# Spike correlation measures that eliminate stimulus effects in response to white noise

Duane Q. Nykamp*

## Abstract

When measured in response to non-repeating white noise, standard covariance measures of two neuronal spike trains contain components due simply to a shared stimulus. We argue that, without stimulus repeats, model-free measures cannot in general remove these stimulus-induced components. We present spike correlation measures that eliminate them when the neural response can be approximated by a linear-nonlinear system. One of these measures fully characterizes the correlations in the special case that all remaining correlations are due to small reciprocal connections between the neurons. In addition, we demonstrate that the proposed measures can give accurate results with a more realistic, integrate-and-fire model of neural response, provided that it is driven like a linear-nonlinear system.

Keywords: neural networks, correlations, Weiner analysis, white noise, correlogram

## 1 Introduction

Data from simultaneous recordings of two spike trains in response to a common stimulus are typically analyzed using the correlation between the spike times of the two neurons. Correlations between the spike times could be induced by a number of factors, including (but not limited to) the presence

*Department of Mathematics, University of California, Los Angeles, Los Angeles, California 90095-1555. Telephone: 310-825-8525, Fax: 310-206-6673, E-mail: nykamp@math.ucla.edu

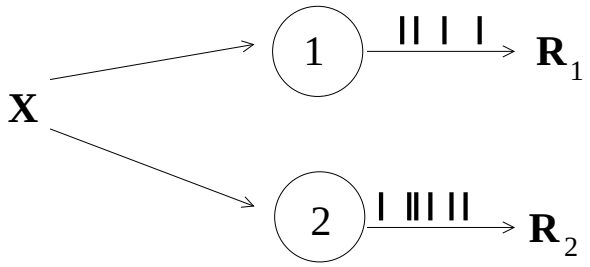


Figure 1: Schematic of two neurons whose responses are independent when conditioned on the stimulus. Given a particular stimulus $\mathbf{X}$, the spike responses ($R_1$ and $R_2$) are independent. However, in general the responses are not independent because the common stimulus introduces dependencies between the two spike trains.

of the common stimulus and the structure of the neural network containing the measured neurons [16, 1, 15, 2].

Correlations due solely to the fact that the neurons are responding to the same stimulus are relatively uninteresting. As explained below, a common stimulus will in general lead to spike correlations. A more useful correlation measure will determine if there are other sources of correlation as well. One may be interested in determining if two neurons respond to common stimulus features due to the connectivity of the neural network. (The neurons may simply be responding to the same stimulus features independently.) Before one can make inferences based on spike correlations about the underlying circuitry, one must, at minimum, remove contributions to the correlations due simply to the common stimulus.

Removal of stimulus-induced correlations is based on the fundamental notion of independence *conditioned on the stimulus*. If two neurons respond to a stimulus $\mathbf{X}$ independently, they are said to be independent conditioned on the stimulus, schematized in Fig. 1. However, since the neurons are responding to a common stimulus, the spike trains will in general be dependent. This dependence typically leads to a correlation between the spike trains.

Mathematically, the responses of two neurons are independent conditioned on the stimulus $\mathbf{X}$ when

$$\Pr\big(R_1^i = 1 \,\&\, R_2^j = 1|\mathbf{X}\big) = \Pr\big(R_1^i = 1|\mathbf{X}\big)\Pr\big(R_2^j = 1|\mathbf{X}\big), \qquad (1)$$

where $R_p^i = 1$ if neuron $p$ spiked at the discrete time point $i$ and is zero otherwise. When the stimulus is known, the probability of a spike pair from

the two neurons is completely determined by the spiking probability of each neuron separately.

When Eq. (1) is satisfied, the neurons are not interacting and one would like the expected value of a correlation measure to be zero. In that case, significantly nonzero values of the correlation measure would indicate that the neurons' responses to the stimulus somehow depend on each other. Since such a correlation measure contains only stimulus-independent correlations (as stimulus-dependent correlations have been removed), we refer to it as a stimulus independent correlation measure (SICM). More precisely, a SICM is a correlation measure whose expected value converges to zero as the sample size increases whenever Eq. (1) is satisfied.

Scientists have long recognized the need for SICMs, and standard SICMs are commonly used to analyze spikes in response to the repetition of an identical stimulus. When the stimulus is repeated, one can estimate $\Pr\left(R_p^i = 1|\mathbf{X}\right)$ by averaging the response to the stimulus (obtaining the peristimulus time histogram or PSTH). Since one can similarly estimate $\Pr\left(R_1^i = 1 \,\&\, R_2^j = 1|\mathbf{X}\right)$ by averaging over the stimulus, one can obtain estimates of all the quantities in Eq. (1). By subtracting the estimates of both sides of Eq. (1), one forms the joint peristimulus time histogram (JPSTH) [16, 1]. This observation is a simple proof that the JPSTH and the shuffle-corrected correlogram (the sum along the diagonals of the JPSTH) are SICMs [15, 2].

A further challenge is to develop SICMs for experiments, such as white noise and related experiments, where one does not repeat a particular realization of the random stimulus. In such experiments, one is sampling a large stimulus space and would like to avoid repeats that, for a given experiment duration, would reduce the size of the sampled space and increase the noise. Unfortunately, when the stimulus is not repeated, one cannot estimate the probabilities in Eq. (1) (other than by the value of the one sample recorded) without specifying additional assumptions on how the neurons' responses depend on the stimulus. For this reason, one cannot derive *model-free* SICMs for such experiments.

One could develop a strategy to remove stimulus-dependent correlations based on the model that stimulus-dependent effects occur at a different time scale than the pertinent correlations. If this model were correct, one could filter out the stimulus effects with a bandpass filter [17]. But, not only won't the resulting measure necessarily be a SICM, the filtering might also remove stimulus-independent correlations.

One could also attempt to develop a modification of the JPSTH based on

the model that a neuron's response is primarily determined by the stimulus during a small window in the past. When one uses a random sequence of a discrete number of stimuli, this model implies that the neuron's response depends only on the previous few stimuli. One could theoretically form a JPSTH for each distinct combination of such stimuli, and this JPSTH would be a SICM, assuming the model were correct.

However, when the stimulus is multidimensional, the number of possible sequences is so large that each combination will rarely occur in a realistically long experiment. For example, even in the subspace experiments designed by Ringach et al. [20], where the stimulus is a sequence of random gratings, one would typically have over $10^8$ possible combinations. With fewer than $10^6$ presentations in a typical experiment, most combinations would not be presented even once[1] [19].

Using standard white noise analysis as a starting point, we demonstrate two SICMs that are based on a simple phenomenological model of a linear-nonlinear system. As with any model, this framework limits the applicability of this analysis to neurons that can be approximated by the model. The motivation for starting with the linear-nonlinear model is its simplicity and the fact that the linear-nonlinear model is commonly used to analyze neurons in the auditory, visual, and somatosensory systems [17, 12, 6, 7, 10, 11, 4, 5, 18, 20, 9, 8]. We demonstrate that the proposed SICMs also give good results with simulations of integrate-and-fire neurons driven like linear-nonlinear systems, indicating broader applicability of the SICMs than to just linear-nonlinear systems. We also demonstrate that modifications are needed to broaden the applicability of the approach since the current form of the method fails when the neural response is fundamentally nonlinear.

## 2 SICMs for linear-nonlinear systems

Let the stimulus $\mathbf{X}$ be a vector of independent Gaussian random variables with zero mean and standard deviation $\sigma$. Without loss of generality, let $\sigma = 1$. The stimulus $\mathbf{X}$ is a discrete approximation to temporal or spatio-temporal white noise.

[1] If one used only 500 different gratings and assumed the neuron responded to only the previous three, one would have $500^3 \approx 10^8$ combinations. If one selected a random grating 50 times a second for an hour, one would have a total of $60^2 \times 50 = 1.8 \times 10^5$ presentations of a grating.

Assume the neurons' responses are independent given the stimulus. Let the probability of a spike of neuron $p$ for $p = 1, 2$ be a linear-nonlinear function of the input,[2]

$$\Pr\left(R_p^i = 1 \middle| \mathbf{X} = \mathbf{x}\right) = g_p\left(\bar{\mathbf{h}}_p^i \cdot \mathbf{x}\right) \tag{2}$$

where $\bar{\mathbf{h}}_p^i$ is the linear kernel of neuron $p$ shifted $i$ units in time (normalized so that $\|\bar{\mathbf{h}}_p^i\| = 1$), and $g_p(\cdot)$ is its output nonlinearity (representing, for example, its spike generating mechanism). See, for example, Ref. [3] for a discussion of the linear-nonlinear model.

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

$$g_p(x) = \frac{\hat{r}_p}{2}\left[1 + \mathrm{erf}\left(\frac{x - \bar{T}_p}{\bar{\epsilon}_p\sqrt{2}}\right)\right], \tag{3}$$

where $\hat{r}_p$ is the maximum firing rate,[3] $\bar{T}_p$ is the threshold, $\bar{\epsilon}_p$ defines the steepness of the nonlinearity, and $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

We chose an error function nonlinearity so that the results could be computed analytically. For other forms of nonlinearities, one could compute numerical results similar to those in Ref. [14]. However, given the low firing rates of neurons in response to white noise stimuli, the precise form of the nonlinearity appears to make little difference [13].

## 2.1 First SICM for linear-nonlinear system

We sketch the derivation of a SICM for the linear-nonlinear systems. Throughout, we use the notation $\langle\cdot\rangle$ to indicate averaging over a (finite) realization of the white noise stimulus and $E\{\cdot\}$ to indicate expected value over all possible realizations. (Note that since the stimulus and the neural response are stationary, the expected values will typically depend on one fewer temporal index than our notation implies.)

[2] We use an overbar to denote model parameters and reserve the unbarred notation for estimates of the parameters from measured statistics.

[3] We allow the maximal firing rate $\hat{r}_p$ to differ from the reciprocal of the temporal discretization interval so that the nonlinearity is not tied to the choice of discretization. The model allows for interspike intervals shorter than $1/\hat{r}_p$, consistent with the fact that the linear-nonlinear model as written implicitly assumes spike trains that are a modulated Poisson process, at least in the limit of small temporal discretization.

It turns out that with a white noise stimulus, the only important geometry of the kernels $\bar{\mathbf{h}}_p^i$ is their inner products (at different temporal shifts of $k$ units of time), $\bar{\mathbf{h}}_p^{i-k} \cdot \bar{\mathbf{h}}_q^i$. Since the kernels are normalized to unit length, their inner products lie between $-1$ and 1 and can be considered a cosine of an angle between the vectors,[4]

$$\cos \bar{\theta}_{pq}^k = \bar{\mathbf{h}}_p^{i-k} \cdot \bar{\mathbf{h}}_q^i. \tag{4}$$

When $\cos \bar{\theta}_{21}^k$ is large, neuron 1 responds to similar stimulus features that neuron 2 responded to $k$ time steps earlier. Small $\cos \bar{\theta}_{21}^k$ indicates a response to dissimilar stimulus features.

In our notation, the measured correlation of a spike in neuron 1 and a spike in neuron 2 occurring $k$ time steps earlier is given by $\langle R_1^i R_2^{i-k} \rangle$. When neural response is given by Eqs. (2) and (3), one can show [13] that

$$E\{\langle R_1^i R_2^{i-k} \rangle\} = \frac{\hat{r}_1 \hat{r}_2}{4} \mathrm{derfc}\Big( \frac{\bar{\delta}_1 \bar{T}_1}{\sqrt{2}}, \frac{\bar{\delta}_2 \bar{T}_2}{\sqrt{2}}, \bar{\delta}_1 \bar{\delta}_2 \cos \bar{\theta}_{21}^k \Big) \tag{5}$$

where

$$\bar{\delta}_p = 1/\sqrt{1 + \bar{\epsilon}_p^2}, \tag{6}$$

and the double complementary error function is defined by the intimidating-looking formula

$$\mathrm{derfc}(a, b, c) = \frac{2}{\sqrt{\pi}} \int_a^\infty e^{-y^2} \mathrm{erfc}\Big( \frac{b - cy}{\sqrt{1 - c^2}} \Big) dy. \tag{7}$$

In fact, the double complementary error function is simply a two-dimensional analogue of the familiar complementary error function $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$.

We assert, leaving the proof to the reader, that the double complementary error function $\mathrm{derfc}(a, b, c)$ is the area under the tail section graphed in Fig. 2 of the surface

$$\frac{4}{\pi} \exp(-[x_1^2 + x_2^2]). \tag{8}$$

[4] With this normalization and a white noise stimulus, the nonlinearity arguments in Eq. (2) are normal random variables with mean zero and variance one, and their covariances are $\cos \bar{\theta}_{pq}^k$.

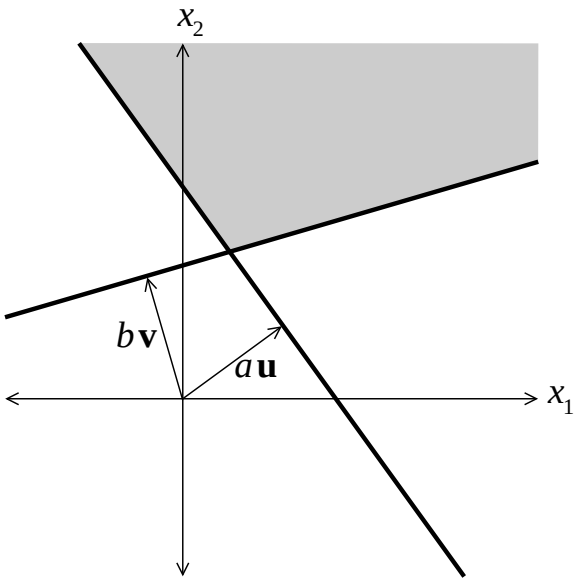


Figure 2: Sketch of (a finite portion of) the tail section defining the double complementary error function. Letting $\mathbf{x}$ be the vector $(x_1, x_2)$, the required tail section (shaded) is defined by $\mathbf{x} \cdot \mathbf{u} > a$ and $\mathbf{x} \cdot \mathbf{v} > b$ where $\mathbf{u}$ and $\mathbf{v}$ can be any unit vectors with $\mathbf{u} \cdot \mathbf{v} = c$ ($c$ is cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$). The thick boundary lines are perpendicular to $\mathbf{u}$ and $\mathbf{v}$ and are a distance of $a$ and $b$ from the origin, respectively. From the radial symmetry of the surface (8), the area depends only on $a$, $b$, and $c$ (and not on the choice of unit vectors $\mathbf{u}$ and $\mathbf{v}$).

Note that $\mathrm{derfc}(a, b, c)$ is symmetric in $a$ and $b$. When $c = 0$, the boundaries of the tail section are perpendicular, and the double complementary error function separates into a product of complementary error functions:

$$\mathrm{derfc}(a, b, 0) = \mathrm{erfc}(a)\mathrm{erfc}(b).$$

This property of the double complementary error function yields an important consequence for the correlation given by Eq. (5). If the two linear kernels happen to be orthogonal at a particular shift $k$ so that $\cos\bar{\theta}_{21}^{k} = 0$, then

$$E\{\langle R_1^i R_2^{i-k}\rangle\} = \Big[\frac{\hat{r}_1}{2}\mathrm{erfc}\Big(\frac{\bar{\delta}_1\bar{T}_1}{\sqrt{2}}\Big)\Big]\Big[\frac{\hat{r}_2}{2}\mathrm{erfc}\Big(\frac{\bar{\delta}_2\bar{T}_2}{\sqrt{2}}\Big)\Big]. \tag{9}$$

Since, as previously shown [14], the expected mean rate of neuron $p$ is

$$E\{\langle R_p^i\rangle\} = \frac{\hat{r}_p}{2}\mathrm{erfc}\Big(\frac{\bar{\delta}_p\bar{T}_p}{\sqrt{2}}\Big),$$

the expected value of the correlation in this case is simply the product of the mean rates

$$E\{\langle R_1^i R_2^{i-k}\rangle\} = E\{\langle R_1^i\rangle\}E\{\langle R_2^{i-k}\rangle\}. \tag{10}$$

The covariance between the response of one neuron and the (delayed) response of the other is

$$\mathcal{C}^k = \langle R_1^i R_2^{i-k}\rangle - \langle R_1^i\rangle\langle R_2^{i-k}\rangle. \tag{11}$$

The covariance is reminiscent of the shuffle-corrected correlogram. Even though the averaging $\langle\rangle$ is over differing stimuli, one might expect that $\mathcal{C}^k$ is close to a SICM since the realizations of the stationary stimulus are statistically identical. Indeed, Eq. (10) implies that the expected value of $\mathcal{C}^k$ is zero (at least for large samples) whenever the kernels are orthogonal.

However, the covariance is not a SICM when the kernels are not orthogonal. By taking a Taylor series of $\mathrm{derfc}(a, b, c)$ around $c = 0$, one can see that the expected value of $\mathcal{C}^k$ is nearly proportional to $\cos\bar{\theta}_{21}^{k}$, at least for small values of the cosine. When the kernels are orthogonal, neuron 1 and neuron 2 are responding to dissimilar features of the stimulus. In this case, there are no stimulus-dependent correlations and its not surprising that the covariance is zero. Whenever the neurons respond to similar stimulus features ($\cos\bar{\theta}_{21}^{k} \neq 0$), the covariance is nonzero.

We propose a correlation measure that takes into account the full correlation calculation in Eq. (5). As described in Ref. [14], one can calculate all the parameters of the model (Eqs. (2), (3) and (6)) from measurable statistics. The only statistics we need to measure are the mean rates $\langle R_p^i \rangle$ and the stimulus-spike correlations $\langle \mathbf{X} R_p^i \rangle$. The kernel for neuron $p$ is proportional to the stimulus-spike correlation in the large sample limit, so we estimate it by

$$\mathbf{h}_p^i = \frac{\langle \mathbf{X} R_p^i \rangle}{|\langle \mathbf{X} R_p^i \rangle|}.$$

We determine $T_p$ and $\delta_p$ (estimates of $\bar{T}_p$ and $\bar{\delta}_p$) from the magnitude of the stimulus-spike correlation $|\langle \mathbf{X} R_p^i \rangle|$ and the mean rate $\langle R_p^i \rangle$ using

$$\langle R_p^i \rangle = \frac{\hat{r}_p}{2} \mathrm{erfc}\Big( \frac{\delta_p T_p}{\sqrt{2}} \Big) \tag{12}$$

and

$$|\langle \mathbf{X} R_p^i \rangle| = \frac{\hat{r}_p \delta_p}{\sqrt{2\pi}} \exp\Big( -\frac{\delta_p^2 T_p^2}{2} \Big). \tag{13}$$

We then define the following correlation measure

$$\mathcal{S}^k = \langle R_1^i R_2^{i-k} \rangle - \nu_{21}^k \tag{14}$$

where $\nu_{21}^k$ is the expected value of $\langle R_1^i R_2^{i-k} \rangle$ given the measured parameters,

$$\nu_{pq}^k = \frac{\hat{r}_p \hat{r}_q}{4} \mathrm{derfc}\Big( \frac{\delta_p T_p}{\sqrt{2}}, \frac{\delta_q T_q}{\sqrt{2}}, \delta_p \delta_q \cos\theta_{pq}^k \Big), \tag{15}$$

and $\cos\theta_{pq}^k$ is the estimate of the kernel inner product $\mathbf{h}_p^{i-k} \cdot \mathbf{h}_q^i$,

$$\cos\theta_{pq}^k = \frac{\langle \mathbf{X} R_p^{i-k} \rangle \cdot \langle \mathbf{X} R_q^i \rangle}{|\langle \mathbf{X} R_p^{i-k} \rangle| |\langle \mathbf{X} R_q^i \rangle|}. \tag{16}$$

(We assume $\hat{r}_1$ and $\hat{r}_2$ are known from other considerations, cf. Ref. [14].)

Eq. (5) implies that $E\{\mathcal{S}^k\} \to 0$ as the sample size increases. $\mathcal{S}$ is our first SICM for linear-nonlinear systems.

## 2.2 Second SICM for linear-nonlinear system

One can further exploit the linear-nonlinear framework to characterize the correlations in $\mathcal{S}$ if one makes further assumptions about the source of the correlations. We demonstrate an example where the two linear-nonlinear neurons are coupled together in a caricature of synaptic coupling. In response to a spike of neuron $q$, the spiking probability of neuron $p$ after $i$ time steps is modified according to the factor $\bar{W}^j_{qp}$. The quantity $\bar{W}^j_{qp}$ is simply added underneath the nonlinearity.

The resulting model of the coupled linear-nonlinear systems is

$$\Pr\left(R^i_p = 1 \middle| \mathbf{X} = \mathbf{x}, \mathbf{R}_q = \mathbf{r}_q\right) = g_p\Big(\mathbf{h}^i_p \cdot \mathbf{x} + \sum_{j\geq 0} \bar{W}^j_{qp} r^{i-j}_q\Big), \tag{17}$$

for $p, q \in \{1, 2\}$, $q \neq p$. We sketch a derivation of a SICM that approximates the coupling terms $\bar{W}^j_{12}$ and $\bar{W}^j_{21}$. This SICM gives an intuitive interpretation of the correlations in $\mathcal{S}$. Moreover, it normalizes the correlations into units of the stimulus standard deviation and removes the temporal filtering of the correlations caused by the structure of the kernels. These properties are demonstrated in section 3.

The derivation of this SICM is given in Ref. [13]. Here, we simply describe the ideas underlying the derivation. The main assumption is that the coupling terms $\bar{W}^j_{12}$ and $\bar{W}^j_{21}$ are relatively small so that we can neglect terms with quadratic and higher powers of the coupling terms. We can then compute an approximation of the expected correlation $E\{\langle R^i_1 R^{i-k}_2\rangle\}$ that is linear in the $\bar{W}$.

From the previous section, we know that the expected value of $\langle R^i_1 R^{i-k}_2\rangle - \nu^k_{21}$ is zero when all $\bar{W}^j_{pq}$ are zero. Since we are linearizing with respect to the $\bar{W}^j_{pq}$, this expected value must be a homogeneous linear function of the $\bar{W}^j_{pq}$, which we write as

$$E\{\langle R^i_1 R^{i-k}_2\rangle - \nu^k_{21}\} \approx \sum_{j\geq 0} A^{kj}_{21} \bar{W}^j_{21} + \sum_{j\geq 0} A^{-kj}_{12} \bar{W}^j_{12}, \tag{18}$$

where $\approx$ indicates the equality within $O((\bar{W}^j_{pq})^2)$ as the sample size increases. If we can determine the values of the $A^{kj}_{pq}$, we are nearly done, needing only to solve Eq. (18) for the $\bar{W}^j_{pq}$.

To determine the $A^{kj}_{pq}$, we calculate $T_p$, $\delta_p$, and $\cos\theta^k_{pq}$ using Eqs. (12), (13), and (16) as before. In this case, these quantities include the effects of

the coupling and are no longer estimates of the model parameters $\bar{T}_p$, $\bar{\delta}_p$, and $\cos\bar{\theta}^k_{pq}$. Instead, $T_p$, $\delta_p$, and $\cos\theta^k_{pq}$ measure the effective parameters for the system treated as uncoupled linear-nonlinear neurons of the form given by Eq. (2). These effective parameters are useful because, if no $\cos\theta^j_{12}$ is close to one or if $\delta_1$ and $\delta_2$ are not close to one, then we can find an approximate formula for the $A$'s in terms of $T_p$, $\delta_p$, and $\cos\theta^k_{pq}$. For completeness, we write the (complicated) formula in Appendix A.

Therefore, if one knows $\hat{r}_p$ and measures $\langle R^i_p\rangle$, $\langle \mathbf{X}R^i_p\rangle$, and $\langle R^i_1 R^{i-k}_2\rangle$ in response to a white noise stimulus, where $k = -N,\ldots,N$, all unknowns from Eq. (18) are determined except $\bar{W}^j_{pq}$. One has $2N-1$ equations for $2N$ unknowns ($\bar{W}^j_{12}$, $\bar{W}^j_{21}$, $j = 0,\ldots N$).

Since our system has more unknowns than equations, we reduce the number of unknowns to $2N-1$ by merging the zero-delay interactions into one quantity, setting $\bar{W}^0_{21} = \bar{W}^0_{12}$. To reflect this reduction and to simplify the notation, we define a total connectivity $\bar{W}$ by

$$\bar{W}^j = \begin{cases} \bar{W}^{-j}_{12} & \text{for } j < 0, \\ \bar{W}^0_{12} + \bar{W}^0_{21} & \text{for } j = 0, \\ \bar{W}^j_{21} & \text{for } j > 0. \end{cases} \tag{19}$$

Similarly, we define a new matrix $A$ by

$$A^{kj} = \begin{cases} A^{k,-j}_{12} & \text{for } j < 0, \\ (A^{k0}_{12} + A^{k0}_{21})/2 & \text{for } j = 0, \\ A^{kj}_{21} & \text{for } j > 0. \end{cases} \tag{20}$$

In this notation, Eq. (18) simplifies to

$$E\{\langle R^i_1 R^{i-k}_2\rangle - \nu^k_{21}\} \approx \sum_j A^{kj}\bar{W}^j. \tag{21}$$

We can now define our second SICM $\mathcal{W}$, which approximates $\bar{W}^j$, by

$$\mathcal{W} = A^{-1}\mathcal{S}, \tag{22}$$

where we use matrix-vector notation and $A^{-1}$ indicates the matrix inverse of $A$. The fact that $\mathcal{W}$ is a SICM is clear because $\mathcal{S}$ is a SICM. Since $E\{\mathcal{W}^j\} \to \bar{W}^j$ (approximately) as the sample size increases, $\mathcal{W}$ can be viewed as a measure of the effective connectivity of the neurons (assuming, of course, that we somehow knew the stimulus-independent correlations were due to mutual coupling).

# 3 Results

We demonstrate the SICMs $\mathcal{S}$ and $\mathcal{W}$ by comparing them to the covariance $C$ from Eq. (11). We first simulate pairs of linear-nonlinear neurons. Then, we simulate pairs of integrate-and-fire neurons to see how well the analysis generalizes to more realistic neuron models.

## 3.1 Linear-nonlinear neuron results

For each example, we simulate two linear-nonlinear neurons with firing probabilities given by Eqs. (17) and (3) with nonlinearity parameters $\hat{r}_1 = \hat{r}_2 = 1$, $\bar{T}_1 = 2$, $\bar{T}_2 = 2.5$, $\epsilon_1 = 0.5$, and $\epsilon_2 = 1.0$. These parameters yield mean spike rates between 0.035 and 0.04 spikes per unit time in response to spatio-temporal white noise.

The linear kernels used in the simulations are two-dimensional in spatial coordinates $\mathbf{j} = (j_1, j_2)$ for each time $t$. For neuron $p = 1, 2$, we used the form

$$\bar{\mathbf{h}}_p(\mathbf{j}, t) = (t - b_p) e^{-\frac{(t-b_p)}{\tau_h}} e^{-\frac{|\mathbf{j}|^2}{40}} \sin((j_1 \cos\phi_p + j_2 \sin\phi_p) f) \qquad (23)$$

for $t > b_p$ and $\bar{\mathbf{h}}_p(\mathbf{j}, t) = 0$ otherwise. We sampled this function on a $20 \times 20 \times 20$ grid in space and time, normalizing the resulting vector to obtain the unit vector $\bar{\mathbf{h}}_p^i$. Units of $t$ and $\mathbf{j}$ are in grid points. We fixed the parameters $f = 0.6$, $b_1 = 0$, $b_2 = 3$, and $\phi_1 = 0$.

The kernels were chosen to be caricatures of the linear receptive fields measured in visual simple cells. As stated in section 2.1, the only important geometry of the kernels is their inner products $\cos\bar{\theta}_{pq}^k = \bar{\mathbf{h}}_p^{i-k} \cdot \bar{\mathbf{h}}_q^i$. In the example simulations, we vary $\tau_h$ and $\phi_2$ to alter the form of $\cos\bar{\theta}_{pq}^k$.

For each simulation, we measure the spikes in response to white noise and calculate the three correlation measures $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{W}$. Each kernel inner product was estimated using bias reduction methods similar to that described in Ref. [14]. For each correlation measure, we also estimate confidence intervals as detailed in Appendix B to provide a gauge on the significance of nonzero values. These confidence intervals approximate one standard error of the measurements. These estimates tend to overestimate the variability compared to estimates from repeated identical simulations (a luxury we have only with simulations), so care should be exercised in their interpretation.

In the first example, we let $\tau_h = 1$ and $\phi_2 = \pi/8$, and keep the neurons uncoupled ($\bar{W}^j = 0$ for all $j$). We simulated the neurons for 100,000 units of

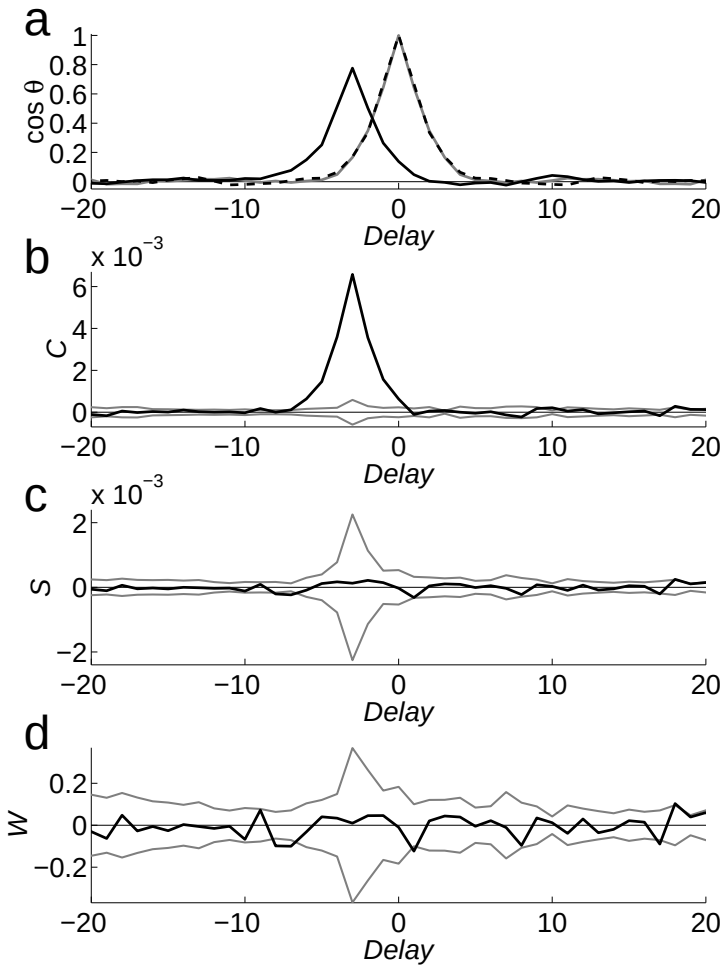


Figure 3: Correlation measures for a pair of uncoupled neurons with similar linear kernels. **(a)** The kernel inner products $\cos\theta_{21}$ (solid black line), $\cos\theta_{11}$ (dashed black line), and $\cos\theta_{22}$ (solid gray line). By definition $\cos\theta^k_{11}$ and $\cos\theta^k_{22}$ are symmetric and one at zero delay ($k = 0$). **(b)** Covariance (black line) showing the high correlation when neuron 2 fires three units of time after neuron 1. At this delay, the kernels match closely ($\cos\theta_{21}$ is large). In panels (b)–(d), the solid gray lines approximate confidence intervals of $\pm 1$ standard error of the measure. Covariance is in units of (spikes/unit time)$^2$. **(c)** The SICM $\mathcal{S}$ (black line) eliminates the stimulus-induced correlation, showing that the neurons are uncoupled. $\mathcal{S}$ is in units of (spikes/unit time)$^2$. **(d)** The SICM $\mathcal{W}$ (black line) also eliminates the stimulus-induced correlation. $\mathcal{W}$ is in units of the stimulus standard deviation. Delay is in units of time and is the spike time of neuron 1 minus the spike time of neuron 2.

time, obtaining 3,000–4,000 spikes per neuron. As shown in Fig. 3a, $\cos\theta_{21}^k$ nearly reaches 0.8 for $k = -3$ because neuron 1 responds to nearly the same stimulus features as neuron 2 did 3 time steps previous. As a consequence, the spike times of the neurons are highly correlated (Fig. 3b) even though the responses are independent when conditioned on the stimulus. This correlation is simply due to the similarity of the two kernels and is completely captured by the angle between them, $\cos\theta_{21}^k$.

The fact that the correlations were due simply to a shared stimulus is revealed by the SICMs $\mathcal{S}$ and $\mathcal{W}$, graphed in Fig. 3c–d. Both $\mathcal{S}$ and $\mathcal{W}$ completely eliminate the correlations. They are zero within the level of the noise as estimated by the confidence intervals.

Next, we let the neurons be coupled by mutual inhibition with a delay of 3 units of time by setting $\bar{W}^j = -0.3$ for $j = -3, 3$. The results from simulating 100,000 units of time, obtaining 3,000–4,000 spikes per neuron, are shown in Fig. 4. As shown in the plot of the covariance $\mathcal{C}$ (Fig. 4b), the connectivity is masked by the stimulus-induced component of the covariance. The inhibition from neuron 1 to neuron 2 occurs at the peak of the stimulus-induced correlation. Although it reduces the size of the stimulus-induced peak, the presence of the inhibition cannot be deduced from the covariance alone. The inhibition from neuron 2 to neuron 1 is visible as a small dip at $t = 3$ but is dwarfed by the peak at $t = -3$. A naive interpretation of the covariance would lead to misleading inferences on the underlying neural circuit.

Both SICMs better reflect the simulated connectivity $\bar{W}^j$ as shown in Fig. 4c–d. The stimulus-induced peak is eliminated, and the measures have two sharp dips at both $t = -3$ and $t = 3$. The dips in $\mathcal{S}$ give the misleading impression that the connectivity at $t = -3$ is stronger. The dips in $\mathcal{W}$, however, are nearly identical. Their magnitudes are even near the simulated values of $-0.3$, though they slightly underestimate the magnitude of the mutual inhibition.

The connectivity-induced dips in $\mathcal{S}$ and $\mathcal{W}$ are only slightly beyond the level of the noise. To demonstrate more clearly the significance of these dips, we doubled the simulation length to 200,000 units of time, obtaining roughly 7,000 spikes per neuron. With more data, the noise level is reduced, and the connectivity-induced dips are well beyond the noise (Fig. 5).

A third example demonstrates how $\mathcal{W}$ accurately reconstructs the connectivity even when the correlation between the spikes is distorted by the temporal structure of $\bar{\mathbf{h}}_p^i$. We increase $\tau_h$ to 5 in order to increase the effect

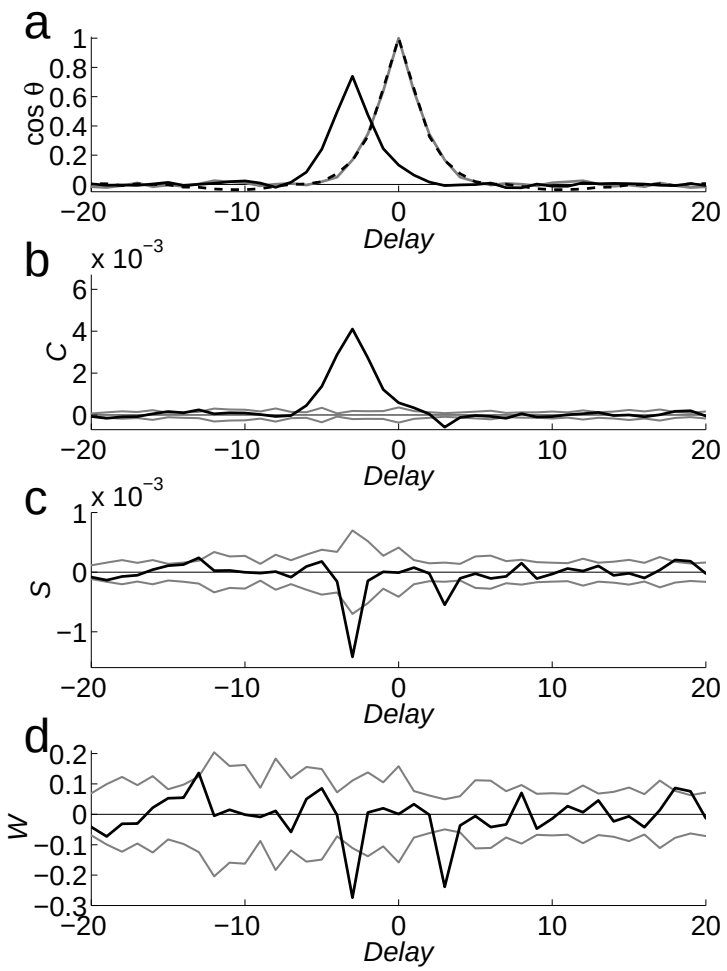


Figure 4: Correlation measures for a pair of mutually inhibited neurons with similar linear kernels. Except for negative coupling of magnitude $-0.3$ at delays of $-3$ and 3, the neurons are identical to those from Fig. 3. Panels are as in Fig. 3. **(a)** The measured kernel inner products are changed little by the connections. **(b)** The stimulus-induced correlation masks the coupling-induced correlation in the covariance. **(c)** Having eliminated the stimulus-induced correlation, the SICM $\mathcal{S}$ reveals the negative coupling between the neurons, although it makes them appear unequal. **(d)** The SICM $\mathcal{W}$ not only eliminates the stimulus-induced correlation but also estimates the magnitudes of the negative coupling relatively accurately.

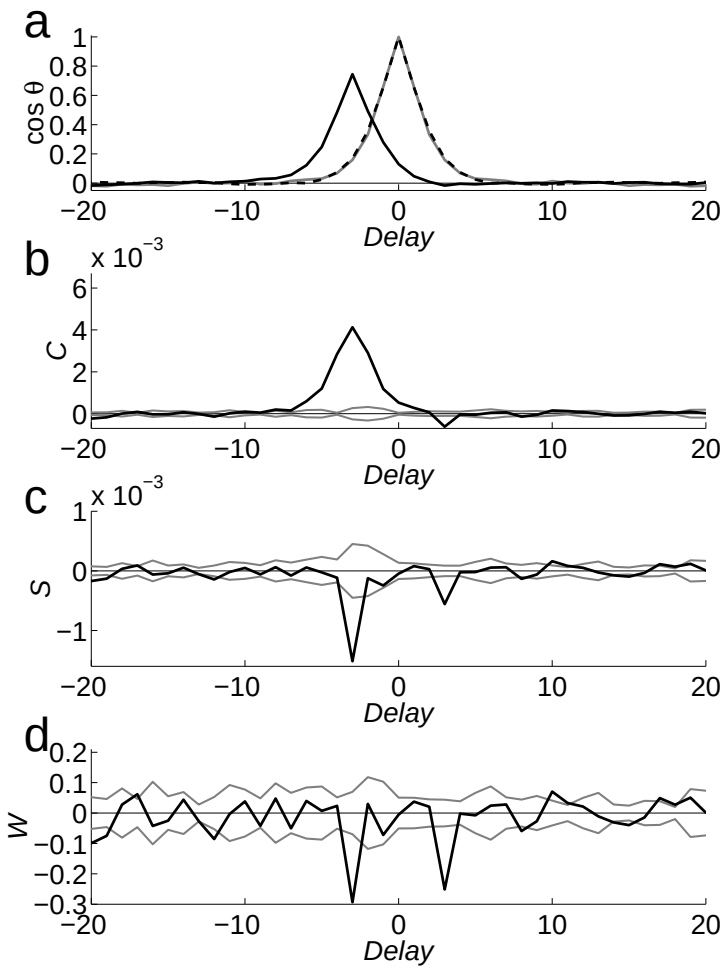


Figure 5: Identical simulation as in Fig. 4 but with a simulation that is twice as long. The negatively coupling estimated by $\mathcal{S}$ and $\mathcal{W}$ is more clearly seen to be significant. Panels are as in Fig. 3.

and set $\phi_2 = \pi/2$ so that the kernels are effectively orthogonal.[5] In this case, the covariance (as well as the SICMs) will be approximately zero when $\bar{W}^j = 0$ because $\cos\theta_{21}^j \approx 0$ (not shown).

We couple the neurons by mutual excitation ($\bar{W}^j = 0.4$ for $j = -3, 3$). To emphasize the relatively subtle effect, we run a longer simulation of 300,000 units of time, obtaining approximately 11,000–12,000 spikes per neuron. The results are shown in Fig. 6. The inner products of the kernels with themselves, $\cos\theta_{11}^j$ and $\cos\theta_{22}^j$ (Fig. 6a), decay only slowly to zero as the shift $j$ is increased. These broad peaks reflect the broad temporal structure of the kernels that will filter connectivity-induced correlations.

Indeed, the covariance shows not only the sharp peaks at $t = -3, 3$ corresponding to the connectivity but also a wide peak centered around $t = 0$ (Fig. 6a). The structure of the wide peak corresponds to the structure in $\cos\theta_{11}^j$ and $\cos\theta_{22}^j$.

The SICM $\mathcal{S}$ reduces some of the wide peak (Fig. 6c). It differs from $\mathcal{C}$ even though $\bar{\mathbf{h}}_1^i$ is approximately orthogonal to $\bar{\mathbf{h}}_2^{i-k}$ only because the mutual excitation increased the measured $\cos\theta_{21}^k$ (Fig. 6a). Though reduced in magnitude, the broad peak is still significant, rising above the confidence intervals around zero delay. The advantage of the SICM $\mathcal{W}$ is that it eliminates the broad peak caused by the filtering. $\mathcal{W}$ even comes down to zero between the two sharp peaks (Fig. 6d).

## 3.2 Tests with integrate-and-fire neurons

### 3.2.1 Integrate-and-fire neuron model

The SICMs we derived were based on the simple linear-nonlinear model. To test the robustness of the measures to deviations from the linear-nonlinear model, we simulated a pair of integrate-and-fire neurons. The evolution of the voltage of neuron $p$ in response to input $g_p(t)$ was given by

$$\tau_m \frac{\mathrm{d}V_p}{\mathrm{d}t} + (V_p - \mathcal{E}_r) + g_p(t)(V_p - \mathcal{E}_s) = 0 \tag{24}$$

[5] The fact that the spatial axes are orthogonal ($\phi_1 = 0$ and $\phi_2 = \pi/2$) does not necessarily imply that the resulting kernels would be orthogonal as vectors ($\bar{\mathbf{h}}_1^i \cdot \bar{\mathbf{h}}_2^{i-k} = 0$). In this case, they are orthogonal because one kernel is odd-symmetric along the same axis that the other kernel is even-symmetric. If, for example, the spatial phase were changed (say, from a sine to a cosine), the inner product would increase.

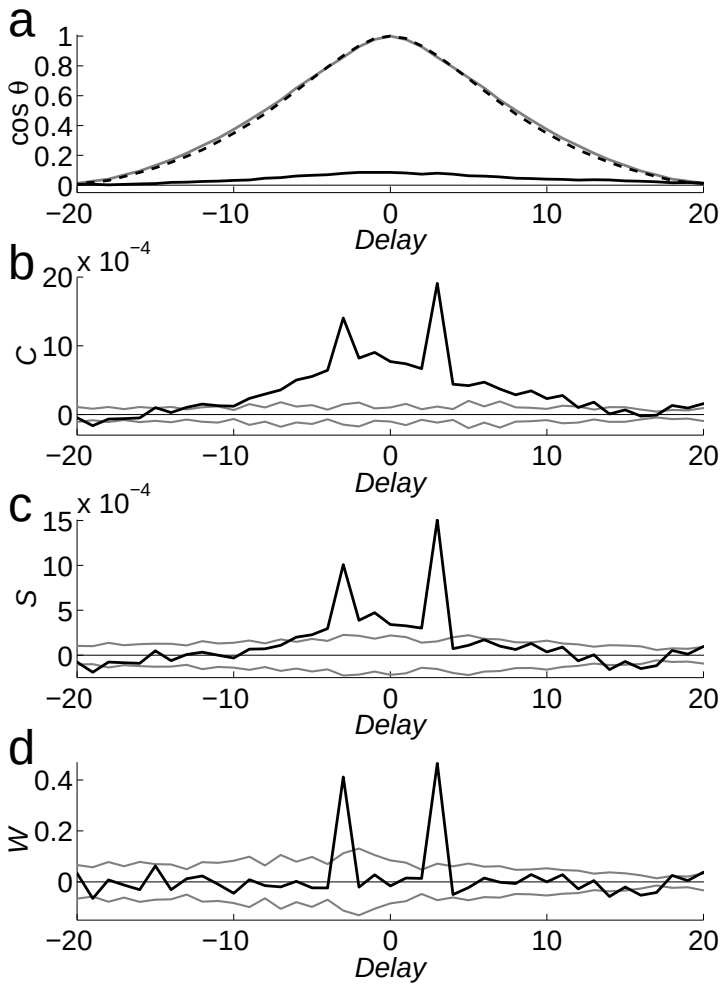


Figure 6: Correlation measures for a pair of neurons with dissimilar linear kernels that are coupled with mutual excitation of magnitude 0.4 at delays of $-3$ and 3. Panels are as in Fig. 3. **(a)** The broad temporal structure of the linear kernels leads to wide peaks in $\cos\theta_{11}$ and $\cos\theta_{22}$. The coupling has introduced a small broad peak in $\cos\theta_{21}$. **(b)** The correlations induced by the connectivity are filtered by the temporal structure of the linear kernels, creating a broad peak in the covariance centered around zero delay. **(c)** The broad peak is smaller in $\mathcal{S}$ but still present. **(d)** $\mathcal{W}$ eliminates the temporal smearing, faithfully reproducing the excitatory coupling.

with membrane time constant $\tau_m$ and equilibrium potentials $\mathcal{E}_r = -65$ mV and $\mathcal{E}_s = 0$ mV. A spike was recorded at times $T_p^j$ when $V_p(T_p^j)$ reached 55 mV. After the spike, the voltage was reset to $-65$ mV and held there for an absolute refractory period of $\tau_{ref} = 2$ ms.

The input conductance $g_p(t)$ is unitless because we have divided through by the resting conductance. The input conductance to neuron $p$ was composed of internal input from the other neuron $g_p^{int}(t)$ and external input $g_p^{ext}(t)$,

$$g_p(t) = g_p^{int}(t) + g_p^{ext}(t).$$

The internal input to neuron 2 was set to zero, $g_2^{int}(t) \equiv 0$. The internal input to neuron 1 was set to reflect an excitatory connection from neuron 2 with a delay of $\Delta = 40$ ms,

$$g_1^{int}(t) = \bar{g}_{int} \sum_j G(t - T_2^j - \Delta), \tag{25}$$

where $\bar{g}_{int}$ gives the maximum of a single unitary conductance. The unitary conductance waveform was given by

$$G(t) = \begin{cases} \frac{e^2}{4}\left(\frac{t}{\tau_s}\right)^2 e^{-t/\tau_s} & \text{for } t > 0, \\ 0 & \text{otherwise,} \end{cases}$$

with time constant $\tau_s$. $G(t)$ is defined so that its maximum is 1 (at $t = 2\tau_s$).

The external input was given by

$$g_p^{ext}(t) = \bar{g}_{ext} \sum_j G(t - T_{ext,p}^j)$$

where $\bar{g}_{ext}$ gives the maximum of a single unitary conductance. The $T_{ext,p}^j$ were drawn from a modulated Poisson process with rate given by a function of the input $r_p(\mathbf{X})$.

In most examples, we set the input rate so that the integrate-and-fire neurons were driven like linear-nonlinear neurons. In particular, the input rates were threshold linear function of the input

$$r_p(\mathbf{X}) = \alpha \left[\bar{\mathbf{h}}_p^i \cdot \mathbf{X}\right]^+ \tag{26}$$

where $\alpha$ determines the average input rate, and $[x]^+ = x$ if $x > 0$ and is zero otherwise. We used the same kernels $\bar{\mathbf{h}}_p^i$ as in the previous section (Eq.

(23)), only with $b_2 = 40$ ms, $\tau_h = 10$ ms, and $\phi_2 = \pi/6$ (all other parameters were unchanged). We sampled the kernel using a temporal discretization of 2 ms, keeping 50 grid points in time and a $20 \times 20$ grid in space.

To demonstrate the limitations of the method when neurons are not driven like linear-nonlinear systems, we simulated an example where the input rate was a fundamentally nonlinear function of the input

$$r_p(\mathbf{X}) = \alpha^+ \left[\bar{\mathbf{h}}_p^i \cdot \mathbf{X}\right]^+ + \alpha^- \left[-\bar{\mathbf{h}}_p^i \cdot \mathbf{X}\right]^+. \tag{27}$$

When $\alpha^+$ and $\alpha^-$ are similar in magnitude, a large component of the input is independent of the sign of the stimulus components. This sign independence is impossible with a linear-nonlinear system, and we would not expect the current form of the method to perform well for such a neuron.

We set the stimulus discretization to be 10 ms, slower than the temporal discretization since a monitor frame rate of 500 $s^{-1}$ is unrealistically fast. We decreased the stimulus magnitude so that the discrete white noise still had an effective power of 1 at the temporal discretization of 2 ms.[6] Assuming we don't attempt to resolve temporal structure finer than 10 ms, this deviation from the model assumptions has little effect.

The evolution of the integrate-and-fire neurons was simulated using a second-order Runge-Kutta method modified for integrate-and-fire neurons [21], obtaining five digit accuracy with a time step of 0.05 ms.

### 3.2.2 Integrate-and-fire neuron results

To test the method for a wide variety of integrate-and-fire parameters, we simulated the pair of neurons using six different combinations of membrane time constant $\tau_m$ and synaptic time constant $\tau_s$, as shown in Table 1. For each time constant combination, we set the coupling strength $\bar{g}_{int}$ so that the magnitude of the stimulus-induced correlation and the magnitude of the connectivity-induced correlation were approximately equal (as measured by the covariance $\mathcal{C}$).

For each time constant combination, we ran simulations with both small and large external unitary conductances $\bar{g}_{ext}$. For small $\bar{g}_{ext}$, we set $\alpha = 1000$ $s^{-1}$ so that the average input rates was nearly 400 inputs per second. We adjusted $\bar{g}_{ext}$ so that the average firing rate of the neurons was between 5 and

[6] To compensate for the stimulus that was 5 times slower than the temporal discretization, we let the standard deviation of each stimulus component be $\sqrt{1/5}$.

| $\tau_m$ | $\tau_s$ | $\bar{g}_{int}$ | $\alpha$ | $\bar{g}_{ext}$ |
|---|---|---|---|---|
| 5 ms | 8 ms | 0.04 | 1000 $s^{-1}$ | 0.005 |
| 5 ms | 8 ms | 0.04 | 200 $s^{-1}$ | 0.023 |
| 5 ms | 4 ms | 0.07 | 1000 $s^{-1}$ | 0.009† |
| 5 ms | 4 ms | 0.07 | 200 $s^{-1}$ | 0.040 |
| 5 ms | 2 ms | 0.10 | 1000 $s^{-1}$ | 0.017* |
| 5 ms | 2 ms | 0.10 | 200 $s^{-1}$ | 0.075 |
| 5 ms | 1 ms | 0.13 | 1000 $s^{-1}$ | 0.034 |
| 5 ms | 1 ms | 0.13 | 200 $s^{-1}$ | 0.130 |
| 20 ms | 2 ms | 0.20 | 1000 $s^{-1}$ | 0.028 |
| 20 ms | 2 ms | 0.20 | 200 $s^{-1}$ | 0.140‡ |
| 1 ms | 2 ms | 0.06 | 1000 $s^{-1}$ | 0.012 |
| 1 ms | 2 ms | 0.06 | 200 $s^{-1}$ | 0.050 |

Table 1: Parameters used for integrate-and-fire simulations. As described in the text, values of $\bar{g}_{int}$ were chosen so that stimulus-induced correlations and connectivity-induced correlations were similar magnitudes, and values of $\bar{g}_{ext}$ were chosen to keep approximately constant mean firing rates. Simulations with parameters flagged by † and ‡ are shown in Figs. 7 and 8, respectively. Parameters based on those flagged by * were used for the highly nonlinear simulation shown in Fig. 10.

10 spikes per second, typical neuronal firing rate in response to white noise. For large $\bar{g}_{ext}$, we set $\alpha = 200$ s$^{-1}$, leading to an average input rate of nearly 80 inputs per second. Again, we adjusted $g_{ext}$ so that the average firing rate was between 5 and 10 spikes per second. The resulting parameter values are given in Table 1. As an illustration of these conductance magnitudes, when $\tau_m = 5$ ms and $\tau_s = 2$ ms, 18 nearly simultaneous inputs would bring the voltage to threshold starting from the rest potential with the small $\bar{g}_{ext}$. With the large $\bar{g}_{ext}$, the voltage would reach threshold after only 4 inputs.

We simulated the neurons in response to 20 minutes of spatio-temporal white noise, recording 6,000–12,000 spikes per neuron. We treated the spikes as though they were generated by a linear-nonlinear system with an error function nonlinearity and computed the correlation measures described above. Since saturation of firing plays little role with the low firing rates, the selection of $\hat{r}_p$ (maximal firing rate) had little effect on the results. We simply set $\hat{r}_p = 1$ ms$^{-1}$ (ignoring the absolute refractory period of 2 ms).

The results from one example simulation are shown in Fig. 7. Given the excitatory connection from neuron 2 to neuron 1 with a delay of $\Delta = 40$ ms, we would expect a positive covariance $\mathcal{C}$ after a delay of 40 ms. Indeed, Fig. 7b shows a peak centered around 50 ms. Moreover, since the kernel $\bar{\mathbf{h}}_2$ is similar to the kernel $\bar{\mathbf{h}}_1$ delayed 40 ms ($b_1 = 0$ and $b_2 = 40$ ms), we would expect a positive covariance around a delay of $-40$ ms. Such a peak is also observed in Fig. 7b.

Ideally, the correlation measures $\mathcal{S}$ and $\mathcal{W}$ would act as SICMs also for the integrate-and-fire neurons since the neurons are driven by the stimulus similar to a linear-nonlinear system. To show that is not always the case, we have started with our worst examples. As shown in Fig. 7c–d, $\mathcal{S}$ and $\mathcal{W}$ remove some, but not all, of the stimulus-induced correlations. In fact, $\mathcal{S}$ removes too much, developing a dip at the location of the stimulus-induced peak in $\mathcal{C}$. This dip is greatly reduced, but still perceptible, in $\mathcal{W}$.

Another example in which $\mathcal{S}$ overcompensates for the stimulus-induced correlation is shown in Fig. 8. Again, $\mathcal{S}$ develops a dip at the location of the stimulus-induced peak in $\mathcal{C}$. This time $\mathcal{W}$ virtually eliminates this dip and essentially contains only the correlations induced by the excitatory connection.

The examples shown in Figs. 7 and 8 were the two simulations in which $\mathcal{S}$ and $\mathcal{W}$ performed most poorly. In some of the other simulations with parameters from Table 1, $\mathcal{S}$ did develop a dip from the stimulus-induced correlation, but in all cases it was smaller, within the confidence intervals

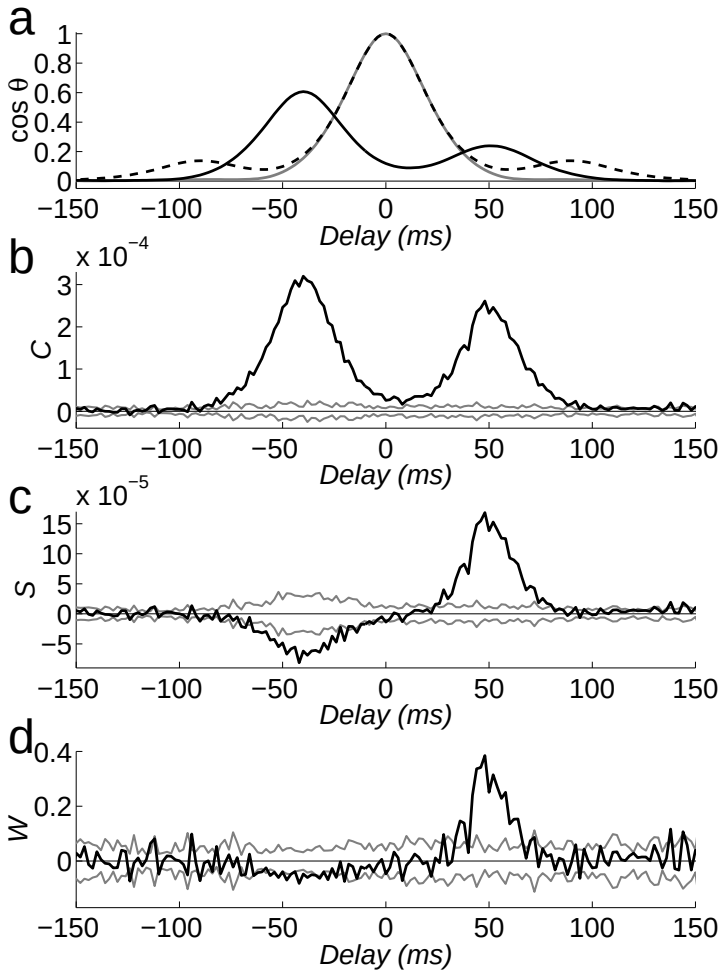


Figure 7: Correlation measures for a pair of integrate-and-fire neurons where neuron 2 has an excitatory connection onto neuron 1. Parameters are those flagged by † in Table 1. Panels are as in Fig. 3, except units of $\mathcal{C}$ and $\mathcal{S}$ are (spikes/second)$^2$. **(a)** The inner product $\cos\theta^k_{21}$ reaches 0.6 at a delay of $-40$ ms due to similarity in the kernels $\mathbf{h}_p$. The peak around 50 ms is due to modification of the kernels by the coupling. Similarly, the small peaks in $\cos\theta^k_{11}$ at absolute delays around 90 ms are due to the coupling. **(b)** The covariance contains a peak centered around a delay of $-40$ ms, due to the kernel similarity, in addition to the connectivity-induced peak around a delay of 50 ms. **(c)** $\mathcal{S}$ overcompensates for the stimulus-induced peak, developing a dip instead, and retains the connectivity-induced peak. **(d)** $\mathcal{W}$ greatly reduces the stimulus-induced dip in $\mathcal{S}$, containing mainly the peak reflecting the excitatory connection from neuron 2 to neuron 1.

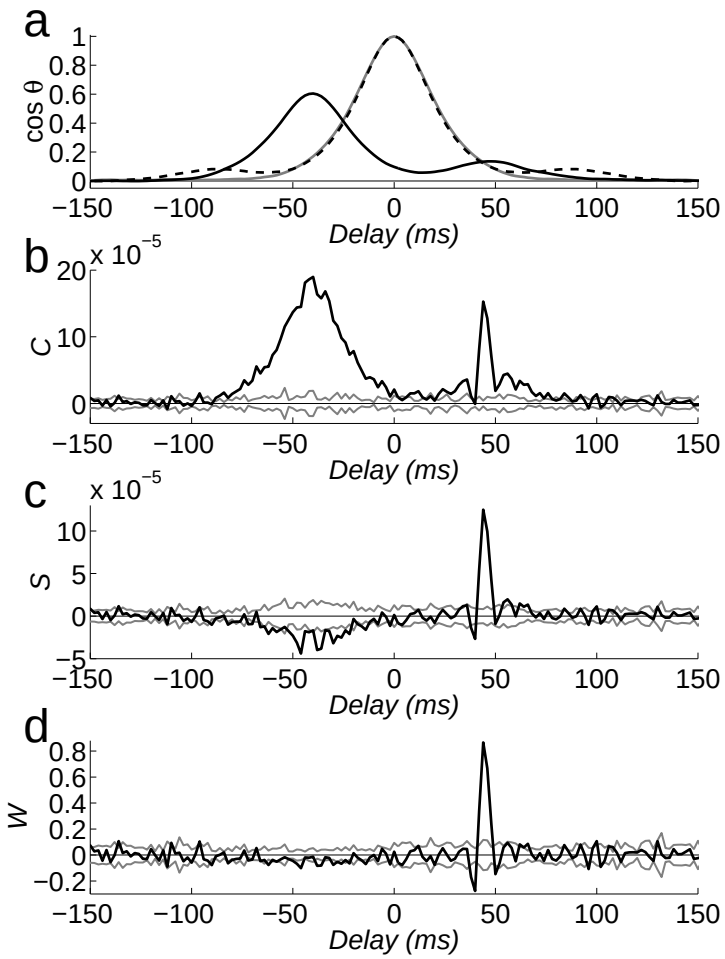


Figure 8: Correlation measures for the pair of integrate-and-fire neurons with parameters that are flagged by ‡ in Table 1. Panels are as in Fig. 7. **(a)** The kernel inner products are similar to those in Fig. 7. **(b)** The covariance shows a stimulus-induced peak around a delay of $-40$ ms and a sharp connectivity-induced peak just before 50 ms. **(c)** As in Fig. 7, $\mathcal{S}$ shows a dip at the location of the stimulus-induced correlation and retains the connectivity-induced correlation. **(d)** $\mathcal{W}$ virtually eliminates the stimulus-induced correlation. The correlation from the connectivity is retained, although a significant dip before the peak develops.

estimated from our 20 minute simulations. In each example, $\mathcal{W}$ virtually eliminated any dip, although in a couple cases (with $\tau_m = 8$ ms and $\bar{g}_{int} = 0.04$) the connectivity-induced peak was barely above the confidence intervals after the 20 minute simulations.

From these examples, it would appear that the correlation measure $\mathcal{S}$ must be interpreted carefully since it frequently overcompensates for stimulus-induced correlations. When $\mathcal{S}$ has the opposite sign of $\mathcal{C}$, the correlation in $\mathcal{S}$ may be only due to the common stimulus. The correlation measure $\mathcal{W}$ does a much better job at eliminating stimulus-induced correlations.

In reality, the measure $\mathcal{S}$ should do better than indicated by these example simulations. In these examples, we have been stressing the model by using kernels $\bar{\mathbf{h}}_1$ and $\bar{\mathbf{h}}_2$ that are identical except for rotation. This choice gives large inner products $\cos\theta_{21}$ and large stimulus-induced correlations that must be eliminated. In general, not only the orientation but also kernel properties such as location, spatial phase, and spatial frequency would differ between neurons. If the kernels differed in these properties, their inner products would be smaller and the task of $\mathcal{S}$ and $\mathcal{W}$ would be easier.

As an example, we took the simulation with the worst results, shown in Fig. 7, and simply changed the spatial phase of kernel 2. We let $\bar{\mathbf{h}}_2$ be even-symmetric rather than odd-symmetric by changing the sine in Eq. (23) to a cosine. Since this modification reduced the magnitude of the stimulus-induced correlation, we reduced $\bar{g}_{ext}$ to 0.05 so that the connectivity-induced correlation was reduced to a similar size.

The results are shown in Fig. 9. Both $\mathcal{S}$ and $\mathcal{W}$ completely eliminated the stimulus-induced correlation evident in $\mathcal{C}$. The proposed correlation measures $\mathcal{S}$ and $\mathcal{W}$ behaved as SICMs just as they did for linear-nonlinear neurons.

We repeated this spatial phase modification for each of the 12 example simulations. In each case, the measures $\mathcal{S}$ and $\mathcal{W}$ were close to SICMs, virtually eliminating the correlations induced by the common stimulus with only slight overcompensation by $\mathcal{S}$ in two cases. The main evidence of the stimulus-induced correlations was a slightly increased variability in $\mathcal{S}$. We achieved similar results by altering the spatial frequency $f$ for neuron 2.

We emphasize that the proposed SICMs $\mathcal{S}$ and $\mathcal{W}$ will work only when a neuron is being driven like a linear-nonlinear system. When a neuron's response to the stimulus is more fundamentally nonlinear, the neuron cannot be approximated as a linear-nonlinear system and the measures $\mathcal{S}$ and $\mathcal{W}$ based on the linear-nonlinear model will not be SICMs.

As an example where the linear-nonlinear approximation breaks down,

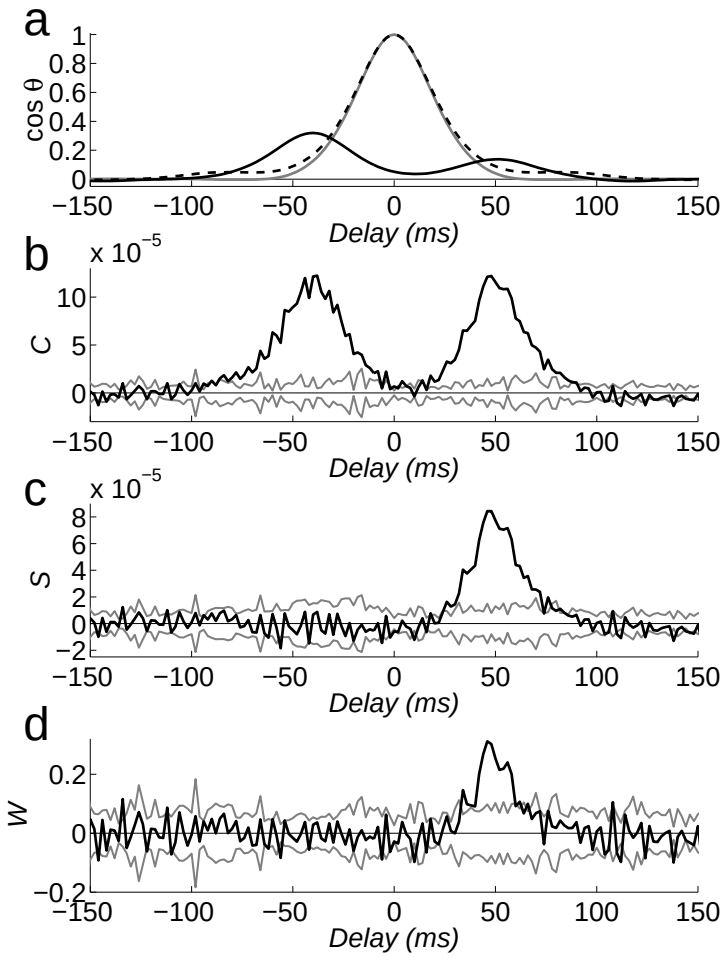


Figure 9: Correlation measures for a pair of integrate-and-fire neurons where the spatial phase of neuron 2 differs from that of neuron 1. Parameters are as in Fig. 7 except for the spatial phase of $\bar{\mathbf{h}}_2$ and $\bar{g}_{ext} = 0.05$. Panels are as in Fig. 7. **(a)** The inner product $\cos\theta_{21}^k$ reaches only 0.3 because the different spatial phases reduces the inner product. **(b)** Both the stimulus-induced correlation and the connectivity-induced correlation are smaller than in Fig. 7. **(c)** $\mathcal{S}$ completely eliminates the stimulus-induced peak. **(d)** $\mathcal{W}$ likewise has no stimulus-induced peak.

we simulated a pair of neurons with input rates given by Eq. (27). We used the simulation parameters marked by an asterisk in Table 1, only setting $\alpha^+ = 0.9$ and $\alpha^- = 0.6$. The results are shown in Fig. 10. Clearly, the measures $\mathcal{S}$ and $\mathcal{W}$ are not SICMs as they still contain substantial peaks due to the stimulus-induced correlations. They reduce the size of the peak because the input rate still contains a linear component. If we set $\alpha^+ = \alpha^-$, then $\mathcal{S}$ and $\mathcal{W}$ would not reduce the stimulus-induced correlations compared to $\mathcal{C}$.

# 4 Discussion

The above simulations demonstrate that the correlation measures $\mathcal{S}$ and $\mathcal{W}$ remove stimulus-induced correlations when the neural responses can be approximated by linear-nonlinear systems (Eq. (17)). With more realistic neuron models, $\mathcal{S}$ and $\mathcal{W}$ may function as approximate SICMs provided that the neurons are driven like a linear-nonlinear system.

The correlation measures $\mathcal{S}$ and $\mathcal{W}$ are not perfect SICMs with integrate-and-fire neurons. When the inner product between the two kernels is large, $\mathcal{S}$ overcompensates for the stimulus-induced correlations. This deficit in $\mathcal{S}$ is mitigated by its signature of flipping the correlation sign so that the likely presence of stimulus-induced correlation can be deduced. Since this overcompensation appears to occur only when the two kernels are very similar, it may not prove to be a problem in practice.

The second SICM $\mathcal{W}$ virtually eliminated the stimulus-induced correlations in almost all cases. The reasons behind its better performance are not completely clear since it is a linear function of $\mathcal{S}$ (Eq. (22)). Apparently, only small changes in $\bar{W}$ are needed for Eq. (17) to adjust to approximate the overcompensation observed in $\mathcal{S}$.

Despite some discrepancies, the integrate-and-fire simulations demonstrate that the applicability of the analysis is broader than linear-nonlinear systems required by the derivation. The relative accuracy of $\mathcal{S}$ and $\mathcal{W}$ in the integrate-and-fire simulations is significant because the integrate-and-fire model differs from a linear-nonlinear in two important ways. First of all, the model has more than one nonlinearity, better approximating biology. The input (Eq. (26)) contains a thresholding nonlinearity and the spiking mechanism of the integrate-and-fire neuron is another nonlinearity. Second, the integration of inputs and the refractory period imply that the firing proba-

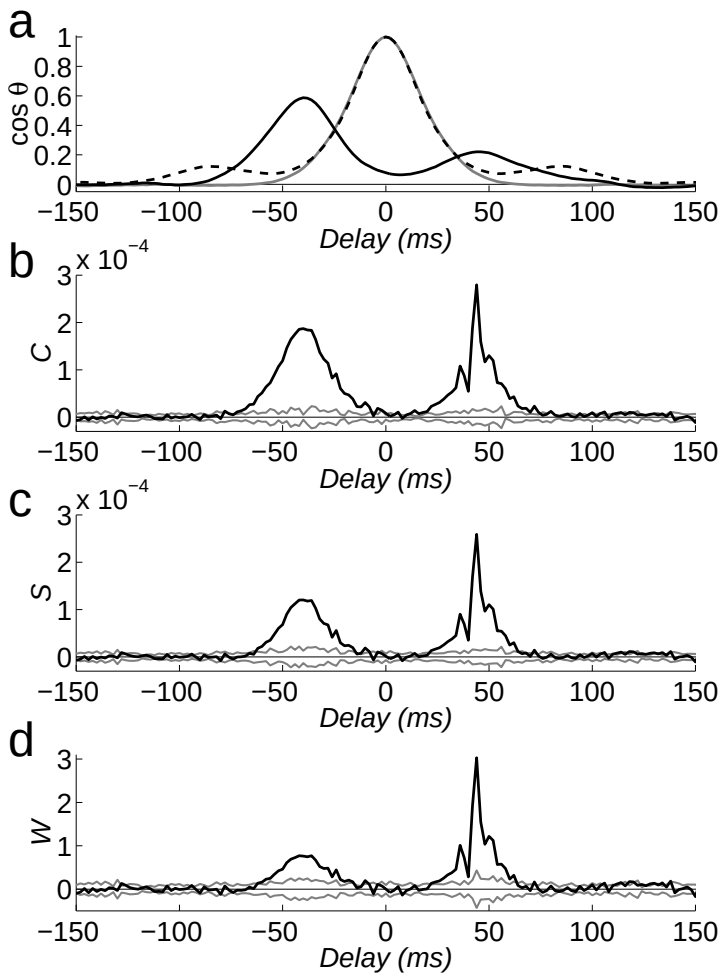


Figure 10: Correlation measures for a pair of integrate-and-fire neurons where the input rates are fundamentally nonlinear functions of the stimulus. Parameters are those flagged by the * in Table 1 except that the input rates are given by Eq. (27) with $\alpha^+ = 0.9$ and $\alpha^- = 0.6$. Panels are as in Fig. 7. **(a)** The calculated inner products account for only the linear portion of the response. **(b)** The covariance contains both stimulus-induced (at negative delays) and connectivity-induced components (at positive delays). **(c)** $\mathcal{S}$ only slightly reduces the stimulus-induced correlation. **(d)** $\mathcal{W}$ reduces the stimulus-induced correlation slightly more but still contains a significant peak at negative delays.

bility depends not only on the history of the stimulus but also the time of the neuron's previous spike. Fortunately, the low firing rate during white noise experiments, which we replicated in our simulations, minimizes the effect of this departure from a linear-nonlinear system. This low firing rate also minimizes the impact of our assumption (made solely for mathematical convenience) that the effective nonlinearity was an error function.

$\mathcal{W}$ and $\mathcal{S}$ are valid only when the neuronal responses can be approximated by linear-nonlinear systems. The integrate-and-fire simulations demonstrate some robustness of these measures to deviations from the linear-nonlinear model. The simulations also show that $\mathcal{W}$ and $\mathcal{S}$ will not be SICMs when the neurons respond to the stimulus in a fundamentally nonlinear fashion. More careful analysis of the limitations of the method is needed so that one can determine the applicability of the method from the measured spike times alone.

The SICM $\mathcal{W}$ provides a useful normalization of the stimulus-independent correlations. It translates the measured correlations into the same units as the stimulus, possibly giving additional meaning to the measurements. When the simulated neurons were linear-nonlinear systems, $\mathcal{W}$ closely matched the connectivity $\bar{W}$ both qualitatively and (assuming sufficient data were collected) quantitatively.

One must exercise care in the interpretation of $\mathcal{W}$, especially since its formulation implies that it is measuring only correlations due to coupling between the two neurons. In reality, it cannot distinguish coupling between the measured neurons and other sources of correlation, such as correlations caused by a shared connection with a third, unmeasured, neuron. The correlation due to this common input can yield a qualitatively similar $\mathcal{W}$ as the correlations due to direct coupling [13]. Further analysis of coupled linear-nonlinear systems is needed to make such a distinction.

# Acknowledgments

The author thanks Dario Ringach for numerous helpful discussions throughout the development of this research and Charlie Peskin and Daniel Tranchina for constructive criticism on an early version of these ideas. This work was supported by a NSF Mathematical Sciences Postdoctoral Research Fellowship.

# A Appendix: Formula for the SICM $\mathcal{W}$

The derivation of the SICM $\mathcal{W}$ is given in Ref. [13]. Here, for completeness, we simply write the formula for $\mathcal{W}$ in terms of the following measurable statistics: the mean rate of neuron $p$ $\langle R_p^i \rangle$, the stimulus-spike correlation of neuron $p$ $\langle \mathbf{X} R_p^i \rangle$, and the correlation between the spikes of neuron 1 and neuron 2 $\langle R_1^i R_2^{i-k} \rangle$.

Since, in matrix notation,

$$\mathcal{W} = A^{-1}\mathcal{S}$$

(Eq. (22)), where

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

(Eq. (20)), we simply need the formula for $A_{pq}^{kj}$.

Subject to the approximation described in section 2.2, $A_{pq}^{kj}$ is given by

$$A_{pq}^{kj} = \mu_q^0 \big[ \tilde{\nu}_{pq}^{kj} - \eta_{pq}^k \eta_{pq}^j + (\cos\theta_{pq}^k \cos\theta_{pq}^j - \cos\theta_{pp}^{k-j}) \mu_{pq}^k \mu_{pq}^j \big], \qquad (28)$$

where

$$\cos\theta_{pq}^k = \frac{\langle \mathbf{X} R_p^{i-k} \rangle \cdot \langle \mathbf{X} R_q^i \rangle}{|\langle \mathbf{X} R_p^{i-k} \rangle| |\langle \mathbf{X} R_q^i \rangle|}$$

(Eq. (16)), $T_p$ and $\delta_p$ are defined by

$$\langle R_p^i \rangle = \frac{\hat{r}_p}{2} \mathrm{erfc}\Big( \frac{\delta_p T_p}{\sqrt{2}} \Big)$$

(Eq. (12)) and

$$|\langle \mathbf{X} R_p^i \rangle| = \frac{\hat{r}_p \delta_p}{\sqrt{2\pi}} \exp\Big( -\frac{\delta_p^2 T_p^2}{2} \Big)$$

(Eq. (13)), and the other parameters are defined by

$$\eta_{pq}^k = \tfrac{\hat{r}_p}{2} \mathrm{erfc}\big( \lambda_{pq}^k / \sqrt{2} \big)$$

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

$$\mu_p^0 = \frac{\hat{r}_p \delta_p}{\sqrt{2\pi}} \exp\left(-\delta_p^2 T_p^2/2\right)$$

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

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

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

Recall that the double complementary error function is defined by

$$\mathrm{derfc}(a, b, c) = \frac{2}{\sqrt{\pi}} \int_a^\infty e^{-y^2} \mathrm{erfc}\left(\frac{b - cy}{\sqrt{1 - c^2}}\right) dy.$$

(Eq. (7)) where

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt.$$

# B Appendix: Calculation of confidence intervals

## B.1 Overview of the method

In order to determine the significance of nonzero correlation measures, one must determine the likelihood of obtaining the observed value by chance even if its expected value were zero. One way to provide a gauge on this likelihood is to estimate the standard deviation of the measurement, computing the standard error of the measured values. From these estimates, one can form confidence intervals as a benchmark for significance.

Computing the standard error of basic measurements such as $\langle R_p^i \rangle$ or $\langle R_1^i R_2^{i-k} \rangle$ is simple. One needs only to divide the data into $M$ subgroups and estimate the measurement for each subgroup. The standard error is simply the standard deviation of their mean.

To calculate the standard error of a nonlinear function (e.g., $\mathcal{S}$) of the basic measurements, one could compute all covariances of the basic measurements (by taking the covariances of subgroup means). One could then use error propagation techniques to estimate the standard error of the nonlinear function. (One can derive typical error propagation methods by assuming the covariances are small and linearizing the nonlinear function around the measured values.)

However, we were unable to derive a suitable generalization of this approach that we could apply to $\mathcal{W}$. First, for large kernels $\mathbf{h}_p^i$, the computation required to calculate the functions of the large number of covariances proved to be unrealistically long. Second, and more importantly, we were unable to find a suitable method to propagate the error through the matrix inversion defining $\mathcal{W}$. Estimates using matrix condition numbers proved to be gross overestimates of the error since they are upper bounds independent of the structure of the error.

Instead, we implemented a Monte Carlo calculation of the standard error. We used this method for all confidence intervals shown in the figures. The method involved calculating the means and covariance matrix (as described above) of a set of base variables from which all statistics could be calculated. We then sampled 50 joint Gaussian random variables with those means and that covariance matrix. From each sample, we calculated the resulting correlation measures ($\mathcal{C}$, $\mathcal{S}$, and $\mathcal{W}$). We estimated the standard errors by the standard deviations of these 50 estimates.

Note that this method does not require simulation of the neurons beyond the simulation that generated the original data set. For this reason, it could be used on data collected from real neurons. It is simply a method to calculate error propagation without a formula.

The rest of this appendix gives technical details about the calculation, including necessary slight modifications from the above outline.

## B.2 The base variables

The correlation measures $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{W}$ are functions of the measured mean rate of neuron $p$ $\langle R_p^i \rangle$, stimulus-spike correlation of neuron $p$ $\langle \mathbf{X} R_p^i \rangle$, and correlation between the spikes of neuron 1 and neuron 2 $\langle R_1^i R_2^{i-k} \rangle$. In practice, we actually measure $\langle \mathbf{X} | R_p^i = 1 \rangle = \langle \mathbf{X} R_p^i \rangle / \langle R_p^i \rangle$ rather than $\langle \mathbf{X} R_p^i \rangle$ as described in Ref. [14].

We could compute the covariances of all these values and base our Monte Carlo simulation on those values, but that would lead to an unmanageably large covariance matrix. The statistics $\langle \mathbf{X}|R_p^i = 1\rangle$ are used only in the form $\langle \mathbf{X}|R_p^i = 1\rangle \cdot \langle \mathbf{X}|R_q^{i-k} = 1\rangle$ (for $|\langle \mathbf{X}R_p^i\rangle|^2$ and $\cos\theta_{pq}^k$). We greatly reduce the number of base variables by using only the inner products of $\langle \mathbf{X}|R_p^i = 1\rangle$ rather than their individual components. If we calculate the statistics for delays $k = -N, \ldots N$ ($2N+1$ delays), then the $6N+6$ base variables are

- $\langle R_1^i\rangle$ and $\langle R_2^i\rangle$ (2 variables),
- $\langle R_1^i R_2^{i-k}\rangle$ for $k = -N, \ldots, N$ ($2N+1$ variables),
- $\langle \mathbf{X}|R_p^i = 1\rangle \cdot \langle \mathbf{X}|R_p^{i-k} = 1\rangle$ for $p = 1, 2$ and $k = 0, \ldots, N$ ($2N+2$ variables), and
- $\langle \mathbf{X}|R_1^i = 1\rangle \cdot \langle \mathbf{X}|R_2^{i-k} = 1\rangle$ for $k = -N, \ldots, N$ ($2N+1$ variables).

Recall that each statistic is independent of time point $i$ since the stimulus and the neural response are stationary.

We divide the data into $M$ equal subgroups and estimate each base variable for each subgroup. We use these estimates to calculate the covariance matrix of the base variables as described above. Since the kernel inner products are inner products of averages (rather than averages of inner products), we make an error by not calculating the covariances of each $\langle \mathbf{X}R_p^i\rangle$ individually. To minimize this error, we make the subgroups large, using only $M = 4$ subgroups.

## B.3 Generating the joint normals

One standard method to generate joint normals $\mathbf{Y}$ with mean zero and covariance matrix $B$ is to first compute the Cholesky decomposition $L$ of the matrix $B$

$$B = LL^T \tag{29}$$

where $L$ is a lower triangular matrix. Then generate a vector $\mathbf{Z}$ of independent standard normal random variables (with mean 0 and standard deviation 1). The variables

$$\mathbf{Y} = L\mathbf{Z} \tag{30}$$

have the covariance matrix $B$.

A covariance matrix $B$ is guaranteed to be symmetric and positive semi-definite if it were computed with 100% accuracy. However, representing the matrix on a computer implies the computer version can be accurate only within machine precision. If, as in the present case, some of the variables are highly correlated, numerical error in the covariance matrix could lead to negative eigenvalues. (Two perfectly correlated variables would create a zero eigenvalue, so highly correlated variables lead to small eigenvalues. If they are sufficiently small, numerical error could turn them negative.)

Since Cholesky decomposition fails if any eigenvalues are negative, a straightforward attempt to generate random variables with our covariance matrix $B$ will fail. To ensure that the covariance matrix is positive definite, we perturb the matrix slightly in the following manner.

First, we normalize the covariance matrix $B$ into a matrix of correlation coefficients (call it the correlation matrix $D$). We store the standard deviation of each base variable (the square root of the diagonal of $B$) and divide each row and column by the standard deviation. This leads to a matrix $D$ with ones on the diagonals and all off-diagonals less than one in magnitude. (Many of the off-diagonals are nearly one, which is the source of the problem.)

Next, we decompose $D$ into eigenvectors and corresponding eigenvalues. We set an eigenvalue cutoff around $10^{-14}$ times the maximal eigenvalue. If any eigenvalues are less than that cutoff, we set them equal to the cutoff. In this way, we ensure that the minimal eigenvalue is no less than the cutoff (hence positive).

We reconstruct $D$ from its eigenvectors and modified eigenvalues. Since $D$ is now positive definite, we can compute its Cholesky decomposition $D = LL^T$ and generate random variables $\mathbf{Y} = L\mathbf{Z}$ with covariances given by $D$. We multiply each component of $\mathbf{Y}$ by the standard deviation of the corresponding base variable and add its mean (the original value of the base variable). The random variables $\mathbf{Y}$ then have the expected value and covariance matrix calculated from the base variables.

## B.4 Computing a linear approximation

The nonlinear calculation of $\mathcal{W}$ can fail if the base variables do not correspond to values achievable by the model of Eq. (17) (the calculation may take a square root of a negative number or the matrix $A$ may become singular). For this reason, computing $\mathcal{W}$ using the random variables $Y$ for the base

variables sometimes fails or nearly fails (leading to arbitrarily large values of $\mathcal{W}$).

To avoid this problem, we look for a linear approximation of the error around the originally measured values of the base variables. Since we are not computing formulas, we can't explicitly linearize the nonlinearities as is done in typical error propagation. Instead, we simply make the deviations from the original base variables smaller.

We divide the standard deviations of the base variables by 10 before multiplying $Y$ by them in the above procedure. The random variables still have the same correlation structure as before, only their variation around the original base variables is 10 times smaller. Since they are so close to the original base variables, the calculation of $\mathcal{W}$ is unlikely to fail.

We compute estimates of $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{W}$ from each sample of $Y$. Our estimate of the standard error is simply the standard deviation of these estimates multiplied by 10. In this way, the error estimates are based on the behavior of $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{W}$ near the original values of the base variables. We've effectively computed error measures based on a linearization around those base values.

# References


[1] A. M. H. J. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm. Dynamics of neuronal firing correlation: Modulation of "effective connectivity". *J. Neurophysiol.*, 61:900–917, 1989.

[2] C. D. Brody. Correlations without synchrony. *Neural. Comput.*, 11:1537–51, 1999.

[3] E. J. Chichilnisky. A simple white noise analysis of neural light responses. *Network: Comput. Neural Syst.*, 12:199–213, 2001.

[4] 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.*, 69:1091–1117, 1993.

[5] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II.

Linearity of temporal and spatial summation. *J. Neurophysiol.*, 69:1118–1135, 1993.

[6] E. DeBoer and P. Kuyper. Triggered correlation. *IEEE Trans. Biomed. Eng.*, 15:169–179, 1968.

[7] R. C. deCharms, D. T. Blake, and M. M. Merzenich. Optimizing sound features for cortical neurons. *Science*, 280:1439–1443, 1998.

[8] J. J. DiCarlo and K. O. Johnson. Velocity invariance of receptive field structure in somatosensory cortical area 3b of the alert monkey. *J. Neurosci.*, 19:401–419, 1999.

[9] 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.*, 18:2626–2645, 1998.

[10] 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.*, 21:4408–4415, 2001.

[11] 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.*, 58:1233–1258, 1987.

[12] P. N. Marmarelis and V. Z. Marmarelis. *Analysis of physiological systems: the white noise approach.* Plenum Press, NewYork, 1978.

[13] D. Q. Nykamp. White noise analysis of coupled linear-nonlinear systems. *SIAM J. Appl. Math.* To appear. (www.math.ucla.edu/~nykamp/pubs).

[14] D. Q. Nykamp and D. L. Ringach. Full identification of a linear-nonlinear system via cross-correlation analysis. *J. Vision*, 2:1–11, 2002.

[15] G. Palm, A. M. H. J. Aertsen, and G. L. Gerstein. On the significance of correlations among neuronal spike trains. *Biol. Cybern.*, 59:1–11, 1988.

[16] D. H. Perkel, G. L. Gerstein, and G. P. Moore. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.*, 7:419–40, 1967.

[17] R. C. Reid and J. M. Alonso. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378:281–284, 1995.

[18] 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.*, 14:1015–1027, 1997.

[19] D. L. Ringach. personal communication.

[20] D. L. Ringach, G. Sapiro, and R. Shapley. A subspace reverse-correlation technique for the study of visual neurons. *Vis. Research*, 37:2455–2464, 1997.

[21] M. J. Shelley and L. Tao. Efficient and accurate time-stepping schemes for integrate-and-fire neuronal networks. *J. Comp. Neurosci.*, 11:111–119, 2001.