LETTER Communicated by Kenneth Harris

Decorrelation by Recurrent Inhibition in Heterogeneous Neural Circuits

Alberto Bernacchia
a.bernacchia@jacobs-university.de
Department of Neurobiology, Yale University, New Haven, CT 06520, and School of Engineering and Science, Jacobs University, Bremen 28759, Germany

Xiao-Jing Wang
xjwang@nyu.edu
Department of Neurobiology, Yale University, New Haven, CT 06520, and Center for Neural Science, New York University, New York, NY 10003, U.S.A.

The activity of neurons is correlated, and this correlation affects how the brain processes information. We study the neural circuit mechanisms of correlations by analyzing a network model characterized by strong and heterogeneous interactions: excitatory input drives the fluctuations of neural activity, which are counterbalanced by inhibitory feedback. In particular, excitatory input tends to correlate neurons, while inhibitory feedback reduces correlations. We demonstrate that heterogeneity of synaptic connections is necessary for this inhibition of correlations. We calculate statistical averages over the disordered synaptic interactions and apply our findings to both a simple linear model and a more realistic spiking network model. We find that correlations at zero time lag are positive and of magnitude $K^{-\frac{1}{2}}$, where K is the number of connections to a neuron. Correlations at longer timescales are of smaller magnitude, of order K^{-1}, implying that inhibition of correlations occurs quickly, on a timescale of $K^{-\frac{1}{2}}$. The small magnitude of correlations agrees qualitatively with physiological measurements in the cerebral cortex and basal ganglia. The model could be used to study correlations in brain regions dominated by recurrent inhibition, such as the striatum and globus pallidus.

1 Introduction

Simultaneous measurements of the activity of multiple neurons have shown significant correlations, and this observation has stimulated the debate on whether and how correlations contribute to neural computation. In principle, correlations allow robust signal processing, because redundancies across neurons can be exploited to separate the signal from the noise (Abbott & Dayan, 1999; Panzeri, Schultz, Treves, & Rolls, 1999). Experimental

studies of the cerebral cortex suggest that correlations improve decoding of stimuli (Graf, Kohn, Jazayeri, & Movshon, 2011), but it remains unclear whether a parsimonious decoder should rely on correlations (Averbeck & Lee, 2003). A challenge to this hypothesis is the observation that correlations are reduced when animal subjects are actively engaged in discrimination (Cohen & Newsome, 2008; Cohen & Maunsell, 2009), and even when they simply start a movement (Poulet & Petersen, 2008). In addition, neurons with similar responses to stimuli show higher correlations (Zohary, Shadlen, & Newsome, 1994; Lee, Port, Kruse, & Georgopoulos, 1998; Maynard et al., 1999; Bair, Zohary, & Newsome, 2001; Constantinidis & Goldman-Rakic, 2002; Averbeck & Lee, 2003; Romo, Hernandez, Zainos, & Salinas, 2003; Kohn & Smith, 2005; Smith & Kohn, 2008; Huang & Lisberger, 2009; Ecker et al., 2010; Komiyama et al., 2010), implying that coding of stimuli should be worsened by correlations (Abbott & Dayan, 1999; Panzeri et al., 1999; Sompolinsky, Yoon, Kang, & Shamir, 2001; Wilke & Eurich, 2002; Averbeck, Latham, & Pouget, 2006; Gutniski & Dragoi, 2008). Another caveat is that the neural code is largely unknown, and if the noise measured in physiological studies encodes some signal, then any correlation would decrease the available information (Nadal & Parga, 1994).

Besides the possible function of correlations in signal and information processing, their physiological causes remain unclear. It has been shown that the correlation between nearby neurons is driven by their correlated synaptic input (Lampl, Reichova, & Ferster, 1999; Poulet & Petersen, 2008). However, a quantitative understanding of the circuit mechanisms regulating correlations between cortical cells is still missing, and the goal of this study is to determine the dependence of correlations on different properties of the neural circuitry. The measured correlation between neurons depends on different factors and varies across studies (Cohen & Kohn, 2011): it increases with the proximity of neuron pairs (Maynard et al., 1999; Constantinidis & Goldman-Rakic, 2002; Smith & Kohn, 2008; Ecker et al., 2010; Komiyama et al., 2010), their activity (de la Rocha, Doiron, Shea-Brown, Josic, & Reyes, 2007), and the temporal window on which action potentials are counted (Bair et al., 2001; Reich, Mechler, & Victor, 2001; Constantinidis & Goldman-Rakic, 2002; Averbeck & Lee, 2003; Kohn & Smith, 2005; Smith & Kohn, 2008; Huang & Lisberger, 2009; Mitchell, Sundberg, & Reynolds, 2010). Figure 1 shows the correlation measured in eight different studies as a function of temporal window for spike counts. Results vary, although correlations are generally found positive and of small magnitude in both the cortex and the basal ganglia (Raz, Vaadia, & Bergman, 2000).

Previous modeling studies of neural circuits have found that the mean correlation between neurons is small, of the order of N^{-1}, where N is the number of neurons in the network. Small correlations have been observed, not surprising, in networks characterized by weak connection strengths (Ginzburg & Sompolinski, 1994; Bernacchia & Amit, 2007). More

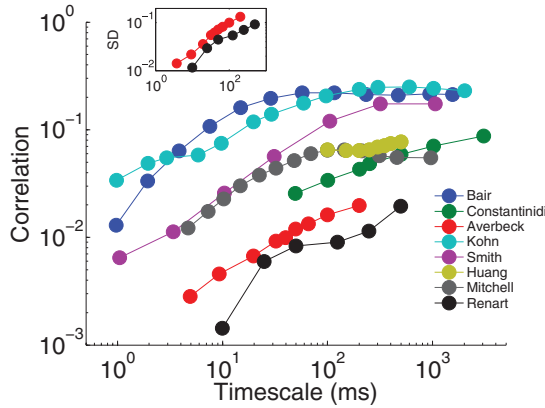

Figure 1: Mean correlation across neuron pairs plotted versus the length of the time window used to count action potentials. Replotted from eight experimental studies of the cortex (legend). Two studies provided not only the mean correlation but also the standard deviation (inset).

surprising, the same result has been obtained in the case of strong connections, such as the high-conductance state (Destexhe, Rudolph, & Paré, 2003; van Vreesvijk & Sompolinski, 1996), provided that the network includes a strong inhibitory feedback (Renart et al., 2010; Hertz, 2010; Tetzlaff, Helias, Einevoll, & Diesmann, 2012). Here we provide an analytical study of correlations in a simple linear model, and we apply our findings to predict correlations in a more realistic spiking network model. We confirm both the observed small correlation and the crucial effect of the inhibitory feedback in reducing it. In addition, we study the effect of the heterogeneity of connection strengths by using random matrix theory and a diagrammatic formalism, and we show that inhibition of correlations crucially depends on such heterogeneity. The model can be compared to brain regions dominated by recurrent inhibition, such as the striatum and globus pallidus.

We find that correlations at zero time lag are of magnitude $K^{-\frac{1}{2}}$, where K is number of connections received by a neuron, while correlations of the activity integrated across time are of order K^{-1}, suggesting that inhibition of correlations operates on a timescale of $K^{-\frac{1}{2}}$. These results are consistent with previous modeling studies, suggesting that a linear approximation is adequate to predict correlations in more realistic spiking models (Renart et al., 2010). In addition, our findings highlight the difference between the effect of the number of neurons N versus the number of connections K on correlations. The small correlations predicted by this and previous modeling studies qualitatively match the small correlations observed in neurons of the cerebral cortex and basal ganglia.

2 Methods

We consider both a linear model and a more realistic spiking network model. In both models, we consider a neural circuit of N neurons, receiving input from N_{ext} external neurons, where each neuron integrates the signal from other neurons weighted by the synaptic connection strength.

The dynamics of the linear model is described by

$$\tau \frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^{N} G_{ij}x_j(t) + \sum_{j=1}^{N_{ext}} G_{ij}^{ext}x_j^{ext}(t), \tag{2.1}$$

where x_i is the activity of neuron i in the local circuit and G_{ij} is the strength of the synaptic connection from neuron j to neuron i. The external (feedforward) input to the circuit is provided by the activities x_j^{ext}, and the synaptic connection from the jth external neuron to the ith local neuron is given by the strength G_{ij}^{ext}. All neuronal activities evolve in time, while the connectivity matrices G and G_{ext} are fixed.

We define the average number of local connections received by a neuron as K and the external connections as K_{ext}. We assume that the connectivity matrices are random, which makes the network akin to a disordered system, characterized by a random but fixed substrate. We consider two scenarios, represented schematically in Figures 2a and 2b:

1. The network is fully connected ($K = N$, $K_{ext} = N_{ext}$) with random connection strengths (all-to-all; see Figure 2a), characterized by a gaussian distribution. The mean and variance of matrix elements are determined by the parameters g and λ for the local connections and g_{ext} and λ_{ext} for the external connections:

$$\langle G_{ij}\rangle = -g/\sqrt{N} \qquad \langle \Delta G_{ij}^2\rangle = \lambda^2/N, \tag{2.2}$$

$$\langle G_{ij}^{ext}\rangle = g_{ext}/\sqrt{N_{ext}} \qquad \langle \Delta G_{ij}^{ext\,2}\rangle = \lambda_{ext}^2/N_{ext}. \tag{2.3}$$

2. The network is sparse; only a fraction of connections exists ($k = K/N$, $k_{ext} = K_{ext}/N_{ext}$), and the others are set to zero (sparse; see Figure 2b). Connections are selected at random but of constant strength, equal to $-g/\sqrt{K}$ for recurrent connections and $g_{ext}/\sqrt{K_{ext}}$ for external connections. The distribution is Bernouillian. The mean and variance of matrix elements are

$$\langle G_{ij}\rangle = -kg/\sqrt{K} \qquad \langle \Delta G_{ij}^2\rangle = k(1-k)g^2/K \tag{2.4}$$

$$\langle G_{ij}^{ext}\rangle = k_{ext}g_{ext}/\sqrt{K_{ext}} \qquad \langle \Delta G_{ij}^{ext\,2}\rangle = k_{ext}(1-k_{ext})g_{ext}^2/K_{ext}. \tag{2.5}$$

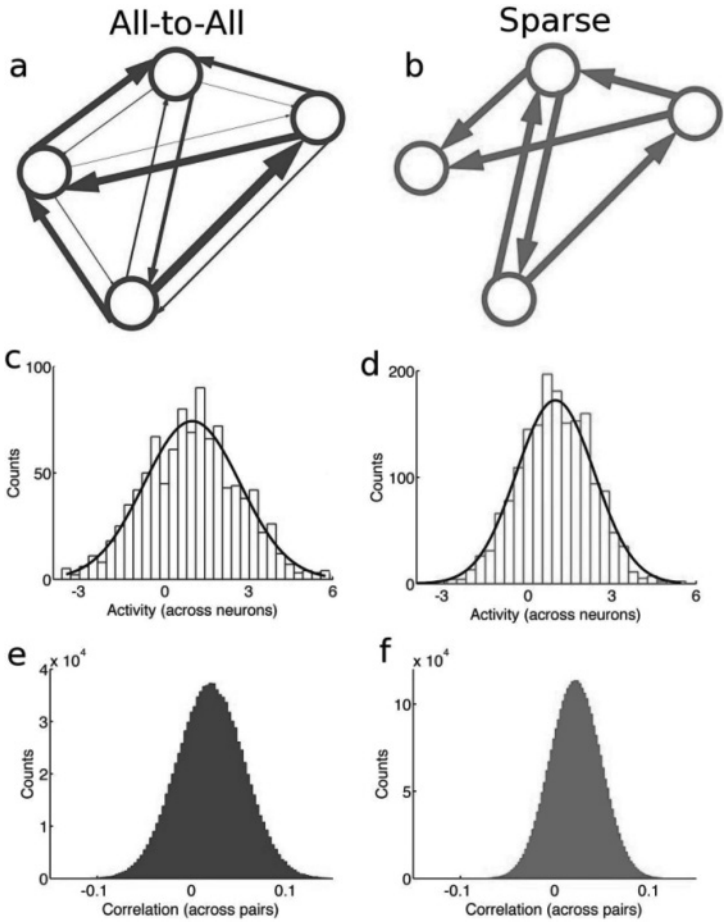

Figure 2: Scheme of the network and distribution of activity and correlations. We study two network architectures: (a) all-to-all connectivity with random strengths and (b) sparse random connections of fixed strength. Connection strength is illustrated by the thickness of edges. The distribution of activity across neurons (c, d) and the distribution of correlations across neuron pairs (e, f) are gaussian for both types of networks (c,e for the all-to-all network and d,f for the sparse network). The parameters used are $g = g_{ext} = 1$, $\overline{x}_{ext} = \overline{\Delta x^2_{ext}} = 1$, $K_{ext} = K$. For the all-to-all network, $K = 1000$, $\lambda_{ext} = 1$, $\lambda = 1/\sqrt{2}$. For the sparse network, $K = 880$ and $k_{ext} = k = 1/2$, which correspond to $\lambda = \lambda_{ext} = 1/\sqrt{2}$.

Angular brackets denote average over the matrix distribution, and Δ indicates variation around the mean. We adopt a single notation for either case, all-to-all or sparse network, by defining the mean and variance of

matrix elements and their scaling with K, N:

$$\langle G_{ij} \rangle = -kg/\sqrt{K} \qquad \left\langle \Delta G_{ij}^2 \right\rangle = \lambda^2/N \tag{2.6}$$

$$\left\langle G_{ij}^{ext} \right\rangle = k_{ext} g_{ext}/\sqrt{K_{ext}} \qquad \left\langle \Delta G_{ij}^{ext\,2} \right\rangle = \lambda_{ext}^2/N_{ext}. \tag{2.7}$$

In the all-to-all network, $k = 1$ and $K = N$. In the sparse network, for convenience of notation, we use the parameters $\lambda^2 = g^2(1-k)$ and $\lambda_{ext}^2 = g_{ext}^2(1-k_{ext})$. The mean connection is negative for G (inhibitory) and positive for G_{ext} (excitatory), since g and g_{ext} are positive. Note that the connections are strong in the sense that the magnitude of the excitatory and inhibitory input to each neuron, which is of order $\sqrt{K}$, is much larger than their sum (which is of order one; see section 3).

Theoretical analysis also considers the case in which local connections can be either excitatory or inhibitory, with two separate populations of excitatory and inhibitory neurons. In general, the analysis considers the case in which the mean and variance of the synaptic strength depend on the presynaptic neuron. We discuss in appendix B that all theoretical results hold provided that the parameters g and λ^2 are substituted by the means across presynaptic neurons. However, we do not show results of simulations for that case since that is outside the scope of this letter.

We assume that the external activity $\mathbf{x}_{ext}(t)$ is a stochastic process uncorrelated in both space and time, that is, a white noise characterized by mean $\overline{x_i^{ext}(t)} = \overline{x}_{ext}$ and covariance $\overline{\Delta x_i^{ext}(t)\Delta x_j^{ext}(t')} = \overline{\Delta x_{ext}^2}\delta_{ij}\delta(t-t')$ (the overbar denotes the average over different realizations of the noise, and δ denotes either the discrete Kronecker or continuous Dirac function). Therefore, equation 2.1 corresponds to a Ornstein-Uhlenbeck stochastic process (Gardiner, 1985).

We test theoretical results by running numerical computer simulations of the linear model. We simulate the dynamics of equation 2.1 with a simple Euler integration method, where each simulation runs for 200,000 time steps and each time step is 0.002 τ. For each set of parameter values, we use a single realization of the external input noise and a single realization of the random connectivity matrix. Since a simulation runs for a long time and the network is composed of a large number of neurons, we do not expect those specific realizations to affect the results significantly (in other words, we expect the system to be ergodic and self-averaging). We calculate sample mean and covariance by averaging across all time steps of a simulation. The correlation is calculated for each neuron pair by the standard Pearson's formula. Finally, we calculate the spatial mean and variance of those quantities across neurons (for the temporal mean) or across neuron pairs (for the covariance and correlation). We also use a semianalytic control by applying the spatial mean and variance on, instead of the full simulation, the theoretical results following the average over temporal noise but

preceding the average over spatial noise, namely, equations A.3 and A.9. The corresponding results are represented by filled symbols in the figures, while results of the full simulations are represented by open symbols.

The spiking network is defined by a current-based integrate-and-fire model. Its dynamics is described by the equations

$$\tau \frac{dI_i(t)}{dt} = -I_i(t) + \sum_{j=1}^{N} G_{ij} S_j(t) + \sum_{j=1}^{N_{ext}} G_{ij}^{ext} S_j^{ext}(t), \tag{2.8}$$

$$C_m \frac{dV_i(t)}{dt} = -g_m(V_i(t) - V_L) + I_i(t). \tag{2.9}$$

These equations are integrated using a simple Euler method with a time step $dt = 0.02$ ms. Equation 2.8 is similar to equation 2.1 of the linear model and describes the dynamics of the total current I_i received by neuron i—both the external excitatory and the recurrent inhibitory input (both types of input are integrated according to the same time constant τ). The matrices describing the synaptic strengths, G and G_{ext}, are defined in the same way as in the case of the linear model, in the fully connected case (see equations 2.2 and 2.3), although in the spiking model, those matrices are given in units of 8 nA·ms. In those units, the parameters are $g = 1$, $\lambda = 0.5$, $g_{ext} = 1$, $\lambda_{ext} = 0.58$. The variable $S_i(t)$ describes whether neuron i emits an action potential at time t or not, respectively, $S_i(t) = 1/dt$ or $S_i(t) = 0$. Equation 2.9 describes the dynamics of the membrane potential V_i of neuron i, which integrates linearly the total current according to the capacitance C_m and conductance g_m of the membrane, where V_L is the resting potential. If the membrane potential V_i exceeds the threshold potential V_{th} at time t, it is set to the reset potential V_{rs} and an action potential is emitted ($S_i(t) = 1/dt$). The variable $S_j^{ext}(t)$ describes the action potentials emitted by the external neurons. Their activity is modeled by a Poisson process characterized by an emission rate ϕ_{ext}, which is constant in time and equal for all external neurons. Parameters used in simulations are $\tau = 10$ ms, $V_{th} = -50$ mV, $V_{rs} = -70$ mV, $V_L = -70$ mV, $\phi_{ext} = 50$ Hz, $C_m = 0.4$ nF, and $g_m = 20$ nS (the time constant of the membrane potential is $C_m/g_m = 20$ ms).

We run 20 s simulations for different values of the network size, from $N = 50$ to $N = 1000$, with all other parameters fixed, each simulating 20 s of network activity (10^6 time steps). For a given network size, $N = 200$, we run additional five simulations at different values of the external input, from $\phi_{ext} = 45$ Hz to $\phi_{ext} = 55$ Hz. We use those five simulations to determine the change of the total current as a function to the change in ϕ_{ext}. This change is approximately linear and is quantified in terms of the four statistics studied in this work; the mean, the spatial variance, the temporal variance, and the covariance (see equations 3.1, 3.2, 3.4, and 3.5). All statistics are calculated with respect to the currents measured at the reference external input, $\phi_{ext} =$ 50 Hz. For example, the spatial variance is calculated by recording the

steady current for each neuron at the reference value and looking at the distribution across neurons of the difference between the reference current and the steady currents measured for the other values of the external input. Linear regression is applied to fit the linear change of the four statistics as a function of the external input, and the effective parameters $(g, \lambda, g_{ext}, \lambda_{ext})$ are determined by inverting the equations of the four statistics given by the linear model—equations 3.1, 3.2, 3.4, and 3.5—where $\overline{x}_{ext}$ and $\overline{\Delta x^2_{ext}}$ are the mean and variance of the change in external rate ($\overline{x}_{ext} = (\phi_{ext} - 50\ \text{Hz})$ and $\overline{\Delta x^2_{ext}} = (\phi_{ext} - 50\ \text{Hz})/\tau$). The effective parameters are used in equation 3.6 to predict correlations in the spiking model at variable network size.

3 Results

We study neural activity and correlations among neurons in a heterogeneous neural circuit model. Local recurrent connections are dominated by inhibition, while external feedforward projections are excitatory. Results are shown for a simple linear model, and at the end of the section, we also include simulations of a more realistic spiking network model. For the linear model we show the results of both theory and simulations, and we conclude by showing that the theory developed for the simple linear model can be used to predict correlations in the spiking network.

We consider two types of circuits: all-to-all connectivity with random strengths (see Figure 2a) and sparse random connections of fixed strengths (see Figure 2b). Results are displayed in a single notation for either case (see section 2). Figures 2c and 2d show the distribution of activity across neurons, and Figures 2e and 2f shows the distribution of correlations across neuron pairs. The purpose of this work is to describe how the mean and variance of those distributions depend on the parameters of the neural circuit. The activity values x are interpreted as deviations from a steady state of the input currents to each neuron, around which the neural dynamics is approximately linear. If we denote the steady current as I_0, the input current is equal to $I = I_0 + x$. As long as the linear approximation is valid, the correlations observed in the model are insensitive to the nature of the steady state (i.e., to the value of I_0).

Due to the linearity of the model, all quantities of interests can be simply calculated. The novel contribution of this work is averaging those quantities over the randomness of the connectivity matrix. Because connections are heterogeneous, different neurons have a different activity, and we compute the sample mean across neurons in order to obtain the spatial average. If the number of neurons N is large, this is independent of the specific realization of the connectivity; therefore we perform its average over the distribution of connections, and we obtain (see equation A.5 in appendix A; angular brackets denote averaging over the random connectivity, overline denotes

temporal average)

$$\langle \overline{x} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} \overline{x}_i \right\rangle = \frac{g_{ext}\sqrt{K_{ext}}}{1 + g\sqrt{K}} \, \overline{x}_{ext}. \tag{3.1}$$

The numerator of this expression is equal to the mean excitatory input received by a neuron, $g_{ext}\sqrt{K_{ext}}\,\overline{x}_{ext}$, while the denominator expresses the recurrent inhibition, whose total postsynaptic strength is $g\sqrt{K}$. Therefore, the strong recurrent inhibition counterbalances the large excitatory input and determines a relatively low activity, regardless of the network size. Note that the numbers of local and external connections, K and K_{ext}, are both large, but they tend to balance in the expression above. Figure 3a shows an example of mean activity as a function of the number of connections. The mean activity is rather insensitive to the number of connections, which are taken equal to the external ones in each simulation ($K = K_{ext}$). The analytical result, equation 3.1, agrees with numerical simulations of the linear dynamics in both the all-to-all and the sparse networks.

Different neurons have different connections and therefore different activity, and the extent to which the activity varies from neuron to neuron is determined by the spatial variance. We calculate this quantity by taking the sample variance across neurons and averaging over the random connectivity, and we obtain (see equation A.8)

$$\langle \Delta\overline{x}^2 \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} \Delta\overline{x}_i^2 \right\rangle = \frac{1}{1-\lambda^2}\left[\langle \overline{x} \rangle^2 \lambda^2 + \overline{x}_{ext}^2 \lambda_{ext}^2 \right]. \tag{3.2}$$

The spatial variance of neural activity increases with the network heterogeneity, expressed by λ and λ_{ext} for, respectively, the recurrent and external connections. Increasing the heterogeneity of connections increases the differences in the total input between neurons and therefore in their activities. The spatial variance is also proportional to the mean activity, local $\langle \overline{x} \rangle$ and external $\overline{x}_{ext}$. Furthermore, increasing the heterogeneity of recurrent connections leads to a divergence of the spatial variance, when λ^2 approaches one. In this case the state $x = 0$ destabilizes, and the linear approximation fails (see appendix A). Figure 3b shows an example of the spatial variance as a function of the variability of the recurrent connections. The analytical result, equation 3.2, agrees with numerical simulations of the linear dynamics in both the all-to-all and the sparse networks.

After looking at the mean and spatial variability of neural activity, we turn to the main theme of our work, the analysis of temporal variability and correlations. The activity of each neuron fluctuates in time due to the fluctuating input, and those temporal fluctuations may be correlated since different neurons receive shared input. We study temporal variability and

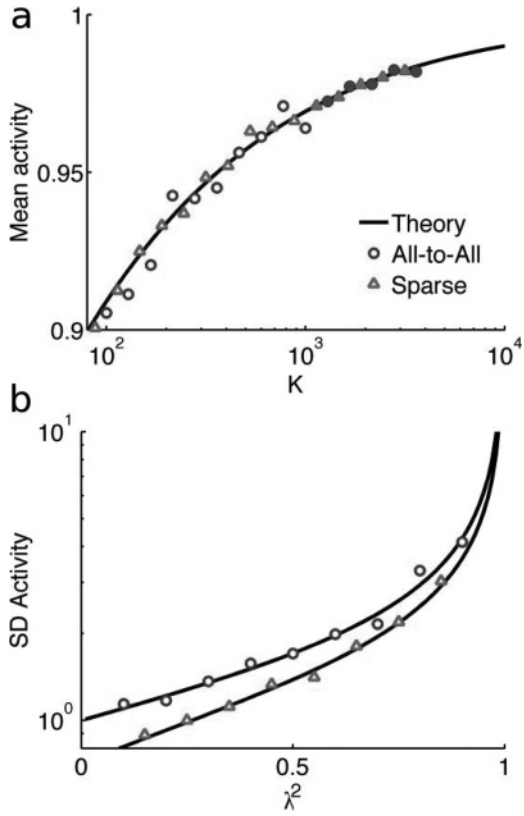

Figure 3: The mean activity has a mild dependence on the number of connections K (a). Its standard deviation (SD) strongly depends on the heterogeneity of connections λ (b). Analytical results (lines) are obtained from equation 3.1 in panel a and equation 3.2 in panel b. Simulation results are shown for the all-to-all (circles) and sparse (triangles) network. Open symbols show simulations of the neural dynamics, equation 2.1. Filled symbols show numerical evaluation of equation A.3. The parameters used are, in both panels, $g = g_{ext} = 1$, $\overline{x}_{ext} = \overline{\Delta x^2_{ext}} = 1$. In panel b, $K_{ext} = K = 500$. For the all-to-all network, $\lambda_{ext} = 1$ and $\lambda = 1/\sqrt{2}$ in panel a. For the sparse network: $k_{ext} = k = 1/2$ in panel a (which correspond to $\lambda = \lambda_{ext} = 1/\sqrt{2}$), while $k_{ext} = k$ is varied in panel b according to the value of λ (see section 2).

correlated fluctuations by calculating the covariance matrix, in particular the instantaneous covariance, at zero time lag. This is defined as

$$Q_{ij} = \overline{\Delta x_i \Delta x_j}. \tag{3.3}$$

First, we look at the on-diagonal elements of this matrix, which are the temporal variances of different neurons. To determine the average temporal variance, we take the sample mean across neurons and we average over the random connectivity, obtaining (see equation A.12)

$$\langle\overline{\Delta x^2}\rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N} Q_{ii} \right\rangle = \frac{\overline{\Delta x^2_{ext}}}{2}\left[\frac{k_{ext}g^2_{ext}}{1+g\sqrt{K}}\,\xi + \frac{\lambda^2_{ext}}{\sqrt{1-\lambda^2}}\right]. \tag{3.4}$$

The temporal variance of neural activity is the sum of two pieces: the first term decreases with the number of connections as $K^{-1/2}$, while the second term remains finite (the factor ξ is close to one; see equation A.13). The first term indicates that recurrent inhibition (g) reduces temporal fluctuations. In fact, the inhibitory feedback not only reduces the mean activity (see equation 3.1), but also cuts down fluctuations by quickly counterbalancing the external excitatory input. This can be verified by calculating the instantaneous covariance between the external excitatory and the local inhibitory input, which is found large and negative, equal to $-k_{ext}g^2_{ext}g\sqrt{K}$. The second term implies that nonzero fluctuations arise even in large networks (large K), and inhibition cannot exert an instantaneous and exact balance for each neuron. However, fluctuations nearly vanish if the external input is homogeneous ($\lambda_{ext} = 0$), in which case the inhibitory feedback would definitively counterbalance the homogeneous drive. Furthermore, as in the case of spatial fluctuations, temporal fluctuations increase with the heterogeneity of connections, recurrent (λ) and external (λ_{ext}). Temporal fluctuations diverge when the network approaches the instability point, when the linear approximation fails ($\lambda \to 1$).

How much of the total variance, expressed in equation 3.4, is independent rather than shared between neurons? To answer this question, we calculate the average covariance, by looking at the off-diagonal elements of the covariance matrix, the pairwise covariances. We take the sample mean across neuron pairs and average over the random connectivity to obtain the average covariance (see equation A.14),

$$\langle\overline{\Delta x'\Delta x''}\rangle = \left\langle \frac{1}{N(N-1)}\sum_{i\neq j}^{1,N} Q_{ij} \right\rangle = \frac{\overline{\Delta x^2_{ext}}}{2}\,\frac{k_{ext}g^2_{ext}}{1+g\sqrt{K}}. \tag{3.5}$$

Notably, this is proportional to the first term in the total variance, equation 3.4, by a factor close to one ($\xi \simeq 1$; see equation A.13), implying that the two terms in the total variance express, respectively, the correlated and uncorrelated fluctuations. Therefore, while the uncorrelated variance remains finite for large K, the correlated variance vanishes. The activities of neuron pairs tend to covary due to their shared external input, but the recurrent inhibition makes the covariance small, of order $K^{-1/2}$. In the sparse

network, the covariance vanishes if the probability of external connections is small ($k_{ext} \to 0$), since the shared external input between neurons tends to zero in that case. In the all-to-all network, the mean covariance vanishes if the mean input connection is zero ($g_{ext} = 0$). In that case, even if neurons receive a shared external input, neuron pairs may weight different inputs with the same or opposite signs, leading to, respectively, positive or negative covariance. Therefore, while the mean covariance across neuron pairs is zero, the covariance of single pairs may be positive or negative.

The mean correlation is obtained by dividing the covariance, equation 3.5, by the variance, equation 3.4 (we assume that variance and covariance are independent):

$$\langle R \rangle = \frac{\langle \overline{\Delta x' \Delta x''} \rangle}{\langle \overline{\Delta x^2} \rangle} = \frac{1}{\xi + \frac{\lambda_{ext}^2}{\sqrt{1-\lambda^2}} \frac{(1+g\sqrt{K})}{k_{ext} g_{ext}^2}}. \tag{3.6}$$

This expression is positive and never exceeds one. It indicates that the mean correlation is small, of order $K^{-1/2}$, despite the strong and shared excitatory input between neurons. However, this result holds only in presence of the local recurrent inhibition ($g > 0$) and provided that external connections are heterogeneous ($\lambda_{ext}^2 \neq 0$). Heterogeneity of local connections (λ) also contributes in decreasing the correlation.

Therefore, the inhibitory feedback and the random connectivity are responsible for the small correlation. If the inhibitory feedback is removed, $g = 0$, the correlation becomes large. If the network heterogeneity is removed, $\lambda = \lambda_{ext} = 0$, the correlation is equal to one, because the network is homogeneous and all neurons get the same input ($\xi = 1$ when $\lambda = 0$; see equation A.13). Figure 4 shows an example of the mean correlation as a function of the number of connections and the heterogeneity of the network. The analytical result, equation 3.6, agrees with numerical simulations of the linear dynamics in both the all-to-all and the sparse network. Insets in Figure 4 show the standard deviation of correlations, which appear to decrease with the number of connections as $K^{-1/2}$ and to increase with the heterogeneity of the network.

The final issue that we address is the timescale of correlations. Neural activity integrates the input on multiple timescales because of the large number of neurons and the heterogeneity of their connections. Which timescales are responsible for correlations? What correlations characterize the activity integrated in time? Note that the mean correlation in equation 3.6 and Figure 4 is the correlation at zero lag; the instantaneous correlation. We investigate the timescale of correlations in Figure 5, where the cross-correlation of neural activity is shown at different time lags. The correlation has a peak at zero lag and shows an exponential decay in time. As we have shown in Figure 4, the correlation at zero lag decreases with the number of connections as $K^{-1/2}$. Figure 5 shows that the timescale of correlation, determining

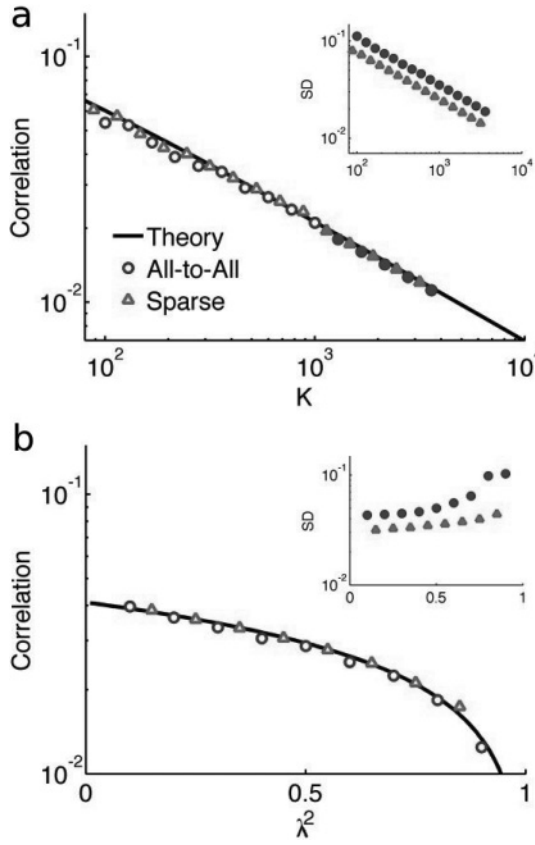

Figure 4: The mean correlation in the activity between neuron pairs decreases with the number of connections K (a) and their heterogeneity λ (b); standard deviation (SD) is shown in the insets. Analytical results (lines) are obtained from equation 3.6. Simulation results are shown for the all-to-all (circles) and sparse (triangles) network. Open symbols show simulations of the neural dynamics, equation 2.1. Filled symbols show the numerical evaluation of equation A.9. The parameters used are, in both panels, $g = g_{ext} = 1$, $\overline{x}_{ext} = \overline{\Delta x^2_{ext}} = 1$. In panel b, $K_{ext} = K = 500$. For the all-to-all network, $\lambda_{ext} = 1$, and $\lambda = 1/\sqrt{2}$ in panel a. For the sparse network, $k_{ext} = k = 1/2$ in panel a, which correspond to $\lambda = \lambda_{ext} = 1/\sqrt{2}$, while $k_{ext} = k$ is varied in panel b according to the value of λ (see section 2).

its rate of decay, also decreases with the number of connections. In fact, we show in appendix A that the integrated correlation across all time lags, namely, the total area of the cross-correlation, is of magnitude K^{-1}. Since the total area is approximately equal to correlation peak times temporal width,

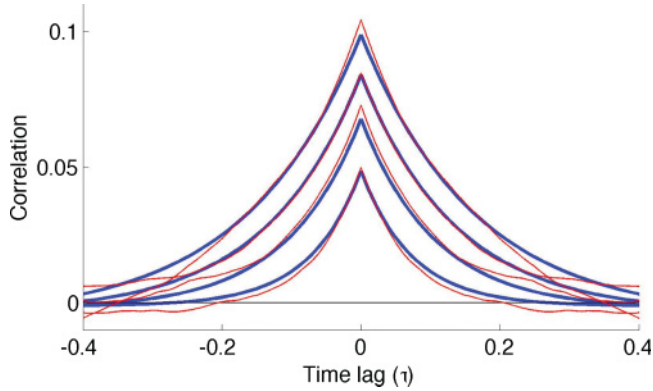

Figure 5: Both peak and width of the mean cross-correlation decrease with the number of connections K. The plot shows the cross-correlation as a function of time lag. Different curves, from top to bottom, correspond to $K = 100, 156, 278, 625$ in the all-to-all network. Consistent with Figure 4, the correlation at zero lag decreases with K. This plot shows that the temporal width of the cross-correlation also decreases with K. Red lines: Simulations of neural dynamics. Blue lines: Numerical evaluation of equation A.16. Parameters: $g = 0.5, g_{ext} = 1, \overline{x}_{ext} = \overline{\Delta x^2_{ext}} = 1, \lambda_{ext} = 1, \lambda = 1/\sqrt{2}$.

the temporal width is of order $K^{-1/2}$. Therefore, inhibition decorrelates on a fast timescale, and integrating neural activity, even for a relatively short time, has the effect of further decreasing the magnitude of correlations (see equation A.22).

It is worth noting that while in other studies, the results are often described in terms of the number of neurons N, here both N and the number of connections K play a role. For the sparse network, it is interesting to note that all the above results depend on the number of neurons N only through the parameter $\lambda^2 = g^2(1 - K/N)$, because N affects the sparsity of connections and therefore also their variance. The order of magnitude of correlations $K^{-\frac{1}{2}}$ holds regardless of the number of neurons, which may be taken even infinite for any fixed value of K. However, the dependence of correlations on the heterogeneity λ, and therefore N, may be quite substantial. Figure 6 shows how the mean correlation varies as a function of either K or N: for a relatively weak inhibition, the mean correlation depends mostly on the number of connections K, while for stronger inhibition, the mean correlation depends mostly on the number of neurons N.

3.1 Spiking Network Simulations. We tested the predictions of the linear model in a more realistic spiking network, described by a current-based integrate-and-fire model (see section 2). The spiking network is

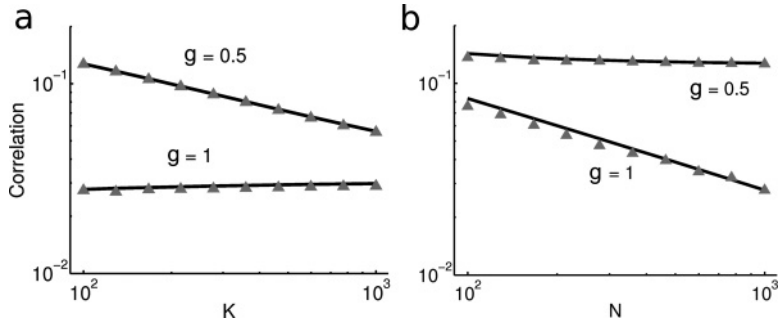

Figure 6: The mean correlation depends primarily on the number of connections K (a) or the number of neurons N (b) depending on the strength of inhibition. If inhibition is weak ($g = 0.5$), correlations depend mostly on K, while if inhibition is stronger ($g = 1$), correlations depend mostly on N. Analytical results (lines) are obtained from equation 3.6. Filled symbols show the numerical evaluation of equation A.9 for the sparse network. (a) $N = 1000$. (b) $K = 100$. Other parameter values are: $g_{ext} = g =$ value in figure, $\overline{x}_{ext} = \overline{\Delta x^2_{ext}} = 1$, $N_{ext} = 1000$, $K_{ext} = 500$, $k_{ext} = 0.5$. The values of λ and λ_{ext} vary according to the values of k, ρ, and ρ_{ext} (see section 2).

characterized by the nonlinear dynamics inherent in the generation of action potentials. However, we tested the hypothesis that this nonlinear system, when displaying small fluctuations around a steady state, may be approximated by a linear system and therefore by the equations derived in the previous section. Figure 7 shows the dynamics of an example neuron's input current and membrane potential, and spike times (rasters) of that neuron and other neurons from the spiking network. Simulation results reproduce qualitatively the phenomenology observed in the cerebral cortex. Since the input current puts neurons close to the firing threshold, neurons are susceptible to noise and fire irregularly, with noisy spike emission times (Shadlen & Newsome, 1998). The distribution of firing rates across neurons is broad, with a higher proportion of neurons displaying low firing rates (Baddeley et al., 1997; Hromadka, DeWeese, & Zador, 2008).

Our goal is not to provide a formal theory for the linear approximation of a spiking model. Instead, we show the results of a quantitative comparison of the two models, and we briefly summarize the theoretical arguments underlying this comparison. Since all results of the linear model are stated in terms of the mean and variance of the synaptic matrix, we hypothesize that the linear response of the spiking network can be described by an effective set of synaptic parameters $g, \lambda, g_{ext}, \lambda_{ext}$. For a given network size ($N = 200$), we probed the response of the spiking network to small changes in the external input, and we used these linear responses to fit the effective parameters of the connectivity. Then we used these parameters to predict

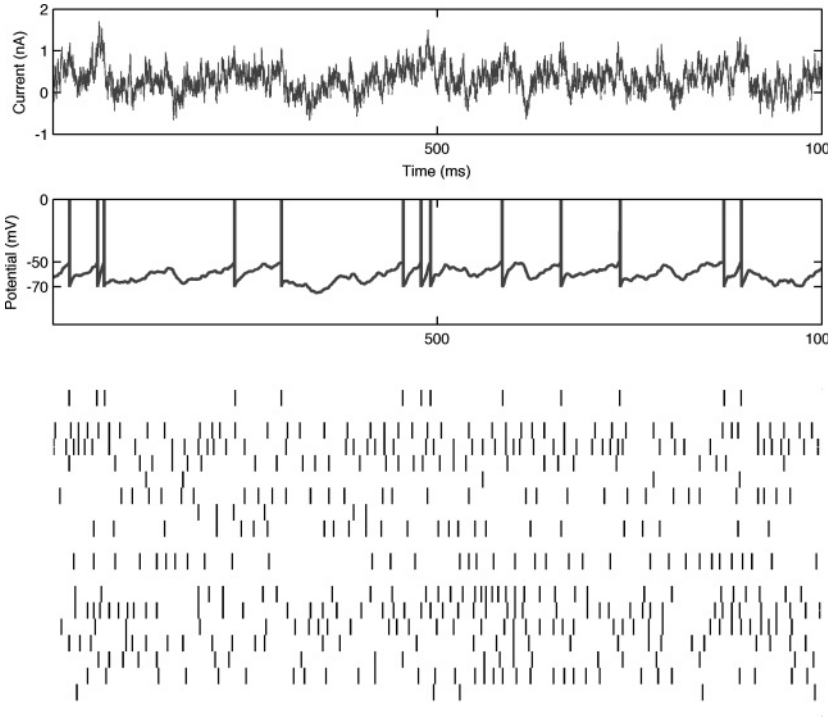

Figure 7: Example of dynamics in the spiking network simulation. (Top) Dynamics of the total input current (excitatory and inhibitory) to one neuron, in a time span of 1 second. (Middle) Dynamics of the membrane potential in the same neuron and temporal window. In the integrate-and-fire model, action potentials are instantaneous and of arbitrary size. (Bottom) Spike times (rasters) of 20 example neurons. Each row represents one neuron, and each tick represents one action potential for the corresponding neuron. The neuron in the top row corresponds to the neuron depicted in the top and middle part of the figure. Parameters used in simulations are $\tau = 10$ ms, $V_{th} = -50$ mV, $V_{rs} = -70$ mV, $V_L = -70$ mV, $\phi_{ext} = 50$ Hz, $C = 0.4$ nF, $g = 20$ nS. Synaptic parameters are $g = 1$, $\lambda = 0.5$, $g_{ext} = 1$, $\lambda_{ext} = 0.58$ (in units of 8 nA·ms).

the mean correlation across a wide range of network sizes (from $N = 50$ to $N = 1000$), without fitting any additional parameter.

Figures 8a to 8d shows the change in the total current (excitatory and inhibitory) as a consequence of the change in the external input rate. This change is approximately linear. The figure shows four statistics of the current: the change in mean current (see Figure 8a), spatial variance (see Figure 8b), temporal variance (see Figure 8c), and covariance (see Figure 8d). In the linear model, those quantities are calculated, respectively, in

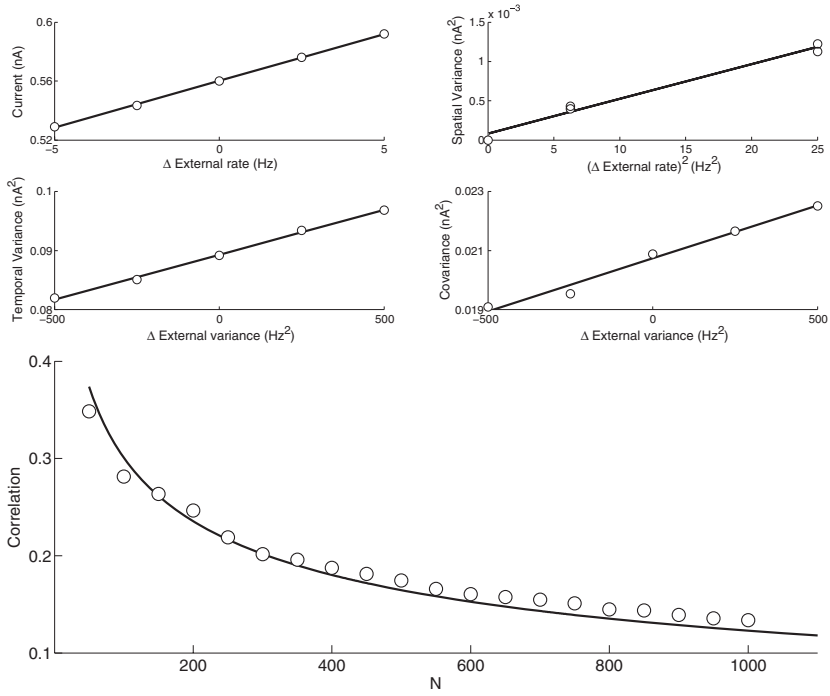

Figure 8: Linear response of the spiking network to external input at fixed $N = 200$ and correlations predicted from the linear response at variable network sizes from $N = 50$ to 1000. Four statistics of the current are presented: mean current (top left), spatial variance (top right), temporal variance (middle left), and covariance (middle right). Linear regression is used in the panels in the top and middle rows to fit effective parameters and predict correlations at variable network sizes in the panel in the bottom row. Circles: spiking network simulations; lines: linear regression fit and analytical prediction of correlation.

equations 3.1, 3.2, 3.4, and 3.5, where x_{ext} corresponds to the change in external input rate. We use linear regression to fit the slopes in Figures 8a to 8d and invert the equations of the linear model to obtain the effective values of the parameters of the connectivity (see section 2). The fitted values are $g = 2.44$, $\lambda = 0.54$, $g_{ext} = 1.99$, $\lambda_{ext} = 0.55$, to be compared with those used to generate the synaptic matrices in the spiking model (see section 2; $g = 1$, $\lambda = 0.5$, $g_{ext} = 1$, $\lambda_{ext} = 0.58$ in units of 8 nA·ms). Then we use these values to predict the mean correlation across a wide range of network sizes by using the formula for the linear model, equation 3.6. The result is plotted in Figure 8e, showing a remarkable agreement with theory. The fact that the theory provides a good fit for $N = 200$ is obvious, since parameters are fit

at that specific network size (although in a different simulation). However, the good fit across a wide range of network sizes suggests that equations of the linear model provide a good instrument to probe spiking networks.

The theoretical arguments underlying the above analysis are based on the dynamics of the total current integrated by the membrane potential of a neuron—both the inhibitory recurrent and the excitatory external current (see Figure 7 top). This dynamics is described by equation 2.8 in section 2, which is similar to the equation describing the linear system, equation 2.1. In the spiking network, the external stimulus is characterized by a sum over N_{ext} independent Poisson spike trains weighted by the matrix G^{ext}. If N_{ext} is large enough, this is approximately equal to a gaussian white noise process (Amit & Tsodyks, 1991) and therefore is equivalent to the external input of the linear model. The main difference between the linear and spiking model is the input from neurons in the recurrent network, determined by the spike trains $S_i(t)$. Those spike trains are nonlinear and noninstantaneous functions of the input currents and also provide an additional source of noise due to the discrete spike times. Nevertheless, we found that the parameters of an effective linear system, determined by the linear response of the spiking network, are able to well predict the correlations.

4 Discussion

We found that inhibitory feedback and heterogeneous connections have important effects on the dynamics of the activity in a neural circuit. The strong excitatory input, shared between neurons, tends to drive the network to a highly active and correlated state. The inhibitory feedback is responsible for balancing the network activity and also for reducing temporal fluctuations, in particular, the correlated fluctuations across neurons. The heterogeneity of couplings plays a crucial role in reducing correlations, since homogeneous connections would determine homogeneous and therefore highly correlated activity. As a consequence, the observed mean correlation is positive and of small magnitude. The fact that mean correlation is positive is obvious, since neurons in a large population cannot be anticorrelated on average.[1] What is not obvious is that the mean correlation is of small magnitude.

[1]The mean correlation must be larger than $-\frac{1}{N-1}$; therefore, it must be nonnegative in infinitely large populations. Proof: Any covariance matrix Q is positive definite; therefore $h^\dagger Q h > 0$ for any vector h. If we choose $h_i = 1/\sqrt{Q_{ii}}$ and define the correlation matrix $R_{ij} = Q_{ij}/\sqrt{Q_{ii}Q_{jj}}$, we have that $\sum_{ij} R_{ij} = N + N(N-1)\langle R\rangle > 0$. Therefore the mean correlation must be $\langle R\rangle > -\frac{1}{N-1}$. Tighter bounds on the mean correlation can be obtained by using $\lambda_{\min} \le h^\dagger Q h/|h|^2 \le \lambda_{\max}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the minimum and maximum eigenvalue of Q. This implies $(\lambda_{\min}/\lambda_{\max} - 1)/(N-1) \le \langle R\rangle \le (\lambda_{\max}/\lambda_{\min} - 1)/(N-1)$.

The main contribution of our work is an analytical calculation of the effect of heterogeneity on correlations in terms of random connectivity or random synaptic strengths. In the presence of heterogeneous connections and inhibitory feedback, the mean correlation at zero time lag is small, and it decreases with the number of connections as $K^{-\frac{1}{2}}$. The mean correlation integrated on large timescales is even smaller, of order K^{-1}, indicating that inhibition downsizes correlations on a timescale of $K^{-\frac{1}{2}}$. Other modeling studies have addressed the issue of correlations in neural circuits. In previous studies (Ginzburg & Sompolinski, 1994; Hertz, 2010; Renart et al., 2010; Tetzlaff et al., 2012), the mean correlation was found to decrease with the number of neurons as N^{-1}. Ginzburg and Sompolinski (1994) and Hertz (2010) studied a network in which connections strengths are of order N^{-1}, which implies a weak interaction between neurons and therefore a weak correlation. More surprisingly, Renart et al. (2010) and Tetzlaff et al. (2012) found weak correlations even in the case of strong interactions, with connection strengths of order $N^{-1/2}$. Both studies found a mean correlation of magnitude N^{-1}, provided that the correlation is integrated across large temporal windows. However, Renart et al. (2010) show that the mean correlation at zero time lag of membrane currents is of magnitude $N^{-1/2}$. These results are consistent with our findings (compare Figure 5 with Figure 2E of Renart et al., 2010, and Figure 3D of Tetzlaff et al., 2012). However, we highlight a potential difference by noting that the main parameter affecting correlations may be the number of connections K rather than the number of neurons N. The exclusive contribution of these two parameters has not been studied in detail in previous studies, and we have shown that it may depend on the network regime.

We also studied a more realistic spiking network model, and we confronted the analytical solutions of the linear model with the simulations of the nonlinear spiking model. We looked at small, linear changes in the current as a function of changes in the external firing rate input to the spiking network and computed the effective parameters of the linear model able to explain those changes. We found that those effective linearized parameters are able to predict correlations accurately even when changing the network size significantly. This suggests that the linear approximation is adequate for studying correlations. We did not consider the problem of a complete linear theory of spiking models, which would address the issues of computing the linearized kernel and the effective interaction matrix. Those issues have been studied, for example, in Lindner, Doiron, and Longtin (2005), Trousdale, Hu, Shea-Brown, and Kresimir (2012), and Tetzlaff et al. (2012).

The small mean correlation observed in our study and previous modeling studies agrees qualitatively with the experimental observations. The model may be useful to investigate correlations in different brain areas, especially those dominated by inhibition, such as the striatum and globus pallidus. Interestingly, large correlations between pallidal neurons have been shown to be correlated with Parkinsonism (Raz et al., 2000; Wilson,

Beverlin, & Netoff, 2011). The model is also consistent with the strong anticorrelations between excitatory and inhibitory inputs observed experimentally (Okun & Lampl, 2008; Cafaro & Rieke, 2010; Salinas & Sejnowski, 2000). However, different experimental studies report quantitative differences in measured correlations. For example, correlations depend on the temporal window on which action potentials are counted to determine a neuron's firing rate (see Figure 1). While neural dynamics occurs on a variety of timescales in our model,[2] as well as in real neurons (Bernacchia, Seo, Lee, & Wang, 2011), additional modeling studies are necessary to capture the wide range of phenomena observed in the experimental measures of correlations, including the effects of distance between neurons, multiple timescales and firing activity (Cohen & Kohn, 2011).

Appendix A: Statistics of Random Networks

In this section, we calculate the averages of neural activity and correlations with respect to both temporal fluctuations (noise) and the spatial variability of the connection strengths (disorder). Due to the linearity of the model, all quantities of interests can be simply calculated. The novel contribution of this work is averaging those quantities over the randomness of the connectivity matrix. The equation of dynamics, equation 2.1, can be expressed in matrix form (the time constant of temporal evolution τ is set to 1):

$$\frac{d\mathbf{x}(t)}{dt} = (G - I)\mathbf{x}(t) + G_{ext}\mathbf{x}_{ext}(t), \tag{A.1}$$

where $\mathbf{x}$ is the vector of local neural activities, $\mathbf{x}_{ext}$ is the vector of external neural activities, the matrices G (of size $N \times N$) and G_{ext} (size $N \times N_{ext}$) express respectively the recurrent connections and the feedforward projections, and I is the identity matrix. The equation of dynamics is linear and, given the interaction matrices G, G_{ext} and the input signal $\mathbf{x}_{ext}$, the neural activity can be expressed as a sum over the external input weighted by an

[2]Since the dynamics is linear (see equation A.1), the timescales of the network are determined by the eigenvalues of the matrix $(I - G)$. A random matrix with gaussian and independent elements has eigenvalues distributed uniformly in a circle in the complex plane centered at 0 and of radius λ (where its elements have variance λ^2/N; see Gudowska-Nowak, Janik, Jurkiewicz, & Nowak, 2003). One isolated eigenvalue is found approximately at mN, where m is the mean of the elements of the matrix and N is its dimension. Finally, the identity matrix translates all eigenvalues by one. Therefore, the eigenvalues of $(I - G)$ are contained in a circle centered at 1 and of radius λ, with one additional eigenvalue approximately equal to $1 + g\sqrt{K}$. Timescales, in units of τ, are equal to the inverse of those eigenvalues; therefore, the fastest timescale is equal to $1/(1 + g\sqrt{K})$, while the remaining timescales are between $(1 + \lambda)^{-1}$ and $(1 - \lambda)^{-1}$.

exponential temporal decay,

$$\mathbf{x}(t) = \int_{-\infty}^{t} dt' e^{(G-I)(t-t')} G_{ext} \mathbf{x}_{ext}(t') = \int_{0}^{+\infty} dt' e^{(G-I)t'} G_{ext} \mathbf{x}_{ext}(t-t'). \quad \text{(A.2)}$$

We assumed that initial conditions have decayed and that the inequality $\lambda < 1$ holds, to prevent network activity from growing in time without bounds. In the limit of large N, the real part of the eigenvalues of G is bounded by λ (Gudowska-Nowak et al., 2003). Therefore, if $\lambda \geq 1$, some eigenvalues of $(G - I)$ have a nonnegative real part, and the integral does not converge. This corresponds to an unstable fixed point at $\mathbf{x} = 0$, and network activity grows in time without bounds.

We start by calculating the mean neural activity. We perform the temporal average of the above expression; therefore, we substitute the external activity $\mathbf{x}_{ext}(t)$ with its average $\overline{x}_{ext}$, and we perform the integral, obtaining (temporal average is denoted by overbar)

$$\overline{\mathbf{x}} = \overline{x}_{ext} \ (I - G)^{-1} G_{ext} \mathbf{1}, \quad \text{(A.3)}$$

where the vector $\mathbf{1}$ has all N_{ext} components equal to one. Because the matrices of connection strengths are heterogeneous, G and G_{ext}, different neurons have a different mean activity. In order to calculate the spatially averaged activity, we compute the sample mean across neurons. For large N, this is independent of the specific realization of the spatial disorder; therefore, we perform its average over the distribution of connectivity strengths, namely,

$$\langle \overline{x} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} \overline{x_i} \right\rangle = \left\langle \frac{\overline{x}_{ext}}{N} \mathbf{1}^{\dagger} (I - G)^{-1} G_{ext} \mathbf{1} \right\rangle \quad \text{(A.4)}$$

The average (angular brackets) is across all possible realizations of the random matrices G and G_{ext}. We denote by $\dagger$ the transpose operation. Note that in expression A.4, the row vector $\mathbf{1}^{\dagger}$ has N components, while the column vector $\mathbf{1}$ has N_{ext}. In the following, we will use the same notation regardless of the dimension of $\mathbf{1}$, since that can be determined by the dimension of the multiplied matrix. Since G and G_{ext} are independent, we can substitute G_{ext} with its mean, $\langle G_{ext} \rangle = \frac{g_{ext}\sqrt{K_{ext}}}{N_{ext}} \mathbf{1}\mathbf{1}^{\dagger}$. Furthermore, we show in appendix B, equation B.15, that $< \mathbf{1}^{\dagger}(I - G)^{-1}\mathbf{1} >= N(1 + g\sqrt{K})^{-1}$. Therefore, the mean activity is equal to

$$\langle \overline{x} \rangle = \frac{g_{ext}\sqrt{K_{ext}}}{1 + g\sqrt{K}} \overline{x}_{ext} \quad \text{(A.5)}$$

This expression is used in the main text (see equation 3.1).

Different neurons have different connections and therefore different activity, and the extent to which the activity varies from neuron to neuron is determined by the spatial variance. We calculate this quantity by taking the sample variance across neurons and averaging over the spatial disorder. We take the scalar product of equation A.3 with itself, and we use again the fact that the sample mean does not depend on the spatial disorder for large N to obtain

$$\langle \Delta \overline{x}^2 \rangle = \left\langle \frac{\mathbf{x}^\dagger \mathbf{x}}{N} \right\rangle - \langle \overline{x} \rangle^2 = \left\langle \frac{\overline{x}_{ext}^2}{N} \mathbf{1}^\dagger G_{ext}^\dagger (I - G^\dagger)^{-1} (I - G)^{-1} G_{ext} \mathbf{1} \right\rangle - \langle \overline{x} \rangle^2. \tag{A.6}$$

We rewrite this expression by using the trace operator and its cyclic invariance. Namely, for any arbitrary matrices A, B, the following equations hold: $\mathbf{1}^\dagger A \mathbf{1} = \mathrm{Tr}(A \mathbf{1}\mathbf{1}^\dagger)$ and $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. We obtain

$$\langle \Delta \overline{x}^2 \rangle = \left\langle \frac{\overline{x}_{ext}^2}{N} \mathrm{Tr}((I - G^\dagger)^{-1} (I - G)^{-1} G_{ext} \mathbf{1}\mathbf{1}^\dagger G_{ext}^\dagger) \right\rangle - \langle \overline{x} \rangle^2. \tag{A.7}$$

Again, since G and G_{ext} are independent, we can average separately the factors involving the two matrices. A simple calculation shows that $\langle G_{ext} \mathbf{1}\mathbf{1}^\dagger G_{ext}^\dagger \rangle = g_{ext}^2 K_{ext} \mathbf{1}\mathbf{1}^\dagger + \lambda_{ext}^2 I$. Furthermore, we show in equations B.27 and B.28 that the following two equalities hold: $\langle \mathrm{Tr}((I - G)^{-1}(I - G^\dagger)^{-1}) \rangle = N(1 - \lambda^2)^{-1}$ and $\langle \mathrm{Tr}((I - G^\dagger)^{-1}(I - G)^{-1} \mathbf{1}\mathbf{1}^\dagger) \rangle = N(1 - \lambda^2)^{-1}(1 + g\sqrt{K})^{-2}$. Using the expression of the mean activity, equation A.5, the spatial variance is equal to

$$\langle \Delta \overline{x}^2 \rangle = \frac{1}{1 - \lambda^2} \left[\langle \overline{x} \rangle^2 \lambda^2 + \overline{x}_{ext}^2 \lambda_{ext}^2 \right]. \tag{A.8}$$

This expression is used in the main text, equation 3.2.

After looking at the spatial variability, we study temporal variability and correlated fluctuations by calculating the covariance matrix. We take the scalar product of equation A.2 with itself and we perform the temporal average, using the fact that the external stimulus is uncorrelated in space and time. This corresponds to the covariance matrix of an Ornstein-Uhlenbeck process (Gardiner, 1985) and is equal to

$$Q = \overline{\Delta \mathbf{x} \Delta \mathbf{x}^\dagger} = \overline{\Delta x_{ext}^2} \int_0^{+\infty} dt \; e^{(G-I)t} G_{ext} G_{ext}^\dagger \; e^{(G^\dagger - I)t}. \tag{A.9}$$

Note that the covariance matrix satisfies the Lyapunov equation,

$$(G - I)Q + Q(G^\dagger - I) + \overline{\Delta x_{ext}^2} G_{ext} G_{ext}^\dagger = 0, \tag{A.10}$$

but this cannot be used for averaging Q, since G and Q are dependent and do not commute. Note also that equation A.9 represents the covariance at zero time lag. We will consider the case of finite time lag at the end of this section.

The on-diagonal elements of the covariance matrix are the temporal variances of different neurons. To determine the average temporal variance, we take the sample mean across neurons and average over the spatial disorder, obtaining

$$\langle\overline{\Delta x^2}\rangle = \left\langle \frac{1}{N}\mathrm{Tr}(Q)\right\rangle = \frac{\overline{\Delta x^2_{ext}}}{N}\int_0^{+\infty} dt \left\langle \mathrm{Tr}\left(e^{(G^\dagger - I)t} e^{(G-I)t} G_{ext} G_{ext}^\dagger\right)\right\rangle, \quad \text{(A.11)}$$

where we applied the trace operator to select the diagonal elements, and we used its cyclic invariance. Again, since G and G_{ext} are independent, we can average separately the factors involving the two matrices. A simple calculation gives $\langle G_{ext} G_{ext}^\dagger\rangle = k_{ext} g_{ext}^2 \mathbf{1}\mathbf{1}^\dagger + \lambda_{ext}^2 I$. Furthermore, using equations B.30 and B.31, we obtain

$$\langle\overline{\Delta x^2}\rangle = \frac{\overline{\Delta x^2_{ext}}}{2}\left[\frac{k_{ext} g_{ext}^2}{1+g\sqrt{K}}\,\xi + \frac{\lambda_{ext}^2}{\sqrt{1-\lambda^2}}\right]. \quad \text{(A.12)}$$

This corresponds to equation 3.4 in the main text. The factor ξ is equal to

$$\xi = \left[1 - \frac{\lambda^2}{1+\sqrt{1-\lambda^2}(1+g\sqrt{K})}\right]^{-1}. \quad \text{(A.13)}$$

It is never smaller than one, and it is very close to one for a wide range of parameters, including for large K and small λ. However, it diverges near the critical point $\lambda \simeq 1$.

Next, we calculate the average covariance by looking at the off-diagonal elements of the matrix in equation A.9. The off-diagonal elements are the pairwise covariances, and the sample mean across neuron pairs can be averaged over the spatial disorder to obtain the average covariance. We use the matrix $(\mathbf{1}\mathbf{1}^\dagger - I)$ to select the off-diagonal elements and obtain

$$\langle\overline{\Delta x' \Delta x''}\rangle = \left\langle \frac{1}{N(N-1)}\mathrm{Tr}((\mathbf{1}\mathbf{1}^\dagger - I)Q)\right\rangle =$$
$$= \frac{\overline{\Delta x^2_{ext}}}{N(N-1)}\int_0^{+\infty} dt \left\langle \mathrm{Tr}\left(e^{(G^\dagger - I)t}(\mathbf{1}\mathbf{1}^\dagger - I) e^{(G-I)t} G_{ext} G_{ext}^\dagger\right)\right\rangle.$$

Again, we used the cyclic invariance of the trace operator, and since G and G_{ext} are independent, we can average separately the factors involving

the two matrices. We use $\langle G_{ext} G_{ext}^{\dagger} \rangle = k_{ext} g_{ext}^2 \mathbf{1}\mathbf{1}^{\dagger} + \lambda_{ext}^2 I$. Furthermore, using equations B.30 to B.32 and neglecting all terms of order N^{-1}, we obtain

$$\langle \overline{\Delta x' \Delta x''} \rangle = \frac{\overline{\Delta x_{ext}^2}}{2} \frac{k_{ext} g_{ext}^2}{1 + g\sqrt{K}}. \tag{A.14}$$

This expression is used in the main text in equation 3.5.

The mean correlation is obtained by dividing the covariance, equation A.14, by the variance, equation A.12 (we assume that variance and covariance are independent):

$$\langle R \rangle = \frac{\langle \overline{\Delta x' \Delta x''} \rangle}{\langle \overline{\Delta x^2} \rangle} = \frac{1}{\xi + \frac{\lambda_{ext}^2}{\sqrt{1-\lambda^2}} \frac{(1+g\sqrt{K})}{k_{ext} g_{ext}^2}}. \tag{A.15}$$

This expression is positive and never exceeds one. This corresponds to equation 3.6 in the main text.

We now turn to calculating the correlations of activity integrated in time. In order to calculate those correlations, we define the covariance at time lag Δt as $Q_x(\Delta t)$. Note that equation A.9 represents the covariance at zero time lag, a special case of the covariance at time lag Δt, namely, $Q = Q_x(0)$. The covariance at finite time lag $\Delta t = t' - t''$ is equal to (Gardiner, 1985)

$$Q_x(\Delta t) = \overline{\Delta \mathbf{x}(t') \Delta \mathbf{x}(t'')^{\dagger}} = \begin{cases} e^{(G-I)\Delta t}\, Q & \text{for } \Delta t \geq 0 \\ Q\, e^{-(G^{\dagger}-I)\Delta t} & \text{for } \Delta t < 0 \end{cases}. \tag{A.16}$$

We define the temporally integrated activity as a linear convolution of the activity, namely,

$$\mathbf{y}(t) = \int_{-\infty}^{+\infty} dt'\, h(t - t') \mathbf{x}(t'), \tag{A.17}$$

where $h(t)$ is a given convolution kernel. Using the Wiener-Khinchin theorem and the convolution theorem, it is straightforward to calculate the covariance of the integrated activity, which is equal to

$$Q_y(\Delta t) = \overline{\Delta \mathbf{y}(t') \Delta \mathbf{y}(t'')^{\dagger}} = \int_{-\infty}^{+\infty} dt\; h_2(\Delta t - t) Q_x(t), \tag{A.18}$$

where the second-order kernel is equal to $h_2(t) = \int_{-\infty}^{+\infty} dt' h(t') h(t' + t)$. If time integration is slow enough, such that the kernels h and h_2 are approximately constant in a time interval in which the covariance Q_x is sensibly

different from zero, the above expression simplifies to

$$Q_y(\Delta t) \simeq h_2(\Delta t)\int_{-\infty}^{+\infty} dt\; Q_x(t) = h_2(\Delta t)\left[(I-G)^{-1}Q + Q(I-G^\dagger)^{-1}\right] =$$
$$= h_2(\Delta t)\overline{\Delta x^2_{ext}}\;(I-G)^{-1}G_{ext}G^\dagger_{ext}(I-G^\dagger)^{-1}$$

In the last two equalities we have, respectively, integrated equation A.16 and used equation A.10, which we have multiplied by $(I-G)^{-1}$ on the left side and by $(I-G^\dagger)^{-1}$ on the right side. The latter expression can be averaged over the network disorder to compute the mean correlation of the integrated activity. We will consider only the case of $\Delta t = 0$, since the case $\Delta t \neq 0$ is straightforward and is not our focus, and we denote $Q_y = Q_y(0)$. In addition, we substitute $h_2(0) = T^{-1}$, where T is defined as the characteristic integration time of the kernel.

The on-diagonal elements of Q_y are the temporal variances of the integrated activity of different neurons. As in the computation of the variance of x, we take the sample mean across neurons and average over the spatial disorder, obtaining

$$\langle\overline{\Delta y^2}\rangle = \left\langle \frac{1}{N}\mathrm{Tr}(Q_y)\right\rangle = \frac{\overline{\Delta x^2_{ext}}}{TN}\left\langle \mathrm{Tr}\left((I-G^\dagger)^{-1}(I-G)^{-1}G_{ext}G^\dagger_{ext}\right)\right\rangle. \quad \text{(A.19)}$$

Again, we applied the trace operator to select the diagonal elements, we used its cyclic invariance, and since G and G_{ext} are independent we can average separately the factors involving the two matrices. We use $\langle G_{ext}G^\dagger_{ext}\rangle = k_{ext}g^2_{ext}\mathbf{1}\mathbf{1}^\dagger + \lambda^2_{ext}I$ and equations B.27 and B.28 to obtain

$$\left\langle\overline{\Delta y^2}\right\rangle = \frac{\overline{\Delta x^2_{ext}}}{T(1-\lambda^2)}\left[\frac{k_{ext}g^2_{ext}}{(1+g\sqrt{K})^2} + \lambda^2_{ext}\right]. \quad \text{(A.20)}$$

Note that the first term in square brackets is small, of order K^{-1}, and could be neglected.

Next, we calculate the average covariance of the integrated activity by looking at the off-diagonal elements of the matrix Q_y. The off-diagonal elements are the pairwise covariances, and the sample mean across neuron pairs can be averaged over the spatial disorder to obtain the average covariance. As in the computation of the covariance of x, we use the matrix $(\mathbf{1}\mathbf{1}^\dagger - I)$ to select the off-diagonal elements, and we obtain

$$\left\langle\overline{\Delta y'\Delta y''}\right\rangle = \left\langle \frac{1}{N(N-1)}\mathrm{Tr}\left((\mathbf{1}\mathbf{1}^\dagger - I)Q_y\right)\right\rangle =$$
$$= \frac{\overline{\Delta x^2_{ext}}}{TN(N-1)}\left\langle \mathrm{Tr}\left((I-G^\dagger)^{-1}(\mathbf{1}\mathbf{1}^\dagger - I)(I-G)^{-1}G_{ext}G^\dagger_{ext}\right)\right\rangle.$$

Again, we used the cyclic invariance of the trace operator, and since G and G_{ext} are independent, we can average separately the factors involving the two matrices. We use $\langle G_{ext} G_{ext}^{\dagger} \rangle = k_{ext} g_{ext}^2 \mathbf{1}\mathbf{1}^{\dagger} + \lambda_{ext}^2 I$ and equations B.27 to B.29. Because the leading term is K^{-1}, here we keep terms of order N^{-1}, and we obtain

$$\overline{\langle \Delta y' \Delta y'' \rangle} = \frac{\overline{\Delta x_{ext}^2}}{T} \left[\frac{k_{ext} g_{ext}^2}{(1 + g\sqrt{K})^2} - \frac{1}{N} \frac{\lambda_{ext}}{1 - \lambda^2} \right]. \tag{A.21}$$

The mean correlation of the integrated activity is obtained by dividing the covariance, equation A.21, by the variance, equation A.20 (we assume that variance and covariance are independent):

$$\langle R_y \rangle = \frac{\overline{\langle \Delta y' \Delta y'' \rangle}}{\overline{\langle \Delta y^2 \rangle}} = \frac{k_{ext} g_{ext}^2}{(1 + g\sqrt{K})^2} \frac{(1 - \lambda^2)}{\lambda_{ext}^2} - \frac{1}{N}. \tag{A.22}$$

Note that we neglected the term of order K^{-1} in using equation A.20. This expression shows that correlations of integrated activity can be negative and are small, of order K^{-1}.

Appendix B: Traces of Random Matrix Products

In this appendix, we introduce the diagrammatic notation to calculate the quenched averages of random matrix products (see, e.g., Gudowska-Nowak et al., 2003). In the context of neural networks, a diagrammatic notation has been also implemented recently by Rangan (2009), Pernice, Staude, Cardanobile, and Rotter (2011), and Trousdale et al. (2012). Theoretical results are obtained for the gaussian distribution, although numerical simulations suggest that they generalize to other distributions with the same mean and variance (e.g., Bernouilli). We consider the case in which the mean of the matrix element is $\sim g/N$ and then recover the scaling studied in the main text by analytical continuation and the substitution $g \to g\sqrt{K}$. We conclude by studying the case of a nonhomogeneous mean (e.g., interconnected excitatory and inhibitory neurons).

We start with the problem of calculating the quenched average of the trace of a power of the random matrix R in the limit of large N (where the size of the matrix is $N \times N$). The matrix R is characterized by independent and normally distributed elements, each element having zero mean and variance N^{-1}:

$$\langle R_{ij} \rangle = 0 \qquad \langle R_{ij}^2 \rangle = \frac{1}{N} \tag{B.1}$$

We start by calculating the second order, that is, the average trace of the square of R. For convenience of notation, we omit the sum over the indices

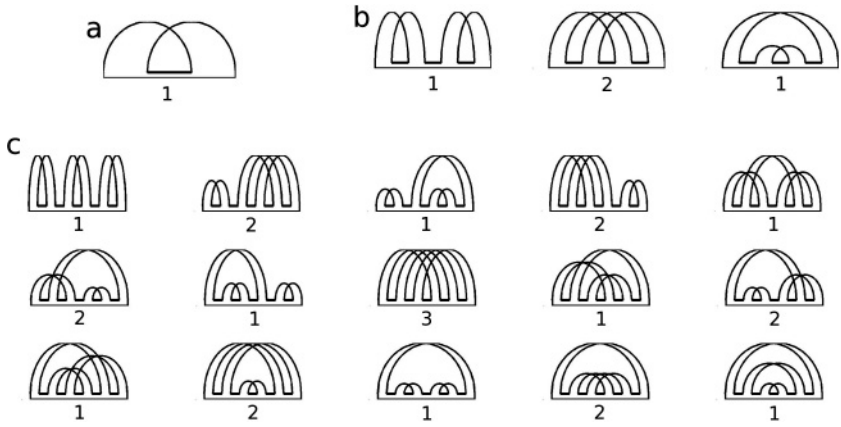

Figure 9: Diagrams of the traces of random matrix powers, described by equation B.5. Below each diagram, the number of its closed loops is indicated. (a) Second order. (b) Fourth order. (c) Sixth order.

(in this case, the sum over the indices a, b, c, d):

$$\left\langle \mathrm{Tr}\left(R^2\right)\right\rangle = \delta_{ad}\delta_{bc}\langle R_{ab}R_{cd}\rangle = N^{-1}\,\delta_{ad}\delta_{bc}\delta_{ac}\delta_{bd} \tag{B.2}$$

The diagram corresponding to this expression is shown in Figure 9a. The diagram is obtained by drawing one node for each one of the four indices a, b, c, d and by drawing an edge for each delta function in the expression, where the two nodes connected by the edge correspond to the two indices of the delta function. Horizontal edges are due to the operations of trace (base edge) and matrix multiplication (middle edge), while arc-shaped edges are due to averaging. The multiple edges determine different paths, and each pair of nodes connected by a path (even if not linked by an edge) corresponds to a pair of indices that must be equal, since they are connected by a sequence of delta functions. Therefore, for each closed loop in the diagram, there is one redundant delta function, which can be eliminated without performing the sum over the corresponding indices. This implies that each closed loop contributes with a factor N due to a free sum over N elements. Since the diagram for the second order has one loop, we have $\langle \mathrm{Tr}\left(R^2\right)\rangle = N^{-1}N = 1$.

Note that all terms of odd order are zero, because $\langle R_{ij}^k\rangle = 0$ for odd k. The next order is therefore the fourth order, which is equal to (again we omit the sum over all indices)

$$\left\langle \mathrm{Tr}\left(R^4\right)\right\rangle = \delta_{ah}\delta_{bc}\delta_{de}\delta_{fg}\langle R_{ab}R_{cd}R_{ef}R_{gh}\rangle = \tag{B.3}$$

$$= N^{-2}\,\delta_{ah}\delta_{bc}\delta_{de}\delta_{fg}\left[\delta_{ac}\delta_{bd}\delta_{eg}\delta_{fh} + \delta_{ae}\delta_{bf}\delta_{cg}\delta_{dh} + \delta_{ag}\delta_{bh}\delta_{ce}\delta_{df}\right]. \tag{B.4}$$

The fourth order has three diagrams—one for each term in the sum, shown in Figure 9b. The middle diagram has two closed loops, and the other two have only one loop. Therefore, the other two terms can be neglected; the middle term contributes with a factor N^2, and the fourth-order gives $\langle \mathrm{Tr}(R^4) \rangle = 1$. The contribution of the fourth-order moment ($\langle R_{ij}^4 \rangle$) can be neglected because the corresponding terms in the sum have quartets of indices with the same value. The number of those terms is smaller by a factor of N^2 with respect to the number of second-order terms. Similar arguments apply for higher order moments.

The 15 diagrams of the sixth order are shown in Figure 9c. Again, we neglect moments higher than the second, and we note that only one diagram contributes with three loops; therefore, $\langle \mathrm{Tr}(R^6) \rangle = 1$. By iterating this procedure, we find that order $2k$ has $(2k-1)!!$ diagrams of which only one has k loops; therefore

$$\left\langle \mathrm{Tr}(R^{2k}) \right\rangle = 1 \tag{B.5}$$

for all values of k.

Note that the elements of the matrix R have zero mean, while the matrices considered in the main text (G and G_{ext}) have nonzero mean. As explained below, in order to calculate the average trace of matrix powers with nonzero mean, we need to compute averages where R is interleaved by the matrix of ones. We denote by $\mathbf{1}$ the column vector of N components all equal to one, by $\mathbf{1}^\dagger$ the row vector, and by $\mathbf{1}\mathbf{1}^\dagger$ the $N \times N$ matrix with all elements equal to one (we denote by $\dagger$ the transpose operation). We consider the two second-order terms:

$$\langle \mathrm{Tr}(R\mathbf{1}\mathbf{1}^\dagger R) \rangle = \delta_{ad} \langle R_{ab} R_{cd} \rangle = N^{-1} \delta_{ad} \delta_{ac} \delta_{bd} = 1, \tag{B.6}$$

$$\langle \mathrm{Tr}\left(R^2 \mathbf{1}\mathbf{1}^\dagger\right) \rangle = \delta_{bc} \langle R_{ab} R_{cd} \rangle = N^{-1} \delta_{bc} \delta_{ac} \delta_{bd} = 1. \tag{B.7}$$

It is not surprising that these two expressions are equal, since the trace is cyclic invariant. The only difference between these expressions and equation B.2 is the absence of a factor δ_{bc} in the former expression and δ_{ad} in the latter. This corresponds to cutting, respectively, the middle and the base horizontal edges in the diagram of Figure 9a. In general, inserting a matrix of ones at a given point of the sequence of R products is equivalent to cutting the horizontal edge at that point in the corresponding diagram. If the edge belongs to a closed loop, the cut has the effect only of removing a redundant delta function; there is no change in the contribution of that diagram to the sum. Conversely, if the edge belongs to an open path, the cut determines an additional N factor, because the delta function removed was not redundant. Since all diagrams have at least one closed loop, inserting a

single matrix of ones has no effect at all orders. Therefore,

$$\langle \mathrm{Tr}(R^{2k-k'}\mathbf{1}\mathbf{1}^\dagger R^{k'})\rangle = 1 \tag{B.8}$$

for all $k' = 0, \ldots, 2k$. Unless more loops are available to cut, inserting more matrices of ones may cut open paths; therefore, the trace may be multiplied by N. An additional N factor is obtained also by multiplying the matrix of ones with itself, which occurs whenever additional matrices are inserted at the same point in the sequence (we have that $\mathbf{1}^\dagger\mathbf{1} = N$ and $(\mathbf{1}\mathbf{1}^\dagger)^k = N^{k-1}\mathbf{1}\mathbf{1}^\dagger$ if $k > 0$).

Using these results, we can calculate the average trace of random matrix powers with nonzero mean and arbitrary variance (provided that the variance is of order N^{-1}). We consider the matrix G equal to

$$G = \frac{g}{N}\mathbf{1}\mathbf{1}^\dagger + \lambda R. \tag{B.9}$$

Note that the mean of this matrix has a different scaling with respect to that considered in the main text, but we will recover the latter by the substitution $g \to -\sqrt{K}g$. A power of G is calculated by multiplying G to itself, and this determines an ordered product of powers of the matrices R and $\mathbf{1}\mathbf{1}^\dagger$. Note that these two matrices do not commute; therefore, the binomial theorem cannot be applied. We consider the average trace

$$\langle \mathrm{Tr}(G^k)\rangle = \sum_{k'=0}^{k} N^{-k'} g^{k'} \lambda^{k-k'} \sum^{\binom{k}{k'}} \langle \mathrm{Tr}(\ldots)\rangle, \tag{B.10}$$

where the trace on the right-hand side is applied to an ordered product of k' matrices $\mathbf{1}\mathbf{1}^\dagger$ and $k - k'$ matrices R, and the sum runs over all the $\binom{k}{k'}$ ordered products for a given k and k'. Using the above results, we find that the contribution of any of those traces is zero for $k - k'$ odd, equal to one for $k' = 0$ (provided that k is even), equal to N^k for $k' = k$, and at most of order $N^{k'-1}$ for $k' = 1, \ldots, k-1$. Therefore, the leading-order terms are $k' = k$ (for any value of k) and $k' = 0$ (for k even); all other terms can be neglected, and we find

$$\langle \mathrm{Tr}(G^k)\rangle = g^k + \lambda^k \delta_{k,even}. \tag{B.11}$$

If the matrix G^k is further multiplied by a matrix of ones, the term $k' = 0$ can also be neglected, and we find that

$$\langle \mathrm{Tr}(G^k\mathbf{1}\mathbf{1}^\dagger)\rangle = Ng^k = \mathrm{Tr}(\langle G\rangle^k\mathbf{1}\mathbf{1}^\dagger) \tag{B.12}$$

for all values of k. Note that if the mean of G has a higher order in N, the result still holds. This expression is particularly useful to compute the average of bracket expressions. Because $\mathrm{Tr}(A\mathbf{x}\mathbf{y}^\dagger)=\mathbf{y}^\dagger A\mathbf{x}$ for any matrix A and vectors $\mathbf{x}$, $\mathbf{y}$, the expression can be rewritten as

$$\langle \mathbf{1}^\dagger G^k \mathbf{1}\rangle = \mathbf{1}^\dagger \langle G\rangle^k \mathbf{1} \tag{B.13}$$

for all values of k. Since any infinitely differentiable function f can be expanded in Taylor series, the above result implies that

$$\langle \mathbf{1}^\dagger f(G)\mathbf{1}\rangle = \mathbf{1}^\dagger f(\langle G\rangle)\mathbf{1}. \tag{B.14}$$

Therefore, the following expression can be calculated and used to compute the mean activity in the main text:

$$\langle \mathbf{1}^\dagger (I-G)^{-1}\mathbf{1}\rangle = \frac{N}{1-g}. \tag{B.15}$$

Note that the substitution $g \to -g\sqrt{K}$ must be applied to recover the scaling studied in the main text.

Next, we calculate the diagrammatic expansion for products of a random matrix with its transpose. Again, all odd orders vanish, and we neglect moments higher than the second at all orders. The second-order term is

$$\langle \mathrm{Tr}(RR^\dagger)\rangle = \delta_{ad}\delta_{bc}\langle R_{ab}R_{dc}\rangle = N^{-1}\,\delta_{aa}\delta_{bc}\delta_{aa}\delta_{bc}. \tag{B.16}$$

The corresponding diagram has two loops and is shown in Figure 10a. Therefore, the loops contribute with a factor N^2, and the second order is $\langle \mathrm{Tr}(RR^\dagger)\rangle = N$. The fourth order is equal to

$$\langle \mathrm{Tr}(R^2R^{2\dagger})\rangle = \delta_{ah}\delta_{bc}\delta_{de}\delta_{fg}\langle R_{ab}R_{cd}R_{fe}R_{hg}\rangle = \tag{B.17}$$

$$= N^{-2}\,\delta_{ah}\delta_{bc}\delta_{de}\delta_{fg}[\delta_{ac}\delta_{bd}\delta_{fh}\delta_{eg} + \delta_{af}\delta_{be}\delta_{ch}\delta_{dg} + \delta_{ah}\delta_{bg}\delta_{cf}\delta_{de}]. \tag{B.18}$$

The three diagrams are shown in Figure 10b. The first two diagrams have one loop, and the third has three. Therefore, that diagram contributes with a factor N^3, and the fourth order is equal to $\langle \mathrm{Tr}(R^2R^{2\dagger})\rangle = N$. The diagrams for the sixth order are shown in Figure 10c: only one diagram has four loops, and no diagram has three; therefore, $\langle \mathrm{Tr}(R^3R^{3\dagger})\rangle = N$. Iterating the procedure, we find that order $2k$ has $(2k-1)!!$ diagrams of which only one has $k+1$ loops; therefore

$$\langle \mathrm{Tr}(R^kR^{k\dagger})\rangle = N \tag{B.19}$$

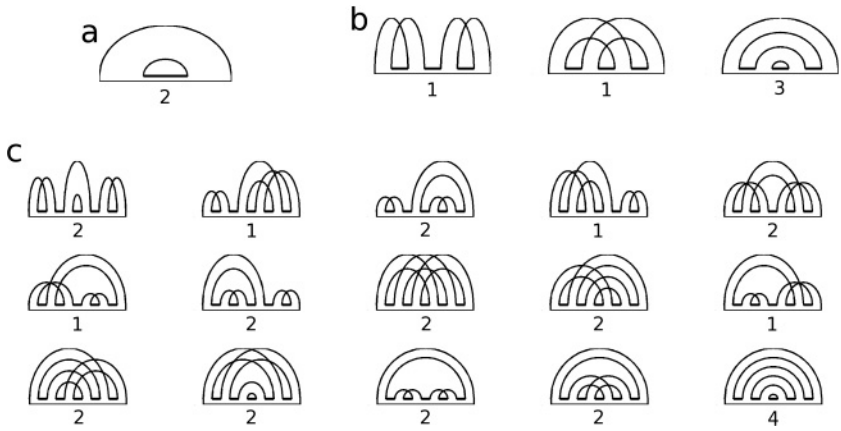

Figure 10: Diagrams of the traces of random matrix powers multiplied by its transpose, described by equation B.19. Below each diagram the number of its closed loops is indicated. (a) Second order. (b) Fourth order. (c) Sixth order.

for all values of k. Other combinations of powers of R and its transpose give a smaller contribution—$\langle \mathrm{Tr}(R^{2k-k'} R^{k'\dagger}) \rangle = o(1)$ for $k' \neq k$.

Inserting matrices of ones in this case has a similar effect as in the case above, equation B.8: each matrix cuts the horizontal edge corresponding to where the matrix is placed. Again, since each diagram has at least one loop, the insertion of a single matrix of ones (and the consequent edge removal) has no effect on the trace at all orders. Therefore,

$$\langle \mathrm{Tr}(R^{k-k'} \mathbf{1}\mathbf{1}^\dagger R^{k'} R^{k\dagger}) \rangle = \langle \mathrm{Tr}(R^k R^{k'\dagger} \mathbf{1}\mathbf{1}^\dagger R^{k-k'\dagger}) \rangle = N. \tag{B.20}$$

An insertion in a term with unequal powers of R and $R^\dagger$ remains of order one. Adding more matrices increases the trace by an order N for each matrix, provided that no further loops are cut.

Using the above expressions, we can compute the average of products of powers of the matrix G and its transpose,

$$\langle \mathrm{Tr}(G^k G^{l\dagger}) \rangle = \sum_{k'=0}^{k} \sum_{l'=0}^{l} N^{-k'-l'} g^{k'+l'} \lambda^{k+l-k'-l'} \sum^{\binom{k}{k'}\binom{l}{l'}} \langle \mathrm{Tr}(\ldots) \rangle, \tag{B.21}$$

where the trace on the right-hand side is applied to an ordered product of $k'+l'$ matrices $\mathbf{1}\mathbf{1}^\dagger$, $k-k'$ matrices R and $l-l'$ matrices $R^\dagger$. If $k'=k$ and $l'=l$, the trace is equal to N^{k+l}, and the term is of order one. If $k'=0$ and $l'=0$, the trace contributes with an order N, provided that $k=l$. If $k-k'=l-l'$, the traces contribute at most with an order $N^{k'+l'}$, and the term is of order one,

while if $k - k' \neq l - l'$, the term is of smaller order. Therefore, the leading order is N, and we have

$$\langle \mathrm{Tr}(G^k G^{l\dagger}) \rangle = N \delta_{kl} \lambda^{k+l}. \tag{B.22}$$

In the case in which matrices of ones are inserted, the term $k' = 0, l' = 0$ is no longer leading, and many other terms have to be considered. Those are the terms for $k - k' = l - l'$, and for which additional inserted matrices cut the same loop. Since the leading diagrams at all orders have one two-node loop in the middle and one at the boundaries, if a matrix of ones is inserted in the middle or at the boundaries, additional matrices must continue to be inserted at the same place in order to cut the same loop. We eliminate one sum and use the index $m = k - k' = l - l'$ in place of k' and l'. We obtain

$$\begin{aligned} &\langle \mathrm{Tr}(G^k G^{l\dagger} \mathbf{1}\mathbf{1}^\dagger) \rangle \\ &= \sum_{m=0}^{\min(k,l)} N^{2m-k-l} g^{l+k-2m} \lambda^{2m} \langle \mathrm{Tr}((\mathbf{1}\mathbf{1}^\dagger)^{k-m} R^m R^{m\dagger} (\mathbf{1}\mathbf{1}^\dagger)^{l-m+1}) \rangle, \end{aligned} \tag{B.23}$$

$$\begin{aligned} &\langle \mathrm{Tr}(G^k \mathbf{1}\mathbf{1}^\dagger G^{l\dagger}) \rangle \\ &= \sum_{m=0}^{\min(k,l)} N^{2m-k-l} g^{l+k-2m} \lambda^{2m} \langle \mathrm{Tr}(R^m (\mathbf{1}\mathbf{1}^\dagger)^{k+l-2m+1} R^{m\dagger}) \rangle. \end{aligned} \tag{B.24}$$

Both expressions are equal to

$$\langle \mathrm{Tr}(G^k G^{l\dagger} \mathbf{1}\mathbf{1}^\dagger) \rangle = \langle \mathrm{Tr}(G^k \mathbf{1}\mathbf{1}^\dagger G^{l\dagger}) \rangle = N \sum_{m=0}^{\min(k,l)} g^{l+k-2m} \lambda^{2m}. \tag{B.25}$$

Furthermore, we calculate the average trace with two inserted matrices. In that case, the leading term is for $k = k'$ and $l = l'$ (or $m = 0$), and we obtain

$$\langle \mathrm{Tr}(G^k \mathbf{1}\mathbf{1}^\dagger G^{l\dagger} \mathbf{1}\mathbf{1}^\dagger) \rangle = N^2 g^{k+l} = \mathrm{Tr}(\langle G \rangle^k \mathbf{1}\mathbf{1}^\dagger \langle G \rangle^{l\dagger} \mathbf{1}\mathbf{1}^\dagger). \tag{B.26}$$

Using the expressions above and the Taylor series expansion of infinitely differentiable functions, we calculate the following traces that are used in appendix A to compute the variance and covariance of the activity

$$\langle \mathrm{Tr}((I - G)^{-1} (I - G^\dagger)^{-1}) \rangle = \frac{N}{1 - \lambda^2}, \tag{B.27}$$

$$\langle \mathrm{Tr}((I - G)^{-1} \mathbf{1}\mathbf{1}^\dagger (I - G^\dagger)^{-1}) \rangle = \frac{N}{(1 - \lambda^2)(1 - g)^2}, \tag{B.28}$$

$$\langle \mathrm{Tr}((I - G)^{-1} \mathbf{1}\mathbf{1}^\dagger (I - G^\dagger)^{-1} \mathbf{1}\mathbf{1}^\dagger) \rangle = \frac{N^2}{(1 - g)^2}, \tag{B.29}$$

$$\int_0^\infty dt\ e^{-2t}\langle \mathrm{Tr}(e^{Gt}e^{G^\dagger t})\rangle = \frac{N}{2\sqrt{1-\lambda^2}}, \tag{B.30}$$

$$\int_0^\infty dt\ e^{-2t}\langle \mathrm{Tr}(e^{Gt}\mathbf{1}\mathbf{1}^\dagger e^{G^\dagger t})\rangle = \frac{N}{2\sqrt{1-\lambda^2}(1-g)}\left[\frac{1+\sqrt{1-\lambda^2}(1-g)}{1+\sqrt{1-\lambda^2}-g}\right], \tag{B.31}$$

$$\int_0^\infty dt\ e^{-2t}\langle \mathrm{Tr}(e^{Gt}\mathbf{1}\mathbf{1}^\dagger e^{G^\dagger t}\mathbf{1}\mathbf{1}^\dagger)\rangle = \frac{N^2}{2(1-g)}. \tag{B.32}$$

Note that the substitution $g \to -g\sqrt{K}$ must be applied to recover the scaling studied in the main text. If K is proportional to N, this substitution may change the order of magnitude of various terms in the summation considered above, possibly modifying the leading terms in each sum. Note that all series converge only for $|g| < 1$, but their sum can be evaluated at $g \to -g\sqrt{K}$ by analytical continuation. Then, approximating the sums by the leading terms described above is accurate under the assumption that all series involving lower-order terms converge to bounded functions of g.

We conclude this appendix by studying the case of nonhomogeneous mean and variance. We have assumed that the mean and variance are homogeneous: they take the same value for different matrix elements: $\langle G_{ij}\rangle = g/N$ and $\langle \Delta G_{ij}^2\rangle = \lambda^2/N$. However, the same methods could be used to analyze the more general case in which the mean and variance are inhomogeneous. In fact, as long as the mean and variances do not depend on N, they do not change the order of different terms in the sums considered above. Therefore, the calculation would consist of taking only the leading terms and recalculating their value according to the new matrices of means and variances. For example, even in the inhomogeneous case, the sums resulting in equations B.15, B.29, and B.32 would still be determined uniquely by the mean $\langle G_{ij}\rangle$, and the sums resulting in equations B.27 and B.30 would be still determined uniquely by the variance $\langle \Delta G_{ij}^2\rangle$. Sums affected by both the mean and variance, such as those resulting in equations B.28 and B.31, would still be calculated by using only the leading terms determined above.

A particularly simple case is when the mean and variance depend only on the presynaptic neuron: $\langle G_{ij}\rangle = g_j/N$ and $\langle \Delta G_{ij}^2\rangle = \lambda_j^2/N$. This includes the case of interconnected excitatory and inhibitory neurons, where g_j is positive for excitatory neurons and negative for inhibitory neurons. In that case, all results above still hold, with the simple substitutions:

$$g \longleftarrow N^{-1}\sum_{j=1}^{N} g_j, \tag{B.33}$$

$$\lambda^2 \longleftarrow N^{-1}\sum_{j=1}^{N} \lambda_j^2. \tag{B.34}$$

Namely, the parameters g and λ^2 now measure the mean connection strength and the mean variance across presynaptic neurons. Simulations suggest that a similar substitution, a mean of g and λ^2 across all matrix entries, works well even for general nonhomogeneous parameters.

Acknowledgment

This study was supported by the U.S. National Institutes of Health grant R01 MH062349 and the Swartz Foundation.

References

Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*, 91–101.

Amit, D. J., & Tsodyks, M. (1991). Quantitative study of attractor neural network retrieving at low spike rates I: Substrate—spikes, rates and neuronal gain. *Network*, *2*, 259–273.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.*, *7*, 358–366.

Averbeck, B. B., & Lee, D. (2003). Neural noise and movement-related codes in the macaque supplementary motor area. *J. Neurosci.*, *23*, 7630–7641.

Baddeley, R., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A., et al. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B*, *264*, 1775–1783.

Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J. Neurosci.*, *21*, 1676–1697.

Bernacchia, A., & Amit, D. J. (2007). Impact of spatiotemporally correlated images on the structure of memory. *Proc. Natl. Acad. Sci. USA*, *104*, 3544–3549.

Bernacchia, A., Seo, H., Lee, D., & Wang, X. J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neurosci.*, *14*, 366–372.

Cafaro, J., & Rieke, F. (2010). Noise correlations improve response fidelity and stimulus encoding. *Nature*, *468*, 964–967.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neurosci.*, *14*, 811–819.

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neurosci.*, *12*, 1594–1600.

Cohen, M. R., & Newsome, W. T. (2008). Context-dependent changes in functional circuitry in visual area MT. *Neuron*, *60*, 162–173.

Constantinidis, C., & Goldman-Rakic, P. S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J. Neurophysiol.*, *88*, 3487–3497.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature*, *448*, 802–806.

Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci.*, *4*, 739–751.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., & Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, *327*, 584–587.

Gardiner, C. W. (1985). *Handbook of stochastic methods: For physics, chemistry and natural sciences* (2nd ed.). New York: Springer.

Ginzburg, I., & Sompolinski, H. (1994). Theory of correlations in stochastic neural networks. *Phys Rev. E*, *50*, 3171–3191.

Graf, A. B. A., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neurosci.*, *14*, 239–245.

Gudowska-Nowak, E., Janik, R. A., Jurkiewicz, J., & Nowak, M. A. (2003). Infinite products of large random matrices and matrix-valued diffusion. *Nuclear Physics B*, *670*, 479–507.

Gutniski, D. A., & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature*, *452*, 220–224.

Hertz, J. (2010). Cross-correlations in high-conductance states of a model cortical network. *Neural Computation*, *22*, 427–447.

Hromadka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol*, *6*(1), e16.

Huang, X., & Lisberger, S. G. (2009). Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J. Neurophysiol.*, *101*, 3012–3030.

Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.*, *25*, 3661–3673.

Komiyama, T., Sato, T. R., O'Connor, D. H., Zhang, Y.-X., Huber, D., Hooks, B. M., et al. (2010). Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature*, *464*, 1182–1186.

Lampl, I., Reichova, I., & Ferster, D. (1999). Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron*, *22*, 361–374.

Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J. Neurosci.*, *18*, 1161–1170.

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

Maynard, E. M., Hatsopoulos, N. G., Ojakangas, C. L., Acuna, B. D., Sanes, J. N., Normann, R. A., et al. (1999). Neuronal interactions improve cortical population coding of movement direction. *J. Neurosci.*, *19*, 8083–8093.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2010). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron*, *63*, 879–888.

Nadal, J. P., & Parga, N. (1994). Non-linear neurons in the low-noise limit: A factorial code maximizes information transfer. *Network: Comp. Neu. Syst.*, *5*, 565–581.

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neurosci.*, *11*, 535–537.

Panzeri, S., Schultz, S. R., Treves, A., & Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proc. R. Soc. Lond. B*, *266*, 1001–1012.

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Comp. Biol.*, *7*, e1002059.

Poulet, J. F. A., & Petersen, C. C. (2008). Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature*, *454*, 881–885.

Rangan, A. V. (2009). Diagrammatic expansion of pulse-coupled network dynamics. *Phys. Rev. Lett.*, *102*, 158101.

Raz, A., Vaadia, E., & Bergman, H. (2000). Firing patterns and correlations of spontaneous discharge of pallidal neurons in the normal and the tremulous 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine vervet model of Parkinsonism. *J. Neurosci.*, *20*, 8559–8571.

Reich, D. S., Mechler, F., & Victor, J. D. (2001). Independent and redundant information in nearby cortical neurons. *Science*, *294*, 2566–2568.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science*, *327*, 587–590.

Romo, R., Hernandez, A., Zainos, A., & Salinas, E. (2003). Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, *38*, 649–657.

Salinas, E., & Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.*, *20*, 6193–6209.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.*, *18*, 3870–3896.

Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.*, *28*, 12591–12603.

Sompolinsky, H., Yoon, H., Kang, K., & Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E*, *64*, 051904.

Tetzlaff, T., Helias, M., Einevoll, G. T., & Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comp. Biol.*, *8*, e1002596.

Trousdale, J., Hu, Y., Shea-Brown, E., & Kresimir, J. (2012). Impact of network structure and cellular response on spike time correlations. *PLoS Comp. Biol.*, *8*, e1002408.

van Vreesvijk, C. A., & Sompolinski, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*, 1724–1726.

Wilke, S. D., & Eurich, C. W. (2002). On the functional role of noise correlations in the nervous system. *Neurocomputing*, *44-46*, 1023–1028.

Wilson, C. J., Beverlin II, B., & Netoff, T. (2011). Chaotic desynchronization as the therapeutic mechanism of deep brain stimulation. *Front. Syst. Neurosci.*, *5*, 50.

Zohary, E., Shadlen, M. N., & Newsome, W. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*, 140–143.

Received March 18, 2012; accepted January 9, 2013.