The correlated state in balanced neuronal networks

Cody Baker,[1] Christopher Ebsch,[1] Ilan Lampl,[2] and Robert Rosenbaum[1,3]

[1] *Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, USA*

[2] *Department of Neurobiology, Weizmann Institute of Science, Rehovot, 7610001, Israel*

[3] *Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, IN, USA*

(Dated: January 29, 2019)

Understanding the magnitude and structure of inter-neuronal correlations and their relationship to synaptic connectivity structure is an important and difficult problem in computational neuroscience. Early studies show that neuronal network models with excitatory-inhibitory balance naturally create very weak spike train correlations, defining the "asynchronous state." Later work showed that, under some connectivity structures, balanced networks can produce larger correlations between some neuron pairs, even when the average correlation is very small. All of these previous studies assume that the local network receives feedforward synaptic input from a population of uncorrelated spike trains. We show that when spike trains providing feedforward input are correlated, the downstream recurrent network produces much larger correlations. We provide an in-depth analysis of the resulting "correlated state" in balanced networks and show that, unlike the asynchronous state, it produces a tight excitatory-inhibitory balance consistent with in vivo cortical recordings.

I. INTRODUCTION

Correlations between the spiking activity of cortical neurons have important consequences for neural dynamics and coding [1–3]. A quantitative understanding of how spike train correlations are generated and shaped by the connectivity structure of neural circuits is made difficult by the noisy and nonlinear dynamics of recurrent neuronal network models [4–7]. Linear response and related techniques have been developed to overcome some of these difficulties [7–15], but their application to networks of integrate-and-fire neurons models typically relies on a diffusion approximation that requires an assumption of sparse and/or weak connectivity and an assumption that neurons receive uncorrelated, feedforward Gaussian white noise input. However, cortical circuits are densely connected and receive spatially and temporally correlated synaptic input from outside the local circuit [16–19].

An alternative approach to analyzing correlated variability in recurrent neuronal network models is motivated in part by the widely observed balance between excitatory and inhibitory synaptic inputs in cortex [20–27]. When synaptic weights are scaled like $1/\sqrt{N}$ where N is the size of a model network, balance between excitation and inhibition arises naturally at large network size, which defines the "balanced state" [28, 29]. A similar scaling of synaptic weights has since been observed in cultured cortical populations [30].

Early work on balanced networks assumed sparse connectivity to produce weak spike train correlations, but it was later shown that keeping connection probabilities $\mathcal{O}(1)$ naturally produces weak, $\mathcal{O}(1/N)$, spike train correlations, defining the "asynchronous state" [31]. While these extremely weak spike train correlations are consistent with some cortical recordings [32], the magnitude of correlations in cortex can depend on stimulus, cortical area, layer, and behavioral or cognitive state, and can be much larger than predicted by the asynchronous state [6, 33–37]. This raises the question of how larger correlation magnitudes can arise in balanced cortical circuits. Later theoretical work showed that larger correlations can be obtained between some cell pairs in densely connected networks with specially constructed connectivity structure [38–42], offering a potential explanation of the larger correlations often observed in recordings.

All of these previous studies of correlated variability in balanced networks assume that the recurrent network receives feedforward synaptic input from an external population of uncorrelated spike trains, so feedforward input correlations arise solely from overlapping feedforward synaptic projections. In reality, feedforward synaptic input to a cortical population comes from thalamic projections, other cortical areas, or other cortical layers in which spike trains could be correlated.

We extend the theory of densely connected balanced networks to account for correlations between the spike trains of neurons in an external, feedforward input layer. We show that correlations between the feedforward synaptic input to neurons in the recurrent network are $\mathcal{O}(N)$ in this model, but cancel with $\mathcal{O}(N)$ correlations between recurrent synaptic input to produce $\mathcal{O}(1)$ spike train correlations on average, defining what we refer to as the "correlated state" in densely connected balanced networks. This correlated state offers an alternative explanation for the presence of moderately large spike train correlations in cortical recordings. We derive a simple, closed form approximation for the average covariance between neurons' spike trains in the correlated state in term of synaptic parameters alone, without requiring the use of linear response theory or any other knowledge of neurons' transfer functions. We show that the tracking of excitatory synaptic input currents by inhibitory currents is more precise and more similar to in vivo recordings [23] in the correlated state than in the asynchronous state.

Our results extend the theory of correlated variability in balanced networks to the biologically realistic assumption that presynaptic neural populations are themselves correlated.

II. MODEL AND BACKGROUND

We consider recurrent networks of N model neurons, N_e of which are excitatory and N_i inhibitory. Neurons are randomly and recurrently interconnected and also receive random feedforward synaptic input from an external population of N_x neurons whose spike trains are homogeneous Poisson processes with rate r_x (Fig. 1A).

The membrane potential of neuron j in population $a = e, i$ obeys the exponential integrate-and-fire (EIF) dynamics

$$C_m \frac{dV_j^a}{dt} = -g_L(V_j^a - E_L) + D_T e^{(V_j^a - V_T)/D_T} + T_j^a(t)$$

with the added condition that each time $V_j^a(t)$ exceeds V_{th}, it is reset to V_{re} and a spike is recorded. Our results do not depend sensitively on the exact neuron model used or the values of neuron parameters. We additionally set a lower bound on the membrane potential at $V_{lb} = -100$mV. Spike trains are represented as a sum of Dirac delta functions,

$$S_j^a(t) = \sum_n \delta(t - t_n^{a,j}),$$

where $t_n^{a,j}$ is the nth spike time of neuron j in population $a = e, i, x$. The total synaptic input current to neuron j in population $a = e, i$ is decomposed as

$$T_j^a(t) = E_j^a(t) + I_j^a(t) + X_j^a(t)$$

where

$$U_j^a(t) = \sum_{k=1}^{N_b} J_{jk}^{ab}(\eta_b * S_j^b)(t) \tag{1}$$

for $U = E, I, X$ and $b = e, i, x$ respectively where $*$ denotes convolution, J_{jk}^{ab} is the synaptic weight from neuron k in population b to neuron j in population a, and $\eta_b(t)$ is a postsynaptic current (PSC) waveform. Without loss of generality, we assume that $\int \eta_b(t) = 1$. We use $\eta_b(t) = \tau_b^{-1} e^{-t/\tau_b} H(t)$ where $H(t)$ is the Heaviside step function, though our results do not depend sensitively on the precise neuron model or PSC kernel used. For calculations, it is useful to decompose the total synaptic input into its recurrent and external sources,

$$T_j^a(t) = R_j^a(t) + X_j^a(t)$$

where

$$R_j^a(t) = E_j^a(t) + I_j^a(t)$$

is the recurrent synaptic input from the local circuit.

Local cortical circuits contain a large number of neurons and individual cortical neurons receive synaptic input from thousands of other neurons within their local circuit and from other layers or areas. Densely connected balanced networks have been proposed to model such large and densely interconnected neuronal networks [31, 40]. In such models, one considers the limit of large N (with N_x, N_e and N_i scaled proportionally) with fixed connection probabilities and where synaptic weights are scaled like $\mathcal{O}(1/\sqrt{N})$ [29, 31]. This scaling naturally captures the balance of mean excitatory and mean inhibitory synaptic input, as well as the tracking of excitation by inhibition, observed in cortical recordings [31]. Recent work in cultured cortical populations shows that similar scaling laws emerge naturally and produce network dynamics consistent with the balanced state [30]. In particular, we consider a random connectivity structure in which

$$J_{jk}^{ab} = \frac{1}{\sqrt{N}} \begin{cases} j_{ab} & \text{with probability } p_{ab} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where connections are statistically independent and $j_{ab}, p_{ab} \sim \mathcal{O}(1)$ for $b = e, i, x$ and $a = e, i$. We furthermore define the proportions

$$q_b = \frac{N_b}{N}$$

which are assumed $\mathcal{O}(1)$. For all examples we consider, $q_e = 0.8$ and $q_i = q_x = 0.2$.

We next introduce notational conventions for quantifying the statistics of spike trains and synaptic inputs in the network. The mean firing rates of neurons in population $a = e, i, x$ is defined by r_a for $a = e, i, x$ and it is useful to define the 2×1 vector, $\boldsymbol{r} = [r_e \;\; r_i]^T$ where $\cdot^T$ denotes the transpose. The mean is technically interpreted as the expectation over realizations of the network connectivity, but for large N it is approximately equal to the sample mean over all neurons the network. Similarly, mean-field synaptic inputs to neurons in populations $a = e, i$ are defined by

$$\overline{U}_a = \text{mean}_j U_j^a(t)$$

for $U = E, I, X, R, T$ and, in vector form, $\overline{\boldsymbol{U}} = [U_e \;\; U_i]^T$

For quantifying correlated variability, we use the cross-spectral density (CSD)

$$\langle U_j^a, Z_k^b \rangle(f) = \int_{-\infty}^{\infty} C_{U_j^a Z_k^b}(\tau) e^{-2\pi i f \tau} d\tau$$

between $U_j^a(t)$ and $Z_k^b(t)$ for $U, Z = E, I, X, S, R, T$ and $a, b = e, i, x$ where

$$C_{U_j^a, Z_k^b}(\tau) = \text{cov}(U_j^a(t), Z_k^b(t + \tau))$$

is the cross-covariance function. The argument, f, is the frequency at which the CSD is evaluated. The CSD is

a convenient measure of correlated variability because it simplifies mathematical calculations due to the fact that $\langle\cdot,\cdot\rangle$ is a Hermitian operator and because most commonly used measures of correlated variability can be written as a function of the CSD. For example, the cross-covariance is the inverse Fourier transform of the CSD. Spike count covariances over large time windows can be written in terms of the CSD by first noting that the spike count is an integral of the spike train [4],

$$\text{spike count over } [0,t_0] = \int_0^{t_0} S_j^a(t)dt.$$

For large t_0, the covariance between two integrals is related to the zero-frequency CSD,

$$\lim_{t_0\to\infty}\frac{1}{t_0}\text{cov}\left(\int_0^{t_0} U_j^a(t)dt, \int_0^{t_0} Z_k^b(t)dt\right) = \langle U_j^a, Z_k^b\rangle(f=0). \tag{3}$$

Hence, the spike count covariance over the interval $[0,t_0]$ is given asymptotically by $t_0\langle S_j^a, S_k^b\rangle(f=0)$ for large t_0. Following this result, we quantify covariability between spike trains and between synaptic currents using the zero-frequency CSD, which we estimate by taking the covariance between integrals as in Eq. (3) using $t_0 = 250$ms. This provides a simple, easily estimated quantity for quantifying covariance.

Most of our computations are performed at the level of population averages, so we define

$$\langle U_a, Z_b\rangle = \text{mean}_{j\neq k}\langle U_j^a, Z_k^b\rangle. \tag{4}$$

which is a scalar function of frequency, f, for each $a,b = e,i,x$ and $U,Z = E,I,X,S,R,T$. It is also convenient to define the 2×2 mean-field matrix form,

$$\langle \boldsymbol{U},\boldsymbol{Z}\rangle = \begin{bmatrix} \langle U_e, Z_e\rangle & \langle U_e, Z_i\rangle \\ \langle U_i, Z_e\rangle & \langle U_i, Z_i\rangle \end{bmatrix}.$$

for $\boldsymbol{U},\boldsymbol{Z} = \boldsymbol{E},\boldsymbol{I},\boldsymbol{X},\boldsymbol{S},\boldsymbol{R},\boldsymbol{T}$. We also define the recurrent and feedforward mean-field connectivity matrices,

$$W = \begin{bmatrix} w_{ee} & w_{ei} \\ w_{ie} & w_{ii} \end{bmatrix} \text{ and } W_x = \begin{bmatrix} w_{ex} \\ w_{ix} \end{bmatrix}$$

where $w_{ab}(f) = p_{ab}j_{ab}q_b\widetilde{\eta}_b(f) \sim \mathcal{O}(1)$ with $\widetilde{\eta}_b(f)$ the Fourier transform of $\eta_b(t)$. For the exponential kernels we use, $\widetilde{\eta}_b(f) = 1/(1+2\pi i f\tau_b)$. The zero-frequency values, $\overline{w}_{ab} = w_{ab}(0) = p_{ab}j_{ab}q_b$, define time-averaged interactions and mean-field connection matrices, $\overline{W} = W(0)$ and $\overline{W}_x = W_x(0)$.

This choice of notation allows us to perform computations on mean-field spike train and input statistics in a mathematically compact way. We first review the mean-field analysis of firing rates in the balanced state [28, 29, 43–45]. Mean external input is given by $\overline{\boldsymbol{X}} = \sqrt{N}\,\overline{W}_x r_x$ and mean recurrent input by $\overline{\boldsymbol{R}} = \sqrt{N}\,\overline{W}\boldsymbol{r}$ so that mean total synaptic input is given by

$$\overline{\boldsymbol{T}} = \sqrt{N}\left[\overline{W}\boldsymbol{r} + \overline{W}_x r_x\right].$$

In the balanced state, $\overline{\boldsymbol{T}} \sim \boldsymbol{r} \sim \mathcal{O}(1)$, which can only be obtained by a cancellation between external and recurrent synaptic inputs. This cancellation requires $\overline{W}\boldsymbol{r} + \overline{W}_x r_x \sim \mathcal{O}(1/\sqrt{N})$ so that [28, 29, 43–45]

$$\lim_{N\to\infty}\boldsymbol{r} = -\overline{W}^{-1}\overline{W}_x r_x \tag{5}$$

in the balanced state. Hence, the balanced state can only be realized when this solution has positive entries, $r_e, r_i > 0$, which requires that [28, 29, 43] $\overline{X}_e/\overline{X}_i > w_{ei}/w_{ii} > w_{ee}/w_{ie}$.

Note that the derivation of Eq. (5) relied on an assumption that $\overline{\boldsymbol{T}} \sim \boldsymbol{r}$, *i.e.*, that the transfer of mean input to firing rates is $\mathcal{O}(1)$. The analysis of correlations in balanced networks requires similar assumptions. Specifically, our analysis of correlations relies on an assumption that transfer of mean-field covariance is $\mathcal{O}(1)$, specifically that $\langle T_a, U_b\rangle \sim \langle S_a, U_b\rangle$ or

$$\frac{\langle T_a, U_b\rangle}{\langle S_a, U_b\rangle} \sim \mathcal{O}(1) \tag{6}$$

for $U = X, S$ as $N\to\infty$. A more precise and slightly weaker assumption that is sufficient for our analysis is that

$$\lim_{N\to\infty}\frac{1}{\sqrt{N}}\frac{\langle T_a, U_b\rangle}{\langle S_a, U_b\rangle} = 0 \tag{7}$$

for $U = X, S$, which is implied by Eq. (6). Note that this does not imply that $\langle S_a, U_b\rangle$ cannot be much larger or smaller than $\langle T_a, U_b\rangle$, but only that their ratio does not converge to zero or diverge to ∞ as $N\to\infty$. The validity of this assumption is discussed in more detail in Appendix A and the end of Section III.

Importantly, we do not need to know the value of the fraction in Eq. (6), the fraction does not need to converge to a limit as $N\to\infty$, and $\langle S_a, U_b\rangle$ need not be linearly related to $\langle T_a, U_b\rangle$, which contrasts to assumptions made by linear response theory.

III. A REVIEW OF THE ASYNCHRONOUS BALANCED STATE

We first review previous work on correlated variability in balanced networks when spike trains in the external population are uncorrelated Poisson processes (Fig. 1A,B),

$$\langle S_x, S_x\rangle = 0.$$

Since spike trains in the external population are uncorrelated, correlations between the external input to neurons in the recurrent network arise solely from overlapping

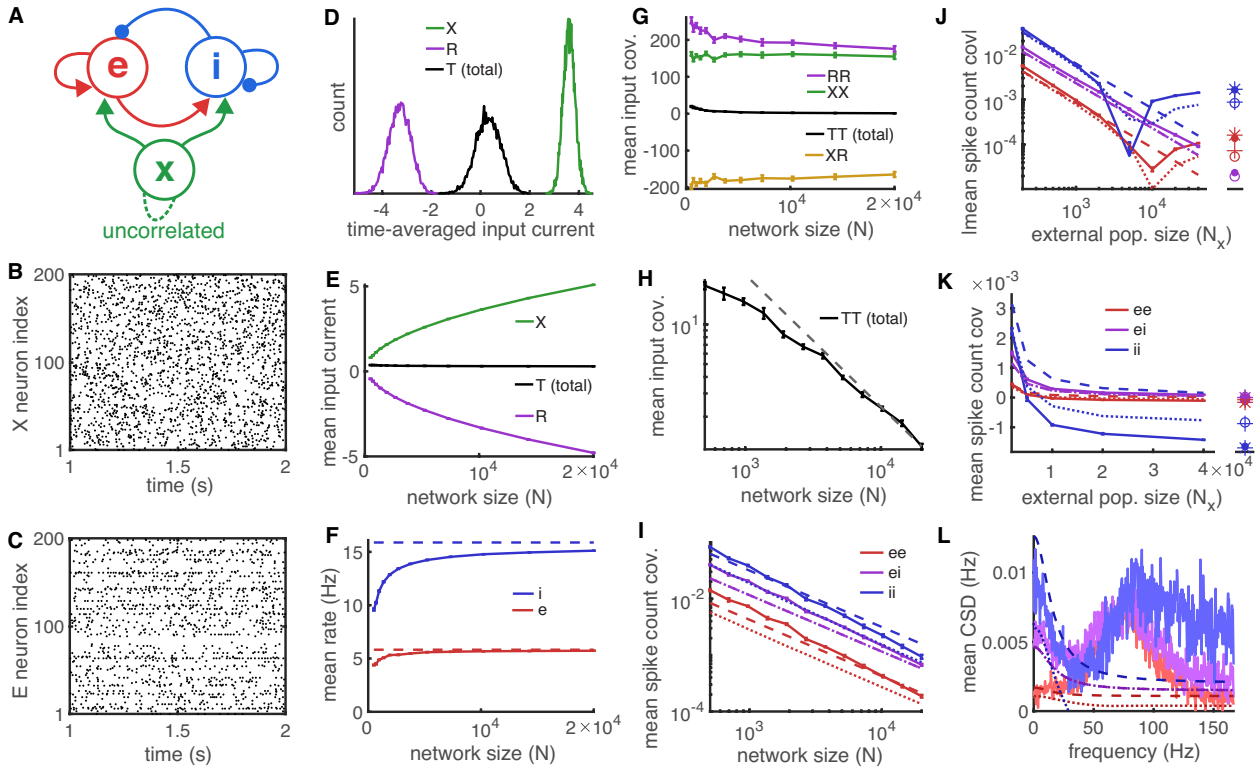

FIG. 1. **The asynchronous state in densely connected balanced networks. A)** Network diagram. An external population, X, of uncorrelated Poisson processes provides feedforward input to a randomly connected recurrent network of excitatory, E, and inhibitory, I, neurons. Feedforward input correlations are solely from overlapping projections from X. **B,C)** Raster plot of 200 randomly selected neurons from population X and E respectively in a network with $N = 10^4$ and $N_x = 2000$. **D)** Histogram of time-averaged external (X, green) recurrent ($R = E + I$, purple) and total ($T = X + E + I$, black) input to all excitatory neurons in a network with $N = 10^4$ and $N_x = 2000$. Currents here and elsewhere are reported in units $C_m V/s$ where C_m is the arbitrary membrane capacitance. **E)** Mean external (green), recurrent (purple), and total (black) input to excitatory neurons for networks of different sizes, N. **F)** Mean excitatory (red) and inhibitory (blue) neuron firing rates for different network sizes. Solid curves are from simulations and dashed curves are from Eq. (5). **G)** Mean covariance between pairs of excitatory neurons' external inputs (green), recurrent inputs (purple), total inputs (black), and mean covariance between the recurrent input to one excitatory neuron and external input to the other (yellow) for different network sizes. Covariances were computed by integrating the inputs over 250ms windows then computing covariances between the integrals, which is proportional to zero-frequency CSD (see Eq. (3) and surrounding discussion). Integrated currents have units $C_m mV$, so their covariances have units $C_m^2 mV^2$. **H)** Zoomed in view of black curve from E on a log-log axis (black) and the function C/N (dashed gray) where C was chosen so that the two curves match at $N = 2 \times 10^4$. **I)** Mean spike count covariance between excitatory neuron spike trains (red), between inhibitory neuron spike trains (blue), and between excitatory-inhibitory pairs of spike trains (purple). In panels I-L, solid curves are from simulations, dashed curves are from the first term in Eq. (9), and dotted curves are from the first two terms. For the dotted curves, the $\{S_e, S_e\}$ and $\{S_i, S_i\}$ terms in Eq. (9) were estimated empirically from simulations. In all panels, counts were computed over 250ms time windows. **J)** Absolute value of mean spike count covariances as a function of the external population size, N_x, when $N = 10^4$ and where feedforward connection probabilities were scaled to keep $p_{ex}N_x = p_{ix}N_x = 200$ fixed as N_x was changed. Filled circles are from simulations with uncorrelated external input (representing $N_x \to \infty$) and open circles are from simulations with deterministic, time-constant external input (external input covariance is zero in both cases). Asterisks and crosses are from evaluating the second term of Eq. (9) using the values of $\{S_e, S_e\}$ and $\{S_i, S_i\}$ estimated from the simulations marked with filled circle and crosses respectively. **K)** Same data as J, but plotted on a linear axis without taking absolute value, and zoomed into to larger values of N_x. **L)** Mean cross-spectral densities between neurons in the recurrent network using the parameters from panels B-D. Solid curves are from simulations, dashed are from the first term of Eq. (9), and dotted are from the first two terms of Eq. (9) using empirically estimated power spectral densities for $\{S_b, S_b\}(f)$. Synaptic time constants were $\tau_e = 8$, $\tau_i = 4$, and $\tau_x = 10$ms in all simulations. In all simulations except those in panels J and K, recurrent and feedforward connection probabilities are all $p_{ab} = 0.1$ for $a = e, i$ and $b = e, i, x$.

feedforward synaptic projections with [31, 40, 46] (see Eq. (A1) in Appendix A for a derivation)

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle = q_x^{-1} W_x r_x W_x^* \sim \mathcal{O}(1). \tag{8}$$

where W_x^* is the conjugate transpose of W_x.

It would at first seem that this $\mathcal{O}(1)$ external input correlation would lead to $\mathcal{O}(1)$ correlations between neurons' spike trains. In the asynchronous state, this is prevented by a cancellation between positive and negative sources of input correlation. In particular, correlations between neurons' recurrent synaptic inputs, $\langle \boldsymbol{R}, \boldsymbol{R} \rangle$, are also positive and $\mathcal{O}(1)$, but these positive sources of input correlations are canceled by negative correlations between neurons' recurrent and external inputs, $\langle \boldsymbol{X}, \boldsymbol{R} \rangle$, in such a way that the total synaptic input correlation is weak,

$$\langle \boldsymbol{T}, \boldsymbol{T} \rangle = \langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{R} \rangle + \langle \boldsymbol{R}, \boldsymbol{X} \rangle + \langle \boldsymbol{R}, \boldsymbol{R} \rangle \sim \mathcal{O}(1/N)$$

where $\langle \boldsymbol{R}, \boldsymbol{X} \rangle = \langle \boldsymbol{X}, \boldsymbol{R} \rangle^*$. In Appendix A, we review a derivation of spike train CSDs in the asynchronous state that gives

$$\begin{aligned} \langle \boldsymbol{S}, \boldsymbol{S} \rangle =& \frac{1}{N} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle W^{-*} - \frac{1}{N} \begin{bmatrix} \frac{\{S_e, S_e\}}{q_e} & 0 \\ 0 & \frac{\{S_i, S_i\}}{q_i} \end{bmatrix} \\ &+ \mathcal{O}\left(N^{-\frac{3}{2}}\right) \end{aligned} \tag{9}$$

where $\{S_b, S_b\}(f)$ is the average power spectral density of spike trains in population $b = e, i$.

Eq. (9) has been previously derived for the integrate-and-fire network models considered here (see equation S.22 in the supplement to [40]) and equivalent expressions have been derived for networks of binary neuron models (see equations 38-39 in the supplement to [31], equation 38 in [46], and equation 23 in [41]) as well as similar equations for similar models in other work [5, 14, 41, 47, 48]. In [46], the $\mathcal{O}(N^{-3/2})$ term is also derived for binary networks. These derivations have also been extended to spatially extended networks with distance-dependent connection probabilities and to networks with several sub-populations [40, 41]. The overall finding in this previous work is that spike train correlations in the recurrent network are $\mathcal{O}(1/N)$. Some exceptions to this $\mathcal{O}(1/N)$ scaling have been demonstrated in spatially extended networks with distance-dependent connection probabilities and in networks with singular mean-field connectivity matrices [39–42, 46], a topic to which we return in Section VI.

To demonstrate these results, we first simulated a network of $N = 10^4$ randomly and recurrently connected neurons receiving feedforward input from a population of $N_x = 2000$ uncorrelated Poisson-spiking neurons (Fig. 1A,B). As predicted, spiking activity in the recurrent network was asynchronous and irregular (Fig. 1C; mean spike count correlation between neurons with rates at least 1 Hz was 5.2×10^{-4}) with approximate balance between external (X) and recurrent (R) synaptic input sources (Fig. 1D). Varying the network size, N, demonstrates the $\mathcal{O}(\sqrt{N})$ growth of mean external ($\overline{\boldsymbol{X}}$) and recurrent ($\overline{\boldsymbol{R}}$) synaptic input currents that cancel to produce $\mathcal{O}(1)$ mean total input current ($\overline{\boldsymbol{T}}$) (Fig. 1E), as predicted by the mean-field theory of balance (Eq. (5) and surrounding discussion). As a result, firing rates converge toward the limiting values predicted by Eq. (5) (Fig. 1F).

As predicted by the analysis of the asynchronous state, the mean covariances between individual sources of synaptic inputs appear $\mathcal{O}(1)$ (Fig. 1G), but cancel to produce total input covariance and spike count covariance that vanish as network size, N, increases (Fig. 1G–I).

Applying the approximation in Eq. (9) requires knowledge of neurons' mean-field power spectral densities, $\{S_e, S_e\}(f)$ and $\{S_i, S_i\}(f)$. While numerical methods for approximating power spectral densities in networks of integrate-and-fire neurons have been developed using Fokker-Planck techniques [8, 9, 11, 12, 49], these approaches require a diffusion approximation that is not justified for our model (see Discussion). Instead, we considered two ways to test Eq. (9) that do not require a diffusion approximation or Fokker-Planck techniques.

First, we noted that the first term, $(1/N)W^{-1}\langle \boldsymbol{X}, \boldsymbol{X} \rangle W^{-*}$, in Eq. (9) only involves known parameters and does not require knowledge of $\{S_e, S_e\}$ or $\{S_i, S_i\}$. Therefore, we were able to directly compare the exact value of this term to spike count covariances computed from simulations, demonstrating a relatively close match (compare solid and dashed blue in Fig. 1I).

Next, we noted that the contribution of the second term in Eq. (9) can be estimated empirically from simulations by computing the average spike count variance in the recurrent network (since $\{S_b, S_b\}(0)$ is proportional to spike count variance over large window sizes). This yields a semi-analytical approach to testing the accuracy of Eq. (9) with only the $\mathcal{O}(N^{-3/2})$ term ignored. This approach improved the approximation to mean inhibitory-inhibitory spike count covariances at large N (compare solid and dotted in Fig. 1I), but left some error even at $N = 2 \times 10^4$.

Previous work showed that intrinsically generated correlations can dominate in large balanced networks [46], but in our simulations external input correlations captured by the first term in Eq. (9) dominated over a range of N values. We conclude that the relative contribution of the three terms in Eq. (9) can depend on model details and parameters. This is clarified by noting that a sufficiently large reduction in the external input covariance, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$, would necessarily prevent the first term in Eq. (9) from dominating because the first term is proportional to $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$, but the second term should remain non-zero even for vanishing $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$. At the extreme, when $\langle \boldsymbol{X}, \boldsymbol{X} \rangle = 0$, the first term is zero, but the second term is non-zero for such networks due to intrinsically generated variability [46].

To demonstrate these ideas, we followed a procedure from [46] by varying N_x while scaling feedforward connection probabilities, p_{ex} and p_{ix}, so as to keep the average

number of external inputs, $p_{ex}N_x = p_{ix}N_x = 200$, fixed. All other parameters, including the size, N, of the recurrent network were kept fixed. In this scenario, an increase in N_x causes a decrease in $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ because it leads to fewer overlapping external inputs (as evidenced by the appearance of $q_x^{-1} = N/N_x$ in Eq. (8); but note that W_x is fixed as we change N_x). In addition, we simulated a version of the network in which each neuron received uncorrelated input from a private external population so that external input had the same univariate statistics as our previous simulations, but these external inputs were uncorrelated, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle = 0$. This represents the $N_x \to \infty$ limit of the previous simulations. Finally, we simulated a deterministic version of the network in which external inputs were replaced by time-constant input, $X_j^a(t) = \overline{X}_a$, which also implies $\langle \boldsymbol{X}, \boldsymbol{X} \rangle = 0$ with mean external input unchanged.

Examining these simulations, the magnitude of spike train covariance is non-monotonic with increasing N_x (Fig. 1J) because it changes sign and becomes increasingly negative at larger values of N_x (Fig. 1K). The first term in Eq. (9) dominates the other two terms (the dashed, solid, and dotted curves are close) for smaller values of N_x because external input covariance is larger in this regime. However, at larger values of N_x, the first term no longer dominates (Fig. 1K, inset, the dashed curve is far from the solid curve) because external input covariance is small. Moreover, even accounting for the second term in Eq. (9) produces a somewhat inaccurate approximation when external input covariance is small (Fig. 1J,K, compare dotted to solid at larger N_x), consistent with previous findings that the $\mathcal{O}(N^{-3/2})$ term can contribute significantly at large but finite N [46].

When external input is noisy but independent, simulations produce weak, but non-zero spike train correlations (Fig. 1J,K, filled circles) representing the $N_x \to \infty$ limit of the previous simulations [46]. In this regime, the first term in Eq. (9) is zero, but the second term produces a relatively accurate approximation to excitatory-excitatory and inhibitory-inhibitory spike count covariances (Fig. 1J,K, compare filled circles to asterisks), but note that the second term in Eq. (9) is zero for excitatory-inhibitory pairs. Similar results are observed when external input is deterministic and constant in time ($X_j^a(t) = \overline{X}_a$; Fig. 1J,K, compare open circles to crosses), but spike count variances (and therefore the $\{S_b, S_b\}(0)$ terms) are smaller when input is deterministic. Repeating the simulation from Fig. 1I,J with different connectivity parameters showed similar overall results (see Appendix B and Fig. 7A,B). See [5, 8–12, 14, 41, 43, 46, 47] for other studies of correlated variability in networks with uncorrelated external input.

So far we have focused on spike count covariance over large window sizes, which is proportional to low-frequency CSD. We next computed the full spike train CSDs from simulations with $N = 10^4$ and $N_x = 2000$. The first two terms in Eq. (9) accurately capture most of the low-frequency CSD for these parameters (Fig. 1I–L), but simulations show a high-frequency peak in the mean CSD that is not captured by these terms (Fig. 1L).

To understand why Eq. (9) becomes inaccurate at higher frequencies for the network considered here, recall that the derivation of Eq. (9) relies on the assumptions made by Eqs. (6) and (7), which posit that neural transfer is $\mathcal{O}(1)$. However, for EIF and many other neuron models, sufficiently high-frequency input fluctuations cannot be reliably be transferred due to the filtering imposed by membrane potential dynamics. More specifically,

$$\lim_{f\to\infty} \frac{\langle T_a, U_b \rangle(f)}{\langle S_a, U_b \rangle(f)} = \infty.$$

Hence, while Eqs. (6) and (7) might be true for any fixed f, the $\mathcal{O}(1)$ term on the right side of Eq. (6) diverges with f. Analogously, when f is large, larger values of N must be considered before the ratio in Eq. (7) is close to zero. Mathematically speaking, the limit in Eq. (7) convergeces pointwise but not uniformly in f. This caveat is technically inconsequential for any fixed f in the $N \to \infty$ limit analyzed in the derivation of Eq. (9), but all networks are finite. For any finite-sized network, the derivation of Eq. (9) is not justified at sufficiently high frequencies. In Appendix A, we show that for finite N our derivations are accurate at frequencies for which

$$\left| \frac{\langle \boldsymbol{T}, \boldsymbol{U} \rangle}{W \langle \boldsymbol{S}, \boldsymbol{U} \rangle} \right| \ll \sqrt{N} \tag{10}$$

for $\boldsymbol{U} = \boldsymbol{X}, \boldsymbol{S}$ where the division is applied element-wise to the matrices. As noted above, for any fixed f, this condition is satisfied in the $N \to \infty$ limit, but for any fixed N the condition is violated at sufficiently large f.

The left-hand-side of Eq. (10) is difficult to compute for networks of integrate-and-fire neurons because we do not know the values of $\langle \boldsymbol{T}, \boldsymbol{U} \rangle$ or $\langle \boldsymbol{S}, \boldsymbol{U} \rangle$, but we next show that linear response theory can provide a useful rough approximation.

In the context of the model considered here, linear response theory is defined by making the approximation [6, 8, 50–55]

$$\langle \boldsymbol{S}, \boldsymbol{U} \rangle \approx A \langle \boldsymbol{T}, \boldsymbol{U} \rangle \tag{11}$$

for $U = S, T, X$ where

$$A(f) = \begin{bmatrix} A_e(f) & 0 \\ 0 & A_i(f) \end{bmatrix}$$

and $A_a(f)$ is the average "susceptibility function" of neurons in population $a = e, i$. Eq. (11) provides an accurate approximation of correlation transfer whenever correlations in the network are not too strong.

Combining Eq. (10) with Eq. (11) gives the condition

$$|WA| \gg 1/\sqrt{N}$$

which can be written element-wise in terms of scalars as

$$|w_{ab}(f)|^2 |A_b(f)|^2 \gg 1/N$$

for $a, b \in \{e, i\}$. For the exponentially decaying synapses considered here, we have $|w_{ab}(f)|^2 = \overline{w}_{ab}^2/(1 + 4\pi^2 f^2 \tau_b^2)$ where $\overline{w}_{ab} = j_{ab} p_{ab} q_b$ and τ_b is the synaptic time constant. However, we do not know the susceptibility functions, $A_b(f)$. Much like the power spectral densities, $\{S_b, S_b\}(f)$, the susceptibility functions, $A_b(f)$, in networks of integrate-and-fire neurons are difficult to compute without a diffusion approximation that is not justified for our model (see Discussion). Due to the linearity of sub-threshold membrane dynamics, however, EIF neurons in a fluctuation dominated regime have susceptibility functions that are approximated by $|A_b(f)|^2 \approx g^2/(4\pi^2 f^2 \tau_m^2)$ at moderately large f [52, 56, 57] where $\tau_m = C_m/g_L$ is the membrane time constant and $g = A_b(0)$ is the gain of the neuron, *i.e.*, the derivative of the neuron's f-I curve. Putting this all together gives the condition

$$4\pi^2 f^2 \tau_b \tau_m \ll g|\overline{w}_{ab}|\sqrt{N}$$

so we can expect that Eq. (9) is accurate for frequencies sufficiently lower than

$$f_0 = \frac{\sqrt{g|\overline{w}_{ab}|}}{2\pi\sqrt{\tau_b \tau_m}} N^{1/4}. \tag{12}$$

To obtain a conservative estimate that will be valid for all combinations of $a, b \in \{e, i\}$, we chose the smallest value of $\overline{w}_{ab}$ used in Fig. 2L, which is $\overline{w}_{ee} = 2C_m$mV/ms (where C_m is the arbitrary membrane capacitance), and the largest value of τ_b, which is $\tau_e = 8$ms. The membrane time constant for the neurons used in our simulations is $\tau_m = 15$ms and $N = 10^4$ in Fig. 1L. The only quantity missing from Eq. (12) is the gain, g. To obtain a rough estimate of g, we fit a rectified quadratic function to the relationship between all neurons' firing rates and mean total inputs (r_j and $\overline{T}_j$), then computed the derivative of the fitted quadratic at the mean input value. The same approach was used in previous work [40, 58] to estimate a mean-field gain. In doing so, we obtained a gain approximation of $g = 0.014(C_m mV)^{-1}$. Combining these values gives an approximate cutoff frequency of $f_0 = 24$Hz. Indeed, Fig. 1L shows that Eq. (9) starts to become inaccurate somewhere just below this value of f, particularly for excitatory-excitatory (red) and excitatory-inhibitory (purple) neuron pairs.

In summary, when spike trains in the external population are uncorrelated, the first two terms in Eq. (9) give an accurate approximation to spike count covariances over long time windows (equivalently, low-frequency CSDs), but the second term can be difficult to compute directly and the equation loses accuracy when evaluating the CSD at higher frequencies. We next extend these results to networks in which spike trains in the external population are correlated.

IV. THE CORRELATED STATE IN BALANCED NETWORKS

Above, we reviewed the asynchronous state in which uncorrelated spike trains in the external layer, $\langle S_x, S_x \rangle = 0$, produce moderate external input covariance, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(1)$, and weak spike train correlations, $\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \mathcal{O}(1/N)$. We next show that moderate correlations between spike trains in the external population (Fig. 2A),

$$\langle S_x, S_x \rangle \sim \mathcal{O}(1),$$

leads to large covariance between neurons' external inputs, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(N)$, and moderate correlations between spike trains in the recurrent network, $\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \mathcal{O}(1)$.

We outline the derivation of correlations in such networks here and give a more detailed derivation in Appendix A. In addition to the assumption made by Eqs. (6) and (7), all of our derivations follow from a few simple arithmetical rules that rely on the bilinearity of the operator $\langle \cdot, \cdot \rangle$. Specifically,

$$\langle \boldsymbol{T}, \boldsymbol{U} \rangle = \langle \boldsymbol{R}, \boldsymbol{U} \rangle + \langle \boldsymbol{X}, \boldsymbol{U} \rangle \tag{13}$$

$$\langle \boldsymbol{R}, \boldsymbol{U} \rangle = \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{U} \rangle \tag{14}$$

$$\langle \boldsymbol{X}, \boldsymbol{U} \rangle = \sqrt{N} W_x \langle S_x, \boldsymbol{U} \rangle \tag{15}$$

$$\langle \boldsymbol{U}, \boldsymbol{Z} \rangle = \langle \boldsymbol{Z}, \boldsymbol{U} \rangle^* \tag{16}$$

for any $\boldsymbol{U}, \boldsymbol{Z} = \boldsymbol{E}, \boldsymbol{I}, \boldsymbol{X}, \boldsymbol{R}, \boldsymbol{S}, S_x, \boldsymbol{T}$ where $\boldsymbol{A}^*$ is the conjugate-transpose of $\boldsymbol{A}$ and where we omit smaller order terms here and below (the derivations in Appendix A keep track of these terms). Eq. (13) follows from the fact that total input is composed of recurrent and external sources, $\boldsymbol{T} = \boldsymbol{R} + \boldsymbol{X}$. Eqs. (14) and (15) follow from the fact that recurrent and external inputs are composed of linear combinations of $\mathcal{O}(N)$ spike trains, *c.f.* Eq. (1), and that synaptic weights are $\mathcal{O}(1/\sqrt{N})$. Eq. (16) is simply a property of the Hermitian cross-spectral operator.

We first derive the CSD between external inputs to neurons in the recurrent network. Applying Eqs. (15) and (16) gives

$$\begin{aligned} \langle \boldsymbol{X}, \boldsymbol{X} \rangle &= \sqrt{N} W_x \langle S_x, \boldsymbol{X} \rangle \\ &= N W_x \langle S_x, S_x \rangle W_x^* \\ &\sim \mathcal{O}(N). \end{aligned}$$

Hence, $\mathcal{O}(1)$ covariance between the spike trains in the external population induces $\mathcal{O}(N)$ covariance between the external input currents to neurons in the recurrent network. This is a result of the effects of "pooling" on covariances, namely that the covariance between two sums of N correlated random variables is typically $\mathcal{O}(N)$ times larger than the covariances between the individual summed variables [31, 59, 60].

We next derive the CSD between spike trains and external inputs. First note that

$$\begin{aligned} \langle \boldsymbol{T}, \boldsymbol{X} \rangle &= \langle \boldsymbol{R}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle \\ &= \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle \end{aligned} \tag{17}$$

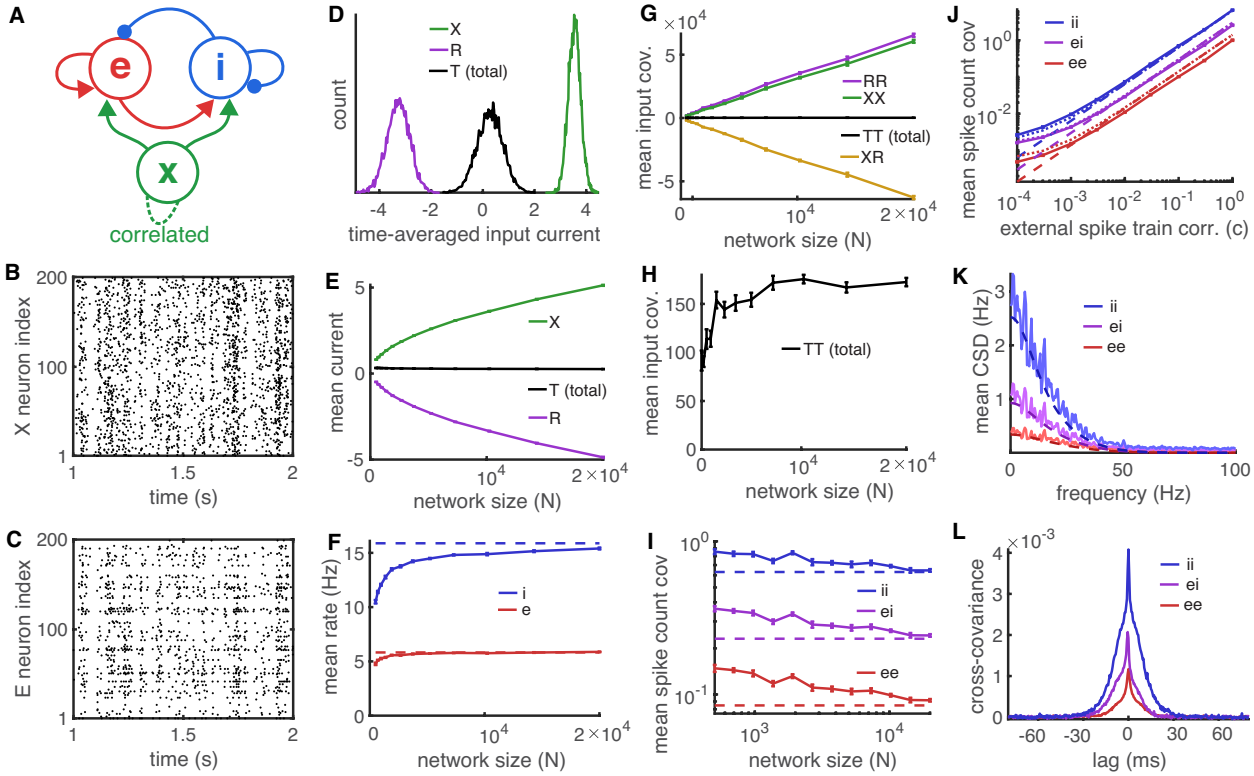

FIG. 2. **The correlated state in densely connected balanced networks. A–I and K)** Same as Fig. 1A–I and L respectively except spike trains in the external population, X, were correlated Poisson processes with spike count correlation $c = 0.1$ and Eq. (19) was used for the dashed lines in panels I and K. **J)** Mean spike count covariance between neurons in the recurrent network as a function of the correlation, c, between spike trains in the external population. Dashed curves are from Eq. (19) and dotted are from Eq. (9). **L)** Mean cross-covariance functions between neurons in the recurrent network (units ms^{-2}; $N = 10^4$). Synaptic time constants were $\tau_e = 8$, $\tau_i = 4$, and $\tau_x = 10$ms in all simulations and the correlation time constant for the spike trains in the external population was $\tau_c = 5$ms. In all simulations, recurrent and feedforward connection probabilities are all $p_{ab} = 0.1$ for $a = e, i$ and $b = e, i, x$.

from Eqs. (13) and (14). It follows from our assumption that neuronal transfer is $\mathcal{O}(1)$ (see Eq. (6)) that $\langle \boldsymbol{T}, \boldsymbol{X} \rangle \sim \langle \boldsymbol{S}, \boldsymbol{X} \rangle$ which, combined with Eq. (17), gives

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle.$$

Because $\langle \boldsymbol{S}, \boldsymbol{X} \rangle$ appears on both sides of this equation with a $\sqrt{N}$ coefficient on one side and not the other, this is only consistent if there is a cancellation between the two terms on the right hand side. Specifically, this cancellation implies that

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle = -\frac{1}{\sqrt{N}} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(\sqrt{N}) \tag{18}$$

since $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(N)$. We can now calculate the CSD between spike trains in the recurrent network. First note that

$$\begin{aligned} \langle \boldsymbol{S}, \boldsymbol{T} \rangle &= \langle \boldsymbol{S}, \boldsymbol{R} \rangle + \langle \boldsymbol{S}, \boldsymbol{X} \rangle \\ &= \sqrt{N} \langle \boldsymbol{S}, \boldsymbol{S} \rangle W^* + \langle \boldsymbol{S}, \boldsymbol{X} \rangle \\ &= \sqrt{N} \langle \boldsymbol{S}, \boldsymbol{S} \rangle W^* - \frac{1}{\sqrt{N}} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle \end{aligned}$$

which follows from Eqs. (13), (14), and (18). The assumption of $\mathcal{O}(1)$ transfer from Eq. (6) implies that $\langle \boldsymbol{S}, \boldsymbol{T} \rangle \sim \langle \boldsymbol{S}, \boldsymbol{S} \rangle$ so we may conclude that

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \sqrt{N} \langle \boldsymbol{S}, \boldsymbol{S} \rangle W^* - \frac{1}{\sqrt{N}} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle,$$

which is only consistent if there is cancellation between the terms on the right hand side. This cancellation can only be realized if

$$\begin{aligned} \langle \boldsymbol{S}, \boldsymbol{S} \rangle &= \frac{1}{N} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle W^{-*} \\ &= W^{-1} W_x \langle S_x, S_x \rangle W_x^* W^{-*} \end{aligned} \tag{19}$$

which is $\mathcal{O}(1)$ and where we have omitted terms of smaller order.

In summary, $\mathcal{O}(1)$ covariance between spike trains in the external population produces $\mathcal{O}(N)$ covariance between neurons' external inputs, but $\mathcal{O}(1)$ covariance between spike trains in the recurrent network on average. We hereafter refer to this state as the "correlated state" since it produces moderately strong spike train correlations in contrast to the asynchronous state characterized by

extremely weak spike train correlations. The reduction from $\mathcal{O}(N)$ external input covariance to $\mathcal{O}(1)$ spike train covariance arises from the same cancellation mechanism that reduces $\mathcal{O}(1)$ external input correlation to $\mathcal{O}(1/N)$ spike train correlations in the asynchronous state.

Notably, evaluating Eq. (19) does not require knowledge of neurons' power spectral densities or spike count variances, but only depends on synaptic parameters and external spike train statistics. This is beneficial because the direct numerical computation of power spectral densities or spike count variances in networks of integrate-and-fire neurons can be difficult (see discussion in Section III). This contrasts to Eq. (9) that depends on mean power spectral densities or spike count variance through the second term. Note that Eq. (9) is also valid in the correlated state (see Appendix A). Specifically, Eq. (19) can be obtained from Eq. (9) by taking $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(N)$ and ignoring smaller order terms. Hence, Eq. (9) is more accurate at finite N, but requires knowledge of mean power spectral densities whereas Eq. (19) is accurate as $N \to \infty$ in the correlated state.

To demonstrate these results, we simulated a network of $N = 10^4$ neurons identical to the network from Fig. 1 except that spike trains in the external population were correlated Poisson processes (Fig. 2A,B) with

$$\langle S_x, S_x \rangle(f) = c r_x e^{-4f^2\pi^2\tau_c^2}. \tag{20}$$

Here, $r_x = 10$Hz is the same firing rate used in Fig. 1, $c = 0.1$ quantifies the spike count correlation coefficient between the spike trains in the external population over large counting windows, and τ_c =5ms quantifies the timescale over which these correlations occur. See Appendix C for a description of the algorithm used to generate the spike trains.

The recurrent network exhibited moderately correlated spike trains in contrast to spike trains in the asynchronous state (Fig. 2C, compare to Fig. 1C; mean spike count correlation between neurons with rates at least 1 Hz was 0.077 in the correlated state). As in the asynchronous state, external and recurrent synaptic input sources approximately canceled (Fig. 2D), as predicted by balanced network theory. Varying N demonstrates that the network exhibits the same cancellation between $\mathcal{O}(\sqrt{N})$ mean external and recurrent synaptic input sources and that Eq. (5) for the mean firing rates is accurate (Fig. 2E,F).

As predicted by the analysis of the correlated state, the covariance between individual sources of input currents appear $\mathcal{O}(N)$ (Fig. 2G), but cancel to produce much smaller, approximately $\mathcal{O}(1)$, total input covariance (Fig. 2G,H). Mean spike count covariances also appear $\mathcal{O}(1)$ and converge toward the limit predicted by Eq. (19) (Fig. 2I). The timescale of correlations between neurons in the recurrent network is quantified by their mean cross-covariance functions (Fig. 2L). Repeating the simulation from Fig. 2I with different connectivity parameters showed similar overall results (see Appendix B and Fig. 7C).

We next investigated the dependence of spike count covariance in the recurrent network on the magnitude, c, of spike train correlations in the external population. The correlated state is characterized by $c \neq 0$ and the asynchronous state by $c = 0$. This transition is continuous in the sense that, for fixed N, sufficiently small values of $c > 0$ generate correlations similar to those in the asynchronous state. However, the transition between asynchronous and correlated spiking with increasing c becomes more abrupt for larger values of N. Specifically (see Appendix A),

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(cN) + \mathcal{O}(1)$$

and

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \mathcal{O}(c) + \mathcal{O}(1/N).$$

Hence, when N is large, even very small values of $c \gg 1/N$ should produce much stronger correlations than those produced in the asynchronous state ($c = 0$).

Indeed, simulations in which $N = 10^4$ is fixed and c is varied between $c = 10^{-4}$ and $c = 1$ show that Eq. (19) is relatively accurate for c larger than 10^{-3} (Fig. 2J, compare solid to dashed). When c is so small that Eq. (19) is inaccurate, Eq. (9) provides a more accurate approximation (Fig. 2J, compare solid to dotted), but requires the empirical estimation of power spectral densities. The approximate linear relationship between spike train correlation and external input correlation (at larger values of $c \gg 1/N$) is consistent with data from cultured cortical populations [30], but those data also show moderate spike train correlations in the absence of external input correlation, in contrast to our model. These could arise from pattern forming intrinsic dynamics (see Discussion).

Mean CSDs between spike trains in the recurrent network closely matched the theoretical predictions from Eq. (19) over a range of frequencies (Fig. 2K) even though our theory only strictly applies to CSDs at sufficiently low frequencies (see Eq. (10) and surrounding discussion). To understand why this is the case, we again turn to linear response theory. Under the assumption that neurons' input CSDs are approximately linearly transferred to their spike train CSDs (*i.e.*, Eq. (11)) one can derive an approximation to spike train CSDs (see [8–14, 41, 46] for derivations of this and similar approximations)

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle \approx [A^{-1} - \sqrt{N}W]^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle [A^{-1} - \sqrt{N}W]^{-*} \tag{21}$$

where recall that $A(f)$ is the matrix of mean-field susceptibility functions. In the $N \to \infty$ limit for any fixed f, Eq. (21) reduces to Eq. (19). At finite N, the two equations are approximately equivalent whenever $\sqrt{N}|W| \gg |A|^{-1}$. Note that this is equivalent to the condition in Eq. (10) and the surrounding discussion, where it was argued that this condition is violated for sufficiently large f because $|A^{-1}(f)| \to \infty$ as $f \to \infty$. Therefore, when f is sufficiently large, $|A^{-1}| \gg \sqrt{N}|W|$ and we can ignore the $\sqrt{N}W$ terms in Eq. (21) to get

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle \approx A \langle \boldsymbol{X}, \boldsymbol{X} \rangle A^*. \tag{22}$$

for sufficiently large f. In other words, at frequencies too high (timescales too fast) for synapses to track, external input CSD is transferred directly to spike train CSD through the neurons' susceptibility functions and recurrent connectivity plays a vanishing role. For the simulations in Fig. 2K, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim e^{-4f^2\pi^2\tau_c^2}/f^2$ at large N and f, but as discussed in Section III, $|A|^2 \sim 1/f^2$ for EIF neurons in the fluctuation dominated regime. Therefore, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ decays to zero much faster than A for large f. Hence, for the larger values of f for which Eq. (22) is accurate, $A\langle \boldsymbol{X}, \boldsymbol{X} \rangle A^* \approx 0$. This explains why the approximation in Eq. (19) does not lose accuracy at higher frequencies in Fig. 2K: because $\langle \boldsymbol{S}, \boldsymbol{S} \rangle \approx 0$ anyways for larger values of f.

V. THE CORRELATED STATE PRODUCES TIGHT BALANCE BETWEEN EXCITATORY AND INHIBITORY INPUT FLUCTUATIONS CONSISTENT WITH CORTICAL RECORDINGS

We have so far considered cancellation between positive and negative sources of input correlations at the mean-field level, *i.e.*, averaged over pairs of postsynaptic neurons (Figs. 1G,H and 2G,H). Previously published *in vivo* intracellular recordings from neurons in rat barrel cortex in reveal that this cancellation occurs even at the level of single postsynaptic neuron pairs [23]. Specifically, paired intracellular recordings of spontaneous neural activity were performed between nearby neurons (distance $< 500\ \mu$m) in the barrel cortex of lightly anesthetized rats in current-clamp mode. When one neuron was clamped near its inhibitory reversal potential and another neuron is clamped near its excitatory reversal potential (spiking suppressed with QX-314), recorded membrane potential fluctuations are approximately mirror images of one another (Fig. 3A, top). Similarly, if both neurons are held near their excitatory reversal potential (Fig. 3A, middle) or both near their inhibitory reversal potential (Fig. 3A, bottom), recorded membrane potential fluctuations are highly correlated. This implies that fluctuations in the excitatory and inhibitory synaptic input to one neuron are strongly correlated with fluctuations in the excitatory and inhibitory input to other nearby neurons (see [23] for more details and interpretation).

To test whether this phenomenon occurred in our simulations, we randomly chose two neurons and decomposed their synaptic input into the total excitatory $(E+X)$ and the inhibitory (I) components. In the asynchronous state, input current fluctuations were fast and largely unshared between neurons or between current sources in the same neuron (Fig. 3B), in contrast to evidence from *in vivo* recordings. Input current fluctuations in the correlated state were larger and highly synchronized between neurons (Fig. 3C), consistent with evidence from *in vivo* recordings. This precise tracking of fluctuations in excitatory and inhibitory synaptic currents is referred to as "tight balance" [61] (as opposed to the "loose balance" of the

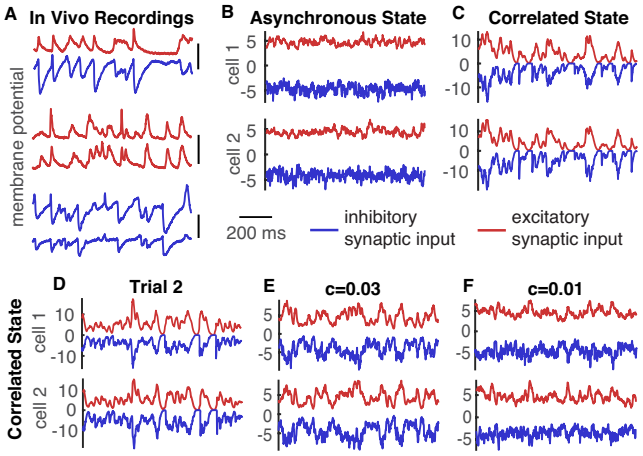

FIG. 3. **Excitatory-inhibitory tracking in vivo and in simulations. A)** In vivo membrane potential recordings from neurons in rat barrel cortex, reproduced from [23]. Each pair of traces are simultaneously recorded membrane potentials. Red traces were recorded in current clamp mode near the reversal potential of inhibition and blue traces near the reversal potential of excitation (with action potentials pharmacologically suppressed), so red traces are approximately proportional to excitatory input current fluctuations and blue traces approximate inhibitory input current fluctuations. Vertical scale bars are 20mV. For ease of comparison with the current traces in panels B–F, the red voltage trace in the top of A was plotted above the corresponding blue trace. **B,C)** Excitatory (red) and inhibitory (blue) synaptic input currents to two randomly selected excitatory neurons in the asynchronous (B) and correlated (C) states. Simulations were the same as those in Figs. 1B-D and 2A-D respectively. **D)** Same as C, but for a second trial with the same connection matrix. **E,F)** Same as C, but correlations between the external population's spike trains changed to $c = 0.03$ and 0.01 respectively (from $c = 0$ in B and $c = 0.1$ in C).

asynchronous state). The results would be similar if we decomposed inputs into their external (X) and recurrent $(R = E + I)$ sources instead of excitatory $(E + X)$ and inhibitory (I). The large fluctuations in synaptic currents in the correlated state are shared between neurons, but exhibit trial-to-trial variability (compare Fig. 3C to D), clarifying that they represent noise correlations instead of signal correlations.

An intuitive way of understanding the tight balance produced in the correlated state is to note that shared fluctuations in neurons' external synaptic input is $\mathcal{O}(cN)+\mathcal{O}(1)$ and these fluctuations are inherited by the recurrent network so that fluctuations between neurons' excitatory $(E + X)$ or inhibitory (I) synaptic inputs are also $\mathcal{O}(cN) + \mathcal{O}(1)$. Hence, shared fluctuations are $\mathcal{O}(N)$ in the correlated state $(c > 0)$ and $\mathcal{O}(1)$ in the asynchronous state $(c = 0)$. Since connectivity in the networks is homogeneous, the strongly correlated input current fluctuations also have a similar magnitude in the correlated state, so they appear to track each other. In the asynchronous state, the more weakly correlated fluctuations are "washed

out" by uncorrelated variability in the network, so currents do not appear to closely track each other. This implies that the transition from the loose balance of the asynchronous state to the tight balance of the correlated state occurs continuously as c is increased from zero, but also that tight balance is realized at small values of $c > 0$ whenever N is large (compare to the discussion of Fig. 2J above). Indeed, simulations demonstrate this continuous transition between tight and loose balance as c is modulated (Fig. 3B,C,E,F).

Despite the striking differences between excitatory and inhibitory synaptic currents in the correlated versus asynchronous states, the distributions of spike count covariances were qualitatively somewhat similar in the two states (Fig. 4A). It is more common in the literature to report spike count correlations instead of spike count covariances and it is common to omit neurons with low firing rates. Therefore, we next computed the distribution of spike count correlations between excitatory neurons with firing rates of at least 1 Hz (Fig. 4B). It is easier to distinguish between the correlated and asynchronous states from these distributions, but the distributions are still somewhat similar. Note especially that, despite the differences between the excitatory and inhibitory synaptic currents between the $c = 0$ and $c = 0.03$ cases (Fig. 3B,E), the distributions of spike count correlations are qualitatively similar (Fig. 4A,B, compare light and dark gray). The mean spike count correlations differed by orders of magnitude across the two states (mean correlation between excitatory neurons' spike trains 2.6×10^{-4}, 2.4×10^{-2}, 6.6×10^{-2} for $c = 0, 0.03, 0.1$ respectively) while the standard deviation of correlations was similar across the states (7.4×10^{-2}, 8.1×10^{-2}, and 1.2×10^{-1} for $c = 0, 0.03, 0.1$ respectively). Hence, while the distributions of correlations are qualitatively somewhat similar, the asynchronous state is distinguished by having a much larger standard deviation than mean correlation value [31, 48].

To further quantify differences between covariances in the asynchronous and correlated states, we first computed the average covariance between the excitatory and inhibitory input to pairs of (excitatory) neurons in the network. These averages have the same dependence on network size, N, as they do when input currents are broken into external and recurrent sources (compare Fig. 4C,D to Figs. 1G and 2G). Specifically, in the asynchronous state, covariances between individual current sources are $\mathcal{O}(1)$ on average, but cancel to produce weak $\mathcal{O}(1/N)$ covariance between the total synaptic input to neurons on average (Fig. 4C). In the correlated state, the average covariance between individual input sources is $\mathcal{O}(N)$ and cancellation produces $\mathcal{O}(1)$ average total input covariance (Fig. 4D).

Hence, despite the precise cancellation of positive and negative sources of input covariance at the mean-field level in the asynchronous state (Fig. 4C), the tracking suggested by this cancellation is apparently not observed at the level of individual neuron pairs (Fig. 3C). To see why this is the case, we computed the distribution of input current covariances across all pairs of excitatory neurons. We found that these distributions were broad and the distribution of total input covariance was highly overlapping with the distributions of individual input current sources (Fig. 4E, the black distribution overlaps with the others). This implies that cancellation does not reliably occur at the level of individual pairs since, for example, the total input covariance for a pair of neurons can be similar in magnitude or even larger than the covariance between those neurons' excitatory inputs.

The distributions of input covariances were strikingly different in the correlated state. The distribution of total input covariances was far narrower than the distributions of individual input current sources and the distributions were virtually non-overlapping (Fig. 4F). Hence, a precise cancellation between positive and negative sources of input covariance must occur for every neuron pair, leading to the tight balance observed in Fig. 3E.

These results are better understood by computing the empirical variance of input covariances across neuron pairs in simulations as N is varied. In the asynchronous state, the empirical variance of input covariances from all sources appear to scale like $\mathcal{O}(1)$ (Fig. 4G). Since the mean input covariance between individual sources are also $\mathcal{O}(1)$ (Fig. 4C), the overlap between distributions in Fig. 4E is expected. In the correlated state, the empirical variances of input covariances appear to scale like $\mathcal{O}(N)$ except for the variance of the total input covariance, which appears to scale like $\mathcal{O}(1)$ (Fig. 4H). If the variances scale like $\mathcal{O}(N)$, then the standard deviations would scale like $\mathcal{O}(\sqrt{N})$. This, combined with the fact that the mean input covariances between individual sources scale like $\mathcal{O}(N)$, implies that the distributions in Fig. 4G will be non-overlapping when N is large. The same conclusions would be reached if we decomposed inputs into their external (X) and recurrent ($R = E + I$) sources instead of total excitatory ($X + E$) and inhibitory (I). Note, that the scaling of the variance of covariances reported here was only computed empirically and we have not derived these scalings analytically. It is possible that, at larger N, the scaling becomes different from how it appears from our simulations.

VI. CORRELATED VARIABILITY FROM SINGULAR MEAN-FIELD CONNECTIVITY STRUCTURE

We have shown that $\mathcal{O}(1)$ spike train correlations can be obtained in balanced networks by including correlations between neurons in an external layer ($\langle S_x, S_x \rangle \sim \mathcal{O}(1)$), defining what we refer to as the "correlated state." Previous work shows that $\mathcal{O}(1)$ spike train correlations can be obtained in the recurrent network with uncorrelated external spike trains ($\langle S_x, S_x \rangle = 0$) when the mean-field connectivity matrix is constructed in such a way that the recurrent network cannot achieve the cancellation required for these states to be realized [39–41]. This

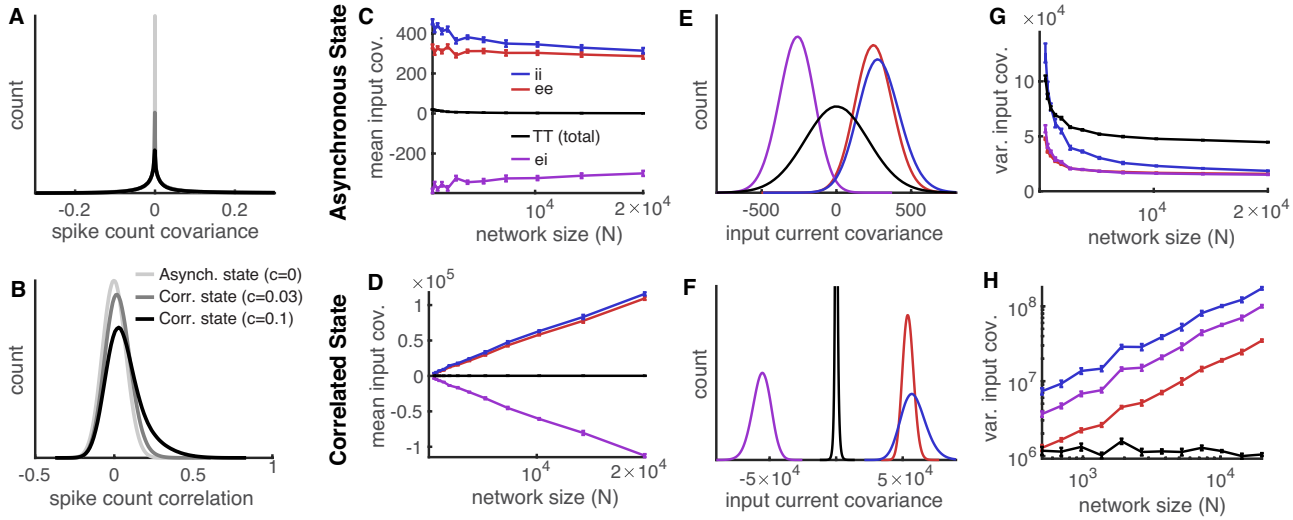

FIG. 4. **The scaling of mean and variance of excitatory and inhibitory input covariance in the asynchronous and correlated states. A,B)** Distributions of spike count covariances (A) and correlations (B) between excitatory neurons in the asynchronous state ($c = 0$, light gray) and in the correlated state ($c = 0.03$, dark gray; $c = 0.1$, black). For spike count correlations, neurons with firing rates less than 1 Hz were omitted from the analysis. **C,D)** Same as Figs. 1G and 2G, except inputs were decomposed into their excitatory ($E + X$), and inhibitory (I) components instead of external and recurrent. Red curves show mean excitatory-excitatory input covariance, blue show inhibitory-inhibitory, purple show excitatory-inhibitory, and black curves show total (same as black curves in Figs. 1G and 2G). **E,F)** Histogram of input current covariances across all excitatory cell pairs for a network of size $N = 10^4$. **G,H)** Same as C,D except we plotted the variance of covariances across cell pairs instead of the mean. As above, integrated currents have units $C_m mV$, so input covariances have units $C_m^2 mV^2$ and the variance of covariances have units $C_m^4 mV^4$ where C_m is the arbitrary membrane capacitance. In panels D,F,H we set $c = 0.1$.

can be achieved using networks with several discrete sub-populations or networks with distance-dependent connectivity. We first review these previous results by considering networks with discrete sub-populations, then show that excitatory-inhibitory tracking is similar to the asynchronous state in this case.

The recurrent networks considered above have two statistically homogeneous sub-populations: one excitatory and one inhibitory and the external population is a single homogeneous population. Suppose instead that there are K sub-populations in the recurrent network, with the kth population containing $N_k = q_k N$ neurons where $\sum_k q_k = 1$. Connectivity is random with p_{jk} denoting the connection probability from population k to j, and $j_{jk}/\sqrt{N}$ denoting the strengths of the connections for $j, k = 1, \dots K$. All neurons in population k are assumed to have the PSC kernel $\eta_k(t)$ which is again assumed to have integral 1. Similarly, suppose that the external network contains K_x sub-populations each with $N_k^x = q_k^x N_x$ neurons where $\sum_k q_k^x = 1$. Feedforward connection probabilities and strengths are given by p_{jk}^x and $j_{jk}^x/\sqrt{N}$ for $j = 1, \dots K$ and $k = 1, \dots, K_x$. Assume that q_k, p_{jk}, j_{jk}, q_k^x, p_{jk}^x, and j_{jk}^x are all $\mathcal{O}(1)$. We then define the $K \times K$ mean-field recurrent connectivity matrix by $[W]_{jk} = p_{jk} j_{jk} q_k \widetilde{\eta}_k$ and the mean-field feedforward connectivity matrix by $[W_x]_{jk} = p_{jk}^x j_{jk}^x q_k^x \widetilde{\eta}_k^x$. For all of the networks considered above, we had $K = 2$ and $K_x = 1$.

When W is an invertible matrix, Eqs. (5), (9), and (19) are equally valid for networks with several subpopulations as they are for the simpler networks considered above. Hence, the mean-field theory of firing rates and correlations extends naturally to networks with several populations [40, 41, 43–45, 58]. However, when W is singular, Eqs. (5), (9), and (19) cannot be evaluated. Instead, Eq. (5) can be re-written as

$$W\boldsymbol{r} = -W_x \boldsymbol{r}_x. \tag{23}$$

When W is singular, this equation only has a solution, $\boldsymbol{r}$, when $\overline{\boldsymbol{X}} = -W_x \boldsymbol{r}_x$ is in the range or "column space" of W. Otherwise, balance is broken. An in-depth analysis of firing rates in such networks is provided in previous work [44, 45, 58] (and extended to spatially continuous networks in [41, 43, 58]), so we hereafter assume that $\overline{\boldsymbol{X}}$ is in the range of W and balance is achieved.

A similar analysis may be applied to spike train CSDs. For simplicity, we assume here that spike trains in the external population are uncorrelated, $\langle S_x, S_x \rangle = 0$, since this is the case considered in previous work and since this is the case in which a singular W breaks the asynchronous state. Eq. (9) can be re-written as

$$NW\langle \boldsymbol{S}, \boldsymbol{S} \rangle W^* + W \mathrm{diag}(\{S_b, S_b\}/q_b) W^* = \langle \boldsymbol{X}, \boldsymbol{X} \rangle. \tag{24}$$

where $\mathrm{diag}(\{S_b, S_b\}/q_b)$ is a diagonal matrix with ratio of the mean power spectral density, $\{S_b, S_b\}$, divided by $q_b = N_b/N$ on the diagonal, and where we have ignored smaller order terms. When W is singular, Eq. (24) is not guaranteed to have a solution, $\langle \boldsymbol{S}, \boldsymbol{S} \rangle$. More precisely, a

solution can only exist when the $K \times K$ matrix, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$, is in the range of the linear operator $\mathcal{L}$ defined by

$$\mathcal{L}U = WUW^*.$$

In that case, Eq. (24) has a solution so that $\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \mathcal{O}(1/N)$ and the asynchronous state is still realized. However, if $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ is not in the range of $\mathcal{L}$, the asynchronous state cannot be realized because Eq. (24) does not have a solution.

Using Eqs. (13), (14), and (16), we can write the mean-field total input CSD as

$$\langle \boldsymbol{T}, \boldsymbol{T} \rangle = NW\langle \boldsymbol{S}, \boldsymbol{S} \rangle W^* + \sqrt{N}(W\langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{S} \rangle W^*) + \langle \boldsymbol{X}, \boldsymbol{X} \rangle. \tag{25}$$

If W is not invertible, then W^* has a non-trivial nullspace. Let $v_1, v_2, \ldots, v_n$ be a basis for the nullspace of W^* and define

$$P = v_1 v_1^* + v_2 v_2^* + \cdots + v_n v_n^*$$

which is a self-adjoint matrix that defines the orthogonal projection onto the nullspace of W^*. Note that P is a Hermitian matrix ($P = P^*$) and $PW = W^*P = 0$ (the zero matrix). Define the projection $A_0 = PAP$ for any matrix A. Unless $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ is carefully constructed otherwise, we can expect that

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle_0 \sim \langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(1).$$

Then take the projection of both sides of Eq. (25) above to get

$$\langle \boldsymbol{T}, \boldsymbol{T} \rangle_0 = \langle \boldsymbol{X}, \boldsymbol{X} \rangle_0 \sim \mathcal{O}(1). \tag{26}$$

Hence, the total input CSD is $\mathcal{O}(1)$ when $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ is not in the range of $\mathcal{L}$, even though it is $\langle \boldsymbol{X}, \boldsymbol{X} \rangle / N$ when W is invertible (*i.e.*, in the asynchronous state). Moreover, the structure of $\langle \boldsymbol{T}, \boldsymbol{T} \rangle$ is given to highest order in N by $\langle \boldsymbol{X}, \boldsymbol{X} \rangle_0 = P\langle \boldsymbol{X}, \boldsymbol{X} \rangle P$, which can be computed exactly from knowledge of the structure of $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ and W.

When neural transfer from $\boldsymbol{T}$ to $\boldsymbol{S}$ is $\mathcal{O}(1)$ (see Eq. (6) and surrounding discussion), this implies that $\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim \mathcal{O}(1)$ so that the asynchronous state is broken when $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ is not in the range of $\mathcal{L}$. While we cannot be certain that $\langle \boldsymbol{S}, \boldsymbol{S} \rangle$ has the same structure as $\langle \boldsymbol{T}, \boldsymbol{T} \rangle$, it should have a similar structure as long as neural transfer of correlations is similar for each sub-population.

Previous work [41] derived similar conditions on the cancellation required for realizing the asynchronous state in networks of binary neurons, extended the analysis to spatially extended networks, and derived analytical expressions for covariances in these networks. Other work [39, 40] analyzed singular connectivity in networks of integrate-and-fire neurons and the extension to spatially extended networks, but only showed when the asynchronous state was broken and did not derive the structure of covariances when it was broken.

To demonstrate these results, we consider the same network from above with re-wired feedforward projections from the external population. Specifically, divide the excitatory, inhibitory, and external populations each into two equal-sized sub-populations, labeled e_1, i_1, x_1, e_2, i_2, and x_2 where population a_k contains $N_a/2$ neurons. Hence the network has the same total number of neurons as before, but we have simply sub-divided the populations. To distinguish this network from the one considered in Figs. 1 and 2, we refer to the previous network as the 3-population network and to this modified network as the 6-population network.

We re-wire the feedforward connections so that x_1 only connects to e_1 and i_1, and x_2 only projects to e_2 and i_2. Specifically, we set the connection probabilities to $p_{a_j x_k} = 2p_{ax}$ if $j = k$ and $p_{a_j x_k} = 0$ if $j \neq k$ for $a, b = e, i$ and $j, k = 1, 2$, where p_{ab} are the connection probabilities for the 3-population network and $p_{a_j b_k}$ for the 6-population network. This re-wiring assures that neurons in the recurrent network receive the same number of feedforward connections on average from the external population. The recurrent connectivity structure is not changed at all. Specifically, we set $p_{a_j b_k} = p_{ab}$ for $a, b = e, i$. All connection strengths are unchanged, $j_{a_j b_k} = j_{ab}$ for $a = e, i$ and $b = e, i, x$ and all neurons in the external population have the same firing rate, r_x, as before. See Fig. 5A for a schematic of this network.

The feedforward mean-field connectivity matrix can be written in block form as

$$W_x = \begin{bmatrix} W_x^{2\times 1} & \mathbf{0} \\ \mathbf{0} & W_x^{2\times 1} \end{bmatrix}$$

where $\mathbf{0}$ is the 2×1 zero-matrix and $W_x^{2\times 1} = [w_{ex} \;\; w_{ix}]^T$ is the 2×1 feedforward connectivity matrix for the 3-population network. Note that W_x is 4×2 since there are 4 populations in the recurrent network and 2 populations in the external population. The recurrent mean-field connectivity matrix is

$$W = \frac{1}{2} \begin{bmatrix} W^{2\times 2} & W^{2\times 2} \\ W^{2\times 2} & W^{2\times 2} \end{bmatrix}$$

where

$$W^{2\times 2} = \begin{bmatrix} w_{ee} & w_{ei} \\ w_{ie} & w_{ii} \end{bmatrix}$$

is the 2×2 recurrent connectivity matrix for the 3-population network. Note that W is 4×4. Here, $w_{ab} = p_{ab} j_{ab} q_b \widetilde{\eta}_b$ are the same values used above for analyzing the 3-population network.

Even though W is non-invertible, $\overline{\boldsymbol{X}} = W_x [r_x \;\; r_x]^T$ is in the range of W for this example, so firing rates in the balanced state can be computed using Eq. (23), and are identical to the firing rates for the 3-population networks considered above.

The nullspace of W^* is spanned by the orthonormal vectors

$$v_1 = \frac{1}{\sqrt{2}} [1 \;\; 0 \;\; -1 \;\; 0]^T$$

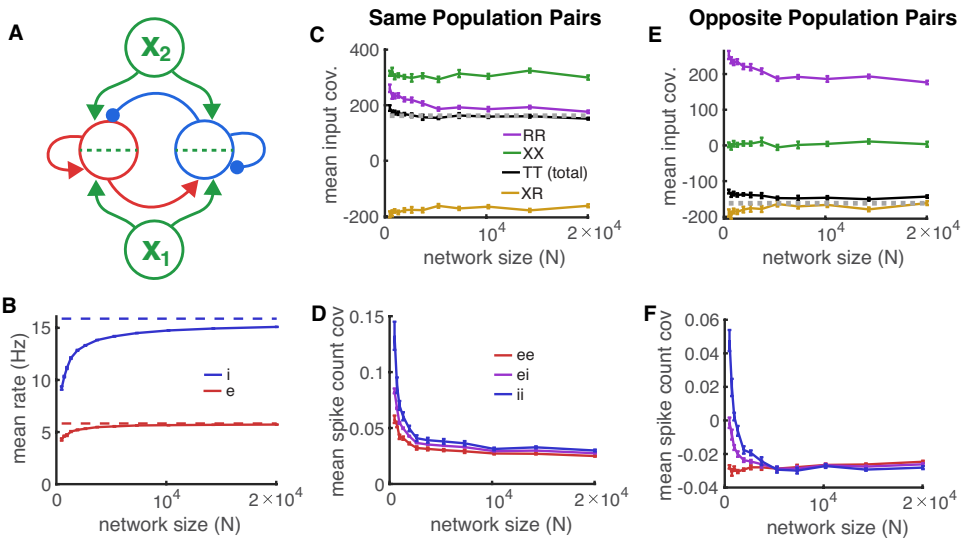

FIG. 5. **Correlated variability in a balanced network with singular mean-field connectivity matrix. A)** Network schematic. The recurrent network is statistically identical to the networks considered previously, but there are two external populations that each connect to a different half of the neurons in the recurrent network. **B)** Same as Fig. 1F, but for the multi-population network from A. **C)** Same as Fig. 1G, but for the network in A and where input covariances are only averaged over postsynaptic neurons in the same group (both postsynaptic cells in e_1 or both in e_2). The dashed gray curve shows the theoretical prediction for total input covariance (the black curve) from Eq. (27). **D)** Same as Fig. 1I, but for the network in A and where spike count covariances are only averaged over postsynaptic neurons in the same group (first cell in a_j and second cell in b_j for $a, b = e, i$ and $j = 1, 2$). **E,F)** Same as C and D, but covariances are computed between cells in opposite groups (one cell in a_1 and the other cell in b_2).

and

$$v_2 = \frac{1}{\sqrt{2}}[0 \;\; 1 \;\; 0 \;\; -1]^T$$

so the projection matrix is given in block form by

$$P = \frac{1}{2}\begin{bmatrix} I_2 & -I_2 \\ -I_2 & I_2 \end{bmatrix}$$

where I_2 is the 2×2 identity matrix.

The external input CSD is determined by the average number of overlapping feedforward projections to any pair of neurons in the recurrent network (multiplied by their connection strength and r_x), which gives (in block form)

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle = 2\begin{bmatrix} \langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} & \mathbf{0} \\ \mathbf{0} & \langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} \end{bmatrix}$$

where $\mathbf{0}$ is the 2×2 zero matrix and $\langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2}$ is the external input CSD from the 3-population network, given by Eq. (8). Therefore, by Eq. (26),

$$\begin{aligned} \langle \boldsymbol{T}, \boldsymbol{T} \rangle_0 &= \langle \boldsymbol{X}, \boldsymbol{X} \rangle_0 = P \langle \boldsymbol{X}, \boldsymbol{X} \rangle P \\ &= \begin{bmatrix} \langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} & -\langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} \\ -\langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} & \langle \boldsymbol{X}, \boldsymbol{X} \rangle^{2\times 2} \end{bmatrix}. \end{aligned} \tag{27}$$

In other words, the mean total input CSD between excitatory neurons in the same subgroup (two neurons in e_1 or two neurons in e_2; diagonal blocks above) is positive and equal to half the mean external input between the same neurons. Hence, the cancellation by the recurrent network only reduces the external input CSD by a factor of $1/2$, as opposed to the $\mathcal{O}(1/N)$ reduction realized in the asynchronous state (when W is invertible). In contrast, the mean total input CSD between excitatory neurons in opposite subgroups (one neuron in e_1 and the other in e_2; off-diagonal blocks above) has the same magnitude as for same-subgroup pairs, but is negative. This represents a competitive dynamic between the two groups since they inhibit one another (recurrent connections are net-inhibitory in balanced networks [28, 45]), but receive different feedforward input noise. Interestingly, the average CSD between all pairs of spike trains is still $\mathcal{O}(1/N)$ in this example, but it is easy to design examples with singular W in which this is not true.

Simulating this network for varying values of N shows that firing rates approach those predicted by the balance equation (23) (Fig. 5B), confirming that balance is realized. Pairs of excitatory neurons in the same group (both neurons in e_1 or both neurons in e_2) receive positively correlated external input and recurrent input (Fig. 5C, purple and green curves) that are partially canceled by negative correlations between their recurrent and excitatory input (Fig. 5C, yellow curve). Because the cancellation is only partial, the correlation between the neurons' total inputs is $\mathcal{O}(1)$ (Fig. 5C, black curve) in contrast to the asynchronous state (compare to Fig. 1G,H where cancellation is perfect at large N). The total input covariance agrees well with the theoretical prediction from Eq. (27) (Fig. 5C,

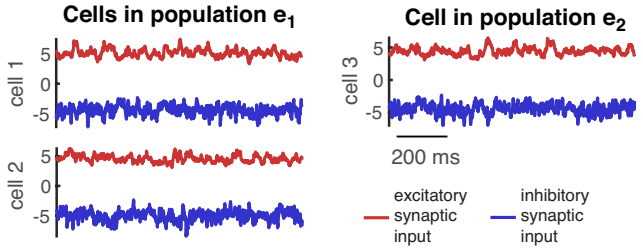

FIG. 6. **Synaptic input currents in a balanced network with correlations from singular mean-field connectivity.** Same as Fig. 3A except for the network from Fig. 5. The left two traces are input currents to two excitatory neurons in population e_1 (cells 1 and 2). The right traces are input currents to an excitatory neuron in population e_2 (cell 3).

dashed gray line). As a result of this lack of cancellation between total input covariance, spike count covariances are also $\mathcal{O}(1)$ and positive between same-group pairs (Fig. 5D). For opposite group pairs (one neuron in e_1 and the other in e_2), cancellation is also imperfect, but this leads to negative total input covariance, in agreement with the theoretical prediction from Eq. (27) (Fig. 5E), and leads to negative spike count covariances between neurons in opposite populations (Fig. 5F).

In summary, we have analyzed two mechanisms to generate $\mathcal{O}(1)$ spike train correlations in balanced networks. For the first mechanism (Fig. 2), spike trains in the external population are correlated so that external input correlations are $\mathcal{O}(N)$. Cancellation is achieved so that spike train correlations are reduced to $\mathcal{O}(1)$. For the other mechanism (Fig. 5), external input correlation is $\mathcal{O}(1)$, but precise cancellation cannot be achieved so that spike trains inherit the $\mathcal{O}(1)$ correlations from the input. How could these two mechanisms be distinguished in cortical recordings? Under the first mechanism, we showed that fluctuations of inhibitory input to individual neurons closely tracks fluctuations of other neurons' excitatory inputs (Fig. 3C). This should not be the case under the second mechanism because precise cancellation is not realized. Indeed, plotting the excitatory and inhibitory input to three excitatory neurons (two in e_1 and one in e_2) shows that input fluctuations are not closely tracked (Fig. 6). This provides a way to distinguish the two mechanisms from paired intracellular recordings. Indeed, the first mechanism (which we refer to as the "correlated state") appears more consistent with the cortical recordings considered here (compare Fig. 3A to Figs. 3C and 6).

VII. SUMMARY AND DISCUSSION

We analyzed correlated variability in recurrent, balanced networks of integrate-and-fire neurons receiving correlated feedforward input from an external population. We showed that correlations between spike trains in the recurrent network are small ($\mathcal{O}(1/N)$) when spike trains in the external population are uncorrelated, consistent with previous work on the asynchronous state [31, 40], but much larger ($\mathcal{O}(1)$) when spike trains in the external population are correlated, giving rise to a "correlated state." In both states, strong correlations in the feedforward input are canceled by recurrent synaptic input due to the excitatory-inhibitory tracking that arises naturally in densely connected balanced networks. In the correlated state, this cancellation allows for the derivation of a concise and accurate closed form expression for mean-field low frequency spike train CSDs and spike count covariances in terms of synaptic parameters alone. Hence spike count covariances in the correlated state are determined predominately by synaptic connectivity structure, not neuronal dynamics. The tracking of excitatory synaptic input by inhibition was observable on a pair-by-pair basis in the correlated state, but not the asynchronous state, suggesting that the correlated state is more consistent with some *in vivo* recordings.

In our analysis of the correlated state ($c > 0$), we only considered recurrent networks with two, statistically homogeneous neural populations: one excitatory and one inhibitory (with the exception of the simple multi-population model analyzed in Section VI). Our analysis can be extended to arbitrarily many subpopulations as long as each sub-population contains $\mathcal{O}(N)$ neurons, and also extends to networks with connection probabilities that depend on distance, orientation tuning, or other continuous quantities. This analysis has been developed for the asynchronous state in previous work [40, 41] and is easily extended to the correlated state as well. The primary difference is that $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ is $\mathcal{O}(N)$ instead of $\mathcal{O}(1)$.

Previous work has shown that networks with multiple sub-populations and networks with distance-dependent connectivity can break the asynchronous state in balanced networks when the network connectivity structure is constructed in such a way that the recurrent network cannot achieve the cancellation required for the asynchronous state [39–41], leading to $\mathcal{O}(1)$ correlations between some cell pairs. We showed that the precise tracking of excitation by inhibition provides an experimentally testable prediction for distinguishing this mechanism from the one underlying the correlated state (see Section VI).

Another alternative mechanism for achieving larger correlations in balanced networks is through instabilities of the balanced state. Such instabilities can create pattern-forming dynamics that produce intrinsically generated spike train correlations [42, 43, 62–68]. Some recordings show that local circuit connectivity can increase correlations [69], which is consistent with internally generated correlations, but inconsistent with the mechanisms that we consider here. In cultured populations of cortical neurons, moderate spike train correlations and excitatory-inhibitory tracking emerge even in the absence of correlated external input [30], suggesting that they are generated intrinsically. Correlations from pattern forming

instabilities can be distinguished from the externally produced correlations considered here in at least two ways. Since instabilities generate correlations internally, they should produce weak correlations between activity in the recurrent network and activity in the external population(s) providing input to that network [42], in contrast to the mechanisms we consider here. Also, the presence of pattern forming instabilities in balanced networks depends on the timescale and spatial extent of excitatory and inhibitory recurrent synaptic projections as well as the strength of external input to the inhibitory population [42, 43, 62, 66, 68]. In cultured populations [30], these parameters can more easily be measured or modified to help determine the role of pattern forming instabilities in generating correlations.

We considered correlations produced by near-synchronous, correlated spiking in an external population (see Appendix C), but the correlated state could also be generated by time-varying firing rates in the external population. This could be modeled, for example, by replacing the homogeneous Poisson processes in the external population by doubly stochastic Poisson processed in which the rate fluctuations were shared across cells. This would produce $\mathcal{O}(N)$ external input covariance to the recurrent network and $\mathcal{O}(1)$ spike train correlation, so the mathematical analysis we considered applies equally well to this model. Future work should consider ways to distinguish correlations arising through shared rate fluctuations from those generated by near-synchronous spiking in a presynaptic layer.

In the correlated state, we found that the mean and empirical standard deviation of correlations is similar in magnitude, but in some recordings, mean correlations are much smaller than standard deviations [32, 35, 48, 70] suggesting that those circuits were in an asynchronous state, not a correlated state, during recording [48]. More generally, some recordings show very small mean correlations consistent with the asynchronous state [32, 48, 70], while others show larger correlation magnitudes [6, 33, 71]. Indeed, the magnitude of mean correlations can depend on many factors [6, 71], including arousal [72], attention [42, 73], stimulus [74], anesthetic state [31, 32, 35, 38], and cortical area and layer [34, 75–78].

In summary, spike train correlations and excitatory-inhibitory tracking likely arise from multiple mechanisms in different cortical areas, cortical layers, and conditions. The external input correlations introduced in this work are one such mechanism. Future studies should work to enumerate these mechanisms and generate experimentally testable predictions that distinguish them.

In the correlated state, spike train correlations in the recurrent network are essentially inherited from correlations between spike trains in the external population. Hence, the $\mathcal{O}(1)$ correlations realized by this mechanism require the presence of another local network with $\mathcal{O}(1)$ correlations. This raises the question of where the $\mathcal{O}(1)$ correlations are originally generated. One possibility is that they could be generated in a presynaptic cortical area or layer through the alternative mechanisms discussed above. Another possibility is that they originate from a network that is not in the balanced state at all. Non-balanced networks can easily achieve $\mathcal{O}(1)$ spike train correlations simply from overlapping synaptic projections. While cortical circuits are commonly believed to operate in a balanced state, correlations could originate in thalamus, retina, or other sub-cortical neural populations then propagate to cortex.

The cancellation between variances of covariances observed empirically in simulations (Fig. 4F,H) is, to our knowledge, a novel observation, but we were unable to prove it analytically. Path integral approaches have recently been applied to compute variances of covariances in recurrent network models with uncorrelated external input [48]. Future work should consider the possibility of extending their analysis to networks with correlated external input.

Many previous studies of correlated variability in recurrent networks rely on linear response theory. In contrast, our derivation of Eqs. (9) and (19) used a weaker assumption that neural transfer of covariance is $\mathcal{O}(1)$ (Eq. (6) and surrounding discussion). Despite the fact that the derivation of Eqs. (9) and (19) do not depend on a linear response assumption, the equations themselves are linear and are the same that one would arrive at using linear response theory. Indeed, Eq. (9) and similar equations have been previously derived for various models using linear response techniques [5, 14, 31, 41, 46–48]. Hence, despite the use of neurons' gains or susceptibility functions in this previous work, the resulting equations for mean-field covariance in the asynchronous state does not depend on the gains in the large N limit (see equations 38-39 in the supplement to [31] and equation 38 in [46]).

Even though the resulting equation is the same, our derivation of Eq. (9) is more general than derivations that rely on linear response theory. Specifically, linear response theory assumes that $\langle S_a, U_b\rangle \approx A_a \langle T_a, U_b\rangle$, but our derivation is valid when relationship between $\langle S_a, U_b\rangle$ and $\langle T_a, U_b\rangle$ is nonlinear, as long as $\langle S_a, U_b\rangle / \langle T_a, U_b\rangle \sim \mathcal{O}(1)$ as $N \to \infty$. Our derivation also applies when $\langle S_a, U_b\rangle / \langle T_a, U_b\rangle$ depends on the identity of U, *e.g.*, it is different for $U = S$ versus $U = X$. Our derivation is also valid when non-linear neural transfer causes cross-frequency coupling, *i.e.*, when $\langle S_a, U_b\rangle(f_0)$ depends on $\langle T_a, U_b\rangle(f)$ for values of $f \neq f_0$. This is an interesting conclusion because frequencies are de-coupled in Eqs. (9) and (19), indicating an asymptotic linearity of the relationship between mean-field external input covariance and mean-field spike count covariance even in a non-linear system. This is possible because synapses in our model are linear and synaptic filtering dominates the network response properties at sufficiently large N and low frequencies (see Eqs. (6), (12), and surrounding discussion).

When input correlations are weak, covariance transfer for integrate-and-fire neurons is approximately linear so linear response approaches are justified [8, 53, 54], but transfer can become nonlinear (but still $\mathcal{O}(1)$) when cor-

relations are stronger [54, 79, 80]. Our results show that, even when the transfer of input covariance to spike train covariance is nonlinear, the mean-field relationship between $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$ and $\langle \boldsymbol{S}, \boldsymbol{S} \rangle$ is still linear to highest order in N and is the same relationship one would obtain if transfer were linear. Note that this linear relationship only applies at the mean-field level. Individual, pairwise CSDs between spike trains might still be non-linearly related to their pairwise external input CSDs. Recent work has called for looking beyond linear analysis of neuronal networks [81]. Our analysis shows that, even in networks where neural transfer of inputs is nonlinear, linear mean-field analysis can still be accurate and useful.

Despite the fact that Eq. (9) does not depend on neurons' susceptibility functions or gains, it does depend on neurons' mean spike count variance or power spectral densities. In contrast, Eq. (19) shows that mean-field covariance in the correlated state does not depend (to highest order in N) on neurons' susceptibility functions, gains, power spectral densities, or spike count variance, but only involves synaptic parameters. This is similar to previous work showing that the dominant frequencies of oscillatory activity in balanced networks depends only on synaptic parameters [82]. Similarly, previous work showed that correlations in binary networks of inhibitory neurons without external drive are independent of connection strengths in the limit of strong connectivity (see equation 30 in [83] and equation 32 in [46]). The independence of spike count covariance on neurons' susceptibilities and power spectral densities is an important conclusion because, for spiking neuron models, power spectral densities and susceptibility functions (or spike count variances and gains) can depend nonlinearly on the parameters of the neuron model as well as first, second, and higher moments of the neurons' input statistics [84–87] and they can be difficult to compute numerically.

Numerical methods for approximating power spectral densities and susceptibility functions in networks of integrate-and-fire neurons have been developed using a diffusion approximation and Fokker-Planck techniques [8, 9, 11, 12, 49, 88], but this approach assumes that neurons' synaptic input is dominated by or well-approximated by Gaussian white noise. Previous work satisfies this assumption by including Gaussian white noise as external input and making input from the recurrent network much weaker than this white noise input (for example, by making recurrent connectivity weak or sparse). In addition, instantaneous synapses ($\eta_b(t) = \delta(t)$) can make the diffusion approximation more accurate (although see [89, 90]). The diffusion approximation is not valid in our model since our external input is not well-approximated by Gaussian white noise, our recurrent input is approximately as strong as external input, and our synapses are not instantaneous. While some specialized approximations have been developed to avoid the assumption of Gaussian white noise input in various models and parameter regimes [1, 15, 55, 90–93], these approaches are outside the scope of our study since our central conclusion is that these approximations are unnecessary for deriving mean-field spike count covariance in the correlated state.

Three unproven assumptions underly our mean-field analysis of the correlated state. The first assumption is that neural transfer is $\mathcal{O}(1)$ (Eq. (6) and surrounding discussion). The third assumption is that mean-field power spectral densities or spike count variances $\{S_a, S_a\}$ are $\mathcal{O}(1)$. The third assumption is that variability connection strengths is not strongly correlated with variability in individual CSD values (see comment in derivation of Claim 2 in Appendix A). These assumptions are made in other work, even if not stated explicitly. We have been unable to rigorously prove these assumptions for the model studied here, leaving an open problem for future work.

In summary, we showed that correlations in balanced networks can be caused by feedforward input from a population of neurons with correlated spike trains, defining the "correlated state." In this state mean low-frequency CSD or spike count covariance is determined by a simple, closed-form equation of known parameters, greatly simplifying the analysis of spike count correlations in such networks. The correlated state predicts a precise balance between the fluctuations in excitatory and inhibitory synaptic input to individual neuron pairs, consistent with some in vivo recordings [23].

ACKNOWLEDGMENTS

We thank Michael S. Okun for helpful comments on a draft of the manuscript and for his contribution to collecting the data used in Fig. 3A. We thank the reviewers for helpful comments that greatly improved the manuscript. RR was supported by National Science Foundation (https://www.nsf.gov/) grants DMS-1517828, DMS-1654268, and Neuronex DBI-1707400. IL was supported by Deutsche Forschungsgemeinschaft Sonderforschungsbereiche (www.dfg.de/sfb) grant number 1089, Israel Science Foundation (https://www.isf.org.il/) grant numbers 326/07 and 1539/17, and Minerva (http://www.minerva.mpg.de/weizmann/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A: Mean-field analysis of correlated variability in the asynchronous and correlated states

Here, we give a detailed derivation of mean-field CSDs that applies to both the asynchronous and correlated states. The only differences between parameters in those states is that $\langle S_x, S_x \rangle = 0$ in the asynchronous state and $\langle S_x, S_x \rangle \sim \mathcal{O}(1)$ in the correlated state. We derive CSDs in terms of $\langle S_x, S_x \rangle$ so that the results can be applied to either state. To more clearly organize the computation, we organize the calculation into the derivation of

several claims that are derived separately. We start with a derivation of the mean-field external input CSD.

Claim 1. *Mean-field external input CSD is given by*

$$\begin{aligned}\langle \boldsymbol{X}, \boldsymbol{X}\rangle =& N W_x \langle S_x, S_x\rangle W_x^* + q_x^{-1} W_x r_x W_x^* \\ &- q_x^{-1} W_x \langle S_x, S_x\rangle W_x^*.\end{aligned} \tag{A1}$$

Proof. We first compute the external input CSD between neuron j in population $a = e, i$ and neuron k in population $b = e, i$ as

$$\begin{aligned}\langle X_j^a, X_k^b\rangle =& \left\langle \sum_{m=1}^{N_x} J_{jm}^{ax}(\eta_x * S_m^x), \sum_{n=1}^{N_x} J_{kn}^{bx}(\eta_x * S_n^x)\right\rangle \\ =& \sum_{m,n=1}^{N_x} J_{jm}^{ax} J_{kn}^{bx} \widetilde{\eta}_x \widetilde{\eta}_x^* \langle S_m^x, S_n^x\rangle \\ =& |\widetilde{\eta}_x|^2 \langle S_x, S_x\rangle \sum_{m\neq n} J_{jm}^{ax} J_{kn}^{bx} \\ &+ |\widetilde{\eta}_x|^2 r_x \sum_{m=1}^{N_x} J_{jm}^{ax} J_{km}^{bx}.\end{aligned}$$

where we used the bilinearity of the Hermitian cross-spectral operator, the fact that $\langle S_m^x, S_n^x\rangle = \langle S_x, S_x\rangle$ for $m \neq n$, and $\langle S_m^x, S_m^x\rangle = r_x$ since external spike trains are Poisson processes. Averaging over j and k and using Eq. (2), we get

$$\langle X_a, X_b\rangle = |\widetilde{\eta}_x|^2 j_{ax} p_{ax} j_{bx} p_{bx} q_x \Big[(N_x - 1)\langle S_x, S_x\rangle + r_x\Big].$$

Writing this in matrix form and using $w_{ax} = \widetilde{\eta}_x j_{ax} p_{ax} q_x$ gives Eq. (A1). □

Writing $\langle S_x, S_x\rangle \sim O(c)$ so that $c = 0$ in the asynchronous state and $c \sim \mathcal{O}(1)$ in the correlated state, Eq. (A1) implies that

$$\langle \boldsymbol{X}, \boldsymbol{X}\rangle \sim \mathcal{O}(cN) + \mathcal{O}(1).$$

This expression helps understand the dependence of correlations on both c and N.

A similar derivation gives the mean-field CSD between neurons' total and external inputs:

Claim 2. *The mean-field CSD between external and total external inputs is given by*

$$\begin{aligned}\langle \boldsymbol{T}, \boldsymbol{X}\rangle =& \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X}\rangle + \langle \boldsymbol{X}, \boldsymbol{X}\rangle \\ &+ \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X}\rangle}{\sqrt{N}}\right)\end{aligned} \tag{A2}$$

where

$$Q = \begin{bmatrix} q_e & 0 \\ 0 & q_i \end{bmatrix},$$

$$\{\boldsymbol{S}, \boldsymbol{X}\} = \begin{bmatrix} \{S_e, X_e\} & 0 \\ 0 & \{S_i, X_i\} \end{bmatrix},$$

and

$$\{S_a, X_a\} = avg_k \langle S_k^a, X_k^a\rangle$$

quantifies the average CSD between neurons' spike trains and their own external inputs.

Proof. First break the total input into its constituent parts $\boldsymbol{T} = \boldsymbol{R} + \boldsymbol{X} = \boldsymbol{E} + \boldsymbol{I} + \boldsymbol{X}$, so

$$\langle \boldsymbol{T}, \boldsymbol{X}\rangle = \langle \boldsymbol{E}, \boldsymbol{X}\rangle + \langle \boldsymbol{I}, \boldsymbol{X}\rangle + \langle \boldsymbol{X}, \boldsymbol{X}\rangle.$$

To compute the $\langle \boldsymbol{E}, \boldsymbol{X}\rangle$ term, we begin with

$$\begin{aligned}\langle E_j^a, X_k^b\rangle &= \left\langle \sum_{m=1}^{N_e} J_{jm}^{ae}(\eta_e * S_m^e), X_k^b\right\rangle \\ &= \widetilde{\eta}_e \sum_{m=1}^{N_e} J_{jm}^{ae} \langle S_m^e, X_k^b\rangle.\end{aligned}$$

Before taking averages, note that whenever $b = e$, we need to treat the $m = k$ term separately since it corresponds to the CSD between the spike train of and external input to the same neuron, which could scale differently than the case when $b = i$ or $m \neq k$. Put another way, note that the definition of the mean-field values, $\langle E_a, X_b\rangle$ and $\langle S_e, X_b\rangle$, assumes that $m \neq k$ or $b \neq e$ (see Eq. (4)). Hence, computing averages separately for each of $b = e, i$ gives

$$\begin{aligned}\langle E_a, X_e\rangle &= \frac{\widetilde{\eta}_e j_{ae} p_{ae}}{\sqrt{N}}(N_e - 1)\langle S_e, X_e\rangle + \{S_e, X_e\} \\ &= \sqrt{N} w_{ae} \langle S_e, X_e\rangle + \frac{w_{ae}}{q_e \sqrt{N}}(\{S_e, X_e\} - \langle S_e, X_e\rangle)\end{aligned}$$

and

$$\langle E_a, X_i\rangle = \frac{\widetilde{\eta}_e j_{ae} p_{ae}}{\sqrt{N}} N_e \langle S_e, X_i\rangle = \sqrt{N} w_{ae} \langle S_e, X_i\rangle.$$

Note that this step requires us to assume that individual values of the random variable, J_{jm}^{ae}, are not strongly correlated with individual values of $\langle S_m^e, X_k^b\rangle$, so that the expectation of their product can be replaced by the product of their expectations. This and similar assumptions are implicit in derivations in other studies, but are not typically proven or even explicitly stated.

An identical calculation for $\langle I_a, X_b\rangle$ gives

$$\langle I_a, X_i\rangle = \sqrt{N} w_{ai} \langle S_i, X_i\rangle + \frac{w_{ai}}{q_i \sqrt{N}}(\{S_i, X_i\} - \langle S_i, X_i\rangle)$$

and

$$\langle I_a, X_e\rangle = \sqrt{N} w_{ai} \langle S_i, X_e\rangle.$$

This allows us to write the CSD between recurrent and total input in matrix form as

$$\begin{aligned}\langle \boldsymbol{R}, \boldsymbol{X}\rangle &= \langle \boldsymbol{E}, \boldsymbol{X}\rangle + \langle \boldsymbol{I}, \boldsymbol{X}\rangle \\ &= \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X}\rangle + \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} \\ &\quad + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X}\rangle}{\sqrt{N}}\right).\end{aligned}$$

Now putting together $\langle \boldsymbol{T}, \boldsymbol{X} \rangle = \langle \boldsymbol{R}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle$ gives Eq. (A2).

□

A similar derivation gives the mean-field CSD between neurons' total and external inputs:

Claim 3. *The mean-field CSD between spike trains and total external inputs is given by*

$$\begin{aligned}\langle \boldsymbol{T}, \boldsymbol{S} \rangle =& \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{S} \rangle + \langle \boldsymbol{X}, \boldsymbol{S} \rangle \\ &+ \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{S} \rangle}{\sqrt{N}}\right)\end{aligned} \tag{A3}$$

where

$$\{\boldsymbol{S}, \boldsymbol{S}\} = \begin{bmatrix} \{S_e, S_e\} & 0 \\ 0 & \{S_i, S_i\} \end{bmatrix}$$

and

$$\{S_a, S_a\} = avg_k \langle S_k^a, S_k^a \rangle$$

quantifies the average power spectral densities of neurons' spike trains.

Proof. The derivation is identical to the derivation of Eq. (A2) above, but with X replaced by S. □

We now use Claim 2 to derive the asymptotic behavior of $\langle \boldsymbol{S}, \boldsymbol{X} \rangle$.

Claim 4. *Under the assumption of $\mathcal{O}(1)$ transfer of mean-field covariance (*i.e.*, Eq. (7)) and under the assumption that $\{\boldsymbol{S}, \boldsymbol{S}\} \sim \mathcal{O}(1)$ (*i.e.*, neurons' mean power spectral densities are $\mathcal{O}(1)$), we have that*

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle = \frac{1}{\sqrt{N}} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}{N}\right) \tag{A4}$$

in the limit of large N. For finite N, this approximation is only accurate whenever

$$\left| \frac{\langle \boldsymbol{T}, \boldsymbol{X} \rangle}{W \langle \boldsymbol{S}, \boldsymbol{X} \rangle} \right| \ll \sqrt{N} \tag{A5}$$

where the division is performed element-wise.

Proof. Combining the assumption of $\mathcal{O}(1)$ transfer of mean-field covariance (Eq. (6)) with Eq. (A2) gives

$$\begin{aligned}\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim& \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle \\ &+ \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X} \rangle}{\sqrt{N}}\right).\end{aligned} \tag{A6}$$

More rigorously, dividing both sides of Eq. (A2) element-wise by $\sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle$ and applying Eq. (7) gives Eq. (A6). At finite N, note that this derivation of Eq. (A6) from Eq. (A2) is accurate whenever Eq. (A5) is true.

We do not know the exact scaling of $\{\boldsymbol{S}, \boldsymbol{X}\}$ with N, but applying the Cauchy-Scwartz inequality gives

$$|\{\boldsymbol{S}, \boldsymbol{X}\}| \leq \sqrt{\{\boldsymbol{S}, \boldsymbol{S}\}\{\boldsymbol{X}, \boldsymbol{X}\}} = \mathcal{O}\left(\sqrt{\{\boldsymbol{X}, \boldsymbol{X}\}}\right) \tag{A7}$$

where we used our assumption that $\{\boldsymbol{S}, \boldsymbol{S}\} \sim \mathcal{O}(1)$. To compute $\{\boldsymbol{X}, \boldsymbol{X}\}$, we can repeat the derivation in Claim 1 with $j = k$ to get

$$\begin{aligned}\{\boldsymbol{X}, \boldsymbol{X}\} =& N W_x \langle S_x, S_x \rangle W_x^* + q_x^{-1} P_x^{-1} W_x r_x W_x^* \\ &- q_x^{-1} W_x \langle S_x, S_x \rangle W_x^*.\end{aligned}$$

where

$$P_x = \begin{bmatrix} p_{ex} & 0 \\ 0 & p_{ix} \end{bmatrix}.$$

Comparing this to Eq. (A1), we may conclude that

$$\{\boldsymbol{X}, \boldsymbol{X}\} \sim \langle \boldsymbol{X}, \boldsymbol{X} \rangle$$

and therefore (using Eq. (A7) and the fact that $\{\boldsymbol{X}, \boldsymbol{X}\} \geq \mathcal{O}(1)$ in both the asynchronous and correlated states), we have that

$$\{\boldsymbol{S}, \boldsymbol{X}\} \leq \mathcal{O}(\langle \boldsymbol{X}, \boldsymbol{X} \rangle)$$

This allows us to rewrite Eq. (A6) as

$$\begin{aligned}\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim& \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle \\ &+ \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{S}, \boldsymbol{X} \rangle}{\sqrt{N}}\right).\end{aligned}$$

The only way that this equation can be self-consistent is if the terms on the right hand side cancel as $N \to \infty$ which implies that $\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim \langle \boldsymbol{X}, \boldsymbol{X} \rangle / \sqrt{N}$ and, more specifically, that Eq. (A4) is satisfied as $N \to \infty$.

□

We now derive the asymptotic values of mean-field spike train CSDs.

Claim 5. *Under the assumptions made in Claim 4, we have that*

$$\begin{aligned}\langle \boldsymbol{S}, \boldsymbol{S} \rangle =& \frac{1}{N} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle W^{-*} + \frac{1}{N} Q^{-1} \{\boldsymbol{S}, \boldsymbol{S}\} \\ &+ \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}{N^{3/2}}\right)\end{aligned} \tag{A8}$$

in the limit of large N. For finite N, this approximation is accurate under Eq. (A5) and

$$\left| \frac{\langle \boldsymbol{T}, \boldsymbol{S} \rangle}{W \langle \boldsymbol{S}, \boldsymbol{S} \rangle} \right| \ll \sqrt{N} \tag{A9}$$

where the division is performed element-wise.

Proof. The proof is similar to that of Claim 4. Combining the assumption of $\mathcal{O}(1)$ transfer of mean-field covariance (Eq. (6)) with Eq. (A3) gives

$$\begin{aligned}\langle \boldsymbol{S}, \boldsymbol{S} \rangle \sim& \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{S} \rangle + \langle \boldsymbol{X}, \boldsymbol{S} \rangle \\ &+ \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{S} \rangle}{\sqrt{N}}\right).\end{aligned} \tag{A10}$$

More rigorously, dividing both sides of Eq. (A3) element-wise by $\sqrt{N}W\langle \boldsymbol{S},\boldsymbol{S}\rangle$ and applying Eq. (7) gives Eq. (A10). At finite N, note that this derivation of Eq. (A10) from Eq. (A3) is accurate whenever Eq. (A9) is true for $\boldsymbol{U}=\boldsymbol{S}$.

Taking the conjugate-transpose of Eq. (A4) gives

$$\langle \boldsymbol{X},\boldsymbol{S}\rangle = \langle \boldsymbol{X},\boldsymbol{S}\rangle^* = \frac{1}{\sqrt{N}}\langle \boldsymbol{X},\boldsymbol{X}\rangle W^{-*} + \mathcal{O}\left(\frac{\langle \boldsymbol{X},\boldsymbol{X}\rangle}{N}\right).$$

Plugging this into Eq. (A10) gives

$$\begin{aligned}\langle \boldsymbol{S},\boldsymbol{S}\rangle \sim &\sqrt{N}W\langle \boldsymbol{S},\boldsymbol{S}\rangle + \frac{1}{\sqrt{N}}\langle \boldsymbol{X},\boldsymbol{X}\rangle W^{-*} \\ &+ \frac{1}{\sqrt{N}}WQ^{-1}\{\boldsymbol{S},\boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S},\boldsymbol{S}\rangle}{\sqrt{N}}\right) \\ &+ \mathcal{O}\left(\frac{\langle \boldsymbol{X},\boldsymbol{X}\rangle}{N}\right).\end{aligned}$$

The only way this equation is self-consistent is if the terms on the right hand side cancel as $N\to\infty$, which implies that $\langle \boldsymbol{S},\boldsymbol{S}\rangle \sim \mathcal{O}(\langle \boldsymbol{X},\boldsymbol{X}\rangle/N + \{\boldsymbol{S},\boldsymbol{S}\}/N)$ and, more specifically, Eq. (A8) is satisfied as $N\to\infty$. For accuracy at finite N, we explicitly needed to assume Eq. (A9), but also implicitly assumed Eq. (A5) when we used Eq. (A4).

□

Appendix B: Results from simulations with different connectivity parameters

So far, all of our simulations used the same connectivity parameters. Here, we consider results with different parameters. Specifically, we set $j_{ee}/C_m = 35$mV, $j_{ei}/C_m = -200$mV, $j_{ie}/C_m = 120$mV, $j_{ii}/C_m = -300$mV, $j_{ex}/C_m = 200$mV, and $j_{ix}C_m = 150$mV. Note that j_{ab} was scaled by $\sqrt{N}$ to produce the true connection strengths, as indicated in Results. We then repeated the simulations from Figs. 1I, 1J, and 2I. The results, shown in Fig. 7A,B, and C respectively, show similar overall findings to those reported for the connectivity parameters used in Results.

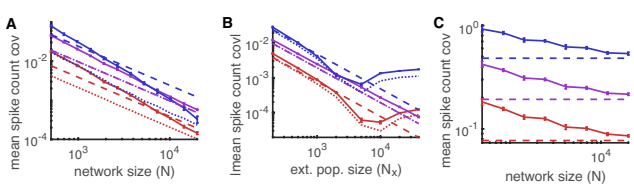

FIG. 7. **Results from simulations with different connectivity parameters A-C)** Same as Figs. 1I, 1J, and 2I respectively except that connectivity parameters were changed to $j_{ee}/C_m = 35$mV, $j_{ei}/C_m = -200$mV, $j_{ie}/C_m = 120$mV, $j_{ii}/C_m = -300$mV, $j_{ex}/C_m = 200$mV, and $j_{ix}C_m = 150$mV.

Appendix C: Generation of correlated spike trains for external inputs.

To generate correlated, Poisson spike trains for the external population in the correlated state we used the multiple interaction process (MIP) method [84] with jittering. Specifically, we generated one shared "mother" process with firing rate $r_m = r_x/c$. Then, for each of the N_x "daughter" processes, we randomly kept each spike in the mother process with probability c. As a result, each daughter process is a Poisson process with firing rate $cr_m = r_x$ and a proportion of c of the spikes are shared between any two daughter processes. To get rid of perfect synchrony between the daughter processes, we jittered each spike time in each daughter process by a normally distributed random variable with mean zero and standard deviation $\tau_c = 5$ms. Upon jittering, the daughter processes remain Poisson and the resulting CSD between daughter processes is given by Eq. (20). Spike count correlations between the daughter processes over large time windows are exactly c. The daughter processes were used as the spike trains, $S_j^x(t)$ in the external population in the correlated state. See [84] for a deeper analysis of this algorithm.

Appendix D: Parameter values and details of computer simulations

All connection probabilities were $p_{ab} = 0.1$ for $a = e, i$ and $b = e, i, x$. Synaptic timescales were $\tau_e = 8$ms, $\tau_i = 4$ms, and $\tau_x = 10$ms. The firing rate of the external population was $r_x = 10$Hz and, in the correlated state, the correlation was $c = 0.1$ with a jitter of $\tau_c = 5$ms. All covariances and correlations were computed by counting spikes or integrating continuous processes over a window of length 250ms. Membrane capacitance, C_m, is arbitrary so we report all current-based parameters in relation to C_m. For convenience, one can therefore set $C_m = 1$. Unscaled connection strengths were $j_{ee}/C_m = 25$mV, $j_{ei}/C_m = -150$mV, $j_{ie}/C_m = 112.5$mV, $j_{ii}/C_m = -250$mV, $j_{ex}/C_m = 180$mV, and $j_{ix}C_m = 135$mV. Note that j_{ab} was scaled by $\sqrt{N}$ to produce the true connection strengths, as indicated in Results. Neuron parameters are $g_L = C_m/15$, $E_L = -72$mV, $V_{th} = -50$mV, $V_{re} = -75$mV, $V_{lb} = -100$mV, $\Delta_T = 1$mV, and $V_T = -55$mV. Synaptic currents in figures are reported in units $C_m V/s$. Covariances between synaptic currents are computed between integrals of the currents (see Eq. (3) and surrounding discussion), so the covariances have units $C_m^2 mV^2$.

The differential equations defining the neuron model were solved using a forward Euler method with time step 0.1ms. Statistics in Figs. 1D, 2C,D, and 4E,F were computed from a simulation of duration 50s. Statistics in Figs. 1E-I, 2E–I, and 4C,D,G,H were computed by repeating a simulation of duration 50s over ten trials for each value of N, then averaging over trials. For each

trial, network connectivity was generated with a different random seed, so the statistics are averaged over time and over realizations of the "quenched" variability of network connectivity.

[1] Rubén Moreno-Bote, Alfonso Renart, and Néstor Parga, "Theory of input spike auto-and cross-correlations and their effect on the response of spiking neurons," Neural Comp **20**, 1651–1705 (2008).

[2] Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham, "Probabilistic brains: knowns and unknowns," Nat Neurosci **16**, 1170–1178 (2013).

[3] Maoz Shamir, "Emerging principles of population coding: in search for the neural code," Curr Op Neurobio **25**, 140–148 (2014).

[4] T Tetzlaff, S Rotter, E Stark, M Abeles, A Aertsen, and M Diesmann, "Dependence of neuronal correlations on filter characteristics and marginal spike train statistics," Neural Comput **20**, 2133–84 (2008).

[5] Tom Tetzlaff, Moritz Helias, Gaute T Einevoll, and Markus Diesmann, "Decorrelation of neural-network activity by inhibitory feedback," PLoS Comput Biol **8**, e1002596 (2012).

[6] B Doiron, A Litwin-Kumar, R Rosenbaum, G K Ocker, and K Josić, "The mechanics of state-dependent neural correlations," Nat Neurosci **19**, 383–393 (2016).

[7] Gabriel Koch Ocker, Yu Hu, Michael A Buice, Brent Doiron, Krešimir Josić, Robert Rosenbaum, and Eric Shea-Brown, "From the statistics of connectivity to the statistics of spike times in neuronal networks," Curr Op Neurobio **46**, 109–119 (2017).

[8] Benjamin Lindner, Brent Doiron, and André Longtin, "Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback," Phys Rev E **72**, 061919 (2005).

[9] S Ostojic, N Brunel, and V Hakim, "How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains," J Neurosci **29**, 10234–10253 (2009).

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

[11] Volker Pernice, Benjamin Staude, Stefano Cardanobile, and Stefan Rotter, "Recurrent interactions in spiking networks with arbitrary topology," Phys Rev E **85**, 031916 (2012).

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

[13] Yu Hu, James Trousdale, Krešimir Josić, and Eric Shea-Brown, "Motif statistics and spike correlations in neuronal networks," J Stat Mech **2013**, P03012 (2013).

[14] Dmytro Grytskyy, Tom Tetzlaff, Markus Diesmann, and Moritz Helias, "A unified view on weakly correlated recurrent networks." Front. Comput. Neurosci. **7**, 131 (2013).

[15] Benjamin Dummer, Stefan Wieland, and Benjamin Lindner, "Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity," Front Comput Neurosci **8**, 104 (2014).

[16] H Ko, S B Hofer, B Pichler, K A Buchanan, P J Sjöström, and T D Mrsic-Flogel, "Functional specificity of local synaptic connections in neocortical networks," Nature **473**, 87–91 (2011).

[17] Elodie Fino and Rafael Yuste, "Dense inhibitory connectivity in neocortex," Neuron **69**, 1188–1203 (2011).

[18] Robert B Levy and Alex D Reyes, "Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex," J Neurosci **32**, 5609–5619 (2012).

[19] Anne-Marie Oswald, Brent Doiron, John Rinzel, and Alex D Reyes, "Spatial profile and differential recruitment of GABAB modulate oscillatory activity in auditory cortex," J Neurosci **29**, 10321–10334 (2009).

[20] Yousheng Shu, Andrea Hasenstaub, and David A McCormick, "Turning on and off recurrent balanced cortical activity," Nature **423**, 288–293 (2003).

[21] Michael Wehr and Anthony M Zador, "Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex," Nature **426**, 442–446 (2003).

[22] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, and David A McCormick, "Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition," J Neurosci **26**, 4535–4545 (2006).

[23] Michael Okun and Ilan Lampl, "Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities," Nat Neurosci **11**, 535–537 (2008).

[24] Anja L Dorrn, Kexin Yuan, Alison J Barker, Christoph E Schreiner, and Robert C Froemke, "Developmental sensory experience balances cortical excitation and inhibition," Nature **465**, 932–936 (2010).

[25] Yujiao J Sun, Guangying K Wu, Bao-hua Liu, Pingyang Li, Mu Zhou, Zhongju Xiao, Huizhong W Tao, and Li I Zhang, "Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development," Nature **465**, 927–931 (2010).

[26] Mu Zhou, Feixue Liang, Xiaorui R Xiong, Lu Li, Haifu Li, Zhongju Xiao, Huizhong W Tao, and Li I Zhang, "Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex," Nat Neurosci **17**, 841–50 (2014).

[27] Peter C Petersen, Mikkel Vestergaard, Kristian H R Jensen, and Rune W Berg, "Premotor spinal network with balanced excitation and inhibition during motor patterns has high resilience to structural division," J Neurosci **34**, 2774–84 (2014).

[28] C van Vreeswijk and H Sompolinsky, "Chaotic balanced state in a model of cortical circuits," Neural Comput **10**, 1321–1371 (1998).

[29] C van Vreeswijk and H Sompolinsky, "Chaos in neuronal networks with balanced excitatory and inhibitory activity," Science **274**, 1724–1726 (1996).

[30] Jérémie Barral and Alex D Reyes, "Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics," Nat Neurosci **19**, 1690 (2016).

[31] A Renart, J de La Rocha, P Bartho, L Hollender, N Parga, A Reyes, and KD Harris, "The Asynchronous State in Cortical Circuits," Science **327**, 587–590 (2010).

[32] AS Ecker, P Berens, GA Keliris, M Bethge, NK Logothetis, and AS Tolias, "Decorrelated Neuronal Firing in Cortical Microcircuits," Science **327**, 584–587 (2010).

[33] Marlene R Cohen and Adam Kohn, "Measuring and interpreting neuronal correlations," Nat Neurosci **14**, 811–819 (2011).

[34] Matthew A Smith, Xiaoxuan Jia, Amin Zandvakili, and Adam Kohn, "Laminar dependence of neuronal correlations in visual cortex," J Neurophysiol **109**, 940–947 (2013).

[35] Alexander S Ecker, Philipp Berens, R James Cotton, Manivannan Subramaniyan, George H Denfield, Cathryn R Cadwell, Stelios M Smirnakis, Matthias Bethge, and Andreas S Tolias, "State dependence of noise correlations in macaque primary visual cortex," Neuron **82**, 235–248 (2014).

[36] Andrew YY Tan, Yuzhi Chen, Benjamin Scholl, Eyal Seidemann, and Nicholas J Priebe, "Sensory stimulation shifts visual cortex from synchronous to asynchronous states," Nature **509**, 226 (2014).

[37] Matthew J McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagha, Cathryn R Cadwell, Andreas S Tolias, Jessica A Cardin, and David A McCormick, "Waking state: rapid variations modulate neural and behavioral responses," Neuron **87**, 1143–1161 (2015).

[38] Gabriela Mochol, Ainhoa Hermoso-Mendizabal, Shuzo Sakata, Kenneth D Harris, and Jaime de la Rocha, "Stochastic transitions into silence cause noise correlations in cortical circuits," Proc Natl Acad Sci USA **112**, 201410509 (2015).

[39] Klaus Wimmer, Albert Compte, Alex Roxin, Diogo Peixoto, Alfonso Renart, and Jaime de la Rocha, "The dynamics of sensory integration in a hierarchical network explains choice probabilities in MT," Nat Commun **6**, 1–13 (2015).

[40] Robert Rosenbaum, Matthew A Smith, Adam Kohn, Jonathan E Rubin, and Brent Doiron, "The spatial structure of correlated neuronal variability," Nature Neurosci **20**, 107 (2017).

[41] Ran Darshan, Carl van Vreeswijk, and David Hansel, "How strong are correlations in strongly recurrent neuronal networks?" , 1–22 (2018), arXiv:1803.00643.

[42] Chengcheng Huang, Douglas A Ruff, Ryan Pyle, Robert Rosenbaum, Marlene R Cohen, and Brent Doiron, "Circuit models of low-dimensional shared variability in cortical networks," Neuron **101**, 337–348 (2018).

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

[44] Itamar D Landau, Robert Egger, Vincent J Dercksen, Marcel Oberlaender, and Haim Sompolinsky, "The impact of structural heterogeneity on excitation-inhibition balance in cortical networks," Neuron **92**, 1106–1121 (2016).

[45] R Pyle and R Rosenbaum, "Highly connected neurons spike less frequently in balanced networks," Phys Rev E **93**, 040302 (2016).

[46] Moritz Helias, Tom Tetzlaff, and Markus Diesmann, "The correlation structure of local neuronal networks intrinsically results from recurrent dynamics," PLoS Comput Biol **10**, e1003428 (2014).

[47] Iris Ginzburg and Haim Sompolinsky, "Correlation Functions in a Large Stochastic Neuronal Network," Adv. Neural Inf. Process. Syst. 6 , 471–476 (1994).

[48] David Dahmen, Sonja Grün, Markus Diesmann, and Moritz Helias, "Two types of criticality in the brain," , 1–15 (2017), arXiv:1711.10930.

[49] DJ Amit and N Brunel, "Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex," Cereb Cortex **7**, 237–252 (1997).

[50] Hannes Risken, "Fokker-planck equation," in *The Fokker-Planck Equation* (Springer, 1996) pp. 63–95.

[51] B Lindner, "Effects of noise in excitable systems," Phys. Rep. **392**, 321–424 (2004).

[52] M Richardson, "Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive," Phys Rev E **76**, 021919 (2007).

[53] Jaime de la Rocha, Brent Doiron, Eric Shea-Brown, Kresimir Josić, and Alex Reyes, "Correlation between neural spike trains increases with firing rate," Nature **448**, 802–6 (2007).

[54] E Shea-Brown, K Josić, J de la Rocha, and B Doiron, "Correlation and Synchrony Transfer in Integrate-and-Fire Neurons: Basic Properties and Consequences for Coding," Phys Rev Lett **100**, 108102 (2008).

[55] Robert Rosenbaum, "A Diffusion Approximation and Numerical Methods for Adaptive Neuron Models with Stochastic Inputs," Front. Comput. Neurosci. **10**, 1–20 (2016).

[56] Nicolas Fourcaud-trocme, David Hansel, Carl Van Vreeswijk, and Nicolas Brunel, "How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs," J Neurosci **23**, 11628–11640 (2003).

[57] Nicolas Fourcaud-Trocmé and Nicolas Brunel, "Dynamics of the instantaneous firing rate in response to changes in input statistics," J computational neuroscience **18**, 311–321 (2005).

[58] Christopher Ebsch and Robert Rosenbaum, "Imbalanced amplification: A mechanism of amplification and suppression from local imbalance of excitation and inhibition in cortical circuits," PLoS Comput Biol **14**, e1006048 (2018).

[59] Robert J Rosenbaum, James Trousdale, and Krešimir Josić, "Pooling and correlated neural activity," Front Comp Neurosci **4**, 1–14 (2010).

[60] Robert Rosenbaum, James Trousdale, and Krešimir Josić, "The effects of pooling on spike train correlations," Front Neurosci **5**, 1–10 (2011).

[61] G Hennequin, E. J. Agnes, and T. P. Vogels, "Inhibitory Plasticity: Balance, Control, and Codependence," Annu. Rev. Neurosci. **40**, 557–579 (2017).

[62] Alex Roxin, Nicolas Brunel, and David Hansel, "Role of delays in shaping spatiotemporal dynamics of neuronal activity in large networks," Phys Rev letters **94**, 238103 (2005).

[63] Birgit Kriener, Moritz Helias, Stefan Rotter, Markus Diesmann, and Gaute T Einevoll, "How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime," Front Comput Neurosci **7** (2014), 10.3389/fncom.2013.00187.

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

[65] Adam Keane and Pulin Gong, "Propagating waves can explain irregular neural dynamics," J Neurosci **35**, 1591–1605 (2015).

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

[67] Ekkehard Ullner, Antonio Politi, and Alessandro Torcini, "Ubiquity of collective irregular dynamics in balanced net-

works of spiking neurons," bioRxiv , 389452 (2018).
[68] Antonio Politi, Ekkehard Ullner, and Alessandro Torcini, "Collective irregular dynamics in balanced networks of leaky integrate-and-fire neurons," arXiv preprint arXiv:1808.03359 (2018).
[69] Katayun Cohen-Kashi Malina, Boaz Mohar, Akiva N Rappaport, and Ilan Lampl, "Local and thalamic origins of correlated ongoing and sensory-evoked cortical activities," Nat Comm **7**, 12740 (2016).
[70] Alexa Riehle, Sarah Wirtssohn, Sonja Grün, and Thomas Brochier, "Mapping the spatio-temporal structure of motor cortical lfp and spiking activities during reach-to-grasp movements," Front neural circuits **7**, 48 (2013).
[71] Adam Kohn, Amin Zandvakili, and Matthew A. Smith, "Correlations and brain states: from electrophysiology to functional imaging," Curr. Opin. Neurobiol. **19**, 434–438 (2009).
[72] James FA Poulet and Carl CH Petersen, "Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice," Nature **454**, 881–885 (2008).
[73] Marlene R Cohen and John HR Maunsell, "Attention improves performance primarily by reducing interneuronal correlations," Nat Neurosci **12**, 1594–1600 (2009).
[74] Adam C Snyder, Michael J Morais, Adam Kohn, and Matthew A Smith, "Correlations in v1 are reduced by stimulation outside the receptive field," J Neurosci **34**, 11222–11227 (2014).
[75] Gayathri Nattar Ranganathan and Helmut Koester, "Correlations decrease with propagation of spiking activity in the mouse barrel cortex," Front neural circuits **5**, 8 (2011).
[76] Bryan J Hansen, Mircea I Chelaru, and Valentin Dragoi, "Correlated variability in laminar cortical circuits," Neuron **76**, 590–602 (2012).
[77] Matthew A Smith and Marc A Sommer, "Spatial and temporal scales of neuronal correlation in visual area v4," J Neurosci **33**, 5422–5432 (2013).
[78] Anirvan S Nandy, Jonathan J Nassi, and John H Reynolds, "Laminar organization of attentional modulation in macaque visual area v4," Neuron **93**, 235–246 (2017).
[79] Robert Rosenbaum, Fabien Marpeau, Jianfu Ma, Aditya Barua, and Krešimir Josić, "Finite volume and asymptotic methods for stochastic neuron models with correlated inputs," J Math Biol **65**, 1–34 (2012).
[80] Taşkın Deniz and Stefan Rotter, "Solving the two-dimensional fokker-planck equation for strongly correlated neurons," Phys Rev E **95**, 012412 (2017).
[81] Tim Herfurth and Tatjana Tchumatchenko, "How linear response shaped models of neural circuits and the quest for alternatives," Curr Op Neurobio **46**, 234–240 (2017).
[82] Nicolas Brunel and Xiao-Jing Wang, "What determines the frequency of fast network oscillations with irregular neural discharges? i. synaptic dynamics and excitation-inhibition balance," J of Neurophysiol **90**, 415–430 (2003).
[83] Néstor Parga, "Towards a self-consistent description of irregular and asynchronous cortical activity," J Stat Mech: Theory and Exp **2013**, P03010 (2013).
[84] Alexandre Kuhn, Ad Aertsen, and Stefan Rotter, "Higher-order statistics of input ensembles and the response of simple model neurons." Neural Comput. **15**, 67–101 (2003).
[85] Benjamin Lindner, André Longtin, and Adi Bulsara, "Analytic expressions for rate and CV of a type I neuron driven by white gaussian noise." Neural Comput. **15**, 1760–87 (2003).
[86] Magnus J E Richardson, "Spike-train spectra and network response functions for non-linear integrate-and-fire neurons," Biol Cybern **99**, 381–92 (2008).
[87] Rafael Vilela and Benjamin Lindner, "Comparative study of different integrate-and-fire neurons: Spontaneous activity, dynamical response, and stimulus-induced correlation," Phys. Rev. E **80**, 031909 (2009).
[88] Magnus JE Richardson, "Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents," Phy Rev E **80**, 021928 (2009).
[89] Magnus JE Richardson and Wulfram Gerstner, "Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance," Neural Comp **17**, 923–947 (2005).
[90] Magnus J E Richardson and Rupert Swarbrick, "Firing-Rate Response of a Neuron Receiving Excitatory and Inhibitory Synaptic Shot Noise," Phys Rev Lett **105**, 178102 (2010).
[91] Rubén Moreno-Bote and Néstor Parga, "Auto-and cross-correlograms for the spike response of leaky integrate-and-fire neurons with slow synapses," Phys Rev letters **96**, 028101 (2006).
[92] Finn Müller-Hansen, Felix Droste, and Benjamin Lindner, "Statistics of a neuron model driven by asymmetric colored noise," Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. **91**, 1–14 (2015).
[93] Tilo Schwalger, Felix Droste, and Benjamin Lindner, "Statistical structure of neural spiking under non-poissonian or other non-white stimulation," J. Comput. Neurosci. **39**, 29–51 (2015).