Motif statistics and spike correlations in neuronal networks

J. Stat. Mech. (2013) P03012

(http://iopscience.iop.org/1742-5468/2013/03/P03012)

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Motif statistics and spike correlations in neuronal networks

Yu Hu[1], **James Trousdale**[2], **Krešimir Josić**[2,3] **and Eric Shea-Brown**[1,4]

[1] Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA
[2] Department of Mathematics, University of Houston, Houston, TX 77204-5001, USA
[3] Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA
[4] Program in Neurobiology and Behavior, University of Washington, Seattle, WA 98195, USA
E-mail: huyu@uw.edu, jrtrousd@math.uh.edu, josic@math.uh.edu and etsb@uw.edu

Received 18 June 2012
Accepted 30 October 2012
Published 12 March 2013

Online at stacks.iop.org/JSTAT/2013/P03012
doi:10.1088/1742-5468/2013/03/P03012

Abstract. Motifs are patterns of subgraphs of complex networks. We studied the impact of such patterns of connectivity on the level of correlated, or synchronized, spiking activity among pairs of cells in a recurrent network of integrate and fire neurons. For a range of network architectures, we find that the pairwise correlation coefficients, averaged across the network, can be closely approximated using only three statistics of network connectivity. These are the overall network connection probability and the frequencies of two second order motifs: diverging motifs, in which one cell provides input to two others, and chain motifs, in which two cells are connected via a third intermediary cell. Specifically, the prevalence of diverging and chain motifs tends to increase correlation. Our method is based on linear response theory, which enables us to express spiking statistics using linear algebra, and a resumming technique, which extrapolates from second order motifs to predict the overall effect of coupling on network correlation. Our motif-based results seek to isolate the effect of network architecture perturbatively from a known network state.

Keywords: neuronal networks (theory), random graphs, networks, computational neuroscience

ArXiv ePrint: 1206.3537

1742-5468/13/P03012+51$33.00

Contents

doi:10.1088/1742-5468/2013/03/P03012

J. Stat. Mech. (2013) P03012

1. Introduction

Neural networks are highly interconnected: a typical neuron in mammalian cortex receives on the order of a thousand inputs. The resulting collective spiking activity is characterized by *correlated* firing of different cells. Such correlations in spiking activity are the focus of a great deal of theoretical and experimental work. This interest arises because correlations can strongly impact the neural coding of information, by introducing redundancy among different (noisy) neurons, allowing noise cancellation effects, or even serving as additional 'channels' of information [1]–[12]. Correlations can also help gate the transfer of signals from one brain area to another [13]–[15], and determine the statistical vocabulary of a network in terms of the likelihood that it will produce a particular spike pattern in its repertoire [16, 17].

Thus, the possible effects of correlations on the encoding and transmission of signals are diverse. Making the situation richer still, correlations can depend on characteristics of the 'input' signals (or *stimuli*) themselves. The result is an intriguing situation in which there is no simple set of rules that determines the role of correlations—beneficial, detrimental, or neutral—in neural computation. Moreover, despite major progress, the pattern and strength of correlations in neuronal networks *in vivo* remain under debate [18, 19]. This demands tools that will let us predict and understand correlations and their impact in specific types of neural circuit.

In this paper, we take one step toward this goal: we determine how simple statistics of network connectivity contribute to the average level of correlations in a neural population. While other factors such as patterns of correlations in upstream neural populations and the dynamical properties of single cells contribute in important ways, we seek to isolate the role of connection statistics in the most direct way possible. We return to the question of how our results on connectivity might be combined with these other factors at several places in the text, and in section 8.

We define the statistics of network connectivity via *motifs*, or subgraphs that are the building blocks of complex networks. While there are many ways to characterize connectivity, we choose the motif-based approach for two main reasons. First, we wish to follow earlier work in theoretical neuroscience in which the frequency of small motifs is used to define low-order statistics of network adjacency matrices [20]. Second, recent

J. Stat. Mech. (2013) P03012

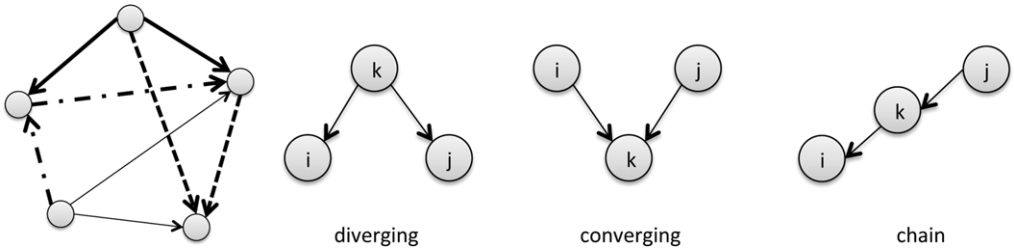

Figure 1. (Left) Counting motifs in a network. An example of a diverging motif in the network is shown with a bold solid line. Similarly, a converging and a chain motif are shown with dashed and dash–dotted lines, respectively. In total, the network has eight connections, six diverging, seven converging and five chain motifs. (Right) The different types of second order motif.

experiments have characterized the frequencies with which different motifs occur in biological neural networks, and found intriguing deviations from what we would expect in the case of unstructured, random connectivity [21, 22].

Figure 1 depicts the motifs we will consider: a single cell projecting to two downstream cells (diverging motif), a single cell receiving inputs from two upstream cells (converging motif), and a series of two connections in and out of a central cell (chain motif). To assess how prevalent these motifs are in a given network, we count their occurrences, and compare the observed motif counts with those expected in a reference graph [21]. This is a *regular* network which has the same total number of connections, and in which each cell has the same, evenly divided number of incoming and outgoing connections—i.e. the same *in and out degrees*. Importantly, in the limit of large network size the (relative) motif counts for such regular graphs agree with those in the classical model of random graphs, the Erdös–Rényi model (in which each possible connection is independently chosen to be present with the same probability). Thus, when we refer to the prevalence of network motifs, this means in comparison to either a regular or large Erdös–Rényi graph.

Figure 2 illustrates the importance of network motifs in determining the average correlation across the network. Here, we simulate excitatory and inhibitory networks of exponential integrate and fire neurons [23]. The cellular dynamics are set to be homogeneous, so that each neuron fires with the same rate, regardless of its connectivity. We explain in more detail below that this is important to isolate the effects of network connectivity alone. The black dots show the average correlation for 90 Erdös–Rényi networks that have different connection probabilities p. As expected, correlations increase with connection probability. Next, the 265 gray dots show the average correlation in networks that all have close to the same connection probability (on average $p = 0.2$, scatter due to sampling), but with a different prevalence of motifs compared to the corresponding Erdös–Rényi model. Interestingly, the range of correlation values obtained at this fixed connection probability p is as large as that obtained in the Erdös–Rényi networks over a range of p values that reaches to roughly twice the connectivity. Thus, motifs—over and above net connectivity—play a strong role in determining the strength of correlations across a network.

But which motifs contribute to this wide range of correlations in non-Erdös–Rényi networks? And are the second order motifs of figure 1 sufficient, or must

J. Stat. Mech. (2013) P03012

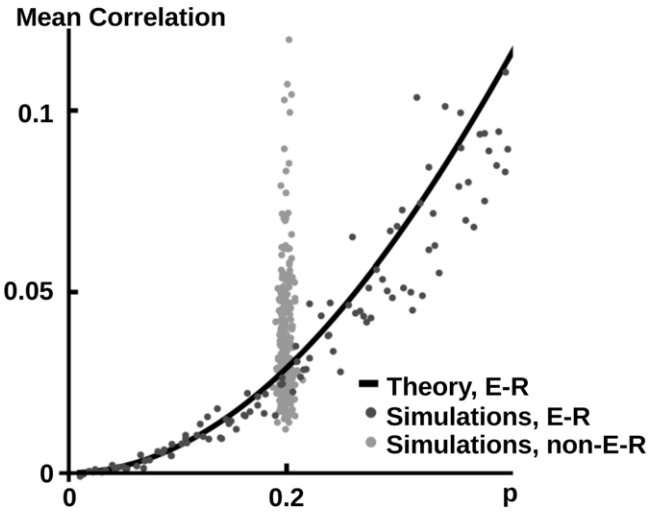

Figure 2. Impact of changing motif frequencies on mean correlation in non-Erdös–Rényi networks (gray dots) compared with the effect of changing connection probability in Erdös–Rényi networks (black dots). Mean correlation coefficients (averaged over all cell pairs in the network) are plotted against connection probability p (equation (10)). Keeping p near 0.2 and varying second order motif frequencies strongly affects average correlations. (The slight scatter in p is due to finite-size effects.) The curve presents the predictions from equation (47). For network parameters see figure 9.

higher order motifs (involving four or more cells) be included to understand average correlations?

Our goal in the balance of this paper is to develop a theory that allows us to answer these questions. First, we show how to derive analytically a relationship between motif counts and network-wide correlation that successfully identifies these trends. We previously built on the work of [24] to derive an explicit expression for the pairwise correlation between cells in a network [25]. Therein, we demonstrated how patterns of network connectivity, together with the dynamical properties of single cells, shape correlations. Here we extend our previous work to show how the same expression can reveal the impact of motifs on network-averaged correlations.

Second, we exhibit that—for a range of different network models (see section 3)—the average network correlation can be closely approximated using the connection probability and frequencies of motifs that involve *only two* connections between cells (see figure 1). Specifically, we show that the prevalence of diverging and chain motifs significantly affects average correlation, while converging motifs have no effect.

The paper is structured as follows. We explain in section 2 our setup and introduce the linear response theory which is the basis of our analysis. section 3 defines the three types of motif and their frequencies, and discusses the classes of networks on which we test and apply our theory. For simplicity, we first demonstrate our analysis in the case of a single population of excitatory cells in section 4, and then generalized to interacting populations of excitatory and inhibitory cells in section 5. We conclude by discussing a crucial assumption of our network model (section 6), and compare our theory with simulations of integrate and fire neuron networks (see section 7). Biological interpretations,

J. Stat. Mech. (2013) P03012

limitations, and extensions are covered in our discussion, and an appendix and table of notation follow (table I.1).

2. Neuron models, cross-spectra, and measures of collective network activity

In this section we describe a model spiking network composed of integrate and fire (IF) neurons. We introduce the measures that we will use to quantify the level of correlations between cell pairs in the network. The analytical approximations of these correlations given in equation (6) will be the basis for the subsequent analysis. While we develop the theory for networks of IF neurons, the main ideas are applicable to many other models (e.g. Hawkes model). We return to this point in section 8.

2.1. Networks of integrate and fire neurons

In a network of *integrate and fire* (IF) units, the dynamics of each cell are described by a single membrane voltage variable v_i, which satisfies

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

Here, $E_{\mathrm{L},i}$ is the leak reversal potential, $\psi(v_i)$ is a spike generating current, and E_i is the mean external synaptic input from sources not modeled. In numerical simulations we use the exponential integrate and fire (EIF) model [23], so that $\psi(v) \equiv \Delta_T \exp[(v - v_T)/\Delta_T]$. Each cell has independent fluctuations due to internal noise and external inputs (e.g. from a surrounding network that is not explicitly modeled). We describe such effects by additive terms, $\sqrt{\sigma_i^2 \tau_i}\xi_i(t)$, which are Gaussian white noise processes [26, 27]. Synaptic input to cell i from other cells in the network is denoted by $f_i(t)$ (see below).

When the membrane potential reaches a threshold, v_{th}, an action potential, or spike, is generated; the membrane potential is then reset to a lower voltage v_{r} and held there for an absolute refractory period τ_{r}. We denote the time at which the jth neuron fires its kth spike as $t_{j,k}$; taken together, these define this neuron's *spike train* $y_j(t) = \sum_k \delta(t - t_{j,k})$. Importantly, synaptic interactions among neurons are initiated at these spike times, so that the total synaptic input to the ith cell is

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

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

In the absence of a synaptic connection from cell j to cell i, we set $\mathbf{W}_{ij} = 0$. Hence, the $N \times N$ matrix of synaptic weights, $\mathbf{W}$, defines a directed, weighted network.

In simulations we used the parameters given in the caption of figure 9, unless stated otherwise.

2.2. Measures of network correlation

Dependences between the responses of cells in the network may be quantified using the spike train auto- and cross-correlation functions [28]. For a pair of spike trains $y_i(t), y_j(t)$,

J. Stat. Mech. (2013) P03012

the cross-covariance function is given by

$$\mathbf{C}_{ij}(\tau) = \mathbf{E}\left[(y_i(t) - r_i)(y_j(t+\tau) - r_j)\right],$$

where r_i and r_j are the firing rates of the two cells. Here and throughout the paper we assume that the spike trains form a (multivariate) stationary process. The auto-correlation function is the cross-correlation of the output of the cell and itself, and $\mathbf{C}(t)$ is the matrix of cross-correlation functions. The matrix $\tilde{\mathbf{C}}(\omega)$ with entries defined by

$$\tilde{\mathbf{C}}_{ij}(\omega) = \mathbf{E}\left[\tilde{y}_i^*(\omega)\tilde{y}_j(\omega)\right],$$

is the matrix of cross-spectra (see next section). Here a^* is the complex conjugate. The cross-spectrum of a pair of cells is equivalent to the Fourier transform of their cross-correlation function [29].

Denote by $N_{y_i}(t_1, t_2) = \int_{t_1}^{t_2} y_i(s)\,\mathrm{d}s$ the spike count of cell i over a time window $[t_1, t_2]$; the spike count correlation $\boldsymbol{\rho}_{ij}(T)$ over windows of length T is defined as

$$\boldsymbol{\rho}_{ij}(T) = \frac{\mathrm{cov}(N_{y_i}(t, t+T), N_{y_j}(t, t+T))}{\sqrt{\mathrm{var}(N_{y_i}(t, t+T))\mathrm{var}\left(N_{y_j}(t, t+T)\right)}}.$$

We will make use of the total correlation coefficient $\boldsymbol{\rho}_{ij}(\infty) = \lim_{T\to\infty}\boldsymbol{\rho}_{ij}(T)$, which captures dependences between the processes y_i, y_j over arbitrarily long timescales, but may also describe well the nature of dependences over reasonably short timescales [30]–[33]. The spike count covariance is related directly to the cross-correlation function by [34]

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

Thus, total correlation may be defined alternatively in terms of the integrated cross-correlations (equivalently, the cross-spectra evaluated at $\omega = 0$):

$$\boldsymbol{\rho}_{ij}(\infty) = \frac{\tilde{\mathbf{C}}_{ij}(0)}{\sqrt{\tilde{\mathbf{C}}_{ii}(0)\tilde{\mathbf{C}}_{jj}(0)}}. \tag{2}$$

Throughout this paper, we will use $\boldsymbol{\rho}_{ij}(\infty)$ as the measure of correlation and use the above equation to calculate it from the cross-spectrum matrix [35]. However, our analysis can be similarly applied to study $\tilde{\mathbf{C}}_{ij}(\omega)$, and hence the entire correlation function in time [25].

2.3. Linear response approximation of cell response covariance

Linear response theory [28, 36] can be used to approximate the response of single cells, and the joint response of cells in a network [37]–[39], [24, 25], and can lead to a linear equation for the covariance matrix of activity [1]. Consider an IF neuron obeying equation (1), but with the mean of the inputs $f(t)$ absorbed into the constant E. We denote the remaining, *zero-mean* input by $\epsilon X(t)$, so that

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

For fixed input fluctuations $\epsilon X(t)$, the output spike train will be different for each realization of the noise $\xi(t)$, and each initial condition $v(0)$. The time-dependent firing rate

J. Stat. Mech. (2013) P03012

is obtained by averaging the resulting spike train over noise realizations and a stationary distribution of initial conditions. For all values of ϵ, this stationary distribution is taken to be the one obtained when $\epsilon = 0$. We denote the resulting averaged firing rate as $r(t) = \langle y(t) \rangle$. Linear response theory approximates this firing rate as

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

where r_0 is the firing rate in the absence of input ($\epsilon = 0$), and the linear response kernel, $A(t)$, characterizes the response to first order in ϵ. The response kernel (and unperturbed power spectrum $\tilde{C}^0(\omega)$, see below) may be calculated numerically by solving an initial value problem derived from the Fokker–Planck equation for the membrane potential [40]. This approximation is remarkably accurate over a wide range of parameters; for example, see [41, 42].

Next, we turn to the problem of approximating the output of a cell on a single trial, rather than the average across trials. We denote the Fourier transform of a function f by $\tilde{f} = \mathcal{F}(f)$. However, for spike trains we adopt a convention that $\tilde{y}_i(\omega)$ is the Fourier transform of the mean subtracted spike train $y_i(t) - r_i$. Following [38, 24, 25], we approximate the spiking output of a cell self-consistently by

$$\tilde{y}_i(\omega) \approx \tilde{y}_i^0(\omega) + \tilde{A}_i(\omega) \left(\sum_j \tilde{\mathbf{J}}_{ij}(\omega) \tilde{y}_j(\omega) \right), \tag{4}$$

where $\tilde{y}_i^0(\omega)$ is a realization of the output of cell i in the absence of synaptic input. Defining the interaction matrix $\tilde{\mathbf{K}}$ with entries $\mathbf{K}_{ij}(t) \equiv (A_i * \mathbf{J}_{ij})(t)$, we can use equation (4) to solve for the vector of Fourier transformed spike train approximations

$$\tilde{y}(\omega) = (\mathbf{I} - \tilde{\mathbf{K}}(\omega))^{-1} \tilde{y}^0(\omega), \tag{5}$$

and matrix of cross-spectra

$$\begin{aligned} \tilde{\mathbf{C}}(\omega) \approx \tilde{\mathbf{C}}^\infty(\omega) &= (\mathbf{I} - \tilde{\mathbf{K}}^*(\omega))^{-1} \langle \tilde{y}^{0*}(\omega) \tilde{y}^{0\mathrm{T}}(\omega) \rangle (\mathbf{I} - \tilde{\mathbf{K}}^{\mathrm{T}}(\omega))^{-1} \\ &= (\mathbf{I} - \tilde{\mathbf{A}}^* \mathbf{W} \tilde{\mathbf{F}}^*)^{-1} \tilde{\mathbf{C}}^0 (\mathbf{I} - \tilde{\mathbf{F}} \mathbf{W}^{\mathrm{T}} \tilde{\mathbf{A}})^{-1}. \end{aligned} \tag{6}$$

Here $\tilde{\mathbf{A}}, \tilde{\mathbf{C}}^0, \tilde{\mathbf{F}}$ are diagonal matrices: $\tilde{\mathbf{A}}_{ii}(\omega) = \tilde{A}_i(\omega)$ is the linear response of cell i, $\tilde{\mathbf{C}}^0_{ii}(\omega) = \tilde{C}^0_i(\omega) = \langle \tilde{y}^{0*}_i(\omega) \tilde{y}^{0\mathrm{T}}_i(\omega) \rangle$ is its 'unperturbed' (i.e. without coupling) power spectrum, and $\tilde{\mathbf{F}}_{ii}(\omega) = \tilde{F}_i(\omega)$ is the Fourier transform of the synaptic coupling kernel from cell i. Again, $\tilde{\mathbf{K}}^*$ is the complex conjugate (without transposition). As noted later, our results and analysis will hold at all frequencies and thus can be used to study correlations at all timescales. The weighted connectivity matrix, $\mathbf{W}$, defines the structure of the network.

To simplify the exposition, we initially assume certain symmetries in the network. For instance, we consider *homogeneous networks* in which cells have identical (unperturbed) power spectra, linear response functions, and synaptic kernels. In this case the diagonal matrices in equation (6) act like scalars. We slightly abuse notation in this case, and replace $\tilde{A}_i(\omega)$ by $\tilde{A}(\omega)$, and $\tilde{C}^0_i(\omega)$ by $\tilde{C}^0(\omega)$. This allows us to disentangle the effects of network structure from the effects of neuronal responses on network activity. The resulting cross-spectrum matrix (evaluated at $\omega = 0$) is

$$\tilde{\mathbf{C}}^\infty = \tilde{C}^0 (\mathbf{I} - \tilde{A}\mathbf{W})^{-1} (\mathbf{I} - \tilde{A}\mathbf{W}^{\mathrm{T}})^{-1} \tag{7}$$

J. Stat. Mech. (2013) P03012

(note that $\tilde{F}_i(0) = 1$ by definition). We use this simpler expression in what follows, and return to heterogeneous networks in section 6. Since we consider only total correlation, we omit the dependence of the spike count correlation coefficient on window size T. Additionally, all spectral quantities are evaluated at $\omega = 0$, so we also suppress the dependence on ω. Finally, we define average network correlation by

$$\rho^{\text{avg}} = \frac{1}{N(N-1)} \sum_{i \neq j}^{N} \boldsymbol{\rho}_{ij}. \tag{8}$$

In subsequent sections, we will examine the average covariance across the network

$$\langle \tilde{\mathbf{C}}^{\infty} \rangle = \frac{1}{N^2} \sum_{ij}^{N} \tilde{\mathbf{C}}^{\infty}_{ij}.$$

This average cannot be directly related to that in equation (8), where individual summands are normalized, and diagonal terms are excluded. The motif-based theory we develop predicts $\langle \tilde{\mathbf{C}}^{\infty} \rangle$, and gives no information about the specific entries $\tilde{\mathbf{C}}^{\infty}_{ij}$. However, ρ^{avg} can be determined approximately from $\langle \tilde{\mathbf{C}}^{\infty} \rangle$ alone. We describe these approximations in appendix A.

2.4. Applicability of linear response theory

Our methods depend on two, related sets of conditions for their validity. First, we take our cells to be driven by a white noise background. This background linearizes the response of the cells to sufficiently weak perturbations, improving the accuracy of the approximation equation (4); its presence is our first condition.

Second, turning to network effects, we assume that the spectral radius $\Psi(\tilde{\mathbf{K}}) < 1$; this gives non-singularity of the approximating processes in equation (4) and allows us to make the series expansion we describe in section 4.2. In practice, we have found that the linear approximation to correlations will cease to provide an accurate approximation before this occurs, likely owing in part to a failure of the perturbative approximation. Furthermore, the approximation seems to be most accurate at weak interaction strengths as characterized by a small radius of the spectrum of $\tilde{\mathbf{K}}$.

For Erdös–Rényi networks, an asymptotic (large N) characterization of the spectral radius of the synaptic weight matrix has been derived [43]. In particular, for an Erdös–Rényi network consisting of only excitatory cells with synaptic weight w, there will be a single eigenvalue at pNw with the remaining eigenvalues distributed uniformly in a circle around the origin of radius $w\sqrt{p(1-p)N}$. In networks with both excitatory and inhibitory populations, there is a single outlier at $pN_Ew_E + pN_{\text{I}}w_{\text{I}}$ and all other eigenvalues will be distributed (non-uniformly) within the circle of radius $\sqrt{p(1-p)(N_Ew_E^2 + N_{\text{I}}w_{\text{I}}^2)}$. We will use these expressions to quantify the strength of interactions given by the asymptotic spectral radius of $\tilde{\mathbf{K}} = \tilde{A}\mathbf{W}$, denoted as

$$\Psi_{\text{ER}}(\tilde{A}\mathbf{W}) = |\tilde{A}| \max\left\{ |pN_{\text{E}}w_{\text{E}} + pN_{\text{I}}w_{\text{I}}|, \sqrt{p(1-p)(N_{\text{E}}w_{\text{E}}^2 + N_{\text{I}}w_{\text{I}}^2)} \right\}. \tag{9}$$

J. Stat. Mech. (2013) P03012

We note that we used IF simulations to directly confirm the accuracy of the linear response approximation for excitatory–inhibitory networks in figure 14. For a complete discussion of the performance of the linear response theory, see [25].

3. Graphical structure of neuronal networks

Our main goal is to determine how the small-scale statistical structure of directed networks influences the collective dynamics they produce—namely, the strength of spike correlations in networks of model neurons. We will quantify network structure using the probability of connections between pairs and among triplets of cells, organized into *network motifs* [21, 20].

3.1. Network motifs and their frequencies

A motif is a subgraph composed of a small number of cells. We classify motifs according to the number of edges they contain. We begin by considering directed networks composed of identical cells. First order motifs contain one connection and hence come in only one type—two cells with a one-way connection. Second order motifs contain two connections, and therefore involve at most three interacting cells. These motifs come in three types: diverging, converging and chain motifs (see figure 1) [21, 20]. (Note that in our definition, a cell can appear twice in the triplet of cells that define a second order motif. For example, the chain motif in figure 1 is equivalent to a bidirectionally coupled pair of cells when $i = j$.)

We will consider mainly the impact of second order motifs, over and above first order effects. The three motifs shown in figure 1 arise naturally in our analysis of correlated spiking activity. In particular, we will show that the *frequency* at which each motif occurs in the network can accurately predict levels of correlation across the network.

We next introduce notation that will allow us to make these ideas precise. Let $\mathbf{W}^0$ be the adjacency matrix, so that $\mathbf{W}^0_{i,j} = 1$ indicates the presence of a directed connection from cell j to cell i, and $\mathbf{W}^0_{i,j} = 0$ indicates its absence. To quantify the frequency of a motif in a given graph, we first count the total number of times the motif occurs, and divide by the total number of possible occurrences in a graph of that size. For first order motifs this definition gives the *empirical* connection probability,

$$p = \left(\sum_{i,j} \mathbf{W}^0_{i,j}\right) \Big/ N^2. \tag{10}$$

The preponderance of second order motifs is measured in two stages. First, we similarly normalize the motif count. Second, we subtract the value expected in a reference graph. The resulting expressions are

$$q_{\mathrm{div}} = \sum_{i,j,k} (\mathbf{W}^0_{i,k}\mathbf{W}^0_{j,k}) \Big/ N^3 - p^2 = \left(\sum_{i,j} (\mathbf{W}^0\mathbf{W}^{0\mathrm{T}})_{i,j}\right) \Big/ N^3 - p^2, \tag{11}$$

$$q_{\mathrm{con}} = \left(\sum_{i,j} (\mathbf{W}^{0\mathrm{T}}\mathbf{W}^0)_{i,j}\right) \Big/ N^3 - p^2, \tag{12}$$

J. Stat. Mech. (2013) P03012

$$q_{\rm ch} = \left(\sum_{i,j} (\mathbf{W}^0\mathbf{W}^0)_{i,j}\right) \Big/ N^3 - p^2, \tag{13}$$

where $\mathbf{W}^{0\rm T}$ denotes the transpose of $\mathbf{W}^0$. Consider the expression defining $q_{\rm div}$: the sum in the first equality simply counts the total number of connections from one cell (k) to two others (i and j), and divides by the total number of possible connections of this type (N^3). This can be written as matrix multiplication followed by a sum over all entries i, j, as shown. In each case we subtract the value p^2, which corresponds to the frequency of the motif in a regular graph, as well as the asymptotic frequency in an Erdös–Rényi graph as the number of cells, N, diverges to infinity. Indeed, for Erdös–Rényi graphs any edge is present with probability p, and any second order motif requires the presence of two edges. Thus, $q_{\rm div}$ corresponds to the propensity for a network to display diverging motifs, over and above expectations from an Erdös–Rényi or regular network. The other measures in equations (12) and (13) have similar interpretations.

The quantities in equations (11)–(13) can also be interpreted as empirical measures of covariance [44, 20]. If we denote by $\mathbf{E}_{\rm e}[\cdot]$ the empirical average over all entries in a given network, then we may write

$$q_{\rm div} = \mathbf{E}_{\rm e}\left[\mathbf{W}^0_{i,k}\mathbf{W}^0_{j,k}\right] - \mathbf{E}_{\rm e}\left[\mathbf{W}^0_{i,k}\right]\mathbf{E}_{\rm e}\left[\mathbf{W}^0_{j,k}\right].$$

We also note that the quantities defined in equations (11)–(13) are not independent (but do have three degrees of freedom). If we first sum over indices i, j in equation (11), we can rewrite the expression in terms of in and out degrees, $\{d^{\rm in}_k\}, \{d^{\rm out}_k\}$. For example,

$$q_{\rm div} = \sum_k (d^{\rm out}_k)^2/N^3 - p^2 = {\rm var}(d^{\rm out})/N^2 \tag{14}$$

is the scaled sample variance of the out degree across the network. Similarly,

$$q_{\rm con} = {\rm var}(d^{\rm in})/N^2, \qquad q_{\rm ch} = {\rm cov}(d^{\rm out}, d^{\rm in})/N^2, \tag{15}$$

where ${\rm cov}(\cdot,\cdot)$ denotes sample covariance [20]. Therefore,

$$q_{\rm div} \geq 0, \qquad q_{\rm con} \geq 0, \qquad |q_{\rm ch}| \leq \sqrt{q_{\rm div} q_{\rm con}}. \tag{16}$$

We further show in appendix B that

$$|q_{\rm div}|, |q_{\rm con}|, |q_{\rm ch}| \leq p(1-p), \tag{17}$$

where equality is attainable for $q_{\rm div}$ and $q_{\rm con}$ (but not simultaneously). Equations (16)–(17) identify the attainable ranges of motif frequencies. In generating our networks, we compare the extent of motif frequencies that we produce using a particular scheme against this maximum possible range (see also section 3.2).

In networks composed of excitatory and inhibitory cells, we can represent interactions between cells using a signed connectivity matrix. Edges emanating from inhibitory neurons are represented by negative entries and those from excitatory neurons by positive entries. In this case, motifs are further subdivided according to their constituent cells. For instance, there are six distinct diverging motifs, since if we list all 2^3 group types of three cells, we see that the motif $E \leftarrow I \rightarrow I$ is the same as $I \leftarrow I \rightarrow E$, and $E \leftarrow E \rightarrow I$ is the same as $I \leftarrow E \rightarrow E$. Similarly, there are six distinct converging motifs, and eight distinct chain motifs, for a total of 20 distinct second order motifs (see figure 3).

J. Stat. Mech. (2013) P03012

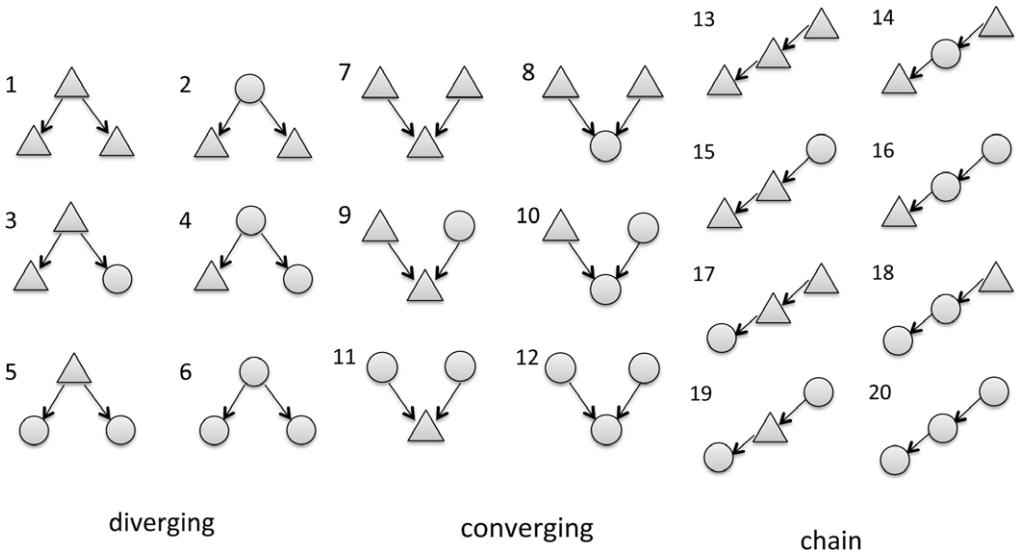

Figure 3. Second order motifs in populations of excitatory and inhibitory cells: There are 20 sub-types of the three main motif types. Triangles represent excitatory neurons and circles for inhibitory neurons.

This will clearly lead to some cumbersome notation! Therefore, while populations of interacting E and I cells are our ultimate goal, we first describe our ideas in a population of cells of a single type.

3.2. Generating graphs with given motif frequency

To numerically examine the impact of motif frequency on dynamics, we need to generate graphs that are equal in connection probability, but differ in the preponderance of second order motifs. The empirical connection probability in network samples will have small fluctuations around the statistical (i.e. expected) value we fixed. We use two ways of generating such graphs. The first is the degree distribution method [45] (related to configuration model [44, 46, 47]). Here, following [20] we use a two-sided power law, with various rising and decreasing exponents, peak locations and truncation limits, as expected in- and out-degree distributions. The other is the *second order network* (SONET) method (for details see [20]). Network samples generated using both methods cover the range of motif frequencies observed experimentally in cortical circuits [21, 20]. Naturally, this experimentally observed range is smaller than the full extent of possibly attainable frequencies (see equations (16)–(17)); however, the SONET method covers this full range as well. Details are given in appendix C.

We use both methods to generate network samples in the excitatory only case and found similar results; here, we only show data generated using the SONET method as it covers a larger range of motif frequencies. In excitatory–inhibitory networks, we use the degree distribution method.

We emphasize that our approaches below do not *a priori* specify any particular way of generating network samples. However, their accuracy will depend on how this is done. Our methods for analyzing the impact of motifs of average correlations are accurate in the

J. Stat. Mech. (2013) P03012

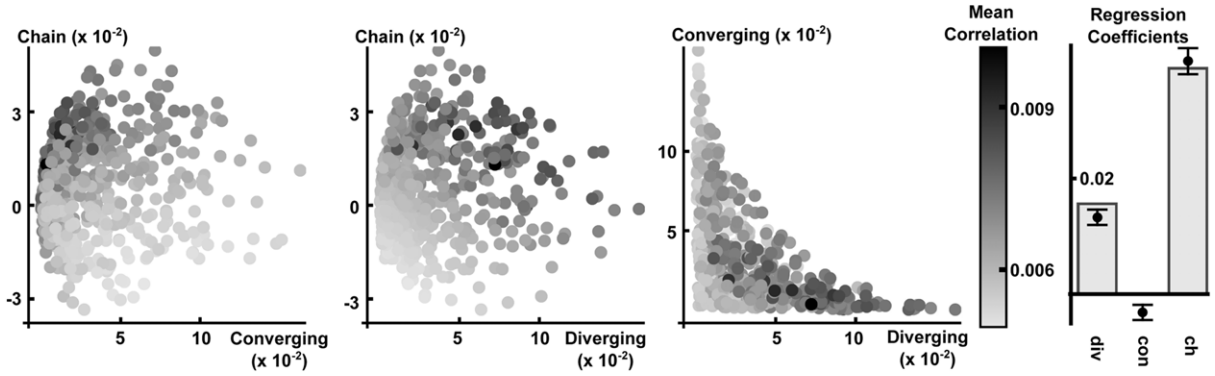

Figure 4. (Left) The relationship between second order motif frequencies and average correlation in purely excitatory networks. Each dot corresponds to a network sample and its shading represents the corresponding average correlation coefficient computed using equation (6). Each axis represents one of the three second order motif types as defined in equations (11)–(13). The effective coupling strength, $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.2$ for all networks (see equation (9)). (Right) Bars show the linear regression coefficients calculated from the resumming theory (see equation (33)) and numerically (points) between motif frequencies and average network correlations. The error bars around each point denote 95% confidence intervals for the regression coefficients. The baseline value in this bar plot is 0. The coefficient of determination R^2 is 0.80. The data were obtained from 512 network samples generated using the SONET algorithm (see section 3.2).

family of networks that we described above. However, they may break down in networks that have significant additional structure, a point to which we will return.

4. Impact of second order motifs in networks of excitatory cells

As we have shown in section 2, linear response theory can be used to approximate cross-correlations between cells in a neuronal network. We next explore how the key expression, given in equation (7), can be applied to relate the frequency of second order motifs to the average correlation across pairs of cells. For simplicity, we first consider networks consisting only of a single class of excitatory cells. The results extend naturally to the case of two interacting populations of excitatory and inhibitory cells, as we show in section 5.

4.1. Results: linear dependence between mean correlation and motif frequencies

Figure 4 illustrates the relationship between second order motif frequency and average correlation in networks of excitatory cells. In all examples in this section, we use networks of $N = 100$ cells, with parameters as given in the caption of figure 9, and with graph structures generated by the SONET algorithm (see section 3.2). Correlation coefficients between cells are computed using the linear response approximation given in equation (6). We find that average correlations depend strongly on the frequency of diverging and especially chain motifs, but only weakly on the frequency of converging motifs. To quantify the linear regression, we use the coefficient of determination between the linear fit (constructed from regression coefficients) and the result of equation (6), obtaining

J. Stat. Mech. (2013) P03012

$R^2 = 0.80$. This high value suggests that the second order motifs are highly predictive of network correlation, in the simplest possible (linear) way (as explained in appendix D; this prediction can be further improved when we compensate for the fluctuations in empirical connection probabilities due to the finite size of network samples). In the balance of this section, we explain why this is the case, derive via a resumming theory a nonlinear predictor of mean correlation in terms of motif frequencies, and extract from this an explicit linear relationship between the probability of observing second order motifs and the mean correlation in the network.

4.2. First theory: second order truncation for network correlation

If the spectral radius $\Psi(\tilde{A}\mathbf{W}) < 1$, then the matrix inverses in equation (7) can be expanded in a power series in $\tilde{A}\mathbf{W}$ as

$$\frac{\tilde{\mathbf{C}}^{\infty}}{\tilde{C}^0} = \sum_{i,j=0}^{\infty} \tilde{A}^{i+j}\mathbf{W}^i(\mathbf{W}^{\mathrm{T}})^j \tag{18}$$

[48]. Terms in this expansion correspond naturally to paths through the graph defining the network [25, 35, 49, 50]. For example, the terms $(\tilde{A}^2\mathbf{W}^2)_{ij} = \tilde{A}^2\sum_k \mathbf{W}_{ik}\mathbf{W}_{kj}$ give the contributions of length two chains between pairs of cells i, j. Entries in the matrices $\tilde{A}^3\mathbf{W}^2\mathbf{W}^{\mathrm{T}}$, $\tilde{A}^3\mathbf{W}(\mathbf{W}^{\mathrm{T}})^2$ give the contributions of diverging motifs of third order—one branch of the motif is of length two, while the other is a direct connection.

Let $\mathbf{1}_{N_1N_2}$ denote the $N_1 \times N_2$ matrix of ones, and define the N-vector $\mathbf{L} = (1/N)\mathbf{1}_{N,1}$. We define the orthogonal projection matrices $\mathbf{H}, \mathbf{\Theta}$ which will play a crucial role in the following analysis:

$$\mathbf{H} = N\mathbf{L}\mathbf{L}^{\mathrm{T}}, \qquad \mathbf{\Theta} = \mathbf{I} - \mathbf{H}. \tag{19}$$

Note that if $\mathbf{X}$ is an $N \times N$ matrix, then $\mathbf{L}^{\mathrm{T}}\mathbf{X}\mathbf{L} =: \langle\mathbf{X}\rangle$ is the empirical average of all entries in $\mathbf{X}$. We first observe that the empirical network connection probability can be obtained from the adjacency matrix, $\mathbf{W}^0$, as

$$p = \mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{L}.$$

We can also express second order motif frequencies in terms of intra-network averages. For instance,

$$\begin{aligned} q_{\mathrm{div}} &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{W}^{0\mathrm{T}}\mathbf{L} - p^2 \\ &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\left(\mathbf{H}+\mathbf{\Theta}\right)\mathbf{W}^{0\mathrm{T}}\mathbf{L} - p^2 \\ &= \left(\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{L}\right)\left(\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{L}\right) + \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{L} - p^2 \\ &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{L}. \end{aligned} \tag{20}$$

Similarly, $q_{\mathrm{con}}, q_{\mathrm{ch}}$ may be expressed as

$$\begin{aligned} q_{\mathrm{con}} &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{W}^0\mathbf{L} - p^2 \\ &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{\Theta}\mathbf{W}^0\mathbf{L}, \end{aligned} \tag{21}$$

J. Stat. Mech. (2013) P03012

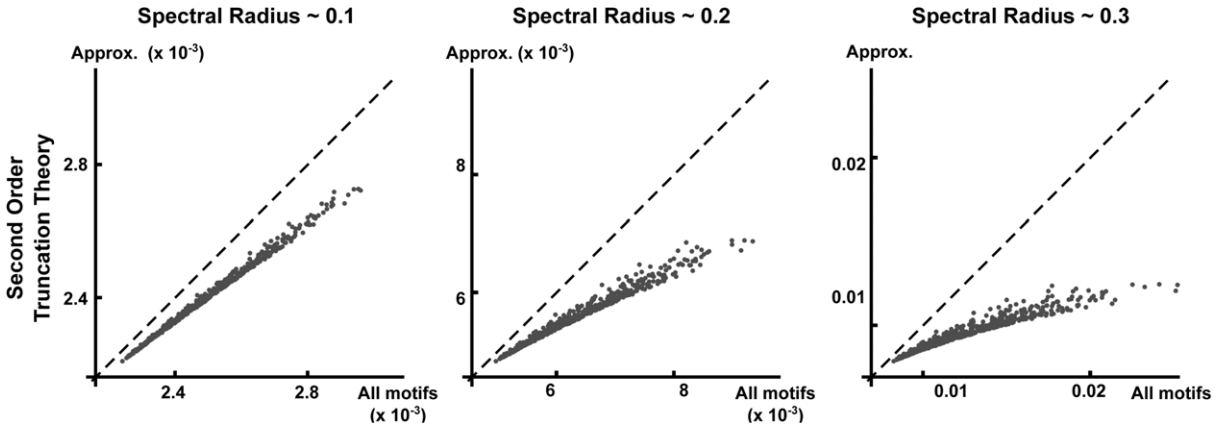

Figure 5. Scatter plot comparing the prediction of average network correlation obtained using equation (7) (horizontal axes) to the second order truncation in equation (24) (vertical axes). The diagonal line $y = x$ is plotted for reference. Each panel corresponds to a different coupling strength in the same set of 512 adjacency matrices. The effective coupling strength $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W})$ is shown at the top of each panel (see equation (9)).

and

$$\begin{aligned} q_{\mathrm{ch}} &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0}\mathbf{W}^{0}\mathbf{L} - p^{2} \\ &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{L} - p^{2} \\ &= \frac{1}{N}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0}\boldsymbol{\Theta}\mathbf{W}^{0}\mathbf{L}. \end{aligned} \tag{22}$$

To relate second order motif frequencies to mean correlations between pairs of cells, we can truncate equation (18) at second order in $(\tilde{A}\mathbf{W})$, giving

$$\frac{\tilde{\mathbf{C}}^{\infty}}{\tilde{C}^{0}} \approx \mathbf{I} + \tilde{A}w\mathbf{W}^{0} + \tilde{A}w\mathbf{W}^{0\mathrm{T}} + (\tilde{A}w)^{2}\mathbf{W}^{0}\mathbf{W}^{0\mathrm{T}} + (\tilde{A}w)^{2}(\mathbf{W}^{0})^{2} + (\tilde{A}w)^{2}(\mathbf{W}^{0\mathrm{T}})^{2}. \tag{23}$$

To obtain the empirical average of pairwise covariances in the network, $\langle\tilde{\mathbf{C}}^{\infty}\rangle$, we multiply both sides of equation (23) on the left and right by $\mathbf{L}^{\mathrm{T}}$ and $\mathbf{L}$, respectively. Making use of equations (20)–(22), we obtain

$$\frac{\langle\tilde{\mathbf{C}}^{\infty}\rangle}{\tilde{C}^{0}} \approx \frac{1}{N} + 2\tilde{A}wp + 3N(\tilde{A}w)^{2}p^{2} + N(\tilde{A}w)^{2}q_{\mathrm{div}} + 2N(\tilde{A}w)^{2}q_{\mathrm{ch}}. \tag{24}$$

How well does this second order truncation predict levels of correlation across different neural networks? Figure 5 shows that the truncation correctly captures general trends in levels of correlation from network to network, but makes a substantial systematic error. Here, we plot correlations predicted with the truncated equation (24) as an approximation of the full expression (i.e. to all orders) for average correlations given by equation (6). Indeed, the truncated expression gives consistent predictions only at very small coupling strengths. We conclude that the terms which were discarded (all terms of order three and

J. Stat. Mech. (2013) P03012

higher in $\tilde{A}\mathbf{W}$) can have an appreciable impact on average network correlation, and will next develop methods to capture this impact.

4.3. Improved theory: resumming to approximate higher order contributions to network correlation

A much better approximation of the average network correlation can be obtained by considering the impact of second order motifs on higher order terms in the expansion given by equation (18). Note that, in an Erdös–Rényi network, every motif of order m occurs with probability p^m (with the exception of motifs which involve the same connection multiple times), so that on average $q_{\rm div} = q_{\rm con} = q_{\rm ch} = 0$. For non-Erdös–Rényi networks, the expected values of $q_{\rm div}, q_{\rm con}$ and $q_{\rm ch}$ are typically not zero. As we show next, the introduction of additional second order structure in the network also affects the frequency of motifs of higher order.

Consider the average covariance as determined from the full order linear response approximation of equation (18):

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^0} = \sum_{i,j=0}^{\infty} (\tilde{A}w)^{i+j} \mathbf{L}^{\rm T} (\mathbf{W}^0)^i (\mathbf{W}^{0\rm T})^j \mathbf{L}. \tag{25}$$

We will first express every term in the sum given by equation (25) approximately in terms of first and second order motif frequencies ($p, q_{\rm div}, q_{\rm con}$ and $q_{\rm ch}$). We illustrate this approximation in two examples, before proceeding to the general calculation.

Consider the term $\mathbf{L}^{\rm T} (\mathbf{W}^0)^3 \mathbf{L}$ corresponding to the average number of length three chains connecting a pair of cells. Using $\mathbf{I} = \mathbf{H} + \mathbf{\Theta}$, we can proceed as in the computation leading to equation (20),

$$\begin{aligned} \mathbf{L}^{\rm T} \left(\mathbf{W}^0\right)^3 \mathbf{L} &= \mathbf{L}^{\rm T}\mathbf{W}^0(\mathbf{H}+\mathbf{\Theta})\mathbf{W}^0(\mathbf{H}+\mathbf{\Theta})\mathbf{W}^0\mathbf{L} \\ &= \mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{H}\mathbf{W}^0\mathbf{H}\mathbf{W}^0\mathbf{L} + \mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{H}\mathbf{W}^0\mathbf{L} \\ &\quad + \mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{H}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L} + \mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L}. \end{aligned} \tag{26}$$

We next replace $\mathbf{H}$ by $N\mathbf{L}\mathbf{L}^{\rm T}$ in the last expression to obtain

$$\begin{aligned} \mathbf{L}^{\rm T}(\mathbf{W}^0)^3\mathbf{L} &= N^2(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L})(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L})(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L}) + N(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L})(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L}) \\ &\quad + N(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L})(\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L}) + \mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L}. \end{aligned} \tag{27}$$

Here, the first three terms are composed of factors that correspond to the connection probability ($\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{L}$) and second order chain motif frequency ($\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L}$). These terms provide an estimate of the frequency of a length three chain in a graph, *in terms of the frequency of smaller motifs that form the chain*. The last term, $\mathbf{L}^{\rm T}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{L}$, gives the frequency of occurrence of the length three chain in addition to that obtained by chance from second order motifs. Such higher order terms will be gathered separately and denoted by *h.o.t.* in the approximation. We therefore obtain

$$\mathbf{L}^{\rm T} \left(\mathbf{W}^0\right)^3 \mathbf{L} = N^2(p^3 + 2pq_{\rm ch}) + \text{h.o.t.} \tag{28}$$

The exact form of this expression can be understood by referring to figure 6: the leading N^2 denotes the number of possible length three chains between a pair of cells (such a chain can pass through N^2 different intermediate pairs of cells) and p^3 represents the probability that one of these length three chains is 'present' in an Erdös–Rényi graph.

J. Stat. Mech. (2013) P03012

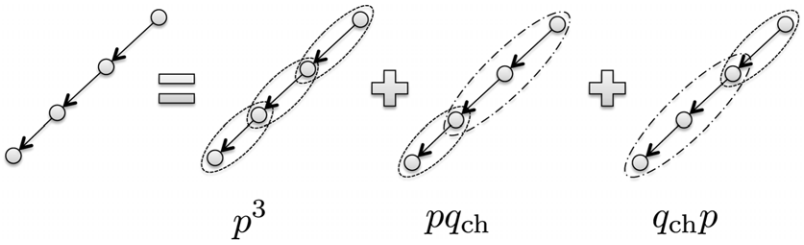

Figure 6. Estimating the number of occurrences of a third order chain motif. The first three terms in equation (27) correspond to different ways in which a chain of length three can be composed from smaller motifs. As explained in the text, each of the three terms represents the estimated probability that smaller motifs occur in one of the arrangements on the right.

Recall that q_{ch} measures the probability that a length two chain is present, above that expected in an Erdös–Rényi network. Therefore, pq_{ch} represents a second order estimate of the probability above (Erdös–Rényi) chance that the first two connections in the chain (q_{ch}) and the last (p) are present simultaneously. The prefactor 2 appears because this may also occur if the first and final two connections are present simultaneously.

As a second example, consider the term $\mathbf{L}^{\mathrm{T}}(\mathbf{W}^0)^2(\mathbf{W}^{0\mathrm{T}})^2\mathbf{L}$, corresponding to indirect diverging motifs with two connections on each branch. A computation similar to that used to obtain equation (28) now gives

$$\mathbf{L}^{\mathrm{T}}(\mathbf{W}^0)^2(\mathbf{W}^{0\mathrm{T}})^2\mathbf{L} = N^3[p^4 + p^2q_{\mathrm{div}} + 2p^2q_{\mathrm{ch}} + q_{\mathrm{ch}}^2] + \mathrm{h.o.t.} \tag{29}$$

The four terms in this sum can be understood with the help of figure 7: p^4 represents the probability of observing the motif in an Erdös–Rényi network. The product p^2q_{div} is the estimated probability above (Erdös–Rényi) chance that the motif is formed by two connections emanating from the source cell, present simultaneously and independently with two connections emanating from the tips of the branches. The term p^2q_{ch} gives the estimated probability, above (Erdös–Rényi) chance, that one branch is present, simultaneously and independently, from two connections which form the other branch (the prefactor 2 concerns the probability of this occurring in each of the two branches). The last term, q_{ch}^2, gives the estimated probability, above the Erdös–Rényi chance level, that two length two chains simultaneously emanate from the root cell.

In equation (29), *h.o.t.* denotes the three distinct terms

$$\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{L}, \qquad N(\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{L})(\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{L}),$$
$$N(\mathbf{L}^{\mathrm{T}}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^0\mathbf{\Theta}\mathbf{W}^{0\mathrm{T}}\mathbf{L})(\mathbf{L}^{\mathrm{T}}\mathbf{W}^{0\mathrm{T}}\mathbf{L}).$$

These terms contain two or more occurrences of $\mathbf{\Theta}$ in one factor, and hence correspond to motifs of higher than second order. In general, a factor which contains m occurrences of $\mathbf{\Theta}$ will depend on the frequency of a motif of order $m+1$, beyond that which is imposed by the frequency of motifs of order m.

The idea behind these two examples extends to all terms in the series in equation (25), assuming absolute convergence. Each term in the resulting series contains a factor of the form $\mathbf{L}^{\mathrm{T}}(\mathbf{W}^0)^i(\mathbf{W}^{0\mathrm{T}})^j\mathbf{L}$ corresponding to a motif of order $i+j$. This motif corresponds to two chains of lengths i and j, respectively, emanating from the same root cell. To understand the impact of second order motifs we need to decompose this motif as

J. Stat. Mech. (2013) P03012

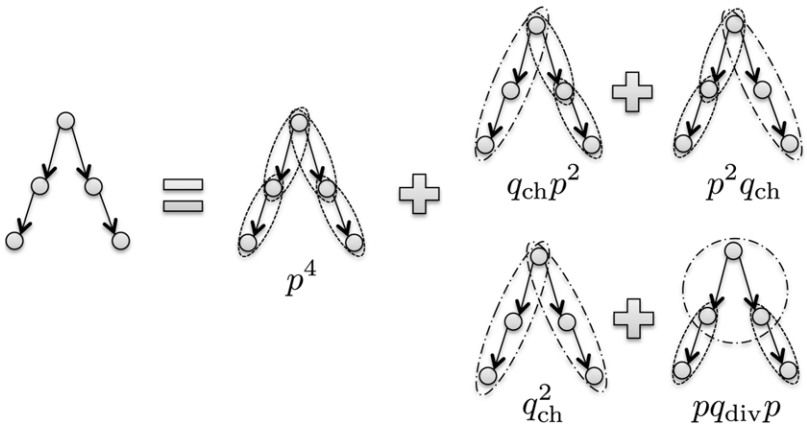

Figure 7. Estimating the number of occurrences of a fourth order motif. Equation (29) can be understood by decomposing this motif into the constituent first and second order motifs.

illustrated in figures 6 and 7. While this is a challenging combinatorial problem, we show that the answer can be obtained by rearranging the terms in equation (25).

Each factor of the form $\mathbf{L}^{\mathrm{T}}(\mathbf{W}^0)^i(\mathbf{W}^{0\mathrm{T}})^j\mathbf{L}$ in equation (25) can be split by inserting $\mathbf{I} = \mathbf{H} + \mathbf{\Theta}$ between each occurrence of $\mathbf{W}^0$ or $\mathbf{W}^{0\mathrm{T}}$, as in equation (26). The resulting expression can be used to identify the impact of motifs of order k on terms in the expansion of order $i+j \geq k$. The following proposition, proved in appendix E, formalizes these ideas.

Proposition 4.1. *Let* $\mathbf{H}$ *be the rank-1 orthogonal projection matrix generated by the unit* N*-vector* $\mathbf{u}$, $\mathbf{H} = \mathbf{u}\mathbf{u}^{\mathrm{T}}$, *and* $\mathbf{\Theta} = \mathbf{I} - \mathbf{H}$. *For any* $N \times N$ *matrix* $\mathbf{K}$, *let*

$$\mathbf{K}_n = (\mathbf{K}\mathbf{\Theta})^{n-1}\mathbf{K} = \underbrace{\mathbf{K}\mathbf{\Theta}\mathbf{K}\cdots\mathbf{\Theta}\mathbf{K}}_{n \text{ factors of } \mathbf{K}}.$$

If the spectral radii $\Psi(\mathbf{K}) < 1$ *and* $\Psi(\mathbf{K}\mathbf{\Theta}) < 1$, *then*

$$\mathbf{u}^{\mathrm{T}}(\mathbf{I}-\mathbf{K})^{-1}(\mathbf{I}-\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{u} = \left(1 - \sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right)^{-1} \times \left(1 + \sum_{n,m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}\right)\left(1 - \sum_{m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}\right)^{-1}. \tag{30}$$

We will use proposition 4.1 to derive a relation between second order motif strengths and mean covariances. Assuming that $\Psi(\tilde{A}w\mathbf{W}^0), \Psi(\tilde{A}w\mathbf{W}^0\mathbf{\Theta}) < 1$, and setting $\mathbf{u} = \sqrt{N}\mathbf{L}$ and $\mathbf{K} = \tilde{A}w\mathbf{W}^0$, applying proposition 4.1 to equation (7) gives

$$\frac{\langle\tilde{\mathbf{C}}^{\infty}\rangle}{\tilde{C}^0} = \frac{1}{N}\left(1 - \sum_{n=1}^{\infty}(N\tilde{A}w)^n\mathbf{L}^{\mathrm{T}}\mathbf{W}_n^0\mathbf{L}\right)^{-1}\left(1 + \sum_{n,m=1}^{\infty}(N\tilde{A}w)^{n+m}\mathbf{L}^{\mathrm{T}}\mathbf{W}_{n,m}^0\mathbf{L}\right) \cdot \left(1 - \sum_{m=1}^{\infty}(N\tilde{A}w)^m\mathbf{L}^{\mathrm{T}}\mathbf{W}_m^{0\mathrm{T}}\mathbf{L}\right)^{-1}, \tag{31}$$

J. Stat. Mech. (2013) P03012

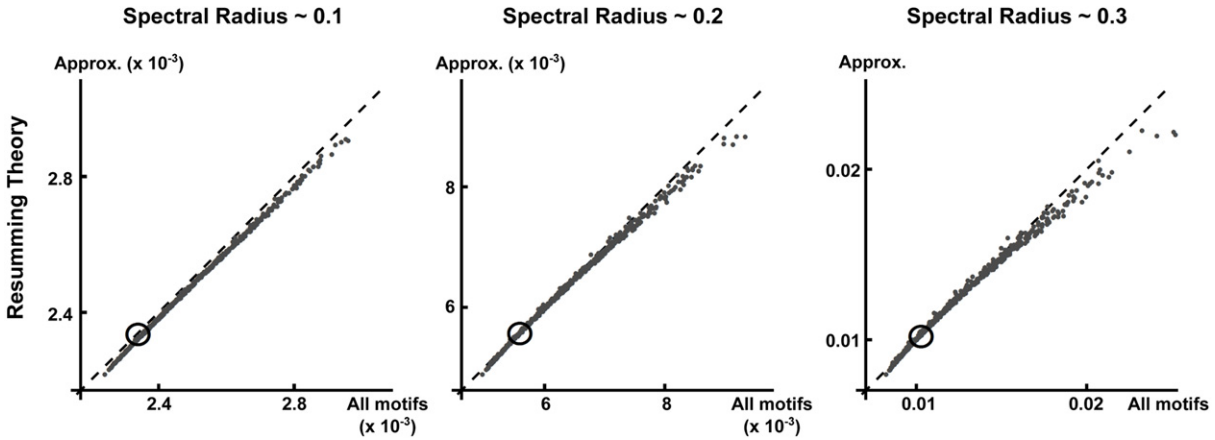

Figure 8. A comparison of the mean correlation obtained using equation (7) (horizontal axes) to the resumming approximation in equation (32) (vertical axes). The diagonal line $y = x$ is plotted for reference. Each panel corresponds to a different scaling of coupling strength for the same set of 512 adjacency matrices as in figure 5, which are sampled from SONET (see section 3.2). The effective coupling strength $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W})$ is recorded at the top of each panel (see equation (9)). Here, the open circle indicates the level of correlation expected from an Erdös–Rényi network with the same overall connection probability and strength.

where

$$\mathbf{W}_n^0 = \frac{1}{N^{n-1}} \underbrace{\mathbf{W}^0 \boldsymbol{\Theta} \mathbf{W}^0 \cdots \boldsymbol{\Theta} \mathbf{W}^0}_{n \text{ factors of } \mathbf{W}^0},$$

$$\mathbf{W}_{n,m}^0 = \frac{1}{N^{n+m-1}} \underbrace{\mathbf{W}^0 \boldsymbol{\Theta} \mathbf{W}^0 \cdots \boldsymbol{\Theta} \mathbf{W}^0}_{n \text{ factors of } \mathbf{W}^0} \boldsymbol{\Theta} \underbrace{\mathbf{W}^{0\mathrm{T}} \boldsymbol{\Theta} \mathbf{W}^{0\mathrm{T}} \cdots \boldsymbol{\Theta} \mathbf{W}^{0\mathrm{T}}}_{m \text{ factors of } \mathbf{W}^{0\mathrm{T}}}.$$

Keeping only terms in equation (25) which can be expressed as polynomials of second order in motif frequency and connection probability is equivalent to keeping only terms involving $\mathbf{W}_1^0$ (connection probability), $\mathbf{W}_2^0$ (chain motifs) and $\mathbf{W}_{1,1}^0$ (diverging motifs) in equation (31). This yields an expression which involves only first and second order motif frequencies:

$$\frac{\langle \tilde{\mathbf{C}}^\infty \rangle}{\tilde{C}^0} = \frac{1}{N} \frac{1 + (N\tilde{A}w)^2 q_{\mathrm{div}}}{[1 - (N\tilde{A}w)p - (N\tilde{A}w)^2 q_{\mathrm{ch}}]^2}. \tag{32}$$

Figure 8 shows that the approximation to the covariance given by equation (32) provides a significant improvement over the second order truncation approximation in equation (24) (see figure 5). We emphasize that it requires only three scalars that summarize the statistics of the entire connection graph: the overall connection probability and the propensity of two second order motifs. We offer a heuristic explanation for the effectiveness of the resumming theory based on spectral analysis of $\mathbf{W}$ in appendix I.

J. Stat. Mech. (2013) P03012

If we expand the denominator in equation (32) as a power series in $[(N\tilde{A}w)p + (N\tilde{A}w)^2 q_{\rm ch}]$, and keep only terms which are linear in $q_{\rm div}, q_{\rm ch}$, we obtain

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^0} = \frac{1}{N(1-N\tilde{A}wp)^2} + \frac{N(\tilde{A}w)^2}{(1-N\tilde{A}wp)^2} q_{\rm div} + \frac{2N(\tilde{A}w)^2}{(1-N\tilde{A}wp)^3} q_{\rm ch} + \text{h.o.t.} \tag{33}$$

We compare the linear coefficients of $q_{\rm div}$, $q_{\rm con}$ and $q_{\rm ch}$ obtained from equation (33) to the empirically estimated coefficients in figure 4 (bar plot).

Finally, we note that similar means may be used to derive versions of equation (32) that (unlike equation (33)) retain a nonlinear dependence on motif frequencies, but keep motifs of either higher or lower order. To approximate the impact of motifs up to order r, we would keep terms with factors $\mathbf{W}_n^0, \mathbf{W}_{n,m}^0$ where $n, n+m \leq r$ in equation (31). For example, if we take $r = 1$, we estimate the mean covariance based only on the probability of occurrence of first order motifs—that is, connection probability. This is equivalent to estimating the covariance in idealized Erdös–Rényi networks where $q_{\rm div}$ and $q_{\rm ch}$, and their higher order analogs are precisely zero. In equation (31), we set all terms involving $\mathbf{W}$ to zero save $\mathbf{W}_1^0$, giving

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^0} = \frac{1}{N(1-N\tilde{A}wp)^2}. \tag{34}$$

This predicted mean correlation is indicated by the open dots in figure 8 (which accurately describes Erdös–Rényi networks as shown in figure 2). This again demonstrates how motif structures exert a large influence over averaged spike correlations.

4.4. Correlations in external input

As is natural for communication among different brain areas or layers [4], or as arises for certain sensory inputs [51], we next consider the case in which the network under study receives an additional noisy input that is *correlated* from cell to cell; in other words, the cells receive common input from upstream sources [31, 11, 52]. We take this input to have total covariance structure ($\omega = 0$) $\tilde{\mathbf{C}}^{\eta} = \sigma_X^2 \mathbf{I} + \sigma_X^2 \rho^{\rm input}(\mathbf{1}_{NN} - \mathbf{I})$, so that the variance of such external input (not to be confused with the implicitly modeled external noise $\xi(t)$ term in equation (3)) to each neuron is fixed (σ_X^2) independent of $\rho^{\rm input}$ and the correlation coefficient of the inputs to all cell pairs is $\rho^{\rm input}$ [24, 53]. Equation (7) then has the form (see [25] for details, and an improvement applicable when the extra noise source is white)

$$\begin{aligned} \tilde{\mathbf{C}}^{\infty} &= (\mathbf{I} - \tilde{A}\mathbf{W})^{-1}(\tilde{C}^0 \mathbf{I} + \tilde{A}^2 \tilde{\mathbf{C}}^{\eta})(\mathbf{I} - \tilde{A}\mathbf{W}^{\rm T})^{-1} \\ &= (\mathbf{I} - \tilde{A}\mathbf{W})^{-1}[(\tilde{C}^0 + \tilde{A}^2\sigma_X^2)\mathbf{I} + (\tilde{A}^2\sigma_X^2\rho^{\rm input})(\mathbf{1}_{NN} - \mathbf{I})](\mathbf{I} - \tilde{A}\mathbf{W}^{\rm T})^{-1}. \end{aligned}$$

In this case, the output variance in the absence of coupling ($\mathbf{W} \equiv 0$) predicted by the linear response theory is $(\tilde{C}^0 + \tilde{A}^2\sigma_X^2)$. Normalizing and multiplying by $\mathbf{L}^{\rm T}, \mathbf{L}$ to arrive at an approximation of average correlation, and applying the ideas of section 4.3, we find that

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^0 + \tilde{A}^2\sigma_X^2} = \frac{B_1[1 + (N\tilde{A}w)^2 q_{\rm div}] + NB_2}{N[1 - (N\tilde{A}w)p - (N\tilde{A}w)^2 q_{\rm ch}]^2},$$

J. Stat. Mech. (2013) P03012

where

$$B_1 = \frac{\tilde{C}^0 + \tilde{A}^2\sigma_X^2(1-\rho^{\text{input}})}{\tilde{C}^0 + \tilde{A}^2\sigma_X^2}, \qquad B_2 = \frac{\tilde{A}^2\sigma_X^2\rho^{\text{input}}}{\tilde{C}^0 + \tilde{A}^2\sigma_X^2}$$

(compare equation (32)). We might ask how the presence of input correlations affects output correlations by calculating the *change in output correlation* resulting from input correlations of size ρ^{input}:

$$\begin{aligned}\Delta\rho^{\text{output}} &= \frac{\langle\tilde{\mathbf{C}}^\infty\rangle}{\tilde{C}^0 + \tilde{A}^2\sigma_X^2} - \frac{\langle\tilde{\mathbf{C}}^\infty\rangle}{\tilde{C}^0 + \tilde{A}^2\sigma_X^2}\bigg|_{\rho^{\text{input}}=0} \\ &= \frac{B_2[N-1-(N\tilde{A}w)^2 q_{\text{div}}]}{N[1-(N\tilde{A}w)p-(N\tilde{A}w)^2 q_{\text{ch}}]^2}. \end{aligned} \tag{35}$$

For simplicity, we can linearize this expression in $q_{\text{div}}, q_{\text{ch}}$ to examine the interaction between second order motifs and input correlations, giving

$$\Delta\rho^{\text{output}} = \frac{B_2}{N}\left[\frac{N-1}{[1-(N\tilde{A}w)p]^2} - \frac{(N\tilde{A}w)^2}{[1-(N\tilde{A}w)p]^2}q_{\text{div}} + \frac{2(N-1)(N\tilde{A}w)^2}{[1-(N\tilde{A}w)p]^3}q_{\text{ch}}\right]. \tag{36}$$

If we assume that $w \sim \mathcal{O}(1/N)$ (so that $\Psi_{\text{ER}}(\tilde{A}\mathbf{W}) = N\tilde{A}w \sim \mathcal{O}(1)$, see section 2.4), then, asymptotically, q_{div} only has an order $\mathcal{O}(1/N)$ (negative) contribution, while q_{ch} has a order $\mathcal{O}(1)$ contribution to $\Delta\rho^{\text{output}}$ (note that the first term in equation (36) is also $\mathcal{O}(1)$, which represents the 'base' response to correlated input in a Erdös–Rényi network). This implies that, in large networks, the chain motif is the most important motif in determining how input correlations will be transferred into network correlations—which will be 'output' to the next area downstream.

5. Impact of second order motifs in networks of excitatory and inhibitory neurons

Biological neuronal networks are composed of excitatory and inhibitory neurons (EI networks). To treat this case, we next show how our theory extends to the case of networks composed of two interacting subpopulations.

5.1. Results: linear dependence between mean correlation and motif frequencies

As in the previous section, we start by a numerical exploration of the contribution of motifs to network-averaged correlation in excitatory–inhibitory networks. In figure 9(A), just as in figure 4 for single population networks, we plot average correlation against motif counts (converging, diverging and chain), without regard for the types of cells involved (excitatory versus inhibitory), or the placement of these types within the motifs (see figure 3). Trends similar to those observed in the case of a single excitatory population are apparent, but they are not as prominent (note the regression coefficients in the right-hand panel of figure 9(A)).

As we will see, the problem is that different sub-types of converging, diverging, and chain motifs (see figure 3) have different contributions to the average correlation, and cannot be simply lumped together. We demonstrate this in figure 10. Here, we evaluated population-averaged correlations (EE, EI, and II) using the full linear response expression

J. Stat. Mech. (2013) P03012

given in equation (7). We then performed a linear regression analysis of the dependence of network correlations on motif frequency in networks of excitatory and inhibitory neurons, for the 20 sub-type second order motifs shown in figure 3. The linear regression gives a reasonable fit ($R^2 \geq 0.69$, see caption). Overall, chain and diverging motifs contribute significantly to the network-averaged correlation, with some motifs having a stronger effect than others. Moreover, mean correlations depend only weakly on converging motifs.

We can now recover the strong dependences on motif counts. In figure 9(B), we demonstrate this for the EE population. Each of the axes is designed as a combined measure of multiple sub-type motif frequencies, which takes into account their distinct contributions to the average correlation (details are given in appendix G). Doing so reveals again a strong positive dependence on chain motifs, a smaller dependence on diverging motifs, and a vanishing dependence on converging motifs.

5.2. First theory: second order truncation for network correlation

To extend the theory developed in section 4 to EI networks, we need to take into account distinct second order motifs. For instance, there are eight different types of three-cell chain (see figure 3). Although this makes the notation more burdensome, the main ideas are the same. Indeed, following a similar approach, the theory can be extended to an arbitrary number of subpopulations.

Consider a network of size $N = N_{\mathrm{E}} + N_{\mathrm{I}}$, consisting of N_{E} excitatory and N_{I} inhibitory neurons. Excitatory (inhibitory) connections have weight w_{E} (w_{I}), so that $w_{\mathrm{E}} > 0$ and $w_{\mathrm{I}} < 0$. The connection probability from class X to class Y is p_{YX}. Note that the connectivity matrix can now be written as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & \mathbf{W}_{EI} \\ \mathbf{W}_{IE} & \mathbf{W}_{II} \end{pmatrix} = \begin{pmatrix} w_{\mathrm{E}}\mathbf{W}^0_{EE} & w_{\mathrm{I}}\mathbf{W}^0_{EI} \\ w_{\mathrm{E}}\mathbf{W}^0_{IE} & w_{\mathrm{I}}\mathbf{W}^0_{II} \end{pmatrix},$$

where $\mathbf{W}_{XY}$ and $\mathbf{W}^0_{XY}$ are, respectively, the weighted and unweighted connection matrices between cells of class Y and cells of class X. The population sizes, weights, connection probabilities, and firing rates together determine the balance between excitatory and inhibitory inputs that cells receive (see Discussion). Throughout this section we use networks with the same parameters and architecture as in figure 10.

First, define the $N \times 2$ block-averaging matrix $\mathbf{L}$ by

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{\mathrm{E}} & 0 \\ 0 & \mathbf{L}_{\mathrm{I}} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{N_{\mathrm{E}},1}/N_{\mathrm{E}} & 0 \\ 0 & \mathbf{1}_{N_{\mathrm{I}},1}/N_{\mathrm{I}} \end{pmatrix},$$

where $\mathbf{1}_{NM}$ is the $N \times M$ matrix of ones. We define the two population analogs of the orthogonal projection matrices $\mathbf{H}$ and $\mathbf{\Theta}$ by

$$\mathbf{H} = \mathbf{L}\mathbf{D}_2\mathbf{L}^{\mathrm{T}}, \qquad \text{and} \qquad \mathbf{\Theta} = \mathbf{I} - \mathbf{H}, \tag{37}$$

where $\mathbf{D}_2$ is the 2×2 matrix

$$\mathbf{D}_2 = \begin{pmatrix} N_{\mathrm{E}} & 0 \\ 0 & N_{\mathrm{I}} \end{pmatrix}.$$

J. Stat. Mech. (2013) P03012

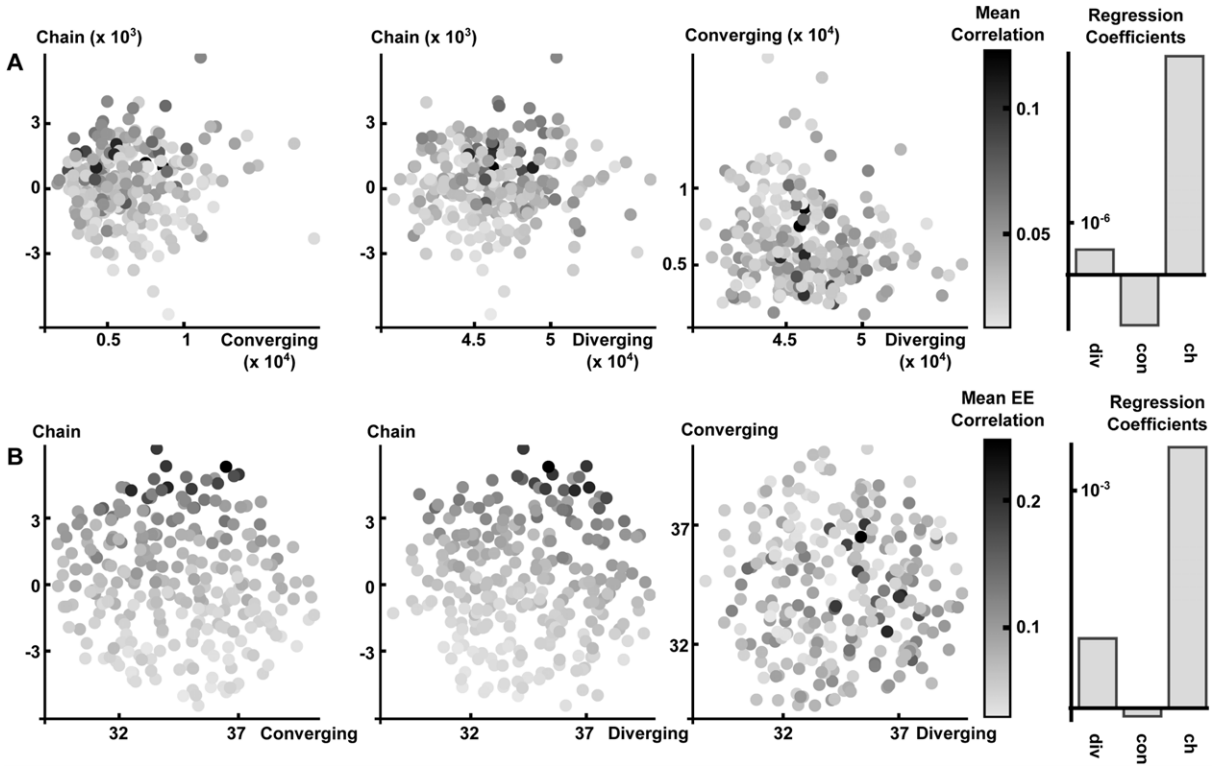

Figure 9. Mean correlation as a function of motif frequency in networks of excitatory and inhibitory neurons. (A) Correlation coefficient averaged over *all* cell pairs as a function of total motif counts. (B) Correlation coefficients averaged over pairs of two excitatory cells, as a function of weighted sums of subgroup motif counts (see text and appendix G for details). The bars on the right show the linear regression coefficients between motif frequencies and average network correlations. The baseline value in these bar plots is 0. The network consists of 51 excitatory neurons and 49 inhibitory neurons with excitatory and inhibitory coupling strengths set at 22.82 or −22.82 mV ms. All neurons have the same uncoupled cellular dynamic parameters (see equation (1)): $\tau_i = 20, v_{\rm th} = 20, v_{\rm r} = -60, \tau_{\rm ref} = 2, E_{{\rm L},i} = -60, E_i = 8, \sigma_i^2 = 12, v_T = -53, \Delta_T = 1.4, \tau_{S,i} = 5, \tau_{D,i} = 10$. Spike count correlation coefficients were calculated using a 500 ms window size. These graphs are re-sampled from 4096 degree distribution generated graphs using Latin hypercube sampling based on parameter space defined by the axis in panel (B) over a reliably sampled region. In panel (A), some of the motifs are given a 'positive' count—i.e. if they involve two excitatory or two inhibitory connections in a chain—and others are given a negative count. For example a converging motif with one excitatory presynaptic cell and one inhibitory presynaptic cell will contribute negatively.

Direct matrix multiplication shows that if $\mathbf{X}$ is a matrix with block form

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{EE} & \mathbf{X}_{EI} \\ \mathbf{X}_{IE} & \mathbf{X}_{II} \end{pmatrix},$$

J. Stat. Mech. (2013) P03012

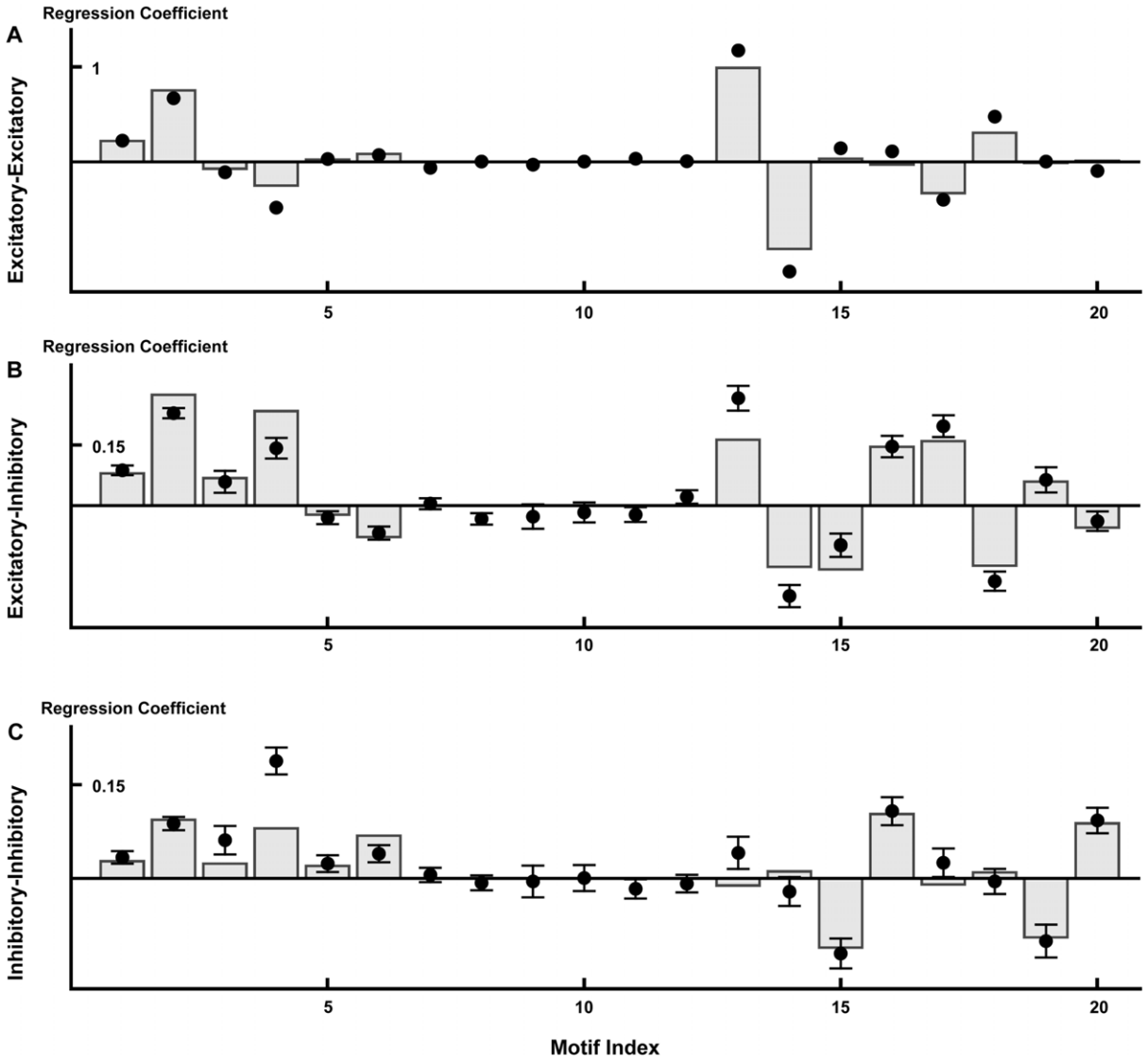

Figure 10. Linear regression coefficients between network-averaged correlations within different subpopulations (vertical axes) and motif frequencies (horizontal axes). The three panels correspond to averages across EE, EI and II cell pairs, from top to bottom. The linear regression coefficient between average correlation and 20 motif frequencies are represented by a dot (see figure 3 for the enumerated list of second order motifs in an EI network). Error bars represent a 95% confidence intervals. In panel (A), confidence intervals are narrower than the dot thickness. The vertical bars were obtained from the resumming theory described in section 5.3. The baseline value in these bar plots is 0. Here $R^2 = 0.91, 0.86, 0.69$ respectively. $\Psi_{\rm ER}(\tilde{A}\mathbf{W}) = 0.25$ (see equation (9)), and $N_{\rm E} = 80$, $N_{\rm I} = 20$, $|w_{\rm I}|/|w_{\rm E}| = 3.707$, so $pN_{\rm E}w_{\rm E} + pN_{\rm I}w_{\rm I} \approx 0$, giving approximate balance between average excitatory and inhibitory inputs. 512 networks are generated using the degree distribution method, as described in section 3.2. Parameters are given in figure 9.

J. Stat. Mech. (2013) P03012

where $\mathbf{X}_{YZ}$ is an $N_Y \times N_Z$ matrix, then $\mathbf{L}^{\mathrm{T}}\mathbf{X}\mathbf{L}$ is a 2×2 matrix of block-wise averages of $\mathbf{X}$, that is,

$$\langle \mathbf{X} \rangle_B \stackrel{\text{def}}{=} \mathbf{L}^{\mathrm{T}}\mathbf{X}\mathbf{L} = \begin{pmatrix} \mathbf{L}_{\mathrm{E}}^{\mathrm{T}}\mathbf{X}_{EE}\mathbf{L}_{\mathrm{E}} & \mathbf{L}_{\mathrm{E}}^{\mathrm{T}}\mathbf{X}_{EI}\mathbf{L}_{\mathrm{I}} \\ \mathbf{L}_{\mathrm{I}}^{\mathrm{T}}\mathbf{X}_{IE}\mathbf{L}_{\mathrm{E}} & \mathbf{L}_{\mathrm{I}}^{\mathrm{T}}\mathbf{X}_{II}\mathbf{L}_{\mathrm{I}} \end{pmatrix} = \begin{pmatrix} \langle \mathbf{X}_{EE} \rangle & \langle \mathbf{X}_{EI} \rangle \\ \langle \mathbf{X}_{IE} \rangle & \langle \mathbf{X}_{II} \rangle \end{pmatrix}.$$

We will make use of the *empirical average connection strength* matrix $\mathbf{M}$ given by

$$\mathbf{M} = \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{L} = \begin{pmatrix} w_{\mathrm{E}}p_{EE} & w_{\mathrm{I}}p_{EI} \\ w_{\mathrm{E}}p_{IE} & w_{\mathrm{I}}p_{II} \end{pmatrix}.$$

To examine the dependence of mean correlations *within a block* on the frequency of second order motifs, we consider the block-wise average of the covariance matrix $\tilde{\mathbf{C}}^{\infty}$,

$$\langle \tilde{\mathbf{C}}^{\infty} \rangle_B = \mathbf{L}^{\mathrm{T}}\tilde{\mathbf{C}}^{\infty}\mathbf{L} = \begin{pmatrix} \langle \tilde{\mathbf{C}}^{\infty}_{EE} \rangle & \langle \tilde{\mathbf{C}}^{\infty}_{EI} \rangle \\ \langle \tilde{\mathbf{C}}^{\infty}_{IE} \rangle & \langle \tilde{\mathbf{C}}^{\infty}_{II} \rangle \end{pmatrix}.$$

Excitatory and inhibitory connection weights need not be equal. It is therefore necessary to consider motif frequencies simultaneously with connection weights. For instance, the contributions of a length two chain passing through an excitatory or inhibitory intermediary cell, such as motifs 13 and 14 in figure 3, are not necessarily equal and opposite in sign. The relative sizes of their contributions are dependent on the ratio $w_{\mathrm{E}}/w_{\mathrm{I}}$. To account for this, we define motif *strength matrices* $\mathbf{Q}_{\mathrm{div}}$, $\mathbf{Q}_{\mathrm{con}}$, and $\mathbf{Q}_{\mathrm{ch}}$ as follows.

The strength of diverging motifs expected in an Erdös–Rényi network is given by

$$\begin{aligned} \mathbf{Q}_{\mathrm{div}}^{\mathrm{ER}} &= \mathbf{M}\mathbf{D}_2\mathbf{M}^{\mathrm{T}} = (\mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{L})\mathbf{D}_2(\mathbf{L}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{L}) \\ &= \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{H}\mathbf{W}^{\mathrm{T}}\mathbf{L} \\ &= \begin{pmatrix} N_{\mathrm{E}}(w_{\mathrm{E}}p_{EE})^2 + N_{\mathrm{I}}(w_{\mathrm{I}}p_{EI})^2 & N_{\mathrm{E}}w_{\mathrm{E}}^2p_{EE}p_{IE} + N_{\mathrm{I}}w_{\mathrm{I}}^2p_{EI}p_{II} \\ N_{\mathrm{E}}w_{\mathrm{E}}^2p_{EE}p_{IE} + N_{\mathrm{I}}w_{\mathrm{I}}^2p_{EI}p_{II} & N_{\mathrm{E}}(w_{\mathrm{E}}p_{IE})^2 + N_{\mathrm{I}}(w_{\mathrm{I}}p_{II})^2 \end{pmatrix}. \end{aligned}$$

Here multiplication by $\mathbf{D}_2$ converts average individual motif strength (e.g. $w_{\mathrm{E}}^2p_{EE}p_{IE}$) to average total motif strength. The empirical average of the strength of diverging motifs is given by

$$\begin{aligned} \mathbf{Q}_{\mathrm{div}}^{\mathrm{total}} &= \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{L} \\ &= \begin{pmatrix} \langle \mathbf{W}_{EE}\mathbf{W}_{EE}^{\mathrm{T}} \rangle + \langle \mathbf{W}_{EI}\mathbf{W}_{EI}^{\mathrm{T}} \rangle & \langle \mathbf{W}_{EE}\mathbf{W}_{IE}^{\mathrm{T}} \rangle + \langle \mathbf{W}_{EI}\mathbf{W}_{II}^{\mathrm{T}} \rangle \\ \langle \mathbf{W}_{IE}\mathbf{W}_{EE}^{\mathrm{T}} \rangle + \langle \mathbf{W}_{II}\mathbf{W}_{EI}^{\mathrm{T}} \rangle & \langle \mathbf{W}_{IE}\mathbf{W}_{IE}^{\mathrm{T}} \rangle + \langle \mathbf{W}_{II}\mathbf{W}_{II}^{\mathrm{T}} \rangle \end{pmatrix}. \end{aligned}$$

Hence, we can write the expected total strength of diverging motifs in excess of that expected in an Erdös–Rényi network as

$$\begin{aligned} \mathbf{Q}_{\mathrm{div}} &= \begin{pmatrix} Q_{\mathrm{div}}^{EE} & Q_{\mathrm{div}}^{EI} \\ Q_{\mathrm{div}}^{IE} & Q_{\mathrm{div}}^{II} \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{Q}_{\mathrm{div}}^{\mathrm{total}} - \mathbf{Q}_{\mathrm{div}}^{\mathrm{ER}} \\ &= \begin{pmatrix} N_{\mathrm{E}}w_{\mathrm{E}}^2q_{\mathrm{div}}^{EE,E} + N_{\mathrm{I}}w_{\mathrm{I}}^2q_{\mathrm{div}}^{EE,I} & N_{\mathrm{E}}w_{\mathrm{E}}^2q_{\mathrm{div}}^{EI,E} + N_{\mathrm{I}}w_{\mathrm{I}}^2q_{\mathrm{div}}^{EI,I} \\ N_{\mathrm{E}}w_{\mathrm{E}}^2q_{\mathrm{div}}^{EI,E} + N_{\mathrm{I}}w_{\mathrm{I}}^2q_{\mathrm{div}}^{EI,I} & N_{\mathrm{E}}w_{\mathrm{E}}^2q_{\mathrm{div}}^{II,E} + N_{\mathrm{I}}w_{\mathrm{I}}^2q_{\mathrm{div}}^{II,I} \end{pmatrix} \\ &= \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{L} - \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{H}\mathbf{W}^{\mathrm{T}}\mathbf{L} \end{aligned}$$

J. Stat. Mech. (2013) P03012

$$= \mathbf{L}^{\mathrm{T}}\mathbf{W}(\mathbf{H}+\mathbf{\Theta})\mathbf{W}^{\mathrm{T}}\mathbf{L} - \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{H}\mathbf{W}^{\mathrm{T}}\mathbf{L}$$
$$= \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{\Theta}\mathbf{W}^{\mathrm{T}}\mathbf{L} \tag{38}$$

where

$$q_{\mathrm{div}}^{XY,Z} = \frac{1}{N_Z}\langle \mathbf{W}_{XZ}^{0}\mathbf{W}_{YZ}^{0\mathrm{T}}\rangle - p_{XZ}p_{YZ}$$

represents the probability of observing a diverging motif to cells of classes X, Y from a cell of class Z in excess of that expected in an Erdös–Rényi network.

It is important to note that the matrix $\mathbf{Q}_{\mathrm{div}}$ contains motif *strengths* in excess of what would be expected in an Erdös–Rényi network (i.e. number of occurrences, scaled by connection weights and probability of occurrence), while the scalars q_{div} still correspond to probabilities.

The strengths and frequencies of converging motifs can be expressed similarly, giving

$$\mathbf{Q}_{\mathrm{con}} = \begin{pmatrix} Q_{\mathrm{con}}^{EE} & Q_{\mathrm{con}}^{EI} \\ Q_{\mathrm{con}}^{IE} & Q_{\mathrm{con}}^{II} \end{pmatrix} \stackrel{\mathrm{def}}{=} \mathbf{L}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{L} - \mathbf{M}^{\mathrm{T}}\mathbf{D}_2\mathbf{M} = \mathbf{L}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{\Theta}\mathbf{W}\mathbf{L}$$
$$= \begin{pmatrix} N_{\mathrm{E}}w_{\mathrm{E}}^2 q_{\mathrm{con}}^{EE,E} + N_{\mathrm{I}}w_{\mathrm{E}}^2 q_{\mathrm{con}}^{EE,I} & N_{\mathrm{E}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{con}}^{EI,E} + N_{\mathrm{I}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{con}}^{EI,I} \\ N_{\mathrm{E}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{con}}^{EI,E} + N_{\mathrm{I}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{con}}^{EI,I} & N_{\mathrm{E}}w_{\mathrm{I}}^2 q_{\mathrm{con}}^{II,E} + N_{\mathrm{I}}w_{\mathrm{I}}^2 q_{\mathrm{con}}^{II,I} \end{pmatrix}, \tag{39}$$

where

$$q_{\mathrm{con}}^{XY,Z} = \frac{1}{N_Z}\langle \mathbf{W}_{XZ}^{0\mathrm{T}}\mathbf{W}_{YZ}^{0}\rangle - p_{ZX}p_{ZY},$$

represents the probability of observing a converging motif to cells of class Z from cells of classes X, Y in excess of that expected in an Erdös–Rényi network.

Finally, for chain motifs,

$$\mathbf{Q}_{\mathrm{ch}} = \begin{pmatrix} Q_{\mathrm{ch}}^{EE} & Q_{\mathrm{ch}}^{EI} \\ Q_{\mathrm{ch}}^{IE} & Q_{\mathrm{ch}}^{II} \end{pmatrix} \stackrel{\mathrm{def}}{=} \mathbf{L}^{\mathrm{T}}\mathbf{W}^2\mathbf{L} - \mathbf{M}\mathbf{D}_2\mathbf{M} = \mathbf{L}^{\mathrm{T}}\mathbf{W}\mathbf{\Theta}\mathbf{W}\mathbf{L}$$
$$= \begin{pmatrix} N_{\mathrm{E}}w_{\mathrm{E}}^2 q_{\mathrm{ch}}^{EEE} + N_{\mathrm{I}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{ch}}^{EIE} & N_{\mathrm{E}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{ch}}^{IEE} + N_{\mathrm{I}}w_{\mathrm{I}}^2 q_{\mathrm{ch}}^{IIE} \\ N_{\mathrm{E}}w_{\mathrm{E}}^2 q_{\mathrm{ch}}^{EEI} + N_{\mathrm{I}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{ch}}^{EII} & N_{\mathrm{E}}w_{\mathrm{E}}w_{\mathrm{I}} q_{\mathrm{ch}}^{IEI} + N_{\mathrm{I}}w_{\mathrm{I}}^2 q_{\mathrm{ch}}^{III} \end{pmatrix}, \tag{40}$$

where

$$q_{\mathrm{ch}}^{ZYX} = \frac{1}{N_Y}\langle \mathbf{W}_{ZY}^{0}\mathbf{W}_{YX}^{0}\rangle - p_{ZY}p_{YX}$$

represents the probability of observing a length two chain motif beginning at a cell of type X and terminating at a cell of type Z, passing through a cell of type Y, in excess of what would be expected in an Erdös–Rényi network.

A truncation of equation (18) at second order gives an initial approximation of the block-wise average $\langle \tilde{\mathbf{C}}(0)\rangle_B$:

$$\langle \tilde{\mathbf{C}}^{\infty}\rangle_B/\tilde{C}^0 = \mathbf{L}^{\mathrm{T}}\left[\mathbf{I} + \tilde{A}(\mathbf{W}+\mathbf{W}^{\mathrm{T}}) + \tilde{A}^2(\mathbf{W}^2 + \mathbf{W}^{2\mathrm{T}} + \mathbf{W}\mathbf{W}^{\mathrm{T}})\right]\mathbf{L} + \text{h.o.t.}$$
$$\approx [\mathbf{L}^{\mathrm{T}}\mathbf{L} + \tilde{A}(\mathbf{M}+\mathbf{M}^{\mathrm{T}}) + \tilde{A}^2(\mathbf{M}\mathbf{D}_2\mathbf{M} + \mathbf{M}^{\mathrm{T}}\mathbf{D}_2\mathbf{M}^{\mathrm{T}} + \mathbf{M}\mathbf{D}_2\mathbf{M}^{\mathrm{T}})$$
$$+ \tilde{A}^2(\mathbf{Q}_{\mathrm{ch}} + \mathbf{Q}_{\mathrm{ch}}^{\mathrm{T}} + \mathbf{Q}_{\mathrm{div}})]. \tag{41}$$

J. Stat. Mech. (2013) P03012

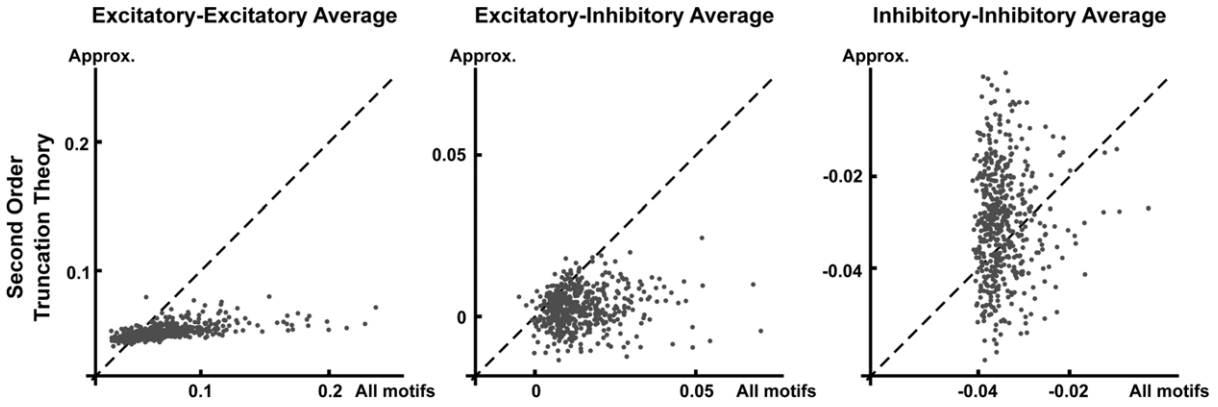

Figure 11. Block-wise average correlations obtained from the second order truncation of equation (41). On the horizontal axis is the average correlation between cells from the given classes (calculated from equation (7)), and the approximate value obtained from the truncation is given on the vertical axis. The diagonal line $y = x$ corresponds to perfect agreement between the true value and the approximation. $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.33$ (see equation (9)); other network parameters are the same as in figure 10, 12 and 13. Here $R^2 = 0.33, 0.02, 0.0005$ respectively for the three panels, confirming that the truncation approach gives a poor prediction of network correlation.

Terms not involving a second order motif matrix correspond to the contributions of motifs up to second order in a purely Erdös–Rényi network—for example, $\mathbf{MD_2M}$ gives the expected contribution of length two chains in an Erdös–Rényi network.

Figure 11 compares the second order truncation given by equation (41) with the mean correlations obtained from the entire series in equation (18). The correlations in the network of inhibitory and excitatory cells can be appreciable, even when the spectral radius of matrix $\tilde{A}\mathbf{W}$ is much smaller than one. However, the second order truncation of equation (41) gives a poor approximation of network-averaged correlations.

5.3. Improved theory: resumming to approximate higher order contributions to network correlation

As in the case of a single population, we can improve our prediction of mean correlations by accounting for the contributions of second order motifs to all orders in connection strength. The equivalent of equation (25) has the form

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle_B}{\tilde{C}^0} = \frac{1}{\tilde{C}^0} \mathbf{L}^{\mathrm{T}} \tilde{\mathbf{C}}^{\infty} \mathbf{L} = \sum_{i,j=0}^{\infty} \tilde{A}^{i+j} \mathbf{L}^{\mathrm{T}} \mathbf{W}^i (\mathbf{W}^{\mathrm{T}})^j \mathbf{L}, \tag{42}$$

where $\mathbf{W}$ and $\mathbf{L}$ are as defined in section 5.2.

First, we generalize proposition 4.1 to the case of two populations (for a full proof, see appendix F)

Proposition 5.1. *Let* $\mathbf{H} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ *be an orthogonal projection matrix generated by an* $N \times M$ *matrix* $\mathbf{U}$*, whose columns are orthonormal vectors. Define* $\mathbf{\Theta} = \mathbf{I} - \mathbf{H}$*. For any*

$N \times N$ matrix $\mathbf{K}$*, define* $\mathbf{K}_n$ *as*

$$\mathbf{K}_n = \underbrace{\mathbf{K}\boldsymbol{\Theta}\mathbf{K}\cdots\boldsymbol{\Theta}\mathbf{K}}_{n \text{ factors of } \mathbf{K}}.$$

If spectral radius $\Psi(\mathbf{K}), \Psi(\mathbf{K}\boldsymbol{\Theta}) < 1$*, we have*

$$\mathbf{U}^{\mathrm{T}}(\mathbf{I}-\mathbf{K})^{-1}(\mathbf{I}-\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{U} = \left(\mathbf{I}-\sum_{n=1}^{\infty}\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\right)^{-1} \times \left(\mathbf{I}+\sum_{n,m=1}^{\infty}\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\boldsymbol{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{U}\right)\left(\mathbf{I}-\sum_{m=1}^{\infty}\mathbf{U}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{U}\right)^{-1}. \tag{43}$$

We now apply this proposition to the expression in equation (42). Let $\mathbf{U} = \mathbf{L}\mathbf{D}_2^{1/2}$ and $\mathbf{U}^{\mathrm{T}}\mathbf{X}\mathbf{U} = \mathbf{D}_2^{1/2}\mathbf{L}^{\mathrm{T}}\mathbf{X}\mathbf{L}\mathbf{D}_2^{1/2}$ for any matrix $\mathbf{X}$, so that $\mathbf{H}$ has the form given in equation (37). In addition, let $\mathbf{K} = \tilde{A}\mathbf{W}$, and assume that $\Psi(\tilde{A}\mathbf{W})$, $\Psi(\tilde{A}\boldsymbol{\Theta}\mathbf{W}) < 1$. Then proposition 5.1 gives

$$\begin{aligned}\langle\tilde{\mathbf{C}}^{\infty}\rangle_B/\tilde{C}^0 &= \mathbf{L}^{\mathrm{T}}(\mathbf{I}-\tilde{A}\mathbf{W})^{-1}(\mathbf{I}-\tilde{A}\mathbf{W}^{\mathrm{T}})^{-1}\mathbf{L} \\ &= \left(\mathbf{I}-\sum_{n=1}^{\infty}\tilde{A}^n\mathbf{L}^{\mathrm{T}}\mathbf{W}_n\mathbf{L}\mathbf{D}_2\right)^{-1}\left(\mathbf{D}_2^{-1}+\sum_{n,m=1}^{\infty}\tilde{A}^{n+m}\mathbf{L}^{\mathrm{T}}\mathbf{W}_n\boldsymbol{\Theta}\mathbf{W}_m^{\mathrm{T}}\mathbf{L}\right) \\ &\quad\cdot\left(\mathbf{I}-\sum_{m=1}^{\infty}\tilde{A}^m\mathbf{D}_2\mathbf{L}^{\mathrm{T}}\mathbf{W}_m^{\mathrm{T}}\mathbf{L}\right)^{-1},\end{aligned} \tag{44}$$

where

$$\mathbf{W}_n = \underbrace{\mathbf{W}\boldsymbol{\Theta}\mathbf{W}\cdots\boldsymbol{\Theta}\mathbf{W}}_{n \text{ factors of } \mathbf{W}}.$$

As in the single population case, we can discard all terms in equation (44) which do not correspond to second order motif frequencies. This means that in the first and third sets of brackets in equation (44) we discard any terms containing $\mathbf{W}_n$ with $n \geq 3$, and for the middle set of brackets we discard terms containing $\mathbf{W}_n\boldsymbol{\Theta}\mathbf{W}_m^{\mathrm{T}}$ for $n+m \geq 3$. This gives

$$\langle\tilde{\mathbf{C}}^{\infty}\rangle_B/\tilde{C}^0 \approx (\mathbf{I}-\tilde{A}\mathbf{M}\mathbf{D}_2-\tilde{A}^2\mathbf{Q}_{\mathrm{ch}}\mathbf{D}_2)^{-1}(\mathbf{D}_2^{-1}+\tilde{A}^2\mathbf{Q}_{\mathrm{div}})(\mathbf{I}-\tilde{A}\mathbf{D}_2\mathbf{M}^{\mathrm{T}}-\tilde{A}^2\mathbf{D}_2\mathbf{Q}_{\mathrm{ch}}^{\mathrm{T}})^{-1}. \tag{45}$$

Figure 12 illustrates that this approximation is a great improvement over that given by truncating the expansion at second order (compare with figure 11). Again, we note that our approximation requires only knowledge of the overall connection probabilities among excitatory and inhibitory cells, and the frequency of second order motifs.

Expanding the inverses in equation (45) in a power series, we can again obtain an approximation to block average correlations to linear order in $\mathbf{Q}_{\mathrm{ch}}$, and $\mathbf{Q}_{\mathrm{div}}$,

$$\begin{aligned}\langle\tilde{\mathbf{C}}^{\infty}\rangle_B/\tilde{C}^0 \approx{}& (\mathbf{I}-\tilde{A}\mathbf{M}\mathbf{D}_2)^{-1}[\mathbf{D}_2^{-1}+\tilde{A}^2\mathbf{Q}_{\mathrm{ch}}(\mathbf{I}-\tilde{A}\mathbf{D}_2\mathbf{M})^{-1}+(\mathbf{I}-\tilde{A}\mathbf{M}^{\mathrm{T}}\mathbf{D}_2)^{-1}\tilde{A}^2\mathbf{Q}_{\mathrm{ch}}^{\mathrm{T}} \\ &+\tilde{A}^2\mathbf{Q}_{\mathrm{div}}](\mathbf{I}-\tilde{A}\mathbf{D}_2\mathbf{M}^{\mathrm{T}})^{-1}.\end{aligned} \tag{46}$$

As in the single population case, each entry of the 2×2 matrix on the right-hand side of equation (46) gives an approximation to the block-averaged correlations expressed in terms of the scalars $q_{\mathrm{ch}}^{ZYX}, q_{\mathrm{div}}^{XY,Z}$, providing an analytical estimate to the regression

J. Stat. Mech. (2013) P03012

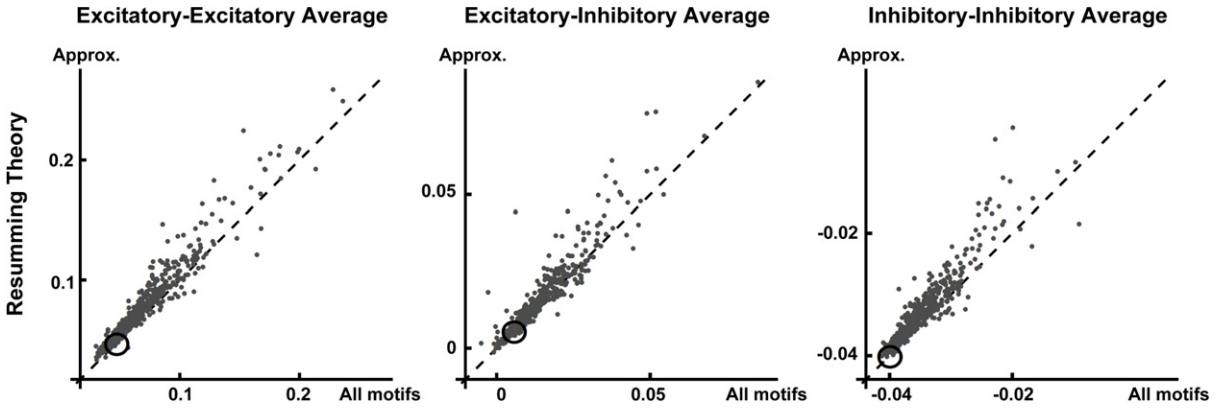

Figure 12. Predicting block-wise average correlations from resumming theory. The horizontal axis is the average correlation (in a certain block) from the original linear response expression of the covariance matrix (equation (7)); the vertical axis the quantity from resumming theory (equation (45)). The diagonal line $y = x$ is plotted for reference. $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.33$ (see equation (9)); other network parameters are the same as in figure 10. Here $R^2 = 0.93, 0.88, 0.87$ respectively for the three panels. The open circle indicates the level of correlation expected from an Erdös–Rényi network with the same overall connection probability and strength.

coefficients plotted in figure 10. An example with uniform connection probability is given in appendix H. We note that in general all sub-types of diverging and converging motifs affect each block-wise average correlation. This is not predicted by the second order truncation in equation (41). Somewhat counterintuitively, $q_{\mathrm{div}}^{EI,E}$ and $q_{\mathrm{div}}^{EI,I}$ (indices 3 and 4 in figure 3) can contribute negatively to $\langle \tilde{\mathbf{C}}_{EE} \rangle$ as shown in figure 10.

As in the single population case, we can also retain contributions to mean correlation which can be expressed as (nonlinear) functions of first order motifs (connection probabilities) only. This follows from setting all $\mathbf{Q}$ terms in equation (45) to zero, yielding the following approximation for mean correlation in the two population analog of Erdös–Rényi networks:

$$\langle \tilde{\mathbf{C}}^{\infty} \rangle_B / \tilde{C}^0 = (\mathbf{I} - \tilde{A}\mathbf{M}\mathbf{D}_2)^{-1} \mathbf{D}_2^{-1} (\mathbf{I} - \tilde{A}\mathbf{D}_2\mathbf{M}^{\mathrm{T}})^{-1}. \tag{47}$$

This predicted mean correlation for an Erdös–Rényi network is shown by an open dot in figure 12; the deviations from this value illustrate that motif structures can both increase and decrease average network correlations (see also figures 2 and 8).

6. Heterogeneous networks

For simplicity in the previous sections we assumed a homogeneous network of neurons composed of cells with identical firing rates, power spectra and response properties. As a result, the diagonal matrices $\tilde{\mathbf{C}}^0$ and $\tilde{\mathbf{A}}$ in our key expression for correlations, equation (6), were scalar matrices, $\tilde{C}^0\mathbf{I}$ and $\tilde{A}\mathbf{I}$, and could be factored—leaving the connectivity structure of the network to determine the correlation value. In particular,

J. Stat. Mech. (2013) P03012

this structure impacts correlations via the matrix products of adjacency matrices in equation (18).

Biological neural networks are heterogeneous. Even if the neurons are identical before they are coupled, any heterogeneities in the coupling structure will lead to different neurons firing with different rates and hence power spectra (cf [20], to which we return in the discussion). Moreover, as a consequence, they will also have different levels of responsivity. In sum, the matrices $\tilde{\mathbf{C}}^0$ and $\tilde{\mathbf{A}}$ will not have identical entries on the diagonal.

Consequently, the equivalent of the expansion equation (18) takes the form

$$\tilde{\mathbf{C}}^{\infty} = \sum_{i,j=0}^{\infty} (\tilde{\mathbf{A}}^{*}\mathbf{W})^{i} \tilde{\mathbf{C}}^{0} (\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{A}})^{j}. \tag{48}$$

As a result, a purely graph theoretic interpretation of the terms in the expansion—that is, one based on the connectivity *alone*—is no longer possible. Recall that, as in equations (11)–(13), motif frequencies are associated with terms such as $\mathbf{W}^{i}(\mathbf{W}^{\mathrm{T}})^{j}$. Here the powers $(\tilde{\mathbf{A}}\mathbf{W})^{i}$ are of a 'weighted' connection matrix (with weights corresponding to the responsivity of different cells). Moreover, $\tilde{\mathbf{C}}^0$ is no longer a scalar matrix and in general does not commute with $\mathbf{W}$, introducing additional complications that we address below.

In this section we discuss how to extend our results to the case of heterogeneous neural populations.

6.1. Performance of the homogeneous approximation

A first attempt at coping with heterogeneity is to hope that it is unimportant, and to apply the homogeneous series expansion of equation (18) and corresponding resumming theory of equation (45) naively. To do this, we need to choose approximate (scalar) values for the power spectrum $\tilde{C}^0$ and responsivity $\tilde{A}$ that we will apply to all of the cells, by plugging into the homogeneous network formula given by equation (45). We choose the $\tilde{C}^0$ as the unadjusted power spectrum (due to the normalization equation (2), the actual value of $\tilde{C}^0$ is not important). For $\tilde{A}$, we use the geometric mean of the $\tilde{A}_i$, which are found from self-consistent equations similar to equation (5) (see [25] for details). The results of this naive application of the homogeneous theory are shown in panel (A) of figure 13. Although general trends are captured, this approach does not give an accurate approximation.

We note that here coupling is relatively strong ($\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.4$; see equation (9)). With weaker coupling ($\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.3$), and hence less network-driven heterogeneity, the homogeneity assumption provides improved approximations in the heterogeneous case (R^2 measure (between resumming theory equation (45) and full linear response equation (6)) 0.78, 0.67 and 0.67 for EE, EI and II average correlation respectively, data not shown). To quantify the heterogeneity of cellular dynamics across the networks, we compute the coefficient of variation CV of $\tilde{A}_i$ and $\tilde{C}_i^0$ averaged over network samples (CV = standard dev./mean). For spectral radius = 0.4 the CV is 0.35 and 0.38 for $\tilde{A}_i$ and $\tilde{C}_i^0$, respectively, while at spectral radius = 0.3 the CV is 0.23 and 0.26 for $\tilde{A}_i$ and $\tilde{C}_i^0$. At such levels of heterogeneity, we clearly need a more systematic approach, and we develop this next.

J. Stat. Mech. (2013) P03012

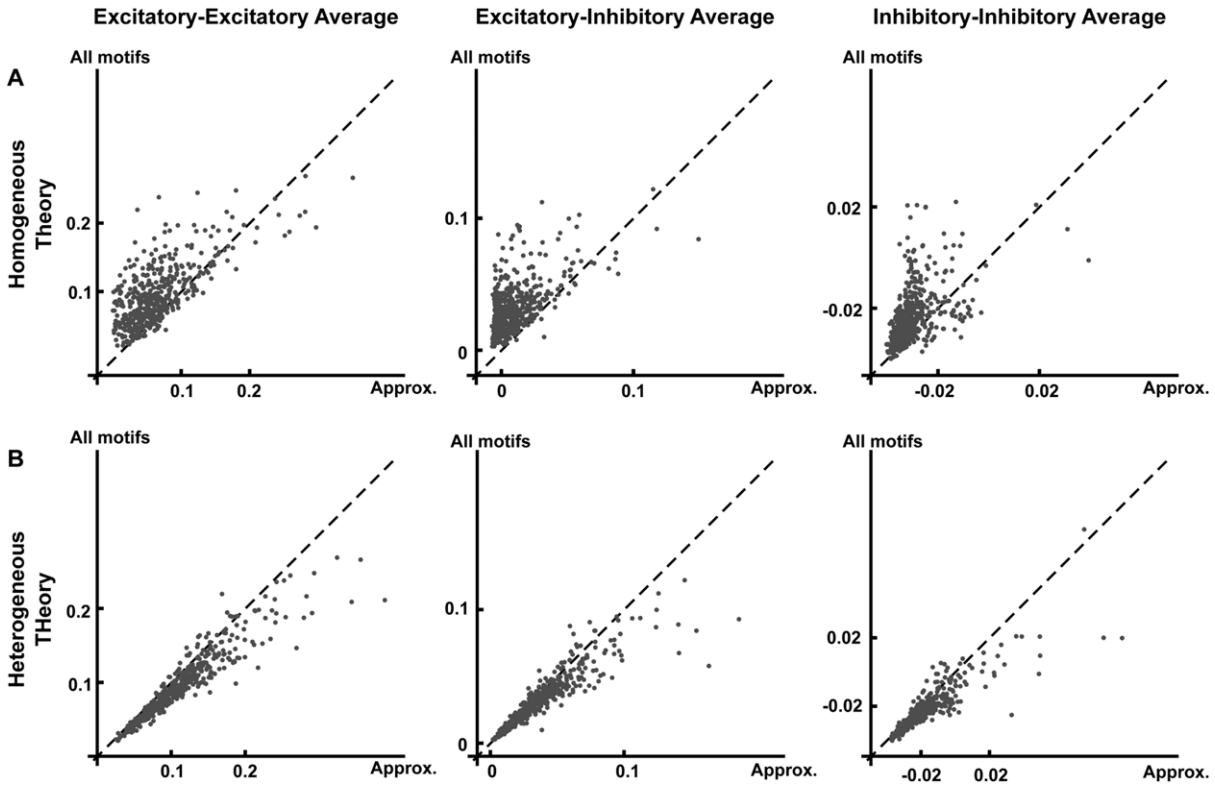

Figure 13. (A) Mean correlation from homogeneous resumming theory (equation (45), horizontal axis) compared with that from full linear response theory (equation (6), vertical axis). (B) Mean correlation from heterogeneous resumming theory (horizontal axis) compared with that from full linear response theory (vertical axis). The diagonal line $y = x$ is plotted for reference. $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.40$ (see equation (9)); other network parameters are given in the caption of figure 10. The coefficients of determination, R^2, are 0.56, 0.44, 0.35 in panel (A) and 0.88, 0.81, 0.73 for panel (B).

6.2. Heterogeneous theory

We are prevented from applying proposition 4.1 to equation (48) in the heterogeneous case, due to the presence of the factor $\tilde{\mathbf{C}}^0$ in the middle of the terms on the right-hand side of equation (48). To deal with this difficulty we use the substitution $\tilde{\mathbf{C}}^0 = (\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{1/2}$, which is possible because the power spectrum is non-negative. We can then rewrite equation (6) as

$$\begin{aligned}\tilde{\mathbf{C}} &\approx (\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{-1/2}(\mathbf{I} - \tilde{\mathbf{A}}\mathbf{W})^{-1}(\tilde{\mathbf{C}}^0)^{1/2}(\tilde{\mathbf{C}}^0)^{1/2}(\mathbf{I} - \mathbf{W}^{\mathrm{T}}\tilde{\mathbf{A}})^{-1}(\tilde{\mathbf{C}}^0)^{-1/2}(\tilde{\mathbf{C}}^0)^{1/2} \\ &= (\tilde{\mathbf{C}}^0)^{1/2}(\mathbf{I} - (\tilde{\mathbf{C}}^0)^{-1/2}\tilde{\mathbf{A}}\mathbf{W}(\tilde{\mathbf{C}}^0)^{1/2})^{-1}(\mathbf{I} - (\tilde{\mathbf{C}}^0)^{1/2}\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{A}}(\tilde{\mathbf{C}}^0)^{-1/2})^{-1}(\tilde{\mathbf{C}}^0)^{1/2},\end{aligned} \tag{49}$$

where we are again evaluating all quantities at $\omega = 0$, so that $\tilde{\mathbf{F}} = \mathbf{I}$.

Let $\tilde{A}_0$ be the geometric mean of $\tilde{A}_i$, which we choose in order to normalize responsivity so that weighted quantities will have the same units (see below). We can then define an

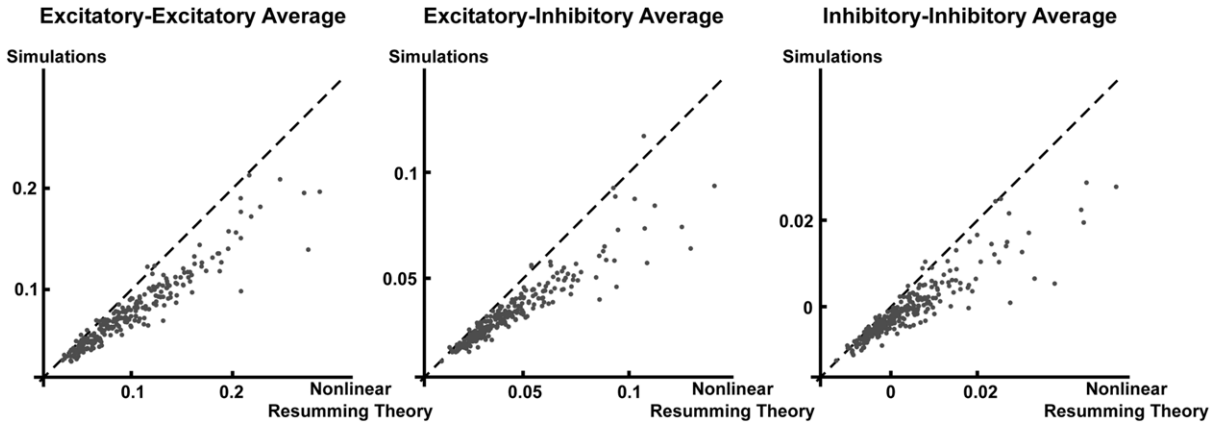

Figure 14. Average correlation from integrate and fire neuron simulations compared with the predictions of equation (45). The vertical axis is the mean correlation coefficient calculated from simulations of the same 265 excitatory–inhibitory networks studied in figure 9. The horizontal axis is the prediction for mean correlation from the resumming theory based on empirical connection probability and second order motif frequencies. The diagonal line $y = x$ is plotted for reference. $\Psi_{\mathrm{ER}}(\tilde{A}\mathbf{W}) = 0.33$ (see equation (9)). Coefficients of determination R^2 are 0.91, 0.87, 0.79 respectively for the three panels.

effective or 'functional' connection matrix, which has the same units as $\mathbf{W}$:

$$\hat{\mathbf{W}} = (\tilde{\mathbf{C}}^0)^{-1/2}\tilde{\mathbf{A}}\mathbf{W}(\tilde{\mathbf{C}}^0)^{1/2}/\tilde{A}_0, \tag{50}$$

so that equation (49) becomes

$$(\tilde{\mathbf{C}}^0)^{1/2}(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}})^{-1}(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}}^{\mathrm{T}})^{-1}(\tilde{\mathbf{C}}^0)^{1/2}\,. \tag{51}$$

This expression will be much easier to study, as the diagonal matrix $\tilde{\mathbf{C}}^0$ no longer appears in the middle. We can thus expand the two terms, $(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}})^{-1}$ and $(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}}^{\mathrm{T}})^{-1}$, to obtain an expression that has a form analogous to that of equation (18).

The only difference in the present case of a heterogeneous network is that the definition of motif frequency and connection probability will involve weighted averages. For example, the entries of $\hat{\mathbf{W}}$ are scaled version of entries of the connection matrix $\mathbf{W}$:

$$\hat{W}_{ij} = \frac{\tilde{A}_i}{\tilde{A}_0}\frac{\sqrt{\tilde{S}_j}}{\sqrt{\tilde{S}_i}}W_{ij}.$$

As an example, a diverging motif $i \leftarrow k \rightarrow j$ has a weighted contribution of the form

$$\frac{\tilde{A}_i\tilde{A}_j}{\tilde{A}_0^2}\frac{\tilde{C}_k^0}{\sqrt{\tilde{C}_i^0\tilde{C}_j^0}}W_{ik}W_{jk}. \tag{52}$$

The ratio $(\tilde{A}_i\tilde{A}_j)/\tilde{A}_0^2$ in equation (52) quantifies the relative responsivity of the post-synaptic cells. Hence, a particular diverging motif will be weighted more strongly (and provide a greater contribution to average correlation) if the recipient cells are more responsive to inputs. Similarly, $\tilde{C}_i^0, \tilde{C}_j^0, \tilde{C}_k^0$ corresponds to the variance of spike counts

J. Stat. Mech. (2013) P03012

in long-time windows for the uncoupled cells: the weight is determined by the variance of the projecting cell divided by the geometric mean of the variance of the post-synaptic cells. These observations agree well with intuition: more responsive cells will be more strongly correlated by a common input, and 'source cells' with larger total variance ($\tilde{C}_k^0$) will lead to a diverging motif with larger impact.

The motif frequencies $q_{\rm div}$ and $q_{\rm ch}$ are defined by equations (20) and (22) upon substituting $\mathbf{W}^0$ with $\hat{\mathbf{W}}^0 = \hat{\mathbf{W}}/w$ in the single population case. In the case of two populations, the matrices $\mathbf{Q}_{\rm div}$ and $\mathbf{Q}_{\rm ch}$ are defined by equations (38) and (40), with $\mathbf{W}$ replaced by $\hat{\mathbf{W}}$. These weighted motif frequencies could be estimated experimentally in an active network, via recordings from neurons known to participate in three-cell motifs. The advantage of our resumming theory remains that everything needed is based only on the statistics of a relatively small number of cell motifs, rather than higher order information about the connectivity graph.

Due to the normalization in equation (2), the two outside diagonal matrix factors in equation (51) cancel in the definition of $\boldsymbol{\rho}$, and we can use the following matrix to calculate correlation coefficients:

$$(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}})^{-1}(\mathbf{I} - \tilde{A}_0\hat{\mathbf{W}}^{\rm T})^{-1}. \tag{53}$$

Equation (53) is the heterogeneous analog of the expression for the correlation $\tilde{\mathbf{C}}^\infty/\tilde{C}^0$ studied in section 4 and 5 (see equation (7)), including diagonal contributions. Applying the motif resumming theory exactly as in these sections yields corresponding approximations of mean correlation for these systems. In particular, for the two population case, the mean correlation $\rho_{EE}^{\rm avg}, \rho_{EI}^{\rm avg}, \rho_{II}^{\rm avg}$ will be estimated based on equation (45) with the substitution of weighted motif counts $\mathbf{Q}_{\rm div}$ and $\mathbf{Q}_{\rm ch}$ (note that one must still adjust the approximation to account for diagonal terms—see appendix A). In panel (B) of figure 13, we plot this prediction of $\rho_{EE}^{\rm avg}, \rho_{EI}^{\rm avg}, \rho_{II}^{\rm avg}$ compared with the full expression for correlation in the heterogeneous case (via equations (2) and (6)). Accounting for dynamical heterogeneity across the network by defining *weighted* second order motifs provides a reasonably accurate prediction of average correlation.

7. Comparisons with IF simulations

In the preceding sections, we have shown how network-wide correlation can be approximated based on coupling probability together with the frequency of second order motifs. In assessing the accuracy of this motif-based theory, we have compared the predictions of the theory with the value of correlation given equation (6), which is exact in the sense that it precisely includes the contribution of motifs at all orders and therefore gives an exact description of linearly interacting point process models [35, 54].

When using our theory to predict the impact of motifs on mean correlation for networks of IF neurons, we are making an additional approximation in describing integrate and fire dynamics with linear response theory [25, 41, 24]. We now directly test the performance of our motif-based predictions in IF networks—thus probing how errors at each of the two layers of approximation combine. Specifically, in figure 14 we compare the block-wise mean correlations from IF simulation and the predications obtained using equation (45). These simulations are for the same networks used in figure 9. We find that our theory gives a good quantitative match with LIF simulations, despite the

J. Stat. Mech. (2013) P03012

multiple approximations involved. Thus, our theory predicts trends in the impact of motif frequencies on network-wide correlation.

8. Discussion

Summary: predicting network-wide correlation from three-cell motifs

We studied the impact of the graphical structure of neural networks—characterized by connectivity motifs—on their dynamical structure, characterized by correlated activity between cells. As shown in figure 2, varying the frequency of such motifs can strongly impact correlation, over and above the overall level of connectivity in a network. Following [20], we focus on the three types of motif that involve two connections each: the diverging, converging and chain motifs.

We chose a standard spiking neuron model, the integrate and fire (IF) neuron, in constructing our recurrent networks (section 2). For IF neurons [24, 25], LNP neurons [12], and other neuron models such as the linearly interacting point process model [54, 55], one can apply linear response approximations (or, in the case of Hawkes models, solve exact equations—see 'Neuron models' below) to get an explicit expression for the pairwise correlations depending explicitly on the connectivity matrix (see equation (6)).

We expand this expression in a series, where each term has clear correspondence to certain graphical structures [35, 49]. In particular, second order terms correspond to second order motifs. Importantly, we show that contributions of higher order terms can be estimated using the frequencies of second order motifs along with the connection probability.

For systems with correlations well approximated by equation (6) and assuming homogeneous cellular dynamics, we find that the frequency of converging motifs, over and above that expected in Erdös–Rényi networks, will have no effect on mean correlation in systems (if the connection probability and other motif frequencies are fixed). Meanwhile, diverging and chain motifs will contribute positively. In networks of excitatory and inhibitory neurons, the three types of motif are subdivided according to the type of the constituent neurons. However, average correlations between cells of a given type are still given in terms of diverging and chain motifs (see figure 9).

We first analyzed networks in which we made a strong homogeneity assumption on the dynamical properties of the uncoupled neurons. The resumming theory we develop approximates the contributions of higher order motifs in terms of the frequency of second order motifs. In section 6, we extended our theory to heterogeneous networks. In such cases, the contribution of one instance of a certain motif to the total motif frequency will be additionally weighted by the relative responsiveness of the neurons composing the motif which *receive* input, as well as the baseline variance of the cells. Overall, our results can be regarded as a general estimator of mean correlation given motif statistics up to second order.

To test our theory numerically, we generated random networks of IF neurons with fixed statistical (i.e. expected) connection probability, but different second order connectivity statistics (see section 3). Simulations show that the theory provides an accurate description of network correlations (see figure 14), despite the additional error introduced by the linear response approximation of activity (equation (4)). We also compared the resumming theory to the direct evaluation of the linear response theory (equation (6)), which takes

J. Stat. Mech. (2013) P03012

account of the full graphical structure. The close match between the two (figures 8 and 12) shows that second order motifs capture much of the dependence of mean correlation on network connectivity. Moreover, figures 2, 8 and 12 demonstrate that variations in the frequency of second order motifs produce changes in network correlation that dominate those expected from the small variation in empirical (i.e. realized) connection probability from one network to the next.

Beyond the resumming theory, we also considered two other ways of simplifying the expansion of equation (6). The first was a truncation in connection strength at second order (in powers of $\tilde{A}$ and $\mathbf{W}$; sections 4.2 and 5.2). This eliminates all contributions due to motifs of length two or less. Except in very weakly connected networks, this is a poor approximation: although the contributions of higher order motifs decay exponentially in interaction strength $\tilde{A}w$, their number also grows exponentially with Np. Thus, motifs of all orders could, in principle, contribute substantially to average correlations. Our second truncation was the Erdös–Rényi approximation of mean correlation. This yielded predictions for mean correlation that included contributions from terms of all orders in expansion equation (18). These predictions therefore depend only on connection probability (see equations (34) and (47)), and should be valid when there is very little structure in the connectivity graph compared to an idealized Erdös–Rényi network.

Neuron models

Our motif-based theory can be applied to a variety of different neuron models. The starting point of our theory, the expression for pairwise cross-spectra given by equation (6), arises in a number of settings.

The main tool in our approach is linear response theory. This connects our methods to IF models (see section 2.3). Importantly, equation (6) also arises as an *exact* expression for linearly interacting point process, or *Hawkes*, models [55, 35, 56]. In this case,

$$\tilde{\mathbf{C}}(\omega) = (\mathbf{I} - \tilde{A}^*(\omega)\mathbf{W})^{-1}\mathbf{Y}(\mathbf{I} - \tilde{A}(\omega)\mathbf{W}^{\mathrm{T}})^{-1} \,,$$

where $\tilde{A}(\omega)$ is the Fourier transform of the interaction filters and $\mathbf{Y}$ is a diagonal matrix of firing rates. Therefore, our analysis can be directly and exactly applied to this setting. We note that the Hawkes process is a linear–Poisson (LP) model. Such models are commonly applied in theoretical neuroscience [57].

Relationship to other studies on motifs and network correlation

Our methodology is very similar to the previous study of [20]. They considered the impact of diverging, converging and chain motifs (along with the reciprocal connection motif) on the ability of an excitatory recurrent network to stay synchronized. They found that prevalence of the converging motif decreases synchrony, prevalence of the chain motif increases synchrony, and that the diverging motif has no significant effect on synchrony. The difference between their results and ours can be understood from the different dynamical regimes considered. In [20], the authors considered perturbations from two extreme cases: perfect synchrony, and evenly distributed asynchronous oscillators. In contrast, we studied the asynchronous regime but allowed the activity of the cells to be correlated. Hence, our methods also differ: we use a linear response approach valid for

J. Stat. Mech. (2013) P03012

strong internal noise, and weak to intermediate coupling, while [20] performed a linear stability analysis of coupled oscillator equations.

Many other studies have also examined the relationship between graphical features and spike correlations in networks. In [44], the author studied how spike-time correlation changes when one interpolates the degree distribution of a network from the binomial distribution (corresponding to Erdös–Rényi networks) to a truncated power law distribution. He found that increasing the variance of the out-degree distribution, but not the in degree, will strongly increase the cross-correlation of input currents to cells (this is a comparable quantity to ρ^{avg}—for more on these metrics see [58]). This finding is consistent with our results for the diverging and converging motifs (see equations (14) and (15)). In [35, 25] and [37], the authors studied the influence of connection structures on pairwise correlation for a number of network classes. In particular [37], used stochastic binary models of large, random networks of excitatory and inhibitory neurons, and derived an explicit expression for spike-time cross-correlation functions using a linear response methodology related to that used here. In [35], the authors used the Hawkes neuron model, which as noted above, leads to very similar expressions for correlation as those studied here and by [25]. Both approaches relate pairwise correlations to certain graphical structures of increasing order (i.e. motif size). In particular [35], obtained an expression for average network correlation in terms of the mean input and common input (equation (22) in [35]) for regular networks with uniform connection probability, while [25] (equation (25) in [25]) considered correlations in networks where only in-degree was fixed. Both are special cases of our resumming theory (note that, according to equations (14) and (15), a fixed in (out) degree is equivalent to $q_{\text{con}} = 0$ ($q_{\text{div}} = 0$) and $q_{\text{ch}} = 0$).

A major contribution of the present study is to show that effects of 'higher order' graphical interactions (i.e. motifs including more than three cells) can be approximated in terms of the frequency of second order motifs and the overall connection probability. This allows a systematic treatment of network-averaged correlation for a broader range of network connectivities.

Limitations

When applied to integrate and fire neuron networks, our analysis relies on the validity of the linear response approximation. In section 7, we demonstrated this validity for a particular firing regime, for the class of random networks studied here. For more on this issue, see [25] and [59]. We note that one avoids this issue entirely when considering Hawkes processes (see above).

We also assumed that the spectral radius of the total coupling matrix is less than one, in order to expand equation (6) into a series, for which each term can be attributed to a different motif. From this point, our methods rely on our ability to predict the impact of higher order motifs on network correlation based on the frequency of second order motifs. We demonstrated that our resumming theory can successfully make this prediction for classes of networks generated in two ways: via two-sided power law distributions, and via the SONET method (see section 3.2). However, for certain connectivity matrices, our resumming theory can produce large errors. An example pointed out to us by Chris Hoffman (personal communication) is $\mathbf{W}^0$ containing a $\sqrt{p}N \times \sqrt{p}N$ block of 1 entries and but taking value 0 everywhere else; here, p is the overall connection probability. Note that this matrix corresponds to a graph with one fully connected group and another fully

J. Stat. Mech. (2013) P03012

isolated group. Such disconnected architecture and strong inhomogeneity may be features that produce large spectral radii $\Psi(\mathbf{W\Theta})$, and hence large errors when applying our theory (see also appendix I; and for a study on the dynamics of inhomogeneous clustered networks [60]).

Three other factors limit the generality of our results. First, in order to separate the effect of network structure from that of cellular dynamics we initially assumed homogeneous firing rates and identical neurons. For many real neural networks, this may be a poor approximation. Thus, in our analysis of heterogeneous networks, we calculate weighted motif strengths, *assuming a priori* knowledge of the (heterogeneous) cellular firing rates and responsivities. A full theory would rather begin by *predicting* this heterogeneity based on network properties. Intriguingly, in [20], the authors show that certain motif frequencies can be a powerful source of such heterogeneity, which is also shown in [44] for the case of in-degree distribution (converging motif). This is a potentially important fact that we neglect.

Second, for simplicity we only considered long-time-window spike count correlation. However, our analysis can be easily generalized to study correlation at any timescale, and for any ω in equation (6). A third, and final limitation of our analysis is that we only studied pairwise correlation coefficients. This may not adequately describe the network dynamics or reveal the full importance of certain motifs.

Extensions and connections with neurobiology and computation

Intriguingly, experiments have shown that motif statistics in actual neural networks deviate from the Erdös–Rényi network model, opening a possible role of multicellular motifs. For example [21], found such deviations in connectivity between excitatory cells in visual cortex. In [61], the authors also suggest that increased frequencies of certain motifs may be a common feature of neural circuits. Moreover [22], studied the different motif statistics from two experimental data of connectivity structures of laminae in cortex, showing deviations from expectations under the Erdös–Rényi model without the laminar structure (implying as a possible origin of non-trivial motif frequencies). Our study can be applied to suggest a possible consequence of these experimental findings for levels of spike-time correlation (in the latter case generalizing it to apply to multi-group networks corresponding to the different laminae).

Building on [62], [63] studied pairwise correlations in balanced excitatory–inhibitory Erdös–Rényi networks, and found that balanced neuronal networks will have average correlations that decrease rapidly to zero with network size. Our results suggest that, with increased propensity of certain second order motifs, EI networks can potentially produce significantly larger correlation values (see figure 2). However, direct comparisons with the results regarding balanced networks [63] are not immediately available. For instance, fluctuations in input currents to neurons in a balanced network in the sense of [63] are self-generated, driven by the asynchronous firing of the population. In our network, however, we essentially make an *a priori* assumption of the existence of asynchronous input via the choice of a white noise background. This may be thought of as input arriving from a larger network which contains the subnetwork we model. In this setting, we define 'balance' as simply an average cancellation across the network of mean synaptic currents [43], which is somewhat different from the concept of dynamic balance in [63]. Further studies are

J. Stat. Mech. (2013) P03012

required before we can comment on the validity of the linear response theory in the large N limit in any setting.

For future studies, we also hope to develop a theory that predicts not just the mean correlation strength across a network, but also its variance across cell pairs. This variance has been studied using theoretical and experimental approaches [63, 64], and it would be interesting to describe how it depends on connection motifs. We will also try to predict the heterogeneity in cellular dynamics caused by motif frequencies, as referred to above—incorporating, for example, the result emphasized in [20] that variability in input to different neurons depends on converging motifs (equation (15)). Finally, we note the important connections between correlations and coding efficiency [2, 6, 13, 11, 16] (see also section 1). The recent work of [12] set up a direct connection between the graphical structure of networks and their coding efficiency, using a similar linear response calculation of covariance matrix for linear–nonlinear–Poisson (LNP) neurons. The present results could be used with the approach of [12] to further link the statistics of motifs to properties of signal transmission in neural networks.

Acknowledgments

We thank Chris Hoffman and Brent Doiron for their helpful insights. This work was supported by NSF grants DMS-0817649 and DMS-1122094 and a Texas ARP/ATP award to KJ, and by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and NSF grants DMS-1056125 and DMS-0818153 to ES-B.

Appendix A. Approximating ρ^{avg} from $\langle\tilde{\mathbf{C}}^{\infty}\rangle$

The average covariance across the network, $\langle\tilde{\mathbf{C}}^{\infty}\rangle$, can be used to approximate ρ^{avg}. Here we describe two such approximations. First, when the uncoupled neurons have equal firing rates, and the perturbation from recurrent coupling is weak, the diagonal terms $\tilde{\mathbf{C}}^{\infty}_{ii}$ will be close to the unperturbed values, $\tilde{C}^0$. In this case, subtracting diagonal terms from the average, we have that

$$\rho^{\mathrm{avg}} \approx \left(\frac{\langle\tilde{\mathbf{C}}^{\infty}\rangle}{\tilde{C}^0} - \frac{1}{N}\right)\cdot\frac{N}{N-1}. \tag{A.1}$$

In a second, more accurate approximation, we assume permutation symmetry between neurons within one population. Also, in our networks, self-connections are allowed and occur with the same probability as other connections. These will lead to identical (marginal) distributions for each entry in the covariance matrix $\tilde{\mathbf{C}}^{\infty}$, excepting the diagonal entries, which are shifted by a constant of $\tilde{C}^0$ due to each neuron's own unperturbed variance (this corresponds to the term proportional to $\mathbf{I}$ if one were to expand equation (7) as a series—see equation (18)). This suggests that $\tilde{\mathbf{C}}^{\infty}$ has the form

$$\tilde{\mathbf{C}}^{\infty} \approx \tilde{C}_0\mathbf{I} + c\mathbf{1}_{NN},$$

J. Stat. Mech. (2013) P03012

where $\mathbf{1}_{NN}$ is the $N \times N$ matrix of all ones, and c is a constant. If this holds, then, using diagonal entries of this matrix as normalization, we obtain

$$\rho^{\text{avg}} \approx \frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle - \tilde{C}_0/N}{\tilde{C}_0 + \langle \tilde{\mathbf{C}}^{\infty} \rangle - \tilde{C}_0/N}. \tag{A.2}$$

The two approximations given in equations (A.1) and (A.2) are approximately equal for small correlations. We will use equation (A.1) to exhibit the linear dependence between mean correlation coefficient and motif frequency (such as linear weights in figure 9(B)) and equation (A.2) for quantitative predications (all numerical plots).

For networks consisting of an excitatory and inhibitory population, we have analogs of equations (A.1) and (A.2) for each of the four population blocks of the covariance matrix,

$$\begin{aligned} \rho^{\text{avg}}_{XX} &\approx \left(\frac{\langle \tilde{\mathbf{C}}^{\infty}_{XX} \rangle}{\tilde{C}^0} - \frac{1}{N_X} \right) \cdot \frac{N_X}{N_X - 1}, \\ \rho^{\text{avg}}_{XY} &\approx \langle \tilde{\mathbf{C}}_{XY}(0) \rangle / \tilde{C}(0), \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} \rho^{\text{avg}}_{XX} &\approx \frac{\langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X}{\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X}, \\ \rho^{\text{avg}}_{XY} &\approx \frac{\langle \tilde{\mathbf{C}}_{XY} \rangle}{\sqrt{(\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{XX} \rangle - \tilde{C}_0/N_X)(\tilde{C}_0 + \langle \tilde{\mathbf{C}}_{YY} \rangle - \tilde{C}_0/N_Y)}}, \end{aligned} \tag{A.4}$$

where $X \neq Y \in \{E, I\}$.

Appendix B. Proof of bound on $q_{\text{div}}, q_{\text{con}}, q_{\text{ch}}$ for one population

We here prove the inequalities in equation (17). Noting that we may write

$$q_{\text{div}} = \mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \mathbf{W}^0_{j,k} \right] - \mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \right] \mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{j,k} \right],$$

(with similar expressions holding for $q_{\text{con}}, q_{\text{ch}}$) it is sufficient to show $\mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \mathbf{W}^0_{j,k} \right] \leq p$. Note

$$\mathbf{H}\mathbf{W}^0(\mathbf{W}^0)^{\text{T}}\mathbf{H} = (N^2 \mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \mathbf{W}^0_{j,k} \right])\mathbf{H}, \tag{B.1}$$

where $\mathbf{H}$ is defined in equation (19). Let $\| \cdot \|_{\text{F}}$ be the Frobenius norm, which is sub-multiplicative. We then have

$$\|\mathbf{H}\mathbf{W}^0(\mathbf{W}^0)^{\text{T}}\mathbf{H}\|_{\text{F}} \leq \|\mathbf{H}\mathbf{W}^0\|_{\text{F}} \|(\mathbf{W}^0)^{\text{T}}\mathbf{H}\|_{\text{F}} \leq \|\mathbf{H}\|_{\text{F}} \|\mathbf{W}^0\|_{\text{F}} \|(\mathbf{W}^0)^{\text{T}}\|_{\text{F}} \|\mathbf{H}\|_{\text{F}}. \tag{B.2}$$

Since $\|\mathbf{H}\|_{\text{F}} = 1$, $\|\mathbf{W}^0\|_{\text{F}} = \sqrt{\sum_{i,j} (W^0_{i,j})^2} = N\sqrt{p}$, the above inequality, together with equation (B.1), gives $\mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \mathbf{W}^0_{j,k} \right] \leq p$. With the fact that $\mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{i,k} \right] \mathbf{E}_{\text{e}} \left[\mathbf{W}^0_{j,k} \right] = p^2$, we have the bound in equation (17).

In the second inequality of equation (B.2), we have used $\|\mathbf{H}\mathbf{W}^0\|_{\text{F}} \leq \|\mathbf{H}\|_{\text{F}} \|\mathbf{W}^0\|_{\text{F}}$. The necessary and sufficient condition for achieving equality is the equality condition in

J. Stat. Mech. (2013) P03012

Table C.1. Range of parameters of the truncated power law degree distribution.

	ρ	γ_1	γ_2	L_1/L_2	L_2/N
Min.	-1	0.25	-2.25	0	0.7
Max.	1	2.25	-0.25	1	1

the following Cauchy inequalities for all i, j:

$$\left(\sum_k H_{i,k} W^0_{k,j}\right)^2 \leq \left(\sum_k H^2_{i,k}\right)\left(\sum_k (W^0_{kj})^2\right).$$

Equality holds exactly when $W^0_{k,j} = W^0_{l,j}$ for all k, l—that is, each column of $\mathbf{W}^0$ has same values (either all ones or all zeros). The equality condition for the first inequality in equation (B.2) is identical. To generate an example graph that achieves equality for a specified overall connection p, we simply set a fraction p of the columns to be all ones, and set the remaining columns to zero.

Similarly, for converging motifs we can show $\mathbf{E}_{\mathrm{e}}\left[\mathbf{W}^0_{k,i}\mathbf{W}^0_{k,j}\right] \leq p$ with a similar equality condition for rows of $\mathbf{W}^0$, and for chain motifs $\mathbf{E}_{\mathrm{e}}\left[\mathbf{W}^0_{i,k}\mathbf{W}^0_{k,j}\right] \leq p$. However, the equalities for $q_{\mathrm{div}}, q_{\mathrm{con}}$ cannot hold simultaneously, and the equality for q_{ch} cannot be achieved.

Appendix C. Graph generation methods

Here we present more details on how we generated network samples with fixed connection probability, but different frequencies of second order motifs. We used two methods.

First, the degree distribution method consists of initially generating a sample of in and out degrees from a truncated power law distribution with density, following [20],

$$f(d) = \begin{cases} C_1 d^{\gamma_1} & 0 \leq d \leq L_1 \\ C_2 d^{\gamma_2} & L_1 \leq d \leq L_2 \\ 0 & \text{otherwise,} \end{cases} \tag{C.1}$$

where d is the in or out degree (see also the configuration model [44, 46, 47]). The two marginal distributions of in and out degree are then coupled using a Gaussian copula with correlation coefficient ρ to generate the in- and out-degree lists. The parameters ρ, $L_1/L_2, L_2$, $\gamma_1 > 0, \gamma_2 < 0$ are independently and uniformly sampled for each network, separately for in and out degrees (their ranges are listed in table C.1). C_1 and C_2 are chosen so that $f(d)$ is continuous at L_1 and the mean of the degree distribution is normalized to Np (the same value for both in and out degrees), where p the fixed connection probability across network samples and N is the network size.

We then use the degree lists to calculate a probability for each possible connection from cell j to cell i, which is proportional to $d_i^{\mathrm{in}} d_j^{\mathrm{out}}$. All these N^2 probabilities are scaled so that the resulting average for the total number of connections is the same as the quantity $N^2 p_{\mathrm{stat}}$. Here p_{stat} is the target connection probability that we aim to achieve in these samples; we recall that the 'empirical' connection probability achieved in a given graph is denoted as p.

J. Stat. Mech. (2013) P03012

Table C.2. Range of motif frequencies in network samples. 'E only' (or 'EI') means excitatory only networks (or excitatory–inhibitory networks). 'SONET' and 'degree' refer to the two methods of generating networks.

		SONET: E only		Degree: E only		Degree: EI	
	Experiment	Min.	Max.	Min.	Max.	Min.	Max.
Diverging	0.033	0.001	0.913	0.015	0.181	0.018	0.295
Converging	0.044	0.001	0.838	0.016	0.189	0.022	0.279
Chain	0.022	−0.192	0.248	−0.068	0.095	−0.082	0.110

For the excitatory–inhibitory case, we generate four degree lists $d_{\mathrm{E}}^{\mathrm{in}}, d_{\mathrm{E}}^{\mathrm{out}}, d_{\mathrm{I}}^{\mathrm{in}}, d_{\mathrm{I}}^{\mathrm{out}}$, again according to the marginal distributions equation (C.1). We therefore need a four dimensional Gaussian copula. Again, parameters for the power law distribution and the correlation coefficient matrix of the copula are randomly chosen. Using these degree lists, we can then generate each of the four blocks (defined by cell type) of the adjacency matrix in the same way as in the single population case. This method allows us to sample from the whole extent of motif parameters (three in the single population case, 20 in the two population case).

For the single population networks studied in section 4, we generated additional network samples via the SONET method [20]. The idea of the algorithm is similar to maximum entropy models. Given only the connection probability and second order motif frequency, we try to generate the most 'random' network satisfying these moment constraints. However, instead of using the Gibbs distribution for the connections as in the maximum entropy method, we use a dichotomized (N^2 dimensional) Gaussian distribution. We then generate samples of excitatory only networks from the complete range of possible motif frequencies, as described in equations (16)–(17), using the SONET algorithm.

Network samples generated using both methods cover the range of motif frequency observed experimentally in cortical circuits [21, 20] as shown by table C.2. Here, we list motif frequency values as $(q_{\mathrm{div}}, q_{\mathrm{con}}, q_{\mathrm{ch}})/(p(1-p))$ for excitatory only networks, and $(q_{\mathrm{div}}^{EE,E}, q_{\mathrm{con}}^{EE,E}, q_{\mathrm{ch}}^{EEE})/(p_{EE}(1-p_{EE}))$ in excitatory–inhibitory networks (since [21] only recorded from excitatory neurons, see the definition at equation (38)).

Appendix D. Compensating for fluctuations in empirical connection probability

In order to isolate the impact of higher order motifs, all network samples should have the same empirical connection probability p, i.e. the same number of connections, as we now explain. The algorithms used to generate networks produced samples with slightly different empirical connection probabilities. These fluctuations impact the linear relationship between motif frequencies and average correlation, as p factors into the regression coefficients (see equation (33)). When attempting to determine the regression coefficients from data, fluctuations in p affect the linear trend.

To address this issue, we can scale motif frequencies by weighting them so that the contribution of a motif to mean correlation does not have a regression coefficient depending

J. Stat. Mech. (2013) P03012

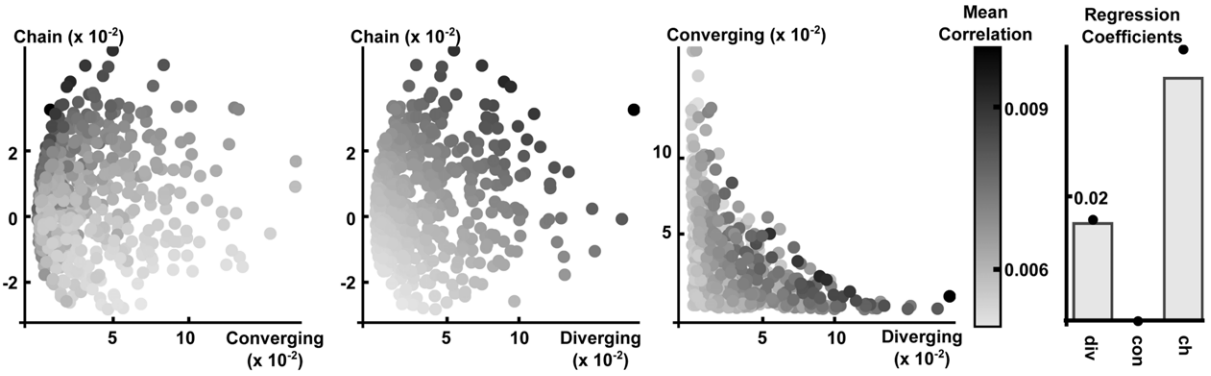

Figure D.1. Reproduction of figure 4 where we have scaled motifs to account for fluctuations in p, as described in the text.

on p. For example, suppose that the theory predicts a linear relation of the form

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^{0}} = f_0(p) + f_{ch}(p) q_{\mathrm{ch}} + f_{div}(p) q_{\mathrm{div}}.$$

Then we may define auxiliary motif frequencies

$$q'_x = \frac{f_x(p)}{f_x(p_{\mathrm{stat}})} q_x,$$

and also replace $f_0(p)$ with $f_0(p_{\mathrm{stat}})$ so that the theoretically predicted regression relationship becomes

$$\frac{\langle \tilde{\mathbf{C}}^{\infty} \rangle}{\tilde{C}^{0}} = f_0(p_{\mathrm{stat}}) + f_{ch}(p_{\mathrm{stat}}) q'_{\mathrm{ch}} + f_{div}(p_{\mathrm{stat}}) q'_{\mathrm{div}}.$$

Here, we have adjusted the definition of a motif frequency in order to account for finite-size fluctuations in the empirical connection probability p. The linear regression fit to the quantities $q'_{\mathrm{ch}}, q'_{\mathrm{div}}$ is much improved, achieving an R^2 measure of 0.99, up from 0.8 in the case where we did not account for such fluctuations. In figure D.1, we show the scatter plots exploring the relationship between motifs and mean correlation after performing this scaling, and observe the same key trends of mean correlation as in the unadjusted figure 4: strong and positive dependence on chain and diverging motifs, and minimal dependence on converging motif. These trends are now presented even more clearly.

Appendix E. Proof of proposition 4.1

We will make use of the following lemma, which may be verified by direct computation.

Lemma E.1. *Let* $\{x_n\}_{n\geq 1}, \{y_m\}_{m\geq 1}, \{z_{nm}\}_{n,m\geq 1}$ *be sequences which converge absolutely when summed, and also satisfy*

$$\left| \sum_{n=1}^{\infty} x_n \right| < 1, \qquad \left| \sum_{m=1}^{\infty} y_m \right| < 1.$$

Then,

$$\sum_{i=1}^{\infty}\sum_{(n_1,\ldots,n_k)\in\{i\}}\left(\prod_{s=1}^{k}x_{n_s}\right)=\sum_{i=1}^{\infty}\left(\sum_{n=1}^{\infty}x_n\right)^i,$$

$$\sum_{i,j=1}^{\infty}\sum_{\substack{(n_1,\ldots,n_{k+1})\in\{i\}\\(m_1,\ldots,m_{l+1})\in\{j\}}}\left[\left(\prod_{s=1}^{k}x_{n_s}\right)z_{n_{k+1}m_{l+1}}\left(\prod_{t=1}^{l}y_{m_t}\right)\right]=\left[\sum_{i=0}^{\infty}\left(\sum_{n=1}^{\infty}x_n\right)^i\right]\times\left(\sum_{n,m=1}^{\infty}z_{nm}\right)\left[\sum_{j=0}^{\infty}\left(\sum_{m=1}^{\infty}y_m\right)^j\right].$$

where the sum over $(n_1,\ldots,n_k)\in\{i\}$ *denotes a sum over all ordered partitions of* i, $(n_1,\ldots,n_k)$, *of length* $1\le k\le i$, *with each* $n_l\ge 1$.

A general result is that, for any matrix $\mathbf{A}$, there exists a sub-multiplicative matrix norm ($\|\mathbf{XY}\|\le\|\mathbf{X}\|\|\mathbf{Y}\|$) arbitrarily approaching the spectral radius $\Psi(\mathbf{A})$, that is $\|\mathbf{A}\|<\Psi(\mathbf{A})+\epsilon$ [48]. Since $\Psi(\mathbf{K\Theta})<1$, we can choose a sub-multiplicative matrix norm $\|\cdot\|_\lambda$ satisfying $\|\mathbf{K\Theta}\|_\lambda<1$. As all matrix norms are equivalent, there also exists a constant c satisfying $\|\cdot\|_2\le c\|\cdot\|_\lambda$ [48]. For $n\ge 1$, we have

$$\begin{aligned}\left|\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right|&=\left|\mathbf{u}^{\mathrm{T}}\left(\mathbf{K\Theta}\right)^{n-1}\mathbf{Ku}\right|\\&\le\|\mathbf{u}\|_2\cdot\|\left(\mathbf{K\Theta}\right)^{n-1}\|_2\cdot\|\mathbf{K}\|_2\|\mathbf{u}\|_2\\&\le c\|\left(\mathbf{K\Theta}\right)^{n-1}\|_\lambda\|\mathbf{K}\|_2\\&\le c\|\mathbf{K}\|_2\|\mathbf{K\Theta}\|_\lambda^{n-1}.\end{aligned}$$

Thus $\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}$ converges absolutely, as it is bounded above in absolute value by a convergent geometric series. Since $\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}=\mathbf{u}^{\mathrm{T}}\mathbf{K}_n^{\mathrm{T}}\mathbf{u}$, this also implies the convergence of $\sum_{m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}$. Absolute convergence of $\sum_{n,m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}$ is proved similarly by noting that, for $n,m\ge 1$,

$$\begin{aligned}|\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}|&=|\mathbf{u}^{\mathrm{T}}(\mathbf{K\Theta})^n\mathbf{K}^{\mathrm{T}}(\mathbf{\Theta}\mathbf{K}^{\mathrm{T}})^{m-1}\mathbf{u}|\\&\le\|\mathbf{u}\|_2\|(\mathbf{K\Theta})^n\|_2\|\mathbf{K}^{\mathrm{T}}\|_2\|(\mathbf{K\Theta})^{(m-1)\mathrm{T}}\|_2\|\mathbf{u}\|_2\\&\le c^2\|\mathbf{K}\|_2\|(\mathbf{K\Theta})^n\|_\lambda\|(\mathbf{K\Theta})^{m-1}\|_\lambda\\&\le c^2\|\mathbf{K}\|_2\|\mathbf{K\Theta}\|_\lambda^{n+m-1}.\end{aligned}$$

Now, suppose that $|\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}|<1$. We will remove this assumption shortly. We can now expand the right-hand side of equation (30) as

$$\left[\sum_{i=0}^{\infty}\left(\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right)^i\right]\left(1+\sum_{i,j\ge1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_i\mathbf{\Theta}\mathbf{K}_j^{\mathrm{T}}\mathbf{u}\right)\left[\sum_{j=0}^{\infty}\left(\sum_{m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}\right)^j\right]. \tag{E.1}$$

J. Stat. Mech. (2013) P03012

Using lemma E.1, we have that

$$
\begin{aligned}
(61) = &\left[\sum_{i=0}^{\infty}\left(\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right)^i\right]\left[\sum_{j=0}^{\infty}\left(\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n^{\mathrm{T}}\mathbf{u}\right)^j\right] \\
&+\left[\sum_{i=0}^{\infty}\left(\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right)^i\right]\left(\sum_{i,j\geq 1}\mathbf{u}^{\mathrm{T}}\mathbf{K}_i\boldsymbol{\Theta}\mathbf{K}_j^{\mathrm{T}}\mathbf{u}\right)\left[\sum_{j=0}^{\infty}\left(\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n^{\mathrm{T}}\mathbf{u}\right)^j\right] \\
= &\left[1+\sum_{i=1}^{\infty}\sum_{(n_1,\ldots,n_k)\in\{i\}}\left(\prod_{s=1}^{k}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_s}\mathbf{u}\right)\right] \\
&\times\left[1+\sum_{j=1}^{\infty}\sum_{(n_1,\ldots,n_l)\in\{j\}}\left(\prod_{t=1}^{l}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_t}^{\mathrm{T}}\mathbf{u}\right)\right] \\
&+\sum_{i,j=1}^{\infty}\sum_{\substack{(n_1,\ldots,n_{k+1})\in\{i\}\\(m_1,\ldots,m_{l+1})\in\{j\}}}\left(\prod_{s=1}^{k}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_s}\mathbf{u}\right)\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_{k+1}}\boldsymbol{\Theta}\mathbf{K}_{m_{l+1}}^{\mathrm{T}}\mathbf{u}\left(\prod_{t=1}^{l}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{m_t}^{\mathrm{T}}\mathbf{u}\right).
\end{aligned}
\tag{E.2}
$$

Next, note that the product term $\prod_{s=1}^{k}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_s}\mathbf{u}$, where $(n_1,\ldots,n_k)\in\{i\}$, is acquired by distributing across sums $\mathbf{H}+\boldsymbol{\Theta}$ in $\mathbf{u}^{\mathrm{T}}[\mathbf{K}(\mathbf{H}+\boldsymbol{\Theta})]^{i-1}\mathbf{K}\mathbf{u}$ and taking the unique term in this where factors of $\mathbf{H}=\mathbf{u}\mathbf{u}^{\mathrm{T}}$ divide the i factors of $\mathbf{K}$ into k blocks $\mathbf{K}_{n_s}$ of size n_s joined by factors of $\boldsymbol{\Theta}$. Therefore, summing over all possible ordered partitions, we have

$$
\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\mathbf{u}=\mathbf{u}^{\mathrm{T}}(\mathbf{K}(\mathbf{H}+\boldsymbol{\Theta}))^{i-1}\mathbf{K}\mathbf{u}=\sum_{(n_1,\ldots,n_k)\in\{i\}}\left(\prod_{s=1}^{k}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_s}\mathbf{u}\right).
\tag{E.3}
$$

Similarly,

$$
\begin{aligned}
\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\boldsymbol{\Theta}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u}&=\mathbf{u}^{\mathrm{T}}(\mathbf{K}(\mathbf{H}+\boldsymbol{\Theta}))^{i-1}\mathbf{K}\boldsymbol{\Theta}\mathbf{K}^{\mathrm{T}}((\mathbf{H}+\boldsymbol{\Theta})\mathbf{K}^{\mathrm{T}})^{j-1}\mathbf{u} \\
&=\sum_{\substack{(n_1,\ldots,n_{k+1})\in\{i\}\\(m_1,\ldots,m_{l+1})\in\{j\}}}\left(\prod_{s=1}^{k}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_s}\mathbf{u}\right)\mathbf{u}^{\mathrm{T}}\mathbf{K}_{n_{k+1}}\boldsymbol{\Theta}\mathbf{K}_{m_{l+1}}^{\mathrm{T}}\mathbf{u}\left(\prod_{t=1}^{l}\mathbf{u}^{\mathrm{T}}\mathbf{K}_{m_t}^{\mathrm{T}}\mathbf{u}\right).
\end{aligned}
\tag{E.4}
$$

Using equations (E.2)–(E.4), we have that

$$
\begin{aligned}
(61)&=\left(\sum_{i=0}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\mathbf{u}\right)\left(\sum_{j=0}^{\infty}\mathbf{u}^{\mathrm{T}}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u}\right)+\sum_{i,j\geq 1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\boldsymbol{\Theta}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u} \\
&=\sum_{i,j\geq 0}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\mathbf{H}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u}+\sum_{i,j\geq 0}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\boldsymbol{\Theta}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u} \\
&=\sum_{i,j\geq 0}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}^i(\mathbf{K}^{\mathrm{T}})^j\mathbf{u}\qquad\text{using }\mathbf{I}=\mathbf{H}+\boldsymbol{\Theta} \\
&=\mathbf{u}^{\mathrm{T}}(\mathbf{I}-\mathbf{K})^{-1}(\mathbf{I}-\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{u},\qquad\text{using }\Psi(\mathbf{K})<1,
\end{aligned}
$$

where we have used $\mathbf{u}^{\mathrm{T}}\boldsymbol{\Theta}(\mathbf{K}^{\mathrm{T}})^j\mathbf{u}=\mathbf{u}^{\mathrm{T}}\mathbf{K}^i\boldsymbol{\Theta}\mathbf{u}=0$ on the second line of the equation.

J. Stat. Mech. (2013) P03012

Lastly, we will eliminate the assumption $|\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}|<1$, which was used to establish equation (30). To see that this can be done, let z be a complex number, and replace $\mathbf{K}$ by $z\mathbf{K}$ in equation (30), giving

$$\mathbf{u}^{\mathrm{T}}(\mathbf{I}-z\mathbf{K})^{-1}(\mathbf{I}-z\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{u}=\left(1-\sum_{n=1}^{\infty}z^n\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right)^{-1}\times\left(1+\sum_{n,m=1}^{\infty}z^{n+m}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\boldsymbol{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}\right)\left(1-\sum_{m=1}^{\infty}z^m\mathbf{u}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}\right)^{-1}. \tag{E.5}$$

For sufficiently small $0<\delta<1$, $|z|<\delta$ we have that

$$\left|\sum_{n=1}^{\infty}z^n\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right|\leq\delta\sum_{n=1}^{\infty}\left|\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}\right|<1.$$

The absolute convergence of the series $\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}$, $\sum_{n=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\mathbf{u}$ and $\sum_{n,m=1}^{\infty}\mathbf{u}^{\mathrm{T}}\mathbf{K}_n\boldsymbol{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{u}$ then guarantees that equation (E.5) holds on $|z|<\delta$, and furthermore that the right-hand side of equation (E.5) is an analytic function of z on $|z|<\delta$.

Finally, note that the left-hand side of equation (E.5) is an analytic function of z on $|z|<1/\Psi(\mathbf{K})$, since the matrix inverses may be expanded as power series on this range. Since the two sides are equal and analytic on $|z|<\delta$, the right-hand side must also be defined and analytic on $|z|<1/\Psi(\mathbf{K})$, and equal the left-hand side on this range. Since $1/\Psi(\mathbf{K})>1$, we may take $z=1$ in equation (E.5), recovering equation (30).

Appendix F. Proof of proposition 5.1

The proof of proposition 5.1 is identical to the one dimensional case proposition 4.1, except that we need an entry-wise argument to eliminate the assumption $\|\sum_{n=1}^{\infty}\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\|_2<1$ that enables the expansion of the right-hand side of equation (43). Let z be a complex number, and replace $\mathbf{K}$ by $z\mathbf{K}$ in equation (43), giving

$$\mathbf{U}^{\mathrm{T}}(\mathbf{I}-z\mathbf{K})^{-1}(\mathbf{I}-z\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{U}=\left(\mathbf{I}-\sum_{n=1}^{\infty}z^n\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\right)^{-1}\times\left(\mathbf{I}+\sum_{n,m=1}^{\infty}z^{n+m}\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\boldsymbol{\Theta}\mathbf{K}_m^{\mathrm{T}}\mathbf{U}\right)\left(\mathbf{I}-\sum_{m=1}^{\infty}z^m\mathbf{U}^{\mathrm{T}}\mathbf{K}_m^{\mathrm{T}}\mathbf{U}\right)^{-1}. \tag{F.1}$$

Note that the series $\sum_{n=1}^{\infty}\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}$ absolutely converges in the 2-norm. This follows from writing

$$\|\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\|_2\leq\|\mathbf{U}^{\mathrm{T}}\|_2\|(\mathbf{K}\boldsymbol{\Theta})^{n-1}\|_2\|\mathbf{K}\|_2\|\mathbf{U}\|_2\leq c\|\mathbf{U}^{\mathrm{T}}\|_2\|\mathbf{K}\boldsymbol{\Theta}\|_{\lambda}^{n-1}\|\mathbf{K}\|_2\|\mathbf{U}\|_2,$$

and using the assumption that $\Psi(\mathbf{K}\boldsymbol{\Theta})<1$ and one associated sub-multiplicative norm $\|\mathbf{K}\boldsymbol{\Theta}\|_{\lambda}<1$. c is a constant such that $\|\cdot\|_2\leq c\|\cdot\|_{\lambda}$. For sufficiently small $0<\delta<1$, $\forall|z|<\delta$, we have that

$$\left\|\sum_{n=1}^{\infty}z^n\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\right\|_2\leq\delta\sum_{n=1}^{\infty}\|\mathbf{U}^{\mathrm{T}}\mathbf{K}_n\mathbf{U}\|_2<1.$$

J. Stat. Mech. (2013) P03012

With this condition, we can establish identity equation (F.1) in the same way as in the one dimensional case proposition 4.1 for $|z| < \delta$. Furthermore, it is similar to the above to show that all of the matrix series on the right-hand side of equation (F.1) are absolutely convergent in the 2-norm and therefore entry-wise absolutely convergent. Hence, we have that each entry of the right-hand side of equation (F.1) is an analytic function of z on $|z| < \delta$.

On the other hand, the left-hand side of equation (F.1) is entry-wise analytic in z (each entry is indeed a rational function of z) on $|z| < 1/\Psi(\mathbf{K})$. Since the two sides are equal and analytic on $|z| < \delta$, the right-hand side must also be defined and analytic on $|z| < 1/\Psi(\mathbf{K})$. Note that $1/\Psi(\mathbf{K}) > 1$, so we may set $z = 1$, yielding equation (43).

Appendix G. Combined measure of multiple sub-type motifs in figure 9(B)

In the 2×2 matrix defined in equation (46), the upper left entry gives a linear relationship between the values Q_{div}, Q_{ch} and $\langle \tilde{\mathbf{C}}_{EE} \rangle / \tilde{C}^0$. (Some explicit calculations are given in equation (H.1).) Recall that the entries of Q_{div} and Q_{ch} are again composed of sub-type motif counts as defined in equations (38)–(40). This leads to

$$\langle \tilde{\mathbf{C}}_{EE} \rangle / \tilde{C}^0 \propto \sum_{XY,Z \in \{E,I\}} k_{\text{div}}^{XY,Z} q_{\text{div}}^{XY,Z} + \sum_{XYZ \in \{E,I\}} k_{\text{ch}}^{XYZ} q_{\text{ch}}^{XYZ}, \tag{G.1}$$

where $k_{\text{div}}^{XY,Z}$, k_{ch}^{XYZ} are constants determined by equation (46). If, for example, we define a combined measure for all six diverging motifs in figure 3 as a weighted sum of the motif frequencies proportional to $k_{\text{div}}^{XY,Z}$, this should reveal a consistent trend with the average EE correlation. It is this combined measure that defines the diverging axis in figure 9(B). We similarly design the chain axis using k_{ch}^{XYZ}. For the converging motifs, the weightings between sub-type motifs are arbitrary, since there are no terms for converging motifs in equation (46). Thus, we calculate $k_{\text{con}}^{XY,Z}$ by introducing a hypothetical term for Q_{con}, $(\mathbf{I} - \tilde{A}\mathbf{D}_2\mathbf{M}^{\text{T}})^{-1}\tilde{A}^2\mathbf{Q}_{\text{con}}(\mathbf{I} - \tilde{A}\mathbf{M}\mathbf{D}_2)^{-1}$, as if it were present in equation (46).

Appendix H. Expression of the linear dependence between $\langle \tilde{\mathbf{C}}_{EE} \rangle$ and motifs

For excitatory–inhibitory networks, the (linearized) relationship between motif frequencies and averaged correlations is given by equation (46). Here, we evaluate the terms in this expression in a special case, to yield an explicit expression for the linear dependence of correlation on motif frequencies. We will express the block-wise average covariances $\langle \tilde{\mathbf{C}}_{EE} \rangle$ in terms of the 20 individual second order motifs. Along with the (approximate) definition of correlation in equation (A.1), this gives the linear weights. We note that this is how we obtain the specific linear weights used in figure 9(B).

For simplicity, assume the special case that all connection probabilities are identical ($p_{XY} = p$ for all $X, Y \in \{E, I\}$). Define the weight of all excitatory (inhibitory) connections into a cell as

$$\mu_{\text{E}} = pN_{\text{E}}w_{\text{E}} \qquad (\text{resp. } \mu_{\text{I}} = pN_{\text{I}}w_{\text{I}})$$

and the new weight of all connections into a cell as

$$\mu = \mu_{\text{E}} + \mu_{\text{I}}.$$

J. Stat. Mech. (2013) P03012

In addition, define η to be the strength of total common input to a cell pair in an Erdös–Rényi network:

$$\eta = N_{\mathrm{E}} w_{\mathrm{E}}^2 p^2 + N_{\mathrm{I}} w_{\mathrm{I}}^2 p^2.$$

Noting that we have

$$(\mathbf{MD}_2)^2 = \mu(\mathbf{MD}_2),$$

it is simple to show that

$$(\mathbf{I} - \tilde{A}\mathbf{MD}_2)^{-1} = \mathbf{I} + \frac{\tilde{A}}{1-\tilde{A}\mu}(\mathbf{MD}_2).$$

Similarly,

$$(\mathbf{I} - \tilde{A}\mathbf{D}_2\mathbf{M})^{-1} = \mathbf{I} + \frac{\tilde{A}}{1-\tilde{A}\mu}(\mathbf{D}_2\mathbf{M}).$$

Then, from equation (46) we get the linear dependence on motifs (treat p as fixed) as

$$\begin{aligned}
\langle \tilde{\mathbf{C}}_{EE} \rangle / \tilde{C}^0 \sim{}& \tilde{A}^2 \left(\frac{1-\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \right)^2 Q_{\mathrm{div}}^{EE} + \tilde{A}^2 \frac{\tilde{A}\mu_{\mathrm{I}}(1-\tilde{A}\mu_{\mathrm{I}})}{(1-\tilde{A}\mu)^2} Q_{\mathrm{div}}^{EI} + \tilde{A}^2 \left(\frac{\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \right)^2 Q_{\mathrm{div}}^{II} \\
&+ 2\tilde{A}^2 \frac{1-\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \frac{(1-\tilde{A}\mu_{\mathrm{I}})^2 + \tilde{A}^2\mu_{\mathrm{I}}^2 (N_{\mathrm{E}}/N_{\mathrm{I}})}{(1-\tilde{A}\mu)^2} Q_{\mathrm{ch}}^{EE} \\
&+ 2\tilde{A}^2 \frac{1-\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \frac{(1-\tilde{A}\mu_{\mathrm{E}})\tilde{A}\mu_{\mathrm{I}} + (1-\tilde{A}\mu_{\mathrm{I}})\tilde{A}\mu_{\mathrm{E}}(N_{\mathrm{I}}/N_{\mathrm{E}})}{(1-\tilde{A}\mu)^2} Q_{\mathrm{ch}}^{EI} \\
&+ 2\tilde{A}^2 \frac{\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \frac{(1-\tilde{A}\mu_{\mathrm{I}})^2 + \tilde{A}^2\mu_{\mathrm{I}}^2 (N_{\mathrm{E}}/N_{\mathrm{I}})}{(1-\tilde{A}\mu)^2} Q_{\mathrm{ch}}^{IE} \\
&+ 2\tilde{A}^2 \frac{\tilde{A}\mu_{\mathrm{I}}}{1-\tilde{A}\mu} \frac{(1-\tilde{A}\mu_{\mathrm{E}})\tilde{A}\mu_{\mathrm{I}} + (1-\tilde{A}\mu_{\mathrm{I}})\tilde{A}\mu_{\mathrm{E}}(N_{\mathrm{I}}/N_{\mathrm{E}})}{(1-\tilde{A}\mu)^2} Q_{\mathrm{ch}}^{II},
\end{aligned} \tag{H.1}$$

where $Q_{\mathrm{div}}^{XY}, Q_{\mathrm{ch}}^{XY}$ were defined in equations (38) and (40) as

$$Q_{\mathrm{div}}^{XY} = N_{\mathrm{E}} w_{\mathrm{E}}^2 q_{\mathrm{div}}^{XY,E} + N_{\mathrm{I}} w_{\mathrm{I}}^2 q_{\mathrm{div}}^{XY,I}, \qquad Q_{\mathrm{ch}}^{XY} = N_{\mathrm{E}} w_X w_{\mathrm{E}} q_{\mathrm{ch}}^{XEY} + N_{\mathrm{I}} w_X w_{\mathrm{I}} q_{\mathrm{ch}}^{XIY}.$$

Since each of the quantities $Q_{\mathrm{div}}^{XY}, Q_{\mathrm{ch}}^{XY}$ is clearly linear in second order motif frequencies, equation (H.1) gives a linear relation between second order motif frequencies and block-wise averaged correlation in the two population network. Equation (H.1) is the two population analog of equation (33) in the single population case.

Appendix I. Intuition for why the resumming approach can produce accurate results

From equations (31) and (44), we see that the error of resumming theory is determined by the tail that we dropped in two series. With respect to the coupling strength order of magnitude (w or $w_{\mathrm{E}}, w_{\mathrm{I}}$), these are geometric series. Therefore the sum of the tail series is controlled by the leading term, that is

$$\tilde{A}^3 \mathbf{L}^{\mathrm{T}} \mathbf{W} \mathbf{\Theta} \mathbf{W} \mathbf{\Theta} \mathbf{W} \mathbf{L}, \qquad \tilde{A}^3 \mathbf{L}^{\mathrm{T}} \mathbf{W} \mathbf{\Theta} \mathbf{W}^{\mathrm{T}} \mathbf{\Theta} \mathbf{W} \mathbf{L}, \qquad \tilde{A}^3 \mathbf{L}^{\mathrm{T}} \mathbf{W} \mathbf{\Theta} \mathbf{W} \mathbf{\Theta} \mathbf{W}^{\mathrm{T}} \mathbf{L},$$

J. Stat. Mech. (2013) P03012

Table I.1. Notation used in the text.

Symbol	Description
$v_i, \tau_i, E_{\mathrm{L},i}$	Membrane potential, membrane time constant and leak reversal potential of cell i
E_i, σ_i	Mean and standard deviation of the background noise for cell i
$v_{\mathrm{th}}, v_{\mathrm{r}}, \tau_{\mathrm{ref}}$	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 [23]
$f_i(t), \eta_i(t)$	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
$\mathbf{W}_{ij}$	The $j \to i$ synaptic weight, proportional to the area under a single post-synaptic current for current-based synapses
$\mathbf{J}_{ij}(t)$	The $j \to i$ synaptic kernel—equals the product of the synaptic weight $\mathbf{W}_{ij}$ and the synaptic filter for outputs of cell j
$\mathbf{C}_{ij}(\tau)$	The cross-correlation function between cells i, j defined by $C_{ij} = \mathrm{cov}(y_i(t), y_j(t+\tau))$
$\tilde{\mathbf{C}} = \tilde{\mathbf{C}}(0)$	The cross-spectrum matrix evaluated at zero frequency. Unless noted otherwise, all spectral quantities are evaluated at zero frequency
$N_{y_i}(t, t+\tau), \rho_{ij}(\tau)$	Spike count for cell i, and spike count correlation coefficient for cells i, j over windows of length τ
$r_i, A_i(t), \mathbf{C}^0_{ii}$	Stationary rate, linear response kernel and uncoupled auto-correlation function for cell i
$\mathbf{K}_{ij}(t)$	The $j \to i$ interaction kernel—describes how the firing activity of cell i is perturbed by an input spike from cell j. It is defined by $\mathbf{K}_{ij}(t) = (A_i * \mathbf{J}_{ij})(t)$
$\mathbf{y}^n_i(t), \mathbf{C}^n_{ij}(t)$	The nth 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 nth 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 $\tilde{g}(\omega) = \mathcal{F}[g](\omega) \equiv \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}\omega t} g(t)\, \mathrm{d}t$
$\mathbf{E}_{\mathrm{e}}[\cdot]$	Empirical average, $\frac{1}{N^2}\sum_{i,j} \cdot$ or $\frac{1}{N^3}\sum_{i,j,k} \cdot$ depending on context
R^2	Coefficient of determination, i.e. the square of correlation coefficient ρ^2
$\mathbf{W}^0$	Adjacency matrix
$\mathbf{H}$	$\mathbf{H} = \frac{1}{N}\mathbf{1}_{NN}$
$\mathbf{\Theta}$	$\mathbf{\Theta} = \mathbf{I} - \mathbf{H}$

This shows that the resumming theory has third order accuracy in the effective interaction strength $\tilde{A}w$.

Another important factor affecting the accuracy of the resumming theory is the spectral radius $\Psi(\mathbf{\Theta W \Theta}) = \Psi(\mathbf{W\Theta}) = \Psi(\mathbf{\Theta W})$ (equalities follow from writing $\mathbf{W}$ under the basis of projections $\mathbf{\Theta}$ and $\mathbf{H}$), which is related to how fast the terms in the series converge to zero. There is a simple intuition of $\Psi(\mathbf{W\Theta})$ for Erdös–Rényi networks. For

J. Stat. Mech. (2013) P03012

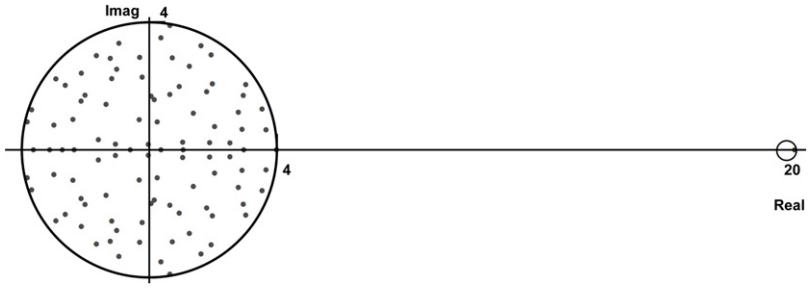

Figure I.1. The spectra of a single population Erdös–Rényi network; the larger circle and small circle on the right are expected locations of the bulk spectra and a single eigenvalue calculated from the asymptotic formula in text above and section 2.4. The excitatory only network has 100 neurons, and $p = 0.2$, $w = 1$.

single population networks, note that (asymptotically, for large N) $\mathbf{W}^0\mathbf{\Theta} = \mathbf{W}^0 - \mathbf{W}^0\mathbf{H}$ is a matrix with i.i.d. entries of mean zero. According to the circular law, all eigenvalues will be asymptotically uniformly distributed within a circle about the origin. Comparing with the spectra of $\mathbf{W}^0$ in section 2.4, multiplying by $\mathbf{\Theta}$ effectively removes the single dominant eigenvalue [43].

The removal of this dominant eigenvalue will reduce the spectral radius of $\mathbf{W}^0\mathbf{\Theta}$ as compared to $\mathbf{W}^0$ by a factor of $\sqrt{N}$ ($w\sqrt{p(1-p)N}$ compared to pNw, see section 2.4). Such reductions of $\Psi(\mathbf{W}\mathbf{\Theta})$ approximately occur in single population networks and at the blocks of $\mathbf{W}\mathbf{\Theta}$ in two population networks, even though those networks are non-Erdös–Rényi. This intuition may help understand why resumming theory works much better than truncation theory: the tails of series in $\mathbf{W}\mathbf{\Theta}$ may be much lighter than those in $\mathbf{W}$.

References

[1] Sejnowski T J, *On the stochastic dynamics of neuronal interaction*, 1976 *Biol. Cybern.* **22** 203

[2] Zohary E, Shadlen M N and Newsome W T, *Correlated neuronal discharge rate and its implication for psychophysical performance*, 1994 *Nature* **370** 140

[3] Singer W and Gray C M, *Visual feature integration and the temporal correlation hypothesis*, 1995 *Annu. Rev. Neurosci.* **18** 555

[4] Shadlen M N and Newsome W T, *The variable discharge of cortical neurons: implications for connectivity, computation, and information coding*, 1998 *J. Neurosci.* **18** 3870

[5] Panzeri S, Schultz S, Treves A and Rolls E, *Correlations and encoding of information in the nervous system*, 1999 *Proc. R. Soc.* B **266** 1001

[6] Abbott L F and Dayan P, *The effect of correlated variability on the accuracy of a population code*, 1999 *Neural Comput.* **11** 91

[7] Sompolinsky H, Yoon H, Kang K and Shamir M", *Population coding in neuronal systems with correlated noise*, 2001 *Phys. Rev.* E **64** 051904

[8] Panzeri S, Petersen R S, Schultz S R, Lebedev M and Diamond M E, *The role of spike timing in the coding of stimulus location in rat somatosensory cortex*, 2001 *Neuron* **29** 769

[9] Schneidman E, Bialek W and Berry M J II, *Synergy, redundancy, and independence in population codes*, 2003 *J. Neurosci.* **23** 11539

[10] Latham P E and Nirenberg S, *Synergy, redundancy, and independence in population codes, revisited*, 2005 *J. Neurosci.* **25** 5195

[11] Josić K, Shea-Brown E, Doiron B and de la Rocha J, *Stimulus-dependent correlations and population codes*, 2009 *Neural Comput.* **21** 2774

J. Stat. Mech. (2013) P03012

[12] Beck J, Bejjanki V R and Pouget A, *Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons*, 2011 *Neural Comput.* **23** 1484

[13] Salinas E and Sejnowski T J, *Impact of correlated synaptic input on output firing rate and variability in simple neuronal models*, 2000 *J. Neurosci.* **20** 6193

[14] Fries P, *A mechanism for cognitive dynamics: neuronal communication through neuronal coherence*, 2005 *Trends Cogn. Sci.* **9** 474

[15] Bruno R M, *Synchrony in sensation*, 2011 *Curr. Opin. Neurobiol.* **21** 701

[16] Schneidman E, Berry M J, Segev R and Bialek W, *Weak pairwise correlations imply strongly correlated network states in a neural population*, 2006 *Nature* **440** 1007

[17] Shlens J, Field G D, Gauthier J L, Grivich M I, Petrusca D, Sher A, Litke A M and Chichilnisky E J, *The structure of multi-neuron firing patterns in primate retina*, 2006 *J. Neurosci.* **26** 8254

[18] Ecker A S, Berens P, Keliris G A, Bethge M, Logothetis N K and Tolias A S, *Decorrelated neuronal firing in cortical microcircuits*, 2010 *Science* **327** 584

[19] Cohen M R and Kohn A, *Measuring and interpreting neuronal correlations*, 2011 *Nature Neurosci.* **14** 811

[20] Zhao L, Beverlin B, Netoff T and Nykamp D Q, *Synchronization from second order network connectivity statistics*, 2011 *Front Comput. Neurosci.* **5** 1

[21] Song S, Sjöström P J, Reigl M, Nelson S and Chklovskii D B, *Highly nonrandom features of synaptic connectivity in local cortical circuits*, 2005 *PLoS Biol.* **3** e68

[22] Haeusler S, Schuch K and Maass W, *Motif distribution, dynamical properties, and computational performance of two data-based cortical microcircuit templates*, 2009 *J. Physiol. Paris* **103** 73

[23] Fourcaud-Trocmé N, Hansel D, van Vreeswijk C and Brunel N, *How spike generation mechanisms determine the neuronal response to fluctuating inputs*, 2003 *J. Neurosci.* **23** 11628

[24] Lindner B, Doiron B and Longtin A, *Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback*, 2005 *Phys. Rev.* E **72** 061919

[25] Trousdale J, Hu Y, Shea-Brown E and Josić K, *Impact of network structure and cellular response on spike time correlations*, 2012 *PLoS Comput. Biol.* **8** e1002408

[26] White J A, Rubinstein J T and Kay A R, *Channel noise in neurons*, 2000 *Trends Neurosci.* **23** 131

[27] Renart A, Brunel N and Wang X J, *Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks*, 2004 *Computational Neuroscience: A Comprehensive Approach* ed J Feng (Boca Raton, FL: CRC Press) p 431

[28] Gabbiani F and Cox S J, 2010 *Mathematics for Neuroscientists* (London: Academic)

[29] Stratonovich R L, 1967 *Topics in the Theory of Random Noise* vol 2 (New York: Gordon and Breach)

[30] Bair W, Zohary E and Newsome W T, *Correlated firing in macaque visual area MT: time scales and relationship to behavior*, 2001 *J. Neurosci.* **21** 1676

[31] de la Rocha J, Doiron B, Shea-Brown E, Josić K and Reyes A, *Correlation between neural spike trains increases with firing rate*, 2007 *Nature* **448** 802

[32] Shea-Brown E, Josić K, de la Rocha J and Doiron B, *Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding*, 2008 *Phys. Rev. Lett.* **100** 108102

[33] Rosenbaum R and Josić K, *Mechanisms that modulate the transfer of spiking correlations*, 2011 *Neural Comput.* **23** 1261

[34] Tetzlaff T, Rotter S, Stark E, Abeles M, Aertsen A and Diesmann M, *Dependence of neuronal correlations on filter characteristics and marginal spike train statistics*, 2008 *Neural Comput.* **20** 2133

[35] Pernice V, Staude B, Cardanobile S and Rotter S, *How structure determines correlations in neuronal networks*, 2011 *PLoS Comput. Biol.* **7** e1002059

[36] Risken H, 1996 *The Fokker–Planck Equation: Methods of Solution and Applications* (Berlin: Springer)

[37] Ginzburg I and Sompolinsky H, *Theory of correlations in stochastic neural networks*, 1994 *Phys. Rev.* E **50** 3171

[38] Lindner B and Schimansky-Geier L, *Transmission of noise coded versus additive signals through a neuronal ensemble*, 2001 *Phys. Rev. Lett.* **86** 2934

[39] Brunel N, Chance F S, Fourcaud N and Abbott L F, *Effects of synaptic noise and filtering on the frequency response of spiking neurons*, 2001 *Phys. Rev. Lett.* **86** 2186

[40] Richardson M J E, *Spike-train spectra and network response functions for non-linear integrate-and-fire neurons*, 2008 *Biol. Cybern.* **99** 381

[41] Ostojic S, Brunel N and Hakim V, *How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains*, 2009 *J. Neurosci.* **29** 10234

[42] Richardson M J E, *Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents*, 2009 *Phys. Rev.* E **80** 021928

J. Stat. Mech. (2013) P03012

[43] Rajan K and Abbott L F, *Eigenvalue spectra of random matrices for neural networks*, 2006 *Phys. Rev. Lett.* **97** 188104
[44] Roxin A, *The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons*, 2011 *Front Comput. Neurosci.* **5** 1
[45] Chung F and Lu L, *Connected components in random graphs with given expected degree sequences*, 2002 *Ann. Comb.* **6** 125
[46] Newman M E J, *The structure and function of complex networks*, 2003 *SIAM Rev.* **45** 167
[47] Newman M E J, Strogatz S H and Watts D J, *Random graphs with arbitrary degree distributions and their applications*, 2001 *Phys. Rev.* E **64** 026118
[48] Horn R A and Johnson C R, 1990 *Matrix Analysis* (Cambridge: Cambridge University Press)
[49] Rangan A V, *Diagrammatic expansion of pulse-coupled network dynamics*, 2009 *Phys. Rev. Lett.* **102** 158101
[50] Rangan A V, *Diagrammatic expansion of pulse-coupled network dynamics in terms of subnetworks*, 2009 *Phys. Rev.* E **80** 036101
[51] Doiron B, Lindner B, Longtin A, Maler L and Bastian J, *Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus*, 2004 *Phys. Rev. Lett.* **96** 048101
[52] Khuc Trong P and Rieke F, *Origin of correlated activity between parasol retinal ganglion cells*, 2008 *Nature Neurosci.* **11** 1343
[53] Marinazzo D, Kappen H J and Gielen S C A M, *Input-driven oscillations in networks with excitatory and inhibitory neurons with dynamic synapses*, 2007 *Neural Comput.* **19** 1739
[54] Hawkes A G, *Spectra of some self-exciting and mutually exciting point processes*, 1971 *Biometrika* **58** 83
[55] Hawkes A G, *Point spectra of some mutually exciting point processes*, 1971 *J. R. Stat. Soc.* B **33** 438
[56] Pernice V, Staude B, Cardanobile S and Rotter S, *Recurrent interactions in spiking networks with arbitrary topology*, 2012 *Phys. Rev.* E **85** 031916
[57] Dayan P and Abbot L F, 2001 *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* 1st edn (Cambridge, MA: MIT Press)
[58] Rosenbaum R and Josić Kšimir, *Membrane potential and spike train statistics depend distinctly on input statistics*, 2011 *Phys. Rev.* E **84** 051902
[59] Ostojic S and Brunel N, *From spiking neuron models to linear–nonlinear models*, 2011 *PLoS Comput. Biol.* **7** e1001056
[60] Watts D J and Strogatz S H, *Collective dynamics of 'small-world' networks*, 1998 *Nature* **393** 440
[61] Sporns O and Kötter R, *Motifs in brain networks*, 2004 *PloS Biol.* **2** e369
[62] van Vreeswijk C and Sompolinsky H, *Chaos in neuronal networks with balanced excitatory and inhibitory activity*, 1996 *Science* **274** 1724
[63] Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A and Harris K D, *The asynchronous state in cortical circuits*, 2010 *Science* **327** 587
[64] Ecker A S, Berens P, Tolias A S and Bethge M, *The effect of noise correlations in populations of diversely tuned neurons*, 2011 *J. Neurosci.* **31** 14272

J. Stat. Mech. (2013) P03012