Strength of Correlations in Strongly Recurrent Neuronal Networks

Ran Darshan,1,2,3 Carl van Vreeswijk,3 and David Hansel3,*

1ELSC, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
2Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA
3Center for Neurophysics, Physiology and Pathology; Cerebral Dynamics, Plasticity and Learning Research Team; CNRS-UMR8119, Paris 75270, France

(Received 24 August 2017; revised manuscript received 15 March 2018; published 17 September 2018)

Spatiotemporal correlations in brain activity are functionally important and have been implicated in perception, learning and plasticity, exploratory behavior, and various aspects of cognition. Neurons in the cerebral cortex are strongly interacting. Their activity is temporally irregular and can exhibit substantial correlations. However, how the collective dynamics of highly recurrent and strongly interacting neurons can evolve into a state in which the activity of individual cells is highly irregular yet macroscopically correlated is an open question. Here, we develop a general theory that relates the strength of pairwise correlations to the anatomical features of networks of strongly coupled neurons. To this end, we investigate networks of binary units. When interactions are strong, the activity is irregular in a large region of parameter space. We find that despite the strong interactions, the correlations are generally very weak. Nevertheless, we identify architectural features, which if present, give rise to strong correlations without destroying the irregularity of the activity. For networks with such features, we determine how correlations scale with the network size and the number of connections. Our work shows the mechanism by which strong correlations can be consistent with highly irregular activity, two hallmarks of neuronal dynamics in the central nervous system.

DOI: 10.1103/PhysRevX.8.031072 Subject Areas: Biological Physics, Complex Systems, Interdisciplinary Physics

I. INTRODUCTION

The central nervous system consists of a huge diversity of neuronal populations divided into excitatory and inhibitory cell types, which differ in their morphology, molecular signature, and in their connectivity. Each neuron, especially in the cortex, receives recurrent inputs from several hundreds [1] to a few thousands [2] of other neurons, resulting in networks with a highly recurrent connectivity. Here, recurrent means that the connectivity has embedded loops. As a result, in a recurrent network the activity reverberates. Functionally, recurrent connections are particularly important, for example, in sharpening of selective response to sensory stimuli [3], associative [4,5] and working memory [6–8], as well as decision-making [9].

In addition to the plethora of neuronal types, recurrent neuronal networks in the brain are also highly structured. In particular, the probability of connection falls off with distance and depends on functional properties of pre- and postsynaptic neurons. For example, in the mouse primary auditory cortex, the probability of excitatory neurons to be connected decays to zero within approximately 300 μm [10]. In the cat primary visual cortex, neurons interact locally on a range of approximately 500 μm, whereas long-range patchy connections are observed up to several millimeters [11,12].

The highly recurrent connectivity, the diversity of the neuronal populations, and the different spatial and functional structures in the brain make the collective dynamics of recurrent cortical neuronal networks highly nontrivial. Like in many other extended dynamical systems [13–18], these collective dynamics are characterized by the activity of single neurons and their correlations. With the rapid advances in technologies [19–21], it is now possible to record simultaneously from many cells. This allows one to estimate not only the temporal variations in the activity of individual cells but also their pairwise correlations. The amount of data made available by current techniques, however, precludes reliable estimation of higher-order statistics.

In recent years, it was shown that correlation strengths depend on the brain area [22,23], the layer in the cortex [24], stimulus conditions, behavioral states [25], and experience [26–28]. A wide range of values for correlation coefficients from negligible [22,29] to substantial [30–35] have been
reported in the last two decades. Correlation coefficients are usually higher for close-by neurons than for neurons that are far apart [10,32,36–38]. In the cortex, they drop significantly over distances of 200–400 µm [10,38]. Recent experimental studies have reported correlations varying nonmonotonically with distance [39] or correlations which are positive for close-by neurons but negative for neurons farther apart [40]. Correlation coefficients also depend on the functional properties of the neurons. Neurons which code for similar features of sensory stimuli are more correlated [22,32,36,38,41].

In this study, we investigate the mechanism by which in recurrent neuronal networks, spatiotemporal correlations in neural activity emerge from the interplay between the nonlinear single-neuron dynamics and the network architecture.

As individual connections between neurons in the cortex can induce postsynaptic potentials in a range of 0.1 mV to several millivolts [10], 50 simultaneous inputs are sufficient to trigger or suppress a spike in a neuron. These facts have been incorporated in cortical network models by assuming strongly recurrent connectivity in which connection strengths are $O(1/\sqrt{K})$, where $K$ is the average number of inputs per neuron. This scaling is in contrast to the one used in standard mean-field models (e.g., Ref. [42]) where connections are $O(1/K)$ and thus are much weaker. Recent experiments in cortical cultures are consistent with a $1/\sqrt{K}$ scaling of connection strength [43].

Previous theoretical studies of strongly recurrent networks have shown that they naturally account for many features of cortical dynamics, such as strong temporal irregularity, trial-to-trial variability [44,45], and spatial heterogeneities [46,47]. It has been shown, however, that correlations between neurons in strongly recurrent and unstructured networks are extremely weak [29]; i.e., these networks operate in an asynchronous state with $O(1/N)$ correlations, where $N$ is the number of neurons in the network [48]. This lack in correlations is because in unstructured and strongly recurrent networks, the dynamics suppress correlations [49]. Thus, while the current theory of strongly recurrent neuronal networks can account for many features of recurrent neuronal networks, it fails to account for experimental reports of irregular neural activity which exhibits coherent fluctuations on a macroscopic scale (i.e., correlated across neurons).

In a previous paper [50], we presented an example of a network model consisting of strongly recurrent populations in which the activity is temporally irregular and spatially correlated. In the present paper, we derive general theorems that relate the strength of pairwise correlations to features of the network architecture. The results presented in this paper answer a long-standing question of how neuronal activity can be both temporally irregular and substantially spatially correlated as observed experimentally in a cortex.

The organization of the paper is as follows. In Sec. II, we define the network architectures and the neuronal model we use. In Sec. III, we establish a set of constraints—the balanced correlation equations—that need to be satisfied in any strongly recurrent network if firing rates do not saturate. We derive in Sec. IV explicit expressions for the Fourier components of the correlations in two-population networks with spatially modulated connectivity. We establish the conditions under which correlations are strong and show that these strong correlations do not violate the balanced correlation equations. Section V is devoted to networks with an arbitrary number of populations. We prove two theorems, which state for which network architectures correlations are $O(1/N)$ and for which they increase with $K$, when $K$ is large. In Sec. VI, we apply these theorems to specific examples. In Sec. VII, we assume that $K = O(N^\gamma)$ and derive a bound on $\gamma$ for which the scaling theorems still apply. The paper closes with a discussion of our results.

II. THE NETWORK MODEL

van Vreeswijk and Sompolinsky [44,51] showed that strongly recurrent networks consisting of populations of neurons, one excitatory ($E$) and one inhibitory ($I$), randomly connected on a directed Erdős-Rényi graph, operate in a state in which the strong excitation is balanced by the strong inhibition. In this balanced regime, neurons receive strong excitatory and inhibitory inputs, each $O(\sqrt{K})$, but due to the recurrent dynamics, these inputs cancel each other at the leading order. This cancellation, which does not require fine-tuning, results in $O(1)$ net inputs into the neurons, whereas spatial heterogeneities and temporal fluctuations in the inputs are also $O(1)$. As a result, the activity of the neurons remains finite and exhibits strong temporal irregularity and heterogeneity [46,52]. In Ref. [44], the authors investigated networks with only two neuronal populations. Here, we extend this approach to a network with an arbitrary number of populations and spatially dependent interactions. As in van Vreeswijk and Sompolinsky [44,51], we study networks of binary interacting neurons [29,42,51,53]. The connection between the results derived using such models to a more biologically relevant model is discussed in Sec. VIII.

A. Architecture

We consider a one-dimensional neuronal network comprising $D$ neuronal populations, some excitatory and others inhibitory (Fig. 1). For simplicity, we assume that all populations have the same number of neurons $N$, and to simplify calculations, we also assume periodic boundary conditions (a ring architecture; see, e.g., Refs. [3,54]). Neuron $i$ in population $\alpha$ [neuron $(i,\alpha)$] is located at angle $\theta_i^\alpha = [(2\pi i)/N]$ with $i = 1,\ldots,N$ and $\alpha = 1,\ldots,D$. The probability $P_{ij}^{\alpha\beta}$ that neuron $(j,\beta)$ projects to neuron $(i,\alpha)$ depends on their distance on the ring [Fig. 1(b)]. We write $P_{ij}^{\alpha\beta} = (K/N) f_{\alpha\beta}(|\theta_i^\alpha - \theta_j^\beta|)$, where $\sum_{i,j} f_{\alpha\beta}(|\theta_i^\alpha - \theta_j^\beta|) = N$. In this paper, we assume a finite number of nonzero Fourier modes in $f_{\alpha\beta}$.
The state of neuron \((i, \alpha)\) is characterized by a binary variable \(S_i^\alpha\). When the neuron is quiescent, \(S_i^\alpha = 0\), while if it is active, \(S_i^\alpha = 1\). The total synaptic current into neuron \((i, \alpha)\) at time \(t\), \(h_i^\alpha(t)\) is the result of all its recurrent interactions with the other neurons in the network, as well as the feed-forward inputs coming from outside the network. It is given by

\[
h_i^\alpha(t) = \sum_{\beta=1}^{D} \sum_{j=1}^{N} J_{ij}^{\alpha\beta} S_j^\beta(t) + I_i^\alpha,
\]

where the external input \(I_i^\alpha = \sqrt{K} I_i(\theta)\) [51] is assumed to be constant in time. Here, \(I_i\) is \(O(1)\) and, thus, \(I_i^\alpha\) is \(O(\sqrt{K})\).

The neurons are updated with Glauber dynamics at zero temperature [13,29,42]. Specifically, the update times for neuron \((i, \alpha)\) are Poisson distributed with the rate \(1/\tau\); if neuron \((i, \alpha)\) is updated at time \(t\), \(S_i^\alpha(t)\) is set to 1 if \(h_i^\alpha(t) \geq T_\alpha\) and to 0 otherwise (for simplicity, we assume the same threshold \(T_\alpha\) for all neurons in population \(\alpha\) and the same update rate for all neurons). Accordingly, the transition probability \(w\) [Fig. 1(c)] can be written as

\[
w[S_i^\alpha(t) \rightarrow 1 - S_i^\alpha(t)] = \frac{1}{\tau} \{S_i^\alpha(t) - \Theta[h_i^\alpha(t) - T_\alpha]\}^2,
\]

where \(\Theta\) is the Heaviside function. We normalize time so that \(\tau = 1\).

### III. Spatiotemporal Profile and Correlations of the Activity for \(N, K \rightarrow \infty\)

We write the state of neuron \((i, \alpha)\) as

\[
S_i^\alpha(t) = m_\alpha(\theta_i^\alpha) + \Delta S_i^\alpha + \delta S_i^\alpha(t),
\]

where \(m_\alpha(\theta_i^\alpha) = \langle S_i^\alpha(t) \rangle_i\) is the average of \(S_i^\alpha(t)\) over realizations of the network connectivity matrix and over time. In Eq. (3), the term \(\Delta S_i^\alpha\) is the quenched disorder \(\Delta S_i^\alpha = \langle S_i^\alpha(t) \rangle_i - m_\alpha(\theta_i^\alpha)\), whereas \(\delta S_i^\alpha(t)\) represents the temporal fluctuations in the activity of the neuron \((i, \alpha)\), \(\delta S_i^\alpha(t) = S_i^\alpha(t) - \langle S_i^\alpha(t) \rangle_i\).

Similarly, we can write

\[
h_i^\alpha(t) = h_\alpha(\theta_i^\alpha) + \Delta h_i^\alpha + \delta h_i^\alpha(t),
\]

where \(h_\alpha(\theta_i^\alpha) = \langle h_i^\alpha(t) \rangle_i\) is a smooth function of its argument. For large \(N\), it is given by

\[
\]
\[ h_\alpha(\theta) = \sqrt{K} \left[ \sum_{\beta=1}^{D} J_{\alpha\beta} \int \frac{d\theta'}{2\pi} f_{\alpha\beta}(\theta' - \theta') m_\beta(\theta') + I_\alpha(\theta) \right]. \]  

(5)

The second term in Eq. (4) is the quenched disorder in the input. It satisfies

\[ \Delta h_i^\alpha = \sum_{\beta=1}^{D} J_{\alpha\beta} \sum_j [\Delta \Lambda_{ij}^{\alpha\beta} m_j(\theta_j^\alpha) + \Lambda_{ij}^{\alpha\beta} \Delta S_j^\alpha], \]  

(6)

where \( \Delta \Lambda_{ij}^{\alpha\beta} = \Lambda_{ij}^{\alpha\beta} - [\Lambda_{ij}^{\alpha\beta}]_j. \)

Finally, the temporal fluctuations in the inputs are

\[ \delta h_i^\alpha(t) = \sum_{\beta=1}^{D} J_{\alpha\beta} \sum_j \Lambda_{ij}^{\alpha\beta} \delta S_j^\alpha(t). \]  

(7)

Because we scale the synaptic strength as \( 1/\sqrt{K} \), these temporal fluctuations are \( O(1) \) when the correlations are weak. At first sight, \( h_i^\alpha(\theta) \) is \( O(\sqrt{K}) \). This limiting behavior of \( h_i^\alpha(\theta) \) implies that depending on whether \( h_i^\alpha(\theta) \) is positive or negative, the neurons fire either at a very high or very low rate. This happens unless the network settles into a state in which excitation and inhibition are, on average, balanced. In this case, the net average inputs to the neurons are in fact \( O(1) \). On the other hand, large spatial and temporal correlations could in principle lead to temporal fluctuations and heterogeneities in the inputs of \( O(\sqrt{K}) \). Thus, in the presence of strong correlations, it is not sufficient to require that the mean input \( h_i^\alpha(\theta) \) is \( O(1) \) for the network to operate in a biologically relevant regime. For that, we need both the mean net inputs and the input fluctuations to be \( O(1) \) at any time and for all neurons. When this happens, we say that the network operates in the balanced regime.

A. Balance equations for the quenched averaged population activities

As in Refs. [55,56], the requirement that the mean input is \( O(1) \) yields

\[ \sum_{\beta=1}^{D} J_{\alpha\beta} \int \frac{d\theta'}{2\pi} f_{\alpha\beta}(\theta' - \theta') m_\beta(\theta') + I_\alpha(\theta) = O(1/\sqrt{K}) \]  

(8)

for all \( \alpha \in \{1, \ldots, D\} \) and all \( \theta \in [0, 2\pi) \). In the large-\( K \) limit, Eq. (8) yields a set of linear equations which determines the functions \( m_\alpha(\theta) \) to leading order. These equations can be written in Fourier space

\[ \sum_{\beta} J_{\alpha\beta}^{(n)} m_\beta^{(n)} + I_\alpha^{(n)} = O\left( \frac{1}{\sqrt{K}} \right), \]  

(9)

where the superscript \( n \) denotes the \( n \)th Fourier mode, and we use the shorthand notation

\[ J_{\alpha\beta}^{(n)} = J_{\alpha\beta} f_{\alpha\beta}^{(n)}, \]  

(10)

which is the \( n \)th Fourier mode of the connectivity matrix. In what follows, we consider \( J_{\alpha\beta}^{(n)} \) as the elements of a \( D \times D \) matrix \( J^{(n)} \).

The spatial average of the activities \( m_\alpha^{(0)} \) must be non-negative for the balanced state to exist. This implies that the parameters \( J_{\alpha\beta} \) and the external inputs \( I_\alpha^{(0)} \) must satisfy a set of inequalities. For example, for networks with two populations, one excitatory (\( \alpha = E \)) and one inhibitory (\( \alpha = I \)), these inequalities are [51]

\[ \frac{I_E^{(0)}}{I_I^{(0)}}, J_{EI} > J_{EE}, J_{IE}. \]  

(11)

In general, Eq. (9) also implies additional inequalities that must be satisfied by \( f_{\alpha\beta}^{(n)} \) to guarantee that \( m_\alpha(\theta) \geq 0 \) for all \( \alpha \) and \( \theta \) [55,56]. In the present work, we focus on the case where the external inputs are spatially homogeneous. Thus, \( m_\alpha^{(0)} = 0 \) for \( n \geq 1 \), and the only condition which is required for the balanced state to exist is \( m_\alpha^{(0)} \geq 0 \).

To study the stability of the homogeneous balanced state, it is useful to introduce the interaction matrix

\[ \tilde{J}_{ij}^{\alpha\beta} = g_\alpha J_{\alpha\beta} f_{\alpha\beta}(\theta_i^\alpha - \theta_j^\beta), \]  

(12)

where \( g_\alpha \) is the population-averaged gain (see Appendices A and D and Refs. [49,51]).

Small perturbations \( \delta m_\alpha(\theta, t) \) of the activity profile around this homogeneous state evolve according to [51]

\[ \frac{d\delta m_\alpha^{(n)}}{dt} = -\delta m_\alpha^{(n)} + \sqrt{K} \sum_{\beta} J_{\alpha\beta}^{(n)} \delta m_\beta^{(n)}, \]  

(13)

where \( \delta m_\alpha^{(n)}(t) \) and \( J_{\alpha\beta}^{(n)} \) are the \( n \)th Fourier modes of \( \delta m_\alpha(\theta, t) \) and \( J_{\alpha\beta} \). Since each row of \( J_{\alpha\beta}^{(n)} \) has \( O(K) \) nonzero elements which are \( O(1/K) \), \( J_{\alpha\beta}^{(n)} \) is \( O(1) \).

The stability of the balanced state with respect to perturbations in \( m_\alpha(\theta) \) requires that for all \( n \) all the eigenvalues of the matrices \( J^{(n)} \) have real parts smaller than 1. For instance, for a two-population (\( E,I \)) network, one must have in the large-\( K \) limit

\[ J_{EE} J_{II} f_{EE}^{(n)} f_{II}^{(n)} - J_{IE} J_{EI} f_{EE}^{(n)} f_{IE}^{(n)} \leq 0 \]  

(14)

for all \( n \). Note that the gain of the neurons \( g_\alpha \) dropped from this equation. The balanced state undergoes a Turing bifurcation when for some \( n \geq 1 \), \( J_{EE} J_{II} f_{EE}^{(n)} f_{II}^{(n)} - J_{IE} J_{EI} f_{EE}^{(n)} f_{IE}^{(n)} \) crosses 0 and becomes positive.
B. Balance equation for pairwise correlations

The time-lagged auto- and cross-correlation functions of the activities of a pair of neurons \((i, \alpha), (j, \beta)\) for \((j, \beta) \neq (i, \alpha)\) are

\[
a_{ij}^{\alpha \beta}(\tau) = \langle \delta S_i^\alpha(t) \delta S_j^\beta(t + \tau) \rangle, \tag{15}
\]

\[
c_{ij}^{\alpha \beta}(\tau) = \langle \delta S_i^\alpha(t) \delta S_j^\beta(t + \tau) \rangle, \tag{16}
\]

In what follows, it is convenient to define \(c_{ii}^{\alpha \beta} = 0\).

Because of the randomness of the connectivity, the number of excitatory and inhibitory inputs varies from neuron to neuron, resulting in heterogeneous firing rates between neurons even when they belong to the same population \([46,51]\). This structural randomness also results in heterogeneity of the autocorrelations of single-neuron activities. It also contributes to the heterogeneity in pair cross-correlations. The latter heterogeneity is further enhanced by the spatial variability in the number of inputs shared by the pair of neurons. A full characterization of the distributions of the auto- and cross-correlations is beyond the scope of this paper. Instead, here we focus on their population averages.

The population-averaged autocorrelation \(A_\alpha\) is given by \(A_\alpha(\tau) = (1/N) \sum_i a_{ii}^\alpha(\tau)\), which in the thermodynamic limit is also \(A_\alpha(\tau) = [a_{ii}^\alpha(\tau)]_J\). With the architecture we use, the probability that neurons \((i, \alpha)\) and \((j, \beta)\) share common inputs from a third neuron \((k, \gamma)\) is

\[
\text{Pr}(\Lambda_{ik}^\alpha = 1 \land \Lambda_{jk}^\beta = 1) = \frac{K^2}{N^2} f_{\gamma \gamma} (\theta^\gamma_k - \theta^\gamma_k)f_{\alpha \beta} (\theta^\alpha_j - \theta^\beta_j). \tag{17}
\]

As is evident from this equation, the fraction of shared inputs averaged over all pairs of neurons separated by the same distance on the ring \(\Delta\) depends only on \(\Delta\). Thus, we define the average cross-correlations as \(C_{ij}^{\alpha \beta}(\Delta, \tau) = \langle c_{ij}^{\alpha \beta}(\tau) \rangle\), where \(\langle \rangle\) denotes the average over pairs with \(|\theta^\alpha_j - \theta^\beta_j| = \Delta\). In the thermodynamic limit, this quantity does not depend on the specific realization of the network connectivity matrix. The Fourier expansion of this function is

\[
C_{ij}(\Delta, \tau) = \sum_{n=0}^{N-1} C_{ij}^{(n)}(\tau) e^{-in\Delta}, \tag{18}
\]

where \(C_{ij}^{(n)}(\tau) = (1/N^2) \sum_{k \neq j} c_{ij}^{\alpha \beta}(\tau) e^{in(\theta^\gamma_k - \theta^\gamma_k)}\).

Equation (9), which determines the population-averaged firing rates, stems from the constraint that the net input in every neuron must be \(O(1)\) when \(K\) is large. The condition that for any pair of neurons the correlation in their inputs is finite in that limit leads to another constraint, which we now derive.

Let us consider the \(ND \times ND\) matrix \(Q\) defined by

\[
Q_{ij}^{\alpha \beta} = \langle [\delta h_i^\alpha(t) \delta h_j^\beta(t)]_J \rangle, \tag{19}
\]

for \((i, \alpha) \neq (j, \beta)\) and \(Q_{ii}^{\alpha \beta} = 0\). Using Eq. (7), one finds

\[
Q_{ij}^{\alpha \beta} = \sum_{\gamma \delta} \sum_{\gamma \delta} J_{\gamma \delta}^{\alpha \beta}/K \Lambda_{ij}^{\gamma \delta} [\delta S_i^{\gamma}(t) \delta S_j^{\delta}(t)]_J, \tag{20}
\]

for \((i, \alpha) \neq (j, \beta)\). Similar to \(C_{ij}^{\alpha \beta}\), \(Q_{ij}^{\alpha \beta}\) is a function of \(\Delta = |\theta^\alpha_i - \theta^\beta_j|\) only.

The Fourier series expansion of Eq. (20) in Fourier space \((Q_{ij}^{(n)} = (1/N^2) \sum_{\gamma \delta} \sum_{\gamma \delta} J_{ij}^{(n)} e^{in(\theta^\gamma_i - \theta^\beta_j)}\) yields

\[
Q_{ij}^{(n)} = K \sum_{\gamma \delta} J_{ij}^{(n)} e^{in(\theta^\gamma_i - \theta^\beta_j)} C_{ij}^{\gamma \delta} + \sum_{\gamma} K J_{ij}^{(n)} e^{in(\theta^\gamma_i - \theta^\beta_j)} A, \tag{21}
\]

which can be written in matrix form as

\[
Q^{(n)} = K J^{(n)} C^{(n)} [J^{(n)}]^\dagger + \frac{K}{N} J^{(n)} A [J^{(n)}]^\dagger. \tag{21}
\]

Here, \(X\) denotes the \(D \times D\) matrix \(X_{\alpha \beta} = A_{\alpha \beta} \delta_{\alpha \beta}\), and the superscript \(\dagger\) denotes Hermitian conjugation \((J_{ij}^{(n)})^\dagger = J_{ij}^{(-n)}\).

Note that the diagonal of the matrix \(Q\) is not the variance of the inputs. This diagonal is (see Appendix E4)

\[
\sigma_{\alpha}^2 = K \sum_{n \neq 0} J_{ij}^{(n)} e^{in(\theta^\gamma_i - \theta^\beta_j)} C_{ij}^{\gamma \delta} + \sum_{\gamma} J_{ij}^{(n)} A. \tag{22}
\]

The requirement that cross-correlations of the inputs into pairs of neurons are at most \(O(1)\) implies that all the quantities \(Q_{ij}^{(n)}\) are also at most \(O(1)\). This requirement yields

\[
J^{(n)} C^{(n)} [J^{(n)}]^\dagger = O(1/K). \tag{23}
\]

We call Eq. (23) the balanced correlation equation. It implies that for all \(n\) for which the matrix \(J^{(n)}\) is invertible and has entries \(O(1)\), \(C^{(n)}\) is smaller or equal to \(O(1/K)\) and is thus weak. For a broad class of network architectures, we show below that the correlations are in fact \(O(1/N)\) and barely depend on \(K\) for large \(K\). When, however, \(J^{(n)}\) is singular for some \(n \geq 0\), the correlations can be larger than \(O(1/K)\). In fact, we show that in those cases, the correlations can even increase with \(K\).

Finally, we note that one can also write a balanced correlation equation for the quenched disorder of the neural activity using the fact that \([\Delta h_i^\alpha \Delta h_j^\beta]_J^{(n)}\) also must be at most \(O(1)\). This requirement leads to the condition

\[
J^{(n)} C^{(n)} [J^{(n)}]^\dagger = O(1/K). \tag{23}
\]
\[ J^{(n)} \Gamma^{(n)} [J^{(n)}]^\dagger = \mathcal{O}(1/K) \]  
(24)

with

\[ \Gamma^{(n)} \equiv [\Delta S^\alpha_i \Delta S^\beta_j]^{(n)}. \]  
(25)

IV. CORRELATIONS IN TWO-POPULATION NETWORKS

In Appendix A, we derive an equation for the spatial Fourier modes of the equal-time quenched average correlation functions, \( C_{\alpha \beta}(\Delta, 0) \). It yields (omitting the second argument)

\[ C^{(n)}_{\Delta} = \frac{A_E}{N} \frac{1}{N} \frac{-2A_E + A_E(T^{(n)} + 2\Delta^{(n)})\sqrt{K} - \{A_E[\Delta^{(n)} + (\Delta^{(n)})^2] + A_I(\Delta^{(n)})^2\} K}{-2 + 3T^{(n)}\sqrt{K} - [(T^{(n)})^2 + 2\Delta^{(n)}]K + T^{(n)}\Delta^{(n)}K^{3/2}}, \]

\[ C^{(n)}_{\Delta} = -\frac{1}{N} \frac{2A_I}{N} \frac{-A_I(\Delta^{(n)})^2 + A_I[\Delta^{(n)} + (\Delta^{(n)})^2] + A_E(\Delta^{(n)})^2\} K}{-2 + 3T^{(n)}\sqrt{K} - [(T^{(n)})^2 + 2\Delta^{(n)}]K + T^{(n)}\Delta^{(n)}K^{3/2}}, \]

\[ C^{(n)}_{\Delta} = -\frac{1}{N} \frac{-2A_I}{N} \frac{-A_I(\Delta^{(n)})^2 + A_I[\Delta^{(n)} + (\Delta^{(n)})^2] + A_E(\Delta^{(n)})^2\} K}{-2 + 3T^{(n)}\sqrt{K} - [(T^{(n)})^2 + 2\Delta^{(n)}]K + T^{(n)}\Delta^{(n)}K^{3/2}}, \]

(27)

where \( T^{(n)} \) and \( \Delta^{(n)} \) are the trace and the determinant of the matrix \( \mathcal{J}^{(n)} \).

The expansion of these expressions for large \( K \) gives

\[ C^{(n)}_{\Delta} = \frac{A_E}{N} \frac{1}{N} \frac{A_I(\Delta^{(n)})^2 + A_I[\Delta^{(n)} + (\Delta^{(n)})^2]}{T^{(n)}\Delta^{(n)}} + \mathcal{O}\left(\frac{1}{NK}\right), \]

\[ C^{(n)}_{\Delta} = \frac{1}{N} \frac{A_E}{N} \frac{A_J(\Delta^{(n)})^2 + A_I[\Delta^{(n)} + (\Delta^{(n)})^2]}{T^{(n)}\Delta^{(n)}} + \mathcal{O}\left(\frac{1}{NK}\right), \]

\[ C^{(n)}_{\Delta} = \frac{A_I}{N} \frac{-A_I(\Delta^{(n)})^2 + A_I[\Delta^{(n)} + (\Delta^{(n)})^2]}{T^{(n)}\Delta^{(n)}} + \mathcal{O}\left(\frac{1}{NK}\right), \]

(28)

Thus, in general, when \( K \) is large \( C^{(n)}_{\Delta} \) and \( C^{(n)}_{\Delta} \) are very small, namely, \( \mathcal{O}(1/N) \), with negative prefactors which do not depend on \( K \). As for \( C^{(n)}_{\Delta} \), it is smaller than \( C^{(n)}_{\Delta} \) by a factor of 1/\( \sqrt{K} \).

It should be noted that to derive Eq. (28), we assume that \( T^{(n)} \neq 0 \) and \( \Delta^{(n)} \neq 0 \). Equation (27) indicates, however, that when \( T^{(n)} = 0 \) and \( \Delta^{(n)} = 0 \), or \( T^{(n)} = 0 \) and \( \Delta^{(n)} = 0 \), \( C_{\Delta}^{(n)} \), \( C_{\Delta}^{(n)} \), and \( C_{\Delta}^{(n)} \) are \( \mathcal{O}(1/N) \).

The situation is different if \( T^{(n)} = 0 \) and \( \Delta^{(n)} = 0 \). Equation (27) shows that in this case it is possible to get correlations which are \( \mathcal{O}(K/N) \).

In the rest of this section, we consider in detail two-population networks in which for the probabilities of connection only the first two Fourier modes are nonzero

\[ P^{\alpha \beta}_{ij} = \frac{K}{N} [1 + 2f^{(1)}(\cos(\theta_i - \theta_j^\beta))], \]

(29)

with \( \alpha = E, I, \beta = E, I \).

We also show that the solution of this equation is a fixed point of the dynamics of the correlations which is always stable when the balanced state is stable with respect to perturbation in the population rates (see Appendix B).

This equation holds for a network with an arbitrary number of neuronal populations. In the case of two populations, it can be solved explicitly, yielding after a straightforward calculation (see Appendix E)

\[ \mathbf{\text{A. \ } T^{(1)} \neq 0 \text{ and } \Delta^{(1)} \neq 0} \]

For \( J^{(1)}_{\alpha \beta} \) such that \( T^{(1)} \neq 0 \) and \( \Delta^{(1)} \neq 0 \), we have in the large-\( N, K \) limit

\[ C^{(1)}_{\Delta} = -\frac{A_E}{N} \frac{1}{N} \frac{1 + 2\cos(\Delta^{(1)}) + \mathcal{O}(1)}{\sqrt{K}}, \]

\[ C^{(1)}_{\Delta} = \frac{1}{N} \frac{1}{N} \frac{C^{(0)} + 2C^{(1)}_{\Delta} \cos(\Delta^{(1)}) + \mathcal{O}(1)}{\sqrt{K}}, \]

\[ C^{(1)}_{\Delta} = -\frac{A_I}{N} \frac{1}{N} \frac{1 + 2\cos(\Delta^{(1)}) + \mathcal{O}(1)}{\sqrt{K}}, \]

(30)

where \( C^{(1)}_{\Delta} = N\sqrt{K}C^{(1)}_{\Delta} \) are \( \mathcal{O}(1) \). Thus, in that limit, the spatial average and modulation of the correlations within the \( E \) and \( I \) populations do not depend on \( K \) at the leading order. Moreover, \( C^{(0)}_{\Delta} \) and \( C^{(1)}_{\Delta} \) depend on the
synaptic strengths only because the autocorrelations $A_E$ and $A_I$ depend on these parameters.

Figure 2 depicts the simulation results for $N = 40,000$ and $K = 2000$. Figure 2(a) plots $C_{EE}(\Delta)$ for $f_{\alpha \beta}^{(1)} = 0.25$ ($\alpha, \beta \in E, I$) and two sets of values for the interaction strengths (solid and dashed lines). For comparison, we also plot the results of a simulation when the connectivity is unstructured ($f_{\alpha \beta}^{(1)} = 0; \alpha, \beta \in E, I$; gray line). In all these cases, $C_{EE}(\Delta)$ is very small (note the scale on the $y$ axis). When the connectivity is spatially modulated, $C_{EE}(\Delta)$ varies with the distance. However, the spatial averages are comparable in the three cases considered $[[1/(2\pi)]] \int C_{EE}(\Delta) d\Delta \sim 0.2 \times 10^{-5}$. This is in agreement with Eq. (30) since, to the leading order, the autocorrelations $A_E$ and $A_I$ are not expected to depend on whether the connectivity is spatially modulated or not.

Note that according to Eq. (30), $C_{EE}(\Delta)$ and $C_{II}(\Delta)$ are negative for close-by neurons ($\Delta$ small) and positive for neurons that are far apart. This is the case for the set of parameters corresponding to the solid red line in Fig. 2(a) [see, also, Figs. 2(b)–2(d)]. However, when finite-$K$ corrections are not negligible, which happens when $|T^{(1)}(\Delta)|$ is sufficiently small, short-range correlations can be positive and longer-range correlations negative [Eq. (30)]. This is the case in Fig. 2(a), dashed line.

Figure 2(b) compares the simulations (circles) and analytical results (solid lines) for the dependence on $N$ of $NC_{\alpha \beta}^I$ [parameters as in Fig. 2(a), red solid line]. It shows that the spatial modulation of the correlations in the simulation is close to the large-$K$ analytical results. Figures 2(c) and 2(d) depict the dependence of $NC_{\alpha \beta}^I$ on $K$. In the whole range of $K$ considered here, the simulations and analytical results are close. For $|C_{\alpha \alpha}|$, the nearby correlations and modulation amplitudes increase with $K$ and are much larger than that of $|C_{EI}|$.

B. $T^{(1)} = \Delta^{(1)} = 0$

We now consider a network in which $f_{EI}^{(1)} \neq 0$ and $f_{EE}^{(1)} = f_{HI}^{(1)} = f_{IE}^{(1)} = 0$. The spatially modulated component of the interaction has, therefore, an explicit feed-forward structure [Fig. 3(a), top panel] and $T^{(1)} = \Delta^{(1)} = 0$.

Solving Eq. (27) shows that the correlations are on average $\mathcal{O}(1/N)$ and that their modulations are

$$
C_{EE}^{(1)} = \frac{KA_I}{N} |\mathcal{J}_{EI}^{(1)}|^2, \\
C_{EI}^{(1)} = -\frac{\sqrt{K}A_I}{N} |\mathcal{J}_{EI}^{(1)}|, \\
C_{HI}^{(1)} = 0.
$$

As a result, the correlations in the $E$ population are spatially modulated and $\mathcal{O}(K/N)$. They are positive for short range and negative for long range. This is in contrast to the correlations in the inhibitory population which are not spatially modulated and $\mathcal{O}(1/N)$, while the $EI$ correlations are spatially modulated and $\mathcal{O}(\sqrt{K}/N)$.

Figure 3 compares the analytical results with simulations. Figure 3(a) plots the simulation results for $C_{\alpha \beta}^{(1)}(\Delta)$ for $f_{EI}^{(1)} = 0.25$. The other parameters are as in Fig. 2(a) (gray solid line). Thus, the locally averaged firing rates are the same as in the simulations in the latter figure. The correlations in the inhibitory neurons (blue) are extremely weak and are not spatially modulated. In contrast, the correlations of excitatory pairs (red) are larger by 2 orders of magnitude compared to those in Fig. 2. The correlations between the $E$ and $I$ neurons (green) are weaker than those between the $E$ neurons. They are negative for nearby neurons, while for nearby excitatory pairs they are positive.
linearly with this parameter. Our simulations are in very good agreement with these analytical results [Fig. 3(d)].

C. $\Delta^{(1)} = 0$, $T^{(1)} \neq 0$

The network investigated in Sec. IV B has an explicit feed-forward structure. Adding any spatial modulation to the $II$ connectivity [$f_{II}^{(1)} \neq 0$, top panel of Fig. 4(a)] destroys this structure and now $T^{(1)} \neq 0$ (while $\Delta^{(1)} = 0$). Solving Eq. (26) for that case, one finds

$$
C_{EE}^{(1)} = \frac{K}{N} A_I \frac{|\bar{J}_{EE}^{(1)}|}{(2 + \sqrt{K} |\bar{J}_{II}^{(1)}|)(1 + \sqrt{K} |\bar{J}_{II}^{(1)}|)} ,
$$

$$
C_{EI}^{(1)} = -\frac{K}{N} A_I \frac{|\bar{J}_{EI}^{(1)}|}{\sqrt{2 + \sqrt{K} |\bar{J}_{II}^{(1)}|}(1 + \sqrt{K} |\bar{J}_{II}^{(1)}|)} ,
$$

$$
C_{II}^{(1)} = \frac{K}{N} A_I \frac{|\bar{J}_{II}^{(1)}|^2}{\sqrt{1 + \sqrt{K} |\bar{J}_{II}^{(1)}|}} .
$$

(32)

FIG. 3. Correlations in a two-population network with an explicit feed-forward structure. Connection probabilities as in Eq. (29). Parameters: $f_{EE}^{(1)} = f_{IE}^{(1)} = f_{II}^{(1)} = 0$. Panels (a)–(c): $f_{EI}^{(1)} = 0.25$. Interaction strengths as in Fig. 2. (a) Top panel: The network architecture. Green: The connections which are spatially modulated ($f_{EI}^{(1)} \neq 0$). Connections in gray are unstructured ($f_{EI}^{(1)} = 0$). Bottom panel: Simulation results for $C_{\alpha\beta}(\Delta)$; $N = 40000$, $K = 2000$. (b), (c) $NC_{\alpha\beta}$ vs $K$. Solid lines: Analytical results, Eq. (31). Simulation results are plotted for $N = 20000$ (plus), $N = 40000$ (circles), and $N = 80000$ (crosses). (b) $NC_{EE}^{(1)}$. (c) Blue, $NC_{II}^{(1)}$. Green, $NC_{EI}^{(1)}$. (d) $NC_{\alpha\beta}$ vs $f_{EI}^{(1)}$. Solid lines: Analytical results, Eq. (31); $N = 40000$, $K = 400$.

All these features are in agreement with our analytical results.

The spatial modulation of the $EE$ correlations increase with $K$ [Fig. 3(b), circles]. There is a quantitative agreement between simulations and theory [Eq. (31)] up to $K \approx 4000$ for $N = 40000$. In this range, $C_{EE}^{(1)}$ in the simulations varies linearly with $K$. For larger values of $K$, $C_{EE}^{(1)}$ is larger than predicted by Eq. (31). This is because this equation is derived by linearizing the dynamics, which is only valid when the correlations are not too large. In fact, the simulation results for fixed $K$ deviate less from Eq. (31) when $N$ is increased [Figs. 3(b) and 3(c)]. When $K$ increases, $C_{EI}^{(1)}$ becomes more negative [Fig. 3(c), green]. Here too, for $N = 40000$ the simulations agree well with Eq. (31) up to $K \approx 4000$ and deviations are smaller when $N$ is larger. The spatial modulation of the $II$ correlations in the simulations is extremely small [Fig. 3(c), blue], as the theory predicts.

Finally, according to Eq. (31), the spatial modulation $C_{EE}^{(1)}$ increases quadratically with $f_{EI}^{(1)}$, whereas $C_{EI}^{(1)}$ varies linearly with this parameter. Our simulations are in very good agreement with these analytical results [Fig. 3(d)].

FIG. 4. Spatial modulation in the $II$ interactions suppresses the correlations in the $E$ populations. Same network as in Fig. 3 but $f_{II}^{(1)} = 0.25$, $N = 40000$. (a) Top panel: The network architecture. Interactions plotted in green are spatially modulated. Bottom panel: Simulation results for $C_{EE}(\Delta)$. $N = 40000$, $K = 2000$. Gray line, $f_{II}^{(1)} = 0$; dashed red, $f_{II}^{(1)} = 0.05$; solid red, $f_{II}^{(1)} = 0.25$. (b) $NC_{\alpha\beta}$ vs $f_{II}^{(1)}$. Circles: Simulations. Solid lines: Solution of Eq. (32). $N = 40000$, $K = 400$. Colors are as in Fig. 3. (c) Left panel: $NC_{EE}^{(1)}$ vs $K$. Solid lines: Eq. (32). Circles: Simulations. Gray line, $f_{II}^{(1)} = 0$; dotted red, $f_{II}^{(1)} = 0.025$; dashed red, $f_{II}^{(1)} = 0.05$; solid red, $f_{II}^{(1)} = 0.25$. Right panel: Same as left panel but in a log-log scale.
In the large-\(N\), \(K\) limit, \(C_{EE}(\Delta)\) and \(C_{II}(\Delta)\) are both \(O(1/N)\) and do not depend on \(K\), whereas \(C_{EI}(\Delta)\) is \(O[1/(N\sqrt{K})]\). Therefore, the addition of a spatial modulation in the \(II\) interactions suppresses the correlations that inhibitory projections induce in the \(E\) population.

To understand further the origin of this suppression, let us consider the quenched average correlations of the inputs \(Q_{\alpha\beta}(\Delta)\) [Eq. (19)]. Using Eq. (21), one finds

\[
Q_{EE}^{(1)} = K(J_{EI}^{(1)})^2 \left( C_{II}^{(1)} + A_I \right) /
N \left( 1 + \sqrt{K} |J_{II}^{(1)}| \right)
\]

When \(J_{II}^{(1)} = 0\), \(C_{II}^{(1)} = 0\), therefore, \(Q_{EE}^{(1)} = K(J_{EI}^{(1)})^2 A_I\).

On the other hand, when \(J_{II}^{(1)} \neq 0\), \(C_{II}^{(1)}\) is given by Eq. (32). Thus, inserting Eq. (32) to the equation above yields:

\[
Q_{EE}^{(1)} = A_I / N \left( 1 + \sqrt{K} |J_{II}^{(1)}| \right)
\]

Thus, when \(II\) interactions are spatially modulated, a cancellation between terms which are \(O(K/N)\) reduces \(Q_{EE}^{(1)}\) by a factor of \(O(1/\sqrt{K})\). As a result, \(C_{EE}^{(1)}\) is much smaller than when \(II\) interactions are not modulated.

A similar argument explains the suppression in \(C_{EI}^{(1)}\).

Figure 4(a) depicts the simulation results for \(f_{EI}^{(1)} = 0.25\) and three values of \(f_{II}^{(1)}\). It demonstrates the suppression of correlations in the excitatory population when \(II\) interactions are also modulated. The dependence of this effect on \(f_{II}^{(1)}\) is depicted in Fig. 4(b). Increasing \(f_{II}^{(1)}\) decreases the modulation of all the correlations, in very good agreement with the analytical results (compare circles and solid lines).

According to Eq. (32), the correlation in the \(E\) population always increases linearly with \(K\) for small \(K\). The crossover between this regime, where \(C_{EE}(\Delta) = O(K/N)\), and the large-\(K\) regime, where \(C_{EE}(\Delta) = O(1/N)\), occurs for \(\sqrt{K} \approx 1/J_{II}^{(1)}\). Figure 4(c) depicts this crossover in numerical simulations. Thus, although in the large-\(N\), \(K\) limit a transition from \(O(K/N)\) to \(O(1/N)\) correlations occurs as soon as \(J_{II}^{(1)} \neq 0\), this qualitative difference is significant only when \(K\) is sufficiently large. In other words, for a moderately large number of inputs per neuron, correlations can exhibit a close to linear increase even if the structure of spatial modulation of the interaction matrix is not completely feed forward.

## V. SCALING CORRELATION THEOREMS

In the previous section, we study networks with two neuronal populations. In this case, it is straightforward to analytically derive explicit expressions for the correlations. These expressions are simple enough to fully classify how the network structure affects the scaling (with \(K\) and \(N\)) of the correlations. For networks consisting of more than two populations, analytical expressions for the correlations can, in principle, be derived. However, dealing with these expressions becomes rapidly impractical as the number of populations increases. In the following, we adopt an alternative approach. We prove two general theorems, which, given the network architecture, allow us to determine how correlations vary with \(K\) without computing these correlations explicitly.

To prove these theorems, we rewrite Eq. (26) in the Jordan basis of \(\hat{J}^{(n)}\). We write

\[
\hat{J}^{(n)} = U^{(n)} J_{\text{for}}^{(n)} U^{(n)\dagger},
\]

\[
[\hat{J}^{(n)}]_{\mu\nu} = V^{(n)} [J_{\text{for}}^{(n)}]_{\mu\nu} V^{(n)\dagger}.
\]

(34)

Here, \(x^\dagger\) denotes the complex conjugate of \(x\), \(U^{(n)} (V^{(n)})\) are matrices whose rows are the generalized eigenvectors of \(\hat{J}^{(n)} ([\hat{J}^{(n)}]_{\mu\nu})^\dagger\), and \(J_{\text{for}}^{(n)}\) is the Jordan normal form of \(\hat{J}^{(n)}\). With these definitions, \([J_{\text{for}}^{(n)}]_{\mu\nu}\) is the Jordan form of \([\hat{J}^{(n)}]_{\mu\nu}\). We can write

\[
[J_{\text{for}}^{(n)}]_{\mu\nu} = \lambda_\mu^{(n)} \delta_{\mu\nu} + \epsilon_\mu^{(n)} \delta_{\mu\nu-1}
\]

with \(\mu, \nu \in 1, \ldots, D\) and \(\epsilon_\mu^{(n)} = 1\) within a Jordan block and zero otherwise.

Equation (26) then yields

\[
\{ 2 - \sqrt{K} (\lambda_\mu^{(n)} + [\lambda_\mu^{(n)}]^\dagger) \} \hat{\lambda}_\mu^{(n)}
\]

\[
= \sqrt{K} (\epsilon_\mu^{(n)} \hat{\lambda}_\mu^{(n)} + \epsilon_{\mu-1}^{(n)} \hat{\lambda}_{\mu-1}^{(n)} + \hat{\lambda}_{\mu-1}^{(n)} - \hat{\lambda}_\mu^{(n)} + [\hat{\lambda}_\mu^{(n)}]^\dagger)\]

\[
\quad + \hat{\lambda}_\mu^{(n)} + \epsilon_\mu^{(n)} \delta_{\mu\nu} + \epsilon_{\mu-1}^{(n)} \delta_{\mu\nu-1} + \hat{\lambda}_\mu^{(n)} - \hat{\lambda}_\nu^{(n)} + [\hat{\lambda}_\nu^{(n)}]^\dagger \hat{\lambda}_\mu^{(n)},
\]

(35)

where we define

\[
\hat{C}^{(n)} = [U^{(n)}]^{-1} C^{(n)} V^{(n)},
\]

\[
\hat{A}^{(n)} = [U^{(n)}]^{-1} A^{(n)} V^{(n)}.
\]

(36)

Note that while the matrix \(C^{(n)}\) is symmetric and the matrix \(A^{(n)}\) is diagonal, this is not in general the case for \(\hat{C}^{(n)}\) and \(\hat{A}^{(n)}\).

031072-9
Let us assume that the network is in a stable balanced state in which the matrix \( \hat{\mathbf{A}}^{(n)} \) has no zero elements. In Appendix B, we prove the following:

**Correlation Theorem 1.** The \( n \)th Fourier mode of the correlation matrix scales as \( \mathcal{O}(1/N) \) if and only if \( \mathbf{J}^{(n)}_{\text{lor}} \) does not have a Jordan block whose real part is a shift matrix (a shift matrix \( \mathbf{S} \) of dimension \( P \) is a \( P \times P \) matrix of the form \( S_{\mu\nu} = \delta_{\mu+1,\nu} \)).

**Correlation Theorem 2.** If \( \mathbf{J}^{(n)} \) has at least one Jordan block whose real part is a shift matrix, the \( n \)th Fourier mode of the correlation matrix is \( \mathcal{O}(K^{P(n)-1}/N) \), where \( P(n) \) is the dimension of the largest block in \( \mathbf{J}^{(n)}_{\text{lor}} \) whose real part is a shift matrix.

**Corollary 1.** The \( n \)th Fourier mode of the correlation matrix is \( \mathcal{O}(K^{D-1}/N) \) if and only if \( \mathbf{J}^{(n)} \) is nilpotent of degree \( D \).

In Appendix B, we also show how to extend these results to a case where \( \hat{\mathbf{A}}^{(n)} \) has zero elements.

The matrix \( \mathbf{U}^{(n)} \) can be viewed as a transformation of the original network of \( D \) populations into a network of \( D \) effective populations. The condition that \( \mathbf{J}^{(n)} \) has a Jordan block which is a shift matrix of size \( P \) can be interpreted as the existence of \( P \) effective populations whose effective interaction matrix is feed forward in its \( n \)th Fourier mode. In other words, the original network exhibits a hidden feed-forward structure, which is embedded in the \( n \)th Fourier mode of its connectivity. Theorems 1 and 2, therefore, imply that only when such a structure exists, the \( n \)th mode of the correlation matrix increases with \( K \). To know which elements in this matrix increase with \( K \), one has to compute the matrices \( \mathbf{U}^{(n)} \) and \( \mathbf{V}^{(n)} \) [see Eq. (36)].

### VI. APPLICATIONS OF THE CORRELATION THEOREMS

In this section, we consider networks comprising \( D \) populations with connection probabilities

\[
P_{\alpha\beta}^{(a\beta)} = \frac{K}{N} \left[ 1 + 2J_{\alpha\beta}^{(a\beta)} \cos(\theta_i^a - \theta_j^b) \right]
\]

with \( \alpha = 1, \ldots, D \) and \( \beta = 1, \ldots, D \).

#### A. Two-population networks

For a network of two populations, the Jordan form of the matrix \( \mathbf{J}^{(1)} \) has the form \( \mathbf{J}^{(1)}_{\text{lor}} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \) if \( \mathbf{J}^{(1)} \) is diagonalizable. Otherwise, it has the form \( \mathbf{J}^{(1)}_{\text{lor}} = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix} \) with \( \alpha \) real. Theorems 1 and 2 imply that only in the second case with \( \lambda = 0 \), some of the correlations \( C_{\alpha\beta}(\Delta) \) are \( \mathcal{O}(K/N) \). Otherwise, all correlations are \( \mathcal{O}(1/N) \). It is equivalent to say that some correlations are \( \mathcal{O}(K/N) \) if and only if \( \mathbf{J}^{(1)} \neq 0 \) and \( T^{(1)} = \Delta^{(1)} = 0 \). We therefore recover the result we derived in Sec. IV without explicit computation of the correlations.

As noted above, the correlation theorems do not tell us which of the elements of the correlation matrix are \( \mathcal{O}(K/N) \). However, since a \( 2 \times 2 \) matrix has such a Jordan form if and only if it is nilpotent \( ([\mathbf{J}^{(1)}]_2^2 = 0) \), we have to consider two types of connectivity:

**Type 1:** The network has an explicit feed-forward structure; i.e., the interaction matrix is either \( \tilde{\mathbf{J}}^{(1)} = \begin{pmatrix} J_{\alpha\beta}^{(1)} & 0 \\ 0 & J_{\beta\gamma}^{(1)} \end{pmatrix} \) or \( \tilde{\mathbf{J}}^{(1)} = \begin{pmatrix} J_{\alpha\beta}^{(1)} & 0 \\ 0 & J_{\beta\gamma}^{(1)} \end{pmatrix} \).

In the former case, the matrix \( \mathbf{U}^{(1)} \) is \( \mathbf{U}^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & J_{\beta\gamma}^{(1)} \end{pmatrix} \), whereas \( \mathbf{V}^{(1)} = \begin{pmatrix} 0 & 1/J_{\alpha\beta}^{(1)} \\ 1/J_{\beta\gamma}^{(1)} & 0 \end{pmatrix} \). Then, Eq. (36) gives [see Appendix B, Eq. (B8)]

\[
\hat{\mathbf{C}}^{(1)} = \begin{pmatrix} \mathcal{O}(\sqrt{K}/N) & \mathcal{O}(K/N) \\ 0 & \mathcal{O}(\sqrt{K}/N) \end{pmatrix}
\]

Using Eq. (36), one finds that \( C_{\lambda\lambda}^{(1)} = \mathcal{O}(K/N) \), whereas \( C_{\alpha\beta}^{(1)} = 0 \), and \( C_{\alpha\beta}^{(1)} = \mathcal{O}(\sqrt{K}/N) \), in agreement with Eq. (31). A similar calculation in the latter case (when the modulation in the \( IE \) interactions) gives \( C_{\lambda\lambda}^{(1)} = 0 \), whereas \( C_{\alpha\beta}^{(1)} = \mathcal{O}(K/N) \), and \( C_{\alpha\beta}^{(1)} = \mathcal{O}(\sqrt{K}/N) \), all in agreement with Eq. (27).

**Type 2:** \( \mathbf{J}^{(1)} = c \begin{pmatrix} 1 & -a/c \\ -a/c & -1 \end{pmatrix} \), where \( a, c > 0 \).

In this case, the network has no explicit feed-forward structure since all four interactions (\( EE, EI, IE, II \)) are spatially modulated. It has, however, a hidden feed-forward structure, as revealed by the Jordan form of the interaction matrix.

For this network, the transformation matrices are \( \mathbf{U}^{(1)} = \begin{pmatrix} a/c & 0 \\ 0 & -c/a \end{pmatrix} \) and \( \mathbf{V}^{(1)} = \begin{pmatrix} -c/a & 0 \\ 0 & a/c \end{pmatrix} \). Using Eqs. (36) and (38), it is clear that the transformation from \( \hat{\mathbf{C}}^{(1)} \) to \( \mathbf{C}^{(1)} \) mixes elements which are 0 and \( \mathcal{O}(\sqrt{K}/N) \) with those which are \( \mathcal{O}(K/N) \). Thus, while in \( \hat{\mathbf{C}}^{(1)} \), only the element \( C_{12}^{(1)} = \mathcal{O}(K/N) \), all the elements of the correlation matrix \( \mathbf{C}^{(1)} \) are \( \mathcal{O}(K/N) \). This is also in line with Eq. (27).

We consider an example of such a network in Fig. 5. The parameters \( J_{\alpha\beta} \) and the external inputs are as in Figs. 2–4. Therefore, to leading order, the population-averaged activities \( m_{\alpha\gamma} \), the autocorrelations \( A_{\alpha\gamma} \), and the population gains \( g_{\alpha\gamma} \) are the same as in Figs. 2–4. The modulation of the connection probability \( J_{\alpha\beta}^{(1)}, J_{\beta\gamma}^{(1)} \) are all nonzero [Fig. 5(a)] and tuned so that

\[
\mathbf{J}^{(1)} = \frac{1}{20} \begin{pmatrix} 1 & -1/2 \\ 2 & -1 \end{pmatrix}
\]

The Jordan form of this matrix is graphically represented in Fig. 5(b). The top panel in Fig. 5(c) depicts the
simulation results for the correlations in this network. They are all positive for close-by neurons, and their spatial modulations increase approximately linearly with $K$ [Fig. 5(c), bottom]. The top panel of Fig. 5(d) shows that most of the power in these correlations results from the element $\tilde{C}_{12}$ (red). The latter correlation element increases linearly with $K$ [Fig. 5(d), bottom], while $\tilde{C}_{11}$, $\tilde{C}_{22} = \mathcal{O}(\sqrt{K}/N)$, and $\tilde{C}_{21}$ is 2 orders of magnitude smaller and does not exhibit significant change with $K$ (note the log-log scale in this figure). These simulations are in quantitative agreement with Eq. (27) up to $K \approx 4000$, and the deviations for larger $K$ decrease with $N$ (as in Fig. 4).

The interaction matrix $\mathcal{J}^{(1)}$ depends on the matrix $J^{(1)}$ and on the population gains $g_E$ and $g_I$. If the connectivity matrix has an explicit feed-forward structure, the interaction matrix has also such a structure independent of the population gains. Thus, although changing the external inputs $I_E$, $I_I$ modifies this gain, this does not destroy the feed-forward structure and, thus, does not change the scaling of the correlations with $K$ and $N$. In contrast, in networks with a hidden feed-forward structure, this scaling is sensitive to perturbations in the external inputs since the hidden feed-forward structure is destroyed unless the ratio $g_E/g_I$ remains constant.

**B. Examples with three populations or more**

The two networks depicted in Fig. 6 consist of one excitatory and two inhibitory populations. The positivity of the spatially averaged activity of these three populations $m_{n}^{(0)} (\alpha = 1, 2, 3)$ constrains the parameters (see Sec. III) $J_{\alpha \beta}$ through a set of inequalities. We leave the calculation of these conditions to the reader.

In both cases, the first Fourier mode of the population-averaged connectivity matrix has the form

$$J^{(1)} = \begin{pmatrix} 0 & J_{12}^{(1)} & J_{13}^{(1)} \\ 0 & J_{22}^{(1)} & J_{23}^{(1)} \\ 0 & 0 & 0 \end{pmatrix} \quad (39)$$

with $J_{12}^{(1)}$, $J_{13}^{(1)}$, $J_{23}^{(1)} < 0$ [left panels of Figs. 6(a)–6(c)].

In the network of Fig. 6(a), $J_{22}^{(1)} = 0$. Therefore $J^{(1)}$ and $\mathcal{J}^{(1)}$ are nilpotent of degree 3. According to Corollary I of Sec. V, the first mode of the correlation matrix is $\mathcal{O}(K^2/N)$.

**FIG. 5.** Correlations in a two-population network with a hidden feed-forward structure. (a) Network architecture. All connections are modulated. Connection probabilities are as in Eq. (29) with $f_{E,E}^{(1)} = 0.373$, $f_{E,I}^{(1)} = 0.0224$, $f_{I,E}^{(1)} = 0.0487$, $f_{I,I}^{(1)} = 0.1623$. Connection strengths are as in Figs. 2 and 3. (b) Hidden feed-forward structure of the mode $n = 1$ as revealed in the Jordan basis. (c) Top panel: Simulation results for $C_{\alpha \beta}(\Delta)$. $N = 40000$, $K = 2000$. Bottom panel: $NC_{\alpha \beta}(\Delta)$ vs $K$. Solid lines: Analytical results [Eq. (27)]. Simulations are plotted for $N = 20000$ (plus), $N = 40000$ (circles), $N = 80000$ (cross), and $N = 160000$ (asterisks). (d) Top panel: Correlation matrix in the Jordan basis of $\mathcal{J}^{(1)}$. Purple, $\tilde{C}_{11}(\Delta)$; brown, $\tilde{C}_{22}(\Delta)$; gray, $\tilde{C}_{21}(\Delta)$; red, $\tilde{C}_{12}(\Delta)$. Bottom panel: $NC_{\alpha \beta}(\Delta)$ vs $K$ in log-log scale. Solid line: Linear fit. Dashed line: Fit to a square root function.

**FIG. 6.** Examples of networks with three populations and their Jordan representations. Probability of connections are as in Eq. (37). (a),(b) Left: A network of three populations, two inhibitory and one excitatory. Gray: Unstructured connections. Green: Spatially modulated connections. Middle: The Jordan representation of the $n = 1$ Fourier mode of the network connectivity (left panel). Right: The scaling of the strongest correlations. (b) $J_{12}^{(1)}$, $J_{13}^{(1)}$, $J_{23}^{(1)} \neq 0$. Other entries of the matrix $J^{(1)}$ are zero (see main text). Same as (a) but with $J_{22}^{(1)} \neq 0$. 

031072-11
The Jordan form of $\mathcal{J}^{(1)}$ is

$$\mathcal{J}^{(1)}_{\text{Jor}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$ 

This is graphically represented in Fig. 6(a), middle panel. The matrix $\hat{\mathcal{C}}^{(1)}$ satisfies [see Appendix B, Eq. (B8)]

$$\hat{\mathcal{C}}^{(1)} = \begin{pmatrix} \mathcal{O}(K/N) & \mathcal{O}(K^{3/2}/N) & \mathcal{O}(K^2/N) \\ \mathcal{O}(K^{1/2}/N) & \mathcal{O}(K/N) & \mathcal{O}(K^{3/2}/N) \\ 0 & \mathcal{O}(K^{1/2}/N) & \mathcal{O}(K/N) \end{pmatrix}.$$ 

Using the transformation matrices [Eq. (34)], one can show that correlations are $\mathcal{O}(K^2/N)$ only within the excitatory population.

In the network in Fig. 6(b), $J^{(1)}_{22} < 0$. Therefore, the interaction matrix $\mathcal{J}^{(1)}$ is not nilpotent. Its Jordan form is

$$\mathcal{J}^{(1)}_{\text{Jor}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & J^{(1)}_{22} \end{pmatrix}.$$ 

The upper Jordan block is a shift matrix of degree 2. The corresponding feed-forward structure is graphically represented in Fig. 6(b) (middle panel). According to Theorem 2, $\hat{\mathcal{C}}^{(1)}$ is $\mathcal{O}(K/N)$. It satisfies [see Appendix B, Eq. (B8)]

$$\hat{\mathcal{C}}^{(1)} = \begin{pmatrix} \mathcal{O}(\sqrt{K}/N) & \mathcal{O}(K/N) & \mathcal{O}(1/N) \\ 0 & \mathcal{O}(\sqrt{K}/N) & \mathcal{O}(1/N) \\ \mathcal{O}(1/N) & \mathcal{O}(1/N) & \mathcal{O}(1/N) \end{pmatrix},$$

and using the transformation matrices, one can show that only $C^{(1)}_{11}$ is $\mathcal{O}(K/N)$. Other correlations are either $\mathcal{O}(\sqrt{K}/N)$ or $\mathcal{O}(1/N)$.

This approach can be generalized to an arbitrary number of populations to classify the scaling of the correlation matrix for different architectures. Examples of networks with four populations are depicted in Fig. 7 together with the graphic representations of their Jordan forms and the maximum order of the correlations.

### VII. Constraint on Scaling of the Number of Inputs with the Network Size

In this section, we assume that $K$ and $N$ scale together

$$K \propto N^\gamma$$

with $0 < \gamma \leq 1$.

Equation (26), which determines the correlations of the neuronal activities, is obtained under the ansatz that the correlations between the inputs $h^{\alpha}(t)$ and $h^{\beta}(t)$ are sufficiently small, namely, $o(1)$ (see Appendix A). This condition is more stringent than the balance correlation equation, Eq. (23). It can be written as

- (a)–(c) Left: The network consists of two coupled $E$-$I$ networks. Gray: Unstructured connections. Green: Spatially modulated connections. Middle: The Jordan representation of the $n = 1$ Fourier mode of the connectivity (network of the left panel). Right: The scaling of the strongest correlation. (d) Same as (a)–(c) with a population-averaged connectivity matrix as in Eq. (46). Middle: The Jordan form is complex [Eq. (47)]. Black lines correspond to complex eigenvalues.
\[
J^{(n)} C^{(n)} [J^{(n)}]^{\dagger} = o(1/N^r). \tag{41}
\]

This condition constrains \( \gamma \) as we now show.

According to Theorem 1, if the Jordan form \( J_{\text{Jor}}^{(n)} \) has no Jordan block whose real part is a shift matrix for any \( n \), the correlations will be \( O(1/N) \). The scaling of \( J^{(n)} C^{(n)} [J^{(n)}]^{\dagger} \) will also be \( O(1/N) \). Therefore, Eq. \( (41) \) requires only \( \gamma < \gamma_{\text{max}} = 1 \).

If the Jordan form \( J_{\text{Jor}}^{(n)} \) contains a Jordan block whose real part is a shift matrix, we have to apply Theorem 2. In this theorem, \( P(n) \) is the dimension of the largest block in \( J_{\text{Jor}}^{(n)} \) with a real part which is a shift matrix (see Sec. V). Let us denote by \( P_{\text{max}} \) the largest \( P(n) \) over all Fourier modes, i.e.,

\[
P_{\text{max}} = \max_n P(n). \tag{42}
\]

Equation \( (41) \) implies

\[
\hat{J}^{(n)} C^{(n)} [\hat{J}^{(n)}]^{\dagger} = o(N^{-r})
\]

for all \( n \). This yields in the Jordan basis

\[
\sum_{m, k} [J_{\text{Jor}}^{(n)}]_{jm} \hat{C}_{mk} ([J_{\text{Jor}}^{(n)}]^{\dagger})_{ke} = o(N^{-r}) \tag{43}
\]

for all \( \mu, \nu \).

By definition of \( P_{\text{max}} \), for at least one Fourier mode \( n \), the matrix \( J_{\text{Jor}}^{(n)} \) has at least one block whose real part is a shift matrix of degree \( P_{\text{max}} \). In general, there can be several such Jordan blocks. For example, in the network depicted in Fig. 7(b) for which \( P_{\text{max}} = 2 \), there are two Jordan blocks with \( P = 2 \).

We first assume that all blocks which are a shift matrix of size \( P_{\text{max}} \) are real. Since for such blocks \( J_{\text{Jor}} \hat{C} J_{\text{Jor}} \) scale the same with \( K \), it is sufficient to consider the case where there is only one such block. We denote it by \( S \) and by \( \hat{C}_{\text{max}} \) the corresponding block in \( \hat{C} \). Equation \( (41) \) then yields

\[
S \hat{C}_{\text{max}} S = o(N^{-r}). \tag{44}
\]

For instance, for \( P_{\text{max}} = 2 \), we have [see Eq. \( (B8) \)]

\[
\hat{C}_{\text{max}} = \begin{pmatrix}
O(N^{r/2-1}) & O(N^{-1}) \\
0 & O(N^{r/2-1})
\end{pmatrix},
\]

and thus,

\[
S \hat{C}_{\text{max}} S = 0.
\]

Therefore, for this block, Eq. \( (41) \) is always satisfied. The latter equation, however, also applies to other Jordan blocks and Fourier modes. This implies that \( \gamma < \gamma_{\text{max}} = 1 \).

For \( P_{\text{max}} = 3 \), we have

\[
\hat{C}_{\text{max}} = \begin{pmatrix}
O(N^{r-1}) & O(N^{3r/2-1}) & O(N^{2r-1}) \\
O(N^{r-1}) & O(N^{-1}) & O(N^{3r/2-1}) \\
0 & O(N^{r/2-1}) & O(N^{-1})
\end{pmatrix}, \tag{45}
\]

and

\[
S \hat{C}_{\text{max}} S = \begin{pmatrix}
0 & O(N^{r/2-1}) & O(N^{-1}) \\
0 & 0 & O(N^{r/2-1}) \\
0 & 0 & 0
\end{pmatrix}.
\]

Thus, Eq. \( (41) \) is satisfied only if \( \gamma < \gamma_{\text{max}} = 1/2 \).

As shown above, \( \hat{C}_{\text{max}} = O(N^{r(P_{\text{max}}-1)-1}) \) and \( S \hat{C}_{\text{max}} S = O(N^{r(P_{\text{max}}-3/2)-1}) \). It is straightforward to also show that \( \hat{C}_{\text{max}} S - S \hat{C}_{\text{max}} = O(N^{r(P_{\text{max}}-3/2)-1}) \). Therefore, \( J_{\text{Jor}} \hat{C}_{\text{max}} [J_{\text{Jor}}^{\dagger}]^* = o^2 \hat{C}_{\text{max}} + S \hat{C}_{\text{max}} S \).

As shown above, \( \hat{C}_{\text{max}} = O(N^{r(P_{\text{max}}-1)-1}) \) and \( S \hat{C}_{\text{max}} S = O(N^{r(P_{\text{max}}-3/2)-1}) \). It is straightforward to also show that \( \hat{C}_{\text{max}} S - S \hat{C}_{\text{max}} = O(N^{r(P_{\text{max}}-3/2)-1}) \). Therefore, \( J_{\text{Jor}} \hat{C}_{\text{max}} [J_{\text{Jor}}^{\dagger}]^* = O(N^{r(P_{\text{max}}-3/2)-1}) \). Equation \( (41) \) is then satisfied only if \( \gamma < \gamma_{\text{max}} \), with

\[
\gamma_{\text{max}} = \frac{1}{P_{\text{max}} - 1}.
\]
\[ \gamma_{\text{max}} = \frac{1}{P_{\text{max}}}. \] (49)

According to Theorem 2, if \( J_{30}^{(n)} \) has a block whose real part is a shift matrix for at least one mode \( n, C = \mathcal{O}(N^{-\alpha}) \) where \( \alpha = 1 - \gamma(P_{\text{max}} - 1). \) If \( \gamma < \gamma_{\text{max}}, \) the correlations in the activity will decrease more slowly than \( 1/N \) when \( N \) is increased. If \( \gamma > \gamma_{\text{max}}, \) the correlations will increase with \( N, \) and the network will not operate in the balanced regime. Finally, if \( \gamma = \gamma_{\text{max}}, \) our theory will give \( \mathcal{O}(1) \) correlations in the input, which is inconsistent with the ansatz in Eq. (41). In this case, substantial corrections to Eq. (26) should be taken into account. A different approach similar to the one in Ref. [29] must be adopted to self-consistently determine these correlations.

VIII. DISCUSSION

We develop a theory for the emergence of correlations in strongly recurrent networks of binary neurons in which connection probabilities solely depend on distance and on the nature of the pre- and postsynaptic populations. The key results of our work are two scaling correlation theorems. The first shows that generically, all the spatial Fourier modes of the quenched average correlations are small when the number of neurons \( N, \) and the average number of connections \( K \) are large. They are of \( \mathcal{O}(1/N) \) and independent of \( K \) to leading order. This is true in the large-\( N \) limit even if we take \( K = pN, \) provided that \( p \) is not too large. However, the second theorem states that there are recurrent network architectures in which some of the Fourier modes of the quenched average correlations increase with \( K. \) These architectures are characterized by an explicit, or a hidden, feed-forward structure in those modes. This structure is revealed by the Jordan form of the interaction matrix averaged over realizations.

Because in the networks we study the correlations can be substantial, the balanced rate equations have to be augmented by balanced correlation equations to define the balanced regime. We derive these balanced correlation equations. To determine the stability of the balanced state, we derive the equations that determine the dynamics of the quenched averaged correlations. We show that, remarkably, these equations do not impose extra constraints on the stability. If the fixed point solution of the rate dynamics is stable so is the fixed point solution of the correlation dynamics.

A. Generality of the results

For simplicity, we assume that synaptic weights depend only on the identities of the populations to which pre- and postsynaptic neurons belong. The results, however, will not change if the weights are heterogeneous with distributions that depend on the pre- and postsynaptic populations, as long as the mean and the variance of all these distributions are finite.

For notational simplicity, we assume that all populations have the same number of neurons \( N, \) and each population receives, on average, inputs from \( K \) neurons in every population. The theory can be easily extended to networks in which population \( \alpha \) has \( N_{\alpha} = \nu_{\alpha} N \) neurons, and the average number of connections from population \( \beta \) to population \( \alpha \) is \( K_{\alpha\beta} = k_{\alpha\beta} K. \) This extension will not affect the scaling of the correlations with \( N \) and \( K. \) Prefactors, however, are different. For instance, assuming 4 times fewer inhibitory than excitatory neurons in the two-population networks of Sec. IV A, without changing the number of connections per neuron, the correlations of the inhibitory neurons will increase by a factor of 4.

We focus on networks with a one-dimensional architecture with periodic boundary conditions and connection probabilities which are solely distance dependent. With such an architecture the analysis of the problem is greatly simplified because when averaged over realizations, the correlations depend solely on distance. Furthermore, because of the linearity of the self-consistent equation for the correlations, the different Fourier modes decouple, allowing us to analyze each mode separately. However, our analytical approach does not require rotational invariance. It can be extended to any network architecture for which the Jordan normal form of the interaction matrix averaged over realizations can be established.

B. Robustness and self-consistency of the results

The theory we present here makes the anzatz that correlations are sufficiently small so that the dynamics of the cross-correlations can be linearized (Appendix A). If this anzatz is correct, the theory is self-consistent. When the theory predicts correlations which are \( \mathcal{O}(1), \) nonlinear terms contribute, and the correlations start to deviate from the theoretical value [see, for example, the simulation results for \( N = 40000 \) in Fig. 3(b)]. Nevertheless, the order of the correlations is still correctly predicted.

In the large-\( N, K \) limit, only networks with feed-forward structures—hidden or explicit—can exhibit correlations that increase with the average number of inputs. However, when \( K \) and \( N \) are only moderately large, as is the case in biological systems, a strict tuning of the architecture is not necessary. This is because there is a crossover between the regimes of strong and weak correlations as \( K \) is increased [see Fig. 4(c)]. As shown in Appendix B 1, the value of \( K \) for which this crossover occurs depends on the eigenvalues of the interaction matrix.

C. Relation to previous work

Noninteracting neurons can exhibit correlations if they share feed-forward inputs [57,58]. For instance, Litvak et al. [59] investigated a chain of layers of integrate-and-fire neurons lacking any recurrent interactions and coupled only feed forwardly. In their model, each neuron in a layer receives inputs from the same number of excitatory and
inhibitory neurons in the previous layer in such a way that their temporal averages exactly balance. They found a buildup of correlations along the chain. This is because the correlations induced by shared feed-forward inputs are not suppressed during the activity propagation since the network lacks any recurrent interactions and is thus purely feed forward.

Cortes and van Vreeswijk [60] studied a chain of strongly recurrent unstructured \(E-I\) subnetworks coupled through excitatory unstructured feed-forward projections. They found a gradual buildup of correlations along the chain. These correlations, however, decrease if the connectivity \(K\) and the subnetworks size \(N\) increase together. This result is in agreement with our theory, which predicts that the correlations are \(O(1/N)\) through the whole chain.

In Ref. [42], Ginzburg and Sompolinsky considered networks of binary neurons with finite-temperature Glauber dynamics, unstructured dense connectivity \((K = pN)\), and weak interactions, i.e., of the order of \(O(1/K)\) and not \(O(1/\sqrt{K})\), as in our work. Mean-field theory shows that in these networks the correlations are \(O(1/N)\) with a prefactor which diverges in the zero-temperature limit. This divergence is in contrast to what happens in the strongly recurrent unstructured networks we consider here where correlations also scale as \(O(1/N)\) but with a prefactor which is finite despite the fact that we assume zero-temperature Glauber dynamics. The latter result is due to the fact that in strongly recurrent network, intrinsic noise emerges from the deterministic dynamics of the network.

Ginzburg and Sompolinsky also demonstrated that in their model the correlations amplify up to \(O(1)\) at Hopf bifurcation onsets [42]. At such onsets, the dynamics exhibit critical slowing-down, and, thus, this amplification is accompanied by a divergence of the decorrelation times. Our work demonstrates a different amplification mechanism: It occurs at a point where the Jordan form of the interaction matrix \(\mathcal{J}\) contains a block which is a shift matrix. Since there is no critical slowing-down at such a point, the decorrelation times are finite and on the order of the update time constant (data not shown). Our theory may thus account for substantial correlations with short time-scales, as frequently observed in the brain [29,32,61].

Renart et al. [29] and Helias et al. [49] investigated strongly recurrent, unstructured, and densely connected binary networks. They found that in these networks, the mean pair correlations were \(O(1/N)\). As they considered dense connectivity, they could not, however, disentangle the dependence on \(N\) and \(K\). Here, we consider different relations between \(K\) and \(N\) and show that in unstructured networks the correlations are \(O(1/N)\), and in practice, they do not depend on \(K\). This last result is remarkable since one would expect correlations to increase with the degree of connectivity. This is not the case: The balance of excitation and inhibition prevents that from occurring in unstructured networks.

On the other hand, and somewhat surprisingly, we find that even if the fraction of common inputs shared by neurons is very small, a buildup of correlations can still occur for some network architectures. For example, in a network of four populations with a feed-forward structure, the correlations will be of \(O(K^3/N)\). Thus, to satisfy the balanced correlation equation, the scaling of \(K\) with the network size can be at most \(K = O(N^{1/3})\). However, with this architecture and scaling, the probability of two neurons to share their inputs is \(O(N^{-4/3})\) [Eq. (17)], and the number of inputs shared by two neurons is, therefore, \(O(N^{-1/3})\). Thus, although the number of shared inputs goes to zero in the large-\(N\) limit, the correlations get amplified up to \(O(1)\) thanks to the feed-forward architecture.

Rosenbaum et al. [40] have recently investigated how feed-forward excitation can drive correlations in spatially structured \(E-I\) networks operating in the balanced regime. The specific architecture they considered is reminiscent of the particular example presented in Fig. 7(a). In their study, the fluctuations which drove the correlated activity were those in the feed-forward inputs, and the contribution to the correlations of the fluctuations generated by the recurrent dynamics was neglected. In contrast, our work focuses on the role of the recurrent dynamics in the emergence of correlations.

When correlations are strong, \(O(K^{(P-1)}/N)\) and \(P \leq 2\), close-by neurons in the same population are positively correlated. This result is model independent. To see that, consider the aggregate activity of \(M\) nearby neurons in the same population. The variance of this aggregate activity is \(MA + M(M - 1)C\), where \(A\) is the average autocorrelation of these neurons and \(C\) is their averaged cross-correlation. Obviously, this sum must be non-negative. Hence, \(C\) must be larger than \(-A/(M - 1)\). Since \(M\) can be \(O(N)\), \(C\) cannot be large and negative (see, also, Fig. 3). This argument does not hold for correlations between nearby neurons in different populations. These correlations can be strong and negative [see, for instance, Eq. (31) and Fig. 3 for the EI correlations]. This is in contrast with the feed-forward model of Ref. [40] where all nearby correlations are positive. However, for architectures in which the correlations are \(O(1/N)\), the nearby neurons can be negatively correlated. An example is depicted in Fig. 2 (solid red line for excitatory neurons).

We recently studied the emergence of correlations in a network consisting of two strongly recurrent \(E-I\) subnetworks [50]. We showed that when one group of neurons in the first subnetwork projects to all excitatory neurons in the second subnetwork, the correlations in the second subnetwork are \(O(1/K)\). Thus, with such an architecture the bound given by the balanced correlation equation is saturated. We also showed that correlations are stronger when several groups of neurons in the first subnetwork project in a topographical way to a macroscopic fraction of neurons in the second subnetwork. While in the first subnetwork the
correlations were weak $O(1/N)$, in the second subnetwork, the correlations in macroscopic subpopulations of excitatory neurons were finite and did not depend on $N$ and $K$. In fact, the architecture in this case is also reminiscent of the example presented in Fig. 7(a). The fact that in Ref. [50] the correlations were independent of $N$ and $K$ [or were $O(1/K)$ when all neurons shared a common input] does not contradict our theory. In the architecture considered in Ref. [50], the subsets of neurons in the first subnetwork are projecting to a macroscopic fraction of neurons in the second subnetwork. This organization generates correlations between elements of the connectivity matrix, unlike in the models considered here. Generalizing our theory to such cases is possible but beyond the scope of this paper.

D. Directions for future work

The theorems of Sec. V tell us how the Fourier modes of the averaged correlations scale for large $N$ and $K$. In the examples considered in Secs. IV, VI, and VII, we focus on connectivities whose Fourier expansion involves only two modes. In these cases, the scaling of the spatial correlations can be immediately deduced from that of the Fourier modes. This is also the case for interactions described by a finite number of Fourier modes. However, if the interactions are described by an infinite number of modes, then inferring the scaling of the correlations from those of their modes can be involved. For instance, if one takes the large-$N$ limit with $K \propto N'$, the convergence of the Fourier series may not be uniform in $N$. In that case, the scaling with $N$ of the spatial correlations may be highly nontrivial. We will address this issue in an upcoming paper.

The present paper focuses on locally averaged two-point correlation functions. Recent progress in experimental techniques will create large data sets of neuronal activities from which distributions of pairwise correlations can be extracted. The power of theoretical approaches to interpret such data will be greatly enhanced if they provide not only locally averaged correlations but also higher-order statistics of their distributions. Thus, it would be interesting to extend our approach to estimate the scaling of higher-order moments of correlations.

Several previous studies investigated EI networks (e.g., Refs. [49,62]) in which inhibition and excitation were unstructured, and their strengths were only a function of the presynaptic neurons, i.e., $J_{EE} = J_{IE}$ and $J_{II} = J_{EI}$. Other studies assumed unstructured connectivity with $J_{ae} = -J_{ad}$, $a \in \{E, I\}$ (e.g., Ref. [63]). In both cases, the interaction parameters are on the edge of the region where the network evolves towards the balanced state [see Eq. (11)]. The network dynamics may be qualitatively different on the edge of this region than inside it. It is thus not clear that the scaling theorems presented here apply to these tuned cases. A further investigation of the correlation structure in such networks is a subject for future research.

Do the conclusions derived here for networks of binary neurons hold for networks with more realistic single-neuron dynamics? To approach this question, we perform extensive numerical simulations of strongly recurrent networks consisting of one inhibitory and one excitatory population of leaky integrate-and-fire neurons. The detailed analysis of these simulations will be presented elsewhere. In brief, we find in our simulations that in these networks the averaged pairwise correlations scale with $K$ and $N$ in a manner that is consistent with the theory presented here for binary networks. It would be very interesting to extend our analytical approach to these types of networks.

E. Concluding remarks

Correlations in neuronal activity impact the ability of the brain to encode information [33,64–66]. Correlations are also functionally important in sensory, motor, or cognitive tasks [67]. For instance, correlated activity has been hypothesized to be involved in visual perception [68] and may underlie the generation of exploratory behavior [50]. Correlations are also important for the self-organization of neuronal networks through activity-dependent plasticity [69]. Our work provides a theoretical account of the mechanisms underlying correlated activity in cortical networks. It gives insights into which features of cortical structure are important to induce functionally relevant correlations.

ACKNOWLEDGMENTS

We thank Gianluigi Mongillo and German Mato for fruitful discussions. We thank Takaumi Arakaki for comments on the manuscript. This work is conducted in the framework of the France Israel Laboratory of Neuroscience (FILN). Supported by the ERC grants No. ANR-14-NEUC-0001-01, No. ANR-13-BSV4-0014-02, and No. ANR-09-SYSC-002-01, the France-Israel High Council of Technology, LIA-FILN and IRN-FICNC.

APPENDIX A: CORRELATIONS IN BINARY NETWORKS

Here we calculate the equilibrium value and the stability of the quenched average correlations.

We define for $(j, \beta) \neq (i, \alpha)$ the out-of-equilibrium auto- and cross-correlations $c_{ij}^{\alpha\beta}(t, \tau)$ as

$$c_{ij}^{\alpha\beta}(t, \tau) \equiv \langle \delta S_i^{\alpha}(t)\delta S_j^{\beta}(t + \tau) \rangle_{\text{init}},$$

$$a_{ii}^{\alpha}(t, \tau) \equiv \langle \delta S_i^{\alpha}(t)\delta S_i^{\alpha}(t + \tau) \rangle_{\text{init}},$$

where $\langle \cdot \rangle_{\text{init}}$ denotes averaging over many initial conditions chosen with a probability measure that, for simplicity, we choose such that $\langle S_i^{\alpha}(0) \rangle_{\text{init}} = \langle S_j^{\alpha}(t) \rangle_{\text{init}}$. It is also convenient for the notation to define $c_{ii}^{\alpha\alpha}(t, \tau) = 0$. In this paper, we focus on the equal-time correlations $c_{ij}^{\alpha\beta}(t) \equiv c_{ij}^{\alpha\beta}(t, 0).$
For networks of binary neurons, the dynamics of the equal-time cross-correlations are given by [13,29,42]

\[
\frac{dc_{ij}^{\alpha\beta}(t)}{dt} = -2c_{ij}^{\alpha\beta}(t) + \langle \delta \Theta[h_i^\alpha(t) - T] \delta S_j^\beta(t) \rangle_{init} + \langle \delta S_i^\alpha(t) \delta \Theta[h_j^\beta(t) - T] \rangle_{init},
\]

where \( \delta \Theta[h_i^\alpha(t) - T] \equiv \Theta[h_i^\alpha(t) - T] - \langle \Theta[h_i^\alpha(t) - T] \rangle_{init} \).

If we make the ansatz that correlations are weak, we can, to leading order, take \( \langle \delta \Theta[h_i^\alpha(t) - T] \delta S_j^\beta(t) \rangle_{init} = g_i^\alpha(t) \delta h_i^\alpha(t) \delta S_j^\beta(t) \rangle_{init} \), where \( g_i^\alpha \) is the gain of neuron \((i, \alpha)\), \( g_i^\alpha = \partial_h^\alpha \langle \Theta[h_i^\alpha(t) - T] \rangle_{init} \), which is, to leading order, independent of the correlations [29,42,49].

Thus,

\[
\frac{dc_{ij}^{\alpha\beta}(t)}{dt} = -2c_{ij}^{\alpha\beta}(t) + \langle g_i^\alpha(t) \delta h_i^\alpha(t) \delta S_j^\beta(t) \rangle_{init} + \langle g_j^\beta(t) \delta S_i^\alpha(t) \delta S_j^\beta(t) \rangle_{init}.
\]

Using \( \langle \delta S_i^\alpha(t) \delta S_j^\beta(t) \rangle_{init} = \epsilon_{ij}^{\alpha\beta}(t) + \delta_{\alpha\beta} \delta_{i,j} a_i^\alpha(t) \) yields

\[
\frac{dc_{ij}^{\alpha\beta}(t)}{dt} = -2c_{ij}^{\alpha\beta}(t) + \sum_{\gamma,k} \{ J_{ik}^{\alpha\beta} \epsilon_{jk}^{\alpha\beta}(t) + J_{jk}^{\alpha\beta} \epsilon_{ik}^{\alpha\beta}(t) \} + \sum_{\gamma} \{ J_{i}^{\alpha\beta} a_i^\alpha(t) + J_{j}^{\alpha\beta} a_i^\beta(t) \},
\]

where \( J_{ik}^{\alpha\beta} = g_i^\alpha g_j^\beta \). Note that for weak correlations, \( \langle S_i^\alpha(t) \rangle_{init} \) does not depend on the correlations so that \( \langle S_i^\alpha(t) \rangle_{init} = \langle S_i^\alpha(t) \rangle \), for our choice of initial conditions \( \langle S_i^\alpha(0) \rangle_{init} = \langle S_i^\alpha(0) \rangle \). Hence, the equal-time autocorrelation \( a_i^\alpha \) is independent of time and given by \( a_i^\alpha = \langle S_i^\alpha(t) \rangle_{t} - \langle S_i^\alpha(t) \rangle$. We now average over the quenched disorder. Because of the rotational symmetry of the connection probabilities \( P_{ij}^{\alpha\beta} \), \( \langle a_i^\alpha \rangle_J \) is a constant

\[
[\epsilon_{ij}^{\alpha\beta}]J = A_{\alpha\beta}.
\]

whereas \( [c_{ij}^{\alpha\beta}]J \) and \( [J_{ij}^{\alpha\beta}]J \) are functions of the difference in the location of neurons \((i, \alpha)\) and \((j, \beta)\),

\[
[c_{ij}^{\alpha\beta}(t)]J = C_{ij}^{\alpha\beta}(\Theta_i^\alpha - \Theta_j^\beta, t),
\]

\[
[J_{ij}^{\alpha\beta}]J = \frac{1}{N} J_{ij}^{\alpha\beta}(\Theta_i^\alpha - \Theta_j^\beta),
\]

where \( J_{ij}^{\alpha\beta}(\Delta) = \sqrt{K} g_{ij} J_{ij} f_{\alpha\beta}(\Delta) \). Here, we assume that the correlations in the quenched disorder in the inputs to the neurons are small, such that, to leading order, the expected value of the gain does not depend on the neuronal position. We comment on this ansatz in Appendix C.

Thus, for large \( N \), the quenched average of Eq. (A1) yields

\[
\frac{d}{dt} C_{\alpha\beta}(\Delta, t) = -2C_{\alpha\beta}(\Delta, t)
\]

\[
+ \sum_{\gamma} \int \frac{d\Delta'}{2\pi} \left\{ J_{\alpha\gamma}(\Delta - \Delta') C_{\gamma\beta}(\Delta', t) + J_{\gamma\beta}(\Delta' - \Delta, t) \right\}
\]

\[
+ J_{\alpha\beta}(\Delta) A_{\beta}^\alpha \frac{A}{N} + J_{\beta\alpha}(\Delta) A_{\alpha}^\beta \frac{A}{N}.
\]

(A2)

Here we do not take into account that \( c_{ij}^{\alpha\beta}(t) = 0 \). It is easy to see, however, that this approximation neglects \( O(1/N^2) \) corrections in the cross-correlations. We show below that these corrections are indeed negligible in the large-\( N \) limit.

The \( n \)th Fourier mode of \( C_{\alpha\beta}(\Delta, t) \), \( C_{\alpha\beta}^{(n)}(t) \) satisfies

\[
\frac{d}{dt} C_{\alpha\beta}^{(n)}(t) = -2C_{\alpha\beta}^{(n)}(t)
\]

\[
+ \sum_{\gamma} \left\{ J_{\alpha\gamma}^{(n)} C_{\gamma\beta}^{(n)}(t) + C_{\gamma\beta}^{(n)}(t) J_{\alpha\gamma}^{(n)} \right\}
\]

\[
+ J_{\alpha\beta}^{(n)} \frac{A_{\beta}^\alpha}{N} + J_{\beta\alpha}^{(n)} \frac{A_{\alpha}^\beta}{N} \left[ J_{\alpha\beta}^{(n)} \right],
\]

where \( J_{\alpha\beta}^{(n)} \) and \( J_{\alpha\beta}^{(n)} \) are the \( n \)th Fourier mode of \( J_{\alpha\beta}(\Delta) \) and \( J_{\alpha\beta}^{(n)}(\Delta) \), respectively.

This equation can be written more compactly as

\[
\frac{d}{dt} \langle n \rangle = -2\langle n \rangle + \langle J^{(n)} \rangle \langle C^{(n)} \rangle + \langle C^{(n)} \rangle \langle J^{(n)} \rangle
\]

\[
+ \langle J^{(n)} \rangle A_{N} + \langle A \rangle \langle J^{(n)} \rangle^2,
\]

(A3)

where \( X \) denotes the \( D \times D \) matrix \( X_{\alpha\beta} \), and \( A \) is the \( D \times D \) matrix \( A_{\alpha\beta} = \delta_{\alpha\beta} A_{\alpha} \). The equilibrium values of the Fourier components of equal-time correlation functions thus satisfy Eq. (26).

**APPENDIX B: CORRELATION THEOREMS**

In Appendix A, we derive \( N^2 D^2 \) coupled linear differential equations that determine the evolution of equal-time cross-correlations for all neuronal pairs in the network. Averaging these equations over the quenched disorder and using the rotational invariance of the connection probabilities yields a set of \( N D^2 \)-coupled equations for the quenched averaged correlations. In Fourier space, these equations lead to \( N \)-independent sets of \( D^2 \)-coupled equations for the correlations. Here we prove Theorems 1 and 2.
correlations as

\( J \) (where we define \( J_{\text{tor}} \) is the Jordan normal form of \( J \)) in the rest of this appendix. Defining

\( \tilde{J} \) can be written as

\[
\tilde{J} = U(J_{\text{tor}} + \Lambda)U^{-1},
\]

where \( J_{\text{tor}} \) is the Jordan normal form of \( J \), and \( U^{-1} \) is the transformation matrix to the Jordan basis.

The matrix \( J_{\text{tor}} \) can be written as

\[
[J_{\text{tor}}]^\mu_\nu = \lambda_\mu^{(n)} \delta_\mu_\nu + \epsilon_\mu^{(n)} \delta_\mu_{\nu+1},
\]

where \( \lambda_\mu^{(n)} \) are the eigenvalues of \( J \) and \( \epsilon_\mu^{(n)} = 1 \) inside a Jordan block and is 0 otherwise (for clarity, in the Jordan basis we use the subscripts \( \mu \) and \( \nu \) rather than \( \alpha \) and \( \beta \) that we use in the original basis).

Importantly, the Jordan form of a matrix and of its Hermitian conjugate are complex conjugate. We thus can write

\[
[J_{\text{tor}}]^\dagger = V(J_{\text{tor}}^\dagger)V^{-1}.
\]

For notational convenience, we suppress the superscript \( (n) \) in the rest of this appendix. Defining \( \tilde{C} \) as

\[
\tilde{C} = U^{-1}CV
\]

and inserting into Eq. (B1) yields

\[
\frac{d}{dt} \hat{C}_{\mu_\nu}(t) = -\Lambda_{\mu\nu} \hat{C}_{\mu_\nu}(t) + \sqrt{K}[e_{\mu_\nu} \hat{C}_{\mu_{\nu+1}}(t) + e_{\nu-1} \hat{C}_{\mu_{\nu-1}}(t)] + \sqrt{K} \frac{\Lambda_{\mu\nu}}{N}[(\lambda_\mu + \lambda_\nu^*) \hat{A}_{\mu\nu} + e_{\mu_\nu} \epsilon_{\mu_{\nu+1}} + e_{\nu-1} \epsilon_{\mu_{\nu-1}}].
\]

We now assume that the connectivity is such that the system is stable to perturbations of the locally averaged rates. This implies that the real part of all eigenvalues of \( J \) is less than \( 1/\sqrt{K} \). The real part of \( \Lambda_{\mu\nu} \) is, therefore, positive and, thus, \( \dot{C}(t) \) converges to an equilibrium value \( \hat{C}_\infty \).

To see that \( \dot{C}(t) \) converges, first consider \( \dot{C}_{D,1}(t) \), which satisfies

\[
\frac{d}{dt} \dot{C}_{D,1}(t) = -\Lambda_{D,1} \dot{C}_{D,1}(t) + \sqrt{K} \frac{\Lambda_{D,1}}{N}[(\lambda_D + \lambda_1^*) \hat{A}_{D,1} + e_{\mu_\nu} \epsilon_{\mu_{\nu+1}} + e_{\nu-1} \epsilon_{\mu_{\nu-1}}].
\]

Therefore, \( \dot{C}_{D,1}(t) \) converges to its equilibrium value.

The evolution equations of \( \dot{C}_{D,2} \) can be written as

\[
\frac{d}{dt} \dot{C}_{D,2}(t) = -\Lambda_{D,2} \dot{C}_{D,2}(t) + M_{D,2} \dot{C}_{D,2}(t),
\]

where \( M_{D,2} \) depends on \( \dot{C}_{D,1}(t) \), \( \dot{A}_{D,1} \), and \( \dot{A}_{D,2} \). Since \( \dot{C}_{D,1}(t) \) converges to its equilibrium value, \( M_{D,2} \) converges to a constant. Because Re(\( \Lambda_{D,2} \)) > 0, \( \dot{C}_{D,2} \) also converges to its equilibrium value \( \hat{C}_\infty^{D,1} \). A similar argument shows that likewise \( \dot{C}_{D-1,1} \) converges to \( \hat{C}_\infty^{D-1,1} \). One then sees by recursion that the whole matrix \( \dot{C} \) converges to its equilibrium value \( \hat{C}_\infty \). Hence, Eq. (26) determines the stable equilibrium values of the correlations.

From here, we consider only the correlations at equilibrium, and, for notational simplicity, we drop the superscript \( \infty \).

In the Jordan basis, the equilibrium values of the correlations satisfy

\[
\Lambda_{\mu\nu} \hat{C}_{\mu_\nu} = \sqrt{K}[e_{\mu_\nu} \hat{C}_{\mu_{\nu+1}} + e_{\nu-1} \hat{C}_{\mu_{\nu-1}}] + \sqrt{K} \frac{\Lambda_{\mu\nu}}{N}[(\lambda_\mu + \lambda_\nu^*) \hat{A}_{\mu\nu} + e_{\mu_\nu} \epsilon_{\mu_{\nu+1}} + e_{\nu-1} \epsilon_{\mu_{\nu-1}}].
\]

Let us now consider the case where \( J \) is diagonalizable so that \( e_{\mu} = 0 \) for all \( \mu \). In this case, \( \hat{C}_{\mu_\nu} = 0 \) for \( \mu \neq \nu \) and

\[
\hat{C}_{\mu_\nu} = \frac{\sqrt{K}(\lambda_\mu + \lambda_\nu^*)}{N[2 - \sqrt{K}(\lambda_\mu + \lambda_\nu^*)]} \hat{A}_{\mu_\nu}.
\]

Thus,

\[
\lim_{K \to \infty} \lim_{N \to \infty} N \hat{C}_{\mu_\nu} = -\hat{A}_{\mu_\nu}
\]

unless \( |\lambda_\mu + \lambda_\nu^*| \) is \( O(1/\sqrt{K}) \), in which case,

\[
\lim_{K \to \infty} \lim_{N \to \infty} N \hat{C}_{\mu_\nu} = 0.
\]
In the first situation, $\hat{C}_{\mu\mu} = \mathcal{O}(1/N)$. In the second situation, $\hat{C}_{\mu\mu} = \mathcal{O}(1/N)$.

Let us now assume that $\mathcal{J}$ is not diagonalizable. Then, the $D \times D$ Jordan form of $\mathcal{J}$ consists of $B$ Jordan blocks $(1 \leq B < D)$ that we denote by $[\mathcal{J}_{j}]^{\nu}$, $i = 1, \ldots, B$. The size of the $i$th block is denoted by $s(i)$. All the diagonal elements of this Jordan block are equal to one of the eigenvalues of $\mathcal{J}$, which we denote by $\hat{\lambda}_i$. The off-diagonal elements $\varepsilon_\mu$ are all equal to 1 except for $\mu = h(i)$ ($i = 1, \ldots, B - 1$) for which $\varepsilon_\mu = 0$.

The matrix $\hat{C}$ consists of $B^2$ sectors that we denote by $S_{ij}$. In the sector $S_{ij}$, $\mu \in \{l(i), \ldots, h(i)\}$ and $\nu \in \{l(j), \ldots, h(j)\}$.

Let us consider Eq. (B3) for $\mu, \nu$ in sector $S_{ij}$. Since $\varepsilon_{h(i)}$ and $\varepsilon_{l(i)-1}$ are zero, the equation in this sector does not depend on elements of $\hat{C}$ outside of it. Thus, we can solve Eq. (B3) recursively to determine all the elements of $\hat{C}$ in this sector. The recursion goes as follows. First, one solves for $\hat{C}_{\mu(i),\nu(j)}$.

$$\hat{C}_{\mu(i),\nu(j)} = \frac{1}{2} \frac{\sqrt{K(\hat{\lambda}_i + \hat{\lambda}_j)}}{2 - \sqrt{K(\hat{\lambda}_i + \hat{\lambda}_j)}} \hat{A}_{\mu(i),\nu(j)}.$$  \hspace{1cm} (B6)

We can solve Eq. (B3) to get $\hat{C}_{\mu\nu}$ for $\mu, \nu = h(i) - 1, l(j)$ and $\mu, \nu = h(i), l(j) + 1$. This process can be repeated until all the elements of $\hat{C}$ in the sector $S_{ij}$ are determined.

1. $\hat{A}_{\mu(i),\nu(j)}$ are all nonzero

A similar recursion can be performed to estimate the order of magnitude of all the elements of $\hat{C}$. First, we obtain

$$\hat{C}_{\mu(i),\nu(j)} = \mathcal{O}\left(\frac{\hat{A}_{\mu(i),\nu(j)}}{N}\right) = \mathcal{O}\left(\frac{1}{N}\right).$$

For $\hat{\lambda}_i + \hat{\lambda}_j = \mathcal{O}(1)$, the recursion shows that we have

$$\hat{C}_{\mu\nu} = \mathcal{O}\left(\frac{1}{N}\right).$$ \hspace{1cm} (B7)

for all $\mu, \nu$ in sector $S_{ij}$.

If $\hat{\lambda}_i + \hat{\lambda}_j = 0$, $\hat{C}_{\mu(i),\nu(j)} = 0$, and by recursion all the other elements in sector $S_{ij}$ are

$$\hat{C}_{\mu\nu} = \mathcal{O}(K^{(\ell+1)/2}/N),$$ \hspace{1cm} (B8)

where $\ell = \nu - \mu + h(i) - l(j)$. Thus, in this case $\hat{C}_{\mu(i),\nu(j)}$ is the largest entry in the sector. It satisfies

$$\hat{C}_{l(i),h(j)} = \mathcal{O}(K^{(s(i)+s(j)-2)/2}/N).$$ \hspace{1cm} (B9)

The scaling of the elements in sector $S_{ij}$ thus depends on $\hat{\lambda}_i + \hat{\lambda}_j$. The stability of the dynamics imposes that $\text{Re}(\hat{\lambda}_i) < 0$ for all blocks [or positive but at most $O(1/\sqrt{K}$); see next subsection, and, thus, the condition $\hat{\lambda}_i + \hat{\lambda}_j = 0$ implies that for large $K$, the real part of $\hat{\lambda}_i$ and $\hat{\lambda}_j$ is zero. In other words, the real parts of $[\mathcal{J}_{j}]^{\nu}$ and $[\mathcal{J}_{i}]^{\mu}$ are shift matrices of size $s(i) \times s(i)$ and $s(j) \times s(j)$, respectively. For the sector $S_{ij}$, it means that the real part of $[\mathcal{J}_{j}]^{\nu}$ is a shift matrix of size $s(i) \times s(i)$.

Proving Correlation Theorem 1 is now straightforward. If the real parts of all the $B$ Jordan blocks of $\mathcal{J}$ are different from a shift matrix, one finds that $\hat{\lambda}_i + \hat{\lambda}_j = \mathcal{O}(1) \neq 0$ for all $i, j \in \{1, \ldots, B\}$. According to Eq. (B7), the correlations are at most $O(1/N)$ in all of the $B^2$ sectors of the matrix $\hat{C}$. As a result, $\hat{C}^{(n)} = \mathcal{O}(1/N)$. On the other hand, if in each of the $B^2$ sectors of the matrix $\hat{C}$, the correlations are at most $O(1/N)$, then this is also the case for the element of $\hat{C}$ in all the sectors. In this case, Eq. (B7) implies that there is no Jordan block in $\mathcal{J}$ for which $\text{Re}(\hat{\lambda}_i) = 0$.

Restoring the index $n$ of the Fourier mode, one sees that if $\mathcal{J}^{(n)}$ has at least one Jordan block whose real part is a shift matrix and denoting by $P(n)$ the size of the largest shift, Eq. (B9) implies that $\hat{C}_{l(i),h(j)} = \mathcal{O}(K^{P(n)-1}/N)$ and, thus, also $\hat{C}^{(n)} = \mathcal{O}(K^{P(n)-1}/N)$. This proves Correlation Theorem 2.

According to Eq. (B9), the scaling of the correlation is $\hat{C}^{(n)} = \mathcal{O}(K^{D-1}/N)$ if and only if $P(n) = D$. When $\mathcal{J}^{(n)}$ is a real matrix (when the probability of connections are symmetric in $\Delta$), it means that the Jordan form of $\mathcal{J}^{(n)}$ is a shift matrix of size $D$. This is equivalent to saying that $\mathcal{J}^{(n)}$ is nilpotent of degree $D$. On the other hand, if $\mathcal{J}^{(n)}$ is a shift matrix of degree $D$, it has a Jordan form which is a shift matrix of degree $D$. According to Eq. (B9), the scaling of the correlation is then $\hat{C}^{(n)} = \mathcal{O}(K^{D-1}/N)$. This proves Corollary 1 in Sec. V.

When $K = \mathcal{O}(1/|\hat{\lambda}_i + \hat{\lambda}_j|^2)$, there is a crossover in sector $S_{ij}$ of the matrix $\hat{C}$ between weak correlations [$\mathcal{O}(1/N)$] and strong correlations [$\mathcal{O}(K^{(s(i)+s(j)-2)/2}/N)$]. To see this, we note that if $K = |\hat{\lambda}_i + \hat{\lambda}_j|^{-1/2}$ for $\Gamma \leq 1/2$,

$$\hat{C}_{\mu\nu} = \mathcal{O}\left(\frac{K^{\ell}}{N}\right),$$

whereas for $1/2 < \Gamma \leq 1$,
\( \hat{C}_{\mu \nu} = O \left( \frac{K^{\ell + 2}}{N} \right) \)

for \( \ell \neq 0 \) and

\( \hat{C}_{\mu \nu} = O \left( \frac{K^{1/2-\gamma}}{N} \right) \)

for \( \ell = 0 \).

### 2. Some elements of \( \hat{A} \) are zero

So far, we assume that all the elements in \( \hat{A} \) are nonzero. The derivation can, however, be extended to include also situations where this is not the case as follows.

Elements \( \hat{A}_{\mu \nu} = 0 \) in a sector where \( \hat{\lambda}_i + \hat{\lambda}_j = O(1) \) might result in some elements of \( \hat{C} \) in that sector to be equal to 0 rather than \( O(1/N) \). However, this will not affect the overall scaling of the correlations.

Elements \( \hat{A}_{\mu \nu} = 0 \) in a sector \( S_{ij} \) where \( \hat{\lambda}_i + \hat{\lambda}_j = 0 \) do not change the scaling in the matrix, provided that \( \hat{A}_{h(i),l(j)} \neq 0 \). If \( \hat{A}_{h(i),l(j)} = 0 \), the effect depends on \( \hat{A}_{h(i)-1,l(j)} \) and \( \hat{A}_{h(i),l(j)+1} \). If at least one of them is nonzero, the order of the largest entry in the sector is decreased by \( \sqrt{K} \). If both of them are zero, the order of this entry decreases at least by a factor of \( K \). As other elements can also be zero, one reaches the following conclusion: Let \( \{\mu, \nu\} \) be the indices \( \{\mu, \nu\} \) in \( S_{ij} \) for which \( \hat{A}_{\mu \nu} \neq 0 \), which maximize \( \mu - \nu \). Then, the maximal order of \( \hat{C}_{\mu \nu} \) in the sector is \( O(K^{P-1}/N) \), where

\[
P_{ij} = 1 + |\mu_{ij} - l(i) + h(i) - \nu_{ij}|/2. \tag{B10}
\]

In conclusion, the highest order in \( \hat{C} \) is \( O(K^{P-1}/N) \), where \( P \) is the maximum value of the \( P_{ij} \) \( (i, j \in \{1, \ldots, B\} \) defined by (1) \( P_{ij} = 1 \) for sector \( S_{ij} \) for which \( \hat{\lambda}_i + \hat{\lambda}_j \neq 0 \), (2) \( P_{ij} = P \) for sector \( S_{ij} \) in which \( \hat{\lambda}_i + \hat{\lambda}_j = 0 \) and all \( \hat{A}_{\mu \nu} \neq 0 \) and (3) \( P_{ij} \) is given by Eq. (B10) if in sector \( S_{ij} \), \( \hat{\lambda}_i + \hat{\lambda}_j = 0 \) but some \( \hat{A}_{\mu \nu} \) are zero.

### APPENDIX C: CORRELATIONS OF THE QUENCHED DISORDER

Let us define the correlation of the quenched disorder in the outputs of the neurons as

\[
\Gamma^{q\beta}_{ij} = [\Delta S^q_i \Delta S^\beta_j]_f
\]

for \( (i, \alpha) \neq (j, \beta) \) and \( \Delta S^q_i \) defined in Eq. (3). A derivation similar to that in Appendix A yields

\[
2\Gamma^{q\beta}_{ij} = \sum_k [J^{q\beta}_{ik} J^{q\beta}_{kj} + J^{q\beta}_{ik} J^{q\beta}_{kj}] + J^{q\beta}_{ij} q_{\beta}/N + J^{q\beta}_{ji} q_{\alpha}/N,
\]

where \( q_{\alpha} = [\langle (\Delta S^q_i)^2 \rangle]_f \). This equation can be solved using the same approach as in Appendixes A and B. This analysis shows that correlations of the quenched disorder and correlations of the temporal fluctuations are of the same order. Thus, when the temporal fluctuations are small, the ansatz in Appendix A, where we neglected the spatial fluctuations in the neuronal gain, is justified. If the correlations are too strong, the ansatz may no longer be satisfied but neither is the linearization assumed to derive Eq. (A2).

### APPENDIX D: SELF-CONSISTENT EQUATIONS FOR THE AUTOCORRELATION AND THE GAIN

We follow the notations from Ref. [51],

\[
m_\alpha = \frac{1}{2} \text{erfc} \left( \frac{T_a - h_a}{\sqrt{2(\sigma^2_a + \sigma^2_{qa})}} \right), \tag{D1}
\]

\[
q_{\alpha} = \frac{1}{2} \int Dx \text{erfc}^2 \left( \frac{T_a - h_a - \sigma_{qa} x}{\sqrt{2\sigma^2_a}} \right), \tag{D2}
\]

where the variance of the input noise \( \sigma^2 = \langle (h^q_i - \langle h^q_i \rangle)^2 \rangle_f \) is

\[
\sigma^2_a = \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} \frac{1}{N^2} \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} f_{ij} \frac{1}{2} \text{erfc} \left( \frac{T_a - h_a}{\sqrt{2(\sigma^2_a + \sigma^2_{qa})}} \right) C^{q\beta}_{ij} (\theta^\alpha_i - \theta^\beta_j) \right) , \tag{D4}
\]

and the quenched disorder in the inputs \( \sigma^2_{qa} = \langle (\langle h^q_i \rangle - \langle h^q_i \rangle)^2 \rangle_f \) is

\[
\sigma^2_{qa} = \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} \frac{1}{N^2} \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} f_{ij} \frac{1}{2} \text{erfc} \left( \frac{T_a - h_a}{\sqrt{2(\sigma^2_a + \sigma^2_{qa})}} \right) + K \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} \frac{1}{N^2} \sum_{j} J^{q\beta}_{ij} J^{q\beta}_{kj} f_{ij} \frac{1}{2} \text{erfc} \left( \frac{T_a - h_a}{\sqrt{2(\sigma^2_a + \sigma^2_{qa})}} \right) \right) \right) , \tag{D5}
\]

Finally, the gain of the neurons is

\[
g_\alpha = \frac{1}{\sqrt{2\pi\sigma^2_a}} \frac{1}{\sqrt{2\pi\sigma^2_a}} e^{-\frac{(T_a - h_a)^2}{2\sigma^2_a} - \frac{q_{\alpha}^2}{2\sigma^2_{qa}}}. \tag{D6}
\]
Equations (D1)–(D3) need to be solved self-consistently. For simplicity, when we solve them we neglect $O(p)$ terms (but see Ref. [49]).

**APPENDIX E: Cross-correlations in two-population networks**

For a two-population network, Eq. (26) yields

$$
B^{(n)} = \frac{1}{N} D^{(n)} \left( \begin{array}{c}
A_E \\
A_I
\end{array} \right)
$$

(E1)

with

$$
C_{EE}^{(n)} = \frac{1}{N} \left[ -2 A_E \bar{J}_{EE}^{(n)} \sqrt{K} + [A_I \bar{J}_{EI}^{(n)}] \Delta^{(n)} + \bar{J}_{EE}^{(n)} \bar{J}_{II}^{(n)} \right] K - A_E T^{(n)} \Delta^{(n)} K^{3/2} / 2 + 3 T^{(n)} \sqrt{K} - \left[ (T^{(n)})^2 + 2 \Delta^{(n)} \right] K + T^{(n)} \Delta^{(n)} K^{3/2},
$$

$$
C_{EI}^{(n)} = \frac{1}{N} \left[ A_E \bar{J}_{EI}^{(n)} \sqrt{K} - A_I \bar{J}_{EE}^{(n)} \bar{J}_{EI}^{(n)} + A_E \bar{J}_{II}^{(n)} \bar{J}_{IE}^{(n)} \right] K
$$

$$
C_{II}^{(n)} = \frac{1}{N} \left[ -2 A_I \bar{J}_{II}^{(n)} \sqrt{K} + A_I \bar{J}_{IE}^{(n)} \bar{J}_{II}^{(n)} T^{(n)} + \Delta^{(n)} + \bar{J}_{EE}^{(n)} \bar{J}_{II}^{(n)} \right] K - A_I T^{(n)} \Delta^{(n)} K^{3/2} / 2 + 3 T^{(n)} \sqrt{K} - \left[ (T^{(n)})^2 + 2 \Delta^{(n)} \right] K + T^{(n)} \Delta^{(n)} K^{3/2},
$$

where $T^{(n)} = \text{Tr} \bar{J}^{(n)}$, $\Delta^{(n)} = \det \bar{J}^{(n)}$. After some algebra, this equation can be rewritten as Eq. (27).

---


[18] T. Tchumatchenko, A. Malyshev, T. Geisel, M. Volgushev, and F. Wolf, 


