j}, t) = \begin{cases} te^{-t/5} \exp\left(-\frac{|\mathbf{j}|^2}{50}\right) \sin(0.5(j_1 \cos \phi_p + j_2 \sin \phi_p)) & \text{for } t > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (6.1)$$

where $\mathbf{j} = (j_1, j_2)$ is the spatial grid point and t is time. We set the spatial axis parameters to be $\phi_1 = 0$ and $\phi_2 = \pi/4$. We sampled $h_p(\mathbf{j}, t)$ on a $32 \times 32 \times 32$ grid and normalized it to form the unit vector \mathbf{h}_p^i . All units are in grid points. The detailed structure of the kernels is insignificant as the only relevant parameters from the kernels are their inner products $\cos \bar{\theta}_{pq}^k$.

In the first example, we set the parameters of the error function nonlinearity (Eq. (5.2)) to $\hat{r}_1 = \hat{r}_2 = 0.5$, $\bar{T}_1 = 1.5$, $\bar{T}_2 = 2.0$, $\epsilon_1 = 0.5$, and $\epsilon_2 = 1.0$. The precise parameter values are arbitrary; we chose them so that the neuron firing rates would be low as observed in white noise experiments. The results are not sensitive to these parameter choices. Just to illustrate the method, we set an artificial coupling of $\bar{W}_{21}^1 = 0.3$, $\bar{W}_{21}^8 = -1.0$, $\bar{W}_{12}^5 = -0.3$, and $\bar{W}_{12}^9 = 1.0$. All other coupling terms were set to zero. We simulated the system for 250,000 units of time, obtaining about 10,000 spikes from each neuron, a realistic number of spikes in white noise experiments [17].

To analyze the results, we assumed we knew that $\hat{r}_p = 0.5$ and calculated all other parameters from the input-output statistics using the proposed method. We focus on the estimate of the simulated connectivity \bar{W} , denoting by \mathcal{W} our estimate of the connectivity. As shown in Fig. 1, the estimate \mathcal{W} captures all the qualitative features of \bar{W} . For the lower magnitude coupling (with $|\bar{W}| = 0.3$), \mathcal{W} also estimates the magnitudes accurately. However, when $|\bar{W}| = 1.0$, the first order approximation breaks down enough to cause \mathcal{W} to overestimate the positive coupling by 20% and underestimate the magnitude of the negative coupling by nearly 40%. (The asymmetry between positive and negative coupling is most likely due to the low average firing rates of 0.04 spikes per unit time; cf. Ref. [14].)

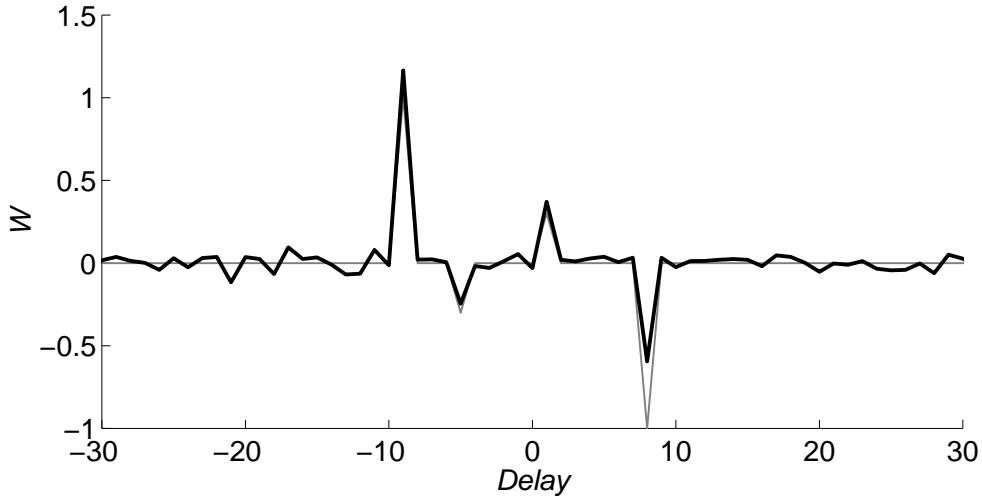


FIG. 2. Estimated connectivity \mathcal{W} (thick black line) when simulated power law nonlinearities are analyzed as error functions. \mathcal{W} agrees with the simulated connectivity \bar{W} (thin gray line) just as well as in the error function case of Fig. 1.

Since the stimulus standard deviation is assumed to be one, we have effectively scaled \mathbf{X} , and likewise \bar{W} , \bar{T}_p , and ϵ_p , by the stimulus standard deviation. When $|\bar{W}^j| = 1$, it is equal in magnitude to the standard deviation of $\mathbf{h}_p^i \cdot \mathbf{X}$. Since in this case the contribution of \bar{W}^j in Eq. (5.1) is the same order of magnitude as the contribution of $\mathbf{h}_p^i \cdot \mathbf{X}$, one cannot expect the first order approximation to be valid. Not only are estimation errors, such as those shown in Fig. 1, possible when the coupling magnitude is sufficiently large, but \mathcal{W} can also show additional peaks due to the second order interactions that we ignored in section 5 (not shown).

For a second example, we demonstrate the robustness of the analysis to deviations in the form of the nonlinearities g_p . We repeat the first example, but rather than using an error function nonlinearity, we use a power law nonlinearity,

$$g_p(y) = \begin{cases} A_p y^{\beta_p} & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

with $A_1 = 0.07$, $A_2 = 0.04$, $\beta_1 = 2.5$, and $\beta_2 = 2.0$ (we truncate so that $g_p(x) \leq 1$). Using the same \bar{W} as above, we simulated the system for 250,000 units of time, obtaining approximately 10,000 spikes from neuron 1 and 5,000 spikes from neuron 2.

We analyze the output of the system identically to the first example. We assume that each nonlinearity was an error function nonlinearity with $\hat{r}_p = 1$ and calculate the error function parameters from $E\{R_p^i\}$ and $|E\{\mathbf{X}R_p^i\}|$. The resulting error function parameters (which include the effects from the connectivity) were $\epsilon_1 = 0.76$, $\epsilon_2 = 1.1$, $T_1 = 2.2$, and $T_2 = 3.0$. As shown in Fig. 2, the method estimated the connectivity just as well as when the simulated nonlinearity really was an error function. The results were not sensitive to the selection of the maximum firing rate parameters, as the calculated \mathcal{W} was virtually identical if we set $\hat{r}_p = 0.5$ or $\hat{r}_p = 2$ and repeated the analysis.

We repeated this test for simulations based on a wide variety of power law parameters A_p and β_p . We were unable to find an example where the calculation of \mathcal{W} was

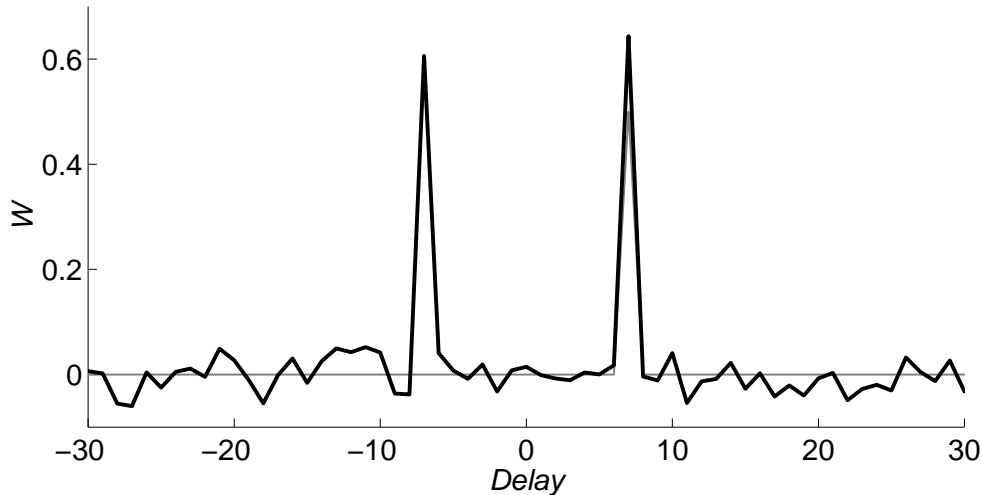


FIG. 3. Estimated connectivity \mathcal{W} (thick black line) when the two neurons receive common input from a third neuron. The peak at a delay of 7 units of time is due to the simulated connectivity \bar{W} (thin gray line). However, the peak of \mathcal{W} at a delay of -7 is not due to connectivity between the two neurons ($\bar{W} = 0$) but rather the common input from the third neuron.

significantly worse than in Fig. 2. Even with $\beta_p < 1$ so that the derivative of $g_p(y)$ was infinite at $y = 0$, the results were similar. The method simply is not sensitive to the detailed form of the nonlinearity.

The measure \mathcal{W} cannot distinguish between correlations caused by the connectivity assumed in Eq. (5.1) and correlations caused by other mechanisms. For example, if the two neurons received common input from a third, unmeasured, neuron, that connectivity would appear in the calculation of \mathcal{W} .

To demonstrate, we simulated three coupled linear-nonlinear neurons analogous to Eq. (5.1). We used Eq. (6.1) with $\phi_3 = \pi/2$ for linear kernel of the third neuron. All three neurons had error function nonlinearities with $\bar{T}_1 = 2$, $\bar{T}_2 = 2.5$, $\bar{T}_3 = 2$, $\epsilon_1 = 0.5$, $\epsilon_2 = 1$, and $\epsilon_3 = 0.7$. We created a connection from neuron 3 to both neurons 1 and 2 as well as a connection from neuron 2 to neuron 1. (We set $\bar{W}_{31}^1 = 1.5$, $\bar{W}_{32}^8 = 1.5$, and $\bar{W}_{21}^7 = 0.5$, leaving the other connectivity terms zero.) We simulated the system for 250,000 units of time, obtaining approximately 12,000–13,000 spikes per neuron, and then analyzed the system as above by ignoring the output of neuron 3.

As shown in Fig. 3, \mathcal{W} has a peak at the delay of 7 corresponding to the connection from neuron 2 to neuron 1 (\bar{W}_{21}^7). However, \mathcal{W} also has a peak at a delay of -7 . This second peak does not correspond to any direct connection between neuron 1 and neuron 2 ($\bar{W}_{12}^7 = 0$). Instead, the peak is created because the connection from neuron 3 to neuron 2 is 7 units of time delayed compared to the connection from neuron 3 to neuron 1. Since \mathcal{W} cannot distinguish between direct connections and common input, it must be interpreted with care. It cannot be viewed as representing the connectivity between the two measured neurons unless one could somehow rule out a mutual connection from any unmeasured neurons.

7. Discussion. We derived a method for analyzing a pair of coupled linear-nonlinear systems driven by white noise. Through measuring first and second order input-output statistics, one can characterize the systems. In particular, one can re-

construct the coupling between the systems if the coupling is assumed to be of a particular form (Eq. (5.1)).

We demonstrated that the method is robust to variations in the detailed form of the nonlinearity. We believe this robustness is due to the smoothing by the white noise input. Each input-output statistic depends on the nonlinearities $g_p(\cdot)$ only through expected values over the white noise. The effect of this smoothing is most clearly seen in the initial expression for each statistic in Appendix A. The $g_p(\cdot)$ appear in the integrals as either $g_p(\mathbf{h}_p^i \cdot \mathbf{x})$ or $g'_p(\mathbf{h}_p^i \cdot \mathbf{x})$. Since the kernels are unit vectors, the arguments of the nonlinearity are standard normals. Only the integrals of the $g_p(\cdot)$ over the probability density function of standard normals, not pointwise evaluation of the $g_p(\cdot)$, affect the input-output statistics. These integrals smooth out minor differences between nonlinearity shapes.

Since the method is a first order approximation in the coupling magnitude, measurements of large \mathcal{W} (on the order of the standard deviation of an input component) must be viewed cautiously. According to our simulation results, the breakdown of the first order approximation typically leads only to deviations in the magnitude of the estimated connectivity. However, in extreme cases, large connectivity could lead to the emergence of second order effects in the form of additional peaks in \mathcal{W} that do not reflect the connectivity \bar{W} .

More importantly, the method cannot distinguish between the assumed mutual coupling of the model and other mechanisms for creating correlations between the responses, such as common input from outside sources. Measurements of \mathcal{W} would be evidence of mutual coupling only if other mechanisms for correlations could be ruled out. Nonetheless, even if the source of \mathcal{W} cannot be definitively determined, measurement of \mathcal{W} still could provide evidence about the time scale and magnitudes of the interactions in the underlying neural network.

The proposed method was developed to analyze multi-electrode recordings of neurons in response to a white noise stimulus. However, the linear-nonlinear model assumed by the analysis is only a crude, phenomenological approximation to the biology. To better interpret the results of the method, one must be able to assign significance to nonzero measurements of \mathcal{W} . One future challenge is to develop methods to identify cases when nonzero measurements of \mathcal{W} are due simply to deviations from the linear-nonlinear model.

Appendix A. Details of derivation for unidirectional coupling.

A.1. Probability of a spike in neuron 1. Under the first order approximation in \bar{W} , we can simplify Eq. (2.2) for neuron 1 to

$$\begin{aligned} \Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_2 = \mathbf{r}_2) &= g_1\left(\mathbf{h}_1^i \cdot \mathbf{x} + \sum_{j \geq 0} \bar{W}_{21}^j r_2^{i-j}\right) \\ &= g_1(\mathbf{h}_1^i \cdot \mathbf{x}) + g'_1(\mathbf{h}_1^i \cdot \mathbf{x}) \sum_{j \geq 0} \bar{W}_{21}^j r_2^{i-j}. \end{aligned} \quad (\text{A.1})$$

The probability of a spike in neuron 1 is then

$$\begin{aligned} \Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}) &= \sum_{\mathbf{r}_2} \Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_2 = \mathbf{r}_2) \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) \\ &= g_1(\mathbf{h}_1^i \cdot \mathbf{x}) \sum_{\mathbf{r}_2} \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) \\ &\quad + g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) \sum_{j \geq 0} \bar{W}_{21}^j \sum_{\mathbf{r}_2} r_2^{i-j} \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) \quad (\text{A.2}) \end{aligned}$$

The sum is over all values of \mathbf{r}_2 where each component $r_2^{\tilde{j}}$ can be either one or zero, i.e., this sum is over all every possible spike combination of neuron 2. The product reflects the assumption that, since $\bar{W}_{12} = 0$, the responses of neuron 2, when conditioned on the stimulus, are independent.

The total probability of any spike combination of neuron 2 must equal one,

$$\sum_{\mathbf{r}_2} \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) = 1. \quad (\text{A.3})$$

Moreover, since $r_2^{i-j} \in \{0, 1\}$, only terms where $r_2^{i-j} = 1$ make a contribution in the coefficient of \bar{W}_{21}^j :

$$\begin{aligned} &\sum_{\mathbf{r}_2} r_2^{i-j} \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) \\ &= \Pr(R_2^{i-j} = 1 | \mathbf{X} = \mathbf{x}) \sum_{\mathbf{r}_2 \text{ except } r_2^{i-j} \neq 1} \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}) \\ &= g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}). \quad (\text{A.4}) \end{aligned}$$

In the last step, we used a generalization of Eq. (A.3) excluding the $i-j$ time interval.

Combining Eqs. (A.2), (A.3), and (A.4), the probability of a spike in neuron 1 is

$$\Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}) = g_1(\mathbf{h}_1^i \cdot \mathbf{x}) + \sum_{j \geq 0} \bar{W}_{21}^j g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}) \quad (\text{A.5})$$

where $=$ indicates equality within $O(\bar{W}^2)$.

A.2. Probability of spike pairs. In the case of unidirectional coupling, the probability of a spike pair is

$$\begin{aligned} \Pr(R_1^i = 1 \&\& R_2^{i-k} = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_2 = \mathbf{r}_2) &= g_1\left(\mathbf{h}_1^i \cdot \mathbf{x} + \sum_{j \geq 0} \bar{W}_{21}^j r_2^{i-j}\right) r_2^{i-k} \\ &= g_1(\mathbf{h}_1^i \cdot \mathbf{x}) r_2^{i-k} + g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) r_2^{i-k} \bar{W}_{21}^k + g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) r_2^{i-k} \sum_{\substack{j \geq 0 \\ j \neq k}} \bar{W}_{21}^j r_2^{i-j}. \quad (\text{A.6}) \end{aligned}$$

Note that $(r_2^{i-k})^2 = r_2^{i-k}$ since $r_2^{i-k} \in \{0, 1\}$.

If we repeat the same procedure as in the previous section,

$$\begin{aligned} \Pr(R_1^i = 1 \&\& R_2^{i-k} = 1 | \mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{r}_2} \Pr(R_1^i = 1 \&\& R_2^{i-k} = 1 | \mathbf{X} = \mathbf{x}, \mathbf{R}_2 = \mathbf{r}_2) \prod_{\tilde{j}} \Pr(R_2^{i-\tilde{j}} = r_2^{i-\tilde{j}} | \mathbf{X} = \mathbf{x}), \quad (\text{A.7}) \end{aligned}$$

the only new term will be

$$\sum_{\mathbf{r}_2} r_2^{i-j} r_2^{i-k} \prod_j \Pr(R_2^{i-j} = r_2^{i-j} | \mathbf{X} = \mathbf{x}) = g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}). \quad (\text{A.8})$$

Therefore,

$$\begin{aligned} \Pr(R_1^i = 1 \& R_2^{i-k} = 1 | \mathbf{X} = \mathbf{x}) &= g_1(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) \\ &+ \bar{W}_{21}^k g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) \\ &+ \sum_{\substack{j \geq 0 \\ j \neq k}} \bar{W}_{21}^j g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}) \end{aligned} \quad (\text{A.9})$$

A.3. Mean rate of neuron 1. The mean rate of neuron 1 (see Eq. (A.5)) is given by

$$\begin{aligned} E\{R_1^i\} &= \frac{1}{(2\pi)^{n/2}} \int \Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \\ &= \frac{1}{(2\pi)^{n/2}} \int g_1(\mathbf{h}_1^i \cdot \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \\ &+ \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{(2\pi)^{n/2}} \int g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \end{aligned} \quad (\text{A.10})$$

The first term is identical to the uncoupled case. For the rest of the terms, we use a different coordinate system for each j . The first unit vector is $\mathbf{e}_1 = \mathbf{h}_1^i$, and the second unit vector is the component of \mathbf{h}_2^{i-j} that is perpendicular to \mathbf{h}_1^i , so that $\mathbf{h}_2^{i-j} = \mathbf{e}_1 \cos \bar{\theta}_{21}^j + \mathbf{e}_2 \sin \bar{\theta}_{21}^j$.

We change variables and integrate by parts (assuming Eq. (3.5)) to simplify the j th term:

$$\begin{aligned} &\frac{\bar{W}_{21}^j}{2\pi} \int g_1'(x_1) g_2(x_1 \cos \bar{\theta}_{21}^j + x_2 \sin \bar{\theta}_{21}^j) e^{-\frac{x_1^2 + x_2^2}{2}} dx_1 dx_2 \\ &= \frac{\bar{W}_{21}^j}{2\pi} \int g_1'(u) g_2(v) \exp\left(-\frac{u^2}{2} - \frac{(v - u \cos \bar{\theta}_{21}^k)^2}{2 \sin^2 \bar{\theta}_{21}^k}\right) \frac{du dv}{\sin \bar{\theta}_{21}^k} \\ &= \frac{\bar{W}_{21}^j}{2\sqrt{2}\pi} \int g_1'(u) g_2'(v) e^{-\frac{u^2}{2}} \operatorname{erfc}\left(\frac{v - u \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du dv. \end{aligned} \quad (\text{A.11})$$

The mean rate of neuron 1 is thus

$$\begin{aligned} E\{R_1^i\} &= \frac{1}{\sqrt{2}\pi} \int g_1(u) e^{-\frac{u^2}{2}} du \\ &+ \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2\sqrt{2}\pi} \int g_1'(u) g_2'(v) e^{-\frac{u^2}{2}} \operatorname{erfc}\left(\frac{v - u \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du dv. \end{aligned} \quad (\text{A.12})$$

A.4. Correlation of spikes of neuron 1 with stimulus. The stimulus-spike correlation of neuron 1 (see Eq. (A.5)) is

$$\begin{aligned} E\{\mathbf{X}R_1^i\} &= \frac{1}{(2\pi)^{n/2}} \int \mathbf{x} \Pr(R_1^i = 1 | \mathbf{X} = \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \\ &= \frac{1}{(2\pi)^{n/2}} \int \mathbf{x} g_1(\mathbf{h}_1^i \cdot \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \\ &\quad + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{(2\pi)^{n/2}} \int \mathbf{x} g_1'(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}) e^{-|\mathbf{x}|^2/2} d\mathbf{x} \end{aligned} \quad (\text{A.13})$$

The first term is identical to the uncoupled case, becoming

$$\frac{1}{\sqrt{2\pi}} \int g_1'(u) e^{-\frac{u^2}{2}} du \mathbf{h}_1^i$$

with an integration by parts. For the rest of the terms, just as in the previous section, we will use a different coordinate system for each j , with $\mathbf{e}_1 = \mathbf{h}_1^i$ and \mathbf{e}_2 being the component of \mathbf{h}_2^{i-j} that is perpendicular to \mathbf{h}_1^i . We will denote this second unit vector by

$$\mathbf{h}_{21}^{\perp ji} = \frac{\mathbf{h}_2^{i-j} - \cos \bar{\theta}_{21}^j \mathbf{h}_1^i}{\sin \bar{\theta}_{21}^j}. \quad (\text{A.14})$$

Note that $\mathbf{h}_2^{i-j} = \mathbf{h}_1^i \cos \bar{\theta}_{21}^j + \mathbf{h}_{21}^{\perp ji} \sin \bar{\theta}_{21}^j$. The j th term thus has two nonzero components,

$$\begin{aligned} &\frac{\bar{W}_{21}^j}{2\pi} \int (x_1 \mathbf{h}_1^i + x_2 \mathbf{h}_{21}^{\perp ji}) g_1'(x_1) g_2(x_1 \cos \bar{\theta}_{21}^j + x_2 \sin \bar{\theta}_{21}^j) e^{-\frac{x_1^2 + x_2^2}{2}} dx_1 dx_2 \\ &= \bar{W}_{21}^j (I_{j,1} \mathbf{h}_1^i + I_{j,2} \mathbf{h}_{21}^{\perp ji}), \end{aligned} \quad (\text{A.15})$$

where the above defines $I_{j,1}$ and $I_{j,2}$. We change variables and integrate by parts (assuming Eq. (3.5)) to simplify the first component:

$$\begin{aligned} I_{j,1} &= \frac{1}{2\pi} \int u g_1'(u) g_2(v) \exp\left(-\frac{u^2}{2} - \frac{(v - u \cos \bar{\theta}_{21}^j)^2}{2 \sin^2 \bar{\theta}_{21}^j}\right) \frac{du dv}{\sin \bar{\theta}_{21}^j} \\ &= \frac{1}{2\sqrt{2\pi}} \int g_1'(u) g_2'(v) u e^{-\frac{u^2}{2}} \operatorname{erfc}\left(\frac{v - u \cos \bar{\theta}_{21}^j}{\sqrt{2} \sin \bar{\theta}_{21}^j}\right) du dv. \end{aligned} \quad (\text{A.16})$$

To simplify $I_{j,2}$, we first integrate by parts in the x_2 variable, then change variables:

$$\begin{aligned} I_{j,2} &= \frac{1}{2\pi} \int g_1'(x_1) g_2'(x_1 \cos \bar{\theta}_{21}^j + x_2 \sin \bar{\theta}_{21}^j) \sin \bar{\theta}_{21}^j e^{-\frac{x_1^2 + x_2^2}{2}} dx_1 dx_2 \\ &= \frac{1}{2\pi} \int g_1'(u) g_2'(v) \exp\left(-\frac{u^2 - 2 \cos \bar{\theta}_{21}^j uv + v^2}{2 \sin^2 \bar{\theta}_{21}^j}\right) du dv. \end{aligned} \quad (\text{A.17})$$

Combining these results, the stimulus-spike correlation of neuron 1 is

$$\begin{aligned}
E\{\mathbf{X}R_1^i\} &= \frac{1}{\sqrt{2\pi}} \left[\int g'_1(u) e^{-\frac{u^2}{2}} du \right. \\
&\quad \left. + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2} \int g'_1(u) g'_2(v) u e^{-\frac{u^2}{2}} \operatorname{erfc}\left(\frac{v - u \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du dv \right] \mathbf{h}_1^i \\
&\quad + \sum_{j \geq 0} \frac{\bar{W}_{21}^j}{2\pi} \int g'_1(u) g'_2(v) \exp\left(-\frac{u^2 - 2 \cos \bar{\theta}_{21}^j uv + v^2}{2 \sin^2 \bar{\theta}_{21}^j}\right) du dv \mathbf{h}_{21}^{\perp j i} \quad (\text{A.18})
\end{aligned}$$

A.5. Correlation between spikes of neuron 1 and 2. The correlation between spikes of neuron 1 and neuron 2 is (see Eq. (A.9))

$$\begin{aligned}
E\{R_1^i R_2^{i-k}\} &= \frac{1}{(2\pi)^{n/2}} \int \Pr(R_1^i = 1 \& R_2^{i-k} = 1 | \mathbf{X} = \mathbf{x}) e^{-\frac{|\mathbf{x}|^2}{2}} d\mathbf{x} \\
&= \frac{1}{(2\pi)^{n/2}} \int \left[g_1(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) \right. \\
&\quad \left. + g'_1(\mathbf{h}_1^i \cdot \mathbf{x}) g_2(\mathbf{h}_2^{i-k} \cdot \mathbf{x}) \left(\bar{W}_{21}^k + \sum_{j \geq 0, j \neq k} \bar{W}_{21}^j g_2(\mathbf{h}_2^{i-j} \cdot \mathbf{x}) \right) \right] e^{-\frac{|\mathbf{x}|^2}{2}} d\mathbf{x}. \quad (\text{A.19})
\end{aligned}$$

The first term is identical to the uncoupled case (Eq. (3.6)). The \bar{W}_{21}^k term is identical to Eq. (A.11).

For the \bar{W}_{21}^j terms with $j \neq k$, we let $\mathbf{e}_1 = \mathbf{h}_1^i$ and $\mathbf{e}_2 = \mathbf{h}_{21}^{\perp ki}$, and let the third unit vector be the component of \mathbf{h}_2^{i-j} perpendicular to both \mathbf{e}_1 and \mathbf{e}_2 so that

$$\mathbf{h}_2^{i-j} = \mathbf{e}_1 \cos \bar{\theta}_{21}^j + \mathbf{e}_2 c_{21}^{kj} \sin \bar{\theta}_{21}^j + \mathbf{e}_3 \sin \bar{\theta}_{21}^j \sqrt{1 - (c_{21}^{kj})^2}$$

where

$$c_{21}^{kj} = \mathbf{h}_{21}^{\perp ki} \cdot \mathbf{h}_{21}^{\perp ji} = \frac{\cos \theta_{22}^{k-j} - \cos \bar{\theta}_{21}^k \cos \bar{\theta}_{21}^j}{\sin \bar{\theta}_{21}^k \sin \bar{\theta}_{21}^j}.$$

Denoting the \bar{W}_{21}^j terms in Eq. (A.19) by $\bar{W}_{21}^j I_{kj}$ and changing variables, we compute

$$\begin{aligned}
I_{kj} &= \frac{1}{(2\pi)^{3/2}} \int g'_1(x_1) g_2(x_1 \cos \bar{\theta}_{21}^k + x_2 \sin \bar{\theta}_{21}^k) \\
&\quad \times g_2\left(x_1 \cos \bar{\theta}_{21}^j + x_2 c_{21}^{kj} \sin \bar{\theta}_{21}^j + x_3 \sin \bar{\theta}_{21}^j \sqrt{1 - (c_{21}^{kj})^2}\right) e^{-\frac{x_1^2 + x_2^2 + x_3^2}{2}} dx_1 dx_2 dx_3 \\
&= \frac{1}{(2\pi)^{3/2}} \int \frac{du_1 du_2 du_3 g'_1(u_1) g_2(u_2) g_2(u_3)}{\sin \bar{\theta}_{21}^k \sin \bar{\theta}_{21}^j \sqrt{1 - (c_{21}^{kj})^2}} \\
&\quad \times \exp\left(-\frac{u_1^2}{2} - \frac{(u_2 - u_1 \cos \bar{\theta}_{21}^k)^2}{2 \sin^2 \bar{\theta}_{21}^k} - \frac{\left[\frac{u_3 - u_1 \cos \bar{\theta}_{21}^j}{\sin \bar{\theta}_{21}^j} - c_{21}^{kj} \frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sin \bar{\theta}_{21}^k}\right]^2}{2[1 - (c_{21}^{kj})^2]}\right).
\end{aligned}$$

Using Eq. (3.5) and integrating by parts twice as in the derivation of Eq. (3.6), we

simplify this expression to

$$I_{kj} = \frac{1}{4\sqrt{2\pi}} \int g'_1(u_1)g'_2(u_2)g'_2(u_3)e^{-\frac{u_1^2}{2}} \times \text{derfc}\left(\frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}, \frac{u_3 - u_1 \cos \bar{\theta}_{21}^j}{\sqrt{2} \sin \bar{\theta}_{21}^j}, c_{21}^{kj}\right) du_1 du_2 du_3. \quad (\text{A.20})$$

The correlation between spikes of neuron 1 and neuron 2 is therefore

$$\begin{aligned} E\{R_1^i R_2^{i-k}\} &= \frac{1}{4} \int g'_1(u_1)g'_2(u_2) \text{derfc}\left(\frac{u_1}{\sqrt{2}}, \frac{u_2}{\sqrt{2}}, \cos \bar{\theta}_{21}^k\right) du_1 du_2 \\ &+ \frac{\bar{W}_{21}^k}{2\sqrt{2\pi}} \int g'_1(u_1)g'_2(u_2) e^{-\frac{u_1^2}{2}} \text{erfc}\left(\frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}\right) du_1 du_2 \\ &+ \sum_{j \geq 0, j \neq k} \frac{\bar{W}_{21}^j}{4\sqrt{2\pi}} \int du_1 du_2 du_3 g'_1(u_1)g'_2(u_2)g'_2(u_3) e^{-\frac{u_1^2}{2}} \\ &\times \text{derfc}\left(\frac{u_2 - u_1 \cos \bar{\theta}_{21}^k}{\sqrt{2} \sin \bar{\theta}_{21}^k}, \frac{u_3 - u_1 \cos \bar{\theta}_{21}^j}{\sqrt{2} \sin \bar{\theta}_{21}^j}, \frac{\cos \theta_{22}^{k-j} - \cos \bar{\theta}_{21}^j \cos \bar{\theta}_{21}^k}{\sin \bar{\theta}_{21}^j \sin \bar{\theta}_{21}^k}\right). \end{aligned} \quad (\text{A.21})$$

Appendix B. Formulas used in derivations. In all formulas, each sine is assumed to be positive.

The formulas

$$\frac{1}{\epsilon_p \sqrt{2\pi}} \iint \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{x^2}{2}\right) dx = \delta_p e^{-\frac{\delta_p^2 T_p^2}{2}} \quad (\text{B.1})$$

and

$$\begin{aligned} \frac{1}{2\pi\epsilon_p\epsilon_q} \iint \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{(y - T_q)^2}{2\epsilon_q^2} - \frac{x^2 - 2xy \cos \theta + y^2}{2 \sin^2 \theta}\right) dx dy \\ = \frac{\delta_p \delta_q \sin \theta}{\sqrt{1 - \delta_p^2 \delta_q^2 \cos^2 \theta}} \exp\left(-\frac{\delta_p^2 T_p^2 - 2\delta_p^2 \delta_q^2 T_p T_q \cos \theta + \delta_q^2 T_q^2}{2(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}\right), \end{aligned} \quad (\text{B.2})$$

where $\delta_q = 1/\sqrt{1 + \epsilon_q^2}$, follow from the application of

$$\int \exp(-[ax^2 + bx + c]) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right)$$

for $a > 0$.

For the following two formulas, change variables in the double integral so that one of the new variables is parallel to the line $u = dx + f$ (where u is the integration variable of the $\text{erfc}(\cdot)$). By completing the square in the resulting integrands, one can derive both

$$\int \exp(-[ax^2 + bx + c]) \text{erfc}(dx + f) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} - c\right) \text{erfc}\left(\frac{2af - bd}{2\sqrt{a(a + d^2)}}\right) \quad (\text{B.3})$$

and

$$\begin{aligned} & \int x \exp(-[ax^2 + bx + c]) \operatorname{erfc}(dx + f) dx \\ &= -\frac{1}{a} \exp\left(\frac{b^2}{4a} - c\right) \left[\frac{d \exp\left(-\frac{(2af - bd)^2}{4a(a+d^2)}\right)}{\sqrt{a+d^2}} + \frac{b\sqrt{\pi}}{2\sqrt{a}} \operatorname{erfc}\left(\frac{2af - bd}{2\sqrt{a(a+d^2)}}\right) \right] \end{aligned} \quad (\text{B.4})$$

for $a > 0$. By applying Eq. (B.3) twice, one can show that

$$\begin{aligned} & \frac{1}{2\pi\epsilon_p\epsilon_q} \iint \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{(y - T_q)^2}{2\epsilon_q^2} - \frac{x^2}{2}\right) \operatorname{erfc}\left(\frac{y - x \cos \theta}{\sqrt{2} \sin \theta}\right) dx dy \\ &= \delta_p e^{-\frac{\delta_p^2 T_p^2}{2}} \operatorname{erfc}\left(\frac{\delta_q T_q - \delta_p^2 \delta_q T_p \cos \theta}{\sqrt{2(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}}\right), \end{aligned} \quad (\text{B.5})$$

and by applying both Eq. (B.3) and Eq. (B.4), one can show that

$$\begin{aligned} & \frac{1}{2\pi\epsilon_p\epsilon_q} \iint x \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{(y - T_q)^2}{2\epsilon_q^2} - \frac{x^2}{2}\right) \operatorname{erfc}\left(\frac{y - x \cos \theta}{\sqrt{2} \sin \theta}\right) dx dy \\ &= \delta_p^3 T_p e^{-\frac{\delta_p^2 T_p^2}{2}} \operatorname{erfc}\left(\frac{\delta_q T_q - \delta_p^2 \delta_q T_p \cos \theta}{\sqrt{2(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}}\right) \\ &+ \frac{2\delta_p \delta_q (1 - \delta_p^2) \cos \theta}{\sqrt{2\pi(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}} \exp\left(-\frac{\delta_p^2 T_p^2 - 2\delta_p^2 \delta_q^2 T_p T_q \cos \theta + \delta_q^2 T_q^2}{2(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}\right). \end{aligned} \quad (\text{B.6})$$

For the following formula, let u be the integration variable of the $\operatorname{erfc}(\cdot)$. Then, change variables in the triple integral so that the first variable is parallel to the line $y = dx + f$ and a linear combination of the first and second variables is parallel to the line $u = gx + h - ky$. By repeatedly completing the square in the integrand, one can derive that

$$\begin{aligned} & \int dx \exp(-[ax^2 + bx + c]) \int_{dx+f}^{\infty} dy e^{-y^2} \operatorname{erfc}(gx + h - ky) \\ &= \sqrt{\frac{\pi}{a}} e^{\left(\frac{b^2}{4a} - c\right)} \int_{\frac{2af - bd}{2\sqrt{a(a+d^2)}}}^{\infty} e^{-u^2} \operatorname{erfc}\left(\frac{(2ha - bg)\sqrt{a+d^2} - 2\sqrt{a}(gd + ka)u}{2a\sqrt{(kd-g)^2 + a+d^2}}\right) du \end{aligned} \quad (\text{B.7})$$

for $a > 0$. Repeated application of Eq. (B.7), combined with extensive algebra, yields

$$\begin{aligned} & \frac{1}{2\pi\epsilon_p\epsilon_q} \iint \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{(y - T_q)^2}{2\epsilon_q^2}\right) \operatorname{derfc}\left(\frac{x}{\sqrt{2}}, \frac{y}{\sqrt{2}}, \cos \theta\right) dx dy \\ &= \operatorname{derfc}\left(\frac{\delta_p T_p}{\sqrt{2}}, \frac{\delta_q T_q}{\sqrt{2}}, \delta_p \delta_q \cos \theta\right) \end{aligned} \quad (\text{B.8})$$

and

$$\begin{aligned} & \frac{1}{(2\pi)^{3/2} \epsilon_p \epsilon_q^2} \iiint \exp\left(-\frac{(x - T_p)^2}{2\epsilon_p^2} - \frac{(y - T_q)^2}{2\epsilon_q^2} - \frac{(z - T_q)^2}{2\epsilon_q^2} - \frac{x^2}{2}\right) \\ & \times \operatorname{derfc}\left(\frac{z - x \cos \theta}{\sqrt{2} \sin \theta}, \frac{y - x \cos \phi}{\sqrt{2} \sin \phi}, \frac{\cos \psi - \cos \theta \cos \phi}{\sin \theta \sin \phi}\right) dx dy dz \\ &= \delta_p e^{-\frac{\delta_p^2 T_p^2}{2}} \operatorname{derfc}\left(\frac{\delta_q T_q - \delta_p^2 \delta_q T_p \cos \theta}{\sqrt{2(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)}}, \frac{\delta_q T_q - \delta_p^2 \delta_q T_p \cos \phi}{\sqrt{2(1 - \delta_p^2 \delta_q^2 \cos^2 \phi)}}, \frac{\delta_q^2 \cos \psi - \delta_p^2 \delta_q^2 \cos \theta \cos \phi}{\sqrt{(1 - \delta_p^2 \delta_q^2 \cos^2 \theta)(1 - \delta_p^2 \delta_q^2 \cos^2 \phi)}}\right) \end{aligned} \quad (\text{B.9})$$

where $\text{derfc}(\cdot)$ is defined by Eq. (3.7).

Acknowledgments. The author thanks Dario Ringach for numerous helpful discussions throughout the development of this research and Charlie Peskin and Dan Tranchina for constructive criticism on an early version of these ideas.

REFERENCES

- [1] Y. DAN, J.-M. ALONSO, W. M. USREY, AND R. C. REID, *Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus*, Nat. Neurosci, 1 (1998), pp. 501–507.
- [2] G. C. DEANGELIS, I. OHZAWA, AND R. D. FREEMAN, *Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development*, J. Neurophysiol., 69 (1993), pp. 1091–1117.
- [3] ———, *Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation*, J. Neurophysiol., 69 (1993), pp. 1118–1135.
- [4] E. DEBOER AND P. KUYPER, *Triggered correlation*, IEEE Trans. Biomed. Eng., 15 (1968), pp. 169–179.
- [5] R. C. DECHARMS, D. T. BLAKE, AND M. M. MERZENICH, *Optimizing sound features for cortical neurons*, Science, 280 (1998), pp. 1439–1443.
- [6] J. J. DICARLO AND K. O. JOHNSON, *Velocity invariance of receptive field structure in somatosensory cortical area 3b of the alert monkey*, J. Neurosci., 19 (1999), pp. 401–419.
- [7] J. J. DICARLO, K. O. JOHNSON, AND S. S. HSIAO, *Structure of receptive fields in area 3b of primary somatosensory cortex in the alert monkey*, J. Neurosci., 18 (1998), pp. 2626–2645.
- [8] R. L. JENISON, J. W. H. SCHNUPP, R. A. REALE, AND J. F. BRUGGE, *Auditory space-time receptive field dynamics revealed by spherical white-noise analysis*, J. Neurosci., 21 (2001), pp. 4408–4415.
- [9] J. P. JONES AND L. A. PALMER, *An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex*, J. Neurophysiol., 58 (1987), pp. 1233–1258.
- [10] S. MARCELJA, *Mathematical description of the responses of simple cortical cells*, J. Opt. Soc. Am., 70 (1980), pp. 1297–1300.
- [11] P. N. MARMARELIS AND V. Z. MARMARELIS, *Analysis of physiological systems: the white noise approach*, Plenum Press, New York, 1978.
- [12] D. Q. NYKAMP, *Spike correlation measures that eliminate stimulus effects in response to white noise*, J. Comp. Neurosci. To appear. (www.math.ucla.edu/~nykamp/pubs).
- [13] D. Q. NYKAMP AND D. L. RINGACH, *Full identification of a linear-nonlinear system via cross-correlation analysis*, J. Vision, 2 (2002), pp. 1–11.
- [14] G. PALM, A. M. H. J. AERTSEN, AND G. L. GERSTEIN, *On the significance of correlations among neuronal spike trains*, Biol. Cybern., 59 (1988), pp. 1–11.
- [15] R. C. REID AND J. M. ALONSO, *Specificity of monosynaptic connections from thalamus to visual cortex*, Nature, 378 (1995), pp. 281–284.
- [16] R. C. REID, J. D. VICTOR, AND R. M. SHAPLEY, *The use of m-sequences in the analysis of visual neurons: linear receptive field properties*, Vis. Neurosci., 14 (1997), pp. 1015–1027.
- [17] D. L. RINGACH. personal communication.
- [18] D. L. RINGACH, M. J. HAWKEN, AND R. SHAPLEY, *Dynamics of orientation tuning in macaque primary visual cortex*, Nature, 387 (1997), pp. 281–284.
- [19] W. M. USREY, J.-M. ALONSO, AND R. C. REID, *Synaptic interactions between thalamic inputs to simple cells in cat visual cortex*, J. Neurosci., 20 (2000), pp. 5461–5467.