Effect of Synaptic Heterogeneity on Neuronal Coordination

Moritz Layer [1,2] Moritz Helias [1,3] and David Dahmen [1,*]

[1] *Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany*

[2] *RWTH Aachen University, Aachen, Germany*

[3] *Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany*

(Received 15 August 2023; accepted 16 February 2024; published 14 March 2024)

Recent advancements in measurement techniques have resulted in an increasing amount of data on neural activities recorded in parallel, revealing largely heterogeneous correlation patterns across neurons. Yet, the mechanistic origin of this heterogeneity is largely unknown because existing theoretical approaches linking structure and dynamics in neural circuits are restricted to population-averaged connectivity and activity. Here we present a systematic inclusion of heterogeneity in network connectivity to derive quantitative predictions for neuron-resolved covariances and their statistics in spiking neural networks. Our study shows that the heterogeneity in covariances is not a result of variability in single-neuron firing statistics but stems from the ubiquitously observed sparsity and variability of connections in brain networks. Linear-response theory maps these features to the effective connectivity between neurons, which in turn determines neuronal covariances. Beyond-mean-field tools reveal that synaptic heterogeneity modulates the variability of covariances and thus the complexity of neuronal coordination across many orders of magnitude.

DOI: 10.1103/PRXLife.2.013013

I. INTRODUCTION

Neuronal networks in the brain display largely heterogeneous activity: common observables such as firing rates [1–3], coefficients of variation (CVs) [4], and pairwise correlations [5–8] are widely distributed across neurons. This has important implications for coding and information processing in the brain, as the coordinated activity across the enormous number of units in neuronal circuits is thought to underlie all complex functions [9–12]. The causes of heterogeneity in neuronal dynamics are diverse: intrinsic neuronal properties, external inputs, and the network connectivity itself are all sources of variability. While these structural and dynamic heterogeneities can be readily observed with modern experimental techniques [13–15], understanding their mechanistic relations requires theoretical tools that are currently still lacking.

In this study, we focus on the effects of connectivity and investigate the influence of heterogeneity in connections on the activity of networks of identical neurons receiving homogeneous external input. Previous work [16] has shown that a considerable fraction of the variance, in the distribution of firing rates across neurons and in the CV of individual neurons' spike trains, in such networks can already be explained by the distributed number of inputs the neurons in a network receive. In this study, we go beyond single-neuron activities and focus on the statistics of pairwise correlations and the related covariances, which measure how strongly the activities of pairs of neurons cofluctuate. Such coordination builds the basis for collective network activity and function.

With the exception of small organisms such as *Caenorhabditis elegans* [17], the microconnectome of most biological neuronal networks is unknown. However, overall connectivity properties and statistics, like the connection probabilities between different cortical areas [18,19] and cell types [15], the distance dependence of connections [15,20], or the statistics of synaptic strengths [15,21–26], are available nowadays. Hence, rather than a one-to-one relation between microconnectome and pairwise covariances [27–32], a relation between connectivity and covariance on a statistical level would readily allow the inclusion of this knowledge. To derive such a relation, common population-level theories [27,33–38] cannot be used because they can only describe population-averaged observables and, in particular, do not capture heterogeneity in covariances within populations. Here, we instead employ mean-field theory on the single-neuron level [30], which we systematically compare to network simulations, and we go beyond mean-field theory by including nontrivial fluctuation terms to obtain the statistics of covariances between individual neuron pairs.

The main difficulty of a single-neuron-level approach is that the predictions of the theory for individual neurons strongly depend on the specific details of the connectivity. To get a description on the level of connectivity statistics,

*d.dahmen@fz-juelich.de

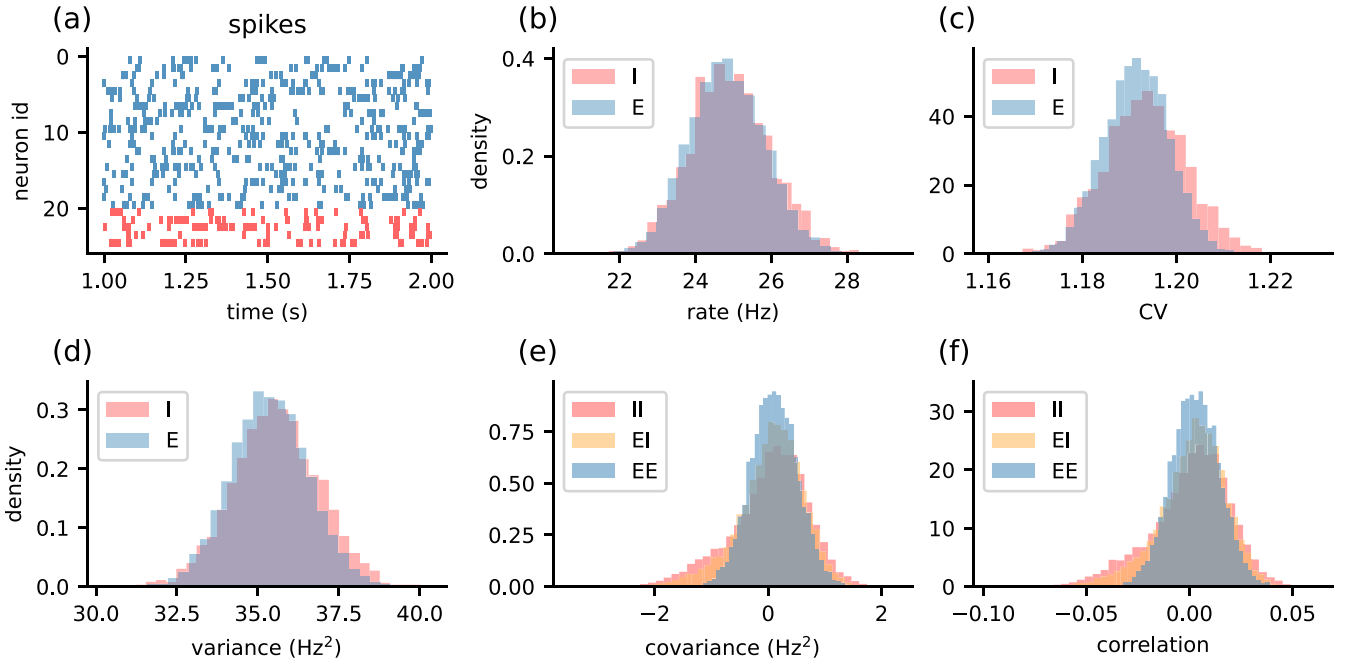

FIG. 1. Simulation of excitatory-inhibitory (E-I) network of leaky integrate-and-fire (LIF) neurons. (a) Spike trains of a sample of 20 excitatory (E, blue) and 5 inhibitory (I, red) neurons. (b) Distributions of firing rates. (c) Distributions of coefficients of variation (CVs). (d) Distributions of variances of spike counts measured in time bins of 1 s [cf. Eq. (7)]. (e) Distributions of spike-count covariances for different pairings of neurons. (f) Distributions of spike-count correlation coefficients for different pairings of neurons. For model details and simulation parameters see Appendix A for spectral radius $r = 0.49$.

we perform a disorder average, a technique originally developed for spin-glass systems [39,40] that allows retaining information about the connectivity statistics while averaging over the realization randomness. As our main results, we show how to systematically calculate higher moments of neuronal activity averaged over the disorder in the connectivity using replica and beyond-mean-field theory, and we use this technique to derive a relation between the mean and variance of covariances and the mean and variance of the network connectivity. First results based on a similar but reduced theoretical approach have already been successfully applied in the neuroscientific context to infer the dynamical regime of cortical networks [7] and to explain spatial properties of coordination structures [8] and dimensionality [41].

To summarize, we investigate the origin of neuronal coordination structures, as experimentally observed across various species and cortical areas, by analyzing covariances in a prototypical network model of cortical dynamics [42], namely, sparsely connected excitatory and inhibitory neurons that operate in the balanced state [34]. In this model, all neurons have identical parameters and receive homogeneous, uncorrelated external input. As in biological cortical networks, the sparsity in the connectivity between neurons [15] as well as the wide distribution in synaptic amplitudes [15,21–26] constitute the source of variability in connections and thereby the dynamics: Rates, CVs, variances, covariances, and hence correlation coefficients are all described by distributions with sizable variance (see Fig. 1).

The following sections investigate the sources of the variance in these quantities. Section II introduces mean-field theory on the single-neuron level. In Sec. III, we derive the main results on how to compute disorder-averaged moments of neuronal activity, and we calculate explicit expressions for the mean and variance of covariances. In Sec. IV, we discuss our findings and their limitations in the context of the existing literature.

II. BACKGROUND: LINEAR-RESPONSE THEORY OF SPIKING NEURONAL NETWORKS ON A SINGLE-NEURON LEVEL

To understand the origin of the distribution of covariances, we start with analyzing a simulated network on a single-neuron level. The example network throughout this study comprises 8000 excitatory (E) and 2000 inhibitory (I) leaky integrate-and-fire (LIF) neurons that make connections according to distinct population-specific statistics. We mimic two abundant features of heterogeneity in connectivity of brain circuits, that are sparse connections and distributed synaptic weights. To do so we consider random sparse connectivity $\boldsymbol{J}$, with connection probability 10%, giving rise to an excitatory indegree $K_\mathrm{E} = 800$ and an inhibitory indegree $K_\mathrm{I} = 200$. To compensate for the imbalance in excitatory and inhibitory neuron count, we follow the work by Brunel [42] and scale the strengths of existing inhibitory connections with respect to excitatory ones by a factor $g = -6$ to obtain an asynchronous irregular dynamic regime. In addition, we distribute synaptic weights of existing connections according to population-specific normal distributions $j_\mathrm{E} \propto \mathcal{N}(j, 0.2j)$ and $j_\mathrm{I} \propto \mathcal{N}(gj, 0.2j)$, such that the overall heterogeneity in

network connectivity is comprised of the random sparseness of connections and the variable strength of connections. For more details of the model, see Appendix A.

Working point. Given the parameters of the simulated network of leaky integrate-and-fire neurons, especially the specific realization of the connectivity matrix $\boldsymbol{J}$, we determine the stationary *working point*, comprising the input statistics $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ and the firing rates $\boldsymbol{\nu}$, as done by Brunel and Hakim [43] and Brunel [42]. To this end, we first neglect correlations between the neurons and approximate the neurons' inputs as independent Gaussian white noise processes. In this *diffusion approximation*, the mean input μ_i and input variance σ_i^2 of neuron i are given by

$$\mu_i = \tau_m \left(\sum_j J_{ij}\nu_j + j\nu_{\mathrm{ext,E}} + gj\nu_{\mathrm{ext,I}} + \frac{I_{\mathrm{ext}}}{C} \right), \tag{1}$$

$$\sigma_i^2 = \tau_m \left(\sum_j J_{ij}^2\nu_j + j^2\nu_{\mathrm{ext,E}} + g^2 j^2 \nu_{\mathrm{ext,I}} \right), \tag{2}$$

with membrane time constant τ_{m}, membrane capacitance C, constant input current I_{ext}, and excitatory and inhibitory external Poisson noise with rates $\nu_{\mathrm{ext,E}}$ and $\nu_{\mathrm{ext,I}}$ which are fed into the system via weights j and gj, respectively. The firing rates are given by the Siegert function [44]

$$\nu_i = \left\{ \tau_{\mathrm{r}} + \tau_{\mathrm{m}}\sqrt{\pi} \int_{y_{\mathrm{r},i}}^{y_{\mathrm{th},i}} \mathrm{d}s\, f(s) \right\}^{-1},$$

$$f(s) = \mathrm{e}^{s^2}[1 + \mathrm{erf}(s)], \tag{3}$$

with refractory period τ_{r}, and rescaled reset and threshold voltages

$$y_{\mathrm{r},i} = \frac{V_{\mathrm{r}} - \mu_i}{\sigma_i}, \quad y_{\mathrm{th},i} = \frac{V_{\mathrm{th}} - \mu_i}{\sigma_i}.$$

These equations can be solved iteratively in a self-consistent manner. Given the working point, we can determine the coefficients of variation using (see Appendix A.1 of Ref. [42]; note that they use different units)

$$\mathrm{CV}_i^2 = 2\pi(\tau_{\mathrm{m}}\nu_i)^2 \int_{y_{\mathrm{r},i}}^{y_{\mathrm{th},i}} \mathrm{d}x\, \mathrm{e}^{x^2} \int_{-\infty}^{x} \mathrm{d}z\, \mathrm{e}^{z^2}[1 + \mathrm{erf}(z)]^2. \tag{4}$$

Linearization. The full dynamics of LIF neurons are nonlinear. However, as covariances measure cofluctuations of neurons around their working points, we can study covariances by analyzing linearized dynamics as long as the fluctuations are sufficiently small. Grytskyy *et al.* (see Sec. 5 of Ref. [31]) show that a network of LIF neurons can be mapped to a linear rate model with output noise

$$x_i(t) = \int_{-\infty}^{t} h(t - t') \sum_j W_{ij}[x_j(t' - d) + \xi_j(t' - d)]\, \mathrm{d}t', \tag{5}$$

with neuronal activity $x_i(t)$, normalized linear-response kernel $h(t)$, synaptic delay d, and uncorrelated Gaussian white noise $\xi_i(t)$, $\langle \xi_i \rangle = 0$, $\langle \xi_i(s)\xi_j(t) \rangle = D_{ij}\delta(s - t)$, with diagonal noise strength matrix $D_{ij} = \delta_{ij}D_{ii}$. The matrix $\boldsymbol{W}$, referred to as *effective connectivity*, combines the connectivity matrix $\boldsymbol{J}$ with the sensitivity of neurons to small fluctuations in their input. It is formally given by the derivative of the stationary firing rate of neuron i [Eq. (3)] with respect to the firing rate of neuron j evaluated at the stationary working point (see Appendix A of Ref. [45])

$$W_{ij} = \frac{\partial \nu_i}{\partial \nu_j} = \alpha_i J_{ij} + \beta_i J_{ij}^2, \tag{6}$$

with

$$\alpha_i = \sqrt{\pi}(\tau_{\mathrm{m}}\nu_i)^2 \frac{1}{\sigma_i}[f(y_{\mathrm{th},i}) - f(y_{\mathrm{r},i})],$$

$$\beta_i = \sqrt{\pi}(\tau_{\mathrm{m}}\nu_i)^2 \frac{1}{2\sigma_i^2}[f(y_{\mathrm{th},i})y_{\mathrm{th},i} - f(y_{\mathrm{r},i})y_{\mathrm{r},i}].$$

Spike-count covariances. In this study we are interested in spike-count covariances in spiking networks,

$$C_{ij} = \frac{1}{T}(\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle), \tag{7}$$

with spike counts n_i occurring within bins of size T, where the average, indicated by the brackets $\langle \cdot \rangle$, is taken across all bins that can be viewed as trials with different realizations of the external input. As shown in Methods and Materials of Ref. [7], for stationary processes and large bin sizes spike-count covariances C_{ij} can be mapped to the time-lag integrated covariances $c_{ij}(\tau)$ between spike trains of neurons i and j (see also Refs. [46,47]; for more details on definitions of covariances see Appendix B):

$$C_{ij} \overset{T\to\infty}{\to} \int_{-\infty}^{\infty} c_{ij}(\tau)\mathrm{d}\tau.$$

In the following the term covariance always refers to C_{ij}. Making use of the Wiener-Khinchin theorem (Appendix C) allows expressing the time-lag integrated covariances in terms of the neuronal activities' Fourier components $X_i(\omega)$ at frequency zero,

$$C_{ij} = \langle X_i(0)X_j(0) \rangle, \tag{8}$$

which can be evaluated by Fourier transforming Eq. (5), yielding

$$\boldsymbol{C} = (\mathbf{1} - \boldsymbol{W})^{-1}\boldsymbol{D}(\mathbf{1} - \boldsymbol{W})^{-\mathrm{T}}. \tag{9}$$

For calculating the covariances, we therefore only need the effective connectivity $\boldsymbol{W}$ and the noise strength $\boldsymbol{D}$. The correlation coefficients follow as

$$\kappa_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}.$$

To estimate the noise strength $\boldsymbol{D}$, we assume that the spike trains are described sufficiently well as renewal processes for which the variances are given by [48]

$$C_{ii} = \mathrm{CV}_i^2 \nu_i. \tag{10}$$

Using that $\boldsymbol{D}$ is by definition diagonal (see Appendix D for limitations on exactly matching simulated spike-count covariances with a linear rate model with uncorrelated white noise

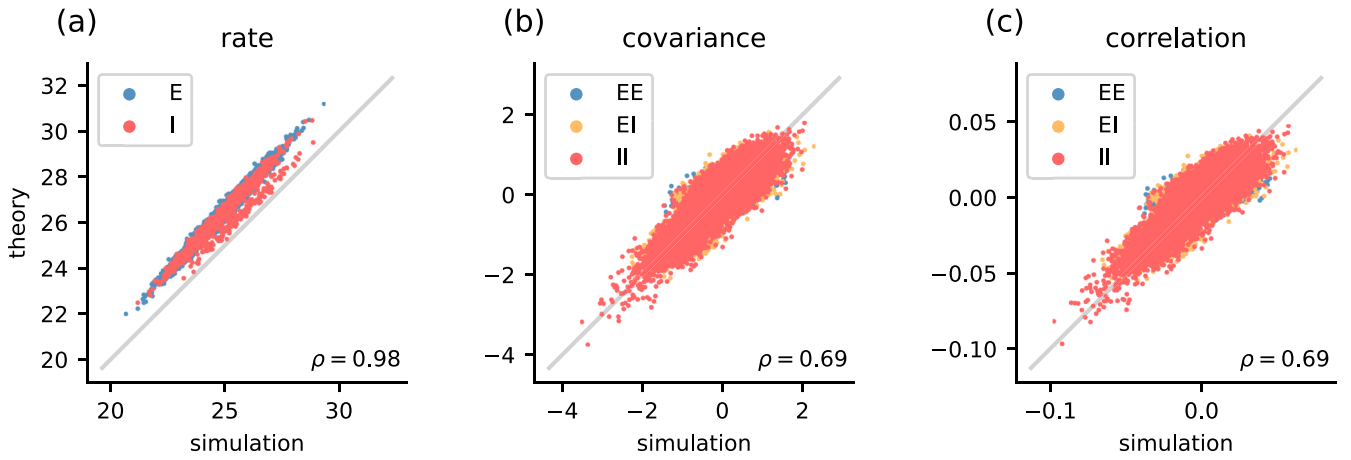

FIG. 2. Simulation results and theoretical estimates for E-I network of LIF neurons. Here ρ denotes the Pearson correlation coefficient. (a) Firing rates $\boldsymbol{\nu}$. (b) Covariances $\boldsymbol{C}$. (c) Correlation coefficients $\boldsymbol{\kappa}$. For model details and simulation parameters see Appendix A for spectral radius $r = 0.49$.

input), we can solve Eq. (9) for $\boldsymbol{D}$, which results in

$$D_{ii} = \sum_j (B^{-1})_{ij} \mathrm{CV}_j^2 \nu_j, \tag{11}$$

with

$$B_{ij} = [(\mathbf{1} - \boldsymbol{W})^{-1}]_{ij}^2. \tag{12}$$

The above expressions can be combined to compute theoretical estimates of the quantities measured in the simulation. To solve the self-consistency equations for the firing rates and to compute the covariances, we make use of the Python package NNMT [49], which includes optimized implementations of the equations introduced above. A comparison of theoretical and simulation results is shown in Fig. 2. For the chosen parameters, simulation and theory correlate strongly, and the theory appears to capture the primary sources of heterogeneity in the rates, covariances, and correlation coefficients. Note that such a good match between theory and simulation cannot be observed in all parameter regimes of the spiking network; the validity of the assumptions made and the resulting theoretical estimates depend on the network state (see Appendix D for further discussion on valid parameter regimes). Figure 2 also reveals some unexplained variance, particularly pronounced in the covariances and correlations. This variance is the result of the finite simulation time and the associated uncertainty in the estimated covariances. As we show in Appendix K, the covariance estimate bias can be significant and it can only be corrected for on a statistical level rather than for individual covariances. Focusing on the statistics of covariances, however, has further advantages: For realistic network sizes, Eq. (9) is a high-dimensional equation that depends on each and every connection in the network. Understanding general mechanisms relating network structure and dynamics is therefore difficult. The covariance statistics instead summarize the most important aspects of covariances and, for large neuron populations, can be assumed to be self-averaging [40,50,51], which makes them less dependent on connectivity details. Second, Eq. (9) cannot be used for inference based on experimentally measured parameters because as of yet it is not possible to determine the effective connectivity or covariances of all neurons in a network. And last, as stated above, we demonstrate that covariance statistics are more robust measures than single-neuron covariances, both with respect to finite measurements as well as to the assumptions made in the derivation above.

III. STATISTICAL DESCRIPTION OF COVARIANCES

Expression (9) reveals that the statistics of the covariances $\boldsymbol{C}$, in particular their heterogeneity, is determined by the statistics and heterogeneity of the effective connectivity matrix $\boldsymbol{W}$ and the external noise strength $\boldsymbol{D}$. Our aim here is to derive a description of the cross-covariance statistics in terms of the statistics of $\boldsymbol{W}$ and $\boldsymbol{D}$. To this end, we derive analytical expressions for the mean and the variance of the time-lag integrated cross-covariances averaged over the heterogeneities of the system.

To do this, simply averaging Eq. (9) is not feasible due to $\boldsymbol{W}$ appearing in the inverse matrix $(\mathbf{1} - \boldsymbol{W})^{-1}$. Performing an average over a random connectivity is, however, a well-known problem in the theory of disordered systems [40,50,52,53], where it is handled on the level of generating functions. To proceed analogously, we start with Eq. (8), which expresses the covariances in terms of the moments of the dynamic variables' Fourier components at frequency zero. This allows us to write the covariances in terms of the moment-generating function $Z(\boldsymbol{J})$ of the zero-frequency Fourier components X_i of the dynamical equation (5) (see Appendix E for more details):

$$C_{ij} = \langle X_i X_j \rangle = \frac{\partial}{\partial J_i} \frac{\partial}{\partial J_j} Z(\boldsymbol{J}) \bigg|_{\boldsymbol{J}=0},$$

with

$$\begin{aligned} Z(\boldsymbol{J}) &= \frac{\widetilde{Z}(\boldsymbol{J})}{\widetilde{Z}(\mathbf{0})} \\ &= |\det(\mathbf{1} - \boldsymbol{W})| \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}} \\ &\quad \times \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1} - \boldsymbol{W})\boldsymbol{X} + \frac{1}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X}\right], \end{aligned} \tag{13}$$

and $\widetilde{Z}(\mathbf{0}) = |\det(1 - \boldsymbol{W})|^{-1}$ the nontrivial normalization of the unnormalized moment-generating function $\widetilde{Z}(\boldsymbol{J})$. Here,

$\widetilde{\boldsymbol{X}}$ are auxiliary variables that can be used to calculate the response function $\langle X_i \widetilde{X}_j \rangle$ of neuron i to a perturbation of neuron j by introducing additional sources $\widetilde{\boldsymbol{J}}$ in the moment-generating function (see Appendix E). Equation (13) shows that calculating the disorder average of the covariances boils down to calculating the disorder average of the moment-generating function. In the following two sections, we use this approach to calculate the mean of the cross-covariances $\langle \boldsymbol{C} \rangle_{\boldsymbol{W},\boldsymbol{D}}$ and subsequently the variance of the cross-covariances $\langle \delta \boldsymbol{C}^2 \rangle_{\boldsymbol{W},\boldsymbol{D}}$, where $\langle \cdot \rangle_{\boldsymbol{W},\boldsymbol{D}}$ refers to the average over the randomness in $\boldsymbol{W}$ and $\boldsymbol{D}$.

A. Mean of cross-covariances

Disorder average. We begin with the mean cross-covariances, focusing first on the average over the ensemble of connectivities, indicated in the following by $\langle \cdot \rangle_{\boldsymbol{W}}$. In the moment-generating function [Eq. (13)], $\boldsymbol{W}$ occurs linearly in the exponent of $\widetilde{Z}(\boldsymbol{J})$, which is advantageous for performing the disorder average. However, the averaging procedure is complicated by two aspects: (1) $\boldsymbol{W}$ contributes to the noise strength $\boldsymbol{D}$ through the variance-rescaling matrix $\boldsymbol{B}^{-1}$, and (2) the normalization $\widetilde{Z}(\boldsymbol{0})$ depends on $\boldsymbol{W}$. However, as illustrated in Fig. 3, in practice the first point does not appear to be a problem: Fig. 3(a) indicates that the specifics of $\boldsymbol{D}$ are largely determined by the details of the variances $\mathbf{C}\mathbf{V}^2\boldsymbol{\nu}$, because a different realization of $\boldsymbol{W}$ essentially yields a similar $\boldsymbol{D}$, and Fig. 3(b) suggests that the effect of the disorder average on $\boldsymbol{D}$ is minimal. For these reasons, we treat $\boldsymbol{D}$ as though it was independent of the explicit realization of $\boldsymbol{W}$. To address the second point, an alternative approach based on the moment-generating functional for the full time-dependent dynamics (see Appendix E) could be utilized. This moment-generating functional has a unit determinant normalization independent of $\boldsymbol{W}$ [54]. The disorder average of its frequency space complement, however, introduces cross-frequency couplings that complicate the further analysis. Here, instead, we follow Dahmen *et al.* [7], and separate the averages over $\widetilde{Z}(\boldsymbol{J})$ and $\widetilde{Z}(\boldsymbol{0})$,

$$\langle Z(\boldsymbol{J}) \rangle_{\boldsymbol{W}} = \left\langle \frac{\widetilde{Z}(\boldsymbol{J})}{\widetilde{Z}(\boldsymbol{0})} \right\rangle_{\boldsymbol{W}} \approx \frac{\langle \widetilde{Z}(\boldsymbol{J}) \rangle_{\boldsymbol{W}}}{\langle \widetilde{Z}(\boldsymbol{0}) \rangle_{\boldsymbol{W}}}, \tag{14}$$

as we find that this factorization approach does yield accurate results. This leaves us with the task of calculating $\langle \widetilde{Z}(\boldsymbol{J}) \rangle_{\boldsymbol{W}}$.[1]

The disorder average only affects the coupling term and can be expressed using the moment-generating functions ϕ_{ij} of W_{ij},

$$\begin{aligned} \left\langle \exp\left(-\widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{X}\right) \right\rangle_{\boldsymbol{W}} &= \left\langle \prod_{i,j} \exp\left(-W_{ij} \widetilde{X}_i X_j\right) \right\rangle_{\boldsymbol{W}} \\ &= \prod_{i,j} \phi_{ij}\left(-\widetilde{X}_i X_j\right), \end{aligned}$$

for independently drawn connections $W_{ij} \sim p_{ij}(W_{ij})$. The moment-generating function can be written in terms of a cumulant expansion $\phi_{ij}(X) = \exp(\sum_{k=1}^{\infty} \kappa_{k,ij} X^k / k!)$, with kth

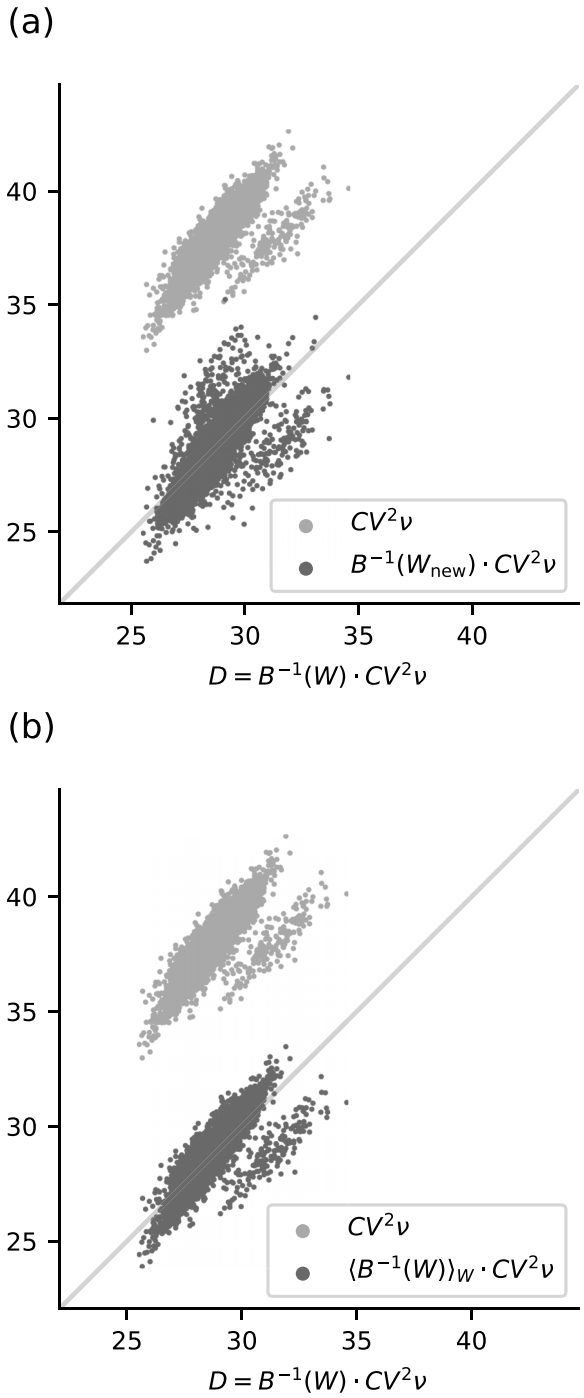

FIG. 3. Effect of averaging noise strength over disorder in $\boldsymbol{W}$ and effect of applying $\boldsymbol{B}^{-1}$ on variances. (a) Noise strength computed using the procedure described above (*noise strength* in the following) vs noise strength computed using a new realization of $\boldsymbol{W}$ (dark gray), and noise strength vs variances (light gray). (b) Noise strength vs noise strength computed using an average over 100 realizations of $\boldsymbol{W}$ (dark gray) and noise strength vs variances (light gray). Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A for spectral radius $r = 0.49$.

[1]Note that a systematic approach to this factorization approximation would be to employ the replica trick $\widetilde{Z}(\boldsymbol{0})^{-1} = \lim_{n\to 0} \widetilde{Z}(\boldsymbol{0})^{n-1}$ and jointly average $\langle \widetilde{Z}(\boldsymbol{J}) \widetilde{Z}(\boldsymbol{0})^{n-1} \rangle_{\boldsymbol{W}}$ in the limit $n \to 0$, or to average the joint moment-generating functional for all time points or frequencies that by construction has a trivial normalization $\widetilde{Z}(\boldsymbol{0}) = 1$ (see Appendix E).

cumulants $\kappa_{k,ij}$. For fixed connection probability, the number of inputs to a neuron scales with the network size N. To keep the input and its fluctuations finite when increasing the network size, we require synaptic weights to scale with $1/\sqrt{N}$ [34,55], such that the cumulant expansion is an expansion in $1/\sqrt{N}$. A truncation at the second cumulant ($\propto N^{-1}$) maps $\boldsymbol{W}$ to a Gaussian connectivity with distribution $\mathcal{N}(\boldsymbol{M}, \boldsymbol{\Delta}/N)$, such that

$$\langle \widetilde{Z}(\boldsymbol{J}) \rangle_{\boldsymbol{W}} = \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}} \times \exp\left[S_0(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) + \boldsymbol{J}^{\mathrm{T}} \boldsymbol{X} \right] \tag{15}$$

$$\times \exp\left[\frac{1}{2N} \sum_{i,j} \Delta_{ij} \widetilde{X}_i \widetilde{X}_i X_j X_j \right], \tag{16}$$

with

$$S_0(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) = \widetilde{\boldsymbol{X}}^{\mathrm{T}} (\mathbf{1} - \boldsymbol{M}) \boldsymbol{X} + \tfrac{1}{2} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{D} \widetilde{\boldsymbol{X}},$$

and mean connection weights $M_{ij} = \mathcal{O}(1/\sqrt{N})$ as well as variances $\Delta_{ij} = \mathcal{O}(1)$. The higher-order cumulants are suppressed by the large network size and therefore do not significantly affect the dynamics of the network. As we show below, our theory with the Gaussian approximation of the connectivity therefore faithfully recovers the correlation statistics in the sparse excitatory-inhibitory spiking network with population-specific connection statistics.

Auxiliary field formulation. To deal with the four-point coupling term in Eq. (15), we define auxiliary variables $Q_i := \frac{1}{N} \sum_j \Delta_{ij} X_j X_j$, which we formally introduce by inserting an identity in the form of a Fourier-transformed delta distribution:

$$1 = \prod_i \int \mathrm{d}Q_i \, \delta\left(\frac{1}{N} \sum_j \Delta_{ij} X_j X_j - Q_i \right)$$

$$= \prod_i N \int_{-\infty}^{\infty} \mathrm{d}Q_i \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\widetilde{Q}_i}{2\pi \mathrm{i}} \times \exp\left[\widetilde{Q}_i \left(\sum_j \Delta_{ij} X_j X_j - N Q_i \right) \right].$$

The auxiliary variables $\widetilde{Q}_i$ are introduced to express the delta distribution as an integral. This leads to

$$\langle \widetilde{Z}(\boldsymbol{J}) \rangle_{\boldsymbol{W}} = \int \mathcal{D}\boldsymbol{Q} \int \mathcal{D}\widetilde{\boldsymbol{Q}} \exp\left(-N \boldsymbol{Q}^T \widetilde{\boldsymbol{Q}} \right) \times \underbrace{\int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}} \exp\left[S_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) + \boldsymbol{J}^{\mathrm{T}} \boldsymbol{X} \right]}_{=: \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J})}, \tag{17}$$

$$S_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) = \widetilde{\boldsymbol{X}}^{\mathrm{T}} (\mathbf{1} - \boldsymbol{M}) \boldsymbol{X} + \frac{1}{2} \widetilde{\boldsymbol{X}}^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})] \widetilde{\boldsymbol{X}} + \boldsymbol{X}^{\mathrm{T}} \mathrm{diag}\left(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta} \right) \boldsymbol{X}. \tag{18}$$

Here $\mathrm{diag}(\boldsymbol{Q})_{ij}$ refers to a diagonal matrix with diagonal elements Q_i. As the action $S_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{X}, \widetilde{\boldsymbol{X}})$ at fixed auxiliary variables describes an auxiliary free theory, Eq. (17) describes the activity of linear rate neurons in a network with disorder-averaged connectivity $\boldsymbol{M}$ that interact with fluctuating external variables $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$. Inserting Eq. (17) into Eq. (14) yields

$$\langle Z(\boldsymbol{J}) \rangle_{\boldsymbol{W}} = \int \mathcal{D}\boldsymbol{Q} \int \mathcal{D}\widetilde{\boldsymbol{Q}} \frac{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J})}{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0})} \times \frac{\exp\left(-N \boldsymbol{Q}^T \widetilde{\boldsymbol{Q}} \right) \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0})}{\int \mathcal{D}\boldsymbol{P} \int \mathcal{D}\widetilde{\boldsymbol{P}} \exp\left(-N \boldsymbol{P}^T \widetilde{\boldsymbol{P}} \right) \widetilde{Z}_{\boldsymbol{P},\widetilde{\boldsymbol{P}}}(\mathbf{0})}$$

$$=: \int \mathcal{D}\boldsymbol{Q} \int \mathcal{D}\widetilde{\boldsymbol{Q}} \, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J}),$$

with joint probability distribution

$$p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) = \frac{\exp\left(-S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \right)}{\int \mathcal{D}\boldsymbol{P} \int \mathcal{D}\widetilde{\boldsymbol{P}} \exp\left(-S(\boldsymbol{P}, \widetilde{\boldsymbol{P}}) \right)},$$

$$S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) = N \boldsymbol{Q}^T \widetilde{\boldsymbol{Q}} - \ln\left[\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0}) \right], \tag{19}$$

and properly normalized moment-generating function $Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J}) = \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J}) / \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0})$. These equations imply that the disorder average of arbitrary moments $\langle X_{i_1} \cdots X_{i_k} \rangle$ can be calculated by determining the corresponding moments $\langle X_{i_1} \cdots X_{i_k} \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}$ with respect to the auxiliary free theory and averaging them over the auxiliary variables:

$$\langle \langle X_{i_1} \cdots X_{i_k} \rangle \rangle_{\boldsymbol{W}} = \int \mathcal{D}\boldsymbol{Q} \int \mathcal{D}\widetilde{\boldsymbol{Q}} \times p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_{i_1} \cdots X_{i_k} \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}},$$

$$\langle X_{i_1} \cdots X_{i_k} \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} = \frac{\partial}{\partial J_{i_1}} \cdots \frac{\partial}{\partial J_{i_k}} Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J}) \Big|_{\boldsymbol{J}=\mathbf{0}}. \tag{20}$$

Saddle-point approximation. Due to the prefactor N in Eq. (19) and the scalar products in $\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0})$ with N contributions, we expect $p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})$ to peak sharply for $N \to \infty$, such that we can perform a saddle-point approximation. To lowest order, we expect $p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \approx \delta(\boldsymbol{Q} - \boldsymbol{Q}^*) \delta(\widetilde{\boldsymbol{Q}} - \widetilde{\boldsymbol{Q}}^*)$, with the saddle-point $\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*$ determined by

$$\frac{\partial}{\partial \widetilde{Q}_i} S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0, \qquad \frac{\partial}{\partial Q_i} S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0,$$

which yields

$$Q_i^* = \frac{1}{N} \sum_j \Delta_{ij} \langle X_j X_j \rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*},$$

$$\widetilde{Q}_i^* = \frac{1}{2N} \langle \widetilde{X}_i \widetilde{X}_i \rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}, \tag{21}$$

with second moments evaluated at the saddle point. The moments can be calculated explicitly by solving the Gaussian integrals (see Appendix F). Using the shorthand $\boldsymbol{R} := (\mathbf{1} - \boldsymbol{M})^{-1}$, we find $\langle X_i X_j \rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = \{ \boldsymbol{R} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q}^*)] \boldsymbol{R}^{\mathrm{T}} \}_{ij}$ and $\langle \widetilde{X}_i \widetilde{X}_i \rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0$, and solving for the saddle point yields

$$Q_i^* = \frac{1}{N} \sum_{j,k,l,m} \left(\mathbf{1} - \frac{1}{N} \Delta \cdot \boldsymbol{R}^{\circ 2} \right)^{-1}_{ij} \Delta_{jk} R_{kl} D_{lm} R_{km},$$

$$\widetilde{Q}_i^* = 0,$$

with $\boldsymbol{R}^{\circ 2} := \boldsymbol{R} \odot \boldsymbol{R}$ and $\odot$ denoting the elementwise (Hadamard) product.

Finally, making use of the Wiener-Khinchin theorem [Eq. (B1)] and inserting the solution of the saddle-point equations into Eqs. (20) yields the mean covariances averaged across the disorder of the connectivity:

$$\begin{aligned}\langle \boldsymbol{C}\rangle_{\boldsymbol{W}} &= \langle \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\\ &= (\mathbf{1}-\boldsymbol{M})^{-1}[\boldsymbol{D}+\mathrm{diag}(\boldsymbol{Q}^*)](\mathbf{1}-\boldsymbol{M})^{-\mathrm{T}}.\end{aligned}$$

Averaging over the disorder in $\boldsymbol{D}$ then yields

$$\langle \boldsymbol{C}\rangle_{\boldsymbol{W},\boldsymbol{D}} = (\mathbf{1}-\boldsymbol{M})^{-1}\{\overline{\boldsymbol{D}}+\mathrm{diag}[\boldsymbol{Q}^*(\boldsymbol{D}=\overline{\boldsymbol{D}})]\}(\mathbf{1}-\boldsymbol{M})^{-\mathrm{T}}. \tag{22}$$

Here $\overline{\boldsymbol{D}}$ denotes the disorder-averaged noise strength [cf. Fig. 3(b) and discussion after Eq. (31)]. Note that the saddle point $\boldsymbol{Q}^*(\boldsymbol{D}=\overline{\boldsymbol{D}})$ together with $\overline{\boldsymbol{D}}$ yields an effective noise strength which shifts average variances and covariances. Importantly, it is only the heterogeneity in the connectivity $\boldsymbol{W}$ that causes this shift. Average covariances are insensitive to heterogeneity in the noise strengths $\boldsymbol{D}$; they only depend on the average $\overline{\boldsymbol{D}}$.

B. Variance of cross covariances

Replica method. Calculating the variances of covariances across the ensemble of possible network connectivities,

$$\langle \delta \boldsymbol{C}^2\rangle_{\boldsymbol{W}} = \langle \boldsymbol{C}^{\circ 2}\rangle_{\boldsymbol{W}} - \langle \boldsymbol{C}\rangle_{\boldsymbol{W}}^{\circ 2}, \tag{23}$$

requires making use of the replica method [40,56] and deriving an expression for the disorder-averaged moment-generating function of the replicated system $\langle Z(\boldsymbol{J})Z(\boldsymbol{K})\rangle_{\boldsymbol{W}}$, as this allows calculating disorder averages of arbitrary squared moments $\langle\langle X_{i_1}\cdots X_{i_k}\rangle^2\rangle_{\boldsymbol{W}}$, which occur in the first term in Eq. (23). The procedure is completely analogous to the previous section's derivations. However, the disorder average now affects the term

$$\begin{aligned}&\left\langle \exp\left(\widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}+\widetilde{\boldsymbol{Y}}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{Y}\right)\right\rangle_{\boldsymbol{W}}\\ &\quad=\prod_{i,j}\exp\left[\sum_{k=1}^{\infty}\frac{\kappa_{k,ij}}{k!}\left(\widetilde{X}_iX_j+\widetilde{Y}_iY_j\right)^k\right],\end{aligned}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ refer to the activity in the first and second replicon, respectively. A cumulant expansion up to second order introduces—along four-point couplings separately in $\boldsymbol{X}$ and $\boldsymbol{Y}$ similar to the one in Eq. (15)—a replica coupling term

$$\exp\left(\frac{1}{N}\sum_{ij}\Delta_{ij}\widetilde{X}_i\widetilde{Y}_iX_jY_j\right).$$

To deal with the four-point couplings, we again introduce auxiliary variables

$$\begin{aligned}Q_{XX,i} &= \frac{1}{N}\sum_j \Delta_{ij}X_jX_j,\\ Q_{YY,i} &= \frac{1}{N}\sum_j \Delta_{ij}Y_jY_j,\\ Q_{XY,i} &= \frac{1}{N}\sum_j \Delta_{ij}X_jY_j,\end{aligned}$$

and obtain a relation similar to Eq. (20),

$$\begin{aligned}&\langle\langle X_{i_1}\cdots X_{i_k}\rangle^2\rangle_{\boldsymbol{W}}\\ &\quad=\int\mathcal{D}\boldsymbol{Q}\int\mathcal{D}\widetilde{\boldsymbol{Q}}\,p(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})\langle X_{i_1}\cdots X_{i_k}Y_{i_1}\cdots Y_{i_k}\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}},\\ &\langle X_{i_1}\cdots X_{i_k}Y_{i_1}\cdots Y_{i_k}\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\\ &\quad=\frac{\partial}{\partial J_{i_1}}\cdots\frac{\partial}{\partial J_{i_k}}\frac{\partial}{\partial K_{i_1}}\cdots\frac{\partial}{\partial K_{i_k}}Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})\Big|_{\boldsymbol{J},\boldsymbol{K}=\boldsymbol{0}},\end{aligned} \tag{24}$$

but with

$$\begin{aligned}p(\boldsymbol{Q},\widetilde{\boldsymbol{Q}}) &= \frac{\exp\left(-S(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})\right)}{\int\mathcal{D}\boldsymbol{P}\int\mathcal{D}\widetilde{\boldsymbol{P}}\,\exp\left(-S(\boldsymbol{P},\widetilde{\boldsymbol{P}})\right)},\\ \mathcal{S}(\boldsymbol{Q},\widetilde{\boldsymbol{Q}}) &= N\boldsymbol{Q}_{XX}^{\mathrm{T}}\widetilde{\boldsymbol{Q}}_{XX}+N\boldsymbol{Q}_{XY}^{\mathrm{T}}\widetilde{\boldsymbol{Q}}_{XY}\\ &\quad+N\boldsymbol{Q}_{YY}^{\mathrm{T}}\widetilde{\boldsymbol{Q}}_{YY}-\ln\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{0},\boldsymbol{0}),\\ \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) &= \int\mathcal{D}\boldsymbol{X}\int\mathcal{D}\widetilde{\boldsymbol{X}}\int\mathcal{D}\boldsymbol{Y}\int\mathcal{D}\widetilde{\boldsymbol{Y}}\\ &\quad\times\exp\left[S_{\boldsymbol{Q}_{XX},\widetilde{\boldsymbol{Q}}_{XX}}(\boldsymbol{X},\widetilde{\boldsymbol{X}})+S_{\boldsymbol{Q}_{YY},\widetilde{\boldsymbol{Q}}_{YY}}(\boldsymbol{Y},\widetilde{\boldsymbol{Y}})\right.\\ &\quad+\widetilde{\boldsymbol{X}}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{Q}_{XY})\widetilde{\boldsymbol{Y}}+\boldsymbol{X}^{\mathrm{T}}\mathrm{diag}\left(\widetilde{\boldsymbol{Q}}_{XY}^{\mathrm{T}}\boldsymbol{\Delta}\right)\boldsymbol{Y}\\ &\quad\left.+\boldsymbol{J}^{\mathrm{T}}\boldsymbol{X}+\boldsymbol{K}^{\mathrm{T}}\boldsymbol{Y}\right],\\ Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) &= \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})/\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{0},\boldsymbol{0}),\end{aligned} \tag{25}$$

where $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$ are shorthand notations denoting all auxiliary variables, and $S_{\boldsymbol{Q}_{XX},\widetilde{\boldsymbol{Q}}_{XX}}(\boldsymbol{X},\widetilde{\boldsymbol{X}})$ and $S_{\boldsymbol{Q}_{YY},\widetilde{\boldsymbol{Q}}_{YY}}(\boldsymbol{Y},\widetilde{\boldsymbol{Y}})$ are given by Eq. (18).

Saddle-point approximation. As in Sec. III A, we approximate $p(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})$ as a delta function at the saddle point $\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*$ (for details see Appendix F), and with Eq. (24) to lowest order we get

$$\begin{aligned}\left\langle C_{ij}^2\right\rangle_{\boldsymbol{W}} &= \langle\langle X_iX_j\rangle^2\rangle_{\boldsymbol{W}}\\ &=\int\mathcal{D}\boldsymbol{Q}\int\mathcal{D}\widetilde{\boldsymbol{Q}}\,p(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})\langle X_iX_jY_iY_j\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\\ &=\int\mathcal{D}\boldsymbol{Q}\int\mathcal{D}\widetilde{\boldsymbol{Q}}\,p(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})\\ &\quad\times\left(\langle X_iX_j\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\langle Y_iY_j\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\right.\\ &\quad+\langle X_iY_i\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\langle X_jY_j\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\\ &\quad\left.+\langle X_iY_j\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\langle X_jY_i\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\right)\\ &\approx\langle X_iX_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle Y_iY_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\\ &\quad+\langle X_iY_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle X_jY_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\\ &\quad+\langle X_iY_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle X_jY_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\\ &=\langle X_iX_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle Y_iY_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\\ &\approx\langle\langle X_iX_j\rangle\rangle_{\boldsymbol{W}}^2\\ &=\langle C_{ij}\rangle_{\boldsymbol{W}}^2,\end{aligned} \tag{26}$$

(27)

where we used Wick's theorem, which is allowed by the fact that, for $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$ given and fixed, $\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})$ describes a Gaussian theory, and the fact that all cross-replica correlators $\langle X_iY_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}$ vanish at the saddle point (see Appendix F).

Fluctuations around the saddle point. Equation (27) implies that the variance of covariances is zero in the saddle-point approximation, and we need to account for Gaussian fluctuations of the auxiliary fields around their saddle points by making a Gaussian approximation of $p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})$. The crucial fluctuations are the ones of $\boldsymbol{Q}_{XY}$ and $\widetilde{\boldsymbol{Q}}_{XY}$, as they can potentially preserve the replica coupling and thus lead to nonvanishing variance contributions of cross-replica correlators $\langle X_i Y_i \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*}$. Away from the saddle points, the correlators in Eq. (26) depend on $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$ in a complicated manner. To render the integrals in Eq. (26) solvable in the Gaussian approximation, we perform a Taylor expansion of the correlators around the saddle points $\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*$, which effectively is an expansion of $\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J}, \boldsymbol{K})$ (see Appendix G for more details). In the first term of Eq. (26), leading-order fluctuations in $\boldsymbol{Q}_{XY}$ and $\widetilde{\boldsymbol{Q}}_{XY}$ depend on correlators with an odd number of variables of each replicon. Therefore, this term cannot yield a contribution to the variance due to fluctuations of $\boldsymbol{Q}_{XY}$ and $\widetilde{\boldsymbol{Q}}_{XY}$. The major replica coupling arises from the second and third terms in Eq. (26). We note that the third term contains off-diagonal elements of correlators which are suppressed by a factor $1/N$ with respect to the diagonal ones. Therefore, we can neglect this term for cross-covariances as well and only keep the second term in Eq. (26) as the leading-order contribution. For autocovariances the second and third terms in Eq. (26) are the same, yielding an additional factor 2. Introducing $\delta\boldsymbol{Q} = \boldsymbol{Q} - \boldsymbol{Q}^*$ and defining $\delta\widetilde{\boldsymbol{Q}}$ equivalently, we obtain

$$\begin{aligned}\langle X_i Y_i \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} &= \sum_k \left\langle X_i Y_i \widetilde{X}_k \widetilde{Y}_k \right\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \delta Q_{XY,k} + \sum_{k,l} \Delta_{kl} \langle X_i Y_i X_l Y_l \rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \delta\widetilde{Q}_{XY,k} + \mathcal{O}\big(|\delta\boldsymbol{Q}|^2, |\delta\widetilde{\boldsymbol{Q}}|^2\big) \\ &= \sum_k \left\langle X_i \widetilde{X}_k \right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \delta Q_{XY,k} + \sum_{k,l} \Delta_{kl} \langle X_i X_l \rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \delta\widetilde{Q}_{XY,k} + \mathcal{O}\big(|\delta\boldsymbol{Q}|^2, |\delta\widetilde{\boldsymbol{Q}}|^2\big), \end{aligned} \tag{28}$$

where we used that cross-replica correlators vanish at the saddle point. Inserting the above fluctuation expansion result around $\boldsymbol{Q}^*_{XY}$ and $\widetilde{\boldsymbol{Q}}^*_{XY}$ into Eq. (26) leads to

$$\begin{aligned}&\int \mathcal{D}\boldsymbol{Q} \int \mathcal{D}\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_i Y_i \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} \langle X_j Y_j \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} \\ &= \sum_{k,l} \left\langle X_i \widetilde{X}_k \right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \left\langle X_j \widetilde{X}_l \right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \langle \delta Q_{XY,k} \delta Q_{XY,l} \rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} + \sum_{k,l,m} \left\langle X_i \widetilde{X}_k \right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \Delta_{lm} \langle X_j X_m \rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \left\langle \delta Q_{XY,k} \delta\widetilde{Q}_{XY,l} \right\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} \\ &\quad + \sum_{k,l,m} \left\langle X_j \widetilde{X}_k \right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \Delta_{lm} \langle X_i X_m \rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \left\langle \delta Q_{XY,k} \delta\widetilde{Q}_{XY,l} \right\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} + \sum_{k,l,m,n} \langle X_i X_m \rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \langle X_j X_n \rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \Delta_{km} \Delta_{ln} \left\langle \delta\widetilde{Q}_{XY,k} \delta\widetilde{Q}_{XY,l} \right\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}. \end{aligned} \tag{29}$$

Next, we consider the Gaussian approximation of $p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})$ with

$$\mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) = \mathcal{S}(\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*) + \frac{1}{2} \big(\delta\boldsymbol{Q}_{XY}, \delta\widetilde{\boldsymbol{Q}}_{XY}\big) \boldsymbol{\mathcal{S}}^{(2)} \begin{pmatrix} \delta\boldsymbol{Q}_{XY} \\ \delta\widetilde{\boldsymbol{Q}}_{XY} \end{pmatrix},$$

where $\boldsymbol{\mathcal{S}}^{(2)}$ contains the second derivatives with respect to the auxiliary fields,

$$\boldsymbol{\mathcal{S}}^{(2)} = \begin{pmatrix} \frac{\partial \mathcal{S}(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})}{\partial \boldsymbol{Q}_{XY} \partial \boldsymbol{Q}_{XY}} \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} & \frac{\partial \mathcal{S}(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})}{\partial \boldsymbol{Q}_{XY} \partial \widetilde{\boldsymbol{Q}}_{XY}} \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \\ \frac{\partial \mathcal{S}(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})}{\partial \widetilde{\boldsymbol{Q}}_{XY} \partial \boldsymbol{Q}_{XY}} \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} & \frac{\partial \mathcal{S}(\boldsymbol{Q},\widetilde{\boldsymbol{Q}})}{\partial \widetilde{\boldsymbol{Q}}_{XY} \partial \widetilde{\boldsymbol{Q}}_{XY}} \Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \end{pmatrix},$$

which allows evaluating the correlators of the auxiliary fields in Eq. (29) (see Appendix H for details). Inserting the results, to leading order we find (see Appendix I for details)

$$\begin{aligned}\big\langle C_{ij}^2 \big\rangle_{\boldsymbol{W}} = (1 + \delta_{ij}) \Bigg[\left(\mathbf{1} - \frac{1}{N} \boldsymbol{R}^{\mathrm{T}\circ 2} \boldsymbol{\Delta} \right)^{-1} \langle \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} \\ \times \left(\mathbf{1} - \frac{1}{N} \boldsymbol{R}^{\mathrm{T}\circ 2} \boldsymbol{\Delta} \right)^{-\mathrm{T}} \Bigg]_{ij} - \delta_{ij} \langle C_{ij} \rangle^2_{\boldsymbol{W}}. \end{aligned} \tag{30}$$

To get the variances rather than the second moments, we subtract the squared mean covariances $\langle C_{ij} \rangle^2_{\boldsymbol{W}}$. However, for the setup that we study here the squared mean cross-covariances are of the order $\mathcal{O}(1/N^2)$ and therefore negligible. Taking into account that $\boldsymbol{R} = (\mathbf{1} - \boldsymbol{M})^{-1} \approx \mathbf{1}$, which holds as long as the network is inhibition dominated,[2] we find the following expression for the disorder-averaged variance of cross-covariances (see Appendix I for full expression):

$$\langle \delta \boldsymbol{C}^2 \rangle_{\boldsymbol{W}} = (\mathbf{1} - \boldsymbol{S})^{-1} (\boldsymbol{D} + \mathrm{diag}[\boldsymbol{Q}^*(\boldsymbol{D})])^2 (\mathbf{1} - \boldsymbol{S})^{-\mathrm{T}}, \tag{31}$$

where we wrote $\boldsymbol{S} = \boldsymbol{\Delta}/N$.

However, if the noise strength $\boldsymbol{D}$ has to be estimated using Eq. (11), this expression is still dependent on the specific realization of $\boldsymbol{W}$, both implicitly through the estimates of the single-neuron rates and CVs described in Sec. II and explicitly through the matrix $\boldsymbol{B}$ [Eq. (12)]. Since the right-hand side of Eq. (31) depends nonlinearly on $\boldsymbol{D}$, averaging over the statistics of $\boldsymbol{D}$ introduces terms depending on the heterogeneity of $\boldsymbol{D}$. However, Fig. 9 in the Appendix shows that heterogeneity in $\boldsymbol{D}$—both via the explicit dependence on $\boldsymbol{W}$ and via the implicit dependence through distributed firing rates and CVs—is negligible for the statistics of cross-covariances. This can be understood by considering the structure of Eq. (35): The matrices $(\mathbf{1} - \boldsymbol{S})^{-1}$ are multiplied with $\boldsymbol{D}$, such that any heterogeneity in $\boldsymbol{D}$ is averaged out. An E-I network is an illustrative example, with $(\mathbf{1} - \boldsymbol{S})^{-1} = \mathbf{1} + \boldsymbol{U}$ with a 2×2 block matrix $\boldsymbol{U}$ whose entries are homogeneous in each population

[2] $\boldsymbol{M}$ scales as $\mathcal{O}(N^{-1/2})$ and, in inhibition-dominated networks, eigenvalues of $\boldsymbol{M}$ are far away from the divergence at 1.

block, such that the matrix product effectively is an average over $\boldsymbol{D}$.

To obtain an average $\boldsymbol{D}$ that is not depending on a specific realization of $\boldsymbol{W}$, we follow Eq. (11) and set

$$\overline{D}_{ii} = \sum_j (\mathbf{1} - \boldsymbol{S})_{ij} \cdot \overline{\mathrm{CV}_j^2 \nu_j}, \tag{32}$$

which inserted into the disorder-averaged expression for the autocovariances [Eq. (22)] yields the correct autocovariances:

$$\begin{aligned}\langle C_{ii}\rangle_{\boldsymbol{W},\boldsymbol{D}} &= [(\mathbf{1}-\boldsymbol{M})^{-1}\{\overline{\boldsymbol{D}} + \mathrm{diag}[\boldsymbol{Q}^*(\overline{\boldsymbol{D}})]\}(\mathbf{1}-\boldsymbol{M})^{-\mathrm{T}}]_{ii} \\ &\approx \{\overline{\boldsymbol{D}} + \mathrm{diag}[(\mathbf{1}-\boldsymbol{S})^{-1}\boldsymbol{S}\cdot\mathrm{diag}(\overline{\boldsymbol{D}})]\}_{ii} \\ &= \sum_j (\mathbf{1}-\boldsymbol{S})^{-1}_{ij}\overline{D}_{jj} \\ &= \overline{\mathrm{CV}_i^2\nu_i}. \end{aligned} \tag{33}$$

Here we used $(\mathbf{1}-\boldsymbol{M})^{-1} \approx \mathbf{1}$ and $\boldsymbol{Q}^* \approx (\mathbf{1}-\boldsymbol{S})^{-1}\boldsymbol{S}\cdot \mathrm{diag}(\boldsymbol{D})$. The realization-independent estimates $\overline{\nu_i}$ and $\overline{\mathrm{CV}_i^2}$ of the rates and CVs, respectively, can be obtained using standard population-resolved mean-field theory [42,43], which only requires knowing the statistics of $\boldsymbol{W}$. A procedure similar to the one described in Sec. II can be used: In the population view, however, the indices i, j no longer denote single neurons but rather populations of equal neurons. In Eq. (1) J_{ij} is replaced by $K_{ij}J_{ij}$ and J_{ij}^2 in Eq. (2) is replaced by $K_{ij}J_{ij}^2$, where K_{ij} is the indegree from population j to population i, and J_{ij} then is interpreted as the mean synaptic weight from population j to population i.

Replacing $\boldsymbol{D}$ in Eq. (31) by Eq. (32) yields a fully realization-independent disorder-averaged estimate of the variance of cross-covariances.

C. Singularities

Next, we discuss the interpretation of the derived formulas. Thereto, we need to have a closer look at the effective noise strength $\boldsymbol{D} + \mathrm{diag}[\boldsymbol{Q}^*(\boldsymbol{D})]$, which occurs in both the mean [Eq. (22)] and the variances [Eq. (31)] of covariances. Using Eq. (32), we find that the impact of heterogeneity on the effective noise cancels:

$$\begin{aligned}\mathrm{diag}(\boldsymbol{D}) + \boldsymbol{Q}^*(\boldsymbol{D}) &\approx \mathrm{diag}(\boldsymbol{D}) + (\mathbf{1}-\boldsymbol{S})^{-1}\boldsymbol{S}\cdot\mathrm{diag}(\boldsymbol{D}) \\ &= (\mathbf{1}-\boldsymbol{S})^{-1}\cdot\mathrm{diag}(\boldsymbol{D}) \\ &= (\mathbf{1}-\boldsymbol{S})^{-1}(\mathbf{1}-\boldsymbol{S})\cdot\boldsymbol{a} \\ &\approx \boldsymbol{a}, \end{aligned} \tag{34}$$

where $a_i = \overline{\mathrm{CV}_i^2}\,\overline{\nu_i}$ is the vector of estimated autocovariances. This is because we specifically choose the noise strength $\boldsymbol{D}$ such that autocovariances match those from the spiking networks: As heterogeneity is increased, external fluctuations get amplified by the factor $(\mathbf{1}-\boldsymbol{S})^{-1}$ in Eq. (34). To achieve that autocovariances do not diverge, external inputs need to be scaled down according to Eq. (32). Hence, the mean and variance of cross-covariances are given by

$$\langle \boldsymbol{C}\rangle_{\boldsymbol{W},\boldsymbol{D}} \approx (\mathbf{1}-\boldsymbol{M})^{-1}\mathrm{diag}(\boldsymbol{a})(\mathbf{1}-\boldsymbol{M})^{-\mathrm{T}}, \tag{35}$$

$$\langle \delta\boldsymbol{C}^2\rangle_{\boldsymbol{W},\boldsymbol{D}} \approx (\mathbf{1}-\boldsymbol{S})^{-1}\mathrm{diag}(\boldsymbol{a}^2)(\mathbf{1}-\boldsymbol{S})^{-\mathrm{T}}. \tag{36}$$

Note that any inverse matrix can be written as $\boldsymbol{A}^{-1} = \det(\boldsymbol{A})^{-1}\mathrm{adj}(\boldsymbol{A})$, where $\mathrm{adj}(\boldsymbol{A})$ denotes the adjugate matrix. As a result, the elements of an inverse matrix $\boldsymbol{A}^{-1}$ diverge if the determinant of the matrix $\boldsymbol{A}$ vanishes, which occurs when at least one eigenvalue of $\boldsymbol{A}$ is zero. Therefore, the divergence behavior of the mean and variance of covariances is determined by the eigenvalues of $\boldsymbol{M}$ and $\boldsymbol{S}$ with real parts close to 1.

Equation (35) reveals that mean cross-covariances are determined by the mean connectivity $\boldsymbol{M}$. By choosing $\boldsymbol{D}$ to match the autocovariances of the spiking model, they are, in particular, unaffected by network heterogeneity, represented by $\boldsymbol{S}$. A range of important network properties, such as population structure determining E-I balance [34,36,37,45,57], spatial structure like distance-dependent connection probabilities [8,58–62], or low-rank structures [63], can be encoded in $\boldsymbol{M}$. Divergences in mean covariances, caused by eigenvalues of $\boldsymbol{M}$ close to 1, can thus be indicative of phenomena like loss of E-I balance with excessive excitation (cf. Fig. 8D of Ref. [57]) or instability of the homogeneously active state in spatially organized networks [64].

Variances of cross-covariances are determined by network heterogeneity [Eq. (36)], encoded in the connectivity variance $\boldsymbol{S}$, and are to leading order independent of the mean connectivity $\boldsymbol{M}$. Note that subleading terms nevertheless can become sizable if eigenvalues of $\boldsymbol{M}$ are close to the instability line at 1. As demonstrated by Aljadeff *et al.* [65], if $\boldsymbol{S}$ is a block structured matrix encoding different populations, its eigenvalue spectrum is circular, with a spectral radius r that is determined by the square root of the maximum eigenvalue of $\boldsymbol{S}$.

For the E-I network studied here, the matrices $\boldsymbol{M}$ and $\boldsymbol{S}$ have nontrivial block structure with one excitatory and one inhibitory block (see Appendix M). In this case, the spectral radius is given by [66]

$$r^2 = N_\mathrm{E}\sigma_\mathrm{E}^2 + N_\mathrm{I}\sigma_\mathrm{I}^2.$$

The spectral radius, a measure of network heterogeneity, increases when the variance of synaptic strength grows, which is controlled by an interplay between the connection probabilities of different populations and the variances of the associated synaptic weights. Intuitively, as explained in Dahmen *et al.* [8], multisynaptic signal transmission is very efficient in a network with a large spectral radius, such that pairs of neurons influence each other via a large number of neuronal pathways, possibly including differing numbers of excitatory and inhibitory neurons. The effects of these various pathways add up, and the large variety of potential pathways results in a broad distribution of covariances.

We see that the effects of $\boldsymbol{M}$ and $\boldsymbol{S}$ are mostly independent of one another, allowing the mean and variance of covariances to vary separately. This, however, applies only to synaptic weights that are identically and independently distributed. If the weights are correlated, such as through chain structures in the connectivity, the respective eigenvalues cannot be changed independently. A more detailed analysis of this behavior is to be published elsewhere. As a final remark, it is worth noting that the independence of the mean covariances of $\boldsymbol{S}$ confirms that previously

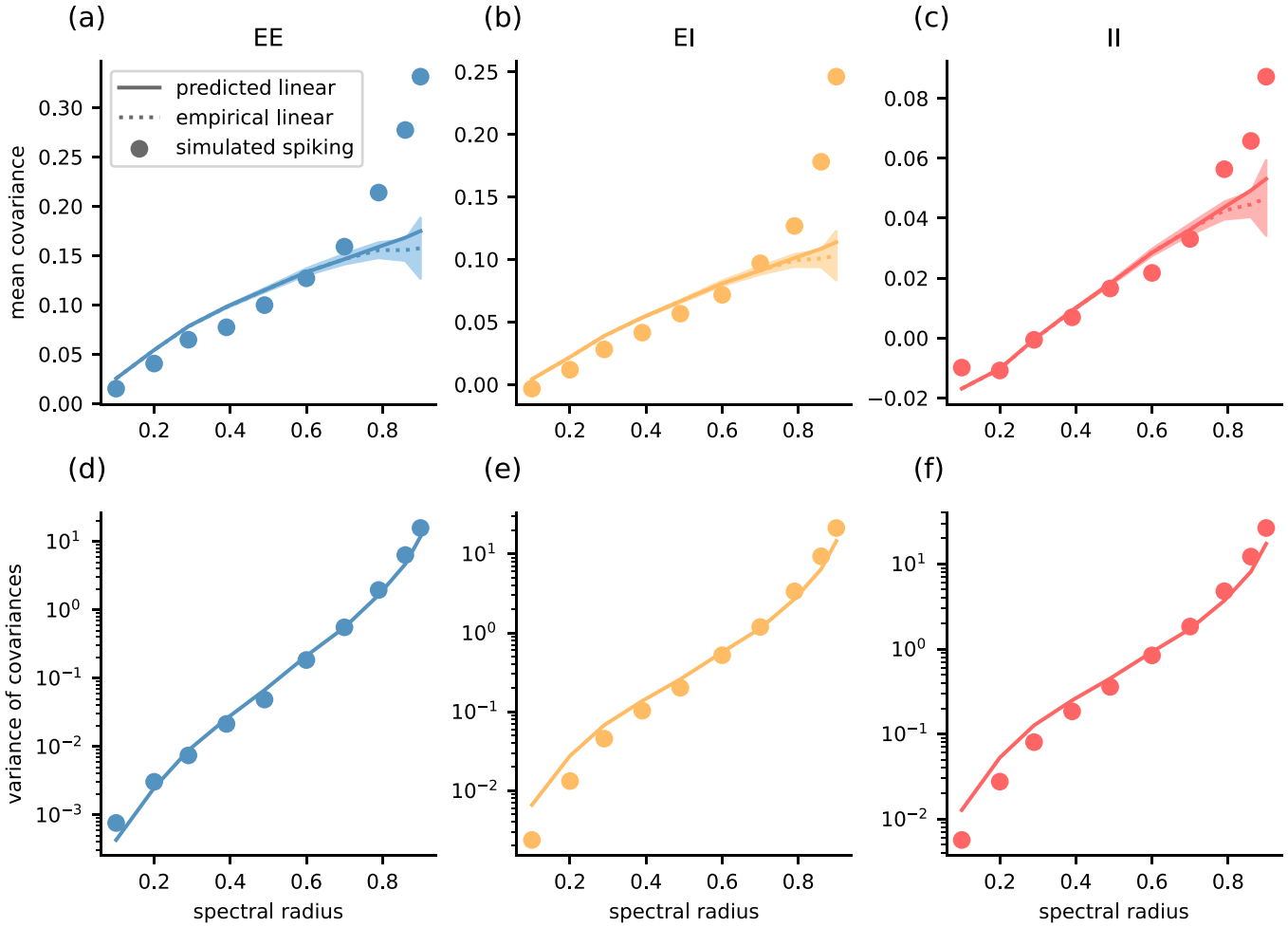

FIG. 4. Statistics of covariances at different spectral radii of the effective connectivity matrix $\boldsymbol{W}$. Covariance statistics are resolved according to population membership of neuron pairs (EE, both neurons excitatory; EI, one neuron excitatory, one neuron inhibitory; II, both neurons inhibitory). Dots show the population-resolved [(a)–(c)] mean or [(d)–(f)] variance of the spike count cross-covariances [Eq. (7)] measured in network simulations of spiking E-I networks. Solid lines show the theoretical predictions of the mean [Eq. (22)] and variance [Eq. (31)] of cross-covariances, respectively, using the noise strength estimate [Eq. (32)]. The dashed lines show the population-resolved empirical mean and variance of cross-covariances from Eq. (9) averaged over 20 realizations of W, and the shaded area depicts a two-standard-deviation range around the mean. The variances computed from the simulation results have been corrected for bias due to finite simulation time (see Appendix K). Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

employed population models [31,36,37,45,57], which neglect the variance of connectivity, are valid for computing mean covariances.

To illustrate how the mean and variance of covariances change as functions of the network heterogeneity, we plot Eqs. (22) and (31) with Eq. (32) for spectral radii between 0 and 1 in Fig. 4 (predicted linear). We kept the working point roughly constant for the different spectral radii by maintaining the mean μ and variance σ^2 of the total input to each neuron while modifying the synaptic efficacy. To compensate for the increased intrinsic input and fluctuations at larger spectral radii, we reduced the mean and fluctuations of the external input.

Confirming the discussion of Eqs. (35) and (36), when the spectral radius is modified, the variances of covariances vary by several orders of magnitude, whereas mean covariances remain in the same order of magnitude. A range of prior research [31,37,57] has shown that a divergence of mean covariances would be observed as a function of E-I balance, e.g., by altering g. Here we focus on network scenarios away from the excitatory instability (fixed $g = -6$) and therefore do not see a divergence of mean covariances. Nevertheless, we observe a change of mean covariances when changing the spectral radius. This is because, in the sparse random network chosen here, the variance of the synaptic weights is not independent from the mean of the weights. Adjusting the spectral radius requires modifying the weights, resulting in the residual change in the mean covariances visible in Figs. 4(a)–4(c). Note that by keeping the working point of the network constant across spectral radii, we also keep the noise strength factor in Eqs. (22) and (31) constant [cf. Eqs. (35) and (36)]. If the external noise strength was instead determined by a fixed external process, i.e., $\boldsymbol{D}$ independent of $\boldsymbol{W}$, then mean covariances would also diverge as a function of the spectral

radius due to the factor $(\mathbf{1}-\boldsymbol{S})^{-1}$, which enters the noise strength term via $\boldsymbol{Q}^*$.

D. Comparison of prediction and measurement of covariance statistics

To check how closely the predictions match the outcomes of spiking network simulations, we ran ten simulations for different spectral radii of the effective connectivity matrix $\boldsymbol{W}$ similar to the one shown in Fig. 1 using the parameters specified in Appendix A. The spectral radius was modified by changing the synaptic weight scale j (see Table II). In the sparse E-I networks that we consider here, this affects both the mean connectivity $\boldsymbol{M}$ and the variance of connections, $\boldsymbol{S}$ [cf. Eq. (A3)]. We ensured that the spiking networks have roughly similar working points for the different spectral radii by adjusting the external input as explained in Appendix D. The theory of Eqs. (22) and (31) yields, for any pair of neurons in the network, predictions for the mean and variance of covariances measured across different network realizations. In brain circuits, one finds statistical similarities between different neurons, for example, given by their population membership. These similarities result in symmetries in the matrices $\boldsymbol{M}$ and $\boldsymbol{S}$. In the current example of the excitatory-inhibitory network, the statistical similarity among all excitatory neurons and the similarity among all inhibitory neurons is reflected in the block structure of the matrices $\boldsymbol{M}$ and $\boldsymbol{S}$. This block structure results in a similar block structure for the matrices for the mean and variance of covariances. Any off-diagonal element of one of the blocks (EE, EI, or II) is representative of the statistics of cross-covariances for this type of neuron pair. If networks and groups of statistically similar neurons are sufficiently large, then—by the self-averaging property [7,40,50,51]—an empirical average across these statistically similar neurons is insensitive to the particular network realization and can be compared to the results for the statistics across network realizations from the theory. We thus computed the empirical mean and variance of the measured covariances for each type of neuron pair (EE, EI, and II), corrected the variances for bias due to finite simulation time (see Appendix K for details), and compared the results to the predictions by Eqs. (22) and (31). The results are displayed in Fig. 4: The top row shows changes in mean covariances with spectral radius of $\boldsymbol{W}$ [via according changes in $\boldsymbol{M}$; see Eq. (22)], and the bottom row shows changes in the variance of covariances with spectral radius of $\boldsymbol{W}$ [via according changes in $\boldsymbol{S}$; see Eq. (31)].

We observe that the order of magnitude of mean and variance are well predicted by Eqs. (22) and (31), which is especially evident for the variances [Figs. 4(d)–4(f)], which span several orders of magnitude. However, there is some quantitative discrepancy between the predictions of the presented linear theory and the results of the simulated spiking network, which is visible in Figs. 4(a)–4(c), indicating that a linear theory cannot fully capture the nonlinear spiking dynamics at high spectral radii, where potential nonrenewal effects of spiking arise [67]. To verify that the discrepancy originates mostly from the linear-response approximation rather than our disorder-average approximations, we plotted the predictions of the linear theory [Eq. (9)] for 20 different network realizations: At small spectral radii, the predicted disorder-average-based mean is equal to the empirical mean of the linear networks, and for large spectral radii, the predicted mean appears to be within the range of two standard deviations around the empirical mean. This shows that the deviations to the spiking network results mostly stem from the linear-response approximation. The remaining difference between the predicted and the empirical mean in linear networks could be explained by the fact that, for high spectral radii, the effective connectivity matrix contributes much more strongly to the noise strength, such that we can no longer disregard its contribution to the noise strength (cf. Fig. 3) and averaging over $\boldsymbol{W}$ and $\boldsymbol{D}$ separately is no longer feasible.

IV. DISCUSSION

In this study, we introduce theoretical tools based on statistical physics of disordered systems to investigate the role of heterogeneous network connectivity in shaping the coordination structure in neural networks. While the presented methods are applicable to arbitrary independent connectivity statistics, for illustration we focus our analysis on the prototypical network model for cortical dynamics by Brunel [42], which is a spiking network of randomly connected excitatory and inhibitory leaky integrate-and-fire neurons receiving uncorrelated external Poisson input. This model has been extensively studied before using mean-field and linear-response methods to understand neuronal spiking statistics such as average firing rates and CVs [30,68] as well as average cross-covariances between populations of neurons [27,37,45,57,58]. In this study, we go beyond the population level and introduce tools from field theory of disordered systems to study the heterogeneity of activity across individual neurons. We show how to turn a linear-response result on the link between covariances and connectivity [27–31] into a field-theoretic problem using moment-generating functions. Then we apply disorder averages, replica theory, and beyond-mean-field approximations to obtain quantitative predictions for the mean and variance of cross-covariances that take into account the statistics of connectivity, but are independent of individual network realizations. We show that this theory can faithfully predict the statistics of cross-covariances of spiking leaky integrate-and-fire networks across the whole linearly stable regime. In doing this, we fixed the statistics of individual neurons according to their theoretical prediction and showed that this one working point, defined by the firing rates of all neurons in the network, can correspond to very distinct correlations structures. Furthermore, we demonstrate that while the heterogeneity in single-neuron activities directly impacts the statistics of neuronal autocovariances, it does not have a sizable impact on the heterogeneity in cross-covariances (cf. Fig. 9). The latter heterogeneity is determined by the heterogeneity in neuronal couplings, quantified by the spectral radius of effective connectivity bulk eigenvalues.

Technically, by employing linear-response theory, we study two systems: the spiking leaky integrate-and-fire network and a network of linear rate neurons. We derive a procedure to set the external input noise of the linear model in such a way that the covariance statistics of the spiking network and the linear network match quantitatively.

This way, the autocovariances are fixed to values determined by single-neuron firing rates and CVs, as predicted by renewal theory for spike trains. Consequently, autocovariances remain finite in the matched rate network even when approaching the point of linear instability. This is achieved by reducing external input fluctuations to account for the increased intrinsically generated fluctuations when increasing the heterogeneity in network connectivity. As a result, also neuronal cross-covariances remain finite close to linear instability. The variance of cross-covariances nevertheless displays a residual divergence, which is why, within the linear regime, mean cross-covariances only vary mildly, while the variance of cross-covariances spans many orders of magnitude when changing the spectral radius of bulk connectivity eigenvalues.

The methods presented here are restricted to the linearly stable network regime, usually referred to as the asynchronous irregular state of the Brunel model [42]. We show that, while mean covariances are low in this state [5,34,36], individual cross-covariances between pairs of neurons can still be large, reflected by the large variance of cross-covariances in strongly heterogeneous network settings. Linear stability can, for example, be realized in excitatory-inhibitory networks if the overall recurrent feedback in the network is inhibition dominated or only marginally positive [37] and if synaptic amplitudes are not too strong. Previous work [67,69] has shown that the here-considered model transitions to a different asynchronous activity state if synaptic amplitudes become larger. This state, however, is not well described by linear-response theory, as slow network fluctuations and nontrivial spike-train autocorrelations emerge, causing deviations from the renewal assumptions on spike trains used here. Note that such slow network fluctuations have not been observed in previous studies on spontaneous activity in macaque motor cortex [7,8] and mouse visual cortex [41] that employed first results of the more general theoretical approach presented here to explain experimentally observed features, such as the large dispersion of covariances, long-range neuronal coordination, a rich repertoire of timescales, and low-dimensional activity. These studies relied on Wick's theorem to calculate the variance of covariances, which is, however, restricted to linear systems. Here we instead employ a more general replica approach that can be straightforwardly applied to nonlinear rate models [52], as extensively studied in the recent theoretical neuroscience literature [65,70–76]. Importantly, the replica theory reveals in a systematic manner that the variance of covariances is an observable that is $\mathcal{O}(1/N)$ in the network size and requires beyond-mean-field methods to be computed. In mean-field or saddle-point approximation, the replica coupling term that yields the nontrivial variance of covariances vanishes. We here calculate the next-to-leading-order Gaussian fluctuations around saddle points that yield good quantitative results across the whole linear regime. The fact that the linear rate model captures the covariance statistics of the spiking leaky integrate-and-fire model further shows that the presented results on the link between connectivity and covariances do not depend on model details and are generally valid in the linear regime, which enables applications to experimental data [7,8,41].

In this paper, we focus on intrinsic mechanisms for heterogeneity and study the first- and second-order statistics of network connectivity. The formalism can be applied to any network topology, as arbitrary connectivity structures can be encoded in the mean and variance matrices that are the central objects of the theory. The results of the formalism are particularly useful for comparing covariance statistics within a single network with groups of statistically equivalent neurons sharing the same connectivity statistics, such as different cell types within neural circuits. Such statistical equivalence imposes symmetries on the structure of the matrices $\boldsymbol{M}$ and $\boldsymbol{S}$ that encode the first- and second-order connectivity statistics. These symmetries allow a dimensionality reduction of the problem and, by the self-averaging property [7,40,50,51], the comparison of theoretical results to empirical averages over covariances of different pairs of neurons within a single circuit. Here we study a network with two populations, excitatory and inhibitory, leading to a block structure in both $\boldsymbol{M}$ and $\boldsymbol{S}$. This example has been chosen as the simplest but relevant setting that goes beyond the fully homogeneous random network in Dahmen *et al.* [7] with a correspondingly simpler homogeneous theory. Extensions to more populations or neuron clusters as well as more complex population-specific connectivity statistics are straightforward by adding more blocks to $\boldsymbol{M}$ and $\boldsymbol{S}$. Another application of the here-derived theory to a more complex scenario, including interneuron distance dependence of connection statistics, has been studied (without derivation) by Dahmen *et al.* [8]. Notably, in our theory we assume that connection weights are independently drawn across different neuron pairs from an arbitrarily complex probability distribution with finite cumulants. The focus on mean M_{ij} and variance S_{ij} of this distribution is justified as long as connection weights scale at least as $\mathcal{O}(1/\sqrt{N})$, a scaling that is often employed to preserve fluctuations in the limit of large networks [34,55]. In this scaling, effects of higher-order-connectivity cumulants are suppressed by the typically large network size, and the Gaussian approximation of the connectivity yields accurate results for network dynamics, as here demonstrated on the example of excitatory-inhibitory networks with population-specific connectivity statistics comprising sparseness (Bernoulli distribution) in addition to distributed (Gaussian) synaptic amplitudes of existing connections. Generalization of dynamic mean-field methods to heavy-tailed connectivity, which cannot be expanded in cumulants, has been proposed for studying single-neuron activity statistics [77,78]. A similar approach may potentially be combined with the methods presented here to investigate cross-covariances. Furthermore, extensions to correlated connection weights, reflecting an over- or under-representation of reciprocal, convergent, divergent, and chain motifs, have been proposed in Ref. [41].

In addition to network connectivity, external inputs can be correlated and heterogeneous and thereby cause heterogeneity in covariances of local circuits. Previous works have shown that external inputs can have a strong impact on local covariances, especially in the limit of infinite network size [36,58,79,80]. For biologically realistic network sizes of local circuits, intrinsically generated covariances via local recurrent activity reverberations, however, make up a substantial

contribution to cross-neuronal coordination [7,57]. This contribution is explainable with the here-presented methods. In general, more research is required to decipher the precise interplay between intrinsic heterogeneity and external inputs to arrive at a complete picture for the mechanistic origin of heterogeneous covariance structures in local circuits.

Many previous studies have linked connectivity and dynamics on an average level, taking into account particular connection pathways between neural populations [19], clustering [81,82], or the spatial dependence of connections [83–85]. In contrast, we here focus on heterogeneity as a key feature of neural network connectivity and show that it yields a wealth of complex coordination patterns that are progressively becoming experimentally accessible via recent advances in measurement techniques [15]. Our theoretical framework to systematically incorporate structural heterogeneity and predict dynamical heterogeneity in biologically plausible neural network models enables the use of this experimental knowledge about neural systems. Likewise, the current framework can be used for the inverse problem of inferring network properties from measured covariances, as we demonstrated in Ref. [7] for the spectral radius in homogeneous random networks (see Appendix M for an extension to E-I networks) and in Ref. [8] for long-range, multisynaptic interactions in networks with spatially organized connectivity. Our work thus opens new avenues for the interpretation of data on network structure and dynamics and proposes a change of focus from population-averaged observables to higher-order statistics that uncover the central role of heterogeneity in biological networks.

All code and data to reproduce the simulations and figures of this study are publicly available under the Ref. [86].

ACKNOWLEDGMENTS

This work was partially supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 945539 (Human Brain Project SGA3) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Grant No. 368482240/GRK2416. Open-access publication is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant No. 491111487. We are grateful to our colleagues in the NEST developer community for continuous collaboration. All network simulations were carried out with NEST [87]. We thank Hannah Bos for the initial numerical implementation of the CVs.

APPENDIX A: NETWORK MODEL AND NEST SIMULATION

We simulate networks of leaky integrate-and-fire neuron models, where the subthreshold dynamics of the membrane potential V_i of neuron i is given by

$$\tau_{\mathrm{m}}\frac{\mathrm{d}V_i(t)}{\mathrm{d}t} = -V_i(t) + RI_i(t), \tag{A1}$$

with total input current $I_i(t)$ that consists of recurrent input via connections with strength J_{ij} and delay d as well as external input:

$$RI_i(t) = \tau_{\mathrm{m}}\left(\sum_j J_{ij}s_j(t-d) + js_{\mathrm{ext,E}}(t) + gjs_{\mathrm{ext,I}}(t) + \frac{I_{\mathrm{ext}}}{C}\right). \tag{A2}$$

Synaptic currents are instantaneous without synaptic filtering [88,89]. The external input is decomposed into a constant current I_{ext} and Poisson spike trains $s_{\mathrm{ext,E}}(t)$ of rate $\nu_{\mathrm{ext,E}}$ and $s_{\mathrm{ext,I}}(t)$ of rate $\nu_{\mathrm{ext,I}}$ that affect neurons with excitatory weight j and inhibitory weight gj, respectively. R and C denote the membrane resistance and capacitance, respectively. More information on the model parameters and their values can be found in Tables I and II.

We consider random sparse connectivity $\boldsymbol{J}$, with connection probability 10 %. Sparse connections are realized with a fixed excitatory indegree $K_{\mathrm{E}} = 800$ and inhibitory indegree $K_{\mathrm{I}} = 200$, with potential self-connections and prohibiting multiple connections between the same pair of neurons. To compensate for the imbalance in excitatory and inhibitory neuron count, we scale the strengths of existing inhibitory connections with respect to excitatory ones by a factor $g = -6$ to obtain an asynchronous irregular dynamic regime [42]. We distribute synaptic amplitudes of existing connections according to population-specific normal distributions $j_{\mathrm{E}} \propto \mathcal{N}(j, 0.2j)$ and $j_{\mathrm{I}} \propto \mathcal{N}(gj, 0.2j)$. The parameter j determining the overall scale of connection strength is varied to modify the overall heterogeneity of connections as measured by the variance

$$\mathrm{Var}(J_{ij}) \propto j^2 \tag{A3}$$

and thereby the spectral radius of bulk connectivity eigenvalues (Table II).

APPENDIX B: TIME-LAG-INTEGRATED COVARIANCES

The cross-covariance function of two stochastic zero-mean processes $x_i(t)$ and $x_j(t)$ is defined as

$$C_{ij}(s,t) = \langle x_i(s)x_j(t)\rangle,$$

where the average is over the ensemble of realizations of the processes. If the stochastic processes are stationary, the cross-covariance function solely depends on the time lag $\tau = t - s$:

$$C_{ij}(\tau) = \langle x_i(s)x_j(s+\tau)\rangle.$$

Here we are considering the time-lag-integrated covariances, as they can be linked to the experimentally accessible spike-count covariances [6,7],

$$\begin{aligned} C_{ij} &\coloneqq \int_{-\infty}^{\infty} C_{ij}(\tau)\mathrm{d}\tau \\ &= \lim_{T\to\infty}\frac{1}{T}(\langle n_i n_j\rangle - \langle n_i\rangle\langle n_j\rangle), \end{aligned}$$

which can be interpreted as a zero-frequency Fourier transform. The Wiener-Khinchin theorem (Appendix C) allows expressing the time-lag-integrated covariances in terms of the time series's Fourier components $X_i(\omega)$ at frequency zero,

$$C_{ij} = \langle X_i(0)X_j(0)\rangle. \tag{B1}$$

TABLE I. Parameters used for NEST simulations and subsequent analysis.

Network parameters		
Neuron type	`iaf_psc_delta`	
Synapse type	`static_synapse`	
Connection rule	`fixed_indegree`	
autapses	`True`	Connections of a neuron to itself
multapses	`False`	Multiple connections between a pair of neurons
N_E	8000	Number of excitatory neurons
N_I	2000	Number of inhibitory neurons
K_E	800	Number of excitatory inputs
K_I	200	Number of inhibitory inputs
C	1 pF	Membrane capacitance
τ_m	20 ms	Membrane time constant
τ_r	2 ms	Refractory period
V_{th}	15 mV	Relative threshold voltage
d	1 ms	Synaptic delay
j	[0.04, 0.38] mV	Excitatory synaptic weight
g	-6	Ratio of inhibitory to excitatory weight
σ_j	20% of j_E	Std of Gaussian distribution of E and I weights
I_{ext}	[5, 125] pA	External DC current
$\nu_{ext,E}$	[800.73, 315049.84] Hz	Rate of external excitatory Poisson noise
$\nu_{ext,I}$	[640.42, 572214.84] Hz	Rate of external inhibitory Poisson noise
Simulation parameters		
dt	0.1 ms	Simulation step size
t_{sim}	10 000 000 ms	Simulation time
Analysis parameters		
T	1000 ms	Bin width for calculating spike-count correlations
T_{init}	1000 ms	Initialization time

APPENDIX C: WIENER-KHINCHIN THEOREM

Here in parts we follow the book by Gardiner [90]. Let $x(t)$ and $y(t)$ be stochastic, stationary processes. Stationary means that for any n-tuple $(t_1, t_2, \ldots, t_n)$ of time points and any real number u the samples $x(t_1), \ldots, x(t_n)$ follow the same distribution as the samples $x(t_1 + u), \ldots, x(t_n + u)$ [91]. Consequently, we may define a raw correlation function as

$$c(\tau) = \langle x(t)y(t+\tau)\rangle,$$

which, due to the assumption of stationarity, does not depend on the time t. The average is over the ensemble of realizations of the processes. If the Fourier transforms of x and y exist, we may calculate the ensemble average over $X(\omega)$ and $Y(\omega)$ as

$$\begin{aligned}\langle X(\omega)Y(\omega')\rangle &= \int \mathrm{d}t\, e^{-\mathrm{i}\omega t}\int \mathrm{d}t'\, e^{-\mathrm{i}\omega' t'}\langle x(t)y(t')\rangle \\ &\overset{\text{subst. } t'=t+\tau}{=} \int \mathrm{d}t\, e^{-\mathrm{i}\omega t}\int \mathrm{d}\tau\, e^{-\mathrm{i}\omega'(t+\tau)} \\ &\quad\times \langle x(t)y(t+\tau)\rangle \\ &= \int \mathrm{d}t\, e^{-\mathrm{i}(\omega+\omega')t}\int \mathrm{d}\tau\, e^{-\mathrm{i}\omega'\tau}\langle x(t)y(t+\tau)\rangle \\ &= 2\pi\delta(\omega+\omega')\int \mathrm{d}\tau\, e^{-\mathrm{i}\omega'\tau}\, c(\tau) \\ &= 2\pi\delta(\omega+\omega')C(\omega'), \end{aligned} \tag{C1}$$

TABLE II. Parameters adjusted for setting different spectral radii while keeping the firing rate constant: The spectral radius of the connectivity is set by choosing different synaptic strengths j of connections. The synaptic strength not only affects the mean connectivity but also the variance of connections, i.e., the heterogeneity of the network, that determines the spectral radius [see Eq. (A3)]. The parameters of the external inputs, which model the total excitatory and inhibitory input from external populations of neurons, are furthermore adjusted to maintain a constant firing rate across different spectral radii. This is done to isolate effects of changing spectral radii on correlations from effects of changing firing rates (see Appendix E and Fig. 6). For small spectral radii and thus weak recurrent input, strong external input is needed to drive the network to moderate firing rates, while for large spectral radii and strong recurrent input, only weak external input is needed for moderate firing rates.

r	0.10	0.20	0.29	0.39	0.49	0.60	0.70	0.79	0.86	0.90
j (mV)	0.04	0.08	0.12	0.16	0.2	0.25	0.29	0.33	0.36	0.38
I_{ext} (pA)	125.0	65.0	40.0	25.0	20.0	15.0	10.0	8.0	6.0	5.0
$\nu_{ext,E}$ (Hz)	315049.84	35406.98	27510.16	32862.34	13335.56	4292.70	6393.05	2149.08	1593.05	800.73
$\nu_{ext,I}$ (Hz)	572214.84	139878.53	58597.12	29923.17	17262.46	9063.65	5147.04	2722.54	1360.93	640.42

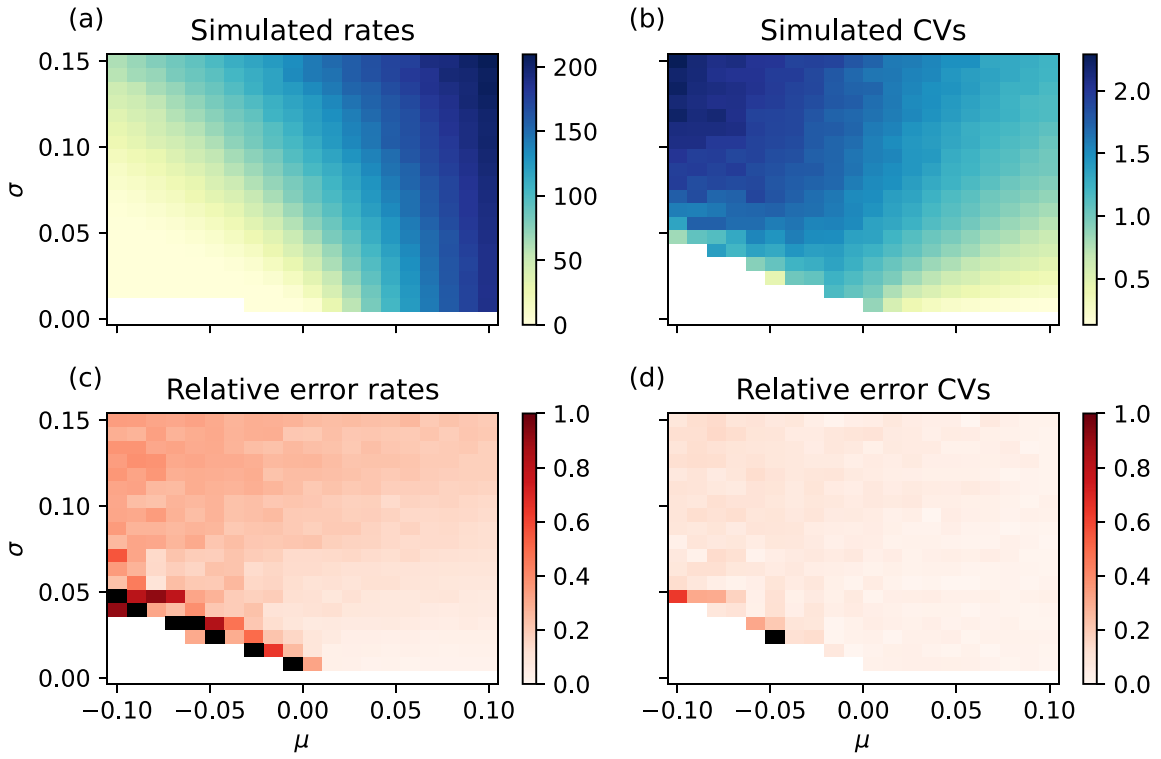

FIG. 5. Validity of firing rate and CV prediction for single LIF neuron with instantaneous synapses provided with two independent Poisson inputs. (a) Simulated rates and (b) simulated CVs at given mean input μ and input variance σ^2. (c) Relative error $\epsilon = |\nu_{\mathrm{sim}} - \nu_{\mathrm{thy}}|/\nu_{\mathrm{sim}}$ of rate and (d) relative error of CV prediction using Eqs. (3) and (4). Black pixels denote error values larger than 1.00.

where we used the identity $\int \mathrm{d}t\, e^{-\mathrm{i}\omega t} = 2\pi\delta(\omega)$ which follows from $\frac{1}{2\pi}\int \mathrm{d}\omega\, e^{\mathrm{i}\omega t} 2\pi\delta(\omega) = 1$, so that $2\pi\delta(\omega)$ is the Fourier transform of the constant function and vice versa. Equation (C1) states that the cross spectrum between two stationary processes vanishes except at those frequencies $\omega = -\omega'$, where it is proportional to a δ distribution times the Fourier transform of the autocorrelation function.

APPENDIX D: VALIDITY OF THEORETICAL PREDICTIONS

In this section, we discuss the conditions under which the theory and simulation described in this paper yield the same results. There are several factors to consider: the limits of the theory we built upon, the limitations of the newly presented theory, and the simulation's constraints.

The estimation of covariances presented in this paper relies on the proper estimation of firing rates and CVs, for which we employ Eqs. (3) and (4) [42], which rely on a diffusion approximation that substitutes spiking input by uncorrelated white noise input. Due to this approximation, these formulas have their own limitations, and they do not yield good estimates in all parameter regimes, as shown in Fig. 5 and Figs. 6(a) and 6(b) for a single neuron receiving Poisson input. Note that a correction due to the finite amplitude of spiking input could in principle be accounted for [92]; the diffusion approximation was shown to typically overestimate firing rates, as can be seen, for example, in Fig. 2(a). In a network context, the diffusion approximation furthermore neglects any nontrivial structure of the autocorrelation of inputs, as well as their cross-correlation structure. Apart from an offset, Fig. 2(a) also shows a clustering of inhibitory firing rates in theory. The difference between the two clusters can be traced back to the presence or absence of self-connections: Firing rates of neurons with self-connections are more in line with simulated rates. This fact, however, does not seem to generalize as networks with other spectral radii do not show any apparent clustering in firing rate predictions (Fig. 7). Because the estimates for firing rates and CVs are used to calculate the effective connectivity matrix and noise strength, a poor estimate has a direct impact on the covariance estimation. Furthermore, the quality of the rate estimates affects how closely the simulated network matches its analytical counterpart due to the way we set the parameters for the simulation: We fix the mean and variance of the single-neuron input, and therefore their firing rates $\boldsymbol{\nu}_{\mathrm{set}}$, and adjust the external input to set the spectral radius r, which we estimate using the result of Rajan and Abbott [66] for random Bernoulli E-I networks,

$$r_{\mathrm{set}} = \sqrt{w^2_{\mathrm{eff,E}}(\boldsymbol{\nu}_{\mathrm{set}})p(1-p)N_{\mathrm{E}} + w^2_{\mathrm{eff,I}}(\boldsymbol{\nu}_{\mathrm{set}})p(1-p)N_{\mathrm{I}}},$$

with connection probability p. The effective weights $w_{\mathrm{eff,E}}(\boldsymbol{\nu})$, $w_{\mathrm{eff,I}}(\boldsymbol{\nu})$ are computed using Eq. (6). Once we simulate the network, we can measure the firing rates, extract the connectivity matrix, and compute the effective connectivity matrix realized in the simulation. Its largest eigenvalue determines the spectral radius r_{sim}. A comparison of r_{set} and r_{sim} is shown in Fig. 6(c). They do not coincide perfectly, which is a direct result of the unreliable estimation of the firing rates, which are slightly overestimated by the theory [see Figs. 2(a), 6(a), 7(a),

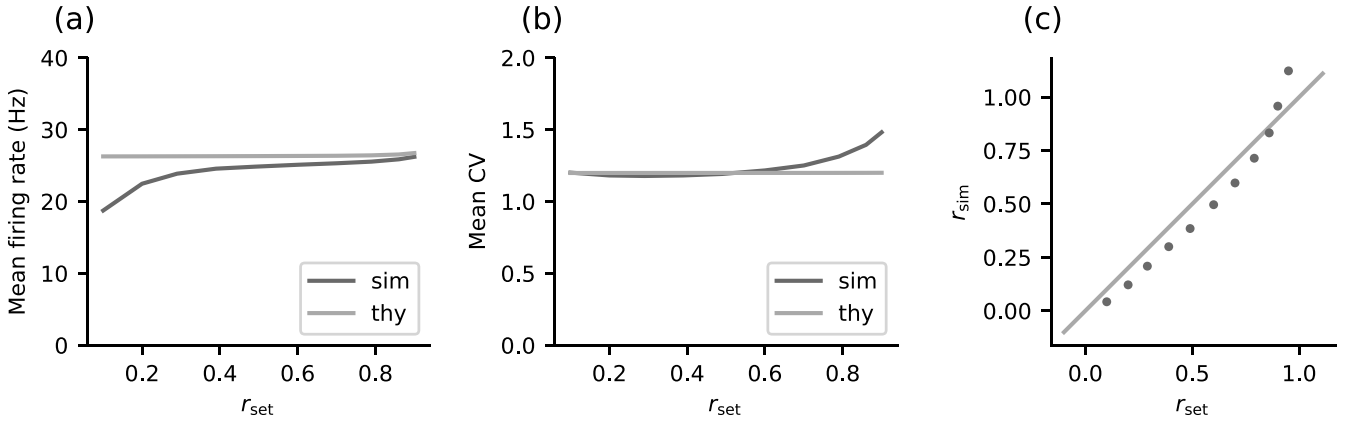

FIG. 6. Prediction accuracy of spiking network simulation properties. Predicted and measured (a) mean firing rates and (b) mean CVs at different spectral radii. (c) Set spectral radius r_{set} vs measured spectral radius r_{sim}. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

7(d), and 7(g)]. To make sure the simulated network is always in a linearly stable regime, we restrict our analysis to spectral radii $r_{set} \leqslant 0.90$, for which $r_{sim} \leqslant 0.94$.

We estimate the noise strength $\boldsymbol{D}$ by computing the variances using Eq. (10), assuming that $\boldsymbol{D}$ is diagonal, and inverting Eq. (9) which yields Eq. (11). First of all, the equation for the variances, Eq. (10), relies on the assumption that the spike trains are well described by renewal processes [48]. Therefore, the noise strength estimate is reliable only if the spike trains are not too bursty. However, even for networks with CV ≈ 1 we observed that for large spectral radii this approach of estimating the noise strength can yield negative values for $\boldsymbol{D}$, which has no physical interpretation. Measuring the covariances in a simulation and inverting Eq. (9) without restricting $\boldsymbol{D}$ to be diagonal yields a matrix that seems to be almost diagonal, shown in Fig. 8(a). Setting the off-diagonal elements to zero and using the result to compute the covariances via Eq. (9), however, reveals that the off-diagonal contribution cannot be neglected [Fig. 8(b)], which means that the external noise sources do have to be correlated to explain the observed covariance. In cases in which the lowest eigenvalue of $\boldsymbol{D}_{\text{full}}$ is negative, we conclude that it is not possible to find a physical linear system (positive-definite $\boldsymbol{D}$) that explains the individual pairwise covariances observed in the spiking network simulation with a large spectral radius. Our theoretical predictions for the mean and variance of cross-covariances, Eqs. (22) and (31), based on $\boldsymbol{D}$ computed with Eq. (11) and its averaged analog, Eq. (32), nevertheless yield quantitatively matching results with respect to the spiking network simulations also in this regime (Fig. 4), because, as we show in Fig. 9, the results only depend on the average of $\boldsymbol{D}$. The theory based on the statistics of connections is therefore found to be more robust than the theory based on individual connectivity realizations. Finally, simulations have one major limitation: their finite simulation time, which results in a biased estimation of the covariances at the single-neuron level. As seen in Figs. 7(b), 7(c), 7(e), 7(f), 7(h), and 7(i), there is some variance in the simulations that is not explained by the theory. This variance is caused by the finite simulation time and vanishes for longer simulations. The relative unexplained variance is larger for small spectral radii, since the firing rates of the neurons are slightly smaller in these networks, leading to poorer estimation of covariances, and overall the covariances are smaller for small spectral radii.

APPENDIX E: DERIVATION OF MOMENT-GENERATING FUNCTION

As discussed in Sec. II, in absence of correlated external input and in the regime of low average covariances, covariances can be understood in linear-response theory [31], where the dynamical equation of LIF neurons describes a model network of Ornstein-Uhlenbeck processes [Eq. (5)]. Grytskyy *et al.* [31] further showed that relation (9) between time-lag-integrated covariances and effective connections is independent of the particular filter kernel $h(t)$ and whether noise is injected in the input or output of neurons. Therefore, we here for simplicity choose Gaussian white noise in the input and $h(t)$ to be an exponential kernel with unit time constant. The stochastic differential equation becomes

$$\mathrm{d}\boldsymbol{x}(t) = -\boldsymbol{x}(t)\mathrm{d}t + \boldsymbol{W}\boldsymbol{x}(t)\mathrm{d}t + \mathrm{d}\boldsymbol{\xi}(t), \tag{E1}$$

with generating functional [7]

$$Z(\boldsymbol{j}) = \int \mathcal{D}\boldsymbol{x} \int \mathcal{D}\widetilde{\boldsymbol{x}} \exp\left[\widetilde{\boldsymbol{x}}^{\mathrm{T}}(\partial_t + 1 - \boldsymbol{W})\boldsymbol{x} + \frac{D}{2}\widetilde{\boldsymbol{x}}^{\mathrm{T}}\widetilde{\boldsymbol{x}} + \boldsymbol{j}^{\mathrm{T}}\boldsymbol{x}\right].$$

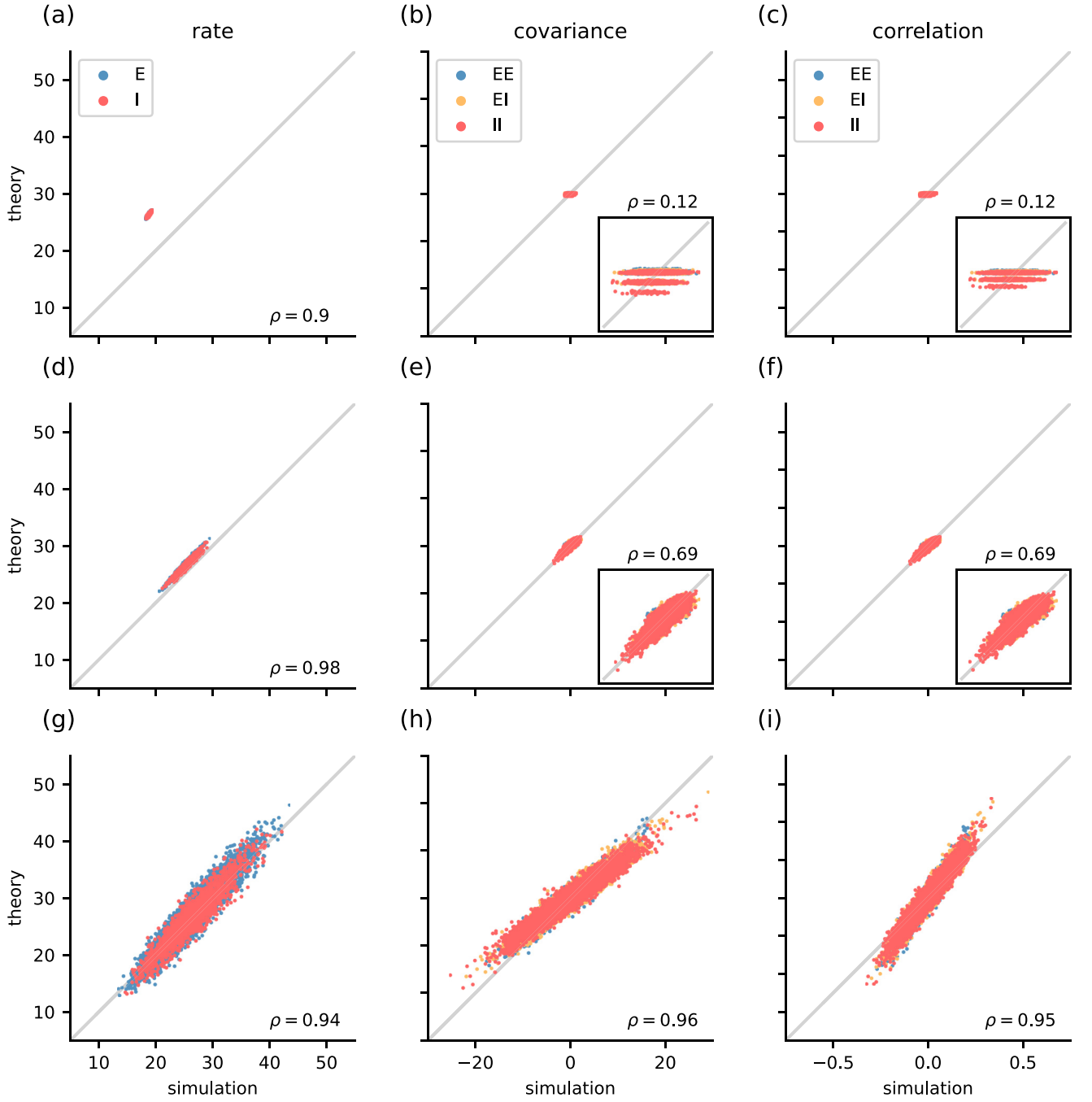

FIG. 7. Simulation results vs theoretical estimates for E-I networks with three different spectral radii [(a)–(c)] $r = 0.10$, [(d)–(f)] $r = 0.49$, and [(g)–(i)] $r = 0.90$. ρ denotes the Pearson correlation coefficient. [(a), (d), (g)] Firing rates. [(b), (e), (h)] Covariances. [(c), (f), (i)] Correlation coefficients. The insets show a closer look at the data points. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

The latter can easily be interpreted in the Fourier domain due to the linearity of Eq. (E1) and the invariance of scalar products under unitary transforms,

$$Z(\boldsymbol{J}) = \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}} \, \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathrm{i}\omega + 1 - \boldsymbol{W})\boldsymbol{X} + \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X}\right],$$

with Fourier-transformed variables denoted by capital letters. The scalar product in the frequency domain reads $\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} = \sum_j \int \mathrm{d}\omega \, \widetilde{X}_j(-\omega)X_j(\omega)$. The generating functional factorizes into generating functions for each frequency ω. As we use Eq. (8) to calculate the time-lag-integrated covariances, we only require the zero-frequency components $\boldsymbol{X}(0)$. In the following, we therefore only discuss zero frequency and omit the frequency argument; i.e., we write $\boldsymbol{X} \equiv \boldsymbol{X}(0)$ and correspondingly for sources

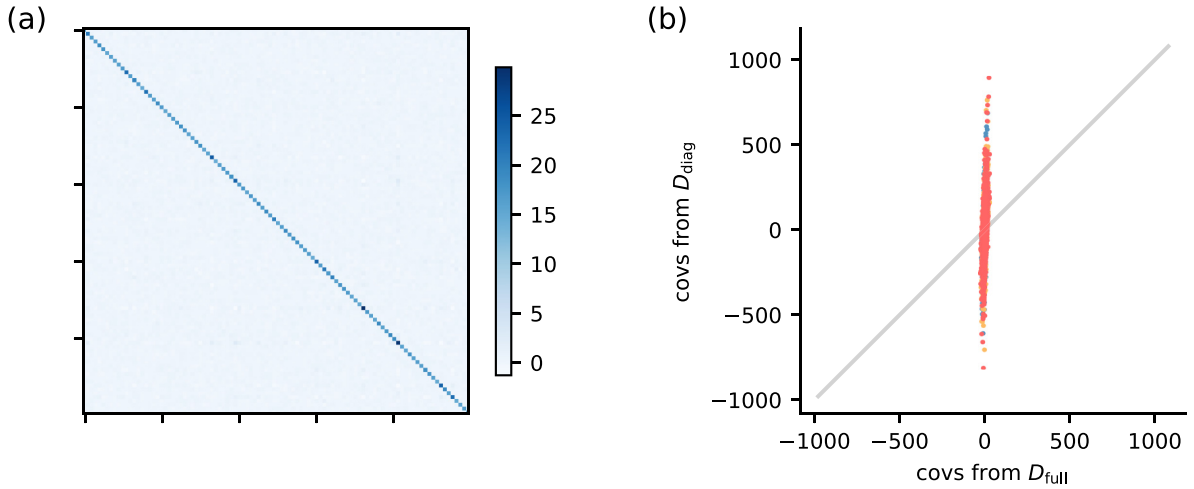

FIG. 8. Noise strength properties. (a) First 100 entries of $\boldsymbol{D}_{\text{full}}$ computed from simulated covariances. (b) Comparison of covariances computed using $\boldsymbol{D}_{\text{full}}$ and using $\boldsymbol{D}_{\text{diag}} = \text{diag}(\boldsymbol{D}_{\text{full}})$. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A for spectral radius $r = 0.9$.

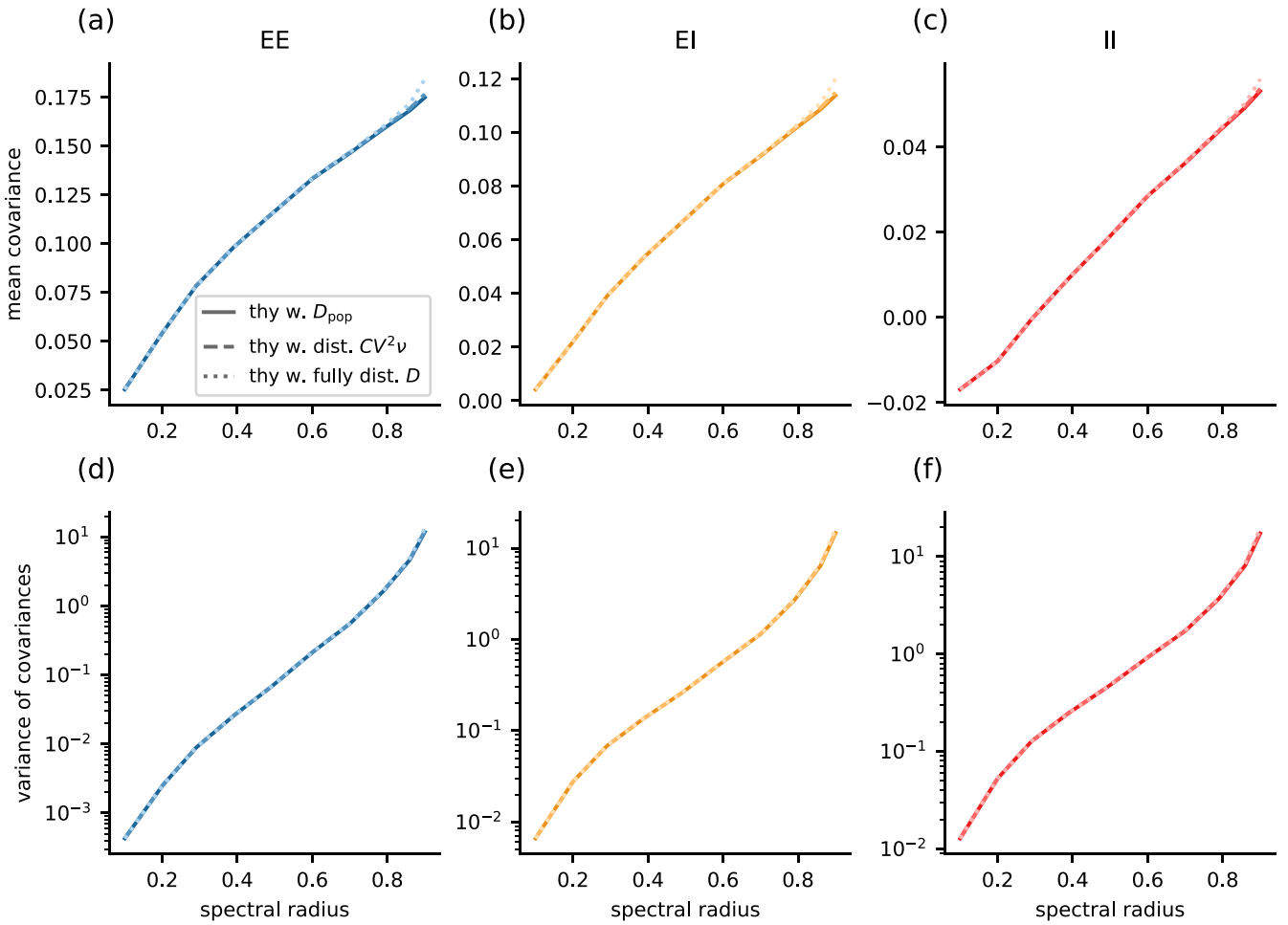

FIG. 9. Dependence of population-resolved covariance statistics on heterogeneity in noise strength $\boldsymbol{D}$. The continuous lines show the results using the realization-independent estimate of $\boldsymbol{D}$ [Eq. (32)]. For the dashed lines, Eq. (32) with the single-neuron resolved estimates of $\text{CV}_i^2\nu_i$ introduced in Sec. II was used, whereas the dotted lines show the results using the full single-neuron resolved estimate of $\boldsymbol{D}$ [Eq. (11)]. [(a)–(c)] Mean cross-covariances. [(d)–(f)] Variance of cross-covariances. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

$\boldsymbol{J}$. After integrating over all nonzero frequencies, we obtain the generating function for zero frequency,

$$Z(\boldsymbol{J}) = \lambda \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}}\, \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})\boldsymbol{X} + \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X}\right], \tag{E2}$$

with the single-frequency scalar product defined as $\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} = \sum_j \widetilde{X}_j\widetilde{X}_j$, integration measures $\int \mathcal{D}\boldsymbol{X} = \prod_j \int_{-\infty}^{\infty} \mathrm{d}X_j$ and $\int \mathcal{D}\widetilde{\boldsymbol{X}} = \prod_j \frac{1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}\widetilde{X}_j$, and normalization prefactor λ. We introduce another source variable $\widetilde{\boldsymbol{J}}$ so that later we can also compute correlators that include $\widetilde{\boldsymbol{X}}$:

$$Z(\boldsymbol{J},\widetilde{\boldsymbol{J}}) = \lambda \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}}\, \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})\boldsymbol{X} + \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X} + \widetilde{\boldsymbol{J}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\right].$$

The Gaussian integrals are solved as follows:

$$\begin{aligned}
Z(\boldsymbol{J},\widetilde{\boldsymbol{J}}) &= \lambda \prod_i \int_{-\infty}^{\infty} \mathrm{d}X_i \prod_j \frac{1}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}\widetilde{X}_j \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})\boldsymbol{X} + \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X} + \widetilde{\boldsymbol{J}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\right] \\
&= \lambda \prod_i \int_{-\infty}^{\infty} \mathrm{d}X_i \prod_j \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\widetilde{X}_j \exp\left[\mathrm{i}\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})\boldsymbol{X} - \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X} + \mathrm{i}\widetilde{\boldsymbol{J}}^{\mathrm{T}}\widetilde{\boldsymbol{X}}\right] \\
&= \lambda\left(\frac{1}{2\pi}\right)^N \prod_i \int_{-\infty}^{\infty} \mathrm{d}X_i \prod_j \int_{-\infty}^{\infty} \mathrm{d}\widetilde{X}_j \exp\left[-\frac{1}{2}\left(\boldsymbol{X}^{\mathrm{T}}, \widetilde{\boldsymbol{X}}^{\mathrm{T}}\right)\begin{pmatrix} 0 & -\mathrm{i}(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix}\begin{pmatrix}\boldsymbol{X}\\ \widetilde{\boldsymbol{X}}\end{pmatrix} + \left(\boldsymbol{J}^{\mathrm{T}}, \mathrm{i}\widetilde{\boldsymbol{J}}^{\mathrm{T}}\right)\begin{pmatrix}\boldsymbol{X}\\ \widetilde{\boldsymbol{X}}\end{pmatrix}\right] \\
&= \lambda\left(\frac{1}{2\pi}\right)^N \sqrt{\frac{(2\pi)^{2N}}{\det\begin{pmatrix} 0 & -\mathrm{i}(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix}}} \exp\left[\frac{1}{2}\left(\boldsymbol{J}^{\mathrm{T}}, \mathrm{i}\widetilde{\boldsymbol{J}}^{\mathrm{T}}\right)\begin{pmatrix} 0 & -\mathrm{i}(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{J}\\ \mathrm{i}\widetilde{\boldsymbol{J}}\end{pmatrix}\right] \\
&= \frac{\lambda}{\sqrt{\det\begin{pmatrix} 0 & -\mathrm{i}(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix}}} \exp\left[\frac{1}{2}\left(\boldsymbol{J}^{\mathrm{T}}, \mathrm{i}\widetilde{\boldsymbol{J}}^{\mathrm{T}}\right)\begin{pmatrix} (\mathbf{1}-\boldsymbol{W})^{-1}D\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right)^{-1} & \mathrm{i}(\mathbf{1}-\boldsymbol{W})^{-1} \\ \mathrm{i}\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right)^{-1} & 0 \end{pmatrix}\begin{pmatrix}\boldsymbol{J}\\ \mathrm{i}\widetilde{\boldsymbol{J}}\end{pmatrix}\right] \\
&= \frac{\lambda}{\sqrt{\det\begin{pmatrix} 0 & -\mathrm{i}(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix}}} \exp\left[\frac{1}{2}\left(\boldsymbol{J}^{\mathrm{T}}, \widetilde{\boldsymbol{J}}^{\mathrm{T}}\right)\begin{pmatrix} (\mathbf{1}-\boldsymbol{W})^{-1}D\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right)^{-1} & -\left(\mathbf{1}-\boldsymbol{W}\right)^{-1} \\ -\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right)^{-1} & 0 \end{pmatrix}\begin{pmatrix}\boldsymbol{J}\\ \widetilde{\boldsymbol{J}}\end{pmatrix}\right].
\end{aligned}$$

The identity matrix and the matrix of ones commute; therefore, we can use $\det\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \det(AD - BC)$, and we get

$$\begin{aligned}
\det\begin{pmatrix} 0 & -\mathrm{i}\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right) \\ -\mathrm{i}(\mathbf{1}-\boldsymbol{W}) & D \end{pmatrix} &= \det\left[-\mathrm{i}^2\left(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}}\right)(\mathbf{1}-\boldsymbol{W})\right] \\
&= \left[\det\left(\mathbf{1}-\boldsymbol{W}\right)\right]^2.
\end{aligned}$$

The normalization condition $Z(\boldsymbol{J}=0) = 1$ yields $\lambda = |\det(1-\boldsymbol{W})|$, and the generating function becomes

$$Z\left(\boldsymbol{J},\widetilde{\boldsymbol{J}}\right) = \exp\left[\frac{1}{2}\left(\boldsymbol{J}^{\mathrm{T}}, \widetilde{\boldsymbol{J}}^{\mathrm{T}}\right)\begin{pmatrix} (\mathbf{1}-\boldsymbol{W})^{-1}D(\mathbf{1}-\boldsymbol{W}^{\mathrm{T}})^{-1} & -(\mathbf{1}-\boldsymbol{W})^{-1} \\ -(\mathbf{1}-\boldsymbol{W})^{-\mathrm{T}} & 0 \end{pmatrix}\begin{pmatrix}\boldsymbol{J}\\ \widetilde{\boldsymbol{J}}\end{pmatrix}\right],$$

or

$$\begin{aligned}
Z\left(\boldsymbol{J},\widetilde{\boldsymbol{J}}=0\right) &= |\det(1-\boldsymbol{W})| \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}}\, \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})\boldsymbol{X} + \frac{D}{2}\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{X}} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X}\right] \\
&= \exp\left(\frac{1}{2}\boldsymbol{J}^{\mathrm{T}}(\mathbf{1}-\boldsymbol{W})^{-1}\boldsymbol{D}(\mathbf{1}-\boldsymbol{W})^{-\mathrm{T}}\boldsymbol{J}\right),
\end{aligned} \tag{E3}$$

respectively. We obtain the time-lag-integrated covariances

$$C_{ij} = \left\langle X_i X_j \right\rangle = \left.\frac{\partial}{\partial J_i}\frac{\partial}{\partial J_j} Z\left(\boldsymbol{J},\widetilde{\boldsymbol{J}}\right)\right|_{\boldsymbol{J},\widetilde{\boldsymbol{J}}=0} = [(\mathbf{1}-\boldsymbol{W})^{-1}\boldsymbol{D}(\mathbf{1}-\boldsymbol{W})^{-\mathrm{T}}]_{ij}.$$

APPENDIX F: SADDLE POINTS AND CORRELATORS OF ACTIVITY FIELDS

The saddle points $\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*$ are given by $\frac{\partial}{\partial Q_i} S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} = 0$ and $\frac{\partial}{\partial \widetilde{Q}_i} S(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} = 0$, which yield

$$\begin{aligned} Q_i^* &= \frac{1}{N} \sum_j \Delta_{ij} \langle X_j X_j \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*}, \\ \widetilde{Q}_i^* &= \frac{1}{2N} \langle \widetilde{X}_i \widetilde{X}_i \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*}, \end{aligned}$$

including correlators evaluated at the saddle point $\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*$. To evaluate them, we need to solve the Gaussian integral in Eq. (17),

$$\widetilde{Z}(\boldsymbol{J}, \widetilde{\boldsymbol{J}}) = \int \mathcal{D}\boldsymbol{X} \int \mathcal{D}\widetilde{\boldsymbol{X}} \, \exp\left[\widetilde{\boldsymbol{X}}^{\mathrm{T}} (\mathbf{1} - \boldsymbol{M}) \boldsymbol{X} + \frac{1}{2} \widetilde{\boldsymbol{X}}^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})] \widetilde{\boldsymbol{X}} + \boldsymbol{X}^{\mathrm{T}} \mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) \boldsymbol{X} + \boldsymbol{J}^{\mathrm{T}} \boldsymbol{X} + \widetilde{\boldsymbol{J}}^{\mathrm{T}} \widetilde{\boldsymbol{X}} \right],$$

where we added the additional source term $\widetilde{\boldsymbol{J}}^{\mathrm{T}} \widetilde{\boldsymbol{X}}$ to allow for the calculation of correlators including $\widetilde{\boldsymbol{X}}$. We can rewrite the equation as

$$\widetilde{Z}(\boldsymbol{J}) = \int \mathcal{D}\boldsymbol{Z} \, \exp\left[-\frac{1}{2} \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{Z} + \boldsymbol{B}^{\mathrm{T}} \boldsymbol{Z} \right] = (2\pi)^{-N} \sqrt{\frac{(2\pi)^{2n}}{\det \boldsymbol{A}}} \exp\left(\frac{1}{2} \boldsymbol{B}^T \boldsymbol{A}^{-1} \boldsymbol{B} \right) = \frac{1}{\sqrt{\det \boldsymbol{A}}} \exp\left(\frac{1}{2} \boldsymbol{B}^T \boldsymbol{A}^{-1} \boldsymbol{B} \right),$$

using

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \widetilde{\boldsymbol{X}} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{J} \\ \mathrm{i}\widetilde{\boldsymbol{J}} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} -2\,\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) & -\mathrm{i}(\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} \\ -\mathrm{i}(\mathbf{1} - \boldsymbol{M}) & [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})] \end{pmatrix},$$

where the prefactor $(2\pi)^{-N}$ comes from the integration measure $\mathcal{D}\widetilde{\boldsymbol{X}}$, such that

$$\begin{aligned} \boldsymbol{A}_{11}^{-1} &= \big\{ -2\,\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) + (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\}^{-1}, \\ \boldsymbol{A}_{12}^{-1} &= \mathrm{i} \big\{ -2\,\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) + (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\}^{-1} (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1}, \\ \boldsymbol{A}_{21}^{-1} &= \mathrm{i} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\{ -2\,\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) + (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\}^{-1}, \\ \boldsymbol{A}_{22}^{-1} &= [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} - [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \\ &\quad \times \big\{ -2\,\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) + (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\}^{-1} (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1}. \end{aligned}$$

Deriving the normalized moment-generating function $Z(\boldsymbol{J}, \widetilde{\boldsymbol{J}}) = \widetilde{Z}(\boldsymbol{J}, \widetilde{\boldsymbol{J}}) / \widetilde{Z}(0, 0)$ twice with respect to $\widetilde{\boldsymbol{J}}$ yields

$$\begin{aligned} \widetilde{\boldsymbol{Q}}^* &= \frac{1}{2N} \big\langle \widetilde{\boldsymbol{X}} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \big\rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \\ &= \frac{1}{2N} A_{22}^{-1} \\ &= \frac{1}{2N} ([\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} - [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\{ -2\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}} \boldsymbol{\Delta}\big) \\ &\quad + (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1} (\mathbf{1} - \boldsymbol{M}) \big\}^{-1} (\mathbf{1} - \boldsymbol{M})^{\mathrm{T}} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})]^{-1}), \end{aligned}$$

which is solved by $\widetilde{\boldsymbol{Q}}^* = 0$. Inserting this result, we find

$$\begin{aligned} \langle \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} &= (\mathbf{1} - \boldsymbol{M})^{-1} [\boldsymbol{D} + \mathrm{diag}(\boldsymbol{Q})] (\mathbf{1} - \boldsymbol{M})^{-\mathrm{T}}, \\ \big\langle \widetilde{\boldsymbol{X}} \boldsymbol{X}^{\mathrm{T}} \big\rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} &= -(\mathbf{1} - \boldsymbol{M})^{-\mathrm{T}}, \\ \langle \widetilde{\boldsymbol{X}} \widetilde{\boldsymbol{X}}^{\mathrm{T}} \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} &= 0. \end{aligned}$$

Inserting the correlators into the saddle-point equations and solving for $\boldsymbol{Q}^*$ yields

$$Q_i^* = \frac{1}{N} \sum_{j,k,l,m} \left(\mathbf{1} - \frac{1}{N} \Delta \cdot \boldsymbol{R}^{\circ 2} \right)^{-1}_{ij} \Delta_{jk} R_{kl} D_{lm} R_{km}, \quad \widetilde{Q}_i^* = 0\,, \tag{F1}$$

with $\boldsymbol{R} = (\mathbf{1} - \boldsymbol{M})^{-1}$.

The saddle point of the auxiliary fields $\boldsymbol{Q}_{XX}, \boldsymbol{Q}_{XY}, \boldsymbol{Q}_{YY}$ in the replica theory are determined by finding the zeros of the first derivative of the action [Eq. (25)]. This yields

$$Q^*_{XX,i} = \frac{1}{N}\sum_j \Delta_{ij}\langle X_j X_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = Q^*_i, \quad Q^*_{YY,i} = \frac{1}{N}\sum_j \Delta_{ij}\langle Y_j Y_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = Q^*_i,$$

$$Q^*_{XY,i} = \frac{1}{N}\sum_j \Delta_{ij}\langle X_j Y_j\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}, \quad \widetilde{Q}^*_{XX,i} = \frac{1}{2N}\langle \widetilde{X}_i\widetilde{X}_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0,$$

$$\widetilde{Q}^*_{YY,i} = \frac{1}{2N}\langle \widetilde{Y}_i\widetilde{Y}_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0, \quad \widetilde{Q}^*_{XY,i} = \frac{1}{N}\langle \widetilde{X}_i\widetilde{Y}_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0,$$

and in a fashion analogous to the derivation above, we find

$$\langle \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = (\mathbf{1}-\boldsymbol{M})^{-1}\mathrm{diag}(\boldsymbol{Q}^*_{XY})(\mathbf{1}-\boldsymbol{M})^{-\mathrm{T}}.$$

Inserting the latter solution into the saddle-point equations again yields a linear self-consistency equation for $\boldsymbol{Q}^*_{XY}$ with the solution

$$Q^*_{XY,i} = 0,$$

such that

$$\langle \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0.$$

APPENDIX G: FLUCTUATIONS AROUND SADDLE POINTS

Here we showcase how to perform a fluctuation expansion of the correlator $\langle X_i Y_i\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}$ around $\boldsymbol{Q}^*_{XY}$ and $\widetilde{\boldsymbol{Q}}^*_{XY}$. Other correlators follow analogously. Following the definition in Eq. (24), the correlator is given by

$$\langle X_i Y_i\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} = \frac{\partial}{\partial J_i}\frac{\partial}{\partial K_i} Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})\Big|_{\boldsymbol{J},\boldsymbol{K}=0}.$$

Now, we expand $Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}$ around the saddle points,

$$Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) \approx Z_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}(\boldsymbol{J},\boldsymbol{K}) + \sum_k \frac{\partial}{\partial Q_{XY,k}} Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})\Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}(Q_{XY,k}-Q^*_{XY,k}) + \sum_k \frac{\partial}{\partial \widetilde{Q}_{XY,k}} Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})\Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\big(\widetilde{Q}_{XY,k}-\widetilde{Q}^*_{XY,k}\big),$$

and use $Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) = \widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})/\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0},\mathbf{0})$ to obtain

$$\frac{\partial}{\partial Q_{XY,k}} Z_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) = \frac{1}{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0},\mathbf{0})}\frac{\partial}{\partial Q_{XY,k}}\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) - \frac{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})}{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0},\mathbf{0})^2}\frac{\partial}{\partial Q_{XY,k}}\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(0,0). \tag{G1}$$

Using the definition of $\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})$ in Eq. (25), its derivative is given by

$$\frac{\partial}{\partial Q_{XY,k}}\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K}) = \int \mathcal{D}\boldsymbol{X}\int \mathcal{D}\widetilde{\boldsymbol{X}}\int \mathcal{D}\boldsymbol{Y}\int \mathcal{D}\widetilde{\boldsymbol{Y}}\,\widetilde{X}_k\widetilde{Y}_k \exp\big[S_{\boldsymbol{Q}_{XX},\widetilde{\boldsymbol{Q}}_{XX}}(\boldsymbol{X},\widetilde{\boldsymbol{X}}) + S_{\boldsymbol{Q}_{YY},\widetilde{\boldsymbol{Q}}_{YY}}(\boldsymbol{Y},\widetilde{\boldsymbol{Y}})$$
$$+ \widetilde{\boldsymbol{X}}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{Q}_{XY})\widetilde{\boldsymbol{Y}} + \boldsymbol{X}^{\mathrm{T}}\mathrm{diag}\big(\widetilde{\boldsymbol{Q}}^{\mathrm{T}}_{XY}\boldsymbol{\Delta}\big)\boldsymbol{Y} + \boldsymbol{J}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{K}^{\mathrm{T}}\boldsymbol{Y}\big],$$

such that normalizing and evaluating at the saddle point and for zero sources yields

$$\frac{\frac{\partial}{\partial J_i}\frac{\partial}{\partial K_i}\frac{\partial}{\partial Q_{XY,k}}\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})|_{\boldsymbol{J},\boldsymbol{K}=0}}{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0},\mathbf{0})}\Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = \langle X_i Y_i \widetilde{X}_k\widetilde{Y}_k\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}.$$

The second term on the right-hand side of Eq. (G1) vanishes at the saddle point and for $\boldsymbol{J}=\boldsymbol{K}=0$

$$\frac{\partial}{\partial J_i}\frac{\partial}{\partial K_i}\frac{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\boldsymbol{J},\boldsymbol{K})}{\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(\mathbf{0},\mathbf{0})^2}\frac{\partial}{\partial Q_{XY,k}}\widetilde{Z}_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}(0,0)\Big|_{\boldsymbol{J},\boldsymbol{K}=0}\Big|_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = \langle X_i Y_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle \widetilde{X}_k\widetilde{Y}_k\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*} = 0.$$

The derivative with respect to $\widetilde{Q}_{XY,k}$ can be computed analogously, with $\widetilde{X}_k\widetilde{Y}_k$ replaced by $\sum_l \Delta_{kl}X_lY_l$. Therefore, the first-order expansion in the replica coupling term reads

$$\langle X_i Y_i\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}} = \underbrace{\langle X_i Y_i\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}}_{=0} + \sum_k \langle X_i Y_i \widetilde{X}_k\widetilde{Y}_k\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}(Q_{XY,k}-Q^*_{XY,k}) + \sum_{k,l}\Delta_{kl}\langle X_i Y_i X_l Y_l\rangle_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\big(\widetilde{Q}_{XY,k}-\widetilde{Q}^*_{XY,k}\big), \tag{G2}$$

which we use in Eq. (28).

APPENDIX H: CORRELATORS OF AUXILIARY FIELDS

We consider the Gaussian approximation of $p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})$ with

$$\mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) = \mathcal{S}(\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*) + \frac{1}{2}(\delta \boldsymbol{Q}_{XY}, \delta \widetilde{\boldsymbol{Q}}_{XY}) \boldsymbol{\mathcal{S}}^{(2)} \begin{pmatrix} \delta \boldsymbol{Q}_{XY} \\ \delta \widetilde{\boldsymbol{Q}}_{XY} \end{pmatrix},$$

where $\boldsymbol{\mathcal{S}}^{(2)}$ contains the second derivatives with respect to the auxiliary fields,

$$\boldsymbol{\mathcal{S}}^{(2)} = \begin{pmatrix} \frac{\partial \mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})}{\partial \boldsymbol{Q}_{XY} \partial \boldsymbol{Q}_{XY}}\Big|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} & \frac{\partial \mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})}{\partial \boldsymbol{Q}_{XY} \partial \widetilde{\boldsymbol{Q}}_{XY}}\Big|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \\ \frac{\partial \mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})}{\partial \widetilde{\boldsymbol{Q}}_{XY} \partial \boldsymbol{Q}_{XY}}\Big|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} & \frac{\partial \mathcal{S}(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}})}{\partial \widetilde{\boldsymbol{Q}}_{XY} \partial \widetilde{\boldsymbol{Q}}_{XY}}\Big|_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \end{pmatrix},$$

with

$$\begin{aligned}
\mathcal{S}^{(2)}_{11,ij} &= 0, \\
\mathcal{S}^{(2)}_{12,ij} &= N\delta_{ij} - \sum_k \Delta_{jk} \big\langle \widetilde{X}_i \widetilde{Y}_i X_k Y_k \big\rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} = N\delta_{ij} - \sum_k \Delta_{jk} R^2_{ki}, \\
\mathcal{S}^{(2)}_{21,ij} &= N\delta_{ij} - \sum_k \Delta_{ik} \big\langle \widetilde{X}_j \widetilde{Y}_j X_k Y_k \big\rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} = N\delta_{ij} - \sum_k \Delta_{ik} R^2_{kj}, \\
\mathcal{S}^{(2)}_{22,ij} &= -\sum_{k,l} \Delta_{ik} \Delta_{jl} \langle X_k Y_k X_l Y_l \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} = -\sum_{k,l} \Delta_{ik} \Delta_{jl} \langle X_k X_l \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \langle Y_k Y_l \rangle_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*}.
\end{aligned}$$

Using

$$\boldsymbol{\mathcal{S}}^{(2)} = \begin{pmatrix} 0 & \boldsymbol{\mathcal{S}}^{(2)\mathrm{T}}_{21} \\ \boldsymbol{\mathcal{S}}^{(2)}_{21} & \boldsymbol{\mathcal{S}}^{(2)}_{22} \end{pmatrix}, \quad \big(\boldsymbol{\mathcal{S}}^{(2)}\big)^{-1} = \begin{pmatrix} -\boldsymbol{\mathcal{S}}^{(2)-1}_{21} \boldsymbol{\mathcal{S}}^{(2)}_{22} \boldsymbol{\mathcal{S}}^{(2)-\mathrm{T}}_{21} & \boldsymbol{\mathcal{S}}^{(2)-1}_{21} \\ \boldsymbol{\mathcal{S}}^{(2)-\mathrm{T}}_{21} & 0 \end{pmatrix},$$

we find

$$\begin{aligned}
\big\langle \delta \boldsymbol{Q}_{XY} \delta \boldsymbol{Q}^{\mathrm{T}}_{XY} \big\rangle &= -\boldsymbol{\mathcal{S}}^{(2)-1}_{21} \boldsymbol{\mathcal{S}}^{(2)}_{22} \boldsymbol{\mathcal{S}}^{(2)-\mathrm{T}}_{21} \\
&= \frac{1}{N^2} \Big[\mathbf{1} - \frac{1}{N} \boldsymbol{\Delta} \cdot \boldsymbol{R}^{\circ 2} \Big]^{-1} \boldsymbol{\Delta} \langle \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \rangle^{\circ 2}_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \boldsymbol{\Delta}^{\mathrm{T}} \Big[\mathbf{1} - \frac{1}{N} \boldsymbol{\Delta} \cdot \boldsymbol{R}^{\circ 2} \Big]^{-\mathrm{T}} \\
\big\langle \delta \boldsymbol{Q}_{XY} \delta \widetilde{\boldsymbol{Q}}^{\mathrm{T}}_{XY} \big\rangle &= \boldsymbol{\mathcal{S}}^{(2)-1}_{21} = \frac{1}{N} \Big[\mathbf{1} - \frac{1}{N} \boldsymbol{\Delta} \cdot \boldsymbol{R}^{\circ 2} \Big]^{-1} \\
\big\langle \delta \widetilde{\boldsymbol{Q}}_{XY} \delta \widetilde{\boldsymbol{Q}}^{\mathrm{T}}_{XY} \big\rangle &= 0.
\end{aligned}$$

APPENDIX I: DISORDER-AVERAGED VARIANCE OF COVARIANCES

Starting with Eq. (26), we find

$$\begin{aligned}
\big\langle C^2_{ij} \big\rangle_{\boldsymbol{W}} &= \langle \langle X_i X_j \rangle^2 \rangle_{\boldsymbol{W}} \\
&= \int D\boldsymbol{Q} \int D\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_i X_j Y_i Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \\
&= \int D\boldsymbol{Q} \int D\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \big(\langle X_i X_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle Y_i Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} + \langle X_i Y_i \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle X_j Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} + \langle X_i Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle X_j Y_i \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \big) \\
&\approx \langle \langle X_i X_j \rangle \rangle^2_{\boldsymbol{W}} + (1 + \delta_{ij}) \int D\boldsymbol{Q} \int D\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_i Y_i \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle X_j Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \\
&= \langle C_{ij} \rangle^2_{\boldsymbol{W}} + (1 + \delta_{ij}) \int D\boldsymbol{Q} \int D\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_i Y_i \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle X_j Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}}.
\end{aligned}$$

Inserting the derived expressions for the correlators into Eq. (29) yields

$$\begin{aligned}
&\int D\boldsymbol{Q} \int D\widetilde{\boldsymbol{Q}}\, p(\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}) \langle X_i Y_i \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \langle X_j Y_j \rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} \\
&= \sum_{k,l} \big\langle X_i \widetilde{X}_k \big\rangle^2_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \big\langle X_j \widetilde{X}_l \big\rangle^2_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \big\langle \delta Q_{XY,k} \delta Q_{XY,l} \big\rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}} + \sum_{k,l,m} \big\langle X_i \widetilde{X}_k \big\rangle^2_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \Delta_{lm} \big\langle X_j X_m \big\rangle^2_{\boldsymbol{Q}^*, \widetilde{\boldsymbol{Q}}^*} \big\langle \delta Q_{XY,k} \delta \widetilde{Q}_{XY,l} \big\rangle_{\boldsymbol{Q}, \widetilde{\boldsymbol{Q}}}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{k,l,m}\left\langle X_j\widetilde{X}_k\right\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\Delta_{lm}\langle X_iX_m\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left\langle\delta Q_{XY,k}\delta\widetilde{Q}_{XY,l}\right\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}+\sum_{k,l,m,n}\langle X_iX_m\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\langle X_jX_n\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\Delta_{km}\Delta_{ln}\left\langle\delta\widetilde{Q}_{XY,k}\delta\widetilde{Q}_{XY,l}\right\rangle_{\boldsymbol{Q},\widetilde{\boldsymbol{Q}}}\\
&=\sum_{k,l}R^2_{ik}\left\{\frac{1}{N^2}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-1}\boldsymbol{\Delta}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\boldsymbol{\Delta}^{\mathrm{T}}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-\mathrm{T}}\right\}_{kl}R^2_{jl}\\
&+\sum_{k,l,m}R^2_{ik}\Delta_{lm}\langle X_jX_m\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left\{\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-1}\right\}_{kl}+\sum_{k,l,m}R^2_{jk}\Delta_{lm}\langle X_iX_m\rangle^2_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left\{\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-1}\right\}_{kl}\\
&=\left\{\boldsymbol{R}^{\circ 2}\frac{1}{N^2}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-1}\boldsymbol{\Delta}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\boldsymbol{\Delta}^{\mathrm{T}}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-\mathrm{T}}\boldsymbol{R}^{\circ 2\mathrm{T}}\right\}_{ij}\\
&+\left\{\boldsymbol{R}^{\circ 2}\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-1}\boldsymbol{\Delta}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\right\}_{ij}+\left\{\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\boldsymbol{\Delta}^{\mathrm{T}}\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-\mathrm{T}}\boldsymbol{R}^{\circ 2\mathrm{T}}\right\}_{ij}\\
&=\left\{\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\frac{1}{N^2}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\boldsymbol{\Delta}^{\mathrm{T}}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\cdot\boldsymbol{R}^{\circ 2}\right]^{-\mathrm{T}}\boldsymbol{R}^{\circ 2\mathrm{T}}\right\}_{ij}\\
&+\left\{\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\right\}_{ij}+\left\{\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\frac{1}{N}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{R}^{\circ 2\mathrm{T}}\right\}_{ij}\\
&=\left\{\left(\mathbf{1}+\frac{1}{N}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\right)\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left(\mathbf{1}+\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\frac{1}{N}\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{R}^{\circ 2\mathrm{T}}\right)\right\}_{ij}-\left[\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\right]_{ij}\\
&=\left\{\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\right\}_{ij}-\left[\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\right]_{ij}\\
&=\left\{\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\right\}_{ij}-\langle C_{ij}\rangle^2_{\boldsymbol{W}},
\end{aligned}
$$

where we used

$$
\begin{aligned}
\boldsymbol{R}^{\circ 2}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}\right]^{-1}\boldsymbol{\Delta}&=\boldsymbol{R}^{\circ 2}\sum_k\left[\frac{1}{N}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}\right]^k\boldsymbol{\Delta}\\
&=\boldsymbol{R}^{\circ 2}\left[\mathbf{1}+\frac{1}{N}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}+\frac{1}{N^2}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}+\cdots\right]\boldsymbol{\Delta}\\
&=\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\left[\mathbf{1}+\frac{1}{N}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}+\frac{1}{N^2}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}+\cdots\right]\\
&=\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\boldsymbol{\Delta}\right]^{-1}.
\end{aligned}
$$

Finally, we obtain the second moment to leading order,

$$
\left\langle C^2_{ij}\right\rangle_{\boldsymbol{W}}=(1+\delta_{ij})\left\{\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\right\}_{ij}-\delta_{ij}\langle C_{ij}\rangle^2_{\boldsymbol{W}},
$$

and for the covariance we find

$$
\begin{aligned}
\left\langle\delta C^2_{ij}\right\rangle_{\boldsymbol{W}}&=\left\langle C^2_{ij}\right\rangle_{\boldsymbol{W}}-\langle C_{ij}\rangle^2_{\boldsymbol{W}}\\
&=(1+\delta_{ij})\left\{\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-1}\langle\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rangle^{\circ 2}_{\boldsymbol{Q}^*,\widetilde{\boldsymbol{Q}}^*}\left[\mathbf{1}-\frac{1}{N}\boldsymbol{R}^{\circ 2}\cdot\boldsymbol{\Delta}\right]^{-\mathrm{T}}\right\}_{ij}-(1+\delta_{ij})\langle C_{ij}\rangle^2_{\boldsymbol{W}}\\
&\approx(1+\delta_{ij})[(\mathbf{1}-\boldsymbol{S})^{-1}[\boldsymbol{D}+\mathrm{diag}(\boldsymbol{Q})](\mathbf{1}-\boldsymbol{S})^{-\mathrm{T}}]_{ij}-(1+\delta_{ij})\langle C_{ij}\rangle^2_{\boldsymbol{W}}.
\end{aligned}
$$

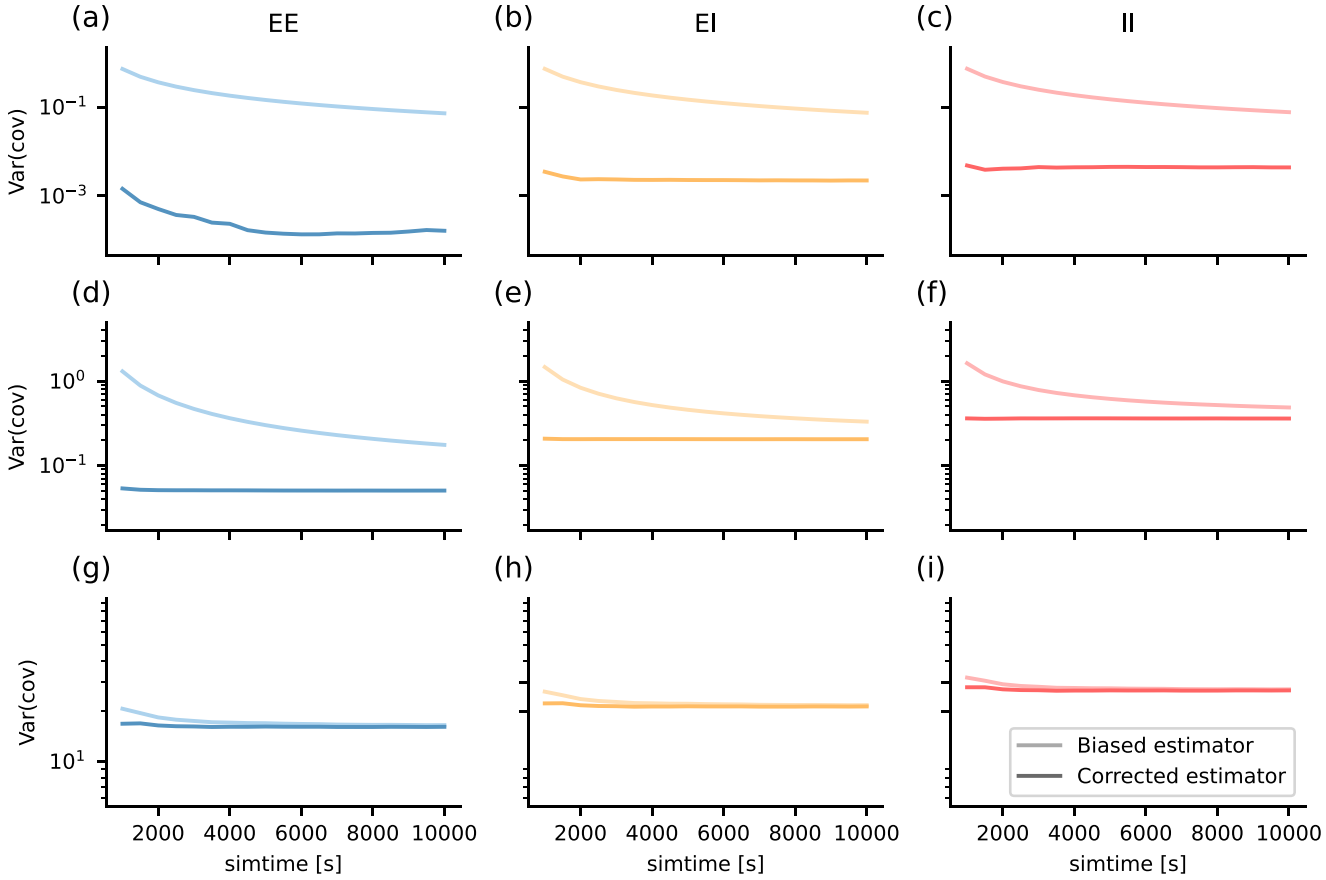

FIG. 10. Bias correction of variance of covariance estimation for different simulation lengths at three different spectral radii r. The light color depicts the biased estimator, the dark color the corrected estimator (K1). [(a)–(c)] $r = 0.10$, [(d)–(f)] $r = 0.49$, and [(g)–(i)] $r = 0.90$. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

APPENDIX J: DEPENDENCE OF POPULATION-RESOLVED COVARIANCE STATISTICS ON HETEROGENEITY IN NOISE STRENGTH D

In this Appendix, we study the dependence of population-resolved covariance statistics on heterogeneity in noise strength. The heterogeneity in noise strength has two sources: variability in CVs and firing rates across neurons, and variability indirectly induced by the connectivity via the matrix $\mathbf{B}$ [Eq. (11)]. We first replace the full single-neuron resolved estimate of $\mathbf{D}$ [Eq. (11)] by an estimate that is independent of the specific connectivity realization in $\mathbf{B}$ (see Eq. (32) but with neuron-resolved firing rates and CVs). Subsequently, we additionally replace the individual neuron firing rates and CVs by their population mean [Eq. (32)]. Both simplifications have hardly any effect on the mean and variance of covariances (Fig. 9), confirming that these statistics are determined by the effective connectivity statistics rather than variability in single-neuron firing statistics.

APPENDIX K: BIAS CORRECTION OF VARIANCE OF COVARIANCES

We utilize Eq. (4) from the supplementary information of Ref. [7] to correct for the bias in the estimation of the variances of covariances due to the finite simulation time. The analogous correction for two populations a and b is given by

$$\delta C_{ab}^2 = \delta \hat{C}_{ab}^2 - \frac{\langle A_a \rangle \langle A_b \rangle - \langle C_{ab} \rangle^2}{N+1}, \tag{K1}$$

with the biased estimator of the variance of cross-covariances $\delta \hat{C}_{ab}^2$, mean autocovariance $\langle A_a \rangle$, mean cross-covariance $\langle C_{ab} \rangle$, and the number of bins the spike trains are divided into, $N = T_{\mathrm{sim}}/T_{\mathrm{bin}}$. Figure 10 illustrates that after a simulation time of 10 000 s, the corrected estimator converges to a fixed value while the biased estimator does not, especially for smaller spectral radii. In contrast, the mean covariance estimator converges much faster for all spectral radii, as shown in Fig. 11.

APPENDIX L: NUMERICAL IMPLEMENTATION OF CVs

For computing the theoretical prediction of the CVs, we make use of the equation found in Appendix A.1 in Ref. [42], which in our units reads

$$\mathrm{CV}^2 = 2\pi(\tau_{\mathrm{m}}\nu)^2 \int_{y_{\mathrm{r}}}^{y_{\mathrm{th}}} \mathrm{d}x\, \mathrm{e}^{x^2} \int_{-\infty}^{x} \mathrm{d}s\, \mathrm{e}^{s^2} [1 + \mathrm{erf}(s)]^2. \tag{L1}$$

However, a naive implementation of Eq. (L1) is numerically unstable due to the diverging integrals. To proceed, we rewrite

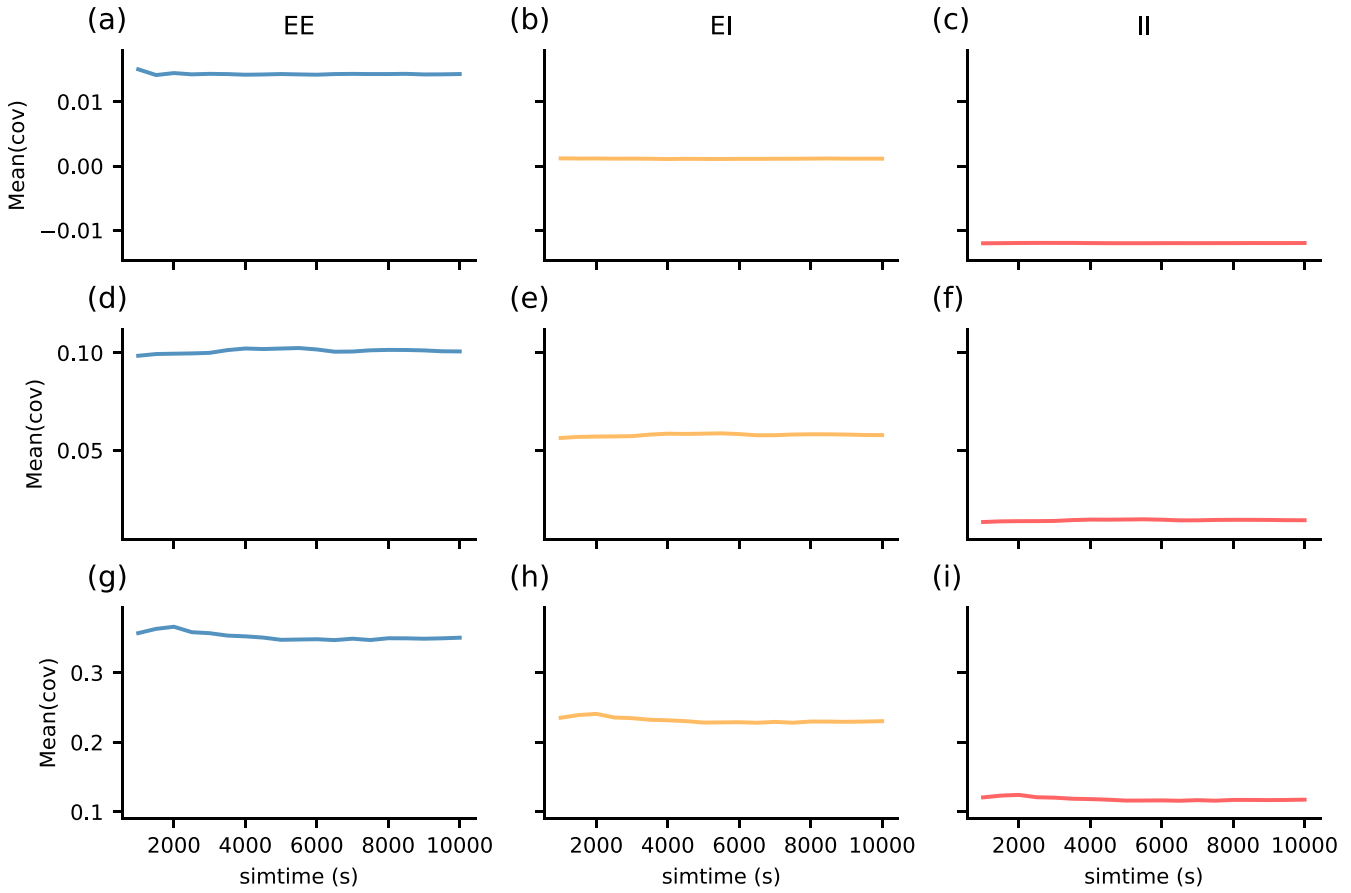

FIG. 11. Mean of covariance estimation for different simulation lengths at three different spectral radii r. [(a)–(c)] $r = 0.10$, [(d)–(f)] $r = 0.49$, [(g)–(i)] $r = 0.90$. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

Eq. (L1) using the following steps:

$$\begin{aligned}\mathrm{CV}^2 &= 2\pi(\tau_\mathrm{m}\nu)^2 \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x\, \mathrm{e}^{x^2} \int_{-\infty}^{x} \mathrm{d}s\, \mathrm{e}^{s^2}(1+\mathrm{erf}(s))^2 \\ &= 2\pi(\tau_\mathrm{m}\nu)^2 \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x \mathrm{e}^{x^2} \int_{-\infty}^{x} \mathrm{d}s\, \mathrm{e}^{s^2} \left(\frac{2}{\sqrt{\pi}} \int_{-\infty}^{s} \mathrm{e}^{-w^2} \mathrm{d}w\right)^2 \\ &= 8(\tau_\mathrm{m}\nu)^2 \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x \int_{-\infty}^{x} \mathrm{d}s \int_{-\infty}^{s} \mathrm{d}v \int_{-\infty}^{s} \mathrm{d}w\, \mathrm{e}^{x^2+s^2-v^2-w^2}.\end{aligned}$$

We make a change of variables $v' = s - v$, $w' = s - w$, where we immediately drop the prime, yielding

$$\begin{aligned}\mathrm{CV}^2 = 8(\tau_\mathrm{m}\nu)^2 &\int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x \int_{0}^{\infty} \mathrm{d}v \int_{0}^{\infty} \mathrm{d}w \\ &\times \int_{-\infty}^{x} \mathrm{d}s\, \mathrm{e}^{x^2-s^2-v^2-w^2+2(v+w)s}.\end{aligned}$$

Another change of variables $s' = x + v + w - s$ gives

$$\mathrm{CV}^2 = 8(\tau_\mathrm{m}\nu)^2 \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x \int_{0}^{\infty} \mathrm{d}v \int_{0}^{\infty} \mathrm{d}w \int_{v+w}^{\infty} \mathrm{d}s\, \mathrm{e}^{-s^2+2vw+2sx}.$$

We switch the order of integration using $\int_0^\infty \mathrm{d}v \int_0^\infty \mathrm{d}w \int_{v+w}^\infty \mathrm{d}s = \int_0^\infty \mathrm{d}s \int_0^s \mathrm{d}v \int_0^{s-v} \mathrm{d}w$, which yields the form we used for the numerical implementation:

$$\begin{aligned}\mathrm{CV}^2 &= 8(\tau_\mathrm{m}\nu)^2 \int_0^\infty \mathrm{d}s \int_0^s \mathrm{d}v \int_0^{s-v} \mathrm{d}w \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x\, \mathrm{e}^{-s^2+2vw+2sx} \\ &= 8(\tau_\mathrm{m}\nu)^2 \int_0^\infty \mathrm{d}s \int_0^s \mathrm{d}v\, \mathrm{e}^{-s^2} \int_0^{s-v} \mathrm{d}w\, \mathrm{e}^{2vw} \int_{y_\mathrm{r}}^{y_\mathrm{th}} \mathrm{d}x\, \mathrm{e}^{2sx} \\ &= 2(\tau_\mathrm{m}\nu)^2 \int_0^\infty \mathrm{d}s \int_0^s \mathrm{d}v\, \mathrm{e}^{-s^2} \frac{1}{v}\left[\mathrm{e}^{2v(s-v)} - 1\right] \\ &\quad \times \frac{1}{s}\left[\mathrm{e}^{2sy_\mathrm{th}} - \mathrm{e}^{2sy_\mathrm{r}}\right] \\ &= 2(\tau_\mathrm{m}\nu)^2 \int_0^\infty \mathrm{d}s\, \frac{1}{s}\left[\mathrm{e}^{2sy_\mathrm{th}} - \mathrm{e}^{2sy_\mathrm{r}}\right] \\ &\quad \times \int_0^s \mathrm{d}v\, \frac{1}{v}\left[\mathrm{e}^{-s^2-2v^2+2sv} - \mathrm{e}^{-s^2}\right].\end{aligned}$$

APPENDIX M: INFERENCE OF CONNECTIVITY FEATURES FROM COVARIANCES

Equations (35) and (36) relate the mean connectivity $\boldsymbol{M}$ and the connectivity's variance $\boldsymbol{S}$ to the mean covariances $\langle \boldsymbol{C} \rangle_{\boldsymbol{W},\boldsymbol{D}}$ and the variance of covariances $\langle \delta \boldsymbol{C}^2 \rangle_{\boldsymbol{W},\boldsymbol{D}}$, respectively. These relations can be inverted to infer features of $\boldsymbol{M}$ and $\boldsymbol{S}$ from the statistics of covariances. Here, we focus on $\boldsymbol{S}$ and, in particular, the spectral radius $r = \sqrt{N_\mathrm{E} S_\mathrm{E} + N_\mathrm{I} S_\mathrm{I}}$, which substantially modulates the size of the variance of covariances, as shown in Fig. 4. In the E-I network considered in this work, $\boldsymbol{S}$ has the structure

$$\boldsymbol{S} = \begin{pmatrix} S_\mathrm{E}\{\mathbf{1}\}_{\mathrm{EE}} & S_\mathrm{I}\{\mathbf{1}\}_{\mathrm{EI}} \\ S_\mathrm{E}\{\mathbf{1}\}_{\mathrm{IE}} & S_\mathrm{I}\{\mathbf{1}\}_{\mathrm{II}} \end{pmatrix}, \tag{M1}$$

with $\{\mathbf{1}\}_{XY}$ being the matrix of ones of dimension $N_X \times N_Y$. The inverse matrix $(\mathbf{1} - \boldsymbol{S})^{-1}$ then follows as

$$(\mathbf{1} - \boldsymbol{S})^{-1} = \begin{pmatrix} \mathbf{1}_{\mathrm{EE}} + \widetilde{S}_E\{\mathbf{1}\}_{\mathrm{EE}} & \widetilde{S}_\mathrm{I}\{\mathbf{1}\}_{\mathrm{EI}} \\ \widetilde{S}_\mathrm{E}\{\mathbf{1}\}_{\mathrm{IE}} & \mathbf{1}_{\mathrm{II}} + \widetilde{S}_\mathrm{I}\{\mathbf{1}\}_{\mathrm{II}} \end{pmatrix}, \tag{M2}$$

with

$$\widetilde{S}_\mathrm{E} = \frac{S_\mathrm{E}}{1 - N_\mathrm{E} S_\mathrm{E} - N_\mathrm{I} S_\mathrm{I}}, \tag{M3}$$

$$\widetilde{S}_\mathrm{I} = \frac{S_\mathrm{I}}{1 - N_\mathrm{E} S_\mathrm{E} - N_\mathrm{I} S_\mathrm{I}}. \tag{M4}$$

From Eq. (36) the variance of covariances follows as $\langle \delta \boldsymbol{C}^2 \rangle_{\boldsymbol{W},\boldsymbol{D}} = (\overline{\mathrm{CV}^2 \nu})^2 \boldsymbol{F}$ with

$$\begin{aligned} \boldsymbol{F} &= (\mathbf{1} - \boldsymbol{S})^{-1} (\mathbf{1} - \boldsymbol{S})^{-\mathrm{T}} \\ &= \mathbf{1} + \begin{pmatrix} 2\widetilde{S}_\mathrm{E}\{\mathbf{1}\}_{\mathrm{EE}} & (\widetilde{S}_\mathrm{E} + \widetilde{S}_\mathrm{I})\{\mathbf{1}\}_{\mathrm{EI}} \\ (\widetilde{S}_\mathrm{E} + \widetilde{S}_\mathrm{I})\{\mathbf{1}\}_{\mathrm{IE}} & 2\widetilde{S}_\mathrm{I}\{\mathbf{1}\}_{\mathrm{II}} \end{pmatrix} \\ &\quad + (\widetilde{S}_\mathrm{E}^2 N_\mathrm{E} + \widetilde{S}_\mathrm{I}^2 N_\mathrm{I})\{\mathbf{1}\}. \end{aligned}$$

Focusing on the off-diagonal elements of each block of $\boldsymbol{F}$ yields quadratic equations

$$\begin{aligned} F_{\mathrm{EE}} &= 2\widetilde{S}_\mathrm{E} + (\widetilde{S}_\mathrm{E}^2 N_\mathrm{E} + \widetilde{S}_\mathrm{I}^2 N_\mathrm{I}), \\ F_{\mathrm{II}} &= 2\widetilde{S}_\mathrm{I} + (\widetilde{S}_\mathrm{E}^2 N_\mathrm{E} + \widetilde{S}_\mathrm{I}^2 N_\mathrm{I}), \\ F_{\mathrm{EI}} &= (\widetilde{S}_\mathrm{E} + \widetilde{S}_\mathrm{I}) + (\widetilde{S}_\mathrm{E}^2 N_\mathrm{E} + \widetilde{S}_\mathrm{I}^2 N_\mathrm{I}) = \frac{1}{2}(F_{\mathrm{EE}} + F_{\mathrm{II}}), \end{aligned}$$

that can be solved for $\widetilde{S}_\mathrm{E}$ and $\widetilde{S}_\mathrm{I}$ and subsequently, using Eqs. (M3) and (M4), for the connectivity parameters S_E and S_I:

$$S_\mathrm{E} = \frac{1}{N}\left(1 - \frac{\sigma\left(1 - \frac{1}{2}(F_{\mathrm{EE}} - F_{\mathrm{II}})N_\mathrm{I}\right)}{\sqrt{1 + N\frac{F_{\mathrm{EE}} - \frac{1}{4}(F_{\mathrm{EE}} - F_{\mathrm{II}})^2 N_\mathrm{I}}{\left[1 - \frac{1}{2}(F_{\mathrm{EE}} - F_{\mathrm{II}})N_\mathrm{I}\right]^2}}}\right), \tag{M5}$$

$$S_\mathrm{I} = \frac{1}{N}\left(1 - \frac{\sigma\left(1 + \frac{1}{2}(F_{\mathrm{EE}} - F_{\mathrm{II}})N_\mathrm{E}\right)}{\sqrt{1 + N\frac{F_{\mathrm{II}} - \frac{1}{4}(F_{\mathrm{II}} - F_{\mathrm{EE}})^2 N_\mathrm{E}}{\left[1 - \frac{1}{2}(F_{\mathrm{II}} - F_{\mathrm{EE}})N_\mathrm{E}\right]^2}}}\right), \tag{M6}$$

with $\sigma(x) = 1$ for $x \geqslant 0$ and $\sigma(x) = -1$ for $x < 0$, $F_{\mathrm{EE}} = \langle \delta C_{\mathrm{EE}}^2 \rangle_{\boldsymbol{W},\boldsymbol{D}} / (\overline{\mathrm{CV}^2 \nu})^2$, and $F_{\mathrm{II}} = \langle \delta C_{\mathrm{II}}^2 \rangle_{\boldsymbol{W},\boldsymbol{D}} / (\overline{\mathrm{CV}^2 \nu})^2$. As shown in Fig. 12, this in particular allows the inference of the spectral radius of bulk connectivity eigenvalues,

$$r_{\mathrm{inf}} = \sqrt{N_\mathrm{E} S_\mathrm{E} + N_\mathrm{I} S_\mathrm{I}}\,, \tag{M7}$$

from the variance of covariances. In principle, a similar procedure could be followed to infer the mean connectivity M_E and M_I from the mean covariances. However, we performed a number of simplifications $(\mathbf{1} - \boldsymbol{M})^{-1} \approx \mathbf{1}$ in calculating the effective noise [see, e.g., Eq. (33)] to arrive at Eq. (35). Therefore, the inference of M_E and M_I based on measured mean covariances is expected to be less accurate and would potentially require a more careful mathematical treatment.

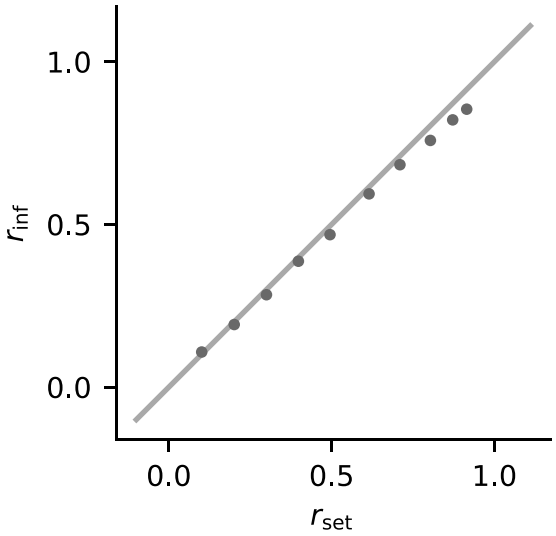

FIG. 12. Set spectral radius r_{set} (defined in Appendix D) vs inferred spectral radius r_{inf}. Same excitatory-inhibitory network model as in previous figures. For model details and simulation parameters see Appendix A.

[1] J. S. Griffith and G. Horn, An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats, J. Physiol. **186**, 516 (1966).

[2] K. W. Koch and J. M. Fuster, Unit activity in monkey parietal cortex related to haptic perception and temporary memory, Exp. Brain Res. **76**, 292 (1989).

[3] P. A. Dąbrowska, N. Voges, M. von Papen, J. Ito, D. Dahmen, A. Riehle, T. Brochier, and S. Grün, On the complexity of resting state spiking activity in monkey motor cortex, Cereb. Cortex Commun. **2**, tgab033 (2021).

[4] S. Shinomoto, K. Shima, and J. Tanji, Differences in spiking patterns among cortical neurons, Neural Comput. **15**, 2823 (2003).

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

[6] M. R. Cohen and A. Kohn, Measuring and interpreting neuronal correlations, Nat. Neurosci. **14**, 811 (2011).

[7] D. Dahmen, S. Grün, M. Diesmann, and M. Helias, Second type of criticality in the brain uncovers rich multiple-neuron dynamics, Proc. Natl. Acad. Sci. USA **116**, 13051 (2019).

[8] D. Dahmen, M. Layer, L. Deutz, P. A. Dąbrowska, N. Voges, M. von Papen, T. Brochier, A. Riehle, M. Diesmann, S. Grün *et al.*, Global organization of neuronal activity only requires unstructured local connectivity, eLife **11**, e68422 (2022).

[9] R. A. da Silveira and M. J. Berry II, High-fidelity coding with correlated neurons, PLOS Comput. Biol. **10**, e1003970 (2014).

[10] R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget, Information-limiting correlations, Nat. Neurosci. **17**, 1410 (2014).

[11] C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, and K. D. Harris, High-dimensional geometry of population responses in visual cortex, Nature (London) **571**, 361 (2019).

[12] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, Computation through neural population dynamics, Annu. Rev. Neurosci. **43**, 249 (2020).

[13] J. J. Jun, N. A. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. A. Anastassiou, A. Andrei, Ç. Aydın, M. Barbic, T. J. Blanche, V. Bonin, J. Couto, B. Dutta, S. L. Gratiy, D. A. Gutnisky, M. Häusser, B. Karsh, P. Ledochowitsch *et al.*, Fully integrated silicon probes for high-density recording of neural activity, Nature (London) **551**, 232 (2017).

[14] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux, S. Chen, J. Colonell, R. J. Gardner, B. Karsh, F. Kloosterman, D. Kostadinov, C. Mora-Lopez, J. O'Callaghan, J. Park, J. Putzeys *et al.*, Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings, Science **372**, eabf4588 (2021).

[15] L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli *et al.*, Local connectivity and synaptic dynamics in mouse and human neocortex, Science **375**, eabj5861 (2022).

[16] A. Roxin, N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk, On the distribution of firing rates in networks of cortical neurons, J. Neurosci. **31**, 16217 (2011).

[17] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, The structure of the nervous system of the nematode *Caenorhabditis elegans*, Philos. Trans. R. Soc. Lond. B Biol. Sci. **314**, 1 (1986).

[18] N. T. Markov, M. M. Ercsey-Ravasz, A. R. Ribeiro Gomes, C. Lamy, L. Magrou, J. Vezoli, P. Misery, A. Falchier, R. Quilodran, M. A. Gariel, J. Sallet, R. Gamanut, C. Huissoud, S. Clavagnier, P. Giroud, D. Sappey-Marinier, P. Barone, C. Dehay, Z. Toroczkai, K. Knoblauch *et al.*, A weighted and directed interareal connectivity matrix for macaque cerebral cortex, Cereb. Cortex **24**, 17 (2014).

[19] S. J. van Albada, A. Morales-Gregorio, T. Dickscheid, A. Goulas, R. Bakker, S. Bludau, G. Palm, C.-C. Hilgetag, and M. Diesmann, Bringing anatomical information into neuronal network models, in *Computational Modelling of the Brain: Modelling Approaches to Cells, Circuits and Networks*, edited by M. Giugliano, M. Negrello, and D. Linaro (Springer International Publishing, Cham, 2022), pp. 201–234.

[20] P. Schnepel, A. Kumar, M. Zohar, A. Aertsen, and C. Boucsein, Physiology and impact of horizontal connections in rat neocortex, Cereb. Cortex **25**, 3818 (2015).

[21] R. Sayer, M. Friedlander, and S. Redman, The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice, J. Neurosci. **10**, 826 (1990).

[22] D. Feldmeyer, V. Egger, J. Lübke, and B. Sakmann, Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single "barrel" of developing rat somatosensory cortex, J. Physiol. **521**, 169 (1999).

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

[24] S. Lefort, C. Tomm, J.-C. F. Sarria, and C. C. H. Petersen, The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex, Neuron **61**, 301 (2009).

[25] Y. Ikegaya, T. Sasaki, D. Ishikawa, N. Honma, K. Tao, N. Takahashi, G. Minamisawa, S. Ujita, and N. Matsuki, Interpyramid spike transmission stabilizes the sparseness of recurrent network activity, Cereb. Cortex **23**, 293 (2013).

[26] Y. Loewenstein, A. Kuras, and S. Rumpel, Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo, J. Neurosci. **31**, 9481 (2011).

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

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

[29] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter, Recurrent interactions in spiking networks with arbitrary topology, Phys. Rev. E **85**, 031916 (2012).

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

[31] D. Grytskyy, T. Tetzlaff, M. Diesmann, and M. Helias, A unified view on weakly correlated recurrent networks, Front. Comput. Neurosci. **7**, 131 (2013).

[32] D. Dahmen, H. Bos, and M. Helias, Correlated fluctuations in strongly coupled binary networks beyond equilibrium, Phys. Rev. X **6**, 031024 (2016).

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

[34] C. van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, Science **274**, 1724 (1996).

[35] M. A. Buice, J. D. Cowan, and C. C. Chow, Systematic fluctuation expansion for neural network activity equations, Neural Comput. **22**, 377 (2010).

[36] A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, The asynchronous state in cortical circuits, Science **327**, 587 (2010).

[37] T. Tetzlaff, M. Helias, G. T. Einevoll, and M. Diesmann, Decorrelation of neural-network activity by inhibitory feedback, PLoS Comput. Biol. **8**, e1002596 (2012).

[38] E. Montbrió, D. Pazó, and A. Roxin, Macroscopic description for networks of spiking neurons, Phys. Rev. X **5**, 021028 (2015).

[39] H. Sompolinsky and A. Zippelius, Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses, Phys. Rev. B **25**, 6860 (1982).

[40] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks* (Springer International Publishing, Cham, 2020), p. 203.

[41] D. Dahmen, S. Recanatesi, X. Jia, G. K. Ocker, L. Campagnola, T. Jarsky, S. Seeman, M. Helias, and E. Shea-Brown, Strong and localized recurrence controls dimensionality of neural activity across brain areas, bioRxiv (2023).

[42] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. **8**, 183 (2000).

[43] N. Brunel and V. Hakim, Fast global oscillations in networks of integrate-and-fire neurons with low firing rates, Neural Comput. **11**, 1621 (1999).

[44] A. J. Siegert, On the first passage time probability problem, Phys. Rev. **81**, 617 (1951).

[45] M. Helias, T. Tetzlaff, and M. Diesmann, Echoes in correlated neural systems, New J. Phys. **15**, 023002 (2013).

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

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

[48] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Methuen's Monographs on Applied Probability and Statistics (Methuen, London, 1966).

[49] M. Layer, J. Senk, S. Essink, A. van Meegen, H. Bos, and M. Helias, NNMT: Mean-field based analysis tools for neuronal network models, Front. Neuroinf. **16**, 835657 (2022).

[50] K. Fischer and J. Hertz, *Spin Gasses* (Cambridge University Press, Cambridge, UK, 1991).

[51] J. A. Hertz, Y. Roudi, and P. Sollich, Path integral methods for the dynamics of stochastic and disordered systems, J. Phys. A **50**, 033001 (2017).

[52] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. **61**, 259 (1988).

[53] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Spectrum of large random asymmetric matrices, Phys. Rev. Lett. **60**, 1895 (1988).

[54] C. De Dominicis, Dynamics as a substitute for replicas in systems with quenched random impurities, Phys. Rev. B **18**, 4913 (1978).

[55] C. van Vreeswijk and H. Sompolinsky, Chaotic balanced state in a model of cortical circuits, Neural Comput. **10**, 1321 (1998).

[56] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, UK, 1996).

[57] M. Helias, T. Tetzlaff, and M. Diesmann, The correlation structure of local cortical networks intrinsically results from recurrent dynamics, PLoS Comput. Biol. **10**, e1003428 (2014).

[58] R. Rosenbaum and B. Doiron, Balanced networks of spiking neurons with spatially dependent recurrent connections, Phys. Rev. X **4**, 021039 (2014).

[59] R. Pyle and R. Rosenbaum, Spatiotemporal dynamics and reliable computations in recurrent spiking neural networks, Phys. Rev. Lett. **118**, 018103 (2017).

[60] R. Darshan, C. van Vreeswijk, and D. Hansel, Strength of correlations in strongly recurrent neuronal networks, Phys. Rev. X **8**, 031072 (2018).

[61] G. B. Smith, B. Hein, D. E. Whitney, D. Fitzpatrick, and M. Kaschube, Distributed network interactions and their emergence in developing neocortex, Nat. Neurosci. **21**, 1600 (2018).

[62] C. Huang, D. A. Ruff, R. Pyle, R. Rosenbaum, M. R. Cohen, and B. Doiron, Circuit models of low-dimensional shared variability in cortical networks, Neuron **101**, 337 (2019).

[63] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).

[64] B. Kriener, M. Helias, S. Rotter, M. Diesmann, and G. T. Einevoll, How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime, Front. Comput. Neurosci. **7**, 1 (2014).

[65] J. Aljadeff, D. Renfrew, M. Vegué, and T. O. Sharpee, Low-dimensional dynamics of structured random networks, Phys. Rev. E **93**, 022302 (2016).

[66] K. Rajan and L. F. Abbott, Eigenvalue spectra of random matrices for neural networks, Phys. Rev. Lett. **97**, 188104 (2006).

[67] S. Ostojic, Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons, Nat. Neurosci. **17**, 594 (2014).

[68] D. J. Amit and N. Brunel, Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex, Cereb. Cortex **7**, 237 (1997).

[69] B. Kriener, H. Enger, T. Tetzlaff, H. E. Plesser, M.-O. Gewaltig, and G. T. Einevoll, Dynamics of self-sustained asynchronous-irregular activity in random networks of spiking neurons with strong synapses. Front. Comput. Neurosci. **8**, 136 (2014).

[70] M. Stern, H. Sompolinsky, and L. F. Abbott, Dynamics of random neural networks with bistable units, Phys. Rev. E **90**, 062710 (2014).

[71] J. Aljadeff, M. Stern, and T. Sharpee, Transition to chaos in random networks with cell-type-specific connectivity, Phys. Rev. Lett. **114**, 088101 (2015).

[72] D. Martí, N. Brunel, and S. Ostojic, Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks, Phys. Rev. E **97**, 062314 (2018).

[73] J. Schuecker, S. Goedeke, and M. Helias, Optimal sequence memory in driven random networks, Phys. Rev. X **8**, 041029 (2018).

[74] A. Crisanti and H. Sompolinsky, Path integral approach to random neural networks, Phys. Rev. E **98**, 062120 (2018).

[75] S. P. Muscinelli, W. Gerstner, and T. Schwalger, How single neuron properties shape chaotic dynamics and signal transmission in random neural networks, PLoS Comput. Biol. **15**, e1007122 (2019).

[76] M. Beiran and S. Ostojic, Contrasting the effects of adaptation and synaptic filtering on the timescales of dynamics in recurrent networks, PLoS Comput. Biol. **15**, e1006893 (2019).

[77] Ł. Kuśmierz, S. Ogawa, and T. Toyoizumi, Edge of chaos and avalanches in neural networks with heavy-tailed synaptic weight distribution, Phys. Rev. Lett. **125**, 028101 (2020).

[78] A. Wardak and P. Gong, Extended Anderson criticality in heavy-tailed neural networks, Phys. Rev. Lett. **129**, 048103 (2022).

[79] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron, The spatial structure of correlated neuronal variability, Nat. Neurosci. **20**, 107 (2017).

[80] C. Baker, C. Ebsch, I. Lampl, and R. Rosenbaum, Correlated states in balanced neuronal networks, Phys. Rev. E **99**, 052414 (2019).

[81] A. Litwin-Kumar and B. Doiron, Slow dynamics and high variability in balanced cortical networks with clustered connections, Nat. Neurosci. **15**, 1498 (2012).

[82] B. A. Brinkman, H. Yan, A. Maffei, I. M. Park, A. Fontanini, J. Wang, and G. La Camera, Metastable dynamics of neural circuits and networks, Appl. Phys. Rev. **9**, 011313 (2022).

[83] P. C. Bressloff, Spatiotemporal dynamics of continuum neural fields, J. Phys. A: Math. Theor. **45**, 033001 (2012).

[84] R. Zeraati, Y.-L. Shi, N. A. Steinmetz, M. A. Gieselmann, A. Thiele, T. Moore, A. Levina, and T. A. Engel, Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity, Nat. Commun. **14**, 1858 (2023).

[85] Y.-L. Shi, R. Zeraati, A. Levina, and T. A. Engel, Spatial and temporal correlations in neural networks with structured connectivity, Phys. Rev. Res. **5**, 013005 (2023).

[86] 10.5281/zenodo.10796382.

[87] M.-O. Gewaltig and M. Diesmann, NEST (NEural Simulation Tool), Scholarpedia **2**, 1430 (2007); http://www.nest-simulator.org.

[88] R. B. Stein, Some models of neuronal variability, Biomed. Pharmacol. J. **7**, 37 (1967).

[89] H. C. Tuckwell, *Introduction to Theoretical Neurobiology*, Vol. 1 (Cambridge University Press, Cambridge, UK, 1988).

[90] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 2nd ed., Springer Series in Synergetics No. 13 (Springer-Verlag, Berlin, 1985).

[91] A. Khintchine, Korrelationstheorie der stationaeren stochastischen prozesse, Math. Ann. **109**, 604 (1934).

[92] M. Helias, M. Deger, S. Rotter, and M. Diesmann, Instantaneous non-linear processing by pulse-coupled threshold units, PLoS Comput. Biol. **6**, e1000929 (2010).