

# Impact of Network Structure and Cellular Response on Spike Time Correlations

James Trousdale<sup>1\*</sup>, Yu Hu<sup>2</sup>, Eric Shea-Brown<sup>2</sup>, Krešimir Josić<sup>1,3</sup>

**1** Department of Mathematics, University of Houston, Houston, Texas, United States of America, **2** Department of Applied Mathematics, Program in Neurobiology and Behavior, University of Washington, Seattle, Washington, United States of America, **3** Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America

## Abstract

Novel experimental techniques reveal the simultaneous activity of larger and larger numbers of neurons. As a result there is increasing interest in the structure of cooperative – or *correlated* – activity in neural populations, and in the possible impact of such correlations on the neural code. A fundamental theoretical challenge is to understand how the architecture of network connectivity along with the dynamical properties of single cells shape the magnitude and timescale of correlations. We provide a general approach to this problem by extending prior techniques based on *linear response theory*. We consider networks of general integrate-and-fire cells with arbitrary architecture, and provide explicit expressions for the approximate cross-correlation between constituent cells. These correlations depend strongly on the operating point (input mean and variance) of the neurons, even when connectivity is fixed. Moreover, the approximations admit an expansion in powers of the matrices that describe the network architecture. This expansion can be readily interpreted in terms of paths between different cells. We apply our results to large excitatory-inhibitory networks, and demonstrate first how precise *balance* – or lack thereof – between the strengths and timescales of excitatory and inhibitory synapses is reflected in the overall correlation structure of the network. We then derive explicit expressions for the average correlation structure in randomly connected networks. These expressions help to identify the important factors that shape coordinated neural activity in such networks.

**Citation:** Trousdale J, Hu Y, Shea-Brown E, Josić K (2012) Impact of Network Structure and Cellular Response on Spike Time Correlations. PLoS Comput Biol 8(3): e1002408. doi:10.1371/journal.pcbi.1002408

**Editor:** Olaf Sporns, Indiana University, United States of America

**Received:** October 21, 2011; **Accepted:** January 11, 2012; **Published:** March 22, 2012

**Copyright:** © 2012 Trousdale et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by NSF grants DMS-0817649, DMS-1122094, and a Texas ARP/ATP award to KJ, as well as NSF Grants DMS-1122106, DMS-0818153, and a Burroughs Wellcome Career Award at the Scientific Interface to ESB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: jtrousd@math.uh.edu

## Introduction

New multielectrode and imaging techniques are revealing the simultaneous activity of neural ensembles and, in some cases, entire neural populations [1–4]. This has thrust upon the computational biology community the challenge of characterizing a potentially complex set of interactions – or *correlations* – among pairs and groups of neurons.

Beyond important and rich challenges for statistical modeling [5], the emerging data promises new perspectives on the neural encoding of information [6]. The structure of correlations in the activity of neuronal populations is of central importance in understanding the neural code [7–13]. However, theoretical [9–11,14–16], and empirical studies [17–19] do not provide a consistent set of general principles about the impact of correlated activity. This is largely because the presence of correlations can either strongly increase or decrease the fidelity of encoded information depending on both the structure of correlations across a population and how their impact is assessed.

A basic mechanistic question underlies the investigation of the role of collective activity in coding and signal transmission: How do single-cell dynamics, connection architecture, and synaptic dynamics combine to determine patterns of network activity? Systematic answers to this question would allow us to predict how

empirical data from one class of stimuli will generalize to other stimulus classes and recording sites. Moreover, a mechanistic understanding of the origin of correlations, and knowledge of the patterns we can expect to see under different assumptions about the underlying networks, will help resolve recent controversies about the strength and pattern of correlations in mammalian cortex [1,20,21]. Finally, understanding the origin of correlations will inform the more ambitious aim of inferring properties of network architecture from observed patterns of activity [22–24].

Here, we examine the link between network properties and correlated activity. We develop a theoretical framework that accurately predicts the structure of correlated spiking that emerges in a widely used model – recurrent networks of general integrate and fire cells. The theory naturally captures the role of single cell and synaptic dynamics in shaping the magnitude and timescale of spiking correlations. We focus on the exponential integrate and fire model, which has been shown to capture membrane and spike responses of cortical neurons [25]; however, the general approach we take can be applied to a much broader class of neurons, a point we return to in the Discussion.

Our approach is based on an extension of linear response theory to networks [24,26]. We start with a linear approximation of a neuron's response to an input. This approximation can be obtained explicitly for many neuron models [27–29], and is

## Author Summary

Is neural activity more than the sum of its individual parts? What is the impact of cooperative, or *correlated*, spiking among multiple cells? We can start addressing these questions, as rapid advances in experimental techniques allow simultaneous recordings from ever-increasing populations. However, we still lack a general understanding of the origin and consequences of the joint activity that is revealed. The challenge is compounded by the fact that both the intrinsic dynamics of single cells and the correlations among them vary depending on the overall state of the network. Here, we develop a toolbox that addresses this issue. Specifically, we show how *linear response theory* allows for the expression of correlations explicitly in terms of the underlying network connectivity and known single-cell properties – and that the predictions of this theory accurately match simulations of a touchstone, nonlinear model in computational neuroscience, the general integrate-and-fire cell. Thus, our theory should help unlock the relationship between network architecture, single-cell dynamics, and correlated activity in diverse neural circuits.

directly related to the spike triggered average [30]. The correlation structure of the network is then estimated using an iterative approach. As in prior work [31–33], the resulting expressions admit an expansion in terms of paths through the network.

We apply this theory to networks with precisely balanced inhibition and excitation in the inputs to individual cells. In this state individual cells receive a combination of excitatory and inhibitory inputs with mean values that largely cancel. We show that, when timescales and strengths of excitatory and inhibitory connections are matched, only local interactions between cells contribute to correlations. Moreover, our theory allows us to explain how correlations are altered when precise tuning balance is broken. In particular, we show how strengthening inhibition may synchronize the spiking activity in the network. Finally, we derive results which allow us to gain an intuitive understanding of the factors shaping average correlation structure in randomly connected networks of neurons.

## Results

Our goal is to understand how the architecture of a network shapes the statistics of its activity. We show how correlations between spike trains of cells can be approximated using response characteristics of individual cells along with information about synaptic dynamics, and the structure of the network. We start by briefly reviewing linear response theory of neuronal responses [28,34,35], and then use it to approximate the correlation structure of a network.

### Network model

To illustrate the results we consider a network of  $N$  nonlinear integrate-and-fire (IF) neurons with membrane potentials modeled by

$$\tau_i \dot{v}_i = -(v_i - E_{L,i}) + \psi(v_i) + E_i + \sqrt{\sigma_i^2} \tau_i \xi_i(t) + f_i(t) + \eta_i(t). \quad (1)$$

Here  $E_{L,i}$  is the leak reversal potential, and  $E_i$  represents the mean synaptic input current from parts of the system not explicitly modeled. A spike-generating current  $\psi(v_i)$  may be included to

emulate the rapid onset of action potentials. Unless otherwise specified, we utilize the exponential IF model (EIF), so that  $\psi(v) \equiv \Delta_T \exp[(v - v_T)/\Delta_T]$  [25]. Cells are subject to internally induced fluctuations due to channel noise [36], and externally induced fluctuations due to inputs not explicitly modelled [37]. We model both by independent, Gaussian, white noise processes,  $\sqrt{\sigma_i^2} \tau_i \xi_i(t)$  [38]. An external signal to cell  $i$  is represented by  $\eta_i(t)$ .

Upon reaching a threshold  $v_{th}$ , an action potential is generated, and the membrane potential is reset to  $v_r$ , where it is held constant for an absolute refractory period  $\tau_{ref}$ . The output of cell  $i$  is characterized by the times,  $t_{i,k}$ , at which its membrane potential reaches threshold, resulting in an output spike train  $y_i(t) = \sum_k \delta(t - t_{i,k})$ . Synaptic interactions are modeled by delayed  $\alpha$ -functions

$$f_i(t) = \sum_j (\mathbf{J}_{ij} * y_j)(t), \quad \text{where} \quad (2)$$

$$\mathbf{J}_{ij}(t) = \begin{cases} \mathbf{W}_{ij} \left( \frac{t - \tau_{D,j}}{\tau_{S,j}^2} \right) \exp \left[ -\frac{t - \tau_{D,j}}{\tau_{S,j}} \right] & t \geq \tau_{D,j} \\ 0 & t < \tau_{D,j} \end{cases}$$

The  $N \times N$  matrix  $\mathbf{J}$  contains the synaptic kernels, while the matrix  $\mathbf{W}$  contains the synaptic weights, and hence defines the network architecture. In particular, if  $g_L$  is the membrane conductance,  $g_L \mathbf{W}_{ij}$  is the area under a post-synaptic current evoked in cell  $j$  by a spike in the presynaptic cell  $i$ , and along with the membrane and synaptic time constants, determines the area under a post-synaptic potential.  $\mathbf{W}_{ij} = 0$  represents the absence of a synaptic connection from cell  $j$  to cell  $i$ .

Table 1 provides an overview of all parameters and variables.

### Linear response of individual cells

Neuronal network models are typically described by a complex system of coupled nonlinear stochastic differential equations. Their behavior is therefore difficult to analyze directly. We will use linear response theory [28,34,35,39] to approximate the cross-correlations between the outputs of neurons in a network. We first review the linear approximation to the response of a single cell. We illustrate the approach using current-based IF neurons, and explain how it can be generalized to other models in the Discussion.

The membrane potential of an IF neuron receiving input  $\epsilon X(t)$ , with vanishing temporal average,  $\langle X(t) \rangle = 0$ , evolves according to

$$\tau \dot{v} = -(v - E_L) + \psi(v) + E + \sqrt{\sigma^2} \tau \xi(t) + \epsilon X(t). \quad (3)$$

The time-dependent firing rate,  $r(t)$ , is determined by averaging the resulting spike train,  $y(t) = \sum_j \delta(t - t_j)$ , across different realizations of noise,  $\xi(t)$ , for fixed  $X(t)$ . Using linear response theory, we can approximate the firing rate by

$$r(t) = r_0 + (A * \epsilon X)(t), \quad (4)$$

where  $r_0$  is the (stationary) firing rate when  $\epsilon = 0$ . The linear response kernel,  $A(t)$ , characterizes the firing rate response to first order in  $\epsilon$ . A rescaling of the function  $A(t)$  gives the spike-triggered average of the cell, to first order in input strength, and is hence equivalent to the optimal Wiener kernel in the presence of the signal  $\xi(t)$ . [39,40]. In Figure 1, we compare the approximate firing rate obtained from Eq. (4) to that obtained numerically from Monte Carlo simulations.

**Table 1.** Notation used in the text.

Symbol	Description
$v_i, \tau_i, E_{L,i}, \sigma_i$	Membrane potential, membrane time constant, leak reversal potential, and noise intensity of cell $i$ .
$E_i, \sigma_i$	Mean and standard deviation of the background noise for cell $i$ .
$V_{th}, V_r, \tau_{ref}$	Membrane potential threshold, reset, and absolute refractory period for cells.
$\psi(v), V_T, \Delta_T$	Spike generating current, soft threshold and spike shape parameters for the IF model [25].
$f_i(t), \eta_i(t)$	Synaptic input from other cells in the network, and external input to cell $i$ .
$\tau_{S,i}, \tau_{D,i}$	Synaptic time constant and delay for outputs of cell $i$ .
$y_i(t)$	Spike train of cell $i$ .
$W_{ij}$	The $j \rightarrow i$ synaptic weight, proportional to the area under a single post-synaptic current for current-based synapses.
$J_{ij}(t)$	The $j \rightarrow i$ synaptic kernel - equals the product of the synaptic weight $W_{ij}$ and the synaptic filter for outputs of cell $j$ .
$C_{ij}(\tau)$	The cross-correlation function between cells $i, j$ defined by $C_{ij}(\tau) = \text{cov}(y_i(t+\tau), y_j(t))$ .
$N_{y_i}(t, t+\tau), \rho_{ij}(\tau)$	Spike count for cell $i$ , and spike count correlation coefficient for cells $i, j$ over windows of length $\tau$ .
$r_i, A_i(t), C_{ii}^0$	Stationary rate, linear response kernel and uncoupled auto-correlation function for cell $j$ .
$K_{ij}(t)$	The $j \rightarrow i$ interaction kernel - describes how the firing activity of cell $i$ is perturbed by an input spike from cell $j$ . It is defined by $K_{ij}(t) = (A_i * J_{ij})(t)$ .
$y_i^n(t), C_{ij}^n(t)$	The $n^{\text{th}}$ order approximation of the activity of cell $i$ in a network which accounts for directed paths through the network graph up to length $n$ ending at cell $i$ , and the cross-correlation between the $n^{\text{th}}$ order approximations of the activity of cells $i, j$ .
$g(t), \tilde{g}(\omega)$	$\tilde{g}(\omega)$ is the Fourier transform of $g(t)$ with the convention $\tilde{g}(\omega) = \mathcal{F}[g](\omega) \equiv \int_{-\infty}^{\infty} e^{-2\pi i \omega t} g(t) dt$

doi:10.1371/journal.pcbi.1002408.t001

The linear response kernel  $A(t)$  depends implicitly on model parameters, but is independent of the input signal,  $\epsilon X(t)$ , when  $\epsilon$  is small relative to the noise  $\sqrt{\sigma^2 \tau} \zeta(t)$ . In particular,  $A(t)$  is sensitive to the value of the mean input current,  $E$ . We emphasize that the presence of the background noise,  $\zeta$ , in Eq. (3) is essential to the theory, as noise linearizes the transfer function that maps input to

output. In addition, when applying linear response methods, there is an implicit assumption that the fluctuations of the input  $X(t)$  do not have a significant effect on the response properties of the cell.

### Linear response in recurrent networks

The linear response kernel can be used to approximate the response of a cell to an external input. However, the situation is more complicated in a network where a neuron can affect its own activity through recurrent connections. To extend the linear response approximation to networks we follow the approach introduced by Lindner et al. [26]. Instead of using the linear response kernel to approximate the firing rate of a cell, we use it to approximate a realization of its output

$$y(t) \approx y^0(t) + (A * X)(t). \quad (5)$$

Here  $y^0(t)$  represents a realization of the spike train generated by an integrate-and-fire neuron obeying Eq. (3) with  $X(t) = 0$ .

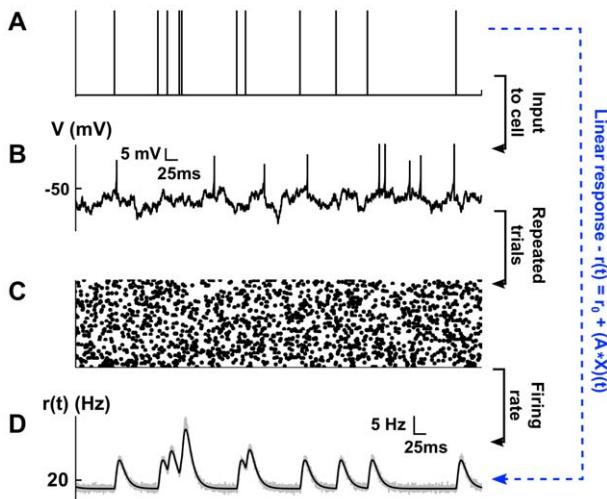
Our central assumption is that a cell acts approximately as a linear filter of its inputs. Note that Eq. (5) defines a mixed point and continuous process, but averaging  $y(t)$  in Eq. (5) over realizations of  $y^0$  leads to the approximation in Eq. (4). Hence, Eq. (5) is a natural generalization of Eq. (4) with the unperturbed output of the cell represented by the point process,  $y^0(t)$ , instead of the firing rate,  $r_0$ .

We first use Eq. (5) to describe spontaneously evolving networks where  $\eta_i(t) = 0$ . Equation (1) can then be rewritten as

$$\tau_i \dot{v}_i = -(v_i - E_{L,i}) + \psi(v_i) + E_i' + \sqrt{\sigma_i^2} \tau_i \zeta_i(t) + (f_i(t) - \mathbf{E}[f_i]), \quad (6)$$

where  $E_i' = E_i + \mathbf{E}[f_i]$  and  $\mathbf{E}[\cdot]$  represents the temporal average.

Lindner et al. used Eq. (5) as an ansatz to study the response of an all-to-all inhibitory network. They postulated that the spiking output  $y_i(t)$  of cell  $i$  in the network, can be approximated in the



**Figure 1. Illustrating Eq. (4).** (A) The input to the post-synaptic cell is a fixed spike train which is convolved with a synaptic kernel. (B) A sample voltage path for the post-synaptic cell receiving the input shown in A) in the presence of background noise. (C) Raster plot of 100 realizations of output spike trains of the post-synaptic cell. (D) The output firing rate,  $r(t)$ , obtained by averaging over realizations of the output spike trains in C). The rate obtained using Monte Carlo simulations (shaded in gray) matches predictions of linear response theory obtained using Eq. (4) (black). doi:10.1371/journal.pcbi.1002408.g001

frequency domain by

$$\tilde{y}_i(\omega) = \tilde{y}_i^0(\omega) + \tilde{A}_i(\omega) \left( \sum_j \tilde{\mathbf{J}}_{ij}(\omega) \tilde{y}_j(\omega) \right),$$

where  $\tilde{y}_i = \mathcal{F}[y_i - r_i]$  are the zero-mean Fourier transforms of the processes  $y_i$ , and  $\tilde{f} = \mathcal{F}(f)$  for all other quantities (see Table 1 for the Fourier transform convention). The term in parentheses is the Fourier transform of the zero-mean synaptic input,  $(f_i(t) - \mathbf{E}[f_i])$ , in Eq. (6), and  $\tilde{y}_i^0(\omega)$  represents a realization of the spiking output of cell  $i$  in the absence of synaptic fluctuations from the recurrent network (*i.e.* assuming  $f_i = \mathbf{E}[f_i]$ ). In matrix form this ansatz yields a simple self-consistent approximation for the firing activities  $\tilde{\mathbf{y}}$  which can be solved to give

$$\tilde{\mathbf{y}}(\omega) = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \tilde{\mathbf{y}}^0(\omega),$$

where the interaction matrix  $\tilde{\mathbf{K}}$  has entries defined by  $\mathbf{K}_{ij}(t) \equiv (A_i * \mathbf{J}_{ij})(t)$ . When averaged against its conjugate transpose, this expression yields an approximation to the full array of cross-spectra in the recurrent network:

$$\langle \tilde{\mathbf{y}}(\omega) \tilde{\mathbf{y}}^*(\omega) \rangle = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \langle \tilde{\mathbf{y}}^0(\omega) \tilde{\mathbf{y}}^{0*}(\omega) \rangle (\mathbf{I} - \tilde{\mathbf{K}}^*(\omega))^{-1}. \quad (7)$$

We next present a distinct derivation of this approximation which allows for a different interpretation of the ansatz given by Eq. (5). We iteratively build to the approximation in Eq. (7), showing how this expression for the correlation structure in a recurrent network can be obtained by taking into account the paths through the network of increasing length.

We start with realizations of spike trains,  $y_i^0$ , generated by IF neurons obeying Eq. (6) with  $f_i(t) = \mathbf{E}[f_i]$ . This is equivalent to considering neurons isolated from the network, with adjusted DC inputs (due to mean network interactions). Following the approximation given by Eq. (5), we use a *frozen* realization of all  $y_i^0$  to find a correction to the output of each cell, with  $X(t)$  set to the mean-adjusted synaptic input,

$$X(t) = f_i(t) - \mathbf{E}[f_i].$$

As noted previously, the linear response kernel is sensitive to changes in the mean input current. It is therefore important to include the average synaptic input  $\mathbf{E}[f_i]$  in the definition of the effective mean input,  $E'_i$ .

The input from cell  $j$  to cell  $i$  is filtered by the synaptic kernel  $\mathbf{J}_{ij}(t)$ . The linear response of cell  $i$  to a spike in cell  $j$  is therefore captured by the interaction kernel  $\mathbf{K}_{ij}$ , defined above as

$$\mathbf{K}_{ij}(t) \equiv (A_i * \mathbf{J}_{ij})(t).$$

The output of cell  $i$  in response to mean-adjusted input,  $y_j^0(t) - r_j$ , from cell  $j$  can be approximated to first order in input strength using the linear response correction

$$y_i^1(t) = y_i^0(t) + \sum_j (\mathbf{K}_{ij} * [y_j^0 - r_j])(t). \quad (8)$$

We explain how to approximate the stationary rates,  $r_j$ , in the Methods.

The cross-correlation between the processes  $y_i^1(t)$  in Eq. (8) gives a first approximation to the cross-correlation function between the cells,

$$\begin{aligned} \mathbf{C}_{ij}(\tau) &\approx \mathbf{C}_{ij}^1(\tau) = \mathbf{E} \left[ (y_i^1(t+\tau) - r_i)(y_j^1(t) - r_j) \right] \\ &= \mathbf{E} \left[ (y_i^0(t+\tau) - r_i)(y_j^0(t) - r_j) \right] \\ &\quad + \sum_k \mathbf{E} \left[ (\mathbf{K}_{ik} * [y_k^0 - r_k])(t+\tau)(y_j^0(t) - r_j) \right] \\ &\quad + \sum_k \mathbf{E} \left[ (y_i^0(t+\tau) - r_i)(\mathbf{K}_{jk} * [y_k^0 - r_k])(t) \right] \\ &\quad + \sum_{k,l} \mathbf{E} \left[ (\mathbf{K}_{ik} * [y_k^0 - r_k])(t+\tau)(\mathbf{K}_{jl} * [y_l^0 - r_l])(t) \right] \end{aligned}$$

which can be simplified to give

$$\begin{aligned} \mathbf{C}_{ij}(\tau) &\approx \mathbf{C}_{ij}^1(\tau) = \delta_{ij} \mathbf{C}_i^0(\tau) + (\mathbf{K}_{ij} * \mathbf{C}_{jj}^0)(\tau) + \\ &(\mathbf{K}_{ji}^- * \mathbf{C}_{ii}^0)(\tau) + \sum_k (\mathbf{K}_{ik} * \mathbf{K}_{jk}^- * \mathbf{C}_{kk}^0)(\tau) \end{aligned} \quad (9)$$

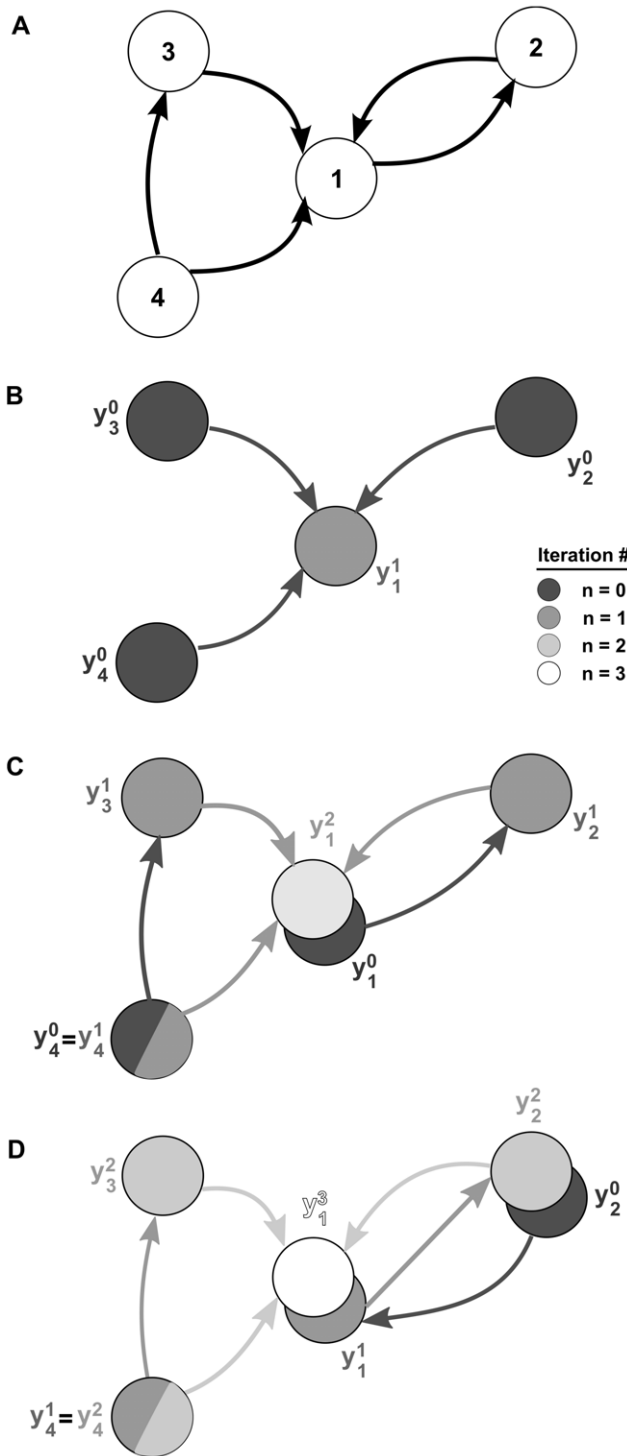
where we used  $f^-(t) = f(-t)$ . Ostojic et al. obtained an approximation closely related to Eq. (9). [24] They first obtained the cross-correlation between a pair of neurons which either receive a common input or share a monosynaptic connection. This can be done using Eq. (4), without the need to introduce the mixed process given in Eq. (5). Ostojic et al. then implicitly assumed that the correlations not due to one of these two submotifs could be disregarded. The correlation between pairs of cells which were mutually coupled (or were unidirectionally coupled with common input) was approximated by the sum of correlations introduced by each submotif individually.

Equation (9) provides a first approximation to the joint spiking statistics of cells in a recurrent network. However, it captures only the effects of direct synaptic connections, represented by the second and third terms, and common input, represented by the last term in Eq. (9). The impact of larger network structures, such as loops and chains are not captured, although they may significantly impact cross-correlations [41–43]. Experimental studies have also shown that local cortical connectivity may not be fully random [44–46]. It is therefore important to understand the effects on network architecture on correlations.

We therefore propose an iterative approach which accounts for successively larger connectivity patterns in the network [32,33]. We again start with  $y_i^0(t)$ , a realization of a single spike train in isolation. Successive approximations to the output of cells in a recurrent network are defined by

$$y_i^{n+1}(t) = y_i^0(t) + \sum_j (\mathbf{K}_{ij} * [y_j^n - r_j])(t), \quad n \geq 0. \quad (10)$$

To compute the correction to the output of a neuron, in the first iteration we assume that its inputs come from a collection of isolated cells: When  $n = 1$ , Eq. (10) takes into account only inputs from immediate neighbors, treating each as disconnected from the rest of the network. The corrections in the second iteration are computed using the approximate cell responses obtained from the first iteration. Thus, with  $n = 2$ , Eq. (10) also accounts for the



**Figure 2. Iterative construction of the linear approximation to network activity.** (A) An example recurrent network. (B)–(D) A sequence of graphs determines the successive approximations to the output of neuron 1. Processes defined by the same iteration of Eq. (11) have equal color. (B) In the first iteration of Eq. (11), the output of neuron 1 is approximated using the unperturbed outputs of its neighbors. (C) In the second iteration the results of the first iteration are used to define the inputs to the neuron. For instance, the process  $y_2^1$  depends on the base process  $y_1^0$  which represents the unperturbed output of neuron 1. Neuron 4 receives no inputs from the rest of the network, and all approximations involve only its unperturbed output,

$y_4^0$ . (D) Cells 3 and 4 are not part of recurrent paths, and their contributions to the approximation are fixed after the second iteration. However, the recurrent connection between cells 1 and 2 implies that subsequent approximations involve contributions of directed chains of increasing length.

doi:10.1371/journal.pcbi.1002408.g002

impact of next nearest neighbors. Successive iterations include the impact of directed chains of increasing length: The isolated output from an independent collection of neurons is filtered through  $n$  stages to produce the corrected response (See Figure 2.)

Notation is simplified when this iterative construction is recast in matrix form to obtain

$$\begin{aligned} \mathbf{y}^{n+1}(t) &= \mathbf{y}^0(t) + (\mathbf{K} * [\mathbf{y}^n - \mathbf{r}])(t) \\ &= \mathbf{y}^0(t) + \sum_{k=1}^{n+1} (\mathbf{K}^{(k)} * [\mathbf{y}^0 - \mathbf{r}])(t), \quad n \geq 0, \end{aligned} \quad (11)$$

where  $\mathbf{y}^n(t) = [y_i^n(t)]$  and  $\mathbf{r} = [r_i]$  are length  $N$  column vectors, and  $\mathbf{K}^{(k)}$  represents a  $k$ -fold matrix convolution of  $\mathbf{K}$  with itself. We define the convolution of matrices in the Methods.

The  $n^{\text{th}}$  approximation to the matrix of cross-correlations can be written in terms of the interaction kernels,  $\mathbf{K}_{ij}$ , and the autocorrelations of the base processes  $\mathbf{y}^0$  as

$$\begin{aligned} C_{ij}(\tau) \approx C^n(\tau) &= \mathbf{E}[(\mathbf{y}^n(t+\tau) - \mathbf{r})(\mathbf{y}^n(t) - \mathbf{r})^T] \\ &= \sum_{k,l=0}^n (\mathbf{K}^{(k)} * \mathbf{C}^0 * (\mathbf{K}^{-})^{(lT)})(\tau), \quad n \geq 0, \end{aligned} \quad (12)$$

where  $\mathbf{K}^{-}(t) = \mathbf{K}(-t)$ ,  $\mathbf{X}^{(kT)} = (\mathbf{X}^{(k)})^T$ , and  $\mathbf{X}^{(k)}$  is the  $k$ -fold matrix convolution of  $\mathbf{X}$  with itself.

Eq. (12) can be verified by a simple calculation. First, Eq. (11) directly implies that

$$\mathbf{y}^n(t) = \mathbf{y}^0(t) + \sum_{k=1}^n (\mathbf{K}^k * [\mathbf{y}^0 - \mathbf{r}])(t), \quad n \geq 0,$$

which we may use to find, for each  $n \geq 0$ ,

$$\begin{aligned} C^n(\tau) &\equiv \mathbf{E}[(\mathbf{y}^n(t+\tau) - \mathbf{r})(\mathbf{y}^n(t) - \mathbf{r})^T] \\ &= \mathbf{E}[(\mathbf{y}^0(t+\tau) - \mathbf{r})(\mathbf{y}^0(t) - \mathbf{r})^T] \\ &\quad + \sum_{k=1}^n \mathbf{E}[(\mathbf{K}^k * [\mathbf{y}^0 - \mathbf{r}])(t+\tau)(\mathbf{y}^0(t) - \mathbf{r})^T] \\ &\quad + \sum_{k=1}^n \mathbf{E}[(\mathbf{y}^0(t+\tau) - \mathbf{r})(\mathbf{K}^k * [\mathbf{y}^0 - \mathbf{r}]^T)(t)] \\ &\quad + \sum_{k,l=1}^n \mathbf{E}[(\mathbf{K}^k * [\mathbf{y}^0 - \mathbf{r}])(t+\tau)(\mathbf{K}^l * [\mathbf{y}^0 - \mathbf{r}]^T)(t)] \\ &= \mathbf{C}^0(\tau) + \sum_{k=1}^n (\mathbf{K}^k * \mathbf{C}^0)(\tau) + \sum_{k=1}^n (\mathbf{C}^0 * (\mathbf{K}^{-})^{kT})(\tau) \\ &\quad + \sum_{k,l=1}^n (\mathbf{K}^k * \mathbf{C}^0 * (\mathbf{K}^{-})^{lT})(\tau). \end{aligned} \quad (13)$$

Since  $\mathbf{K}_{ij}^0(t) = \delta_{ij}\delta(t)$ , Eq. (13) is equivalent to Eq. (12).

If we apply the Fourier transform, to Eq. (12), we find that for each  $\omega$ ,

$$\begin{aligned}\tilde{\mathbf{C}}^n(\omega) &= \mathbf{E}[\tilde{\mathbf{y}}^n(\omega)\tilde{\mathbf{y}}^{n*}(\omega)] = \sum_{k,l=0}^n \tilde{\mathbf{K}}^k(\omega)\mathbf{E}[\tilde{\mathbf{y}}^0(\omega)\tilde{\mathbf{y}}^{0*}(\omega)](\tilde{\mathbf{K}}^*)^l(\omega) \\ &= \left( \sum_{k=0}^n \tilde{\mathbf{K}}^k(\omega) \right) \mathbf{E}[\tilde{\mathbf{y}}^0(\omega)\tilde{\mathbf{y}}^{0*}(\omega)] \\ &\quad \left( \sum_{l=0}^n (\tilde{\mathbf{K}}^*)^l(\omega) \right),\end{aligned}\quad (14)$$

where  $\mathbf{X}^*$  denotes the conjugate transpose of the matrix  $\mathbf{X}$ . As before, the zero-mean Fourier transforms  $\tilde{y}_i^n$  of the processes  $y_i^n$  are defined by  $\tilde{y}_i^n = \mathcal{F}[y_i^n - r_i]$ , and  $\tilde{f} = \mathcal{F}(f)$  for all other quantities.

Defining  $\Psi(\mathbf{X})$  to be the spectral radius of the matrix  $\mathbf{X}$ , when  $\Psi(\tilde{\mathbf{K}}) < 1$ , we can take the limit  $n \rightarrow \infty$  in Eq. (14) [47,48], to obtain an approximation to the full array of cross-spectra

$$\begin{aligned}\tilde{\mathbf{C}}(\omega) &\approx \tilde{\mathbf{C}}^\infty(\omega) = \lim_{n \rightarrow \infty} \tilde{\mathbf{C}}^n(\omega) \\ &= (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \tilde{\mathbf{C}}^0(\omega) (\mathbf{I} - \tilde{\mathbf{K}}^*(\omega))^{-1}.\end{aligned}\quad (15)$$

As noted previously, this generalizes the approach of Lindner et al. [26] (also see [13]). In the limit  $n \rightarrow \infty$ , directed paths of arbitrary length contribute to the approximation. Equation (15) therefore takes into account the full recurrent structure of the network. Note that Eq. (15) may be valid even when  $\Psi(\tilde{\mathbf{K}}) > 1$ . However, in this case the series in Eq. (14) do not converge, and hence the expansion of the correlations in terms of paths through the network is invalid. We confirmed numerically that  $\Psi(\tilde{\mathbf{K}}) < 1$  for all of the networks and parameters we considered.

Finally, consider the network response to external signals,  $\eta_i(t)$ , with zero mean and finite variance. The response of the neurons in the recurrent network can be approximated iteratively by

$$\mathbf{y}^{n+1} = \mathbf{y}^0 + \mathbf{K} * [\mathbf{y}^n - \mathbf{r}] + \mathbf{A} * \boldsymbol{\eta},$$

where  $\mathbf{A} = \text{diag}(A_i)$  and  $\boldsymbol{\eta}(t) = [\eta_i(t)]$ . External signals and recurrent synaptic inputs are both linearly filtered to approximate a cell's response, consistent with a generalization of Eq. (4). As in Eq. (12), the  $n^{\text{th}}$  approximation to the matrix of correlations is

$$\begin{aligned}\mathbf{C}(\tau) &\approx \mathbf{C}^n(\tau) = \sum_{k,l=0}^n (\mathbf{K}^{(k)} * \mathbf{C}^0 * (\mathbf{K}^-)^{(lT)})(\tau) + \\ &\quad \sum_{k,l=0}^{n-1} (\mathbf{K}^{(k)} * \mathbf{A} * \mathbf{C}^n * (\mathbf{A}^-) * (\mathbf{K}^-)^{(lT)})(\tau),\end{aligned}$$

where  $\mathbf{C}^h(\tau) = \mathbf{E}[\boldsymbol{\eta}(t+\tau)\boldsymbol{\eta}(t)^T]$  is the covariance matrix of the external signals. We can again take the Fourier transform and the limit  $n \rightarrow \infty$ , and solve for  $\tilde{\mathbf{C}}(\omega)$ . If  $\Psi(\tilde{\mathbf{K}}) < 1$ ,

$$\tilde{\mathbf{C}}^\infty(\omega) = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} (\tilde{\mathbf{C}}^0(\omega) + \tilde{\mathbf{A}}(\omega)\tilde{\mathbf{C}}^n(\omega)\tilde{\mathbf{A}}^*(\omega)) (\mathbf{I} - \tilde{\mathbf{K}}^*(\omega))^{-1}. \quad (16)$$

When the signals comprising  $\boldsymbol{\eta}$  are white (and possibly correlated) corrections must be made to account for the change in spectrum and response properties of the isolated cells [26,49,50] (See Methods).

We note that Eq. (11), which is the basis of our iterative approach, provides an approximation to the network's output which is of higher than first order in connection strength. This may seem at odds with a theory that provides a linear correction to a cell's

response, cf. Eq. (4). However, Eq. (11) does not capture nonlinear corrections to the response of individual cells, as the output of each cell is determined linearly from its input. It is the input that can contain terms of any order in connection strength stemming from directed paths of different lengths through the network.

We use the theoretical framework developed above to analyze the statistical structure of the spiking activity in a network of IF neurons described by Eq. (1). We first show that the cross-correlation functions between cells in two small networks can be studied in terms of contributions from directed paths through the network. We use a similar approach to understand the structure of correlations in larger all-to-all and random networks. We show that in networks where inhibition and excitation are tuned to exactly balance, only local interactions contribute to correlations. When such balance is broken by a relative elevation of inhibition, the result may be increased synchrony in the network. The theory also allows us to obtain averages of cross-correlation functions conditioned on connectivity between pairs of cells in random networks. Such averages can provide a tractable yet accurate description of the joint statistics of spiking in these networks.

The correlation structure is determined by the response properties of cells together with synaptic dynamics and network architecture. Network interactions are described by the matrix of synaptic filters,  $\mathbf{J}$ , given in Eq. (2), while the response of cell  $i$  to an input is approximated using its linear response kernel  $A_i$ . Synaptic dynamics, architecture, and cell responses are all combined in the matrix  $\mathbf{K}$ , where  $\mathbf{K}_{ij}$  describes the response of cell  $i$  to an input from cell  $j$  (See Eq. (1)). The correlation structure of network activity is approximated in Eq. (15) using the Fourier transforms of the interaction matrix,  $\mathbf{K}$ , and the matrix of unperturbed autocorrelations  $\mathbf{C}^0$ .

### Statistics of the response of microcircuits

We first consider a pair of simple microcircuits to highlight some of the features of the theory. We start with the three cell model of feed-forward inhibition (FFI) shown in Figure 3A [51]. The interaction matrix,  $\tilde{\mathbf{K}}(\omega)$ , has the form

$$\tilde{\mathbf{K}}(\omega) = \begin{pmatrix} 0 & 0 & 0 \\ \tilde{K}_{E_2E_1}(\omega) & 0 & \tilde{K}_{E_2I}(\omega) \\ \tilde{K}_{IE_1}(\omega) & 0 & 0 \end{pmatrix},$$

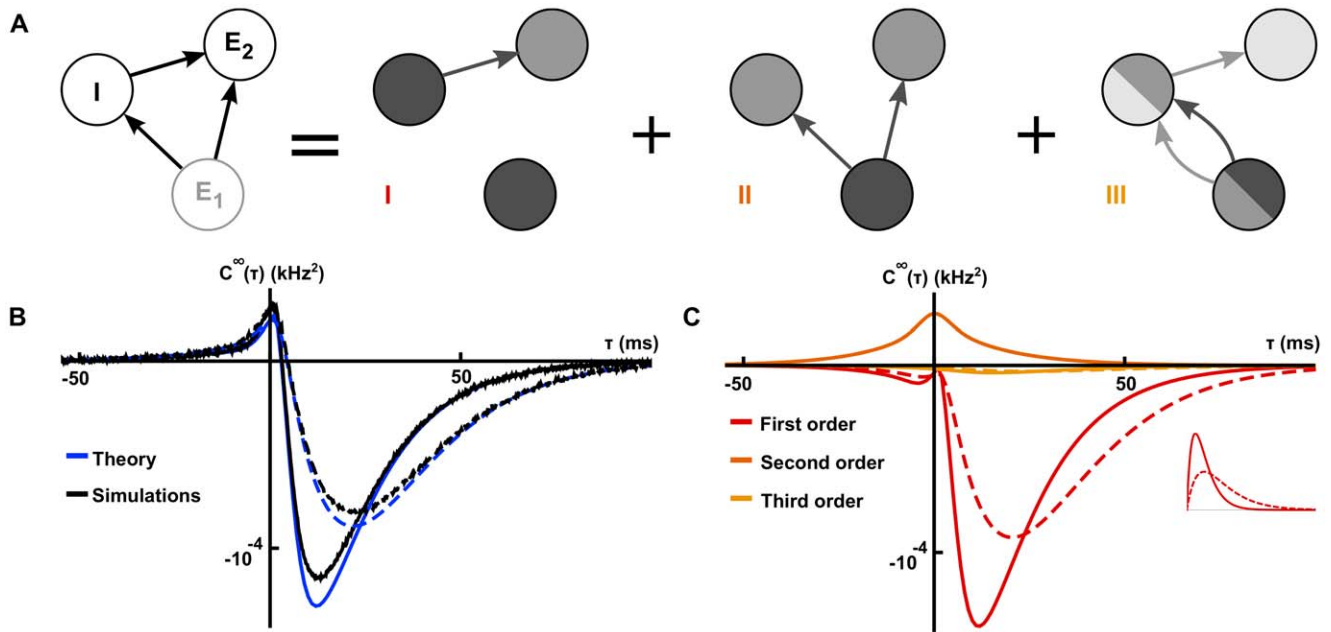
where cells are indexed in the order  $E_1, E_2, I$ . To simplify notation, we omit the dependence of  $\tilde{\mathbf{K}}(\omega)$  and other spectral quantities on  $\omega$ .

Note that  $\tilde{\mathbf{K}}$  is nilpotent of degree 3 (that is,  $\tilde{\mathbf{K}}^3 \equiv 0$ ), and the inverse of  $(\mathbf{I} - \tilde{\mathbf{K}})$  may be expressed as

$$(\mathbf{I} - \tilde{\mathbf{K}})^{-1} = (\mathbf{I} + \tilde{\mathbf{K}} + \tilde{\mathbf{K}}^2) = \begin{pmatrix} 1 & 0 & 0 \\ \tilde{K}_{E_2E_1} + \tilde{K}_{E_2I}\tilde{K}_{IE_1} & 1 & \tilde{K}_{E_2I} \\ \tilde{K}_{IE_1} & 0 & 1 \end{pmatrix}. \quad (17)$$

Substituting Eq. (17) into Eq. (15) (and noting that a similar equation as Eq. (17) holds for  $(\mathbf{I} - \tilde{\mathbf{K}}^*)^{-1}$ ) yields an approximation to the matrix of cross-spectra. For instance,

$$\begin{aligned}\tilde{\mathbf{C}}_{E_2I}^\infty &= \tilde{K}_{E_2I}\tilde{\mathbf{C}}_I^0 + \tilde{K}_{E_2E_1}\tilde{K}_{IE_1}^*\tilde{\mathbf{C}}_{E_1}^0 + \tilde{K}_{E_2I}|\tilde{K}_{IE_1}|^2\tilde{\mathbf{C}}_{E_1}^0 \\ &= \underbrace{(\tilde{A}_{E_2}\tilde{J}_{E_2I})\tilde{\mathbf{C}}_I^0}_{\text{I}} + \underbrace{(\tilde{A}_{E_2}\tilde{J}_{E_2E_1})(\tilde{A}_I\tilde{J}_{IE_1})^*\tilde{\mathbf{C}}_{E_1}^0}_{\text{II}} \\ &\quad + \underbrace{(\tilde{A}_{E_2}\tilde{J}_{E_2E_1})|\tilde{A}_I\tilde{J}_{IE_1}|^2\tilde{\mathbf{C}}_{E_1}^0}_{\text{III}}.\end{aligned}\quad (18)$$



**Figure 3. The relation between correlation structure and response statistics in a feed-forward inhibitory microcircuit.** (A) The FFI circuit (left) can be decomposed into three submotifs. Equation (18) shows that each submotif provides a specific contribution to the cross-correlation between cells  $E_2$  and  $I$ . (B) Comparison of the theoretical prediction with the numerically computed cross-correlation between cells  $E_2$  and  $I$ . Results are shown for two different values of the inhibitory time constant,  $\tau_I$  ( $\tau_I = 5$  ms, solid line,  $\tau_I = 10$  ms, dashed line). (C) The contributions of the different submotifs in panel A are shown for both  $\tau_I = 5$  ms (solid) and  $\tau_I = 10$  ms (dashed). Inset shows the corresponding change in the inhibitory synaptic filter. The present color scheme is used in subsequent figures. Connection strengths were  $\pm 40$  mV·ms for excitatory and inhibitory connections. In each case, the long window correlation coefficient  $\rho(\infty)$  between the two cells was  $\approx -0.18$ . doi:10.1371/journal.pcbi.1002408.g003

Figure 3B shows that these approximations closely match numerically obtained cross-correlations.  $\tilde{C}_X^0$  is the uncoupled power spectrum for cell  $X$ .

Equation (18) gives insight into how the joint response of cells in this circuit is shaped by the features of the network. The three terms in Eq. (18) are directly related to the architecture of the microcircuit: Term I represents the correlating effect of the direct input to cell  $E_2$  from cell  $I$ . Term II captures the effect of the common input from cell  $E_1$ . Finally, term III represents the interaction of the indirect input from  $E_1$  to  $E_2$  through  $I$  with the input from  $E_1$  to  $I$  (See Figure 3C). A change in any single parameter may affect multiple terms. However, the individual contributions of all three terms are apparent.

To illustrate the impact of synaptic properties on the cross-correlation between cells  $E_2$  and  $I$  we varied the inhibitory time constant,  $\tau_I$  (See Figure 3B and C). Such a change is primarily reflected in the shape of the first order term, I: Multiplication by  $\tilde{J}_{E_2 I}$  is equivalent to convolution with the inhibitory synaptic filter,  $J_{E_2 I}$ . The shape of this filter is determined by  $\tau_I$  (See Eq. (2)), and a shorter time constant leads to a tighter timing dependency between the spikes of the two cells [24,52–55]. In particular, Ostojic et al. made similar observations using a related approximation. In the FFI circuit, the first and second order terms, I and II, are dominant (red and dark orange, Figure 3B). The relative magnitude of the third order term, III (light orange, Figure 3B), is small. The next example shows that even in a simple recurrent circuit, terms of order higher than two may be significant.

More generally, the interaction matrices,  $\tilde{\mathbf{K}}$ , of recurrent networks are not nilpotent. Consider two reciprocally coupled excitatory cells,  $E_1$  and  $E_2$  (See Figure 4A, left). In this case,

$$\tilde{\mathbf{K}} = \begin{pmatrix} 0 & \tilde{K}_{E_1 E_2} \\ \tilde{K}_{E_2 E_1} & 0 \end{pmatrix}$$

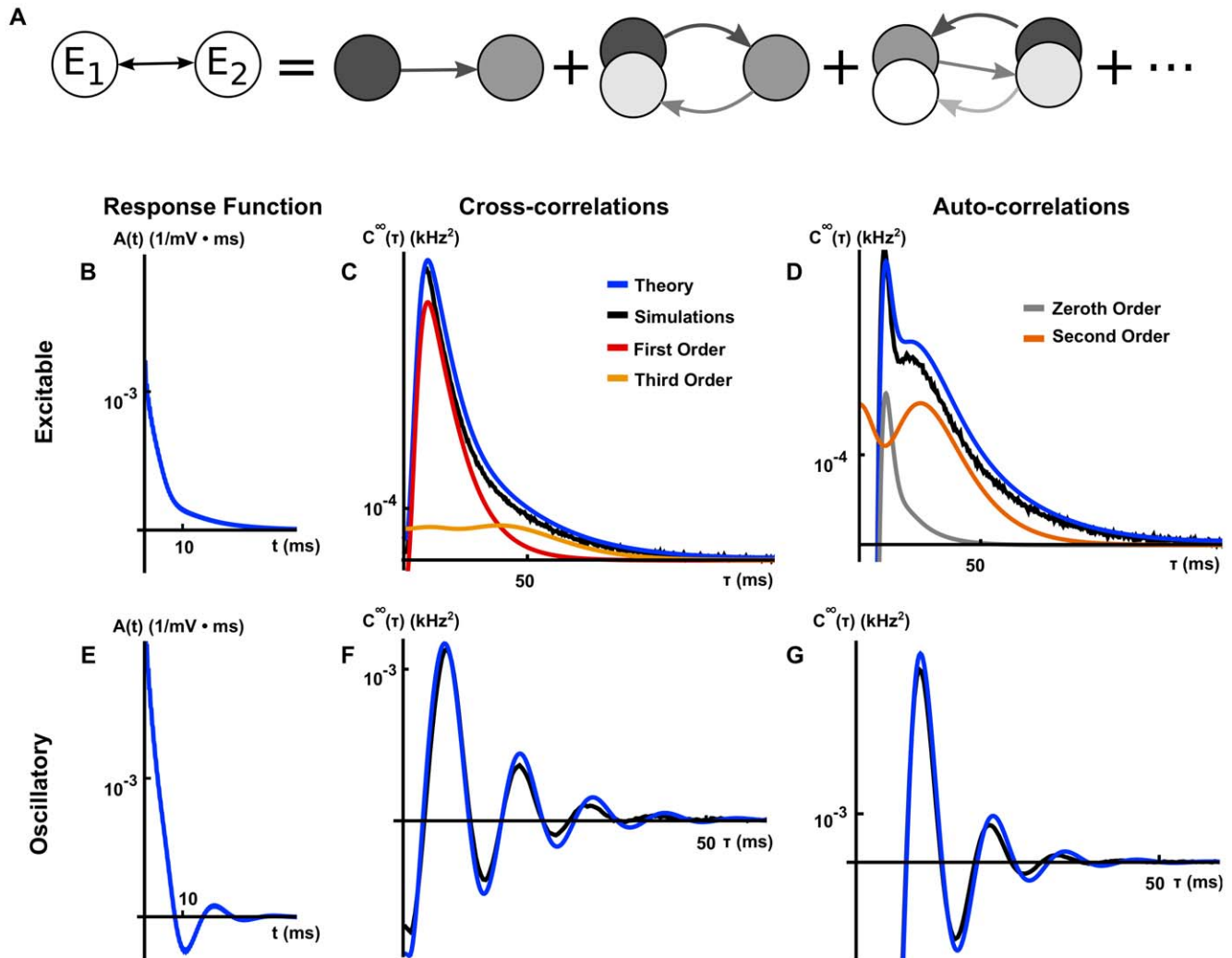
so that

$$(\mathbf{I} - \tilde{\mathbf{K}})^{-1} = \frac{1}{1 - \tilde{K}_{E_1 E_2} \tilde{K}_{E_2 E_1}} (\mathbf{I} + \tilde{\mathbf{K}}).$$

Equation (15) gives the following approximation to the matrix of cross-spectra

$$\begin{aligned} \tilde{\mathbf{C}}^\infty &= \frac{1}{|1 - \tilde{K}_{E_1 E_2} \tilde{K}_{E_2 E_1}|^2} (\mathbf{I} + \tilde{\mathbf{K}}) \begin{pmatrix} \tilde{C}_{E_1}^0 & 0 \\ 0 & \tilde{C}_{E_2}^0 \end{pmatrix} (\mathbf{I} + \tilde{\mathbf{K}}^*) \\ &= \frac{1}{|1 - \tilde{K}_{E_1 E_2} \tilde{K}_{E_2 E_1}|^2} \\ &\quad \begin{pmatrix} \tilde{C}_{E_1}^0 + |\tilde{K}_{E_1 E_2}|^2 \tilde{C}_{E_2}^0 & \tilde{K}_{E_2 E_1}^* \tilde{C}_{E_1}^0 + \tilde{K}_{E_1 E_2} \tilde{C}_{E_2}^0 \\ \tilde{K}_{E_2 E_1} \tilde{C}_{E_1}^0 + \tilde{K}_{E_1 E_2}^* \tilde{C}_{E_2}^0 & \tilde{C}_{E_2}^0 + |\tilde{K}_{E_2 E_1}|^2 \tilde{C}_{E_1}^0 \end{pmatrix}. \end{aligned} \quad (19)$$

In contrast to the previous example, this approximation does not terminate at finite order in interaction strength. After expanding, the cross-spectrum between cells  $E_1$  and  $E_2$  is approximated by



**Figure 4. The relation between correlation structure and response statistics for two bidirectionally coupled, excitatory cells.** (A) The cross-correlation between the two cells can be represented in terms of contributions from an infinite sequence of submotifs (See Eq. (20)). Though we show only a few “chain” motifs in one direction, one should note that there will also be contributions to the cross-correlation from chain motifs in the reverse direction in addition to indirect common input motifs (See the discussion of Figure 5). (B), (E) Linear response kernels in the excitable (B) and oscillatory (E) regimes. (C), (F) The cross-correlation function computed from simulations and theoretical predictions with first and third order contributions computed using Eq. (19) in the excitable (C) and oscillatory (F) regimes. (D), (G) The auto-correlation function computed from simulations and theoretical predictions with zeroth and second order contributions computed using Eq. (19) in the excitable (D) and oscillatory (G) regimes. In the oscillatory regime, higher order contributions were small relative to first order contributions and are therefore not shown. The network’s symmetry implies that cross-correlations are symmetric, and we only show them for positive times. Connection strengths were 40 mV·ms. The long window correlation coefficient  $\rho(\infty)$  between the two cells was  $\approx 0.8$  in the excitable regime and  $\approx 0.5$  in the oscillatory regime. The ISI CV was approximately 0.98 for neurons in the excitable regime and 0.31 for neurons in the oscillatory regime. doi:10.1371/journal.pcbi.1002408.g004

$$\tilde{C}_{E_1 E_2}^{\infty} = \sum_{k,l=0}^{\infty} (\tilde{K}_{E_1 E_2}^* \tilde{K}_{E_2 E_1})^k (\tilde{K}_{E_1 E_2}^* \tilde{K}_{E_2 E_1})^l \quad (20)$$

$$(\tilde{K}_{E_2 E_1}^* \tilde{C}_{E_1}^0 + \tilde{K}_{E_1 E_2}^* \tilde{C}_{E_2}^0).$$

Directed paths beginning at  $E_1$  and ending at  $E_2$  (or vice-versa) are of odd length. Hence, this approximation contains only odd powers of the kernels  $\tilde{K}_{E_i E_j}$ , each corresponding to a directed path from one cell to the other. Likewise, the approximate power spectra contain only even powers of the kernels corresponding to directed paths that connect a cell to itself (See Figure 4A).

The contributions of different sub-motifs to the cross- and auto-correlations are shown in Figures 4C, D when the isolated cells are in a near-threshold excitable state (CV  $\approx 0.98$ ). The auto-correlations are significantly affected by network interactions. We also note that chains of length two and three (the second and third submotifs in Figure 4A) provide significant contributions. Earlier approximations do not capture such corrections [24].

The operating point of a cell is set by its parameters ( $\tau_i, E_{L,i}$ , etc.) and the statistics of its input ( $E_i, \sigma_i$ ). A change in operating point can significantly change a cell’s response to an input. Using linear response theory, these changes are reflected in the response functions  $A_i$ , and the power spectra of the isolated cells,  $\tilde{C}^0$ . To highlight the role that the operating point plays in the approximation of the correlation structure given by Eq. (15), we



elevated the mean and decreased the variance of background noise by increasing  $E_i$  and decreasing  $\sigma_i$  in Eq. (1). With the chosen parameters the isolated cells are in a super-threshold, low noise regime and fire nearly periodically ( $CV \approx 0.31$ ). After the cells are coupled, this oscillatory behavior is reflected in the cross- and auto-correlations where the dominant contributions are due to first and zeroth order terms, respectively (See Figures 4F,G).

**Orders of coupling interactions.** It is often useful to expand Eq. (15) in terms of powers of  $\tilde{\mathbf{K}}$  [31]. The term  $\tilde{\mathbf{K}}^n \tilde{\mathbf{C}}^0 (\tilde{\mathbf{K}}^*)^m$  in the expansion is said to be of order  $n+m$ . Equivalently, in the expansion of  $\tilde{\mathbf{C}}_{ij}^0$ , the order of a term refers to the sum of the powers of all constituent interaction kernels  $\tilde{\mathbf{K}}_{ab}$ . We can also associate a particular connectivity submotif with each term. In particular,  $n^{\text{th}}$  order terms of the form

$$\tilde{\mathbf{K}}_{ia_{n-1}} \tilde{\mathbf{K}}_{a_{n-1}a_{n-2}} \cdots \tilde{\mathbf{K}}_{a_1 j} \tilde{\mathbf{C}}_{jj}^0$$

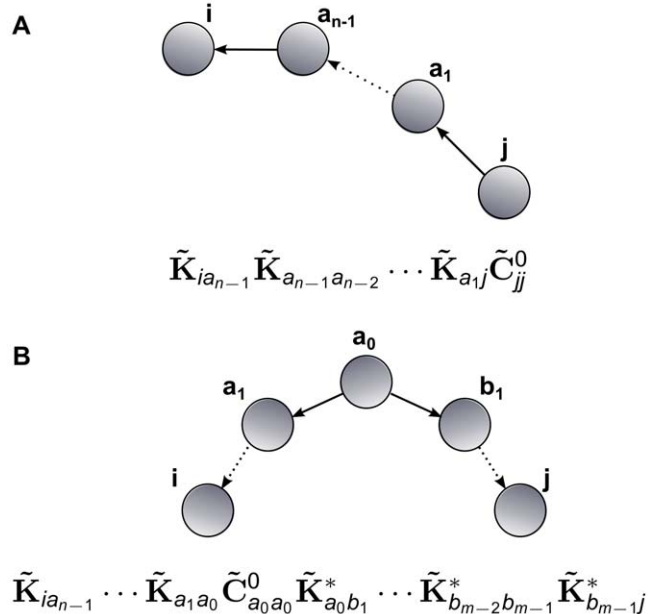
are associated with a directed path  $j \rightarrow a_1 \rightarrow \cdots \rightarrow a_{n-2} \rightarrow a_{n-1} \rightarrow i$  from cell  $j$  to cell  $i$ . Similarly, the term  $\tilde{\mathbf{C}}_{ii}^0 \tilde{\mathbf{K}}_{ia_1}^* \cdots \tilde{\mathbf{K}}_{a_{n-2}a_{n-1}}^* \tilde{\mathbf{K}}_{a_{n-1}j}^*$  corresponds to a  $n$ -step path from cell  $i$  to cell  $j$ . An  $(n+m)^{\text{th}}$  order term of the form

$$\tilde{\mathbf{K}}_{ia_{n-1}} \tilde{\mathbf{K}}_{a_{n-1}a_{n-2}} \cdots \tilde{\mathbf{K}}_{a_1 a_0} \tilde{\mathbf{C}}_{a_0 a_0}^0 \tilde{\mathbf{K}}_{a_0 b_1}^* \cdots \tilde{\mathbf{K}}_{b_{m-2}b_{m-1}}^* \tilde{\mathbf{K}}_{b_{m-1}j}^*$$

represents the effects of an indirect common input  $n$  steps removed from cell  $i$  and  $m$  steps removed from cell  $j$ . This corresponds to a submotif of the form  $i \leftarrow a_{n-1} \leftarrow \cdots \leftarrow a_0 \rightarrow b_1 \rightarrow \cdots \rightarrow b_{m-1} \rightarrow j$  consisting of two branches originating at cell  $a_0$ . (See Figure 5, and also Figure 6A and the discussion around Eqs. (18,20).)

### Statistics of the response of large networks

The full power of the present approach becomes evident when analyzing the activity of larger networks. We again illustrate the theory using several examples. In networks where inhibition and excitation are tuned to be precisely balanced, the theory shows



**Figure 5. The motifs giving rise to terms in the expansion of Eq. (15).** (A) Terms containing only unconjugated (or only conjugated) interaction kernels  $\tilde{\mathbf{K}}_{ab}$  correspond to directed chains. (B) Terms containing both unconjugated and conjugated interaction kernels  $\tilde{\mathbf{K}}_{ab}$  correspond to direct or indirect common input motifs.

doi:10.1371/journal.pcbi.1002408.g005

that only local interactions contribute to correlations. When this balance is broken, terms corresponding to longer paths through the network shape the cross-correlation functions. One consequence is that a relative increase in inhibition can lead to elevated network synchrony. We also show how to obtain tractable and accurate approximation of the average correlation structure in random networks.

**A symmetric, all-to-all network of excitatory and inhibitory neurons.** We begin with an all-to-all coupled network of  $N$  identical cells. Of these cells,  $N_E$  make excitatory, and  $N_I$  make inhibitory synaptic connections. The excitatory cells are assigned indices  $1, \dots, N_E$ , and the inhibitory cells indices  $N_E + 1, \dots, N$ . All excitatory (inhibitory) synapses have weight  $W_E = \frac{G_E}{N_E}$  ( $W_I = \frac{G_I}{N_I}$ ), and timescale  $\tau_E$  ( $\tau_I$ ). The interaction matrix  $\tilde{\mathbf{K}}$  may then be written in block form,

$$\tilde{\mathbf{K}} = \tilde{\mathbf{A}} \tilde{\mathbf{J}}, \quad \text{where} \quad \tilde{\mathbf{J}} = \begin{pmatrix} \tilde{\mathbf{J}}_E \mathbf{1}_{N_E N_E} & \tilde{\mathbf{J}}_I \mathbf{1}_{N_E N_I} \\ \tilde{\mathbf{J}}_E \mathbf{1}_{N_I N_E} & \tilde{\mathbf{J}}_I \mathbf{1}_{N_I N_I} \end{pmatrix}.$$

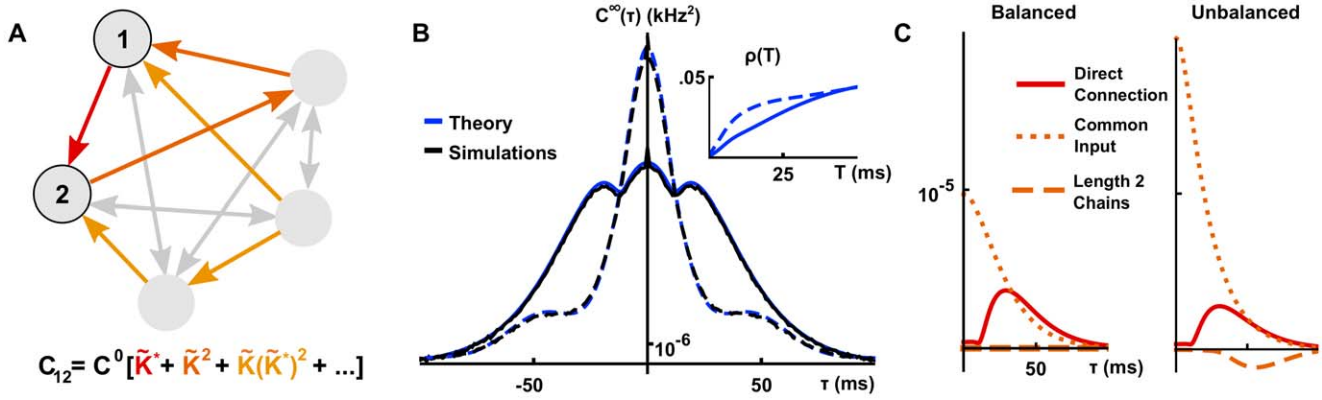
Here  $\mathbf{1}_{N_1 N_2}$  is the  $N_1 \times N_2$  matrix of ones,  $\tilde{\mathbf{J}}_X$  is the weighted synaptic kernel for cells of class  $X$  (assumed identical within each class), and  $\tilde{\mathbf{A}}$  is the susceptibility function for each cell in the network. Although the effect of autaptic connections (those from a cell to itself) is negligible (See Figure S2 in Text S1), their inclusion significantly simplifies the resulting expressions.

We define  $\tilde{\mu}_E = N_E \tilde{\mathbf{J}}_E$ ,  $\tilde{\mu}_I = N_I \tilde{\mathbf{J}}_I$ , and  $\tilde{\mu} = \tilde{\mu}_E + \tilde{\mu}_I$ . Using induction, we can show that

$$\tilde{\mathbf{K}}^k = \tilde{\mathbf{A}}^k \tilde{\mu}^{k-1} \tilde{\mathbf{J}}.$$

Direct matrix multiplication yields

$$\tilde{\mathbf{J}} \tilde{\mathbf{J}}^* = \tilde{\mu}_c \mathbf{1}_{NN} \quad \text{where} \quad \tilde{\mu}_c = N_E |\tilde{\mathbf{J}}_E|^2 + N_I |\tilde{\mathbf{J}}_I|^2,$$



**Figure 6. All-to-all networks and the importance of higher order motifs.** (A) Some of the submotifs contributing to correlations in the all-to-all network. (B) Cross-correlations between two excitatory cells in an all-to-all network ( $N_E = 80, N_I = 20$ ) obtained using Eq. (21) (Solid – precisely tuned network with  $\tilde{\mu} \equiv 0$  [ $G_E = -G_I = 140 \text{ mVms}, \tau_E = \tau_I = 10 \text{ ms}$ ], dashed – non-precisely tuned network with  $\tilde{\mu} \neq 0$  [ $G_E = 168 \text{ mVms}, G_I = -210 \text{ mVms}, \tau_E = 10 \text{ ms}, \tau_I = 5 \text{ ms}$ ]). (C) Comparison of first and second order contributions to the cross-correlation function in panel A in the precisely tuned (left) and non-precisely tuned (right) network. In both cases, the long window correlation coefficient  $\rho(\infty)$  was 0.05. doi:10.1371/journal.pcbi.1002408.g006

which allows us to calculate the powers  $\tilde{\mathbf{K}}^k \tilde{\mathbf{K}}^{l*}$  when  $k, l \neq 0$ ,

$$\tilde{\mathbf{K}}^k \tilde{\mathbf{K}}^{l*} = \tilde{\mathbf{A}}^k (\tilde{\mathbf{A}}^*)^l \tilde{\mu}^{k-1} (\tilde{\mu}^*)^{l-1} \tilde{\mu}_c \mathbf{1}_{NN}.$$

An application of Eq. (15) then gives an approximation to the matrix of cross-spectra:

$$\tilde{\mathbf{C}}^\infty = \tilde{\mathbf{C}}^0 \sum_{k,l=0}^{\infty} \tilde{\mathbf{K}}^k \tilde{\mathbf{K}}^{l*} = \tilde{\mathbf{C}}^0 \left[ \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right) \tilde{\mathbf{J}} + \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right)^* \tilde{\mathbf{J}}^* + \left| \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right|^2 \tilde{\mu}_c \mathbf{1}_{NN} + \mathbf{I}_N \right] \quad (21)$$

The cross-spectrum between two cells in the network is therefore given by

$$\tilde{\mathbf{C}}_{ij}^\infty]_{i \in X, j \in Y} = \tilde{\mathbf{C}}^0 \left[ \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right) \frac{\tilde{\mu}_Y}{N_Y} + \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right)^* \frac{\tilde{\mu}_X^*}{N_X} + \left| \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\mu}} \right|^2 \tilde{\mu}_c + \delta_{ij} \right], \quad (22)$$

where  $X, Y \in \{E, I\}$ . In Eq. (22) the first two terms represent the effects of all unidirectional chains originating at cell  $j$  and terminating at cell  $i$ , and vice versa. To see this, one should expand the denominators as power series in  $\tilde{\mathbf{A}}\tilde{\mu}$ . The third term represents the effects of direct and indirect common inputs to the two neurons, which can be seen by expanding this denominator as a product of power series in  $\tilde{\mathbf{A}}\tilde{\mu}$  and  $(\tilde{\mathbf{A}}\tilde{\mu})^*$ . In Figure 6A, we highlight a few of these contributing motifs.

Interestingly, when excitation and inhibition are tuned for precise balance (so that the mean excitatory and inhibitory synaptic currents cancel, and  $\tilde{\mu} = \tilde{\mu}_E + \tilde{\mu}_I = 0$ ). Using  $\tilde{\mu} = 0$  in Eq. (22) yields

$$[\tilde{\mathbf{C}}^\infty]_{i \in X, j \in Y} = \tilde{\mathbf{C}}^0 \left[ \tilde{\mathbf{A}} \frac{\tilde{\mu}_Y}{N_Y} + \tilde{\mathbf{A}}^* \frac{\tilde{\mu}_X^*}{N_X} + |\tilde{\mathbf{A}}|^2 \tilde{\mu}_c + \delta_{ij} \right]. \quad (23)$$

Effects of direct connections between the cells are captured by the first two terms, while those of direct common inputs to the pair are captured by the third term. Contributions from other paths do not

appear at any order. In other words, *in the precisely balanced case only local interactions contribute to correlations.*

To understand this cancellation intuitively, consider the contribution of directed chains originating at a given excitatory neuron,  $j$ . For  $\tau > 0$ , the cross-correlation function,  $\mathbf{C}_{ij}(\tau)$ , is determined by the change in firing rate of cell  $i$  at time  $\tau$  given a spike in cell  $j$  at time 0. By the symmetry of the all-to-all connectivity and stationarity, the firing of cell  $j$  has an equal probability of eliciting a spike in any excitatory or inhibitory cell in the network. Due to the precise synaptic balance, the postsynaptic current generated by the elicited spikes in the excitatory population will cancel the postsynaptic current due to elicited spikes in the inhibitory population on average. The contribution of other motifs cancel in a similar way.

In Figure 6B, we show the impact of breaking this excitatory-inhibitory balance on cross-correlation functions. We increased the strength and speed of the inhibitory synapses relative to excitatory synapses, while holding constant, for sake of comparison, the long window correlation coefficients  $\rho(\infty)$  between excitatory pairs (note that, by symmetry, all excitatory pairs should have the same correlation coefficient). Moreover, the degree of network synchrony, characterized by the short window correlation coefficients, is increased (See Figure 6B inset). Intuitively, a spike in one of the excitatory cells transiently increases the likelihood of spiking in all other cells in the network. Since inhibition in the network is stronger and faster than excitation, these additional spikes will transiently decrease the likelihood of spiking in twice removed cells.

Linear response theory allows us to confirm this heuristic observation, and quantify the impact of the imbalance on second order statistics. Expanding Eq. (22) for two excitatory cells to second order in coupling strength, we find

$$\tilde{\mathbf{C}}_{E_i E_j}^\infty = \tilde{\mathbf{C}}^0 \left[ \tilde{\mathbf{A}} \frac{\tilde{\mu}_E}{N_E} + \tilde{\mathbf{A}}^* \frac{\tilde{\mu}_E^*}{N_E} + \underbrace{|\tilde{\mathbf{A}}|^2 \tilde{\mu}_c + \tilde{\mathbf{A}}^2 \tilde{\mu}_c \frac{\tilde{\mu}_E}{N_E}}_{\text{}} + \underbrace{(\tilde{\mathbf{A}}^*)^2 \tilde{\mu}^* \frac{\tilde{\mu}_E^*}{N_E}}_{\text{}} + \delta_{ij} \right] + \mathcal{O}(|\tilde{\mathbf{K}}|^3). \quad (24)$$

Compared to the balanced case, there is no longer a complete cancellation between contributions of chains involving excitatory and inhibitory cells, and the two underlined terms appear as a result (compare with Eq. (23)). These terms capture the effects of

all length two chains between cells  $E_i$  or  $E_j$ , starting at one and terminating at the other. The relative strengthening of inhibition implies that chains of length two provide a negative contribution to the cross-correlation function at short times (*cf.* [56], see the dashed orange lines in Figure 6C). Additionally, the impact of direct common input to cells  $E_i$  and  $E_j$  on correlations is both larger in magnitude (because we increased the strength of both connection types) and sharper (the faster inhibitory time constant means common inhibitory inputs induce sharper correlations). These changes are reflected in the shape of the second order, common input term  $|\bar{A}|^2 \bar{\mu}_c$  in Eq. (24) (see dotted orange lines in Figure 6C).

In sum, unbalancing excitatory and inhibitory connections via stronger, faster inhibitory synapses enhances synchrony, moving a greater proportion of the covariance mass closer to  $\tau=0$  (See Figure 6B). To illustrate this effect in terms of underlying connectivity motifs, we show the contributions of length two chains and common input in both the precisely tuned and non-precisely tuned cases in Figure 6C. A similar approach would allow us to understand the impact of a wide range of changes in cellular or synaptic dynamics on the structure of correlations across networks.

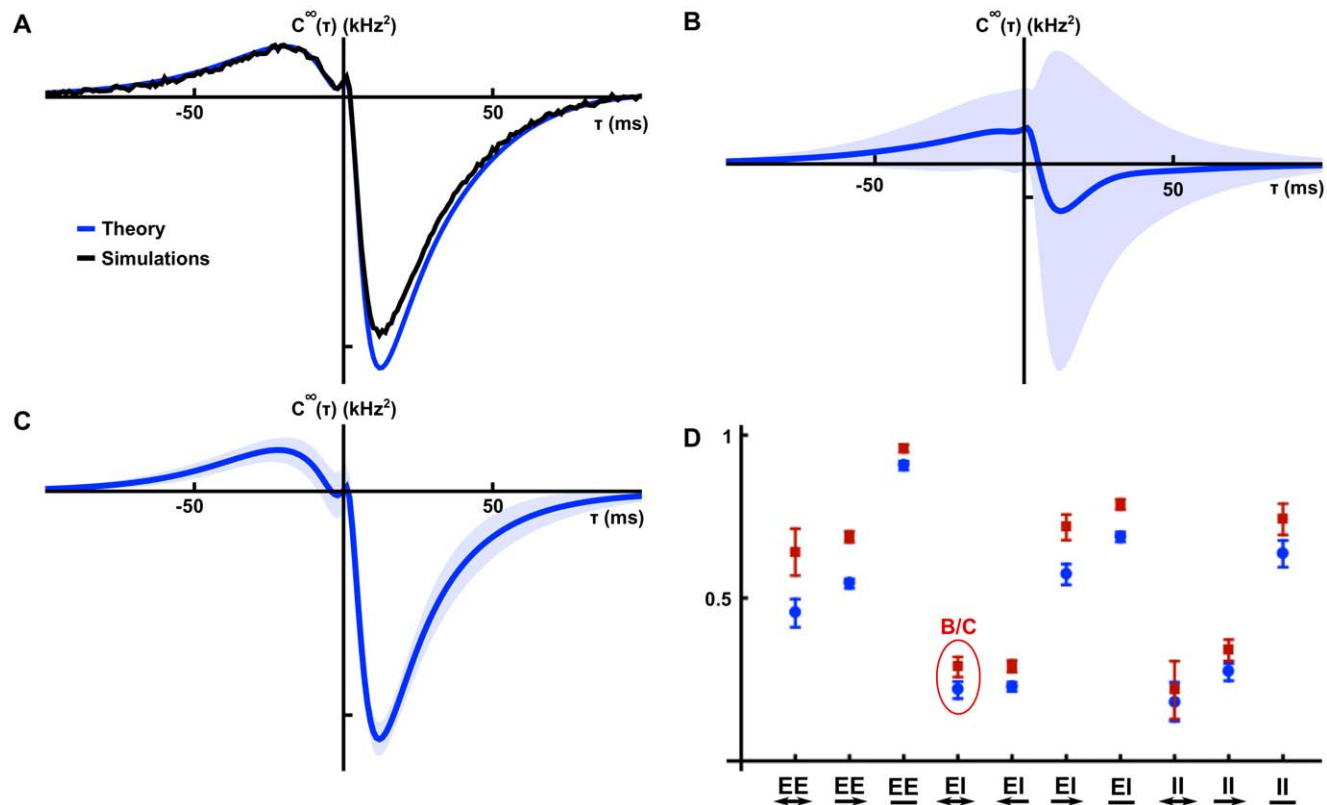
**Random, fixed in-degree networks of homogeneous excitatory and inhibitory neurons.** Connectivity in cortical

neuronal networks is typically sparse, and connection probabilities can follow distinct rules depending on area and layer [57]. The present theory allows us to consider arbitrary architectures, as we now illustrate.

We consider a randomly connected network of  $N_E$  excitatory and  $N_I$  inhibitory cells coupled with probability  $p$ . To simplify the analysis, every cell receives exactly  $pN_E$  excitatory and  $pN_I$  inhibitory inputs. Thus, having fixed in-degree (that is, the number of inputs is fixed and constant across cells), each cell receives an identical level of mean synaptic input. In addition, we continue to assume that cells are identical. Therefore, the response of each cell in the network is described by the same linear response kernel. The excitatory and inhibitory connection strengths are  $G_E/(pN_E)$  and  $G_I/(pN_I)$ , respectively. The timescales of excitation and inhibition may differ, but are again identical for cells within each class.

The approximation of network correlations (Eq. (15)) depends on the realization of the connectivity matrix. For a fixed realization, the underlying equations can be solved numerically to approximate the correlation structure (See Figure 7A). However, the cross-correlation between a pair of cells of given types has a form which is easy to analyze when only leading order terms in  $1/N$  are retained.

Specifically, the average cross-spectrum for two cells of given types is (See Section 1 in Text S1)



**Figure 7. Correlations in random, fixed in-degree networks.** (A) A comparison of numerically obtained excitatory-inhibitory cross-correlations to the approximation given by Eq. (26). (B) Mean and standard deviation for the distribution of correlation functions for excitatory-inhibitory pairs of cells. (Solid line – mean cross-correlation, shaded area – one standard deviation from the mean, calculated using bootstrapping in a single network realization). (C) Mean and standard deviation for the distribution of cross-correlation functions conditioned on cell type and first order connectivity for a reciprocally coupled excitatory-inhibitory pair of cells. (Solid line – mean cross-correlation function, shaded area – one standard deviation from the mean found by bootstrapping). (D) Average reduction in  $L^2$  error between cross-correlation functions and their respective first-order conditioned averages, relative to the error between the cross-correlations and their cell-type averages. Blue circles give results for a precisely tuned network, and red squares for a network with stronger, faster inhibition. Error bars indicate two standard errors above and below the mean.  $G_E, G_I, \tau_E, \tau_I$  for panels A-C are as in the precisely tuned network of Figure 6, and the two networks of panel D are as in the networks of the same figure. doi:10.1371/journal.pcbi.1002408.g007

$$\mathbf{E}\left\{\tilde{\mathbf{C}}_{ij}^{\infty}\right\}_{i \in X, j \in Y} = \tilde{\mathbf{C}}^0 \left[ \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right) \frac{\tilde{\boldsymbol{\mu}}_Y}{N_Y} + \left( \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right)^* \frac{\tilde{\boldsymbol{\mu}}_X^*}{N_X} + \left| \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right|^2 \tilde{\boldsymbol{\mu}}_c \right] + \mathcal{O}(1/N^2), \quad (25)$$

when  $i \neq j$ . This shows that, to leading order in  $1/N$ , the mean cross-spectrum between two cells in given classes equals that in the all-to-all network (see Eq. (22)). Therefore our previous discussion relating network architecture to the shape of cross-correlations in the all-to-all network extends to the average correlation structure in the random network for large  $N$ .

Pernice et al. [31] derived similar expressions for the correlation functions in networks of interacting *Hawkes processes* [58,59], which are linear, self-exciting point processes with history-dependent intensities. They assumed that either the network is regular (i.e., both in- and out-degrees are fixed) or has a sufficiently narrow degree distribution. Our analysis depends on having fixed in-degrees, and we do not assume that networks are fully regular. Both approaches lead to results that hold approximately (for large enough  $N$ ) when the in-degree is not fixed.

**Average correlations between cells in the random network conditioned on first order connectivity.** As Figure 7B shows there is large variability around the mean excitatory-inhibitory cross-correlation function given by the leading order term of Eq. (25). Therefore, understanding the average cross-correlation between cells of given types does not necessarily provide much insight into the mechanisms that shape correlations on the level of individual cell pairs. Instead, we examine the average correlation between a pair of cells conditioned on their first order (direct) connectivity.

We derive expressions for first order conditional averages correct to  $\mathcal{O}(1/N^2)$  (See Section 2 in Text S1). The average cross-spectrum for a pair of cells with indices  $i \neq j$ , conditioned on the value of the direct connections between them is

$$\mathbf{E}\left\{\tilde{\mathbf{C}}_{ij}^{\infty} \tilde{\mathbf{J}}_{ij} \tilde{\mathbf{J}}_{ji}\right\}_{i \in X, j \in Y} = \tilde{\mathbf{C}}^0 \left[ \underline{\tilde{\mathbf{A}}\tilde{\mathbf{J}}_{ij} + \tilde{\mathbf{A}}^* \tilde{\mathbf{J}}_{ji}^*} + \left( \frac{\tilde{\mathbf{A}}^2 \tilde{\boldsymbol{\mu}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right) \frac{\tilde{\boldsymbol{\mu}}_Y}{N_Y} + \left( \frac{\tilde{\mathbf{A}}^2 \tilde{\boldsymbol{\mu}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right)^* \frac{\tilde{\boldsymbol{\mu}}_X^*}{N_X} + \left| \frac{\tilde{\mathbf{A}}}{1 - \tilde{\mathbf{A}}\tilde{\boldsymbol{\mu}}} \right|^2 \tilde{\boldsymbol{\mu}}_c \right] + \mathcal{O}(1/N^2). \quad (26)$$

Here we set  $\tilde{\mathbf{J}}_{ij} = \mathbf{0}$  if we condition on the absence of a connection  $j \rightarrow i$ , and  $\tilde{\mathbf{J}}_{ij} = \tilde{\mathbf{J}}_Y/p$  if we condition on its presence. The term  $\tilde{\mathbf{J}}_{ji}$  is set similarly.

Although Eq. (26) appears significantly more complicated than the cell-type averages given in Eq. (25), they only differ in the underlined, first order terms. The magnitude of expected contributions from all higher order motifs is unchanged and coincides with those in the all-to-all network.

Figure 7C shows the mean cross-correlation function for mutually coupled excitatory-inhibitory pairs. Taking into account the mutual coupling significantly reduces variability (Compare with Figure 7B). To quantify this reduction, we calculate the mean reduction in variability when correlation functions are computed conditioned on the connectivity between the cells. For a single network, the relative decrease in variability can be quantified using

$$\mu_{\text{error}} = \frac{1}{N_T} \sum_{\substack{(i,j) \in T \\ i > j}} \frac{\|\mathbf{C}_{ij}(\tau) - C_T^{\text{FOC}}(\tau)\|_2}{\|\mathbf{C}_{ij}(\tau) - C_T^{\text{CT}}(\tau)\|_2},$$

where  $T$  represents pairs of cells of a given type and connection (in the present example these are reciprocally coupled excitatory-inhibitory pairs),  $N_T$  is the number of pairs of that type in the network,  $C_T^{\text{CT}}(\tau)$  is the leading order approximation of average correlations given only the type of cells in  $T$  (as in Eq. (25)), and  $C_T^{\text{FOC}}(\tau)$  the leading order approximation to average correlations conditioned on the first order connectivity of class  $T$  (as in Eq. (26)). We make use of the norm  $\|\cdot\|_2$  defined by  $\|f\|_2 = (\int |f|^2)^{1/2}$ . Figure 7D shows  $\mu_{\text{error}}$  averaged over twenty networks. In particular, compare the reduction in variability when conditioning on bidirectional coupling between excitatory-inhibitory pairs shown in Figures 7B,C, with the corresponding relative error in Figure 7D (circled in red).

## Discussion

We have extended and further developed a general theoretical framework that can be used to describe the correlation structure in a network of spiking cells. The application of linear response theory allows us to find tractable approximations of cross-correlation functions in terms of the network architecture and single cell response properties. The approach was originally used to derive analytical approximations to auto- and cross-spectra in an all-to-all inhibitory network in order to study the population response of the electrosensory lateral line lobe of weakly electric fish [26]. The key approximation relies on the assumption that the activity of cells in the network can be represented by a mixed point and continuous stochastic process, as given in Eq. (9). This approximation may be viewed as a generalization of classic Linear-Poisson models of neural spiking: the crucial difference is the replacement of the stationary firing rate by a realization of an integrate-and-fire spiking process. This allows for the retention of the underlying IF spiking activity while additionally posing that neurons act as perfect linear filters of their inputs. An iterative construction then leads to the expressions for approximate cross-correlations between pairs of cells given by Eq. (15).

The linear response framework of Lindner et al. [26] was extended by Marinazzo et al. [60] to somewhat more complex networks, and compared with other studies in which networks exhibit collective oscillations. In addition, other works [13,61,62] used linear response techniques to study information in the collective response of cells in a network. More recently, Ostojic et al. [24] obtained formulas for cross-correlations given in Eq. (9), which correspond to the first step in the iterative construction. Their approach captures corrections due to direct coupling (first order terms) and direct common input (second order terms involving second powers of interaction kernels; see also [49,63]). Our approach can be viewed as a generalization that also accounts for length two directed chains, along with all higher order corrections. As Figure 4 illustrates, these additional terms can be significant. The present approach also allows us to calculate corrected auto-correlations, in contrast with that of Ostojic et al.

Our work is also closely related to that of Pernice et al. [31], who analyzed the correlation structure in networks of interacting *Hawkes processes* [58,59]. Both studies represent correlations between cell pairs in terms of contributions of different connectivity motifs. However, our methods also differ: while their expressions are exact for Hawkes processes, Pernice et al. did not compare their results to those obtained using physiological models, and did not account for the response properties of individual cells (though it is possible that both can be achieved approximately by using appropriate kernels for the Hawkes processes). Moreover, for simplicity Pernice et al. examined only “total” spike count covariances, which are the integrals of the cross-correlation

functions. However, as they note, their approach can be extended to obtain the temporal structure of cross-correlations. Similarly, Toyozumi et al. [64] derive expressions for cross-correlations in networks of interacting point process models in the Generalized Linear Model (GLM) class. These are very similar to Hawkes processes, but feature a static nonlinearity that shapes the spike emission rate.

To illustrate the power of the present linear response theory in analyzing the factors that shape correlations, we considered a number of simple examples for which the approximation given by Eq. (15) is tractable. We showed how the theory can be used both to gain intuition about the network and cell properties that shape correlations, and to quantify their impact. In particular, we explained how only local connections affect correlations in a precisely tuned all-to-all network, and how strengthening inhibition may synchronize spiking activity. In each case, we use comparisons with integrate-and-fire simulations to show that linear response theory makes highly accurate predictions.

It may be surprising that *linear* response theory can be used to provide corrections to cross-correlations of arbitrary order in network connectivity. The key to why this works lies in the accuracy of the linearization. A more accurate approximation could be obtained by including second and higher order corrections to the approximate response of a single cell, as well as corrections to the joint response. While including such terms is formally necessary to capture all contributions of a given order in network connectivity [32,33], the success of linear response theory suggests that they are small for the cases at hand. In short, the present approximation neglects higher-order corrections to the approximate response of individual cells, along with all corrections involving joint responses, but accounts for paths through the network of arbitrary length.

As expected from the preceding discussion, simulations suggest that, for IF neurons, our approximations become less accurate as cells receive progressively stronger inputs. The physical reasons for this loss of accuracy could be related to interactions between the “hard threshold” and incoming synaptic inputs with short timescales. Additionally, while the theory will work for short synaptic timescales, it will improve for slower synaptic dynamics, limiting towards being essentially exact in the limit of arbitrarily long synaptic time constants (note the improvement in the approximation for the FFI circuit for the slower timescale exhibited in Figure 3). Another important factor is background noise, which is known to improve the accuracy of the linear description of single cell responses. We assume the presence of a white noise background, although it is possible to extend the present methods to colored background noise [25,65].

We found that linear response theory remains applicable in a wide range of dynamical regimes, including relatively low noise, superthreshold regimes where cells exhibit strong oscillatory behavior. Moreover, the theory can yield accurate approximations of strong correlations due to coupling: for the bidirectionally coupled excitatory circuit of Figure 4, the approximate cross-correlations match numerically obtained results even when correlation coefficients are large ( $\rho_{E_1 E_2}(\infty) \approx 0.8$  in the excitable regime,  $\approx 0.5$  in the oscillatory regime). Additional discussion of the limits of applicability of linear response to the computation of correlations in networks can be found in the Supplementary Information. There, we show that the approximation is valid over a range of physiological values in the case of the all-to-all network, and that the theory gives accurate predictions in the presence of low firing rates (see Figures S3, S4 in Text S1).

The limits of linear response approximations of time-dependent firing activity and correlations have been tested in a number of

other studies. Ostojic and Brunel [66] examined this accuracy in the relatively simple case of a neuron receiving filtered Gaussian noise in addition to a white background. Chacron et al. [61] noted that linear response approaches applied to networks of perfect integrators begin to display significant errors at larger connection strengths. Marinazzo et al. [60] remarked on the errors induced by network effects in linear response approximations to correlations in a delayed feedback loop. In particular, these errors were attributed to network effects such as synchrony in the excitatory population. The authors noted that such activity can not be correctly modeled by a linear approach.

Although we have demonstrated the theory using networks of integrate-and-fire neurons, the approach is widely applicable. The linear response kernel and power spectrum for a general integrate and fire neuron model can be easily obtained [29]. In addition, it is also possible to obtain the rate, spectrum, and susceptibility for modulation of the mean conductance in the case of conductance-based (rather than current-based) synapses (See [67] and Section 3 in Text S1). As the linear response kernel is directly related to the spike triggered average [24,30], the proposed theoretical framework should be applicable even to actual neurons whose responses are characterized experimentally.

The possibilities for future applications are numerous. For example, one open question is how well the theory can predict correlations in the presence of adaptive currents [67]. In addition, the description of correlations in terms of architecture and response properties suggests the possibility of addressing the difficult inverse problem of inferring architectural properties from correlations [22–24,64]. Ostojic et al. applied linear response methods to the latter problem. It is our hope that the present approach will prove a valuable tool in moving the computational neuroscience community towards a more complete understanding of the origin and impact of correlated activity in neuronal populations.

## Methods

### Measures of spike time correlation

We quantify dependencies between the responses of cells in the network using the spike train auto- and cross-correlation functions [39]. For a pair of spike trains,  $y_i(t), y_j(t)$ , the cross-correlation function  $\mathbf{C}_{ij}(\tau)$  is defined as

$$\mathbf{C}_{ij}(\tau) = \text{cov}(y_i(t+\tau), y_j(t)).$$

The auto-correlation function  $\mathbf{C}_{ii}(t)$  is the cross-correlation between a spike train and itself, and  $\mathbf{C}(t)$  is the matrix of cross-correlation functions. Denoting by  $N_{y_i}(t_1, t_2) = \int_{t_1}^{t_2} y_i(s) ds$  the number of spikes over a time window  $[t_1, t_2]$ , the spike count correlation,  $\rho_{ij}(T)$ , over windows of length  $\tau$  is defined as,

$$\rho_{ij}(T) = \frac{\text{cov}(N_{y_i}(t, t+T), N_{y_j}(t, t+T))}{\sqrt{\text{var}(N_{y_i}(t, t+T)) \text{var}(N_{y_j}(t, t+T))}}.$$

We assume stationarity of the spiking processes (that is, the network has reached a steady state) so that  $\rho_{ij}(T)$  does not depend on  $t$ . We also use the total correlation coefficient  $\rho_{ij}(\infty) = \lim_{T \rightarrow \infty} \rho_{ij}(T)$  to characterize dependencies between the processes  $y_i$  and  $y_j$  over arbitrarily long timescales.

The spike count covariance is related to the cross-correlation function by [7,68]

$$\text{cov}\left(N_{y_i}(t, t+T), N_{y_j}(t, t+T)\right) = \int_{-T}^T \mathbf{C}_{ij}(\tau)(T-|\tau|)d\tau.$$

We can interpret the cross-correlation as the conditional probability that cell  $i$  spikes at time  $t+\tau$  given that cell  $j$  spiked at time  $t$ . The conditional firing rate,

$$\mathbf{H}_{ij}(\tau) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \Pr\left(N_{y_i}(t+\tau, t+\tau+\Delta t) > 0 | N_{y_j}(t, t+\Delta t) > 0\right),$$

is the firing rate of cell  $i$  conditioned on a spike in cell  $j$  at  $\tau$  units of time in the past, and  $\mathbf{C}_{ij}(\tau) = r_j(\mathbf{H}_{ij}(\tau) - r_i)$ .

Define the Fourier transform of a function  $f(t)$  as  $\tilde{f}(\omega) = \mathcal{F}[f(t)](\omega) \equiv \int_{-\infty}^{\infty} f(t)e^{-2\pi i \omega t} dt$ . We will often make use of the cross-spectrum between the output of cells  $i, j$ , given by  $\tilde{\mathbf{C}}_{ij}(\omega) = \mathbf{E}[\tilde{y}_i(\omega)\tilde{y}_j^*(\omega)]$ , which is the Fourier transform of the cross-correlation function of cells  $i, j$ . The power spectrum  $\tilde{\mathbf{C}}_{ii}(\omega)$  is the cross-spectrum between a cell and itself, and is the Fourier transform of the auto-correlation function.

**Numerical methods.** Simulations were run in C++, and the stochastic differential equations were integrated with a standard Euler method with a time-step of 0.01 ms. General parameter values were as follows:  $\tau_i = 20\text{ms}$ ,  $E_{L,i} + E_i = -54\text{mV}$ ,  $\sigma_i = \sqrt{12}\text{mV}$ ,  $v_{th} = 20\text{mV}$ ,  $v_r = -54\text{mV}$ ,  $\tau_{ref} = 2\text{ms}$ ,  $V_T = -52.5\text{mV}$ ,  $\Delta_T = 1.4\text{mV}$ ,  $\tau_E = 10\text{ms}$ ,  $\tau_I = 5\text{ms}$ ,  $\tau_{D,i} = 1\text{ms}$ . Marginal statistics (firing rates, uncoupled power spectra and response functions) were obtained using the threshold integration method of [29] in MATLAB. We have posted a package of code which contains examples of all the numerical methods used in this paper (both simulations and theory) at <http://www.math.uh.edu/~josic/myweb/software.html>. Additional code is available upon request.

**Calculation of stationary rates in a recurrent network.** The stationary firing rate of an IF neuron can be computed as a function of the mean and intensity of internal noise  $(E_i, \sigma_i)$  and other cellular parameters  $(\tau_i, E_{L,i}, \dots)$  [69]. Denote the stationary firing rate of cell  $i$  in the network by  $r_i$ , and by  $r_{i,0}(E, \sigma)$  the stationary firing rate in the presence of white noise with mean  $E$  and variance  $\sigma^2$ . We keep the dependencies on other parameters are implicit. The stationary rates,  $r_i$ , in the recurrent network without external input are determined self-consistently by

$$r_i = r_{i,0}(E_i', \sigma_i) = r_{i,0}(E_i + \sum_j \mathbf{W}_{ij} r_j, \sigma_i) \quad i = 1, \dots, N,$$

where we used  $\mathbf{E}[f_i] = \sum_j \mathbf{W}_{ij} \mathbf{E}[y_j] = \sum_j \mathbf{W}_{ij} r_j$ . This equality holds because the synaptic kernels,  $\mathbf{J}_{ij}$ , were normalized to have area  $\mathbf{W}_{ij}$ . These equations can typically be solved by fixed-point iteration.

Note that this provides an effective mean input,  $E_i'$ , to each cell, but does not give adjustments to the variance,  $\sigma_i$ . We assume that the major impact of recurrent input is reflected in  $E_i'$ , and ignore corrections to the cell response involving higher order statistics of the input. This approach is valid as long as fluctuations in the recurrent input to each cell are small compared to  $\sigma_i$ , and may break down otherwise [27].

**Correction to statistics in the presence of an external white noise signals.** Expression (16) can be used to compute

## References

1. Cohen M, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819.
2. Ohki K, Chung S, Ch'ng Y, Kara P, Reid R (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433: 597–603.

the statistics of the network response to inputs  $\eta_i(t)$  of finite variance. As noted by [26], when inputs have infinite variance additional corrections are necessary. As a particular example, consider the case where the processes are correlated white noise, i.e., when  $\eta_i(t) = \sqrt{c}x_c(t) + \sqrt{1-c}x_i(t)$ , where  $x_c, x_i$  are independent white noise processes with variance  $\sigma_e$ . Then each  $\eta_i$  is also a white noise process with intensity  $\sigma_i^e$ , but  $\mathbf{E}[\eta_i(t+\tau)\eta_j(t)] = [\delta_{ij}\delta(\tau) + (1-\delta_{ij})c\delta(\tau)]\sigma_i^e$ . The firing rate of cell  $i$  in response to this input is  $r_i = r_0(E_i', \sqrt{(\sigma_i)^2 + (\sigma_i^e)^2})$ , and the point around which the response of the cell is linearized needs to be adjusted.

Finally, we may apply an additional correction to the linear response approximation of autocorrelations. For simplicity, we ignore coupling in Eq. (16) (so that  $\tilde{\mathbf{K}} = 0$ ). Linear response predicts that  $\tilde{\mathbf{C}}_{ii}(\omega) = \tilde{\mathbf{C}}_{ii}^0(\omega; \sigma_i^2) + (\sigma_i^e)^2 |\tilde{A}_i(\omega)|^2$ , where we have introduced explicit dependence on  $\sigma_i^2$ , the variance of white noise being received by an IF neuron with power spectrum  $\tilde{\mathbf{C}}_{ii}^0(\omega; \sigma_i^2)$ , in the absence of the external signal. The approximation may be improved in this case by making the following substitution in Eq. (16) [26,50]:

$$\tilde{\mathbf{C}}_{ii}^0(\omega; \sigma_i^2) + (\sigma_i^e)^2 |\tilde{A}_i(\omega)|^2 \rightarrow \tilde{\mathbf{C}}_{ii}^0(\omega; \sigma_i^2 + (\sigma_i^e)^2)$$

The response function  $A$  should be adjusted likewise.

**Convolution of matrices.** Let  $\mathbf{X}(t) = [X_{ij}(t)]$  and  $\mathbf{Y}(t) = [Y_{ij}(t)]$  be  $n_1 \times n_2$  and  $n_2 \times n_3$  matrices of functions, respectively. We define the convolution of matrices  $(\mathbf{X} * \mathbf{Y})(t)$  to be the  $n_1 \times n_3$  matrix of functions with entries defined by

$$(\mathbf{X} * \mathbf{Y})_{ij}(t) = \sum_k (X_{ik} * Y_{kj})(t).$$

Expectations and convolutions commute for matrix convolutions as matrix expectations are taken entry-wise. Each entry of a matrix convolution is a linear combination of scalar convolutions which commute with expectations. Additionally, we adopt the convention that the zeroth power of the interaction matrix,  $\mathbf{K}_{ij}^0(t)$ , is the diagonal matrix with  $\mathbf{K}_{ij}^0(t) = \delta(t)$  when  $i=j$ . Hence  $\mathbf{K}_{ij}^0(t)$  acts as the identity matrix under matrix convolution.

## Supporting Information

**Text S1** Supplementary information file containing derivations and additional content, such as an exploration of the error of the theory. Supporting information figures were included in this file (and not separately). (PDF)

## Acknowledgments

The authors thank Robert Rosenbaum, Brent Doiron and Srdjan Ostojic for many helpful discussions.

## Author Contributions

Conceived and designed the experiments: JT YH ESB KJ. Performed the experiments: JT. Analyzed the data: JT. Contributed reagents/materials/analysis tools: JT YH ESB KJ. Wrote the paper: JT YH ESB KJ.

3. Field G, Gauthier J, Sher A, Greschner M, Machado T, et al. (2010) Functional connectivity in the retina at the resolution of photoreceptors. *Nature* 467: 673–677.
4. Jia H, Rochefort N, Chen X, Konnerth A (2010) In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons. *Nat Protoc* 6: 28–35.
5. Brown EN, Kass RE, Mitra P (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci* 7: 456–61.
6. Averbeck B, Latham P, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366.
7. Shadlen M, Newsome W (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870–3896.
8. Panzeri S, Schultz S, Treves A, Rolls E (1999) Correlations and encoding of information in the nervous system. *P Roy Soc Lond B Bio* 266: 1001–1012.
9. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature* 370: 140–143.
10. Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11: 91–101.
11. Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E* 64: 051904.
12. Panzeri S, Petersen R, Schultz S, Lebedev M, Diamond M (2001) The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* 29: 769–777.
13. Beck J, Bejjanki V, Pouget A (2011) Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural Comput* 23: 1484–1502.
14. Josić K, Shea-Brown E, Doiron B, de La Rocha J (2009) Stimulus-dependent correlations and population codes. *Neural Comput* 21: 2774–2804.
15. Schneidman E, Bialek W, Berry M (2003) Synergy, Redundancy, and Independence in Population Codes. *J Neurosci* 23: 11539–11553.
16. Latham PE, Nirenberg S (2005) Synergy, Redundancy, and Independence in Population Codes, Revisited. *J Neurosci* 25: 5195–5206.
17. Nirenberg S, Carcieri SM, Jacobs AL, Latham PE (2001) Retinal ganglion cells act largely as independent encoders. *Nature* 411: 698–701.
18. Cohen M, Newsome W (2009) Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J Neurosci* 29: 6635–6648.
19. Romo R, Hernandez A, Zainos A, Salinas E (2003) Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38: 649–657.
20. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical circuits. *Science* 327: 587–590.
21. Ecker A, Berens P, Keliris G, Bethge M, Logothetis N, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327: 584–587.
22. Paninski L, Ahmadian Y, Ferreira D, Koyama S, Rahnema Rad K, et al. (2010) A new look at state-space models for neural data. *J Comput Neurosci* 29: 107–126.
23. Nykamp D (2007) A mathematical framework for inferring connectivity in probabilistic neuronal networks. *Math Biosci* 205: 204–251.
24. Ostojic S, Brunel N, Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J Neurosci* 29: 10234–10253.
25. Fourcaud-Trocmé N, Hansel D, Van Vreeswijk C, Brunel N (2003) How spike generation mechanisms determine the neuronal response to actuating inputs. *J Neurosci* 23: 11628–11640.
26. Lindner B, Doiron B, Longtin A (2005) Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys Rev E* 72: 061919.
27. Brunel N, Hakim V (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput* 11: 1621–1671.
28. Lindner B, Schimansky-Geier L (2001) Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys Rev Lett* 86: 2934–2937.
29. Richardson M (2007) Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive. *Phys Rev E* 76: 021919.
30. Gabbiani F, Koch C (1998) Principles of spike train analysis. In: Segev I, Koch C, eds. *Methods in Neuronal Modeling*. Cambridge: MIT Press. pp 313–360.
31. Pernice V, Staude B, Cardanobile S, Rotter S (2011) How structure determines correlations in neuronal networks. *PLoS Comput Biol* 7: e1002059.
32. Rangan A (2009) Diagrammatic expansion of pulse-coupled network dynamics. *Phys Rev Lett* 102: 158101.
33. Rangan A (2009) Diagrammatic expansion of pulse-coupled network dynamics in terms of subnetworks. *Phys Rev E* 80: 036101.
34. Risken H (1996) *The Fokker-Planck equation: Methods of solution and applications*. Berlin: Springer Verlag. 488 p.
35. Brunel N, Chance F, Fourcaud N, Abbott L (2001) Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys Rev Lett* 86: 2186–2189.
36. White JA, Rubinstein JT, Kay AR (2000) Channel noise in neurons. *Trends Neurosci* 23: 131–137.
37. Renart A, Brunel N, Wang X (2004) Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In: Feng J, ed. *Computational Neuroscience: A Comprehensive Approach*. Boca Raton: CRC Press. pp 431–490.
38. Burkitt A (2006) A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biol Cybern* 95: 1–19.
39. Gabbiani F, Cox S (2010) *Mathematics for Neuroscientists*. London: Academic Press. 498 p.
40. Barreiro A, Shea-Brown E, Thilo E (2010) Time scales of spike-train correlation for neural oscillators with common drive. *Phys Rev E* 81: 011916.
41. Roxin A (2011) The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. *Front Comput Neurosci* 5: 1–15.
42. Zhao L, Beverlin B, Netoff T, Nykamp DQ (2011) Synchronization from second order network connectivity statistics. *Front Comput Neurosci* 5: 1–16.
43. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10: 186–198.
44. Song S, Sjöström P, Reigl M, Nelson S, Chklovskii D (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol* 3: e68.
45. Oswald A, Doiron B, Rinzel J, Reyes A (2009) Spatial profile and differential recruitment of gabab modulate oscillatory activity in auditory cortex. *J Neurosci* 29: 10321.
46. Perin R, Berger T, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. *Proc Natl Acad Sci U S A* 108: 5419–5424.
47. Katō T (1995) *Perturbation Theory for Linear Operators* Springer-Verlag. 640 p.
48. Horn R, Johnson C (1990) *Matrix Analysis*. Cambridge: Cambridge University Press. 575 p.
49. de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. *Nature* 448: 802–806.
50. Vilela R, Lindner B (2009) Comparative study of different integrate-and-fire neurons: Spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys Rev E* 80: 031909.
51. Kremkow J, Perrinet L, Masson G, Aertsen A (2010) Functional consequences of correlated excitatory and inhibitory conductances in cortical networks. *J Comp Neurosci* 28: 579–594.
52. Veredas F, Vico F, Alonso J (2005) Factors determining the precision of the correlated firing generated by a monosynaptic connection in the cat visual pathway. *J Physiol* 567: 1057.
53. Kirkwood P (1979) On the use and interpretation of cross-correlations measurements in the mammalian central nervous system. *J Neurosci Meth* 1: 107.
54. Fetz E, Gustafsson B (1983) Relation between shapes of post-synaptic potentials and changes in firing probability of cat motoneurons. *J Physiol* 341: 387.
55. Herrmann A, Gerstner W (2001) Noise and the PSTH response to current transients: I. general theory and application to the integrate-and-fire neuron. *J Comput Neurosci* 11: 135–151.
56. Vreeswijk C, Abbott L, Bard Ermentrout G (1994) When inhibition not excitation synchronizes neural firing. *J Comput Neurosci* 1: 313–321.
57. Shepherd G (1991) *Foundations of the Neuron Doctrine*. Oxford: Oxford University Press. 352 p.
58. Hawkes A (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58: 83–90.
59. Hawkes A (1971) Point spectra of some mutually exciting point processes. *J Roy Stat Soc B Met* 33: 438–443.
60. Marinazzo D, Kappen H, Gielen S (2007) Input-driven oscillations in networks with excitatory and inhibitory neurons with dynamic synapses. *Neural Comput* 19: 1739–1765.
61. Chacron M, Longtin A, Maler L (2005) Delayed excitatory and inhibitory feedback shape neural information transmission. *Phys Rev E* 72: 051917.
62. Åkerberg O, Chacron M (2009) Noise shaping in neural populations. *Phys Rev E* 79: 011914.
63. Shea-Brown E, Josić K, de La Rocha J, Doiron B (2008) Correlation and synchrony transfer in integrate-and-fire neurons: Basic properties and consequences for coding. *Phys Rev Lett* 100: 108102.
64. Toyozumi T, Rad K, Paninski L (2009) Mean-field approximations for coupled populations of generalized linear model spiking neurons with markov refractoriness. *Neural Comput* 21: 1203–1243.
65. Alijani A, Richardson M (2011) Rate response of neurons subject to fast or frozen noise: From stochastic and homogeneous to deterministic and heterogeneous populations. *Phys Rev E* 84: 011919.
66. Ostojic S, Brunel N (2011) From spiking neuron models to linear-nonlinear models. *PLoS Comput Biol* 7: e1001056.
67. Richardson M (2009) Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents. *Phys Rev E* 80: 021928.
68. Bair W, Zohary E, Newsome W (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior. *J Neurosci* 21: 1676–1697.
69. Ricciardi L, Sacerdote L (1979) The ornstein-uhlenbeck process as a model for neuronal activity. *Biol Cybern* 35: 1–9.